Probability Theory Hybrid Notes

Course 20604, Bocconi University

Giancola Simone Maria[1†]

[†]Bocconi University, Milan

January 20, 2023

[1]simonegiancola09@gmail.com

Contents

List of Symbols

$\overline{\mathbb{N}} = \mathbb{N} \cup \infty$	naturals and infinity
$\mathbb{N}^* = \mathbb{N} \setminus \{0\}$	naturals without zero
$n, N, \mathbb{N}$	countable collection
$I, \mathbb{T}$	uncountable collection
$\subset$	subset proper and not
E	sample space
$\mathscr{A}$	algebra
$\mathcal{C}$	p-system
$\sigma(\cdot)$	generated sigma algebra
$d(\cdot, \cdot)$	distance
$\mathcal{B}(\cdot)$	Borel sigma algebra
$(E, \mathcal{E})$	measurable space
μ	measure
$\delta_x(\cdot)$	Dirac measure
$(E, \mathcal{E}, \mu)$	measure space
Leb	Lebesgue measure
$(\Omega, \mathcal{F}, \mathbb{P})$	probability space
X	random variable
$\mathcal{P}_X$	distribution r.v.
F_X	cumulative distribution function
$\mathbb{E}$	expectation
$\widehat{\mathcal{P}}.(\mathrm{t})$	laplace transform
$\Phi.(t)$	characteristic function
$\mathcal{G}(\cdot, \cdot)$	Erdős Renyi random graph
$\mathcal{P}^{\star}(\cdot)$	convolution
$\star$	convolution operation
$\xrightarrow{\mathcal{P}}$	convergence in probability
$\xrightarrow{\mathcal{L}_p}$	convergence in in Lp
$\xrightarrow{w}, \xrightarrow{d}$	weak convergence in distribution
$\overline{X}_{\mathcal{G}}$	best predictor conditional
X^{∞}	infinite countable sequence
$\mathcal{C}_n(\cdot)$	finite dimensional base cylinders
$\mathbf{\Gamma}$	stochastic matrix
$\mathcal{C}_T(\cdot)$	measurable base cylinders
$\mathscr{P}$	class of probability distributions
$\theta.$	shift operator
$\mathscr{B}$	semialgebra

Chapter 0

Read Me

Content The following document contains:

- redacted notes from a semester course I attended at Bocconi University
- interesting results I found along the way
- expansions I added to prove results not proved in class, and understand better the flow of lectures

For these reasons, it is *slightly* more than a course, while being mostly based on the course books [Çin11], [Ver18] for Chapter 7 and [BZ99] for Chapter 22. Occasionally exercises are taken from [Dur19]. Some results come from random internet sources.

Why hybrid I am used to taking notes by hand, but I love LaTeX, so this is a hybrid document with scans of the original notes for the proofs of the third part.

Structure Each Chapter is highly schematic. The reasons are many:

- I like to think schematically
- the handwritten part is the result of a rewriting of a lecture but needs to be studied, not presented
- it is easier to cite results *atomically*

About the last point, I have implemented a LaTeX-like numbering for handwriting, which is mirrored in the PC-typed part of each Chapter, to allow the reader to go back and forth and find proofs easily. At the same time, I can cite with references any result inequivocably.
In terms of *choice* of boxes, the first Chapters make a heavy use of Theorems instead of Propositions. After reading about the approach of the author of [Çin11], I chose instead to highlight as Theorems **only the most important results**.

Colors Writing by hand, I am used to thinking in colors. In the LaTeX-part Definitions, Theorems, Observations and Examples are distinguished by a color and a symbol in front, in the scanned part titles are highlighted with the same colors. For the sake of avoiding too many colors, I did not highlight Corollaries and Lemmas in LaTeX.

Appendix In the appendix there are only results which were not covered in class but end up being useful for an improved understanding of concepts. For the sake of time they will most likely have proofs in the handwritten version even at the end.

How this doc evolves If time permits, I will add more and more parts to study for the exam. Probably, the proofs of the appendix will only be on paper.

Part I

Measure Theoretic Probability

Chapter 1

Classes of Sets

1.1 Sigma algebras, Borel sets

♠ **Definition 1.1** (Sample Space E)**.** *All possible events of a random experiment.*

♠ **Definition 1.2** (Partition)**.** *A partition for a set E is a collection of disjoint subsets that covers the set. Namely $\{A_i\}_{i\in I}$ such that:*

1. $A_i \neq \emptyset \; \forall i$
2. $A_i \cap A_j = \emptyset \; \forall i, j$
3. $\bigcup_{i\in I} A_i = E$

♢ **Observation 1.3** (Why sets?)**.** *When doing probability we need structure.*

♠ **Definition 1.4** (Algebra $\mathscr{A}$)**.** *An algebra of a set E is a collection of subsets $\{A_i\}_{i\in I}$ $A_i \subset E$ such that:*

1. $E \in \mathscr{A}$
2. *closed under complements* $A \in \mathscr{A} \iff A^c \in \mathscr{A}$
3. *closed under finite intersection*

$$\{A_k\}_{k=1}^n \subset \mathscr{A},\; n \in \mathbb{N} \setminus \{\infty\} \implies \bigcap_{k=1}^n A_k \in \mathscr{A}$$

♣ **Theorem 1.5** (Finite union in algebra closedness)**.** *An algebra is closed under finite unions.*

$$\mathscr{A} \text{ algebra } \{A_i\}_{i=1}^n,\; n \in \mathbb{N} \implies \bigcup_{i=1}^n A_i \in \mathscr{A}$$

Proof. By De Morgan's laws

$$\bigcup_{i=1}^n A_i = \left(\bigcap_{i=1}^n A_i^c\right)^c$$

Where $\bigcap_{i=1}^n A_i^c \in \mathscr{A}$ by Def. 1.4#3 (finite intersection) and $\left(\bigcap_{i=1}^n A_i^c\right)^c \in \mathscr{A}$ by Def. 1.4#2 (complements). □

♠ **Definition 1.6** (Sigma Algebra)**.** *A sigma algebra is an algebra as in Definition 1.4 that allows for countable unions closedness, extending the finite union of requirement #3 to possibly infinite but countable. Namely:*

$$\mathcal{E} \;\sigma\text{-algebra} \quad (A_n) \subset \mathcal{E} \implies \bigcup_n A_n \in \mathcal{E}$$

Lemma 1.7 (Countable intersection in σ-algebra closedness)**.**

$$(A_n) \subset \mathcal{E} \implies \bigcap_n A_n \in \mathcal{E}$$

Proof. Similarly to Theorem 1.5, we have that countable intersections are allowed once we allow for countable unions. □

♠ **Definition 1.8** (p-system, also π-system $\mathcal{C}$)**.** *A p-system of a set E is a collection of subsets that is closed under intersections.*

$$\forall A, B \in \mathcal{C} \implies A \cap B \in \mathcal{C}$$

♡ **Example 1.9** (Basic algebras)**.** *The following examples are found throughout textbooks.*

- *(discrete σ-algebra) for E, set $\mathcal{E} = 2^E$, its power set. There are situations in which the construction reaches paradoxical results, but we will not see them. Most of the times we do not work with it.*
- *(trivial σ-algebra) for E, choose $\mathcal{E} = \{\emptyset, \mathcal{E}\}$*

♣ **Theorem 1.10** (σ-algebra intersection)**.** *Let $\{\mathcal{E}_i\}_{i\in I}$ be a collection of σ-algebras on the same space E, then their intersection $\mathcal{E} = \bigcap_{i\in I} \mathcal{E}_i$ is a sigma algebra as in Definition 1.6.*

Proof. We check the requirements of Definition 1.6.

1. $E \in \mathcal{E}_i \forall i \implies E \in \mathcal{E}$
2. $A^c \in \mathcal{E}_i \forall i \iff A \in \mathcal{E}_i \forall i \iff A \in \mathcal{E}$ which means $\iff A^c \in \mathcal{E}$
3. $(A_n) \subset \mathcal{E}_i \forall i \implies \bigcup_n A_n \in \mathcal{E}_i \forall i \implies \bigcup_n A_n \in \mathcal{E}$

So that $\mathcal{E}$ is a σ-algebra. □

♠ **Definition 1.11** (Generated σ-algebra $\sigma(\cdot)$)**.** *A generated σ-algebra for a collection of sets $\mathcal{C}$ is the smallest σ-algebra containing it. Namely:*

$$\sigma(\mathcal{C}) := \bigcap_{i\in I} \mathcal{E}_i \quad \textit{where } \forall i \in I \ \mathcal{C} \subset \mathcal{E}_i$$

We may occasionally denote it as $\mathscr{A}(\mathcal{C})$ as well. Note that by Theorem 1.10, $\sigma(\mathcal{C})$ is a well defined σ-algebra.

♣ **Theorem 1.12** (Properties of generated algebras)**.** *consider a collection of subsets $\mathcal{C}$ for E, then:*

1. *there always exists a generated σ-algebra*

$$\exists \sigma(\mathcal{C}) \ \forall \mathcal{C}$$

2. *if $\mathcal{C}$ is a σ-algebra, then its generated one is itself*

$$\mathcal{C} \ \sigma\textit{-algebra} \implies \sigma(\mathcal{C}) = \mathcal{C}$$

Proof. **(Claim #1)(basic)** let $E \neq \emptyset$. In any case, we have that $\mathcal{C} \subset 2^E$ the power set. So it always holds that $\mathscr{A}(\mathcal{C}) = \sigma(\mathcal{C}) \subset 2^E$ guaranteeing existance.
(Claim #1)(advanced) we did not consider the case in which $E = \emptyset$ and trivially $\mathcal{C} = \emptyset$ again. There, we make use of two basic facts:

- if we are within $E = \emptyset$, it tautologically holds that $\emptyset^c = \emptyset$
- the power set of the empty set is the empty set set, i.e. $\{\emptyset\}$. This holds since $S \subset \emptyset \implies S = \emptyset$ (i.e. empty set is subset of all sets), which means $S \in 2^\emptyset \implies S = \emptyset$

and eventually we have $\sigma(\mathcal{C}) = \sigma(\{\emptyset\}) = \{\emptyset, E^c\} = \{\emptyset, \emptyset\}$, which is the trivial σ-algebra.
(Claim #2) let $\mathcal{C}$ be a σ-algebra. Then for the set I where elements $\mathcal{E}_i$ are such that $\mathcal{C} \subset \mathcal{E}_i$:

$$\sigma(\mathcal{C}) = \bigcap_{i\in I} \mathcal{E}_i = \left(\bigcap_{i\in I'} \mathcal{E}_i \right) \cap \mathcal{C} = \mathcal{C}$$

□

♢ **Observation 1.13** (Why $\sigma(\mathcal{C})$?)**.** *We will show that a property that holds for $\mathcal{C}$ will hold for $\sigma(\mathcal{C})$ up to reasonable conditions.*

♡ **Example 1.14** (Easy $\sigma(\mathcal{C})$). *The trivial σ-algebra is $\sigma(\mathcal{C}) = \{\emptyset, E\}$ generated by $\mathcal{C} = \{E\}$. The easiest possible structure which is not trivial is generated by $\mathcal{C} = B$ for $B \subset E$ and is:*

$$\sigma(\mathcal{C}) = \Big\{\emptyset, B, B^c, E\Big\}$$

For a partition of E $\mathcal{C} = \{A, B, C\}$ we have instead that:

$$\sigma(\mathcal{C}) = \Big\{\emptyset, E, A, B, C, A \cup B, B \cup C, A \cup C\Big\}$$

Where intersections are ok as it is a partition. And complements are unions of other elements, namely:

- $C = (A \cup B)^c$
- $C^c = A \cup B$
- *and so on ...*

♠ **Definition 1.15** (Distance $d(\cdot,\cdot)$). *A distance is a function $d : E \times E \to \mathbb{R}_+$ satisfying $\forall x, y, z \in E$:*

1. *symmetry $d(x, y) = d(y, x)$*
2. *uniqueness of elements $d(x, y) = 0 \iff x = y$*
3. *triangle inequality $d(x, y) \leq d(x, z) + d(z, y)$*

♠ **Definition 1.16** (Open set). *A set A is open if for each element belonging to it there exists a neighborhood contained in A.*

$$A \text{ open} \iff \forall x \in A\, \exists \mathcal{N}_x \subset A$$

Throughout the course, we will focus on metric spaces, and the neighborhood will be induced by the distance d from Definition 1.15.

♠ **Definition 1.17** (Borel σ-algebra $\mathcal{B}(\cdot)$). *For a collection of open sets in E, a Borel sigma algebra is the sigma algebra generated by them.*

$$\mathcal{B}(E) = \sigma(\{A_i\}_{i\in I}) \quad \{A_i\}_{i\in I} \text{ open},\ A_i \subset E$$

Lemma 1.18 (Strength of algebras). *It holds:*

$$\text{Borel } \sigma\text{-algebra} \subset \sigma\text{-algebra} \subset \text{algebra}$$

Proof. Trivial □

Lemma 1.19 (Open sets of $\mathbb{R}$ representation). *We have:*

$$U \subset \mathbb{R} \text{ open} \implies U = \bigcup_{i\in I}(a_i, b_i) \quad I \text{ countable}$$

Proof. Let $U \subset \mathbb{R}$ be open, with $x \in U$, which is irrational or rational.
If x is rational define $I_x = \bigcup_{I \text{ open}, x\in I\subset U} I$, which is an open interval subset of U.
If x is irrational, since U is open $\exists \epsilon > 0 : (x - \epsilon, x + \epsilon) \subset U$ where $y \in \mathbb{Q}$ falls inside it. With the same construction, it holds $x \in I_y$.
For x arbitrary, $x \in I_q$ where $q \in U \cap \mathbb{Q}$ so that:

$$U \subset \bigcup_{q\in U\cap\mathbb{Q}} I_q \tag{1.1}$$

Yet $I_q \subset U \forall q \in U \cap \mathbb{Q}$ so:

$$U \supset \bigcup_{q\in U\cap\mathbb{Q}} I_q \tag{1.2}$$

Equality holds by double inclusion, and $U \cap \mathbb{Q}$ is countable. □

♡ **Example 1.20** (Borel σ of $\mathbb{R} = E$ generators). *Define a collection of subsets:*

$$\mathcal{C} = \left\{ U \subset \mathbb{R} \mid U \text{ open} \right\} \ : \ \sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$$

Which follows by Definition 1.17.
Now build a second collection of subsets:

$$\mathcal{I} := \left\{ (a,b) \ : \ -\infty < a < b < \infty \right\} \subset \mathcal{B}(\mathbb{R}) \quad \text{as } (a,b) \in \mathcal{B}(\mathbb{R}) \forall a,b \implies \sigma(\mathcal{I}) \subset \sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$$

On the other hand, for $U \in \sigma(\mathcal{C})$ notice that by Lemma 1.19:

$$U = \bigcup_{i \in I} (a_i, b_i) \in \sigma(\mathcal{I}) \quad I \text{ countable}$$

By I being a countable set. Hence, we have $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R}) \subset \sigma(\mathcal{I})$.
Eventually, $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{I})$.

♣ **Proposition 1.21** (Borel σ-algebra generators). *One can show that letting $\mathcal{I}$ be either of:*

$$\begin{gathered} \{(a,b) : -\infty < a < b < \infty\} \\ \{(a,b] : -\infty < a \leq b < \infty\} \\ \{[a,b] : -\infty < a < b < \infty\} \\ \{(-\infty, x] : -\infty < x < \infty\} \end{gathered}$$

Are all such that:

$$\sigma(\mathcal{I}) = \mathcal{B}(\mathbb{R})$$

♢ **Observation 1.22** (About $\mathcal{I}^*$). *The interval $\mathcal{I}^* = \{(-\infty, x] : -\infty < x < \infty\}$ of Proposition 1.21 is a p-system (Def. 1.8). For such p-system we see that as for the others $\sigma(\mathcal{I}^*) = \mathcal{B}(\mathbb{R})$. However, it comes with even nicer properties. These are explored further in Appendix A, and used in later claims.*

♢ **Observation 1.23** (Algebra and p-systems). *Notice that $\mathcal{C}$ is not necessarily an algebra (Def. 1.4). Indeed, it is not guaranteed that the first requirement ($E \in \mathcal{I}$) holds.*

♢ **Observation 1.24** (About unions of σ-algebras). *In Theorem 1.10, we show that intersections of σ-algebras are σ-algebras. For what concerns unions, this is **not** guaranteed.*

♡ **Example 1.25** (Two counterexamples to justify Observation 1.24). *Let $A_i = \{j\}_{j=1}^{i}$, and $\mathcal{E}_i = \sigma(\{A_i\})$ for $i \in \mathbb{N}^*$. Assume by contradiction that $\mathcal{E} = \bigcup_{i \in \mathbb{N}^*} \mathcal{E}_i$ is a σ-algebra. We have that:*

- $\forall i \ \{i\} \in \mathcal{E}_i, \mathcal{E}_i \subset \mathcal{E} \implies \{i\} \in \mathcal{E}$
- $\mathbb{N}^* = \bigcup_{i=1}^{\infty} \{i\} \in \mathcal{E}$ *by countable union* $\implies \exists i : \mathbb{N}^* \in \mathcal{E}_i$ *since it must come from some of the algebras used in the union for which* $\mathbb{N}^* \subset \mathcal{E}_i$.

However, we reached a contradiction since $\mathbb{N}^*$ is never inclued in any $\mathcal{E}_i$.
An even easier example is for $\Omega = \{1,2,3\}$ and the two σ-algebras:

$$\mathcal{E} = \{\emptyset, \{1\}, \{2,3\}, \Omega\} \quad \mathcal{F} = \{\emptyset, \{2\}, \{1,3\}, \Omega\}$$

Their intersection is not a σ-algebra since:

$$\{1,2\} \notin \mathcal{E} \cup \mathcal{F} = \{\emptyset, \{1\}, \{2\}, \{1,3\}, \{2,3\}, \Omega\}$$

Chapter Summary

Objects:

- algebras and σ-algebras as collection of subsets
  - with the sample space
  - closed under complements
  - closed under countable/uncountable union
- p-systems as collections of subsets closed under intersection
- generated σ-algebra, the smallest σ-algebra containing a set
- open sets induced by a distance
- Borel σ-algebra, a σ-algebra generated by open sets

Results:

- closedness under countable/uncountable union is equivalent to closedness under countable/uncountable intersection for algebras/σ-algebras
- an arbitrary intersection of σ-algebras is a σ-algebra
- the generated σ-algebra always exists and is equivalent to the generating set if the generating set is a σ-algebra
- a countable collection of open sets covers any open set of $\mathbb{R}$
- the Borel σ-algebra or $\mathbb{R}$ is generated by collections of intervals, in particular by the p-system:

$$\mathcal{I} = \{(-\infty, x] : x \in \mathbb{R}\}$$

- a p-system is not necessarily an algebra
- unions of σ-algebras are not necessarily σ-algebras

Chapter 2

Measures & Probablity Spaces

2.1 Probability & other measures

♠ **Definition 2.1** (Measurable Space $(E, \mathcal{E})$). *A measurable space is a tuple $E, \mathcal{E}$ where $\mathcal{E}$ is a σ-algebra of E (Def. 1.6). Whenever possible, we consider a Borel σ algebra 1.17, denoted as $\mathcal{B}(E)$.*

♠ **Definition 2.2** (Measure μ). *For a measurable space $(E, \mathcal{E})$ a measure is a map $\mu : \mathcal{E} \to \mathbb{R}_+$ such that:*

1. $\mu(\emptyset) = 0$
2. *countable additivity of disjoint collections*

$$\forall (A_n) \subset \mathcal{E}\ A_i \cap A_j = \emptyset \text{ countable disjoint} \quad \mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

*We say a measure is **finite** when $\mu(E) < \infty$. If a measure has total weight 1, then it is a **probability measure** (shorthand is p.m.).*

Next, we provide some examples which will be useful in the future.

♠ **Definition 2.3** (Point mass, dirac measure $\delta_x(\cdot)$). *Define:*

$$\forall x \in E \quad \forall A \in \mathcal{E} \quad \mu(A) := \delta_x(A) = \begin{cases} 1 & x \in A \\ 0 & else \end{cases}$$

♠ **Definition 2.4** (Counting measure). *Let $D \subset E$ be a countable subset of the sample space. Then let the measure be the sum of Dirac measures restricted to the countable subset:*

$$\forall A \in \mathcal{E}\ \nu(A) := |A \cap D| = \sum_{x \in D} \delta_x(A)$$

♠ **Definition 2.5** (Discrete measure). *Let $D \subset E$ be a countable subset. Set $m : D \to \mathbb{R}_+$ to be the mass at a point, independent from the set where it lies. A discrete measure is constructed as:*

$$\mu(A) = \sum_{x \in D} m(x)\delta_x(A) \quad \forall A \in \mathcal{E}$$

We will formalize this notion in Chapter 5.
Observe that for a counting measure $\nu(A)$ (def. 2.4) and a discrete measure μ, we have that:

$$\forall A \in \mathcal{E} \quad \nu(A) = 0 \implies \mu(A) = 0$$

This peculiarity is given a definition below.

♠ **Definition 2.6** (Absolutely continuous measure $\ll$)**.** *For two measures μ, ν we say that μ is absolutely continuous with respect to ν and write $\mu \ll \nu$ if:*

$$\forall A \in \mathcal{E} \quad \nu(A) = 0 \implies \mu(A) = 0$$

♠ **Definition 2.7** (Measure Space $(E, \mathcal{E}, \mu)$)**.** *We call a measure space a triplet $(E, \mathcal{E}, \mu)$ where the tuple $(E, \mathcal{E})$ is measurable (Def. 2.1) and μ is a measure on $\mathcal{E}$ (Def. 2.2).*

♠ **Definition 2.8** (Lebesgue measure *Leb*)**.** *The practical definition for a Lebesgue measure is that of a measure (Def. 2.2) defined on the measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ for which:*

$$Leb((a, b)) = Leb([a, b)) = Leb([a, b]) = Leb((a, b]) = b - a \quad \forall a \leq b$$

The more involved Definition is made via the concept of outer Lebesgue measure Leb^* (Def. 4.14), which for sets E that satisfy the Carathéodory criterion:

$$Leb^*(A) = Leb^*(A \cap E) + Leb^*(A \cap E^c) \quad \forall A \subset \mathbb{R}$$

builds a σ-algebra of those sets E which are Lebesgue measurable with measure $Leb^*(E) = Leb(E)$.
It is rather easy to check that with the former definition it is a well defined measure, and that the extensions follow as areas. For the $\mathbb{R}^2$ case we have for example:

$$\mu((a_1, b_1) \times (a_2, b_2)) = (b_1 - a_1)(b_2 - a_2)$$

♠ **Definition 2.9** (Probability space $(\Omega, \mathcal{F}, \mathbb{P})$)**.** *A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a measure space where the measure is a probability measure. Namely we have that $\mathcal{F}$ is a σ-algebra of Ω (Def. 1.6) and $\mathbb{P}$ is a measure with total mass 1 meaning:*

1. $\mathbb{P}[A] \in [0, 1] \; \forall A \in \mathcal{F}$
2. $\mathbb{P}(\Omega) = 1$
3. *countable additivity of disjoint collections:*

$$(A_n) \subset \mathcal{F} \text{ disjoint} \implies \mathbb{P}\left[\bigcup_n A_n\right] = \sum_n \mathbb{P}[A_n]$$

2.2 First properties

Most of the results we introduce extend to general measure, and are shown in Appendix A.

♣ **Theorem 2.10** (Monotonicity)**.** *For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we have:*

$$\forall A, B \in \mathcal{F} \; A \subset B \quad \mathbb{P}(A) \leq \mathbb{P}(B)$$

Proof. Notice that by $A \subset B \implies B = A \cup (A^c \cap B)$. Then by Definition 2.9#3 we have:

$$\mathbb{P}[B] = \mathbb{P}[A] + \mathbb{P}[A^c \cap B] \geq \mathbb{P}[A]$$

Where in the inequality we use the fact that the p.m. assigns positive measure to any Borel set (Def. 2.9#1). □

♣ **Theorem 2.11** (Inclusion Exclusion formula)**.** *Finite size unions (**NB** not necessarily disjoint) of collections of sets $(A_n) \subset \mathcal{F}$ where $n \in \mathbb{N}$ in a probability space satisfy the relation:*

$$\mathbb{P}\left[\bigcup_{i=1}^n A_i\right] = \sum_{i=1}^n \mathbb{P}[A_i] - \sum_{i<j} \mathbb{P}[A_i \cap A_j] + \sum_{i<j<k} \mathbb{P}[A_i \cap A_j \cap A_k] + \ldots + (-1)^{n-1}\mathbb{P}\left[\bigcap_{i=1}^n A_i\right]$$

Proof. We proceed by induction.
(base case) for $n = 2$ the statement is trivial as $\{A, B\}$ are such that:

$$A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B) = (A \cap B^c) \cup (B \cap A^c) \cup (A \cap B)$$

Where the events are disjoint. Additionally notice that:

$$A = (A \cap B^c) \cup (A \cap B) \quad B = (B \cap A^c) \cup (B \cap A)$$

again, disjoint. Using Definition 2.9#3 we have that when switching to the p.m.:

$$\begin{aligned}\mathbb{P}[A \cup B] &= \mathbb{P}[A \cap B^c] + \mathbb{P}[B \cap A^c] + \mathbb{P}[A \cap B] \\ &= \mathbb{P}[A] - \mathbb{P}[A \cap B] + \mathbb{P}[B] - \mathbb{P}[A \cap B] + \mathbb{P}[A \cap B] \\ &= \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]\end{aligned}$$

(induction assumption) assume the result is true $\forall n$.
(inductive step) wts the claim holds for $n+1$. Now:

$$\begin{aligned}\mathbb{P}\left[\bigcup_{i=1}^{n+1} A_i\right] &= \mathbb{P}\left[\bigcup_{i=1}^{n} A_i \cup A_{n+1}\right] \\ &= \mathbb{P}\Big[B \cup A_{n+1}\Big] && \text{where } B = \bigcup_{i=1}^{n} A_i \\ &= \mathbb{P}[B] + \mathbb{P}[A_{n+1}] - \mathbb{P}[B \cap A_{n+1}] && \text{base case}\end{aligned}$$

The last term is inspected below:

$$\begin{aligned}\mathbb{P}\left[A_{n+1} \cap \bigcup_{i=1}^{n} A_i\right] &= \mathbb{P}\left[\bigcup_{i=1}^{n} A_{n+1} \cap A_i\right] \\ &= \sum_{i=1}^{n} \mathbb{P}[A_{n+1} \cap A_i] - \sum_{i<j} \mathbb{P}[(A_{n+1} \cap A_i) \cap A_j] + \cdots + (-1)^{n-1}\mathbb{P}[(A_{n+1} \cap A_1) \cap A_2 \cdots \cap A_n]\end{aligned}$$

Using the fact that intersection distributes over unions, and applying the inductive hypothesis over n for the intersections. Plugging this back into the main calculation, we collect it with the result of:

$$\mathbb{P}[B] = \mathbb{P}\left[\bigcup_{i=1}^{n} A_i\right]$$

Which itself is split according to the inductive hypothesis. For a general index r we have:

$$(-1)^{r-1} \sum_{I \subset [1,\dots,n], |I|=s} \mathbb{P}\left[\bigcap_{i=1}^{r} A_i\right] + (-1)^{r-2} \sum_{J \subset [1,\dots,n-1], |J|=s-1} \mathbb{P}\left[\bigcap_{i=1}^{r-1} A_i \cap A_{n+1}\right]$$

Where the sums range over all possible choices of indices in the brackets where the collection has the size specified. This turns out being equal to:

$$(-1)^{r-1} \sum_{I \subset [1,\dots,n+1], |I|=r} \mathbb{P}\left[\bigcap_{i=1}^{r} A_i\right]$$

Which is exactly what we need to complete the claim. □

♢ **Observation 2.12** (Limits of sequences of sets). *Recall that:*

- (A_n) *non decreasing* $\implies \exists \lim_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} A_n = A$
- (A_n) *non increasing* $\implies \exists \lim_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} A_n = A$

♣ **Theorem 2.13** (Continuity of $\mathbb{P}$). *In a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *a non decreasing sequence of events has a limiting measure.*

$$(A_n) \subset \mathcal{F} \; (A_n) \nearrow A \implies \lim_{n\to\infty} \mathbb{P}[A_n] = \mathbb{P}[A]$$

Proof. We proceed using the additivity of $\mathbb{P}$ (Def. 2.9#3), aiming to build a sequence of disjoint sets. Imagine a sequence $(B_n)_{n\in\mathbb{N}}$ such that:

$$B_1 = A_1, \quad B_2 = A_2 \setminus B_1 = A_2 \cap B_1^c, \quad B_3 = A_3 \setminus B_2, \quad B_n = A_n \setminus B_{n-1}$$

We have that such sets are all pairwise disjoint $\mathcal{F}$-sets:

$$B_i \cap B_j = \emptyset \forall i \neq j \quad and \quad A_n, A_{n-1} \in \mathcal{F} \forall n \implies B_n \in \mathcal{F}$$

Additionally:

$$\begin{aligned} A_n &= \bigcup_{k=1}^{n} B_k \quad \forall n \\ A &= \lim_{n\to\infty} A_n = \bigcup_n A_n \\ &= \lim_{n\to\infty} \bigcup_{k=1}^{n} B_k = \bigcup_{k=1}^{\infty} B_k \end{aligned}$$

Which, passed to the probability measure:

$$\begin{aligned} \mathbb{P}[A] &= \mathbb{P}[\lim_{n\to\infty} A_n] \\ &= \mathbb{P}\left[\bigcup_n B_n\right] \\ &= \sum_{k=1}^{\infty} \mathbb{P}[B_k] && \text{disjoint sets 2.9\#3} \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}[B_i] \\ &= \lim_{n\to\infty} \mathbb{P}\left[\bigcup_{i=1}^{n} B_i\right] && \text{again disjoint sets} \\ &= \lim_{n\to\infty} \mathbb{P}[A_n] \end{aligned}$$

□

Lemma 2.14 (Continuity of $\mathbb{P}$ II). *A non increasing sequence of events has a limit as well.*

$$(A_n) \subset \mathcal{F} \ (A_n) \searrow A \implies \lim_{n\to\infty} \mathbb{P}[A_n] = \mathbb{P}[A]$$

Proof. Notice that $(A_n)_{n\in\mathbb{N}}$ non increasing $\iff$ $(A_n^c)_{n\in\mathbb{N}}$ non decreasing. So that we apply Theorem 2.13 to the complements to get:

$$\lim_{n\to\infty} \mathbb{P}[A_n^c] = \mathbb{P}[A^c] \implies \lim_{n\to\infty} \mathbb{P}[A_n] = \lim_{n\to\infty} 1 - \mathbb{P}[A_n^c] = 1 - \mathbb{P}[A] = \mathbb{P}[A^c]$$

□

♡ **Example 2.15** (Limits of sets). *Two instructive facts are:*

- $A_n = (0, 1 + \frac{1}{n}] \to (0, 1] = \bigcap_{n=1}^{\infty} (0, 1 + \frac{1}{n}]$
- $A_n = (0, 1 - \frac{1}{n}] \to (0, 1) = \bigcup_{n=1}^{\infty} (0, 1 - \frac{1}{n}]$

◇ **Observation 2.16** (Kolmogorov's approach in reality). *This construction is in high contrast with subjective probabilities, especially considering the continuity property of Theorem 2.13. It is a nice mathematical construction for improved tractability though.*

♣ **Theorem 2.17** (Boole's Inequality (subadditivity property)). *Consider a countable collection of subsets of E in the σ-algebra $(A_n) \subset \mathcal{F}$, then:*

$$\mathbb{P}\left[\bigcup_n A_n\right] \leq \sum_n \mathbb{P}[A_n]$$

Proof. Again, we aim to construct pairwise disjoint sets.
Let $C_1 = A_1$, $C_n = A_n \cap A_{n-1}^c \cap \cdots \cap A_1^c \quad \forall n \geq 2$. Clearly the sets are disjoint and decreasing and $C_n \subset A_n \forall n$. Also:

$$\bigcup_{k=1}^{n} C_k = \bigcup_{k=1}^{n} A_k$$

Then:

$$\begin{aligned}
\mathbb{P}\left[\bigcup_{k=1}^{n} C_k\right] &= \mathbb{P}\left[\bigcup_{k=1}^{n} A_k\right] \\
&= \sum_{k=1}^{n} \mathbb{P}[C_k] && \text{disjoint sets, Def. 2.9\#3} \\
&\leq \sum_{k=1}^{n} \mathbb{P}[A_k] && \text{monotonicity, Thm. 2.10}
\end{aligned}$$

So that applying the limit:

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\left[\bigcup_{k=1}^{n} A_k\right] &= \mathbb{P}\left[\bigcup_{n} A_n\right] && \text{Continuity on } (C_n)\text{, Thm. 2.13} \\
&\leq \lim_{n\to\infty} \sum_{k=1}^{n} \underbrace{\mathbb{P}[A_k]}_{\geq 0 \forall i} && \text{previous argument} \\
&= \sum_{k=1}^{\infty} \mathbb{P}[A_k] && \text{by sum of positive elements is defined at limit}
\end{aligned}$$

□

Lemma 2.18 (Boole's plus inclusion exclusion)**.** *Using Theorems 2.11, 2.17 we can say that for all finite collections of subsets* $\{A_i\}_{i=1}^{n} \subset \mathcal{F}$*, where finiteness comes from the inclusion exclusion requirement, we have:*

$$-\sum_{i<j} \mathbb{P}[A_i \cap A_j] + \sum_{i<j<k} \mathbb{P}[A_i \cap A_j \cap A_k] + \ldots + (-1)^{n+1}\mathbb{P}\left[\bigcap_{i=1}^{n} A_i\right] \leq 0$$

Chapter Summary

Objects:

- the measurable space $(E, \mathcal{E}, \mu)$ for a sample space, a σ-algebra on it and a valid measure
- dirac measure, counting measure, discrete measure, and Lebesgue measure
- notion of limit of monotonic sequences of sets

Results

- monotonicity

$$\forall A, B \in \mathcal{F}\ A \subset B \quad \mu(A) \leq \mu(B)$$

- inclusion-exclusion formula

$$\mathbb{P}\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{i=1}^{n} \mathbb{P}[A_i] - \sum_{i<j} \mathbb{P}[A_i \cap A_j] + \sum_{i<j<k} \mathbb{P}[A_i \cap A_j \cap A_k] + \ldots + (-1)^{n-1}\mathbb{P}\left[\bigcap_{i=1}^{n} A_i\right]$$

  trick: induction
- continuity of Probability

$$(A_n) \subset \mathcal{F}\ (A_n) \nearrow A \implies \lim_{n\to\infty} \mathbb{P}[A_n] = \mathbb{P}[A]$$

  trick: $B_n = B_{n-1} \setminus A_n$
- Boole's inequality

$$\mathbb{P}\left[\bigcup_n A_n\right] \leq \sum_n \mathbb{P}[A_n]$$

  trick: $C_n = A_n \cap A_{n-1}^c \cap \cdots \cap A_1^c$ sequence

Chapter 3

Random Variables

3.1 Realizations over the domain, distribution, the pdf, the cdf

Assumption 3.1 (Setting)**.** *The probability space is always* $(\Omega, \mathcal{F}, \mathbb{P})$*.*

♠ **Definition 3.2** (Random variable X, r.v.)**.** *For a measurable space* $(E, \mathcal{E})$ *(Def. 2.1), a random variable is a function from the probability space to* E*,* $X : \Omega \to E$ *such that it satisfies the* ***measurability condition****:*

$$X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\} \in \mathcal{F} \quad \forall A \in \mathcal{E} \tag{3.1}$$

♣ **Theorem 3.3** (p-systems sufficiency)**.** *Let* $\mathcal{C}$ *be a p-system (Def. 1.8) such that* $\sigma(\mathcal{C}) = \mathcal{E}$*. Then:*

$$X \text{ r.v.} \iff X^{-1}(A) \in \mathcal{F}\ \forall A \in \mathcal{C}$$

Namely, the ***measurability condition*** *of Equation 3.1 needs to be checked for the sets belonging to the p-system only, where clearly* $\mathcal{C} \subset \mathcal{E}$*.*

Proof. In general, Section A.1, in particular, Theorem A.4, Proposition A.8. □

Assumption 3.4 (Measurable spaces in the course)**.** *Throughout the course, we will restrict to cases in which* E *is one of the following:*

- *random variables* $E = \mathbb{R}$
- *random vectors* $E = \mathbb{R}^d$
- *discrete time stochastic processes* $E = \mathbb{R}^\infty$
- *continuous time stochastic processes* $E = \mathbb{R}^T$

♡ **Example 3.5** (Random variables check I, constant)**.** *Consider* $E = \mathbb{R}, \mathcal{E} = \mathcal{B}(\mathbb{R})$ *so that the p-system is:*

$$\mathcal{C} = \{(-\infty, x], x \in \mathbb{R}\} \quad \sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$$

We know the condition is:

$$X : \Omega \to \mathbb{R} \text{ r.v.} \iff X^{-1}((-\infty, x]) = \{\omega \in \Omega \ : \ X(\omega) \in (-\infty, x]\} \in \mathcal{F} \quad \forall x \in \mathbb{R}$$

Applying it to the easiest case where $X(\omega) = k \quad \forall \omega \in \Omega$ *where* $k \in \mathbb{R} = E$ *we have that:*

$$X^{-1}((-\infty, x]) = \{\omega \in \Omega \ : \ X(\omega) \le x\} = \begin{cases} \emptyset & x < k \\ \Omega & x \ge k \end{cases}$$

So that $\forall x \in \mathbb{R}$ *it holds* $X^{-1}((-\infty, x]) \in \{\emptyset, \Omega\}$ *where both* $\emptyset, \Omega \in \mathcal{F}$ *making it* ***always*** *a r.v. with respect to any* σ*-algebra of the space* Ω*.*

♡ **Example 3.6** (Random variables check II)**.** *In a similar fashion, we consider more elaborate examples:*

- *(indicator function) Let $A \subset \Omega$ and $X = \mathbb{1}_A$ so that:*

$$X(\omega) = \mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & otherwise \end{cases}$$

Using the condition derived before for $E = \mathbb{R}$ we check that:

$$X^{-1}((-\infty, x]) = \begin{cases} \emptyset & x \leq 0 \\ A^c & x \in [0,1) \\ \Omega & x \geq 1 \end{cases}$$

Where as before $\emptyset, \Omega \in \mathcal{F}$ for any σ-algebra, while we have that $A^c \in \mathcal{F} \iff A \in \mathcal{F}$ by the very Definition 1.4#2. An indicator r.v. is valid if and only if the event is in the σ-algebra of Ω considered.

- *(simple random variable) Suppose $\{A_1, \dots, A_n\}$ is a partition of Ω into $\mathcal{F}$-sets, namely:*

$$A_i \in \mathcal{F} \forall i, \quad A_i \cap A_j = \emptyset \forall i \neq j, \quad \bigcup_{i=1}^{n} A_i = \Omega$$

then for different realizations $x_1 \neq x_2 \neq \dots \neq x_n$ set:

$$X(\omega) = \sum_{k=1}^{n} x_k \mathbb{1}_{A_k}(\omega) \implies \forall \omega \in A_k \quad X(\omega) = x_k \iff X^{-1}(\{x_k\}) = \{\omega \in \Omega : X(\omega) = x_k\}$$

using the approach we outlined:

$$X^{-1}((-\infty, x]) = \bigcup_{k : x_k \leq x} A_k = \begin{cases} \emptyset & x < x_k \forall k \\ \Omega & x \geq x_k \forall k \\ a\ finite\ union & otherwise \end{cases}$$

Again, $\Omega, \emptyset \in \mathcal{F}$, and finite unions belong by definition to a σ-algebra provided that the elements $A_k \in \mathcal{F}$.

♠ **Definition 3.7** (Simple random variable)**.** *A random variable is said to be simple when it can be decomposed as a linear combination of indicators of a **finite** partition (Def. 1.2) of the Ω space.*

$$X(\omega) = \sum_{k=1}^{n} x_k \mathbb{1}_{A_k}(\omega) \quad where \quad \{A_k\}_{k=1}^{n}\ partition\ \Omega$$

♡ **Example 3.8** (Random variables check III)**.** *Let $\Omega = \{H, T\}^\infty$ where $\omega = (\omega_1, \dots)$ such that $\omega_n \in \{H, T\} \forall n \geq 1$. The σ-algebra is:*

$$\mathcal{F} = \sigma\Big(\{\omega \in \Omega : \omega_n = w\},\ n \in \mathbb{N},\ w \in \{H, T\}\Big)$$

For all $n \geq 1$ the map:

$$\omega \to X_n(\omega) = \begin{cases} 1 & \omega_n = H \\ 0 & \omega_n = T \end{cases}$$

is a random variable since

$$\forall A = (-\infty, x] \qquad X_n^{-1}(A) = \begin{cases} \emptyset & A \cap \{0,1\} = \emptyset \\ \Omega & A \cap \{0,1\} = \{0,1\} \\ \{\omega \in \Omega : \omega_n = H\} & A \cap \{0,1\} = \{1\} \\ \{\omega \in \Omega : \omega_n = T\} & A \cap \{0,1\} = \{0\} \end{cases}$$

Which are all in $\mathcal{F}$ since they generate it.

♣ **Theorem 3.9** (Operations of random variables)**.** *For two random variables X, Y on $E = \mathbb{R}$, we have that:*

1. *$X + Y$ is a r.v.*
2. *$min\{X, Y\}$ is a r.v.*

Proof. The strategy is proving the measurability condition of Equation 3.1.
(Claim #1) we have that:

$$\{X+Y\geq x\} = \{X > x-Y\} = \bigcup_{q\in\mathbb{Q}}\{X > q > x-Y\}$$
$$= \bigcup_{q\in\mathbb{Q}}\Big(\{X>q\}\cap\{Y>x-q\}\Big) = \bigcup_{q\in\mathbb{Q}}\Big(\underbrace{\{X\leq q\}^c}_{\in\mathcal{F}}\cap\underbrace{\{Y\leq x-q\}^c}_{\in\mathcal{F}}\Big)\in\mathcal{F}$$

where we are allowed to express the union over rationals since rationals are dense in $\mathbb{R}$ (Prop. 18.15).
We have a **countable** union of intersections of sets in $\mathcal{F}$, so by Theorem 1.10 and the definition of σ-algebra itself (Def. 1.6) we have that the element is in $\mathcal{F}$ making $X+Y$ a r.v.
(Claim #2) by direct computation:

$$\Big\{\omega\in\Omega : \min\{X(\omega),Y(\omega)\}\leq x\Big\} = \underbrace{\{\omega\in\Omega : X(\omega)\leq x\}}_{\in\mathcal{F}}\cap\underbrace{\{\omega\in\Omega : Y(\omega)\leq x\}}_{\in\mathcal{F}}\in\mathcal{F}\quad\forall x\in\mathbb{R}$$

where we have just checked the measurability condition for arbitrary $x\in\mathbb{R}$. ☐

♢ **Observation 3.10** (new random variables). *Using Theorem 3.9 we can say that the following are r.v.s:*

- $S_n(\omega) = \sum_{k=1}^{n} X_k(\omega) \quad \forall n$
- $\lim_{n\to\infty}\dfrac{S_n(\omega)}{n}$

♢ **Observation 3.11** (A taste of the Law of Large Numbers, LLN). *Following Obs. 3.10 can equivalently say that* $\limsup_n X_n$ *and* $\liminf_n X_n$ *are r.v.s. And that the set:*

$$\Lambda = \left\{\omega\in\Omega \mid \lim_n\frac{S_n}{n} = p\in[0,1]\right\}\in\mathcal{F}$$
$$= \left\{\omega\in\Omega \mid \limsup_n\frac{S_n}{n} = \liminf_n\frac{S_n}{n}\right\}$$

Has, under precise conditions, limiting measure $\mathbb{P}[\Lambda]=1$

♠ **Definition 3.12** (Distribution of a random variable, pushforward measure $\mathcal{P}_X = \mathbb{P}\circ X^{-1}$). *Let X be a r.v. (Def. 3.2) from the probability space $(\Omega,\mathcal{F},\mathbb{P})$ to the measurable space $(E,\mathcal{E})$. Then define the pushforward measure $\mathcal{P}_X=\mathbb{P}\circ X^{-1}$ as:*

$$\mathcal{P}_X(A) := \mathbb{P}\left[\{\omega\in\Omega\mid X(\omega)\in A\}\right] = \mathbb{P}[X\in A] = \mathbb{P}\left[X^{-1}(A)\right]\quad A\in\mathcal{E}$$

♢ **Observation 3.13** ($\mathbb{P}$ vs $\mathcal{P}$). *We recognize that:*

- $\mathbb{P}$ *is a p.m. on* $(\Omega,\mathcal{F})$
- $\mathcal{P}_X$ *is a p.m. on* $(E,\mathcal{E})$

♠ **Definition 3.14** (Equality in Distribution $\overset{d}{=}$). *We say two r.v.s X,Y are equal distribution when their distributions are equal for all the sets of the σ-algebra over which they are defined.*

$$X\overset{d}{=}Y \iff \mathcal{P}_X(A)=\mathcal{P}_Y(A)\;\forall A\in\mathcal{E}$$

♣ **Theorem 3.15** (Sufficient condition for Equality in distribution). *Given two random variables X,Y taking values on $(E,\mathcal{E})$, and a p-system $\mathcal{C}$ (Def. 1.8) such that: $\sigma(\mathcal{C})=\mathcal{E}$ we establish that:*

$$\mathcal{P}_X(A)=\mathcal{P}_Y(A)\;\forall A\in\mathcal{C}\implies X\overset{d}{=}Y$$

Where the opposite direction is trivial, and equality over the p-system is equivalent to equality over the whole σ-algebra.

Proof. Proposition A.29. ☐

♠ **Definition 3.16** (Cumulative distribution function F_X). *For $E = \mathbb{R}$ in the simplest case, and a p-system of intervals:*

$$\mathcal{C} = \{(-\infty, x] \mid x \in \mathbb{R}\} \quad \textit{where} \quad \sigma(\mathcal{C}) \overset{Prop.1.21}{=} \mathcal{B}(\mathbb{R})$$

Define the map $F_X : \overline{\mathbb{R}} \to \mathbb{R}_+$ as:

$$F_X(x) := \mathcal{P}_X((-\infty, x]) = \mathbb{P}[X \leq x]$$

By virtue of Theorem 3.15, we can check equality in distribution for the generating p-system $\mathcal{C}$ and assert:

$$X \overset{d}{=} Y \iff F_X(x) = F_Y(x)\ \forall x \in \mathbb{R}$$

♣ **Theorem 3.17** (Properties of F_X). *A cdf as in Def. 3.16 has the following properties:*

1. *monotone non decreasing*
$$F_X(x) \leq F_X(y) \quad \forall x \leq y$$
2. *has limits:*
$$\lim_{n \to -\infty} F_X(x) = 0 \qquad \lim_{n \to \infty} F_X(x) = 1$$
3. *right continuity*
$$\lim_{h \downarrow 0} F_X(x+h) = F_X(x)\ \forall x \in \mathbb{R}$$

Proof. **(Claim #1)** observe that $x < y$ correspond to sets of the p-system such that $A_x \subset A_y$. By monotonicity (Thm. 2.10) we have that:

$$F_X(x) = \mathbb{P}[A_x] \leq \mathbb{P}[A_y] = F_X(y)$$

(Claim #2) The following limits on Ω subsets hold:

$$\{\omega \in \Omega : X(\omega) \leq x\} \overset{x \to -\infty}{\to} \emptyset \quad \{\omega \in \Omega : X(\omega) \leq x\} \overset{x \to \infty}{\to} \Omega$$

With corresponding p.m. transferred to the distribution $F_X \to 0, F_X \to 1$.
(Claim #3) Consider a the set where $x_n = x + h : h \downarrow 0$:

$$A_n = \{\omega \in \Omega : X(\omega) \leq x_n\}, \quad x_n \downarrow z \implies \bigcap_n A_n = \{\omega \in \Omega : X(\omega) \leq x\}$$

Here it holds that:

- $x_n \downarrow x \implies X(\omega) \leq x, x_n \searrow x$
- $X(\omega) \leq x \implies X(\omega) \leq x_n\ \forall n$

So that:

$$\begin{aligned} F_X(x_n) &= \mathcal{P}_X((-\infty, x_n]) = \mathbb{P}[X \leq x_n] = \mathbb{P}[\omega \in A_n] \\ &\overset{n \to \infty}{\to} \mathbb{P}\left[\bigcap_n A_n\right] = \mathbb{P}[X \leq x] = \mathcal{P}_X((-\infty, x]) = F_X(x) \end{aligned}$$

by the continuity of $\mathbb{P}$ (Thm. 2.13). □

♢ **Observation 3.18** (Intuition for no left continuity). *Claim #2 2 of Theorem 3.17 does not necessarily mean that the distribution function is also left continuous. Consider* $\lim_{h \to 0^+} F_X(x-h)$ *and* $F_X(x)$. *The two are not necessarily equal, since we might have that for $\omega \in A_n$ where $x_n \to a^-$ the countable union never attains the value. Namely:*

$$a_n \notin \bigcup_n A_n = (-\infty, a_n)$$

♠ **Definition 3.19** (Probability of random variable realization). *By Theorem 3.17 and Observation 3.18 we write:*

$$\mathbb{P}[X = x] := F_X(x) - F_X(x^-) \quad F_X(x^-) := \lim_{h \uparrow 0} F_X(x-h)$$

Notice that if F_X is continuous ***(right and left)*** *at x then $\mathbb{P}[X = x] = 0$, while if it is not, such value is positive by Theorem 3.17#1.*

♢ **Observation 3.20** (Journey so far). *We have shown that we can construct a chain:*

$$X \text{ r.v.} \rightsquigarrow \mathcal{P}_X \text{ on } (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightsquigarrow F_X \text{ on } \mathbb{R}$$

How do we define a p.m. on $E = \mathbb{R}$, is it unique? Often r.v.s are artifacts and we do not have them at the start of the process; what happens more often is that we have a distribution. The next set of results sheds a light on the relations between these objects.

♣ **Theorem 3.21** ($F, \mathcal{P}$ identification). *Let $F : \overline{\mathbb{R}} \to \mathbb{R}_+$ satisfy the properties of Theorem 3.17, then:*

$$\implies \exists! \mathcal{P} \text{ on } (\mathbb{R}, \mathcal{B}(\mathbb{R})) \ : \ \mathcal{P}((-\infty, x]) = F(x) \ \forall x \in \mathbb{R}$$

Proof. We are working with a finite measure so we can safely assume that is is σ-finite. Then, Example C.18 is a direct proof using arguments from Section C.2. □

♣ **Theorem 3.22** (Identification F, X). *For F satisfying the properties of Theorem 3.17 we have:*

$$\implies \exists (\Omega, \mathcal{F}, \mathbb{P}) \ X : \Omega \to \mathbb{R} \quad : \quad F(x) = F_X(x) \ \forall x \in \mathbb{R}$$

where we take simply $\Omega = [0, 1]$, its borel σ-algebra $\mathcal{B}([0, 1])$ and the Lebesgue measure for the r.v.:

$$X(\omega) = \inf\{z \in \mathbb{R} : F(z) \geq \omega\}$$

Proof. With the claimed form, we notice that:

$$\{\omega \in [0, 1] : X(\omega) \leq x\} = \{\omega \in [0, 1] : \omega \leq F(x)\}$$

Implying that:

$$\begin{aligned} F_X(x) = \mathbb{P}[X \leq x] &= Leb\Big(\{\omega \in [0, 1] : \omega \leq F(x)\}\Big) \\ &= Leb([0, F(x)]) \\ &= F(x) - 0 = F(x) \quad \forall x \in \mathbb{R} \end{aligned}$$

□

♢ **Observation 3.23** (Takeaway). *We have shown:*

- $F \rightsquigarrow F_X \rightsquigarrow \mathcal{P}_X \rightsquigarrow X$ *by Theorems 3.21, 3.22*
- $X \rightsquigarrow \mathcal{P}_X \rightsquigarrow F_X \rightsquigarrow F$ *with the properties of Theorem 3.17*

Therefore, it is enough to use $F(x)$ to refer to a random variable.

Chapter Summary

Objects:

- random variable $X : \Omega \to E$ such that the measurability condition is verified:

$$X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\} \in \mathcal{F} \quad \forall A \in \mathcal{E}$$

- pushforward measure $\mathcal{P}_X = \mathbb{P} \circ X^{-1}$
- equality in distribution

$$X \overset{d}{=} Y \iff \mathcal{P}_X(A) = \mathcal{P}_Y(A) \quad \forall A \in \mathcal{E}$$

- cumulative distribution function $F_X(x) = \mathcal{P}_X((-\infty, x]) \quad \forall x \in \mathbb{R}$

Results:

- checking generator p-system is enough for $\overset{d}{=}$
- checking generator p-system is enough for measurability
- defining properties of F_X:
  - monotone non decreasing

$$F_X(x) \leq F_X(y) \quad \forall x \leq y$$

  - has limits:

$$\lim_{n\to-\infty} F_X(x) = 0 \qquad \lim_{n\to\infty} F_X(x) = 1$$

  - right continuity

$$\lim_{h\downarrow 0} F_X(x+h) = F_X(x) \; \forall x \in \mathbb{R}$$

- F valid identifies a unique $\mathcal{P}$
- F valid identifies a random variable and a measurable space

The conclusion is that we can work with F satisfying the defining properties.

Chapter 4

Expectation

We aim to describe the uncertain realization of a r.v. with a reasonable interpretation. By the results of Section 3, for a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ over which we define a r.v. $X : \Omega \to E$ on $(E, \mathcal{E})$ we can have two notions of integral:

$$\int_\Omega X(\omega)\mathbb{P}(d\omega) \quad or \quad \int_E x\mathcal{P}_X(dx)$$

Where the latter is equivalent by the very last results of the previous Chapter.

We refer to both as expectations of r.v.s and will use a constructive approach to build their most general form.

4.1 Building the expectation step by step

♠ **Definition 4.1** (Expectation of a simple random variable). *For a simple r.v. (Def. 3.7) we naturally define its integral as:*

$$\int_\Omega X(\omega)\mathbb{P}(d\omega) := \sum_{k=1}^{n} x_k \mathbb{P}[A_k]$$

*Where $\{A_k\}_{k=1}^n \subset \mathcal{F}$ is a **finite** partition of Ω.*

♠ **Definition 4.2** (Expectation of non negative random variable). *For a r.v. $X : \Omega \to \mathbb{R}_+$ we can always find a monotone sequence $(X_n)(\omega) \nearrow X(\omega)$ $\forall \omega \in \Omega$ where X_n is simple $\forall n$ as in Def. 3.7. This result is shown in Theorem A.18. Then:*

$$\forall n \geq 1 \quad \begin{cases} \int_\Omega X_n(\omega)\mathbb{P}(d\omega) = \sum_{k=1}^n x_k \mathbb{P}[A_k] & \textit{Def. 4.1} \\ \int_\Omega X_n(\omega)\mathbb{P}(d\omega) \leq \int_\Omega X_{n+1}(\omega)\mathbb{P}(d\omega) & X_n \leq X_{n+1} \end{cases}$$

Which identifies a non decreasing sequence of integrals that has a limit (see Obs 2.12)

$$\exists \lim_{n\to\infty} \left\{ \left(\int_\Omega X_n(\omega)\mathbb{P}(d\omega) \right)_n \right\}$$

For the expectation of a non negative random variable, we use such limit:

$$\int_\Omega X(\omega)\mathbb{P}(d\omega) := \lim_{n\to\infty} \left\{ \left(\int_\Omega X_n(\omega)\mathbb{P}(d\omega) \right)_n \right\}$$

♢ **Observation 4.3** (Getting to the general formula, properties of functions). *Observe that for a function, and thus any r.v. X, we have that $X = X^+ - X^-$ where:*

$$X^+ = \max\{0, X\} = X \wedge 0$$
$$X^- = -\min\{0, X\} = X \vee 0$$

*Where both are **positive** r.v.s.*

♠ **Definition 4.4** (Expectation of a random variable $\mathbb{E}$)**.** *Using the construction of Obs. 4.3 we can eventually state that:*

$$\mathbb{E}[X] := \int_\Omega X(\omega)\mathbb{P}(d\omega) = \int_\Omega X^+(\omega)\mathbb{P}(d\omega) - \int_\Omega X^-(\omega)\mathbb{P}(d\omega)$$

Which is ***well defined whenever the integrability condition holds****, namely that at least one of the two integrals is* ***not*** ∞*. If this condition was not verified, we would have an undefined form* $\infty - \infty$*. Notice that if one of the two diverges, and the other does not, the integral is well defined.*
Use as shorthand for simplicity:

$$\int_\Omega X d\mathbb{P} \quad \textit{and} \quad \int_\mathbb{R} x d\mathcal{P}$$

♠ **Definition 4.5** (Space of integrable r.v.s $\mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$)**.** *We say* $X : \Omega \to E$ *defined on the probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and taking values in* $(E, \mathcal{E})$ *is integrable if:*

$$\mathbb{E}\left[|X|\right] < \infty$$

We denote the collection of such r.v.s as $\mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$*. Notice that this is stronger than the existance of the integral, which in the class* $\mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ *is always finite.*

♢ **Observation 4.6** (Expected value and integrability)**.** *Notice that the condition is equivalent to requiring that the positive and negative parts do not diverge together as in Definition 4.4. Namely:*

$$\begin{cases} \nexists \mathbb{E}[X] & if \ \int_\Omega X^+ d\mathbb{P} = \infty \ \wedge \ \int_\Omega X^- d\mathbb{P} = \infty \\ \mathbb{E}[X] = \infty & if \ \int_\Omega X^+ d\mathbb{P} = \infty \ \wedge \ \int_\Omega X^- d\mathbb{P} < \infty \\ \mathbb{E}[X] = -\infty & if \ \int_\Omega X^+ d\mathbb{P} < \infty \ \wedge \ \int_\Omega X^- d\mathbb{P} = \infty \\ X \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}) & else \end{cases}$$

4.2 Properties of Expectation and integral

♣ **Theorem 4.7** (Properties of $\mathbb{E}$)**.** *Let* X *be a r.v. (Def. 3.2). The conditions can hold in the a.s. case or* $\forall \omega \in \Omega$*, depending on the formulation.*

1. *positivity for positive r.v.*

$$\mathbb{P}[X \geq 0] = 1 \implies \mathbb{E}[X] \geq 0$$

2. *monotonicity*

$$\begin{cases} \mathbb{P}[X \geq Y] = 1 \\ \exists \mathbb{E}[X], \mathbb{E}[Y] \\ Y \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}) \end{cases} \implies \mathbb{E}[X] \geq \mathbb{E}[Y]$$

3. *linearity*

$$\begin{cases} \exists \mathbb{E}[X] \\ Y \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}) \end{cases} \implies \mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y] \quad \forall a, b \in \mathbb{R}$$

4. *equality respect*

$$\begin{cases} \exists \mathbb{E}[X], \mathbb{E}[Y] \\ \mathbb{P}[X = Y] = 1 \end{cases} \implies \mathbb{E}[X] = \mathbb{E}[Y]$$

Proof. For all the results we have the a.s. hypothesis which allows us to reduce integrals over Ω to integrals over the set where the hypothesis holds. This is the basic trick to prove all the statements. We can use the approximation via simple functions and thus express the integral as we wish.
(Claim #1) We have $X = X^+$ so that $\int X d\mathbb{P} = \int X^+ d\mathbb{P} \geq 0$

(Claim #2) It holds:

$$\begin{aligned}\int X d\mathbb{P} - \int Y d\mathbb{P} &= \mathbb{E}[X] - \mathbb{E}[Y] \\ &= \lim_{n\to\infty} \int X_n^+ d\mathbb{P} - \lim_{n\to\infty} \int X_n^- d\mathbb{P} - \lim_{n\to\infty} \int Y_n^+ d\mathbb{P} + \lim_{n\to\infty} \int Y_n^- d\mathbb{P} && \text{Def. 4.4} \\ &= \lim_{n\to\infty} \int (X_n^+ - Y_n^+) d\mathbb{P} - \lim_{n\to\infty} (X_n^- - Y_n^-) d\mathbb{P} \\ &= \int X - Y d\mathbb{P} \\ &\geq 0\end{aligned}$$

Where the last passage follows from Claim #1 with the fact that $\mathbb{P}[X - Y \geq 0] = 1$. We could have
(Claim #3) with arguments similar to #2 we have:

$$\mathbb{E}[aX + bY] = a\mathbb{E}[x] + b\mathbb{E}[Y]$$

This time we require the integrability of at least one to make the integral not a undecidable form.
(Claim #4) by observing that $\mathbb{P}[X = Y] = 1 \iff \mathbb{P}[X - Y = 0] = 1$ use claims #1,#3 to get the claim:

$$\mathbb{E}[X] - \mathbb{E}[Y] = \mathbb{E}[X - Y] = \int X - Y d\mathbb{P} = \int 0 d\mathbb{P} = 0$$

□

♠ **Definition 4.8** (Expectations of functions of random variables). *Recall $\mathcal{P}_X = \mathbb{P} \circ X^{-1}$ induced on $(E, \mathcal{E})$. Consider $h : \mathbb{R} \to \mathbb{R}$ where $h(X) = Y$ and h is measurable (satisfies Eqn. 3.1). Y is a r.v. with expectation:*

$$\mathbb{E}[h(X)] = \int_\Omega h(X(\omega))\mathbb{P}(d\omega) := \int_\Omega h^+(X(\omega))\mathbb{P}(d\omega) + \int_\Omega h^-(X(\omega))\mathbb{P}(d\omega)$$

Which is meaningful if the expectation of the positive and the negative part are not both infinite. When h is also integrable, we may state it as $h \in \mathcal{L}_1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}_X)$ or $h(X) \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$. The extension to h mapping to other measurable spaces is trivial.

♣ **Theorem 4.9** (Law of the unconscious statistician). *Let $X : \Omega \to E$ be a r.v. Then for a measurable map h:*

$$\exists \mathbb{E}\left[h(X)\right] \implies \mathbb{E}[h(X)] = \int_\Omega h(X(\omega))\mathbb{P}(d\omega) = \int_E h(x)\mathcal{P}_X(dx)$$

Additionally, one can prove:

$$h \in \mathcal{L}_1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}_X) \iff h \circ X \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$$

Proof. In this case, the result holds by a more refined version of the change of variable method in integrals for a pushforward measure. □

♠ **Definition 4.10** (Notation for expectation of h). *We let:*

$$\mathcal{P}_X(h) := \mathbb{E}[h(X)] = \int_E h(x)\mathcal{P}_X(dx)$$

which is meaningful if h is measurable, namely:

$$h^{-1}(A) = \{x \in \mathbb{R} \mid h(x) \in A\} \in \mathcal{F} \quad \forall A \in \mathcal{B}(\mathbb{R})$$

establishing a chain of measurable pushforwards.

♣ **Theorem 4.11** (Equality in probability law characterization). *In the context of probability and Borel measures over X in a metric space:*

$$\forall h \in C_b^+(\mathbb{R}) \quad \mathcal{P}_X(h) = \mathcal{P}(h) \implies \mathcal{P}_X = \mathcal{P}$$

The result can be extended, but since we work with probability measures, it is sufficient.

Proof. By assumption for continuous and bounded functions h:

$$\mathcal{P}_X(h) = \int_{\mathbb{R}} h(x)\mathcal{P}_X(dx) = \int_{\mathbb{R}} h(x)\mathcal{P}(dx) = \mathcal{P}(h)$$

Notice also that $\forall a < b$ it holds that:

$$\mathbb{1}_{(a,b)}(x) = \lim_{n\to\infty} h_n(x) \qquad h_n(x) = n(x-a)\mathbb{1}_{(a,a+\frac{1}{n})}(x) + \mathbb{1}_{(a+\frac{1}{n},b-\frac{1}{n})} + n(b-x)\mathbb{1}_{(b-\frac{1}{n},b)}(x)$$

Where h_n is continuous and bounded $\forall n$, approaching the indicator of any interval. Our assumption on the integrals also holds there and we can state that $\mathcal{P}_X(h_n) = \mathcal{P}(h_n)$ for all n. More precisely, the former, by using the results of Theorem 4.7 takes form:

$$\mathcal{P}_X(h_n) = n\int_{(a,a+\frac{1}{n}]}(x-a)\mathcal{P}_X(dx) + \mathcal{P}_X\left[\left(a+\frac{1}{n}, b-\frac{1}{n}\right)\right] + n\int_{[b-\frac{1}{n},b)}(b-x)\mathcal{P}_X(dx)$$

In the above Equation, we have that:

- $\left(a, a+\frac{1}{n}\right] \searrow \emptyset$
- $\left[b-\frac{1}{n}\right) \nearrow \emptyset$
- h is continuous
- $\forall x \in \left(a, a+\frac{1}{n}\right] \quad n(x-a) < 1$
- $\forall x \in \left[b-\frac{1}{n}\right) \quad n(b-x) < 1$

These facts, which hold by construction and assumption, imply with continuity Thm. 2.13 that:

$$n\int_{(a,a+\frac{1}{n}]}(x-a)\mathcal{P}_X(dx) \leq \int_{(a,a+\frac{1}{n}]}\mathcal{P}_X(dx) = \mathcal{P}_X\left(\left(a, a+\frac{1}{n}\right]\right) \overset{n\to\infty}{\to} 0$$
$$n\int_{[b-\frac{1}{n},b)}(b-x)\mathcal{P}_X(dx) \leq n\int_{[b-\frac{1}{n},b)}\mathcal{P}_X(dx) = \mathcal{P}_X\left(\left[b-\frac{1}{n}, b\right)\right) \overset{n\to\infty}{\to} 0$$

Thus, in the limit the expression approaches $\mathcal{P}_X((a,b))$ and makes it such that :

$$\lim_{n\to\infty}\int h_n(x)\mathcal{P}_X(dx) = \lim_{n\to\infty}\mathcal{P}_X(h_n) = \mathcal{P}_X((a,b)) = \mathcal{P}((a,b)) = \lim_{n\to\infty}\mathcal{P}(h_n(x)) = \lim_{n\to\infty}\int h_n(x)\mathcal{P}(dx)$$

We have just proved that any non negative continuous and bounded function, expressed as the limit of our bar function, reduces to evaluating probabilities of simple intervals for both measures (red). That is, if two p.m.s are equal over h functions as hypothesized, then they are equal over all intervals (a,b) by the arbitrariness of a, b. Eventually, over the σ-algebra:

$$\sigma(\{(a,b) : -\infty < a < b < \infty\}) = \mathcal{B}(\mathbb{R})$$

the two measures agree, and by Theorem 3.15 they are equal. □

Lemma 4.12 (Full Characterization of equality in distribution)**.** *Let X, Y be positive r.v.s on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. The following are equivalent:*

1. $X \overset{d}{=} Y$
2. $\mathbb{E}\left[e^{-rX}\right] = \mathbb{E}\left[e^{-rY}\right] \quad \forall t \in \mathbb{R}_+$*, the Laplace transforms, which we will introduce in the next chapter*
3. $\mathbb{E}[f \circ X] = \mathbb{E}[f \circ Y] \quad \forall f \in C_b(\mathbb{R})$
4. $\mathbb{E}[f \circ X] = \mathbb{E}[f \circ Y] \quad \forall f$ *bounded Borel*
5. $\mathbb{E}[f \circ X] = \mathbb{E}[f \circ Y] \quad \forall f$ *positive Borel*

Proof. [Çin11], Example II.2.34. □

♡ **Example 4.13** (Examples of $\mathbb{E}$)**.** *We provide two basic cases:*

- *For $D \subset \Omega$ countable such that $D = \{\omega_1, \omega_2, \dots\}$ with law $\mathbb{P} = \sum_n \pi_n \delta_{\omega_n}$ such that $\sum_n \pi_n = 1$ and $\pi_n \geq 0 \forall n$ if we define*

$$X : \Omega \to \mathbb{R} \quad \omega_n \to x_n \in \mathbb{R} \implies \mathcal{P}_X = \sum_n \pi_n \delta_{x_n}$$

which for $h : \mathbb{R} \to \mathbb{R}$ has expectation:

$$\mathbb{E}[h(X)] = \mathcal{P}_X(h) = \sum_n h(X(\omega_n))\pi_n = \sum_n h(x_n)\pi_n$$

the expectation of a discrete r.v.

- *For $\Omega = [0,1]$ instead, with law $\mathbb{P} = Leb$ it holds:*

$$\mathbb{E}[h(X)] = \int h(X(\omega))d\mathbb{P}(d\omega) = \int_{[0,1]} h(x)Leb(dx) = \int_0^1 h(x)dx \quad \text{if Riemann integral exists}$$

♠ **Definition 4.14** (Outer Lebsegue measure)**.** ***This is better explained in appendix C***
Let $E \subset \mathbb{R}$ and define:

$$Leb^*(E) = \inf \sum_{k=1}^{\infty} l(I_k)$$

Where $\{I_k\}$ is countable and covers E, namely $E \subset \bigcup_1^\infty I_k$

Lemma 4.15 (Countable sets have $Leb^* = 0$)**.** *If E is countable then $Leb(E) = 0$*

Proof. Let $E = \{e_k\}_{k=1}^\infty$ countable and $\epsilon > 0$. Then:

$$\forall k \in \mathbb{N} \quad I_k = \left(e_k - \frac{\epsilon}{2^k}, e_k + \frac{\epsilon}{2^k}\right)$$

Such collection is countable and covers E, namely $\bigcup_k I_k \supset E$, by dyadics (rationals) being dense in $\mathbb{R}$ (again Prop. 18.15). Clearly:

$$0 \leq Leb(E) < \sum_k \ell(I_k) = \sum_k \frac{\epsilon}{2^k} = \epsilon \sum_k \frac{1}{2^k} = \epsilon \frac{\frac{1}{2}}{1 - \frac{1}{2}} = \epsilon \quad \forall \epsilon > 0$$

Implying that $Leb(E) = 0$. □

Corollary 4.16 (Countable sets and outer measure)**.** *establish that:*

1. *any outer measure giving zero to singletons satisfies Lemma 4.15*
2. *$Leb^*(\mathbb{Q}) = 0$*

Proof. Both claims follow directly by Lemma 4.15. □

◇ **Observation 4.17** (Lebesgue vs Riemann integral)**.** *Observe that Riemann integrable $\implies$ Lebesgue integrable and not the opposite. A counterexample is:*

$$h(x) = \begin{cases} 1 & x \in \mathbb{Q} \cap [0,1] \\ 0 & else \end{cases}$$

*Then, the discontinuities are **dense** in $[0,1]$ (see Prop. D.3), so the Riemann sum does not converge. Contrarily:*

$$\int_0^1 h(x)Leb(dx) = \int_{q \in \mathbb{Q}} h(x)Leb(dx) = 0$$

By Corollary 4.16.

4.3 Notions of convergence and results

♠ **Definition 4.18** (Almost sure $a.s.$ convergence $\xrightarrow{a.s.}$). *Let $(X_n)_{n\geq 1}$ be a countable series on $(\Omega, \mathcal{F}, \mathbb{P})$. The diverging limit $\lim\limits_{n\to\infty} X_n$ has many notions since $X_n(\omega)$ requires a qualification.*
*We say $X_n \xrightarrow{a.s.} X$ if $\mathbb{P}[\lim\limits_{n\to\infty} X_n = X] = 1$. However, notice that we **implicitly** assume that:*

$$\mathbb{P}[\lim_{n\to\infty} X_n = X] = \mathbb{P}\left[\{\omega \in \Omega \mid X_n(\omega) = X(\omega)\}\right]$$

is well defined, which holds if and only if:

$$\{\omega \in \Omega \mid X_n(\omega) = X(\omega)\} \in \mathcal{F}$$

To explain this, we start from the definition of limit:

$$\forall \epsilon > 0 \quad \exists n_0 = n_0(\epsilon) \mid \forall n_{>} n_0 \quad |X_n - X| < \epsilon$$

Then in set notation we say:

$$\bigcap_{\epsilon>0} \bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \underbrace{\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| < \epsilon\}}_{\in \mathcal{F} \text{ as } X_n, X \text{ r.v.}}$$

$$\bigcap_{\epsilon>0} \underbrace{\bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| < \epsilon\}}_{\in \mathcal{F} \text{ as } n_0, n \text{ countable}}$$

$$\underbrace{\bigcap_{\epsilon>0} \bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| < \epsilon\}}_{\notin \mathcal{F} \text{ as } \epsilon>0 \text{ uncountable}}$$

*Where, however, by the rationals being **dense** in $\mathbb{R}_+$ (again, see Prop. 18.15), we can replace the last intersection with one ranging over rationals $\frac{1}{k} \in \mathbb{Q}$, $k \in \mathbb{N}_+$, eventually concluding:*

$$\bigcap_{k\in\mathbb{N}_+} \bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \left\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| < \frac{1}{k}\right\} \in \mathcal{F}$$

♢ **Observation 4.19** (Interpreting $\xrightarrow{a.s.}$). *We can see $X_n \xrightarrow{a.s.} X$ as X_n being convergent to X for all ω but those that have null measure (negligible).*

♠ **Definition 4.20** (Increasingly and decreasingly $a.s.$ notation). *We denote:*

1. *$X_n(\omega) \leq X_{n+1}(\omega)$ $\forall\omega$ and $X_n \xrightarrow{a.s.} X$ with the symbol $X_n \nearrow X$ a.s.*
2. *$X_n(\omega) \geq X_{n+1}(\omega)$ $\forall\omega$ and $X_n \xrightarrow{a.s.} X$ with the symbol $X_n \searrow X$ a.s.*

♣ **Theorem 4.21** (Monotone Convergence Theorem). *It holds that in a monotone convergence regime we can **exchange the limit and the integral**:*

$$\begin{cases} (X_n)_{n\geq 1} \ : \ \mathbb{P}[X_n \geq 0] = 1 \ \forall n \\ X_n \nearrow X \ a.s. \end{cases} \implies \lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X] \quad a.e.$$

Proof. See Theorem A.46. □

Corollary 4.22 (Countable sum monotone convergence). *With the assumptions of Theorem 4.21 we can say:*

$$\mathbb{E}\left[\sum_n X_n\right] = \sum_n \mathbb{E}[X_n] \quad a.e.$$

Proof. Let $S_n = \sum_{k=1}^{n} X_k$. By $X_k \geq 0$ we have that $S_{n+1} \geq S_n \ \forall n$. By assumption, the sequence is strictly increasing to a limit.

$$S_n \nearrow S = \sum_{k=1}^{\infty} X_k$$

Moving to the expectation:

$$\begin{aligned}
\mathbb{E}[\lim_{n\to\infty} S_n] &= \mathbb{E}[S] \\
&= \lim_{n\to\infty} \mathbb{E}[S_n] && \text{Monotone conv Thm. 4.21} \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}[X_k] && \text{linearity Thm. 4.7\#3} \\
&= \sum_{k=1}^{\infty} \mathbb{E}[X_k]
\end{aligned}$$

Eventually:

$$\mathbb{E}[S] = \mathbb{E}\left[\sum_n X_n\right] = \sum_n \mathbb{E}[X_n]$$

□

♢ **Observation 4.23** (Commenting Corollary 4.22)**.** *Thanks to the Monotone convergence Theorem result, we can expand the linearity of integrals to infinite instances subject to convergence of the sum.*

♣ **Theorem 4.24** (Dominated Convergence Theorem)**.** *Let* $(X_n)_{n\geq 1}$ *be such that:*

$$\begin{cases} \mathbb{P}[X_n \leq Y] = 1 & \forall n \geq 1 \\ Y \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}) \\ X_n \overset{a.s.}{\to} X \end{cases}$$

Then:

1. $X_n, X \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$
2. $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$

Proof. Both claims follow by Theorem A.51. □

♢ **Observation 4.25** (On Theorem 4.21)**.** *A sketch of the proof is:*

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}[X_n] &= \mathbb{E}[\lim_{n\to\infty} X_n] \\
&= \int_\Omega \lim_{n\to\infty} X_n(\omega)\mathbb{P}[d\omega] \\
&= \int_{\{\omega\in\Omega: X_n(\omega)=X(\omega)\}} \lim_{n\to\infty} X_n(\omega)\mathbb{P}[d\omega] \\
&\quad + \underbrace{\int_{\{\omega\in\Omega: X_n(\omega)\neq X(\omega)\}} \lim_{n\to\infty} X_n(\omega)\mathbb{P}[d\omega]}_{=0} && \textit{by a.s. conv. hypothesis} \\
&= \int_{\{\omega\in\Omega: X_n(\omega)=X(\omega)\}} \lim_{n\to\infty} X_n(\omega)\mathbb{P}[d\omega] \\
&= \int_{\{\omega\in\Omega: X_n(\omega)=X(\omega)\}} X(\omega)\mathbb{P}[d\omega] \\
&= \int_\Omega X(\omega)\mathbb{P}[d\omega] \\
&= \mathbb{E}[X]
\end{aligned}$$

Corollary 4.26 (Bounded Convergence Theorem)**.** *Let* $(X_n)_{n\geq 1}$ *be such that:*

$$\begin{cases} \mathbb{P}[X_n \leq b] = 1 \quad \forall n \geq 1 \\ b < \infty \\ X_n \overset{a.s.}{\to} X \end{cases}$$

Then precisely:

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$$

Proof. Apply Theorem 4.24 with $Y = b \in \mathbb{R}$ a constant r.v. □

Chapter Summary

Objects:

- constructive definition of the integral:
  - simple functions
  - positive functions using monotone approximation
  - general functions using $X = X^+ - X^-$ decomposition
- space of integrable random variables $\mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$
- the $\mathcal{P}(h)$ notation
- almost sure convergence, as equality almost everywhere

$$X_n \overset{a.s.}{\to} X \iff \mathbb{P}\left[\lim_{n\to\infty} X_n = X\right] = 1$$

Results:

- properties of expectations:
  - positivity
  - monotonicity
  - linearity
  - equality a.s. implies equality of integrals

  tricks: be careful with the hypothesis, and use the decomposition and the a.s. conditions
- law of the unconscious statistician
- equality in distribution is implied by equality over C_b^+ functions for Borel probability measures in a metric space
- monotone convergence Theorem

  tricks: can exchange if the sequence converges, used to prove linearity of expectation for countable sums
- dominated convergence Theorem

  trick: need to assume integrability of the almost sure dominator, and convergence to something
- bounded convergence Theorem

Chapter 5

Density Functions

5.1 Radon-Nikodym Perspective

♢ **Observation 5.1** (Setting and goal). *We aim to provide an alternative formulation of $\mathcal{P}_X$. For simplicity recall the definition of absolute continuity (Def. 2.6 and of sigma finite measure (Def. A.26).*

♡ **Example 5.2** (Absolutely continuous measures). *To refresh the concept of $\ll$ (Def. 2.6) we provide an example. Let $\mu = \sum_{j\geq 0} \delta_j$ and $\nu = \delta_0 + \delta_1$. Then:*

$$\forall A \in \mathcal{B}(\mathbb{R}) \quad \mu(A) = 0 \implies \{0, 1, \dots\} \cap A = \emptyset \implies \nu(A) = 0 \implies \nu \ll \mu$$

While the opposite does not hold. A trivial counterexample is the set $A = \{2\} \in \mathcal{B}(\mathbb{R})$.

♠ **Definition 5.3** (Mutual Absolute Continuity $\sim$). *If $\mu \ll \nu$ and $\nu \ll \mu$ then we say $\mu \sim \nu$.*

♡ **Example 5.4** (Mutual absolute continuity w.r.t. *Leb*). *Consider two measures:*

$$\nu = Leb \quad \mu = e^{-|x|} Leb \;\; \forall x \in \mathbb{R}$$

Then:

$$\mu(A) = \int_A e^{-|x|} Leb(dx) = 0 \iff \nu(A) = \int_A Leb(dx) = 0 \quad \forall A \in \mathcal{B}(\mathbb{R})$$

which holds by $e^{-|x|} > 0 \forall x \in \mathbb{R}$. We then say $\mu \sim \nu$.

♡ **Example 5.5** (*Leb* is σ-finite but not finite). *The Lebesgue measure restricted to the positive half line $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ is such that $Leb(\mathbb{R}_+) = \infty$, so it is not finite.*
However, letting $E_n = [n-1, n) \quad \forall n \in \mathbb{N}^*$ we have:

$$\bigcup_{n\in\mathbb{N}^*} E_n = \mathbb{R}_+ \quad E_i \cap E_j = \emptyset \; \forall i \neq j \quad Leb(E_n) = 1 \; \forall n \in \mathbb{N}^*$$

So it is σ-finite in the sense of Definition A.26, since it has a countable partition of finite measures decomposition. It is rather trivial to show that this also holds for the Lebesgue measure on $\mathbb{R}$.

♡ **Example 5.6** (Counting measure is σ-finite). *Take $E = \mathbb{R}$ for simplicity. The counting measure (Def. 2.4) is constructed from countable set $D \subset \mathbb{R}$. We have by definition:*

$$\nu(A) = |A \cap D| \leq |D| \quad \forall A \in \mathcal{B}(\mathbb{R}) \tag{5.1}$$

The sets $|A \cap D|$ are countable. We now take for the space $\mathbb{R} = D \cup (\mathbb{R} \setminus D)$ two subpartitions:

- *D is partitioned into singletons since it is countable, call them (A_n)*
- *$\mathbb{R} \setminus D$, is partitioned into intervals of the form $[m, m+1)$, which cover it by countable union. We call these (B_m)*

It holds that

$$\forall m, n \quad \nu(A_n) = 1 \quad \nu(B_m) = 0 \tag{5.2}$$

All are finite and the union of the partitions is

$$\left(\bigcup_n A_n \right) \cup \left(\bigcup_m B_m \right) = \mathbb{R} \tag{5.3}$$

Thus the counting measure is σ-finite provided that D is countable, which is a requirement of the definition we gave (Def. 2.4).

♣ **Theorem 5.7** (Radon-Nikodym Theorem)**.** *Let μ, ν be two σ-finite measures (Def. A.26) on $(E, \mathcal{E})$, and ν be such that $\nu \ll \mu$ (Def. 2.6). Then:*

$$\exists p : E \to \mathbb{R}_+ \; measurable \; function \; s.t. \; \nu(A) = \int_A p(x)\mu(dx) \quad \forall A \in \mathcal{E} \tag{5.4}$$

Where p is ***almost everywhere unique*** *on the μ measure, meaning that that for another measurable function q satisfying Equation 5.4 it holds:*

$$\mu\left(\{x \in E \mid q(x) \neq p(x)\}\right) = 0$$

♢ **Observation 5.8** (About the assumptions)**.** *Notice that we require both measures to be σ-finite since it is not obvious that a measure absolutely continuous wrt a σ-finite measure is itself σ-finite. Counterexamples are found easily. Let $\mu = Leb$ and $\nu \ll \mu$ where:*

$$\nu(A) = \begin{cases} 0 & \mu(A) = 0 \\ \infty & else \end{cases}$$

The measure ν is not σ-finite since:

- *we can cover null-sets of Leb with a countable union, thus having finite (null) measures for each null set of Leb*
- *we cannot cover the whole $\mathbb{R}$ with finite measures since each other set has infinite measure by construction*

However, in the context of probability, since we work with ν a probability measure, we get σ-finiteness for free! In other words, this comment is just stated to specify that the Theorem is well defined, but in practical terms for the course we never have to check σ-finiteness of the measure we want to express.

♠ **Definition 5.9** (Radon Nikodym derivative of ν w.r.t μ, aka density function p)**.** *In the context of Theorem 5.7 we often denote the resulting density function as:*

$$p(x) = \frac{d\nu}{d\mu}(x) \tag{5.5}$$

This is not a formal definition but rather a symbol.

♢ **Observation 5.10** (Link with pmf)**.** *Recall from Example 5.5 that for $D \subset E$ a countable subset the counting measure (Def. 2.4):*

$$\mu(\cdot) = \sum_{x \in D} \delta_x(\cdot)$$

is σ-finite. Then for a r.v. X taking values inside D we have that:

$$\mathcal{P}_X \ll \mu \overset{Rad.Nyk.Thm}{\underset{Thm.5.7}{\Longrightarrow}} \exists p : E \to \mathbb{R}_+$$

satisfying:

$$\begin{aligned} \mathcal{P}_X(A) &= \int_A p(x)\mu(dx) \quad \forall A \in \mathcal{B}(E) \\ &= \mathbb{P}[X \in A] \\ &= \sum_{x \in A \cap D} p(x) \end{aligned}$$

Where:

$$p(x) = \frac{d\mathcal{P}_X}{d\mu}(x) = \mathbb{P}[X = x]$$

With the usual pmf properties. Namely:

$$\begin{cases} p(x) \geq 0 & \forall x \\ p(x) = 0 & \forall x \notin D \\ \sum p(x) = 1 \end{cases}$$

♠ **Definition 5.11** (Discrete random variable characterization)**.** *We say a random variable $X : \Omega \to E$ is **discrete** when it is absolutely continuous with respect to the counting measure (Def. 2.4).*

♢ **Observation 5.12** (Link with *pdf*)**.** *Similarly to what we did in Obs. 5.10, for a random variable X with probability distribution $\mathcal{P}_X \ll Leb$ we can say by the Radon Nikodym Theorem (Thm. 5.7) that there is an almost everywhere unique function p such that:*

$$\mathcal{P}_X(A) = \int_A p(x) Leb(dx) \quad \forall A \in \mathcal{B}(E)$$

With classic properties of a pdf:

$$\begin{cases} p(x) \geq 0 & \forall x \in \mathbb{R} \\ \int_{\mathbb{R}} p(x)dx = 1 \end{cases}$$

♠ **Definition 5.13** (Continuous random variable characterization)**.** *$X : \Omega \to E$ is said to be continuous when it is absolutely continuous with respect to the Lebesgue measure (Def. 2.8).*

♡ **Example 5.14** (Degenerate distribution)**.** *Let $x_0 \in E$ and $X : \Omega \to E$ with pushforward measure:*

$$\mathcal{P}_X(A) = \mathbb{P} \circ X^{-1}(A) = \delta_{x_0}(A) \quad \forall A \in \mathcal{E}$$

We say X is a degenerate at x_0 r.v.
For $E = \mathbb{R}$ it holds that:

$$\mathcal{P}_X((-\infty, x]) = F_X(x) = \mathbb{1}_{[x_0,\infty)}(x)$$

Notice that $\mathcal{P}_X \ll \mu = \sum_{j\geq 0} \delta_j$ the counting measure. By Theorem 5.7 we can derive a Radon Nikodym derivative:

$$p(x) = \frac{d\mathcal{P}_X}{d\mu}(x) = \begin{cases} 1 & x = x_0 \\ 0 & x \neq x_0 \end{cases} \qquad \mathcal{P}_X(A) = \int_A p(x)\mu(dx) \quad \forall A \in \mathcal{B}(\mathbb{R})$$

Where:

$$\mathcal{P}_X(\mathbb{R}) = \int_{\mathbb{R}} p(x)\mu(dx) = 1 \cdot \delta_{x_0} + 0 \cdot \delta_1 + 0 \cdot \delta_2 + \cdots$$

So that in general the expectation of a function $h : \mathbb{R} \to \mathbb{R}$ where $h(X) \in \mathcal{L}_1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}_X)$ is:

$$\begin{aligned} \mathbb{E}[h(X)] &= \int_{\mathbb{R}} h(X)\mathcal{P}_X(dx) && \mathcal{P}_X(dx) = p(x)\mu(dx) \\ &= \int_{\mathbb{R}} h(x)p(x)\mu(dx) && p(x) = \delta_{x_0} \\ &= \int_{\mathbb{R}} h(x)\delta_{x_0} \otimes \underbrace{(\delta_{x_0}(dx) + \delta_1(dx) + \delta_2(dx) + \cdots)}_{=\mu(dx)} \\ &= \int_{\mathbb{R}} h(x) \cdot 1 \cdot \delta_{x_0} \\ &= h(x_0) \end{aligned}$$

♡ **Example 5.15** (Poisson distribution)**.** *Let $\nu = \sum_{x\in\mathbb{N}^*} \delta_x$ be the counting measure. A r.v. X is saif to be Poisson with mean $\lambda > 0$ if:*

$$\mathcal{P}_X \ll \nu \quad \mathbb{P}[X = x] = p(x) = \frac{d\mathcal{P}_x}{d\nu}(x) = \frac{\lambda^x e^{-\lambda}}{x!}\mathbb{1}_{\mathbb{N}^*}(x)$$

So that:

$$\forall A \in \mathcal{B}(\mathbb{R}) \quad \mathcal{P}_X(A) = \int_A p(x)\nu(dx) = \sum_{x \in A\cap\mathbb{N}^*} \frac{\lambda^x e^{-\lambda}}{x!}$$
$$\forall h : \mathbb{R} \to \mathbb{R},\ h(X) \in \mathcal{L}_1 \quad \mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x)\mathcal{P}_X(dx) = \int_{\mathbb{R}} h(x)p(x)\nu(dx) = \sum_{x=0}^{\infty} h(x)\frac{\lambda^x e^{-\lambda}}{x!}$$

♡ **Example 5.16** (Binomial distribution)**.** *Let $n \in \mathbb{N}, \pi \in [0,1]$. Consider $\nu_n = \sum_{x=0}^{n} \delta_x$, the counting measure (Def. 2.4) on $D_n = \{0,\dots,n\}$. A r.v. X has binomial distribution, and we say $X \sim \mathcal{B}inom(n,\pi)$ if $\mathcal{P}_X \ll \nu_n$ and:*

$$\mathbb{P}[X = x] = p(x) = \frac{d\mathcal{P}_X}{d\nu_n}(x) = \binom{n}{x}\pi^x(1-\pi)^{n-x}\mathbb{1}_{D_n}(x)$$

where for Borel sets $A \in \mathcal{B}(\mathbb{R})$:

$$\mathcal{P}_X(A) = \int_A p(x)d\nu_n(x) = \sum_{x\in A\cap\{0,\dots,n\}} \binom{n}{x}\pi^x(1-\pi)^{n-x}$$

♡ **Example 5.17** (Gamma distribution)**.** *Let $\alpha, \beta > 0$. A r.v. X has distribution $\mathcal{G}amma(\alpha,\beta)$ where α is the shape and β is the scale if $\mathcal{P}_X \ll Leb$ and:*

$$p(x) = \frac{d\mathcal{P}_X}{dLeb}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}\mathbb{1}_{(0,\infty)}(x)$$

where at the denominator we are using the Gamma function

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx \quad \alpha > 0$$

of which the nicest property is $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$.
The probability law is recovered as:

$$\mathcal{P}_X(A) = \int_A p(x)dx = \int_{A\cap(0,\infty)} \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}dx \quad \forall A \in \mathcal{B}(\mathbb{R})$$

It is also recognized that:

- *for $\alpha = 1$ we recover the exponential distribution $\mathcal{E}xp(\beta)$*
- *for $\alpha = \frac{n}{2}, \beta = \frac{1}{2}$ we obtain the chi-square distribution with n degrees of freedom χ^2_n*

♡ **Example 5.18** (Normal/Gaussian distribution)**.** *Let $m \in \mathbb{R}$ and $\sigma > 0$. A r.v. X is normally distributed with mean m and variance σ^2 if $\mathcal{P}_X \ll Leb$ and:*

$$p(x) = \frac{d\mathcal{P}_X}{dLeb}(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}} \quad x \in \mathbb{R}$$

with law:

$$\mathcal{P}_X(A) = \int_A p(x)dx = \int_A \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}dx \quad \forall A \in \mathcal{B}(\mathbb{R}) \tag{5.6}$$

♠ **Definition 5.19** (Absolutely continuous function)**.** *In the context of functions, we establish absolute continuity for $F : \mathbb{R} \to \mathbb{R}$ when:*

$$\forall \epsilon > 0\ \exists \delta > 0 \quad s.t. \quad \begin{cases} \forall\{(a_i,b_i] : i = 1,\dots,k\} \text{ disjoint} \\ \sum_{i=1}^k (b_i - a_i) < \delta \end{cases} \implies \sum_{i=1}^k |F(b_i) - F(a_i)| < \epsilon$$

♣ **Proposition 5.20** (Distribution function absolute continuity)**.** *For $\mathcal{P}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mathcal{P} \ll Leb$ with Radon Nikodym derivative p it holds that:*

1. *F the distribution function is absolutely continuous (Def. 5.19) and:*

$$F(b) - F(a) = \int_a^b p(x)dx \quad \forall a < b$$

2. *F is differentiable almost everywhere with respect to the Lebesgue measure:*

$$Leb\left(\{x \in \mathbb{R} \mid \nexists F'(x)\}\right) = 0$$

3. *F has derivative almost everywhere p namely:*

$$F' = p\ a.e. \quad s.t. \quad Leb\left(\{x \in \mathbb{R} \mid F'(x) \neq p(x)\}\right) = 0$$

♠ **Definition 5.21** (Transformations of random variables)**.** *For a r.v. X on $(E, \mathcal{E})$ and a measurable function $f : E \to H$ (Def. A.7) mapping to another measurable space $(H, \mathcal{H})$ we can define the counterimage of f as:*

$$f^{-1}(B) = \{x \in E : f(x) \in B\} \in \mathcal{E} \quad \forall B \in \mathcal{H}$$

So the map $\omega \to Y(\omega) = f(X(\omega))$ is a r.v. taking values in $(H, \mathcal{H})$ and we can evaluate:

$$\mathcal{P}_Y(B) = \mathbb{P}[Y \in B] = \mathbb{P}[f(X) \in B] = \mathbb{P}[X \in f^{-1}(B)] = \mathcal{P}_X(f^{-1}(B)) \quad \forall B \in \mathcal{H}$$

Eventually concluding that $\mathcal{P}_Y = \mathcal{P}_X \circ f^{-1}$ where f^{-1} is the ***counterimage*** *and* ***not properly the inverse*** *of f.*
When we work in a Euclidean space $E = \mathbb{R}^d$ the formulae to make this change of variable are well known.

Chapter Summary

Objects:

- absolute continuity
- σ-finite measures
- Radon Nikodym derivative $p(x) = \dfrac{\mathcal{P}_X}{\mu}(x)$
- discrete and continuous random variables
- absolute continuity
- transformations of random variables

Results:

- the Lebsgue measure and the counting measure are σ-finite
- Radon Nikodym Theorem, for two σ-finite measures with $\nu \ll \mu$:

$$\exists p : E \to \mathbb{R}_+ \ measurable\ function\ s.t.\ \nu(A) = \int_A p(x)\mu(dx) \quad \forall A \in \mathcal{E}$$

  almost everywhere unique with respect to μ
- a probability measure absolutely continuous to the Lebesgue measure has an absolutely continuous cdf, differentiable almost everywhere, with a.e. derivative p, the Radon Nikodym derivative

Chapter 6

Random Vectors, Transforms

6.1 Multivariable approach

♠ **Definition 6.1** (Random vector). *We let $n = 2$, the definition naturally extends for $n > 2$. A random vector is a tuple of r.v.s (X, Y) on $(\Omega, \mathcal{F}, \mathbb{P})$ taking values on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that taken jointly they are measurable. Namely:*

$$(X,Y) : \Omega \to \mathbb{R}^2 \quad (X,Y)^{-1}(A) = \{\omega \in \Omega \mid (X(\omega), Y(\omega)) \in A\} \in \mathcal{F} \quad \forall A \in \mathcal{B}(\mathbb{R}^2)$$

♣ **Proposition 6.2** (Inherited properties of random vectors). *Drawing from Theorem 3.21 we have that:*

1. *Recalling that like in Prop. 1.21 it holds: $\mathcal{C} = \{(-\infty, x] \times (-\infty, y],\ x, y \in \mathbb{R}\}\ :\ \sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R}^2)$ we can establish validity by checking the p-system only:*

$$(X,Y)\ r.v. \iff (X,Y)^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \le x, Y(\omega) \le y\} \in \mathcal{F} \quad \forall A \in \mathcal{C}$$

2. *Unique identification of p.m. and cdf:*

$$A \to \underbrace{\mathcal{P}_{X,Y}(A)}_{on\ (\mathbb{R}, \mathcal{B}(\mathbb{R}^2))} = \underbrace{\mathbb{P}[\{\omega \in \Omega \mid (X(\omega), Y(\omega)) \in A\}]}_{on\ (\Omega, \mathcal{F})} \iff \mathcal{P}_{X,Y}((-\infty, x] \times (-\infty, y]) = \underbrace{\mathbb{P}[X \le x, Y \le y]}_{= F_{X,Y}(x,y)}$$

♢ **Observation 6.3** (About $\mathcal{P}_{X,Y}$). *Also the marginals are uniquely identified by the joint as:*

$$\begin{aligned} \mathcal{P}_X(A) &= \mathcal{P}_{X,Y}(A \times \mathbb{R}) && \forall A \in \mathcal{B}(\mathbb{R}) \\ \mathcal{P}_Y(A) &= \mathcal{P}_{X,Y}(\mathbb{R} \times A) && \forall A \in \mathcal{B}(\mathbb{R}) \end{aligned}$$

In general, the opposite is not true, as we need a well specified dependence structure to conclude how they mix.

♣ **Proposition 6.4** (Cumulative distribution function $F_{X,Y}$ properties inherited). *Similarly to Theorem 3.17, we have that given a tuple (X, Y) and their cdf $F_{X,Y} : \mathbb{R}^2 \to \mathbb{R}_+$:*

1. $\lim_{x \to -\infty} F_{X,Y} = 0$ *and* $\lim_{y \to -\infty} F_{X,Y} = 0$
2. $\lim_{x \to \infty} F_{X,Y} = F_Y$ *and* $\lim_{y \to \infty} F_{X,Y} = F_X$
3. $\lim_{x \to \infty} \lim_{y \to \infty} F_{X,Y} = \lim_{y \to \infty} \lim_{x \to \infty} F_{X,Y} = \lim_{x,y \to \infty} F_{X,Y} = 1$

Proof. All Claims follow by reasoning as in the one dimensional case, using the fact that the Borel sets either go to $\emptyset, X, Y, \Omega$. □

♣ **Proposition 6.5** (Cumulative distribution right continuity component wise). *Right continuity holds component wise:*

$$\forall (x,y) \in \mathbb{R}^2 \quad \lim_{h \downarrow 0} F_{X,Y}(x+h, y) = \lim_{h \downarrow 0} F_{X,Y}(x, y+h) = F_{X,Y}(x,y)$$

Proof. Again, the proof is the same as in the one dimensional case, Theorem 3.17#3. □

♣ **Proposition 6.6** (Cumulative distribution quasi monotonicity)**.** *Instead of being increasing, with more than one r.v. the cdf becomes quasi motononic, meaning that for (X, Y) we have:*

$$F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_2) + F_{X,Y}(x_1, y_1) \geq 0 \quad \forall x_2 \geq x_1,\ y_2 \geq y_1$$

Proof. The most intuitive proof is graphical. Consider Figure 6.1. For a rectangle in $\mathcal{B}(\mathbb{R}^2)$ the area must be non negative (since we are working with a measure), and we establish that:

$$F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_2) + F_{X,Y}(x_1, y_1) \geq 0 \quad \forall x_2 \geq x_1,\ y_2 \geq y_1$$

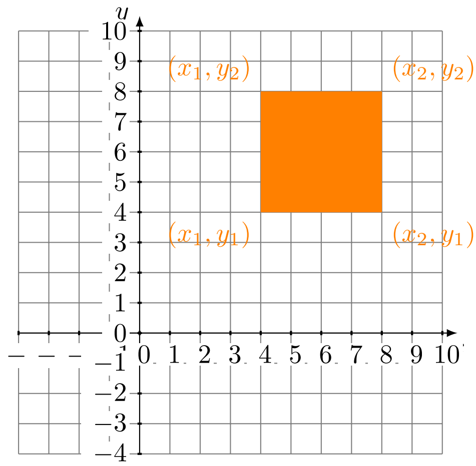

Figure 6.1: An example of measurable rectangle in $\mathcal{B}(\mathbb{R}^2)$

□

◇ **Observation 6.7** (Extending to bigger random vectors and the identification issue)**.** *We can easily extend these results to $n > 2$ sizes, either countably or uncountably ∞ as well. In Theorems 3.21,3.22 we proved that a general distribution function F satisfying certain properties identifies a unique probability law and a probability space with a random variable, can we do the same for general dimensions?*

♣ **Theorem 6.8** (Random vector identification via F)**.** *A function $F : \mathbb{R}^2 \to \mathbb{R}_+$ satisfying Propositions 6.4, 6.5 and 6.6 has a unique identified probability distribution $\mathcal{P}_X$ for a r.v. X on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$*

Proof. This is the same as Theorem 3.21 for random variables. □

♠ **Definition 6.9** (Independence $\perp\!\!\!\perp$)**.** *We say that two r.v.s (X, Y) respectively on $(E, \mathcal{E})$ and $(H, \mathcal{H})$ are independent if $\mathcal{P}_{X,Y} = \mathcal{P}_X \times \mathcal{P}_Y$. Namely:*

$$\mathcal{P}_{X,Y}[A \times B] = \mathbb{P}[X \in A, Y \in B] = \mathbb{P}[X \in A]\mathbb{P}[Y \in B] = \mathcal{P}_X(A)\mathcal{P}_Y(B) \quad \forall A \in \mathcal{E}, B \in \mathcal{H}$$

We write explicitly $X \perp\!\!\!\perp Y$. Similarly a collection of r.v.s $\{X_i, i \in I\}$ is an independency if:

$$\forall k \geq 2\ i_1 \neq \ldots \neq i_k \quad \mathcal{P}_{X_{i_1},\ldots,X_{i_k}} = \bigtimes_{j=1}^{k} \mathcal{P}_{X_{i_j}}$$

◇ **Observation 6.10** (About independence)**.** *We can easily check by Theorem 6.8 that:*

$$X \perp\!\!\!\perp Y \iff F_{X,Y}(x, y) = F_X(x)F_Y(y) \quad \forall x, y \in \mathbb{R}$$

In case of independence, we can also recover inequivocably the joint distribution by multiplying the marginals:

$$X \perp\!\!\!\perp Y \implies \exists! \mathcal{P}_{X,Y} \text{ on } (E \times H, \mathcal{E} \otimes \mathcal{H})$$

6.2 Transforms

♠ **Definition 6.11** (Laplace transform $\widehat{\mathcal{P}}_{\cdot}(t)$)**.** *For a positive r.v.* $X : \mathbb{P}[X \geq 0] = 1$ *we let:*

$$\widehat{\mathcal{P}}_X(t) = \mathbb{E}[e^{-tX}] = \int_{\mathbb{R}_+} e^{-tx}\mathcal{P}_X(dx) \quad \widehat{\mathcal{P}}_X(0) = 1 \quad \forall t \geq 0$$

For a non a.s. positive r.v., the definition is called two-sided Laplace transform and involves a modulus of t*.*

♣ **Theorem 6.12** (Laplace transform distribution characterization)**.** *It holds:*

$$\mathcal{P}_X = \mathcal{P}_Y \iff \widehat{\mathcal{P}}_X(t) = \widehat{\mathcal{P}}_Y(t) \quad \forall t \geq 0$$

♢ **Observation 6.13** (Getting back $\mathcal{P}_X$)**.** *There are inverse transformations from the laplace to the probability law.*

♣ **Proposition 6.14** (Independence sum factorizes in Laplace)**.** *For independent r.v.s satisfying the conditions of Def. 6.11 we have:*

$$X \perp\!\!\!\perp Y \implies \widehat{\mathcal{P}}_{X+Y}(t) = \widehat{\mathcal{P}}_X(t)\widehat{\mathcal{P}}_Y(t)$$

Proof. We proceed by just applying the Definition of Laplace transform.

$$\widehat{\mathcal{P}}_{X+Y}(t) = \mathbb{E}\left[e^{-t(X+Y)}\right] = \mathbb{E}\left[e^{-tX}e^{-tY}\right] = \mathbb{E}\left[e^{-tX}\right]\mathbb{E}\left[e^{-tY}\right] = \widehat{\mathcal{P}}_X(t)\widehat{\mathcal{P}}_Y(t)$$

Where we splitted the expectations since we have by hypothesis $X \perp\!\!\!\perp Y$. □

Lemma 6.15 (A useful identity in calculus)**.** *For a positive r.v.* $X : \mathbb{P}[X \geq 0] = 1$*, realized at* $x \geq 0$ *we have:*

$$e^{-tx} = \int_t^{\infty} xe^{-xw}dw$$

Proof. This is just a nice observation in calculus.

$$e^{-tx} = -e^{-xw}\Big|_{w=t}^{w=\infty} = x\left(\underbrace{-\frac{1}{x}e^{-xw}\Big|_{w=t}^{w=\infty}}_{\text{fund Thm. calculus}}\right) = x\int_t^{\infty} e^{-xw}dw = \int_t^{\infty} xe^{-xw}dw$$

where we used the fact that e^{-xw} is continuous. □

♣ **Theorem 6.16** (Connection of Laplace & moments)**.** *Consider an integrable positive random variable* X : $\mathbb{P}[X \geq 0] = 1,\ X \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ *the moments can be recovered as:*

1. *mean*
$$\frac{d}{dt}\widehat{\mathcal{P}}_X(t)\Big|_{t=0} = -\mathbb{E}[X]$$
2. *higher moments up to higher integrability* $\mathbb{E}[X^n] < \infty$
$$\frac{d^n}{dt^n}\widehat{\mathcal{P}}_X(t)\Big|_{t=0} = (-1)^n\mathbb{E}[X^n]$$

Proof. **(Claim #1)** We proceed step by step.

$$\begin{aligned}
\widehat{\mathcal{P}}_X(t) = \mathbb{E}[e^{-tX}] &= \int_{\mathbb{R}_+} e^{-tx}\mathcal{P}_X(dx) \\
&= \int_{\mathbb{R}_+}\int_t^{\infty} xe^{-xw}dw\mathcal{P}_X(dx) && \text{Lem. 6.15} \\
&= \int_t^{\infty}\int_{\mathbb{R}_+} xe^{-xw}\mathcal{P}_X(dx)dw && \text{Fubini Thm. B.30} \\
&= \int_t^{\infty} \mathbb{E}[Xe^{-wX}]dw
\end{aligned}$$

Now by $X \in \mathcal{L}_1$ we can say that $\exists d\widehat{\mathcal{P}}_X$ for *a.e.* $t \in (0, \infty)$ wrt *Leb.*
We would like to use again the fundamental Theorem of Calculus. Since we will show continuity in Lemma 6.26#1, it is possible to make the classic trick:

$$\int_t^{\infty} \mathbb{E}[Xe^{-wX}]dw = -\int_{\infty}^{t} \mathbb{E}[Xe^{-wX}]dw = -\left(\int_{\infty}^{c} \mathbb{E}[Xe^{-wX}]dw + \int_c^t \mathbb{E}[Xe^{-wX}]dw\right) \quad c \in (t, \infty)$$

where the first term is constant when differentiating by t. We are now in the position to apply the fundamental Theorem of calculus, slightly adatped to Measure theory[1]:

$$\frac{d\widehat{\mathcal{P}}_X(t)}{dt} = -\mathbb{E}[Xe^{-tX}]$$

evaluated at $t = 0$ is by continuity:

$$\begin{aligned}
\frac{d}{dt}\widehat{\mathcal{P}}_X(t)\bigg|_{t=0} &= \lim_{t\downarrow 0} \frac{d}{dt}\widehat{\mathcal{P}}_X(t) \\
&= \lim_{t\downarrow 0} -\mathbb{E}\left[Xe^{-tX}\right] && Xe^{-tX} \leq X \iff e^{-tX} \leq 1,\ X \geq 0 \quad \forall t \geq 0 \\
&= -\mathbb{E}\left[\lim_{t\downarrow 0} Xe^{-tX}\right] && \text{dom. conv. Thm. 4.24 as } X \in \mathcal{L}_1 \\
&= -\mathbb{E}[X]
\end{aligned}$$

(Claim #2) similarly obtained by the same arguments recursively. □

♠ **Definition 6.17** (Characteristic function $\Phi_\cdot(t)$)**.** *For a real r.v. X on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and any argument $t \in \mathbb{R}$ define:*

$$\Phi_X(t) = \mathbb{E}\left[e^{itX}\right] = \int_{\mathbb{R}} e^{itx}\mathcal{P}_X(dx) = \int_{\mathbb{R}} \cos(tx)\mathcal{P}_X(dx) + i\int_{\mathbb{R}} \sin(tx)\mathcal{P}_X(dx)$$

Where in the last passage we use Euler's identity $e^{itx} = \cos(tx) + i\sin(tx)$

♣ **Theorem 6.18** (Characteristic function distribution characterization)**.**

$$\mathcal{P}_X = \mathcal{P}_Y \iff \Phi_X(t) = \Phi_Y(t)\ \forall t \in \mathbb{R}$$

♣ **Theorem 6.19** (Existance of $\Phi_X(t)$)**.** *The characteristic function always exists $X \implies \exists \Phi_X(t)$*

Proof. Notice that $|\Phi_X(t)|$ is such that:

$$\begin{aligned}
|\Phi_X(t)| &= \left|\mathbb{E}[e^{itX}]\right| \\
&= \left|\int_{\mathbb{R}} e^{itX}\mathcal{P}_X(dx)\right| \\
&\leq \int_{\mathbb{R}} |e^{itX}|\mathcal{P}_X(dx) \\
&=\leq \int_{\mathbb{R}} \mathcal{P}_X(dx) && |e^{itx}| = |\cos(tx) + i\sin(tx)| = \sqrt{\cos^2(tx) + \sin^2(tx)} = 1 \\
&= 1
\end{aligned}$$

Which means that the characteristic function is finite and always exists by being a finite expectation in terms of the probability law. □

♣ **Proposition 6.20** ($\Phi_X(t)$ real condition)**.** *A symmetric random variable has real Characteristic function:*

$$X\ :\ \mathcal{P}_X(A) = \mathcal{P}_X(-A)\ \forall A \in \mathcal{B}(\mathbb{R}) \implies \Phi_X(t) \in \mathbb{R}\ \forall t \in \mathbb{R}$$

[1] we did not go much deep into this

Proof. We aim to evaluate the characteristic function of a symmetric r.v.:

$$\begin{aligned}\Phi_X(t) &= \int_{\mathbb{R}} e^{itx}\mathcal{P}_X(dx) \\ &= \int_{\mathbb{R}} \cos(tx)\mathcal{P}_X(dx) + i\int_{\mathbb{R}} \sin(tx)\mathcal{P}_X(dx) && \text{Euler's formula and linearity}\end{aligned}$$

Where the complex part is:

$$\begin{aligned}\int_{\mathbb{R}} \sin(tx)\mathcal{P}_X(dx) &= \int_{\mathbb{R}_-} \sin(tx)\mathcal{P}_X(dx) + \int_{\mathbb{R}_+} \sin(tx)\mathcal{P}_X(dx) \\ &= \int_{\infty}^{0} \sin(-tx)\mathcal{P}_{-X}(-dx) + \int_0^{\infty} \sin(tx)\mathcal{P}_X(dx) && \text{Ch. integration index} \\ &= \int_{\infty}^{0} -\sin(tx)\mathcal{P}_{-X}(-dx) + \int_0^{\infty} \sin(tx)\mathcal{P}_X(dx) && \sin(-tx) = -\sin(tx) \\ &= \int_{\infty}^{0} -\sin(tx)(-\mathcal{P}_{-X}(dx)) + \int_0^{\infty} \sin(tx)\mathcal{P}_X(dx) && \mathcal{P}_{-X}(-dx) = -\mathcal{P}_{-X}(dx) \\ &= -\int_0^{\infty} \sin(tx)(\mathcal{P}_{-X}(dx)) + \int_0^{\infty} \sin(tx)\mathcal{P}_X(dx) && \text{Ch. integration index} \\ &= -\int_0^{\infty} \sin(tx)\mathcal{P}_X(dx) + \int_0^{\infty} \sin(tx)\mathcal{P}_X(dx) && \text{Symmetry } -X \overset{d}{=} X \\ &= 0\end{aligned}$$

So that $\Phi_X(t) \in \mathbb{R} \quad \forall t \in \mathbb{R}$. □

♣ **Proposition 6.21** (Properties of $\Phi_X(t)$ similar to $\widehat{\mathcal{P}}_X(t)$)**.** *We have that:*

1. $X \perp\!\!\!\perp Y \implies \Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t)$
2. $\frac{d^n}{dt^n}\Phi_X(t) = i^n\mathbb{E}[X^n]\ \forall n \geq 1$

Proof. Both Claims are as those of Theorem 6.16 and Proposition 6.14. □

♠ **Definition 6.22** (Probability Generating Function, PGF)**.** *For a discrete r.v. (Def. 5.11)* $X : \Omega \to \mathbb{N}\cup\{\infty\} = \overline{\mathbb{N}}$ *we say the PGF is:*

$$\mathbb{E}\left[z^X\right] = \sum_{n=0}^{\infty} z^n\mathbb{P}[X = n]$$

Which uniquely determines $\mathcal{P}_X$ *since it is the power series expansion of the coefficients of* $\mathbb{P}[X = n]$

♡ **Example 6.23** ($\mathcal{G}$amma distribution)**.** *Let* $X \sim \mathcal{G}amma(\alpha, \beta) \quad \alpha > 0, \beta > 0$.
(△ ***moments closed form****) wts that* $\exists\mathbb{E}[X^p] \quad \forall p > -\alpha$.

$$\begin{aligned}\mathbb{E}[X^p] &= \int_{\mathbb{R}_+} x^p \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}e^{-\beta x}dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}\int_{\mathbb{R}_+} x^p x^{\alpha-1}e^{-\beta x}dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha+p)}{\beta^{\alpha+p}} \underbrace{\int_{\mathbb{R}_+} \frac{\beta^{\alpha+p}}{\Gamma(\alpha+p)} x^{p+\alpha-1}e^{-\beta x}dx}_{\text{gamma density}} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha+p)}{\beta^{\alpha+p}} \\ &= \frac{\Gamma(\alpha+p)}{\Gamma(\alpha)}\beta^{-p} && \text{domain is } \alpha + p > 0 \iff p > -\alpha\end{aligned}$$

Where if $\alpha, p \in \mathbb{N}$ *can be further simplified by the well known property that* $\alpha\Gamma(\alpha) = \Gamma(\alpha+1)\forall\alpha \in \mathbb{N}$:

$$p = n \in \mathbb{N}, \alpha \in \mathbb{N} \implies \mathbb{E}[X^n] = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\beta^{-p} = \frac{\prod_{j=0}^{n-1}\alpha + j}{\beta^p}$$

*(□ **rescaling**) a new r.v. of the form $Y = \beta X$ can be characterized in its distribution using Theorem 4.11 via bounded continuous positive borel functions and their expectations:*

$$h : \mathbb{R} \to \mathbb{R}_+ \quad \mathbb{E}[h(Y)]$$

If such integral is equal in $h \in C_b$ to a well known distribution, we are done.

$$\begin{aligned}
\mathbb{E}[h(Y)] &= \mathcal{P}_Y(h) \\
&= \int_{\mathbb{R}} h(y)\mathcal{P}_Y(dy) \\
&= \int_{\mathbb{R}} h(\beta x)\mathcal{P}_X(dx) \\
&= \int_{\mathbb{R}_+} h(\beta x)\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}dx && \text{let } y = \beta x, dy = \beta dx \\
&= \int_{\mathbb{R}_+} \frac{\beta^\alpha}{\Gamma(\alpha)}h(y)\left(\frac{y}{\beta}\right)^{\alpha-1} e^{-y}\frac{1}{\beta}dy \\
&= \int_{\mathbb{R}_+} \frac{1}{\Gamma(\alpha)}h(y)y^{\alpha-1}e^{-y}dy
\end{aligned}$$

Where we find the expectation of a $\mathcal{B}eta(\alpha, 1) = \mathcal{G}amma(\alpha, 1)$ distribution.

♢ **Observation 6.24** (About Example 6.23)**.** *The result $Y = \beta X \sim \mathcal{G}amma(\alpha, 1)$ justifies why we call the β parameter scale.*

♡ **Example 6.25** (Independent $\mathcal{G}$ammas)**.** *Consider $X \sim \mathcal{G}amma(\alpha_1, 1), Y \sim \mathcal{G}amma(\alpha_2, 1)$ with $X \perp\!\!\!\perp Y \iff \mathcal{P}_X\mathcal{P}_Y = \mathcal{P}_{X,Y}$. Define the joint distribution of two new r.v.s:*

$$(S, Z) = \left(X + Y, \frac{X}{X + Y}\right)$$

Usually, their distribution is extracted with transforms, now we wish to do so using the result of Theorem 4.11 with bounded continuous positive Borel functions $h : \mathbb{R} \to \mathbb{R}_+$. Namely, we check:

$$\mathcal{P}_{S,Z}(h) = \int_{\mathbb{R}^2} h(s, z)\mathcal{P}_{S,Z}(ds, dz) \quad s = x + y, z = \frac{x}{x + y}, \quad h \in C_b(\mathbb{R})$$

Applying a change of variable we see that:

$$\begin{cases} x = sz \\ y = s(1 - z) \end{cases} \qquad \mathcal{J} = \begin{bmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial z} \end{bmatrix} \text{ s.t. } |\mathcal{J}| = |-s| = |s|$$

Further, notice that:

$$(X, Y) \in \mathbb{R}_+ \times \mathbb{R}_+ \implies (S, Z) \in \mathbb{R}_+ \times [0, 1] \implies |s| = s$$

So that the expectation becomes:

$$\begin{aligned}
\mathcal{P}_{S,Z}(h) &= \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} h\left(x+y, \frac{x}{x+y}\right)\mathcal{P}_{X,Y}(dx,dy) \qquad \mathcal{P}_{X,Y} = \mathcal{P}_X\mathcal{P}_Y \\
&= \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} h\left(x+y, \frac{x}{x+y}\right)\frac{1}{\Gamma(\alpha_1)}x^{\alpha_1-1}e^{-x}\frac{1}{\Gamma(\alpha_2)}y^{\alpha_2-1}e^{-y}dxdy \\
&= \int_{\mathbb{R}_+}\int_0^1 h(s,z)\frac{1}{\Gamma(\alpha_1)}(sz)^{\alpha_1-1}e^{-sz}\frac{1}{\Gamma(\alpha_2)}(s(1-z))^{\alpha_2-1}e^{-s(1-z)}s\, dsdz \\
&= \int_{\mathbb{R}_+}\int_0^1 h(s,z)\frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)}s^{\alpha_1-1}z^{\alpha_1-1}e^{-sz-sz+s}s^{\alpha_2-1}(1-z)^{\alpha_2-1}s\, dsdz \\
&= \int_{\mathbb{R}_+}\int_0^1 h(s,z)\frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\underbrace{z^{\alpha_1-1}(1-z)^{\alpha_2-1}}_{\mathcal{B}eta\ Kernel}\underbrace{s^{\alpha_1+\alpha_2-1}e^{-s}}_{\mathcal{G}amma\ kernel}\, dsdz \\
&= \int_{\mathbb{R}_+}\int_0^1 h(s,z)\frac{1}{\Gamma(\alpha_1+\alpha_2)}s^{\alpha_1+\alpha_2-1}e^{-s}\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}z^{\alpha_1-1}(1-z)^{\alpha_2-1}\, dsdz \\
&= \int_{\mathbb{R}_+}\int_0^1 h(s,z)\mathcal{P}_S(ds)\mathcal{P}_Z(dz)
\end{aligned}$$

Where we see that $S \perp\!\!\!\perp Z$ as the laws factorize and $S \sim \mathcal{G}amma(\alpha_1+\alpha_2, 1), Z \sim \mathcal{B}eta(\alpha_1, \alpha_2)$.

Lemma 6.26 (Properties of $\widehat{\mathcal{P}}_X(t)$)**.** *Consider an integrable r.v. $X \in \mathcal{L}_1$. For a Laplace transform as in Def. 6.11:*

1. *$\widehat{\mathcal{P}}_X(t)$ is decreasing and continuous in t*
2. $\lim_{t\to\infty}\widehat{\mathcal{P}}_X(t) = 0$
3. $\lim_{t\to 0}\widehat{\mathcal{P}}_X(t) = 1$

Proof. We always use the $X \geq 0$ a.s. condition.
(Claim #1) follows as $e^{-tx} \leq 1 \forall t \geq 0$, so by dominated convergence (Thm. 4.24):

$$\lim_{h\downarrow 0}\widehat{\mathcal{P}}_X(t+h) = \widehat{\mathcal{P}}_X(t)\forall t > 0$$

For $h \uparrow 0$ observe that $e^{-tx} \geq 0 \forall x, t$ and apply the same reasoning.
Moreover, by e^{-tx} being decreasing in t, the integral is decreasing in t by monotonicity (Thm. 4.7#2).
(Claim #2, #3) Again by dominated convergence:

$$\begin{aligned}
\lim_{t\to 0}\mathbb{E}\left[e^{-tX}\right] &= \mathbb{E}\left[\lim_{t\to 0}e^{-tX}\right] = \mathbb{E}[1] = 1 \\
\lim_{t\to\infty}\mathbb{E}\left[e^{-tX}\right] &= \mathbb{E}\left[\lim_{t\to\infty}e^{-tX}\right] = \mathbb{E}[0] = 0
\end{aligned}$$

□

♣ **Theorem 6.27** (Laplace transform and seemingly exponential distribution connection)**.** *It holds:*

$$X\ :\ \mathbb{P}[X \geq 0] = 1 \implies For\quad \widehat{\mathcal{P}}_X(t)\quad \exists\mathcal{P}_T \ll Leb\quad p(t) = \int_{(0,\infty)} xe^{-xt}\mathcal{P}_X(dx)$$

Proof. Let $F(t) = 1 - \widehat{\mathcal{P}}_X(t)$, by Lemma 6.26, $F(t)$ satifies the properties of a distribution function, and we can say it identifies a unique probability law by Theorem 3.21. The distribution indexed by T with such cdf is:

$$\mathbb{P}[T > t] = \begin{cases} 1 & t < 0 \\ \widehat{\mathcal{P}}_X(t) & t \geq 0 \end{cases}$$

Where the random variable exists by Theorem 3.22.
It is also useful to recall the result of Lemma 6.15, namely $e^{-tx} = \int_t^\infty xe^{-sx}ds$. For positive times $t \geq 0$ we then

say:

$$\begin{aligned}\mathbb{P}[T>t] &= \mathbb{E}[e^{-tX}] \\ &= \int_{(0,\infty)}\int_t^\infty xe^{-sx}ds\mathcal{P}_X(dx) \\ &= \int_t^\infty\int_{(0,\infty)} xe^{-sx}\mathcal{P}_X(dx)ds && \text{Fubini Thm. B.30} \\ &= -\int_\infty^t\int_{(0,\infty)} xe^{-sx}\mathcal{P}_X(dx)ds\end{aligned}$$

Which is nothing but the representation of Radon Nykodym theorem (Thm. 5.7) for $\mathcal{P}_T \ll Leb$ with derivative:

$$\frac{d\mathcal{P}_T}{dLeb}(t) = \int_{(0,\infty)} xe^{-tx}\mathcal{P}_X(dx)$$

□

♣ **Theorem 6.28** (Laplace transform is a random variable)**.** *By the result of Theorem 6.27 we conclude:*

$$X\ :\ \mathbb{P}[X\geq 0]=1 \implies T=\frac{Y}{X}\ \textit{identified by}\ \widehat{\mathcal{P}}_X(t), X\sim\mathcal{P}_X, Y\sim\mathcal{E}xp(1), X\perp\!\!\!\perp Y$$

Proof. We restart from what we had at the end of Theorem 6.27 to get:

$$\begin{aligned}\mathcal{P}_T(h) &= \int_{\mathbb{R}_+} h(t)\mathcal{P}_T(dt) && h:\mathbb{R}\to\mathbb{R}_+ \text{ positive bounded continuous} \\ &= \int_{\mathbb{R}_+} h(t)\int_{(0,\infty)} xe^{-tx}\mathcal{P}_X(dx)dt && \text{let } tx=y, dt=dy\frac{1}{x} \\ &= \int_{\mathbb{R}_+}\int_{(0,\infty)} h\left(\frac{y}{x}\right)xe^{-y}\frac{1}{x}\mathcal{P}_X(dx)dy\end{aligned}$$

that implies:

$$\mathcal{P}_T(h) = \mathcal{P}_{X,Y}(h) = \int_{\mathbb{R}^2_+} h\left(\frac{y}{x}\right)e^{-y}\mathcal{P}_X(dx)dy$$

meaning that $X \perp\!\!\!\perp Y$ and $Y\sim\mathcal{N}eg\mathcal{E}xp(1)$. □

♡ **Example 6.29** (Laplace transform of $\mathcal{G}$amma random variable has $\mathcal{P}$areto distribution)**.** *Let $X\geq 0$ a.s. and $X\sim\mathcal{G}amma(\alpha,\beta)$, $\quad\alpha,\beta>0$. The Laplace transform is:*

$$\begin{aligned}\widehat{\mathcal{P}}_X(t) &= \mathbb{E}\left[e^{-tX}\right] \\ &= \int e^{-tx}\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}\int x^{\alpha-1}e^{-(\beta+t)x}dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha)}{(\beta+t)^\alpha} && \textit{using gamma density identity} \\ &= \left(\frac{\beta}{\beta+t}\right)^\alpha\end{aligned}$$

Which, by the identification via Laplace function (Thm. 6.12) suggests that T is a r.v. with ***Pareto distribution****, i.e. $T\sim\mathcal{P}areto(\alpha,\beta)$.*

♡ **Example 6.30** (Independent $\mathcal{G}$ammas and Laplace)**.** *Let $X\sim\mathcal{G}amma(\alpha,\beta), Y\sim\Gamma(\gamma,\beta),\ \ X\perp\!\!\!\perp Y$. The Laplace transform of their sum is:*

$$\begin{aligned}\widehat{\mathcal{P}}_{X+Y}(t) &= \widehat{\mathcal{P}}_X(t)\widehat{\mathcal{P}}_Y(t) && \textit{Prop. 6.14} \\ &= \left(\frac{\beta}{\beta+t}\right)^\alpha\left(\frac{\beta}{\beta+t}\right)^\gamma && \textit{Ex. 6.29} \\ &= \left(\frac{\beta}{\beta+t}\right)^{\alpha+\gamma}\end{aligned}$$

So that again, by Laplace characterization (Thm. 6.12), the sum of two $\mathcal{G}$amma independent r.v.s is $\mathcal{P}$areto distributed as $T \sim \mathcal{P}areto(\alpha+\gamma,\beta)$ and clearly $X+Y \sim \mathcal{G}amma(\alpha+\gamma,\beta)$

♡ **Example 6.31** (Theorem 6.27 for a $\mathcal{G}$amma gives $\mathcal{P}$areto by Theorem 6.28)**.** *Let $X \sim \mathcal{G}amma(\alpha,\beta)$ so that $T=\frac{X}{Y}$ is by Theorem 6.27:*

$$\mathcal{P}_T(dt)=\int_{(0,\infty)} xe^{-tx}\mathcal{P}_X(dx)dt=\frac{\beta^\alpha}{\Gamma(\alpha)}\int_{(0,\infty)} xe^{-xt}x^{\alpha-1}e^{-\beta x}dxdt=\frac{\alpha\beta^\alpha}{(\beta+t)^{\alpha+1}}\mathbb{1}_{(0,\infty)}(t)\,dt$$

Which is the pdf of a heavy tailed $\mathcal{P}$areto distribution.

♠ **Definition 6.32** (Integral transforms of random vectors)**.** *Let $X=(X_1,\dots,X_d)^T$ be a r.v. on $(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$. It is clear that $\forall t\in\mathbb{R}^d$ the inner product is:*

$$t^TX=\sum_{j=1}^d t_jX_j$$

so that there is a natural definition of characteristic function and Laplace transform:

- $\Phi_X(t)=\mathbb{E}\left[e^{it^TX}\right]=\int_{\mathbb{R}^d} e^{it^Tx}\mathcal{P}_X(dx)\quad t\in\mathbb{R}$
- *provided that* $\mathbb{P}[X_i\geq 0]=1\forall i$ *then* $\widehat{\mathcal{P}}_X(t)=\mathbb{E}\left[e^{-t^TX}\right]=\int_{\mathbb{R}_+^{\times d}} e^{-t^Tx}\mathcal{P}_X(dx)\quad t\in\mathbb{R}_+$

♢ **Observation 6.33** (Characteristic function of random Gaussian)**.** *Recall that for $X\sim\mathcal{N}(m,\sigma^2)$ the characteristic function is:*

$$\Phi_X(t)=e^{-itm-\frac{\sigma^2}{2}t^2}$$

♠ **Definition 6.34** (Multivariate Gaussian distribution)**.** *We say $X\sim\mathcal{N}^d(m,\Sigma)$ where $m\in\mathbb{R}^d$ and $\Sigma\in\mathbb{R}^{d\times d}$ is non negative definite if and only if:*

$$\forall t\in\mathbb{R}^d\begin{cases}\Phi_X(t)=\mathbb{E}\left[e^{it^TX}\right]=\exp\left\{im_t-\frac{\sigma_t^2}{2}\right\}\\ m_t=m^Tt=\sum_{j=1}^d m_jt_j\\ \sigma_t^2=t^T\Sigma t=\sum_j^d\sum_i^d\sigma_{ij}t_it_j\end{cases}$$

♢ **Observation 6.35** (Comments about the Definition)**.** *Notice that:*

- $\sigma_{ii}=V[X_i],\sigma_{ij}=CoV[X_i,X_j],m_i=\mathbb{E}[X_i]\quad\forall i,j$
- $\Sigma\succeq 0$ *(non negative definite) means either of two cases:*
  - $\Sigma\succ 0\implies\lambda^T\Sigma\lambda>0\quad\forall\lambda\iff\exists\Sigma^{-1}$ *so that $\mathcal{P}_X\ll Leb^d$ and we have a representation via Radon Nikodym (Thm. 5.7):*

    $$\frac{d\mathcal{P}_X}{dLeb^d}(x)=\frac{\det(\Sigma)^{-\frac{1}{2}}}{2\pi^{\frac{d}{2}}}\exp\left\{-\frac{1}{2}(x-m)^T\Sigma^{-1}(x-m)\right\}$$

  - $\Sigma\succeq 0\implies\nexists\Sigma^{-1}\implies\mathcal{P}_X$ *is concentrated on a hyperplane of dimension $d'<d$*

♣ **Proposition 6.36** (Orthogonality characterizes independence in Gaussian random vectors)**.** *Let $X\sim\mathcal{N}^d(m,\Sigma)$. Then*

$$\Sigma=\sigma^TI_d\iff X_i\perp\!\!\!\perp X_j\quad\forall i\neq j$$

Proof. By the above observation we notice that:

$$\Sigma=\sigma^TI_d\implies\exists\Sigma^{-1}=(\sigma^{-1})^TI_d\implies\exists\frac{d\mathcal{P}_X}{dLeb^d}$$

We can write down the characteristic transform as:

$$\Phi_X(t)=\exp\left\{i\sum m_jt_j-\frac{1}{2}\sum_j t_j^2\sigma_{jj}\right\}=\prod_j\exp\left\{im_jt_j-\frac{1}{2}t_j^2\sigma_{jj}\right\}=\prod_j\Phi_{X_j}(t)$$

so that by Proposition 6.21#1 and its trivial opposite the claim is verified. Namely we use: integral transform decouples implies independence and the opposite. □

Chapter Summary

Objects:

- random vectors via joint measurability
- independence as factorization of laws
- Laplace transform for positive r.v.s

$$\widehat{\mathcal{P}}_X(t) = \mathbb{E}[e^{-tX}] = \int_{\mathbb{R}_+} e^{-tx}\mathcal{P}_X(dx) \quad \widehat{\mathcal{P}}_X(0) = 1 \quad \forall t \geq 0$$

- characteristic function for r.v.s on $\mathbb{R}, t \in \mathbb{R}$

$$\Phi_X(t) = \mathbb{E}\left[e^{itX}\right] = \int_{\mathbb{R}} e^{itx}\mathcal{P}_X(dx) = \int_{\mathbb{R}} \cos(tx)\mathcal{P}_X(dx) + i\int_{\mathbb{R}} \sin(tx)\mathcal{P}_X(dx)$$

- probability generating function
- extension of characteristic function and Laplace transform to the multidimensional case via inner products
- multivariate Gaussian distribution via Characteristic function

Results:

- random vectors inherit most of the properties of random variables
- the Laplace transform:
  - characterizes distribution functions
  - factorizes sums of independent variables
  - its n^{th}-derivative gives a closed form expression of the n^{th}-moment
  - is decreasing and continuous in t
  - has $\lim_{t\to 0}$ equal to 1 and $\lim_{t\to\infty}$ equal to 0
  - has a a.e. unique distribution function T which is absolutely continuous to the Lebesgue measure and has a Radon Nikodym derivative involving $\mathcal{P}_X$
  - can be seen as a random variable which is the fraction of the original random variable and a negative exponential with parameter 1.
- the characteristic function:
  - characterizes distributions
  - always exists
  - is real valued for symmetric random variables
  - factorizes sums of independent variables
  - its n^{th}-derivative gives a closed form expression of the n^{th}-moment
- independence in Gaussian random vectors is characterized by a diagonal variance-covariance matrix

Chapter 7

Uniform Integrability & Inequalities

This Chapter is mostly based on [Ver18].

7.1 More requirements for integrability

Lemma 7.1 (integrability characterization)**.**

$$\begin{aligned} X \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}) &\iff \lim_{k\to\infty} \mathbb{E}\left[|X| \mathbb{1}_{(k,\infty)}(|X|)\right] = 0 \\ &\iff \forall \epsilon > 0 \exists k_0 = k_0(\epsilon) \mid \forall k > k_0 \quad \mathbb{E}\left[|X| \mathbb{1}_{(k,\infty)}(|X|)\right] < \epsilon \end{aligned}$$

Proof. ($\implies$) let $X \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ which means $\mathbb{E}[|X|] < \infty$. Notice that two trivial facts are:

$$|X| \mathbb{1}_{(k,\infty)}(|X|) \le |X| \quad \forall k > 0 \qquad \lim_{k\to\infty} \mathbb{1}_{(k,\infty)}(|X|) = 0$$

Using these two:

$$\begin{aligned} \lim_{k\to\infty} \mathbb{E}[|X| \mathbb{1}_{(k,\infty)}(|X|)] &= \mathbb{E}[|X| \lim_{k\to\infty} \mathbb{1}_{(k,\infty)}(|X|)] && \text{dominated conv. Thm. 4.24} \\ &= \mathbb{E}[|X| \cdot 0] \\ &= 0 \end{aligned}$$

($\impliedby$) We notice another trivial fact:

$$|X| = \mathbb{1}_{(-\infty,k)}(|X|)\,|X| + \mathbb{1}_{[k,\infty)}(|X|)\,|X| \le k + \mathbb{1}_{[k,\infty)}(|X|)\,|X|$$

So that by monotonicity (Thm. 4.7#2):

$$\mathbb{E}[|X|] \le k + \underbrace{\mathbb{E}[|X| \mathbb{1}_{[k,\infty)}(|X|)]}_{\to 0 \text{ by hyp}} < \infty \implies X \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$$

The second inequality statement is an implication of the hypothesis of finite limit. □

♢ **Observation 7.2** (About Lemma 7.1)**.** *Consider an uncountable collection $\mathcal{C} = \{X_i, i \in I\}$ for an arbitrary set I. In such a case, it might be that $k_0^{(i)}$ is divergent for some i.*

♠ **Definition 7.3** (Uniform Integrability)**.** *For a collection of r.v.s $\mathcal{C} = (X_i)_{i\in I}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we say it is uniformly integrable if:*

$$\forall \epsilon > 0 \quad \exists K > 0 \sup_{X\in\mathcal{C}} \left\{\mathbb{E}\left[|X| \mathbb{1}_{(K,\infty)}(|X|)\right]\right\} < \epsilon$$

More precisely:

- $|\mathcal{C}| < \infty \wedge X_i \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})\ \forall i \implies \mathcal{C}$ *u.i. with* $K^* = \max_{\mathcal{C}}\{K_X\}$
- *If* $\mathcal{C} = (X_n)_{n\in\mathbb{N}}$ *is countably or uncountably infinite, then we reformulate the definition as:*

$$\lim_{k\to\infty} \sup_{n\geq 1} \mathbb{E}\left[|X_n| \mathbb{1}_{(k,\infty)}(|X_n|)\right] = 0$$

Remark 1 (Norm bounds VS uniform integrability Def. 7.3)**.** *Notice that a u.i. collection of r.v.s is* $\mathcal{L}_1$*-bounded. Indeed:*

$$\begin{aligned}
\mathbb{E}[|X|] &= \mathbb{E}\left[|X|\mathbb{1}_{(0,k]}(|X|)\right] + \mathbb{E}\left[|X|\mathbb{1}_{(k,\infty))}(|X|)\right] \\
&\leq k + h(k) && h(k) = \mathbb{E}\left[|X|\mathbb{1}_{(k,\infty)}(|X|)\right] \\
\implies \sup_{\mathcal{C}} \mathbb{E}|X| &\leq \sup_{\mathcal{C}} k + h(k) \\
&\leq 1 + k && \text{for } k \mid h(k) \leq 1 \\
&< \infty
\end{aligned}$$

Where we exploit the fact that $h(k) \to 0$ *as* $k \to \infty$ *which is obtained by the assumption of uniform Integrability (Def. 7.3). This means that the whole collection of variables has an* $\mathcal{L}_1$ *norm that is finite for each norm, and we can choose a value such that all of the collection norms are below it.*

7.2 Concentration inequalities

♣ **Theorem 7.4** (Markov's inequality)**.** *For a r.v.* X *from* $(\Omega, \mathcal{F}, \mathbb{P})$ *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$*, an increasing measurable function* $f : \mathbb{R} \to \mathbb{R}_+$ *and* $b \in \mathbb{R}$ *with* $f(b) \neq 0$*, we have that:*

$$\mathbb{P}[X > b] \leq \frac{1}{f(b)} \mathbb{E}[f(X)]$$

Proof. We go on by simple computation:

$$\begin{aligned}
\mathbb{P}[X > b] = \int_b^\infty \mathcal{P}_X(dx) &\leq \int_b^\infty \frac{f(x)}{f(b)} \mathcal{P}_X(dx) && \frac{f(x)}{f(b)} \geq 1 \forall x \geq b \text{ \& monotonicity 4.7\#2} \\
&= \frac{1}{f(b)} \int_b^\infty f(x)\mathcal{P}_X(dx) \\
&\leq \frac{1}{f(b)} \int_{\mathbb{R}_+} f(x)\mathcal{P}_X(dx) && X \geq 0 \text{ a.s.} \\
&= \frac{1}{f(b)} \mathbb{E}[f(x)]
\end{aligned}$$

□

Corollary 7.5 (Chebyshev's inequality)**.** *Let* X *be such that* $\mathbb{E}[X] = m$ *and* $V[X] = \sigma^2$*, then:*

$$\mathbb{P}[|X - m| > \epsilon] \leq \frac{1}{\epsilon^2}\sigma^2$$

Proof.

$$\begin{aligned}
\mathbb{P}[|X - m| > \epsilon] = \mathbb{P}[(X - m)^2 > \epsilon^2] &\leq \frac{\mathbb{E}[(X - m)^2]}{\epsilon^2} && \text{Markov's Thm. 7.4} \\
&= \frac{V[X]}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2}
\end{aligned}$$

□

Lemma 7.6 (Convexity Implication by Hardy, Littlewood, Polya). *We characterize convexity of a function $f : X \to \mathbb{R}$ by:*

$$\begin{cases} f(tx_1 + (1-t)x_2) \le tf(x_1) + (1-t)f(x_2) \\ \forall x_1, x_2 \in X,\ \forall t \in (0,1) \end{cases} \iff \begin{cases} f(x) \ge f(x_0) + (x - x_0)f_r'(x_0) \\ f_r'(x_0) := \lim_{h \downarrow 0} \frac{f(x_0+h) - f(x_0)}{h} \\ \forall x \in X, x_0 \in X \end{cases}$$

♣ **Theorem 7.7** (Jensen's Inequality). *For a convex function $f : \mathbb{R} \to \mathbb{R}$ and integrable r.v.s $X, f(X) \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ the following inequality holds:*

$$\mathbb{E}[f(X)] \ge f(\mathbb{E}[X])$$

Proof. Using Lemma 7.6 with $x_0 = f(\mathbb{E}[X])$ we have:

$$\begin{aligned} \mathbb{E}[f(x)] &\ge \mathbb{E}[f(\mathbb{E}[X])] + \underbrace{\mathbb{E}[X - \mathbb{E}[X]]}_{=\mathbb{E}[X]-\mathbb{E}[X]=0} f_r'(\mathbb{E}[X]) \\ &= f(\mathbb{E}[X]) \end{aligned}$$

Notice that we required also $f(X) \in \mathcal{L}_1$ otherwise the expression $\mathbb{E}[f_r'(\mathbb{E}[f(X)])]$ might have led to an undecidable form when multiplied with $X - \mathbb{E}[X]$. □

Corollary 7.8 (Concave Jensen's Inequality). *For $f : \mathbb{R} \to \mathbb{R}$ concave the opposite sign of Theorem 7.7 is verified:*

$$\mathbb{E}[f(X)] \le f(\mathbb{E}[X])$$

♢ **Observation 7.9** (Intro to concentration inequalities). *We wish to find an upper bound for the Probability of deviating from the mean. The desired result is similar to the negative exponential convergence rate of the CLT, but at finite n. Indeed we know that at the limit any mean distribution will converge to a normal, but wish to do so at a finite sample size.*

♣ **Theorem 7.10** (Hoeffding's Inequality, One sided, symmetric Bernoullis). *For symmetric iid Bernoulli r.v.s:*

$$X_i \overset{iid}{\sim} \mathcal{B}ern_{\pm 1}\left(\frac{1}{2}\right) \implies \forall t \ge 0 \quad \mathbb{P}\left[\frac{1}{\sqrt{n}} \sum X_i \ge t\right] \le e^{-\frac{t^2}{2}}$$

Proof. **(△ idea and a simple fact)** Notice that:

$$\mathbb{E}[X_i] = 0 \forall i \implies \mathbb{E}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i\right] = 0$$

The strategy is using Markov's inequality, a monotonic transformation and a Taylor hyperbolic expansion.
(□ **preliminary computation)** We have:

$$\begin{aligned}
\mathbb{P}\left[\frac{1}{\sqrt{n}}\sum X_i \geq t\right] &= \mathbb{P}\left[e^{\frac{1}{\sqrt{n}}\sum X_i} \geq e^t\right] && x \to e^x \text{ increasing} \\
&= \mathbb{P}\left[e^{\lambda\frac{1}{\sqrt{n}}\sum X_i} \geq e^{\lambda t}\right] && \forall\, \lambda > 0 \\
&\leq e^{-\lambda t}\mathbb{E}\left[e^{\lambda\frac{1}{\sqrt{n}}\sum X_i}\right] && \text{Markov's Thm. 7.4} \\
&= e^{-\lambda t}\prod_{i=1}^{n}\mathbb{E}\left[e^{\lambda\frac{1}{\sqrt{n}}X_i}\right] && X_i \perp\!\!\!\perp X_j\ \forall i \neq j \quad iid \\
&= e^{-\lambda t}\prod_{i=1}^{n} e^{\lambda\frac{1}{\sqrt{n}}}\underbrace{\frac{1}{2}}_{=\mathbb{P}[X_i=1]} + e^{-\lambda\frac{1}{\sqrt{n}}}\underbrace{\frac{1}{2}}_{=\mathbb{P}[X_i=0]} \\
&= e^{-\lambda t}\prod_{i=1}^{n}\frac{1}{2}\left(e^{\lambda\frac{1}{\sqrt{n}}} + e^{-\lambda\frac{1}{\sqrt{n}}}\right) \\
&= e^{-\lambda t}\prod_{i=1}^{n}\cosh\frac{\lambda}{\sqrt{n}} && \cosh x = \frac{1}{2}(e^x + e^{-x}) \\
&= e^{-\lambda t}\left(\cosh\frac{\lambda}{\sqrt{n}}\right)^n
\end{aligned}$$

(◯ **Taylor hyperbolic computation)** notice that the Taylor series of the hypoerbolic cosine is, using the simple inequality $(2n)! \geq 2^n n!$:

$$\cosh x = \sum_{n=0}^{\infty}\frac{x^{2n}}{(2n)!} \leq \sum_{n=0}^{\infty}\frac{x^{2n}}{2^n(n)!} = \sum_{n=0}^{\infty}\frac{\left(\frac{x^2}{2}\right)^n}{n!} = e^{\frac{x^2}{2}}$$

Where in the last passage we recognize the Taylor series of the exponential.
(▽ **inequality)** Applying ◯ to □ we get:

$$\mathbb{P}\left[\frac{1}{\sqrt{n}}\sum X_i \geq t\right] \leq e^{-\lambda t}\left(e^{\frac{\lambda^2}{2n}}\right)^n = \exp\left\{-\lambda t + \frac{\lambda^2}{2}\right\} \quad \forall\lambda > 0$$

Where the RHS depends on λ. To find the tighthest bound, we minimize it wrt λ to find:

$$\begin{aligned}
&\min_{\lambda>0}\exp\left\{-\lambda t + \frac{\lambda^2}{2}\right\} = \min_{\lambda>0} -\lambda t + \frac{\lambda^2}{2} && \text{FOC } \frac{\partial}{\partial\lambda} = 0 \iff t = \lambda \\
\implies &\min_{\lambda>0}\exp\left\{-\lambda t + \frac{\lambda^2}{2}\right\} = \exp^{-t^2+\frac{t^2}{2}} = e^{-\frac{t^2}{2}} \\
\implies &\mathbb{P}\left[\frac{1}{\sqrt{n}}\sum X_i \geq t\right] \leq e^{-\frac{t^2}{2}}
\end{aligned}$$

□

♢ **Observation 7.11** (About Theorem 7.10)**.** *We are restricting our analysis to one sided identically distributed r.v.s! It is possible to generalize further the bound.*

Corollary 7.12 (Two sided Hoeffding's inequality, symmetric Bernoulli)**.** *Extending the result of Theorem 7.10:*

$$X_i \overset{iid}{\sim} \mathcal{B}ern_{\pm 1}\left(\frac{1}{2}\right) \implies \forall t \geq 0 \quad \mathbb{P}\left[\frac{1}{\sqrt{n}}\left|\sum X_i\right| \geq t\right] \leq 2e^{-\frac{t^2}{2}}$$

Proof. Observe that $\{X_i\}$ symmetric $\implies -X_i \overset{d}{=} X_i \quad \forall i$ and in general $|X| = X^+ + X^-$. Then:

$$\mathbb{P}\left[\frac{1}{\sqrt{n}}\left|\sum X_i\right| \geq t\right] \leq \mathbb{P}\left[\frac{1}{\sqrt{n}}\sum |X_i| \geq t\right] = \mathbb{P}\left[\frac{1}{\sqrt{n}}\sum X_i \geq t\right] + \mathbb{P}\left[\frac{1}{\sqrt{n}}\sum -X_i \geq t\right] \leq 2e^{-\frac{t^2}{2}}$$

Where the first passage is by monotonicity of probability (Thm. 2.10) and the triangle inequality, and the second passage is using the general fact to split the sum into disjoint events that both the positive and the negative sum are less than t. This last event has a greater probability than their sum being less than t. The last inequality follows by using the symmetry argument of Hoeffding's inequality for the second term, and the usual inequality for the first term. □

◇ **Observation 7.13** (About Hoeffding's inequality Thm. 7.10 and 7.12)**.** *Additionally:*

- *Hoeffding's Inequality tells how much mass is farther than t*
- *is tighter than Chebyshev's bound from Cor. 7.5*
- *it is possible to ignore the 2 coefficient in front of the two sided inequality as $t \to \infty$*

What about going beyond Bernoulli r.v.s?

Assumption 7.14 (Boundedness of random variables)**.** *We assume there exists finite m, M such that:*

$$\mathbb{P}[m \leq X \leq M] = 1$$

Which is an a.s. bound for any r.v. considered.

Lemma 7.15 (Moments of buonded random variables)**.**

$$X \text{ as in Ass. } 7.14 \implies \forall n \in \mathbb{N} \quad \exists \mathbb{E}[X^n]$$

Proof. Note that the moments are always bounded in powers $[m^n, M^n]$ by monotonicity of expectation (Thm. 4.7#1), which means that they are always finite. □

Lemma 7.16 (Symmetrization argument)**.** *A r.v. X satisfying Ass. 7.14 is such that:*

$$\mathbb{E}\left[e^{\lambda(X-\mathbb{E}[X])}\right] \leq e^{\frac{\lambda^2(M-m)^2}{2}} \quad \forall \lambda > 0$$

Proof. **(△ independent copies approach)** Let $X' \overset{d}{=} X$, $X' \perp\!\!\!\perp X$ be an independent copy. By Lemma 7.15 it always has moments. Then:

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda(X-\mathbb{E}[X])}\right] &= \mathbb{E}\left[e^{\lambda(X-\mathbb{E}[X'])}\right] && \lambda > 0,\ X' \overset{d}{=} X \\
&= \mathbb{E}\left[f(\mathbb{E}[X'])\right] && f(y) = e^{\lambda(X-y)} \text{ convex} \\
&\leq \mathbb{E}\left[\underbrace{\mathbb{E}[f(X')]}_{\text{deterministic}}\right] && \text{Jensen's Thm. 7.7} \\
&= \mathbb{E}[f(X')] \\
&= \mathbb{E}\left[e^{\lambda(X-X')}\right]
\end{aligned}$$

(□ introducing a Bernoulli) Now notice that $X \overset{d}{=} X' \implies X - X' \overset{d}{=} X' - X \overset{d}{=} -(X - X')$ so that the difference of independent copies is symmetric. Thanks to this, if $\varepsilon \sim \mathcal{B}ern_{\pm 1}\left(\frac{1}{2}\right)$ is a symmetric Bernoulli then:

$$X - X' \overset{d}{=} \varepsilon(X - X')$$

Plugging this back into the inequality of △:

$$\mathbb{E}\left[e^{\lambda(X-\mathbb{E}[X])}\right] \leq \mathbb{E}\left[e^{\lambda(X-X')}\right] = \mathbb{E}\left[e^{\lambda\varepsilon(X-X')}\right] = \mathbb{E}\left[\mathbb{E}[e^{\lambda\varepsilon(X-X')}|X, X']\right]$$

Where we anticipate the intuitive use of the iterated law of Expectation.
Notice that inside the expectation we are **fixing** X, X', so that they are not random. This is useful since we can apply the hyperbolic reasoning in the proof of Hoeffing's inequality (Thm. 7.12) to the last form wrt ε:

$$\mathbb{E}\left[e^{\lambda\varepsilon(X-X')}|X, X'\right] = \frac{1}{2}e^{\lambda(X-X')} + \frac{1}{2}e^{-\lambda(X-X')} \leq e^{\frac{\lambda^2(X-X')^2}{2}}$$

where the last inequality follows again by the reasoning of the proof cited above with Taylor's expansions.
(○ **boundedness finalization)** Up to now, we have not exploited Assumption 7.14. Notice that:

$$\mathbb{P}[m \leq X \leq M] = \mathbb{P}[m \leq X' \leq M] \iff \mathbb{P}[|X - X'| \leq M - m] = 1 = \mathbb{P}[(X - X')^2 \leq (M - m)^2]$$

Which means that the maximum distance between X and X' is necessarily less than $|M - m|$. Plugging this into the result of □ it is trivial to conclude that:

$$\mathbb{E}\left[e^{\lambda(X-\mathbb{E}[X])}\right] \leq \mathbb{E}\left[\exp\left\{\frac{\lambda^2(X - X')^2}{2}\right\}\right] \leq e^{\frac{\lambda^2(M-m)^2}{2}}$$

□

♣ **Theorem 7.17** (Hoeffding's general inequality for bounded random variables)**.** *Assume independence (Def. 6.9) and that Ass. 7.14 holds. Then:*

$$\forall t \geq 0 \quad \mathbb{P}\left[\sum^n X_i - \mathbb{E}[X_i] \geq t\right] \leq exp\left\{-\frac{t}{2\sum^n (M_i - m_i)^2}\right\}$$

Proof. **(△ improving the parametric bound)** Recalling the proof of Hoeffding's one sided Inequality (Thm. 7.10) we have:

$$\begin{aligned}
\mathbb{P}\left[\sum X_i - \mathbb{E}[X_i] \geq t\right] &\leq e^{-\lambda t}\mathbb{E}\left[e^{\lambda(\sum X_i - \mathbb{E}[X_i])}\right] && \forall \lambda > 0\\
&= e^{-\lambda t}\prod_{i=1}^n \mathbb{E}\left[e^{\lambda(X_i - \mathbb{E}[X_i])}\right] && \text{independence}\\
&\leq e^{-\lambda t}\prod_{i=1}^n \exp\left\{\frac{\lambda^2(M_i - m_i)^2}{2}\right\} && \text{symmetrization, Lem. 7.16}\\
&= \exp\left\{-\lambda t + \frac{1}{2}\lambda^2\sum_{i=1}^n (M_i - m_i)^2\right\} && \forall \lambda > 0
\end{aligned}$$

(□ best overall λ) To obtain the tightest bound we search:

$$\lambda^* = \arg\min_{\lambda>0} \exp\left\{-\lambda t + \frac{1}{2}\lambda^2\sum_{i=1}^n (M_i - m_i)^2\right\} = \arg\min_{\lambda>0} -\lambda t + \frac{1}{2}\lambda^2\sum_{i=1}^n (M_i - m_i)^2 = \arg\min_{\lambda>0} F$$

The FOC suggests:

$$\frac{\partial}{\partial\lambda}F = -t + \lambda(\sum(M_i - m_i)^2) = 0, \quad \frac{\partial^2}{\partial\lambda^2}F \geq 0 \implies \lambda^* = \frac{t}{\sum(M_i - m_i)^2}$$

(○ non parametric bound) Plugging the λ^* from □ into the inequality of △:

$$\mathbb{P}\left[\sum X_i - \mathbb{E}[X_i] \geq t\right] \leq \exp\left\{-\frac{t^2}{\sum(M_i - m_i)^2} + \frac{t^2}{2\sum(M_i - m_i)^2}\right\} = \exp\left\{-\frac{t^2}{2\sum(M_i - m_i)^2}\right\} \quad \forall t \geq 0$$

□

◇ **Observation 7.18** (About Thm. 7.17)**.** *Dependence on n in the RHS is highlighted by $\sum^n (M_i - m_i)^2$. Indeed, we did not normalize by $\frac{1}{\sqrt{n}}$.*

Lemma 7.19 (Hoeffding's Lemma)**.** *Let X be a bounded r.v.:*

$$X : \mathbb{E}[X] = 0 \; \mathbb{P}[m \leq X \leq M] = 1, \; m, M \in \mathbb{R} \implies \forall \lambda \in \mathbb{R} \quad \mathbb{E}\left[e^{\lambda X}\right] \leq exp\left\{\frac{\lambda^2(M - m)^2}{8}\right\}$$

Proof. (△ **preliminary bound**) Notice that $V[X] = \mathbb{E}[X^2] \leq \mathbb{E}[(X-c)^2] \quad \forall c \in [m, M]$. In particular, for the midpoint $c = \frac{M+m}{2}$ we obtain:

$$V[X] \leq \mathbb{E}\left[\left(X - \frac{M+m}{2}\right)^2\right] \leq \frac{(M-m)^2}{4}$$

By the simple fact that for $X \in [m, M]$ we have:

$$X^2 - (m+M)X \leq 0 \implies \left(X - \frac{M+m}{2}\right)^2 = X^2 - (M+m)X + \frac{(m+M)^2}{4} \leq \frac{(M-m)^2}{4}$$

Simply since the maximum of $X^2 - (m+M)X$ is at zero in the midpoint.
(□ **another preliminary bound**) let $f(\lambda) = \log \mathbb{E}[e^{\lambda X}]$. Then:

$$f'(\lambda) = \frac{\mathbb{E}Xe^{\lambda X}}{\mathbb{E}e^{\lambda X}} \quad f''(\lambda) = \frac{\mathbb{E}X^2e^{\lambda X}}{\mathbb{E}e^{\lambda X}} - \frac{(\mathbb{E}Xe^{\lambda X})^2}{(\mathbb{E}e^{\lambda X})^2}$$

Also, $f''(\lambda) = V[U]$ where U is such that:

$$\mathcal{P}_U(du) = \frac{e^{\lambda u}\mathcal{P}_X(du)}{\int_m^M e^{\lambda x}\mathcal{P}_X(dx)} \quad \mathbb{P}[m \leq U \leq M] = 1 \implies f''(\lambda) = V[U] \leq \frac{(M-m)^2}{4}$$

(◯ **final computation**) Notice further that $f(0) = f'(0) = 0$ and by the fundamental Theorem of calculus:

$$f(\lambda) = f(0) + \int_0^\lambda f'(\mu)d\mu = \int_0^\lambda \int_0^\mu f''(s)dsd\mu \leq \frac{\lambda^2(M-m)^2}{8}$$

Which helps us in computing:

$$e^{f(\lambda)} = e^{\log \mathbb{E}e^{\lambda x}} = \mathbb{E}e^{\lambda x} \leq e^{\frac{\lambda^2(M-m)^2}{8}}$$

□

Corollary 7.20 (Hoeffding's Bounded inequality improved)**.** *Using Lemma 7.19 we improve the result of Theorem 7.17 as:*

$$\mathbb{P}\left[\sum X_i - \mathbb{E}[X_i] \geq t\right] \leq exp\left\{-\frac{2t^2}{\sum(M_i - m_i)^2}\right\}$$

Proof. By direct application of Hoeffding's Lemma (Lem. 7.19) above. □

♡ **Example 7.21** (Algorithm runs, applying Hoeffding's inequality)**.** *Consider an algorithm $\mathcal{A}$ for a decision problem. For each stage, the decision $d \in \{0, 1\}$ is positive or negative. Assume further we have a better than guessing performance:*

$$\mathbb{P}[succes] = \frac{1}{2} + \delta \quad \delta \in \left(0, \frac{1}{2}\right)$$

We run such procedure many times, and make the final decision based on a majority vote. The question we ask is:

Fix $\epsilon > 0$, how many runs n are needed to have $\mathbb{P}[correct] \geq 1 - \epsilon$?

*Notice that success and fail denote the **decision** of $\mathcal{A}$ while correct and wrong denote the ground truth. We proceed by setting:*

$$X_i = \begin{cases} 1 & w.p. \ \frac{1}{2} - \delta \quad (fail) \\ 0 & w.p. \ \frac{1}{2} + \delta \quad (success) \end{cases} \implies S_n = \sum_{i=1}^n X_i = \#fails$$

With $X_i \sim \mathcal{B}ern\left(\frac{1}{2} - \delta\right)$ so that $\mathbb{E}[X_i] = \frac{1}{2} - \delta \forall i$ and the natural bounds could be $m_i = 0, M_i = 1 \forall i$. Applying Hoeffding's (Cor. 7.20) we get that:

$$\mathbb{P}\left[\sum X_i - \mathbb{E}\left[X_i\right] \geq t\right] \leq \exp\left\{-\frac{2t^2}{n}\right\} \quad \forall t > 0$$

Then for all positive t:

$$\mathbb{P}\left[\frac{1}{n}\sum X_i \geq \frac{1}{2} - \delta + \frac{t}{n}\right] \leq \exp\left\{-\frac{2t^2}{n}\right\}$$

since we just exploited the fact that $\mathbb{E}[X_i] = \dfrac{1}{2} - \delta$. *Picking* $t = n\delta > 0$ *we can recover the estimate on the probability that the majority of the votes is fail as:*

$$\mathbb{P}\left[\frac{1}{n}\sum X_i \geq \frac{1}{2} - \delta + \delta\right] \leq \exp\left\{-\frac{2t^2}{n}\right\}$$

We want to find a value of runs n such that $\frac{1}{n}S_n \geq \frac{1}{2}$ has a small probability of occurrence ϵ. This holds if and only if:

$$n : \frac{1}{n}(n - S_n) \geq \frac{1}{2}$$

has high $1-\epsilon$ probability. In other words, for $t = n\delta$:

$$\mathbb{P}\left[\frac{1}{n}\sum X_i \geq \frac{1}{2}\right] = \mathbb{P}\left[\frac{1}{n}S_n \geq \frac{1}{2}\right] \leq \exp\left\{-\frac{2t^2}{\underbrace{n}_{t=n\delta}}\right\} < \epsilon$$

is equivalent, after adapting for our carefully chosen t, to:

$$\exp\left\{\frac{-2n^2\delta^2}{n}\right\} < \epsilon \iff n > \frac{1}{2\delta^2}\log\left[\frac{1}{\epsilon}\right] = -\frac{1}{2\delta^2}\log[\epsilon]$$

♣ **Theorem 7.22** (Chernoff's bound). *For an independency (Def. 6.9) of finite Bernoulli r.v.s $\{X_i\}_{i=1}^n$, where $X_i \sim \mathcal{B}ern(p_i)$ assign the symbols $S_n \coloneqq \sum X_i$, $\mu = \mathbb{E}[S_n] = \sum p_i$. Then:*

$$\mathbb{P}[S_n \geq t] \leq e^{-\mu}\left(\frac{e\mu}{t}\right)^t \quad \forall t \geq \mu$$

Proof. We have:

$$\begin{aligned}
\mathbb{P}[S_n \geq t] &= \mathbb{P}\left[e^{\lambda S_n} \geq e^{\lambda t}\right] && \forall \lambda > 0, x \to e^{f(x)} \text{ increasing} \\
&\leq e^{-\lambda t}\mathbb{E}\left[e^{\lambda S_n}\right] && \text{Markov's Thm. 7.4} \\
&= e^{-\lambda t}\prod_{i=1}^n \mathbb{E}\left[e^{\lambda X_i}\right] && \text{independence} \\
&= e^{-\lambda t}\prod_{i=1}^n p_i e^{\lambda} + 1 - p_i && \text{Bernoulli} \\
&= e^{-\lambda t}\prod_{i=1}^n 1 + p_i(e^{\lambda} - 1) && \\
&\leq e^{-\lambda t}\prod_{i=1}^n e^{p_i(e^{\lambda}-1)} && 1 + x \leq e^x \forall x \quad x = p_i(e^{\lambda} - 1) \\
&= e^{-\lambda t + \mu(e^{\lambda}-1)} && \forall \lambda > 0
\end{aligned}$$

Now notice:

$$0 < \lambda^* = \log\frac{t}{\mu} = \arg\min_{\lambda \in \mathbb{R}_+} e^{-\lambda t + \mu(e^{\lambda}-1)}$$

and conclude that:

$$\begin{aligned}
\mathbb{P}[S_n \geq t] &\leq \exp\left\{-\log\left[\frac{t}{\mu}\right]t + \mu\frac{t}{\mu} - \mu\right\} \\
&= \exp\{-\mu\}\exp\left\{\left(\log\left[\frac{\mu}{t}\right] + 1\right)t\right\} \\
&= \exp\{-\mu\}\left(\frac{e\mu}{t}\right)^t && \forall t \geq \mu
\end{aligned}$$

□

♢ **Observation 7.23** (About Chernoff's bound Thm. 7.22). *The bound is even more than exponential with a* $e^{-t\log(t)}$ *rate.*

♢ **Observation 7.24** (There is a two sided bound). *For now, we just considered the case* $t \geq \mu$*, and not the other side.*

♣ **Theorem 7.25** (Two sided Chernoff's inequality). *With the assumptions of Theorem 7.22 we further conclude that for* $\{X_i\}_{i=1}^n,\ X_i \overset{ind}{\sim} \mathcal{B}ern(p_i),\ S_n = \sum X_i,\ \mu = \mathbb{E}[S_n]$*:*

$$\implies \exists c > 0\ :\ \forall \delta \in (0,1) \quad \mathbb{P}[|S_n - \mu| \geq \delta\mu] \leq 2e^{-c\frac{\mu\delta^2}{2}}$$

Proof. We first split the sum into two terms:

$$\mathbb{P}[|S_n - \mu| \geq \delta\mu] = \mathbb{P}[S_n \geq (1+\delta)\mu] + \mathbb{P}[S_n \leq (1-\delta)\mu]$$

(△ first term) a direct application of Chernoff's bound (Thm. 7.22) is possible.
(□ second term) The focus is moved to the second term.

$$\begin{aligned}
\mathbb{P}[S_n \leq (1-\delta)\mu] &= \mathbb{P}[-S_n \geq -(1-\delta)\mu] \\
&= \mathbb{P}[-\lambda S_n \geq -\lambda(1-\delta)\mu] && \forall \lambda > 0 \\
&= \mathbb{P}\left[e^{-\lambda S_n} \geq e^{-\lambda(1-\delta)\mu}\right] \\
&\leq e^{\lambda(1-\delta)\mu}\mathbb{E}\left[e^{-\lambda S_n}\right] && \text{Markov's Thm. 7.4} \\
&= e^{\lambda(1-\delta)\mu}\prod_{i=1}^n \mathbb{E}\left[e^{-\lambda X_i}\right] && \text{indepedence} \\
&= e^{\lambda(1-\delta)\mu}\prod_{i=1}^n p_i e^{-\lambda} + 1 - p_i \\
&\leq e^{\lambda(1-\delta)\mu}e^{-\mu(1-e^{\lambda})} && \text{Like proof of Thm. 7.22}
\end{aligned}$$

As a minimizer, we choose $\lambda^* = -\log(1-\delta)$ so that eventually:

$$\begin{aligned}
\mathbb{P}[S_n \leq (1-\delta)\mu] &\leq \exp\{-\log[1-\delta](1-\delta)\mu - \mu(1-1+\delta)\} \\
&= \exp\{(1-\delta)\mu(-\log[1-\delta]) - \mu\delta\} \\
&= \exp\{-\delta\mu\}\left(\exp\log\left\{\frac{1}{1-\delta}\right\}\right)^{(1-\delta)\mu} \\
&= e^{-\delta\mu}\left(\frac{1}{1-\delta}\right)^{(1-\delta)\mu}
\end{aligned}$$

(◯ back to main problem) collecting the results of △, □ we get to:

$$\mathbb{P}[|S_n - \mu| \geq \delta\mu] \leq e^{-\mu}\left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu} + e^{-\delta\mu}\left(\frac{1}{1-\delta}\right)^{(1-\delta)\mu} = \frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}} + \frac{e^{-\delta\mu}}{(1-\delta)^{(1-\delta)\mu}}$$

Where it is possible to notice that:

$$\begin{aligned}
\frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}} &\leq e^{\delta\mu}e^{-\mu(1+\delta)}\frac{\delta}{1+\frac{\delta}{2}} && \log(1+x) \geq \frac{x}{1+\frac{x}{2}},\ x = \delta \\
&= e^{-\frac{-\mu\delta^2}{2+\delta}}
\end{aligned}$$

Reasoning similarly with the inequality $\log(1-\delta) \geq \frac{-\delta+\frac{\delta^2}{2}}{1-\delta} \quad \forall \delta \in (0,1)$ we get:

$$\frac{e^{-\delta\mu}}{(1-\delta)^{(1-\delta)\mu}} \leq e^{\frac{-\mu\delta^2}{2}}$$

Recollecting again the conclusion follows:

$$\exists c > 0\ :\ \forall \delta \in (0,1) \quad \mathbb{P}[|S_n - \mu| \geq \delta\mu] \leq 2e^{-c\frac{-\mu\delta^2}{2}}$$

Where the c is extracted since we could make the bound tighter in principle. □

7.3 A Graph Theoretic Application

♠ **Definition 7.26** (Erdős-Renyi model for random graphs $\mathcal{G}(\cdot,\cdot)$). *Consider a set V of vertices where $|V| = n$ and connections between edges $(i,j) \in V \times V$ taking place with propability p independently. We refer to this model as $\mathcal{G}(n,p)$ and further notice that the single connections have a Bernoulli distribution:*

$$X_j^{(i)} = \begin{cases} 1 & if i \sim j \\ 0 & else \end{cases} \quad \forall i \neq j$$

♠ **Definition 7.27** (Degree d_i and average degree $\overline{d}$). *For a graph $\mathcal{G}(n,p)$ define:*

- *the number of incident to node i edges as $d_i = \sum_{j \neq i} X_j^{(i)}$ where $d_i \sim Bin(n-1,p)$*
- *the average degree $\overline{d} := \mathbb{E}[d_i] = (n-1)p$*

♠ **Definition 7.28** (Dense random graph). *A random graph $\mathcal{G}(n,p)$ is dense when:*

$$\overline{d} \geq C\log(n) \quad C \in \mathbb{R}$$

♠ **Definition 7.29** (Almost regular random graph). *A random graph $\mathcal{G}(n,p)$ is almost regular when:*

$$\mathbb{P}\left[\bigcap_{i=1}^{n} \{|d_i - \overline{d}| > \epsilon\}\right] \leq 1 - \epsilon \quad \forall \epsilon > 0$$

In plain text, there is a high probability that the degrees of nodes are close to their mean.

Lemma 7.30 (Dense random graphs almost regularity). *Consider a dense (Def. 7.28) random graph $\mathcal{G}(n,p)$ (Def. 7.26), then:*

$$\forall \epsilon > 0, \delta > 0 \; \exists C \in \mathbb{R} \; : \; \overline{d} \geq C\log(n), \quad \mathbb{P}\left[\bigcap_{i=1}^{n} \{|d_i - \overline{d}| > \delta\overline{d}\}\right] \leq 1 - \epsilon$$

With C satisfying:

$$C > \frac{1}{c\delta^2} + \log\left(\frac{2}{\epsilon}\right)\frac{1}{c\delta^2\log(n)}$$

Where the small c comes from the Chernoff Bound (Theorem 7.25).

Proof. **(△ setting)** We make a direct application of the two sided Chernoff's bound from Theorem 7.25, where $X_i \rightsquigarrow d_i$ and $\mu \rightsquigarrow \overline{d}$.
(□ starting point and another fact) it follows that for some $c > 0$, $\forall \delta \in (0,1)$

$$\mathbb{P}[|d_i - \overline{d}| \geq \delta\overline{d}] \leq 2e^{-\frac{\overline{d}\delta^2}{2}}$$

Additionally, by De Morgan's laws:

$$\left(\bigcap_{i=1}^{n} \{|d_i - \overline{d}| \geq \delta\overline{d}\}\right)^c = \bigcup_{i=1}^{n} \{|d_i - \overline{d}| < \delta\overline{d}\}$$

And we aim to find a ϵ bound on $\mathbb{P}$.
(○ **finalizing the bound)** in the context of □ it holds:

$$\begin{aligned}
\mathbb{P}\left[\bigcup_{i=1}^{n} |d_i - \overline{d}| \geq \delta\overline{d}\right] &\leq \sum_{i=1}^{n} \mathbb{P}\left[|d_i - \overline{d}| \geq \delta\overline{d}\right] && \text{Boole's Thm. 2.17} \\
&\leq \sum_{i=1}^{n} 2\exp\left\{-c\frac{\overline{d}\delta^2}{2}\right\} && \text{Chernoff's Thm. 7.25} \\
&= 2n\exp\left\{-c\frac{\overline{d}\delta^2}{2}\right\} \\
&\leq 2n\exp\left\{-c\frac{\delta^2}{2}C\log[n]\right\} && \text{Dense } \mathcal{G} \text{ s.t. } \exists C : \overline{d} \leq C\log[n] \\
&= 2n\frac{1}{n^{cC\frac{\delta^2}{2}}} \\
&= 2n^{1-cC\frac{\delta^2}{2}}
\end{aligned}$$

Now, setting this $< \epsilon \quad \forall\epsilon > 0$, we note that such threshold is $\epsilon = \delta\overline{d}$ and we want it to be controlled by C which is our coefficient of interest (we want to prove its existance):

$$2n^{1-cC\frac{\delta^2}{2}} < \epsilon \iff C > \frac{1}{c\delta^2} + \log\left[\frac{2}{\epsilon}\right]\frac{1}{c\delta^2\log[n]}$$

Existance of a C means that the graph is almost regular.
Summarizing we proved existance of C with constant c from Chernoff's bound and fixed ϵ, δ. □

Chapter Summary

Objects:

- uniform integrability for an arbitrary collection $\mathcal{C} = \{X_i, i \in I\}$

$$\forall \epsilon > 0 \quad \exists K > 0 \lim_{k \to \infty} \sup_{X \in \mathcal{C}} \left\{ \mathbb{E}\left[|X| \mathbb{1}_{(K,\infty)} (|X|) \right] \right\} = 0$$

- Graph Theory:
  - random graphs, degree of a node and average degree of a graph
  - dense random graphs

$$\mathcal{G}(n,p) \quad : \quad \overline{d} \geq C \log(n) \quad C \in \mathbb{R}$$

  - almost regular random graphs

$$\mathcal{G}(n,p) \quad : \quad \mathbb{P}\left[\bigcap_{i=1}^{n} \{ |d_i - \overline{d}| > \epsilon \} \right] \leq 1 - \epsilon \quad \forall \epsilon > 0$$

Results:

- integrability characterization:

$$X \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}) \iff \lim_{k \to \infty} \mathbb{E}\left[|X| \mathbb{1}_{(k,\infty)}(|X|) \right] = 0$$

- Markov's inequality

$$f \text{ } increasing, \text{ } f(b) \neq 0 \implies \mathbb{P}[X > b] \leq \frac{1}{f(b)} \mathbb{E}[f(X)]$$

  trick: we almost always start from this inequality with the increasing function $e^{\lambda x}$ for some $\lambda > 0$
- Chebyshev's inequality
- Jensen's inequality

$$f \text{ } convex, \text{ } X \in \mathcal{L}_1 \implies \mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$$

- Hoeffding's inequality

$$X_i \overset{iid}{\sim} \mathcal{B}ern_{\pm 1}\left(\frac{1}{2} \right) \implies \forall t \geq 0 \quad \mathbb{P}\left[\frac{1}{\sqrt{n}} \sum X_i \geq t \right] \leq e^{-\frac{t^2}{2}}$$

  trick: Markov's + assumptions + hyperbolic trigonometry + Taylor's + find the best λ.
- Hoeffding's two sided inequality
  trick: triangle inequality + symmetry of Bernoulli
- symmetrization argument for bounded random variables
  trick: Jensen's on an independent copy inside the expectation, attach a Bernoulli, optimize for λ
- Hoeffding's general inqequality for bounded random variables
  trick: just apply previous facts
- Hoeffding's inequality improved via Hoeffding's Lemma
  trick: just apply the Lemma
- Chernoff's bound
- two sided Chernoff's bound
- dense random graphs are almost regular

Chapter 8

Independence & Convolutions

8.1 Sigma algebra approach to independence

♠ **Definition 8.1** (σ-algebra generated by a random variable)**.** *Let X be a r.v. (Def. 3.2) on $(\Omega, \mathcal{F}, \mathbb{P})$ taking values on the measurable space $(E, \mathcal{E})$. The idea of a generated σ-algebra (Def. 1.11) is that of the smallest σ-algebra containing the generator. We extend this to r.v.s as follows:*

- *consider a collection of σ-algebras $\{\mathcal{G}_i,\ i \in I\}$ such that $\mathcal{G}_i \subset \mathcal{F}\ \forall i$*
- *these include all the possible σ-algebras that satisfy the measurability condition of Eqn. 3.1:*

$$\forall i,\ \forall A \in \mathcal{E} \quad X^{-1}(A) \in \mathcal{G}_i$$

Then we naturaly define the σ algebra generated by X as:

$$\sigma(X) = \bigcap_{i \in I} \mathcal{G}_i$$

Where the arbitrary intersection of σ-algebras is a σ-algebra by Theorem 1.10.

♣ **Theorem 8.2** ($\sigma(X)$ characterization)**.** *We identify $\sigma(X)$ as the set of counterimages of X in $\mathcal{E}$:*

$$\sigma(X) = \{X^{-1}(A),\ A \in \mathcal{E}\}$$

Proof. Let $\mathcal{G} = \{X^{-1}(A) : A \in \mathcal{E}\}$ which is a σ-algebra. We prove $\mathcal{G} \subset \sigma(X)$ and $\mathcal{G} \supset \sigma(X)$.
($\subset$) If $\mathcal{F}'$ is a σ-algebra making X measurable it then $X^{-1}(A) \in \mathcal{F}' \forall A \in \mathcal{E}$. Clearly then $\mathcal{G} \subset \mathcal{F}' \implies \mathcal{G} \subset \sigma(X)$.
($\supset$) $\sigma(X)$ is necessarily included in the counterimages $\mathcal{G}$ by the definition of $\sigma(X)$. □

♢ **Observation 8.3** (Intuition for $\sigma(X)$ of Def. 8.1)**.** *We can intepret $\mathcal{F}$ as full information on events, and $\sigma(X) \subset \mathcal{F}$ as the information learnt from X.*

It may be useful to compare the following examples with what we derived in Chapter 3 when constructing random variables.

♡ **Example 8.4** (Intuitive $\sigma(X)$)**.** *We provide two examples.*
(trivial σ-algebra) *Let $X(\omega) = k \forall \omega \in \Omega$, this is a non informative r.v. $X : \Omega \to \mathbb{R}$. Indeed:*

$$\forall B \in \mathcal{B}(\mathbb{R}) \quad X^{-1}(B) = \begin{cases} \emptyset & k \notin B \\ \Omega & k \in B \end{cases} \implies \sigma(X) = \{X^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\} = \{\Omega, \emptyset\}$$

(simple σ-algebra) *Let $c_1, c_2 \in \mathbb{R}, A \in \mathcal{F}$ and $X = \mathbb{1}_A c_1 + \mathbb{1}_{A^c} c_2$. Then, $\forall B \in \mathcal{B}(\mathbb{R})$ we have:*

$$X^{-1}(B) = \begin{cases} \emptyset & c_1, c_2 \notin B \\ \Omega & c_1, c_2 \in B \\ A & c_1 \in B, c_2 \notin B \\ A^c & c_1 \notin B, c_2 \in B \end{cases} \implies \sigma(X) = \{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\} = \{\emptyset, \Omega, A, A^c\}$$

♡ **Example 8.5** (More advanced $\sigma(X)$)**.** *Let $X(\Omega) = \{x_1, x_2, \ldots\}$ be a countable collection of discete realizations $x_n \in \mathbb{R} \forall n$. Then, the sets $\{A_n\} = \{X^{-1}(x_n)\}$ form a partition of Ω into $\mathcal{F}$-sets. It follows:*

$$\forall B \in \mathcal{B}(\mathbb{R}) \quad X^{-1}(B) = \bigcup_{n: x_n \in B} A_n \quad \{n : x_n \in B\} \neq \emptyset \implies X^{-1}(B) \neq \emptyset$$

And the σ-algebra generated is:

$$\sigma(X) = \{X^{-1}(B), B \in \mathcal{B}(\mathbb{R})\} = \left\{ \bigcup_{n: x_n \in B} A_i \; \forall B \in \mathcal{B}(\mathbb{R}) \right\}$$

Lemma 8.6 (σ-algebra by partitions)**.** *Let $\mathcal{C} = \{C_i, \; i \in I\}$ be a countable partition of Ω into $\mathcal{F}$-sets. By Definition 1.11 and Example 8.5 we can think of its generated σ-algebra as collections of unions of sets C_i. The claim is that this holds if and only if X takes a single value in each of the partitioning sets C_i:*

$$X \; : \; \sigma(X) = \sigma(\mathcal{C}) \underset{Ex.8.5}{\implies} \forall B \in \mathcal{B}(\mathbb{R}) \; \exists I_B \subset I \mid X^{-1}(B) = \bigcup_{i \in I_B} C_i \iff X(\omega) = x_i \forall \omega \in C_i, \; \forall i$$

Proof. **($\implies$ by contradiction)** suppose $\exists C \in \mathcal{C} : \omega_1 \neq \omega_2 \in C, \; c_1 = X(\omega_1) \neq X(\omega_2) = c_2$. Then:

$$A_1 = X^{-1}(\{c_1\}) \in \sigma(X) = \sigma(\mathcal{C}) \implies A_1 = \bigcup_{i \in I_{A_1}} C_i \implies A_1 \supset C$$

So that A_1 is single valued $A_1 = X^{-1}(\{c_1\})$ but covers the whole C. We have found a contradiction.
($\impliedby$) let X be such that $\forall C_i \in \mathcal{C}, i \in I$ the maps $X(C_i) = \{x_i\}$ are disjoint atoms. We prove both directions of inclusion:

- **($\subset$)** $\forall B \in \mathcal{B}(\mathbb{R}) \quad X^{-1}(B) = \bigcup_{i: x_i \in B} C_i \in \sigma(\mathcal{C}) \implies \sigma(X) \subset \sigma(\mathcal{C})$
- **($\supset$)** $X^{-1}(\{x_i\}) = c_i \forall i \implies \mathcal{C} \subset \sigma(X) \implies \sigma(\mathcal{C}) \subset \sigma(X)$
  Where the first implication is by $\mathcal{C}$ being a part of the options in $\sigma(X)$ and the second is by trivial implication.

□

♣ **Theorem 8.7** (Characterizing r.v. measurability by $\sigma(X)$)**.** *Let X, Y be r.v.s on $(\Omega, \mathcal{F}, \mathbb{P})$ taking values on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then:*

$$Y \text{ measurable w.r.t. } \sigma(X) \text{ i.e. } \forall A \in \mathcal{E} \; Y^{-1}(A) \in \sigma(X) \iff Y = f(X) \; f : \mathbb{R} \to \mathbb{R} \text{ deterministic}$$

Proof. [Çin11](Thm. II.4.4). See Proposition 12.40 for an idea. □

◇ **Observation 8.8** (Statistical models & Theorem 8.7)**.** *Classical assumptions for regression aim at applying Theorem 8.7. In other words, while it is not possible to check measurability over the whole $\sigma(X)$, with $Y = f(X) + \epsilon$ we assume measurability up to a ϵ relaxation.*

♠ **Definition 8.9** (Independence of σ-algebras)**.** *For $n \geq 1$ a finite collection $\{\mathcal{G}_1, \ldots, \mathcal{G}_n\}$ where $\mathcal{G}_k \subset \mathcal{F} \; \forall k$, is an independency (a collection of independent objects) when:*

$$\forall A_k \in \mathcal{G}_k, \; \forall i \quad \mathbb{P}\left[\bigcap_{k=1}^{n} A_k\right] = \prod_{k=1}^{n} \mathbb{P}[A_k]$$

For an uncountable collection $\{\mathcal{G}_t, \; t \in \mathbb{T}\}$ such that $\mathcal{G}_t \subset \mathcal{F} \; \forall t$ we say it is an independency if $\forall n \geq 1$ and distinct times $t_1 \neq \ldots \neq t_n$ the finite subcollection $\{\mathcal{G}_{t_i}\}_{i=1}^{n}$ is an indepencency in the above sense. Namely:

$$\forall A_{t_i} \in \mathcal{G}_{t_i}, \; \forall i \quad \mathbb{P}\left[\bigcap_{i=1}^{n} A_{t_i}\right] = \prod_{i=1}^{n} \mathbb{P}[A_{t_i}]$$

Lemma 8.10 (Independence characterizations by p-systems)**.** *Let $\mathcal{G}_1, \mathcal{G}_2 \subset \mathcal{F}$, consider their generating p-systems (Def. 1.8) $\mathcal{C}_1, \mathcal{C}_2 \; : \; \sigma(\mathcal{C}_1) = \mathcal{G}_1$ and $\sigma(\mathcal{C}_2) = \mathcal{G}_2$. Then:*

$$\mathcal{G}_1 \perp\!\!\!\perp \mathcal{G}_2 \iff \mathcal{C}_1 \perp\!\!\!\perp \mathcal{C}_2$$
$$\text{i.e.} \quad \mathbb{P}[I \cap H] = \mathbb{P}[I]\mathbb{P}[H] \; \forall I \in \mathcal{C}_1, H \in \mathcal{C}_2$$

Proof. Use Proposition A.29. □

♠ **Definition 8.11** (Independence of random variables revisited). *A collection $\{X_i\}_{i\in I}$ is an independence if $\{\sigma(X_i)\}_{i\in I}$ is an independence in the sense of Definition 8.9. The next proposition recovers this result.*

♣ **Proposition 8.12** (Independence is factorization, old is new). *For a finite collection:*

$$\{\sigma(X_i)\}_{i=1}^n \text{ } independencency \iff \mathbb{P}[X_i \le x_i\ \forall i] = \prod_{i=1}^n \mathbb{P}[X_i \le x_i] \quad \forall x_i \in \mathbb{R} \forall i$$

Proof. **($\iff$ together)** Let $\{X_i\}_{i=1}^n$ be independent and measurable. Then:

$$\begin{aligned}
&\forall A_1 \in \sigma(X_1), \dots, \forall A_n \in \sigma(X_n) \quad \mathbb{P}\left[\bigcap_{i=1}^n A_i\right] = \prod_{i=1}^n \mathbb{P}[A_i] \\
\iff & A_i = X_i^{-1}(B_i) \ : \ B_i \in \mathcal{B}(\mathbb{R})\ \forall i && \text{Thm. 8.2} \\
\iff & \mathbb{P}[X_1 \in B_1 \cap \dots \cap X_n \in B_n] = \prod_{i=1}^n \mathbb{P}[X_i \in B_i] \\
\iff & \mathcal{P}_{X_1,\dots,X_n} = \bigtimes_{i=1}^n \mathcal{P}_{X_i} \\
\iff & \mathbb{P}[X_1 \le x_1, \dots, X_n \le x_n] = \prod_{i=1}^n \mathbb{P}[X_i \le x_i] \quad \forall x_1, \dots, x_n \in \mathbb{R} && \text{Lem. 8.10}
\end{aligned}$$

□

♣ **Proposition 8.13** (Measurable functions independence). *Consider a finite collection of independent r.v.s $X_i \perp\!\!\!\perp X_j\ \forall i \ne j, i = 1, \dots, n$ and correspondent measurable functions f_i where $f_i(X_i) = Y_i \forall i$. Then:*

$$Y_i \perp\!\!\!\perp Y_j \quad \forall i \ne j,\ i = 1, \dots, n$$

Proof. Apply the approach of Proposition 8.12. □

♡ **Example 8.14** (Some independence structures). *We provide two examples.*

- *Let $(A_n) \subset \mathcal{F}, \mathcal{G}_n = \{\emptyset, A_n, A_n^c, \Omega\} \forall n$ so that we could say $\mathcal{G}_n = \sigma(\mathbb{1}_{A_n}) \forall n$. To check independence of the sequence of events it is sufficient to check that all the indicators are since:*

$$(A_n)_n \quad independency \overset{Thm.\ 8.12}{\iff} (\mathcal{G}_n)_n \quad independency \overset{Lem.\ 8.10}{\iff} (\mathbb{1}_{A_n})_n \quad independency$$

- *to establish that $X \perp\!\!\!\perp (Y_t)_{t\in\mathbb{T}}$ a stochastic process on an arbitrary index set $\mathbb{T}$ it suffices to check:*

$$\sigma(X) \perp\!\!\!\perp \sigma\left(\bigcup_{i=1}^n \sigma(Y_i)\right) \quad \forall \{Y_i\}_{i=1}^n \subset (Y_t)_{t\in\mathbb{T}},\ \forall n$$

  Which is the second case of Definition 8.9.

8.2 Convolutions and Radon-Nykodym again

♠ **Definition 8.15** (Convolution $\mathcal{P}^\star(\cdot), \star$). *Given two **independent** r.v.s X, Y taking values in a general Euclidean Borel space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ we define their convolution as:*

$$\begin{aligned}
\mathcal{P}^\star(H) &= (\mathcal{P}_X \star \mathcal{P}_Y)(H) && \forall H \in \mathcal{B}(\mathbb{R}^d) \\
&= \int_{\mathbb{R}^d} \mathcal{P}_Y[H - x]\mathcal{P}_X(dx) && H - x := \{y \mid x + y \in H\} = \{y \mid \exists h \in H \ : \ y = h - x\}
\end{aligned}$$

♣ **Proposition 8.16** (Convolution is independent sum in $\mathbb{P}$ space)**.** *An intepretation of convolution is:*

$$\mathcal{P}^*[H] = \mathbb{P}[X+Y \in H] = \mathcal{P}_{X+Y}[H] \quad \forall H \in \mathcal{B}(\mathbb{R}^d)$$

Proof. By simply carrying out computations one gets:

$$\begin{aligned}
\mathcal{P}_{X+Y}(H) &= \int_H d\mathcal{P}_{X+Y} \\
&= \int_B \mathcal{P}_{X,Y}(dx,dy) && \text{simplex } B = \{(x,y) : x+y \in H\} \\
&= \int_{\mathbb{R}^d} \int_{\{y: x+y \in H\}} \mathcal{P}_Y(dy)\mathcal{P}_X(dx) && \text{independence} \\
&= \int_{\mathbb{R}^d} \mathcal{P}_Y(H-x)\mathcal{P}_X(dx) && H = \{(x,y) : x+y \in H\} \\
&= \mathcal{P}^*_{X,Y}(H) && \forall H \in \mathcal{B}(\mathbb{R}^d)
\end{aligned}$$

□

♣ **Proposition 8.17** (Symmetry of convolution)**.** *The $\star$ operation is symmetric for independent r.v.s:*

$$(\mathcal{P}_X \star \mathcal{P}_Y)(H) = (\mathcal{P}_Y \star \mathcal{P}_X)(H) \quad \forall H \in \mathcal{B}(\mathbb{R}^d)$$

Proof. Again, by independence:

$$\begin{aligned}
(\mathcal{P}_X \star \mathcal{P}_Y)(H) &= \int_B \mathcal{P}_{X,Y}(dx,dy) && B = \{(x,y) : x+y \in H\} \\
&= \int_{\mathbb{R}^d} \int_{\{x: x+y \in H\}} \mathcal{P}_X(dx)\mathcal{P}_Y(dy) \\
&= \int_{\mathbb{R}^d} \mathcal{P}_X(H-y)\mathcal{P}_Y(dx) \\
&= (\mathcal{P}_Y \star \mathcal{P}_X)(H) && \forall H \in \mathcal{B}(\mathbb{R}^d)
\end{aligned}$$

□

Corollary 8.18 (Implications of Radon Nikodym Theorem and F, P identification)**.** *Using Theorems 3.21, 3.22, Radon Nikodym (Thm. 5.7), Proposition 5.20, symmetry of convolution (Prop. 8.17) and Fubini's Theorem B.30 we reach the following conclusions:*

1. *Radon Nikodym absolute continuous convolution representation*

$$\mathcal{P}_X \ll \lambda,\ \mathcal{P}_Y \ll \mu,\ p = \frac{d\mathcal{P}_X}{d\lambda},\ q = \frac{\mathcal{P}_Y}{\mu} \implies \mathcal{P}_{X+Y}(H) = \int_{\mathbb{R}} p(x) \int_H q(y-x)\mu(dy)\lambda(dx)$$

2. *Sum cdf representation as a convolution in trivial euclidean space*

$$d = 1,\ H = (-\infty, z] \implies F_{X+Y}(z) = \mathbb{P}[X+Y \le z] = \int_{\mathbb{R}} F_Y(z-x)\mathcal{P}_X(dx)$$

3. *Claim* 1 *in the Lebesgue case*

$$d = 1,\ \mathcal{P}_X \ll Leb,\ \mathcal{P}_Y \ll Leb \implies \mathcal{P}^* \ll Leb, \quad p^*(z) = \int_{\mathbb{R}} q(z-x)p(x)dx$$

where we could also state the same result for measures absolutely continuous to the counting measure

Proof. **(Claim #1)** we make an instrumental application of Radon Nikodym Theorem, granted by the assumption of absolute continuity. Indeed, we have a representation of the convolution as:

$$\begin{aligned}
\mathcal{P}_{X+Y}(H) = (\mathcal{P}_X \star \mathcal{P}_Y)(H) &= \int_{\mathbb{R}^d} \mathcal{P}_Y(H-x)\mathcal{P}_X(dx) \\
&= \int_{\mathbb{R}^d} \int_H \mathcal{P}_Y(d(y-x))\mathcal{P}_X(dx) = \int_{\mathbb{R}^d} p(x) \int_H q(y-x)\mu(dy)\lambda(dx)
\end{aligned}$$

(Claim #2)

$$\begin{aligned}
F_{X+Y}(z) = \mathcal{P}_{X+Y}((-\infty, z]) &= \int_{\mathbb{R}} p(x) \int_{-\infty}^{z} q(y-x)\mu(dy)\lambda(dx) \\
&= \int_{\mathbb{R}} p(x) \int_{-\infty}^{z-x} q(y)\mu(dy)\lambda(dx) && \text{ch. variable} \\
&= \int_{\mathbb{R}} p(x) F_Y(z-x)\lambda(dx) \\
&= \int_{\mathbb{R}} F_Y(z-x)\mathcal{P}_X(dx) && \text{Rad. Nyk. representation}
\end{aligned}$$

(Claim #3) we inspect the cdf of the sum to get that:

$$\begin{aligned}
F_{X+Y}(z) &= \int_{\mathbb{R}} \int_{-\infty}^{z-x} q(y)dy p(x)dx && \ll \textit{Leb} \text{ hypothesis and previous Claims} \\
&= \int_{\mathbb{R}} \int_{-\infty}^{z} \underbrace{q(y-z)}_{\geq 0} \underbrace{p(x)}_{\geq 0} dydx && \text{ch. var.} \\
&= \int_{-\infty}^{z} \int_{\mathbb{R}} q(y-z)p(x)dxdy && \text{Fubini Thm. B.30}
\end{aligned}$$

So that $\mathcal{P}_{X+Y} \ll Leb$ clearly by being an integral wrt $dy = Leb$ and we can eventually conclude that the density of the convolution is:

$$p^*(z) = \frac{\mathcal{P}_{X+Y}}{dLeb}(z) = F'_{X+Y}(z) = \int_{\mathbb{R}} q(z-x)p(x)dx$$

almost everywhere, by Proposition 5.20#1,#2,#3. □

♢ **Observation 8.19** (About Corollary 8.18)**.** *In Claim* 3 *we conclude that the sum of absolutely continuous independent random variable distribution is absolutely continuous and has a precise density.*

♡ **Example 8.20** (Triangular distribution sum of uniforms)**.** *Let* $X, Y \overset{ind}{\sim} \mathcal{U}nif(0,1)$. *Clearly* $\mathcal{P}_X \ll Leb, \mathcal{P}_Y \ll Leb$ *and by Radon Nikodym Thm. 5.7 the densities take form:*

$$p(x) = \mathbb{1}_{[0,1]}(x) \quad q(y) = \mathbb{1}_{[0,1]}(y)$$

By the just proved Corollary 8.18#3 we can further say that:

$$\begin{aligned}
p^*(z) = \int_{\mathbb{R}} q(z-x)p(x)dx &= \int_{\mathbb{R}} \mathbb{1}_{[0,1]}(z-x)\, \mathbb{1}_{[0,1]}(x)\, dx \\
&= \int_{\mathbb{R}} \mathbb{1}_{[z-1,z]}(x)\, \mathbb{1}_{[0,1]}(x) && 0 \leq z-x \leq 1 \iff x \geq z-1, x \leq z \\
&= \int_0^1 \mathbb{1}_{[z-1,z]}(x)
\end{aligned}$$

Four cases can be recognized:

- $z < 0 \implies p^*(z) = 0$
- $z > 2 \implies p^*(z) = 0$
- $z \in [0,1] \implies p^*(z) = \int_0^z dx = z$
- $z \in [1,2] \implies p^*(z) = \int_{z-1}^1 dx = 2-z$

The density has a representation:

$$\begin{aligned}
p^*(z) &= \mathbb{1}_{(-\infty,0]}(z) \cdot 0 + \mathbb{1}_{[0,1]}(z) \cdot z + \mathbb{1}_{[1,2]}(z) \cdot (2-z) + \mathbb{1}_{(2,\infty)}(z) \cdot 0 \\
&= \mathbb{1}_{[0,1]}(z) \cdot z + \mathbb{1}_{[1,2]}(z) \cdot (2-z)
\end{aligned}$$

Which is the shape of a triangular distribution centered at 1.

♡ **Example 8.21** (Poisson distribution convoluted)**.** *Let $X \perp\!\!\!\perp Y, X \sim \mathcal{P}o(\lambda), Y \sim \mathcal{P}o(\mu)$. Their sum is trivially absolutely continous wrt the counting measure, namely $X + Y \ll \nu = \sum_{j\geq 0} \delta_j$. The density of the convolution (equivalently, sum) is the probability law at a singleton[1]:*

$$p^*(z) = \mathcal{P}_{X+Y}(\{z\}) = \int_{\mathbb{R}} p(x)q(z-x)\nu(dx) \quad Z = X + Y \in \mathbb{N}$$

We keep an indicator for natural numbers that restricts the density to its domain in the background. This value is regarded as $\mathbb{1}_{\{0,1,\ldots\}}(z)$.
The density becomes:

$$\begin{aligned}
p^*(z) &= \mathbb{1}_{\{0,1,\ldots\}}(z) \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} \frac{\mu^{z-x} e^{-\mu}}{(z-x)!} \mathbb{1}_{\{0,1,\ldots\}}(z-x) \\
&= \mathbb{1}_{\{0,1,\ldots\}}(z) \sum_{x=0}^{\infty} \mathbb{1}_{[0,z]}(x) \frac{\lambda^x \mu^{z-x} e^{-(\mu+\lambda)}}{x!(z-x)!} && \mathbb{1}_{\{0,1,\ldots\}}(z-x) = \mathbb{1}_{[0,z]}(x) \\
&= \mathbb{1}_{\{0,1,\ldots\}}(z) \frac{e^{-(\mu+\lambda)}}{z!} \sum_{x=0}^{z} \frac{\lambda^x \mu^{z-x} z!}{x!(z-x)!} \\
&= \mathbb{1}_{\{0,1,\ldots\}}(z) \frac{e^{-(\mu+\lambda)}}{z!} \sum_{x=0}^{z} \underbrace{\frac{z!}{x!(z-x)!}}_{=\binom{z}{x}} \lambda^x \mu^{z-x} && \sum_{x=0}^{z} \binom{z}{x} \lambda^x \mu^{z-x} = (\lambda+\mu)^z
\end{aligned}$$

Which implies that $X + Y \sim \mathcal{P}p(\lambda + \mu)$.

♡ **Example 8.22** (Gamma convolutions)**.** *Let $X \perp\!\!\!\perp Y, X \sim \mathcal{G}amma(\alpha, \gamma), Y \sim \mathcal{G}amma(\beta, \gamma)$. Now $X+Y \ll Leb$ and the density of the convolution takes form:*

$$\begin{aligned}
p^*(z) &= \int_{\mathbb{R}} p(x)q(z-x)dx \\
&= \int_{\mathbb{R}} \frac{\gamma^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\gamma x} \mathbb{1}_{(0,\infty)}(x) \frac{\gamma^\beta}{\Gamma(\beta)} (z-x)^{\beta} e^{-\gamma(z-x)} \mathbb{1}_{(0,\infty)}(z-x)\, dx \\
&= \frac{\gamma^\alpha}{\Gamma(\alpha)} \frac{\gamma^\beta}{\Gamma(\beta)} \int_{\mathbb{R}} x^{\alpha-1} (z-x)^{\beta-1} e^{-\gamma(z-x)} e^{-\gamma x} \mathbb{1}_{(0,\infty)}(x)\, \mathbb{1}_{(0,\infty)}(z-x)\, dx \\
&= \frac{\gamma^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \int_{\mathbb{R}} x^{\alpha-1}(z-x)^{\beta-1} e^{-\gamma z} \mathbb{1}_{(0,z]}(x)\, dx \\
&= \frac{\gamma^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\gamma z} \int_0^z x^{\alpha-1}(z-x)^{\beta-1} dx && \text{let } w = \frac{x}{z}, dw = \frac{dx}{z} \\
&= \frac{\gamma^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\gamma z} \int_0^1 (wz)^{\alpha-1}(z(1-w))^{\beta-1} z dw \\
&= \frac{\gamma^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\gamma z} z^{\alpha+1-1+\beta-1} \int_0^1 \underbrace{(w)^{\alpha-1}(1-w)^{\beta-1}}_{\mathcal{B}eta\ kernel} dw \\
&= \frac{\gamma^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\gamma z} z^{\alpha+\beta-1} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \\
&= \frac{\gamma^{\alpha+\beta}}{\Gamma(\alpha+\beta)} e^{-\gamma z} z^{\alpha+\beta-1}
\end{aligned}$$

So that $X + Y \sim \mathcal{G}amma(\alpha + \beta, \gamma)$.

♡ **Example 8.23** (Bernoulli trials convolutions)**.** *Let $X \sim B(p), p \in [0,1]$. Below are some important information about Bernoulli r.v.s*

$$\begin{aligned}
p(x) = \mathcal{P}_X(\{x\}) = p^x(1-p)^{1-x} \mathbb{1}_{[0,1]}(x) \quad \mathbb{E}[X] = \mathbb{E}[X^2] = \cdots = p, \\
V[X] = p(1-p) \quad \widehat{\mathcal{P}}_X(t) = pe^{-t} + 1 - p, \quad \mathbb{E}[z^X] = 1 - p + pz
\end{aligned}$$

[1] discrete measures have non zero measure at singletons

We consider $X_1 \overset{iid}{\sim} B(p)$. The density of the convolution (sum) is:

$$\begin{aligned}
p^*(z) &= \mathcal{P}_{X+Y}(\{z\}) = \sum_x q(z-x)p(x) = \sum_{x=0}^{1} q(z-x)p^x(1-p)^{1-x} \\
&= q(z-1)p + q(z)(1-p) \qquad\qquad q(\cdot) \text{ is still a function!} \\
&= \underbrace{p^{z-1}(1-p)^{2-z}\mathbb{1}_{\{0,1\}}(z-1)}_{=q(z-1)}\, p + \underbrace{p^z(1-p)^{1-z}\mathbb{1}_{\{0,1\}}(z)}_{=q(z)}(1-p) \\
&= p^z(1-p)^{2-z}\mathbb{1}_{\{1,2\}}(z) + p^z(1-p)^{2-z}\mathbb{1}_{\{0,1\}}(z) \\
&= \binom{2}{z} p^z(1-p)^{2-z}\mathbb{1}_{\{0,1,2\}}(z)
\end{aligned}$$

Which implies that $Z \sim \mathcal{B}inom(2,p)$. Induction is straigthforward and we add that:

$$\{X_i\}_{i=1}^n \quad independency \implies S_n = \sum_{i=1}^n X_i \sim \mathcal{B}inom(n,p)$$

Chapter Summary

Objects:

- σ algebra generated by a random variable, the smallest σ-algebra containing all the measurable sets
- sigma algebras are independent if the laws of finite subcollections factorize
- convolution, $\mathcal{P}^*(H) = (\mathcal{P}_X \star \mathcal{P}_Y)(H) = \int_{\mathbb{R}^d} \mathcal{P}_Y(H-x)\mathcal{P}_X(dx)$ where $H - x = \{y : x + y \in H\}$

Results:

- $\sigma(X) = \{X^{-1}(A), A \in \mathcal{E}\}$ is the σ-algebra generated by the random variable X
- a random variable Y is measurable with respect to a σ-algebra $\sigma(X)$ if and only if it is a deterministic function of X
- random variables form an independency in the σ-algebra sense if and only if laws of random variables factorize
- $\mathcal{P}^*(H) = \mathcal{P}_{X+Y}(H) \quad \forall H \in \mathcal{B}(\mathbb{R}^d)$
- convolution is symmetric
- Radon Nikodym + convolution:
  - Radon Nikodym absolute continuous convolution representation

    $$\mathcal{P}_X \ll \lambda,\ \mathcal{P}_Y \ll \mu,\ p = \frac{d\mathcal{P}_X}{d\lambda},\ q = \frac{\mathcal{P}_Y}{\mu} \implies \mathcal{P}_{X+Y}(H) = \int_{\mathbb{R}} p(x) \int_H q(y-x)\mu(dy)\lambda(dx)$$

  - Sum cdf representation as a convolution in trivial euclidean space

    $$d = 1,\ H = (-\infty, z] \implies F_{X+Y}(z) = \mathbb{P}[X+Y \le z] = \int_{\mathbb{R}} F_Y(z-x)\mathcal{P}_X(dx)$$

  - Claim 1 in the Lebesgue case

    $$d = 1,\ \mathcal{P}_X \ll Leb,\ \mathcal{P}_Y \ll Leb \implies \mathcal{P}^* \ll Leb, \quad p^*(z) = \int_{\mathbb{R}} q(z-x)p(x)dx$$

    similar for the counting measure

Chapter 9

Borel Cantelli Lemmas & Convergence

9.1 Borel-Cantelli Lemmas

♢ **Observation 9.1** (Recap on Limits of sets). *Recall Observation 2.12:*

- $A_n \subset A_{n+1} \forall n \implies \lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$
- $A_n \supset A_{n+1} \forall n \implies \lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$

♠ **Definition 9.2** ($\limsup_n$ infinitely often *i.o.*). *For a sequence of sets, we establish the equivalence:*

$$\limsup_n A_n = True \iff \infty \; many \; A_n \; occurr \; (A_n \; i.o.)$$

Namely:

$$\limsup_n A_n = \bigcap_{N\geq 1} \bigcup_{n=N}^{\infty} A_n = \{A_n \; i.o.\}$$

Or $\forall N \; \exists n > N : A_n \; occurs$

♠ **Definition 9.3** ($\liminf_n$ eventually). *Similarly, for a sequence of sets:*

$$\liminf_n A_n = True \iff all \; A_n \; but \; finitely \; many \; occurr \quad (A_n \; eventually)$$

Namely:

$$\liminf_n A_n = \bigcup_{N\geq 1} \bigcap_{n=N}^{\infty} A_n$$

Or $\exists N \;:\; \forall n \geq N \; A_n \; occurs.$

♠ **Definition 9.4** (Conciliating the usual limit). *Clearly:*

1. $\limsup_n A_n \neq \liminf_n A_n \implies \nexists \lim_n A_n$
2. $\limsup_n A_n = \liminf_n A_n = A \implies \exists \lim_n A_n = A$

♣ **Proposition 9.5** (Facts about lim sup & lim inf). *Definitions 9.2 and 9.3 are nested and make sense in the σ-algebra construction*

1. $\exists \liminf_n A_n \implies \exists \limsup_n A_n$ *i.e.* $\limsup_n A_n \supset \liminf_n A_n$
2. *the events are well defined:*

$$(A_n)_{n\geq 1} \subset (\Omega, \mathcal{F}, \mathbb{P}) \implies \liminf_n A_n \in \mathcal{F} \; \limsup_n A_n \in \mathcal{F}$$

Proof. **(Claim #1)** trivial, lim inf is i.o. after N in Definition 9.3.
(Claim #2) holds by countable intersections and unions closedness in $\mathcal{F}$ (Def. 1.6, Lem. 1.7). □

♣ **Theorem 9.6** (First Borell Cantelli Lemma, BC1)**.** *Consider a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and a sequence of sets* $(A_n)_{n\geq 1} \in \mathcal{F}$*. Then:*

$$\sum_{n\geq 1} \mathbb{P}[A_n] < \infty \implies \mathbb{P}[\limsup_n A_n] = 0$$

Proof. $\forall N > 1$ let $G_N = \bigcup_{n=N}^{\infty} A_n \implies G_N \supset G_{N+1}$. By properties of sets then:

$$\exists G = \lim_{N\to\infty} G_N = \bigcap_{N\geq 1} G_N = \bigcap_{N\geq 1} \bigcup_{n=N}^{\infty} A_n = \limsup_n A_n$$

And further at finite N:

$$\mathbb{P}[G_N] = \mathbb{P}\left[\bigcup_{n=N}^{\infty} A_n\right] \leq \sum_{n=N}^{\infty} \mathbb{P}[A_n] \quad \text{Boole's Thm. 2.17}$$

Moving to the limit then:

$$\begin{aligned}
\mathbb{P}[\limsup_n A_n] &= \mathbb{P}\left[\lim_{N\to\infty} G_N\right] \\
&= \lim_{N\to\infty} \mathbb{P}[G_N] && \text{continuity Thm. 2.13} \\
&\leq \lim_{N\to\infty} \sum_{n=N}^{\infty} \mathbb{P}[A_n] \\
&= 0 && \sum_{n=1}^{\infty} \mathbb{P}[A_n] < \infty \iff \lim_{N\to\infty} \sum_{n=N}^{\infty} \mathbb{P}[A_n] = 0
\end{aligned}$$

□

Lemma 9.7 (BC1 equivalent statement)**.** *For Theorem 9.6 we could equivalently conclude:*

$$\mathbb{P}[\limsup_n A_n] = 0 \iff \mathbb{P}[\liminf A_n^c] = 1$$

Proof. It holds that $\mathbb{P}[\limsup_n A_n] = 0 \iff \mathbb{P}[(\limsup_n A_n)^c] = 1$ so:

$$\begin{aligned}
\mathbb{P}\left[\left(\bigcup_{N\geq 1} \bigcap_{n=N}^{\infty} A_n\right)^c\right] &= 1 && \limsup \text{ Def. } 9.2 \\
&= \mathbb{P}\left[\bigcup_{N\geq 1} \bigcap_{n=N}^{\infty} A_n^c\right] && \text{De Morgan's Laws} \\
&= \mathbb{P}\left[\liminf A_n^c\right] && \liminf \text{ Def. } 9.3
\end{aligned}$$

□

♣ **Theorem 9.8** (Second Borel Cantelli Lemma, BC2)**.** *Consider* $(\Omega, \mathcal{F}, \mathbb{P})$*,* $(A_n)_{n\geq 1} \in \mathcal{F}$*. Then:*

$$(A_n)_{n\geq 1} \underbrace{\textit{independent}}_{Def. 8.9} \quad \sum_{n\geq 1} \mathbb{P}[A_n] = \infty \implies \mathbb{P}[\limsup_n A_n] = 1$$

Proof. **(△ aim)** wts $\mathbb{P}[(\limsup_n A_n)^c] = 0$ which is equivalent.
(□ the inner intersection) Recall that by Definition of lim inf:

$$\liminf_n A_n^c = \bigcup_{N\geq 1} \bigcap_{n=N}^{\infty} A_n^c$$

We focus on the intersection:

$$\begin{aligned}
\mathbb{P}\left[\bigcap_{n=N}^{N+j} A_n^c\right] &= \prod_{n=N}^{N+j} \mathbb{P}[A_n^c] && \text{indepedence} \\
&= \prod_{n=N}^{N+j} 1 - \mathbb{P}[A_n] \\
&\leq \prod_{n=N}^{N+j} e^{-\mathbb{P}[A_n]} && 1 - x \leq e^{-x} \forall x \quad x = \mathbb{P}[A_n] \\
&= \exp\left\{-\sum_{n=N}^{N+j} \mathbb{P}[A_n]\right\}
\end{aligned}$$

Clearly then $\forall N \geq 1$ at the limit:

$$\begin{aligned}
\mathbb{P}\left[\bigcap_{n=N}^{\infty} A_n^c\right] &= \lim_{j\to\infty} \mathbb{P}\left[\bigcap_{n=N}^{N+j} A_n^c\right] && \text{continuity Thm. 2.13} \\
&\leq \lim_{j\to\infty} \exp\left\{\sum_{n=N}^{N+j} -\mathbb{P}[A_n]\right\} \\
&= 0 && \sum_{n=1}^{\infty} \mathbb{P}[A_n] = \infty
\end{aligned}$$

$$\implies 0 \leq \mathbb{P}\left[\bigcap_{n=N}^{\infty} A_n^c\right] \leq 0 \implies \mathbb{P}\left[\bigcap_{n=N}^{\infty} A_n^c\right] = 0$$

(◯ **final claim)** By the result of □ we conclude:

$$\begin{aligned}
0 \leq \mathbb{P}\left[\liminf_n A_n^c\right] & && \text{positivity of probability} \\
= \mathbb{P}\left[\bigcup_{N\geq 1} \bigcap_{n=N}^{\infty} A_n^c\right] & && \liminf \text{ Def. } 9.3 \\
\leq \sum_{N\geq 1} \mathbb{P}\left[\bigcap_{n=N}^{\infty} A_n^c\right] & && \text{Boole's Thm. 2.17} \\
= 0 & && \text{sum of zeroes}
\end{aligned}$$

Which, by squeezing, proves the objective in △. □

9.2 Convergence revisited

♡ **Example 9.9** (Almost sure convergence by BC1)**.** *Consider a sequence $(X_n)_{n\in\mathbb{N}}$ and a r.v. X where:*

$$\forall \epsilon > 0 \quad \sum_n \mathbb{P}\left[|X_n - X| \geq \epsilon\right] < \infty$$

Let $A_n = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\}$, by BC1 (Thm. 9.6) and Lemma 9.7 it holds that:

$$\mathbb{P}\left[\limsup_n A_n\right] = 0 \quad \mathbb{P}\left[\liminf_n A_n^c\right] = 1$$

The latter is for fixed $\epsilon > 0$. Letting the property for all $\epsilon > 0$ we have an equivalent expression:

$$\iff \mathbb{P}\left[\bigcup_{N\geq 1}\bigcap_{n=N}^{\infty}\underbrace{\{\omega\in\Omega : |X_n(\omega)-X(\omega)|<\epsilon\}}_{=A_n^c}\right] \qquad \forall\epsilon>0 \iff \forall\epsilon\in\mathbb{Q}_+,\ \textit{see below}$$

$$\mathbb{P}\left[\bigcap_{\epsilon\in\mathbb{Q}_+}\bigcup_{N\geq 1}\bigcap_{n=N}^{\infty}\{\omega\in\Omega : |X_n(\omega)-X(\omega)|<\epsilon\}\right]$$

$$=\mathbb{P}\left[\bigcap_{\epsilon>0}\bigcup_{N\geq 1}\bigcap_{n=N}^{\infty}\{\omega\in\Omega : |X_n(\omega)-X(\omega)|<\epsilon\}\right] \qquad \textit{rationals are dense in } \mathbb{R}_+ \textit{ Prop. 18.15}$$

Which means that we have proved an $\iff$ relation between the first and the third statement. Additionally, this is also equivalent to $X_n \overset{a.s.}{\to} X$, being in line with Definition 4.18.

♡ **Example 9.10** (Negative exponential random variables). *Let $(X_n)_{n\in\mathbb{N}}$ be iid and such that $\mathcal{P}_{X_n} \ll Leb$ with density:*

$$p(x) = \lambda e^{-\lambda x}\mathbb{1}_{\mathbb{R}_+}(x) \implies X_n \sim \mathcal{N}eg\mathcal{E}xp(\lambda)$$

in the usual sense.
(△ aim) *we want to find for a fixed $\alpha > 0$ the i.o. occurence of:*

$$\mathbb{P}\left[\limsup_n\left\{\frac{X_n}{\log n} > \alpha\right\}\right]$$

(□ solution) *define $A_n = \left\{\omega\in\Omega : \dfrac{X_n(\omega)}{\log n} > \alpha\right\}$ which are independent sets by construction. Then:*

$$\mathbb{P}[A_n] = \mathbb{P}[X_n > \alpha\log n] = \int_{\alpha\log n}^{\infty}\lambda e^{-\lambda x}dx = e^{-\alpha\lambda\log n} = \frac{1}{n^{\alpha\lambda}}$$

and sum of these events is parameter dependent

$$\sum_n \mathbb{P}[A_n] = \sum_n\frac{1}{n^{\alpha\lambda}} = \begin{cases} <\infty & \alpha\lambda>1 \\ \infty & \alpha\lambda\leq 1\end{cases}$$

So that in conclusion:

$$\mathbb{P}\left[\limsup_n\left\{\frac{x_n}{\log n} > \alpha\right\}\right] = \begin{cases} 0 & \alpha\lambda>1 \quad \textit{BC1 Thm. 9.6} \\ 1 & \alpha\lambda\leq 1 \quad \textit{BC2 Thm. 9.8}\end{cases}$$

♡ **Example 9.11** (Coin tossing). *Let $\mathbb{P}[H]\in(0,1) \implies \mathbb{P}[T] > 0$. Consider as space $\Omega = \{H,T\}^\infty$ and a realization $s\in\{H,T\}^k$ for some $k > 0$. Clearly $s = (s_1,\ldots,s_k)$.*
(△ aim) *define an event that checks for correspondence at arbitrary n:*

$$A_n = \{\omega\in\Omega : (\omega_n,\ldots,\omega_{n+k-1}) = s\}$$

We look for the probability that this event happens i.o.
(□ solution) *Notice that the sequence $(A_n)_{n\in\mathbb{N}}$ is not independent since the events overlap. Contrarily:*

$$\begin{aligned} B_1 &= \{\omega\in\Omega : (\omega_1,\ldots,\omega_k) = s\}, \\ B_2 &= \{\omega\in\Omega : (\omega_{k+1},\ldots,\omega_{k+k+1}) = s\}, \\ \ldots &= \ldots \\ B_j &= \{\omega\in\Omega : (\omega_{(j-1)k+1},\ldots,\omega_{jk}) = s\}\end{aligned}$$

Are non overlapping, independent and such that:

$$\{B_n \textit{ i.o.} \implies A_n \textit{ i.o.}\} \implies \limsup_n B_n \subset \limsup_n A_n \implies \mathbb{P}\left[\limsup_n B_n\right] \leq \mathbb{P}\left[\limsup_n A_n\right]$$

Where the last passage is by monotonicity (Thm. 2.10). Let:

$$p_{s_j} = \begin{cases} \mathbb{P}[H] & s_j = H \\ \mathbb{P}[T] & s_j = T \end{cases} \implies p_{s_j} > 0 \quad \forall s_j$$

Then:

$$\begin{aligned} \mathbb{P}[B_n] &= \prod_{j=1}^{k} p_{s_j} > 0 && \forall n, \forall k, \ Kost \ wrt \ n \\ \implies \sum_n \mathbb{P}[B_n] &= \sum_n kost = \infty \\ \implies \mathbb{P}[\limsup_n B_n] &= 1 && BC2 \ Thm. \ 9.8 \\ &\leq \mathbb{P}[\limsup_n A_n] = 1 && by \ above \ result \end{aligned}$$

and we have proved $\triangle$.

♢ **Observation 9.12** (Limit notions in $\mathbb{P}$ spaces)**.** *Recall that the usual definition of limits in metric spaces is:*

$$x_n \to x \ as \ n \to \infty \quad if \quad \forall \epsilon > 0 \ \exists n_0 = n_0(\epsilon) : \forall n > n_0 \quad |x_n - x| < \epsilon$$

In a $\mathbb{P}$ space this can have different meanings, we already mentioned one in Definition 4.18.

♠ **Definition 9.13** (Convergence in Probability $\overset{\mathcal{P}}{\to}$)**.** *For* $(X_n)_{n\geq 1}, X$ *on* $(\Omega, \mathcal{F}, \mathbb{P})$ *taking values on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *we say that* X_n *converges in Probability to* X *and write* $X_n \overset{\mathcal{P}}{\to} X$ *when:*

$$\forall \epsilon > 0 \ \lim_{n\to\infty} \mathbb{P}[|X_n - X| > \epsilon] = 0$$

While for a general space $(E, \mathcal{E})$ *we use the suitable* $d(X_n, X)$ *notion.*

♣ **Proposition 9.14** (Almost sure & probability convergence)**.** *Compare Definitions 4.18 and 9.13 for* $(X_n)_{n\geq 1}, X$*. Then:*

$$X_n \overset{a.s.}{\to} X \implies X_n \overset{\mathcal{P}}{\to} X$$

*And **not the opposite***

Proof. ($\implies$) recall that by the results of Example 9.9 and the definition of $\overset{a.s.}{\to}$ (Def. 4.18):

$$X_n \overset{a.s.}{\to} X \iff \mathbb{P}[|X_n - X| > \epsilon \ i.o.] = 0$$

Where we notice that by squeezing

$$\begin{aligned} 0 &= \mathbb{P}[|X_n - X| > \epsilon \ i.o.] \\ &= \mathbb{P}\left[\limsup_n \{|X_n - X| > \epsilon\}\right] && \limsup \text{ Def. } 9.2 \\ &\geq \limsup_n \mathbb{P}[|X_n - X| > \epsilon] && \text{Fatou's Lem. A.49} \\ &\geq \liminf_n \mathbb{P}[|X_n - X| > \epsilon] && \limsup \geq \liminf \\ &\geq 0 && \text{positivity} \end{aligned}$$

So that $\limsup = \liminf = \lim$ and they are all null. Clearly then:

$$\forall \epsilon > 0 \quad \lim_{n\to\infty} \mathbb{P}[|X_n - X| > \epsilon] = 0 \overset{\text{Def. } 9.13}{\iff} X_n \overset{\mathcal{P}}{\to} X$$

(opposite by counterexample) Let $\Omega = [0,1], \mathcal{F} = \mathcal{B}([0,1]), \mathbb{P} = Leb$. In such a probability space, define:

$$A_1 = [0,1], \ A_2 = \left[0, \frac{1}{2}\right], \ A_3 = \left[\frac{1}{2}, 1\right], A_4 = \left[0, \frac{1}{3}\right], \ A_5 = \left[\frac{1}{3}, \frac{2}{3}\right], \ldots$$

By construction $A_n \to \emptyset \implies Leb(A_n) \to 0$, yet $\forall \omega \in \Omega$ such $\omega \in \infty$-many A_n and ∞-many A_n^c. Then:

$$\mathbb{P}[\limsup_n A_n] = \mathbb{P}[\limsup_n A_n^c] = 1 = Leb([0,1])$$

And for $X_n = \mathbb{1}_{A_n}$ by $Leb(A_n) \to 0$ we have:

$$\mathbb{P}[|X_n| > \epsilon] = \mathbb{P}[A_n] = Leb(A_n) \to 0 \implies X_n \xrightarrow{\mathcal{P}} 0$$

but by the previous point $X_n = 0$ ∞-many times and 1 ∞-many times so:

$$\begin{aligned} &\mathbb{P}[X_n = 1 \; i.o.] = 1 && X_n(\omega) \to 1 \\ &\mathbb{P}[X_n = 0 \; i.o.] = 1 && X_n(\omega) \to 0 \end{aligned}$$

And we cannot have $X_n \xrightarrow{a.s.} 0$. □

♣ **Proposition 9.15** (Continuous functions and probability convergence)**.**

$$(X_n)_{n\geq 1}, X : X_n \xrightarrow{\mathcal{P}} X, \; f : E \to E \; continuous \implies f(X_n) \xrightarrow{\mathcal{P}} f(X)$$

Proof. Fix $\epsilon > 0$, and $\forall \delta > 0$ consider:

$$B_\delta = \{x \in E, x \notin D_f \mid \exists y \in E : |x - y| < \delta, \; |f(x) - f(y)| > \epsilon\}$$

Namely, points mapping outside ϵ in f and inside δ in the X r.v. realization. 2Here D_f are discontinuity points. By the continuity of f we have $B_\delta \to \emptyset$ as $\delta \to 0$.
Now let $|f(x) - f(x_n)| > \epsilon$. Then, either:

- $|X - X_n| \geq \delta$
- $x \in D_f$
- $x \in B_\delta$

So that:

$$\mathbb{P}[|f(X) - f(X_n)| > \epsilon] \leq \mathbb{P}[|X_n - X| > \delta] + \mathbb{P}[X \in B_\delta] + \mathbb{P}[X \in D_f] = 0$$

Since:

- $\mathbb{P}[|X_n - X| > \delta] \overset{n\to\infty}{\to} 0 \forall\delta$ by $X_n \xrightarrow{\mathcal{P}} X$
- $\mathbb{P}[X \in B_\delta] \overset{\delta\to 0}{\to} 0$ by $B_\delta \to \emptyset$
- $\mathbb{P}[X \in D_f] = 0$ by $D_f = \emptyset$ (f is continuous)

We conclude that:

$$\forall \epsilon > 0 \quad \lim_{n\to\infty} \mathbb{P}[|f(X) - f(X_n)| > \epsilon] = 0 \implies f(X_n) \xrightarrow{\mathcal{P}} f(X)$$

□

9.3 Inner product space and Orthogonal Projection Theorem

♠ **Definition 9.16** ($\mathcal{L}_p$ space and its pseudonorm)**.** *Extend Definition 4.5 to:*

$$\forall p \geq 1 \quad \mathcal{L}_p(\Omega, \mathcal{F}, \mathbb{P}) = \left\{ X : \Omega \to E : \mathbb{E}[|X|^p] = \int_E |x|^p \mathcal{P}_x(dx) < \infty \right\}$$

And endow it with a pseudonorm $\|\cdot\|_p : \mathcal{L}_p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}_+$ *such that:*

$$\|X\|_p = (\mathbb{E}[|X|^p])^{\frac{1}{p}} \quad where \quad \|X\|_p = 0 \iff \mathbb{P}[X = 0] = 1 \; in \; line \; with \; Ass.9.17$$

Assumption 9.17 (Quotiented norm space)**.** *The (pseudo)normed space* $(\mathcal{L}_p, \|\cdot\|_p)$ *is such that there is a quotient* $\mathbb{Y}$ *(thus the pseudo in front) which established the uniqueness property of the norm outside of negligible sets. Namely:*

$$\mathcal{L}_p(\Omega, \mathcal{F}, \mathbb{P}) \setminus \mathbb{Y} \; : \; \mathbb{Y} = \{X : \mathbb{P}[X = 0] = 1 \; i.e. \; X \overset{a.s.}{=} 0\}$$

So that the r.v.s are split into classes (collections) and:

$$\begin{aligned} X, Y \in \textit{ same class} &\iff \|X - Y\|_p = 0 \\ &\iff \mathbb{P}[X - Y = 0] = 1 \\ &\iff X \overset{a.s.}{=} Y \\ &\not\Rightarrow X(\omega) = Y(\omega) \forall \omega \in \Omega \end{aligned}$$

♠ **Definition 9.18** ($\mathcal{L}_p$ convergence $\overset{\mathcal{L}_p}{\to}$)**.** *Consider* $(X_n)_{n \geq 1}, X$ *where* $X \in \mathcal{L}_p(\Omega, \mathcal{F}, \mathbb{P})$*. Define:*

$$\begin{aligned} X_n \overset{\mathcal{L}_p}{\to} &\iff \forall n \geq 1 \quad \|X_n - X\|_p \to 0 \; as \; n \to \infty \\ &\iff \lim_{n \to \infty} \mathbb{E}\left[|X_n - X|^p\right] = 0 && \textit{Def. 9.16} \end{aligned}$$

Notice that $\mathcal{L}_p$ *convergence implies convergence of the means by the reverse triangle inequality. Indeed*

$$|\, \|X_n\|_{\mathcal{L}_p} - \|X\|_{\mathcal{L}_p} | \leq \|X_n - X\|_{\mathcal{L}_p} \overset{n \to \infty}{\to} 0 \implies \mathbb{E}\left[|X_n|^p\right] \overset{n \to \infty}{\to} \mathbb{E}\left[|X|^p\right]$$

We also specify that $\mathcal{L}_p$ *convergence implies that the limit is almost surely unique, by the quotiented space discussion of Assumption 9.17.*

♣ **Proposition 9.19** ($\mathcal{L}_p$ and probability convergence)**.**

$$X_n \overset{\mathcal{L}_p}{\to} X \implies X_n \overset{\mathcal{P}}{\to} X$$

And ***not the opposite***

Proof. ($\implies$) We trivially have that:

$$\begin{aligned} 0 \leq \mathbb{P}[|X_n - X| > \epsilon] & && \forall \epsilon > 0 \\ \leq \frac{\mathbb{E}[|X_n - X|^p]}{\epsilon^p} & && \text{Markov's Thm. 7.4, } g(x) = x^p, p \geq 1 \\ \overset{n \to \infty}{\to} 0 & && \text{as } n \to \infty \text{ by hypothesis } \mathcal{L}_p \text{ convergence} \end{aligned}$$
$$\implies \lim_{n \to \infty} \mathbb{P}[|X_n - X| > \epsilon] = 0 \forall \epsilon > 0 \implies X_n \overset{\mathcal{P}}{\to} X$$

(opposite by counterexample) let $\Omega = [0,1], \mathcal{F} = \mathcal{B}([0,1]), \mathbb{P} = Leb$. In this probability space define events:

$$E_n = \left[0, \frac{1}{n}\right], \quad X_n = n\mathbb{1}_{E_n} \implies Leb(E_n) = \frac{1}{n} \to 0$$

It holds

$$\mathbb{P}[|X_n - 0| > \epsilon] = \mathbb{P}[n\mathbb{1}_{E_n} > \epsilon] = \mathbb{P}[E_n] = Leb(E_n) \overset{n \to \infty}{\to} 0 \implies X_n \overset{\mathcal{P}}{\to} X = 0$$

Where the red equality follows from:

$$\begin{aligned}
\mathbb{P}[n\mathbb{1}_{E_n} > \epsilon] &= \mathbb{P}\left[\{\omega \in \Omega : n\mathbb{1}_{E_n}(\omega) > \epsilon\}\right] \\
&= \mathbb{P}\left[\{\omega \in E_n : n\,\underbrace{\mathbb{1}_{E_n}(\omega)}_{=1} > \epsilon\} \cup \{\omega \in E_n^c : n\,\underbrace{\mathbb{1}_{E_n}(\omega)}_{=0} > \epsilon\}\right] && \text{disjoint} \\
&= \mathbb{P}\left[\{\omega \in E_n : n\,\underbrace{\mathbb{1}_{E_n}(\omega)}_{=1} > \epsilon\}\right] + \underbrace{\mathbb{P}\left[\{\omega \in E_n^c : n\,\underbrace{\mathbb{1}_{E_n}(\omega)}_{=0} > \epsilon\}\right]}_{=0} && \text{since } \epsilon > 0 \\
&= \mathbb{P}\left[\{\omega \in E_n : n\,\underbrace{\mathbb{1}_{E_n}(\omega)}_{=1} > \epsilon\}\right] \\
&= \mathbb{P}\left[\{\omega \in E_n : n > \epsilon\}\right] \\
&= \mathbb{P}[n\mathbb{1}_{E_n} > \epsilon] \\
&= \mathbb{P}[E_n] && \epsilon > 0 \text{ arbitrary}
\end{aligned}$$

However, in norm:

$$\|X_n - 0\|_1 = \mathbb{E}[X_n] = n\mathbb{P}[E_n] = n\mathbb{E}[\mathbb{1}_{E_n}] = n\frac{1}{n} = 1 \neq 0 \implies X_n \text{ not} \stackrel{\mathcal{L}_p}{\to} 0$$

□

♢ **Observation 9.20** (The $\mathcal{L}_2$ special case). *If $p = 2$ square integrable r.v.s are of great interest since we can naturally identify an inner product. The converse is not true in general. The sufficient condition for the inner product existance to hold in $\mathcal{L}_2$ is that it satisfies the parallelogram law. We do not go much deeper into this matter, and just take it as granted.*

♠ **Definition 9.21** (Inner product on $\mathcal{L}_2$ quotiented space $\langle \cdot, \cdot \rangle$). *The inner product is a function $\langle \cdot, \cdot \rangle : \mathcal{L}_2 \times \mathcal{L}_2 \to \mathbb{R}$ defined as:*

$$\langle X, Y \rangle = \int_{E \times E} xy \mathcal{P}_{X,Y}(dxdy) = \mathbb{E}[XY]$$

♣ **Proposition 9.22** (Basic properties of $\langle \cdot, \cdot \rangle$). *Some trivial properties are reported below for $X, Y, Z \in \mathcal{L}_2$ and $c_1, c_2 \in \mathbb{R}$.*

1. *symmetry* $\langle X, Y \rangle = \langle Y, X \rangle$
2. *positivity and nullity in quotient space* $\langle X, X \rangle \geq 0$ *and* $\langle X, X \rangle = 0 \iff \mathbb{P}[X = 0] = 1$
3. *linearity* $\langle c_1 X + c_2 Y, Z \rangle = c_1 \langle X, Z \rangle + c_2 \langle Y, Z \rangle$

Proof. **(Claim #1)** trivial by Definition 9.21.
(Claim #2) we have:

$$\langle X, X \rangle = \int_E x^2 \mathcal{P}_X(dx) := \|X\|_2^2 \geq 0$$

Which is null $\iff \mathbb{P}[X = 0] = 1$ by Definition 9.18 and Assumption 9.17.
(Claim #3) this follows by linearity of the integral. Indeed:

$$\begin{aligned}
\langle c_1 X + c_2 Y, Z \rangle &= \int_{E \times E} (c_1 x + c_2 y) z \mathcal{P}_{X+Y,Z}(dx, dy, dz) \\
&= c_1 \int_{E \times E} xz \mathcal{P}_{X,Z}(dx, dz) + c_2 \int_{E \times E} yz \mathcal{P}_{Y,Z}(dy, dz) \\
&= c_1 \langle X, Z \rangle + c_2 \langle Y, Z \rangle
\end{aligned}$$

□

♠ **Definition 9.23** (Norm induced by inner product). *The inner product and the norm in $\mathcal{L}_2$ follow the relation:*

$$\sqrt{\langle X, X \rangle} = \|X\|_2 = \sqrt{\mathbb{E}[X^2]}$$

♣ **Proposition 9.24** (Joint properties of $\|\cdot\|_2$ and $\langle\cdot,\cdot\rangle$)**.** *Definitions 9.21, 9.23 allow to obtain nice results* $\forall X, Y \in \mathcal{L}_2$*:*

1. *Cauchy-Schwarz inequality* $\langle X, Y\rangle \le \|X\|_2 \|Y\|_2$
2. *triangle inequality* $\|X+Y\|_2 \le \|X\|_2 + \|Y\|_2$
3. *distance* $d(X,Y) = \|X-Y\|_2 = \sqrt{\langle X-Y, X-Y\rangle}$
4. *variance covariance relation:*

$$\mathbb{E}[X] = \mathbb{E}[Y] = 0 \implies \begin{cases} \langle X, Y\rangle = CoV(X,Y) \\ \|X\|_2 = \sqrt{V[X]} \\ \|Y\|_2 = \sqrt{V[Y]} \end{cases}$$

5. *angle (correlation) Pythagora's Theorem*

$$\begin{cases} \mathbb{E}[X] = \mathbb{E}[Y] = 0 \\ \|X\|_2 \neq 0 \\ \|Y\|_2 \neq 0 \end{cases} \implies \cos(\theta) = \frac{\langle X, Y\rangle}{\|X\|_2 \|Y\|_2} = corr(X,Y)$$

Proof. **(Claims #1, #2)** easy, there are many proofs online. Consider the inner product:

$$\langle X - tY, X - tY\rangle = \|X - tY\|_2^2 \ge 0 \quad \forall t \ge 0$$

positive by Prop. 9.22#2. Expanding the expression by linearity (Prop. 9.22#3) we get:

$$\begin{aligned} &\langle X - tY, X - tY\rangle = \langle X, X\rangle - 2t\langle X, Y\rangle + t^2\langle Y, Y\rangle \ge 0 && \forall t \\ \iff\ & 4\langle X, Y\rangle - 4\langle X, X\rangle\langle Y, Y\rangle \le 0 && \text{positivity of parabula } \forall t \text{ is negative delta} \\ \iff\ & \langle X, Y\rangle - \langle X, X\rangle\langle Y, Y\rangle \le 0 \\ \iff\ & \langle X, Y\rangle \le \|X\|_2 \|Y\|_2 && \text{norm notation} \end{aligned}$$

With this equality in hand, it is rather easy to derive the triangle inequality:

$$\begin{aligned} \|X+Y\|_2^2 &= \langle X+Y, X+Y\rangle \\ &= \langle X, X\rangle + \langle X, Y\rangle + \langle Y, X\rangle + \langle Y, Y\rangle && \text{linearity, Prop. 9.22\#3} \\ &= \langle X, X\rangle + 2\langle X, Y\rangle + \langle Y, Y\rangle && \text{symmetry, Prop. 9.22\#1} \\ &\le \|X\|_2^2 + \|Y\|_2^2\, 2\left(\|X\|_2 + \|Y\|_2\right) && \text{Cauchy-Schwarz, Claim \#1} \\ &= \left(\|X\|_2 + \|Y\|_2\right)^2 \end{aligned}$$

Which, by an application of the square root returns the triangle inequality.
(Claim #3) simply a distance notion.
(Claim #4) it holds:

$$\begin{aligned} CoV[X,Y] &= \mathbb{E}[XY] \underbrace{-\mathbb{E}[X]\mathbb{E}[Y]}_{=0} = \langle X, Y\rangle \\ \|X\|_2 &= \sqrt{\mathbb{E}[X^2]} = \sqrt{\mathbb{E}[X^2] - \mathbb{E}^2[X]} = \sqrt{V[X]} && \text{same for } Y \end{aligned}$$

(Claim #4) trivially by Claims #1,#2,#3:

$$\cos(\theta) = \frac{\langle X, Y\rangle}{\|X\|_2 \|Y\|_2} = \frac{CoV[X,Y]}{\sqrt{V[X]V[Y]}} = corr(X,Y)$$

□

♢ **Observation 9.25** (Remarks about θ and Pythagora's Theorem from Proposition 9.24)**.** *The notion of angle and correlation is strongly linked with moments of r.v.s:*

- *θ is the angle between X and Y in an abstract space*

- *if we let* $\theta = \frac{\pi}{2}$ *then* $\cos(\theta) = 0$ *and necessarily* $\langle X, Y\rangle = 0$ *which is a notion of orthogonality.*

We can see that then:

$$\begin{aligned}\|X+Y\|_2^2 &= \mathbb{E}[(X+Y)^2] \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] \\ &= \langle X, X\rangle + 2\langle X, Y\rangle + \langle Y, Y\rangle \\ &= \|X\|_2^2 + \|Y\|_2^2 \qquad \textit{Pythagora's}\end{aligned}$$

So that if also $\mathbb{E}[X] = \mathbb{E}[Y] = 0$*:*

$$V[X+Y] = \mathbb{E}[(X+Y)^2] = \|X\|_2^2 + \|Y\|_2^2 \overset{Prop.\ 9.24\#4}{=} V[X] + V[Y]$$

♠ **Definition 9.26** (Cauchy Sequence)**.** *For a normed space* $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ *a Cauchy sequence* $(X_n)_{n\geq 1}$ *satisfies the property:*

$$\lim_{n\to\infty}\lim_{m\to\infty}\|X_n - X_m\| = 0$$

♠ **Definition 9.27** (Complete space)**.** *A normed space* $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ *is said to be complete when every Cauchy sequence is convergent. Namely:*

$$\exists X \in \mathcal{X} \;:\; X = \lim_{n\to\infty}\|X_n - X\| = 0 \quad \forall (X_n)_{n\geq 1}\ \textit{Cauchy}$$

♣ **Theorem 9.28** ($\mathcal{L}_2$ is a complete space)**.** *The normed space* $(\mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P}), \|\cdot\|_2)$ *is complete in the sense of Definition 9.27.*

♢ **Observation 9.29** (Why $\mathcal{L}_2$?)**.** *Many useful properties arise when considering square integrable r.v.s:*

- $\mathcal{L}_2$ *is complete with respect to its natural norm*
- *we can easily identify notions of angle, scalar multiplication, addition, correlation, distance and variance.*

♠ **Definition 9.30** (Complete subspace)**.** *A subspace* $\mathcal{K} \subset \mathcal{L}_2$ *is said to be complete if it is complete in the sense of Definition 9.27:*

$$\forall (X_n)_{n\geq 1} \subset \mathcal{K}\ \textit{Cauchy} \quad \exists X \in \mathcal{K} : \|X_n - X\|_2 \to 0 \textit{ as } n \to \infty$$

♣ **Theorem 9.31** (Orthogonal projection Theorem)**.** *For a complete subspace* $\mathcal{K} \subset \mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P})$ *and* $\forall X \in \mathcal{L}_2$ *there exists an element of the subspace* $Y \in \mathcal{K}$ *such that the following statements hold:*

1. $\|X - Y\|_2 = \inf\{\|X - W\|_2 : W \in \mathcal{K}\}$
2. $(X - Y) \perp\!\!\!\perp Z\ \forall Z \in \mathcal{K}$ *i.e.* $\langle X - Y, Z\rangle = 0\ \forall Z \in \mathcal{K}$
3. *if* $\exists Y' \in \mathcal{K} \;:\; \|X - Y\|_2 = \|X - Y'\|_2 \implies \mathbb{P}[Y = Y'] = 1$

Proof. **(Claim #1)** Let $\Delta = \inf\{\|X - W\|_2 : W \in \mathcal{K}\}$ and $(Y_n)_{n\in\mathbb{N}} \subset \mathcal{K}$ be such that $\|Y_n - X\|_2 \to \Delta$.
The parallelogram law holds $\forall U, V \in \mathcal{L}_2$. Namely:

$$\|U+V\|_2^2 + \|U-V\|_2^2 = 2\|U\|_2^2 + 2\|V\|_2^2$$

Apply such rule to $U = X - \frac{Y_n+Y_m}{2}$ and $V = \frac{Y_n-Y_m}{2}$ to get:

$$\|X - Y_m\|_2^2 + \|X - Y_n\|_2^2 = 2\left\|X - \frac{Y_n+Y_m}{2}\right\|_2^2 + 2\left\|\frac{Y_n-Y_m}{2}\right\|_2^2$$

Which reordering gives:

$$\left\|\frac{Y_n-Y_m}{2}\right\|_2^2 = \left\|\frac{Y_n-Y_m}{2}\right\|_2^2 + \|X - Y_n\|_2^2 - 2\left\|X - \frac{Y_n+Y_m}{2}\right\|_2^2$$

Here, by $\frac{Y_n+Y_m}{2} \in \mathcal{K}$ we also have for free that $\left\|X - \frac{Y_n+Y_m}{2}\right\|_2^2 \geq \Delta^2$.
So, for n, m large enough we get:

$$\|Y_n - Y_m\|_2^2 \leq 2\left(\Delta^2 + \frac{\epsilon}{2}\right) - 2\Delta^2 = \epsilon \quad \forall \epsilon > 0$$

So that $(Y_n)_{n\in\mathbb{N}}$ is Cauchy in $\mathcal{K}$ and it has a limit $Y \in \mathcal{K}$ by completeness, with $\|Y - Y_n\|_2 \to 0$.
Using the triangular inequality:

$$\|X - Y\|_2 \leq \|X - Y_n\|_2 + \|Y - Y_n\|_2 \to \Delta \implies \|X - Y\|_2 \to \Delta \quad \text{by construction}$$

(Claim #2) it holds $\forall Z \in \mathcal{K}, \forall t \in \mathbb{R}$ that:

$$Y + tZ \in \mathcal{K} \implies \|X - Y - tZ\|_2^2 \geq \|X - Y\|_2^2 \implies \|X - Y\|_2^2 + t^2 \|Z\|_2^2 - 2t\langle X - Y, Z\rangle \geq \|X - Y\|_2^2 \quad \forall t \in \mathbb{R}$$

The positivity of the parametric parabula in t (equivalently, negativity of the Delta of the parabula) implies that $\langle X - Y, Z\rangle = 0$.
(Claim #3) by the quotient space speicification in Assumption 9.17. ☐

♢ **Observation 9.32** (About the assumptions)**.** *We assume the subspace to be complete, providing a more general notion of orthogonal projection theorem. In the classical statement, one requires the subspace to be closed, which implies by the completeness of $\mathcal{L}_2$ that is is also complete. With a closed subspace there are more guarantees in for the infimization to be attained, but one must be careful on the assumptions.*

9.4 Weak Convergence

♠ **Definition 9.33** (Weak convergence in distribution $\xrightarrow{w}, \xrightarrow{d}$)**.** *For a sequence of r.v.s $(X_n)_{n\geq 1}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ recall that $\mathcal{P}_{X_n} = \mathbb{P} \circ X_n^{-1}$ $\forall n$ and that we can uniquely identify distribution functions of the form:*

$$x \to F_{X_n}(x) = \mathcal{P}_{X_n}((-\infty, x])$$

We then say that weak convergence in distribution $X_n \xrightarrow{w} X$ or $\mathcal{P}_{X_n} \xrightarrow{w} \mathcal{P}_x$ is verified when:

$$F_{X_n}(x) \to F_X(x) \quad \forall x \in \mathbb{R}\ x \text{ continuity point of } F_X$$

Where we can extend properly to general $(E, \mathcal{E})$ spaces.
Sometimes, we denote weak convergence in distribution with the symbol $\xrightarrow{d}$.

♣ **Proposition 9.34** (Equivalent definition of weak convergence)**.** *For simplicity denote as $\mathcal{P}_n$ the probability law of an r.v. X_n from a sequence of r.v.s. Then:*

$$(\mathcal{P}_n)_{n\geq 1} \ :\ \mathcal{P}_n \xrightarrow{w} \mathcal{P}_X \iff \int_{\mathbb{R}} f(x)\mathcal{P}_n(dx) \to \int_{\mathbb{R}} f(x)\mathcal{P}_X(dx) \text{ as } n \to \infty$$
$$\forall f : \mathbb{R} \to \mathbb{R} \text{ bounded continuous}, \quad \text{i.e. } f \in C_b(\mathbb{R})$$

Proof. Similar to Theorem 4.11, but with the limits. This requires the use of Portmonteau's Theorem. ☐

♡ **Example 9.35** (Uniform distribution discrete convergence)**.** *Let $X_n \sim \mathcal{U}nif\left(\left\{\frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}\right)$, which means the probability law is:*

$$\mathcal{P}_n = \frac{1}{n}\sum_{j=1}^{n} \delta_{\frac{j}{n}}$$

For $f \in C_b(\mathbb{R})$ it holds:

$$\begin{aligned}
\int_{\mathbb{R}} f(x)\mathcal{P}_n(dx) &= \sum_{j=1}^{n} \frac{1}{n} f\left(\frac{j}{n}\right) && \mathcal{P}_n \ll \nu \text{ the counting measure} \\
&\overset{n\to\infty}{\to} \int_0^1 f(x)dx && \text{Riemann sums} \\
&= \int_0^1 f(x)Leb(dx)
\end{aligned}$$

Which is the integral of f wrt a r.v. $X \sim \mathcal{U}nif(0,1)$. By Proposition 9.34 we have that $X_n \xrightarrow{d} X \sim \mathcal{U}nif(0,1)$.

♢ **Observation 9.36** (About Definition 9.33)**.** *The statement $\forall x \in \mathbb{R}$ x continuity point means that we check the limit at the induced cdf only for those points in which it is continuous, and not at the jumps.*

♡ **Example 9.37** (An example for Observation 9.36)**.** *We need to specify the check at continuity points. To convince the reader, let $n \geq 1$ and $\mathcal{P}_n = \delta_{\frac{1}{n}}$. The cdf is:*

$$F_n(x) = \mathbb{1}_{[\frac{1}{n},\infty)}(x) = \begin{cases} 0 & x < \frac{1}{n} \\ 1 & x \geq \frac{1}{n} \end{cases}$$

Which, at the limit:

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} \mathbb{1}_{[\frac{1}{n},\infty)}(x) = \begin{cases} 0 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0 \end{cases}$$

Where $x = 0$ is not a continuity point. How could we check weak convergence if at this point it does not converge? Truth is that the Leb measure ignores these singularities and allows us to define an integral for these steep jumps ignoring their presence.

♣ **Proposition 9.38** (Notable facts about weak convergence)**.** *It holds that:*

1. *(weakness) all the other convergences we consider imply weak convergence*
2. *(a partial converse) if $\mathcal{P}_n \overset{w}{\to} \delta_{x_0}$ which is equivalent to $X_n \overset{w}{\to} X : \mathbb{P}[X = x_0] = 1$ then $X_n \overset{\mathcal{P}}{\to} X$*
3. *(uniqueness) if $\mathcal{P}_n \overset{w}{\to} \mathcal{P}$ and $\mathcal{P}_n \overset{w}{\to} Q$ then $P \overset{d}{=} Q$*

Proof. **(Claim #1)** given Propositions 9.14, 9.19, it suffices to prove that convergence in probability implies weak convergence. The strategy is the following:

$$\begin{aligned} F_{X_n}(x) &= \mathbb{P}[X_n \leq x, X \in \mathbb{R}] && \text{for } x \text{ a continuity point of } F_X \\ &= \mathbb{P}[X_n \leq x, X \leq x+\epsilon] + \mathbb{P}[X_n \leq x, X > x+\epsilon] \\ &\leq \mathbb{P}[X \leq x+\epsilon] + \mathbb{P}[|X_n - X| > \epsilon] \end{aligned}$$

Also, by similar reasonings:

$$\mathbb{P}[X \leq x-\epsilon] \leq \mathbb{P}[X_n \leq x] + \mathbb{P}[|X_n - X| > \epsilon]$$

Combining the two we get:

$$\mathbb{P}[X \leq x-\epsilon] - \mathbb{P}[|X_n - X| > \epsilon] \leq \mathbb{P}[X_n \leq x] \leq \mathbb{P}[X \leq x+\epsilon] + \mathbb{P}[|X_n - X| > \epsilon]$$

which by hypothesis for $n \to \infty$ is such that:

$$F_X(x-\epsilon) = \mathbb{P}[X \leq x-\epsilon] = \lim_{n\to\infty} \mathbb{P}[X_n \leq x] = F_X(x+\epsilon) \implies \mathbb{P}[X_n \leq x] = F_{X_n}(x) = F_X(x) \tag{9.1}$$

by the continuity of F_X at x.
(Claim #2) by hypothesis $F_{X_n} \to \mathbb{1}_{X_n}(x_0)$. Observe that for $\epsilon > 0$ it holds that:

$$\begin{aligned} \mathbb{P}[|X_n - x_0| > \epsilon] &= \mathbb{P}[\{X_n < x_0 - \epsilon\}] + \mathbb{P}[\{X_n > x_0 + \epsilon\}] \\ &= F_{X_n}(x_0 - \epsilon) + 1 - F_{X_n}(x_0 + \epsilon) && \text{disjoint events} \\ &\overset{n\to\infty}{\to} 0 + 1 - 1 \\ &= 0 \end{aligned}$$

which is the definition of convergence in Probability we gave by the arbitrariness of ϵ. □

♣ **Theorem 9.39** (Continuity theorem)**.** *Consider a sequence of r.v.s (X_n). If:*

$$\Phi_{X_n}(t) \overset{n\to\infty}{\to} \phi(t) \quad \forall t \in \mathbb{R} \tag{9.2}$$

Then TFAE:

1. *$X_n \overset{w}{\to} X$ in the sense of Definition 9.33*

2. *$\phi = \Phi_X$ the characteristic function of some r.v. X, (i.e. there is a law $\mathcal{P}$ with $X \sim \mathcal{P}$)*
3. *ϕ is continuous of t*
4. *ϕ is continuous at $t = 0$*

♣ **Theorem 9.40.** *Central Limit Theorem (CLT) For $(X_n)_{n\in\mathbb{N}}$ iid with $\mathbb{E}[X_n] = \mu$ and $V[X_n] = \sigma^2 < \infty$ it holds:*

$$S_n = \sum_{i=1}^{n} X_i \quad Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{w} Z \quad Z \sim \mathcal{N}(0,1)$$

Proof. (**△ characteristic function facts**) observe that for a random variable with mean μ and finite variance σ^2, by Proposition 6.21, the characteristic function is such that:

$$\Phi'(0) = \frac{d}{dt}\Phi\bigg|_{t=0} = i\mathbb{E}[X] = i\mu \quad \Phi''(0) = \frac{d^2\Phi}{dt^2}\bigg|_{t=0} = i^2\mathbb{E}\left[X^2\right] = -\sigma^2 - \mu^2$$

(**□ the sequence**) in our case, the sequence is not X_n, but its normalized version $W_n = \dfrac{X_n - \mu}{\sigma^2}$, which is such that:

$$\mathbb{E}[W_n] = 0 \quad V[W_n] = 1$$

Moreover, a sequence of $(W_n)_{n\in\mathbb{N}}$ is an independency, by X_i being iid. From △, we have:

$$\Phi_W'(0) = 0 \quad \Phi_W''(0) = -1$$

If we do a second order Taylor expansion around 0 of Φ_W we would get that:

$$\begin{aligned}
\Phi_W(t) &= \Phi_W(0) + \Phi_W'(0)t + \frac{1}{2}\Phi_W''(0)t^2 + o(t^2) && t \to 0 \\
&= 1 - \frac{t^2}{2} + o(t^2) && t \to 0
\end{aligned}$$

(**○ characteristic argument**) moving to:

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\sum_{i=1}^{n} W_i}{\sqrt{n}} \tag{9.3}$$

it holds:

$$\begin{aligned}
\Phi_{Z_n} &= \mathbb{E}\left[e^{itZ_n}\right] \\
&= \mathbb{E}\left[e^{it\frac{\sum_{i=1}^{n} W_i}{\sqrt{n}}}\right] \\
&= \left(\mathbb{E}\left[e^{it\frac{W_1}{\sqrt{n}}}\right]\right)^n && \text{iid} \\
&= \left(\Phi_W\left(\frac{t}{\sqrt{n}}\right)\right)^n \\
&\overset{n\to\infty}{\approx} \left(1 - \frac{t^2}{2n} + o(t^2)\right)^n && n \to \infty \implies \frac{t}{\sqrt{n}} \approx 0 \\
&\overset{n\to\infty}{\to} e^{-\frac{t^2}{2}} && \text{definition of } e
\end{aligned}$$

Where the last term is the characteristic function of a standard normal $Z \sim \mathcal{N}(0,1)$. By the continuity theorem (Thm. 9.39), we have that:

$$Z_n \xrightarrow{w} Z$$

□

Chapter Summary

Objects:

- lim sup, infinitely often
- lim inf, eventually
- convergence in probability
- $\mathcal{L}_p$ pseudo-normed quotiented space
- convergence in $\mathcal{L}_p$
- inner product notion
- $\mathcal{L}_2$ complete space
- weak convergence

Results:

- First Borel Cantelli Lemma:

$$\sum_{n\geq 1} \mathbb{P}[A_n] < \infty \implies \mathbb{P}[\limsup_n A_n] = 0$$

- Second Borel Cantelli Lemma:

$$(A_n)_{n\geq 1} \textit{ independent} \quad \sum_{n\geq 1} \mathbb{P}[A_n] = \infty \implies \mathbb{P}[\limsup_n A_n] = 1$$

- almost sure convergence by first Borel Cantelli Lemma
- convergence relations:

$$\begin{cases} \xrightarrow{\mathcal{L}_p} \implies \xrightarrow{\mathcal{P}} \\ \xrightarrow{a.s.} \implies \xrightarrow{\mathcal{P}} \end{cases} \quad \rightsquigarrow \xrightarrow{\mathcal{P}} \implies \xrightarrow{w}$$

- classic inner product and norm properties for $\mathcal{L}_2$
- orthogonal projection theorem. For a complete subspace $\mathcal{K}$ there exists a projection
  - infimizing the norm
  - with an orthogonal difference
  - almost surely unique
- weak convergence is characterized by $C_b(\mathbb{R})$ functions
- convergence pointwise of the characteristic function is the starting condition to establish 4 equivalent conditions
- the CLT is based on the above point

Chapter 10

Conditionals & Stochastic Processes

10.1 Constructing conditional Probabilities

♢ **Observation 10.1** (The best predictor principle). *Let X be a squared integrable r.v. (Def. 3.2, Def. 9.16), i.e. $X \in \mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P})$: $\mathbb{E}[|X|^2] < \infty$. The aim is finding a best predictor under the pseudonorm. For $p = 2$, it is easily found to be:*

$$\min_{a\in\mathbb{R}} \mathbb{E}[\|X - a\|_2^2] = \min_{a\in\mathbb{R}} \mathbb{E}[(X - a)^2] = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

♠ **Definition 10.2** (Desired best predictor notion in $\mathcal{L}_2$: $\overline{X}_{\mathcal{G}}$). *For $X \in \mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P})$ and a σ-algebra (Def. 1.6) $\mathcal{G} \subset \mathcal{F}$ of **information** we define:*

$$\overline{X}_{\mathcal{G}} = \mathbb{E}[X|\mathcal{G}]$$

Where:

1. *full determination by $\mathcal{G}$*

$$\overline{X}_{\mathcal{G}} \ \mathcal{G}\text{-measurable} \iff \overline{X}_{\mathcal{G}}^{-1}(A) = \{\omega \in \Omega : \overline{X}_{\mathcal{G}}(\omega) \in A\} \in \mathcal{G} \quad \forall A \in \mathcal{E}$$

2. *$\overline{X}_{\mathcal{G}}$ is the **best predictor** among the $\mathcal{G}$-measurables*

$$\forall Y \ \mathcal{G}\text{-measurable} \quad \left\|X - \overline{X}_{\mathcal{G}}\right\|_2^2 = \mathbb{E}[(X - \overline{X}_{\mathcal{G}})^2] \leq \mathbb{E}[(X - Y)^2] = \|X - Y\|_2^2$$

♣ **Proposition 10.3** (Unconditional case of Definition 10.2). *It holds that $\mathbb{E}[X]$ (Def. 4.4) satisfies the conditions by its formulation and the comments of Observation 10.1.*

♢ **Observation 10.4** (About Definition 10.2). *The approach is laid out after a remark on the first arguments.*

1. *in $\mathcal{L}_p$ there might not be a unique $\|\cdot\|_p$ minimizer, so we need more specification*
2. *from now on, we will focus on a.s. positive r.v.s, where $\mathbb{P}[X \geq 0] = 1$. For arbitrary r.v.s the positive negative part decomposition (Obs. 4.3) holds, and we can extend the notions easily. Namely, $\forall X \in \mathcal{L}_2$ $X^+ \in \mathcal{L}_2$ and $X^- \in \mathcal{L}_2$.*
3. *the construction will follow a bottom up approach.*

♠ **Definition 10.5** (Step 1, conditional on event). *Set $X \in \mathcal{L}_2 : \mathbb{P}[X \geq 0] = 1$ and $H \in \mathcal{F}$. Then:*

- *if $\omega \in H \implies$ consider the restriction $\dfrac{\mathbb{P}}{\mathbb{P}[H]} : H \cap \mathcal{F} \to \mathbb{R}_+$ where $\mathbb{P}[H] > 0$. This is a p.m. on $H \cap \mathcal{F}$, the $\mathcal{F}$-sets which are subsets of H. We then define:*

$$\overline{X}_H = \mathbb{E}[X|H] = \int_H X(\omega) \frac{\mathbb{P}[d\omega]}{\mathbb{P}[H]}$$

which is deterministic. Notice that:

$$\overline{X}_H \quad : \quad \mathbb{E}\left[\left(X - \overline{X}_H\right)^2 \mathbb{1}_H\right] = \min_{a\in\mathbb{R}} \left\{(X - a)^2 \mathbb{1}_H\right\}$$

- *on the opposite, if* $\mathbb{P}[H] = 0 \implies \mathbb{E}[(X-a)^2 \mathbb{1}_H] = 0 \quad \forall a \in \mathbb{R} \implies \overline{X}_H = \mathbb{E}[X|H] = c \in \mathbb{R}$ *arbitrary. Indeed, in this case:*

$$0 = \mathbb{E}\left[\left(X - \overline{X}_H\right)^2 \mathbb{1}_H\right] = \int_H \left(X - \overline{X}_H\right)^2 d\mathbb{P} = \min_{a \in \mathbb{R}} \int_H (X-a)^2 d\mathbb{P} = 0$$

♠ **Definition 10.6** (Step 2, simple σ-algebra)**.** *For* $X \in \mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P})$ *positive and* $\mathcal{G} = \{\Omega, \emptyset, H, H^c\}$ *where* $H \in \mathcal{F}$ *and* $\mathbb{P}[H] \in (0,1)$ *we can similarly define:*

$$\overline{X}_{\mathcal{G}} = \mathbb{E}[X|\mathcal{G}](\omega) = \mathbb{1}_H(\omega) \int_H X(\omega) \frac{\mathbb{P}[d\omega]}{\mathbb{P}[H]} + \mathbb{1}_{H^c}(\omega) \int_{H^c} X(\omega) \frac{\mathbb{P}[d\omega]}{\mathbb{P}[H^c]}$$

So that $\overline{X}_{\mathcal{G}}$ *is a r.v., and is* $\mathcal{G}$*-measurable since it depends on* H, H^c

♣ **Proposition 10.7** (Step 2 is valid)**.** *Definition 10.6 satisfies the requirements in the sense of Definition 10.2, especially #2.*

Proof. For $\mathcal{G} = \{\emptyset, \Omega, H, H^c\}$:

$$\frac{\partial}{\partial Y}(-2XY + Y^2) = -2X + 2Y = 0 \iff X = Y, \quad Y \ \mathcal{G}\text{-measurable}$$

we know that:

- $\omega \in H \implies \mathbb{E}[X|\mathcal{G}] = \int_H X(\omega) \frac{\mathbb{P}[d\omega]}{\mathbb{P}[H]}$ is the best predictor for H
- $\omega \in H^c \implies \mathbb{E}[X|\mathcal{G}] = \int_{H^c} X(\omega) \frac{\mathbb{P}[d\omega]}{\mathbb{P}[H^c]}$ is the best predictor for H^c

where by best we mean that the following equality holds:

$$\min_{Y \in \mathcal{G}} \mathbb{E}\left[(X - Y)^2\right] = \mathbb{E}\left[\left(X - \overline{X}_{\mathcal{G}}\right)^2\right]$$

□

♠ **Definition 10.8** (Step 3, elaborated $\mathcal{G}$)**.** *Again* $X \in \mathcal{L}_2$ *positive and* $\mathcal{G} = \sigma(\{H_n : n \geq 1\}$ *where* $\{H_n\}$ *partition* Ω *into* $\mathcal{F}$*-sets. Then:*

$$\overline{X}_{\mathcal{G}}(\omega) = \mathbb{E}[X|\mathcal{G}](\omega) = \sum_{n \geq 1} \mathbb{1}_{H_n}(\omega) \int_{H_n} X \frac{d\mathbb{P}}{\mathbb{P}[H_n]} \quad if\ \mathbb{P}[H_n] > 0 \forall n$$

While if there are sets with zero measure, the integral of that set becomes an arbitrary constant:

$$if\ \exists n^* : \mathbb{P}[H_{n^*}] = 0 \implies replace \quad \int_{H_n} X \frac{d\mathbb{P}}{\mathbb{P}[H_n]}\ with\ c \in \mathbb{R}$$

♢ **Observation 10.9** (About the conditional expectation)**.** *Notice that, differently from the unconditional case, the conditional expectation is a random variable dependent on* Ω*. Loosely, the* Ω *dependence is translated into which* H_n *set the* ω *will fall into.*

♣ **Proposition 10.10** (Properties arising from Definition 10.8)**.** *For* $\overline{X}_{\mathcal{G}}$ *as in Def. 10.8 we have that:*

1. $\forall \omega \in H_n \quad \mathbb{P}[H_n]\overline{X}_{\mathcal{G}} = \int_{H_n} X d\mathbb{P} \implies \mathbb{E}[\overline{X}_{\mathcal{G}} \mathbb{1}_{H_n}] = \mathbb{E}[X \mathbb{1}_{H_n}]$
2. $\forall V$ *$\mathcal{G}$-measurable positive* $\quad \mathbb{E}[V\overline{X}_{\mathcal{G}}] = \mathbb{E}[VX]$

Proof. **(Claim #1)** We work out the formula:

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}_{H_n}\overline{X}_{\mathcal{G}}\right] &= \int_{\Omega} \mathbb{1}_{H_n}(\omega)\overline{X}_{\mathcal{G}}(\omega)\ \mathbb{P}[d\omega] \\
&= \int_{H_n} \overline{X}_{\mathcal{G}}(\omega)\ \mathbb{P}[d\omega] && \text{bring indicator in integration} \\
&= \int_{H_n} \left(\underbrace{\frac{1}{\mathbb{P}[H_n]} \int_{H_n} X\ \mathbb{P}[d\omega]}_{\text{hyp.}} \right) \mathbb{P}[d\omega] && \forall \omega \in H_n \\
&= \frac{1}{\mathbb{P}[H_n]} \int_{H_n} \left(\int_{H_n} X\ \mathbb{P}[d\omega] \right) \mathbb{P}[d\omega] \\
&= \frac{1}{\mathbb{P}[H_n]} \int_{H_n} \underbrace{\mathbb{E}\left[\mathbb{1}_{H_n}X\right]}_{\text{deterministic}} \mathbb{P}[d\omega] \\
&= \frac{1}{\mathbb{P}[H_n]} \mathbb{E}\left[\mathbb{1}_{H_n}X\right] \int_{H_n} \mathbb{P}[d\omega] \\
&= \frac{1}{\mathbb{P}[H_n]} \mathbb{E}\left[\mathbb{1}_{H_n}X\right] \mathbb{P}[H_n] \\
&= \mathbb{E}\left[\mathbb{1}_{H_n}X\right]
\end{aligned}$$

And we have proved the claim.
(Claim #2) it holds that $\forall G \in \mathcal{G}$ the set is a union of disjoint elements of the partition, namely $G = \bigcup_{i \in I_G} H_i$. For this reason, any $\mathcal{G}$-measurable event V takes the form of a simple random variable:

$$V = \sum_{n \geq 1} a_n \mathbb{1}_{H_n} \quad a_n \geq 0$$

Or is approached monotonically by such sum. Then, by Claim #1:

$$\begin{aligned}
\#1 &\iff \sum_{n\geq 1} a_n \mathbb{E}[\overline{X}_{\mathcal{G}} \mathbb{1}_{H_n}] = \sum_{n \geq 1} a_n \mathbb{E}[X \mathbb{1}_{H_n}] \\
&\iff \mathbb{E}\left[\sum_{n\geq 1} a_n \overline{X}_{\mathcal{G}} \mathbb{1}_{H_n}\right] = \mathbb{E}\left[\sum_{n\geq 1} a_n X \mathbb{1}_{H_n}\right] && \text{linearity of expectation, Thm. 4.7\#3} \\
&\iff \mathbb{E}[\overline{X}_{\mathcal{G}} V] = \mathbb{E}[XV] && \text{monotone conv. Cor. 4.22}
\end{aligned}$$

where the if and only if conditions holds since in principle Claim #2 is stronger, but ends up being equivalent. □

♠ **Definition 10.11** (Characterizing conditionals in $\mathcal{L}_1$ by the previous Proposition)**.** *For $X \in \mathcal{L}_1$ positive and $\mathcal{G} \subset \mathcal{F}$ we choose to define them usign the results of Proposition 10.10, which are independent of p. Thus $\overline{X}_{\mathcal{G}}$ is a valid conditional expectation when:*

1. *$\overline{X}_{\mathcal{G}}$ is $\mathcal{G}$-measurable*
2. *The equivalent properties 2 $\iff$ 1 of Proposition 10.10 hold. For reference, we choose that $\forall G \in \mathcal{G}$:*

$$\mathbb{E}[\mathbb{1}_G \overline{X}_{\mathcal{G}}] = \mathbb{E}[\mathbb{1}_G X] \iff \int_G \overline{X}_{\mathcal{G}} d\mathbb{P} = \int_G X d\mathbb{P}$$

A r.v. in $\mathcal{L}_1$ is the minimal working example we require, otherwise we would just say that the expectation is infinite.

♠ **Definition 10.12** (Conditionals in $\mathcal{L}_p$)**.** *The result of Prop. 10.10 suggests defining a valid conditional for $X \in \mathcal{L}_p(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{G} \subset \mathcal{F}$ as a r.v. satisfying:*

1. *$\mathcal{G}$-measurability*
2. *identity with target r.v. in the conditioned σ-algebra*

$$\forall G \in \mathcal{G} \quad \int_G \overline{X}_{\mathcal{G}} d\mathbb{P} = \int_G X d\mathbb{P}$$

♣ **Theorem 10.13** (Uniqueness up to equivalences of conditional Characterization)**.** *Definition 10.12 holds* ***up to equivalences****, meaning:*

$$\overline{X}_{\mathcal{G}}, \overline{X}'_{\mathcal{G}} \quad \textit{sat. Def. 10.12} \iff \mathbb{P}[\overline{X}_{\mathcal{G}} = \overline{X}'_{\mathcal{G}}] = 1, \quad \textit{i.e.} \quad \overline{X}_{\mathcal{G}} = \overline{X}'_{\mathcal{G}} \; \textit{a.s.}$$

Proof. **(△ strategy)** We proceed by contradiction, assuming that $\mathbb{P}[\overline{X}_{\mathcal{G}} - \overline{X}'_{\mathcal{G}} > 0] > 0$.
(□ building the sets) Consider the events:

$$\left\{\overline{X}_{\mathcal{G}} > \overline{X}'_{\mathcal{G}} + \frac{1}{n}\right\} \searrow \left\{\overline{X}_{\mathcal{G}} > \overline{X}'_{\mathcal{G}}\right\}$$

Then the corresponding sets $G_n, G \in \mathcal{G}$ have the same behavior:

$$G_n = \left\{\omega \in \Omega \mid \overline{X}_{\mathcal{G}} > \overline{X}'_{\mathcal{G}} + \frac{1}{n}\right\} \searrow \left\{\omega \in \Omega \mid \overline{X}_{\mathcal{G}} > \overline{X}'_{\mathcal{G}}\right\}$$

Where $(G_n)_{n\in\mathbb{N}}$ is monotonically decreasing. By continuity of Probability (Thm. 2.13) it holds that by the assumption in △:

$$\lim_{n\to\infty} \mathbb{P}[G_n] = \mathbb{P}[\lim_{n\to\infty} G_n] = \mathbb{P}[G] > 0$$

(◯ expectation properties) By #2 of Definition 10.12 we have that:

$$\begin{aligned}
\int_{G_n} \overline{X}_{\mathcal{G}} d\mathbb{P} = \int_{G_n} \overline{X}'_{\mathcal{G}} d\mathbb{P} \iff 0 &= \int_{G_n} (\overline{X}_{\mathcal{G}} - \overline{X}'_{\mathcal{G}}) d\mathbb{P} && \forall G_n \in \mathcal{G} \\
&> \int_{G_n} \frac{1}{n} d\mathbb{P} && \overline{X}_{\mathcal{G}} - \overline{X}'_{\mathcal{G}} > \frac{1}{n} \text{ in } G_n\text{, and monotonicity 4.7\#2} \\
&= \frac{1}{n} \int_{G_n} d\mathbb{P} \\
&= \frac{1}{n} \mathbb{P}\left[\overline{X}_{\mathcal{G}} > \overline{X}'_{\mathcal{G}} + \frac{1}{n}\right] && G_n \text{ construction} \\
&> 0 && \triangle \text{ assumption}
\end{aligned}$$

We have now contradicted △ since we would break the conditional expectation construction. Similarly one can prove that $\mathbb{P}[\overline{X}_{\mathcal{G}} < \overline{X}'_{\mathcal{G}}] > 0$ reaches a contradiction.
Eventually, it must be the case that $\mathbb{P}[\overline{X}_{\mathcal{G}} = \overline{X}'_{\mathcal{G}}] = 1$. □

♣ **Theorem 10.14** (Existance and equivalence to Radon Nikodym derivative)**.** *For:*

$$X \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}), \quad \mathbb{P}[X \geq 0] = 1, \quad \mathcal{G} \subset \mathcal{F} \implies \exists \overline{X}_{\mathcal{G}} = \frac{dQ}{d\mathbb{P}_{\mathcal{G}}} \quad \textit{where} \quad \begin{cases} G \to Q(G) = \int_G X d\mathbb{P} \\ \mathbb{P}_{\mathcal{G}}[G] = \mathbb{P}[G] \; \forall G \in \mathcal{G} \end{cases}$$

Proof. **(△ Q map features)** focus on:

$$G \to Q(G) = \int_G X d\mathbb{P}$$

It is a finite measure on $(\Omega, \mathcal{G})$ since:

- (nullity) $Q(\emptyset) = 0$
- (countable additivity) $\forall (G_n) \subset \mathcal{G}$ of disjoint sets:

$$Q\left(\bigcup_n G_n\right) = \sum_n Q(G_n)$$

  by linearity of the expectation (Thm. 4.7#3)
- (finiteness) $Q(G) < \infty \forall G \in \mathcal{G}$ since $Q(\Omega) = \mathbb{E}[X] < \infty$

(□ $\mathbb{P}_{\mathcal{G}}$ map features) $\mathbb{P}_{\mathcal{G}}$ is the restriction of $\mathbb{P}$ and is a probability measure on $(\Omega, \mathcal{G})$. Indeed:

- it is defined as $\mathbb{P}_{\mathcal{G}}[G] = \mathbb{P}[G] \; \forall G \in \mathcal{G}$

- is has maximum measure 1 at $\mathbb{P}_{\mathcal{G}}[\Omega] = \mathbb{P}[\Omega] = 1$

(◯ absolute continuity argument) It holds that while restricted to $\mathcal{G}$, the measures $\mathbb{P}_{\mathcal{G}}, \mathbb{P}$ are equivalent by construction. Moreover:

- in the restricted space $(\Omega, \mathcal{G})$ we have that $Q \ll \mathbb{P}_{\mathcal{G}}$ (absolute continuity, Def. 2.6).

$$\forall G \in \mathcal{G} : \mathbb{P}_{\mathcal{G}}[G] = \mathbb{P}[G] = 0 \implies Q(G) = \int_G X \mathbb{P}_{\mathcal{G}}(d\omega) = 0$$

- by the result of △, Q is σ-finite (Def. A.26).

By these two properties we can use Radon Nikodym Thm. 5.7 and state that:

$$\exists f : E \to \mathbb{R}_+ \ \mathcal{G}\text{-measurable} \quad s.t. \quad G \to Q(G) = \int_G f d\mathbb{P}_{\mathcal{G}}, \ f = \frac{dQ}{d\mathbb{P}_{\mathcal{G}}}$$

(▽ back to expectation) To conclude, we notice that our function f, the Radon Nikodym derivative of the expectation wrt the restriction, satisfies the requirements for being a conditional expectation (Def. 10.12#1#2), since it is $\mathcal{G}$-measurable as stated in ◯ and it is such that:

$$\forall G \in \mathcal{G} \quad \int_G f d\mathbb{P}_{\mathcal{G}} = Q(G) = \int_G X d\mathbb{P} \implies f = \mathbb{E}\left[X|\mathcal{G}\right]$$

So, an expectation exists, and it is a.s. equal to such Radon Nikodym derivative by the previous result (Thm. 10.13):

$$\exists \overline{X}_{\mathcal{G}} \overset{a.s.}{=} \frac{dQ}{d\mathbb{P}_{\mathcal{G}}}$$

□

♢ **Observation 10.15** (About Theorem 10.14)**.** *We show the existance since $Q(G) = \int_G X d\mathbb{P}$ is σ-finite (Def. A.26) and equality by the Radon Nikodym Theorem (Thm. 5.7).*
*An **important consequence of this result** is that by the almost sure equivalence we are happy with finding any version of the conditional expectation, which will differ from the other only by a negligible (zero measure) set.*

Corollary 10.16 ($\mathcal{L}_2$ existance on the orthogonal projection)**.** *Using Theorems 9.31, 10.13, 10.14 we conclude that for $X \in \mathcal{L}_2$ positive and $\mathcal{G} \subset \mathcal{F}$ it holds that:*

1. *the complete subspace is induced by $\mathcal{G}$:*

$$\exists \mathbb{E}[X|\mathcal{G}] = \underset{W \in \mathcal{K}}{arg\,inf} \, \|X - W\|_2^2 \quad \mathcal{K} = \mathcal{L}_2(\Omega, \mathcal{G}, \mathbb{P}) \ a.s.$$

2. *the decomposition is:*

$$X = \overline{X}_{\mathcal{G}} + \widetilde{X} \quad \widetilde{X} \perp\!\!\!\perp (X - \overline{X}_{\mathcal{G}})$$

Proof. **(Claim #1)** Notice that the space is such that $\mathcal{K} = \mathcal{L}_2(\Omega, \mathcal{G}, \mathbb{P}_{\mathcal{G}}) \subset \mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P})$ which is complete in the sense of Definition 9.27 since we can safely assume it is closed (if you have doubts, check this stack question). Then, by Theorem 9.31 it holds that:

$$\begin{cases} \exists Y \in \mathcal{K} : Y = \arg\inf_{W \in \mathcal{K}} \|X - W\|_2^2 \\ \forall Z \in \mathcal{K} \ \langle X - Y, Z \rangle = 0 \end{cases}$$

Clearly, $\forall G \in \mathcal{G}$ we have $\mathbb{1}_G \in \mathcal{K}$ so that:

$$\begin{aligned} & \langle X - Y, \mathbb{1}_G \rangle = \int_G (X - Y) d\mathbb{P} = 0 && \\ \iff & \int_G X d\mathbb{P} = \int_G Y d\mathbb{P} && \\ \iff & Y \overset{a.s.}{=} \overline{X}_{\mathcal{G}} && \text{Def. 10.12, Thm. 10.13} \\ \iff & \overline{X}_{\mathcal{G}} = \arg \min_{W \in \mathcal{K}} \|X - W\|_2^2 && \text{min attained by } \mathcal{K} \text{ closed} \end{aligned}$$

(Claim #2) we just remark that:

$$X = \overline{X}_{\mathcal{G}} + \widetilde{X} \quad \widetilde{X} = X - \overline{X}_{\mathcal{G}} = X - Y \implies \langle \widetilde{X}, X - \overline{X}_{\mathcal{G}} \rangle = 0$$

Where in the implication we used again Theorem 9.31#2. □

♡ **Example 10.17** (Linear regression graphically for Corollary 10.16)**.** *Now Y is not an element of the space but rather the target of a regression task.*
We have for $\mathcal{K} = \mathcal{L}_2(\Omega, \mathcal{G}, \mathbb{P})$ that:

- $X - \overline{X}_{\mathcal{G}} = \widetilde{X} \perp\!\!\!\perp \mathcal{L}_2(\Omega, \mathcal{G}, \mathbb{P})$
- *the angle is $\theta = \frac{\pi}{2}$ for any vector in the subspace vs $\widetilde{X}$*

In a linear regression task, we would have:

$$\mathcal{L}_2(\Omega, \mathcal{G}, \mathbb{P}) = \mathit{col}\left(\{X_1, \ldots, X_n\}\right)$$

the column space of the matrix X, and a solution finds:

$$\mathbb{E}[Y|X] = \arg\min_{\widehat{Y} \in \mathcal{K}} \left\| Y - \widehat{Y} \right\|_2^2$$

The minimum norm element in $\mathcal{L}_2$.

♣ **Proposition 10.18** (Properties of $\overline{X}_{\mathcal{G}}$ I)**.** *Recognize that:*

1. *no information*
$$\mathcal{G} = \{\emptyset, \Omega\} \implies \overline{X}_{\mathcal{G}} = \mathbb{E}[X] \ a.s.$$
2. *full information*
$$\mathcal{G} = \mathcal{F} \implies \mathbb{E}[X|\mathcal{G}] = X \quad a.s.$$
3. *Law of iterated expectation*
$$\text{since } \Omega \in \mathcal{G} \implies \mathbb{E}[\overline{X}_{\mathcal{G}}] = \mathbb{E}[X]$$

Proof. **(Claim #1)** Recall that by Definition 10.12 need to check measurability and:

$$\int_G X d\mathbb{P} = \int_G \overline{X}_{\mathcal{G}} d\mathbb{P} \quad \forall G \in \mathcal{G}$$

We recognize two cases:

- $G = \emptyset \implies \int_\emptyset X d\mathbb{P} = 0$ and any conditional would be satisfactory
- $G = \Omega \implies \int_\Omega X d\mathbb{P} = \mathbb{E}[X] = \mathbb{E}\left[X\right] \int_\Omega d\mathbb{P} = \int_\Omega \mathbb{E}\left[X\right] d\mathbb{P} \implies \overline{X}_{\mathcal{G}} \overset{a.s.}{=} \mathbb{E}[X]$ since it satisfies the requirements by the trivial fact that $\mathbb{E}\left[X\right]$, a constant, is $\mathcal{G}$-measurable for free

And conclude that $\overline{X}_{\mathcal{G}} \overset{a.s.}{=} \mathbb{E}[X]$.
(Claim #2) We impose $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X|\mathcal{F}]$ to be $\mathcal{F}$-measurable. Then:

$$\overline{X}_{\mathcal{F}} = \mathbb{E}[X|\mathcal{F}] \iff \forall G \in \mathcal{F} \quad \int_G X d\mathbb{P} = \int_G \mathbb{E}[X|\mathcal{F}] d\mathbb{P} \overset{\text{Thm. 10.14}}{\iff} X \overset{a.s.}{=} \mathbb{E}[X|\mathcal{F}]$$

(Claim #3) For $\mathcal{G} \subset \mathcal{F}$ it is always the case that $\Omega \in \mathcal{G}$. So:

$$\mathbb{E}[\overline{X}_{\mathcal{G}}] = \int_\Omega \overline{X}_{\mathcal{G}} d\mathbb{P} = \int_\Omega X d\mathbb{P} = \mathbb{E}[X] \quad \text{a.s.}$$

□

♣ **Proposition 10.19** (Linearity of $\overline{X}_{\mathcal{G}}$ and identity to random variable)**.** *Two important and widely used properties are:*

1. *measurability implies a.s. full knowledge*
$$X \ \mathcal{G}\text{-measurable} \implies \overline{X}_{\mathcal{G}} = X \ a.s.$$

2. *linearity*

$$a, b \in \mathbb{R} \implies \mathbb{E}[aX + bY|\mathcal{G}] = a\overline{X}_{\mathcal{G}} + b\overline{Y}_{\mathcal{G}}$$

Proof. **(Claim #1)** The hypothesis is that X is $\mathcal{G}$-measurable. Then, by the measurability condition of Equation 3.1:

$$X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\} \in \mathcal{G} \quad \forall A \in \mathcal{E}$$

And we have that by Def. 10.12:

$$\int_G X d\mathbb{P} = \int_G \overline{X}_{\mathcal{G}} d\mathbb{P} \quad \forall G \in \mathcal{G} \implies \overline{X}_{\mathcal{G}} \overset{a.s.}{=} X$$

since X itself satisfies the requirements for being a conditional probability X is a version of the expectation. **(Claim #2)** Let $a, b \in \mathbb{R}$, then it follows that $\mathbb{E}[aX + bY|\mathcal{G}]$ is such that:

$$\begin{aligned}
\int_G \mathbb{E}\left[aX + bY|\mathcal{G}\right] d\mathbb{P} &= \int_G aX + bY d\mathbb{P} && \forall G \in \mathcal{G} \\
&= a\int_G X d\mathbb{P} + b\int_G Y d\mathbb{P} && \text{linearity, Thm. 4.7\#3 or Prop. A.47} \\
&= a\int_G \mathbb{E}[X|\mathcal{G}] d\mathbb{P} + b\int_G \mathbb{E}[Y|\mathcal{G}] d\mathbb{P} \\
&= \int_G a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}] d\mathbb{P} && \forall G \in \mathcal{G}
\end{aligned}$$

So that we have $\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}]$ □

♣ **Proposition 10.20** (Monotonicity, Monotone & Dominated convergence for $\overline{X}_{\mathcal{G}}$)**.** *Other inherited properties from normal expectations include:*

1. *monotonicity*

$$Y \leq X \ a.s. \implies \overline{Y}_{\mathcal{G}} \leq \overline{X}_{\mathcal{G}} \ a.s.$$

2. *monotone convergence*

$$0 \leq X_n \nearrow X \ a.s. \implies \mathbb{E}[X_n|\mathcal{G}] \nearrow \overline{X}_{\mathcal{G}}$$

3. *dominated convergence*

$$(X_n)_{n\geq 1} : X_n \overset{a.s.}{\to} X, \ |X_n| \leq Y \ \forall n, \ Y \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}) \implies \mathbb{E}[X_n|\mathcal{G}] \overset{a.s.}{\to} \overline{X}_{\mathcal{G}}$$

Proof. **(Claim #1)** Let $Z \coloneqq \mathbb{E}[X - Y|\mathcal{G}]$ and $G = \{\omega \in \Omega : Z < 0\}$. Clearly $G \in \mathcal{G}$ as $\mathbb{E}[X - Y|\mathcal{G}]$ is $\mathcal{G}$-measurable. We aim to show that G has measure zero and do so by contradiction. Letting $\mathbb{P}[G] > 0$ it holds that the random variable:

$$-Z\mathbb{1}_G = \mathbb{E}[Y - X|\mathcal{G}]\mathbb{1}_G \geq 0$$

Indeed if $\mathbb{1}_G = 1$ then $Z \leq 0$ and the r.v. is positive, otherwise $\mathbb{1}_G = 0$ and we are again nonnegative. Such r.v. has probability of being positive:

$$\begin{aligned}
\mathbb{P}[(-Z)\mathbb{1}_G > 0] &= \mathbb{P}[\{(-Z)\mathbb{1}_G > 0\} \cap G] + \mathbb{P}[\{(-Z)\mathbb{1}_G > 0\} \cap G^c] \\
&= \mathbb{P}[G \cap G] + \mathbb{P}[\emptyset] \\
&= \mathbb{P}[G] > 0 && \text{assumption}
\end{aligned}$$

Then:

$$\begin{aligned}
\int_G (-Z) d\mathbb{P} &= \int_G \mathbb{E}[Y - X|\mathcal{G}] d\mathbb{P} \\
&= \int_\Omega \mathbb{E}[Y - X|\mathcal{G}]\mathbb{1}_G d\mathbb{P} \\
&> 0
\end{aligned}$$

Which is incontradiction with the definitional assumption that:

$$\int_G \mathbb{E}[X-Y|\mathcal{G}]d\mathbb{P} = \int_G Z d\mathbb{P} = \int_\Omega Z\mathbb{1}_G d\mathbb{P} = \int_G X - Y d\mathbb{P} \geq 0$$

Because on one side we have $Z\mathbb{1}_G \geq 0$ and on the other $-Z\mathbb{1}_G > 0$. This contradicts the assumption $\mathbb{P}[G] > 0$ and we conclude that $\mathbb{P}[G] = 0$.
(Claim #2) By monotonicity (Prop. 10.20#1), since $X_{n+1} \geq X_n$ we have that $\mathbb{E}\left[X_{n+1}|\mathcal{G}\right] \geq \mathbb{E}\left[X_n|\mathcal{G}\right] \quad \forall n$, with a limit by the very monotonicity:

$$\overline{X}_{candidate} = \lim_{n\to\infty} \mathbb{E}\left[X_n|\mathcal{G}\right]$$

We have measurability for free since $\overline{X}_{candidate}$ is the limit of measurable functions (Thm. A.16). By Definition 10.12#2 we further have that:

$$\int_G \mathbb{E}\left[X_n|\mathcal{G}\right] d\mathbb{P} = \int_G X_n d\mathbb{P} \quad \forall n \geq 1$$

Which leads to the following chain of equalities for arbitrary $G \in \mathcal{G}$:

$$\begin{aligned}
\lim_{n\to\infty}\int_G \mathbb{E}\left[X_n|\mathcal{G}\right] d\mathbb{P} &= \int_G \lim_{n\to\infty} \mathbb{E}\left[X_n|\mathcal{G}\right] d\mathbb{P} && \text{mon. conv. Thm. 4.21 for } \mathbb{E}\left[X_n|\mathcal{G}\right] \\
&= \int_G \overline{X}_{candidate} d\mathbb{P} \\
&= \int_G X d\mathbb{P} \\
&= \int_G \lim_{n\to\infty} X_n d\mathbb{P} && \text{hypothesis } X_n \nearrow X \\
&= \lim_{n\to\infty} \int_G X_n d\mathbb{P} && \text{reverse mon. conv.}
\end{aligned}$$

Where, the first and the last term are equal and squeeze the central ones. We conclude that necessarily it holds $\overline{X}_{candidate} = \mathbb{E}\left[X_n|\mathcal{G}\right] \nearrow \mathbb{E}\left[X|\mathcal{G}\right]$ almost surely. **(Claim #3)** similar to Claim #2. □

♣ **Theorem 10.21** (Towering property)**.** *A peculiar property of conditionals is:*

$$\mathcal{H} \subset \mathcal{G} \subset \mathcal{F} \quad \sigma\text{-algebras} \implies \mathbb{E}\left[\mathbb{E}[X|\mathcal{G}]\Big|\mathcal{H}\right] = \mathbb{E}\left[\mathbb{E}[X|\mathcal{H}]\Big|\mathcal{G}\right] = \mathbb{E}[X|\mathcal{H}] \quad a.s.$$

Proof. **(△ strategy)** denoting the triple equality as ① = ② = ③ we prove ① = ③, ② = ③ $\implies$ ① = ②.
(② = ③) Consider:

$$\mathbb{E}\left[\mathbb{E}[X|\mathcal{H}]\Big|\mathcal{G}\right]$$

There $\mathcal{H} \subset \mathcal{G}$ and $\mathbb{E}[X|\mathcal{H}]$ is $\mathcal{H}$-measurable by Definition 10.12#1. So:

$$\left(\mathbb{E}[X|\mathcal{H}]\right)^{-1}(A) = \{\omega \in \Omega : \mathbb{E}[X|\mathcal{H}](\omega) \in A\} \in \mathcal{H} \quad \forall A \in \mathcal{E} \overset{\mathcal{H}\subseteq\mathcal{G}}{\Longrightarrow} \left(\mathbb{E}[X|\mathcal{H}]\right)^{-1}(A) \in \mathcal{G} \forall A \in \mathcal{E}$$

So that $\mathbb{E}[X|\mathcal{H}]$ is $\mathcal{G}$-measurable and by Proposition 10.18#3 we get:

$$\mathbb{E}[X|\mathcal{H}] = \mathbb{E}\left[\mathbb{E}[X|\mathcal{H}]\Big|\mathcal{G}\right]$$

(① = ③) by definition of conditional expectation:

$$\mathbb{E}\left[\mathbb{E}[X|\mathcal{G}]\Big|\mathcal{H}\right] \iff \begin{cases} \int_H \mathbb{E}\left[\mathbb{E}[X|\mathcal{G}]\Big|\mathcal{H}\right] d\mathbb{P} = \int_H \mathbb{E}[X|\mathcal{G}] d\mathbb{P} & \forall H \in \mathcal{H} \\ \mathbb{E}\left[\mathbb{E}[X|\mathcal{G}]\Big|\mathcal{H}\right] & \mathcal{H}\text{-measurable} \end{cases}$$

Notice that $H \in \mathcal{H} \implies H \in \mathcal{G}$ so that:

$$\int_H \mathbb{E}\left[\mathbb{E}[X|\mathcal{G}]\middle|\mathcal{H}\right] d\mathbb{P} = \int_H X d\mathbb{P} \quad \forall H \in \mathcal{H} \qquad \text{Def. 10.12\#2 on } \mathbb{E}[X|\mathcal{G}]$$
$$= \int_H \mathbb{E}[X|\mathcal{H}] d\mathbb{P} \quad \forall H \in \mathcal{H} \qquad \text{Def. 10.12\#2 on } X$$

Since the arguments of the first and third integral are $\mathcal{H}$-measurable by definition 10.12#1 we conclude that:

$$\mathbb{E}\left[\mathbb{E}[X|\mathcal{G}]\middle|\mathcal{H}\right] \overset{a.s.}{=} \mathbb{E}[X|\mathcal{H}]$$

□

♢ **Observation 10.22** (About the equalities with conditionals). *Notice that all the = statements are almost sure $\overset{a.s.}{=}$ whenever we condition on $\mathcal{G} \subset \mathcal{F}$. Sometimes this will be omitted.*
Additionally, notice that the symbol $\subset$ does not necessarily mean strictly a subset, since in all the reasoning of the course there would be no distinction.

♣ **Proposition 10.23** (Properties of $\overline{X}_\mathcal{G}$ (II)). *Other important results are:*

1. *conditional determinism*

$$X\ \mathcal{G}\text{-measurable},\ \mathcal{G} \subset \mathcal{F},\ Y : YX \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}) \implies \mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}] \quad a.s.$$

2. *X ignores $\mathcal{F}$-sets to which it does not belong:*

$$\mathcal{H} \perp\!\!\!\perp \sigma(\sigma(X) \cup \mathcal{G}) \implies \mathbb{E}[X|\sigma(\mathcal{G} \cup \mathcal{H})] = \mathbb{E}[X|\mathcal{G}]$$

Proof. **(Claim #1)** Let Y, X be a.s. positive for simplicity. The general case follows.
Suppose $X \in \mathcal{G}_+$ (i.e. positive $\mathcal{G}$-measurable function) and $Y \in \mathcal{F}_+$. Then $\overline{Y}_\mathcal{G} = \mathbb{E}[Y|\mathcal{G}]$ is $\mathcal{G}$-measurable positive as well and:

$$\mathbb{E}[V \cdot (XY)] = \mathbb{E}[(V \cdot X)Y] = \mathbb{E}\left[(V \cdot X)\overline{Y}_\mathcal{G}\right] = \mathbb{E}\left[V \cdot (X\overline{Y}_\mathcal{G})\right]$$

Where the middle equality sign follows from an application of Proposition 10.10#2 after noticing that $V \cdot X \in \mathcal{G}_+$, and the definitional property of $\overline{Y}_\mathcal{G}$ to be equal to Y under the integral over each $\mathcal{G}$-set. Hence:

$$X\overline{Y}_\mathcal{G} = X\mathbb{E}[Y|\mathcal{G}]$$

is an almost sure version of $\mathbb{E}[XY|\mathcal{G}]$. □

♠ **Definition 10.24** (Specifying conditional probabilities). *Let $X = \mathbb{1}_A, A \in \mathcal{F}$, then:*

$$\mathbb{P}[A|\mathcal{G}] := \mathbb{E}[\mathbb{1}_A|\mathcal{G}]$$

Where by Definition 10.12

1. *$\mathbb{1}_A$ $\mathcal{G}$-measurable*
2. *$\forall G \in \mathcal{G}$ $\quad \int_G \mathbb{E}[\mathbb{1}_A|\mathcal{G}] d\mathbb{P} = \int_G \mathbb{1}_A d\mathbb{P} = \mathbb{P}[A \cap G]$*

Meaning that:

$$\mathbb{P}[A|\mathcal{G}] \ :\ \forall G \in \mathcal{G} \quad \int_G \mathbb{P}[A|\mathcal{G}] d\mathbb{P} = \mathbb{P}[A \cap G]$$

♣ **Proposition 10.25** (Trivial properties of conditional probability). *Notice that for Definition 10.24:*

1. *almost sure positivity $\mathbb{P}[A|\mathcal{G}] \geq 0$ a.s. $\forall A \in \mathcal{F}$*
2. *almost sure normalization $\mathbb{P}[\Omega|\mathcal{G}] = 1$ a.s.*
3. *almost sure countable additivity:*

$$\{A_n\}\ \text{pairwise disjoint } \mathcal{F}\text{-sets} \implies \mathbb{P}\left[\bigcup_{n\geq 1} A_n \middle| \mathcal{G}\right] = \sum_{n \geq 1} \mathbb{P}[A_n|\mathcal{G}]$$

Proof. All claims are trivial by the construction $\mathbb{P}[A|\mathcal{G}] := \mathbb{E}[\mathbb{1}_A|\mathcal{G}]$. □

♢ **Observation 10.26** (Link with usual definition)**.** *Definition 10.24 looks like* $\mathbb{P}[A|B] = \frac{\mathbb{P}[A\cap B]}{\mathbb{P}[B]}$

♡ **Example 10.27** (Linking old & new)**.** *For* $\mathcal{G} = \sigma(B) = \{\emptyset, \Omega, B, B^c\}$ *consider a function* $f : \Omega \to \mathbb{R}$. *Moreover:*

$$\omega \to f(\omega) = \mathbb{1}_B(\omega)\frac{\mathbb{P}[A\cap B]}{\mathbb{P}[B]} + \mathbb{1}_{B^c}(\omega)\frac{\mathbb{P}[A\cap B^c]}{\mathbb{P}[B^c]} \quad \forall B : \mathbb{P}[B] > 0$$

Now recognize that f is $\mathcal{G}$-measurable since:

$$\begin{cases} \int_B f d\mathbb{P} = \int_B \frac{\mathbb{P}[A\cap B]}{\mathbb{P}[B]} d\mathbb{P} = \mathbb{P}[A\cap B] \\ \int_{B^c} f d\mathbb{P} = \int_{B^c} \frac{\mathbb{P}[A\cap B^c]}{\mathbb{P}[B^c]} d\mathbb{P} = \mathbb{P}[A\cap B^c] \end{cases}$$

So that $f \overset{a.s.}{=} \mathbb{P}[A|\mathcal{G}]$ *by Definition 10.12#2.*
Contrarily, if $\mathbb{P}[B] = 0$ *we just replace the function with:*

$$\omega \to f(\omega) = c\mathbb{1}_B(\omega) + \mathbb{P}[A\cap B^c]\mathbb{1}_{B^c}(\omega)$$

and again

$$\begin{cases} \int_B f d\mathbb{P} = c\mathbb{P}[B] = 0 = \mathbb{P}[A\cap B] \\ \int_{B^c} f d\mathbb{P} = \mathbb{P}[A\cap B^c]\mathbb{P}[B^c] = \mathbb{P}[A\cap B^c] \end{cases}$$

Guaranteeing $f \overset{a.s.}{=} \mathbb{P}[A|\mathcal{G}]$.

♣ **Proposition 10.28** (Easy conditional probability properties)**.** *Other results that can be quickly recovered include:*

1. $\mathcal{G} = \mathcal{F} \implies \mathbb{P}[A|\mathcal{G}] = \mathbb{1}_A$ *a.s.*
2. $A \in \mathcal{F},\ A \perp\!\!\!\perp \mathcal{G} \implies \mathbb{P}[A|\mathcal{G}] = \mathbb{P}[A]$ *a.s.*

Proof. **(Claim #1)** let $\mathcal{G} = \mathcal{F}$ so that:

$$\mathbb{P}[A|\mathcal{G}] = \mathbb{P}[A|\mathcal{F}]\ \forall A \in \mathcal{F}$$

And the condition $\int_G \mathbb{1}_A d\mathbb{P} = \int_G \mathbb{P}[A|\mathcal{F}]d\mathbb{P}$, after noticing that $\mathbb{1}_A$ is $\mathcal{F}$-measurable ensures that:

$$\mathbb{P}[A|\mathcal{F}] = \mathbb{E}[\mathbb{1}_A|\mathcal{F}] \overset{a.s.}{=} \mathbb{1}_A \quad \text{Prop. 10.19\#1}$$

(Claim #2) By independence, probabilities decouple $A \perp\!\!\!\perp \mathcal{G} \implies \mathbb{P}[A\cap G] = \mathbb{P}[A]\mathbb{P}[G]\ \forall G \in \mathcal{G}$. Then:

$$\mathbb{P}[A\cap G] = \mathbb{P}[A]\mathbb{P}[G] = \int_G \mathbb{P}[A]d\mathbb{P} = \int_G \mathbb{P}[A|\mathcal{G}] \quad \forall G \in \mathcal{G} \implies \mathbb{P}[A] \overset{a.s.}{=} \mathbb{P}[A|\mathcal{G}]d\mathbb{P}$$

□

♢ **Observation 10.29** (Is $\mathbb{P}(\cdot|\mathcal{G})$ a probability measure?)**.** *Is* $A \to \mathbb{P}[A|\mathcal{G}]$ *a p.m. on* $(\Omega, \mathcal{F})$*?*
Proposition 10.25 suggests yes but almost surely.
For this reason, we consider $N \in \mathcal{F} : \mathbb{P}[N] = 0$ *and state that* $\forall \omega \in N^c \quad \mathbb{P}[\cdot|\mathcal{G}](\omega)$ *is a p.m. on* $(\Omega, \mathcal{F})$*. This allows to define:*

$$\mathcal{N} = \{A \in \mathcal{F} : \ either\ 1, 2, 3\ Prop.10.25\ fail\ on\ A\}$$

Such set may be uncountable, which would mean that $\bigcup_{A\in\mathcal{N}} A \notin \mathcal{F}$ *or* $\mathbb{P}\left[\bigcup_{A\in\mathcal{N}} A\right] \leq 0$*, and we wish to* ***avoid these pathological cases****, even though they* ***almost never happen****.*

♣ **Theorem 10.30** (Avoidance of pathological negligible sets)**.** *For Y a r.v. on* $(\Omega, \mathcal{F}, \mathbb{P})$ *taking values on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *and* $\mathbb{P}_Y[B|\mathcal{G}] = \mathbb{P}[Y^{-1}(B)|\mathcal{G}] = \mathbb{P}[\{\omega \in \Omega : Y(\omega) \in B\}|\mathcal{G}]\ \forall B \in \mathcal{B}(\mathbb{R})$ *we have that there exists* $\exists Q_{\mathcal{G}}(B;\omega)$ *such that:*

1. $\omega \to Q_{\mathcal{G}}(B;\omega)$ *measurable* $\forall B \in \mathcal{B}(\mathbb{R})$
2. $B \to Q_{\mathcal{G}}(B;\omega)$ *is a p.m. on* $(\mathbb{R}, \mathcal{B}(\mathbb{R})) \quad \forall \omega \in \Omega$

3. $Q_{\mathcal{G}}(B;\omega) = \mathbb{P}_Y[B|\mathcal{G}]$ *a.s.* $\forall B \in \mathcal{B}(\mathbb{R})$ *which holds if and only if:*

$$\int_A Q_{\mathcal{G}}(B;\omega)\mathbb{P}[d\omega] = \mathbb{P}[Y^{-1}(B) \cap A] \quad \forall A \in \mathcal{F},\ \forall B \in \mathcal{B}(\mathbb{R})$$

4. $Q_{\mathcal{G}}(\cdot;\omega)$ *is the conditional distribution of* Y *given* $\mathcal{G}$

Corollary 10.31 (Enlarging Theorem 10.30)**.** *We can state the same results also if Y is on a complete (Def. 9.27) separable metric space $(\mathbb{Y}, \mathcal{B}(\mathbb{Y}))$. We do not expand on this topic, but hint that it is a complete space with distinguishable elements and a metric.*

10.2 The infinite dimensional case for stochastic processes

♠ **Definition 10.32** (Infinite countable sequence of random variables X^∞)**.** *We define $X^\infty = (X_n)_{n\geq 1}$ where $X^\infty : \Omega \to \mathbb{R}^\infty$ and:*

1. *measurability holds:*
$$\{\omega \in \Omega : X^\infty(\omega) \in B\} \in \mathcal{F} \quad \forall B \in \mathcal{B}(\mathbb{R}^\infty)$$
2. *probability*
$$\mathcal{P}_{X^\infty}(B) = \mathbb{P}\left[\{\omega \in \Omega : X^\infty(\omega) \in B\}\right] \quad \forall B \in \mathcal{B}(\mathbb{R}^\infty)$$

♢ **Observation 10.33** (Characterizing X^∞)**.** *We aim to find a way to construct a valid probability measure on $\mathbb{R}^\infty$ without starting from the abstract space $(\Omega, \mathcal{F}, \mathbb{P})$. While for finite dimensions $N < \infty$ we could use the cdf on $\mathbb{R}^N$ and Theorems 3.21, 3.22, in the case of ∞ dimensions we make use of* ***projections****.*

♠ **Definition 10.34** (Cylinders with finite dimensional base $\mathcal{C}_n(\cdot)$)**.** *Fix the first n coordinates, and make a projection with respect to this finite dimensional set of $\mathbb{R}^\infty$:*

$$\mathcal{C}_n(B) = \{x = (x_1, \ldots) \in \mathbb{R}^\infty : (x_1, \ldots, x_n) \in B\} \quad \forall n \geq 1, \forall B \in \mathcal{B}(\mathbb{R}^n)$$

Lemma 10.35 (Generating set of infinite Borel set)**.**

$$\mathcal{B}(\mathbb{R}^\infty) = \sigma\left(\bigcup_{n\geq 1} \{\mathcal{C}_n(B) : B \in \mathcal{B}(\mathbb{R}^n)\}\right)$$

In other words, the generating σ-algebra is composed of measurable finite base cylinders.

Lemma 10.36 (Induced finite probability measure by $\mathcal{P}$ on the cylinder)**.** *for $\mathcal{P}$ a p.m. on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ it holds that:*

1. $\forall B \in \mathcal{B}(\mathbb{R}^n) \quad \mathcal{P}_n(B) = \mathcal{P}(\mathcal{C}_n(B))$ *induced is a p.m. on* $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \quad \forall n$
2. $\mathcal{P}$ *induces a sequence of p.m.s* $(\mathcal{P}_n)_{n\geq 1}$

Proof. **(Claim #1)** a map of the form $B \to \mathcal{P}_n(B) = \mathcal{P}(\mathcal{C}_n(B))$ for all $B \in \mathcal{B}(\mathbb{R}^n)$ is always a valid p.m.
(Claim #2) let $B \in \mathcal{B}(\mathbb{R}^n)$ then:

$$\begin{aligned}
\mathcal{P}_{n+1}(B \times \mathbb{R}) &= \mathcal{P}(\mathcal{C}_{n+1}(B \times \mathbb{R})) && \text{Claim \#1}\\
&= \mathcal{P}\left(\{x \in \mathbb{R}^\infty : (x_1, \ldots, x_n) \in B, x_{n+1} \in \mathbb{R}\}\right) && \text{Def. 10.34}\\
&= \mathcal{P}\left(\{x \in \mathbb{R}^\infty : (x_1, \ldots, x_n) \in B\}\right) && x_{n+1} \text{ is arbitrary}\\
&= \mathcal{P}(\mathcal{C}_n(B))\\
&= \mathcal{P}_n(B) && \text{Claim \#1}
\end{aligned}$$

□

♠ **Definition 10.37** (Kolmogorov Consistency)**.** *For $(\mathcal{P}_n)_{n\geq 1}$ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ such that $\mathcal{P}_{n+1}(B\times\mathbb{R}) = \mathcal{P}_n(B) \quad \forall B \in \mathcal{B}(\mathbb{R}^n)\ \forall n \geq 1$.*
Namely, we have that the second condition of Lemma 10.36 is satisfied by construction on $\mathcal{C}_n(B)$.

♢ **Observation 10.38** (About Definition 10.37 and Lemma 10.36). *The question we try to answer is if starting from this consistency condition is enough to define a probability measure at the limit that is non ambiguous.*

♣ **Theorem 10.39** (Kolmogorov Extension Theorem). *Consider a sequence of r.v.s $(X_n)_{n\in\mathbb{N}}$ on a standard measurable space $(E,\mathcal{E})$ (Def. A.22). Then:*

$$(\mathcal{P}_n)_{n\geq 1} \text{ consistent (Def. 10.37)} \implies \exists!\mathcal{P} \text{ on } \left(\bigtimes_n E_n, \bigotimes_n \mathcal{E}_n\right)$$

$$\text{such that} \quad \mathcal{P}(\mathcal{C}_n(B)) = \mathcal{P}_n(B) \quad \forall B \in \bigotimes_{i=1}^{n} \mathcal{E}_n, \forall n \geq 1$$

♢ **Observation 10.40** (About the hypothesis). *We assumed the space to be standard (Def. A.22). Thank to this, we can include in the result any space which is equal to a Euclidean space up to isomorphisms, thus the discussion about cylinders in Euclidean spaces. In the more general setting of a measurable space the statement becomes existance only.*

♢ **Observation 10.41** (Advantage of the extension). *Now $\mathcal{P}_n(B) = \mathbb{P}[(x_1,\dots,x_n)\in B]$ is useful and a shared consistency for all n means that we work with a unique measure in the background. This sufficient condition is nice and there are many laws that obey it.*

♡ **Example 10.42** (Exchangeable sequence). *Consider a sequence $(X_n)_{n\in\mathbb{N}}$ in $\{0,1\}^\infty$ and a probability measure Q on $([0,1],\mathcal{B}([0,1]))$. Assume further that $\forall n\geq 1$ $\mathcal{P}_n$ is a probability law on $\{0,1\}^\infty$ defined as:*

$$\mathcal{P}_n(\{x_1,\dots,x_n\}) = \int_{[0,1]} \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\mathbb{1}_{\{0,1\}^n}(x_1,\dots,x_n)\,Q(d\theta)$$

The sequence $(\mathcal{P}_n)_{n\in\mathbb{N}}$ is consistent in the sense of Definition 10.37 and by Theorem 10.39 we have:

$$\exists!\mathcal{P} \quad \text{on } \left(\{0,1\}^\infty, \mathcal{B}\left(\{0,1\}^\infty\right)\right) : \mathcal{P}(\mathcal{C}_n(B)) = \mathcal{P}_n(B) \quad \forall n\geq 1, \forall B\in\mathcal{B}(\{0,1\}^n)$$

Now notice that the elements of $(X_n)_{n\in\mathbb{N}}$ are conditionally on θ independent. Namely, specifying θ they are independent. The law $\mathcal{P}$ is said to be exchangeable since:

$$(X_1,\dots,X_n) \overset{d}{=} (X_{\pi(1)},\dots,X_{\pi(n)}) \quad \forall\pi\in\{permutations\}$$

And the dependence structure is:

$$\begin{cases} \mathbb{P}[X_1 = x_1,\dots,X_n = x_n | \theta = \widetilde{\theta}] = \widetilde{\theta}^{\sum x_i}(1-\widetilde{\theta})^{n-\sum x_i} \\ \widetilde{\theta} \sim Q \end{cases}$$

♡ **Example 10.43** (The iid case). *Let $\Omega = \{0,1\}^\infty, \theta\in(0,1)$ and:*

$$\mathcal{P}_n(\{(x_1,\dots,x_n\}) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\mathbb{1}_{\{0,1\}^n}(x_1,\dots,x_n)$$

We know $\mathcal{P}_n$ is a p.m., and wish to check its consistency:

$$\begin{aligned} \mathcal{P}_{n+1}(A\times\{0,1\}) &= \sum_{(x_1,\dots,x_{n+1})\in A\times\{0,1\}} \theta^{\sum^{n+1} x_i}(1-\theta)^{n+1-\sum^{n+1} x_i} \\ &= \sum_A \theta^{\sum^n x_i}(1-\theta)^{n-\sum x_i} \sum_{x_{n+1}=0, x_{n+1}=1} \theta^{x_{n+1}}(1-\theta)^{1-x_{n+1}} \\ &= \mathcal{P}_n(A) \qquad\qquad \forall A\in\mathcal{B}(\{0,1\}^n), \forall n \end{aligned}$$

Thanks to this consistency, we apply Kolmogorov's extension Thm. 10.39 to conclude:

$$\exists!\mathcal{P} \text{ on } \{0,1\}^\infty \quad : \quad \mathcal{P}(\mathcal{C}_n(A)) = \mathcal{P}_n(A) \quad \forall A\in\mathcal{B}(\{0,1\}^n), \forall n$$

♠ **Definition 10.44** (Stochastic Matrix $\mathbf{\Gamma}$). *For $E = \{\xi_1,\dots,\xi_N\}$ and $\mathcal{E} = Pow(E) = 2^E$, a stochastic matrix is a matrix $\boldsymbol{\Gamma}$:*

$$\mathbb{R}^N\times\mathbb{R}^N \ni \boldsymbol{\Gamma} = \begin{bmatrix} p_{1,1} & \dots & p_{1,N} \\ \vdots & \vdots & \vdots \\ p_{N,1} & \dots & p_{N,N} \end{bmatrix} \quad \begin{cases} p_{ij}\geq 0 & \forall i,j \\ \sum_j p_{ij} = 1 & \forall j \end{cases}$$

♠ **Definition 10.45** (Probability sequence in $(E, \mathcal{E})$). *Let π be a p.m. on E such that $\pi(\xi_i) = \pi_i$ and $\vec{\pi} = (\pi_1, \dots, \pi_N)^{\mathbb{T}}$. Then, $(\mathcal{P}_n)_{n\geq 1}$ is such that $\mathcal{P}_n$ is a p.m. on $\bigtimes_{i=1}^{n} E\ \forall n$ where:*

$$\begin{cases} \mathcal{P}_1(\{\xi_i\}) = \pi_i & \forall i \\ \mathcal{P}_n(\{(\xi_{i_1}, \dots, \xi_{i_n})\}) = \pi_{i_1} p_{i_1,i_2} \cdot \dots \cdot p_{i_{n-1},i_n} \end{cases}$$

For $n \geq 2$ and $\{i_1, \dots, i_n\} \in \{1, \dots, N\}^n$

Lemma 10.46 (Applying Kolmogorov Extension Theorem). *Using Theorem 10.39 we conclude that for a sequence of probabilities $(\mathcal{P}_n)_{n\geq 1}$ as in Definition 10.45 it holds that:*

1. *positivity $\mathcal{P}_n \geq 0 \forall n$*
2. *normalization*

$$\sum_{(i_1,\dots,i_N)\in\{1,\dots,N\}^n} \mathcal{P}_n(\{\xi_{i_1}, \dots, \xi_{i_n})\}) = 1$$

3. *consistency is satisfied (Def. 10.37)*
   $\forall A \in \mathcal{E}^n$ let $D_{n,A} = \{(i_1, \dots, i_N) : (\xi_{i_1}, \dots, \xi_{i_n}) \in A\}$ then

$$\mathcal{P}_{n+1}(A \times E) = \sum_{\vec{i}\in D_{n,A}} \sum_{i_{n+1}=1}^{N} \pi_{i_1} p_{i_1,i_2} \cdot \dots \cdot \pi_{i_n,i_{n+1}} = \mathcal{P}_n(A)$$

4. *(Kolmogorov's Theorem)*

$$\exists ! \mathcal{P} \text{ on } \left(\bigtimes_n E_n, \bigotimes_n \mathcal{E}n \right) \quad \text{such that} \quad \mathcal{P}(\mathcal{C}_n(A)) = \mathcal{P}_n(A)\ \forall A \in \bigotimes_{i=1}^{n} \mathcal{E}_i$$

5. *(Direct implication of 4)*

$$\exists (X_n)_{n\geq 1} : \Omega \to E^{\infty}\ :\ \mathbb{P}[X_1 = \xi_{i_1}, \dots, X_n = \xi_{i_n}] = \mathcal{P}_n(\{\xi_{i_1}, \dots, \xi_{i_n}\})$$

Proof. **(Claims #1, #2)** follows by $\vec{\pi}, \mathbf{\Gamma}$ construction where $\mathbf{\Gamma} = \{p_{ij}\}_{i=1,\dots,N, j=1,\dots,N}$
(Claim #3) algebra
(Claims #4,#5) just an application of the Theorem and a direct implication of it. □

♢ **Observation 10.47** (Interpreting Lemma 10.46). *We have that:*

$$\underbrace{\pi_i = \mathbb{P}[X_1 = \xi_i]}_{initial\ state}\ \forall i \quad \underbrace{p_{ij}^n = \mathbb{P}[X_n = \xi_j | X_{n-1} = \xi_i]}_{one\ step\ transitions}\ \forall n \geq 2, \forall i, j$$

And $(X_n)_{n\geq 1}$ is the classic Markov Chain!

♣ **Proposition 10.48** (Markov property). *For a Markov Chain it holds:*

$$\forall n, i \quad \mathbb{P}[X_{n+1} = \xi_{i_{n+1}} | X_1 = \xi_{i_1}, \dots, X_n = \xi_{i_n}] = \mathbb{P}[X_{n+1} = \xi_{i_{n+1}} | X_n = \xi_{i_n}]$$

Proof. just observe that:

$$\mathbb{P}[X_{n+1} = \xi_{i,n+1} | (X_1, \dots, X_n) = \vec{\xi_i}] = \frac{\pi_{i_1} p_{i_1,i_2} \cdots p_{i_n,i_{n+1}}}{\pi_{i_1} p_{i_1,i_2} \cdots p_{i_{n-1},i_n}} = p_{i_n,i_{n+1}}$$

□

♣ **Theorem 10.49** (Chapman Kolmogorov equations). *For a Markov Chain $(X_n)_{n\geq 1}$ with stochastic matrix (Def. 10.44) $\mathbf{\Gamma}$ we have that:*

$$\vec{q}_n = \begin{bmatrix} \mathbb{P}[X_n = \xi_1] \\ \vdots \\ \mathbb{P}[X_n = \xi_N] \end{bmatrix} \quad \text{is such that} \quad \begin{cases} \vec{q}_{n+1} = \mathbf{\Gamma} \vec{q}_n & \forall n \\ \mathbb{P}[X_{n+1} = \xi_j] = \sum_{\ell=1}^{N} p_{\ell,j} \mathbb{P}[X_n = \xi_\ell] & \forall j \end{cases}$$

Proof. For arbitrary n, i, j it holds:

$$p_{ij}^n = \mathbb{P}[X_{n+1} = \xi_j | X_1 = \xi_1] = \mathbb{P}[X_{n+k} = \xi_j | X_k = \xi_k] \quad \forall k \geq 2, k \in \mathbb{N}$$

Which we could also express equivalently

$$p_{ij}^n = \sum_I p_{i,i_2} \cdots p_{i_n,j} \quad I = \left\{(i_2, \ldots, i_n) \in \{1, \ldots, N\}^{n-1}\right\}$$

The last form induces a recursive representation:

$$\mathbb{P}[X_{n+1} = \xi_j | X_n = \xi_i] = \sum_{\ell=1}^{N} p_{i\ell} p_{\ell j}$$

And iterating:

$$\mathbb{P}[X_{n=1} = \xi_j | X_1 = \xi_i] = \sum_{\ell=1}^{n} p_{i,\ell}^n p_{\ell,j} \implies \mathbb{P}[X_{n=1} = \xi_j] = \sum_{\ell=1}^{n} p_{\ell,j} \mathbb{P}[X_n = \xi_\ell]$$

Which is equivalent to:

$$\vec{\mathbf{q}}_{n+1} = \mathbf{\Gamma}^T \vec{\mathbf{q}}_n$$

□

♠ **Definition 10.50** (The $\left(\mathbb{R}^{\mathbb{T}}, \mathcal{B}\left(\mathbb{R}^{\mathbb{T}}\right)\right)$ space)**.** *Let $\mathbb{T}$ be an arbitrary index, typically $\mathbb{T} \subset \mathbb{R}$. Then we define the tuple $\left(\mathbb{R}^{\mathbb{T}}, \mathcal{B}\left(\mathbb{R}^{\mathbb{T}}\right)\right)$ as always recalling that:*

$$\mathbb{R}^{\mathbb{T}} = \prod_{t \in \mathbb{T}} \mathbb{R} = \{f : \mathbb{T} \to \mathbb{R}\}$$

Such a space is potentially ***uncountably infinite****.*

♠ **Definition 10.51** (Class of cylinders with measurable base $\mathcal{C}_T(\cdot)$)**.** *Just like in Definition 10.34 we let:*

$$C_{\vec{t}}(B) = \left\{(x_t)_{t \in \mathbb{T}} : (x_{t_1}, \ldots, x_{t_n}) \in B,\ \vec{t} = (t_1, \ldots, t_n)\right\} \quad \forall t \in \mathbb{T}^n, \forall n \geq 1, \forall B \in \mathcal{B}(\mathbb{R}^n)$$

$$\mathcal{C}_{\mathbb{T}} = \left\{\mathcal{C}_{\vec{t}}(B)\ B \in \mathcal{B}(\mathbb{R}^n), \forall n \quad where \quad \vec{t} = (t_1, \ldots, t_n) : t_k \in \mathbb{T},\ |\vec{t}| = n\right\}$$

Lemma 10.52 (Generating set of ∞ Borel uncountable is countable)**.** *Just like Lemma 10.35*

$$\sigma(\mathcal{C}_{\mathbb{T}}) = \mathcal{B}(\mathbb{R}^{\mathbb{T}}) \quad \forall \mathbb{T}$$

♠ **Definition 10.53** (Class of consistent Probability distributions $\mathscr{P}$)**.**

$$\mathscr{P} = \{\mathcal{P}_{t_1, \ldots, t_n} : n \geq 1,\ \ t_1, \ldots t_n \in \mathbb{T}\}$$

where $\mathcal{P}_{t_1,\ldots,t_n}$ satisfies the consistency condition of Definition 10.54.

♠ **Definition 10.54** (Kolmogorov Consistency)**.** *Like in Definition 10.37 conclude that the fundamental condition is:*

$$\begin{aligned} &\forall n \geq 1,\ t_1, \ldots, t_n \in \mathbb{T} \quad t_1', \ldots, t_k' \in \{t_1, \ldots, t_n\} \quad : \quad \mathcal{C}_{t_1,\ldots,t_n}(B_n) = \mathcal{C}_{t_1',\ldots,t_k'}(B_k) \\ &\textit{it holds } \mathcal{P}_{t_1,\ldots,t_n}(B_n) = \mathcal{P}_{t_1',\ldots,t_k'}(B_k) \end{aligned}$$

We are enforcing equivalence of probability laws by the projections up to permutations of the indexes, aiming to see if this is a sufficient condition for having a unique probability measure in the background.

♣ **Theorem 10.55** (Kolmogorov Extension Theorem, uncountable version)**.** *Consider a sequence of r.v.s $(X_n)_{n \in \mathbb{N}}$ on a standard measurable space $(E, \mathcal{E})$ (Def. A.22). Then, the uncountable versions for spaces as in Definition 10.50 of Theorem 10.39 is stated as follows.*
A class $\mathscr{P}$ as in Def. 10.53 satisfying the consistency version of Def. 10.54 is such that:

1. *unique infinite law*

$$\exists ! \mathcal{P} \text{ on } \sigma(\mathcal{C}_{\mathbb{T}}) = \mathcal{B}(\mathbb{R}^{\mathbb{T}}) \quad \forall B_n \in \mathcal{B}(\mathbb{R}^n), \forall n \geq 1, \ \forall t_1, \dots, t_n \in \mathbb{T} \quad \text{such that } \mathcal{P}(\mathcal{C}_{t_1,\dots,t_n}(B_n)) = \mathcal{P}_{t_1,\dots,t_n}(B_n)$$

2. *unique stochastic process*

$$\exists ! \text{stochastic process } (X_t)_{t\in\mathbb{T}} \quad : \quad \mathbb{P}[(X_{t_1}, \dots, X_{t_n}) \in B] = \mathcal{P}_{t_1,\dots,t_n}(B)$$

♢ **Observation 10.56** (About the hypothesis). *We assumed the space to be standard (Def. A.22). In the more general setting of a measurable space the statement becomes existance only.*

♣ **Theorem 10.57** (Ionescu-Tulcea Theorem). *Another important result is that of the uniqueness of the law on Ω given some almost always verified conditions. For measurable spaces $(E_n, \mathcal{E}_n)_{n\in\mathbb{N}}$ consider*

- *a p.m. μ on $(E_0, \mathcal{E}_0)$*
- *Transition Kernels K_n for $n = 1, \dots$ where $\forall n \in \mathbb{N}^*$ we have:*

$$K_{n+1} : \left(F_n^0, \mathcal{F}_n^0\right) = (E_0 \times \cdots \times E_n, \mathcal{E}_0 \otimes \cdots \otimes \mathcal{E}_n) \to (E_{n+1}, \mathcal{E}_{n+1})$$

Denoting:

- *the current space:*

$$(F_n^0, \mathcal{F}_n^0) = \left(\bigtimes_0^n E_t, \bigotimes_0^n \mathcal{E}_t \right)$$

- *the total space:*

$$(\Omega, \mathcal{F}) = \bigotimes_{n\in\mathbb{N}} (E_n, \mathcal{E}_n)$$

- *the n^{th} outcome*

$$\omega = (x_n)_{n\in\mathbb{N}} \in \Omega : \quad X_n : \Omega \to E_n \quad X_n(\omega) = x_n$$

- *the chain of outcomes up to the n^{th}*

$$Y_n = (X_0, \dots, X_n) : \Omega \to F_n^0$$

Then, the distribution of Y_n is given by the concatenation of the kernels:

$$\pi_n(dx_0, \dots, dx_n) = \mu(dx_0) K_1(x_0; dx_1) \cdots K_n(x_{n-1}; dx_n)$$

where we identify cylinders with base $B \in \mathcal{F}_n^0$ by the following principle:

$$\mathcal{F}_n = \sigma\left((Y_n)_{n\in\mathbb{N}}\right) \quad s.t. \quad \mathcal{F}_n \ni H = \{Y_n \in B\} = B \times E_{n+1} \times \cdots$$

Looking for a law $\mathbb{P}$ on the infinite space $(\Omega, \mathcal{F})$, we find that:

$$\exists ! \mathbb{P} \quad s.t. \quad \mathbb{P}[H] = \pi_n(B) \quad \forall H \in \mathcal{F}_n, \text{ base } B \in \mathcal{F}_n^0$$

Where the regularity condition is sufficient for uniqueness, and reasonable to ensure non ambiguity at the n^{th} trial.
For a more articulated argument, see [Çin11](VI.4).

We are now in the position to loosely define a stochastic process.

♠ **Definition 10.58** (Stochastic process). *For a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a measurable space $(E, \mathcal{E})$ a stochastic process is a collection of r.v.s indexed by an arbitrary index $\mathbb{T}$. If we can endow $\mathbb{T}$ with an order we will have a sequence:*

$$(X_t)_{t\in\mathbb{T}} \quad X_t : \Omega \to E \quad \forall t \in \mathbb{T}$$

*The most common case is $\mathbb{T} \subset \mathbb{R}$ where we interpret the index as **Time or Position**.*

♠ **Definition 10.59** (Continuous time Markov Process). *Let $T = (0, \infty)$ and π be a p.m. on $\mathbb{R}$. For $(s, t) \in T$ with $s < t$ define the Markov transition kernels on $\mathbb{R} \times \mathcal{B}(\mathbb{R})$ (Def. B.13, with weight 1) with the symbols $P_{s,t}$. We have by definition:*

- *$x \to Ps, t(x, A)$ is $\mathcal{B}(\mathbb{R})$ measurable for any $A \in \mathcal{B}(\mathbb{R})$*
- *$A \to P_{s,t}(x, A)$ is a p.m. on $\mathbb{R}$ for any $x \in \mathbb{R}$*

For distinct ordered times $0 = t_0 < t_1 < \cdots < t_n$ we set:

$$\mathcal{P}_{t_0=0,t_1,\ldots,t_n}(dx_0, \ldots, dx_n) = \pi(dx_0)P_{0,t_1}(x_0, dx_1)\cdots P_{t_{n-1},t_n}(x_{n-1}, dx_n)$$

Where the transition probabilities satisfy the Chapman-Kolmogorov equation just like the discrete case (Thm. 10.49):

$$P_{s,u}(x, B) = \int_{\mathbb{R}} P_{s,t}(x, dy)P_{t,u}(y, B)$$

The family

$$\mathscr{P} = \{\mathcal{P}_{t_1,\ldots,t_n} : n \geq 1,\ t_1, \ldots, t_n \in (0, \infty) = T\}$$

is consistent in the sense of Definition 10.54, and by Kolmogorov extension (Thm. 10.55) corresponds to a unique p.m. $\mathcal{P}$ on $\left(\mathbb{R}^{(0,\infty)}, \mathcal{B}\left(\mathbb{R}^{(0,\infty)}\right)\right)$.
Moreover, on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ there exists a process (Def. 10.58) $(X_t)_{t\in\mathbb{R}_+}$ with law $\mathcal{P}$ such that:

- *for any $A \in \mathcal{B}(\mathbb{R})$*

$$\mathbb{P}[X_0 \in A] = \mathcal{P}\left(\left\{(x_t)_{t\in\mathbb{R}_+} : x_0 \in A\right\}\right) = \pi(A)$$

- *for the algebra generated by the past items of the process, namely $\mathcal{F}_t = \sigma(\{X_s : 0 \leq s \leq t\})$ for positive $t \in \mathbb{R}_+$ is holds for any $s < t$ that:*

$$\mathbb{P}[X_t \in A|\mathcal{F}_s] = \mathbb{P}[X_t \in A|X_s] = P_{s,t}(X_s, A) \quad \forall A \in \mathcal{B}(\mathbb{R})$$

We call such stochastic process with state space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with transition kernels $(P_{s,t})_{0\leq s<t}$ a ***Markov process****.*

♠ **Definition 10.60** (Homogeneous Markov Process)**.** *A Markov Process as in the above definition with $\mathcal{P}_{s,t} = \mathcal{P}t - s$. That is, the kernels depend only on the difference between the times considered.*

♠ **Definition 10.61** (Gaussian process)**.** *For functions*

- *$m : \mathbb{T} \to \mathbb{R}$*
- *$k : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$ is a positive definite function. That is, $\forall p \geq 1$, all $\boldsymbol{c} \in \mathbb{R}^p$ and $t_1, \ldots, t_p \in \mathbb{T}$ it holds:*

$$\sum_{i,j} c_i c_j k(t_j, t_j) > 0$$

A stochastic process $(X_t)_{t\in\mathbb{T}}$ is Gaussian and we write $(X_t)_{t\in\mathbb{T}} \sim \mathcal{G}P(m, k)$ if:
$\forall n \geq 1, \boldsymbol{t} = (t_1, \ldots, t_n) \subset \mathbb{T}$ the vector:

$$\boldsymbol{X} = (X_{t_1}, \ldots, X_{t_n}) \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{K}) \quad \boldsymbol{m} = \boldsymbol{m}(t) = \begin{bmatrix} m(t_1) \\ \vdots \\ m(t_n) \end{bmatrix} \quad \boldsymbol{K} = \boldsymbol{K}(\boldsymbol{t}, \boldsymbol{t}) = \begin{bmatrix} k(t_1, t_1) & k(t_1, t_2) & \cdots & k(t_1, t_n) \\ k(t_2, t_1) & k(t_2, t_2) & \cdots & k(t_2, t_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(t_n, t_1) & k(t_n, t_2) & \cdots & k(t_n, t_n) \end{bmatrix}$$

which means that for any finite collection of time points the vector arising from the process is a multivariate normal with parameters depending on the function specified. The positive definiteness of the function is enforced to ensure that the covariance matrix is well defined.

Chapter Summary

Objects:

- conditional expectation $\overline{X}_{\mathcal{G}}$ for $X \in \mathcal{L}_1$ at least and $\mathcal{G} \subset \mathcal{F}$:
  - $\mathcal{G}$-measurability
  - $\mathbb{E}[\mathbb{1}_G \overline{X}_{\mathcal{G}}] = \mathbb{E}[\mathbb{1}_G X] \iff \int_G \overline{X}_{\mathcal{G}} d\mathbb{P} = \int_G X d\mathbb{P}$ for all $G \in \mathcal{G}$
- conditional probability as $\mathbb{P}[A|\mathcal{G}] = \mathbb{E}[\mathbb{1}_A|\mathcal{G}]$
- infinite sequences, infinite countable base cylinders in $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$
- Kolmogorov Consistency in $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ as projection consistency
- infinite sequences and infinite countable base cylinders in $\left(\mathbb{R}^{\mathbb{T}}, \mathcal{B}\left(\mathbb{R}^{\mathbb{T}}\right)\right)$
- Kolmogorov consistency in $\left(\mathbb{R}^{\mathbb{T}}, \mathcal{B}\left(\mathbb{R}^{\mathbb{T}}\right)\right)$ is projection and permutations
- a Stochastic process as a collection of r.v.s on $(\Omega, \mathcal{F}, \mathbb{P})$ indexed by $\mathbb{T}$
- Gaussian process, uniquely identified by functions m, k

Results:

- conditional expectation
  - uniqueness almost surely
  - given $X \in \mathcal{L}_p(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{P}[X \geq 0] = 1$, $\mathcal{G} \subset \mathcal{F}$ existance and a.s. equivalence to Radon Nikodym derivative of:
$$\exists \overline{X}_{\mathcal{G}} = \frac{dQ}{d\mathbb{P}_{\mathcal{G}}} \quad where \quad \begin{cases} G \to Q(G) = \int_G X d\mathbb{P} \\ \mathbb{P}_{\mathcal{G}}[G] = \mathbb{P}[G] \ \forall G \in \mathcal{G} \end{cases}$$
  - orthogonal projection decomposition as $X = \overline{X}_{\mathcal{G}} + \widetilde{X}$ where $X - \overline{X}_{\mathcal{G}} \perp\!\!\!\perp \mathcal{G}$
  - no information, full information, iterated expectation
  - measurability implies full knowledge
  - linearity, monotonicity, monotone convergence, dominated convergence
  - towering property:
$$\mathcal{H} \subset \mathcal{G} \subset \mathcal{F} \quad \sigma\text{-algebras} \implies \mathbb{E}\left[\mathbb{E}[X|\mathcal{G}]\middle|\mathcal{H}\right] = \mathbb{E}\left[\mathbb{E}[X|\mathcal{H}]\middle|\mathcal{G}\right] = \mathbb{E}[X|\mathcal{H}] \quad a.s.$$
  - conditional determinism
  - conditional probability is an almost sure (random) probability measure
- processes construction:
  - Kolmogorov extension on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ to unique $\mathbb{P}$ on standard measurable spaces
  - Kolmogorov extension on $\left(\mathbb{R}^{\mathbb{T}}, \mathcal{B}\left(\mathbb{R}^{\mathbb{T}}\right)\right)$ to unique $\mathbb{P}$ on standard measurable spaces
  - Ionescu-Tulcea Theorem
  - in simple words, we can always refer to a unique probability measure at the infinite space when talking about processes during the course

Part II

Stochastic Processes

Chapter 11

Martingales & Stopping Times

11.1 Filtrations, Stopping times and easy notions

♠ **Definition 11.1** (Background notation). *Assume that we are now in a probability space (Def. 10.45)* $(\Omega, \mathcal{H}, \mathbb{P})$*. We will index sequences by a countable collection* $\mathbb{N} = \{0, 1, 2, \ldots\}$ *or an uncountable collection such as* $\mathbb{R}_+$ *or a generic* $\mathbb{T}$*.*
A stochastic process will be an indexed sequence of random variables (Def. 10.58).
Occasionally, we might denote measurable functions over a measurable space $(E, \mathcal{E})$ *with the symbol* $f \in \mathcal{E}$ *and accordingly with signs* $\pm$ *for negative and positive cases.*

♠ **Definition 11.2** (Filtration). *For an index set* $\mathbb{T}$ *a filtration is a sequence* $(\mathcal{F}_t)_{t \in \mathbb{T}}$ *such that:*

1. $\mathcal{F}_t \subset \mathcal{H}\ \forall t$ *and* $\mathcal{F}_t$ *is a* σ*-algebra (Def. 1.6)* $\forall t$
2. $\mathcal{F}_s \subset \mathcal{F}_t\ \forall s < t$

Intuitively, it is a sequence of increasing information in the probability space.

♠ **Definition 11.3** (Filtration generated by a random variable). *Given a stochastic process* $(X_t)_{t \in \mathbb{T}}$ *the filtration generated by it is denoted as:*

$$\mathcal{F}_t = \sigma(\{X_s : s \leq t\})$$

Which can be seen as a flow of information accumulated at each time point.

♢ **Observation 11.4** (About Definition 11.3). *Recall that a* σ*-algebra is a synonym of information, any random variable generating it is clearly measurable with respect to it. For an intuition, come back to Theorem 8.2.*
For this reason, $\mathcal{F}_t$ *could be seen as the collection of random variables* V *such that* $\forall \omega \in \Omega\ V(\omega)$ *is known* ***at latest by time*** $t \in \mathbb{T}$*.*

♡ **Example 11.5** (A random experiment). *Let* $\Omega = \{\omega = (\omega_n)_{n \in \mathbb{N}} : \omega_n \in \{A, B, C, D, E\} \forall n\}$*. We could express this as:*

$$\Omega = \{A, B, C, D, E\}^{\mathbb{N}}$$

Further define $X_n \coloneqq X_n(\omega) = \omega_n \forall n$ *and the realization space becomes* $E = \{A, B, C, D, E\}$*, the outcome of the* n^{th} *trial. While the* ***master*** σ*-algebra would be* $\mathcal{H} = \sigma\left((X_n)_{n \in \mathbb{N}}\right)$ *at* $n = 3$ *the information is* $X_1(\omega) = \omega_1, X_2(\omega) = \omega_2, X_3(\omega) = \omega_3$ *so that the filtration at* $n = 3$ *is a collection*

$$\mathcal{F}_3 = \{V(\omega) = f(\omega_1, \omega_2, \omega_3)\} = \{V(\omega) = f(X_1, X_2, X_3)\} \quad f : E \times E \times E \to \mathbb{R}$$

i.e. all the deterministic functions of random information solely concerning $\omega_1, \omega_2, \omega_3$*. An example of such functions could be a counter of how many vowels after* $n = 3$ *have occurred.*

♠ **Definition 11.6** (Finer, coarser filtration). *Consider* $\mathcal{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}, \mathcal{G} = (\mathcal{G}_t)_{t \in \mathbb{T}}$ *to be two filtrations. We say* $\mathcal{F}$ *is finer (coarser) than* $\mathcal{G}$ *is* $\forall t \in \mathbb{T}\ \mathcal{F}_t \supset$ *(respectively,* $\subset$*)* $\mathcal{G}_t$*.*

♠ **Definition 11.7** (Stochastic process adapted to filtration). *Consider a filtration* $\mathcal{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$ *and a stochastic process* $X = (X_t)_{t \in \mathbb{T}}$ *taking values on* $(E, \mathcal{E})$*. We say that* X *is adapted to* $\mathcal{F}$ *if* $\forall t\ X_t$ *measurable w.r.t.* $\mathcal{F}$&$\mathcal{E}$

♣ **Proposition 11.8** (Equivalent statements for filtrations and adaptedness). *Consider a stochastic process $X = (X_t)_{t\in\mathbb{T}}$ and a filtration $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$. Then:*

1. *X adapted to $\mathcal{F}$ (Def. 11.7) $\iff \forall t, s \le t,\ f \in \mathcal{E} \quad f \circ X_s \in \mathcal{F}_t$*
2. *since $\mathcal{G} = \sigma(X) \implies X$ adapted $\mathcal{G}$ we have that:*

$$X \text{ adapted } \mathcal{F} \iff \mathcal{F} \text{ finer } \mathcal{G}$$

Proof. **(Claim #1)**($\implies$) let $f \in \mathcal{E}$ so that $f : \mathcal{E} \to \mathcal{E}$ and $f \circ X_s : \Omega \to \mathcal{E}$. By the fact that f is deterministic and $X_s \in \mathcal{F}_s \subset \mathcal{F}_t \forall s \le t$ we have that trivially $f \circ X_s \in \mathcal{F}_t \forall s \le t.$,
($\impliedby$) let $f \circ X_s = X_s$ then $X_s \in \mathcal{F}_t \forall s \le t, \forall \mathcal{E}$ and Definition 11.7 holds.
(Claim #2) Assuming $\mathcal{G} = \sigma(X)$.
($\implies$) X adapted to $\mathcal{F} \iff \forall t, X_t \in \mathcal{F}_t$ so that $\forall t \mathcal{F}_t \supset \mathcal{G}_t \implies \mathcal{F}$ is finer than $\mathcal{G}$.
($\impliedby$) $\mathcal{F}$ finer than $\mathcal{G}$ means $\forall t \mathcal{F}_t \supset \mathcal{G}_t$ so the assumption that X is adapted to $\mathcal{G}$ trivially transfers to being also adapted to $\mathcal{F}$. □

♠ **Definition 11.9** (Stopping times). *For a filtration $\mathcal{F}$ a stopping time with respect to it is a random function $T : \Omega \to \mathbb{T} \cup \{\infty\}$ such that:*

$$\{T \le t\} \in \mathcal{F}_t \quad \forall t \in \mathbb{T}$$

Which is equivalent to requiring the process $Z_t = \mathbb{1}_{\{T \le t\}} \in \mathcal{F}_t$ for all $t \in \mathbb{T}$.
*In the **special case** in which $\mathbb{T} = \mathbb{N}$ or $\overline{\mathbb{N}}$ the condition reduces for $\widehat{Z}_n = Z_n - Z_{n-1}$ to:*

$$\widehat{Z}_n = \mathbb{1}_{\{T=t\}} \quad \forall t \in \mathbb{T}$$

◇ **Observation 11.10** (Interpreting stopping times). *A random time seen as signal of occurence of a random event. The term stopping comes from the measurability with respect to the filtration. Namely, the information flow allows to detect whether the event has happened or not at any time point.*

♡ **Example 11.11** (The alarm clock metaphor). *Recall Example 11.5 and let:*

$$T(\omega) := \inf \{n \in \mathbb{N} : X_n(\omega) \in A\}, \quad \omega \in \Omega$$

$T(\omega)$ can be interpreted as the first time of entrance in A. T is a stopping time since $\forall n \in \mathbb{N}$ if we assume X is adapted to $\mathcal{F}$ then we have:

$$\{T \le n\} = \bigcup_{k=0}^{n} \underbrace{\{X_k \in A\}}_{\in \mathcal{F}_k \subset \mathcal{F}_n} \in \mathcal{F}_n$$

Instead $L = \max\{0, \sup\{n \le 5 : X_n(\omega) \in A\}\}$ is not a stopping time since $X_4(\omega) \in A, X_t(\omega) \notin A$ mean $L = 4$ but $\mathcal{F}_4$ is not sufficient to conclude at $t = 4$.

♠ **Definition 11.12** (Notation for max and min). *Implement a widely used notation for the aggregators:*

$$\min\{a, b\} = a \wedge b$$
$$\max\{a, b\} = a \vee b$$

♠ **Definition 11.13** (Counting process $(N_t)_{t\in\mathbb{T}}$). *Let $) < T_1 < T_2 < \dots$ be random times of the form $T_n : \Omega \to \mathbb{T} = \mathbb{R}_+$ such that $\lim_{n\to\infty} T_n = +\infty$.*
*These can be seen as a sequence of **distinct arrival times**.*
A counting process is a stochastic process (Def. 10.58) of the form:

$$N_t = \sum_n \mathbb{1}_{[0,t]}(T_n)$$

♣ **Proposition 11.14** (Properties of $(N_t)_{t\in\mathbb{T}}$). *The map $t \to N_t$ is:*

1. *right continuous*
2. *increasing in t*
3. *has jumps of size 1*
4. *$N_0 = 0$, $N_t < \infty \forall t \in \mathbb{R}_+$, $\lim_{t\to\infty} N_t = \infty$*

Proof. Trivial, but recognize that:

- the jumps of size one property is due to $\mathbb{T} = \mathbb{R}_+$ so that the simultaneity of random times has negligible probability
- the finiteness for finite times is due to the fact that to have countably infinite arrival times (i.e. $N_t = +\infty$) one must have $t = \infty \in \overline{\mathbb{R}}_+$, violating the assumption that $t \in \mathbb{R}_+$.

□

♡ **Example 11.15** (Some stopping & not stopping times). *we provide three examples:*

- *Let $\mathcal{F} = \sigma(\{N_t\})$. If we denote as T_k the k^{th} occurrence time in $[0,t]$ we can safely say that it is a stopping time of $\mathcal{F}$ since:*
$$\forall k \geq 1, k \in \mathbb{N}, \forall t \in \mathbb{R}_+ \quad \{T_k \leq t\} = \{N_t \geq k\} \in \mathcal{F}_t$$
since N is adapted to $\mathcal{F}$ by construction
- *The first time that an interval a passes without an arrival, namely:*
$$T = \inf\{t \geq a : N_t = N_{t-a}\} \quad a > 0$$
Needs the formalism of stopped filtration (Def. 11.19) and we will show it is a stopping time in Example 11.27.
- *instead a random time such as the time of last arrival before $b > 0$:*
$$L = \inf\{t : N_t = N_b\} \quad b > t$$
is not a stopping time since we need the information from the interval $[t,b]$ to establish what occurs at time t.

♢ **Observation 11.16** (About stopping times). *Recall Definition 11.9. The maps are of the form $T : \Omega \to \mathbb{T} \cup \{\infty\}$ so the realization may be unbounded, and for some ω it might be that $T(\omega) = \infty$. Consider Example 11.11. It could be that $\nexists \omega : X_n(\omega) \in A \forall n \in \mathbb{N} \implies T(\omega) = \infty$.*

♠ **Definition 11.17** (End of time information $\mathcal{F}_\infty$, extended filtration $(\mathcal{F}_t)_{t\in\overline{\mathbb{T}}}$). *We define $\mathcal{F}_\infty = \lim_{t\to\infty} \mathcal{F}_t = \bigvee_t \mathcal{F}_t$, where the union symbol is different as it is **over σ-algebras**.*
Then, the extended filtration is a filtration which accounts for $\mathbb{P}[T = \infty] > 0$:

$$(\mathcal{F}_t)_{t\in\overline{\mathbb{T}}} \quad \overline{\mathbb{T}} = \mathbb{T} \cup \{\infty\}$$

♢ **Observation 11.18** (About end of time and extended filtrations). *Notice that:*

1. *$(\mathcal{F}_t)_{t\in\overline{\mathbb{T}}}$ and T a stopping time $\iff$ T stopping for $(\mathcal{F}_t)_{t\in\mathbb{T}}$*
2. *X a stochastic process adapted (Def. 11.7) to $\mathcal{F}$ is extended onto $\overline{\mathbb{T}}$ with any $X_\infty \in \mathcal{F}_\infty$*

♠ **Definition 11.19** (Stopped filtration $\mathcal{F}_T$ at T, past until T). *For $\mathcal{F}$ a filtration on $\mathbb{T}$, extended to $\overline{\mathbb{T}}$, and T a stopping time, the stopped filtration is defined as:*

$$\mathcal{F}_T = \left\{H \in \mathcal{H} : H \cap \{T \leq t\} \in \mathcal{F}_t \; \forall t \in \overline{\mathbb{T}}\right\}$$

Lemma 11.20 (Properties of $\mathcal{F}_T$). *A stopped filtration $\mathcal{F}_T$ is such that:*

1. *$\mathcal{F}_T$ is a σ-algebra (Def. 1.6)*
2. *$\mathcal{F}_T \subset \mathcal{F}_\infty \subset \mathcal{H} \; \forall t$*

Proof. **(Claim #1)** we check the requirements according to the definition. Indeed $\mathcal{F}$ is a sequence of σ-algebras and T is a stopping time for the filtration.

- the whole sample space is included
$$\underbrace{\Omega}_{\in\mathcal{F}_t} \cap \underbrace{\{T \leq t\}}_{\in\mathcal{F}_t} \in \mathcal{F}_t$$

- complements are included

$$\begin{cases} H \in \mathcal{F}_T \iff H \cap \{T \le t\} \in \mathcal{F}_t \forall t \in \overline{\mathbb{T}} \\ H^c \in \mathcal{F}_T \iff H^c \cap \{T \le t\} \in \mathcal{F}_t \forall t \in \overline{\mathbb{T}} \end{cases} \implies H \cap \{T \le t\},\ H^c \cap \{T \le t\} \in \mathcal{F}_t \forall t$$

since if $H \in \mathcal{F}_T$ then the intersection with $\{T \le t\}$ is either the whole set, an empty set or a subset for $r < t$. So that in the three cases ordered:
  - H^c has a null intersection
  - H^c is the whole set
  - H^c is the remaining set $\{T \in [0, r \wedge t]\}$

  By $\mathcal{F}_t$ being a σ-algebra and T being a stopping time they all belong to $\mathcal{F}_t$ and so to $\mathcal{F}_T$.
- countable unions are shown following the same arguments so that

$$\forall t \in \mathbb{T} \quad \bigcap_n H_n \cap \{T \le t\} \in \mathcal{F}_t$$

(Claim #2) For $T : \Omega \to \mathbb{T} \cup \{\infty\}$ and $\mathcal{F}_\infty = \lim_{t\to\infty} \mathcal{F}_t$ we can safely say that by $\mathcal{F}$ a filtration and T a stopping time:

$$\mathbb{P}[T \le \infty] = 1,\ \mathcal{F}_s \subset \mathcal{F}_t\ \forall t \in \overline{\mathbb{T}} \implies \mathcal{F}_T \subset \mathcal{F}_\infty$$

And also:

$$\mathcal{F}_\infty = \bigcup_t \underbrace{\mathcal{F}_t}_{\subset \mathcal{H} \forall t \in \mathbb{T}} \subset \mathcal{H}$$

since a union of subsets of $\mathcal{H}$ is again a subset of $\mathcal{H}$ □

♡ **Example 11.21** (Events in stopped filtrations). *Given $T : \Omega \to \overline{T}$ an event $H = \{T \le r\}$ is such that:*

$$\forall t \in \overline{\mathbb{T}} \quad H \cap \{T \le t\} = \{T \le r\} \cap \{T \le t\} = \{T \le (r \wedge t)\} \in \mathcal{F}_{r\wedge t} \subset \mathcal{F}_r$$

And we can further say that T is $\mathcal{F}_T$-measurable as $\forall r$ we have $H \in \mathcal{F}_r$, with $H \in \mathcal{F}_T$ as well.

♢ **Observation 11.22** (σ-algebras as collections of measurable random variables). *See that:*

$$\mathcal{F}_T = \{V : \Omega \to \mathbb{R} \mid \forall \omega \in \Omega\ V(\omega) = f(T(\omega)),\ f \text{ deterministic}\}$$

♣ **Theorem 11.23** (Formalizing Observation 11.22). *Drawing from the previous comment, for a stopped filtration $\mathcal{F}_T$, with stopping time T, index $\mathbb{T}$, and filtration $\mathcal{F}$:*

1. *stopped filtration filtration identification*

$$V \in \mathcal{F}_T \iff V\mathbb{1}_{T\le t} \in \mathcal{F}_t \forall t \in \overline{\mathbb{T}}$$

2. *stopped filtration identification for a discrete process*

$$\overline{\mathbb{N}} = \overline{\mathbb{T}}\ \ V \in \mathcal{F}_T \iff V\mathbb{1}_{T=t} \in \mathcal{F}_t \forall t \in \overline{\mathbb{T}}$$

Which are both an extension of the comments of adaptedness from Definition 11.7 for deterministic times.

Proof. **(Claim #1)** let $V \ge 0$ a.s. wlog with $X_t = V\mathbb{1}_{\{T\le t\}}$. Then:

$$\forall r \in \mathbb{R}_+, \forall t \in \mathbb{T} \quad \{V > r\} \cap \{T \le t\} = \{X_t > r\}$$

and we have that

$$\{V > r\} \in \mathcal{F}_T \forall r \in \mathbb{R}_+ \overset{\text{Def. 11.7}}{\iff} \{X_t > r\} \in \mathcal{F}_t \forall r, \forall t$$
$$\iff V \in \mathcal{F}_T \iff X_t \in \mathcal{F}_t \forall t \in \overline{\mathbb{T}}$$

(Claim #2) for $\overline{\mathbb{T}} = \overline{\mathbb{N}}$ it holds:

$$V\mathbb{1}_{\{T=n\}} = \begin{cases} X_n - X_{n-1} & n \in \mathbb{N} \\ X_\infty - \sum_n X_n - X_{n-1} & n = \infty \end{cases}$$

Which gives us:

$$V \in \mathcal{F}_T \iff V\mathbb{1}_{\{T=n\}} \in \mathcal{F}_n \quad \forall n \in \mathbb{N}$$

□

♠ **Definition 11.24** ($\mathcal{F}$ processes collection)**.** *Using the same notation for positive measurable functions as $f \in \mathcal{E}$ we let:*

$$\mathcal{F} = \{\text{right continuous processes on } \overline{\mathbb{T}} \text{ adapted to } \mathcal{F}\}$$

Where $\mathcal{F}$ is extended to $\overline{\mathbb{T}}$.
This means that $X \in \mathcal{F}$ whenever:

1. *$X = (X_t)_{t\in\mathbb{T}}$ is adapted to $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$*
2. *$t \to X_t(\omega)$ where $X_t : \overline{\mathbb{T}} \to \overline{\mathbb{R}}$ is right continuous $\forall\omega \in \Omega$*

♢ **Observation 11.25** (Justifying the $X \in \mathcal{F}$ notation)**.** *Let $\overline{\mathbb{N}} = \overline{\mathbb{T}}$, then requirement #2 from Definition 11.24 holds since $n \to X_n(\omega)$ is continuous on the discrete topology $\mathbb{N}$. This allows us to conclude that:*

$$X \in \mathcal{F} \iff X_n \in \mathcal{F}_n \ \forall n \in \overline{\mathbb{N}}$$

♣ **Theorem 11.26** (Comparing different stopping times)**.** *Let S, T be stopping times of a filtration $\mathcal{F}$ (Def. 11.9), where $S \le T$ almost surely, meaning $S(\omega) \le T(\omega)\forall\omega \in \Omega$. Then:*

1. *$S \wedge T$, $S \vee T$ are stopping times of $\mathcal{F}$*
2. *specifically $S \le T \implies \mathcal{F}_S \subset \mathcal{F}_T$*
3. *$\mathcal{F}_{S\wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$*
4. *$V \in \mathcal{F}_S \implies \begin{cases} V\mathbb{1}_{S\le T} \in \mathcal{F}_{S\wedge T} \\ V\mathbb{1}_{S=T} \in \mathcal{F}_{S\wedge T} \\ V\mathbb{1}_{S<T} \in \mathcal{F}_{S\wedge T} \end{cases}$*

Proof. **(Claim #1)** Assume S, T are stopping times for $\mathcal{F}$. Then, by the fact that $\mathcal{F}_t$ is a σ-algebra:

$$\begin{cases} \{S \wedge T \le t\} = \{S \le t\} \cup \{T \le t\} \in \mathcal{F}_t \\ \{S \vee T \le t\} = \{S \le t\} \cap \{T \le t\} \in \mathcal{F}_t \end{cases}$$

and both are stopping times.
(Claim #2) Let $V \in \mathcal{F}_S$, by Theorem 11.23#1 it holds that $V\mathbb{1}_{\{S\le t\}} = X_t$ is a process with index $t \in \overline{\mathbb{T}}$. X is adapted to $\mathcal{F}$ and is right continuous, so we can write $X \in \mathcal{F}$ in the sense of Definition 11.24. By $S \le T$ we can say that $X_T = V\mathbb{1}_{\{S\le T\}} = V$ is such that $X_T \in \mathcal{F}_T$ by [Çin11](Thm. V.1.14). Then $S \le T$ is a sufficient condition for $\mathcal{F}_S \subset \mathcal{F}_T$.
(Claim #4) the claim is equivalent to proving:

$$H \in \mathcal{F}_S \implies H \cap \{S \le T\}, H \cap \{S < T\}, H \cap \{S = T\} \in \mathcal{F}_{S\wedge T}$$

By Claim #1 $S \wedge T$ is a stopping time so $X_{S\wedge T} \in \mathcal{F}_{S\wedge T}$ [Çin11](Thm. V.1.14). Moreover:

$$X_{S\wedge T} = V\mathbb{1}_{\{S\le S\wedge T\}} \implies V\mathbb{1}_{\{S\le T\}} \in \mathcal{F}_{S\wedge T}$$

For $V = 1$ it holds that:

$$\{S \le T\} \in \mathcal{F}_{S\wedge T}, \quad \{S \ge T\} \in \mathcal{F}_{S\wedge T} \quad \textit{by symmetry}$$

So that all of the following belong to $\mathcal{F}_{S\wedge T}$ with the indicators:

$$\{S = T\} = \{S \le T\} \cap \{S \ge T\}, \quad \{S < T\} = \{S \le T\} \setminus \{S = T\}, \quad \{S > T\} = \{S \ge T\} \setminus \{S = T\}$$

(Claim #3)(⊂) since $S \le T$ we have $S \wedge T \le S, T$ so that by Claim #2:

$$\mathcal{F}_{S\wedge T} \subset \mathcal{F}_S, \mathcal{F}_T \implies \mathcal{F}_{S\wedge T} \subset \mathcal{F}_S \cap \mathcal{F}_T$$

(⊃) conversely, let $H \in (\mathcal{F}_S \cap \mathcal{F}_T)$. By Claim #4 it holds:

$$\begin{cases} H \in \mathcal{F}_S \implies H \cap \{S \le T\} \in \mathcal{F}_{S\wedge T} \\ H \in \mathcal{F}_T \implies H \cap \{T \le S\} \in \mathcal{F}_{S\wedge T} \end{cases} \implies (H \cap \{S \le T\}) \cup (H \cap \{T \le S\}) = H \in \mathcal{F}_{S\wedge T} \implies \mathcal{F}_S \cap \mathcal{F}_T \subset \mathcal{F}_{S\wedge T}$$

Eventually we proved by double inclusion and $\mathcal{F}_S \cap \mathcal{F}_T = \mathcal{F}_{S\wedge T}$. □

♡ **Example 11.27** (Counting process from Definition 11.13). *This Example will be expanded across the lectures.* ***(△ setting)*** *Consider the counting process on $\mathbb{T} = \mathbb{R}_+$ from Definition 11.13. If we consider $H \cap \{S < T\}$ we could tell if H and $S < T$ happened in $\mathcal{F}_{S \wedge T}$. For any t it holds that:*

$$\exists k : \; T_k(\omega) \leq t < T_{k+1}(\omega)$$

Recall also that all of these T_k are stopping times of $\mathcal{F} = \sigma(\{N_t\}_{t \geq 0})$.
We set $T_0 = 0$ for convenience, and consider the random time:

$$\tau = \inf \{t \geq a : N_t = N_{t-a}\} \quad a > 0$$

denoted in blue for convenience. Before this Example, the symbol T was used, but here we wish to distinguish many objects that are similar in notation. After this Example, we will not use the symbol τ.
We want to show that τ is a stopping time in the sense of Definition 11.9.
(□ solution) *Notice that $\forall \omega$ we have for some k:*

$$\tau(\omega) = T_k(\omega) + a \quad k \in \mathbb{N}^* \iff \{\tau \leq t\} = \bigcup_{k \geq 1} \{\{\tau = T_k + a\} \cap \{\tau \leq t\}\}$$

Where the union over k statement comes from the fact that we have $\exists k$ as a condition. Recall the objects of Theorem 11.23, visualizing them as $T = T_{k+1}, S = T_k + a : S \leq T$. Be careful as it may lead to confusion, left is old and red, right is this Example. It holds:

$$\tau \text{ stopping} \iff \{\tau = T_k + a\} \cap \{\tau \leq t\} \in \mathcal{F}_t \quad k \in \mathbb{N}^*$$

Using Theorem 11.23#1, we need to check that $V = \{\tau = T_k + a\} \in \mathcal{F}_{T_k + a} \subset \mathcal{F}_T$ to conclude.
To clear out why, recognize that it is equivalent to $\{\tau = T_k + a\}\mathbb{1}_{\{T_k + a \leq t\}} = \{\tau \leq t\} \in \mathcal{F}_t \forall t \in \overline{\mathbb{T}}$ by the very Theorem invoked.
For this purpose, observe that for any k:

$$\{\tau = T_k + a\} = \underbrace{\{T_1 - T_0 \leq a, \ldots, T_k - T_{k-1} \leq a\}}_{\in \mathcal{F}_S, S = T_k + a} \cap \underbrace{\{T_k + a < \overbrace{T_{k+1}}^{T = T_{k+1}}\}}_{S < T}$$

By Theorem 11.26#4 we will have that:

$$H \in \mathcal{F}_S \implies H \cap \mathbb{1}_{\{S<T\}} \in \mathcal{F}_{S \wedge T}$$

Where $S = T_k + a, T = T_{k+1}, H = \{T_1 - T_0 \leq a, \ldots, T_k - T_{k-1} \leq a\}$. Eventually:

$$\begin{aligned} \{\tau = T_k + a\} \in \mathcal{F}_{T_k + a \wedge T_{k+1}} = \mathcal{F}_{T_k + a} = \mathcal{F}_\tau &\iff \{\tau = T_k + a\} \cap \{\tau \leq t\} \in \mathcal{F}_t && \forall k, \forall t \\ &\iff \{\tau \leq t\} \in \mathcal{F}_t && \forall t \end{aligned}$$

11.2 Random Expectation and Martingales

◇ **Observation 11.28** (Recapping conditional expectation). *For $X \in \mathcal{L}_1(\Omega, \mathcal{H}, \mathbb{P})$ (Def. 4.5) and $\mathcal{F}$ a σ-algebra (Def. 1.6) where $\mathcal{F} \subset \mathcal{H}$ we defined:*

$$\overline{X}_\mathcal{F} = \mathbb{E}[X|\mathcal{F}] = \mathbb{E}_\mathcal{F}[X]$$

where by Definition 10.12:

1. *$\overline{X}_\mathcal{F}$ is $\mathcal{F}$-measurable (Eqn. 3.1)*
2. *$\mathbb{E}[VX] = \mathbb{E}[V\overline{X}_\mathcal{F}]$ $\forall V$ $\mathcal{F}$-measurable positive*

Where $\#2, \#1 \iff \overline{X}_\mathcal{F}$ is the best approximation in $\mathcal{L}_2(\Omega, \mathcal{H}, \mathbb{P})$, namely:

$$\mathbb{E}\left[(X - \overline{X}_\mathcal{F})^2\right] \leq \mathbb{E}\left[(X - Y)^2\right] \quad \forall Y \; \mathcal{F}\text{-measurable}$$

◇ **Observation 11.29** (Properties of $\mathbb{E}_\mathcal{F}$ inherited by $\mathbb{E}$). *Comparing Chapter 4 with Chapter 10, notice that some properties carried over by the fact that $\mathcal{F}$ is simply a σ-algebra:*

- *monotonicity (Prop. 10.20, #1)*
- *linearity (Prop. 10.19, #2)*
- *monotone convergence (Prop. 10.20, #2)*
- *dominated convergence (Prop. 10.20, #3)*
- *Fatou's lemma (Lem. A.49)*

♢ **Observation 11.30** (Peculiar properties of $\mathbb{E}_{\mathcal{F}}$)**.** *There are also properties which are specific to the construction of conditional expectation, among which we saw:*

- *unconditioning (Prop. 10.18, #3)*
- *conditional determinism (Prop. 10.23, #1)*
- *towering property (repeated conditioning) (Thm. 10.21)*

♠ **Definition 11.31** (Expectation in Filtration $\mathbb{E}_t$)**.** *Given a filtration $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ (Def. 11.2) use as notation:*

$$\mathbb{E}_t[X] \coloneqq \mathbb{E}[X|\mathcal{F}_t] = \mathbb{E}_{\mathcal{F}_t}[X] = \overline{X}_{\mathcal{F}_t} \quad t \in \mathbb{T}$$

♣ **Proposition 11.32** (Repeated conditioning of $\mathbb{E}_t$)**.** *For $X \geq 0$ a.s. it holds that:*

$$\mathbb{E}_t\left[\mathbb{E}_s[X]\right] = \mathbb{E}_{s\wedge t}[X] \quad \forall s,t \in \mathbb{T}$$

Proof. Notice wlog $s \leq t \implies \mathcal{F}_s \subset \mathcal{F}_t$ and the result is an application of the towering property (Thm. 10.21). □

♠ **Definition 11.33** (Expectation with respect to stopped filtration $\mathbb{E}_T$)**.** *Given $\mathcal{F}_T$ a stopped filtration (Def. 11.19), recalling that $\mathcal{F}_T$ is a σ-algebra (Lem. 11.20, #1), simply define:*

$$\mathbb{E}_T[X] \coloneqq \mathbb{E}[X|\mathcal{F}_T] = \mathbb{E}_{\mathcal{F}_T}[X] = \overline{X}_{\mathcal{F}_T}$$

♣ **Theorem 11.34** (Properties of $\mathbb{E}_T$)**.** *Consider $X, Y, W \geq 0$ a.s. and S, T stopping times (Def. 11.9) of a filtration $\mathcal{F}$ (Def. 11.2). Then:*

1. *Projection defining property*
$$\mathbb{E}_T[X] = Y \iff Y \in \mathcal{F}_T\ \mathbb{E}[VX] = \mathbb{E}[VY]\ \forall V\ \mathcal{F}_T\text{-measurable positive}$$
2. *unconditioning*
$$\mathbb{E}\left[\mathbb{E}_T[X]\right] = \mathbb{E}[X]$$
3. *repeated conditioning/towering*
$$\mathbb{E}_S\mathbb{E}_T[X] = \mathbb{E}_{S\wedge T}X$$
4. *conditional determinism*
$$\mathbb{E}_T[WX] = W\mathbb{E}_T[X] \quad \forall W\ \mathcal{F}_T\text{-measurable}$$

Proof. **(Claims #1#2#4)** hold trivially by Observations 11.29, 11.30.
(Claim #3) for $S \leq T$ by Theorem 11.26#2 it holds $\mathcal{F}_S \subset \mathcal{F}_T$. However, we are not given this condition and we want to show in general that:

$$\mathbb{E}_T\mathbb{E}_S X = \mathbb{E}_S\mathbb{E}_T X = \mathbb{E}_{S\wedge T}X$$

For arbitrary S, T, wlog we know $S \wedge T \leq T$ so that $\mathbb{E}_{S\wedge T}\mathbb{E}_T X = \mathbb{E}_{S\wedge T}X$ by Towering. For this reason, we set $Y \coloneqq \mathbb{E}_T X$ and the claim is equivalent to showing:

$$\mathbb{E}_S Y = \mathbb{E}_{S\wedge T}Y \quad \mathbb{E}_S\mathbb{E}_T X = \mathbb{E}_{S\wedge T}\mathbb{E}_T X = \mathbb{E}_{S\wedge T}X$$

Where the RHS of the first is well defined if $\mathbb{E}_{S\wedge T}Y \in \mathcal{F}_{S\wedge T} \subset \mathcal{F}_S$ so that Def. 10.12#1 is satisfied. We are left to show that Claim #1 holds, namely:

$$\mathbb{E}[VY] = \mathbb{E}[V\mathbb{E}_{S\wedge T}Y] \quad \forall V \in \mathcal{F}_S^+$$

Fix $V \in \mathcal{F}_S^+$, by Theorem 11.26#4 it holds that $V\mathbb{1}_{\{S\leq T\}} \in \mathcal{F}_{S\wedge T}$. Moreover, the defining property of expectation gives:

$$\mathbb{E}\left[V\mathbb{1}_{\{S\leq T\}}Y\right] = \mathbb{E}\left[V\mathbb{1}_{\{S\leq T\}}\mathbb{E}_{S\wedge T}[Y]\right]$$

Additionally, by $Y \in \mathcal{F}_T$ we also get again by Theorem 11.26#4 that :

$$\mathbb{E}\left[V \mathbb{1}_{\{T \leq S\}} Y\right] = \mathbb{E}\left[V \mathbb{E}_{S \wedge T}[Y \mathbb{1}_{\{T \leq S\}}]\right] = \mathbb{E}\left[V \mathbb{1}_{\{T \leq S\}} \mathbb{E}_{S \wedge T}[Y]\right]$$

adding the last two results together gives the claim:

$$\mathbb{E}[VY] = \mathbb{E}[V \mathbb{E}_{S \wedge T}[Y]] \quad \forall V \in \mathcal{F}_S^+$$

□

♠ **Definition 11.35** (Martingales). *For $\mathbb{T} = \mathbb{R}_+$ and $\mathcal{F}$ a filtration, possibly extended to $\overline{\mathbb{T}}$ a $\mathcal{F}$-martingale is a stochastic process $X = (X_t)_{t \in \mathbb{T}}$ such that:*

1. *X is adapted to $\mathcal{F}$ (Def. 11.7)*
2. *$\forall t$ X_t is integrable $\iff \mathbb{E}[|X_t|] < \infty \forall t \iff X_t \in \mathcal{L}_1 \; \forall t$*
3. *martingale equality*

$$\mathbb{E}_s[X_t - X_s] = 0 \quad \forall s < t, \forall t$$

♠ **Definition 11.36** (Submartingale, supermartingale). *we recognize two additional options for the last property:*

- *a **submartingale** satisfies Definition 11.35 but has $\geq$ in the martingale equality*
- *a **supermartingale** satisfies Definition 11.35 but has $\leq$ in the martingale equality*

♢ **Observation 11.37** (About martingales). *If X is an $\mathcal{F}$-submartingale then it tends to increase over time. The other options follow trivially.*

♣ **Proposition 11.38** (Best guess of future is present characterizes martingale equality). *It holds that:*

$$\textit{Def. 11.35\#3} \iff \mathbb{E}_s[X_t] = X_s \quad \forall s < t$$

Proof. By linearity of expectation and conditional determinism $X_s \in \mathcal{F}_s$ we have $\mathbb{E}_S[X_t] = \mathbb{E}_s[X_s] = X_s$. □

♣ **Proposition 11.39** (Martingale implies stationarity).

$$X \; \mathcal{F}\text{-martingale} \implies \mathbb{E}[X_t] = \mathbb{E}[X_0] \forall t \in \mathbb{T}$$

Proof. Use the martingale equality (Def. 11.35#3):

$$\begin{aligned} \mathbb{E}\left[\mathbb{E}_s[X_t - X_s]\right] &= \mathbb{E}\left[X_t - X_s\right] && \text{unconditioning Prop. 10.18\#3} \\ &= \mathbb{E}[X_t] - \mathbb{E}[X_s] && \text{linearity} \\ &= 0 && \mathbb{E}[0] \text{ martingale equality} \end{aligned}$$

So that $\mathbb{E}[X_t] = \mathbb{E}[X_0] \forall t \in \mathbb{T}$. □

♣ **Proposition 11.40** (Discrete time martingale check). *For a discrete process over $\mathbb{T} = \mathbb{N}$ it is sufficient to check for one step forward the martingale equality if the other two conditions are satisfied (**integrability** and **adaptedness**):*

$$\mathbb{E}_n[X_{n+k} - X_n] = 0 \; \forall k > 0, \forall n \in \mathbb{N} \iff \mathbb{E}_n[X_{n+1} - X_n] = 0 \; \forall n$$

Proof. ($\implies$) trivial direction.
($\impliedby$) we work by induction on k with the assumption that:

$$\mathbb{E}_n[X_{n+1} - X_n] = 0 \forall n$$

(Base case) for $k = 1$ it holds:

$$\begin{aligned} \mathbb{E}_n[X_{n+2}] - \underbrace{\mathbb{E}_n[X_n]}_{=X_n} = \mathbb{E}_n[X_{n+1+1} - X_n] &= \mathbb{E}_n[X_{n+1+1} - X_{n+1} + X_{n+1} - X_n] \\ &= \mathbb{E}_n[X_{n+2} - X_{n+1}] - \underbrace{\mathbb{E}_n[X_{n+1} - X_n]}_{=0} \\ &= \mathbb{E}_n[X_{n+2}] - \mathbb{E}_n[X_{n+1}] \end{aligned}$$

Observe that by the towering property and $n < n+1$:

$$\begin{aligned} \mathbb{E}_n[X_{n+2} - X_{n+1}] &= \mathbb{E}_n\left[\mathbb{E}_{n+1}[X_{n+2} - X_{n+1}]\right] \\ &= \mathbb{E}_n[0] && \text{hypothesis} \\ &= 0 \end{aligned}$$

which holds by the arbitrariness of n.
(induction assumption) assume it is true $\forall k$.
(inductive step) for $k+1$, doing the same trick we have:

$$\begin{aligned} \mathbb{E}_n[X_{n+k+1}] - \mathbb{E}_n[X_n] = \mathbb{E}_n[X_{n+k+1} - X_n] &= \mathbb{E}_n[X_{n+k+1} - X_{n+k} + X_{n+k} - X_n] \\ &= \mathbb{E}_n[X_{n+k+1} - X_{n+k}] + \underbrace{\mathbb{E}_n[X_{n+k} - X_n]}_{=0} \\ &= \mathbb{E}_n\left[\mathbb{E}_{n+k}[X_{n+k+1}] - \mathbb{E}_n[X_{n+k}]\right] \\ &= \mathbb{E}_n[0] && \text{hypothesis} \end{aligned}$$

and we have proved the claim. □

Corollary 11.41 (Jensen's for martingales)**.** *This is a Corollary of Theorem 7.7. For a martingale X and a convex function $f : \mathbb{R} \to \mathbb{R}$ then:*

$$f \circ X_t \text{ integrable } \forall t \in \mathbb{T} \implies f \circ X \text{ submartingale}$$

Proof. The process $(Y_t)_{t\in\mathbb{T}} = (f(X_t))_{t\in\mathbb{T}}$ is such that by Jensen's Inequality (Thm. 7.7) $\mathbb{E}_t[f(X)] \geq f(\mathbb{E}_t[X])$ so that:

$$\mathbb{E}_s[f(X_t)] \geq f(\mathbb{E}_s[X_t]) = f(X_s) \quad \forall s < t, \forall t$$

Where the first is an application of Jensen's and the second is the martingale property. So $\mathbb{E}_s[Y_t] \geq Y_s$ and the process $(Y_t)_{t\in\mathbb{T}} = (f \circ X_t)_{t\in\mathbb{T}}$ is a submartingale since it is trivially adapted to $\mathcal{F}$ and integrable by hypothesis. □

♡ **Example 11.42** (Some submartingales applying Corollary 11.41)**.**

$$(|X_t|)_{t\in\mathbb{T}}, \quad \left(X_t^+\right)_{t\in\mathbb{T}}, \quad \left(X_t^-\right)_{t\in\mathbb{T}}, \quad \mathbb{E}[|X_t|^p] < \infty \forall t \implies (|X_t|^p)_{t\in\mathbb{T}}$$

♢ **Observation 11.43** (Submartingales linearity)**.** *Observe that X, Y $\mathcal{F}$-submartingales $\implies$ $aX + bY$, $(X_t \wedge Y_t)_{t\in\mathbb{T}}$ are submartingales.*

♡ **Example 11.44** (Sum of independent random variables martingale)**.** *Let $(X_n)_{n\in\mathbb{N}}$ be an independency where $\mathbb{E}[X_n] = 0 \forall n$. The sum r.v. is such that $S_0 = 0, S_n = S_{n-1} + X_n \forall n \geq 1$, and the underlying filtration is generated by the process itself $\mathcal{F} = \sigma\left((X_n)_{n\in\mathbb{N}}\right)$. Then:*

- *$(S_n)_{n\in\mathbb{N}} = S$ is adapted to $\mathcal{F}$ trivially*
- *$\mathbb{E}[S_n] = \mathbb{E}\left[\sum_{k=1}^n X_k\right] = 0 \forall n \implies \mathbb{E}[|S_n|] < \infty$ so that $S_n \in \mathcal{L}_1(\Omega, \mathcal{H}, \mathbb{P})$ for all n*
- *using Proposition 11.40 we only check the martingale for one step forward:*

$$\begin{aligned} \mathbb{E}_n[S_{n+1} - S_n] &= \mathbb{E}_n[X_{n+1}] && \textit{recursion} \\ &= \mathbb{E}[X_{n+1}] && \textit{independence hyp.} \\ &= 0 && \textit{hypothesis} \end{aligned}$$

So $(S_n)_{n\in\mathbb{N}}$ is a martingale.

♢ **Observation 11.45** (A counterintuitive process failing to be a martingale)**.** *The process $\left(\frac{1}{n}S_n\right)_{n\in\mathbb{N}}$ has mean 0 but is not a martingale in general, we will see conditions for it to satisfy the other two requirements of Definition 11.35, namely adaptedness and integrability.*

♡ **Example 11.46** (Product of independent random variables martingale)**.** *Let:*

- *$R_1, R_2, \ldots$ be independent and such that $\mathbb{E}[R_k] = 1$ and $V[R_k] < \infty \ \forall k$*
- *$M_0 = 1$ and $M_n = M_{n-1}R_n = M_0 R_1 R_2 \cdots R_n$*

We check that M is a martingale with respect to its natural filtration according to Definition 11.35.
(adaptedness)*Clearly $(M_n)_{n\in\mathbb{N}}$ is adapted to $\mathcal{F} = \sigma((M_n)_{n\in\mathbb{N}})$.*
(integrability) *Observe that:*

$$\mathbb{E}[|M_n|] = \mathbb{E}[|M_{n-1}R_n|] = \mathbb{E}\left[\left|M_0 \prod_{k=1}^{n} R_k\right|\right]$$

Where by induction we can show that:

- $\mathbb{E}[|M_1|] = \mathbb{E}[|M_0 R_1|] = \mathbb{E}[|R_1|] < \infty$ *by hypothesis*
- *one step forward*

$$\begin{aligned}
\mathbb{E}[|M_2|] &= \mathbb{E}[|M_0 R_1 R_2|] && M_0R_1R_2 = M_1R_2 \\
&\leq \sqrt{\mathbb{E}[|M_1^2|]\mathbb{E}[|R_2^2|]} && \textit{Cauchy-Schwartz} \\
&= \sqrt{\mathbb{E}[R_1^2]\mathbb{E}[R_2^2]} && \\
&< \infty && \textit{by } V[R_k] < \infty \forall k
\end{aligned}$$

- *naturally iterate*

So that $\mathbb{E}[M_n] < \infty \forall n$.
(martingale equality) *Using Proposition 11.40 we check only for $k = 1$ increments:*

$$\begin{aligned}
\mathbb{E}_n[M_{n+1}] = \mathbb{E}_n[M_n R_{n+1}] &= M_n \mathbb{E}_n[R_{n+1}] && \textit{since } M_n \in \mathcal{F}_n \textit{ and deterministic conditioning} \\
&= M_n \mathbb{E}[R_{n+1}] && \textit{remove } n \textit{ since } R_{n+1} \perp\!\!\!\perp R_n \\
&= M_n \cdot 1 && \textit{hypothesis}
\end{aligned}$$

And the claim holds: $(M_n)_{n\in\mathbb{N}}$ is a martingale.

♢ **Observation 11.47** (Interpreting the model of the example). *If $R_n > 0$ a.s. $\forall n$ then $(M_n)_{n\in\mathbb{N}}$ can be thought of as a stock price. Accordingly, the return is $R_{n+1} = \frac{M_{n+1}}{M_n}$ and the growth is $R_{n+1} - 1$.*
If $(M_n)_{n\in\mathbb{N}}$ is not a martingale then it is either a submartingale or a supermartingale. With the assumption that prices are perfect ***this does not make sense****, as it would imply a rush to buy/sell, inducing a trend in $(M_n)_{n\in\mathbb{N}}$. It is then possible to characterize market equilibrium for a process as it being a martingale.*

11.3 Uniform integrability of martingales

♠ **Definition 11.48** (Uniformly integrable martingale). *We define e u.i. martingale as in Definition 7.3 with as arbitrary index set $\mathbb{T}$. Namely:*

$$(X_t)_{t\in\mathbb{T}} \quad \lim_{b\to\infty} \sup_t \mathbb{E}|X_t| \mathbb{1}_{[b,\infty)}(|X_t|) = 0$$

♢ **Observation 11.49** (Improving the remark below Def. 7.3). *Recall that:*

$$X = (X_t)_{t\in\mathbb{T}} \subset \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}) \not\Leftarrow X \text{ u.i} \implies \sup_t \mathbb{E}[|X_t|] < \infty.$$

Since:

$$X \text{ u.i.} \overset{\text{Def. 7.3}}{\iff} \lim_{k\to\infty} \sup_{t\geq 1} \mathbb{E}\left[|X_t| \mathbb{1}_{(k,\infty)}(|X_t|)\right] = 0 \implies \sup_t \mathbb{E}[|X_t|] < \infty$$

Where in the last implication we applied the remark below Def. 7.3.

Remark 2 (About uniformly integrable martingales, ctd). *If a martingale is over a finite set $n \leq N$, denoted as $(X_n)_{n\leq N}$, then all elements are integrable by Definition 11.35#2 and the process is trivially uniformly integrable.*

♣ **Proposition 11.50** (Uniformly integrable martingale by integrable random variable). *Let $Z \in \mathcal{L}_1(\Omega, \mathcal{H}, \mathbb{P})$ and $\mathcal{F}$ a filtration.*

$$\implies X = (X_t)_{t\in\mathbb{T}} : X_t = \mathbb{E}_t[Z] \ \forall t \in \mathbb{T} \quad \textit{uniformly integrable martingale}$$

Proof. **($\triangle$ setting)** we aim to prove the three requirements for a martingale in Definition 11.35 and uniform integrability at last.
$X_t \in \mathcal{F}_t$ since $X_t = \mathbb{E}_t[Z] \forall t \in \mathbb{T}$. This satisfies the adaptedness requirement.
($\square$ integrability) consider X_t, then:

$$\begin{aligned} |X_t| &= |\mathbb{E}_t[Z]| = f\left(\mathbb{E}_t[Z]\right) && f \text{ convex} \\ &\leq \mathbb{E}_t[|Z|] && \text{Jensen's Thm. 7.7} \end{aligned}$$

so that in expectation, reapplying Jensen's and using the monotonicity (Thm. 2.10) the unconditioning properties (Thm. 10.18#3) we get:

$$\mathbb{E}[|X_t|] \leq \mathbb{E}[\mathbb{E}_t[|Z|]] = \mathbb{E}[|Z|] < \infty \quad \text{by hypothesis}$$

($\bigcirc$ martingale equality) we check that by the towering property (repeated conditioning, Thm. 10.21):

$$\mathbb{E}_s[X_t] = \mathbb{E}_s[\mathbb{E}_t[Z]] = \mathbb{E}_{s \wedge t}[Z] = \mathbb{E}_s[Z] = X_s$$

eventually, by $\triangle, \square, \bigcirc$ the process $(X_t)_{t \in \mathbb{T}}$ is a martingale in the sense of Definition 11.35.
Uniform integrability follows by the below Lemma. $\square$

To prove this result, we need a Theorem from the book, which is reported in the Appendix.

Lemma 11.51 (A more general result)**.** *It actually holds that:*

$$Z \in \mathcal{L}_1(\Omega, \mathcal{H}, \mathbb{P}) \implies \mathcal{K} = \{X \mid X = \mathbb{E}_{\mathcal{G}}[Z],\ \mathcal{G} \subset \mathcal{H}\} \quad uniformly\ integrable$$

Proof. Start with observing that if $Z \in \mathcal{L}_1$ then trivially the collection $\{Z\}$ is uniformly integrable.
By Theorem C.1 there exists a convex positive increasing coercive function in $\mathbb{R}$, namely:

$$f : \mathbb{R} \to \mathbb{R}_+ \quad convex,\ increasing,\ \lim_{x \to \infty} \frac{f(x)}{x} = \infty$$

such that $\mathbb{E}[f \circ |Z|] < \infty$.
($\triangle$ aim) we want to show that the collection $\mathcal{K}$ is uniformly intergrable using this fact.
For $X = \mathbb{E}_{\mathcal{G}}[Z]$ where $\mathcal{G} \subset \mathcal{H}$ it holds by Jensen's (Thm. 7.7) on the modulus:

$$|X| = |\mathbb{E}_{\mathcal{G}}[Z]| \leq \mathbb{E}_{\mathcal{G}}[|Z|]$$

moreover by f being convex and increasing and applying Jensen's to f:

$$f \circ |X| \leq f \circ \mathbb{E}_{\mathcal{G}}[Z] \leq \mathbb{E}_{\mathcal{G}}[f \circ |Z|]$$

Eventually:

$$\begin{aligned} \mathbb{E}[f \circ |X|] &\leq \mathbb{E}\left[\mathbb{E}_{\mathcal{G}}[f \circ |Z|]\right] && \text{Monotonicity 2.10} \\ &= \mathbb{E}[f \circ |Z|] && \text{unconditioning 10.18\#3} \\ &< \infty && \text{hypothesis} \end{aligned}$$

So that $\mathcal{K}$ is uniformly integrable since its modulus is bounded by above by an integrable r.v. $\square$

♡ **Example 11.52** (A uniformly integrable martingale: Bayes mean estimation)**.** *Let* $Z_i \overset{iid}{\sim} \mathcal{N}(0,1)$*, and* $\theta \in \mathcal{L}_1(\Omega, \mathcal{H}, \mathbb{P})$ *such that* $\theta \perp\!\!\!\perp Z_i \forall i$.
Define $Y_i = Z_i + \theta \overset{iid}{\sim} \mathcal{N}(\theta, 1)$ *and aim to infer* θ *from a set of observations* $\boldsymbol{Y} = \{Y_i\}_{i=1}^n$.
Why is θ *random? We use a bayesian approach and assign a prior* $\pi(A) = \mathbb{P}[\theta \in A]$ *such that* $Y_i | \theta \overset{iid}{\sim} \mathcal{N}(\theta, 1)$.
Further assume the joint distribution $(\boldsymbol{Y}, \theta)$ *is absolutely continuous (Def. 2.6) wrt Leb and that* $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$.

Using Bayes theorem we can estimate:

$$\pi_n(A) = \mathbb{P}[\theta \in A | Y_1 = y_1, \ldots, Y_n = y_n]$$

By Ionescu-Tulcea Thm. 10.57 construct also the (unique) space:

$$(\mathbb{R}^\infty \times \Theta, \mathcal{B}(\mathbb{R}^\infty) \otimes \mathcal{B}(\Theta), \mathbb{P})$$

so that it is possible to work with a filtration $\mathcal{F} = \sigma((Y_n)_{n\in\mathbb{N}})$. *By Proposition 11.50 since* $\theta \in \mathcal{L}_1$*:*

$$\widehat{\theta}_n = \mathbb{E}[\theta|\,\boldsymbol{Y}] = \mathbb{E}_n[\theta] \quad \textit{such that} \quad \left(\widehat{\theta}_n\right)_{n\in\mathbb{N}} \; \textit{uniformly integrable}$$

Thanks to the $\theta \sim \mathcal{N}$ *assumption we can explicitly compute the posterior distribution as:*

$$\begin{aligned}
\pi_n(\theta) &\propto \pi(\theta) \prod p(y_i|\theta) \\
&\propto exp\left\{-\frac{1}{2\sigma_n^2}(\theta - \mu_n)^2\right\} \\
\implies \theta|\,\boldsymbol{Y} &\sim \mathcal{N}(\mu_n, \sigma_n^2) \\
\mu_n &= \sigma_n^2\left(\frac{\mu_0}{\sigma_0^2} + n\overline{y}\right) \\
\sigma_n^2 &= \left(\frac{1}{\sigma_0^2} + n\right)^{-1}
\end{aligned}$$

Where $\widehat{\theta}_n = \sigma_n^2\left(\frac{1}{\sigma_0^2}\mu_0 + n\overline{y}\right)$ and $\widehat{\theta}_0 = \mu_0$, with the martingale equality satisfied, meaning $\mathbb{E}[\widehat{\theta}_n] = \mu_0 \;\forall n$. We prove uniform integrability in the next exercise.

♡ **Example 11.53** (Uniform integrability of $\widehat{\theta}$ process)**.** *Using Lemma 11.51. We have:*

$$\widehat{\theta}_n = \sigma_n^2\left(\frac{\mu_0}{\sigma_0^2} + n\overline{Y}\right) \quad \overline{Y} \sim \mathcal{N}\left(\theta, \frac{1}{n}\right)$$

and for $f(x) = x^2$ *convex positive increasing and coercive:*

$$\begin{aligned}
\mathbb{E}\left[f(|\widehat{\theta}_n|)\right] = \mathbb{E}\left[\widehat{\theta}_n^2\right] &= V\left[\widehat{\theta}_n\right] - \left(\mathbb{E}\left[\widehat{\theta}_n\right]\right)^2 \\
&= V\left[\widehat{\theta}_n\right] - \mu_0^2 \\
&< \infty \iff V\left[\widehat{\theta}_n\right] < \infty
\end{aligned}$$

where we aim to find an upper bound for the variance. First notice that by the variance decomposition:

$$\overline{Y}^{(n)}|\theta \sim \mathcal{N}\left(\theta, \frac{1}{n}\right) \quad V\left[\overline{Y}^{(n)}\right] = \mathbb{E}\left[V\left[\overline{Y}^{(n)}\right]\right] + V\left[\mathbb{E}\left[\overline{Y}^{(n)}\right]\right] = \mathbb{E}\left[\frac{1}{n}\right] + V[\theta] = \frac{1}{n} + \sigma_0^2$$

Such variance is by the first term in the addition being constant:

$$\begin{aligned}
V[\widehat{\theta}_n] &= n^2\sigma_n^4 V\left[\overline{Y}^{(n)}\right] & V\left[\overline{Y}^{(n)}\right] &= \frac{1}{n} + \sigma_0^2 \\
&= n^2\sigma_n^4\left(\frac{1}{n} + \sigma_0^2\right) \\
&= n^2\sigma_n^4\left(\frac{1+\sigma_0^2 n}{n}\right) \\
&= n\sigma_n^4\left(1+\sigma_0^2 n\right) & \sigma_n^2 &= \left(\frac{1}{\sigma_0^2} + n\right)^{-1} = \frac{\sigma_0^2}{n\sigma_0^2 + 1} \\
&= \frac{n(1+\sigma_0^2 n)\sigma_0^2}{(1+n\sigma_0^2)^2} = \frac{n\sigma_0^2}{1+n\sigma_0^2} \le \sigma_0^2 = V[\theta_0]
\end{aligned}$$

So that the variance is finite $\forall n$ *and* $\mathbb{E}[f(\widehat{\theta}_n)] < \infty$ *for* f *convex and positive. Then,* $\left(\widehat{\theta}_n\right)_{n\in\mathbb{N}}$ *is a uniformly integrable martingale.*

♡ **Example 11.54** (Branching Process, a uniformly integrable martingale). *The following is a discrete time biological model for population evolution. We interpret Z_n as the size of a population, which starts at $Z_0 = 1$, has no overlapping generations and lifetimes of unit one. At $n+1$, the population is an offspring of the n^{th} generation only. We denote:*

$$Z_0 = 1, \quad Z_1 = \xi_1^{(1)}, \quad Z_2 = \sum_{i=1}^{Z_1} \xi_i^{(2)}$$

And assume:

$$\left\{\xi_i^{(n)}, i \geq 1, n \geq 1\right\} \text{ iid } \mathbb{E}[\xi_i^{(n)}] = \mu \geq 0, \quad p_k := \mathcal{P}\left[\xi_i^{(n)} = k\right], \ k \geq 0$$

where p_k is referred to as the offspring distribution. The underlying filtration is generated by the sizes of past families as $\mathcal{F}_n = \sigma\left(\left\{\xi_i^{(m)}, i \geq 1, m \leq n\right\}\right)$.

*(△ **aim**) we want to show that $\left(\frac{Z_n}{\mu^n}\right)_{n\in\mathbb{N}}$ is a martingale.*

*(□ **solution**) This is equivalent to showing that another process satisfies the martingale equality:*

$$\left(\frac{Z_n}{\mu^n}\right)_{n\in\mathbb{N}} \iff \mathbb{E}_n[Z_{n+1}] = \mu Z_n$$

Which follows by simple computation. Adaptedness and integrability are trivial. Maybe it is useful to notice that $\mathbb{E}\left[|\xi_i^{(n)}|\right] = \mathbb{E}\left[\xi_i^{(n)}\right]$ by positivity. The above formula can be checked for one time step only by Proposition 11.40. Then:

$$\begin{aligned}
\mathbb{E}_n[Z_{n+1}] &= \mathbb{E}_n\left[\left(\xi_1^{(n+1)} + \cdots + \xi_{Z_n}^{(n+1)}\right)\mathbb{1}_{\{Z_n>0\}}\right] && \textit{recursion hypothesis}\\
&= \mathbb{E}_n\left[\sum_{k=1}^{\infty}\left(\xi_1^{(n+1)} + \cdots + \xi_{Z_n}^{(n+1)}\right)\mathbb{1}_{\{Z_n=k\}}\right]\\
&= \sum_{k=1}^{\infty}\mathbb{E}_n\left[\left(\xi_1^{(n+1)} + \cdots + \xi_{Z_n}^{(n+1)}\right)\underbrace{\mathbb{1}_{\{Z_n=k\}}}_{\in\mathcal{F}_n}\right] && \textit{linearity, Prop. 10.19\#2}\\
&= \sum_{k=1}^{\infty}\mathbb{1}_{\{Z_n=k\}}\mathbb{E}_n\left[\left(\xi_1^{(n+1)} + \cdots + \xi_k^{(n+1)}\right)\right] && \textit{conditional determ. Prop. 10.23\#1}\\
&= \sum_{k=1}^{\infty}\mathbb{1}_{\{Z_n=k\}}\left(\mathbb{E}_n\left[\xi_1^{(n+1)}\right] + \cdots + \mathbb{E}_n\left[\xi_k^{(n+1)}\right]\right) && \textit{linearity}\\
&= \sum_{k=1}^{\infty}\mathbb{1}_{\{Z_n=k\}}k\mu && \mathbb{E}\left[\xi_i^{(n+1)}\right] = \mu \ \forall i\\
&= \mu\sum_{k=1}^{\infty}\mathbb{1}_{\{Z_n=k\}}k\\
&= \mu Z_n && \forall n
\end{aligned}$$

Clearly $(Z_n)_{n\in\mathbb{N}}$ is a martingale for $\mu = 1$, a submartingale for $\mu < 1$ and a supermartingale for $\mu > 1$. For free, we also get that $\left(\frac{Z_n}{\mu^n}\right)_{n\in\mathbb{N}}$ is a martingale since:

$$\mathbb{E}_n\left[\frac{Z_{n+1}}{\mu^{n+1}}\right] = \frac{1}{\mu^{n+1}}\mu Z_n = \frac{Z_n}{\mu^n} = \mathbb{E}_n\left[\frac{Z_n}{\mu^n}\right]$$

11.4 Wiener Processes

♠ **Definition 11.55** (Wiener process W). *A stochastic process $W = (W_t)_{t\in\mathbb{R}_+}$ is Wiener with respect to the filtration $\mathcal{F}$ if:*

1. *W is adapted to $\mathcal{F}$ (Def. 11.7)*

2. *Gaussian intervals*
$$\mathbb{E}_s[f(W_{s+t} - W_s)] = \int f(x)\frac{1}{\sqrt{2\pi t}}e^{-\frac{1}{2t}x^2}dx \quad \forall s,t,\ \forall f \in \mathcal{E}_+$$
Where $\mathcal{E}_+$ is to be intended as ***positive Borel functions mapping to*** *$\mathbb{R}$.*
3. $W_0 = 0$

♣ **Proposition 11.56** (Definitional implications of Wiener process). *We have that by requirement 2 of a Wiener process W:*

1. *Markov*
$$W_{s+t} - W_s \perp\!\!\!\perp \mathcal{F}_s\ \forall s$$
2. *stationarity*
$$W_{s+t} - W_s \perp\!\!\!\perp s\ \forall s$$
3. *normality*
$$W_{s+t} - W_s \sim \mathcal{N}(0,t)\ \forall t \neq 0$$

Proof. **(Claims #1, #2, #3)** the statement is $\forall f \in \mathcal{E}_+$, this includes indicator functions. Using indicators, we can state:

$$\begin{aligned}
\mathcal{P}_{W_{t+s}-W_s}(A) &= \mathbb{E}\left[\mathbb{1}_A \circ (W_{t+s} - W_s)\right] \\
&= \mathbb{E}\left[\mathbb{E}_s\left[\mathbb{1}_A \circ (W_{t+s} - W_s)\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}_{x\sim\mathcal{N}(0,1)}\left[\mathbb{1}_A\right]\right] && \text{Def. 11.55\#2} \\
&= \mathbb{E}\left[\int_A d\mathcal{P}_{X(t)}\right] && X(t) \sim \mathcal{N}(0,t) \\
&= \int_A d\mathcal{P}_{X(t)} && \forall A \in \mathcal{E}
\end{aligned}$$

This means that the integrals agree on every Borel set, and the two are equal in distribution in the sense of Definition 3.14. This proves claim #3 directly, and indirectly ensures independence from s and $\mathcal{F}_s$ since we notice that the filtration has no influence on the distribution. □

♢ **Observation 11.57** (About the third result). *For $0 < t_0 < t_1 < \ldots < t_n$ we implement Kolmogorov's Extension 10.39 to identify a single probability law, and decorate this result with Proposition 11.56:*

$$\begin{aligned}
\mathcal{P}\left((W_t)_{t\in\mathbb{T}}\right) \iff \mathcal{P}(W_{t_1},\ldots,W_{t_n}) &\iff \mathcal{P}(W_{t_1} - W_{t_0},\ldots,W_{t_n} - W_{t_{n-1}}) && \textit{Claim } 3 \\
&\iff \prod_{k=1}^{n-1}\mathcal{P}(W_{t_{k+1}-W_{t_k}}) && \textit{Claim } 2
\end{aligned}$$

Where for each interval we have that $\mathcal{N}(0, t_{i+1} - t_i)$ is the distribution.

♢ **Observation 11.58** (About the proof). *Notice that in principle having $\forall f \in \mathcal{E}_+$ makes the result definitional. Indicators are measurable and we recover the definition of equality in probability law. Using Theorem 4.11 is not as fast since we would need to check that continuous bounded functions are in $\mathcal{E}_+$ and then assessing equality of the integrals.*

♣ **Proposition 11.59** (Martingale characterization of Wiener Process, exponential).

$$W = (W_t)_{t\in\mathbb{R}_+}\ \underbrace{\textit{Wiener}}_{\textit{Def. 11.55}} \iff M_t = e^{rW_t - \frac{1}{2}r^2t}\ \underbrace{\mathcal{F}\textit{-martingale}}_{\textit{Def. 11.35}} \quad \forall r \in \mathbb{R}$$

Proof. ($\implies$) Let W be a Wiener process. The adaptedness of $(M_t)_{t\in\mathbb{T}}$ is trivial. We check the other two requirements of Definition 11.35.

(△ **integrability)** we get:

$$\begin{aligned}
\mathbb{E}[|M_t|] &= \mathbb{E}[M_t] && M_t > 0 \forall t \\
&= \mathbb{E}\left[\exp\left\{rW_t - \frac{1}{2}r^2 t\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{r(W_t - W_0) - \frac{1}{2}r^2 t\right\}\right] && W_0 = 0 \\
&= \\
&= e^{-\frac{1}{2}r^2 t}\mathbb{E}\left[\exp\left\{r(W_t - W_0)\right\}\right] \\
&= e^{-\frac{1}{2}r^2 t}\mathbb{E}\left[\mathbb{E}_0\left[\exp\left\{r(W_t - W_0)\right\}\right]\right] && \text{use Def. 11.55\#2} \\
&= e^{-\frac{1}{2}r^2 t}\int_{\mathbb{R}} e^{rx}\frac{1}{\sqrt{2\pi t}}e^{-\frac{1}{2t}x^2}dx && \text{is the mgf, general case: } mgf(r) = e^{\mu r + \frac{\sigma^2 r^2}{2}} \\
&= e^{-\frac{1}{2}r^2 t}e^{\frac{1}{2}r^2 t} = 1 < \infty
\end{aligned}$$

using the Mgf of a normal distribution.
(◯ **martingale equality)** an equivalent solution for $s < t$ is:

$$\mathbb{E}_s[M_t] = M_s \iff M_s\mathbb{E}_s\left[\frac{M_t}{M_s}\right] = M_s \iff \mathbb{E}_s\left[\frac{M_t}{M_s}\right] = 1$$

Which if we expand properly

$$\begin{aligned}
\mathbb{E}_s\left[\frac{M_t}{M_s}\right] &= \mathbb{E}_s\left[\exp\left\{rW_t - \frac{1}{2}r^2 t - rW_s + \frac{1}{2}r^2 s\right\}\right] \\
&= e^{-\frac{1}{2}r^2(t-s)}\mathbb{E}_s\left[\exp\left\{r(W_t - W_s)\right\}\right] \\
&= e^{-\frac{1}{2}r^2(t-s)}\mathbb{E}_s\left[f(W_t - W_s)\right] \\
&= e^{-\frac{1}{2}r^2(t-s)}\mathbb{E}_s\left[e^{rZ}\right] && Z \sim \mathcal{N}(0, t-s) \text{ by Wiener Def. 11.55\#2} \\
&= e^{-\frac{1}{2}r^2(t-s)}e^{+\frac{1}{2}r^2(t-s)} && \text{mgf of normal} \\
&= 1
\end{aligned}$$

which proves the martingale equality.
($\Longleftarrow$) to prove the reverse, let M be a martingale. Then $\mathbb{E}_s\left[\frac{M_s}{M_t}\right] = 1 \forall s < t$ so that:

$$\mathbb{E}_s\left[\exp\left\{r(W_{s+t} - W_s)\right\}\right] = e^{\frac{1}{2}r^2 t} \quad \forall r \in \mathbb{R} \iff \mathbb{E}_s[f(W_{t+s} - W_s)] = \mathbb{E}\left[f(X)\right] \quad X \sim \mathcal{N}(0, t)$$

As per Def. 11.55. The if and only if condition holds by a Theorem similar to the Laplace characterization (Thm. 6.12) for moment generating functions. □

♢ **Observation 11.60** (About W_t)**.** *Notice that by Definition 11.55#3 it is always safe to assume that* $W_t \sim \mathcal{N}(0, t)$ *by Prop. 11.56#3. We will avoid doing this calculation every time.*

♢ **Observation 11.61** (Exponential Martingale as stock)**.** *Let* $r \in \mathbb{R}$*, the exponential martingale we proved is the continuous time version of Example 11.46 with:*

$$R_n = \exp\left\{r\left(W_{n+1} - W_n\right) - \frac{1}{2}r^2(n+1-n)\right\} = \exp\left\{r\left(W_{n+1} - W_n\right) - \frac{1}{2}r^2\right\}$$

♣ **Proposition 11.62** (Wiener processes are martingales)**.**

$$W = (W_t)_{t\in\mathbb{R}_+} \underbrace{\textit{Wiener}}_{\textit{Def. 11.55}} \implies W \underbrace{\mathcal{F}\textit{-martingale}}_{\textit{Def. 11.35}}$$

Proof. Adaptedness follows by Definition 11.55#1.
(△ **integrability)** trivially

$$\mathbb{E}[|W_t|] < \infty$$

by the fact that it reduces to a $\mathcal{N}(0, t)$ random variable.
(□ martingale equality) for $s < t$:

$$\begin{aligned}
\mathbb{E}_s[W_t - W_s] &= \mathbb{E}[W_t - W_s] && W_t - W_s \perp\!\!\!\perp \mathcal{F}_s \text{ Def. } 11.56\#1 \\
&= 0 && W_t - W_s \sim \mathcal{N}(0, t-s) \text{ Prop. } 11.56\ \#3
\end{aligned}$$

So that W is an $\mathcal{F}$-martingale. □

♣ **Proposition 11.63** (Martingale characterization of Wiener process, square)**.**

$$W = (W_t)_{t\in\mathbb{R}_+} \underbrace{\textit{Wiener}}_{\textit{Def. } 11.55} \implies Y_t = W_t^2 - t \quad \underbrace{\mathcal{F}\textit{-martingale}}_{\textit{Def. } 11.35}$$

Proof. **(** $\implies$ **)(△ adaptedness)** Y_t is adapted to $\mathcal{F}_t$ trivially.
(□ integrability) follows by the triangle inequality:

$$\mathbb{E}[|Y_t|] = \mathbb{E}[|W_t^2 - t|] \leq \mathbb{E}[|W_t^2|] + \mathbb{E}[|t|] < \infty$$

Where the first term is finite since the variance and mean of the Wiener process is well defined via its normal distribution.
(◯ martingale equality) let $s < t$. The aim is reducing the expression to differences $W_t - W_s$ to exploit the property of Wiener processes. This is easily done as:

$$\begin{aligned}
\mathbb{E}_s\left[Y_t - Y_s\right] &= \mathbb{E}_s\left[W_t^2 - W_s^2 - t + s\right] \\
&= \mathbb{E}_s\left[W_t^2 - W_s^2 - t + s \pm 2W_sW_t \pm 2W_s^2\right] \\
&= \mathbb{E}_s\left[\left(W_t - W_s\right)^2 + 2W_s\left(W_t - W_s\right) - t + s\right] \\
&= \mathbb{E}_s\left[\left(W_t - W_s\right)^2\right] + 2W_s\mathbb{E}_s\left[W_t - W_s\right] - t + s && \text{linearity and conditional determinism} \\
&= \mathbb{E}\left[\left(W_{t-s}\right)^2\right] + 2W_s\mathbb{E}_s\left[W_t - W_s\right] - t + s && \mathbb{E}\left[(W_{t-s})^2\right] = V\left[W_{t-s}\right] - \left(\mathbb{E}\left[W_{t-s}\right]\right)^2 \\
&= t - s + 2W_s\underbrace{\mathbb{E}\left[W_t - W_s\right]}_{=0} - t + s && W_t - W_s \sim \mathcal{N}(0, t-s) \perp\!\!\!\perp \mathcal{F}_s \\
&= 0
\end{aligned}$$

Following the Wiener properties of Proposition 11.56, linearity of the expectation, and conditional determinism (Prop. 10.23#1).
By △, □, ◯ the process $(Y_t)_{t\in\mathbb{T}}$ is a $\mathcal{F}$-martingale. □

♣ **Theorem 11.64** (Combination of Wiener martingale characterization)**.** *It actually holds that W is Wiener if and only if:*

1. *W is an $\mathcal{F}$-martingale*
2. *$Y_t = W_t^2 - t$ is an $\mathcal{F}$-martingale*

Namely, the ***two previous results together*** *characterize Wiener processes.*

Chapter Summary

Objects:

- filtrations and random times:
  - filtration $\mathcal{F}$ a sequence of increasing σ-algebras and its properties
  - stopping times, measurable random functions with respect to a filtration
  - the counting process $N_t = \sum_n \mathbb{1}_{[0,t]}(T_n)$ for $(T_n)_{n\in\mathbb{N}}$ arrival random times
  - the end of time filtration
  - the stopped filtration, for T a stopping time and $\mathcal{F}$ a filtration extended to $\overline{\mathbb{T}}$

$$\mathcal{F}_T = \left\{H \in \mathcal{H} : H \cap \{T \leq t\} \in \mathcal{F}_t \ \forall t \in \overline{\mathbb{T}}\right\}$$

- martingales
  - conditional expectation with respect to a filtration recap
  - $\mathbb{E}_t$ notation, $\mathbb{E}_T$ notation
  - martingale as an adapted integrable process that tends to be stable over time
  - uniformly integrable martingales

$$(X_t)_{t\in\mathbb{T}} \quad \lim_{b\to\infty} \sup_t \mathbb{E}|X_t| \mathbb{1}_{[b,\infty)}(|X_t|) = 0$$

  - Wiener process, a stochastic process adapted to $\mathcal{F}$ and starting at 0 such that:

$$\mathbb{E}_s[f(W_{s+t} - W_s)] = \int f(x) \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}x^2} dx \quad \forall s,t,\ \forall f \in \mathcal{E}_+$$

Results:

- filtrations and random times
  - the stopped filtration is a σ-algebra and $\mathcal{F}_T \subset \mathcal{F}_\infty \subset \mathcal{H}$
  - it holds $V \in \mathcal{F}_T \iff V\mathbb{1}_{\{T\leq t\}} \in \mathcal{F}_t \forall t \in \overline{\mathbb{T}}$ and just the equality indicator needs to be checked for discrete processes.
  - the algebra of deterministic times extends to random stopping times (measurability grants this)
  - the counting process with arrival stopping times is such that $T = \inf\{t \geq a : N_t = N_{t-a}\}$ for $a > 0$ is a stopping time.
- martingales
  - all the properties of conditional expectation naturally extend to $\mathbb{E}_t$
  - for the stopped filtration expectation, we only need to prove the projection property since $S \leq T$ is more articulate
  - martingale implies stationary
  - Jensen's for martingales: for f convex and $f \circ X$ integrable the process $f \circ X$ is a submartingale
  - if $Z \in \mathcal{L}_1$ then $\mathbb{E}_t[Z]$ is a u.i. martingale
  - Wiener processes are stationary, Markovian and normal at interval differences
  - W is Wiener if and only if the exponential process is a martingale
  - W is Wiener if and only if both W and $W_t^2 - t$ are a martingale, but with the only if direction we can say that a Wiener process is a martingale

Chapter 12

More Processes & Integration

12.1 Poisson processes

◇ **Observation 12.1** (Recap of Counting process). *Recall Definition 11.13. $(N_t)_{t\in\mathbb{T}}$ starts at $N_0 = 0$, is right continuous, increasing, with jumps of size 1. Using the notion of distinct arrival times $0 < T_1 < T_2 < \dots$ it can be seen as:*

$$N_t = \sum_{n=1}^{\infty} \mathbb{1}_{[0,t]}(T_n)$$

♠ **Definition 12.2** (Poisson Process $\mathcal{P}ois(c)$). *A counting process $(N_t)_{t\in\mathbb{T}}$ is Poisson with rate $c > 0$ with respect to a filtration $\mathcal{F}$ (Def. 11.2) when:*

1. *N is adapted to $\mathcal{F}$ (Def. 11.7)*
2. *increments are Poisson distributed in expectation:*

$$\mathbb{E}_s[f(N_{s+t} - N_s)] = \sum_{k=0}^{\infty} \frac{e^{-ct}(ct)^k}{k!} f(k) \quad \forall s,t,\ f \in \mathcal{E}^+$$

Where, as usual, by $f \in \mathcal{E}^+$ we mean a positive measurable function in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the space where the process takes values on.

♣ **Proposition 12.3** (Definitional Properties of $\mathcal{P}ois(c)$). *Definition 12.2 has some direct implications. For $N \sim \mathcal{P}ois(c)$ it holds that:*

1. *markov property*
$$N_{s+t} - N_s \perp\!\!\!\perp \mathcal{F}_s \quad \forall s,t$$
2. *stationarity*
$$(N_t)_{t\in\mathbb{T}} \perp\!\!\!\perp t$$
3. *Poisson increments*
$$N_{t+s} - N_s \sim \mathcal{P}o(ct)$$

Proof. Follow the approach of Proposition 11.56. □

♣ **Theorem 12.4** ($\mathcal{P}ois(c)$ characterization). *For a counting process N (Def. 11.13) and a filtration $\mathcal{F}$ over which it is a Poisson process we can see that:*

$$N \sim \mathcal{P}ois(c) \iff (N_t - ct)_{t\in\mathbb{T}} \quad \mathcal{F}\text{-martingale (Def. 11.35)}$$

Proof. **($\implies$)($\triangle$ adaptedness)** we have $N_t - ct \in \mathcal{F}_t \forall t \in \mathbb{T}$ since $N_t \in \mathcal{F}_t$ by the Poisson process being adapted to the filtration (Def. 12.2#1) so adaptedness is verified.
($\square$ integrability) by the triangle inequality and the integrability of N_t we have:

$$\mathbb{E}[|N_t - ct|] \leq \mathbb{E}[|N_t|] + \mathbb{E}|ct| < \infty$$

Which follows by $\mathbb{E}[|N_t|] = \mathbb{E}[N_t] < \infty$, a result of Proposition 11.14.
(◯ martingale equality) By Proposition 12.3#3 we have $\mathbb{E}_s[N_{t+s} - N_s] = ct$ so that for $s < t$:

$$\begin{aligned}\mathbb{E}_s[N_t + ct - N_s - cs] &= \mathbb{E}_s[N_t - N_s] - c(t-s) \\ &= c(t-s) - c(t-s) && \text{Prop. 12.3\#3} \\ &= 0\end{aligned}$$

By △, □, ◯ The process $(N_t - ct)_{t\in\mathbb{T}}$ is a $\mathcal{F}$-martingale.
($\Longleftarrow$) shown in Proposition C.28. □

12.2 Stochastic Integrals

♠ **Definition 12.5** (Predictable process). *A natural process $(F_n)_{n\in\mathbb{N}}$ is predictable with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ when $F \in \mathcal{F}_0$ and $F_{n+1} \in \mathcal{F}_n \ \forall n$, where by $\in$ we mean measurable with respect to (see Def. 11.1 for context).*

♢ **Observation 12.6** (Interpreting integration of Example 11.46 and Observation 11.47). *For $(R_n)_{n\in\mathbb{N}}$ independent and positive with $\mathbb{E}[R_n] = 1$ and $V[R_n] < \infty \ \forall n$, recalling Observation 11.47 we can characterize a stock price in equilibrium as a martingale where:*

$$M_0 = 1 \quad M_n = M_0 \prod_{k=1}^{n} R_k \quad R_{n+1} = \frac{M_{n+1}}{M_n}$$

Further, denote F_{n+1} as the number of shares over the interval $(n, n+1]$ so that:

$$(M_{n+1} - M_n)F_{n+1} \coloneqq \text{profit or loss over } (n, n+1]$$

The martingale $(M_n)_{n\in\mathbb{N}}$ with respect to $\mathcal{F}$ satisfies the martingale equality $\mathbb{E}_n[M_{n+1} - M_n] = 0 \ \forall n$.
If we assume further that $(F_n)_{n\in\mathbb{N}}$ is predictable wrt $\mathcal{F}$ (Def. 12.5) letting:

- *$X_0 = M_0F_0$ be the initial price times number of shares*
- *$(X_n)_{n\in\mathbb{N}} \ : \ \forall n \quad X_n = X_0M_0 + \sum_{m=1}^{n}(M_m - M_{m-1})F_m$*

*We can see this as a model for the portfolio value, as **stochastic** integral of $(F_n)_{n\in\mathbb{N}}$ with respect to $(M_n)_{n\in\mathbb{N}}$.*

♠ **Definition 12.7** (Steltjes-Lebesgue integral). *For $(F_n)_{n\in\mathbb{N}}$ a random function and $(M_n)_{n\in\mathbb{N}}$ a signed measure with mass $M_n - M_{n-1} \ \forall n$ and $M_0 = 1$ we define:*

$$\begin{aligned} X = (X_n)_{n\in\mathbb{N}} \quad &: \quad X = \int F dM \\ &\iff X_n = \int_{[0,n]} F dM = M_0F_0 + \sum_{m=1}^{n}(M_m - M_{m-1})F_m\end{aligned}$$

A series of increasing in n integrals.

♡ **Example 12.8** (An investment strategy). *Let T be the random time to exit the market. Assume T is a stopping time wrt $(\mathcal{F}_n)_{n\in\mathbb{N}}$ as per Definition 11.9.*
Let $F_n = \mathbb{1}_{[0,T]}(n)$ be a random indicator. Then:

$$\begin{aligned} X_n &= \int_{[0,n]} F dM \\ &= \int_{[0,n]} \mathbb{1}_{[0,T]} dM \\ &= \int \mathbb{1}_{[0,n]} \mathbb{1}_{[0,T]} dM \\ &= \int \mathbb{1}_{[0,n\wedge T]} dM \\ &= M_{n\wedge T} \\ &= \begin{cases} M_n & n < T \\ M_T & n \geq T \end{cases}\end{aligned}$$

Namely, the simplest strategy one can think of invest everything up to time T, and sell right after. With perfectly shared information, there is no profit or loss.
We will show in Corollary 12.10 that up to reasonable conditions this is again a martingale.

♣ **Theorem 12.9** (Martingality of integral for bounded processes)**.** *For X as in Definition 12.7 with $(F_n)_{n\in\mathbb{N}}$ bounded (i.e. $\mathbb{P}[|F_n| \le b] = 1$ for some $b \in \mathbb{R}$) it holds that:*

1. *$(M_n)_{n\in\mathbb{N}}$ martingale $\Longrightarrow$ X martingale*
2. *$(M_n)_{n\in\mathbb{N}}$ submartingale, $(F_n)_{n\in\mathbb{N}}$ positive $\forall n \implies X$ submartingale*

Proof. **(Claim #1)** recall $X = \int FdM$ and proceed as follows.
(△ integral expression) for all n it holds:

$$X_n = \int_{[0,n]} FdM = \sum_{m=1}^{n} (M_m - M_{m-1})F_m + M_0F_0$$

where $\{F_0, \dots, F_n\} \subset \mathcal{F}_n$ by the fact that $\mathcal{F}$ is a filtration and F is predictable (Def. 12.5). The same holds for $\{M_m\}_{m=0}^n$.
(□ boundedness implication) let $|F_n| < b$ a.s. for some $b \in \mathbb{R}$. Then:

$$\begin{aligned}
\mathbb{E}[|X_n|] &= \mathbb{E}\left[\left|M_0F_0 + \sum_{m=1}^{n}(M_m - M_{m-1})F_m\right|\right] \\
&\le \mathbb{E}\left[|M_0F_0|\right] + \mathbb{E}\left[\sum_{m=1}^{n}|(M_m - M_{m-1})F_m|\right] \\
&\le b\left(\underbrace{\mathbb{E}\left[|M_0|\right]}_{<\infty} + \mathbb{E}\left[\underbrace{\sum_{m=1}^{n}|M_m - M_{m-1}|}_{\le \sum_m \mathbb{E}[|M_m|] + \sum_m \mathbb{E}[|M_{m-1}|] < \infty}\right]\right) \\
&< \infty
\end{aligned}$$

where the last bounds follow by the fact that M is a martingale. **(◯ martingale equality)** By Proposition 11.40 we only check for one time step and get:

$$\begin{aligned}
\mathbb{E}_n[X_{n+1} - X_n] &= \mathbb{E}_n[(M_{n+1} - M_n)F_{n+1}] && \text{construction} \\
&= F_{n+1}\mathbb{E}_n[M_{n+1} - M_n] && F_{n+1} \in \mathcal{F}_n \text{ predictable process} \\
&= 0 && (M_n)_{n\in\mathbb{N}} \text{ martingale}
\end{aligned}$$

By △, □, ◯ $(X_n)_{n\in\mathbb{N}}$ is a martingale.
(Claim #2) similar to #1. □

Corollary 12.10 (Stopped time process martingality)**.** *Let T be a stopping time (Def. 11.9) and $(X_n)_{n\in\mathbb{N}}$ with $X_n = M_{n\wedge T}$ as in Example 12.8. Then:*

1. *$(M_n)_{n\in\mathbb{N}}$ martingale $\Longrightarrow$ X martingale*
2. *$(M_n)_{n\in\mathbb{N}}$ submartingale $\Longrightarrow$ X submartingale*

Notice that by the result of Example 12.8 this means that $M_{n\wedge T}$ is a martingale/submartingale since $X_n = M_{n\wedge T}$ for all $n \in \mathbb{N}$.

Proof. **(Claims #1#2)** same as Theorem 12.9 noting that:

$$X_n = \int_{[0,n]} FdM \quad : \quad F_n = \mathbb{1}_{[0,T]}(m) \le 1 \quad \forall T$$

So that the process $(F_n)_{n\in\mathbb{N}}$ is bounded, positive and predictable. □

◇ **Observation 12.11** (From deterministic to random times)**.** *Recall the martingale equality (Def. 11.35#3):*

$$\mathbb{E}_s[M_t - M_s] = 0 \quad s < t \overset{Prop.\ 11.39}{\Longrightarrow} \mathbb{E}\left[\mathbb{E}_s[M_t - M_s]\right] = \mathbb{E}\left[M_t - M_s\right] = 0 \ s < t \quad \& \quad \mathbb{E}[M_t] = \mathbb{E}[M_0] \ \forall t$$

What about $\mathbb{E}_S[M_T - M_S]$ for S, T stopping times?

♡ **Example 12.12** (Random times alone do not satisfy martingale equality in a symmetric random walk)**.** *Let:*

$$S_0 = 1, \ S_n = S_{n-1} + \xi_n \quad \xi_n \sim \mathcal{B}ern_{\pm 1}\left(\frac{1}{2}\right), \ \xi_n \in \{-1, +1\} \quad \forall n$$

Assign $X_n = S_{T \wedge n}$. To see that $(S_n)_{n\in\mathbb{N}}$ is a martingale, refer to Example 11.44. To see that X_n is a nonnegative martingale, use Theorem 12.9. In particular $\mathbb{E}[S_n] = \mathbb{E}\left[\underbrace{\mathbb{E}_0[S_n]}_{=\mathbb{E}_0[S_0]}\right] = \mathbb{E}\left[1\right] = 1 \forall n$. Consider the random time:

$$T = \inf\left\{k : S_k = 0\right\}$$

T is a stopping time wrt $\mathcal{F} = \sigma\left((S_n)_{n\in\mathbb{N}}\right)$ (for this recover Example 11.27) or observe that:

$$\{T \le t\} \subset \bigcup_{k \le t, k\in\mathbb{N}} \{S_k = 0\} \in \mathcal{F}_t$$

Namely, the sum being equal to zero is included in the event that at least one of the times before the sum has reached zero which is in the increasing filtration. Clearly, $S_T = 0$ with probability 1 since at time T the martingale will be certainly null but:

$$0 = \mathbb{E}[S_T] \neq \mathbb{E}[S_0] = 1$$

12.3 Doob's results and Martingale Convergence

♣ **Proposition 12.13** (Doob's Theorem I)**.** *Let $(M_n)_{n\in\mathbb{N}}$ be a martingale, T a stopping time (Defs. 11.35, 11.9), with T bounded $\mathbb{P}[T \le k] = 1$ for some $k \in \mathbb{R}$. Then*

$$\mathbb{E}[M_0] = \mathbb{E}[M_T] = \mathbb{E}[M_k]$$

Proof. Consider M_T and $(X_n)_{n\in\mathbb{N}} = (M_{n\wedge T})_{n\in\mathbb{N}}$ like in Example 12.8. By Corollary 12.10 X is a martingale. Moreover:

$$\begin{aligned}
\mathbb{E}\left[M_0\right] &= \mathbb{E}\left[M_{0\wedge T}\right] && 0 = 0 \wedge T \\
&= \mathbb{E}\left[M_{k\wedge T}\right] && \text{martingale eq. at } k \\
&= \mathbb{E}\left[M_T\right] && \mathbb{P}[T \le k] = 1
\end{aligned}$$

We want to show that $\mathbb{E}\left[M_T\right] = \mathbb{E}\left[M_k\right]$. For this purpose, consider $(F_n)_{n\in\mathbb{N}} = \left(\mathbb{1}_{(T,\infty)}(n)\right)_{n\in\mathbb{N}}$ so that F is predictable (this is shown in the second item of Example 12.17 with $V = 1$) and bounded. Then by Theorem 12.9 the integral $X' = \int F dM$ is a martingale. Additionally notice that $\mathbb{1}_{(T,\infty)}(n) = 1 - \mathbb{1}_{[0,T]}(n)$.
By these last two facts we have that the process:

$$X'_n = \int_{[0,n]} \left(1 - \mathbb{1}_{[0,T]}\right) dM = M_n - M_{n\wedge T}$$

has expectation:

$$\begin{aligned}
\mathbb{E}\left[M_n - M_{n\wedge T}\right] &= \mathbb{E}\left[M_k - M_{k\wedge T}\right] && \text{martingale equality at } k \\
&= \mathbb{E}\left[M_0 - M_{0\wedge T}\right] && \text{martingale equality at zero} \\
&= \mathbb{E}\left[M_0 - M_0\right] = 0
\end{aligned}$$

Thus, taking the red terms and the final equality one gets:

$$0 = \mathbb{E}\left[M_k - M_{k\wedge T}\right] = \mathbb{E}\left[M_k\right] - \mathbb{E}\left[M_{k\wedge T}\right] = \mathbb{E}\left[M_k\right] - \mathbb{E}\left[M_T\right] \iff \mathbb{E}\left[M_k\right] = \mathbb{E}\left[M_T\right]$$

and eventually $\mathbb{E}\left[M_0\right] = \mathbb{E}\left[M_T\right] = \mathbb{E}\left[M_k\right]$. □

Corollary 12.14 (Double stopping Time Doob's Theorem I). *Let $(M_n)_{n\in\mathbb{N}}$ be a martingale, T a stopping time (Defs. 11.35, 11.9), with T bounded $\mathbb{P}[T \leq k] = 1$ for some $k \in \mathbb{R}$ as before. If $S \leq T$ is another stopping time:*

$$\mathbb{E}[M_S] = \mathbb{E}[M_T]$$

Proof. The boundedness of T a.s. implies the boundedness of S almost surely. An application of Proposition 12.13 for S has the same result of T. Clearly:

$$\mathbb{E}[M_0] = \mathbb{E}[M_T] = \mathbb{E}[M_k] = \mathbb{E}[M_S]$$

□

♣ **Theorem 12.15** (Doob's decomposition). *Let $(X_n)_{n\in\mathbb{N}}$ be adapted to $\mathcal{F}$ and integrable $X_n \in \mathcal{L}_1$. The following statements are true:*

1. *decomposition, with M a martingale, $M_0 = 0$, and $(A_n)_{n\in\mathbb{N}}$ a predictable process, $A_0 = 0$ (Defs. 11.35, 12.5)*
$$X_n = X_0 + M_n + A_n \quad \forall n \in \mathbb{N}$$
2. *the decomposition at point 1 is unique up to equivalence*
3. *if $(X_n)_{n\in\mathbb{N}}$ is a submartingale, $(A_n)_{n\in\mathbb{N}}$ is an increasing predictable process, if $(X_n)_{n\in\mathbb{N}}$ is a supermartingale, $(A_n)_{n\in\mathbb{N}}$ is decreasing predictable*

Proof. **(Claim #1)** Let $M_0 = A_0 = 0$. We define M, A via increments:

$$A_{n+1} - A_n = \mathbb{E}_n[X_{n+1} - X_n] \quad M_{n+1} - M_n = (X_{n+1} - X_n) - (A_{n+1} - A_n) \quad \forall n \in \mathbb{N}$$

Then:

$$X_n = X_0 + M_n + A_n \quad \forall n \in \mathbb{N} \tag{12.1}$$

and M is a martingale since $\mathbb{E}_n[M_{n+1} - M_n] = \mathbb{E}_n[X_{n+1} - X_n] - \mathbb{E}_n[X_{n+1} - X_n] = 0$, with $(A_n)_{n\in\mathbb{N}}$ predictable since $A_{n+1} = \mathbb{E}_n[X_{n+1} - X_n] + A_n \in \mathcal{F}_n$ by the expectation being constructed as to be measurable. The decomposition is valid.
(Claim #3) for X a submartingale, it holds that the expectation of the difference is positive, making $(A_n)_{n\in\mathbb{N}}$ increasing. The reverse holds for a supermartingale.
(Claim #2) assume there is another decomposition $X = X_0 + M' + A'$, it holds that:

$$B = A - A' = M - M'$$

is a predictable (by A) martingale (by M). With the predictability, we can measure with $\mathbb{E}_n$ also B_{n+1}, with the martingality, we can use the martingale equality. Then:

$$B_{n+1} - B_n = \mathbb{E}_n[B_{n+1} - B_n] = 0 \quad \forall n \in \mathbb{N} \implies B_n \overset{a.s.}{=} 0 \implies A \overset{a.s.}{=} A',\ M \overset{a.s.}{=} M'$$

□

♢ **Observation 12.16** (About Doob's decomposition). *We can see that:*

$$X_{n+1} - X_n = \underbrace{A_{n+1} - A_n}_{\textit{prediction process}} + \underbrace{M_{n+1} - M_n}_{\textit{innovation process}}$$

Where on the right hand side, the first process is known at n, and the second is the extra information provided by the martingale.

♡ **Example 12.17** (Some predictable processes and their integral martingales). *we present four easy examples with $S \leq T$ almost surely two stopping times and $V \in \mathcal{F}_S$.*
(one extreme) *let T be a stopping time. Then the process $F_n = \mathbb{1}_{[0,T]}(n)$ is such that :*

$$F_{n+1} = \mathbb{1}_{[0,T]}(n+1) = \mathbb{1}_{\{n+1 \leq T\}} = \mathbb{1}_{\{T \leq n\}^c} \in \mathcal{F}_n$$

Since T is a stopping time. Clearly the process $(F_n)_{n\in\mathbb{N}}$ is predictable.
(other extreme) *Let $F_n = V\mathbb{1}_{(S,\infty)}(n)$ for $S \le T$ two stopping times, and $V \in \mathcal{F}_S$. Then:*

$$V \in \mathcal{F}_S \implies F_{n+1} = V\mathbb{1}_{(S,\infty)}(n+1) = V\mathbb{1}_{\{n+1\ge S\}} = V\mathbb{1}_{\{S\le n\}^c} \in \mathcal{F}_n$$

Where we applied Theorem 11.23#1. The process is predictable.
(two extremes) *for $F_n = \mathbb{1}_{(S,T]}(n) = \mathbb{1}_{[S,\infty)}(n)\cdot\mathbb{1}_{[0,T)}(n)$ the product of two predictable processes, the process is predictable.*
(two extremes + V) *let $F_n = V\mathbb{1}_{(S,T]}(n)$, the result is trivial by the previous ones.*
For all three cases, we have a stochastic integral:

$$X_n = \int_{[0,n]} FdM$$

Where $(X_n)_{n\in\mathbb{N}}$ is a martingale by Theorem 12.9.

♣ **Theorem 12.18** (Doob's Theorem II, fully general)**.** *For a process $(M_n)_{n\in\mathbb{N}}$ adapted to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ the following are equivalent:*

1. *$(M_n)_{n\in\mathbb{N}}$ is a martingale*
2. *for bounded stopping times $S \le T$ M_S and M_T are integrable and $\mathbb{E}_S[M_T - M_S] = 0$*
3. *for bounded stopping times $S \le T$ M_S and M_T are integrable and $\mathbb{E}[M_T - M_S] = 0$*

Proof. **(△ strategy)** denoting the triple equivalence as ①②③ we prove ① $\implies$ ②, ② $\implies$ ③, ③ $\implies$ ①. We also have in hand ① $\implies$ ③ which is Corollary 12.14.
(**①** $\implies$ **②**) by hypothesis M is a martingale and $S(\omega) \le T(\omega) \le n$ for non negligible $\omega \in \Omega$ and $n \in \mathbb{N}$. Let $V \in \mathcal{F}_S$ be bounded and positive, the process $F = \mathbb{1}_{(S,T]}$ is predictable as shown in Example 12.17. The stochastic integral is:

$$X_n = \int_{[0,n]} FdM = \int_{[0,n]} V\mathbb{1}_{(S,T]}dM = V\cdot(M_T - M_S) + X_0 \implies X_n - X_0 = V\cdot(M_T - M_S)$$

By Theorem 12.9#1 or better Corollary 12.10, since V is bounded so that F is predictable bounded, the process $(X_n)_{n\in\mathbb{N}}$ is a martingale. We show integrability of M_S and M_T by choosing the particular cases:

- $V = 1, S = 0 \implies M_T \in \mathcal{L}_1$
- $V = 1, T = n \implies M_S \in \mathcal{L}_1$

Eventually:

$$\begin{aligned}
\mathbb{E}\left[V\mathbb{E}_S[M_T - M_S]\right] &= \mathbb{E}\left[V(M_T - M_S)\right] && \text{expectation defining property, Def. 10.12\#2}\\
&= \mathbb{E}\left[X_n - X_0\right] && \text{above result}\\
&= 0 && \text{martingale equality}
\end{aligned}$$

By the arbitrariness of $V \in \mathcal{F}_S$ positive and bounded, we have as result that $\mathbb{E}_S[M_T - M_S] = 0$.
(**②** $\implies$ **③**) taking expectations:

$$\mathbb{E}\left[\underbrace{\mathbb{E}_S[M_T - M_S]}_{=0}\right] = \mathbb{E}\left[M_T - M_S\right] = 0$$

(**③** $\implies$ **①**) for $T = n$, M_n is clearly integrable by M_T begin integrable and $n = T$ as choice. Adaptedness holds by hypothesis, and the martingale equality reads:

$$\mathbb{E}_m[M_n - M_m] = 0 \forall m, n \in \mathbb{N},\ n \ge m \iff \mathbb{E}\left[\mathbb{1}_H \mathbb{E}_m[M_n - M_m]\right] = 0 \forall m, n \in \mathbb{N},\ n \ge m \quad H \in \mathcal{F}_m$$

For this purpose, we fix $m, n \in \mathbb{N}, H$ and for $\omega \in \Omega$ we let:

$$S(\omega) = m, \quad T(\omega) = n\mathbb{1}_H(\omega) + m\mathbb{1}_{\Omega\setminus H}(\omega)$$

Which makes sense since by fixing ω random times have become deterministic. S is a fixed time and $T \geq S$ is a stopping time. By the fact that $H \in \mathcal{F}_S, T \geq S$ the time T is foretold at $S = m$. It clearly holds $S \leq T \leq n$. Eventually:

$$M_T - M_S = \mathbb{1}_H \cdot (M_n - M_m)$$

by construction. The hypothesis of (3) implies the claim of (1). □

♢ **Observation 12.19** (Why Theorem 12.18 does not work in Example 12.12). *The random time T maps to $\overline{\mathbb{R}}$ and fails to be bounded in probability, i.e. $\mathbb{P}[T > k] > 0\ \forall k$.*

♢ **Observation 12.20** (Recap about concepts needed). *Recall the min max convention and the useful decompositions:*

$$x^+ = \max\{x, 0\} = x \vee 0$$
$$x^- = -\min\{x, 0\} = -(x \wedge 0)$$

So that:

$$x = x^+ - x^- \qquad |x| = x^+ + x^- = 2x^+ - x$$

We will also make use of:

- *Fatou's Lemma (Lem. A.49)*

$$(X_n)_{n\in\mathbb{N}} \textit{ nonnegative} \implies \mathbb{E}[\liminf_n X_n] \leq \liminf_n \mathbb{E}[X_n]$$

- *Borel Cantelli Lemma 1 (Thm. 9.6)*

$$(H_n)_{n\in\mathbb{N}} \quad \sum \mathbb{P}[H_n] < \infty \implies \mathbb{P}[\limsup_n H_n] = \mathbb{P}[H_n \; i.o.] = 0$$

- *Borel Cantelli Lemma 1 almost sure version (Ex. 9.9)*

$$(X_n)_{n\in\mathbb{N}}, X \quad \sum_{n\geq 1} \mathbb{P}[|X_n - X| > \epsilon] < \infty \quad \forall \epsilon > 0 \implies X_n \overset{a.s.}{\to} X$$

♠ **Definition 12.21** (Upcrossing, downcrossing and counter). *Let $(M_n)_{n\in\mathbb{N}}$ be adapted to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ (Def. 11.7), $a < b$ and $T_0 = -1$ for convenience.*
For all natural $k \geq 1$ define:

$$S_k := \inf\{n \geq T_{k-1} : M_n \leq a\}$$
$$T_k := \inf\{n \geq S_k : M_n \geq b\}$$

By the adaptedness of $(M_n)_{n\in\mathbb{N}}$ we can say that $\{S_1, T_1, S_2, T_2, \ldots\}$ is an increasing sequence of stopping times (Def. 11.9).
S_k can be seen as the k^{th} downcrossing of the interval (a, b), while T_k is the k^{th} upcrossing of the interval (a, b). We then define the number of upcrossings of (a, b) as:

$$U_n(a, b) = \sum_{k=1}^{\infty} \mathbb{1}_{[0,n]}(T_k)$$

♡ **Example 12.22** (Visualizing the definition). *Consider Figure 12.1. This could describe a buy/sell strategy for stocks with price $(M_n)_{n\in\mathbb{N}}$.*

♠ **Definition 12.23** (F_n formalism). *The number of buy/sell cycles in $[0, n]$ is exactly $U_n(a, b)$. In this context we let:*

- *F_n be such that:*

$$\begin{cases} F_n = \sum_{n=1}^{\infty} \mathbb{1}_{(S_k, T_k]}(n) \\ F_0 = 0 \end{cases} = \begin{cases} 1 & if\ \exists k : S_k < n \leq T_k \\ 0 & else \end{cases}$$

So that F represents the number of stocks owned at $(n, n+1]$

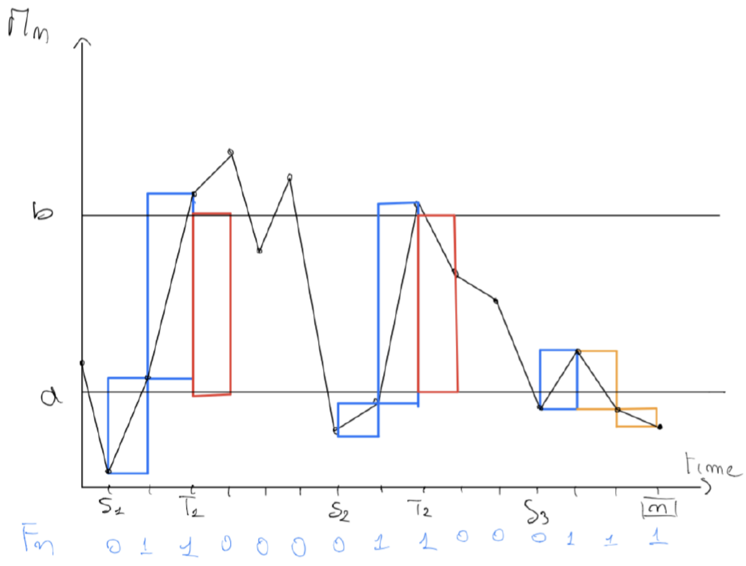

Figure 12.1: Upcrossings of an imaginary stock price

- *we already saw that the value of the portfolio is formalized as:*

$$\begin{cases} X_n = \int_{[0,n]} F dM \\ X_0 = 0 \end{cases}$$

With this context, the profit is in general:

$$X_n - X_0 \geq (b-a)U_n(a,b)$$

where we put $\geq$ instead of $=$ since it could be that the price at the end is less than the price at the start! See the plot of Example 12.22 for reference.

♣ **Proposition 12.24** (Upcrossing inequality)**.**

$$(M_n)_{n\in\mathbb{N}} \quad \textit{submartingale} \implies (b-a)\mathbb{E}[U_n(a,b)] \leq \mathbb{E}[(M_n-a)^+ - (M_0-a)^+]$$

Proof. The upcrossings of (a,b) by $(M_n)_{n\in\mathbb{N}}$ are equivalent to the upcrossing of $(0, b-a)$ by $((M_n-a)^+)_{n\in\mathbb{N}}$. With this new formulation using the fact that max is convex and Jensen's equality for martingales with $f(x) = (x-a)^+$ (Cor. 11.41):

$$((M_n-a)^+)_{n\in\mathbb{N}} \text{ submartingale } \implies \mathbb{E}\left[(M_n-a)^+ - (M_m-a)^+\right] \geq 0 \quad \forall n \geq m$$

Without loss of generality assume $a = 0$ so that $M_n \geq 0 \forall n$. We eventually want to show:

$$b\mathbb{E}\left[U_n(0,b)\right] \leq \mathbb{E}\left[M_n - M_0\right]$$

($\triangle$ first result) Use for this purpose $(F_n)_{n\in\mathbb{N}}$ from Definition 12.23 and $X_n = \int_{[0,n]} F dM$. Then:

$$M_{S_k} = 0 \forall k \implies bU_n(0,b) \leq X_n - X_0$$

Since we are always priced more. We now want to show:

$$\mathbb{E}\left[X_n - X_0\right] \leq \mathbb{E}\left[M_n - M_0\right]$$

($\square$ second result) Notice further that F is defined as:

$$F_n = \sum_n \mathbb{1}_{(S_k, T_k]}(n)$$

which is predictable by being a sum of predictable processes. Additionally, it is positive and bounded by 1 since we only keep one share at a time. Then, by Theorem 12.9#2 $X = (X_n)_{n\in\mathbb{N}}$ is a submartingale.
(◯ **conclusion)** it holds:

$$\begin{aligned} 0 \le \mathbb{E}_k\{X_{k+1}-X_k\} &= \mathbb{E}_k\{F_{k+1}(M_{k+1}-M_k)\} \\ &= F_{k+1}\mathbb{E}_k\{M_{k+1}-M_k\} && F_{k+1}\in\mathcal{F}_k \text{ predictable} \\ &\le \mathbb{E}_k\{M_{k+1}-M_k\} && F_{k+1}\le 1 \text{ bounded} \end{aligned}$$

giving as result:

$$\begin{aligned} \sum_k \mathbb{E}\left[\mathbb{E}_k\left[X_{k+1}-X_k\right]\right] &\le \sum_k \mathbb{E}\left[\mathbb{E}_k\left[M_{k+1}-M_k\right]\right] \\ \sum_k \mathbb{E}\left[X_{k+1}-X_k\right] &\le \sum_k \mathbb{E}\left[M_{k+1}-M_k\right] && \text{unconditioning Prop. 10.18} \\ \mathbb{E}[X_n-X_0] &\le \mathbb{E}\left[M_n-M_0\right] && \text{telescopic sum} \end{aligned}$$

Where by the result of △ we get the claim since $b\mathbb{E}\left[U_n(a,b)\right] \le \mathbb{E}\left[X_n-X_0\right] \le \mathbb{E}\left[M_n-M_0\right]$. □

♢ **Observation 12.25** (About the Proposition)**.** *two conclusions are drawn*

- *We derive that the upcrossings are controlled by the positive moments of the process.*
- *The integrability of the submartingale gives a bound on $U_n(a,b)$. We use this to investigate the **pathwise convergence** of $M_n(\omega)$ $\forall\omega\in\Omega, n\in\mathbb{N}$*

Lemma 12.26 (Finiteness of crossings for pathwise convergence idea)**.** *We aim to prove pathwise convergence up to having finite $U_n(a,b)$ as $n\to\infty$ diverges.*
Recall that a series $(m_n)_{n\in\mathbb{N}}$ on the real line $m_n\in\mathbb{R}$ is such that:

$$\nexists \lim_{n\to\infty} m_n \iff \liminf_n m_n < \limsup_n m_n$$

An example is $m_n=\sin\left(\frac{\pi}{4}n\right)$ which satisfies $\liminf_n m_n = -1 \neq \limsup_n m_n = 1$.
A finite number of upcrossings should avoid this incosistency of the path.

♣ **Theorem 12.27** (Martingale Convergence Theorem, MCT)**.** *For a submartingale $(X_n)_{n\in\mathbb{N}}$ (Def. 11.36):*

$$\sup_n \mathbb{E}[X_n^+] < \infty \implies X_n \overset{a.s.}{\to} X_\infty, \quad X_\infty\in\mathcal{L}_1$$

*So that we can establish an **almost sure limiting distribution**.*

Proof. **(△ strategy)** Reason pathwise for $\omega\in\Omega : X_n(\omega)$ is not convergent. We then want to show:

$$\mathbb{P}\left[\{\omega\in\Omega : X_n(\omega) \text{ not conv}\}\right] = 0 \iff X_n \overset{a.s.}{\to} X_\infty$$

(□ upcrossings diverge) For such ω the absence of a convergence indicates that:

$$\liminf_n X_n(\omega) \le \limsup_n X_n(\omega) \iff \exists a<b \ : \ \lim_{n\to\infty}[U_n(a,b)](\omega) = \infty \quad \text{for } \omega$$

Where we could update the condition of △ with :

$$\mathbb{P}\left[\bigcup_{a<b}\{U_n(a,b)=\infty\}\right] = 0 \qquad \bigcup_{a<b}\left\{\omega\in\Omega : \lim_{n\to\infty} U_n(a,b)(\omega)=\infty\right\} := \bigcup_{a<b}\{U_n(a,b)=\infty\}$$

(○ density and De Morgan) While $a < b$ is uncountable, we could equivalently restate the set via a sequence of operations:

$$
\begin{aligned}
\mathbb{P}\left[\bigcup_{a<b}\{U_n(a,b)=\infty\}\right] = \mathbb{P}\left[\bigcup_{a<b,a,b\in\mathbb{Q}}\{U_n(a,b)=\infty\}\right] &= 0 && \text{dense rationals Prop. 18.15} \\
\iff 1 = \mathbb{P}\left[\bigcap_{a<b,a,b\in\mathbb{Q}}\{U_n(a,b)<\infty\}\right] & && \text{De Morgan's Laws} \\
= \mathbb{P}\left[\lim_{n\to\infty} U_n(a,b) < \infty \ \forall a,b\in\mathbb{Q}\right] & \\
\iff \mathbb{E}\left[\lim_{n\to\infty} U_n(a,b)\right] < \infty &
\end{aligned}
$$

Now by noticing that $(U_n(a,b))_{n\in\mathbb{N}}$ is nondecreasing in n we apply:

$$
\begin{aligned}
(b-a)\mathbb{E}\left[\lim_{n\to\infty} U_n(a,b)\right] &= (b-a)\lim_{n\to\infty}\mathbb{E}\left[U_n(a,b)\right] && \text{Monotone convergence (Thm. 4.21)} \\
&\leq \sup_n \mathbb{E}\left[(X_n-a)^+ - (X_0-a)^+)\right] && \text{Upcrossing ineq. Prop. 12.24} \\
&\leq \sup_n \mathbb{E}\left[(X_n-a)^+\right] && (X_0-a)^+ \geq 0 \\
&\leq \mathbb{E}\left[(X_n)^+\right] + |a| && (X_n-a)^+ \leq (X_n)^+ + |a| \\
&< \infty && \text{hypothesis}
\end{aligned}
$$

(♠ integrability) e have integrability by:

$$
\begin{aligned}
\mathbb{E}\left[|X_\infty|\right] &= \mathbb{E}\left[\liminf_n |X_n|\right] && \liminf = \limsup \\
&\leq \liminf_n \mathbb{E}\left[|X_n|\right] && \text{Fatou's Lem. A.49} \\
&\leq \limsup_n \mathbb{E}\left[|X_n|\right] && \liminf \leq \limsup \ \text{ in general} \\
&< \sup_n \mathbb{E}\left[|X_n|\right] && \limsup < \sup \\
&= \sup_n \mathbb{E}\left[2(X_n)^+ - X_n\right] && |X_n| = 2(X_n)^+ - X_n \\
&\leq \sup_n \mathbb{E}\left[2(X_n)^+ - X_0\right] && (X_n)_{n\in\mathbb{N}} \text{ is submartingale and submart. ineq.} \\
&= 2\sup_n\left\{\mathbb{E}\left[(X_n)^+\right]\right\} - \mathbb{E}\left[X_0\right] \\
&< \infty && \text{hypothesis}
\end{aligned}
$$

Which proves $X_\infty \in \mathcal{L}_1$. □

Corollary 12.28 (An equivalent sufficient condition)**.** *We can restate the problem in terms of a more useful condition noting that:*

$$
\begin{cases}\sup_n \mathbb{E}[X_n^+] < \infty \\ X_n^+ \in \mathcal{L}_1 & \forall n\end{cases} \iff \begin{cases}\sup_n \mathbb{E}[|X_n|] < \infty \\ |X_n| \in \mathcal{L}_1 & \forall n\end{cases}
$$

*Namely, an $\mathcal{L}_1$ bound on the martingale, **not** $\mathcal{L}_1$ **convergence!***

Proof. We retake what we showed in the previous and notice that using $|X_n| = 2(X_n)^+ - X_n \leq 2(X_n)^+ - X_0$:

$$
\mathbb{E}\left[X_n^+\right] \leq \mathbb{E}\left[|X_n|\right] \leq 2\mathbb{E}\left[(X_n)^+\right] - \mathbb{E}\left[X_0\right]
$$

So that the two values bound each other whenever $\mathbb{E}\left[X_0\right] \in \mathbb{R}$. □

♢ **Observation 12.29** (Interpreting the Theorem and the Corollary)**.** *Recall that by Jensen's for martingales (Cor. 11.41):*

$$
(X_n)_{n\in\mathbb{N}} \quad \textit{martingale} \implies \left(X_n^+\right)_{n\in\mathbb{N}} \quad \textit{submartingale}
$$

*And that the expectation $\mathbb{E}[X_n^+]$ is **increasing** in n.*

Corollary 12.30 (Special cases). *We recognize a number of familiar situations in which the requirements are easily verified:*

1. $(X_n)_{n\in\mathbb{N}}$ *non positive submartingale*
2. $(X_n)_{n\in\mathbb{N}}$ *non negative supermartingale*
3. $(X_n)_{n\in\mathbb{N}}$ *non positive or non negative martingale*
4. $(X_n)_{n\in\mathbb{N}}$ *bounded above or below by an integrable random variable*

Proof. all cases satisfy the requirement on the supremum. □

♡ **Example 12.31** (Branching process, Example 11.54 continued). *In the previous example, we showed that* $\left(\frac{Z_n}{\mu^n}\right)_{n\in\mathbb{N}}$ *is a martingale with* $\mu = \mathbb{E}\left[\xi_1^{(1)}\right]$. *We also have that:*

$$\mathbb{E}_n\left[\frac{Z_{n+1}}{\mu^{n+1}}\right] = \frac{Z_n}{\mu^n} \implies \mathbb{E}_n[Z_{n+1}] = \mu Z_n = \begin{cases} < Z_n, \mu < 1 & \textit{supermartingale} \\ = Z_n, \mu = 1 & \textit{martingale} \\ > Z_n, \mu > 1 & \textit{submartingale} \end{cases}$$

Here, using the MCT (Thm. 12.27) we want to show this as for Corollary 12.30. For this purpose, let $\mu = 1, p_1 < 1$. *Then* $(Z_n)_{n\in\mathbb{N}}$ *is a positive martingale an by the MCT Corollary:*

$$\exists Z_\infty = \lim_{n\to\infty} Z_n \implies Z_n = Z_\infty \textit{ eventually Def. 9.3}$$

*(△ **aim**)In this context, we want to show that:*

$$Z_\infty = 0 \iff \mathbb{P}[Z_n = k, \forall n \geq N] = 0 \quad \forall k \in \mathbb{N}, N \textit{ sufficiently large}$$

*(□ **solution**) compute the following:*

$$\begin{aligned} \mathbb{P}[Z_n = k, \forall n > N] &= \mathbb{P}[Z_n = k, Z_{n+1} = k, \ldots] && \textit{stable limit} \\ &= \mathbb{P}\left[\sum_{i=1}^{k} \xi_i^{(m+N)} = k \quad m = 1, 2, \ldots, Z_n, Z_n = k\right] && \textit{hypothesis} \\ &\leq \mathbb{P}\left[\sum_{i=1}^{k} \xi_i^{(m+N)} = k \quad m = 1, 2, \ldots\right] && \mathbb{P}[A \cap B] \leq \mathbb{P}[A] \\ &= \prod_{m=1}^{\infty} \mathbb{P}\left[\sum_{i=1}^{k} \xi_i^{(m+N)} = k\right] && \textit{independence} \\ &\overset{m\to\infty}{\to} \lim_{m\to\infty}\left(\mathbb{P}\left[\sum_{i=1}^{k} \xi_1^{(1)} = k\right]\right)^m && \textit{indentically distr.} \end{aligned}$$

By △ we need it to be null, this is the same as:

$$\iff \mathbb{P}\left[\sum_{i=1}^{k} \xi_1^{(1)} = k\right] < 1$$

from which we get:

$$\begin{aligned} \mathbb{P}\left[\sum_{i=1}^{k} \xi_i^{(1)} = k\right] &\leq \mathbb{P}\left[\sum_{i=1}^{k} \xi_1^{(1)} > 0\right] \\ &= 1 - \mathbb{P}\left[\sum_{i=1}^{k} \xi_1^{(1)} = 0\right] \\ &= 1 - p_0^k && \textit{iid} \\ &< 1 && \mu = 1, p_1 < 1 \implies p_0 > 0 \end{aligned}$$

By the same arguments for $\mu < 1$ we could show that $(Z_n)_{n\in\mathbb{N}}$ is a positive supermartingale and that by the Corollary we have a limit which is almost sure. We call it Z_∞. Then, by similar arguments, one can show for $k > 0$ and $N > 0$ arbitrary that:

$$\mathbb{P}[Z_n = k, \forall n \geq N] = 0$$

since $\mu < 1 \implies p_0 > 0$. We then conclude $Z_\infty \stackrel{a.s.}{=} 0$

♢ **Observation 12.32** (Bounds on random variables norms). *Recognize that:*

- *for a martingale, we require integrablity $\mathbb{E}[|X_n|] < \infty$, according to Definition 11.35#2*
- *for almost sure convergence of a martingale, we require $\sup_n \mathbb{E}[|X_n|] < \infty$, according to the MCT (Thm. 12.27, Cor. 12.28)*
- *For a uniformly integrable martingale, we require an almost sure $\mathcal{L}_1$ bound of the p norm (Def. 11.48*

How are these linked together?

♡ **Example 12.33** (Symmetric random walk (Example 12.12 ctd.)). *We show that for a symmetric RW the MCT can be used, but it is not $\mathcal{L}_1$ convergent, namely $\stackrel{a.s.}{\to} \not\Longrightarrow \mathcal{L}_1$ at the same limit.*
Recover the previous setting, where $T = \inf\{m \in \mathbb{N} : S_n = 0\}$. We have that $\mathbb{E}[X_i] = 0 \forall i$ and $S = (S_n)_{n\in\mathbb{N}}$ is a martingale with $\mathbb{E}[S_n] = 1 \forall n$. The stopped martingale $(S_{n\wedge T})_{n\in\mathbb{N}}$ is such that $S_{n\wedge T} \geq 0$ and by the MCT (Thm. 12.27, Cor. 12.30) there exists an almost sure limiting process S_∞. Now obviously $S_{n\wedge T} \stackrel{a.s.}{\to} S_\infty = 0$ since convergence to $k > 0$ is impossible as it would mean that $S_n = k > 0 \implies S_{n+1} \in \{k-1, k+1\}$, i.e. no convergence. However, there is no $\mathcal{L}_1$ convergence. Indeed by Proposition 12.13:

$$\mathbb{E}[S_{n\wedge T}] = \mathbb{E}[S_{0\wedge T}] = \mathbb{E}[S_0] = 1$$

but:

$$\mathbb{E}[|S_{n\wedge T} - S_\infty|] = \mathbb{E}[|S_{n\wedge T} - 0|] = \mathbb{E}[|S_{n\wedge T}|] = \mathbb{E}[S_{n\wedge T}] = 1 \neq 0 \quad \forall n \in \mathbb{N}$$

Lemma 12.34 (A quick Lemma for $\mathcal{L}_1$ convergence). *If $(X_n)_{n\in\mathbb{N}} \stackrel{\mathcal{L}_1}{\to} X$ then:*

$$\lim_{n\to\infty} \mathbb{E}[X_n Y] = \mathbb{E}[XY] \quad \forall Y \text{ bounded a.s.}$$

Proof. By hypothesis $|Y| \leq b$ almost surely, so that:

$$|\mathbb{E}[X_n Y] - \mathbb{E}[XY]| \leq \mathbb{E}[|X_n Y - XY|] \leq b\mathbb{E}[X_n - X] \stackrel{n\to\infty}{\to} 0$$

□

♣ **Theorem 12.35** (Uniform Integrability vs a.s. $\mathcal{L}_1$ characterization). *For a martingale $(M_n)_{n\in\mathbb{N}}$ it holds:*

1. *Same convergence by uniformity*

$$\begin{cases} M_n \stackrel{a.s.}{\to} M_\infty \\ M_n \stackrel{\mathcal{L}_1}{\to} M_\infty \end{cases} \iff (M_n)_{n\in\mathbb{N}} \quad uniformly\ integrable$$

2. *Martingale equality extends at infinity as a martingale $\overline{X}$ if $M_n = \mathbb{E}_n[Z]$ for $Z \in \mathcal{L}_1$:*

$$M_n = \mathbb{E}_n[Z], Z \in \mathcal{L}_1 \implies M_\infty = \lim_{n\to\infty} M_n \ : \ \overline{X} = (X_n)_{n\in\overline{\mathbb{N}}}$$

Proof. **(Claim #1)($\Longleftarrow$)** Let $M = (M_n)_{n\in\mathbb{N}}$ be a u.i. martingale.
(△ MCT application) Assign:

$$k(b) := \sup_n \mathbb{E}\left[|M_n| \mathbb{1}_{(b,\infty)}(|M_n|)\right] \to 0 \quad as\ b \to \infty$$

Notice that:

$$\mathbb{E}[|M_n|] = \mathbb{E}\left[|M_n| \mathbb{1}_{(-\infty,b]}(|M_n|)\right] + \mathbb{E}\left[|M_n| \mathbb{1}_{(b,\infty)}(|M_n|)\right] \leq b + k(b)$$

So that for b large enough it holds $k(b) \leq 1$ and $(M_n)_{n\in\mathbb{N}}$ is $\mathcal{L}_1$ bounded. By the MCT (Thm. 12.27) $\exists M_\infty$ integrable with $M_n \overset{a.s.}{\to} M_\infty$.
($\square$ norm convergence) we now wts that $M_n \overset{\mathcal{L}_1}{\to} M_\infty$ this is equivalent to:

$$\forall\epsilon > 0 \exists N \in \mathbb{N} \quad : \quad \forall n \geq N \, \mathbb{E}\left[|M_n - M_\infty\right] < 2\epsilon$$

Fix now $\epsilon > 0$ and let $H_n := \{|M_n - M_\infty| > \epsilon\}$. It holds:

$$\mathbb{E}\left[|M_n - M_\infty|\right] \leq \epsilon + \underbrace{\mathbb{E}\left[|M_n - M_\infty|\mathbb{1}_{H_n}\right]}_{>\epsilon}$$

($\bigcirc$ final computation) For X_∞ integrable and $X = (X_n)_{n\in\mathbb{N}}$ u.i. it holds that $(X_n - X_\infty)$ is u.i.. Let $k^*(b)$ be as $k(b)$. Notice that:

$$|M_n - M_\infty|\mathbb{1}_H \leq b\mathbb{1}_H + |M_n - M_\infty|\mathbb{1}_{\{|M_n - M_\infty|>b\}} \quad \forall H, \forall b$$

Then:

$$\mathbb{E}\left[|M_n - M_\infty|\mathbb{1}_{H_n}\right] \leq b\mathbb{P}[H_n] + k^*(b) \quad k^*(b) \overset{b\to\infty}{\to} 0 \implies \forall\epsilon\exists b < \infty : k^*(b) \leq \frac{\epsilon}{2}$$

So if we set $\delta = \frac{\epsilon}{2b}, \mathbb{P}[H_n] \leq \delta$ then:

$$\mathbb{E}\left[|M_n - M_\infty|\mathbb{1}_{H_n}\right] \leq \epsilon \tag{12.2}$$

The result of $\square$ and Equation 12.2 imply that:

$$\forall\epsilon > 0 \exists\delta > 0 : \mathbb{P}[H_n] < \delta \implies \mathbb{E}\left[|M_n - M_\infty|\right] \leq 2\epsilon$$

And since $\mathbb{P}[H_n] = \mathbb{P}[|M_n - M_\infty| > \epsilon] \to 0 \quad \forall\epsilon$ (i.e. almost sure implies conv. in probability, Prop. 9.14) for n sufficiently large we obtain $M_n \overset{\mathcal{L}_1}{\to} M_\infty$.
($\implies$) [Çin11](III.4.6).
(Claim #2) It holds that the martingale is uniformly integrable by Proposition 11.50, moreover, the condition we want to prove can be reformulated :

$$\mathbb{E}_n[M_\infty] = M_n \iff \mathbb{E}_n[M_\infty - M_n] = 0 \forall n$$

Fix $m \in \mathbb{N}, H \in \mathcal{F}_m$. Then $\forall n \geq m$ the martingale equality implies:

$$\begin{aligned} \mathbb{E}\left[\mathbb{1}_H(M_n - M_m)\right] &= \mathbb{E}\left[\mathbb{1}_H \mathbb{E}_m[M_n - M_m]\right] && \text{unconditioning and } H \in \mathcal{F}_n \\ &= 0 && \text{martingale equality} \end{aligned}$$

Since by #1 $M_n - M_m \overset{\mathcal{L}_1}{\to} M_\infty - M_n$ as $n \to \infty$ we also have that:

$$\mathbb{E}\left[\mathbb{1}_H(M_\infty - M_m)\right] = \lim_{n\to\infty} \mathbb{E}\left[\mathbb{1}_H(X_n - X_m)\right] = 0$$

So that $X_n \overset{\mathcal{L}_1}{\to} X$ is such that $\lim_n \mathbb{E}\left[X_n Y\right] = \mathbb{E}\left[XY\right]$ for all Y bounded by Lemma 12.34. By the arbitrariness of $H \in \mathcal{F}_n$, the claim is proved. $\square$

♡ **Example 12.36** (Branching process, (Ex. 12.31 ctd.))**.** *We know $\mu = 1, p_1 < 1$ are such that $(Z_n)_{n\in\mathbb{N}} \overset{a.s.}{\to} Z_\infty = 0$ but not such that $Z_n \overset{\mathcal{L}_1}{\to} Z_\infty$. Yet we also argued that $Z_n \neq \mathbb{E}_n[Z_\infty] = 0$ since $Z_n > 0 \forall n$ with positive probability. So, we cannot conclude that $\mathbb{E}_n[M_\infty] = M_n$.*

♢ **Observation 12.37** (Building $\mathcal{L}_1$ convergent martingales)**.** *Build the following chain:*

$$Z_n \in \mathcal{L}_1 \quad \forall n \underset{Prop.\ 11.50}{\implies} (M_n)_{n\in\mathbb{N}} : M_n = \mathbb{E}[Z_n] \quad uniformly\ integrable\ martingale \underset{Thm.\ 12.35}{\iff} M_n \underset{\mathcal{L}_1}{\overset{a.s.}{\to}} M_\infty$$

What is the limiting filtration? We can use the extended filtration notion from Definition 11.17:

$$\mathcal{F}_\infty = \bigvee_n \mathcal{F}_n$$

Corollary 12.38 (Appying Theorem 12.35 to characterize Observation 12.37)**.** *Conclude that:*

1. $\forall Z : \mathbb{E}[|Z|] < \infty$ *it holds* $\mathbb{E}_n[Z] \underset{\mathcal{L}_1}{\overset{a.s.}{\to}} \mathbb{E}_\infty[Z] = \mathbb{E}[Z|\mathcal{F}_\infty]$
2. $Z \in \mathcal{F}_\infty \implies \mathbb{E}_n[Z] \underset{\mathcal{L}_1}{\overset{a.s.}{\to}} Z$

So that Z is eventually revealed.

Proof. **(Claim #1)** by Proposition 11.50 we have a u.i. martingale $M_n = \mathbb{E}_n[Z] = \mathbb{E}[M_\infty]$ and by Theorem 12.35#2 for all n almost surely and in $\mathcal{L}_1$. Then:

$$\begin{aligned}
\lim_{n\to\infty} M_n &= \lim_{n\to\infty} \mathbb{E}_n[Z] \\
&= \lim_{n\to\infty} \mathbb{E}_n[M_\infty] && \in \mathcal{F}_\infty \text{ by construction} \\
&= \mathbb{E}_\infty[M_\infty] \\
&= M_\infty \\
&= \mathbb{E}_\infty[Z] && \text{a.s. } \mathcal{L}_1
\end{aligned}$$

(Claim #2) if $Z \in \mathcal{F}_\infty$ then we can remove the expectation:

$$M_n = \mathbb{E}_n[Z] \underset{\mathcal{L}_1}{\overset{a.s.}{\to}} \mathbb{E}_\infty[Z] = Z$$

□

♡ **Example 12.39** (Bayesian mean estimation, Corollary 12.38 (Ex. 11.52 ctd.))**.** *Recall that Z_i are iid standard normals and $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$ is independent from Z_i, integrable and finite.*
For $Y_i = \theta + Z_i \implies Y_i|\theta \overset{iid}{\sim} \mathcal{N}(\theta, 1)$ we have that for observables $\boldsymbol{Y} = \{Y_i\}_{i=1}^n$:

$$\pi_n(A) = \mathbb{P}[\theta \in A|\,\boldsymbol{Y} = \boldsymbol{y}] \quad \mathcal{F} = \sigma(\{\,\boldsymbol{Y}\})$$

Then $\widehat{\theta}_n = \mathbb{E}_n[\theta] = \mathbb{E}[\theta|\mathcal{F}_n]$ is a uniformly integrable martingale by Proposition 11.50.
Further, by Corollary 12.38#1 we conclude:

$$\widehat{\theta}_n \underset{\mathcal{L}_1}{\overset{a.s.}{\to}} \mathbb{E}_\infty[\theta] = \mathbb{E}[\theta|\mathcal{F}_\infty]$$

Moreover, if the condition $\theta \in \mathcal{F}_\infty$ holds, we apply Corollary 12.38#2 and further state that:

$$\theta \in \mathcal{F}_\infty \implies \widehat{\theta}_n \underset{\mathcal{L}_1}{\overset{a.s.}{\to}} \theta$$

We prove a sufficient condition for this to be true in Proposition 12.40.

♣ **Proposition 12.40** (Frequentist validation of Bayesian mean estimator)**.** *This Proposition is also useful for context in Theorem 8.7.*
Recall the setting of Example 12.39

$$identifiability \quad : \quad \mathcal{P}_\theta(A) = \mathbb{P}[\,\boldsymbol{Y} \in A|\theta] : \mathcal{P}_\theta(\cdot) \neq \mathcal{P}_{\theta'}(\cdot) \forall \theta \neq \theta' \implies \theta \in \mathcal{F}_\infty$$

In other words, identifiability is a sufficient condition for the true value to be revealed at the end of time.

Proof. Consider $\theta \in \mathcal{F}_\infty$. This holds if and only if:

$$\exists h \; measurable: \; h : \mathbb{R}^\infty \to \Theta, \quad h(Y_1, \ldots) = \theta$$

If $Y_n \overset{iid}{\sim} p_\theta$ then by the SLLN we could obtain $\mathcal{P}_\theta$ and subsequently θ by the identifiability principle. What is missing is measurability. Using instead a consistent estimator $\overline{Y}_n$ we would get:

$$\mathcal{P}_\theta^\infty \left(|\overline{Y}_n - \theta| > \epsilon\right) \overset{n\to\infty}{\to} 0 \quad \forall \epsilon > 0$$

Where $\mathcal{P}_\theta^\infty = \bigtimes_n \mathcal{P}_\theta^n$ the infinite product of the probability laws. Then:

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}[|\overline{Y}_n - \theta| > \epsilon] &= \lim_{n\to\infty} \int \underbrace{\mathcal{P}_\theta^\infty(|\overline{Y}_n - \theta| > \epsilon)}_{\leq 1} \pi(d\theta) \\ &= \int \underbrace{\lim_{n\to\infty} \mathcal{P}_\theta^\infty(|\overline{Y_n} - \theta| > \epsilon)}_{\to 0 \text{ as } n\to\infty} \pi(d\theta) \qquad \text{dominated conv. Thm. 4.24} \\ &= 0\end{aligned}$$

Which implies that $\overline{Y}_n \xrightarrow{\mathcal{P}} \theta$. Then, by [Çin11](Thm. III.3.3-(b)) there exists a subsequence $(n(k))_{k\in\mathbb{N}}$ going to ∞ such that $\overline{Y}_{n(k)} \xrightarrow{a.s.} \theta$ in $\mathbb{P}$. By the fact that $\overline{Y}_{n(k)} \in \mathcal{F}_{n(k)} \subset \mathcal{F}_\infty$ we have that θ is expressible as the limit of functions which are measurable wrt $\mathcal{F}_\infty$. eventually, $\theta \in \mathcal{F}_\infty$ and $\exists h : \mathbb{R}^\infty \to \Theta$ such that $h(Y_1, Y_2, \ldots) = \theta$. □

Corollary 12.41 (Frequentist perspective validation)**.** *$\forall \theta_0$ in almost sure sets of $\pi(\cdot)$:*

$$Y_i \overset{iid}{\sim} \mathcal{P}_{\theta_0} = \mathbb{P}[Y_i \in \cdot|\theta_0] \implies \widehat{\theta}_n \to \theta_0 \quad \textit{in } \mathcal{P}_{\theta_0}^\infty \quad \textit{a.s.}$$

♣ **Theorem 12.42** (Levy's 0-1 law)**.**

$$A \in \mathcal{F}_\infty \implies \mathbb{E}_n[\mathbb{1}_A] \xrightarrow{a.s.} \mathbb{1}_A$$

Proof. For any A, $\mathbb{1}_A$ is bounded and such that $\mathbb{1}_A \sim \mathcal{B}ern(\mathbb{P}[A])$. Then $\mathbb{1}_A$ is integrable, and its integral is, by Theorem 12.35 and Corollary 12.38#2:

$$\mathbb{E}_n[\mathbb{1}_A] \underset{\mathcal{L}_1}{\overset{a.s.}{\to}} \mathbb{E}_\infty[\mathbb{1}_A] = \mathbb{1}_A$$

□

♢ **Observation 12.43** (Uniform integrability and martingales summary)**.** *We recognize the main results of this Chapter and the previous ones.*

- *$Z \in \mathcal{L}_1$ integrable $\implies X_t = \mathbb{E}_t[Z]$ is uniformly integrable (Prop. 11.50)*
- *by the MCT a uniformly integrable martingale has a limit X_∞ (Thm. 12.27)*
- *contrarily, for a characterization we need also $\mathcal{L}_1$ convergence (Thm. 12.35#1), and can extend the martingale equality at ∞ if it is the expectation of an integrable random variable (Thm. 12.35#2)*
- *If the value of Z is revealed at the end of time $Z \in \mathcal{F}_\infty$ then the martingale converges to the actual value (Cor. 12.38#2)*

♡ **Example 12.44** (Branching Process, Ex. 12.36 ctd.)**.** *Before we had $\mu < 1$ or $\mu = 1$ and $p_1 < 1$ so that* $Z_n \xrightarrow{a.s.} 0$.
*(△ **aim**) The assumptions now become $p_0 \in (0,1) \implies \mu < 1$ and we want to show that:*

$$Z_n \xrightarrow{a.s.} 0 \quad \textit{or} \quad Z_n \xrightarrow{a.s.} \infty$$

Namely, if there is no extinction, then the population explodes. Another possible formulation is:

$$\mathbb{P}\left[\lim_{n\to\infty} Z_n \in \{0, \infty\}\right] = 1$$

*(□ **characterizing extinction**) we have:*

$$\left\{\lim_{n\to\infty} Z_n = 0\right\} = \{Z_n = 0 \; \forall n \textit{ sufficiently large}\}$$

while $D := \{Z_n = 0 \textit{ for some } n\} = \bigcup_n \{Z_n = 0\}$. Then exctinction is characterized by:

$$D = \bigcup_n \{Z_n = 0\} \in \mathcal{F}_\infty$$

By the Lévy 0-1 law (Thm. 12.42) this would mean that:

$$\mathbb{E}_n[\mathbb{1}_D] = \mathbb{P}[D|\mathcal{F}_n] \xrightarrow{a.s.} \mathbb{1}_D$$

*(◯ **characterizing explosion**) we aim to see if $D^c = \{Z_n > x \text{ i.o.}\}$ for any $x \in \mathbb{R}$ implies that $Z_n \to \infty$ necessarily.*
*(♢ **solution**) it holds:*

$$\begin{aligned}
\mathcal{P}[D|\mathcal{F}_n] &= \mathcal{P}_n[D] \\
&= \mathcal{P}_n(D \cap \{Z_n \leq x\}) + \mathcal{P}_n(D \cap \{Z_n > x\}) \\
&\geq \mathcal{P}_n(D \cap \{Z_n \leq x\}) \\
&\geq \mathcal{P}_n(Z_{n+1} = 0, Z_n \leq x) && D \supset \{Z_{n+1} = 0\} \\
&= \sum_{k=0}^{x} \mathcal{P}_n(Z_{n+1} = 0, Z_n = k) \\
&= \sum_{k=0}^{x} p_0 \mathbb{1}_{\{Z_n = k\}} && \text{iid generations} \\
&= p_0^x \sum_{k=0}^{x} \mathbb{1}_{\{Z_n = k\}} \\
&= p_0^x \mathbb{1}_{\{Z_n \leq x\}} && p_0 \in (0,1)
\end{aligned}$$

In this setting, we recognize that to have no extinction:

$$\begin{aligned}
\mathcal{P}[D|\mathcal{F}_n] = p_0^x \mathbb{1}_{\{Z_n \leq x\}} \overset{a.s.}{\to} 0 &\overset{p_0 \in (0,1)}{\iff} \mathbb{1}_{\{Z_n \leq x\}} \overset{a.s.}{\to} 0 \\
&\iff \{Z_n \geq x,\ \forall x \in \mathbb{R},\ i.o\} \overset{a.s.}{\to} 1 \\
&\iff D^c \overset{a.s.}{\to} 1
\end{aligned}$$

On the contrary, on D^c, the population explodes, namely $Z_n \overset{a.s.}{\to} \infty$.

12.4 Exercise Session

The following examples are instructive and allow to apply the concepts shown in Chapters 11 and 12.

♡ **Example 12.45** (Occupancy problem)**.** *Consider n independent bins and m balls. We are interested in the number of empty bins, denoted as Z.*

(△ setting) *We set:*

$$C_i := \text{bin chosen at } i^{th} \text{ ball} \quad : \quad \mathbb{P}[C_i = j] = \frac{1}{n} \quad j = 1, \dots, n$$

Which are iid random variables.

(□ Azuma inequality by martingales) *let $(\mathcal{F}_n)_{n\in\mathbb{N}} = \sigma(\{C_i\}_{i=1}^m)$ and $Z_t := \mathbb{E}_t[Z]$. In this setting, Z_t is the estimate of the number of empty bins at the end having observed t throws.*
Using Proposition 11.50 we have:

$$Z \in [0, n] \text{ bounded} \implies (Z_n)_{n\in\mathbb{N}} \quad uniformly\ integrable$$

Then set $Z_0 = \mathbb{E}_0[Z] = \mu =$ by the martingale equality (Def. 11.35#3).
Notice that $Z \in \mathcal{F}_m \implies Z_t = Z \forall t \geq m$, meaning that after having thrown all the balls (m throws), Z belongs to the σ-algebra. This is rather intuitive.

(◯ Azuma inequality applied) *We have that:*

$$\square \implies \mathbb{P}[|Z - \mu| > \delta\mu] = \mathbb{P}[|Z_m - Z_0| > \delta\mu]$$

And using Azuma Inequality (Thm. 12.49), by $|Z_{t+1} - Z_t| \leq 1$ we set $c = 1$ and get that:

$$\implies \mathbb{P}\left[|Z_t - Z_0| > \lambda\sqrt{t}\right] \leq 2e^{-\frac{\lambda^2}{2}} \quad : \quad c = 1$$

With

$$\lambda\sqrt{t} = \delta\mu \implies \lambda = \frac{\delta\mu}{c\sqrt{m}}$$

$$\implies \mathbb{P}[|Z - \mu| > \delta\mu] \leq 2exp\left\{-\frac{-\delta^2\mu^2}{2m}\right\} \quad c = 1 \tag{12.3}$$

($\diamondsuit$ finding μ) *letting $X_j :=$ # balls in $j \in \{1, \ldots, n\}$ we get:*

$$Z = \sum_{j=1}^{n} \mathbb{1}\{X_j = 0\} \quad : \quad X_1, \ldots, X_n \sim \mathcal{M}ultinom\left(m, \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)\right)$$

$$X_j \sim \mathcal{B}inom\left(m, \frac{1}{n}\right)$$

So that:

$$\begin{aligned}
\mathbb{E}[Z] = Z_0 = \mu &= \sum \mathbb{E}\left[\mathbb{1}\{X_j = 0\}\right] && \textit{linearity of integral} \\
&= \sum \mathbb{P}(\{X_j = 0\}) \\
&= n\mathbb{P}[X_1 = 0] && \textit{iid assumption} \\
&= n\left(1 - \frac{1}{n}\right)^m \\
&= np && p := \left(1 - \frac{1}{n}\right)^m
\end{aligned}$$

And Equation 12.3 in $\square$ becomes:

$$\begin{aligned}
\mathbb{P}[|Z - \mu| > \delta\mu] &\leq 2exp\left\{-\frac{1}{2}\frac{\delta^2 n^2 p^2}{m}\right\} \\
&= exp\left\{-\frac{1}{2}\delta^2 np^2\right\} && \textit{if } n = m
\end{aligned}$$

Notice also that as $m = n \to \infty$ we also have that $p \to e^{-1}$

($\triangledown$ informal Chernoff's bound) *ignoring the dependency let $\delta \in (0,1), c > 0, \mu = np$ and derive a much more restrictive bound on the probability by Theorem 7.25:*

$$\mathbb{P}[|Z - \mu| > \delta\mu] \leq 2e^{-\frac{1}{2}cnp\delta^2}$$

♡ **Example 12.46** (Averages)**.** *We show that the average process needs a specific condition to be a martingale, as discussed earlier.*
Assume a discrete process $(X_n)_{n\in\mathbb{N}}$ is adapted to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ and is integrable. Then let:

$$\overline{X_n} = \frac{1}{n}\sum^{n} X_i \quad \textit{assume} \quad \mathbb{E}_n[X_{n+1}] = \overline{X_n}$$

($\triangle$ aim) *we want to show that $\left(\overline{X}_n\right)_{n\in\mathbb{N}}$ an $\mathcal{F}$-martingale according to Definition 11.35.*

($\square$ adaptedness) *as $\overline{X}_n = \frac{1}{n}\sum^n X_i : X_i \in \mathcal{F}_n \; \forall i \leq n$ adapteness is trivial.*

(◯ integrability) *Notice that* $\mathbb{E}[|X_n|] \leq \frac{1}{n}\sum \mathbb{E}[|X_i|] < \infty$, *by trivial application of linearity,Jensen's inequality and the hypothesis of integrability.*

(♢ martingale equality) *we proceed by manipulation:*

$$\begin{aligned}
\mathbb{E}_n\left[\overline{X_{n+1}} - \overline{X_n}\right] &= \frac{1}{n(n+1)}\mathbb{E}_n\left[n\left(\sum^{n+1} X_i\right) - (n+1)\left(\sum^{n} X_i\right)\right] \\
&= \frac{1}{n(n+1)}\mathbb{E}_n\left[n\left(\sum^{n+1} X_i\right) - n\left(\sum^{n} X_i\right) - \left(\sum^{n} X_i\right)\right] \\
&= \frac{1}{n(n+1)}\mathbb{E}_n\left[nX_{n+1} - \left(\sum^{n} X_i\right)\right] \\
&= \frac{1}{n(n+1)}\mathbb{E}_n\left[n\frac{\sum X_i}{n} - \left(\sum^{n} X_i\right)\right] && \text{by hypothesis} \\
&= 0
\end{aligned}$$

And the equality holds. By □, ◯, ♢ $\implies$ △ *claim is verified.*
Notice however that $\overline{X}_n \neq 0$ *a.s. since* $X_{n+1} \perp\!\!\!\perp \mathcal{F}_n$ *so that:*

$$\mathbb{E}[X_{n+1}] = \overline{X_n} = 0 \iff X_i = 0\ \forall i$$

♡ **Example 12.47** (Poisson Process)**.** *Let* N *be a counting process (Def. 11.13) such that* $N_t = \sum_{k=0}^{\infty} \mathbb{1}_{[0,t]}(T_k)$ *is adapted to* $\mathcal{F}$ *(Def. 11.7).*

(△ aim) *we want to show that:*

$$N \sim \mathcal{P}ois(c) \quad \text{Def. 12.2} \implies M_t = exp\{-rN_t + ct - cte^{-r}\} \quad \mathcal{F}\text{-martingale} \quad \forall r \in \mathbb{R}_+ \quad \text{Def. 11.35}$$

The relation is actually $\iff$ *but we only show one side.*

(□ Laplace approach) *recall that by Definition 6.11 for a Poisson random variable we have that:*

$$X \sim Po(\lambda) \iff \widehat{\mathcal{P}}_X(r) = \mathbb{E}[e^{-rX}] = exp\{-\lambda(1-e^{-r})\}$$

From this simple fact we could show that for any time point the martingale equality holds by noticing that from Definition 12.2 we have $N_t \sim Po(ct)$ *and* $N_t - N_s|\mathcal{F}_s \sim Po(c(t-s))$ *by Proposition 12.3#3. This allows us to say that the Laplace transform of* $N_t - N_s$ *is:*

$$\mathbb{E}_s\left[exp\left\{-r(N_t - N_s)\right\}\right] = exp\left\{-c(t-s)(1-e^{-r})\right\} \tag{12.4}$$

And we could instead check that:

$$\begin{aligned}
\mathbb{E}_s\left[\frac{M_t}{M_s}\right] &= \mathbb{E}_s\left[exp\{-r(N_t - N_s + c(t-s)(1-e^{-r})\}\right] \\
&= \mathbb{E}_s\left[exp\{-r(N_t - N_s\}\right] exp\{c(t-s)(1-e^{-r})\} \\
&= exp\{-c(t-s)(1-e^{-r})\}exp\{c(t-s)(1-e^{-r})\} && \text{Eqn. 12.4} \\
&= 1
\end{aligned}$$

(▽ adaptedness and integrability) *we have:*

$$M_t = exp\{-r\underbrace{N_t}_{\in\mathcal{F}_t} + c\underbrace{t}_{\in\mathcal{F}_t} - \underbrace{cte^{-r}}_{\in\mathcal{F}_t}\} \in \mathcal{F}_t \quad \forall t$$

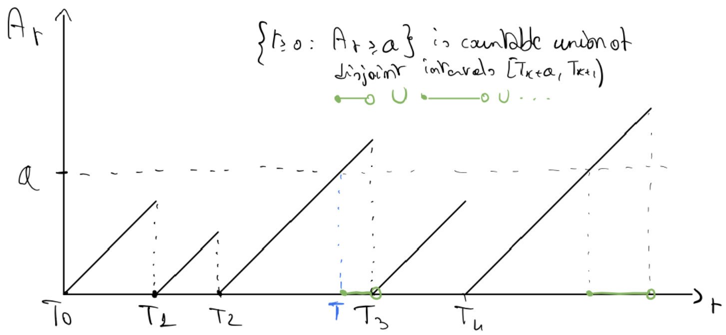

Figure 12.2: A visualization of $(A_t)_{t\in\mathbb{T}}$

Which proves adaptedness. Concerning integrability:

$$\begin{aligned}
\mathbb{E}[|M_t|] = \mathbb{E}[M_t] &= e^{ct} exp\left\{-cte^{-r}\right\}\mathbb{E}[e^{-rN_t}] \\
&= e^{ct} exp\left\{-cte^{-r}\right\}\widehat{\mathcal{P}}_X(r) \\
&= e^{ct} exp\left\{-cte^{-r}\right\}\exp\{-ct(1-e^{-r})\} && \text{by } X_t \sim Po(ct) \\
&= 1 < \infty
\end{aligned}$$

♡ **Example 12.48** (Counting process: age perspective)**.** *We propose a different view on the counting process (Def. 11.13).*

(△ setting) *let* $0 < T_1 < T_2 < \dots$ *be such that* $\lim_{n\to\infty} T_n = +\infty$ *and:*

$$N_t = \sum_{n=1}^{\infty} \mathbb{1}_{[0,t]}(T_n) \qquad \mathcal{F} = \sigma((N_t)_{t\in\mathbb{T}})$$

See N_t *as the number of replacements of some object. Then, the duration of the* k^{th} *object can be formalized as:*

$$A_t(\omega) := t - T_k(\omega) \quad if \;\; T_k(\omega) \le t \le T_{k+1}(\omega)$$

Where the map $t \to A_t$ *is:*

- *strictly increasing in each interval*
- *right continuous at each jump*

See Figure 12.2 for an intuition. We can further define for $a > 0$*:*

$$T := \inf\{t \ge 0 : A_t \ge a\}$$

As the first time the age of a replacement is at least a*.*

(□ A is adapted) *if* $t < a \implies A_t = \emptyset \in \mathcal{F}_t \forall t$ *and the statement is trivial.*
Else consider:

$$\begin{aligned}
\{A_t \ge a\} \in \mathcal{F}_t &\iff A_t \in \mathcal{F}_t \\
&\iff \{t - T_k \ge a\} && = \{T_k < t < T_{k+1}\} \cap \{A_t \ge a\} \\
& && = \underbrace{\{t < T_{k+1}\}}_{\mathcal{F}_t} \cap \underbrace{\{t - T_k \le a\}}_{\mathcal{F}_{t-a}\subset\mathcal{F}_t}
\end{aligned}$$

So that by closedness under countable intersections (Lem. 1.7):

$$\{A_t \ge a\} = \bigcup_k (\{T_k < t < T_{k+1}\} \cap \{A_t \ge a\}) \in \mathcal{F}_t \quad \forall t$$

(◯ equivalence to counting) *we aim to show that:*

$$\inf\{t \geq 0 : N_t = N_{t-a}\} = \inf\{t \geq 0 : A_t \geq a\}$$

The time above a is the union of disjoint $[\cdot, \cdot)$ intervals such that $T_{k+1} - T_k \geq a$ by construction, implying that:

$$\{t \geq 0 : A_t \geq a\} = \bigcup_{k : T_{k+1} - T_k \geq a} [T_k + a, T_{k+1})$$

which infimized:

$$\begin{aligned} &\implies \inf\{t \geq 0 : A_t \geq a\} = \min_k \{T_k + a : T_{k+1} - T_k \geq a\} \\ &\quad \iff \{T = T_k + a\} = \{T_1 - T_0 < a, \ldots, T_k - T_{k-1} < a\} \cap \{T_{k+1} > T_k + a\} \end{aligned}$$

(♢ T is a stopping time) *we eventually show that T is again a stopping time. Differently from Example 11.27 with respect to $\mathcal{G} = \sigma((A_t)_{t \in \mathbb{T}})$.*

$$\begin{aligned} \{T \leq t\} &= \bigcup_{s < t} \{A_s \geq a\} \\ &= \bigcup_{s \in \mathbb{Q}, s < t} \{A_s \geq a\} && \text{By the continuity in } \triangle \text{ unless } A_t = 0 \\ &= \bigcup_{s \in \mathbb{Q}, s < t} \{\underbrace{A_s}_{\in \mathcal{F}_s} \geq a\} && \text{where } \mathcal{G}_s \subset \mathcal{G}_t \forall s < t \\ &\in \mathcal{G}_t && \text{by countable unions (Thm. 1.5)} \end{aligned}$$

The discussions of $\triangle, \diamondsuit \implies \{T \leq t\} \in \mathcal{G}_t$ and T is a stopping time in the sense of Definition 11.9.

♣ **Theorem 12.49** (Azuma Inequality)**.** *For a process $X = (X_t)_{t \in \mathbb{T}}$:*

$$|X_t - X_{t-1}| \leq c \quad \forall t \implies \forall \lambda > 0 \quad \mathbb{P}[|X_t - X_0| > \lambda c \sqrt{t}] \leq 2e^{-\frac{\lambda^2}{2}}$$

Chapter Summary

Objects:

- Poisson process with rate $c > 0$, a counting process adapted to a filtration such that:

$$\mathbb{E}_s\left[f(N_{s+t} - N_s)\right] = \sum_{k=0}^{\infty} \frac{e^{-ct}(ct)^k}{k!} f(k) \quad \forall s, t,\ f \in \mathcal{E}^+$$

- predictable processes $(F_n)_{n\in\mathbb{N}}$ to construct the Steltjes-Lebesgue integral series:

$$X = \int F dM, \quad X_n = \int_{[0,n]} F dM = M_0 F_0 + \sum_{m=1}^{n} (M_m - M_{m-1}) F_m$$

- upcrossings $S_k := \inf\{n \geq T_{k-1} : M_n \leq a\}$, downcrossings $T_k := \inf\{n \geq S_k : M_n \geq b\}$ and interval (a, b) counter $U_n(a, b) = \sum_{k=1}^{n} \mathbb{1}_{[0,t]}(T_k)$
- F_n formalism to represent the number of stocks at $(n, n+1]$ with portfolio value $X_n = \int_{[0,n]} F dM$ and $X_0 = 0$ and profit in general: $X_n - X_0 \geq (b-a) U_n(a, b)$

Results:

- Poisson processes are stationary, Markovian and have Poisson distributed intervals with rate ct
- $N \sim \mathcal{P}ois(c) \iff (N_t - ct)_{t\in\mathbb{T}}$ is an $\mathcal{F}$-martingale
- the Steltjes Lebesgue integral of a (sub)martingale and a predictable process is a (sub)martingale
- the Steltjes Lebesgue integral of a (sub)martingale and a stopping time indicator is a (sub)martingale
- Doob's:
  - we proved that a martingale and a **bounded** stopping time result in $\mathbb{E}\left[M_0\right] = \mathbb{E}\left[M_T\right] = \mathbb{E}\left[M_k\right]$ so that the martingale equality extends to stopping times
  - for $S \leq T$ a.s. stopping times, it also holds that $\mathbb{E}\left[X_S - X_T\right] = 0$
  - any adapted integrable X process has a decomposition into a deterministic starting point X_0, predictable process A (prediction) and a martingale M (innovation). This looks like the orthogonal projection on the filtration
  - the general result says that for an adapted process $(M_n)_{n\in\mathbb{N}}$ the equivalent conditions are:
    * $(M_n)_{n\in\mathbb{N}}$ is a martingale
    * for bounded stopping times $S \leq T$ M_S and M_T are integrable and $\mathbb{E}_S[M_T - M_S] = 0$
    * for bounded stopping times $S \leq T$ M_S and M_T are integrable and $\mathbb{E}[M_T - M_S] = 0$
  - the upcrossing inequality:

$$(M_n)_{n\in\mathbb{N}} \quad submartingale \implies (b-a)\mathbb{E}[U_n(a, b)] \leq \mathbb{E}[(M_n - a)^+ - (M_0 - a)^+]$$

  - the Martingale Convergence Theorem for submartingales:

$$\sup_n \mathbb{E}[X_n^+] < \infty \implies X_n \overset{a.s.}{\to} X_\infty, \quad X_\infty \in \mathcal{L}_1$$

    where the condition is equivalent to an $\mathcal{L}_1$ bound on the norm
  - Uniform integrability of a martingale is characterized by both convergence almost sure and in $\mathcal{L}_1$ norm. If the process is the expectation of an integrable random variable, the martingale equality also extends at ∞
  - identifiability ensures in the Bayesian mean estimation that $\theta \in \mathcal{F}_\infty$
  - Levy's 0-1 law is $A \in \mathcal{F}_\infty \implies \mathbb{E}_n[\mathbb{1}_A] \overset{a.s.}{\to} \mathbb{1}_A$

Chapter 13

Poisson Random Measures

13.1 Random Measures

♠ **Definition 13.1** (Random Measure $M(\cdot,\cdot)$, r.m.). *The concept is equivalent to that of a Transition Kernel (Def. B.13) from $(\Omega, \mathcal{H})$ onto $(E, \mathcal{E})$. Consider a probability space $(\Omega, \mathcal{H}, \mathbb{P})$ and a measurable space $(E, \mathcal{E})$. A random measure on $(E, \mathcal{E})$ is a mapping:*

$$M : \Omega \times \mathcal{E} \to \overline{\mathbb{R}}_+$$

Such that:

1. *$\omega \to M(\omega, A)$ is a r.v. $\forall A \in \mathcal{E}$ denoted as $M(A)$, which is $\mathcal{H}$-measurable and takes values on $(E, \mathcal{E})$*
2. *$A \to M(\omega, A)$ is a measure on $(E, \mathcal{E})$ denoted as $M_\omega(dx)$ for all $\omega \in \Omega$*

♠ **Definition 13.2** (Measure description of M). *The measure in M denoted as $M_\omega(dx)$ can be atomic or diffuse (Def. A.32), finite, σ-finite or Σ-finite (Defs. A.26, A.27).*

♠ **Definition 13.3** (Random counting measure). *$M(dx)$ such that $M_\omega(dx)$ atomic and with weight 1 a.s. is a random counting measure. It is the equivalent of a counting measure after fixing ω.*

♡ **Example 13.4** (Counting process as Definition 11.13). *For ordered distinct arrival times $0 < T_1 < \dots$ the counting process $N_t = \sum_{n=1}^{\infty} \mathbb{1}_{[0,t]}(T_n)$ can be seen as the measure arising from a random measure:*

$$N_t = M([0,t]) \quad E = \mathbb{R}_+,\ A = [0,t]$$

♠ **Definition 13.5** (Recap of integral notation). *Let $f : E \to \mathbb{R}$ be a Borel function and assume we wish to integrate wrt $M(dx)$. Recalling that for a fixed measure ν we have $\nu f = \int f(x)\nu(dx)$ then:*

$$Mf : E \to \mathbb{R} \mid Mf := \int_E f(x)M(dx) \quad \textit{is an r.v.}$$

Notice also that:

$$M(A) = \int_A M(dx) = \int_E \mathbb{1}_A M(dx) = M\mathbb{1}_A \quad \forall A \in \mathcal{E}$$

Remark 3 (About the transition kernel). *We aim to make clear why we are going to express certain objects as random variables or measures. This is due to the result of Theorem B.15, thanks to which for a transition kernel M it holds that:*

- *$Mf(\omega) = \int_E M(\omega, dx)f(x)$ is a random variable governed by $\omega \in \Omega$ on $(E, \mathcal{E})$ for each $f \in \mathcal{E}_+$*
- *$\mathbb{E}[M(A)] = \int_\Omega M(\omega, A)d\mathbb{P}[\omega]$ is a measure on $(E, \mathcal{E})$ which assigns weight to the set $A \in \mathcal{E}$*

♠ **Definition 13.6** (Expected version of random measure, mean measure). *For a random measure as in Definition 13.1 we refer to the mean measure ν when considering the measure such that:*

1. *$\nu(A) = \mathbb{E}[M(A)] \quad \forall A \in \mathcal{E}$*

2. *equivalently* $\nu f = \mathbb{E}[Mf] \quad \forall f \in \mathcal{E}_+$

In particular:

$$\nu(A) = \mathbb{E}\left[M(A)\right] = \int_\Omega M(\omega, A)\mathbb{P}[d\omega]$$

Where we are integrating out the ω *of the random measure over the underlying probability space.*

13.2 Stones in a Field and Poisson Random Measures

Lemma 13.7 (Mean in terms of tail)**.**

$$X \geq 0 \quad a.s. \implies \mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}[X \geq i] = \sum_{i=0}^{\infty} \mathbb{P}[X > i]$$

Proof. Wlog let X discrete. This is sufficient for our needs. Then:

$$\begin{aligned}
\mathbb{E}\left[X\right] &= \sum_{x=0}^{\infty} x\mathbb{P}[X = x] \\
&= \sum_{x=1}^{\infty} x\mathbb{P}[X = x] \\
&= \sum_{x=1}^{\infty} \sum_{i=1}^{x} 1 \cdot \mathbb{P}[X = x] \\
&= \sum_{x=1}^{\infty} \sum_{i=1}^{\infty} \mathbb{1}_{\{i \leq x\}} \cdot \mathbb{P}[X = x] \\
&= \sum_{i=1}^{\infty} \sum_{x=1}^{\infty} \mathbb{1}_{\{i \leq x\}} \mathbb{P}[X = x] && \text{Fubini Thm. B.30} \\
&= \sum_{i=1}^{\infty} \mathbb{P}[X \geq i] \\
&= \sum_{i=0}^{\infty} \mathbb{P}[X > i] && \text{ch. variable}
\end{aligned}$$

□

♡ **Example 13.8** (The "stones in a field" perspective)**.** *Let* $K \sim \mathcal{P}o(c)$ *be Poisson distributed. Consider* K *to be the random number of stones in a field* $E \subset \mathbb{R}^2$*. This throwing process is done always with the same mechanism with no regard to total or previous positions (i.e.* ***independence****).*

$$\mathbb{P}[K = k] = \frac{e^{-c}c^k}{k!} \mathbb{1}_{[0,1,\ldots]}(k)$$

Let X_i *be the* i^{th} *stone position.* $X_i \sim \lambda(d\vec{x})$ *is a distribution over* $E \subset \mathbb{R}^2$*.*
Assume $K \perp\!\!\!\perp \{X_i\}$*, as argued before.*
The random measure $M(dx)$ *assigns the number of stones to the* $A \subset E$ *region, mathematically:*

$$M(A) = \sum_{i=1}^{K} \mathbb{1}_A(X_i)$$

Is the number of stones in region A*.*
We will show that $M(dx)$ *is atomic counting whenever* λ *is diffuse, i.e. no two stones are in the same position (i.e. Thm. 14.10).*

♡ **Example 13.9** (Stones in a field mean measure)**.** *In the "Stones in a field" formalism, the mean measure is:*

$$c\lambda \quad c = \mathbb{E}[K]$$

We can see this as follows.
For $f = \mathbb{1}_A$ it holds:

$$Mf = M\mathbb{1}_A = \sum_{i=1}^{K} \mathbb{1}_A(X_i)$$

Similarly for $f \in \mathcal{E}_+$:

$$Mf = \int_E f(x)M(dx) = \sum_{i=1}^{K} f(X_i) = \sum_{i=1}^{\infty} f(X_i)\mathbb{1}_{\{K \geq i\}}$$

namely, a sum of images under a random number of K atoms. The last form is for convenience. Then, applying the Definition of mean measure (Def. 13.6):

$$\begin{aligned}
\mathbb{E}[Mf] &= \mathbb{E}\left[\sum_{i=1}^{\infty} f(X_i)\mathbb{1}_{\{K\geq i\}}\right] \\
&= \sum_{i=1}^{\infty} \mathbb{E}\left[f(X_i)\mathbb{1}_{\{K\geq i\}}\right] && \textit{linearity} \\
&= \sum_{i=1}^{\infty} \mathbb{E}[f(X_i)]\,\mathbb{E}\left[\mathbb{1}_{\{K\geq i\}}\right] && \textit{independence \& Fubini Thm. B.30} \\
&= \mathbb{E}[f(X_1)] \sum_{i=1}^{\infty} \underbrace{\mathbb{E}\left[\mathbb{1}_{\{K\geq i\}}\right]}_{=\mathbb{P}[K\geq i]} && \textit{iid} \\
&= \mathbb{E}[f(X_1)]\,\mathbb{E}[K] && \textit{Lem. 13.7} \\
&= \int_E f(x)\lambda(dx)\cdot c \\
&= c(\lambda f) && \textit{integral notation} \\
&= (c\lambda)f
\end{aligned}$$

Eventually, the mean measure is $\nu(dx) = c\lambda(dx)$ where $c = \mathbb{E}[K]$ as claimed.

♠ **Definition 13.10** (Laplace functional)**.** *This definition resembles that of Def. 6.11.*
For a random measure M and a positive Borel function $f \in \mathcal{E}_+$ we define the Laplace functional as:

$$\widehat{\mathcal{P}}_M(f) = \mathbb{E}\left[e^{-Mf}\right]$$

Which can be seen as the Laplace transform of Mf, which is a r.v., evaluated at $r = 1$.

♡ **Example 13.11** ("Stones in a field" Laplace functional)**.** *For $c = \mathbb{E}[K]$ in Ex. 13.8 it holds that:*

$$\widehat{\mathcal{P}}_M(f) = exp\left\{-c(\lambda(1 - e^{-f}))\right\}$$

(□ **solution)** *We perform the following long computation:*

$$
\begin{aligned}
\mathbb{E}\left[e^{-Mf}\right] &= \mathbb{E}\left[\exp\left\{-\sum_{i=1}^{K} f(X_i)\right\}\right] && \textit{Ex. 13.9}\\
&= \mathbb{E}\left[\prod_{i=1}^{K}\exp\left\{-f(X_i)\right\}\right]\\
&= \mathbb{E}\left[\mathbb{E}_K\left[\prod_{i=1}^{K}\exp\left\{-f(X_i)\right\}\right]\right] && \textit{unconditioning}\\
&= \mathbb{E}\left[\prod_{i=1}^{K}\mathbb{E}\left[\exp\left\{-f(X_i)\right\}\right]\right] && \textit{independence}\\
&= \mathbb{E}\left[\left(\mathbb{E}\left[\exp\left\{-f(X_1)\right\}\right]\right)^K\right] && \textit{iid}\\
&= \mathbb{E}\left[\left(\int_E e^{-f(x)}\lambda(dx)\right)^K\right]\\
&= \mathbb{E}\left[\left(\lambda(e^{-f})\right)^K\right]\\
&= \sum_{k=0}^{\infty}\underbrace{\left(\lambda e^{-f}\right)^k \overbrace{\frac{e^{-c}c^k}{k!}}^{\mathbb{P}[K=k]}}_{pgf\ of\ K\ \ at\ t\,=\,\lambda e^{-f}} && \lambda(e^{-f})\ \textit{is a number}\\
&= \exp\left\{-c\left(1-\lambda e^{-f}\right)\right\} && \textit{pgf closed form } X\sim \mathcal{P}o(\lambda) \implies pgf(s)=\sum_{x\geq 0}\mathbb{P}[X=x]s^x = e^{-\lambda(1-s)}\\
&= \exp\left\{-c\left(\int_E \lambda(dx) - \lambda e^{-f}\right)\right\}\\
&= \exp\left\{-c\left(\lambda(1)-\lambda e^{-f}\right)\right\}\\
&= \exp\left\{-c\left(\lambda(1-e^{-f})\right)\right\} && \textit{linearity}
\end{aligned}
$$

◇ **Observation 13.12** (Counting process as random measure)**.** *The random variable* $Mf = \sum_{i=1}^{\infty} f(X_i)$ *is the prototype of atomic random measure.*
The connection with the counting process is done via $N_t = M([0,t])$*, where the atoms of* $M(dx)$ *can be seen as the arrival times of* N_t *when* M *is atomic, and no two arrival times happen at the same time.*

♠ **Definition 13.13** (Poisson random measure, p.r.m.)**.** $N(dx) \sim \mathcal{P}ois(\nu(dx))$ *is a Poisson random measure (Def. 13.1) with mean measure* $\nu(dx)$ *when:*

1. $N(A) \sim \mathcal{P}o(\nu(A)) \quad \forall A \in \mathcal{E}$
2. *For* $\{A_i\}_{i=1}^n \subset \mathcal{E}$ *disjoint* $\implies \{N(A_i)\}_{i=1}^n$ *is an independency (Def. 6.9)*

♡ **Example 13.14** ("Stones in a field" is a Poisson Random measure)**.** $N(dx)$ *as in Ex. 13.8 is a p.r.m. in Definition 13.13 sense.*
(□ **solution)** (△ **setup)** *wts for* $\{A_i\}_{i=1}^n \subset \mathcal{E}$ *disjoint it holds:*

$$
\begin{cases}
\mathbb{P}[N(A_1)=i_1,\dots,N(A_n)=i_n] = \dfrac{e^{-\nu(A_1)}(\nu(A_1))^{i_1}}{i_1!}\cdots\dfrac{e^{-\nu(A_n)}(\nu(A_n))^{i_n}}{i_n!}\\
\nu = c\lambda : X_i \overset{iid}{\sim} \lambda, c = \mathbb{E}[K]
\end{cases}
$$

(□ **baseline)** *wlog let* $n=2$ *and* $A_1 \cap A_2 = \emptyset$ *with* $A_3 = (A_1 \cup A_2)^c$*. The collection* $\{A_1, A_2, A_3\}$ *is a partition of* E *and we might show △ there. Indeed:*

$$
\begin{cases}
\lambda(A_1)+\lambda(A_2)+\lambda(A_3) = 1\\
i_1+i_2+i_3 = k
\end{cases}
$$

where we call the former $\square(1)$ *and the latter* $\square(2)$*, with* k *a realization of the r.v.* K*.*
($\bigcirc$ ***work)*** *it holds that:*

$$\begin{aligned}\mathbb{P}[N(A_1)=i_1,N(A_2)=i_2,N(A_3)=i_3] &= \mathbb{P}[N(A_1)=i_1,N(A_2)=i_2,N(A_3)=i_3,K=k]\\ &[\square(2)]\\ &= \mathbb{P}[N(A_1)=i_1,N(A_2)=i_2,N(A_3)=i_3|K=k]\\ &[\textit{distribution is } \mathcal{M}ultinom\left(3,(\lambda(A_i))_{i=1}^3\right)]\\ &= \frac{e^{-c}c^k}{k!}\frac{k!}{i_1!i_2!i_3!}\left(\lambda(A_1)\right)^{i_1}\left(\lambda(A_2)\right)^{i_2}\left(\lambda(A_3)\right)^{i_3}\\ &= \frac{e^{-(\lambda(A_1)+\lambda(A_2)+\lambda(A_3))}c^{i_1+i_2+i_3}}{i_1!i_2!i_3!}\left(\lambda(A_1)\right)^{i_1}\left(\lambda(A_2)\right)^{i_2}\left(\lambda(A_3)\right)^{i_3}\\ &[\square(1),\square(2)]\\ &= \frac{e^{-c\lambda(A_1)}\left(\lambda(A_1)\right)^{i_1}}{i_1!}\frac{e^{-c\lambda(A_2)}\left(\lambda(A_2)\right)^{i_2}}{i_2!}\frac{e^{-c\lambda(A_3)}\left(\lambda(A_3)\right)^{i_3}}{i_3!}\end{aligned}$$

13.3 Properties of Poisson Random Measures

♡ **Example 13.15** (Homogeneous counting measure and Weibull)**.** *Let* $N(dx,dy)$ *be a p.r.m. on* $E=\mathbb{R}^2$*, with mean measure* $\nu(dx,dy)=cLeb(dx,dy)$*. It holds that* N *is invariant to translations and rotations (i.e. homogeneous). Let* R *be the distance of the closest atom of* N *from the origin* $\boldsymbol{0}=(0,0)$*. We describe* R *via its probability distribution* $\mathbb{P}[R>r]$*. It turns out that this is equivalent to a ball having null mass:*

$$B_r(\boldsymbol{0})=\{(x,y):x^2+y^2\le r^2\}:N(B_r(\boldsymbol{0}))=0\quad\forall r>0$$

This can be seen as:

$$\begin{aligned}\mathbb{P}[R>r] &= \mathbb{P}[N(B_r(\boldsymbol{0}))=0]\\ &= e^{-\nu(B_r(\boldsymbol{0}))} && N(B_r(\boldsymbol{0}))\sim\mathcal{P}o(\nu(B_r(\boldsymbol{0})))\\ &= \exp\{-c\cdot Leb(B_r(\boldsymbol{0}))\}\\ &= \exp\{-c\cdot Area(B_r(\boldsymbol{0}))\}\\ &= \exp\left\{-c\pi r^2\right\}\end{aligned}$$

Which is the well known ***Weibull*** *distribution.*

♡ **Example 13.16** (Homogeneous Poisson random measure visibility)**.** *Let the atoms of* N *have radius* $a\approx 0$*. We interpret the model as a forest with density* $c=\mathbb{E}[K]$ *and mean measure* $\nu=Leb$*. For simplicity, we ignore the overlapping trees. By construction,* N *is homogeneous, and the horizontal direction is as good as any by rotation invariance. We refer to the distance between the origin and the closest tree as a measure of* ***visibility****. An atom with radius* a *intersects* $y=0$ *if and only if the distance between* y *and the center is* $\le a$*. Then:*

$$\{V\ge x\}=\{N(D_x)=0\}\quad D_x=[0,x]\times[-a,a]$$

is the expression in terms of sets of the visibility being greater than x*. We describe the r.v. in terms of its distribution as:*

$$\begin{aligned}\mathbb{P}[V\ge x] &= \mathbb{P}[N(D_x)=0] && N(D_x)\sim\mathcal{P}o(\nu(D_x))\\ &= e^{-\nu(D_x)}\\ &= \exp\{-cLeb([0,x]\times[-a,a])\}\\ &= \exp\{-c(2ax)\}\end{aligned}$$

♣ **Proposition 13.17** (Mean Variance for sets of Poisson random measure)**.** *For* $N(dx)\sim\mathcal{P}ois(\nu(dx))$ *such that* $\nu(A)<\infty\quad\forall A\in\mathcal{E}$*:*

1. $\mathbb{E}[N(A)]=\nu(A)$

2. $V[N(A)] = \nu(A)$
3. If $\nu(A) = \infty \implies \mathbb{E}[N(A)] = \infty$ a.s. and $V[N(A)]$ is undefined a.s.

Proof. **(Claims #1, #2)** follow by Definition 13.13#1 directly as:

$$N(A) \sim \mathcal{P}o(\nu(A)) \implies \mathbb{E}[N(A)] = V[N(A)] = \nu(A) \quad \forall A \in \mathcal{E}$$

(Claim #3) again trivial since:

$$V[N(A)] = \mathbb{E}\left[(N(A))^2\right] - \mathbb{E}[N(A)]^2 = \infty - \infty \implies undefined$$

□

♣ **Proposition 13.18** (Mean and variance for functions, Poisson random measure). *Let N be a p.r.m. and $f \in \mathcal{E}_+$:*

1. $\mathbb{E}[N(f)] = \nu(f)$
2. $V[N(f)] = \nu(f^2)$ if $\nu f < \infty$

Proof. **(Claim #1)** it holds by Definition 13.13 and $N(A) \sim \mathcal{P}o(\nu(A)) \forall A \in \mathcal{E}$ that $\mathbb{E}[Nf] = \nu f$.
(Claim #2) we have:

$$\begin{aligned} V[Nf] &= \mathbb{E}\left[(Nf)^2\right] - \mathbb{E}[Nf]^2 \\ &= \nu(f^2) + (\nu f)^2 - (\nu f)^2 \qquad f_n = \sum a_i \mathbb{1}_{A_i} \nearrow f, \quad f_n : \mathbb{E}\left[(Nf_n)^2\right] = \nu(f_n^2) + (\nu f)^2 \\ &= \nu(f^2) \end{aligned}$$

□

♣ **Theorem 13.19** (Laplace functional of Poisson random measure characterization). *Using the theory of Laplace transforms, for a random measure N on $(E, \mathcal{E})$ (Def. 13.1) with mean measure ν:*

$$N \sim \mathcal{P}ois(\nu) \quad (Def.\ 13.13) \iff \mathbb{E}[e^{-Nf}] = e^{-\nu(1-e^{-f})} \quad \forall f \in \mathcal{E}_+$$

Proof. **(△ strategy)** we first show the claim for f_n simple and then use a $f_n \nearrow f$ argument.
($\implies$) choose $a \in \mathbb{R}_+, A \in \mathcal{E} : \nu(A) < \infty$. Then:

$$\begin{aligned} \mathbb{E}\left[\exp\{-aN(A)\}\right] &= \sum_0^\infty \frac{e^{-\nu(A)}(\nu(A))^k}{k!} e^{-ak} \\ &= \sum_0^\infty \frac{e^{-\nu(A)}(\nu(A)e^{-a})^k}{k!} \\ &= \exp\left\{-\nu(A)(1-e^{-a})\right\} \qquad \text{Poisson pgf} \end{aligned}$$

Let f_n be simple, namely $f_n = \sum a_i \mathbb{1}_{A_i}$ with A_i disjoint. It holds that $Nf = \sum a_i N(A_i)$ and by the A_i being disjoint they are independent:

$$\implies \mathbb{E}\left[e^{-Nf}\right] = \prod_{i=1}^n \mathbb{E}\left[e^{-a_i N(A_i)}\right] = \exp\left\{-\sum \nu(A_i)(1-e^{-a_i})\right\}$$

(□ $f \in \mathcal{E}_+$ arbitrary) For $(f_n)_{n\in\mathbb{N}} \subset \mathcal{E}_+$ such that $f_n \nearrow f$, by the continuity of Lemma 13.20#1 we have that:

$$\mathbb{E}\left[e^{-Nf}\right] = \lim_{n\to\infty} \mathbb{E}\left[e^{-N(f_n)}\right] = \lim_{n\to\infty} \exp\left\{-\nu(1-e^{-f_n})\right\}$$

Where $g_n = 1 - e^{-f_n} \nearrow 1 - e^{-f} = g \implies \nu(g_n) \nearrow \nu(g)$ by monotone convergence (Thm. 4.24), giving us the equality needed.
($\impliedby$) immediate by ($\implies$) and Lemma 13.20#2. □

Lemma 13.20 (Laplace functional uniqueness and continuity). *The Laplace functional mapping $f \to \mathbb{E}[e^{-Mf}]$ for $f \in \mathcal{E}_+$ is such that:*

1. $(f_n) \subset \mathcal{E}, f_n \nearrow f \implies \lim_{n\to\infty} \mathbb{E}\left[e^{-Mf_n}\right] = \mathbb{E}\left[e^{-Mf}\right]$
2. $N = M$ *on* $(E, \mathcal{E})$ *random measures* $\iff \widehat{\mathcal{P}}_M(f) = \widehat{\mathcal{P}}_N(f)\ \forall f \in \mathcal{E}_+$

Proof. **(Claim #1)** take $f_n \nearrow f$, by Monotone convergence (Thm. 4.21) it holds $Mf_n \nearrow Mf$ for all M_ω pathwise in ω. Then, by bounded convergence (Cor. 4.26) also:

$$\mathbb{E}\left[e^{-Mf_n}\right] \searrow \mathbb{E}\left[e^{-Mf}\right]$$

Eventually we have the following chain of equalities:

$$\widehat{\mathcal{P}}_M(f) = \widehat{\mathcal{P}}_M(\lim_{n\to\infty} f_n) = \mathbb{E}\left[e^{-Mf}\right] = \mathbb{E}\left[\lim_{n\to\infty} e^{-Mf_n}\right] = \lim_{n\to\infty} \mathbb{E}\left[e^{-Mf_n}\right] = \lim_{n\to\infty} \widehat{\mathcal{P}}_M(f_n)$$

(Claim #2) is analogous to Theorem 6.12. □

Corollary 13.21 (Extending the results of Theorem 13.19). *Clearly:*

$$\widehat{\mathcal{P}}_M(f) = \widehat{\mathcal{P}}_N(f) \quad \forall f \in \mathcal{E}_+ \iff M = N \quad a.s. \iff M \text{ r.m.} \quad \text{specified by } \nu \text{ only}$$

Proof. The first characterization holds by Theorem 13.19.
For the second, notice that:

$$\widehat{\mathcal{P}}_M(f) = \exp\left\{-\nu(1 - e^{-f})\right\} \quad \forall f \in \mathcal{E}_+$$

So M is completely specified by ν. □

♣ **Proposition 13.22** (Laplace function of $N(A)$). *We provide quickly an intuition of the $\implies$ direction in the Proof of Theorem 13.19 for the simplest case possible.*
We can show for $r = 1$ that :

$$\mathbb{E}\left[e^{-1\cdot N(A)}\right] = \exp\{-\nu(1 - e^{-\mathbb{1}_A})\}\}$$

and then reason by simple functions approximation.

Proof. observe that:

$$1 - e^{-\mathbb{1}_A} = \begin{cases} 1 - e^{-1} & x \in A \\ 1 - 1 & x \notin A \end{cases} = (1 - e^{-1})\mathbb{1}_A$$

which eases out the main computation:

$$\begin{aligned}
\mathbb{E}\left[e^{-1\cdot N(A)}\right] &= \exp\{-\nu(A)(1 - e^{-1})\} && \text{Laplace transform at } r = 1 \\
&= \exp\left\{-\int_A (1 - e^{-1})\nu(dx)\right\} \\
&= \exp\left\{-\int_E (1 - e^{-1})\mathbb{1}_A\nu(dx)\right\} \\
&= \exp\left\{-\int_E (1 - e^{-\mathbb{1}_A})\nu(dx)\right\} && \text{previous equation} \\
&= \exp\left\{-\nu(1 - e^{-\mathbb{1}_A})\right\}
\end{aligned}$$

□

Chapter Summary

Objects:

- random measure, a transition kernel between $(\Omega, \mathcal{H})$ and $(E, \mathcal{E})$
- a counting process is the realization of a random counting measure $N_t = M((0,t])$ for any $t \in \mathbb{T}$
- $Mf(\omega) = \int_E M(\omega, dx) f(x)$ is a random variable on $(E, \mathcal{E})$. $M(A)$ is a random variable in the indicator version
- $M_\omega(dx)$ is the underlying measure, dependent on ω. To have no randomness we can say that $\mathbb{E}[M(A)] = \nu(A) = \int_A M(\omega, A) d\mathbb{P}$ is the mean measure on $(E, \mathcal{E})$ with $\nu f = \mathbb{E}[Mf]$ for any $f \in \mathcal{E}$
- the Laplace functional $\widehat{\mathcal{P}}_M(f) = \mathbb{E}\left[e^{-Mf}\right]$ for $f \in \mathcal{E}_+$
- Poisson random measures as random counting measures with $N(A) \sim \mathcal{P}o(\nu(A))$ for any $A \in \mathcal{E}$ and independence of $\{N(A_k)\}_{k=1}^n$ for any disjoint set $\{A_k\}_{k=1}^n$

Results:

- mean measure of stones in a field if $c\lambda(dx)$ for $K \sim \mathcal{P}o(c)$.
  trick: expression as sum, isolate K, use tail probability trick proved by Fubini
- stones in a field Laplace functional is $\exp\{-c\lambda(1 - e^{-f})\}$
- stones in a field is a p.r.m.
- mean and variance for $N(A)$ and $N(f)$ p.r.m.s
- Laplace functional of p.r.m. is like stones in a field and is $\exp\{-\nu(1 - e^{-f})\}$ for mean ν

Chapter 14

Atomic View of Poisson Random Measures

♢ **Observation 14.1** (Recalling atom and atomic measures). *We have that:*

- $x \in E$ *is an atom for* $\mu \iff \mu(\{x\}) = 0$*, Def. A.31*
- μ *is atomic* $\iff D = \{x \text{ atoms}\}$ *is such that* $\mu(E \setminus D) = 0$*, Def. A.32*

♠ **Definition 14.2** (Proper random variable for random measure). *Given* $f \in \mathcal{E}_+$ *we say* $Mf = \int_E f(x)M(dx)$ *is proper when* $\mathbb{P}[Mf < \infty] = 1$*, namely* $Mf \overset{a.s.}{=} 1$*.*

Lemma 14.3 (Finiteness of random variable by Laplace function).

$$X \geq 0 \quad a.s. \implies \mathbb{P}[X < \infty] = \lim_{r \to 0} \widehat{\mathcal{P}}_X(r)$$

Proof. Proposition C.19. □

14.1 Other Properties of Poisson Random Measures

♣ **Proposition 14.4** (Finiteness of Poisson random measure). *Let* $f \in \mathcal{E}_+$ *and* $N \sim Pois(\nu)$*. Then:*

$$\nu(f \wedge 1) < \infty \implies Nf < \infty \quad a.s.$$

Else $Nf = \infty$ *a.s.*

Proof. **(△ strategy)** we use Lemma 14.3 and $f \in \mathcal{E}_+$ to show that $Nf \geq 0$ a.s. as a r.v. This means showing:

$$\mathbb{P}[Nf < \infty] = \lim_{r \to 0} \mathbb{E}\left[e^{-rNf}\right] = 1$$

This is also equivalent, using continuity (Lem. 13.20) and the Laplace functional Theorem 13.19 to:

$$\lim_{r \to 0} \mathbb{E}\left[e^{-rNf}\right] = \lim_{r \to 0} \mathbb{E}\left[e^{-Nrf}\right] = \lim_{r \to 0} \exp\left\{-\nu(1 - e^{-rf})\right\} = 1$$

(□ another simplification) looking at the last result, it also holds:

$$\begin{aligned}\lim_{r \to 0} \exp\left\{-\nu(1 - e^{-rf})\right\} = 1 &\iff \lim_{r \to 0} -\nu(1 - e^{-rf}) = 0 \\ &\iff \int 1 - e^{-rf(x)}\nu(dx) \overset{r \to 0}{\to} 0 \\ &\iff 1 - e^{-rf(x)} \overset{r \to 0}{\to} 0 \qquad \text{as } 1 - e^{-rf(x)} \geq 0 \forall x\end{aligned}$$

Here it holds that $1 - e^{-rf(x)} \leq (f(x) \wedge 1) \quad \forall x$ so that by Dominated Convergence (Thm. 4.24):

$$\nu(f \wedge 1) < \infty \implies \lim_{r \to 0} \int 1 - e^{-rf(x)}\nu(dx) = \int \lim_{r \to 0} 1 - e^{-rf(x)}\nu(dx) = 0 \implies \mathbb{P}[Nf < \infty] = 1$$

(◯ opposite) Contrarily, if $\nu(f \wedge 1) = \infty$ we conclude $Nf = \infty$ almost surely. Indeed by:

$$1 - e^{-t} \geq (1 - e^{-1})(t \wedge 1) \quad \forall t \geq 0 \quad \mathbb{E}\left[e^{-Nf}\right] = \exp\{-\nu(1 - e^{-f})\}$$

we can notice that

$$\begin{aligned}
\nu(1 - e^{-f}) &= \int (1 - e^{-f(x)})\nu(dx) \\
&\geq \int (1 - e^{-1})(f \wedge 1)\nu(dx) \\
&= (1 - e^{-1}) \int (f \wedge 1)\nu(dx) \\
&= +\infty \qquad\qquad \nu(f \wedge 1) = \infty \\
\implies \mathbb{E}\left[e^{-Nf}\right] &= \exp\{-\nu(1 - e^{-f})\} \\
&\leq \exp\{-\underbrace{\nu((1 - e^{-1})(f \wedge 1))}_{=\infty}\} \\
&= 0 \\
\implies \mathbb{E}\left[e^{-Nf}\right] &= 0
\end{aligned}$$

which, by an application of the inverse of Jensen's inequality for concave functions $f(x) = x^r$ with $r < 1$ (namely, Cor. 7.8) suggests that:

$$\mathbb{E}\left[e^{-rNf}\right] = \mathbb{E}\left[(e^{-Nf})^r\right] \leq \left(\mathbb{E}\left[e^{-Nf}\right]\right)^r = 0 \quad \forall r > 0$$

which in the limit means:

$$\mathbb{P}[Nf < \infty] = \lim_{r \to 0} \mathbb{E}\left[e^{-rNf}\right] = 0 \implies \mathbb{P}[Nf = \infty] = 1$$

and $Nf = \infty$ almost surely. □

♢ **Observation 14.5** (Sketch for proving the Proposition)**.** *Basically we implement the bound:*

$$\begin{cases} 1 - e^{-ry} \leq ry < y & r < 1 \\ y \wedge 1 & \textit{bounds the whole graph since 1 is asymptote} \end{cases}$$

♠ **Definition 14.6** (Independent random measures)**.** *Two random measures N, M are such that $N \perp\!\!\!\perp M$ when $N(A) \perp\!\!\!\perp M(A)\ \forall A \in \mathcal{E}$*

♢ **Observation 14.7** (Finiteness of measures recap)**.** *Recall that:*

- *ν finite: $\nu(E) < \infty$*
- *ν σ-finite if there exists a partition of the sample space which covers it and each elements has finite measure*
- *ν Σ-finite: $\exists(\nu_n)$ such that $\nu_n(E) < \infty \forall n$ and $\nu = \sum_n \nu_n$*

Do we have a p.r.m. N given a mean measure ν? Theorem 14.8 shows this for Σ-finite measures.

♣ **Theorem 14.8** (Poisson random measure existance)**.** *Let ν be Σ-finite on $(E, \mathcal{E})$. Then:*

$$\exists(\Omega, \mathcal{H}, \mathbb{P}) \quad \& \quad N(\omega, \cdot) \textit{ on } (E, \mathcal{E}) : N \sim Pois(\nu) \quad \forall \omega \in \Omega$$

Proof. **(△ ν finite)** let $c = \nu(E) < \infty$ and set $\lambda(dx) = \frac{\nu(dx)}{c}$. λ is a probability measure. Let $\pi \sim \mathcal{P}o(c)$ and construct the probability space:

$$(\Omega, \mathcal{H}, \mathbb{P}) = (\mathbb{N}, 2^{\mathbb{N}}, \pi) \times (E, \mathcal{E}, \lambda)^{\mathbb{N}}$$

Where an event takes form

$$\vec{\omega} \in \mathbb{N} \times E^{\mathbb{N}}, \quad \vec{\omega} = (\omega_0, \omega_1, \omega_2, \dots)$$

By Ionescu-Tulcea (Thm. 10.57), such space exists. Let $K(\omega) = \omega_0$ and $X_i(\omega) = \omega_i \forall i \in \mathbb{N}^*$. It holds:

$$K \perp\!\!\!\perp X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp \cdots \quad K \sim \mathcal{P}o(c), \quad X_i \overset{iid}{\sim} \lambda$$

Then, the Random measure:

$$N(\omega,\cdot)\ :\ N(\omega,A)=\sum_{i=1}^{K(\omega)}\mathbb{1}_A\left(X_i(\omega)\right)$$

Is a p.r.m. by analogy given Example 13.14.
(□ Σ-finite ν) For (ν_n) all finite such that $\sum_n \nu_n=\nu$ we construct the same probability space and p.r.m. as △:

$$(\Omega_n,\mathcal{H}_n,\mathbb{P}_n)\quad N_n$$

And then set their product space:

$$(\Omega,\mathcal{H},\mathbb{P})=\bigotimes_n(\Omega_n,\mathcal{H}_n,\mathbb{P}_n)\quad\forall\boldsymbol{\omega}\in\Omega\quad\boldsymbol{\omega}=(\vec{\omega_1},\vec{\omega_2},\ldots),\vec{\omega_n}\in\Omega_n\forall n$$

Define now the objects:

$$\widetilde{N}_n(\omega,A)=N(\vec{\omega_n},A),\quad N(\omega,A)=\sum_n\widetilde{N}_n(\omega,A)$$

The tensor product ensures that $\widetilde{N}_n\perp\!\!\!\perp\widetilde{N}_m\forall m,n$, and all are defined on $(\Omega,\mathcal{H},\mathbb{P})$. Then, for $f:\Omega\to E, f\in\mathcal{E}_+$ it holds that for a finite collection:

$$\mathbb{E}\left[\exp\left\{-\sum_{i=1}^n\widetilde{N}_if\right\}\right]=\prod_{i=1}^n\exp\left\{-\nu_i(1-e^{-f})\right\}=\exp\left\{-\sum_{i=1}^n\nu_i(1-e^{-f})\right\}$$

So that for a countable collection:

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}\left[\exp\left\{-\sum_{i=1}^n\widetilde{N}_if\right\}\right]&=\lim_{n\to\infty}\exp\left\{-\sum_{i=1}^n\nu_i(1-e^{-f})\right\}\\&=\exp\left\{-\sum_{i=1}^\infty\nu_i(1-e^{-f})\right\}\\&=\exp\left\{-\nu(1-e^{-f})\right\}\qquad\qquad\sum_n\nu_n=\nu\end{aligned}$$

Which means that N is a p.r.m. with mean ν by Theorem 13.19 since it also holds:

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}\left[\exp\left\{-\sum_{i=1}^n\widetilde{N}_if\right\}\right]&=\lim_{n\to\infty}\prod_{i=1}^n\mathbb{E}\left[\exp\left\{-\widetilde{N}_if\right\}\right]\\&=\prod_{i=1}^\infty\mathbb{E}\left[\exp\left\{-\widetilde{N}_if\right\}\right]\\&=\mathbb{E}\left[\exp\left\{-\sum_{i=1}^\infty\widetilde{N}_if\right\}\right]\\&=\mathbb{E}\left[\exp\{-Nf\}\right]\end{aligned}$$

□

14.2 Simulation

♢ **Observation 14.9** (Monte Carlo simulation of Poisson random measure)**.** *We construct $N_\omega(dx)$ for $\omega\in\Omega$ of a p.r.m. $N(dx)$ (Def. 13.13).*
The aim is to do so from a set of uniform variables over $(0,1)$.
Let $E=\mathbb{R}_+\times\mathbb{R}_+$ and $\nu=cLeb^2$, so that $\nu(dx,dy)=cdxdy$. For simplicity set $c=1$. Observe that ν is σ-finite.

- *pick the a-sized square $E_0=[0,a]\times[0,a]$ and generate $K\sim Po(a^2)$ using u_0 (classic simulation of an r.v.)*
- *assign to its realization k the entry ω_0.*
- *form $(\omega_1,\ldots\omega_{2k})$ as pairs $(au_1,au_2),\ldots,(au_{2k-1},au_{2k})$ atoms with unit weight*

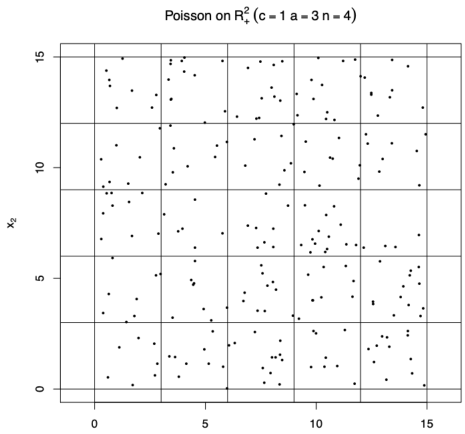

Figure 14.1: Poisson Random Measure on $\mathbb{R}_+ \times \mathbb{R}_+$

- *repeat for different squares to get a p.r.m. realization in finite time*

See Figure 14.1 for a plot.

♣ **Theorem 14.10** (Random counting measure and diffusivity of Poisson random measure)**.** *Let N be a p.r.m. on $(E, \mathcal{E})$ according to Definition 13.13, with Σ-finite mean measure ν. Then:*

$$N \quad \textit{random counting measure (Def. 13.3)} \iff \nu \quad \textit{diffuse (Def. A.32)}$$

Proof. **($\implies$)(△ setting)** fix $x \in E$, let $c = \nu(\{x\})$, we want to show that $c = 0$.
(□ nullity of singletons) for $c = \nu(\{x\})$ it holds that the set $\{N(\{x\}) \geq 2\}$ has measure zero (is negligible) since otherwise N would not be a counting measure. Then, by N being a p.r.m.:

$$N(\{x\}) \sim \mathcal{P}o(\nu(\{x\})) = \mathcal{P}o(c)$$

where the probability of the negligible set is:

$$\mathbb{P}\left[\{N(\{x\}) \geq 2\}\right] = 1 - \mathbb{P}\left[\{N(\{x\}) < 2\}\right] = 1 - \mathbb{P}\left[\{N(\{x\}) = 0\}\right] - \mathbb{P}\left[\{N(\{x\}) = 1\}\right] = 1 - e^{-c} - ce^{-c} = 0$$

which holds if and only if $c = 0$.
($\impliedby$) let ν be diffuse and Σ-finite.
(◯ construct N) by Corollary 13.21 the mean measure ν completely characterizes the p.r.m. N. We aim to build N as in Theorem 14.8. This leads us to the decomposition of N into:

$$N = \sum_n N_n \quad N_n = \sum_{i \leq K_n} \mathbb{1}_{\{X_{n,i}, \cdot\}}$$

for $X = \{X_{n,i} : n \geq 1, i \geq 1\}$ an independency. Every element of X is from a diffuse distribution by ν being diffuse. Then, the sets:

$$\{X_{n,i} = X_{m,j}\} \quad \text{for some } n, i, m, j$$

are negligible and so is their union by being countable:

$$\Omega_0 = \bigcup_{m\in\mathbb{N}} \bigcup_{n\in\mathbb{N}} \bigcup_{i\in\mathbb{N}} \bigcup_{j\in\mathbb{N}} \{X_{n,i} = X_{m,j}\}$$

clearly, N is a random counting measure, since $\forall\omega$ non negligible (i.e. out of Ω_0) N_ω is a counting measure □

Corollary 14.11 (Extension to special case)**.** *Let $N \sim Pois(\nu)$ on $E = \mathbb{R}_+ \times \mathbb{R}_+$ and $\mathcal{E} = \mathcal{B}(E)$.*
Let $\nu = Leb \times \lambda$, with

- $\lambda(\{0\}) = 0$
- $\lambda((\epsilon, \infty)) < \infty \quad \forall \epsilon > 0$

We can interpret $N(t, z)$ for a time of arrival t of an object of size z. Then:

1. *for a.e. $\omega \in \Omega$ N_ω is a counting measure that:*
2. *(no simultaneity) has not atom at $t = 0$, no atom of size $z = 0$, i.e. no simultaneity of $X_i, X_j : t_i = t_j$*
3. *(finite big activity) $\forall t < \infty, \epsilon > 0$ there are finitely many atoms before t with size $z > \epsilon$*
4. *(infinite small activity) claim #3 holds for $\epsilon = 0$ if λ is finite. Otherwise there are ∞ many atoms of size $z \leq \epsilon \quad \forall \epsilon > 0$*

Proof. **(Claim #1)** Theorem 14.10.
(Claim #2) the p.r.m. N is such that:

- by $Leb(\{0\}) \overset{a.s.}{=} 0$ there is almost surely no mass on $\{0\} \times \mathbb{R}_+$
- by $\lambda(\{0\}) \overset{a.s.}{=} 0$ there is no mass on $\mathbb{R}_+ \times \{0\}$

Let $\Omega_0 = (\{0\} \times \mathbb{R}_+) \cup (\mathbb{R}_+ \times \{0\})$, the set we just described. Fix $\epsilon > 0$. The random variable $M(A) = N(A \times (\epsilon, \infty))$ on $\mathbb{R}_+$ for arbitrary $A \in \mathcal{B}(\mathbb{R})$ is:

- a p.r.m. with mean $\mu = \lambda((\epsilon, \infty))$
- a counting random measure since Leb is diffuse and we can apply Theorem 14.10

Then:

$$\exists \Omega_\epsilon : \mathbb{P}[\Omega_\epsilon] = 1, \ \forall \omega \in \Omega_\epsilon \quad (t, z), (t', z') : z > \epsilon, z' > \epsilon$$

Which are the atoms of a realization N_ω, with the peculiarity that $t \neq t'$ since M is a random counting measure. Let:

$$\Omega_a = \Omega_0 \cap \left(\bigcap_{\epsilon > 0} \Omega_\epsilon \right) = \Omega_0 \cap \left(\bigcap_{\epsilon \in \mathbb{Q}} \Omega_\epsilon \right)$$

and Ω_a is the almost sure set where Claim #2 holds.
(Claim #3) ν puts mass $t \cdot \lambda((\epsilon, \infty)) < \infty$ on the set $[0, t] \times (\epsilon, \infty)$. Clearly, there exists an almost sure set $\Omega_{t,\epsilon}$ such that N has finite atoms there. Let:

$$\Omega_b = \bigcap_{t \in \mathbb{N}} \bigcap_{\epsilon \in \mathbb{Q}} \Omega_{t,\epsilon}$$

the set is almost sure and Claim #3 is true for every ω in it.
(Claim #4) for $\lambda(\mathbb{R}_+) < \infty$ so that the λ measure is finite, construct the set:

$$\Omega_c = \bigcap_{t \in \mathbb{N}} \Omega_{t,0}$$

for λ not finite, the mean measure of N is such that:

$$\nu((t, t + \delta) \times (0, \epsilon]) = \delta \lambda((0, \epsilon]) = \infty$$

By the hypothesis that $\lambda((\epsilon, \infty)) < \infty$, there is an almost sure event $\Omega_{t,\delta,\epsilon}$ with finitely many atoms in $(t, t + \delta) \times (0, \epsilon]$.
We then let:

$$\Omega_c = \bigcap_{t \in \mathbb{N}} \bigcap_{\epsilon \in \mathbb{Q}} \bigcap_{\delta \in \mathbb{Q}} \Omega_{t,\epsilon,\delta}$$

which is an almost sure event where Claim #4 holds.
All the statements hold almost surely in the set:

$$\Omega_a \cap \Omega_b \cap \Omega_c$$

□

♠ **Definition 14.12** (Image of N under h, $N \circ h^{-1}$)**.** *This is equivalent to Definition B.1.*
Let N be a p.r.m. on $(E, \mathcal{E})$, and $h : E \to F$ a measurable map (satifies Eqn. 3.1). The image of N under h is a random measure on $(\Omega, \mathcal{H}, \mathbb{P}), (F, \mathcal{F})$ (Def. 13.1) defined as:

$$N \circ h^{-1} \quad : \quad (N \circ h^{-1})(B) = N \circ (h^{-1}(B)) \quad \forall B \in \mathcal{F}$$

the last expression is:

$$\begin{aligned} N(h^{-1}(B)) &= \int_E \mathbb{1}_{h^{-1}(B)}(x)\, N(dx) \\ &= \int_{x:h(x)\in B} N(dx) \\ &= \int_E \mathbb{1}_B(h(x))\, N(dx) \\ &= N(\mathbb{1}_B \circ h) \end{aligned}$$

Where we infer that instead for a borel map $f : F \to \mathbb{R}$:

$$(N \circ h^{-1})(f) = N(h^{-1}(f)) = N(f \circ h)$$

which by $Nf = \sum f(X_i)$ for (X_i) atoms of N suggests that:

$$(N \circ h^{-1})(f) = N(f \circ h) = \sum_{i=1}^{K} f(h(X_i)) = \sum_{i=1}^{K} f(Y_i)$$

For (Y_i) the atoms of $N \circ h^{-1}$.

♣ **Proposition 14.13** (Image measure is a Poisson random measure). *$N \circ h^{-1}$ on $(\Omega, \mathcal{H}, \mathbb{P}), (F, \mathcal{F})$ satisfies the requirements of Definition 13.13 and has mean $\mu = \nu \circ h^{-1}$.*

$$N \sim \mathcal{P}ois(\nu), \quad h : E \to F \implies N \circ h^{-1} \sim \mathcal{P}ois(\nu \circ h^{-1})$$

Proof. For all $B \in \mathcal{F}$ it holds that $(N \circ h^{-1})(B) \sim \mathcal{P}o(\nu \circ h^{-1}(B))$. If we take $\{B_i\}_{i=1}^n \subset \mathcal{F}$ disjoint, it is also the case that $\{(N \circ h^{-1})(B_i)\}_{i=1}^n$ is an independent set. □

♢ **Observation 14.14** (Mean measure and integrals recap). *we quickly refresh already introduced notation:*

- $(N \circ h^{-1})(f) = N(f \circ h)$
- $\mu(f) = (\nu \circ h^{-1})(f) = \nu(f \circ h) = \int_E f(h(x))\nu(dx)$

♢ **Observation 14.15** (What we want & need). *We aim to simulate $N \sim Pois(\nu)$ for any mean measure ν. By Observation 14.9 we can only do so for $\nu = cLeb$.*
Eventually, we will implement a $N \circ h^{-1}$ image construction to relate to the simulation we can perform.

14.3 Arrival Process

♠ **Definition 14.16** (Arrival process formalism). *Let $N(dx)$ be a p.r.m. on $E = \mathbb{R}_+$ with diffuse mean $\nu(dx)$, such that $c(t) = \nu((0, t]) < \infty\ \forall t$.*
By Theorem 14.10 we know that ν is diffuse $\iff$ N is a random counting measure.
With this premise we can interpret $(T_k)_{k\geq 1}$ as distinct ordered arrival times. We want to simulate this random measure.

♣ **Proposition 14.17** (Arrival process simulation by inverse image). *For N as in Definition 14.16 let $h : \mathbb{R}_+ \to \mathbb{R}_+$ be such that in the arrival process formalism:*

$$h(u) = t \iff c(t) = u$$

Namely, the inverse of the cdf. Then for $\widetilde{N}$ a p.r.m. with mean measure Leb:

1. *$\nu = Leb \circ h^{-1}$*
2. *$(u_i)_{i\geq 1}$ are the atoms and $(h(u_i))_{i\geq 1}$ are the atoms of N*

Proof. **(Claim #1)** consider $\widetilde{N} \sim \mathcal{P}ois(Leb)$, h as hypothesized. By Proposition 14.13, $\widetilde{N} \circ h^{-1}$ is a p.r.m. as well with mean:

$$\begin{aligned}
(Leb \circ h^{-1})(A) = Leb(h^{-1}(A)) &= \int_{x:h(x)\in A} dx && \text{wlog let } A = [a, b] \\
&= \int_{c(a)}^{c(b)} dx && a < h(x) \le b \iff c(a) < x \le c(b) \\
&= c(b) - c(a) \\
&= \nu((0, b]) - \nu((0, a]) && \text{Def. 14.16} \\
&= \nu([a, b]) \\
&= \nu(A) && \forall A = [a, b]
\end{aligned}$$

It is evident that $N = \widetilde{N} \circ h^{-1}$ is the p.r.m. we want to simulate as it has the same mean over any Borel set. Indeed, by the classic result of Proposition A.29 we have $\nu = Leb \circ h^{-1}$.
(Claim #2) let $(U_i)_{i\ge1} = (u_i)_{i\ge1}$ be the atoms of $\widetilde{N}$, according to Definition B.1 the atoms of $N = \widetilde{N} \circ h^{-1}$ with $N(B) = \widetilde{N}(h^{-1}(B)), Nf = \widetilde{N}(f \circ h)$ are $(h(u_i))_{i\ge1}$ and we can simulate $Nf = \sum_{i\ge1} f(h(u_i))$. Notice that the Poisson random measure N has the desired mean ν. □

♠ **Definition 14.18** (Trace of random measure, also restriction)**.** *For $D \subset E$ and a random measure M on E we call restriction the measure M_D characterized as:*

$$M_D(B) := M(B \cap D) \quad \forall B \in \mathcal{E}$$

Which has mean $\mu_D(B) = \mu(B \cap D)\ \forall B \in \mathcal{E}$

♠ **Definition 14.19** (Intensity or expected arrival time r)**.** *In the context of Definition 14.16 further let ν be σ-finite and such that $\nu \ll Leb$. By Radon Nikodym Theorem (Thm. 5.7) we have that:*

$$\exists r \text{ Leb-measurable}, \quad \nu(A) = \int_A r(t)dt$$

We call $r(t) = \frac{d\nu}{dLeb}(t)$ the Radon-Nykodym derivative also with the term intensity.
Recall the discussion we did in Chapter 5. It is not granted that the measure ν will be σ-finite once it is absolutely continuous to the Lebesgue measure. The observation we did when introducing the Radon-Nykodim theorem made it precise that this requirement was lifted for probability measures, but ν in principle could be just a measure. This comment can be ignored in most of the cases.

♣ **Proposition 14.20** (Arrival process simulation by intensity)**.** *Using the interpretation of Definition 14.19 for an intensity r we also let:*

- $h(t, z) = t$
- $D = \{(t, z) : z \leq r(t)\} \subset \mathbb{R}_+ \times \mathbb{R}_+$
- M_D *be the trace of the p.r.m.* M *on* $\mathbb{R}_+ \times \mathbb{R}_+$ *with mean Leb, so that it is a p.r.m. with mean* $\mu_D \ll$ *Leb. The mean measure* μ_D *is also* σ*-finite since it is just a restriction of Leb inside the set* D

Then:

1. $N = M_D \circ h^{-1}$ *is a p.r.m. with mean* $\nu = \mu_D \circ h^{-1}$. N *here is the counting measure on* $\mathbb{R}_+$ *whose atoms are arrival times* T_i *with size* $Z_i \leq r(T_i)$*, according to the restriction* D.
2. *can simulate* $(T_i, Z_i)_{i \geq 1}$ *from* M *and set* $Nf = \sum_{i:Z_i \leq r(T_i)} f(T_i)$

Proof. **(Claim #1)** let D, h as hypothesized. The p.r.m. $M_D \circ h^{-1} = N$ has mean $\nu = \mu_D \circ h^{-1}$ since for

$$D_t = \{(s, z) \in D : s \leq t, z \leq r(s)\} = \{(s, z) \in D : 0 \leq h(s, z) \leq t\} = h^{-1}((0, t]) \cap D$$

we have:

$$\begin{aligned}
(\mu_D \circ h^{-1})((0, t]) = \mu_D(h^{-1}((0, t]))) &= Leb(D \cap h^{-1}((0, t]))) && \text{Def. 14.18} \\
&= Leb(D_t) && \text{above argument} \\
&= \int_0^t r(s)ds && D_t \text{ construction} \\
&= \nu((0, t]) && \text{intensity notion}
\end{aligned}$$

(Claim #2) it trivially follows that we can simulate N by:

1. simulating from M atoms (T_i, Z_i)
2. setting $Nf = \sum_{i:Z_i \leq r(T_i)} f(T_i)$

Where Nf is the projection via h of the restriction M_D. □

Chapter Summary

Objects:

- proper random variables of random measures are almost surely finite random variables
- independent random measures are such that all the random variables are mutually independent
- the image measure $N \circ h^{-1}$ for N a random measure on $(E, \mathcal{E})$ and $h : E \to F$ a measurable map is defined as:

$$(N \circ h^1)(B) = N \circ (h^{-1}(B)) = N(\mathbb{1}_B \circ h) \quad \forall B \in \mathcal{F}, \quad (N \circ h^1)(f) = N(f \circ h)$$

  and it is a random measure on $(F, \mathcal{F})$
- the arrival process formalism for a p.r.m. on $E = \mathbb{R}_+$ with mean ν diffuse such that the cumulative $c(t) = \nu((0,t]) < \infty \forall t$ is a random counting measure that can be simulated
- trace of a random measure for $D \subset E$ we set $M_D(B) := M(B \cap D)$
- intensity of random measure, the Radon Nikodym derivative of the mean measure

Results:

- Poisson Random Measures
  - $\nu(f \wedge 1) < \infty$ is sufficient for Nf to be proper
  - a Σ-finite mean measure is sufficient for the existance of a p.r.m. $N \sim \mathcal{P}ois(\nu)$ on a probability space
  - N is a random counting measure if and only if the mean measure is diffuse
  - infinite activity: for $N \sim \mathcal{P}ois(Leb \times \lambda)$ with $\lambda(\{0\}) = 0, \lambda((\epsilon, \infty)) < \infty \forall \epsilon$ we established N_ω is a random counting measure by Leb making it diffuse such that:
    * no simultaneity in time, no events at null times or space
    * finitely many atoms before t with size $z > \epsilon$ for arbitrary ϵ
    * for λ finite the finite activity extends to $\epsilon = 0$ otherwise ∞-many atoms of size at most ϵ in an interval $[0, t]$
- Simulation
  - Monte Carlo for Lebesgue p.r.m.
    * aim: construct $N_\omega(dx)$ for $\omega \in \Omega$ of a p.r.m. $N(dx)$ from a set of uniform variables over $(0, 1)$ where N is simple
    * setting: $E = \mathbb{R}_+ \times \mathbb{R}_+$ and $\nu = cLeb^2$, so that $\nu(dx, dy) = cdxdy$. For simplicity set $c = 1$. Observe that ν is σ-finite.
    * pick the a-sized square $E_0 = [0, a] \times [0, a]$ and generate $K \sim Po(a^2)$ using u_0 (classic simulation of an r.v.)
    * assign to its realization k the entry ω_0.
    * form $(\omega_1, \dots \omega_{2k})$ as pairs $(au_1, au_2), \dots, (au_{2k-1}, au_{2k})$ atoms with unit weight
    * repeat for different squares to get a p.r.m. realization in finite time
  - Arrival process by inverse image:
    * aim: simulate arbitrary mean measure p.r.m. N, where N is an arrival process
    * setting: $h : \mathbb{R}_+ \to \mathbb{R}_+$ with $h(u) = t \iff c(t) = u$
    * consider $\widetilde{N}$ a p.r.m. with mean measure Leb
    * $\nu = Leb \circ h^{-1}$
    * $(u_i)_{i \geq 1}$ are the atoms and $(h(u_i))_{i \geq 1}$ are the atoms of N
  - Arrival process by intensity
    * aim: simulate arbitrary mean measure p.r.m. using intensity notion
    * setting: measurable map $h(t, z) = t$ and:
      · $D = \{(t, z) : z \leq r(t)\} \subset \mathbb{R}_+ \times \mathbb{R}_+$
      · M_D the trace of the p.r.m. M on $\mathbb{R}_+ \times \mathbb{R}_+$ with mean Leb, so that it is a p.r.m. with σ-finite mean $\mu_D \ll Leb$
    * $N = M_D \circ h^{-1}$ is a p.r.m. with mean $\nu = \mu_D \circ h^{-1}$. N here is the counting measure on $\mathbb{R}_+$ whose atoms are arrival times T_i with size $Z_i \leq r(T_i)$, according to the restriction D.
    * can simulate $(T_i, Z_i)_{i \geq 1}$ from M and set $Nf = \sum_{i: Z_i \leq r(T_i)} f(T_i)$

Chapter 15

Transformations & Increasing Lévy Processes

Assumption 15.1 (Setting for transformations)**.** *We consider measurable spaces* $(E, \mathcal{E}), (F, \mathcal{F})$*, and collections* $\{X_i : i \in I\}, \{Y_i : i \in I\}$*.*
N is a p.r.m. on $(E, \mathcal{E})$ with mean ν (Def. 13.13) $\Longrightarrow$ $Nf = \sum_{i \in I} f \circ X_i \quad f \in \mathcal{E}_+$*.*
For a measurable map $h : E \to F$, satisfying Equation 3.1, we set $Y_i = h \circ X_i$ and derive the new p.r.m. $N \circ h^{-1}$ using Proposition 14.13.
Y_i is ultimately the random transform associated to the kernel (Def. B.13):

$$Y_i \in B \text{ w.p. } Q(x, B) \quad \text{if } X_i = x \iff \mathbb{P}[Y \in B | X = x] = Q(x, B) \quad \forall B \in \mathcal{F}$$

Where $Q : E \times \mathcal{F} \to E$

♣ **Theorem 15.2** (Transformation independence poissonity)**.** *For a measure ν on $(E, \mathcal{E})$, and a kernel Q from $(E, \mathcal{E})$ to $(F, \mathcal{F})$ such that:*

- *X is a p.r.m. with mean ν*
- $Y_i | X \overset{ind}{\sim} Q(X_i, \cdot)$

It holds:

1. *Y is a p.r.m. on $(F, \mathcal{F})$ with mean $\pi(Q) : \pi(Q(B)) = \int_F \nu(dx) Q(x, B) \quad \forall B \in \mathcal{F}$ or in other terms* $\pi(dy) = \int_F Q(x, dy) \nu(dx)$
2. *(X, Y) is a p.r.m. on $(E \times F, \mathcal{E} \otimes \mathcal{F})$ with mean $\mu = \nu \times Q$ so that:*

$$\mu(dx, dy) = \nu(dx) Q(x, dy)$$

Proof. **($\triangle$ setting)** let N be the r.m. defined by X on $(E, \mathcal{E})$ and M be the r.m. defined by (X, Y) on the product space. Since Y defines a random measure as the image of M under the projection map $h(x, y) = y$ it holds that $\#2 \implies \#1$.
($\square$ focus on $\#2$) note that $f \in (\mathcal{E} \otimes \mathcal{F})_+$ implies by independence across I that:

$$e^{-Mf} = \prod_{i \in I} e^{-f} \circ (X_i, Y_i)$$

Choosing:

$$e^{-g(x)} := \int_F Q(x, dy) e^{-f(x,y)} \implies e^{-Mf} | X = \prod_{i \in I} \int_F Q(x, dy) e^{-f} \circ (X_i, y) = \prod_{i \in I} e^{-g} \circ X_i = e^{-Ng}$$

Which means that in expectation:

$$\begin{aligned} \mathbb{E}\left[e^{-Mf}\right] &= \mathbb{E}\left[e^{-Ng}\right] \\ &= \exp\left\{-\nu(1 - e^{-g})\right\} && \text{Thm. 13.19 for } N \sim \mathcal{P}ois(\nu) \end{aligned}$$

(◯ finalization) noticing that $Q(x, F) = 1$ we will have that:

$$\begin{aligned}\nu(1-e^{-g}) &= \int_E \nu(dx) \int_F Q(x,dy)(1-e^{-f(x,y)}) \\ &= (\nu \times Q)(1-e^{-f}) \\ &\iff M \sim \mathcal{P}ois(\nu \times Q) \qquad \text{Thm. 13.19}\end{aligned}$$

And the claim is proved. □

Corollary 15.3 (Special case Kernel is probability measure)**.** *For $X \sim \mathcal{P}ois(\nu)$ on $(E, \mathcal{E})$ and $Y \perp\!\!\!\perp X$ such that $Y \sim \pi$ on $(F, \mathcal{F})$:*

$$\implies (X,Y) \sim \mathcal{P}ois(\mu) \quad on \quad (E \times F, \mathcal{E} \otimes \mathcal{F}) \quad \mu = \nu \times \pi : \mu(dx,dy) = \nu(dx)\pi(dy)$$

Proof. Same as Theorem above. □

♡ **Example 15.4** (Poisson compound process, customers in a store)**.** *Consider a sequence of arrival times $(T_i)_{i\geq 1}$ from a p.r.m. $N \sim \mathcal{P}ois(cLeb)$. We can visualize a sequence of customers spending random money $Y \perp\!\!\!\perp T$ where $Y \sim \pi$ has mean a and variance b^2.*
Applying Corollary 15.3 we can safely say (T, Y) is a p.r.m. such that:

$$(T,Y) \sim \mathcal{P}ois(cLeb \times \pi) \quad on \quad \mathbb{R}_+ \times \mathbb{R}_+$$

Where for a fixed time $t \geq 0$ we have that the amount of money spent is:

$$\begin{aligned}Z_t = \sum_{T_i \leq t} Y_i = \sum_{i=1}^{\infty} Y_i \mathbb{1}_{[0,t]}(T_i) &= \sum_{i=1}^{\infty} f(T_i, Y_i) \qquad f(x,y) := y\mathbb{1}_{[0,t]}(x) \\ &= \int_{[0,t]\times\mathbb{R}_+} \widetilde{N}(dx,dy)y \\ &= \widetilde{N}f\end{aligned}$$

where $\widetilde{N} = (T, Y)$ is a Poisson Random measure.
We can use the previous results for p.r.m.s from Chapter 13 and 14. The new mean is $\mu = cLeb \times \pi$ with $\mu(dx, dy) = cdx\pi(dy)$ and:

$$\begin{aligned}\mathbb{E}[Z_t] = \mathbb{E}[\widetilde{N}f] &= \mu f \qquad \textit{Prop. 13.18\#1} \\ &= \int_{\mathbb{R}_+\times\mathbb{R}_+} f(x,y)\mu(dx,dy) \\ &= \int_{\mathbb{R}_+\times\mathbb{R}_+} y\mathbb{1}_{[0,t]}(x)\, cdx\pi(dy) \\ &= ct\int_{\mathbb{R}_+} y\pi(dy) \\ &= cta \qquad \textit{by } a = \mathbb{E}[Y]\end{aligned}$$

Similarly the variance is:

$$\begin{aligned}V[Z_t] = V[\widetilde{N}f] &= \mu f^2 \qquad \textit{Prop. 13.18\#2} \\ &= \int_{\mathbb{R}_+\times\mathbb{R}_+} (y\mathbb{1}_{[0,t]}(x))^2 cdx\pi(dy) \\ &= ct(a^2+b^2) \qquad \textit{by } a^2+b^2 = \mathbb{E}[Y]^2 + V[Y] = \mathbb{E}[Y^2]\end{aligned}$$

Concluding with the Laplace transform:

$$\begin{aligned}\widehat{\mathcal{P}}_{Z_t}(r) = \widehat{\mathcal{P}}_{\widetilde{N}}(rf) &= \mathbb{E}[e^{-r\widetilde{N}f}] \\ &= exp\left\{-\mu(1-e^{-rf})\right\} && \text{Thm. 13.19} \\ &= \exp\left\{-\int_{\mathbb{R}_+\times\mathbb{R}_+} 1-e^{-r(y\mathbb{1}_{[0,t]}(x))}cdx\pi(dy)\right\} \\ &= \exp\left\{-ct\int_{\mathbb{R}_+} 1-e^{-ry}\pi(dy)\right\}\end{aligned}$$

Notice that we used the random variable version with r instead of the functional version since Z_t is a random variable and not the underlying random measure.

♠ **Definition 15.5** (Compound Poisson process $(S_t)_{t\geq 0}$). *We give a precise definition of the object presented in the above example.*
For arrival times $T_1 < T_2 < \dots$ atoms of a p.r.m. on $\mathbb{R}_+$ with mean $cdx = \nu(dx)$ we consider a sequence of random variables $Y_i \overset{iid}{\sim} \pi$ on $\mathbb{R}$ where $Y \perp\!\!\!\perp T$.
The compound Poisson process that arises is the continuous time process of the random sum:

$$\begin{aligned}(S_t)_{t\geq 0} \ : \ S_t = \sum_{i:T_i\leq t} Y_i &= \sum_i \mathbb{1}_{[0,t]}(T_i)\, Y_i \\ &= \int_{\mathbb{R}_+\times\mathbb{R}} y\mathbb{1}_{[0,t]}(x)\, N(dx,dy)\end{aligned}$$

Where by Theorem 15.2 N is a p.r.m. and the expression makes sense.

♠ **Definition 15.6** (Borel version of compound Poisson process). *$(S_t)_{t\in\mathbb{R}_+}$ can be seen as a cumulative version of a r.m. (Def. 13.1) on $\mathbb{R}_+$:*

$$S_t = L((0,t]) \quad L(dx) \text{ r.m. } : \ L(A) = \int_{A\times\mathbb{R}} yN(dx,dy)$$

Indeed the Laplace transform of S_t would be:

$$\begin{aligned}\mathbb{E}\left[e^{-rL((0,t])}\right] = \mathbb{E}\left[e^{-rS_t}\right] &= \exp\left\{-ct\int_{\mathbb{R}_+}(1-e^{-ry})\pi(dy)\right\} \\ &= \exp\left\{\int_{(0,t]\times\mathbb{R}_+}(1-e^{-ry})dxc\pi(dy)\right\}\end{aligned}$$

Which we write for general A below the Observation that follows.

♢ **Observation 15.7** (Setting and aims). *we reorder ideas for the next results:*

- *y in notation is changed to z*
- *$\lambda(dz)$ is **not necessarily** finite*
- *we will try to express the Laplace functional of random measures in terms of the underlying p.r.m. using:*
  - *the result of Theorem 15.2 which allows us to do so*
  - *the intensity notion (Def. 14.19)*
  - *Theorem 13.19 for a closed form formula*

We eventually inspect:

$$\widehat{\mathcal{P}}_{L(A)}(r) = \mathbb{E}[e^{-rL(A)}] = exp\left\{\int_{A\times\mathbb{R}_+}(1-e^{-rz})\ \overbrace{dx\ \underbrace{\lambda(dz)}_{=c\pi(dy)}}^{mean\ N(dx,dz)}\right\}$$

♠ **Definition 15.8** (Additive random measure). *A random measure (Def 13.1) M is said to be additive when for disjoint sets $\{A_i\}_{i=1}^n \subset \mathcal{E}$ the set of random variables $\{M(A_i)\}_{i=1}^n$ is an independency according to Definition 6.9.*

♢ **Observation 15.9** (Additive random measure vs Independent random measure)**.** *A potentially confusing fact is that additive random measures are independent* ***within*** *themselves, while two random measures are independent* ***to each other****. Both rely on a choice of countable disjoint sets in $\mathcal{E}$. In the former, we compare random variables arising from the same measure on different sets. In the latter, we compare random variables arising from two different measures on the same sets.*

♣ **Proposition 15.10** (Compound Poisson process has underlying additive measure)**.** *L as in Definition 15.5 is an additive random measure.*

Proof. Consider a disjoint set $A_1, \dots A_n$ in $\mathcal{B}(\mathbb{R}_+)$. Then $\{A_i \times \mathbb{R}\}_{i=1}^n$ are disjoint in $\mathcal{B}(\mathbb{R}_+ \times \mathbb{R})$ and we can say that the Laplace functional transform of the underlying measure takes form:

$$\mathbb{E}\left[e^{-rL(A)}\right] = \exp\left\{-\int_{A\times\mathbb{R}} \underbrace{\nu(dx,dz)}_{dxc\pi(dz)}(1-e^{-rz})\right\}$$

By the Definition of p.r.m. (Def. 13.13), the restrictions on $A_i \times \mathbb{R}_+$ are independent random measures, so the above expression becomes:

$$\begin{aligned}
\widehat{\mathcal{P}}_{L(A)}(r) &= exp\left\{\sum_{i=1}^n \int_{A_i\times\mathbb{R}_+} (1-e^{-rz}) \overbrace{dx\ \underbrace{\lambda(dz)}_{=c\pi(dy)}}^{mean\ N(dx,dz)}\right\} \\
&= \prod_{i=1}^n exp\left\{\int_{A_i\times\mathbb{R}_+} (1-e^{-rz}) \overbrace{dx\ \underbrace{\lambda(dz)}_{=c\pi(dy)}}^{mean\ N(dx,dz)}\right\} \\
&= \prod_{i=1}^n \widehat{\mathcal{P}}_{L(A_i)}(r) \\
&\implies \{L(A_i)\}_{i=1}^n \quad independency
\end{aligned}$$

Where we used Theorem 6.12, since the random variables are identified by disjont Laplace transforms as a product, then they are independent. ☐

Lemma 15.11 (Automatic additive random measure)**.** *For a countable set $D \subset E$ and an independency of positive random variables $\{W_x : W_x \geq 0\ x \in D\}$ the random measure:*

$$K(\omega, A) = \sum_{x\in D} W_x(\omega)\mathbb{1}_A(x) \quad \omega \in \Omega, A \in \mathcal{E}$$

is additive.

Proof. $K(\omega, A)$ is additive clearly by the $\mathbb{1}_A(x)$ construction. ☐

♣ **Theorem 15.12** (A form of additive random measure decomposition)**.** *Consider a measure α on $(E, \mathcal{E})$, a random measure K as in Lemma 15.11, purely atomic with fixed atoms, and a random measure L as in Proposition 15.10, namely:*

$$L(A) = \int_{A\times\mathbb{R}_+} yN(dx,dy) \quad N \sim \mathcal{P}ois(\nu)$$

Then:

1. *any additive r.m. (Def. 15.8) can be decomposed in a sum $M = \alpha + K + L$*
2. *if M is a Σ-bounded kernel (Def. B.23) the same decomposition holds and if additionally α is diffuse, and the mean measure of K $\nu(\cdot \times \mathbb{R}_+)$ is diffuse the decomposition is unique*

Proof. Clear by Proposition 15.10, Lemma 15.11 and α being a measure. ☐

♢ **Observation 15.13** (Comments on Theorem). *Atoms are at fixed points with random weights. Ignoring the countable fixed randomness arising from $K(dx)$ from now onwards we concentrate on:*

$$S_t = M((0,t]) = \alpha((0,t]) + \int_{[0,t]\times\mathbb{R}_+} zN(dx,dz)$$

In the $\mathbb{R}_+ \times \pi(dy)$ case of Definition 15.6.

♠ **Definition 15.14** (Increasing Lévy process). *A process $S = (S_t)_{t\in\mathbb{R}_+}$ is increasing Lévy when it is such that:*

1. *independence of increments:*

$$S_{t_1} - S_{t_0}, \ldots, S_{t_n} - S_{t_{n-1}} \perp\!\!\!\perp \quad \forall n \geq 2, 0 \leq t_0 < t_1 < \ldots < t_n$$

2. *stationarity of increments*

$$S_{t+u} - S_u \overset{d}{=} S_t \quad \forall u, t \in \mathbb{R}_+$$

3. *increasing, right continuous and starting at $S_0 = 0$*

Assumption 15.15 (Structure of compound Poisson process revisited). *We know by Proposition 15.10 that the underlying random measure of a compound Poisson process is additive. We now impose that:*

- *$S_t(\omega) = M(\omega, [0,t])$ for M an additive r.m., so that S_t is increasing and right continuous*
- *$S_t < \infty$ a.s.$\forall t$ which will ensure independence by the additivity of M*
- *$\alpha(dx) = bdx \quad b \in \mathbb{R}_+$ to ensure linearity, which will guarantee stationarity of increments*

♠ **Definition 15.16** (Candidate Poisson additive random measure). *We present here the r.m. we will feed to the following results, carefully constructed according to Assumption 15.15 and Observation 15.13:*

$$S_t = bt + \int_{[0,t]\times\mathbb{R}_+} zN(dx,dz) = M(\omega,[0,t]) \quad b,t \in \mathbb{R}_+, M \text{ additive}$$

for N a Poisson random measure with mean $\nu(dx,dz) = Leb \times \lambda(dz)$.

♣ **Proposition 15.17** (Candidate compound Poisson with weak integrability is increasing Lévy). *Let $b \in \mathbb{R}_+$, N a p.r.m. on $\mathbb{R}_+ \times \mathbb{R}_+$ with mean $\nu = Leb \times \lambda$. If the integrability condition:*

$$\int_{\mathbb{R}_+} \lambda(dz)(z \wedge 1) = \lambda(z \wedge 1) < \infty$$

is satisfied then:

1. *(Lévyness) $(S_t)_{t\in\mathbb{R}_+}$ as in Definition 15.16 is an increasing Lévy process in the sense of Definition 15.14*
2. *(characterization) the Laplace transform is:*

$$\mathbb{E}[e^{-rS_t}] = \exp\left\{-t\left[br + \int_{\mathbb{R}_+} \lambda(dz)(1-e^{-rz})\right]\right\} \quad r \in \mathbb{R}_+$$

Proof. **(Claim #1)** Consider L on $E = \mathbb{R}_+$ as in Proposition 15.10. Then $M = \alpha + L$ is additive for $\alpha = Leb$.
(△ basics) $S = (S_t)_{t\in\mathbb{R}_+}$ is increasing, right continuous and starting at $S_0 = 0$ clearly.
(□ finiteness and independence of increments) Moreover:

$$S_t = bt + \sum_{i\geq 1} z_i \mathbb{1}_{[0,t]}(T_i) = bt + \sum_{i:T_i \geq t} Z_i = bt + Nf \quad f(x,z) = z\mathbb{1}_{[0,t]}(x)$$

By hypothesis:

$$\nu(f \wedge 1) = \int_{\mathbb{R}_+\times\mathbb{R}_+} \nu(dx,dz)(f \wedge 1) = t\int_{\mathbb{R}_+} \lambda(dz)(z \wedge 1) < \infty$$

So that by finiteness of p.r.m. (Prop. 14.4) we have $Nf < \infty$ *a.s.* $\implies S_t < \infty$ *a.s.*.
Independence of increments follows by N being a p.r.m. as in Proposition 15.10.
(Claim #2) the Laplace transform, given △, □ and Theorem 13.19 takes form:

$$\mathbb{E}\left[e^{-rS_t}\right] = \exp\left\{-t(br) - t\int_{\mathbb{R}_+} \lambda(dz)(1-e^{-rz})\right\}$$

(Claim #1)(◯ stationarity Def. 15.14#2) using the just derived Laplace transform:

$$\begin{aligned}
\mathbb{E}\left[e^{-r(S_{u+t}-S_u)}\right] &= \mathbb{E}\left[e^{-r(\alpha((u,u+t])-Nf)}\right] && f(x,z) = z\mathbb{1}_{(u,u+t]}(x) \\
&= e^{-r\alpha((u,u+t])}\mathbb{E}\left[e^{-rNf}\right] \\
&= \exp\left\{-r(b)(u+t-u)\right\}\exp\left\{-\int_{(u,u+t]\times\mathbb{R}_+}(1-e^{-rz})\lambda(dz)\right\} \\
&= \exp\left\{-rbt - t\int_{\mathbb{R}_+} 1-e^{-rz}\lambda(dz)\right\} \\
&= \exp\left\{-t\left(rb + \int_{\mathbb{R}_+}(1-e^{-rz})\lambda(dz)\right)\right\} \\
&= \mathbb{E}\left[e^{-rS_t}\right] && \text{Claim \#2}
\end{aligned}$$

So that $(S_t)_{t\in\mathbb{R}_+}$ is an increasing Lévy process. □

♢ **Observation 15.18** (Usefulness of the result)**.** *By proving that Poisson compound processes are increasing Lévy processes we are certain that the latter exists. The question now becomes if this form is the only one, and so any Lévy process is Poisson compound, or there are other forms.*

♠ **Definition 15.19** (Lévy process terminology)**.** *We say that:*

- $b \in \mathbb{R}_+$ *is the drift*
- λ *is the Lévy measure*

Where the two of them uniquely identify S via the Laplace transform of Proposition 15.17#2.

Lemma 15.20 (Finite measures Lévyness)**.** *It is rather easy to check for $E \subset \mathbb{R}_+$ that:*

$$\lambda : \lambda(E) < \infty \implies \lambda(z \wedge 1) < \infty \quad \textit{can apply Prop. 15.17}$$

Proof. Trivial. □

15.1 Stable and Gamma Processes

♢ **Observation 15.21** (Interesting cases for further results)**.** *Until the end of the Chapter, set $b = 0$ and consider measures λ not finite. We will establish a connection with the claims of Corollary 14.11#2,#3 through examples and generalized results. The Poisson compound process we consider has form:*

$$S_t = \int_{\mathbb{R}_+\times\mathbb{R}_+} f(x,z)N(dx,dz) \quad N \sim \mathcal{P}ois(Leb\times\lambda),\ f(x,z) = \mathbb{1}_{[0,t]}(x)\,z$$

Namely, the c constant in the Lebesgue measure is ignored. It is a simplified version of the candidate of Definition 15.16.

♡ **Example 15.22** (Gamma process)**.** *Consider the (soon to be) Lévy measure:*

$$\lambda(dz) = a\frac{e^{-cz}}{z}dz \quad z \in \mathbb{R}_+, a \in (0,1), c > 0$$

We call the arising compound Poisson process (Def. 15.5) $S = (S_t)_{t\in\mathbb{R}_+}$ a Gamma process, and aim to show that it is also an increasing Lévy process (Def. 15.14) with the construction just explained, according to the setting of Observation 15.21.
(△ integrability) *we want to show that $\int \lambda(dz)(z\wedge 1) < \infty$. This holds since:*

- $\int_1^\infty \lambda(dz)(z\wedge 1) = \int_1^\infty \lambda(dz) = \int_1^\infty a\frac{e^{-cz}}{z}dz \to 0$ *as $z \to \infty$ sufficiently fast (we take this for granted)*
- $\int_0^1 \lambda(dz)(z\wedge 1) = \int_0^1 \lambda(dz)z = \int_0^1 a\frac{e^{-cz}}{z}zdz = \int_0^1 ae^{-cz}dz < \infty$

Given that the condition of Proposition 15.17 is satisfied, we conclude that S is an increasing Lévy process.
(□ why Gamma?) *wts $(S_t)_{t\in\mathbb{R}_+}$ is such that $S_t \overset{d}{=} X_t \sim Gamma(at, c) \quad \forall t$*
We do this by using the Laplace functional. We recall that a Gamma distribution is such that:

$$\widehat{\mathcal{P}}_{X_t}(r) = \left(\frac{c}{r+c}\right)^{at} \tag{15.1}$$

For the Gamma process at a fixed $t \in \mathbb{R}_+$:

$$\begin{aligned}
\mathbb{E}[e^{-rS_t}] &= exp\left\{-t\int_0^\infty (1-e^{-rz})\lambda(dz)\right\} && \textit{Prop. 15.17\#2}\\
&= exp\left\{-t\int_0^\infty (1-e^{-rz})a\frac{e^{-cz}}{z}dz\right\}\\
&= exp\left\{-at\int_0^\infty \frac{e^{-cz}-e^{-(c+r)z}}{z}dz\right\}
\end{aligned}$$

(○ blue integral) *We focus on the highlighted part for a moment and observe that the inside can be seen as the integral in dt:*

$$\begin{aligned}
\int_0^\infty \frac{e^{-cz}-e^{-(c+r)z}}{z}dz &= \int_0^\infty \left.\frac{-e^{-tz}}{z}\right|_{t=c}^{c+r} dz\\
&= \int_0^\infty \int_c^{c+r} \frac{-d}{dt}\frac{e^{-tz}}{z}dtdz\\
&= \int_0^\infty \int_c^{c+r} e^{-tz}dtdz && \textit{deriving}\\
&= \int_c^{c+r} \int_0^\infty e^{-tz}dzdt && \textit{Fubini Thm. B.30}\\
&= \int_c^{c+r} \left.\frac{-e^{-tz}}{t}\right|_{z=0}^{\infty}\\
&= \int_c^{c+r} \frac{1}{t}dt\\
&= \log t\Big|_{t=c}^{c+r}\\
&= \log\left(\frac{c+r}{c}\right)
\end{aligned}$$

(♢ back to Laplace) *we plut the result of ○ into □ and conclude that:*

$$\begin{aligned}
\mathbb{E}[e^{-rS_t}] &= \exp\left\{-at\log\left(\frac{c+r}{c}\right)\right\}\\
&= \exp\left\{\log\left(\frac{c+r}{c}\right)^{-at}\right\}\\
&= \left(\frac{c}{c+r}\right)^{at}
\end{aligned}$$

Which is equal to the Laplace transfom of $X_t \sim Gamma(at, c)$. By Theorem 6.12 this means that the two variables are equivalent. This holds $\forall t \in \mathbb{R}_+$.

♡ **Example 15.23** (Stable process of index α). *Consider the (soon to be) Lévy measure:*

$$\lambda(dz) = \frac{1}{\Gamma(1-a)}acz^{-1-a}dz \quad z\in\mathbb{R}_+, a\in(0,1), c>0$$

We will see that the arising process $S = (S_t)_{t\in\mathbb{R}_+}$ with the form of Observation 15.21 is an increasing Lévy process (Def. 15.14) and has some nice properties.

*(△ **integrability**) we aim to show that the integrability condition for being increasing Lévy holds. For this purpose, notice that:*

$$\Gamma(x) := \int_0^\infty t^{x-1}e^{-t}dt \quad \Gamma(n) = (n-1)!\,\forall n \in \mathbb{N}$$

So that:

- $\int_1^\infty \lambda(dz)(z \wedge 1) = \int_1^\infty \lambda(dz) = \frac{1}{\Gamma(1-a)}ac\int_1^\infty z^{-1-a}dz = \frac{1}{\Gamma(1-a)}ac\frac{1}{a} - z^{-a}\Big|_{z=1}^{\infty} = \frac{1}{\Gamma(1-a)}c\frac{1}{a} < \infty$
- *by* $a \in (0,1)$

$$\int_0^1 \lambda(dz)(z \wedge 1) = \frac{1}{\Gamma(1-a)}ac\int_0^1 z^{-1-a}z = \frac{1}{\Gamma(1-a)}ac\int_0^1 z^{-a}$$
$$= \frac{1}{\Gamma(1-a)}ac\frac{1}{1-a}z^{1-a}\Big|_{z=0}^{1} = \frac{1}{\Gamma(1-a)}\frac{ac}{1-a} < \infty$$

Making their sum finite. By $\int_0^\infty \lambda(dz)(z \wedge 1) < \infty$ *we can apply Proposition 15.17#1 and conclude that* S *is an increasing Lévy process.*

♢ **Observation 15.24** (About the stable process)**.** *Notice that even though* $S_t < \infty$ *a.s. the process has no expectation. Infact:*

$$\begin{aligned}
\mathbb{E}[S_t] &= \mathbb{E}\left[\int_{(0,t]\times\mathbb{R}_+} z\, N(dx,dz)\right] && \textit{Obs. 15.21, } b=0\\
&= \mathbb{E}\left[Nf\right] && f(x,z) := \mathbb{1}_{[0,t]}(x)\, z\\
&= \nu f && \textit{Def. 13.6}\\
&= \int f\nu(dx)\\
&= \int \mathbb{1}_{[0,t]}(x)\, z dx\lambda(dz)\\
&= \int_0^t\int_0^\infty z\lambda(dz)dx\\
&= t\int_0^\infty z\lambda(dz)\\
&= t\int_0^\infty z\frac{1}{\Gamma(1-\alpha)}acz^{-1-a}dz\\
&= \frac{tca}{\Gamma(1-\alpha)}\int_0^\infty z^{-a}dz && \textit{improper integral}
\end{aligned}$$

Where the improper integral diverges at ∞*, and* S_t *has no expectation. For the sake of completeness, we report the calculation here below. An improper integral of this form can be calculated considering the discontinuity at zero and the divergent limit on the other side:*

$$\int_0^\infty z^{-a}dz = \int_0^1 z^{-a}dz + \int_1^\infty z^{-a}dz = \lim_{b\to 0}\int_b^1 z^{-a}dz + \lim_{c\to\infty}\int_1^c z^{-a}dz$$

while the indefinite integral is easily found as $\frac{1}{1-a}z^{-a+1} + K, K \in \mathbb{R}$*. Ignoring the constant which is positive since* $a \in (0,1)$ *by construction and* $1-a > 0$*, we get:*

$$\lim_{b\to 0} z^{-a+1}\Big|_{z=b}^{1} = 1 - \lim_{b\to 0} b^{1-a} = 1 < \infty$$

But

$$\lim_{c\to\infty} z^{1-a}\Big|_{z=1}^{c} = \lim_{c\to\infty} c^{1-a} - 1 = \infty$$

and the sum diverges. All the constants are positive and the claim is proved.

♣ **Proposition 15.25** (Link integrability & infinite activity). *Consider a measure on $\mathbb{R}_+$ which is not finite. We link Proposition 15.17#1 and Corollary 14.11#2,#3 by:*

$$\int_0^\infty (z \wedge 1)\lambda(dz) < \infty \implies \lambda((\epsilon,\infty)) < \infty \quad \forall \epsilon > 0$$
$$\textit{but still} \quad \lambda((\epsilon,\infty)) \overset{\epsilon\to 0}{\to} \infty$$

Proof. ($\triangle$ $\epsilon > 1$) it holds:

$$\lambda((\epsilon,\infty)) < \lambda((1,\infty)) = \int_1^\infty \lambda(dz) \leq \int_0^\infty (z \wedge 1)\lambda(dz) < \infty$$

($\square$ $\epsilon \leq 1$) a little more elaborate since:

$$\lambda((\epsilon,\infty)) = \int_\epsilon^1 \lambda(dz) + \int_1^\infty \lambda(dz) \qquad := (I_1 + I_2)$$

Where $I_2 < \infty$ by $\triangle$.
($\bigcirc$ $I_1 < \infty$) we have:

$$\begin{aligned}
\int_\epsilon^1 \lambda(dz) &\leq \int_\epsilon^1 \frac{z}{\epsilon}\lambda(dz) && \epsilon < z < 1 \implies \frac{z}{\epsilon} > 1 \\
&\leq \frac{1}{\epsilon}\int_0^1 z\lambda(dz) && z = z \wedge 1 \quad \forall z \in [0,1] \\
&< \infty && \textit{by } \triangle
\end{aligned}$$

(♠ $\epsilon \to 0$) as $\epsilon \to 0$ all the results hold but asymptotically as claimed. Since the measure is on $\mathbb{R}_+$ and it is not finite, we necessarily have $\lambda((\epsilon,\infty)) \overset{\epsilon\to 0}{\to} \infty$ $\square$

♢ **Observation 15.26** (A side result). *Let λ be not finite. Since a compound Poisson process is increasing Lévy if it satisfies the integrability condition, it will also be the case that it has infinite activity.*

♡ **Example 15.27** (The stability of the stable process from Example 15.23). *$S = (S_t)_{t\in\mathbb{R}_+}$ from Example 15.23 is stable in the sense that:*

$$S_{ut} \overset{d}{=} u^{\frac{1}{a}} S_t \quad \forall u,t \in \mathbb{R}_+ \quad i.e. \quad S_t \overset{d}{=} t^{\frac{1}{a}} S_1 \quad \forall t \in \mathbb{R}_+$$

*($\triangle$ **Laplace approach**) use the Laplace transform from Proposition 15.17#2.*

$$\begin{aligned}
\mathbb{E}\left[e^{-rS_t}\right] &= \exp\left\{-t\int_0^\infty (1-e^{-rz})\lambda(dz)\right\} \\
&= \exp\left\{-t\int_0^\infty (1-e^{-rz})\frac{ac}{\Gamma(1-a)}z^{-1-a}dz\right\} \\
&= \exp\{-tcr^a\} && \textit{proved below in } \square
\end{aligned}$$

Where the last equality is $\int_0^\infty (1-e^{-rz})az^{-1-a}dz = r^a\Gamma(1-a)$.
($\square$ ***missing equality***) *by direct computation:*

$$\begin{aligned}
\int_0^\infty (1-e^{-rz})az^{-1-a}dz &= \int_0^\infty (1-e^{-t})a\left(\frac{t}{r}\right)^{-1-a}\frac{dt}{r} && t=rz \quad dt=rdz\\
&= r^a\int_0^\infty (1-e^{-t})at^{-1-a}dt\\
&= -r^a\int_0^\infty \underbrace{(1-e^{-t})}_{g}\underbrace{(-at^{-1-a})}_{f'}dt && \text{integrate by parts}\\
&= -r^a\left(\underbrace{(1-e^{-t})t^{-a}\Big|_0^\infty}_{=0} - \int_0^\infty e^{-t}t^{-a}dt\right) && t^{-a}=t^{1-a-1}\\
&= -r^a\left(-\int_0^\infty e^{-t}t^{1-a-1}dt\right) && \text{Gamma integral at } 1-a\\
&= r^a\Gamma(1-a)
\end{aligned}$$

($\spadesuit$ ***back to Laplace***) *by* $\triangle$ *the general form at time ut is:*

$$\begin{aligned}
\mathbb{E}\left[e^{-rS_{ut}}\right] &= \exp\{-utcr^a\}\\
&= \exp\left\{-ct\left(u^{\frac{1}{a}}r\right)^a\right\}\\
&= \mathbb{E}\left[e^{-u^{\frac{1}{a}}rS_t}\right]\\
&= \mathbb{E}\left[e^{-r(u^{\frac{1}{a}}S_t)}\right]\\
\implies S_{ut} &\overset{d}{=} u^{\frac{1}{a}}S_t && \forall u,t
\end{aligned}$$

$\heartsuit$ **Example 15.28** (Wiener process is stable)**.** *We showed that a Wiener process* $W=(W_t)_{t\in\mathbb{R}_+}$ *(Def. 11.55) is such that:*

$$W_t = \sqrt{t}W_1, W_t \sim \mathcal{N}(0,t), W_1 \sim \mathcal{N}(0,1)$$

which is the result of Proposition 11.56#3 for $s=0$.
This is equivalent to saying that the process is stable as that of Example 15.23.

$\spadesuit$ **Definition 15.29** (Inverse Gaussian distribution)**.** *We consider a stable process* $(S_t)_{t\in\mathbb{R}_+}$ *as in Example 15.23 with* $a=\frac{1}{2}, c=\sqrt{2}$. *The Lévy measure becomes:*

$$\lambda(dz) = \frac{1}{\sqrt{2\pi z^3}}dz$$

The density associated to such measure is available in closed form:

$$f(z) = \frac{t}{\sqrt{2\pi z^3}}e^{-\frac{t^2}{2z}} \quad z\in\mathbb{R}_+$$

We know that this is the density function of an inverse gaussian distribution, so we can safely say that $S_t \overset{d}{=} \frac{a^2}{Z^2}$ *for* $Z\sim\mathcal{N}(0,1)$ *and write* $S_t\sim\mathcal{IN}(a)$.

$\diamondsuit$ **Observation 15.30** (Simulating infinite activity measures)**.** *We know that:*

$$\begin{aligned}
S_t &= \int_{[0,t]\times\mathbb{R}_+} zN(dx,dz) && \text{mean } dx\lambda(dz)\\
&= \sum_{i=1}^\infty Z_i\mathbb{1}_{[0,t]}(T_i) && \text{jumps of size } Z_i
\end{aligned}$$

However, both a Gamma process (Ex. 15.22) and a stable process (Ex. 15.23) have ∞ *activity since* $\lambda((\epsilon,\infty)) \overset{\epsilon\to 0}{\to} \infty$. *We cannot apply Proposition 14.17 directly since the premises of Def. 14.16 are not respected. We need some form of* ***truncation***. *For this purpose, we will reconsider:*

- $\widetilde{N}(dx, du)$ *on* $\mathbb{R}_+ \times \mathbb{R}_+$ *with mean* $dxdu$
- $N(dx, dz)$ *simulated via transform using Proposition 14.20 assuming that it has an intensity notion (Def. 14.19)*

we will do so for the stable process of Example 15.23.

♠ **Definition 15.31** (Generalized inverse j)**.** *To implement the truncation, we make use of:*

$$j(u) = \inf \{\epsilon > 0 \ : \ \lambda((\epsilon, \infty)) < u\}$$

Where the inf *accounts for possible discontinuities. Notice that* j *is decreasing since* λ *is decreasing in* ϵ*.*

♣ **Proposition 15.32** (Generalized inverse properties)**.** *We have that:*

1. $\lambda(A) = (Leb \circ j^{-1})(A) = Leb(\mathbb{1}_A \circ j) \quad \forall A \in \mathcal{E}$
2. $\lambda(f) = Leb(f \circ j)$
3. $S_t = \sum_{i=1}^{\infty} j(U_i)\mathbb{1}_{[0,t]}(T_i) = \sum_{i:T_i \le t} j(U_i)$ *for* $((U_i, T_i))_{i\ge1} \sim N(dx, du)$

Proof. **(Claim #1)** wlog for $A = (a, b]$ observe that:

$$\begin{aligned}
\lambda(A) = \lambda((a, b]) &= \lambda((a, \infty)) - \lambda([b, \infty)) \\
&= Leb\left(\{u : u \in (\lambda([b, \infty)), \lambda((a, \infty))]\}\right) \\
&= Leb\left(\{u : j(u) \in A\}\right) \\
&= Leb\left(j^{-1}(A)\right) \\
&= Leb(\mathbb{1}_A \circ j) && j^{-1}(A) = \mathbb{1}_A \circ j
\end{aligned}$$

(Claim #2) more in general:

$$\lambda(f) = \int f(z)\lambda(dz) = \int f(j(u))du = Leb(f \circ j)$$

(Claim #3) The integrability condition becomes:

$$\int_{\mathbb{R}_+} (j(u) \wedge 1)du < \infty$$

and if $\widetilde{N}$ is a p.r.m. with mean $Leb \times Leb$ on $\mathbb{R}_+ \times \mathbb{R}_+$ with atoms (T_i, U_i) then $(T_i, j(U_i))$ are the atoms of N and we recover S_t via:

$$S_t = \sum_{i\ge1} j(U_i)\mathbb{1}_{[0,t]}(T_i) = \sum_{i:T_i\le t} j(U_i)$$

□

♢ **Observation 15.33** (About the results)**.** *More comments can be made:*

- $j(\cdot)$ *decreasing* $\implies (U_i)_{i\ge1}$ *arrival times are in increasing order*
- S_t *has decreasing increments*
- *it will be shown in Theorem 16.2 that* $(U_i)_{i\ge1} \overset{d}{=} (G_i)_{i\ge1}$ *where:*

$$G_1 = E_1, \dots, G_i = \sum_{k=1}^{i} E_k \quad E_k \sim \mathcal{E}xp(1), G_i = \mathcal{G}amma(i, 1)$$

- $S_1 = \sum_{i=1}^{\infty} j(G_i)$ *but for* $t \ne 1$ *it holds* $S_t = \sum_{i=1}^{\infty} j(\frac{1}{t}G_i)$

For these reasons, we need an appropriate j *generalized inverse. We will prove such results as a continuation of Examples 15.22, 15.23.*

♡ **Example 15.34** (Stable process j, Ex. 15.23 ctd.)**.** *Remember that* $\lambda(dz) = \frac{ac}{\Gamma(1-a)} z^{-1-a}dz$ *and:*

$$\lambda((\epsilon, \infty)) = \frac{c}{\Gamma(1-a)}\epsilon^{-a}$$

Using Definition 15.31 for j we have that the solution in ϵ to the infimization is:

$$j(u)\ :\ \lambda((\epsilon,\infty)) = u \implies j(u) = \left(\frac{c}{\Gamma(1-a)}\right)^{\frac{1}{a}} u^{-\frac{1}{a}}$$

Using Observation 15.33 we can safely say that by the exponential distribution of the $G_i \sim \mathcal{E}xp(1)$ with $\frac{1}{t}\sum G_i \sim Exp(1)$ it is the case that:

$$\begin{cases} S_t = \sum_{i:T_i\le t} \widehat{c}(U_i)^{-\frac{1}{a}} = \sum_{i=1}^{\infty} \widehat{c}\left(\frac{1}{t}G_i\right)^{-\frac{1}{a}} \\ \widehat{c} = \left(\frac{c}{\Gamma(1-a)}\right)^{\frac{1}{a}} \end{cases}$$

If we take $t = 1$, (U_i) forms a p.r.m. with unit intensity. In particular the arrival times of the allied counting process (that is U_i in increasing order) are equal in distribution to:

$$G_1 = E_1,\ G_2 = E_1 + E_2, \dots, G_k = E_1 + \cdots E_k$$

Where (E_i) are exponential iid of unit rate. Hence:

$$S_1 = \sum_{i=1}^{\infty} \widehat{c} G_i^{-\frac{1}{a}}$$

while in general:

$$S_t = \sum_{i=1}^{\infty} \widehat{c}\left(\frac{1}{t}G_i\right)^{-\frac{1}{a}} \qquad \frac{1}{t}G_i \overset{iid}{\sim} \mathcal{E}xp(1)$$

♡ **Example 15.35** (Gamma process j approximation, Ex. 15.22 ctd)**.** *For a Lévy density as that of the gamma process the integral:*

$$\lambda((\epsilon,\infty)) = \int_\epsilon^\infty a\frac{e^{-cz}}{z}dz$$

is not available in closed form. To simulate from it, we resort to the notion of incomplete Gamma function (Def. 15.36) and the result of Lemma 15.37. Indeed:

$$\Gamma(0,x) = \gamma_0(x) = \int_x^\infty u^{-1}e^{-u}du \quad \Gamma_1(x) \overset{x\to 0}{\to} \infty$$

And we can express the Lévy measure as:

$$\begin{aligned} \lambda((\epsilon,\infty)) &= \int_\epsilon^\infty a\frac{e^{-cz}}{z}dz \\ &= \int_{c\epsilon}^\infty a\frac{e^{-x}}{xc}cdz && \textit{let } x = cz, dx = cdz \\ &= \int_{c\epsilon}^\infty a e^{-x}x^{-1}dx \\ &= a\gamma_0(c\epsilon) \end{aligned}$$

So that the following chain holds:

$$a\gamma_0(c\epsilon) = u \iff c\epsilon = \gamma_0^{-1}\left(\frac{u}{a}\right) \iff j(u) = \frac{1}{c}\gamma_0^{-1}\left(\frac{u}{a}\right)$$

And eventually:

$$\begin{aligned} S_t &= \sum_{i=1}^{\infty} j\left(\frac{1}{t}G_i\right) && \textit{Obs. 15.33} \\ &= \sum_{i=1}^{\infty} \frac{1}{c}\gamma_0^{-1}\left(\frac{G_i}{at}\right) \end{aligned}$$

And we know how to approximate the inverse of the incomplete Gamma function (Lem. 15.37).

♠ **Definition 15.36** (Incomplete Gamma function $\Gamma(s,x)$). *Also known as upper incomplete gamma function:*

$$\gamma_s(x) = \Gamma(s,x) = \int_x^\infty t^{s-1}e^{-t}dt$$

Where for $s=0$ *we see that* $\Gamma(0,x) = \gamma_0(x) \overset{x\to 0}{\to} \infty$.

Lemma 15.37 (Incomplete gamma-χ^2 link). *Let* $\chi^2_{d,(qt)} :=$ *upper quantile of the chi-square distribution such that* $\mathbb{P}[\chi^2_d > \chi^2_{d,(\alpha)}] = \alpha$. *Then:*

1. $\gamma_0(u) = \Gamma(0,u) \overset{d\to 0}{\approx} \frac{1}{2}\chi^2_{d,(\frac{du}{2})}$
2. *Accordingly:*

$$S_t \overset{d\to 0}{\approx} \sum_{i=1}^\infty \frac{1}{c}\frac{1}{2}\chi^2_{d,(\frac{d}{2}\frac{G_i}{at})}$$

Proof. **(Claim #1)(△ a basic fact)** notice that $\chi^2_d = \mathcal{G}amma\left(\frac{d}{2},\frac{1}{2}\right)$ with density:

$$f(x) = \frac{1}{\Gamma\left(\frac{d}{2}\right)}\left(\frac{1}{2}\right)^{\frac{d}{2}} x^{\frac{d}{2}-1}e^{-\frac{x}{2}}$$

(□ approximating) we extract the inequality:

$$\begin{aligned}
\frac{2}{d}\mathbb{P}[\chi^2_d > 2x] &= \frac{2}{d}\int_{2x}^\infty \frac{1}{\frac{d}{2}\Gamma\left(\frac{d}{2}\right)}\left(\frac{1}{2}\right)^{\frac{d}{2}} u^{\frac{d}{2}-1}e^{-\frac{u}{2}}du \\
&= \int_{2x}^\infty \frac{1}{\Gamma\left(\frac{d}{2}\right)}\left(\frac{1}{2}\right)^{\frac{d}{2}} u^{\frac{d}{2}-1}e^{-\frac{u}{2}}du \\
&\overset{d\to 0}{\approx} \int_{2x}^\infty \left(\frac{1}{2}\right)^{\frac{d}{2}} u^{\frac{d}{2}-1}e^{-\frac{u}{2}}du && a\Gamma(a) \overset{a\to 0}{\to} 1 \\
&= \int_x^\infty y^{-1}e^{-y}dy && \text{ch. var } y = \frac{u}{2} \\
&= \gamma_0(x) && \text{Def. 15.36}
\end{aligned}$$

So that the expression tends to the law of the incomplete Gamma function as $d \to 0$.
(◯ inverting) we invert the relation to derive a generalized inverse in the sense of Definition 15.31.

$$\frac{2}{d}\mathbb{P}[\chi^2_d > 2x] = u \iff \mathbb{P}[\chi^2_d > 2x] = u\frac{d}{2} \iff \chi^2_{d,(\frac{d}{2}u)} = 2x \iff x = \frac{1}{2}\chi^2_{d,(\frac{d}{2}u)}$$

which means that $E_1^{-1}(u) \overset{d\to 0}{\approx} \frac{1}{2}\chi^2_{d,(\frac{d}{2}u)}$.
(Claim #2) we just use the result of Example 17.12 to conclude that:

$$S_t = \sum_{i=1}^\infty \frac{1}{c}\frac{1}{2}\chi^2_{d,\left(\frac{dG_i}{2at}\right)} \overset{d\to 0}{\approx} \sum_{i=1}^\infty \frac{1}{c}E_1^{-1}\left(\frac{G_i}{at}\right)$$

□

Chapter Summary

Objects:

- we consider the arising Poisson compound process from a Poisson random measure (T, Y) where $(T_n)_{n\in\mathbb{N}}$ are arrival time atoms of a p.r.m. on $\mathbb{R}_+$ with mean $cdx = \nu(dx)$ and $Y_i \overset{iid}{\sim} \pi, Y \perp\!\!\!\perp T$

$$S_t = \sum_{i:T_i\leq t} Y_i = \int_{\mathbb{R}_+\times\mathbb{R}_+} y\mathbb{1}_{[0,t]}(T_i)Y_i = L((0,t]) \quad L(A) = \int_{A\times\mathbb{R}} yN(dx,dy)$$

- additive random measures are random measures with independencies of random variables over disjoint sets
- increasing Lévy process is:
  - increasing, right continuous, $S_0 = 0$
  - indepedendent increments
  - stationary increments
- the candidate Poisson compound process with underlying additive random measure to be increasing Lévy is:

$$S_t = bt + \int_{[0,t]\times\mathbb{R}_+} zN(dx,dz) = M(\omega,[0,t]), \quad N \sim \mathcal{P}(cdx \times \lambda(dz))$$

  where M is an additive random measure, and N is a Poisson random measure
- we introduce the drift and Lévy measure of a precise form of increasing Lévy processes
- the inverse Gaussian distribution has measure & denstity wrt Lebesgue:

$$\lambda(dz) = \frac{1}{\Gamma(1-a)}acz^{-1-a}dz = \frac{1}{\sqrt{2\pi z^3}}dz \quad \text{for } \alpha = \frac{1}{2}, c = \sqrt{2}, \quad f(z) = \frac{t}{\sqrt{2\pi z^3}}e^{-\frac{t^2}{2z}}$$

  where the density is that of $\frac{a^2}{Z^2} \sim \mathcal{IN}(a)$
- to simulate infinite activity processes, we need the generalized inverse notion:

$$j(u) = \inf\{\epsilon > 0 : \lambda((\epsilon,\infty)) < u\}$$

Results:

- for $X \sim \mathcal{P}ois(\nu)$ on $(E,\mathcal{E})$ and $Y \perp\!\!\!\perp X$ such that $Y \sim \pi$ on $(F,\mathcal{F})$ it holds:

$$(X,Y) \sim \mathcal{P}ois(\mu), \quad \mu = \nu\times\pi, \quad (X,Y) : (E\times F, \mathcal{E}\otimes\mathcal{F}) \to \mathbb{R}_+$$

- the compound Poisson process S_t has an underlying additive measure L
- any additive random measure is decomposed as $M = \alpha + K + L$, where α is deterministic and K is random with fixed positions
- the candidate compound Poisson process with mean $Leb\times\lambda$ is Lévy increasing if the Lévy measure:

$$\lambda(f\wedge 1) = \int_{\mathbb{R}_+}(f\wedge 1)\lambda(dz) < \infty$$

- the integrability condition implies infinite activity, namely $\lambda((\epsilon,\infty)) < \infty$ for all $\epsilon > 0$
- Simulation of infinite activity process:
  - aim: for a compound Poisson process with weak integrability (i.e. an increasing Lévy process), we have infinite activity, and cannot fall under the arrival process formalism
  - setting: λ with weak integrability for a p.r.m. $N \sim \mathcal{P}ois(Leb\times\lambda)$, $\widetilde{N}(dx,du) \sim \mathcal{P}ois(Leb\times Leb))$,
  - look for j in closed form assuming that $U_i = G_i \sim \mathcal{E}xp(1)$ to conclude
  - then use the intensity simulation of before
  - the generalized inverse of an infinite activity process is such that $\lambda(A) = Leb(\mathbb{1}_A\circ j), \lambda(f) = Leb(f\circ j), S_t = \sum_{i\geq 1} j(U_i)$

Chapter 16

Poisson Processes

♢ **Observation 16.1** (Setting & previous results)**.** *We briefly recall some facts that will be useful in this Chapter.*

- *A Poisson process* $N = (N_t)_{t\in\mathbb{R}_+} \sim \mathcal{P}ois(c) \quad c > 0$ *(Def. 12.2) is such that:*
  1. *N is adapted to a filtration* $\mathcal{F}$ *(Def. 11.7)*
  2. $\mathbb{E}_s[f(N_{s+t} - N_s)] = \sum_{k=0}^{\infty} f(k)\frac{e^{-ct}(ct)^k}{k!}$
- *N can be seen as a counting process (Def. 11.13) by definition with map* $t \to N_t$*:*
  1. *starting at zero* $N_0 = 0$
  2. *increasing, right continuous and with jumps of size* 1
- *the representation of a counting process for* $0 < T_1 < \dots$; $\lim_{k\to\infty} T_k = \infty$ *arrival times is:*

$$N_t = \sum_{k=1}^{\infty} \mathbb{1}_{[0,t]}(T_k) \quad t \in \mathbb{R}_+$$

- *a filtration of the form* $(\mathcal{F}_t)_{t\in\mathbb{R}_+} = \sigma(\{N_s, s \leq t\}) = \sigma(\{M(A) : A \in \mathcal{B}([0,t])\}$*, where* $N_t(\omega) = M(\omega, [0,t])$ *makes* T_k *a stopping time* $\forall k$ *in the sense of Definition 11.9, by the result of Example 11.27 since:*

$$\{T_k \leq t\} = \{N_t \geq k\} \quad \forall k$$

- *by Theorem 12.4 we know that* $(N_t - ct)_{t\geq 0}$ *is a martingale, in the sense of Definition 11.35*
- *ultimately, we observe that* $(T_k)_{k\geq 1}$ *identifies a random counting measure (Def. 13.3) on* $\mathbb{R}_+$ *which we call* $M(dx)$ *and for* $f \in \mathcal{E}_+$ *extended to* ∞ *with* $f(\infty) = 0$ *we write:*

$$Mf = \sum_{k=1}^{\infty} f(T_k)$$

- *such* $M(dx)$ *is a p.r.m. (Def. 13.13) with mean* $\nu(dx) = cdx$ *allowing us to write:*

$$N_t = M((0,t])$$

Formal conclusions in this structure are made in the following Theorem.

♣ **Theorem 16.2** (Poisson process, random measure & counting process equivalence)**.** *Let* $c > 0$*, TFAE:*

1. *M is a p.r.m. (Def. 13.13) with mean* $\mu = cLeb$
2. *N is a poisson (counting) process (Defs. 12.2, 11.13) with rate c*
3. *N is a counting process (Def. 11.13) and* $\widetilde{N} = (N_t - ct)_{t\geq 0}$ *is an* $\mathcal{F}$*-martingale (Def. 11.35)*
4. $(T_k)_{k\geq 1}$ *is an increasing sequence of* $\mathcal{F}$*-stopping times (Def. 11.9) and:*

$$T_1, T_2 - T_1, \dots \overset{iid}{\sim} \mathcal{E}xp(c)$$

Proof. **(△ strategy)** we show the chain of implications:

$$① \implies ② \iff ③ \implies ④ \implies ①$$

((1) $\implies$ (2)) by the Lebesgue measure being diffuse we can apply Theorem 14.10 to conclude that M is a Poisson counting measure with mean $\mu([0,t]) = ct < \infty$, so that N is a counting process. The independence and the Poisson distribution requirements of Definition 12.2 follow by the fact that M is not only diffuse but also a Poisson random measure.
((2) $\iff$ (3)) this is Theorem 12.4
((3) $\implies$ (4)) assume #2, #3 holds and recall the discussion of the above observation.
(△ stopping) for all $k \geq 1$ it holds $\{T_k \leq t\} = \{N_t \geq k\} \in \mathcal{F}_t \forall t$ by the filtration we considered. Then T_k is always a stopping time in the sense of Definition 11.9.
(□ domain check) by N being a counting process, it holds $\lim_{t\to\infty} N_t = M(\mathbb{R}_+) = \infty$ almost surely with ordered arrival times:

$$0 < T_1(\omega) < \cdots < T_k(\omega) \quad if\ T_k(\omega) < \infty, \qquad \lim_{k\to\infty} T_k = \infty$$

(○ stochastic integrals approach) we want to show:

$$\mathbb{E}_{T_k}\left[\exp\{-r(T_{k+1} - T_k)\right] = \frac{c}{c+r} \quad \forall r \in \mathbb{R}_+ \tag{16.1}$$

the Steltjes integral of the martingale $\widetilde{N}$ in this case is, by Corollary C.25:

$$\mathbb{E}_S \int_{(S,T]} F_t dN_t = \mathbb{E}_S \int_{(S,T]} F_t c dt$$

Let $T_k = S, T_{k+1} = T, S \leq T$ almost surely for fixed $k \in \mathbb{N}$. We have:

$$\begin{cases} N_t = K & S < t < T \implies dN_t = 0 \\ N_t = k+1 & t = T \implies dN_t = 1 \end{cases}$$

choosing $F_t = re^{-rt}$ with fixed $r > 0$ we get:

$$LHS = r\mathbb{E}_S\left[e^{-rT}\right] = c\mathbb{E}_S\left[e^{-rS} - e^{-rT}\right] = RHS$$

reordering gives us the equivalent claim we eventually want to show:

$$\mathbb{E}_S[e^{-r(T-S)}] = \frac{c}{c+r} \iff r\mathbb{E}_S[e^{-r(T-S)}] = c - c\mathbb{E}_S[e^{-r(T-S)}] \tag{16.2}$$

(▽ big computation) the result is as follows:

$$\begin{aligned}
r\mathbb{E}_S\left[e^{-r(T-S)}\right] &= r\mathbb{E}_S\left[e^{-rT}\right]\mathbb{E}_S\left[e^{rS)}\right] \\
&= e^{rS}\mathbb{E}_S\left[re^{-rT}\right] && \text{conditional determinism} \\
&= e^{rS}\mathbb{E}_S\left[\int_{(S,T]} F_t dN_t\right] && \square, \bigcirc \\
&= e^{rS}\mathbb{E}_S\left[\underbrace{\int_{(S,T]} F_t d(N_t - ct)}_{=0} + \int_{(S,T]} F_t c dt\right] && \text{Thm. 12.18\#2} \\
&= e^{rS}\mathbb{E}_S\left[\int_{(S,T]} F_t c dt\right] \\
&= e^{rS}\mathbb{E}_S\left[\int_{(S,T]} re^{-rt}\right] \\
&= c\mathbb{E}_S\left[c(e^{-rS} - e^{-rT})\right] \\
&= c\mathbb{E}_S\left[1 - e^{-r(T-S)}\right] \\
&= c - c\mathbb{E}_S\left[e^{-r(T-S)}\right]
\end{aligned}$$

which means that $T - S = T_{k+1} - T_k$ is independent of $\mathcal{F}_S = T_k$ and has exponential distribution with parameter c, as it was to be proved.
(#4 $\Longrightarrow$ #1) [Çin11](VII.5.5). $\square$

♣ **Theorem 16.3** (Poisson increasing Lévy characterization)**.** *For a counting process N (Def. 11.13) we conclude that:*

$$N \quad \textit{increasing Lévy (Def. 15.14)} \iff N \quad \textit{Poisson (Def. 12.2)}$$

Proof. **($\Longleftarrow$)** trivial, a Poisson process is a counting process (thus, Prop. 11.14 holds), with independent increments and stationary distribution by the definitional properties (Prop. 12.3), the two together cover the requirements of an increasing Lévy process.
($\Longrightarrow$)($\triangle$ aim) wts $N = (N_t)_{t\in\mathbb{R}_+} \sim \mathcal{P}ois$ with mean ct $\forall t$ and $c > 0$.
($\square$ first step) wts:

$$q(t) = \mathbb{P}[\{N_t = 0\}] = e^{-ct} \quad \forall t \in \mathbb{R}_+, c > 0 \; fixed$$

By N begin Lévy, it holds that $N_{s+t} - N_s \stackrel{d}{=} N_t$ and:

$$\begin{aligned} \mathbb{P}[\{N_{s+t} = 0\}] &= \mathbb{P}[\{N_s = 0, N_{s+t} - N_s = 0\}] && N_s \perp\!\!\!\perp N_{s+t} - N_s \\ &= \mathbb{P}[\{N_t = 0\}]\mathbb{P}[\{N_s = 0\}] \end{aligned}$$

So that $q(s+t) = q(t)q(s)$ decouples for any $s, t \in \mathbb{R}_+$. Also $q(0) = 1$. Lastly, $q(t) = \mathbb{E}\left[\mathbb{1}_{\{0\}} \circ N_t\right]$ is right continuous by:

- the map $t \to \mathbb{1}_{\{0\}} \circ N_t$ being a.s. right continuous (simply, N_t is a.s. right continuous)
- the bounded convergence Theorem (Cor. 4.26):

$$\begin{aligned} \lim_{t\downarrow s} q(t) &= \lim_{t\downarrow s} \mathbb{E}\left[\underbrace{\mathbb{1}_{\{0\}} \circ N_t}_{\leq 1}\right] \\ &= \mathbb{E}\left[\lim_{t\downarrow s} \mathbb{1}_{\{0\}} \circ N_t\right] \\ &= \mathbb{E}\left[\mathbb{1}_{\{0\}} \circ N_s\right] && \text{right continuity of } N \\ &= q(s) \end{aligned}$$

the solution this problem is unique for positive c:

$$c > 0, \quad \exists! q : q(s+t) = q(s)q(t), \quad q(t) = e^{-ct}$$

and is in the form claimed at the beginning of $\square$.
($\bigcirc$ corner case) for $c = 0$ $N_t = 0$ almost surely $\forall t$ and N is a Poisson process as well.
(♠ second part) we now consider M a random counting measure such that $N_t = M((0,t])$ and see that for a partition of $A = (0, t]$ into equal length subspaces $\{A_k\}_{k=1}^n$ of the form $(\cdot, \cdot]$ we have that for:

$$X_i = \mathbb{1}_{\{M(A_i)\geq 1\}} \quad S_n = X_1 + \cdots + X_n$$

by N being Lévy it holds that $\{M(A_k)\}_{k=1}^n$ is an independency (Def. 17.1) and all are iid having equal size. We denote them as $M(A_k) = N \forall k$. Then:

$$X_i \sim B(p) \quad p = 1 - q\left(\frac{t}{n}\right) = 1 - e^{\frac{-ct}{n}}$$

implying:

$$\begin{aligned} \forall k \in \mathbb{N} \quad \mathbb{P}[\{S_n = k\}] &= \frac{n!}{k!(n-k)!}\left(1 - e^{-\frac{ct}{n}}\right)^k \left(e^{-\frac{ct}{n}}\right)^{n-k} \\ &= \frac{e^{-ct}}{k!} n(n-1)\cdots(n-k+1)\left(e^{-\frac{ct}{n}} - 1\right)^k \end{aligned}$$

(♠ **third part**) for almost every ω M_ω has a finite number of atoms in A. Given $\delta(\omega)$ the minumum distance if $n > \frac{t}{\delta(\omega)}$ we will have that:

$$S_n(\omega) = M(\omega, A) = N_t(\omega) \implies S_n \overset{n\to\infty}{\to} N_t \implies \mathbb{P}[N_t = k] = \lim_{k\to\infty} \mathbb{P}[S_n = k] = \frac{e^{-ct}(ct)^k}{k!}$$

□

♢ **Observation 16.4** (Side note on the conclusion). *In Chapter 15, we introduced increasing Lévy processes and showed that a way to obtain them was via compound Poisson processes (Prop. 15.17. This ensured existance with a specific form, but not uniqueness. This result instead,* ***relates all increasing Lévy processes and Compound Poisson processes as equivalent objects under the requirement that jumps are of unitary size****.*

♢ **Observation 16.5** (Comments about the Theorem). *We inspect for $\widetilde{N} \sim \mathcal{P}ois(cdx\lambda(dz))$ on $\mathbb{R}_+ \times \mathbb{R}_+$ and:*

$$N_t = \int_{[0,t]\times\mathbb{R}_+} z\widetilde{N}(dx, dz)$$

the form of $\lambda(dz)$. Since N_t is a counting measure we derived that:

$$\begin{aligned} N_t = \sum_{i=1}^{\infty} \mathbb{1}_{[0,t]}(T_i) &= \sum_{i=1}^{\infty} 1 \cdot \mathbb{1}_{[0,t]}(T_i) \\ &= \sum_{i=1}^{\infty} Z_i \cdot \mathbb{1}_{[0,t]}(T_i) \qquad\qquad Z_i \equiv 1 \end{aligned}$$

So that $\lambda(dz) = \delta_1(dz)$ has mass 1 for all Z_i.

♣ **Proposition 16.6** (Strong Markov Property of Poisson Processes). *We establish independence of future events from the past even when the present is a stopping time.*
For a Poisson process $N \sim \mathcal{P}ois(c)$ and a stopping time S:

$$\mathbb{E}_S[f(N_{S+t} - N_S)\mathbb{1}_{\{S<\infty\}}] = \sum_{k=0}^{\infty} f(k)\frac{e^{-ct}(ct)^k}{k!}\mathbb{1}_{\{S<\infty\}}$$

Proof. By definition of Poisson Process (Def. 12.2) we extend to stopping times using Proposition C.27 replacing s with S and t with $S + t$ and proceed by Applying Corollary C.25. The result follows after some work. Note that deterministically for $t > 0$ it holds $t + S > S$, even though S is random. □

♡ **Example 16.7** (Counting process, ctd). *Consider the random time:*

$$S = \inf\{t \geq a : N_t = N_{t-a}\}$$

Where $T_k + a \quad k > 0$ is equivalent to having the first k interarrivals of size at most a and the $(k+1)^{th}$ exceeding a. It holds that $S < \infty$ almost surely, since the union over k of the events has probability one.
Let T be the next jump, and note that S falls in $a + (T - S)$. We ask the following question:

Is it true that $a \to \infty \implies T - S \to 0$?

This is ***False****. Indeed, notice that:*

$$\{T - S > t\} = \{N_{S+t} - N_S = 0\} \perp\!\!\!\perp \mathcal{F}_S \implies T - S \sim \mathcal{E}xp(ct)$$

Where we exploited the loss of memory property, namely the second set being strong Markovian. The probability is:

$$\begin{aligned}
\mathcal{P}(T-S \geq t) &= \mathbb{E}\left[\mathbb{1}_{\{T-S>t\}}\right] \\
&= \mathbb{E}\left[\mathbb{E}_S\left[\mathbb{1}_{\{T-S>t\}}\right]\right] && \textit{unconditioning} \\
&= \mathbb{E}\left[\mathbb{E}_S\left[\mathbb{1}\{N_{S+t}-N_S=0\}\right]\right] && \textit{set equivalence above} \\
&= \mathbb{E}_S\left[f(N_{S+t}-N_S)\right] && f(x)=1\cdot\mathbb{1}_{\{x=0\}} \\
&= \mathbb{E}\left[\mathbb{1}_{\{x=0\}}\right] && \textit{Strong Markov Prop. 16.6} \\
&= \mathcal{P}(X=0) && \textit{distr as } \mathcal{P}o(ct) \\
&= e^{-ct} && \perp\!\!\!\perp \mathcal{F}_S, a
\end{aligned}$$

And the distribution is completely independent of a.

♣ **Proposition 16.8** (Total unpredictability of jumps)**.** *This result is mirroring that of Proposition 19.10, which will be proved later.*
Consider a Poisson process N, of which the first jump is $T=T_1$, and a stopping time S wrt $\mathcal{F}$. Then:

$$0 \leq S < T \quad a.s. \implies S=0 \quad a.s.$$

Namely, we cannot find a sequence of stopping times that would approximate T.

Proof. As in the previous Example, we establish with an application of the strong Markov property (Prop. 16.6) that:

$$\{T-S>t\} = \{N_{S+t}-N_S\} \perp\!\!\!\perp \mathcal{F}_S \quad wp\ e^{-ct}$$

Having the Poisson distribution, we derive using the unconditioning property (Prop. 10.18)

$$\mathbb{E}_S\left[T-S\right] = \frac{1}{c} \implies \mathbb{E}\left[\mathbb{E}_S\left[T-S\right]\right] = \mathbb{E}\left[T\right]-\mathbb{E}\left[S\right] = \mathbb{E}\left[\frac{1}{c}\right] = \frac{1}{c}$$

Where $T \sim \mathcal{P}ois(c)$ ensures that the arrival times are exponentially distributed and $\mathbb{E}\left[T\right]=\frac{1}{c}$ implying $\mathbb{E}\left[S\right]=0$. By the hypothesis $S \geq 0$ almost surely, and the fact that we have null expectation, it must be that $S \overset{a.s.}{=} 0$. □

♡ **Example 16.9** (Shot Noise, Ornstein Uhlenbeck process)**.** *The following is a descriptive discussion of a famous process, which will be generalized in Example 22.28.*
*(△ $\mathbb{R}$ **case**) We aim to describe a p.r.m. N on the real line $\mathbb{R}$ with mean $\nu(dx)=cdx$. The arrival times in this case are:*

$$\ldots < T_{-2} < T_{-1} < T_0 < 0 < T-1 < \ldots$$

This could model the arrivals to an anode of electrons producing a current intensity g decreasing as a function of the elapsed time $u \geq 0$. We assume for simplicity that currents are additive.
The total current at time t is then modelled as:

$$\begin{aligned}
X_t &= \sum_{n:T_n \leq t} g(t-T_n) \\
&= \sum_{n=\infty}^{\infty} g(t-T_n)\mathbb{1}_{(-\infty,t]}(T_n) \\
&= \int g(t-x)\mathbb{1}_{(-\infty,t]}(x)\,N(dx) \\
&= Nf && f(x) \coloneqq g(t-x)\mathbb{1}_{[0,t]}(x)
\end{aligned}$$

We wish to describe the moments and the Laplace functional of this process on $\mathbb{R}$:

$$\begin{aligned}
\mathbb{E}[X_t] &= \mathbb{E}[Nf] = \nu f && \text{Prop. 13.18\#1}\\
&= \int_{\mathbb{R}} g(t-x)\mathbb{1}_{(-\infty,t]}(x)\, cdx\\
&= c\int_{-\infty}^{t} g(t-x)dx && \text{let } u=t-x, du=-dx\\
&= c\int_{\infty}^{0} -g(u)du\\
&= c\int_{0}^{\infty} g(u)du
\end{aligned}$$

Which is $\perp\!\!\!\perp t$ once we integrate. Moving on to the variance:

$$\begin{aligned}
V[X_t] &= V[Nf] = \nu f^2 && \text{Prop. 13.18\#2}\\
&= \int_{\mathbb{R}} \left(g(t-x)\mathbb{1}_{(-\infty,t]}(x)\right)^2 cdx\\
&= c\int_{-\infty}^{t} [g(t-x)]^2dx && \text{let } u=t-x, du=-dx\\
&= c\int_{\infty}^{0} -[g(u)]^2du\\
&= c\int_{0}^{\infty} [g(u)]^2du
\end{aligned}$$

Again $\perp\!\!\!\perp t$. Lastly, the Laplace transform is:

$$\begin{aligned}
\mathbb{E}[e^{-rX_t}] &= \mathbb{E}[e^{-rNf}] = \exp\left\{-\nu(1-e^{-rf})\right\} && \text{Thm. 13.19}\\
&= \exp\left\{-\int_{\mathbb{R}}(1-e^{-rg(t-x)\mathbb{1}_{(\infty,t]}(x)})cdx\right\}\\
&= \exp\left\{-c\int_{\mathbb{R}}(1-e^{-rg(t-x)})\mathbb{1}_{(-\infty,t]}(x)\,dx\right\} && \text{move indicator out}\\
&= \exp\left\{-\int_{-\infty}^{t}(1-e^{-rg(t-x)})dx\right\} && \text{let } u=t-x, du=-dx\\
&= \exp\left\{-c\int_{0}^{\infty}(1-e^{-rg(u)})du\right\} && (16.3)
\end{aligned}$$

Since the independence from t carries over to the Laplace functional, we can safely say that by Theorem 13.19 we have that:

$$X_t \overset{d}{=} \widetilde{X}_0 \ \forall t \quad \widetilde{X}_0 \text{ with transform as above}$$

*(☐ $(0,\infty)$ **case**) consider now a more realistic p.r.m. on $\mathbb{R}_+$, which would allow a researcher to simulate the phenomenon[1]. We consider as intensity function*

$$g(u) = ae^{-bu} \quad a>0, b>0$$

and set a starting current to our X_t amount:

$$\begin{cases}
X_t = e^{-bt}X_0 + \sum_{n=1}^{\infty} \underbrace{ae^{-b(t-T_n)}}_{=g(t-T_n)} \mathbb{1}_{[0,t]}(T_n)\\
X_0 \perp\!\!\!\perp T_1 < T_2 < \dots
\end{cases}$$

[1]indeed, a physicist has to start somewhere, but the process in reality does not have a starting point itself.

In this context, we want to show that $X_t \xrightarrow{d} \widetilde{X}_0$ *as before. Proceeding in the same way, we inspect moments and Laplace functional:*

$$\begin{aligned}
\mathbb{E}[X_t] &= \mathbb{E}[e^{-bt}X_0 + Nf] && f(x) := g(t-x)\mathbb{1}_{[0,t]}(x) \\
&= e^{-bt}\mathbb{E}[X_0] + \int_0^t g(t-x)c\,dx && \textit{again by Prop. 13.18\#1} \\
&= e^{-bt}\mathbb{E}[X_0] + c\int_0^t g(u)du && \textit{same ch. variable}
\end{aligned}$$

which is ***dependent*** *on* t*. For what concerns the variance*

$$V[X_t] = \mathbb{E}\left[(e^{-bt}X_0 + Nf)^2\right] - \mathbb{E}\left[e^{-bt}X_0 + Nf\right]^2$$

the result is the same. Moving to the Laplace transform:

$$\begin{aligned}
\mathbb{E}[e^{-rX_t}] &= \mathbb{E}\left[e^{-re^{-bt}X_0}\right]\mathbb{E}\left[e^{-rNf}\right] && \textit{by } X_0 \perp\!\!\!\perp t \\
&= \mathbb{E}\left[e^{-re^{-bt}X_0}\right]\exp\left\{-c\int_0^t \left(1-e^{-rg(t-x)}\right)dx\right\} && \textit{let } u = t-x, du = -dx \\
&= \underbrace{\mathbb{E}\left[e^{-re^{-bt}X_0}\right]}_{\to 1 \text{ as } t\to\infty}\exp\left\{c\int_0^t \left(1-e^{-rg(u)}\right)du\right\} \\
&\overset{t\to\infty}{\to} \exp\left\{c\int_0^t \left(1-e^{-rae^{-bu}}\right)du\right\}
\end{aligned}$$

In the limit, the laplace transform converges pointwise to that of $\widetilde{X}_0$ *of Equation 16.3. Laplace pointwise convergence ensures that* $X_t \xrightarrow{d} \widetilde{X}_0$ *(Thm. 9.39).*

(◯ ***stationarity)*** *we aim to show* $X_0 \overset{d}{=} \widetilde{X}_0 \implies X_t \overset{d}{=} \widetilde{X}_0\ \forall t \in \mathbb{R}_+$*. This would mean that the distribution is stationary around the realistic one over* $\mathbb{R}$*, but feasible for experimentation as argued in □. We start with a split:*

$$\mathbb{E}[e^{-rX_t}] = \underbrace{\mathbb{E}[e^{-re^{-bt}\widetilde{X}_0}]}_{=\boldsymbol{A}}\underbrace{\exp\left\{-c\int_0^t e^{-rg(u)}du\right\}}_{=\boldsymbol{B}} \qquad (\star)$$

$\boldsymbol{A}$ *is the Laplace transform of* $\widetilde{X}_0$ *at* $r' = re^{-bt} > 0$*. Using Theorem 13.19 together with the explicit form in Equation 16.3 we get that:*

$$\begin{aligned}
\boldsymbol{A} &= \exp\left\{-c\int_0^\infty \left(1-e^{-re^{-bt}g(u)}\right)du\right\} \\
&= \exp\left\{-c\int_0^\infty \left(1-e^{-re^{-bt}ae^{-bu}}\right)du\right\} \\
&= \exp\left\{-c\int_0^\infty \left(1-e^{-rae^{-b(t+u)}}\right)du\right\} && \textit{let } x = t+u, dx = du \\
&= \exp\left\{-c\int_t^\infty \left(1-e^{-rae^{-bx}}\right)dx\right\} && \textit{for clearness, let } x = u \\
&= \exp\left\{-c\int_t^\infty \left(1-e^{-rae^{-bu}}\right)du\right\}
\end{aligned}$$

So that $(\star)$ *becomes:*

$$\begin{aligned}
\mathbb{E}[e^{-rX_t}] &= \exp\left\{-c\int_t^\infty \left(1-e^{-rae^{-bu}}\right)du\right\}\exp\left\{-c\int_0^t e^{-rg(u)}du\right\} \\
&= \exp\left\{-c\int_0^\infty \left(1-e^{-rg(u)}\right)du\right\} \perp\!\!\!\perp t \\
\implies X_t &\overset{d}{=} X_0 \overset{d}{=} \widetilde{X}_0 \quad \forall t \in \mathbb{R}_+
\end{aligned}$$

*(◊ **stochastic differential equation**) we want to show that such a process satisfies the SDE:*

$$X_t = X_0 - b\int_0^t X_s ds + aN([0,t])$$

Also written as $dX_t = -bX_t dt + aN(dt)$*. This is equivalent to:*

$$\iff X_t = e^{-bt}X_0 + \int_0^t g(t-x)N(dx) = X_0 - b\int_0^t X_s ds + N([0,t])$$

Where the form we have is the LHS and the form we want is the RHS. Inspecting the integral in the RHS with the result of the LHS [2]*:*

$$\begin{aligned} \int_0^t X_s ds &= \int_0^t e^{-bs}X_0 ds + \int_0^t \int_0^s g(s-x)N(dx)ds && \text{where } s \le x \le t \\ &= \frac{1}{b}X_0(1-e^{-bt}) + \int_0^t \int_x^t g(s-x)dsN(dx) && \text{order change in accordance with } s \le x \le t \end{aligned}$$

Where the blue integral is precisely

$$\int_0^t \int_x^t g(s-x)dsN(dx) = \int_0^t -\frac{a}{b}e^{-b(s-x)}\Big|_{s=x}^t N(dx) = \int_0^t \frac{a}{b}\left(1-e^{-b(t-x)}\right)N(dx)$$

Eventually substituting in the RHS one gets:

$$\begin{aligned} X_0 - b\int_0^t X_s ds + aN([0,t]) &= X_0 + e^{-bt}X_0 - X_0 - a\int_0^t N(dx) + a\int_0^t e^{-b(t-x)}N(dx) \\ &= e^{-bt}X_0 + \int_0^t g(t-x)N(dx) \end{aligned}$$

Which is the LHS.

♡ **Example 16.10** (Continuous time Yule Branching process, Ex. 12.44 ctd)**.** *Consider* $Z_t := \#$ *individuals at time* t*, with* $Z_0 = 1$*. Assume death is not possible and the chance of birth is* dt*, independently for each individual. Namely, one child in the interval* $(t, t+dt]$*, with no influence within the population.*
*(△ **aim**) We show that for each individual the number of descendants is an independent copy of the counting Yule process, upon time shifts to restart it.*
*(□ **exponential interbirths premise**) let* Y_k *be the* k^{th} *inter-birth time. It holds that:*

$$Y_k \overset{ind}{\sim} \mathcal{E}xp(k)$$

*Since the waiting time for the first birth is a unit rate with exponential variable, given the **linear** chance of birth. For general* k*, there are* $k-1$ *individuals plus one ancestor, each birthing at a rate* dt*. The first birth is the minumum of* k *exponential unit random variables. We show that this is again exponential. Observe that* $Y_2 = \min\{E_1, E_2\}, \dots, Y_k = \min\{E_1, \dots, E_k\}$ *where for each* $E_i \sim \mathcal{E}xp(1)$*. We easily conclude:*

$$\begin{aligned} \mathcal{P}(Y_k > t) &= \mathcal{P}(E_1 > t, \dots, E_k > t) \\ &= (\mathcal{P}(E_1 > t))^k && iid \\ &= e^{-kt} \\ &= \mathcal{P}(E > t) && E \sim \mathcal{E}xp(k) \end{aligned}$$

*(○ **first result**)wts*

$$Z_t \sim \mathcal{G}eom(e^{-t}) \iff \mathcal{P}(Z_t = x) = e^{-t}(1-e^{-t})^{x-1} \quad x = 1, 2, \dots$$

Notice that the interarrivals denoted with Y_k *allow to define the arrivals process* $S = (S_n)_{n\in\mathbb{N}}$ *as*

$$S_n = \sum_{k=1}^n Y_k$$

[2]this is slightly informal to say

Which is equivalent to:

$$\iff \{S_n \le t\} = \{Z_t - 1 \ge n\} \iff \{S_n \le t\} = \{Z_t > n\} \tag{16.4}$$

Meaning that the arrival times are stopping times for the underlying counting process Z *in the usual sense (Def. 11.9).*
Notice that in the p.r.m. case of the compound Poisson process (Def. 15.5) we had that S_n *was a sum of unit exponentials, returning a* $\mathcal{G}amma(n,1)$ *distribution (Thm. 16.2). Here instead:*

$$S_n = \sum_{k=1}^{n} \underbrace{Y_k}_{\sim \mathcal{E}xp(k)} = \sum_{k=1}^{n} \frac{1}{k} \underbrace{E_1}_{\sim \mathcal{E}xp(1)} \stackrel{d}{=} \max\{E_1, \dots, E_k\}$$

By a reverse time heuristic argument or a mgf argument. Eventually:

$$\begin{aligned}
\mathcal{P}(S_n < t) &= \mathcal{P}(E_1 \le t, \dots, E_n \le t) && E_k \stackrel{iid}{\sim} \mathcal{E}xp(1) \\
&= (\mathcal{P}(E_1 < t))^k && iid \\
&= (1 - e^{-t})^k \\
&= \mathcal{P}(Z_t > n) && \text{Eqn. 16.4} \\
\implies \mathcal{P}(Z_t = n) &= e^{-t}(1 - e^{-t})^k \implies Z_t \sim \mathcal{G}eom(e^{-t})
\end{aligned}$$

*($\triangledown$ **growth rate**)wts*

$$\frac{Z_t}{\mathbb{E}[Z_t]} \stackrel{a.s.}{\to} W \sim \mathcal{E}xp(1)$$

First of all, observe that the unnormalized rate would explode exponentially fast:

$$\mathbb{E}[Z_t] = \frac{1}{e^{-t}} = e^t \nearrow \infty$$

thus a normalized version. Let $W_t = e^{-t} Z_t = \dfrac{Z_t}{\mathbb{E}[Z_t]}$ *and inspect the process* $(W_t)_{t \in \mathbb{R}_+}$.
*($\triangledown\spadesuit$ **subpoint, W is a martingale**) we show for* $\mathcal{F} = \sigma(Z)$ *that* W *is a martingale according to Definition 11.35. Adaptedness and integrability are easily verified. The martingale equality holds since:*

$$\begin{aligned}
\mathbb{E}_s[W_t] &= \mathbb{E}[W_t | Z_s] && Z \text{ only determinator} \\
&= e^{-t}\mathbb{E}[Z_t | Z_s] \\
&= e^{-t}\mathbb{E}[Z_{t-s} | Z_0 = Z_s] \\
&= e^{-t} Z_s \mathbb{E}[Z_{t-s} | Z_0 = 1] \\
&= e^{-t} Z_s e^{t-s} && \text{previous results} \\
&= e^{-s} Z_s = W_s
\end{aligned}$$

*($\triangledown\clubsuit$ **subpoint, Gumbell limit identity**) recall that for* $S = (S_n)_{n \in \mathbb{N}} = \max\{E_1, \dots, E_n\}$ *it holds that:*

$$S_n - \log(n) \stackrel{d}{\to} \Psi \sim \mathcal{G}umbell \quad \mathbb{P}[\Psi = x] = e^{-e^{-x}}$$

*($\triangledown\heartsuit$ **subpoint, limiting distribution**) we close the task of* $\triangledown$*. Recall that* $\mathcal{P}(Z_t > n) = \mathcal{P}(S_n \le t)$ *by Equation 16.4 in* $\triangle$*. Now consider:*

$$\begin{aligned}
\mathcal{P}(e^{-t} Z_t < x) &= \mathcal{P}(W_t < x) && \triangledown \\
&= \mathcal{P}(Z_t < e^t x) \\
&= \mathcal{P}(S_{e^t x} > t) && \text{if necessary take integer part} \\
&= \mathcal{P}\left(S_{e^t x} - \log[e^{tx}] > t - \log[e^{tx}]\right) \\
&= \mathcal{P}\left(S_{e^t x} - \log[e^{tx}] > -\log[x]\right) && \text{apply } \triangledown, \clubsuit \\
&\stackrel{t \to \infty}{\to} e^{tx} && \mathcal{P}(\Psi < -\log[x]) = e^{-e^{-\log(x)}} = e^{-x} \\
\implies W_t &\stackrel{d}{\to} \mathcal{E}xp(1)
\end{aligned}$$

Now we can apply the MCT (Thm. 12.27, especially Corollary 12.30#2 for positive martingales) to conclude that we have an almost sure limiting distribution such that $W_t \xrightarrow{a.s.} W_\infty$. *By the existance of an almost sure limit, it coincides with that in distribution (Props. 9.14, 9.38) and:*

$$\exists W_\infty : W_t \xrightarrow{a.s.} W_\infty, \quad W_t \xrightarrow{d} \mathcal{E}xp(1) \implies W_t \xrightarrow{a.s.} W_\infty \sim \mathcal{E}xp(1)$$

Chapter Summary

Objects:

- summary of previous objects including Poisson processes, counting processes in general, with representation via arrival times, which are stopping times for the natural filtration of N, itself generated by the underlying random measure M
- N the counting process is also a martingale when normalized by its current expectation and the underlying measure is random counting and Poisson with mean $\nu(dx) = cdx$

Results:

- for $c > 0$ the following are equivalent:
  - M is a p.r.m. (Def. 13.13) with mean $\mu = cLeb$
  - N is a Poisson (counting) process with rate c
  - N is a counting process and $\widetilde{N} = (N_t - ct)_{t\geq 0}$ is an $\mathcal{F}$-martingale
  - $(T_k)_{k\geq 1}$ is an increasing sequence of $\mathcal{F}$-stopping times and:

$$T_1, T_2 - T_1, \ldots \overset{iid}{\sim} \mathcal{E}xp(c)$$

- a counting process is increasing Lévy if and only if it is Poisson
- Poisson processes (rate c) follows the strong Markov property, conditioning at a stopping time S makes them Markovian in the sense that:

$$\mathbb{E}_S[f(N_{S+t} - N_S)\mathbb{1}_{\{S<\infty\}}] = \sum_{k=0}^{\infty} f(k)\frac{e^{-ct}(ct)^k}{k!}\mathbb{1}_{\{S<\infty\}}$$

- jumps of Poisson processes are totally unpredictable, if T is the first jump of N and $S \leq T$ then $S \overset{a.s.}{=} 0$ almost surely, and there is no sequence of stopping times approximating T

Chapter 17

Lévy Processes

♠ **Definition 17.1** (Lévy process)**.** *A process $X = (X_t)_{t\in\mathbb{R}_+}$ is Lévy wrt a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ if:*

1. *(adaptedness) it is adapted to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ (Def. 11.7)*
2. *(right continuity and starts at zero) for a.e. $\omega \in \Omega$ the path $t \to X_t$ is right continuous and $X_0(\omega) = 0$*
3. *(stationary and independent increments) $\forall s, t \geq 0$ $X_{s+t} - X_s \perp\!\!\!\perp \mathcal{F}_s \overset{d}{=} X_t$*

♢ **Observation 17.2** (Differences with increasing Lévy processes)**.** *Compare this Definition with Definition 15.14. We are* ***removing*** *the term* ***increasing****.*

♢ **Observation 17.3** (About Lévy processes)**.** *trivially:*

- *cX_t is a Lévy process*
- *$\sum^n X_t^{(i)}$ is a Lévy process*

♠ **Definition 17.4** (Infinite divisibility)**.** *We express a Lévy process $(X_t)_{t\in\mathbb{R}_+}$ as $\sum_{i=1}^n X_i = X_t \quad \forall n, t$ where the elements are all Lévy processes:*

$$\delta = \frac{t}{n} \implies X_t = X_t \mathbb{1}_{[0,\delta]}(t) + X_t \mathbb{1}_{[\delta,2\delta]}(t) + \cdots + X_t \mathbb{1}_{[\delta(n-1),\delta n]}(t)$$

where the increments are independent and identically distributed.

♠ **Definition 17.5** (Characteristic exponent $\psi(r)$)**.** *This is a direct result of the infinite divisibility, which makes the process decompose into independent processes. A Lévy process can be described by the characteristic exponent, a complex valued function such that:*

$$\Phi_{X_t}(r) = \mathbb{E}[e^{irX_t}] = e^{t\psi(r)} \quad t \in \mathbb{R}_+, r \in \mathbb{R}$$

Where $\psi : \mathbb{R} \to \mathbb{C}$ is complex valued.

♡ **Example 17.6** (Easy Lévy processes)**.** *recognize that:*

- *a Wiener process (Def. 11.55) is such that $X_t = bt + cW_t$ is Lévy*
- *a Poisson process (Def. 12.2) $N = (N_t)_{t\in\mathbb{R}_+}$ is Lévy*
- *a compound Poisson process (Def. 15.5) $X_t = \sum_{n=1}^\infty Y_n \mathbb{1}_{[0,t]}(T_n)$ for arrival times (T_n) is increasing Lévy if it satisfies the integrability condition on the Lévy measure:*

$$\int \lambda(dz)(z \wedge 1) < \infty \overset{Thm 16.3}{\implies} X_t = bt + \int_{[0,t]\times\mathbb{R}_+} zN(dx, dz)$$

where $N \sim \mathcal{P}ois(Leb \times \lambda)$. Provided that this holds, $(X_t)_{t\in\mathbb{T}}$ is an (increasing) Lévy process

♠ **Definition 17.7** (Pure jump process)**.** *Consider on $\mathbb{R}$ the Lévy process*

$$X_t = \sum_{s\in[0,t]\cap D_\omega} \Delta X_s \qquad \forall t \quad \Delta X_s = X_s(\omega) - X_{s-}(\omega),\ D_\omega = \{t > 0 : \Delta X_t(\omega) \neq 0\}$$

Then:

- *the jumps are positive or negative, we could see $X_t = X_t^+ + X_t^-$ where both are increasing Lévy*
- *if the jumps are countable (e.g. arising from arrival times (T_n)) then we can evaluate the sum, we do so by intersecting the time interval with D_ω*

We call this a pure jump process, notice that it is not necessarily ***increasing****.*

♠ **Definition 17.8** (Total Variation V_t of the pure jump)**.** *We give a first definition of total variation of a path of a pure jump process $t \to X_t$ as:*

$$V_t = \sum_{s \in [0,t] \cap D_\omega} |\Delta X_s| \quad \forall t \in \mathbb{R}_+$$

♣ **Proposition 17.9** (General representation & existance conditions of Lévy process)**.** *For a p.r.m. M on $\mathbb{R}_+ \times \mathbb{R}^d$ with mean $Leb \times \lambda$ and $\lambda(\{0\}) = 0$ if:*

$$\lambda(|x| \wedge 1) = \int_{\mathbb{R}^d} \lambda(dx)(|x| \wedge 1) < \infty \tag{17.1}$$

then:

1. *for a.e. ω the process arising from the integral $X_t(\omega) = \int_{[0,t]\times\mathbb{R}^d} M_\omega(ds, dx)x$ converges absolutely $\forall t$ and it has bounded total variation $V_t < \infty \quad \forall t$*
2. *X is a pure jump Lévy process with characteristic exponent*

$$\psi(r) = \lambda(e^{ir\cdot x} - 1) = \int_{\mathbb{R}^d} \lambda(dx)(e^{ir\cdot x} - 1) \quad \forall r \in \mathbb{R}$$

Proof. Let:

- $\widehat{M}$ be the image of M under $(s, x) \to (s, |x|)$ from $\mathbb{R}_+ \times \mathbb{R}^d$ to $\mathbb{R}_+ \times \mathbb{R}_+$
- $\widehat{\lambda}$ be the image of λ under $x \to |x|$

Then M is a p.r.m. on $\mathbb{R}_+ \times \mathbb{R}_+$ with mean $Leb \times \widehat{\lambda}$ by Proposition 14.13. Note that $\widehat{\lambda}(\{0\}) = 0$ so that the integrability condition becomes:

$$\int_{\mathbb{R}_+} \widehat{\lambda}(dv)(v \wedge 1) < \infty$$

Which implies that $\widehat{\lambda}((\epsilon, \infty)) < \infty \forall \epsilon > 0$. By Corollary 14.11 we can choose Ω' an almost sure event such that:

$$\forall \omega \in \Omega' \quad \widehat{M}_\omega \textit{ counting}$$

according to Definition 2.4, with:

- no atoms in $\mathbb{R}_+ \times \{0\}$
- at most one atom in $\{t\} \times \mathbb{R}_+ \quad \forall t$

On the other hand, for each time t:

$$V_t = \int_{[0,t]\times\mathbb{R}_+} \widehat{M}(ds, dv)v = \int_{[0,t]\times\mathbb{R}^d} M(ds, dx)|x| \tag{17.2}$$

is positive and in $\mathbb{R}$ almost surely by the integrability condition and Proposition 14.4. Let Ω_t be the almost sure event for each t. Additionally:

$$\Omega'' = \bigcup_{t\in\mathbb{N}} \Omega_t$$

Fix $\omega \in \Omega' \cap \Omega''$. The map:

$$t \to V_t(\omega) \quad f : \mathbb{R}_+ \to \mathbb{R}_+$$

is:

- right continuous
- increasing
- $V_0(\omega) = 0$
- jump of size v at time $s \iff M_\omega$ has an atom (s, x) with $|x| = v$

From this the integral of Claim #1 converges absolutely for every t and

$$\sum_{s\leq t}|\Delta X_s(\omega)| = V_t \quad \sum_{s\leq t}\Delta X_s(\omega) = X_t(\omega)$$

So that X is of pure-jump type with total variation over $[0,t]$ identified by V_t.
From the integral of Claim #1 and the Poisson character of M the process X has independent and stationary increments. Its form, the fact that we can use a characteristic exponent (Def. 17.5) and Theorem 13.19 suggest that:

$$\psi(r) = \int_{\mathbb{R}^d}\lambda(dx)(e^{irx}-1) \quad \forall r\in\mathbb{R}$$

□

♢ **Observation 17.10** (A side note on the result)**.** *The integrability condition on the Lévy measure is very similar to that of compound Poisson processes. The naïve difference is that we need to bound negative values as well, while before the process was only increasing.*

♢ **Observation 17.11** (About the Proposition)**.** *Many comments can be made:*

- *the Lévy measure λ regulates the jumps and characterizes both M and X.*
- *we have $\forall A\subset\mathbb{R}$ such that $\lambda(A)<\infty$ that the jump sizes in A form a counting process with path $t\to M([0,t]\times A)$, where the latter has Poisson rate $\lambda(A)$.*
- *the condition $\lambda(\{0\})=0$ is to ensure that M_ω and $X(\omega)$ determine each other uniquely almost surely*
- *trivially, a measure λ finite satisfies the requirements, resulting in a Poisson compound process with jumps $(Y_n)_{n\geq1}$ and $Y_n \overset{iid}{\sim} \pi(dz)$ in the sense of Definition 15.5. Notice however that differently from Example 15.4, we are allowing jumps to take place on $\mathbb{R}$ and not $\mathbb{R}_+$ only*
- *infinite Lévy measures that satisfy the integrability condition come with ∞ activity but finitely many jumps exceeding $\varepsilon>0$ arbitrary in absolute value since $\lambda((-\infty,-\epsilon))+\lambda((\epsilon,\infty))<\infty$ for any interval $(s,t): s<t$*
- *another characterization of total variation is Equation 17.2*
- *the total variation V_t is a pure jump increasing Lévy process with Lévy measure as the image of λ under $x\to|x|$. The path $X(\omega)$ jumps with size x at time t if and only if $V(\omega)$ jumps at time t with size $|x|$*

♡ **Example 17.12** (Symmetric Gamma process, Ex. 15.35 ctd)**.** *Recall that a Gamma process is an increasing Lévy process (Def. 15.14) with measure and distribution:*

$$\lambda(dz) = \frac{ae^{-cz}}{z}dz,\ z\in\mathbb{R}_+, \quad X_t\sim \mathcal{G}amma(at,c)\ \forall t\in\mathbb{R}_+$$

Let $X_t^+ \perp\!\!\!\perp X_t^-$ be independent copies, and set $X_t = X_t^+ - X_t^-$. Then, X_t is a pure jump Lévy process according to Definition 17.7 with measure:

$$\lambda(dz) = \frac{ae^{-c|z|}}{|z|}dz$$

We aim to evaluate its characteristic function to see if it coincides with some known distribution.

$$\begin{aligned}
\mathbb{E}\left[e^{irX_t}\right] &= \mathbb{E}\left[e^{irX_t^+}\right]\mathbb{E}\left[e^{i(-r)X_t^-}\right] && \textit{independence}\\
&= \left(\frac{c}{c+ir}\right)^{at}\left(\frac{c}{c-ir}\right)^{at} && \textit{Ex. 15.35}\\
&= \left(\frac{c^2}{(c+ir)(c-ir)}\right)^{at}\\
&= \left(\frac{c^2}{c^2+r^2}\right)^{at}
\end{aligned}$$

Which means that the characteristic exponent is real:

$$\psi(r) = \frac{1}{t}\log\left[\mathbb{E}[e^{irX_t}]\right] = a\log\left[\frac{c^2}{c^2+r^2}\right]\in\mathbb{R}$$

While X_t has no known distribution, it can be shown that the total variation $V = X^+ + X^-$ is such that:

$$V_t \sim Gamma(2at,c) \quad \forall t\in\mathbb{R}_+$$

♡ **Example 17.13** (Isotropic stable process, Ex. 15.34 ctd). *Let $X_t^+ \perp\!\!\!\perp X_t^-$ be independent copies of the stable process from Example 15.34. The density of $X_t = X_t^+ - X_t^-$ is:*

$$\lambda(dz) = \frac{ac}{\Gamma(1-a)}|z|^{1-a}dz, \quad z \neq 0, a \in (0,1)$$

Such a process is pure jump Lévy according to Definitions 17.1, 17.7 and has Laplace transform:

$$\begin{aligned}\mathbb{E}[e^{irX_t}] &= \exp\left\{t\int_{\mathbb{R}}(e^{irz}-1)\lambda(dz)\right\}\\ &= \exp\left\{tc\cos\left(\frac{1}{2}\pi a\right)|r|^a\right\}\end{aligned}$$

With characteristic exponent:

$$\psi(r) = \int_{\mathbb{R}}(e^{irz}-1)\lambda(dz) = -c\cos\left(\frac{1}{2}\pi a\right)|r|^a$$

Which is stable since $X_t \overset{d}{=} t^{\frac{1}{a}}X_t$ $\forall t$.

17.1 Compensated sum of jumps from relaxed integrability conditions

♢ **Observation 17.14** (Infinite variation process). *We consider a Lévy process such that $V_t = \infty$ for each positive path $t > 0$. Then:*

$$X_{t-}(\omega) = \lim_{s\uparrow t} X_s(\omega) \quad X_{t+}(\omega) = \lim_{s\downarrow t} X_s(\omega)$$

By right continuity of the Lévy process, we have for free that $X_{t+}(\omega) = X_t(\omega)$. If $X_{t-}(\omega) \neq X_{t+}(\omega)$ then we will have a jump at time t denoted as $\Delta X_t(\omega)$. Let:

$$D_\omega = \{t > 0 \ : \ \Delta X_t(\omega) \neq 0\}$$

be the set of times where a jump takes place. Then, $\forall \epsilon > 0$ there are infinitely many jumps in $D_\omega \cap (s,u)$ such that $\Delta X_t(\omega) > \epsilon$. Otherwise, by Bolzano-Weierstrass there would exist a countable subsequence $(t_n) \subset D_\omega$ where $n \to t_n \in [s,u]$ such that either $X_{t_n-}(\omega)$ or $X_{t_n+}(\omega)$ do not exist.
However, does an infinite V_t Lévy process exist at all?
We aim to build a Lévy process driven by a p.r.m. N with mean $dx\lambda(dz)$ even when the integrability condition is not satisfied. For simplicity, we concentrate on jumps that have size less than one in absolute value.

♠ **Definition 17.15** (Basis notation). *denote:*

- $\mathbb{B} = \{x \in \mathbb{R} \ : \ |x| \leq 1\}$
- $\mathbb{B}_\epsilon = \{x \in \mathbb{R} \ : \ \epsilon < |x| \leq 1\}$ *for* $\epsilon \in (0,1)$

♣ **Theorem 17.16** (Infinite total variation Lévy existance as compensated sum of jumps). *Let $M \sim \mathcal{P}ois(Leb\times\lambda)$ (Def. 13.13) on $\mathbb{R}_+ \times \mathbb{B}$ where $\lambda(\{0\}) = 0$ and:*

$$\lambda(|x|^2\mathbb{1}_{\mathbb{B}}) = \int_{\mathbb{B}}\lambda(dx)|x|^2 < \infty \tag{17.3}$$

For $\epsilon \in (0,1)$ consider:

$$X_t^\epsilon(\omega) = \int_{[0,t]\times\mathbb{B}_\epsilon} xM_\omega(ds,dx) - t\int_{\mathbb{B}_\epsilon}\lambda(dx)x \quad \omega \in \Omega, t \in \mathbb{R}_+$$

Then:

1. *$\exists X$ Lévy such that $\lim_{\epsilon\downarrow 0} X_t^\epsilon(\omega) \overset{a.s.}{=} X_t(\omega)$ uniformly convergent over bounded intervals*
2. *$\psi(r) = \int_{\mathbb{B}}\lambda(dx)(e^{irx}-1-irx) \quad r \in \mathbb{R}$*

♢ **Observation 17.17** (Technicalities of the theorem). *some comments are:*

1. $(e^{irx} - 1 - irx) \leq \frac{1}{2}(irz)^2 = \frac{1}{2}r^2z^2$ *which is a well defined quantity under the integrability condition of Equation 17.3*
2. $X_t^\epsilon = Y_t^\epsilon + a_\epsilon t$ *where* Y_t^ϵ *is compound Poisson (Def. 15.5) and* a_ϵ *is a fixed drift. Additionally denote:*

$$b_\epsilon = \int_{\mathbb{B}_\epsilon} \lambda(dx)|x| \quad c_\epsilon = \int_{\mathbb{B}_\epsilon} \lambda(dx)|x|^2, \quad \epsilon \in [0,1)$$

notice that by[1] $\epsilon^2\lambda(\mathbb{B}_\epsilon) \leq \epsilon b_\epsilon \leq c_\epsilon \leq c_0$ *the integrability condition of Equation 17.3 implies that* $c_0 < \infty$ *guaranteeing that all such quantities are finite:*

$$\lambda(\mathbb{B}_\epsilon) < \infty \quad b_\epsilon < \infty \quad \forall \epsilon \in [0,1)$$

so that $\lambda(\mathbb{B}_\epsilon) < \infty \implies X_t(\omega)$ *from Thm. 17.16 is convergent and the second integral in the expression* a_ϵ *is a vector in* $\mathbb{R}^d$*.*

3. *the uniform convergence over bounded intervals claim is equivalent to:*

$$X \ : \ for\ a.e.\ \omega \quad \lim_{\epsilon \downarrow 0} \sup_{0 \leq t \leq u} |X_t^\epsilon(\omega) - X_t(\omega)| = 0 \quad \forall u \in \mathbb{R}_+$$

4. *if it also holds that:*

$$b_0 < \infty \implies \lambda \ sat.\ Eqn.\ 17.1 \overset{Prop.\ 17.9}{\implies} Y_t = \lim_{\epsilon \downarrow 0} Y_t^\epsilon = \int_{[0,t]\times\mathbb{B}} M(ds,dx)x \quad \forall t \in \mathbb{R}_+$$

with Y_t *a pure jump process (Def. 17.7) and:*
- $a = \lim_{\epsilon\downarrow 0} a_\epsilon = \int_{\mathbb{B}} \lambda(dx)x$ *well defined*
- $X_t = Y_t - at \quad \forall t \in \mathbb{R}_+$

We refer to X_t *as the stochastic integral:*

$$X_t = \int_{[0,t]\times\mathbb{B}} x\,[M(ds,dx) - ds\lambda(dx)]$$

called a compensated sum of jumps, compensated in that every X_t^ϵ *is the sum of the sizes of the jumps during* $[0,t]$ *larger than* ϵ *in magnitude minus the expected value at* a_t^ϵ *so to get a martingale.*

♢ **Observation 17.18** (Why compensated jumps & a Theorem)**.** *We try to justify the use of a Theorem to present the results.*

1. *Recall Observation 17.17, let* $b_0 = \infty$ *and* $c_0 < \infty$*. Then Theorem 17.16 can be applied since Equation 17.3 holds but Proposition 17.9 cannot be applied since Equation 17.1 does not. So:*
   - a_ϵ, Y_t^ϵ *are not convergent as* $\epsilon \to 0$
   - *but* $X_t^\epsilon = Y_t^\epsilon - a_\epsilon$ *converges!*
   - X *has infinite variation over every time interval* $(s,t) : s < t$
2. X_t^ϵ *is a compensated jumps martingale in the sense that:*

$$\mathbb{E}[Y_t^\epsilon] = a_\epsilon t \implies \mathbb{E}[X_t^\epsilon] = 0 \quad \forall \epsilon \in (0,1), \forall t \in \mathbb{R}_+$$

where the sum of the sizes of the jumps is Y_t^ϵ *compensated by* $a_\epsilon t$*.*

♡ **Example 17.19** (Standard Cauchy process)**.** *Let the Lévy measure be:*

$$\lambda(dz) = \frac{1}{\pi z^2}dz \quad z \in \mathbb{R}_+$$

[1] which in turn holds as $\epsilon^2 \leq \epsilon|x| \leq |x|^2$ for $x \in \mathbb{B}_\epsilon$

It holds that (be careful with the second as it is a bit tricky):

$$\begin{aligned}\int_{-1}^{1} z^2\lambda(dz) &= \int_{\mathbb{B}} z^2\lambda(dz) = \frac{2}{\pi} < \infty \\ \int_{\mathbb{B}} |z|\lambda(dz) &= \int_{-1}^{1} |z|\frac{1}{\pi z^2}dz \\ &= \int_{-1}^{0} -z\frac{1}{\pi z^2}dz + \int_{0}^{1} z\frac{1}{\pi z^2}dz \qquad \textit{basically without modulus undefined} \\ &= \frac{1}{\pi}\left(\int_0^1 \frac{1}{z}dz - \int_{-1}^{0}\frac{1}{z}dz\right) \\ &= \frac{1}{\pi}\left(\ln(|z|)\Big|_0^1 - \ln(|z|)\Big|_{-1}^{0}\right) \\ &= \frac{1}{\pi}(\infty + \infty) \\ &= +\infty\end{aligned}$$

So we can apply Theorem 17.16 having infinite total variation.
Let $X_t = X_t^d + X_t^e$ where:

$$\begin{aligned}X_t^d &= \int_{[0,t]\times\mathbb{B}} z\left(N(dx,dz) - dx\lambda(dz)\right) \qquad \textit{jumps size} \le 1 \\ X_t^e &= \int_{[0,t]\times\mathbb{B}^c} zN(dx,dz) \qquad \textit{jumps size} > 1\end{aligned}$$

Here X_t^e is such that $\int_1^\infty \lambda(dz) < \infty$.
The characteristic exponent is:

$$\psi(r) = \int_{\mathbb{B}}(e^{irz} - 1 - irz)\lambda(dz) + \int_{\mathbb{B}^c}(e^{irz} - 1)\lambda(dz)$$

since we can split the process into the infinite variation part and the finite one. Recalling Observation 17.17 we also have that:

$$X_t^d = \lim_{\epsilon\downarrow 0} X_t^{d,\epsilon} \quad X_t^{d,\epsilon} = \int_{[0,t]\times\mathbb{B}_\epsilon} zN(dx,dz) - t\underbrace{\int_{\mathbb{B}_\epsilon} z\lambda(dx,dz)}_{=0 \textit{ by symmetry}}$$

so that $X_t^{d,\epsilon}$ requires no compensation and we eventually get:

$$\begin{aligned}X_t &= \lim_{\epsilon\downarrow 0}\int_{[0,t]\times\mathbb{B}_\epsilon} zN(dx,dz) + \int_{[0,t]\times\mathbb{B}^c} zN(dx,dz) \\ &= \lim_{\epsilon\downarrow 0}\int_{[0,t]\times\mathbb{R}_\epsilon} zN(dx,dz) + \int_{[0,t]\times\mathbb{B}^c} zN(dx,dz) \qquad \textit{Dominated conv. Thm. A.51}\end{aligned}$$

where $\mathbb{R}_\epsilon = \{x : |x| > \epsilon\}$ dominates $\mathbb{B}_\epsilon$. Notice that this is ***not a pure jump process*** *in the sense of Definition 17.7. Nevertheless, the Laplace transform is:*

$$\begin{aligned}\mathbb{E}[e^{irX_t}] &= \exp\left\{t\lim_{\epsilon\downarrow 0}\int_{\mathbb{R}_\epsilon}(e^{irz} - 1)\frac{1}{\pi z^2}dz\right\} \\ &= \exp\left\{t\int_{\mathbb{R}}(e^{irz} - 1)\frac{1}{\pi z^2}dz\right\} \\ &= \exp\left\{t\int_{\mathbb{R}}(\cos(rz) + i\sin(rz) - 1)\frac{1}{\pi z^2}dz\right\} \\ &= \exp\left\{t\int_{\mathbb{R}}(\cos(rz) - 1)\frac{1}{\pi z^2}dz\right\} \qquad \frac{\sin(rz)}{\pi z^2} \textit{ symm around } 0 \\ &= \exp\{t|r|\} \qquad \textit{Prop. 17.21\#2}\end{aligned}$$

And we have that $X_t \overset{d}{=} t^{\frac{1}{1}} X_1$ *(stability with index* 1*). Moreover by:*

$$X_1 \overset{d}{=} \frac{Z_1}{Z_2}, \quad Z_1, Z_2 \sim \mathcal{N}(0,1), \quad X_1 \sim \mathcal{C}auchy(1) \text{ (Prop. 17.21\#1)}, \ f(x) = \frac{1}{\pi(1+x)^2}$$

We have that:

$$X_t \ : \ f(x) = \frac{t}{\pi(t^2+x^2)} \quad \forall x \in \mathbb{R}, \forall t \in \mathbb{R}_+$$

Lemma 17.20 (An identity for distribution and Φ)**.** *This is presented here but proved in Lemma C.32. For X a r.v. on $\mathbb{R}$ with density $f(x)$:*

$$\Phi_X(r) = \int_{\mathbb{R}} e^{irx} f(x)dx \iff f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-irx}\Phi_X(r)dr$$

Proof. Lemma C.32. □

♡ **Example 17.21** (Concluding Example 17.19)**.** *In the context of the standard Cauchy process we add that:*

1. $X_1 \sim \mathcal{C}auchy(1)$
2. $\int_{\mathbb{R}}(\cos(rz)-1)\frac{1}{\pi z^2}dz = |r|$

(Claim #1) *use Lemma 17.20 and the fact that the Characteristic function of a $\mathcal{C}auchy(1)$ distribution is $\Phi(r) = e^{-|r|}$. We have:*

$$\begin{aligned}
\frac{1}{2\pi}\int_{\mathbb{R}} e^{-irx}e^{-|r|}dr &= \frac{1}{\pi}\int_{\mathbb{R}} e^{ir(-x)} \underbrace{\frac{1}{2}e^{-|r|}}_{\text{density Laplace}} dr \\
&= \frac{1}{\pi}\frac{1}{1+(-x)^2} && \text{characteristic function Laplace} \\
&= \frac{1}{\pi(1+x^2)} \\
&= f(x) && X_1
\end{aligned}$$

(Claim #2) *by the symmetry of the integrated function, the claim is equivalent to:*

$$2\int_0^\infty (1-\cos(rz))\frac{1}{\pi(z^2)}dz = |r|$$

(△ first step) *for a triangular distribution $U+U'-1$ where $U \sim \mathcal{U}nif(0,1)$ the density is*

$$f(x) = 1-|x| \quad |x| < 1$$

with characteristic function:

$$\begin{aligned}
\Phi(r) &= \int_{-1}^{1} e^{-rx}(1-|x|)dx \\
&= \int_{-1}^{1}(\cos(rx)+i\sin(rx))(1-|x|)dx \\
&= \int_{-1}^{1}\cos(rx)(1-|x|)dx && \text{symmetry of second term} \\
&= 2\int_0^1 \underbrace{\cos(rx)}_{f'}\underbrace{(1-x)}_{g}dx && \cos(-x) = \cos(x) \quad \forall x \\
&= \underbrace{2\frac{-\sin(rx)(1-x)}{r}\bigg|_{x=0}^{1}}_{=0} - 2\int_0^1 \frac{-\sin(rx)(-1)}{r}dx && \text{integration by parts} \\
&= -\frac{2}{r^2}\cos(rx)\bigg|_{x=0}^{1} \\
&= \frac{2}{r^2}(1-\cos(r)) && \cos' = -\sin',\ \sin' = \cos
\end{aligned}$$

*(□ **density**) using Lemma 17.20 for the triangular distribution of △ we have*

$$f(x) = (1 - |x|)\mathbb{1}_{[-1,1]}(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-irx}\frac{2}{r^2}(1-\cos(r))dr \implies 1 = \frac{1}{\pi}\int_{\mathbb{R}}\frac{1-\cos(r)}{r^2}dr \quad \text{at } x = 0$$

where, using $r = u,\ u = rx : du = |r|dx$:

$$\int_{\mathbb{R}}\frac{1-\cos(u)}{\pi u^2}du = \int_{\mathbb{R}}\frac{1-\cos(rx)}{\pi r^2 x^2}|r|dx = \int_{\mathbb{R}}\frac{1-\cos(rx)}{\pi|r|x^2}dx = 1$$

So that:

$$\int_{\mathbb{R}}\frac{1-\cos(rx)}{\pi x^2}dx = |r|$$

and we have proved the claim.

Chapter Summary

Objects:

- Lévy processes are:
  - adapted
  - right continuous, starting at 0 almost surely
  - with stationary and indepedent increments
- the concept of infinite divisibility allows for the formulation of the characteristic exponent $\psi(r)$: $\mathbb{E}\left[e^{irX_t}\right] = e^{t\psi(r)}$ for all $t \in \mathbb{R}_+, r \in \mathbb{R}$
- a pure jump process is a countable sum of sizes of its jumps over time:

$$X_t = \sum_{s\in[0,t]\cap D_\omega}\Delta X_s \quad \forall t \quad \Delta X_s = X_s(\omega) - X_{s-}(\omega),\ D_\omega = \{t > 0 : \Delta X_t(\omega) \neq 0\}$$

  their total variation is $V_t = \sum_{s\in[0,t]\cap D_\omega}|\Delta X_s|$. For X_t to exist, V_t must be finite
- to ease out computations we define $\mathbb{B} = \{x : |x| \leq 1\}$, $\mathbb{B}_\epsilon = \{x : \epsilon < |x| < 1\}$ for some $\epsilon \in (0,1)$

Results:

- a pure jump Lévy process exists when it is defined as the integral of a Poisson random measure $M \sim \mathcal{P}ois(Leb \times \lambda)$ where the size measure satisfies:

$$\lambda(\{0\}) = 0 \qquad \lambda(|x| \wedge 1) < \infty \tag{17.4}$$

  so that:
  - for almost every ω $X_t(\omega) = \int_{[0,t]\times\mathbb{R}^d} M_\omega(ds,dx)x$ is absolutely convergent with bounded total variation
  - $X = (X_t)_{t\in\mathbb{T}}$ is of pure jump time as desired with characteristic exponent

$$\psi(r) = \lambda(e^{ir\cdot x} - 1) \quad \forall r \in \mathbb{R}$$

- in the case in which the total variation is infinite, we can still devise a compensated Lévy process if the size measure λ satisfies:

$$\lambda(\{0\}) = 0 \quad \lambda(|x|^2\mathbb{1}_{\mathbb{B}}) < \infty$$

  so that:
  - the process:

$$X_t^\epsilon(\omega) = \int_{[0,t]\times\mathbb{B}_\epsilon} xM_\omega(ds,dx) - t\int_{\mathbb{B}_\epsilon}\lambda(dx)x \quad \omega \in \Omega, t \in \mathbb{R}_+$$

  - is uniformly convergent as $\epsilon \downarrow 0$ almost surely to a Lévy process. That is, $\lim_{\epsilon\downarrow 0} X_t^\epsilon(\omega) \overset{a.s.}{=} X_t(\omega)$ uniformly convergent over bounded intervals
  - the characteristic exponent is $\psi(r) = \int_{\mathbb{B}}\lambda(dx)(e^{irx} - 1 - irx) \quad r \in \mathbb{R}$

Chapter 18

Brownian Motion

♢ **Observation 18.1** (So far and setting). *We will work on a probability space* $(\Omega, \mathcal{H}, \mathbb{P})$*, and consider:*

- *Wiener processes (Def. 11.55), martingales (Prop. 11.62), with stationary independent Gaussian increments*
- *Lévy processes (Def. 17.1), right continuous, starting at zero, with stationary independent increments*

Recall also that a Wiener process is stable of order 2*, meaning:*

$$W_t \overset{d}{=} \sqrt{t} W_1, \quad W_1 \sim \mathcal{N}(0,1), \quad W_{ut} \overset{d}{=} \sqrt{u} W_t \quad \forall u, t \in \mathbb{R}_+$$

♠ **Definition 18.2** (Brownian motion). *A process* $X = (X_t)_{t \in \mathbb{R}_+}$ *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *such that:*

1. *the path* $t \to X_t$ *is continuous*
2. *it has stationary and independent increments*

♢ **Observation 18.3** (Connection Brownian-Lévy). *straightforward:*

$$(X_t)_{t \in \mathbb{R}_+} \text{ Brownian} \implies (X_t - X_0)_{t \in \mathbb{R}_+} \quad \text{continuous Lévy}$$

But could we say more?

18.1 A Different Perspective of Wiener Processes

♣ **Theorem 18.4** (Lévy characterization as Wiener). *As a first step, notice that:*

1. $X_t = at + bW_t$ *continuous Lévy* $\implies W_t$ *Wiener*
2. W_t *Wiener* $\implies X_t = at + bW_t$ *continuous Lévy*

Which establish an $\iff$ *relation*

Proof. **(Claim #1)** Theorem D.1.
(Claim #2) Example 17.6 □

Corollary 18.5 (Applying Theorem in the Brownian-Wiener-Lévy context). *Combine the results to obtain:*

$$X_t \text{ Brownian} \overset{Obs.\ 18.3}{\implies} X_t - X_0 \text{ Lévy} \overset{Thm.\ 18.4}{\iff} X_t - X_0 = bt + cW_t \ : \ W_t \text{ Wiener}$$

♠ **Definition 18.6** (Brownian motion decomposition). *We build a Brownian motion from Definition 18.2 as:*

$$X_t = X_0 + bt + cW_t$$

for a drift coefficient b*, a volatility coefficient* c *and a Wiener process* W*.*

♠ **Definition 18.7** (Wiener process as Brownian motion revisited). *According to our results, a Brownian motion $W = (W_t)_{t\in\mathbb{R}_+}$ with $W_0 = 0, \mathbb{E}[W_t] = 0, V[W_t] = t$ (namely, $X_0 = 0, b = 0, c = 1$) is also a Wiener process! We will see in the next result that a Wiener process is a Gaussian process with continuity. However, while constructing a Gaussian process is immediate, as it is only required to specify the functions m, K (see Def. 10.61), it is not granted that there exists a probability space where such process is continuous. We will eventually see that this condition is satisfied, but the question at the moment is proving that Wiener processes exist in the Brownian formulation.*

Lemma 18.8 (Gaussian Transformation linearity). *Quickly recall that:*

$$X \sim \mathcal{N}^d(\mu, \Sigma), \quad A \in \mathbb{R}^p \times \mathbb{R}^d \implies Y = AX \sim \mathcal{N}^d(A\mu, A\Sigma A^T)$$

♢ **Observation 18.9** (Wiener Gaussianity). *By the very definition of a Wiener process W we have that:*

- $W_{s+t} - W_s \perp\!\!\!\perp \mathcal{F}_s$ *(Markovianity)*
- $\mathcal{P}[(W_{s+t} - W_s) \in B] = \mathcal{P}[W_t \in B] = \int_B dx \frac{e^{-\frac{x^2}{2t}}}{\sqrt{2\pi t}}$ *(gaussianity)*

So for distinct times $0 < t_1 < t_2 < \cdots < t_n$ the joint distribution is:

$$(W_{t_1}, \ldots, W_{t_n}) \sim \mathcal{N}^n(\boldsymbol{0}, \Sigma) \quad \Sigma = \{t_i \wedge t_j\}_{ij}$$

Since we use Lemma 18.8 with:

$$A = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ -1 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & 0 & -1 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

Basically a triangular matrix, so that:

$$\begin{bmatrix} t_1 \\ t_2 - t_1 \\ \vdots \\ t_n - t_{n-1} \end{bmatrix} = A \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_{n-1} \end{bmatrix}$$

and we can check that $A\Sigma A^T$ returns the desired result.

♢ **Observation 18.10** (Gaussian Wienerity). *the process $(W_{\boldsymbol{t}}) = (W_{t_1}, \ldots, W_{t_n})$ Gaussian, with null expectation $\mathbb{E}[W_{t_i}] = 0 \forall i$ and Covariance of the form $CoV[W_{t_i}, W_{t_j}] = t_i \forall i \leq j$ can be used to recover a Wiener process making use of A^{-1} where A is from the above Observation and Lemma 18.8 so that we get and independecy:*

$$(W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}}) \quad \perp\!\!\!\perp \text{ mutually} \quad W_{t_i} - W_{t_{i-1}} \sim \mathcal{N}(0, t_i - t_{i=1})$$

behaving as a Wiener process.

♣ **Theorem 18.11** (Wiener-Gaussian characterization). *The previous observations suggest a useful conclusion. For $W = (W_t)_{t\in\mathbb{R}_+}$ a process on $\mathbb{R}$ we establish:*

$$W \text{ Wiener} \iff \begin{cases} W & \text{continuous} \\ W \sim \mathcal{GP}(m, k) & \text{Gaussian Def. 10.61} \\ E[W_t] = 0, CoV[W_s, W_t] = s \wedge t \end{cases}$$

where $m(t) \equiv 0$ and $k(s, t) = s \wedge t$.

Proof. Observations 18.9, 18.10. □

♢ **Observation 18.12** (About the Theorem). *Does Brownian motion exist at all? We have not yet assessed continuity in the Gaussian process, as we only have seen that expectation and variance coincide, up to transformation. So far, we know the easiest instance of Brownian motion, with the adjusted Definition of Wiener process (Def. 18.7) should work. We are missing a formal continuity of the path proof.*

Lemma 18.13 (Kolmogorov's maximal inequality). *Assume $\{X_i\}_{i=1}^n$ is an independency where $\mathbb{E}[X_i] = 0 \forall i$. Setting $S_n = \sum^n X_i$:*

$$a^2 \mathbb{P}\left[\max_{k \leq n} |S_k| > a\right] \leq V[S_n]$$

Proof. Lemma D.2. □

♠ **Definition 18.14** (Recap about Dyadic rationals). *Dyadic rationals are also discussed in Lemma A.17. Here we denote them as:*

$$D = \{x \in \mathbb{R}_+ \ : \ x = k2^{-m}, \ k, m \in \mathbb{N}\}$$

♣ **Proposition 18.15** (Dyadics are dense in $\mathbb{R}$). *This is a very important result. It is reported here to reference it when needed.*

$$\forall t \in \mathbb{R}, \forall \epsilon > 0 \quad \exists k, m \in \mathbb{N} : t \in (k2^{-m}, (k+1)2^{-m}], \ t - k2^{-m} < \epsilon$$

which is the exact definition of dense set.

Proof. Lemma D.3. □

♣ **Theorem 18.16** (Wiener Process properties, Brownian formulation). *Let $W = (W_t)_{t \in \mathbb{R}_+}$ be a Wiener process according to Definition 18.7. Then:*

1. *(symmetry) $(-W_t)_{t \in \mathbb{R}_+}$ is Wiener*
2. *(scaling) $(W_{ct})_{ct \in \mathbb{R}_+}$ Wiener $\implies \widehat{W} = (\sqrt{c} W_t)_{t \in \mathbb{R}_+}$ is Wiener $\forall c \in (0, \infty)$*
3. *(time inversion) setting $\widetilde{W}_0 = 0$ for convention, the process $\widetilde{W}_t = t W_{\frac{1}{t}}$ is Wiener*

Proof. **(Claim #1)** trivial by Definition 18.7.
(Claim #2) holds if and only if we have stability of order 2, something verified by Definition 18.7.
(Claim #3)(△ $t > 0$) for $t > 0$ the process $\widetilde{W}_t$ is continuous and such that $\mathbb{E}\left[\widetilde{W}_t\right] = \mathbb{E}\left[t W_{\frac{1}{t}}\right] = 0$. Moreover:

$$\begin{aligned} CoV\left[\widetilde{W}_s, \widetilde{W}_t\right] &= CoV\left[s W_{\frac{1}{s}}, t W_{\frac{1}{t}}\right] \\ &= st CoV\left[W_{\frac{1}{s}}, W_{\frac{1}{t}}\right] && \text{bilinearity} \\ &= st\left(\frac{1}{s} \wedge \frac{1}{t}\right) && \text{Thm. 18.11} \\ &= st \cdot \begin{cases} \frac{1}{s} & t < s \\ \frac{1}{t} & s < t \end{cases} \\ &= s \wedge t \end{aligned}$$

So that $\widetilde{W}_t$ satisfies Theorem 18.11 for $t > 0$.
(□ $t = 0$) wts:

$$\lim_{t \to 0} t W_{\frac{1}{t}} = \lim_{t \to 0} \widetilde{W}_t \overset{a.s.}{=} 0 \iff \lim_{t \to \infty} \frac{1}{t} W_t \overset{a.s.}{=} 0$$

(□○ **subpoint,** $n \in \mathbb{N}$ **argument**) by the SLLN:

$$W_n \overset{d}{=} Z_1 + \ldots + Z_n \quad Z_i \overset{iid}{\sim} \mathcal{N}(0,1) \implies \lim_{n \to \infty} \frac{1}{n} W_n = \mathbb{E}[W_1] \overset{a.s.}{=} 0$$

(□▽ **subpoint,** $t \in \mathbb{R}_+$ **argument)** let $n < t \leq n+1$ so that $\frac{1}{t} < \frac{1}{n}$. Then:

$$\begin{aligned} \left|\frac{1}{t} W_t\right| \leq \frac{1}{n}\left|W_t\right| &= \frac{1}{n}\left|W_n + W_t - W_n\right| \\ &\leq \frac{1}{n}|W_n| + \frac{1}{n}|W_t - W_n| \\ &= \underbrace{\frac{1}{n}|W_n|}_{\overset{n\to\infty}{\to} 0 \text{ by } \square\bigcirc} + \frac{1}{n}\sup_{s\in[0,1]}|W_{n+s} - W_n| \end{aligned}$$

So the condition becomes:

$$wts \quad \frac{1}{n}\sup_{s\in[0,1]}|W_{n+s} - W_n| \overset{n\to\infty}{\to} 0 \quad a.s. \tag{$\star$}$$

we denote it as $\star$.
(♣ **objects involved)** we will use Kolmogorov's maximal inequality (Lem. 18.13) and the BC1 (Thm. 9.6), which we recall below:

$$\sum_n \mathbb{P}[|X_n| > \epsilon] < \infty \quad \forall \epsilon > 0 \implies X_n \overset{a.s.}{\to} 0$$

(♠ sup **discretization)** let $X_k = W_{n+k2^{-m}} - W_{n+(k-1)2^{-m}}$ for $k = 1, \ldots, 2^m$, where $m \in \mathbb{N}, m \geq 1$. This is a dyadic construction as Definition 18.14. As $m \to \infty$ for a finer and finer grid of subintervals we let:

$$\begin{cases} S_k = X_1 + \ldots X_k = W_{n+2^{-m}k} - W_n \\ S_{2^m} = W_{n+1} - W_n \\ a = \epsilon n \end{cases}$$

observe that $X_1, \ldots, X_k$ are independent and have mean zero by construction. Then:

$$\begin{aligned} \mathcal{P}\left(\frac{1}{n}\max_{0<k<2^m}|W_{n+2^{-m}k} - W_n| > \epsilon\right) &\leq \frac{1}{n^2\epsilon^2} V\left[\underbrace{|W_{n+1} - W_n|}_{=S_{2^m}}\right] \\ &\leq \frac{1}{n^2\epsilon^2}\cdot 1 \qquad \forall m \end{aligned}$$

By Lemma 18.13 and the variance being unitary as the process is Wiener.
(♡ **finalization)** By Proposition 18.15 as $m \to \infty$ it holds that max $\approx$ sup monotonically and with arbitrary precision:

$$\max_{0<k<2^m}|W_{n+2^{-m}k} - W_n| \uparrow \sup_{s\in[0,1]}|W_{n+s} - W_n|$$

which is granted by the map $t \to W_t$ being right continuous, since W is also a Lévy process (Theorem 18.4) and Lévy processes are right continuous (Def. 17.1).
We are in the position to apply monotone convergence (Thm. 4.21) to conclude that:

$$\begin{aligned} \mathcal{P}\left(\frac{1}{n}\sup_{s\in[0,1]}|W_{n+s} - W_n| > \epsilon\right) &= \mathcal{P}\left(\frac{1}{n}\lim_{m\to\infty}\max_{0<k<2^m}|W_{n+2^{-m}k} - W_n| > \epsilon\right) && \text{Prop. 18.15} \\ &= \lim_{m\to\infty}\mathcal{P}\left(\frac{1}{n}\max_{0<k<2^m}|W_{n+2^{-m}k} - W_n| > \epsilon\right) && \text{Mon. conv Thm. 4.21} \\ &\leq \frac{1}{n^2\epsilon^2} && \text{above inequality} \end{aligned}$$

Recalling the classic identity $\sum_n \frac{1}{n^2} = \frac{\pi^2}{6}$ we eventually have that:

$$\sum_n \frac{1}{n^2\epsilon^2} < \infty \quad \forall \epsilon > 0 \overset{\text{BC1 (Thm. 9.6)}}{\Longrightarrow} \frac{1}{n}\sup_{s\in[0,1]}|W_{n+s} - W_n| \overset{a.s.}{\to} 0$$

where in particular we used the interpretation of Borel Cantelli for almost sure convergence (see Example 9.9). □

18.2 Continuity

In the previous proof we make use of right continuity only. We are missing the continuity stated in Definition 18.7 for Wiener processes, which would ensure existance of the process $W = (W_t)_{t\in\mathbb{R}_+}$. We do so by Kolmogorov extension. For a finite collection of ordered times $\{t_i\}_{i=1}^n$ the multivariate Gaussian with mean zero and covariance $\{t_i \wedge t_j\}$ is denoted as $\mu_{t_1,\dots,t_n}$. Such distribution is a consistent family, in the sense of Definition 10.54. For $\Omega = \{\omega : \mathbb{R}_+ \to \mathbb{R}\}$ a set of functions put:

$$W_t(\omega) = \omega(t) \quad \mathcal{H} = \sigma\{\omega \in \Omega : \omega(t) \in A, A \in \mathcal{B}(\mathbb{R})\}$$

Using Kolmogorov extension Thm. 10.55:

$$\exists!\mathbb{P} \; on \; (\Omega, \mathcal{H}) \quad s.t. \quad \mathbb{P}[\omega : \omega(t_i) \in A_i \; \forall i = 1,\dots,n] = \mu_{t_1,\dots,t_n}(A_1 \times \dots \times A_n)$$

Yet to evaluate continuity, we would need to check the paths $t \to \omega(t)$ at uncountable time points, while $\mathcal{H}$ is constructed from countably many coordinates. The collection of continuous functions is not a priori measurable. What we do is a **modification** of the process $(W_t)_{t\in\mathbb{R}_+}$ into $\left(\widetilde{W}_t\right)_{t\in\mathbb{R}_+}$ so that:

$$\forall t \exists \Omega_t : W_t = \widetilde{W}_t, \quad \widetilde{W}_t \; continuous$$

where Ω_t is an almost sure set.
Precisely, we establish Hölder continuity instead of continuity.

♠ **Definition 18.17** (Uniform continuity). *the map $t \to W_t$ is uniformly continuous if:*

$$\forall \epsilon > 0 \exists \delta = \delta(\epsilon) \quad |t-s| < \delta \implies |W_t - W_s| < \epsilon \quad \forall t, s$$

Where δ depends on ϵ only.

♠ **Definition 18.18** (Hölder continuity). *The path $t \to W_t$ is Hölder continuous when:*

$$\exists C \in \mathbb{R} \quad |W_t - W_s| \le C|t-s|^\alpha, \quad for \; \alpha \in [0,1] \quad \forall t, s$$

Which means uniform continuity for $\delta = (\frac{\epsilon}{C})^{\frac{1}{\alpha}}$.

♢ **Observation 18.19** (About Hölder continuity). *At $\alpha = 1$ we recover Lipschitz continuity, which guarantees continuous and bounded derivatives, and differentiability.*

Lemma 18.20 (Kolmogorov moment condition). *For a process $X = (X_t)_{t\in[0,1]}$ on $\mathbb{R}$ and D the dyadic set of Definition 18.14 if:*

$$\exists c, p, q \in (0,1) \quad : \quad \mathbb{E}[|X_t - X_s|^p] \le c|t-s|^{\frac{1}{q}} \quad \forall s, t \in [0,1]$$

Then:

1. $\forall \alpha \in [0, \frac{q}{p}) \exists K$ *r.v. such that:*
   (a) $\mathbb{E}[K^p] < \infty$
   (b) $|X_t - X_s| \le k|t-s|^\alpha$ *for $s, t \in D$*
2. *if X is continuous then #1 holds $\forall s, t \in [0,1]$*

Proof. **(Claim #1)(△ K aim)** for $\alpha \in \left[0, \frac{q}{p}\right)$ we let:

$$K := \sup_{s,t\in D, s\ne t} \frac{|X_t - X_s|}{|t-s|^\alpha}$$

where by $D \times D$ being countable K is a random variable.
(□ proving #1(b) and #2) by the way K is defined, and the fact that D are dense in $\mathbb{R}$ (Prop. 18.15) we can easily conclude that both hold in the genereal setting.
(◯ proving #1(a)) wts $\mathbb{E}[K^p] < \infty$

(▽ **dyadic argument**) let $M_n = \sup_{D_n} |X_t - X_s|$ where $D_n = \{s,t : t-s = 2^{-n}\}$, by the equatio in the hypothesis we can say that:

$$\mathbb{E}\left[M_n^p\right] \leq 2^n \cdot c \cdot (2^{-n})^{1+q} = c2^{-nq}$$

since $2^n = |D_n|$.
(♠ **general** D) let $s,t \in D$ and:

$$s_n = \inf\{D_n \cap [s,1]\}, \quad t_n = \inf\{D_n \cap [0,t]\}$$

trivially $s_n \searrow s, t_n \nearrow t$ and for n large both are equal to their limits. Then:

$$X_t - X_s = \sum_{n\geq m} \left(X_{t_{n+1}-X_{t_n}}\right) + X_{t_m} - X_{s_m} + \sum_{n\geq m}\left(X_{s_n} - X_{s_{n+1}}\right)$$

with this setting, if $0 < t-s < 2^{-m}$ either $t_m - s_m = 0$ or $t_m - s_m = 2^{-m}$. We can then derive the bound:

$$|X_t - X_s| \leq \sum_{m\geq n} M_{n+1} - M_m + \sum_{n\geq m} M_{n+1} \leq 2\sum_{n\geq m} M_n \tag{18.1}$$

(♣ **final computation**) for K as in ▽, take the sup over s,t with the condition that $2^{-m-1} < |t-s| \leq 2^{-m}$. Then, the sup condition can be seen in terms of m and by Equation 18.1 we get that:

$$\begin{aligned} K &\leq \sup_m (2^{m+1})^\alpha \cdot 2\sum_{n\geq m} M_n && (2^{-m-1})^{-1} = ((t-s)^\alpha)^{-1} \text{ in } \triangle \\ &\leq 2^{1+\alpha}\sum_{n\geq 0} 2^{n\alpha} M_n \end{aligned}$$

for $p \geq 1$ and the $\mathcal{L}_p$ norm denoted as $\|\cdot\|$ we eventually get that:

$$\begin{aligned} \|K\| &\leq \left\| 2^{1+\alpha}\sum_{n\geq 0} 2^{n\alpha} M_n \right\| \\ &\leq 2^{1+\alpha}\sum_{n\geq 0} 2^{n\alpha} c^{\frac{1}{p}} 2^{-\frac{nq}{p}} \\ &= 2^{1+\alpha} c^{\frac{1}{p}} \sum_{n\geq 0} 2^{n\alpha - \frac{nq}{p}} && n\alpha - \frac{nq}{p} = n\left(\alpha - \frac{q}{p}\right) \leq 0 \text{ hypothesis} \\ &< \infty \end{aligned}$$

□

◇ **Observation 18.21** (What is missing?). *Recall Definition 18.7 and Observations 18.9, 18.10. We* ***need*** *to show that the map $t \to W_t$ is continuous. Notice that we can also use the result of Theorem 18.16 since in the proof we only use the right continuity of the process, which is granted by the Wiener-Lévy connection (Thm. 18.4), and does not use assumptions brought by the Brownian formulation.*

♣ **Theorem 18.22** (Brownian continuity of Wiener process). *We prove the local Höldercontinuity of the paths of a Wiener process, closely linked to the Gaussian process (see Thm. 18.11). This result is parallel to the existance of a p.r.m. (Thm. 14.8).*
(Def. 10.61) If $W = (W_t)_{t\in\mathbb{R}_+}$ is a Wiener process with:

- $\mathbb{E}[W_{t_i}] = 0 \forall i$
- $CoV[W_{t_i}, W_{t_j}] = t_i \wedge t_j \quad \forall i,j$

Then there exists an almost sure version of W that is locally Hölder continuous:

$$t \to \widetilde{W}_t \quad \textit{Hölder continuous} \quad W \overset{a.s.}{=} \widetilde{W}$$

Proof. Let $0 < t_1 < \cdots < t_n$, and consider the law $\mu_{t_1,\ldots,t_n}$, which is by hypothesis Gaussian.
(△ **extension by consistency**) we first build the space. $\mu_{t_1,\ldots,t_n}$ is Kolmogorov consistent in the sense of

Definition 10.54. Additionally let:

$$\begin{cases} \Omega = \{\omega : \mathbb{R}_+ \to \mathbb{R}\} \\ W_t(\omega) = \omega(t) & \omega \in \Omega, t \geq 0 \\ \mathcal{H} = \sigma\left(\{\omega \in \Omega : \omega(t) \in A,\ A \in \mathcal{B}(\mathbb{R})\}\right) \end{cases}$$

where the algebra is finitely generated. By Kolmogorov Extension (Thm. 10.55) there exists a unique $\mathbb{P}$ on $(\Omega, \mathcal{H})$ such that:

$$W_{t_1}, \ldots, W_{t_n} \sim \mu_{t_1,\ldots,t_n} \quad \mathbb{P}[\omega : \omega(t_i) \in A_i,\ \forall i = 1, \ldots, n] = \mu_{t_1,\ldots,t_n}(A_1 \times \cdots \times A_n)$$

(□ a problem) the set $\{t \to \omega(t) : \textit{continuous}\} = \mathcal{C}$ is not measurable wrt $\mathcal{H}$ since $H \in \mathcal{H}$ is generated by countably many coordinates and we cannot say $\mathbb{P}[\mathcal{C}] = 1$ directly (i.e. we cannot check all continuity points).
(▽ redirected aim) given □ we consider finding $\widetilde{W}_t$, a version of W_t such that:

$$\forall t \quad \exists \Omega_t : \mathbb{P}[\Omega_t] = 1 \quad \widetilde{W}_t = W_t,\ \widetilde{W}_t \textit{ continuous}$$

(♣ first facts) by scaling and stationarity of $W = (W_t)_{t \in \mathbb{R}_+}$ we recover for all $s, t \in (0, 1)$ and $m \geq 1$ that:

$$\begin{aligned} \mathbb{E}\left[|W_t - W_s|^m\right] &\leq \mathbb{E}\left[|W_{t-s}|^{2m}\right] && \text{stationarity} \\ &= \mathbb{E}\left[|t-s|^m |W_1|^{2m}\right] && W_1 \overset{d}{=} \sqrt{t-s} W_{t-s} \\ &= |t-s|^m \mathbb{E}\left[|W_1|^{2m}\right] \\ &= |t-s|^m C_m \end{aligned}$$

Where C_m exists since the Gaussian process W_t always has moments.
(♠ Kolmogorov moment) use Lemma 18.20 with $p = 2m, 1 + q = m$ to get:

$$|W_t - W_s| \leq K|t-s|^\alpha \quad \forall s, t \in D \cap [0, 1]$$

Where $K \in \mathcal{L}_p$ and $\alpha \in [0, \frac{m-1}{2m})$.
For $\Omega_0 \coloneqq \{K < \infty\}$ we have that $\mathbb{P}[\Omega_0] = 1$ and W_t is Hölder continuous according to Definition 18.18.
(♡ defining $\widetilde{W}_t$) we let

$$\widetilde{W}_t = \begin{cases} 0 & \omega \notin \Omega_0 \\ \lim_{s \to t} W_s(\omega) & t \in [0, 1], \omega \in \Omega_0 \end{cases}$$

such process, as per ♠, is Hölder continuous for all $\omega \in \Omega$ and for α arbitrary close to $\frac{1}{2}$ by the fact that $[0, \frac{m-2}{2m}) \overset{m \to \infty}{\to} [0, \frac{1}{2})$.
We conclude that $\forall t \in [0, 1]$ the process $W_t = \widetilde{W}_t$ is almost surely continuous by the construction of the Ω_0 almost sure set.
(𓄿 iteration) use the previous steps repeatedly to build a series of chained $(W_t)_{t \in [n, n+1]}$ that respect the relation $\forall n \in \mathbb{N}$. By $\bigcup_n [n, n+1) = \mathbb{T} = \mathbb{R}_+$ we can safely say that $(W_t)_{t \in \mathbb{T}}$ is a Wiener process in the sense of Definition 18.7, since its Gaussianity already implies continuity, and we can apply Theorem 18.11. □

◇ **Observation 18.23** (About further properties and conclusion). *The results we have shown are:*

- *Wiener implies Gaussian continuous, Theorem 18.11*
- *Wiener is continuous, Theorem 18.22*
- *Brownian motion automatically exists, via Definition 18.7*

So we could see a Brownian motion as:

$$\underbrace{X_t}_{Brown} = X_0 + bt + c \underbrace{W_t}_{Wien}, \quad X_t - X_0 \textit{ Lévy}$$

Existsance is a consequence of the very existance of continuous Wiener processes in the Brownian formulation. It is also possible to prove that the path is nowhere differentiable (i.e. very wiggly). This is somewhat hinted by the $\alpha \leq \frac{1}{2}$ Hölder continuity rate, which is less than Lipschitz continuity (and thus differentiability).

Chapter Summary

Objects:

- Brownian motion is a process on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that:
  - $t \to X_t$ is continuous
  - Markovian and stationary increments
  - decomposed as $X_0 + bt + cW_t$ where W_t is a Wiener process, assumed to be continuous
- uniform continuity of a map $t \to W_t$

$$\forall \epsilon > 0 \exists \delta = \delta(\epsilon) \quad |t - s| < \delta \implies |W_t - W_s| < \epsilon$$

- Hölder continuity of a map $t \to W_t$

$$\exists C \in \mathbb{R} \quad |W_t - W_s| \leq C|t - s|^{\alpha}, \quad for\ \alpha \in [0, 1]$$

Results:

- $(X_t)_{t\in\mathbb{R}_+}$ Brownian $\implies (X_t - X_0)_{t\in\mathbb{R}_+}$ is Lévy
- $X_t = at + bW_t$ continuous Lévy $\iff W_t$ Wiener (assuming it is continuous)
- $W = (W_t)_{t\in\mathbb{R}_+}$ Wiener $\iff W$ is continuous Gaussian Process with $\mathbb{E}[W_{t_i}] = 0, CoV[W_{t_i}, W_{t_j}] = t_i \wedge t_j$
- dyadics are dense in $\mathbb{R}$
- Wiener process properties:
  - (symmetry) $(W_t)_{t\in\mathbb{R}_+}$ Wiener $\implies (-W_t)_{t\in\mathbb{R}_+}$ is Wiener
  - (scaling) stable of order 2
  - (time inversion) with $\widetilde{W}_0 = 0$, it holds $\widetilde{W}_t = tW_{\frac{1}{t}}$ is Wiener
- Wiener process is Hölder continuous, thus Brownian motion exists

Chapter 19

Arcsine laws, Hitting times

19.1 Augmentations and Hitting Times

♠ **Definition 19.1** (Right continuous augmentation of filtration). *For this object and its properties, refer also to Appendix D, especially from Definition D.4 onwards.*
For a filtration $\left(\mathcal{F}_t^0\right)_{t\in\mathbb{R}_+}$ we let it be generated by the process itself. Namely $\left(\mathcal{F}_t^0\right)_{t\in\mathbb{R}_+} = \sigma(\{X_s : s \le t\})$. In this context, we define the right continuous augmentation as:

$$(\mathcal{F}_t)_{t\in\mathbb{R}_+} \quad : \quad \forall t\ \mathcal{F}_t = \bigcap_{s\ge 0} \mathcal{F}^0_{t+s} = \lim_{s\downarrow 0} \mathcal{F}^0_{t+s}$$

Since $\left(\mathcal{F}_t^0\right)_{t\in\mathbb{R}_+}$ is increasing by Definition of filtration. This new filtration can be interpreted as a peek into the future. From now on, when referring to $\mathcal{F}$ it will be the ***augmented filtration****.*

♠ **Definition 19.2** (Hitting time of barrier a, T_a). *For a random time $T : \Omega \to \overline{\mathbb{R}}_+$ we define:*

$$T_a \coloneqq \inf\{t \ge 0 : X_t > a\}$$

Namely, the entrance time in the interval (a, ∞). Notice that we intuitively assume T_a to be almost surely finite for a Wiener process.

♣ **Theorem 19.3** (Hitting times are stopping in augmentation). *Let $\mathcal{F}$ be augmented as in Definition 19.1, a process X on E be right continuous and adapted to $\mathcal{F}$. Then:*

$$\forall B \in \mathcal{B}(E) \quad T_B = \inf\{t \in \mathbb{R}_+ : X_t \in B\} \quad \textit{stopping time wrt } \mathcal{F}$$

According to the usual stopping time knowledge (Def. 11.9).

Proof. More context is given in Appendix D, especially from Theorem D.8 onwards. □

♠ **Definition 19.4** (Shift operator $\theta_\cdot$). *For a collection of continuous maps $\mathscr{C} = \{t \to w(t) \mid \textit{continuous}\}$ we define an operator:*

$$\theta_s : \mathscr{C} \to \mathscr{C},\ s \in \mathbb{R} \quad (\theta_s \circ w)(t) \coloneqq w(s+t)$$

♣ **Theorem 19.5** (Markov property of Lévy processes). *This is Theorem VII.3.5 in [Çin11].*
For $X = (X_t)_{t\in\mathbb{R}_+}$ a Lévy process (Def. 17.1), for any time t, the process $X \circ \theta_t$ is independent of $\mathcal{F}_t$ and has the same law of X. Equivalently:

$$\mathbb{E}_t\left[V \circ \theta_t\right] = \mathbb{E}\left[V\right] \quad t \in \mathbb{R}_+, \quad \forall V \in \mathcal{G}_\infty \textit{ bounded}$$

where the boundedness of V is used to ensure existance of the expectation, but the result is extended to positive or integrable random variables in $\mathcal{G}_\infty$ the underlying end of time filtration generated by the process.
In terms of Wiener processes, which are Lévy by Theorem 18.4 for $a = 0, b = 1$ we have that:

$$\forall s \ge 0 \quad (W \circ \theta_s) = \left(\widetilde{W}_t\right)_{t\in\mathbb{R}_+} = (W_{s+t} - W_s)_{t\in\mathbb{R}_+} \perp\!\!\!\perp \mathcal{F}_s \quad \textit{Wiener law}$$

Proof. **(△ comments only)** recall that a Wiener process is Lévy by Theorem 18.4, then [Çin11] Theorem VII.3.5 states exactly this for Lévy processes. □

♣ **Proposition 19.6** (Conciliating time shifts to Wiener and Brownian). *While a Wiener process (on a stochastic base [Çin11]) is such that:*

$$W_t \circ \theta_s = W_{s+t} - W_s$$

A Brownian motion as in Definition 18.2 is such that:

$$X_t \circ \theta_s = X_{t+s}$$

Proof. For $X_t = X_0 + W_t$ taken for simplicity, we have that:

$$\begin{aligned}
X_{s+t} &= \underbrace{X_{s+t} - X_s}_{\perp\!\!\!\perp \mathcal{F}_s} + X_s && \text{Markov Thm. 19.5} \\
&= X_0 + W_{s+t} - X_0 - W_s + X_s && \text{independent terms explicit} \\
&= W_{s+t} - W_s + X_s && \\
&\overset{d}{=} \widetilde{W}_t + X_s && \widetilde{W}_t = W_{s+t} - W_s = W_t \circ \theta_s
\end{aligned}$$

□

♢ **Observation 19.7** (θ_s interpretation). *We have that:*

$$X_{s+t} = \begin{cases} X_t \circ \theta_s & \text{as a function of } X_t \\ (\theta_s \circ X)(t) & \text{as } X \text{ a random function, a process} \end{cases}$$

So that:

$$\begin{aligned}
X_{s+t} = X_t \circ \theta_s &= (X_0 + W_t) \circ \theta_s && \textit{easiest Brownian form} \\
&= X_0 \circ \theta_s + W_t \circ \theta_s && \textit{linearity} \\
&= X_s + W_{s+t} - W_s && \textit{Prop. 19.6} \\
&= X_s + \widetilde{W}_t && \textit{makes sense}
\end{aligned}$$

♡ **Example 19.8** (Recurrence times). *Define G_t as the last time at zero before t and D_t as the first time at zero after t. Namely:*

$$G_t := \sup\{s \in [0,t] : W_s = 0\} \quad D_t = \inf\{u \in (t,\infty) : W_u = 0\}$$

Accordingly, the forward recurrence time is $R_t = D_t - t$ and the backward recurrence time is $Q_t = t - G_t$. By the process $(W_t)_{t\in\mathbb{R}_+}$ being such that $W_t \sim \mathcal{N}(0,t) \forall t$ we have $\mathbb{P}[W_t = 0] = 0$ a.s. by the diffusivity of a normal distribution. Then:

$$G_t < t < D_t \text{ a.s.} \quad \& \quad s \in [0,t] \implies \{G_t < s\} = \{D_s > t\}$$

an intuition is given in Figures 19.1, 19.2. So, if $W_t = a > 0 \implies R_t$ is the hitting time from above of the barrier $-a$ of the rescaled process:

$$(W_u \circ \theta_t)_{u\geq 0} = (W_{t+u} - W_t)_{u \geq 0}$$

By the markov property of Wiener processes $\widetilde{W}_u = W_{t+u} - W_t \perp\!\!\!\perp \mathcal{F}_t$ is again Wiener and we can see that:

$$R_t = \inf\{u > 0 : \widetilde{W}_u < -a\}$$

Additionally, by symmetry (Thm. 18.16#1) we have:

$$\widetilde{W}_u < -a \iff -\widetilde{W}_u > a \iff \widetilde{W}_u > a$$

So that:

$$R_t \overset{d}{=} T_a = \inf\{u > 0 : W_u \circ \theta_t = \widetilde{W}_u > a\} \quad a = W_t,\ T_a \perp\!\!\!\perp W_t$$

Which means that if T_a is known $\forall a > 0 \implies R_t$ is known and so is $D_t = R_t + t$ and G_t via $\mathbb{P}[G_t < s] = \mathbb{P}[D_s > t]$, as well as $Q_t = t - G_t$.

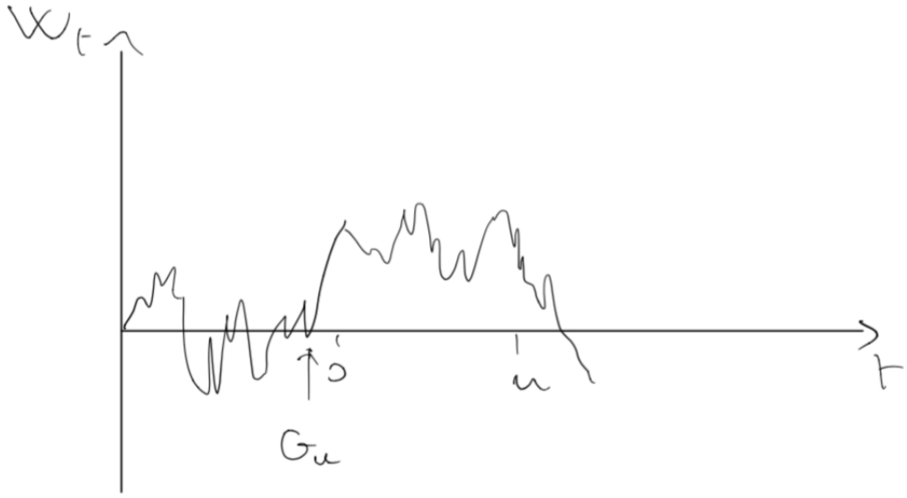

Figure 19.1: Recurrence times of a Wiener process

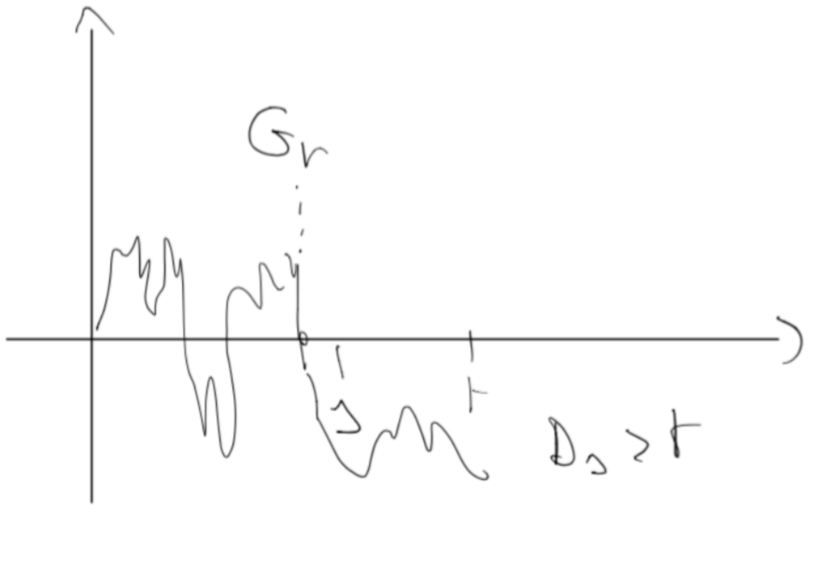

Figure 19.2: Recurrence times of a Wiener process

Corollary 19.9 (Blumenthal's 0-1 law). *It holds that:*

$$\mathcal{F} \quad \textit{augmented, right continuous} \implies \forall A \in \mathcal{F}_0 \ \mathbb{P}[A] \in \{0,1\}$$

Namely, a suitable augmented filtration makes every event in the infinitesimal peek in the future at the start either certain or null.

Proof. Corollary D.18 □

♣ **Proposition 19.10** (Hitting zero almost surely). *This result is like that of Proposition 16.8. For* $T_0 = \inf\{t > 0 : W_t > 0\}$ *as in Definition 19.2 we have:*

$$T_0 = 0 \quad a.s.$$

Proof. We have that $\mathcal{F}_0$ is right continuous (Def. 19.1). An application of Corollary 19.9 gives:

$$\{T_0 = 0\} \in \mathcal{F}_0 \implies \mathbb{P}[\{T_0 = 0\}] \in \{0,1\}$$

Notice that $\mathbb{P}[\{W_t > 0\}] = \frac{1}{2}$ for all $t > 0$ since the Gaussian distribution is centered at mean zero. By Theorem 19.3 it holds

$$\{W_t > 0\} \subset \{T_0 < t\} \implies \mathbb{P}[\{T_0 < t\}] > 0 \quad \forall t > 0$$

And as $t \to 0$ it holds that:

$$\mathbb{P}[\{T_0 = 0\}] > 0 \implies \mathbb{P}[\{T_0 = 0\}] = 1 \quad a.s.$$

□

Corollary 19.11 (Highly oscillatory behavior of W_t at zero). *There are* ∞ *many crossings for any time interval starting from zero. Namely:*

$$for\ a.e.\ \omega\ \exists u_1 > t_1 > s_1, u_2 > t_2 > s_2, \ldots \to 0 \quad s.t. \quad W_{u_n}(\omega) > 0, W_{t_n}(\omega) = 0, W_{s_n}(\omega) < 0$$

Proof. **(△ start)** for all $\epsilon > 0$ there exists $u < \epsilon$ such that $W_u(\omega) > 0$ in Proposition 19.10.
(□ mid) there exists $s < \epsilon$ such that $W_s(\omega) < 0$ by symmetry (Thm. 18.16#1).
By △, □ and continuity of the map $t \to W_t$ (Thm. 18.22) it holds that:

$$\forall \epsilon > 0 \exists 0 < s < t < u \quad : \quad W_s(\omega) < 0,\ W_t(\omega) = 0,\ W_u(\omega) > 0$$

□

Corollary 19.12 (Highly oscillatory behavior of W_t at infinity).

$$for\ a.e.\ \omega \exists u_1 > t_1 > s_1, u_2 > t_2 > s_2, \ldots \to \infty \quad s.t. \quad \lim_{n\to\infty} W_{s_n} = -\infty, \quad \lim_{n\to\infty} W_{u_n} = +\infty$$

Proof. Use Corollary 19.11 and time inversion (Thm. 18.16#3) with $W_t = tW_{\frac{1}{t}}$ to show that :

$$\begin{cases} u_1 > t_1 > s_1 \to 0 \\ W_{u_1}, W_{t_1}, W_{s_1} \end{cases} \iff \begin{cases} \dfrac{1}{u_1} < \dfrac{1}{t_1} < \dfrac{1}{s_1} \\ W_{\frac{1}{u_1}}, W_{\frac{1}{t_1}}, W_{\frac{1}{s_1}} \quad W_{\frac{1}{t}} = \underbrace{\dfrac{1}{t}}_{\to\infty} \underbrace{W_t}_{\in\mathbb{R}} \quad in\ general \end{cases}$$

Giving us that:

$$\begin{cases} W_{\frac{1}{u}} = \underbrace{\dfrac{1}{u}}_{\to\infty} \underbrace{W_u}_{>0} = +\infty \\ W_{\frac{1}{t}} = \underbrace{\dfrac{1}{t}}_{\to\infty} \underbrace{W_t}_{=0} = 0 \\ W_{\frac{1}{s}} = \underbrace{\dfrac{1}{s}}_{\to\infty} \underbrace{W_s}_{<0} = -\infty \end{cases}$$

□

◇ **Observation 19.13** (About the hitting time process $(T_a)_{a\geq 0}$). *Analyze $(T_a)_{a\geq 0}$ as a process. By Proposition 19.10 $T_0 = 0$ a.s. like a Lévy process (Def. 17.1). In the next arguments, we will aim to show that it an increasing stable process (Ex. 15.23) with index $\alpha = \frac{1}{2}$ and $c = \sqrt{2}$. Next we show it is an increasing Lévy process (Def. 15.14) with density:*

$$\lambda(dz) = \frac{1}{\sqrt{2\pi z^3}}dz \quad \text{by } \Gamma\left(1 - \frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$$

So that $(T_a)_{a\geq 0}$ is such that $T_a \in \mathcal{IN}(a)$ with density:

$$f_a(z) = \frac{a}{\sqrt{2\pi z^3}} e^{-\frac{a}{2z}}$$

with the classic result that $\frac{a^2}{Z^2} \sim \mathcal{IN}(a)$ when $Z \sim \mathcal{N}(0,1)$.

♣ **Theorem 19.14** (Strong Markov property of Lévy and Wiener processes). *For T a stopping time wrt $\mathcal{F}^0$ (the **not augmented** filtration) and X a Lévy process we have that:*

1. *$\forall V$ bounded in $\overline{\mathcal{G}}_\infty$:*
$$\mathbb{E}_T[V \circ \theta_T \mathbb{1}_{\{T<\infty\}}] = \mathbb{E}[V \mathbb{1}_{\{T<\infty\}}]$$
independent of $\mathcal{F}_T$ and Lévy
2. *shift operator version:*
$$(X_t \circ \theta_T)_{t\geq 0} = (X_{t+T})_{t\geq 0} \perp\!\!\!\perp \mathcal{F}_T, \text{ Lévy}$$

The claims are also valid for Wiener processes by Theorem 18.4.

Proof. **(Claim #1)** see [Çin11], Theorem VII.3.10 and Theorem VIII.1.13 with the subsequent discussions. The latter is just a reformulation of the former. We know Wiener processes are a special case of Lévy processes.
(Claim #2) is a specific case of #1 provided that $T < \infty$ so that the indicator disappears. □

♣ **Theorem 19.15** (A property of Wiener functions & stopping times). *Assume:*

- *T is stopping wrt $\mathcal{F}$*
- *$U : \Omega \to \mathbb{R}_+$ is such that $U \in \mathcal{F}_T$*
- *$W = (W_t)_{t\in\mathbb{R}_+}$ is a Wiener process*
- *f is a bounded Borel function on $\mathbb{R}$*
- *$g(u) := \mathbb{E}[f \circ W_u], \quad u \in \mathbb{R}_+$*

Then:

$$\mathbb{E}_T\left[f(W_{T+u} - W_T)\mathbb{1}_{\{T<\infty\}}\right] = g(U)\mathbb{1}_{\{T<\infty\}}$$

Proof. **(△ monotone class Theorem)** the functions f for which the claim hold for a monotone class, in the sense of Definition A.19. Thus, we can show the claim holds for $f \in C_b$ and use Theorem A.20.
(□ continuity and bounded convergence) assuming $f \in C_b$, g is necessarily bounded and continuous as well by the continuity of the W (Thm. 18.22) and the bounded convergence Theorem for the expectation (Cor. 4.26).
(◯ simple U) let U be simple, in the sense that $U : \Omega \to D \subset \mathbb{R}_+$ where $|D| < \infty$ is finite. Since $U \in \mathcal{F}_T$ (measurable), the event $\{U = u\}$ is in $\mathcal{F}_T$ for all $u \in D$. We derive:

$$\mathbb{E}_T\left[f(W_{T+U} - W_T)\mathbb{1}_{\{U=u,T<\infty\}}\right] = \mathbb{E}_T\left[f(W_{T+u} - W_T)\mathbb{1}_{\{U=u,T<\infty\}}\right] = g(u)\mathbb{1}_{\{U=u,T<\infty\}}$$

where the last equality is an application of the strong Markov property (Thm. 19.14).
The sum $\sum_{u\in D}$ gives us the desired claim for U simple.
(▽ U arbitrary) for U arbitrary, it holds $U = \lim_{n\to\infty} U_n$ where $(U_n) \subset \mathcal{F}_T$ by assumption. We know that for each n, we can use the result of ◯, we do so taking the limit as $n \to \infty$.

$$\begin{aligned}
\mathbb{E}_T\left[f(W_{T+U} - W_T)\mathbb{1}_{\{U=u,T<\infty\}}\right] &= \lim_{n\to\infty} \mathbb{E}_T\left[f(W_{T+U_n} - W_T)\mathbb{1}_{\{U_n=u_n,T<\infty\}}\right] \\
&\quad [W \text{ is continuous, } f \in C_b\text{, bounded convergence Cor. 4.26}] \\
&= \lim_{n\to\infty} g(u_n)\mathbb{1}_{\{U_n=u_n,T<\infty\}} && \text{by } \bigcirc \\
&= g(u)\mathbb{1}_{\{U=u,T<\infty\}} \\
&= g(U)\mathbb{1}_{\{T<\infty\}}
\end{aligned}$$

□

♣ **Theorem 19.16** (Hitting time (Lévy) distribution by reflection principle)**.** *As per Observation 19.13 we have:*

$$(T_a)_{a\geq 0} \quad : \quad T_a \sim \mathcal{IN}(a)\ \forall a \in \mathbb{R}_+$$

Meaning that $(T_a)_{a\geq 0}$ *has the same distribution of a stable Lévy process for each time point. Notice that we* ***are not proving*** *that it is a Lévy process, something we will show in Theorem 19.25.*

Proof. **(△ premise)** notice that if W_t hits a at $s < t$ then

$$W_t = W_t - W_s + W_s = (W_t - W_{T_a}) + W_{T_a} \quad s = T_a$$

(□ strong Markov) we wish to apply the strong Markov property of Wiener processes (Thm. 19.14). To do so, we first have to reduce $u = t - T_a$ to the right form. This is viable thanks to Theorem 19.15 applied to T_a. We then have that:

$$(\widetilde{W}_u)_{u\geq 0} = (W_u \circ \theta_{T_a})_{u\geq 0} \overset{\text{Prop. 19.6}}{=} (W_{T_a+u} - W_{T_a})_{u\geq 0} \perp\!\!\!\perp \mathcal{F}_{T_a}, W_{T_a}, T_a$$

is Wiener.
(◯ joint) We will prove next that $T_a = T_{a-}$ (Prop. 19.26), where the latter is the time at which the barrier a is hit. This fact, together with the continuity of W (Thm. 18.22), suggest that at the barrier a we have $a = W_{T_a}$. With this in mind, examinate:

$$\begin{aligned}
\mathcal{P}(T_a < t, W_t > a) &= \mathcal{P}(T_a < t, W_{T_a+t-T_a} - W_{T_a} > 0) && \triangle, a = W_{T_a} \\
&= \mathcal{P}\left(T_a < t, \widetilde{W}_{t-T_a} > 0\right) && \square \\
&= \mathcal{P}(T_a < t)\,\mathcal{P}\left(\widetilde{W}_{t-T_a} > 0\right) && \text{independence from } \square \\
&= \mathcal{P}(T_a < t)\,\mathcal{P}\left(\widetilde{W}_{t-T_a} < 0\right) && \text{symmetry Thm. 18.16\#1} \\
&= \mathcal{P}\left(T_a < t, \widetilde{W}_{t-T_a} < 0\right) && \text{independence from } \square \\
&= \mathcal{P}(T_a < t, W_t > a) && \triangle, a = W_{T_a}
\end{aligned}$$

so that paths above and below have the same probability.
(▽ marginal) by $\mathcal{P}(\widetilde{W}_u = 0) = 0 \forall u > 0$ i.e. $\mathcal{P}(W_t = a) = 0 \forall a$ we have that:

$$\mathcal{P}(T_a < t, W_t > a) + \mathcal{P}(T_a < t, W_t < a) = \mathcal{P}(T_a < t)$$

Which by the discussion of ◯ also suggests that:

$$\mathcal{P}(T_a < t) = 2\mathcal{P}(T_a < t, W_t > a)$$

Notice that $W_t > a$ means that the process hits a before t, namely the events are such that $\{W_t > a\} \subset \{T_a < t\}$. Eventually:

$$\begin{aligned}
\mathcal{P}(T_a < t) &= 2\mathcal{P}(W_t > a) \\
&= 2\mathcal{P}(|W_t| > a) && \text{Gaussian symmetry} \\
&= \mathcal{P}\left(W_t^2 > a^2\right) \\
&= \mathcal{P}\left((\sqrt{t}W_1)^2 > a^2\right) && \text{scaling of Wiener} \\
&= \mathcal{P}\left(tZ^2 > a^2\right) \\
&= \mathcal{P}\left(t > \frac{a^2}{Z^2}\right)
\end{aligned}$$

Which implies that $\dfrac{a^2}{Z^2}$ has an inverse Gaussian distribution $T_a \sim \mathcal{IN}(a)$ with density:

$$f_a(z) = \frac{a}{\sqrt{2\pi z^3}} e^{-\frac{a^2}{2z}} \quad z \in \mathbb{R}_+$$

For a visualization of the reflection, refer to Figure 19.3. □

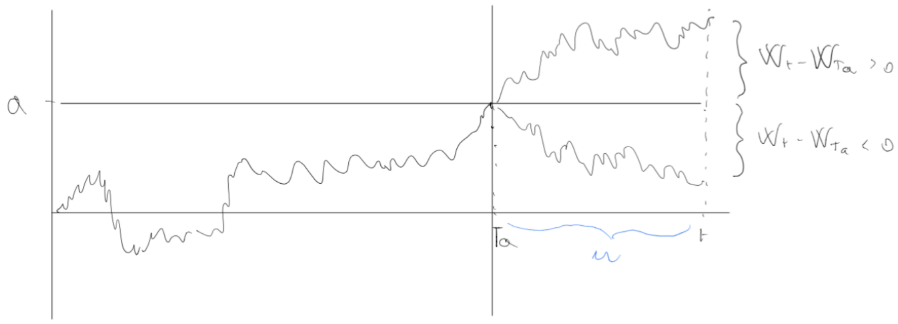

Figure 19.3: Reflection principle for the Wiener process

19.2 Arcsine Laws

Lemma 19.17 (Elementary facts of Gamma, Beta and Cauchy distributions)**.** *For $Z_1 \perp\!\!\!\perp Z_2, Z_1 \stackrel{d}{=} Z_2 \sim \mathcal{N}(0,1)$:*

1. *(Chi-square) $Z_1^2 \stackrel{d}{=} Z_2^2 \sim \chi_1^2 = \mathcal{G}amma(\frac{1}{2}, \frac{1}{2})$*
2. *(Beta) $A = \frac{Z_1^2}{Z_1^2+Z_2^2} \sim \mathcal{B}eta(\frac{1}{2}, \frac{1}{2})$ with density:*

$$f(u) = \frac{\Gamma(\frac{1}{2}+\frac{1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2})} u^{-\frac{1}{2}}(1-u)^{-\frac{1}{2}} = \frac{1}{\pi}\frac{1}{\sqrt{u(1-u)}}$$

   by $\Gamma(1) = 1, \Gamma(\frac{1}{2}) = \sqrt{\pi}$
3. *(Beta cdf) A has cumulative distribution*

$$F_A(u) = \mathbb{P}[A \leq u] = \frac{2}{\pi} \arcsin \sqrt{u}$$

4. *(Cauchy-Beta link) for $C = \frac{Z_1}{Z_2} \sim \mathcal{C}auchy$:*

$$\implies A \stackrel{d}{=} \frac{Z_1^2}{Z_2^2 + Z_1^2} \stackrel{d}{=} \frac{1}{1 + \frac{Z_2^2}{Z_1^2}} \stackrel{d}{=} \frac{1}{1+C^2}$$

   where C has density $f(x) = \frac{1}{\pi(1+x^2)}$ for $x \in \mathbb{R}$ (full support)

♣ **Proposition 19.18** (Arcsine law of G_t and D_t)**.** *Let A be as in Lemma 19.17. Define, as in Example 19.8:*

$$G_t := \sup\{s \in [0,t] : W_t = 0\} \quad D_t := \{u \in (t, \infty) : W_u = 0\}$$

Then:

$$\forall t \in \mathbb{R}_+ \quad G_t \stackrel{d}{=} tA \quad D_t \stackrel{d}{=} \frac{t}{A}$$

Proof. (△ **previous results**) for $t > 0$ we have by Proposition 19.10 that $W_t \neq 0$ almost surely. Then:

$$G_t < t < D_t \quad a.s.$$

Corollaries 19.11, 19.12 suggest that:

$$0 < G_t \quad D_t < \infty$$

While a hitting time process with Theorem 19.16 implies that $T_a \sim \mathcal{IN}(a)$.
(□ **graphical shifts**) let $Z_1, Z_2 \stackrel{iid}{\sim} \mathcal{N}(0,1)$. Consider $R_t = D_t - t > 0$ a.s., which is the hitting time of the barrier $W_t(\omega) = a$. It could be seen as the hitting time for:

$$u \to W_u \circ \theta_t = W_{t+u} - W_t \quad T_{-a} = \inf\{t > 0 : W_t < -a\}$$

of the barrier $-a$ from above when $a > 0$ or $-a$ from below when < 0. By the Markov property the process $W \circ \theta_t \perp\!\!\!\perp \mathcal{F}_t$ and so is $-W \circ \theta_t$ by symmetry (Thm. 18.16#1). We assign $\left(\widetilde{W}\right)_{u\geq 0} = (W_u \circ \theta_t)_{u\geq 0}$ and observe that for arbitrary $b > 0$:

$$\begin{aligned} T_{-b} &= \inf\{u \geq 0 : \widetilde{W}_u < -b\} = \inf\{u \geq 0 : -\widetilde{W}_u < -b\} \\ &= \inf\{u \geq 0 : \widetilde{W}_u > b\} \\ &= T_b \end{aligned}$$

which holds again by symmetry. Hence the time R_t is distributed as $R_t \stackrel{d}{=} T_a$ for $a = |W_t|$. By the reflection principle (Thm. 19.16):

$$\{W_t > a\} \subset \{T_a < t\} \implies \mathcal{P}(T_a < t) = 2\mathcal{P}(W_t > a) \quad T_a \stackrel{d}{=} \frac{a^2}{Z^2} \sim \mathcal{IN}(a), \quad Z \sim \mathcal{N}(0,1)$$

Which means that at position $a = W_t$ it holds that

$$\begin{aligned} R_t &\stackrel{d}{=} \frac{W_t^2}{Z_2^2} \\ &\stackrel{d}{=} t\frac{Z_1^2}{Z_2^2} && W_t \stackrel{d}{=} \sqrt{t}Z_1 \quad Z_1 \sim \mathcal{N}(0,1) \end{aligned}$$

Moving on to D_t, we just recognize that by $D_t = t + R_t$ it holds:

$$D_t \stackrel{d}{=} t\frac{Z_1^2 + Z_2^2}{Z_1^2} \stackrel{d}{=} \frac{t}{A}$$

As per G_t, for $s \in [0,t]$ we have by $\{G_t < s\} = \{D_s > t\}$:

$$\mathbb{P}[G_t < s] = \mathbb{P}[D_s > t] = \mathbb{P}\left[\frac{s}{A} > t\right] = \mathbb{P}[tA < s] \implies G_t \stackrel{d}{=} tA$$

□

Corollary 19.19 (More results from the Proposition)**.** *We could also get:*

1. $R_t \stackrel{d}{=} tC^2$ *(by direct application of Lemma 19.17#4) for* $C \sim \mathcal{C}auchy$
2. G_t *has density* $f(x) = \frac{2}{\pi}\arcsin\sqrt{\frac{s}{t}}$
3. $Q_t \stackrel{d}{=} G_t \stackrel{d}{=} tA$

Proof. **(Claim #1)** trivial.
(Claim #2) from $G_t \stackrel{d}{=} tA$ using Lemma 19.17#3 we get:

$$\mathcal{P}(G_t < s) = \mathcal{P}(tA < s) = \mathcal{P}\left(A < \frac{s}{t}\right) = \frac{2}{\pi}\arcsin\sqrt{\frac{s}{t}}$$

(Claim #3) consider $Q_t = t - G_t$ where:

$$\begin{aligned} \mathbb{P}[Q_t < s] &= \mathbb{P}[t - G_t < s] \\ &= \mathbb{P}[t - tA < s] \\ &= \mathbb{P}[t(1-A) < s] && 1 - A \stackrel{d}{=} A \\ &= \mathbb{P}[tA < s] \\ &= \mathbb{P}[G_t < s] \end{aligned}$$

We proved $Q_t \stackrel{d}{=} G_t$, a shift of proportions. □

♡ **Example 19.20** (Another arcsine law)**.** *We provide two examples.*

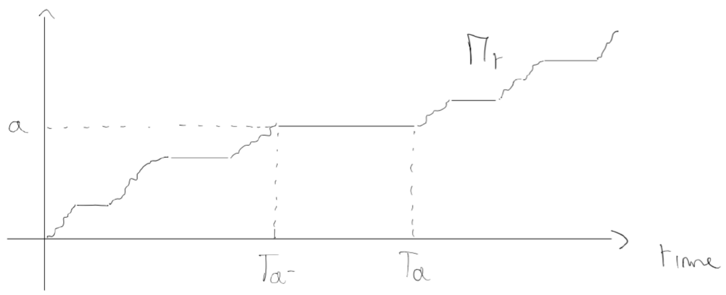

Figure 19.4: Running Maximum Process

1. *W such that no touch on $[s,u]$, namely $\{W_t \neq 0, t \in [s,u]\}$. Then:*

$$\{W_t \neq 0 \forall t \in [s,u]\} = \{G_u < s\} \implies \mathbb{P}[\{W_t \neq 0 \forall t \in [s,u]] = \frac{2}{\pi} arc \sin\left(\sqrt{\frac{s}{u}}\right) \quad s \leq u$$

2. *the occupation time, by [Çin11](Thm. VIII.6.22) could be seen as:*

$$A_t = \int_{[0,t]} \mathbb{1}_{\{W_s>0\}} ds \stackrel{d}{=} tA$$

19.3 Running Maximum and Poisson Jumps to interpret Hitting Times

♠ **Definition 19.21** (Running maximum of process)**.** *Consider a process $(W_t)_{t\in\mathbb{R}_+}$, we define for later use:*

$$M_t(\omega) := \max_{0\leq s\leq t} W_s(\omega) \quad t \in \mathbb{R}_+,\ \omega \in \Omega$$

♣ **Proposition 19.22** (Running maximum vs hitting time)**.** *Recognize that:*

1. *M_t is continuous and increasing in $\mathbb{R}$, with $M_0(\omega) = 0$ and $\lim_{t\to\infty} M_t(\omega) = \infty$*
2. *$(M_t)_{t\geq 0}$ and $(T_a)_{a\geq 0}$ the hitting time process (Def. 19.2) are functional inverses:*

$$\{M_t > a\} = \{T_a < t\}$$

Proof. **(Claim #1)** trivial.
(Claim #2) it holds:

$$T_a = \inf\{t > 0 : W_t > a\} = \inf\{t > 0 : M_t > a\} \implies \{T_a < t\} \implies \{M_t > a\}$$

Where the direction $\{M_t > a\} \implies \{T_a < t\}$ is rather trivial. Proof by words is also useful to understand. □

♢ **Observation 19.23** (Why running maximum and the general idea)**.** *In Theorem 19.14 we used $\{W_t > a\} \subset \{T_a > t\}$. Now we use the more precise equivalence of sets $\{M_t > a\} = \{T_a < t\}$.*
We also use in Proposition 19.26 the hitting time of point a, constructed as:

$$T_{a-} = \inf\{t > 0 : W_t = a\} = \inf\{t > 0 : W_t \geq a\} = \lim_{u\uparrow 0} T_{a-u} = \lim_{u\uparrow 0} \inf\{t > 0 : W_t > a-u\} \quad T_a = \inf\{t > 0 : W_t > a\}$$

*One problem is that while M_t is **continuous**, the path $a \to T_a$ is only **right continuous**, as is shown below.*

♡ **Example 19.24** (Graphical interpretation of the Observation)**.** *See Figure 19.4.*

♣ **Theorem 19.25** (Hitting time process is stable Lévy)**.** *The process $T = (T_a)_{a\geq 0}$ is a strictly increasing pure jump Lévy process (Defs. 17.1, 17.7) with index $\frac{1}{2}$ and Lévy density:*

$$\lambda(dz) = \frac{1}{\sqrt{2\pi z^3}} dz \quad z \in \mathbb{R}_+$$

Proof. **($\triangle$ aim)** wts that the definitions are verified.
($\square$ $a+b$ split) fix $a, b \in (0, \infty)$the process W hits $a+b$ if it hits a and the shifted process $\widetilde{W} = W \circ \theta_{T_a}$ hits b. This means that:

$$T_{a+b} = T_a + T_b \circ \theta_{T_a}$$

By the strong Marjov property (Thm. 19.14) and the fact that $T_a < \infty$ almost surely (i.e. version without indicators) we get:

$$\widetilde{W} = W \circ \theta_{T_a} \perp\!\!\!\perp \mathcal{F}_{T_a}, T_a, \quad Wiener \implies T_{a+b} - T_a = T_b \circ \theta_{T_a} \overset{d}{=} T_b$$

We have independence and stationarity, which, together with $T_0 \overset{a.s.}{=} 0$ (Prop. 19.10) and right continuity means $(T_a)_{a \geq 0}$ is an increasing Lévy process.
($\bigcirc$ density) the density is that of a stable process by Theorem 19.16. The process is of pure jump type (Def. 17.7), stable of index $\frac{1}{2}$. From $T_a = a^2 T_1$ and this discussion, every Lévy process of this form has density $\lambda(dz) = dz\frac{c}{t^{\frac{3}{2}}}$ by [Çin11](Ex. VII.2.1), which in our case, by a Laplace transform argument is precisely:

$$c = \frac{1}{\sqrt{2\pi}} \quad \mathbb{E}\left[e^{-pT_a}\right] = \exp\left\{-a \int_{\mathbb{R}_+} \lambda(dz)(1 - e^{-rt})\right\} = \exp\left\{-a\sqrt{2p}\right\}$$

$\square$

♣ **Proposition 19.26** (Properties of T_{a-})**.** *Recall Observation 19.23 with $T_{a-} = \lim_{u \uparrow 0} T_{a-u} = \inf\{t > 0 : W_t = a\}$. Then:*

1. *T_{a-} is a stopping time for $\mathcal{F}^0$ the **not augmented** filtration, while T_a is a stopping time wrt $\mathcal{F}$ the **augmented filtration***
2. *sojourne time is zero almost surely*

$$T_{a-} = T_a \quad a.s.$$

Proof. **(Claim #1)** the claim holds by construction.
(Claim #2) by Theorem 19.16 we have that $T_a \sim \mathcal{IN}(a)$, which means that $T_a < \infty$ almost surely. Clearly $T_{a-} \leq T_a$, and we have that $T_{a-} < \infty$ almost surely. By Claim #1 we can also apply the strong Markov property (Thm. 19.14):

$$\begin{aligned}
T_a &= T_{a-} + \inf\{u > 0 : \underbrace{W_{T_{a-}+u} - W_{T_{a-}}}_{= W \circ \theta_{T_{a-}}} > 0\} && W \circ \theta_{T_{a-}} \perp\!\!\!\perp \mathcal{F}_{T_{a-}}, T_{a-} \\
&= T_{a-} + \underbrace{T_0 \circ \theta_{T_{a-}}}_{\perp\!\!\!\perp \mathcal{F}_{T_{a-}}, T_{a-}} && T_0 \circ \theta_{T_{a-}} \overset{d}{=} T_0 \\
&\overset{d}{=} T_{a-} + \underbrace{T_0}_{=0\ a.s.} && \text{Prop. 19.10} \\
&= T_{a-} \quad a.s.
\end{aligned}$$

$\square$

♡ **Example 19.27** (Plots of Poisson Jumps)**.** *Consider Figure 19.5. Atoms of $N(dx, dz)$ are marked with little circles, corresponding to the atom (a, z) there is a jump of size z from T_{a-} to $T_{a-} + z = T_a$. The path $t \to M_t$ stays constant at level a during $[T_{a-}, T_a]$, an interval of length z. Since $N(dx, dz)$ has only countably many atoms, the situation occurs at countably many levels a only. Since there are infinitely many atoms in the strip $[a, a+b] \times \mathbb{R}_+$, the path $t \to M_t$ stays flat at infinitely many levels on its way from a to $a+b$. However, for every $\epsilon > 0$, only finitely many of those sojourns exceed ϵ in duration. The situation at fixed a is simpler. For $a > 0$ almost surely, there are no atoms on the line $\{a\} \times \mathbb{R}_+$, therefore $T_{a-} \overset{a.s.}{=} T_a = 0$.*

♢ **Observation 19.28** (Poisson jump structure)**.** *By the Itô-Lévy decomposition [Çin11](Thm. VII.5.2) we have that $(T_a)_{a \geq 0}$ is an increasing Lévy process (Def. 15.14). For a general one we have $S = (S_t)_{t \in \mathbb{R}_+}$ with Lévy measure satisfying:*

$$\int (1 \wedge z)\lambda(dz) < \infty$$

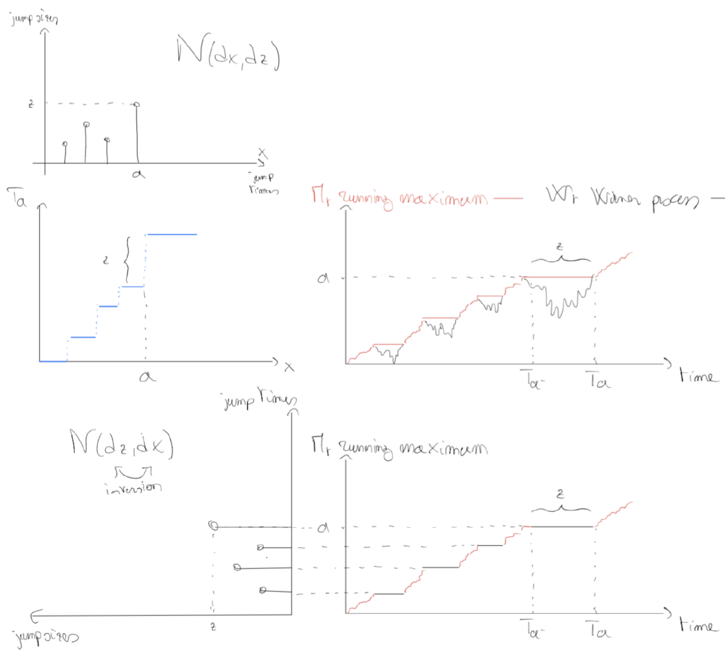

Figure 19.5: Poisson Jumps final plot

which is described in the sense of Definition 15.16 as an integral wrt the underlying p.r.m. on $\mathbb{R}_+ \times \mathbb{R}_+$ *with mean* $dx\lambda(dz)$*:*

$$S_t = \int_{[0,t]\times\mathbb{R}_+} zN(dx,dz) = \sum_{i:X_i\leq t} Z_i$$

In our specific case we obtain the jump structure by means of a p.r.m. $N(dx,dz)$*:*

$$N(B) = \sum_a \mathbb{1}_B\left(a, T_a - T_{a-}\right) \quad B \in \mathcal{B}(\mathbb{R}_+\times\mathbb{R}_+) \quad mean \;\; dx\lambda(dz) = dx\frac{1}{\sqrt{2\pi z^3}}$$

If (a,z) *is an atom then the map* $a \to T_a$ *controls the abstract jumps of size* a *at time* z*. With:*

$$T_a = \int_{[0,a]\times\mathbb{R}_+} zN(dx,dz) = \sum_{i:X_i\leq a} z_i$$

Chapter Summary

Objects:

- right continuous augmented filtration, a peek in the future with all negligible sets from the start

$$\mathcal{F}_{t+} = \cap_{\epsilon>0}\mathcal{F}_{t+\epsilon}$$

- hitting time of barrier a is $T_a = \inf\{t > 0 : X_t > a\}$
- shift operator for a class of continuous maps $\mathscr{C}$ is θ_s such that $(\theta_s \circ w)(t) = w(s+t)$ for all $s \in \mathbb{R}, t \in \mathbb{R}_+$
- running maximum of a process $M_t(\omega) = \max_{0\le s\le t} W_s(\omega)$

Results:

- hitting times are stopping times in the augmented right continuous filtration
- Lévy processes X (and thus Wiener processes W) are Markovian

$$\mathbb{E}_t\left[V \circ \theta_t\right] = \mathbb{E}\left[V\right] \quad t \in \mathbb{R}_+, \quad \forall V \in \mathcal{G}_\infty \textit{ bounded}$$

  which is an equivalent characterization with $\mathcal{G} = \sigma(X)$
- time shifts on Wiener process $W_t \circ \theta_s = W_{t+s} - W_s$
- time shifts on Brownian motion $X_t \circ \theta_s = X_{t+s}$
- Blumenthal's 0-1 law

$$\mathcal{F} \quad \text{augmented, right continuous} \implies \forall A \in \mathcal{F}_0 \; \mathbb{P}[A] \in \{0,1\}$$

- the hitting time process $(T_a)_{a\ge0}$ is such that $T_0 \overset{a.s.}{=} 0$
- Wiener highly oscillatory behavior at zero, there are ∞-many crossing for any time interval starting from zero
- Wiener highly oscillatory behavior at infinity
- Lévy (and thus Wiener) processes are strongly Markovian in the Brownian formulation for stopping times in the natural filtration $\mathcal{F}^0$ (the **not augmented** filtration). Namely:
  - $\forall V$ bounded in $\overline{\mathcal{G}}_\infty$, where $\mathcal{G} = \sigma(X)$:

$$\mathbb{E}_T[V \circ \theta_T \mathbb{1}_{\{T<\infty\}}] = \mathbb{E}[V \mathbb{1}_{\{T<\infty\}}] \perp\!\!\!\perp \mathcal{F}_T, \text{ Lévy}$$

    independent of $\mathcal{F}_T$ and Lévy
  - shift operator version:

$$(X_t \circ \theta_T)_{t\ge0} = (X_{t+T})_{t\ge0} \perp\!\!\!\perp \mathcal{F}_T, \text{ Lévy}$$

- for T a stopping time wrt $\mathcal{F}$ augmented, $U : \Omega \to \mathbb{R}_+$ such that $U \in \mathcal{F}_T$, $W = (W_t)_{t\in\mathbb{R}_+}$ a Wiener process, f a bounded Borel function on $\mathbb{R}$ it holds:

$$\mathbb{E}_T\left[f(W_{T+u} - W_T)\mathbb{1}_{\{T<\infty\}}\right] = g(U)\mathbb{1}_{\{T<\infty\}} \quad g(u) := \mathbb{E}[f \circ W_u], \quad u \in \mathbb{R}_+$$

- reflection principle, $(T_a)_{a\ge0}$ is inverse Gaussian distributed $T_a \sim \mathcal{IN}(a)\; \forall a \in \mathbb{R}_+$
- recap about elementary connections of Gamma, Beta, Cauchy distributions
- arcsine law of G_t and D_t derived in the exercises is for $A \sim \mathcal{B}eta\left(\frac{1}{2}, \frac{1}{2}\right)$ and $C \sim \mathcal{C}auchy, C = \frac{Z_1}{Z_2}$:

$$G_t \overset{d}{=} tA, \quad D_t \overset{d}{=} \frac{t}{A}, \quad R_t tC^2, \quad Q_t G_t : f(x) = \frac{2}{\pi}\arcsin\sqrt{\frac{s}{t}}$$

- running maximum is continuous and increasing in $\mathbb{R}$, starting at zero
- running maximum is the functional inverse of the hitting time process
- the hitting time process is a pure jump Lévy process stable of order 2 with density:

$$\lambda(dz) = \frac{1}{\sqrt{2\pi z^3}}dz \quad z \in \mathbb{R}_+$$

- the pre hitting time T_{a-} is such that $T_a = T_{a-}$ almost surely

Chapter 20

Path Properties of Wiener processes

20.1 Variation

♠ **Definition 20.1** (Subdivision & mesh)**.** *We denote for an interval $[a, b]$ a subdivision as a finite collection of intervals whose union is the interval itself (ignoring the start).*

$$\mathcal{A} := \{(s, t]\}, \qquad \bigcup_{(s,t]\in\mathcal{A}} (s, t] = (a, b]$$

Given a subdivision, we also identify the mesh as:

$$\|\mathcal{A}\| := \sup\{t - s : (s, t] \in \mathcal{A}\}$$

♠ **Definition 20.2** (True p-variation, total variation, quadratic variation)**.** *For a function $f : \mathbb{R}_+ \to \mathbb{R}$, right continuous, an interval $[a, b] \in \mathbb{R}_+$, and a positive coefficient $p > 0$ the true p-variation is the quantity:*

$$\sup_{\mathcal{A}} \sum_{(s,t]\in\mathcal{A}} |f(t) - f(s)|^p$$

*Where for $p = 1$ we call it total variation and for $p = 2$ **true** quadratic variation. It turns out that this formulation is not very well suited for random processes, as it is infinite in both $p = 1, p = 2$. Below we prove the results for a reasonable surrogate.*

♠ **Definition 20.3** (A dyadic subdivision)**.** *In the fashion of Definition 18.14, and Lemma A.17 we could construct a subdivision of equally spaced intervals using $t_k = \frac{kt}{2^n}$ for $k = 0, \dots, 2^n$.*

♣ **Theorem 20.4** (Wiener probabilistically finite $\mathcal{L}^2$ quadratic variation)**.** *For a Wiener process $(W_t)_{t\in\mathbb{R}_+}$ and a sequence of subdivisions $(\mathcal{A}_n)_{n\in\mathbb{N}}$ with $\|\mathcal{A}_n\| \to 0$ we have:*

1. $V_n = \sum_{(s,t]\in\mathcal{A}_n} |W_t - W_s|^2 \xrightarrow{\mathcal{L}^2} b - a$
2. $V_n \xrightarrow{\mathcal{P}} b - a$

*We call V_n the quadratic variation, not to be confused with the **true** quadratic variation. It is rather a probabilistic version of the latter. In some terms, the* sup *can be replaced with the* lim sup *of a sequence of meshes with size decreasing to zero (i.e. $\|\mathcal{A}_n\| \to 0$).*

Proof. **(Claim #1)(△ expectation)** by the scaling property (Thm. 18.16#2) it holds:

$$W_t - W_s \overset{d}{=} \sqrt{t-s}Z \quad Z \sim \mathcal{N}(0, 1) \implies |W_t - W_s|^2 \overset{d}{=} (t - s)Z^2 : \begin{cases} Z^2 \sim \chi^2_1 = \mathcal{G}amma\left(\frac{1}{2}, \frac{1}{2}\right) \\ \mathbb{E}\left[Z^2\right] = 1 \\ V[Z^2] = 2 \end{cases}$$

So that the expectation of the quadratic variation is:

$$\begin{aligned}
\mathbb{E}[V_n] &= \sum_{(s,t]\in\mathcal{A}_n} \mathbb{E}\left[(W_t - W_s)^2\right] && \text{linearity} \\
&= \sum_{(s,t]\in\mathcal{A}_n} (t-s)\mathbb{E}\left[Z^2\right] \\
&= \sum_{(s,t]\in\mathcal{A}_n} (t-s) \\
&= b - a && \text{telescopic sum}
\end{aligned}$$

(□ variance) it holds:

$$\begin{aligned}
V[V_n] &= \sum_{(s,t]\in\mathcal{A}_n} V[|W_t - W_s|^2] && [t,s) \text{ are disjoint increments of a Wiener process} \\
&= \sum_{(s,t]\in\mathcal{A}_n} V[(t-s)Z^2] \\
&= \sum_{(s,t]\in\mathcal{A}_n} (t-s)^2 V[Z^2] \\
&\le 2\,\|\mathcal{A}_n\| \sum_{(s,t]\in\mathcal{A}_n} (t-s) \\
&= 2\,\|\mathcal{A}_n\|\,(b-a) && \text{telescopic sum} \\
&\overset{n\to\infty}{\to} 0
\end{aligned}$$

(◯ convergence) by △, □ it holds:

$$\mathbb{E}\left[|V_n - (b-a)|^2\right] = V[V_n] \overset{n\to\infty}{\to} 0 \implies V_n \overset{\mathcal{L}_2}{\to} b - a$$

(Claim #2) Just an application of Proposition 9.19. □

♣ **Proposition 20.5** (Almost sure dyadic subdivision for quadratic variation)**.** $\forall n \in \mathbb{N}$ *let* $\mathcal{A}_n$ *be a subdivision of the form presented in Definition 20.3. Then, with the hypothesis of Theorem 20.4 it also holds that:*

$$V_n = \sum_{(s,t]\in\mathcal{A}_n} |W_t - W_s|^2 \overset{a.s.}{\to} b - a$$

Proof. From the previous results we proved $\mathbb{E}[V_n] = b - a$. In this case, the size of the subdivision is $\|\mathcal{A}_n\| = \frac{1}{2^n}$ $\forall n$, by the dyadic construction. Clearly, by the subdivisions having equal length:

$$\begin{aligned}
V[V_n] &= 2 \cdot \sum_{(s,t]\in\mathcal{A}_n} (t-s)^2 && \text{as in Thm. 20.4 with } = \text{ instead of } \le \\
&= 2 \sum_{k=0}^{2^n-1} \left(\frac{b-a}{2^n}\right)^2 \\
&= 2 \cdot 2^n \frac{(b-a)^2}{2^{2n}} \\
&= 2\frac{(b-a)^2}{2^n}
\end{aligned}$$

By Chebychev's inequality (Cor. 7.5) we obtain the bound:

$$\mathcal{P}\left(|V_n - \underbrace{(b-a)}_{\mathbb{E}[V_n]}| > \epsilon\right) \le \frac{V[V_n]}{\epsilon^2} = \frac{1}{\epsilon^2}\left(\frac{2}{2^n}(b-a)^2\right) \quad \forall \epsilon > 0$$

From which we are in the position to apply BC1 (Thm. 9.6) as in Example 9.9:

$$\forall \epsilon > 0 \quad \sum_n \mathcal{P}\left(|V_n - (b-a)| > \epsilon\right) \leq \sum_n \frac{1}{\epsilon^2}\left(\frac{2}{2^n}(b-a)^2\right) = \frac{(b-a)^2}{\epsilon^2} \underbrace{\sum_n \frac{1}{2^{n-1}}}_{=\frac{1}{2}} < \infty \implies V_n \overset{a.s.}{\to} b-a$$

□

♣ **Proposition 20.6** (Infinite total variation of Wiener process). *For a Wiener process $(W_t)_{t\in\mathbb{R}_+}$ we have:*

$$TV = \sup_{\mathcal{A}} \sum_{(s,t]\in\mathcal{A}} |W_t - W_s| = \infty \quad \textit{almost surely}$$

over any interval $[a,b]$

Proof. **(△ setting)** We make use of the sequence of subdivisions $(\mathcal{A}_n)_{n\in\mathbb{N}}$ with $\|\mathcal{A}_n\| \to 0$ from Prop. 20.5 and the result of Theorem 20.4. Let Ω_{ab} be the a.s. set in which $V_n \overset{a.s.}{\to} b-a$. Denote the total variation as $v^* \leq +\infty$ over $[a,b]$.
(□ calculation) Then for $\omega \in \Omega_{ab}$:

$$V_n(\omega) = \sum_{(s,t]\in\mathcal{A}_n} |W_t - W_s|^2 \leq \sup_{(s,t]\in\mathcal{A}_n} \{W_t - W_s\} \sum_{(s,t]\in\mathcal{A}_n} |W_t - W_s| = \sup_{(s,t]\in\mathcal{A}_n} \{W_t - W_s\} \cdot v^*$$

(◯ argument) We know the LHS is $b-a$ as $n \to \infty$ since we are in Ω_{ab}. The first term of the RHS is by the Hölder continuity of W (Thm. 18.22) is:

$$|W_t - W_s| \leq C|t-s|^\alpha = C\,\|\mathcal{A}_n\|^\alpha \overset{n\to\infty}{\to} 0 \implies \sup\{|W_t - W_s|\} \leq C\,\|\mathcal{A}_n\|^\alpha$$

Which means that the second term, v^*, necessarily diverges to ∞.
Let $\Omega_0 = \bigcap_{a,b:0\leq a<b} \Omega_{ab}$, the event is almost sure, and the total variation is infinite in it. □

♢ **Observation 20.7** (Consequences of infinite total variation). *By Proposition 20.6 For a fixed path $\omega \in \Omega$ we almost surely cannot define an integral with respect to W_t in the Steltjes-Lebesgue Riemaniann way (Def. 12.7), as there is no limit of subdivisions in the Riemann sense and $\nexists \int f(s)dW_s$.*

Chapter Summary

Objects:

- subdivision, a finite collection of intervals covering $[a, b]$ without the start

$$\mathcal{A} = \{(s,t]\}, \qquad \bigcup_{(s,t]\in\mathcal{A}} (s,t] = (a,b]$$

- mesh of a subdivision

$$\|\mathcal{A}\| = \sup\{t - s : (s,t] \in \mathcal{A}\}$$

- true p-variation of a right continuous function $f : \mathbb{R}_+ \to \mathbb{R}$ on an interval $[a, b] \in \mathbb{R}_+$

$$\sup_{\mathcal{A}} \sum_{(s,t]\in\mathcal{A}} |f(t) - f(s)|^p \quad p \in \mathbb{R}_+$$

  Where for $p = 1$ we call it total variation and for $p = 2$ **true** quadratic variation. Both are infinite for a Wiener process

- dyadic subdivision, a subdivision of equally spaced intervals using $t_k = \frac{kt}{2^n}$ for $k = 0, \ldots, 2^n$

Results:

- the Wiener process $(W_t)_{t\in\mathbb{R}_+}$ has a probabilistically bounded quadratic variation when $(\mathcal{A}_n)_{n\in\mathbb{N}}$ with $\|\mathcal{A}_n\| \to 0$

$$V_n = \sum_{(s,t]\in\mathcal{A}_n} |W_t - W_s|^2 \xrightarrow{\mathcal{L}^2} b - a \quad V_n \xrightarrow{\mathcal{P}} b - a$$

  We call V_n the quadratic variation, not to be confused with the **true** quadratic variation. It is rather a probabilistic version of the latter.

- the Wiener process has an almost sure dyadic subdivision limit of the quadratic variation

$$V_n = \sum_{(s,t]\in\mathcal{A}_n} |W_t - W_s| \xrightarrow{a.s.} b - a$$

- the Wiener process has infinite total variation

$$TV = \sup_{\mathcal{A}} \sum_{(s,t]\in\mathcal{A}} |W_t - W_s| = \infty \quad \textit{almost surely}$$

- we cannot define a Steltjes-Riemann-Lebesgue integral with respect to a Wiener process after fixing the path $\omega \in \Omega$

Part III

Additional Material

Chapter 21

Recap of Part II

The following is a meta Chapter to ease help organize ideas.

21.1 Results Collection

Throughout the second part of the course (Chapters 11-20), many results were reported on different occasions, the purpose of this Section is collecting them under the same discussion. It is just a copy paste. No proofs are reported.

21.1.1 Martingales and Processes

Definition (Background notation)**.** *Assume that we are now in a probability space (Def. 10.45)* $(\Omega, \mathcal{H}, \mathbb{P})$. *We will index sequences by a countable collection* $\mathbb{N} = \{0, 1, 2, \ldots\}$ *or an uncountable collection such as* $\mathbb{R}_+$ *or a generic* $\mathbb{T}$.
A stochastic process will be an indexed sequence of random variables (Def. 10.58).
Occasionally, we might denote measurable functions over a measurable space $(E, \mathcal{E})$ *with the symbol* $f \in \mathcal{E}$ *and accordingly with signs* $\pm$ *for negative and positive cases.*

Definition (Filtration)**.** *For an index set* $\mathbb{T}$ *a filtration is a sequence* $(\mathcal{F}_t)_{t\in\mathbb{T}}$ *such that:*

1. $\mathcal{F}_t \subset \mathcal{H}\ \forall t$ *and* $\mathcal{F}_t$ *is a* σ*-algebra (Def. 1.6)* $\forall t$
2. $\mathcal{F}_s \subset \mathcal{F}_t\ \forall s < t$

Intuitively, it is a sequence of increasing information in the probability space.

Definition (Filtration generated by a random variable)**.** *Given a stochastic process* $(X_t)_{t\in\mathbb{T}}$ *the filtration generated by it is denoted as:*

$$\mathcal{F}_t = \sigma(\{X_s : s \leq t\})$$

Which can be seen as a flow of information accumulated at each time point.

Definition (Finer, coarser filtration)**.** *Consider* $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}, \mathcal{G} = (\mathcal{G}_t)_{t\in\mathbb{T}}$ *to be two filtrations. We say* $\mathcal{F}$ *is finer (coarser) than* $\mathcal{G}$ *is* $\forall t \in \mathbb{T}\ \mathcal{F}_t \supset$ *(respectively,* $\subset$*)* $\mathcal{G}_t$.

Definition (Stochastic process adapted to filtration)**.** *Consider a filtration* $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ *and a stochastic process* $X = (X_t)_{t\in\mathbb{T}}$ *taking values on* $(E, \mathcal{E})$. *We say that* X *is adapted to* $\mathcal{F}$ *if* $\forall t\ X_t$ *measurable w.r.t.* $\mathcal{F}\&\mathcal{E}$

Proposition (Equivalent statements for filtrations and adaptedness)**.** *Consider a stochastic process* $X = (X_t)_{t\in\mathbb{T}}$ *and a filtration* $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$. *Then:*

1. X *adapted to* $\mathcal{F}$ *(Def. 11.7)* $\iff \forall t, s \leq t,\ f \in \mathcal{E}\quad f \circ X_s \in \mathcal{F}_t$

2. *since* $\mathcal{G} = \sigma(X) \implies X$ *adapted* $\mathcal{G}$ *we have that:*

$$X \text{ adapted } \mathcal{F} \iff \mathcal{F} \text{ finer } \mathcal{G}$$

Definition (Stopping times). *For a filtration* $\mathcal{F}$ *a stopping time with respect to it is a random function* $T : \Omega \to \mathbb{T} \cup \{\infty\}$ *such that:*

$$\{T \leq t\} \in \mathcal{F}_t \quad \forall t \in \mathbb{T}$$

Which is equivalent to requiring the process $Z_t = \mathbb{1}_{\{T \leq t\}} \in \mathcal{F}_t$ *for all* $t \in \mathbb{T}$.
In the ***special case*** *in which* $\mathbb{T} = \mathbb{N}$ *or* $\overline{\mathbb{N}}$ *the condition reduces for* $\widehat{Z}_n = Z_n - Z_{n-1}$ *to:*

$$\widehat{Z}_n = \mathbb{1}_{\{T=t\}} \quad \forall t \in \mathbb{T}$$

Definition (Counting process $(N_t)_{t\in\mathbb{T}}$). *Let* $) < T_1 < T_2 < \dots$ *be random times of the form* $T_n : \Omega \to \mathbb{T} = \mathbb{R}_+$ *such that* $\lim_{n\to\infty} T_n = +\infty$.
These can be seen as a sequence of ***distinct arrival times***.
A counting process is a stochastic process (Def. 10.58) of the form:

$$N_t = \sum_n \mathbb{1}_{[0,t]}(T_n)$$

Proposition (Properties of $(N_t)_{t\in\mathbb{T}}$). *The map* $t \to N_t$ *is:*

1. *right continuous*
2. *increasing in* t
3. *has jumps of size* 1
4. $N_0 = 0$, $N_t < \infty \forall t \in \mathbb{R}_+$, $\lim_{t\to\infty} N_t = \infty$

Definition (End of time information $\mathcal{F}_\infty$, extended filtration $(\mathcal{F}_t)_{t\in\overline{\mathbb{T}}}$). *We define* $\mathcal{F}_\infty = \lim_{t\to\infty} \mathcal{F}_t = \bigvee_t \mathcal{F}_t$, *where the union symbol is different as it is* ***over σ-algebras***.
Then, the extended filtration is a filtration which accounts for $\mathbb{P}[T = \infty] > 0$*:*

$$(\mathcal{F}_t)_{t\in\overline{\mathbb{T}}} \quad \overline{\mathbb{T}} = \mathbb{T} \cup \{\infty\}$$

Definition (Stopped filtration $\mathcal{F}_T$ at T, past until T). *For* $\mathcal{F}$ *a filtration on* $\mathbb{T}$, *extended to* $\overline{\mathbb{T}}$, *and* T *a stopping time, the stopped filtration is defined as:*

$$\mathcal{F}_T = \left\{H \in \mathcal{H} : H \cap \{T \leq t\} \in \mathcal{F}_t \ \forall t \in \overline{\mathbb{T}}\right\}$$

Lemma (Properties of $\mathcal{F}_T$). *A stopped filtration* $\mathcal{F}_T$ *is such that:*

1. $\mathcal{F}_T$ *is a σ-algebra (Def. 1.6)*
2. $\mathcal{F}_T \subset \mathcal{F}_\infty \subset \mathcal{H} \ \forall t$

Theorem (Formalizing Observation 11.22). *Drawing from the previous comment, for a stopped filtration* $\mathcal{F}_T$, *with stopping time* T, *index* $\mathbb{T}$, *and filtration* $\mathcal{F}$*:*

1. *stopped filtration filtration identification*

$$V \in \mathcal{F}_T \iff V\mathbb{1}_{T\leq t} \in \mathcal{F}_t \forall t \in \overline{\mathbb{T}}$$

2. *stopped filtration identification for a discrete process*

$$\overline{\mathbb{N}} = \overline{\mathbb{T}} \ \ V \in \mathcal{F}_T \iff V\mathbb{1}_{T=t} \in \mathcal{F}_t \forall t \in \overline{\mathbb{T}}$$

Which are both an extension of the comments of adaptedness from Definition 11.7 for deterministic times.

Definition ($\mathcal{F}$ processes collection). *Using the same notation for positive measurable functions as* $f \in \mathcal{E}$ *we let:*

$$\mathcal{F} = \left\{\text{right continuous processes on } \overline{\mathbb{T}} \text{ adapted to } \mathcal{F}\right\}$$

Where $\mathcal{F}$ *is extended to* $\overline{\mathbb{T}}$.
This means that $X \in \mathcal{F}$ *whenever:*

1. *$X = (X_t)_{t\in\mathbb{T}}$ is adapted to $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$*
2. *$t \to X_t(\omega)$ where $X_t : \overline{\mathbb{T}} \to \overline{\mathbb{R}}$ is right continuous $\forall\omega \in \Omega$*

Theorem (Comparing different stopping times)**.** *Let S, T be stopping times of a filtration $\mathcal{F}$ (Def. 11.9), where $S \leq T$ almost surely, meaning $S(\omega) \leq T(\omega)\forall\omega \in \Omega$. Then:*

1. *$S \wedge T$, $S \vee T$ are stopping times of $\mathcal{F}$*
2. *specifically $S \leq T \implies \mathcal{F}_S \subset \mathcal{F}_T$*
3. *$\mathcal{F}_{S\wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$*
4. *$V \in \mathcal{F}_S \implies \begin{cases} V\mathbb{1}_{S\leq T} \in \mathcal{F}_{S\wedge T} \\ V\mathbb{1}_{S=T} \in \mathcal{F}_{S\wedge T} \\ V\mathbb{1}_{S<T} \in \mathcal{F}_{S\wedge T} \end{cases}$*

Definition (Expectation in Filtration $\mathbb{E}_t$)**.** *Given a filtration $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ (Def. 11.2) use as notation:*

$$\mathbb{E}_t[X] \coloneqq \mathbb{E}[X|\mathcal{F}_t] = \mathbb{E}_{\mathcal{F}_t}[X] = \overline{X}_{\mathcal{F}_t} \quad t \in \mathbb{T}$$

Proposition (Repeated conditioning of $\mathbb{E}_t$)**.** *For $X \geq 0$ a.s. it holds that:*

$$\mathbb{E}_t\left[\mathbb{E}_s[X]\right] = \mathbb{E}_{s\wedge t}[X] \quad \forall s, t \in \mathbb{T}$$

Definition (Expectation with respect to stopped filtration $\mathbb{E}_T$)**.** *Given $\mathcal{F}_T$ a stopped filtration (Def. 11.19), recalling that $\mathcal{F}_T$ is a σ-algebra, simply define:*

$$\mathbb{E}_T[X] \coloneqq \mathbb{E}[X|\mathcal{F}_T] = \mathbb{E}_{\mathcal{F}_T}[X] = \overline{X}_{\mathcal{F}_T}$$

Theorem (Properties of $\mathbb{E}_T$)**.** *Consider $X, Y, W \geq 0$ a.s. and S, T stopping times (Def. 11.9) of a filtration $\mathcal{F}$ (Def. 11.2). Then:*

1. *Projection defining property*

$$\mathbb{E}_T[X] = Y \iff Y \in \mathcal{F}_T \ \mathbb{E}[VX] = \mathbb{E}[VY] \ \forall V \ \mathcal{F}_T\text{-measurable positive}$$

2. *unconditioning*

$$\mathbb{E}\left[\mathbb{E}_T[X]\right] = \mathbb{E}[X]$$

3. *repeated conditioning/towering*

$$\mathbb{E}_S\mathbb{E}_T[X] = \mathbb{E}_{S\wedge T}X$$

4. *conditional determinism*

$$\mathbb{E}_T[WX] = W\mathbb{E}_T[X] \quad \forall W \ \mathcal{F}_T\text{-measurable}$$

Definition (Martingales)**.** *For $\mathbb{T} = \mathbb{R}_+$ and $\mathcal{F}$ a filtration, possibly extended to $\overline{\mathbb{T}}$ a $\mathcal{F}$-martingale is a stochastic process $X = (X_t)_{t\in\mathbb{T}}$ such that:*

1. *X is adapted to $\mathcal{F}$ (Def. 11.7)*
2. *$\forall t$ X_t is integrable $\iff \mathbb{E}[|X_t|] < \infty\forall t \iff X_t \in \mathcal{L}_1 \ \forall t$*
3. *martingale equality*

$$\mathbb{E}_s[X_t - X_s] = 0 \quad \forall s < t, \forall t$$

Definition (Submartingale, supermartingale)**.** *we recognize two additional options for the last property:*

- *a **submartingale** satisfies Definition 11.35 but has $\geq$ in the martingale equality*
- *a **supermartingale** satisfies Definition 11.35 but has $\leq$ in the martingale equality*

Proposition (Best guess of future is present characterizes martingale equality)**.** *It holds that:*

$$\text{Def. 11.35\#3} \iff \mathbb{E}_s[X_t] = X_s \quad \forall s < t$$

Proposition (Martingale implies stationarity)**.**

$$X \ \mathcal{F}\text{-martingale} \implies \mathbb{E}[X_t] = \mathbb{E}[X_0]\forall t \in \mathbb{T}$$

Proposition (Discrete time martingale check)**.** *For a discrete process over $\mathbb{T} = \mathbb{N}$ it is sufficient to check for one step forward the martingale equality if the other two conditions are satisfied (**integrability** and **adaptedness**):*

$$\mathbb{E}_n[X_{n+k} - X_n] = 0 \; \forall k > 0, \forall n \in \mathbb{N} \iff \mathbb{E}_n[X_{n+1} - X_n] = 0 \; \forall n$$

Corollary (Jensen's for martingales)**.** *This is a Corollary of Jensen's inequality.*
For a martingale X and a convex function $f : \mathbb{R} \to \mathbb{R}$ then:

$$f \circ X_t \text{ integrable } \forall t \in \mathbb{T} \implies f \circ X \text{ submartingale}$$

Definition (Uniformly integrable martingale)**.** *We define e u.i. martingale as in Definition 7.3 with as arbitrary index set $\mathbb{T}$. Namely:*

$$(X_t)_{t \in \mathbb{T}} \quad \lim_{b \to \infty} \sup_t \mathbb{E}|X_t| \mathbb{1}_{[b,\infty)}(|X_t|) = 0$$

Proposition (Uniformly integrable martingale by integrable random variable)**.** *Let $Z \in \mathcal{L}_1(\Omega, \mathcal{H}, \mathbb{P})$ and $\mathcal{F}$ a filtration.*

$$\implies X = (X_t)_{t \in \mathbb{T}} : X_t = \mathbb{E}_t[Z] \; \forall t \in \mathbb{T} \quad uniformly\ integrable\ martingale$$

To prove this result, we need a Theorem from the book, which is reported in the Appendix.

Lemma (A more general result)**.** *It actually holds that:*

$$Z \in \mathcal{L}_1(\Omega, \mathcal{H}, \mathbb{P}) \implies \mathcal{K} = \{X \mid X = \mathbb{E}_\mathcal{G}[Z], \; \mathcal{G} \subset \mathcal{H}\} \quad uniformly\ integrable$$

Definition (Wiener process W)**.** *A stochastic process $W = (W_t)_{t \in \mathbb{R}_+}$ is Wiener with respect to the filtration $\mathcal{F}$ if:*

1. *W is adapted to $\mathcal{F}$ (Def. 11.7)*
2. *Gaussian intervals*
$$\mathbb{E}_s[f(W_{s+t} - W_s)] = \int f(x) \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2t}x^2} dx \quad \forall s, t, \; \forall f \in \mathcal{E}_+$$
*Where $\mathcal{E}_+$ is to be intended as **positive Borel functions mapping to** $\mathbb{R}$.*
3. *$W_0 = 0$*

Proposition (Definitional implications of Wiener process)**.** *We have that by requirement 2 of a Wiener process W:*

1. *Markov*
$$W_{s+t} - W_s \perp\!\!\!\perp \mathcal{F}_s \; \forall s$$
2. *stationarity*
$$W_{s+t} - W_s \perp\!\!\!\perp s \; \forall s$$
3. *normality*
$$W_{s+t} - W_s \sim \mathcal{N}(0, t) \; \forall t \neq 0$$

Proposition (Martingale characterization of Wiener Process, exponential)**.**

$$W = (W_t)_{t \in \mathbb{R}_+} \underbrace{Wiener}_{Def.\ 11.55} \iff M_t = e^{rW_t - \frac{1}{2}r^2 t} \underbrace{\mathcal{F}\text{-}martingale}_{Def.\ 11.35} \quad \forall r \in \mathbb{R}$$

Proposition (Wiener processes are martingales)**.**

$$W = (W_t)_{t \in \mathbb{R}_+} \underbrace{Wiener}_{Def.\ 11.55} \implies W \underbrace{\mathcal{F}\text{-}martingale}_{Def.\ 11.35}$$

Proposition (Martingale characterization of Wiener process, square)**.**

$$W = (W_t)_{t \in \mathbb{R}_+} \underbrace{Wiener}_{Def.\ 11.55} \implies Y_t = W_t^2 - t \quad \underbrace{\mathcal{F}\text{-}martingale}_{Def.\ 11.35}$$

Theorem (Combination of Wiener martingale characterization). *It actually holds that W is Wiener if and only if:*

1. *W is an $\mathcal{F}$-martingale*
2. *$Y_t = W_t^2 - t$ is an $\mathcal{F}$-martingale*

Namely, the ***two previous results together*** *characterize Wiener processes.*

Definition (Poisson Process $\mathcal{P}ois(c)$). *A counting process $(N_t)_{t\in\mathbb{T}}$ is Poisson with rate $c > 0$ with respect to a filtration $\mathcal{F}$ (Def. 11.2) when:*

1. *N is adapted to $\mathcal{F}$ (Def. 11.7)*
2. *increments are Poisson distributed in expectation:*

$$\mathbb{E}_s\left[f(N_{s+t} - N_s)\right] = \sum_{k=0}^{\infty} \frac{e^{-ct}(ct)^k}{k!} f(k) \quad \forall s, t,\ f \in \mathcal{E}^+$$

Where, as usual, by $f \in \mathcal{E}^+$ we mean a positive measurable function in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the space where the process takes values on.

Proposition (Definitional Properties of $\mathcal{P}ois(c)$). *Definition 12.2 has some direct implications. For $N \sim \mathcal{P}ois(c)$ it holds that:*

1. *markov property*
$$N_{s+t} - N_s \perp\!\!\!\perp \mathcal{F}_s \quad \forall s, t$$
2. *stationarity*
$$(N_t)_{t\in\mathbb{T}} \perp\!\!\!\perp t$$
3. *Poisson increments*
$$N_{t+s} - N_s \sim \mathcal{P}o(ct)$$

Theorem ($\mathcal{P}ois(c)$ characterization). *For a counting process N (Def. 11.13) and a filtration $\mathcal{F}$ over which it is a Poisson process we can see that:*

$$N \sim \mathcal{P}ois(c) \iff (N_t - ct)_{t\in\mathbb{T}} \quad \mathcal{F}\text{-martingale (Def. 11.35)}$$

Definition (Predictable process). *A natural process $(F_n)_{n\in\mathbb{N}}$ is predictable with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ when $F \in \mathcal{F}_0$ and $F_{n+1} \in \mathcal{F}_n\ \forall n$, where by $\in$ we mean measurable with respect to (see the background notation for context).*

Definition (Steltjes-Lebesgue integral). *For $(F_n)_{n\in\mathbb{N}}$ a random function and $(M_n)_{n\in\mathbb{N}}$ a signed measure with mass $M_n - M_{n-1}\ \forall n$ and $M_0 = 1$ we define:*

$$X = (X_n)_{n\in\mathbb{N}} \quad : \quad X = \int F dM$$
$$\iff X_n = \int_{[0,n]} F dM = M_0 F_0 + \sum_{m=1}^{n} (M_m - M_{m-1}) F_m$$

A series of increasing in n integrals.

Theorem (Martingality of integral for bounded processes). *For X a Steltjes-Lebesgue integral as in Definition 12.7 with $(F_n)_{n\in\mathbb{N}}$ bounded (i.e. $\mathbb{P}[|F_n| \leq b] = 1$ for some $b \in \mathbb{R}$) it holds that:*

1. *$(M_n)_{n\in\mathbb{N}}$ martingale $\Longrightarrow$ X martingale*
2. *$(M_n)_{n\in\mathbb{N}}$ submartingale, $(F_n)_{n\in\mathbb{N}}$ positive $\forall n$ $\Longrightarrow$ X submartingale*

Corollary (Stopped time process martingality). *Let T be a stopping time (Def. 11.9) and $(X_n)_{n\in\mathbb{N}}$ with $X_n = M_{n\wedge T}$ as in Example 12.8. Then:*

1. *$(M_n)_{n\in\mathbb{N}}$ martingale $\Longrightarrow$ X martingale*
2. *$(M_n)_{n\in\mathbb{N}}$ submartingale $\Longrightarrow$ X submartingale*

Notice that by the result of Example 12.8 this means that $M_{n\wedge T}$ is a martingale/submartingale since $X_n = M_{n\wedge T}$ for all $n \in \mathbb{N}$.

Proposition (Doob's Theorem I). *Let $(M_n)_{n\in\mathbb{N}}$ be a martingale, T a stopping time (Defs. 11.35, 11.9), with T bounded $\mathbb{P}[T \le k] = 1$ for some $k \in \mathbb{R}$. Then*

$$\mathbb{E}[M_0] = \mathbb{E}[M_T] = \mathbb{E}[M_k]$$

Corollary (Double stopping Time Doob's Theorem I). *Let $(M_n)_{n\in\mathbb{N}}$ be a martingale, T a stopping time (Defs. 11.35, 11.9), with T bounded $\mathbb{P}[T \le k] = 1$ for some $k \in \mathbb{R}$ as before. If $S \le T$ is another stopping time:*

$$\mathbb{E}\left[M_S\right] = \mathbb{E}\left[M_T\right]$$

Theorem (Doob's decomposition). *Let $(X_n)_{n\in\mathbb{N}}$ be adapted to $\mathcal{F}$ and integrable $X_n \in \mathcal{L}_1$. The following statements are true:*

1. *decomposition, with M a martingale, $M_0 = 0$, and $(A_n)_{n\in\mathbb{N}}$ a predictable process, $A_0 = 0$ (Defs. 11.35, 12.5)*
$$X_n = X_0 + M_n + A_n \quad \forall n \in \mathbb{N}$$
2. *the decomposition at point 1 is unique up to equivalence*
3. *if $(X_n)_{n\in\mathbb{N}}$ is a submartingale, $(A_n)_{n\in\mathbb{N}}$ is an increasing predictable process, if $(X_n)_{n\in\mathbb{N}}$ is a supermartingale, $(A_n)_{n\in\mathbb{N}}$ is decreasing predictable*

Theorem (Doob's Theorem II, fully general). *For a process $(M_n)_{n\in\mathbb{N}}$ adapted to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ the following are equivalent:*

1. *$(M_n)_{n\in\mathbb{N}}$ is a martingale*
2. *for bounded stopping times $S \le T$ M_S and M_T are integrable and $\mathbb{E}_S[M_T - M_S] = 0$*
3. *for bounded stopping times $S \le T$ M_S and M_T are integrable and $\mathbb{E}[M_T - M_S] = 0$*

Definition (Upcrossing, downcrossing and counter). *Let $(M_n)_{n\in\mathbb{N}}$ be adapted to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ (Def. 11.7), $a < b$ and $T_0 = -1$ for convenience.*
For all natural $k \ge 1$ define:

$$S_k := \inf\{n \ge T_{k-1} : M_n \le a\}$$
$$T_k := \inf\{n \ge S_k : M_n \ge b\}$$

By the adaptedness of $(M_n)_{n\in\mathbb{N}}$ we can say that $\{S_1, T_1, S_2, T_2, \ldots\}$ is an increasing sequence of stopping times (Def. 11.9).
S_k can be seen as the k^{th} downcrossing of the interval (a, b), while T_k is the k^{th} upcrossing of the interval (a, b). We then define the number of upcrossings of (a, b) as:

$$U_n(a,b) = \sum_{k=1}^{\infty} \mathbb{1}_{[0,n]}\left(T_k\right)$$

Definition (F_n formalism). *The number of buy/sell cycles in $[0, n]$ is exactly $U_n(a, b)$. In this context we let:*

- *F_n be such that:*
$$\begin{cases} F_n = \sum_{n=1}^{\infty} \mathbb{1}_{(S_k,T_k]}(n) \\ F_0 = 0 \end{cases} = \begin{cases} 1 & if\ \exists k : S_k < n \le T_k \\ 0 & else \end{cases}$$
So that F represents the number of stocks owned at $(n, n+1]$
- *we already saw that the value of the portfolio is formalized as:*
$$\begin{cases} X_n = \int_{[0,n]} F dM \\ X_0 = 0 \end{cases}$$

With this context, the profit is in general:

$$X_n - X_0 \ge (b-a)U_n(a,b)$$

where we put $\ge$ instead of $=$ since it could be that the price at the end is less than the price at the start! See the plot of Example 12.22 for reference.

Proposition (Upcrossing inequality).

$$(M_n)_{n\in\mathbb{N}} \quad submartingale \implies (b-a)\mathbb{E}[U_n(a,b)] \leq \mathbb{E}[(M_n-a)^+ - (M_0-a)^+]$$

Theorem (Martingale Convergence Theorem, MCT). *For a submartingale $(X_n)_{n\in\mathbb{N}}$ (Def. 11.36):*

$$\sup_n \mathbb{E}[X_n^+] < \infty \implies X_n \overset{a.s.}{\to} X_\infty, \quad X_\infty \in \mathcal{L}_1$$

*So that we can establish an **almost sure limiting distribution**.*

Corollary (An equivalent sufficient condition). *We can restate the problem in terms of a more useful condition noting that:*

$$\begin{cases} \sup_n \mathbb{E}[X_n^+] < \infty \\ X_n^+ \in \mathcal{L}_1 & \forall n \end{cases} \iff \begin{cases} \sup_n \mathbb{E}[|X_n|] < \infty \\ |X_n| \in \mathcal{L}_1 & \forall n \end{cases}$$

*Namely, an $\mathcal{L}_1$ bound on the martingale, **not** $\mathcal{L}_1$ **convergence!***

Corollary (Special cases). *We recognize a number of familiar situations in which the requirements are easily verified:*

1. $(X_n)_{n\in\mathbb{N}}$ *non positive submartingale*
2. $(X_n)_{n\in\mathbb{N}}$ *non negative supermartingale*
3. $(X_n)_{n\in\mathbb{N}}$ *non positive or non negative martingale*
4. $(X_n)_{n\in\mathbb{N}}$ *bounded above or below by an integrable random variable*

Lemma (A quick Lemma for $\mathcal{L}_1$ convergence). *If $(X_n)_{n\in\mathbb{N}} \overset{\mathcal{L}_1}{\to} X$ then:*

$$\lim_{n\to\infty} \mathbb{E}[X_n Y] = \mathbb{E}[XY] \quad \forall Y \text{ bounded a.s.}$$

Theorem (Uniform Integrability vs a.s. $\mathcal{L}_1$ characterization). *For a martingale $(M_n)_{n\in\mathbb{N}}$ it holds:*

1. *Same convergence by uniformity*

$$\begin{cases} M_n \overset{a.s.}{\to} M_\infty \\ M_n \overset{\mathcal{L}_1}{\to} M_\infty \end{cases} \iff (M_n)_{n\in\mathbb{N}} \quad uniformly\ integrable$$

2. *Martingale equality extends at infinity as a martingale $\overline{X}$ if $M_n = \mathbb{E}_n[Z]$ for $Z \in \mathcal{L}_1$:*

$$M_n = \mathbb{E}_n[Z], Z \in \mathcal{L}_1 \implies M_\infty = \lim_{n\to\infty} M_n \ : \ \overline{X} = (X_n)_{n\in\overline{\mathbb{N}}}$$

Corollary (Appying Theorem the almost sure $\mathcal{L}_1$ limit for u.i. martingales to characterize Observation 12.37). *Conclude that:*

1. $\forall Z : \mathbb{E}[|Z|] < \infty$ *it holds* $\mathbb{E}_n[Z] \underset{\mathcal{L}_1}{\overset{a.s.}{\to}} \mathbb{E}_\infty[Z] = \mathbb{E}[Z|\mathcal{F}_\infty]$
2. $Z \in \mathcal{F}_\infty \implies \mathbb{E}_n[Z] \underset{\mathcal{L}_1}{\overset{a.s.}{\to}} Z$

So that Z is eventually revealed.

Proposition (Frequentist validation of Bayesian mean estimator). *Recall the setting of Example 12.39*

$$identifiability \quad : \quad \mathcal{P}_\theta(A) = \mathbb{P}[\boldsymbol{Y} \in A|\theta] : \mathcal{P}_\theta(\cdot) \neq \mathcal{P}_{\theta'}(\cdot) \forall \theta \neq \theta' \implies \theta \in \mathcal{F}_\infty$$

In other words, identifiability is a sufficient condition for the true value to be revealed at the end of time.

Corollary (Frequentist perspective validation). $\forall \theta_0$ *in almost sure sets of* $\pi(\cdot)$*:*

$$Y_i \overset{iid}{\sim} \mathcal{P}_{\theta_0} = \mathbb{P}[Y_i \in \cdot|\theta_0] \implies \widehat{\theta}_n \to \theta_0 \quad in\ \mathcal{P}_{\theta_0}^\infty \quad a.s.$$

Theorem (Levy's 0-1 law).

$$A \in \mathcal{F}_\infty \implies \mathbb{E}_n[\mathbb{1}_A] \overset{a.s.}{\to} \mathbb{1}_A$$

21.1.2 Random Measures and Processes

Definition (Random Measure $M(\cdot,\cdot)$, r.m.)**.** *The concept is equivalent to that of a Transition Kernel (Def. B.13) from $(\Omega, \mathcal{H})$ onto $(E, \mathcal{E})$. Consider a probability space $(\Omega, \mathcal{H}, \mathbb{P})$ and a measurable space $(E, \mathcal{E})$. A random measure on $(E, \mathcal{E})$ is a mapping:*

$$M : \Omega \times \mathcal{E} \to \overline{\mathbb{R}}_+$$

Such that:

1. *$\omega \to M(\omega, A)$ is a r.v. $\forall A \in \mathcal{E}$ denoted as $M(A)$, which is $\mathcal{H}$-measurable and takes values on $(E, \mathcal{E})$*
2. *$A \to M(\omega, A)$ is a measure on $(E, \mathcal{E})$ denoted as $M_\omega(dx)$ for all $\omega \in \Omega$*

Definition (Measure description of M)**.** *The measure in M denoted as $M_\omega(dx)$ can be atomic or diffuse (Def. A.32), finite, σ-finite or Σ-finite (Defs. A.26, A.27).*

Definition (Random counting measure)**.** *$M(dx)$ such that $M_\omega(dx)$ atomic and with weight 1 a.s. is a random counting measure. It is the equivalent of a counting measure after fixing ω.*

Definition (Recap of integral notation)**.** *Let $f : E \to \mathbb{R}$ be a Borel function and assume we wish to integrate wrt $M(dx)$. Recalling that for a fixed measure ν we have $\nu f = \int f(x)\nu(dx)$ then:*

$$Mf : E \to \mathbb{R} \mid Mf := \int_E f(x)M(dx) \quad \textit{is an r.v.}$$

Notice also that:

$$M(A) = \int_A M(dx) = \int_E \mathbb{1}_A M(dx) = M\mathbb{1}_A \quad \forall A \in \mathcal{E}$$

Definition (Expected version of random measure, mean measure)**.** *For a random measure as in Definition 13.1 we refer to the mean measure ν when considering the measure such that:*

1. *$\nu(A) = \mathbb{E}[M(A)] \quad \forall A \in \mathcal{E}$*
2. *equivalently $\nu f = \mathbb{E}[Mf] \quad \forall f \in \mathcal{E}_+$*

In particular:

$$\nu(A) = \mathbb{E}\left[M(A)\right] = \int_\Omega M(\omega, A)\mathbb{P}[d\omega]$$

Where we are integrating out the ω of the random measure over the underlying probability space.

Lemma (Mean in terms of tail)**.**

$$X \geq 0 \quad a.s. \implies \mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}[X \geq i] = \sum_{i=0}^{\infty} \mathbb{P}[X > i]$$

Definition (Laplace functional)**.** *This definition resembles that of Def. 6.11.*
For a random measure M and a positive Borel function $f \in \mathcal{E}_+$ we define the Laplace functional as:

$$\widehat{\mathcal{P}}_M(f) = \mathbb{E}\left[e^{-Mf}\right]$$

Which can be seen as the Laplace transform of Mf, which is a r.v., evaluated at $r = 1$.

Definition (Poisson random measure, p.r.m.)**.** *$N(dx) \sim \mathcal{P}ois(\nu(dx))$ is a Poisson random measure (Def. 13.1) with mean measure $\nu(dx)$ when:*

1. *$N(A) \sim \mathcal{P}o(\nu(A)) \quad \forall A \in \mathcal{E}$*
2. *For $\{A_i\}_{i=1}^n \subset \mathcal{E}$ disjoint $\implies \{N(A_i)\}_{i=1}^n$ is an independency (Def. 6.9)*

Proposition (Mean Variance for sets of Poisson random measure)**.** *For $N(dx) \sim \mathcal{P}ois(\nu(dx))$ such that $\nu(A) < \infty \quad \forall A \in \mathcal{E}$:*

1. *$\mathbb{E}[N(A)] = \nu(A)$*
2. *$V[N(A)] = \nu(A)$*
3. *If $\nu(A) = \infty \implies \mathbb{E}[N(A)] = \infty$ a.s. and $V[N(A)]$ is undefined a.s.*

Proposition (Mean and variance for functions, Poisson random measure). *Let N be a p.r.m. and $f \in \mathcal{E}_+$:*

1. $\mathbb{E}[N(f)] = \nu(f)$
2. $V[N(f)] = \nu(f^2)$ *if* $\nu f < \infty$

Theorem (Laplace functional of Poisson random measure characterization). *Using the theory of Laplace transforms, for a random measure N on $(E, \mathcal{E})$ (Def. 13.1) with mean measure ν:*

$$N \sim \mathcal{P}ois(\nu) \quad (\text{Def. } 13.13) \iff \mathbb{E}[e^{-Nf}] = e^{-\nu(1-e^{-f})} \quad \forall f \in \mathcal{E}_+$$

Lemma (Laplace functional uniqueness and continuity). *The Laplace functional mapping $f \to \mathbb{E}[e^{-Mf}]$ for $f \in \mathcal{E}_+$ is such that:*

1. $(f_n) \subset \mathcal{E}, f_n \nearrow f \implies \lim_{n\to\infty} \mathbb{E}\left[e^{-Mf_n}\right] = \mathbb{E}\left[e^{-Mf}\right]$
2. $N = M$ *on* $(E, \mathcal{E})$ *random measures* $\iff \widehat{\mathcal{P}}_M(f) = \widehat{\mathcal{P}}_N(f)\ \forall f \in \mathcal{E}_+$

Corollary (Extending the results of Theorem). *Clearly:*

$$\widehat{\mathcal{P}}_M(f) = \widehat{\mathcal{P}}_N(f) \quad \forall f \in \mathcal{E}_+ \iff M = N \quad a.s. \iff M \ r.m. \quad specified\ by\ \nu\ only$$

Proposition (Laplace function of $N(A)$). *We provide quickly an intuition of the $\implies$ direction in the Proof of Theorem for the simplest case possible.*
We can show for $r = 1$ that :

$$\mathbb{E}\left[e^{-1\cdot N(A)}\right] = \exp\{-\nu(1 - e^{-\mathbb{1}_A})\}\}$$

and then reason by simple functions approximation.

Definition (Proper random variable for random measure). *Given $f \in \mathcal{E}_+$ we say $Mf = \int_E f(x)M(dx)$ is proper when $\mathbb{P}[Mf < \infty] = 1$, namely $Mf \overset{a.s.}{=} 1$.*

Lemma (Finiteness of random variable by Laplace function).

$$X \geq 0 \quad a.s. \implies \mathbb{P}[X < \infty] = \lim_{r\to 0} \widehat{\mathcal{P}}_X(r)$$

Proposition (Finiteness of Poisson random measure). *Let $f \in \mathcal{E}_+$ and $N \sim Pois(\nu)$. Then:*

$$\nu(f \wedge 1) < \infty \implies Nf < \infty \quad a.s.$$

Else $Nf = \infty$ a.s.

Definition (Independent random measures). *Two random measures N, M are such that $N \perp\!\!\!\perp M$ when $N(A) \perp\!\!\!\perp M(A)\ \forall A \in \mathcal{E}$*

Theorem (Poisson random measure existance). *Let ν be Σ-finite on $(E, \mathcal{E})$. Then:*

$$\exists(\Omega, \mathcal{H}, \mathbb{P}) \quad \& \quad N(\omega, \cdot)\ on\ (E, \mathcal{E}) : N \sim Pois(\nu) \quad \forall \omega \in \Omega$$

Theorem (Random counting measure and diffusivity of Poisson random measure). *Let N be a p.r.m. on $(E, \mathcal{E})$ according to Definition 13.13, with Σ-finite mean measure ν. Then:*

$$N \quad random\ counting\ measure\ (\text{Def. } 13.3) \iff \nu \quad diffuse\ (\text{Def. } A.32)$$

Corollary (Extension to special case). *Let $N \sim Pois(\nu)$ on $E = \mathbb{R}_+ \times \mathbb{R}_+$ and $\mathcal{E} = \mathcal{B}(E)$.*
Let $\nu = Leb \times \lambda$, with

- $\lambda(\{0\}) = 0$
- $\lambda((\epsilon, \infty)) < \infty \quad \forall \epsilon > 0$

We can interpret $N(t, z)$ for a time of arrival t of an object of size z. Then:

1. *for a.e. $\omega \in \Omega$ N_ω is a counting measure that:*
2. *(no simultaneity) has not atom at $t = 0$, no atom of size $z = 0$, i.e. no simultaneity of $X_i, X_j : t_i = t_j$*
3. *(finite big activity) $\forall t < \infty, \epsilon > 0$ there are finitely many atoms before t with size $z > \epsilon$*

4. *(infinite small activity) claim #3 holds for* $\epsilon = 0$ *if* λ *is finite. Otherwise there are* ∞ *many atoms of size* $z \leq \epsilon \quad \forall \epsilon > 0$

Definition (Image of N under h, $N \circ h^{-1}$)**.** *This is equivalent to Definition B.1.*
Let N *be a p.r.m. on* $(E, \mathcal{E})$*, and* $h : E \to F$ *a measurable map (satifies Eqn. 3.1). The image of* N *under* h *is a random measure on* $(\Omega, \mathcal{H}, \mathbb{P}), (F, \mathcal{F})$ *(Def. 13.1) defined as:*

$$N \circ h^{-1} \quad : \quad (N \circ h^{-1})(B) = N \circ (h^{-1}(B)) \quad \forall B \in \mathcal{F}$$

the last expression is:

$$\begin{aligned} N(h^{-1}(B)) &= \int_E \mathbb{1}_{h^{-1}(B)}(x) \, N(dx) \\ &= \int_{x: h(x) \in B} N(dx) \\ &= \int_E \mathbb{1}_B(h(x)) \, N(dx) \\ &= N(\mathbb{1}_B \circ h) \end{aligned}$$

Where we infer that instead for a borel map $f : F \to \mathbb{R}$*:*

$$(N \circ h^{-1})(f) = N(h^{-1}(f)) = N(f \circ h)$$

which by $Nf = \sum f(X_i)$ *for* (X_i) *atoms of* N *suggests that:*

$$(N \circ h^{-1})(f) = N(f \circ h) = \sum_{i=1}^{K} f(h(X_i)) = \sum_{i=1}^{K} f(Y_i)$$

For (Y_i) *the atoms of* $N \circ h^{-1}$*.*

Proposition (Image measure is a Poisson random measure)**.** $N \circ h^{-1}$ *on* $(\Omega, \mathcal{H}, \mathbb{P}), (F, \mathcal{F})$ *satisfies the requirements of Definition 13.13 and has mean* $\mu = \nu \circ h^{-1}$*.*

$$N \sim \mathcal{P}ois(\nu), \quad h : E \to F \implies N \circ h^{-1} \sim \mathcal{P}ois(\nu \circ h^{-1})$$

Definition (Arrival process formalism)**.** *Let* $N(dx)$ *be a p.r.m. on* $E = \mathbb{R}_+$ *with diffuse mean* $\nu(dx)$*, such that* $c(t) = \nu((0, t]) < \infty \; \forall t$*.*
By Theorem the previous Theorem we know that ν *is diffuse* $\iff$ N *is a random counting measure.*
With this premise we can interpret $(T_k)_{k \geq 1}$ *as distinct ordered arrival times. We want to simulate this random measure.*

Proposition (Arrival process simulation by inverse image)**.** *For* N *as in Definition the arrival process formalism of Definition 14.16 let* $h : \mathbb{R}_+ \to \mathbb{R}_+$ *be such that in the arrival process formalism:*

$$h(u) = t \iff c(t) = u$$

Namely, the inverse of the cdf. Then for $\widetilde{N}$ *a p.r.m. with mean measure Leb:*

1. $\nu = Leb \circ h^{-1}$
2. $(u_i)_{i \geq 1}$ *are the atoms and* $(h(u_i))_{i \geq 1}$ *are the atoms of* N

Definition (Trace of random measure, also restriction)**.** *For* $D \subset E$ *and a random measure* M *on* E *we call restriction the measure* M_D *characterized as:*

$$M_D(B) \coloneqq M(B \cap D) \quad \forall B \in \mathcal{E}$$

Which has mean $\mu_D(B) = \mu(B \cap D) \; \forall B \in \mathcal{E}$

Definition (Intensity or expected arrival time r). *In the context of the arrival process formalism of Definition 14.16 further let ν be σ-finite and such that $\nu \ll Leb$. By Radon Nikodym Theorem we have that:*

$$\exists r \text{ Leb-measurable}, \quad \nu(A) = \int_A r(t)dt$$

We call $r(t) = \frac{d\nu}{dLeb}(t)$ the Radon-Nykodym derivative also with the term intensity.
Recall the discussion we did in Chapter 5. It is not granted that the measure ν will be σ-finite once it is absolutely continuous to the Lebesgue measure. The observation we did when introducing the Radon-Nykodim theorem made it precise that this requirement was lifted for probability measures, but ν in principle could be just a measure. This comment can be ignored in most of the cases.

Proposition (Arrival process simulation by intensity). *Using the interpretation of Definition 14.19 for an intensity r we also let:*

- $h(t, z) = t$
- $D = \{(t, z) : z \leq r(t)\} \subset \mathbb{R}_+ \times \mathbb{R}_+$
- *M_D be the trace of the p.r.m. M on $\mathbb{R}_+ \times \mathbb{R}_+$ with mean Leb, so that it is a p.r.m. with mean $\mu_D \ll Leb$. The mean measure μ_D is also σ-finite since it is just a restriction of Leb inside the set D*

Then:

1. *$N = M_D \circ h^{-1}$ is a p.r.m. with mean $\nu = \mu_D \circ h^{-1}$. N here is the counting measure on $\mathbb{R}_+$ whose atoms are arrival times T_i with size $Z_i \leq r(T_i)$, according to the restriction D.*
2. *can simulate $(T_i, Z_i)_{i\geq 1}$ from M and set $Nf = \sum_{i:Z_i \leq r(T_i)} f(T_i)$*

Assumption 21.1 (Setting for transformations). *We consider measurable spaces $(E, \mathcal{E}), (F, \mathcal{F})$, and collections $\{X_i : i \in I\}, \{Y_i : i \in I\}$.*
N is a p.r.m. on $(E, \mathcal{E})$ with mean ν (Def. 13.13) $\implies Nf = \sum_{i\in I} f \circ X_i \quad f \in \mathcal{E}_+$.
For a measurable map $h : E \to F$, satisfying Equation 3.1, we set $Y_i = h \circ X_i$ and derive the new p.r.m. $N \circ h^{-1}$ using the Proposition in which we proce the image measure is a random measure.
Y_i is ultimately the random transform associated to the kernel (Def. B.13):

$$Y_i \in B \text{ w.p. } Q(x, B) \quad if\ X_i = x \iff \mathbb{P}[Y \in B|X = x] = Q(x, B) \quad \forall B \in \mathcal{F}$$

Where $Q : E \times \mathcal{F} \to E$

Theorem (Transformation independence poissonity). *For a measure ν on $(E, \mathcal{E})$, and a kernel Q from $(E, \mathcal{E})$ to $(F, \mathcal{F})$ such that:*

- *X is a p.r.m. with mean ν*
- $Y_i|X \overset{ind}{\sim} Q(X_i, \cdot)$

It holds:

1. *Y is a p.r.m. on $(F, \mathcal{F})$ with mean $\pi(Q) : \pi(Q(B)) = \int_F \nu(dx)Q(x, B) \quad \forall B \in \mathcal{F}$ or in other terms $\pi(dy) = \int_F Q(x, dy)\nu(dx)$*
2. *(X, Y) is a p.r.m. on $(E \times F, \mathcal{E} \otimes \mathcal{F})$ with mean $\mu = \nu \times Q$ so that:*

$$\mu(dx, dy) = \nu(dx)Q(x, dy)$$

Corollary (Special case Kernel is probability measure). *For $X \sim \mathcal{P}ois(\nu)$ on $(E, \mathcal{E})$ and $Y \perp\!\!\!\perp X$ such that $Y \sim \pi$ on $(F, \mathcal{F})$:*

$$\implies (X, Y) \sim \mathcal{P}ois(\mu) \quad on \quad (E \times F, \mathcal{E} \otimes \mathcal{F}) \quad \mu = \nu \times \pi : \mu(dx, dy) = \nu(dx)\pi(dy)$$

Definition (Compound Poisson process $(S_t)_{t\geq 0}$). *We give a precise definition of the object presented in the above example.*
For arrival times $T_1 < T_2 < \dots$ atoms of a p.r.m. on $\mathbb{R}_+$ with mean $cdx = \nu(dx)$ we consider a sequence of random variables $Y_i \overset{iid}{\sim} \pi$ on $\mathbb{R}$ where $Y \perp\!\!\!\perp T$.

The compound Poisson process that arises is the continuous time process of the random sum:

$$(S_t)_{t\geq 0} \ : \ S_t = \sum_{i:T_i\leq t} Y_i = \sum_i \mathbb{1}_{[0,t]}(T_i)\, Y_i$$
$$= \int_{\mathbb{R}_+\times\mathbb{R}} y\mathbb{1}_{[0,t]}(x)\, N(dx,dy)$$

Where by the previous Theorem N is a p.r.m. and the expression makes sense.

Definition (Borel version of compound Poisson process)**.** *$(S_t)_{t\in\mathbb{R}_+}$ can be seen as a cumulative version of a r.m. (Def. 13.1) on $\mathbb{R}_+$:*

$$S_t = L((0,t]) \quad L(dx)\ r.m.\ :\ L(A) = \int_{A\times\mathbb{R}} yN(dx,dy)$$

Indeed the Laplace transform of S_t would be:

$$\mathbb{E}\left[e^{-rL((0,t])}\right] = \mathbb{E}\left[e^{-rS_t}\right] = \exp\left\{-ct\int_{\mathbb{R}_+}(1-e^{-ry})\pi(dy)\right\}$$
$$= \exp\left\{\int_{(0,t]\times\mathbb{R}_+}(1-e^{-ry})dxc\pi(dy)\right\}$$

Which we write for general A below the Observation that follows.

Definition (Additive random measure)**.** *A random measure (Def 13.1) M is said to be additive when for disjoint sets $\{A_i\}_{i=1}^n \subset \mathcal{E}$ the set of random variables $\{M(A_i)\}_{i=1}^n$ is an independency according to Definition 6.9.*

Proposition (Compound Poisson process has underlying additive measure)**.** *L as in Definition 15.5 is an additive random measure.*

Lemma (Automatic additive random measure)**.** *For a countable set $D \subset E$ and an independency of positive random variables $\{W_x\ :\ W_x \geq 0\ x \in D\}$ the random measure:*

$$K(\omega, A) = \sum_{x\in D} W_x(\omega)\mathbb{1}_A(x) \quad \omega\in\Omega, A\in\mathcal{E}$$

is additive.

Theorem (A form of additive random measure decomposition)**.** *Consider a measure α on $(E,\mathcal{E})$, a random measure K as in the Lemma for the automatic random measure, purely atomic with fixed atoms, and a random measure L as in the previous Proposition, namely:*

$$L(A) = \int_{A\times\mathbb{R}_+} yN(dx,dy) \quad N \sim \mathcal{P}ois(\nu)$$

Then:

1. *any additive r.m. (Def. 15.8) can be decomposed in a sum $M = \alpha + K + L$*
2. *if M is a Σ-bounded kernel (Def. B.23) the same decomposition holds and if additionally α is diffuse, and the mean measure of K $\nu(\cdot\times\mathbb{R}_+)$ is diffuse the decomposition is unique*

Definition (Increasing Lévy process)**.** *A process $S = (S_t)_{t\in\mathbb{R}_+}$ is increasing Lévy when it is such that:*

1. *independence of increments:*
$$S_{t_1} - S_{t_0}, \ldots, S_{t_n} - S_{t_{n-1}} \perp\!\!\!\perp \quad \forall n\geq 2, 0\leq t_0 < t_1 < \ldots < t_n$$
2. *stationarity of increments*
$$S_{t+u} - S_u \overset{d}{=} S_t \quad \forall u,t\in\mathbb{R}_+$$
3. *increasing, right continuous and starting at $S_0 = 0$*

Assumption 21.2 (Structure of compound Poisson process revisited)**.** *We know by the previous Proposition in which we proved that the compound Poissoin process is additive, that the underlying random measure of a compound Poisson process is additive. We now impose that:*

- $S_t(\omega) = M(\omega, [0,t])$ *for* M *an additive r.m., so that* S_t *is increasing and right continuous*
- $S_t < \infty \quad a.s. \forall t$ *which will ensure independence by the additivity of* M
- $\alpha(dx) = bdx \quad b \in \mathbb{R}_+$ *to ensure linearity, which will guarantee stationarity of increments*

Definition (Candidate Poisson additive random measure)**.** *We present here the r.m. we will feed to the following results, carefully constructed according to Assumption 15.15 and Observation 15.13:*

$$S_t = bt + \int_{[0,t]\times\mathbb{R}_+} zN(dx,dz) = M(\omega,[0,t]) \quad b,t \in \mathbb{R}_+, M \text{ additive}$$

for N *a Poisson random measure with mean* $\nu(dx,dz) = Leb \times \lambda(dz)$*.*

Proposition (Candidate compound Poisson with weak integrability is increasing Lévy)**.** *Let* $b \in \mathbb{R}_+$*,* N *a p.r.m. on* $\mathbb{R}_+ \times \mathbb{R}_+$ *with mean* $\nu = Leb \times \lambda$*. If the integrability condition:*

$$\int_{\mathbb{R}_+} \lambda(dz)(z \wedge 1) = \lambda(z \wedge 1) < \infty$$

is satisfied then:

1. *(Lévyness)* $(S_t)_{t\in\mathbb{R}_+}$ *the candidate of Definition 15.16 is an increasing Lévy process in the sense of Definition 15.14*
2. *(characterization) the Laplace transform is:*

$$\mathbb{E}[e^{-rS_t}] = \exp\left\{-t\left[br + \int_{\mathbb{R}_+} \lambda(dz)(1-e^{-rz})\right]\right\} \quad r \in \mathbb{R}_+$$

Definition (Lévy process terminology)**.** *We say that:*

- $b \in \mathbb{R}_+$ *is the drift*
- λ *is the Lévy measure*

Where the two of them uniquely identify S *via the Laplace transform of the previous Proposition.*

Lemma (Finite measures Lévyness)**.** *It is rather easy to check for* $E \subset \mathbb{R}_+$ *that:*

$$\lambda : \lambda(E) < \infty \implies \lambda(z \wedge 1) < \infty$$

and we can say that the candidate Poisson process is Lévy.

Proposition (Link integrability & infinite activity)**.** *Consider a measure on* $\mathbb{R}_+$ *which is not finite. We link the above Proposition and the Corollary for diffusivity of the counting measure by:*

$$\int_0^\infty (z\wedge 1)\lambda(dz) < \infty \implies \lambda((\epsilon,\infty)) < \infty \quad \forall \epsilon > 0$$

$$\text{but still} \quad \lambda((\epsilon,\infty)) \overset{\epsilon\to 0}{\to} \infty$$

Definition (Inverse Gaussian distribution)**.** *We consider a stable process* $(S_t)_{t\in\mathbb{R}_+}$ *as in Example 15.23 with* $a = \frac{1}{2}, c = \sqrt{2}$*. The Lévy measure becomes:*

$$\lambda(dz) = \frac{1}{\sqrt{2\pi z^3}}dz$$

The density associated to such measure is available in closed form:

$$f(z) = \frac{t}{\sqrt{2\pi z^3}}e^{-\frac{t^2}{2z}} \quad z \in \mathbb{R}_+$$

We know that this is the density function of an inverse gaussian distribution, so we can safely say that $S_t \overset{d}{=} \frac{a^2}{Z^2}$ *for* $Z \sim \mathcal{N}(0,1)$ *and write* $S_t \sim \mathcal{IN}(a)$*.*

Definition (Generalized inverse j). *To implement the truncation, we make use of:*

$$j(u) = \inf\{\epsilon > 0 \ : \ \lambda((\epsilon, \infty)) < u\}$$

Where the inf *accounts for possible discontinuities. Notice that j is decreasing since λ is decreasing in ϵ.*

Proposition (Generalized inverse properties). *We have that:*

1. $\lambda(A) = (Leb \circ j^{-1})(A) = Leb(\mathbb{1}_A \circ j) \quad \forall A \in \mathcal{E}$
2. $\lambda(f) = Leb(f \circ j)$
3. $S_t = \sum_{i=1}^{\infty} j(U_i)\mathbb{1}_{[0,t]}(T_i) = \sum_{i:T_i \le t} j(U_i)$ *for* $((U_i, T_i))_{i\ge1} \sim N(dx, du)$

Definition (Incomplete Gamma function $\Gamma(s, x)$). *Also known as upper incomplete gamma function:*

$$\gamma_s(x) = \Gamma(s, x) = \int_x^{\infty} t^{s-1} e^{-t} dt$$

Where for $s = 0$ we see that $\Gamma(0, x) = \gamma_0(x) \overset{x\to0}{\to} \infty$.

Lemma (Incomplete gamma-χ^2 link). *Let $\chi^2_{d,(qt)}$:= upper quantile of the chi-square distribution such that $\mathbb{P}[\chi^2_d > \chi^2_{d,(\alpha)}] = \alpha$. Then:*

1. $\gamma_0(u) = \Gamma(0, u) \overset{d\to0}{\approx} \frac{1}{2}\chi^2_{d,(\frac{du}{2})}$
2. *Accordingly:*

$$S_t \overset{d\to0}{\approx} \sum_{i=1}^{\infty} \frac{1}{c}\frac{1}{2}\chi^2_{d,(\frac{d}{2}\frac{G_i}{at})}$$

Theorem (Poisson process, random measure & counting process equivalence). *Let $c > 0$, TFAE:*

1. *M is a p.r.m. (Def. 13.13) with mean $\mu = cLeb$*
2. *N is a poisson (counting) process (Defs. 12.2, 11.13) with rate c*
3. *N is a counting process (Def. 11.13) and $\widetilde{N} = (N_t - ct)_{t\ge0}$ is an $\mathcal{F}$-martingale (Def. 11.35)*
4. *$(T_k)_{k\ge1}$ is an increasing sequence of $\mathcal{F}$-stopping times (Def. 11.9) and:*

$$T_1, T_2 - T_1, \ldots \overset{iid}{\sim} \mathcal{E}xp(c)$$

Theorem (Poisson increasing Lévy characterization). *For a counting process N (Def. 11.13) we conclude that:*

$$N \quad \text{increasing Lévy (Def. 15.14)} \iff N \quad \text{Poisson (Def. 12.2)}$$

Proposition (Strong Markov Property of Poisson Processes). *We establish independence of future events from the past even when the present is a stopping time.*
For a Poisson process $N \sim \mathcal{P}ois(c)$ and a stopping time S:

$$\mathbb{E}_S[f(N_{S+t} - N_S)\mathbb{1}_{\{S<\infty\}}] = \sum_{k=0}^{\infty} f(k)\frac{e^{-ct}(ct)^k}{k!}\mathbb{1}_{\{S<\infty\}}$$

Proposition (Total unpredictability of jumps). *This result is mirroring that of the Proposition in which we prove that Brownian motion's hitting time process hits zero almost surely, which will be proved later.*
Consider a Poisson process N, of which the first jump is $T = T_1$, and a stopping time S wrt $\mathcal{F}$. Then:

$$0 \le S < T \quad a.s. \implies S = 0 \quad a.s.$$

Namely, we cannot find a sequence of stopping times that would approximate T.

21.1.3 Continuous Time Processes and Path Properties

Definition (Lévy process). *A process $X = (X_t)_{t\in\mathbb{R}_+}$ is Lévy wrt a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ if:*

1. *(adaptedness) it is adapted to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ (Def. 11.7)*

2. *(right continuity and starts at zero) for a.e.* $\omega \in \Omega$ *the path* $t \to X_t$ *is right continuous and* $X_0(\omega) = 0$
3. *(stationary and independent increments)* $\forall s, t \geq 0 \ X_{s+t} - X_s \perp\!\!\!\perp \mathcal{F}_s \overset{d}{=} X_t$

Definition (Infinite divisibility)**.** *We express a Lévy process* $(X_t)_{t\in\mathbb{R}_+}$ *as* $\sum_{i=1}^n X_i = X_t \quad \forall n, t$ *where the elements are all Lévy processes:*

$$\delta = \frac{t}{n} \implies X_t = X_t \mathbb{1}_{[0,\delta]}(t) + X_t \mathbb{1}_{[\delta,2\delta]}(t) + \cdots + X_t \mathbb{1}_{[\delta(n-1),\delta n]}(t)$$

where the increments are independent and identically distributed.

Definition (Characteristic exponent $\psi(r)$)**.** *This is a direct result of the infinite divisibility, which makes the process decompose into independent processes. A Lévy process can be described by the characteristic exponent, a complex valued function such that:*

$$\Phi_{X_t}(r) = \mathbb{E}[e^{irX_t}] = e^{t\psi(r)} \quad t \in \mathbb{R}_+, r \in \mathbb{R}$$

Where $\psi : \mathbb{R} \to \mathbb{C}$ *is complex valued.*

Definition (Pure jump process)**.** *Consider on* $\mathbb{R}$ *the Lévy process*

$$X_t = \sum_{s\in[0,t]\cap D_\omega} \Delta X_s \qquad \forall t \quad \Delta X_s = X_s(\omega) - X_{s-}(\omega),\ D_\omega = \{t > 0 : \Delta X_t(\omega) \neq 0\}$$

Then:

- *the jumps are positive or negative, we could see* $X_t = X_t^+ + X_t^-$ *where both are increasing Lévy*
- *if the jumps are countable (e.g. arising from arrival times* (T_n)*) then we can evaluate the sum, we do so by intersecting the time interval with* D_ω

We call this a pure jump process, notice that it is not necessarily ***increasing****.*

Definition (Total Variation V_t of the pure jump)**.** *We give a first definition of total variation of a path of a pure jump process* $t \to X_t$ *as:*

$$V_t = \sum_{s\in[0,t]\cap D_\omega} |\Delta X_s| \quad \forall t \in \mathbb{R}_+$$

Proposition (General representation & existance conditions of Lévy process)**.** *For a p.r.m.* M *on* $\mathbb{R}_+ \times \mathbb{R}^d$ *with mean* $Leb \times \lambda$ *and* $\lambda(\{0\}) = 0$ *if:*

$$\lambda(|x| \wedge 1) = \int_{\mathbb{R}^d} \lambda(dx)(|x| \wedge 1) < \infty \tag{21.1}$$

then:

1. *for a.e.* ω *the process arising from the integral* $X_t(\omega) = \int_{[0,t]\times\mathbb{R}^d} M_\omega(ds, dx)x$ *converges absolutely* $\forall t$ *and it has bounded total variation* $V_t < \infty \quad \forall t$
2. X *is a pure jump Lévy process with characteristic exponent*

$$\psi(r) = \lambda(e^{ir\cdot x} - 1) = \int_{\mathbb{R}^d} \lambda(dx)(e^{ir\cdot x} - 1) \quad \forall r \in \mathbb{R}$$

Definition (Basis notation)**.** *denote:*

- $\mathbb{B} = \{x \in \mathbb{R} \ : \ |x| \leq 1\}$
- $\mathbb{B}_\epsilon = \{x \in \mathbb{R} \ : \ \epsilon < |x| \leq 1\}$ *for* $\epsilon \in (0, 1)$

Theorem (Infinite total variation Lévy existance as compensated sum of jumps)**.** *Let* $M \sim \mathcal{P}ois(Leb \times \lambda)$ *(Def. 13.13) on* $\mathbb{R}_+ \times \mathbb{B}$ *where* $\lambda(\{0\}) = 0$ *and:*

$$\lambda(|x|^2 \mathbb{1}_{\mathbb{B}}) = \int_{\mathbb{B}} \lambda(dx)|x|^2 < \infty \tag{21.2}$$

For $\epsilon \in (0, 1)$ *consider:*

$$X_t^\epsilon(\omega) = \int_{[0,t]\times\mathbb{B}_\epsilon} xM_\omega(ds, dx) - t\int_{\mathbb{B}_\epsilon} \lambda(dx)x \quad \omega \in \Omega, t \in \mathbb{R}_+$$

Then:

1. $\exists X$ *Lévy such that* $\lim_{\epsilon\downarrow 0} X_t^\epsilon(\omega) \overset{a.s.}{=} X_t(\omega)$ *uniformly convergent over bounded intervals*
2. $\psi(r) = \int_{\mathbb{B}} \lambda(dx)(e^{irx} - 1 - irx) \quad r \in \mathbb{R}$

Lemma (An identity for distribution and Φ). *For* X *a r.v. on* $\mathbb{R}$ *with density* $f(x)$*:*

$$\Phi_X(r) = \int_{\mathbb{R}} e^{irx} f(x)dx \iff f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-irx}\Phi_X(r)dr$$

Definition (Brownian motion). *A process* $X = (X_t)_{t\in\mathbb{R}_+}$ *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *such that:*

1. *the path* $t \to X_t$ *is continuous*
2. *it has stationary and independent increments*

Theorem (Lévy characterization as Wiener). *As a first step, notice that:*

1. $X_t = at + bW_t$ *continuous Lévy* $\implies W_t$ *Wiener*
2. W_t *Wiener* $\implies X_t = at + bW_t$ *continuous Lévy*

Which establish an $\iff$ *relation*

Corollary (Applying Theorem in the Brownian-Wiener-Lévy context). *Combine the results to obtain:*

$$X_t \text{ Brownian} \overset{Obs.\ 18.3}{\implies} X_t - X_0 \text{ Lévy} \overset{Thm.\ Lévy\text{-}Wiener}{\iff} X_t - X_0 = bt + cW_t \ : \ W_t \text{ Wiener}$$

Definition (Brownian motion decomposition). *We build a Brownian motion from Definition 18.2 as:*

$$X_t = X_0 + bt + cW_t$$

for a drift coefficient b*, a volatility coefficient* c *and a Wiener process* W*.*

Definition (Wiener process as Brownian motion revisited). *According to our results, a Brownian motion* $W = (W_t)_{t\in\mathbb{R}_+}$ *with* $W_0 = 0, \mathbb{E}[W_t] = 0, V[W_t] = t$ *(namely,* $X_0 = 0, b = 0, c = 1$*) is also a Wiener process!*
We will see in the next result that a Wiener process is a Gaussian process with continuity. However, while constructing a Gaussian process is immediate, as it is only required to specify the functions m, K *(see Def. 10.61), it is not granted that there exists a probability space where such process is continuous. We will eventually see that this condition is satisfied, but the question at the moment is proving that Wiener processes exist in the Brownian formulation.*

Lemma (Gaussian Transformation linearity). *Quickly recall that:*

$$X \sim \mathcal{N}^d(\mu, \Sigma), \quad A \in \mathbb{R}^p \times \mathbb{R}^d \implies Y = AX \sim \mathcal{N}^d(A\mu, A\Sigma A^T)$$

Theorem (Wiener-Gaussian characterization). *The previous observations suggest a useful conclusion. For* $W = (W_t)_{t\in\mathbb{R}_+}$ *a process on* $\mathbb{R}$ *we establish:*

$$W \text{ Wiener} \iff \begin{cases} W & \text{continuous} \\ W \sim \mathcal{GP}(m, k) & \text{Gaussian Def. 10.61} \\ E[W_t] = 0, CoV[W_s, W_t] = s \wedge t \end{cases}$$

where $m(t) \equiv 0$ *and* $k(s,t) = s \wedge t$*.*

Lemma (Kolmogorov's maximal inequality). *Assume* $\{X_i\}_{i=1}^n$ *is an independency where* $\mathbb{E}[X_i] = 0 \forall i$*. Setting* $S_n = \sum^n X_i$*:*

$$a^2\mathbb{P}\left[\max_{k\le n}|S_k| > a\right] \le V[S_n]$$

Definition (Recap about Dyadic rationals). *Dyadic rationals are also discussed in Lemma A.17. Here we denote them as:*

$$D = \{x \in \mathbb{R}_+ \ : \ x = k2^{-m}, \ k, m \in \mathbb{N}\}$$

Proposition (Dyadics are dense in $\mathbb{R}$). *This is a very important result. It is reported here to reference it when needed.*

$$\forall t \in \mathbb{R}, \forall \epsilon > 0 \quad \exists k, m \in \mathbb{N} : t \in (k2^{-m}, (k+1)2^{-m}], \ t - k2^{-m} < \epsilon$$

which is the exact definition of dense set.

Theorem (Wiener Process properties, Brownian formulation). *Let $W = (W_t)_{t\in\mathbb{R}_+}$ be a Wiener process according to the updated Definition of Wiener process as Brownian motion. Then:*

1. *(symmetry) $(-W_t)_{t\in\mathbb{R}_+}$ is Wiener*
2. *(scaling) $(W_{ct})_{ct\in\mathbb{R}_+}$ Wiener $\implies \widehat{W} = (\sqrt{c}W_t)_{t\in\mathbb{R}_+}$ is Wiener $\forall c \in (0,\infty)$*
3. *(time inversion) setting $\widetilde{W}_0 = 0$ for convention, the process $\widetilde{W}_t = tW_{\frac{1}{t}}$ is Wiener*

In the previous proof we make use of right continuity only. We are missing the continuity stated in the Wiener-Brownian Definition for Wiener processes, which would ensure existance of the process $W = (W_t)_{t\in\mathbb{R}_+}$. We do so by Kolmogorov extension. For a finite collection of ordered times $\{t_i\}_{i=1}^n$ the multivariate Gaussian with mean zero and covariance $\{t_i \wedge t_j\}$ is denoted as $\mu_{t_1,\dots,t_n}$. Such distribution is a consistent family, in the sense of Definition 10.54. For $\Omega = \{\omega : \mathbb{R}_+ \to \mathbb{R}\}$ a set of functions put:

$$W_t(\omega) = \omega(t) \quad \mathcal{H} = \sigma\{\omega \in \Omega : \omega(t) \in A, A \in \mathcal{B}(\mathbb{R})\}$$

Using Kolmogorov extension Theorem:

$$\exists!\mathbb{P} \; on \; (\Omega, \mathcal{H}) \quad s.t. \quad \mathbb{P}[\omega : \omega(t_i) \in A_i \; \forall i = 1,\dots,n] = \mu_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n)$$

Yet to evaluate continuity, we would need to check the paths $t \to \omega(t)$ at uncountable time points, while $\mathcal{H}$ is constructed from countably many coordinates. The collection of continuous functions is not a priori measurable. What we do is a **modification** of the process $(W_t)_{t\in\mathbb{R}_+}$ into $\left(\widetilde{W}_t\right)_{t\in\mathbb{R}_+}$ so that:

$$\forall t \exists \Omega_t : W_t = \widetilde{W}_t, \quad \widetilde{W}_t \; continuous$$

where Ω_t is an almost sure set.
Precisely, we establish Hölder continuity instead of continuity.

Definition (Uniform continuity). *the map $t \to W_t$ is uniformly continuous if:*

$$\forall \epsilon > 0 \exists \delta = \delta(\epsilon) \quad |t-s| < \delta \implies |W_t - W_s| < \epsilon \quad \forall t, s$$

Where δ depends on ϵ only.

Definition (Hölder continuity). *The path $t \to W_t$ is Hölder continuous when:*

$$\exists C \in \mathbb{R} \quad |W_t - W_s| \leq C|t-s|^\alpha, \quad for \; \alpha \in [0,1] \quad \forall t, s$$

Which means uniform continuity for $\delta = (\frac{\epsilon}{C})^{\frac{1}{\alpha}}$.

Lemma (Kolmogorov moment condition). *For a process $X = (X_t)_{t\in[0,1]}$ on $\mathbb{R}$ and D the dyadic set of the Definition just introduced if:*

$$\exists c, p, q \in (0,1) \quad : \quad \mathbb{E}[|X_t - X_s|^p] \leq c|t-s|^{\frac{1}{q}} \quad \forall s, t \in [0,1]$$

Then:

1. *$\forall \alpha \in [0, \frac{q}{p}) \exists K$ r.v. such that:*
   (a) *$\mathbb{E}[K^p] < \infty$*
   (b) *$|X_t - X_s| \leq k|t-s|^\alpha$ for $s, t \in D$*
2. *if X is continuous then #1 holds $\forall s, t \in [0,1]$*

Theorem (Brownian continuity of Wiener process). *We prove the local Höldercontinuity of the paths of a Wiener process, closely linked to the Gaussian process (see Wiener Gaussian characterization Thm.). This result is parallel to the existance of a p.r.m. (Existance of p.r.m. Thm.).*
(Def. 10.61) If $W = (W_t)_{t\in\mathbb{R}_+}$ is a Wiener process with:

- $\mathbb{E}[W_{t_i}] = 0 \forall i$
- $CoV[W_{t_i}, W_{t_j}] = t_i \wedge t_j \quad \forall i, j$

Then there exists an almost sure version of W that is locally Hölder continuous:

$$t \to \widetilde{W}_t \quad \textit{Hölder continuous} \quad W \overset{a.s.}{=} \widetilde{W}$$

Definition (Right continuous augmentation of filtration)**.** *For this object and its properties, refer also to Appendix D, especially from Definition D.4 onwards.*
For a filtration $\left(\mathcal{F}_t^0\right)_{t\in\mathbb{R}_+}$ we let it be generated by the process itself. Namely $\left(\mathcal{F}_t^0\right)_{t\in\mathbb{R}_+} = \sigma(\{X_s : s \le t\})$. In this context, we define the right continuous augmentation as:

$$(\mathcal{F}_t)_{t\in\mathbb{R}_+} \quad : \quad \forall t \; \mathcal{F}_t = \bigcap_{s\ge 0} \mathcal{F}^0_{t+s} = \lim_{s\downarrow 0} \mathcal{F}^0_{t+s}$$

Since $\left(\mathcal{F}_t^0\right)_{t\in\mathbb{R}_+}$ is increasing by Definition of filtration. This new filtration can be interpreted as a peek into the future. From now on, when referring to $\mathcal{F}$ it will be the ***augmented filtration****.*

Definition (Hitting time of barrier a, T_a)**.** *For a random time $T : \Omega \to \overline{\mathbb{R}}_+$ we define:*

$$T_a \coloneqq \inf\{t \ge 0 : X_t > a\}$$

Namely, the entrance time in the interval (a, ∞). Notice that we intuitively assume T_a to be almost surely finite for a Wiener process.

Theorem (Hitting times are stopping in augmentation)**.** *Let $\mathcal{F}$ be augmented and right continuous as in the above Definition, a process X on E be right continuous and adapted to $\mathcal{F}$. Then:*

$$\forall B \in \mathcal{B}(E) \quad T_B = \inf\{t \in \mathbb{R}_+ : X_t \in B\} \quad \textit{stopping time wrt } \mathcal{F}$$

According to the usual stopping time knowledge (Def. 11.9).

Definition (Shift operator $\theta_\cdot$)**.** *For a collection of continuous maps $\mathscr{C} = \{t \to w(t) \mid \textit{continuous}\}$ we define an operator:*

$$\theta_s : \mathscr{C} \to \mathscr{C}, \; s \in \mathbb{R} \quad (\theta_s \circ w)(t) \coloneqq w(s+t)$$

Theorem (Markov property of Lévy processes)**.** *This is Theorem VII.3.5 in [Çin11].*
For $X = (X_t)_{t\in\mathbb{R}_+}$ a Lévy process (Def. 17.1), for any time t, the process $X \circ \theta_t$ is independent of $\mathcal{F}_t$ and has the same law of X. Equivalently:

$$\mathbb{E}_t\left[V \circ \theta_t\right] = \mathbb{E}\left[V\right] \quad t \in \mathbb{R}_+, \quad \forall V \in \mathcal{G}_\infty \textit{ bounded}$$

where the boundedness of V is used to ensure existance of the expectation, but the result is extended to positive or integrable random variables in $\mathcal{G}_\infty$ the underlying end of time filtration generated by the process.
In terms of Wiener processes, which are Lévy by the Lévy Wiener characterization Theorem for $a = 0, b = 1$ we have that:

$$\forall s \ge 0 \quad (W \circ \theta_s) = \left(\widetilde{W}_t\right)_{t\in\mathbb{R}_+} = (W_{s+t} - W_s)_{t\in\mathbb{R}_+} \perp\!\!\!\perp \mathcal{F}_s \quad \textit{Wiener law}$$

Proposition (Conciliating time shifts to Wiener and Brownian)**.** *While a Wiener process (on a stochastic base [Çin11]) is such that:*

$$W_t \circ \theta_s = W_{s+t} - W_s$$

A Brownian motion as in Definition 18.2 is such that:

$$X_t \circ \theta_s = X_{t+s}$$

Corollary (Blumenthal's 0-1 law)**.** *It holds that:*

$$\mathcal{F} \quad \textit{augmented, right continuous} \implies \forall A \in \mathcal{F}_0 \; \mathbb{P}[A] \in \{0, 1\}$$

Namely, a suitable augmented filtration makes every event in the infinitesimal peek in the future at the start either certain or null.

Proposition (Hitting zero almost surely). *This result is like that of the total unpredictability of jumps Proposition for discrete times.*
For $T_0 = \inf\{t > 0 : W_t > 0\}$ *the hitting time of the barrier* 0 *as in Definition 19.2 we have:*

$$T_0 = 0 \quad a.s.$$

Corollary (Highly oscillatory behavior of W_t at zero). *There are* ∞ *many crossings for any time interval starting from zero. Namely:*

$$for\ a.e.\ \omega\ \exists u_1 > t_1 > s_1, u_2 > t_2 > s_2, \ldots \to 0 \quad s.t. \quad W_{u_n}(\omega) > 0, W_{t_n}(\omega) = 0, W_{s_n}(\omega) < 0$$

Corollary (Highly oscillatory behavior of W_t at infinity).

$$for\ a.e.\ \omega \exists u_1 > t_1 > s_1, u_2 > t_2 > s_2, \ldots \to \infty \quad s.t. \quad \lim_{n\to\infty} W_{s_n} = -\infty, \quad \lim_{n\to\infty} W_{u_n} = +\infty$$

Theorem (Strong Markov property of Lévy and Wiener processes). *For* T *a stopping time wrt* $\mathcal{F}^0$ *(the* ***not augmented*** *filtration) and* X *a Lévy process we have that:*

1. $\forall V$ *bounded in* $\overline{\mathcal{G}}_\infty$*:*
$$\mathbb{E}_T[V \circ \theta_T \mathbb{1}_{\{T<\infty\}}] = \mathbb{E}[V \mathbb{1}_{\{T<\infty\}}]$$
independent of $\mathcal{F}_T$ *and Lévy*
2. *shift operator version:*
$$(X_t \circ \theta_T)_{t\geq 0} = (X_{t+T})_{t\geq 0} \perp\!\!\!\perp \mathcal{F}_T, \text{ Lévy}$$

The claims are also valid for Wiener processes by the Lévy-Wiener characterization.

Theorem (A property of Wiener functions & stopping times). *Assume:*

- T *is stopping wrt* $\mathcal{F}$
- $U : \Omega \to \mathbb{R}_+$ *is such that* $U \in \mathcal{F}_T$
- $W = (W_t)_{t\in\mathbb{R}_+}$ *is a Wiener process*
- f *is a bounded Borel function on* $\mathbb{R}$
- $g(u) := \mathbb{E}[f \circ W_u], \quad u \in \mathbb{R}_+$

Then:

$$\mathbb{E}_T\left[f(W_{T+u} - W_T)\mathbb{1}_{\{T<\infty\}}\right] = g(U)\mathbb{1}_{\{T<\infty\}}$$

Theorem (Hitting time (Lévy) distribution by reflection principle). *As per Observation 19.13 we have:*

$$(T_a)_{a\geq 0} \quad : \quad T_a \sim \mathcal{IN}(a)\ \forall a \in \mathbb{R}_+$$

Meaning that $(T_a)_{a\geq 0}$ *has the same distribution of a stable Lévy process for each time point. Notice that we* ***are not proving*** *that it is a Lévy process, something we will show in a subsequent Theorem.*

Lemma (Elementary facts of Gamma, Beta and Cauchy distributions). *For* $Z_1 \perp\!\!\!\perp Z_2, Z_1 \stackrel{d}{=} Z_2 \sim \mathcal{N}(0,1)$*:*

1. *(Chi-square)* $Z_1^2 \stackrel{d}{=} Z_2^2 \sim \chi_1^2 = \mathcal{G}amma(\frac{1}{2}, \frac{1}{2})$
2. *(Beta)* $A = \frac{Z_1^2}{Z_1^2+Z_2^2} \sim \mathcal{B}eta(\frac{1}{2}, \frac{1}{2})$ *with density:*
$$f(u) = \frac{\Gamma(\frac{1}{2}+\frac{1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2})} u^{-\frac{1}{2}}(1-u)^{-\frac{1}{2}} = \frac{1}{\pi}\frac{1}{\sqrt{u(1-u)}}$$
by $\Gamma(1) = 1, \Gamma(\frac{1}{2}) = \sqrt{\pi}$
3. *(Beta cdf)* A *has cumulative distribution*
$$F_A(u) = \mathbb{P}[A \leq u] = \frac{2}{\pi}\arcsin\sqrt{u}$$
4. *(Cauchy-Beta link) for* $C = \frac{Z_1}{Z_2} \sim \mathcal{C}auchy$*:*
$$\implies A \stackrel{d}{=} \frac{Z_1^2}{Z_2^2 + Z_1^2} \stackrel{d}{=} \frac{1}{1+\frac{Z_2^2}{Z_1^2}} \stackrel{d}{=} \frac{1}{1+C^2}$$
where C *has density* $f(x) = \frac{1}{\pi(1+x^2)}$ *for* $x \in \mathbb{R}$ *(full support)*

Proposition (Arcsine law of G_t and D_t)**.** *Let A be as in the previous Lemma. Define, as in Example 19.8:*

$$G_t := \sup\{s \in [0,t] : W_t = 0\} \quad D_t := \{u \in (t,\infty) : W_u = 0\}$$

Then:

$$\forall t \in \mathbb{R}_+ \quad G_t \overset{d}{=} tA \quad D_t \overset{d}{=} \frac{t}{A}$$

Corollary (More results from the Proposition)**.** *We could also get:*

1. $R_t \overset{d}{=} tC^2$ *(by direct application of the Lemma above) for* $C \sim \mathcal{C}auchy$
2. G_t *has density* $f(x) = \frac{2}{\pi} \arcsin \sqrt{\frac{s}{t}}$
3. $Q_t \overset{d}{=} G_t \overset{d}{=} tA$

Definition (Running maximum of process)**.** *Consider a process $(W_t)_{t\in\mathbb{R}_+}$, we define for later use:*

$$M_t(\omega) := \max_{0\le s\le t} W_s(\omega) \quad t \in \mathbb{R}_+,\ \omega \in \Omega$$

Proposition (Running maximum vs hitting time)**.** *Recognize that:*

1. M_t *is continuous and increasing in* $\mathbb{R}$, *with* $M_0(\omega) = 0$ *and* $\lim_{t\to\infty} M_t(\omega) = \infty$
2. $(M_t)_{t\ge0}$ *and* $(T_a)_{a\ge0}$ *the hitting time process (Def. 19.2) are functional inverses:*

$$\{M_t > a\} = \{T_a < t\}$$

Theorem (Hitting time process is stable Lévy)**.** *The process $T = (T_a)_{a\ge0}$ is a strictly increasing pure jump Lévy process (Defs. 17.1, 17.7) with index $\frac{1}{2}$ and Lévy density:*

$$\lambda(dz) = \frac{1}{\sqrt{2\pi z^3}} dz \quad z \in \mathbb{R}_+$$

Proposition (Properties of T_{a-})**.** *Recall Observation 19.23 with $T_{a-} = \lim_{u\uparrow0} T_{a-u} = \inf\{t > 0 : W_t = a\}$. Then:*

1. T_{a-} *is a stopping time for* $\mathcal{F}^0$ *the* ***not augmented*** *filtration, while* T_a *is a stopping time wrt* $\mathcal{F}$ *the* ***augmented filtration***
2. *sojourne time is zero almost surely*

$$T_{a-} = T_a \quad a.s.$$

Definition (Subdivision & mesh)**.** *We denote for an interval $[a,b]$ a subdivision as a finite collection of intervals whose union is the interval itself (ignoring the start).*

$$\mathcal{A} := \{(s,t]\}, \qquad \bigcup_{(s,t]\in\mathcal{A}} (s,t] = (a,b]$$

Given a subdivision, we also identify the mesh as:

$$\|\mathcal{A}\| := \sup\{t - s : (s,t] \in \mathcal{A}\}$$

Definition (True p-variation, total variation, quadratic variation)**.** *For a function $f : \mathbb{R}_+ \to \mathbb{R}$, right continuous, an interval $[a,b] \in \mathbb{R}_+$, and a positive coefficient $p > 0$ the true p-variation is the quantity:*

$$\sup_{\mathcal{A}} \sum_{(s,t]\in\mathcal{A}} |f(t) - f(s)|^p$$

Where for $p = 1$ we call it total variation and for $p = 2$ ***true*** *quadratic variation. It turns out that this formulation is not very well suited for random processes, as it is infinite in both $p = 1, p = 2$. Below we prove the results for a reasonable surrogate.*

Definition (A dyadic subdivision)**.** *In the fashion of the dyadics Definition, and their properties in Lemma A.17 we could construct a subdivision of equally spaced intervals using $t_k = \frac{kt}{2^n}$ for $k = 0, \dots, 2^n$.*

Theorem (Wiener probabilistically finite $\mathcal{L}^2$ quadratic variation)**.** *For a Wiener process* $(W_t)_{t\in\mathbb{R}_+}$ *and a sequence of subdivisions* $(\mathcal{A}_n)_{n\in\mathbb{N}}$ *with* $\|\mathcal{A}_n\| \to 0$ *we have:*

1. $V_n = \sum_{(s,t]\in\mathcal{A}_n} |W_t - W_s|^2 \overset{\mathcal{L}^2}{\to} b-a$
2. $V_n \overset{\mathcal{P}}{\to} b-a$

We call V_n *the quadratic variation, not to be confused with the* ***true*** *quadratic variation. It is rather a probabilistic version of the latter. In some terms, the* sup *can be replaced with the* lim sup *of a sequence of meshes with size decreasing to zero (i.e.* $\|\mathcal{A}_n\| \to 0$*).*

Proposition (Almost sure dyadic subdivision for quadratic variation)**.** $\forall n \in \mathbb{N}$ *let* $\mathcal{A}_n$ *be a dyadic subdivision of the form presented in Definition 20.3. Then, with the hypothesis of the previous Theorem it also holds that:*

$$V_n = \sum_{(s,t]\in\mathcal{A}_n} |W_t - W_s|^2 \overset{a.s.}{\to} b-a$$

Proposition (Infinite total variation of Wiener process)**.** *For a Wiener process* $(W_t)_{t\in\mathbb{R}_+}$ *we have:*

$$TV = \sup_{\mathcal{A}} \sum_{(s,t]\in\mathcal{A}} |W_t - W_s| = \infty \quad \textit{almost surely}$$

over any interval $[a,b]$

21.2 Examples Collection

Throughout the second part of the course (Sections 11-19), many examples were reported on different occasions, the purpose of this Section is collecting them under the same discussion. It is just a copy paste with a horizontal line every time the Example changes.

21.2.1 Counting process

Counting process and Stopping Times

Some stopping & not stopping times we provide three examples:

- Let $\mathcal{F} = \sigma(\{N_t\})$. If we denote as T_k the k^{th} occurrence time in $[0,t]$ we can safely say that it is a stopping time of $\mathcal{F}$ since:

  $$\forall k \geq 1, k \in \mathbb{N}, \forall t \in \mathbb{R}_+ \quad \{T_k \leq t\} = \{N_t \geq k\} \in \mathcal{F}_t$$

  since N is adapted to $\mathcal{F}$ by construction
- The first time that an interval a passes without an arrival, namely:

  $$T = \inf\{t \geq a : N_t = N_{t-a}\} \quad a > 0$$

  Needs the formalism of stopped filtration (Def. 11.19) and we will show it is a stopping time in the next points.
- instead a random time such as the time of last arrival before $b > 0$:

  $$L = \inf\{t : N_t = N_b\} \quad b > t$$

  is not a stopping time since we need the information from the interval $[t,b]$ to establish what occurs at time t.

Counting Process (△ **setting**) Consider the counting process on $\mathbb{T} = \mathbb{R}_+$ from Definition 11.13. If we consider $H \cap \{S < T\}$ we could tell if H and $S < T$ happened in $\mathcal{F}_{S \wedge T}$. For any t it holds that:

$$\exists k: \ T_k(\omega) \leq t < T_{k+1}(\omega)$$

Recall also that all of these T_k are stopping times of $\mathcal{F} = \sigma(\{N_t\}_{t \geq 0})$.
We set $T_0 = 0$ for convenience, and consider the random time:

$$\tau = \inf\{t \geq a : N_t = N_{t-a}\} \quad a > 0$$

denoted in blue for convenience. Before this Example, the symbol T was used, but here we wish to distinguish many objects that are similar in notation. After this Example, we will not use the symbol τ.
We want to show that τ is a stopping time in the sense of Definition 11.9.
(□ **solution**) Notice that $\forall \omega$ we have for some k:

$$\tau(\omega) = T_k(\omega) + a \quad k \in \mathbb{N}^* \iff \{\tau \leq t\} = \bigcup_{k \geq 1} \{\{\tau = T_k + a\} \cap \{\tau \leq t\}\}$$

Where the union over k statement comes from the fact that we have $\exists k$ as a condition. Recall the objects of Theorem 11.23, visualizing them as $T = T_{k+1}, S = T_k + a : S \leq T$. Be careful as it may lead to confusion, left is old and red, right is this Example. It holds:

$$\tau \ stopping \iff \{\tau = T_k + a\} \cap \{\tau \leq t\} \in \mathcal{F}_t \quad k \in \mathbb{N}^*$$

Using Theorem 11.23#1, we need to check that $V = \{\tau = T_k + a\} \in \mathcal{F}_{T_k + a} \subset \mathcal{F}_T$ to conclude.
To clear out why, recognize that it is equivalent to $\{\tau = T_k + a\}\mathbb{1}_{\{T_k + a \leq t\}} = \{\tau \leq t\} \in \mathcal{F}_t \forall t \in \overline{\mathbb{T}}$ by the very Theorem invoked.
For this purpose, observe that for any k:

$$\{\tau = T_k + a\} = \underbrace{\{T_1 - T_0 \leq a, \ldots, T_k - T_{k-1} \leq a\}}_{\in \mathcal{F}_S, S = T_k + a} \cap \underbrace{\{T_k + a < \overbrace{T_{k+1}}^{T = T_{k+1}}\}}_{S < T}$$

By Theorem 11.26#4 we will have that:

$$H \in \mathcal{F}_S \implies H \cap \mathbb{1}_{\{S < T\}} \in \mathcal{F}_{S \wedge T}$$

Where $S = T_k + a, T = T_{k+1}, H = \{T_1 - T_0 \leq a, \ldots, T_k - T_{k-1} \leq a\}$. Eventually:

$$\begin{aligned} \{\tau = T_k + a\} \in \mathcal{F}_{T_k + a \wedge T_{k+1}} = \mathcal{F}_{T_k + a} = \mathcal{F}_\tau \iff & \{\tau = T_k + a\} \cap \{\tau \leq t\} \in \mathcal{F}_t && \forall k, \forall t \\ \iff & \{\tau \leq t\} \in \mathcal{F}_t && \forall t \end{aligned}$$

Counting process: age perspective We propose a different view on the counting process (Def. 11.13).

(△ **setting**) let $0 < T_1 < T_2 < \ldots$ be such that $\lim_{n \to \infty} T_n = +\infty$ and:

$$N_t = \sum_{n=1}^{\infty} \mathbb{1}_{[0,t]}(T_n) \quad \mathcal{F} = \sigma((N_t)_{t \in \mathbb{T}})$$

See N_t as the number of replacements of some object. Then, the duration of the k^{th} object can be formalized as:

$$A_t(\omega) := t - T_k(\omega) \quad if \ T_k(\omega) \leq t \leq T_{k+1}(\omega)$$

Where the map $t \to A_t$ is:

- strictly increasing in each interval
- right continuous at each jump

See Figure 21.1 for an intuition. We can further define for $a > 0$:

$$T := \inf\{t \geq 0 : A_t \geq a\}$$

As the first time the age of a replacement is at least a.

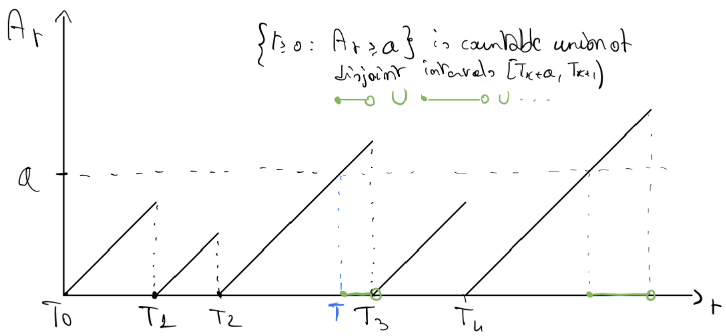

Figure 21.1: A visualization of $(A_t)_{t \in \mathbb{T}}$

($\square$ A is adapted) if $t < a \implies A_t = \emptyset \in \mathcal{F}_t \forall t$ and the statement is trivial.
Else consider:

$$\{A_t \geq a\} \in \mathcal{F}_t \iff A_t \in \mathcal{F}_t$$

$$\iff \{t - T_k \geq a\} \qquad = \{T_k < t < T_{k+1}\} \cap \{A_t \geq a\}$$

$$= \underbrace{\{t < T_{k+1}\}}_{\mathcal{F}_t} \cap \underbrace{\{t - T_k \leq a\}}_{\mathcal{F}_{t-a} \subset \mathcal{F}_t}$$

So that by closedness under countable intersections (Lem. 1.7):

$$\{A_t \geq a\} = \bigcup_k (\{T_k < t < T_{k+1}\} \cap \{A_t \geq a\}) \in \mathcal{F}_t \quad \forall t$$

($\bigcirc$ equivalence to counting) we aim to show that:

$$\inf\{t \geq 0 : N_t = N_{t-a}\} = \inf\{t \geq 0 : A_t \geq a\}$$

The time above a is the union of disjoint $[\cdot, \cdot)$ intervals such that $T_{k+1} - T_k \geq a$ by construction, implying that:

$$\{t \geq 0 : A_t \geq a\} = \bigcup_{k : T_{k+1} - T_k \geq a} [T_k + a, T_{k+1})$$

which infimized:

$$\implies \inf\{t \geq 0 : A_t \geq a\} = \min_k \{T_k + a : T_{k+1} - T_k \geq a\}$$

$$\iff \{T = T_k + a\} = \{T_1 - T_0 < a, \dots, T_k - T_{k-1} < a\} \cap \{T_{k+1} > T_k + a\}$$

($\diamondsuit$ T is a stopping time) we eventually show that T is again a stopping time. Differently from the paragraph where we already showed it, with respect to $\mathcal{G} = \sigma((A_t)_{t \in \mathbb{T}})$.

$$\{T \leq t\} = \bigcup_{s < t} \{A_s \geq a\}$$

$$= \bigcup_{s \in \mathbb{Q}, s < t} \{A_s \geq a\} \qquad \text{By the continuity in } \triangle \text{ unless } A_t = 0$$

$$= \bigcup_{s \in \mathbb{Q}, s < t} \{\underbrace{A_s}_{\in \mathcal{F}_s} \geq a\} \qquad \text{where } \mathcal{G}_s \subset \mathcal{G}_t \forall s < t$$

$$\in \mathcal{G}_t \qquad \text{by countable unions (Thm. 1.5)}$$

The discussions of $\triangle, \diamondsuit \implies \{T \leq t\} \in \mathcal{G}_t$ and T is a stopping time in the sense of Definition 11.9.

Counting process random measure For ordered distinct arrival times $0 < T_1 < \dots$ the counting process $N_t = \sum_{n=1}^{\infty} \mathbb{1}_{[0,t]}(T_n)$ can be seen as the measure arising from a random measure:

$$N_t = M([0,t]) \quad E = \mathbb{R}_+,\ A = [0,t]$$

Interarrivals of stopping times distribution Consider the random time:

$$S = \inf\{t \geq a : N_t = N_{t-a}\}$$

Where $T_k + a \quad k > 0$ is equivalent to having the first k interarrivals of size at most a and the $(k+1)^{th}$ exceeding a. It holds that $S < \infty$ almost surely, since the union over k of the events has probability one.
Let T be the next jump, and note that S falls in $a + (T - S)$. We ask the following question:

$$\textit{Is it true that } a \to \infty \implies T - S \to 0?$$

This is **False**. Indeed, notice that:

$$\{T - S > t\} = \{N_{S+t} - N_S = 0\} \perp\!\!\!\perp \mathcal{F}_S \implies T - S \sim \mathcal{E}xp(ct)$$

Where we exploited the loss of memory property, namely the second set being strong Markovian. The probability is:

$$\begin{aligned}
\mathcal{P}(T - S \geq t) &= \mathbb{E}\left[\mathbb{1}_{\{T-S>t\}}\right] & \\
&= \mathbb{E}\left[\mathbb{E}_S\left[\mathbb{1}_{\{T-S>t\}}\right]\right] & \text{unconditioning} \\
&= \mathbb{E}\left[\mathbb{E}_S\left[\mathbb{1}\{N_{S+t} - N_S = 0\}\right]\right] & \text{set equivalence above} \\
&= \mathbb{E}_S\left[f(N_{S+t} - N_S)\right] & f(x) = 1 \cdot \mathbb{1}_{\{x=0\}} \\
&= \mathbb{E}\left[\mathbb{1}_{\{x=0\}}\right] & \text{Strong Markov Prop. 16.6} \\
&= \mathcal{P}(X = 0) & \text{distr as } \mathcal{P}o(ct) \\
&= e^{-ct} & \perp\!\!\!\perp \mathcal{F}_S, a
\end{aligned}$$

And the distribution is completely independent of a.

21.2.2 Random Walk

Sum of independent random variables martingale Let $(X_n)_{n\in\mathbb{N}}$ be an independency where $\mathbb{E}[X_n] = 0 \forall n$. The sum r.v. is such that $S_0 = 0, S_n = S_{n-1} + X_n \forall n \geq 1$, and the underlying filtration is generated by the process itself $\mathcal{F} = \sigma\left((X_n)_{n\in\mathbb{N}}\right)$. Then:

- $(S_n)_{n\in\mathbb{N}} = S$ is adapted to $\mathcal{F}$ trivially
- $\mathbb{E}[S_n] = \mathbb{E}\left[\sum_{k=1}^{n} X_k\right] = 0 \forall n \iff \mathbb{E}[|S_n|] < \infty$ so that $S_n \in \mathcal{L}_1(\Omega, \mathcal{H}, \mathbb{P})$
- using Proposition 11.40 we only check the martingale for one step forward:

$$\begin{aligned}
\mathbb{E}_n[S_{n+1} - S_n] &= \mathbb{E}_n[X_{n+1}] & \text{recursion} \\
&= \mathbb{E}[X_{n+1}] & \text{independence hyp.} \\
&= 0 & \text{hypothesis}
\end{aligned}$$

So $(S_n)_{n\in\mathbb{N}}$ is a martingale.

Random times alone do not satisfy martingale equality in a symmetric random walk Let:

$$S_0 = 1,\ S_n = S_{n-1} + \xi_n \quad \xi_n \sim \mathcal{B}ern_{\pm 1}\left(\frac{1}{2}\right),\ \xi_n \in \{-1,+1\} \quad \forall n$$

Assign $X_n = S_{T\wedge n}$. To see that $(S_n)_{n\in\mathbb{N}}$ is a martingale, refer to the discussion above. To see that X_n is a nonnegative martingale, use Theorem 12.9. In particular $\mathbb{E}[S_n] = \mathbb{E}\left[\underbrace{\mathbb{E}_0[S_n]}_{=\mathbb{E}_0[S_0]}\right] = \mathbb{E}[1] = 1\forall n$. Consider the random time:

$$T = \inf\{k : S_k = 0\}$$

T is a stopping time wrt $\mathcal{F} = \sigma\left((S_n)_{n\in\mathbb{N}}\right)$ (for this recover the paragraph of the previous Section in which we prove T is a stopping time), or observe that:

$$\{T \le t\} \subset \bigcup_{k\le t, k\in\mathbb{N}} \{S_k = 0\} \in \mathcal{F}_t$$

Namely, the sum being equal to zero is included in the event that at least one of the times before the sum has reached zero which is in the increasing filtration. Clearly, $S_T = 0$ with probability 1 since at time T the martingale will be certainly null but:

$$0 = \mathbb{E}[S_T] \ne \mathbb{E}[S_0] = 1$$

Symmetric random walk We show that for a symmetric RW the MCT can be used, but it is not $\mathcal{L}_1$ convergent, namely $\overset{a.s.}{\to} \not\Longrightarrow \mathcal{L}_1$ at the same limit.
Recover the previous setting, where $T = \inf\{m \in \mathbb{N} : S_n = 0\}$. We have that $\mathbb{E}[X_i] = 0\forall i$ and $S = (S_n)_{n\in\mathbb{N}}$ is a martingale with $\mathbb{E}[S_n] = 1\forall n$. The stopped martingale $(S_{n\wedge T})_{n\in\mathbb{N}}$ is such that $S_{n\wedge T} \ge 0$ and by the MCT (Thm. 12.27, Cor. 12.30) there exists an almost sure limiting process S_∞. Now obviously $S_{n\wedge T} \overset{a.s.}{\to} S_\infty = 0$ since convergence to $k > 0$ is impossible as it would mean that $S_n = k > 0 \implies S_{n+1} \in \{k-1, k+1\}$, i.e. no convergence. However, there is no $\mathcal{L}_1$ convergence. Indeed by Proposition 12.13:

$$\mathbb{E}[S_{n\wedge T}] = \mathbb{E}[S_{0\wedge T}] = \mathbb{E}[S_0] = 1$$

but:

$$\mathbb{E}[|S_{n\wedge T} - S_\infty|] = \mathbb{E}[|S_{n\wedge T} - 0|] = \mathbb{E}[|S_{n\wedge T}|] = \mathbb{E}[S_{n\wedge T}] = 1 \ne 0 \quad \forall n \in \mathbb{N}$$

21.2.3 Bayesian Mean Estimation

A uniformly integrable martingale Let $Z_i \overset{iid}{\sim} \mathcal{N}(0,1)$, and $\theta \in \mathcal{L}_1(\Omega, \mathcal{H}, \mathbb{P})$ such that $\theta \perp\!\!\!\perp Z_i \forall i$.
Define $Y_i = Z_i + \theta \overset{iid}{\sim} \mathcal{N}(\theta, 1)$ and aim to infer θ from a set of observations $\mathbf{Y} = \{Y_i\}_{i=1}^n$.
Why is θ random? We use a bayesian approach and assign a prior $\theta(A) = \mathbb{P}[\theta \in A]$ such that $Y_i|\theta \overset{iid}{\sim} \mathcal{N}(\theta, 1)$.
Further assume the joint distribution $(\mathbf{Y}, \theta)$ is absolutely continuous (Def. 2.6) wrt *Leb* and that $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$.

Using Bayes theorem we can estimate:

$$\pi_n(A) = \mathbb{P}[\theta \in A | Y_1 = y_1, \ldots, Y_n = y_n]$$

By Ionescu-Tulcea Thm. 10.57 construct also a space:

$$(\mathbb{R}^\infty \times \Theta, \mathcal{B}(\mathbb{R}^\infty) \otimes \mathcal{B}(\Theta), \mathbb{P})$$

So that for a filtration $\mathcal{F} = \sigma((Y_n)_{n\in\mathbb{N}})$ by Proposition 11.50 since $\theta \in \mathcal{L}_1$:

$$\widehat{\theta}_n = \mathbb{E}[\theta|\mathbf{Y}] = \mathbb{E}_n[\theta] \quad \textit{such that} \quad \left(\widehat{\theta}_n\right)_{n\in\mathbb{N}} \ \textit{uniformly integrable}$$

Thanks to the $\theta \sim \mathcal{N}$ assumption we can explicitly compute the posterior distribution as:

$$\begin{aligned}
\pi_n(\theta) &\propto \pi(\theta) \prod p(y_i|\theta) \\
&\propto exp\left\{ -\frac{1}{2\sigma_n^2}(\theta - \mu_n)^2 \right\} \\
\implies \theta|\mathbf{Y} &\sim \mathcal{N}(\mu_n, \sigma_n^2) \\
\mu_n &= \sigma_n^2 \left(\frac{\mu_0}{\sigma_0^2} + n\overline{y} \right) \\
\sigma_n^2 &= \left(\frac{1}{\sigma_0^2} + n \right)^{-1}
\end{aligned}$$

Uniform integrability of $\widehat{\theta}$ process Using Lemma 11.51. We have:

$$\widehat{\theta}_n = \sigma_n^2 \left(\frac{\mu_0}{\sigma_0^2} + n\overline{Y} \right) \quad \overline{Y} \sim \mathcal{N}\left(\theta, \frac{1}{n} \right)$$

and for $f(x) = x^2$ convex positive increasing and coercive:

$$\begin{aligned}
\mathbb{E}\left[f(|\widehat{\theta}_n|) \right] = \mathbb{E}\left[\widehat{\theta}_n^2 \right] &= V\left[\widehat{\theta}_n \right] - \left(\mathbb{E}\left[\widehat{\theta}_n \right] \right)^2 \\
&= V\left[\widehat{\theta}_n \right] - \mu_0^2 \\
&< \infty \iff V\left[\widehat{\theta}_n \right] < \infty
\end{aligned}$$

where we aim to find an upper bound for the variance. First notice that by the variance decomposition:

$$\overline{Y}^{(n)}|\theta \sim \mathcal{N}\left(\theta, \frac{1}{n} \right) \quad V\left[\overline{Y}^{(n)} \right] = \mathbb{E}\left[V\left[\overline{Y}^{(n)} \right] \right] + V\left[\mathbb{E}\left[\overline{Y}^{(n)} \right] \right] = \mathbb{E}\left[\frac{1}{n} \right] + V[\theta] = \frac{1}{n} + \sigma_0^2$$

Such variance is by the first term in the addition being constant:

$$\begin{aligned}
V[\widehat{\theta}_n] &= n^2 \sigma_n^4 V\left[\overline{Y}^{(n)} \right] & V\left[\overline{Y}^{(n)} \right] = \frac{1}{n} + \sigma_0^2 \\
&= n^2 \sigma_n^4 \left(\frac{1}{n} + \sigma_0^2 \right) \\
&= n^2 \sigma_n^4 \left(\frac{1 + \sigma_0^2 n}{n} \right) \\
&= n \sigma_n^4 \left(1 + \sigma_0^2 n \right) & \sigma_n^2 = \left(\frac{1}{\sigma_0^2} + n \right)^{-1} = \frac{\sigma_0^2}{n\sigma_0^2 + 1} \\
&= \frac{n(1 + \sigma_0^2 n)\sigma_0^2}{(1 + n\sigma_0^2)^2} = \frac{n\sigma_0^2}{1 + n\sigma_0^2} \leq \sigma_0^2 = V[\theta_0]
\end{aligned}$$

So that the variance is finite $\forall n$ and $\mathbb{E}[f(\widehat{\theta}_n)] < \infty$ for f convex and positive. Then, $\left(\widehat{\theta}_n \right)_{n \in \mathbb{N}}$ is a uniformly integrable martingale.

Bayesian mean estimation, Corollary 12.38 Recall that Z_i are iid standard normals and $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$ is independent from Z_i, integrable and finite.
For $Y_i = \theta + Z_i \implies Y_i|\theta \overset{iid}{\sim} \mathcal{N}(\theta, 1)$ we have that for observables $\mathbf{Y} = \{Y_i\}_{i=1}^n$:

$$\pi_n(A) = \mathbb{P}[\theta \in A|\mathbf{Y} = \mathbf{y}] \quad \mathcal{F} = \sigma(\{\mathbf{Y}\})$$

Then $\widehat{\theta}_n = \mathbb{E}_n[\theta] = \mathbb{E}[\theta|\mathcal{F}_n]$ is a uniformly integrable martingale by Proposition 11.50.
Further, by Corollary 12.38#1 we conclude:

$$\widehat{\theta}_n \underset{\mathcal{L}_1}{\overset{a.s.}{\to}} \mathbb{E}_\infty[\theta] = \mathbb{E}[\theta|\mathcal{F}_\infty]$$

Moreover, if the condition $\theta \in \mathcal{F}_\infty$ holds, we apply Corollary 12.38#2 and further state that:

$$\theta \in \mathcal{F}_\infty \implies \widehat{\theta}_n \underset{\mathcal{L}_1}{\overset{a.s.}{\to}} \theta$$

We prove a sufficient condition for this to be true in Proposition 12.40.

21.2.4 Branching Process

Branching Process, a uniformly integrable martingale The following is a discrete time biological model for population evolution. We interpret Z_n as the size of a population, which starts at $Z_0 = 1$, has no overlapping generations and lifetimes of unit one. At $n+1$, the population is an offspring of the n^{th} generation only. We denote:

$$Z_0 = 1, \quad Z_1 = \xi_1^{(1)}, \quad Z_2 = \sum_{i=1}^{Z_1} \xi_i^{(2)}$$

And assume:

$$\left\{\xi_i^{(n)}, i \geq 1, n \geq 1\right\} \; iid \; \mathbb{E}[\xi_i^{(n)}] = \mu \geq 0, \quad p_k := \mathcal{P}\left[\xi_i^{(n)} = k\right], \; k \geq 0$$

where p_k is referred to as the offspring distribution. The underlying filtration is generated by the sizes of past families as $\mathcal{F}_n = \sigma\left(\left\{\xi_i^{(m)}, i \geq 1, m \leq n\right\}\right)$.
(△ aim) we want to show that $\left(\frac{Z_n}{\mu^n}\right)_{n\in\mathbb{N}}$ is a martingale.
(□ solution) This is equivalent to showing that another process satisfies the martingale equality:

$$\left(\frac{Z_n}{\mu^n}\right)_{n\in\mathbb{N}} \iff \mathbb{E}_n[Z_{n+1}] = \mu Z_n$$

Which follows by simple computation. Adaptedness and integrability are trivial. Maybe it is useful to notice that $\mathbb{E}\left[|\xi_i^{(n)}|\right] = \mathbb{E}\left[\xi_i^{(n)}\right]$ by positivity. The above formula can be checked for one time step only by Proposition 11.40. Then:

$$\begin{aligned}
\mathbb{E}_n[Z_{n+1}] &= \mathbb{E}_n\left[\left(\xi_1^{(n+1)} + \cdots + \xi_{Z_n}^{(n+1)}\right)\mathbb{1}_{\{Z_n>0\}}\right] && \text{recursion hypothesis} \\
&= \mathbb{E}_n\left[\sum_{k=1}^{\infty}\left(\xi_1^{(n+1)} + \cdots + \xi_{Z_n}^{(n+1)}\right)\mathbb{1}_{\{Z_n=k\}}\right] \\
&= \sum_{k=1}^{\infty}\mathbb{E}_n\left[\left(\xi_1^{(n+1)} + \cdots + \xi_{Z_n}^{(n+1)}\right)\underbrace{\mathbb{1}_{\{Z_n=k\}}}_{\in\mathcal{F}_n}\right] && \text{linearity, Prop. 10.19\#2} \\
&= \sum_{k=1}^{\infty}\mathbb{1}_{\{Z_n=k\}}\mathbb{E}_n\left[\left(\xi_1^{(n+1)} + \cdots + \xi_k^{(n+1)}\right)\right] && \text{conditional determ. Prop. 10.23\#1} \\
&= \sum_{k=1}^{\infty}\mathbb{1}_{\{Z_n=k\}}\left(\mathbb{E}_n\left[\xi_1^{(n+1)}\right] + \cdots + \mathbb{E}_n\left[\xi_k^{(n+1)}\right]\right) && \text{linearity} \\
&= \sum_{k=1}^{\infty}\mathbb{1}_{\{Z_n=k\}}k\mu && \mathbb{E}\left[\xi_i^{(n+1)}\right] = \mu \;\forall i \\
&= \mu\sum_{k=1}^{\infty}\mathbb{1}_{\{Z_n=k\}}k \\
&= \mu Z_n && \forall n
\end{aligned}$$

Clearly $(Z_n)_{n\in\mathbb{N}}$ is a martingale for $\mu = 1$, a submartingale for $\mu < 1$ and a supermartingale for $\mu > 1$. For free, we also get that $\left(\frac{Z_n}{\mu^n}\right)_{n\in\mathbb{N}}$ is a martingale since:

$$\mathbb{E}_n\left[\frac{Z_{n+1}}{\mu^{n+1}}\right] = \frac{1}{\mu^{n+1}}\mu Z_n = \frac{Z_n}{\mu^n} = \mathbb{E}_n\left[\frac{Z_n}{\mu^n}\right]$$

Martingale Convergence Theorem for Branching In the previous example, we showed that $\left(\frac{Z_n}{\mu^n}\right)_{n\in\mathbb{N}}$ is a martingale with $\mu = \mathbb{E}\left[\xi_1^{(1)}\right]$. We also have that:

$$\mathbb{E}_n\left[\frac{Z_{n+1}}{\mu^{n+1}}\right] = \frac{Z_n}{\mu^n} \implies \mathbb{E}_n[Z_{n+1}] = \mu Z_n = \begin{cases} < Z_n, \mu < 1 & \textit{supermartingale} \\ = Z_n, \mu = 1 & \textit{martingale} \\ > Z_n, \mu > 1 & \textit{submartingale} \end{cases}$$

Here, using the MCT (Thm. 12.27) we want to show this as for Corollary 12.30. For this purpose, let $\mu = 1, p_1 < 1$. Then $(Z_n)_{n\in\mathbb{N}}$ is a positive martingale an by the MCT Corollary:

$$\exists Z_\infty = \lim_{n\to\infty} Z_n \implies Z_n = Z_\infty \textit{ eventually} \text{ Def. 9.3}$$

($\triangle$ aim) In this context, we want to show that:

$$Z_\infty = 0 \iff \mathbb{P}[Z_n = k, \forall n \geq N] = 0 \quad \forall k \in \mathbb{N}, N \text{ sufficiently large}$$

($\square$ solution) compute the following:

$$\begin{aligned}
\mathbb{P}[Z_n = k, \forall n > N] &= \mathbb{P}[Z_n = k, Z_{n+1} = k, \ldots] && \text{stable limit} \\
&= \mathbb{P}\left[\sum_{i=1}^{k} \xi_i^{(m+N)} = k \quad m = 1, 2, \ldots, Z_n, Z_n = k\right] && \text{hypothesis} \\
&\leq \mathbb{P}\left[\sum_{i=1}^{k} \xi_i^{(m+N)} = k \quad m = 1, 2, \ldots\right] && \mathbb{P}[A \cap B] \leq \mathbb{P}[A] \\
&= \prod_{m=1}^{\infty} \mathbb{P}\left[\sum_{i=1}^{k} \xi_i^{(m+N)} = k\right] && \text{independence} \\
&\overset{m\to\infty}{\to} \lim_{m\to\infty}\left(\mathbb{P}\left[\sum_{i=1}^{k} \xi_1^{(1)} = k\right]\right)^m && \text{identically distr.}
\end{aligned}$$

By $\triangle$ we need it to be null, this is the same as:

$$\iff \mathbb{P}\left[\sum_{i=1}^{k} \xi_1^{(1)} = k\right] < 1$$

from which we get:

$$\begin{aligned}
\mathbb{P}\left[\sum_{i=1}^{k} \xi_i^{(1)} = k\right] &\leq \mathbb{P}\left[\sum_{i=1}^{k} \xi_1^{(1)} > 0\right] \\
&= 1 - \mathbb{P}\left[\sum_{i=1}^{k} \xi_1^{(1)} = 0\right] \\
&= 1 - p_0^k && \text{iid} \\
&< 1 && \mu = 1, p_1 < 1 \implies p_0 > 0
\end{aligned}$$

By the same arguments for $\mu < 1$ we could show that $(Z_n)_{n\in\mathbb{N}}$ is a positive supermartingale and that by the Corollary we have a limit which is almost sure. We call it Z_∞. Then, by similar arguments, one can show for $k > 0$ and $N > 0$ arbitrary that:

$$\mathbb{P}[Z_n = k, \forall n \geq N] = 0$$

since $\mu < 1 \implies p_0 > 0$. We then conclude $Z_\infty \overset{a.s.}{=} 0$

More about martingale convergence for Branching We know $\mu = 1, p_1 < 1$ are such that $(Z_n)_{n\in\mathbb{N}} \overset{a.s.}{\to} Z_\infty = 0$ but not such that $Z_n \overset{\mathcal{L}_1}{\to} Z_\infty$. Yet we also argued that $Z_n \neq \mathbb{E}_n[Z_\infty] = 0$ since $Z_n > 0 \forall n$ with positive probability. So, we cannot conclude that $\mathbb{E}_n[M_\infty] = M_n$.

Simplest model is extinction or explosion Before we had $\mu < 1$ or $\mu = 1$ and $p_1 < 1$ so that $Z_n \overset{a.s.}{\to} 0$.
(△ aim) The assumptions now become $p_0 \in (0,1) \implies \mu < 1$ and we want to show that:

$$Z_n \overset{a.s.}{\to} 0 \quad or \quad Z_n \overset{a.s.}{\to} \infty$$

Namely, if there is no extinction, then the population explodes. Another possible formulation is:

$$\mathbb{P}\left[\lim_{n\to\infty} Z_n \in \{0, \infty\}\right] = 1$$

Continuous time Yule Branching process Consider $Z_t := \#$ individuals at time t, with $Z_0 = 1$. Assume death is not possible and the chance of birth is dt, independently for each individual. Namely, one child in the interval $(t, t + dt]$, with no influence within the population.
(△ aim) We show that for each individual the number of descendants is an independent copy of the counting Yule process, upon time shifts to restart it.
(□ exponential interbirths premise) let Y_k be the k^{th} inter-birth time. It holds that:

$$Y_k \overset{ind}{\sim} \mathcal{E}xp(k)$$

Since the waiting time for the first birth is a unit rate with exponential variable, given the **linear** chance of birth. For general k, there are $k - 1$ individuals plus one ancestor, each birthing at a rate dt. The first birth is the minumum of k exponential unit random variables. We show that this is again exponential. Observe that $Y_2 = \min\{E_1, E_2\}, \ldots, Y_k = \min\{E_1, \ldots, E_k\}$ where for each $E_i \sim \mathcal{E}xp(1)$. We easily conclude:

$$\begin{aligned}
\mathcal{P}(Y_k > t) &= \mathcal{P}(E_1 > t, \ldots, E_k > t) \\
&= (\mathcal{P}(E_1 > t))^k && \text{iid} \\
&= e^{-kt} \\
&= \mathcal{P}(E > t) && E \sim \mathcal{E}xp(k)
\end{aligned}$$

(○ first result)wts

$$Z_t \sim \mathcal{G}eom(e^{-t}) \iff \mathcal{P}(Z_t = x) = e^{-t}(1 - e^{-t})^{x-1} \quad x = 1, 2, \ldots$$

Notice that the interarrivals denoted with Y_k allow to define the arrivals process $S = (S_n)_{n\in\mathbb{N}}$ as

$$S_n = \sum_{k=1}^{n} Y_k$$

Which is equivalent to:

$$\iff \{S_n \le t\} = \{Z_t - 1 \ge n\} \iff \{S_n \le t\} = \{Z_t > n\} \tag{21.3}$$

Meaning that the arrival times are stopping times for the underlying counting process Z in the usual sense (Def. 11.9).
Notice that in the p.r.m. case of the compound Poisson process (Def. 15.5) we had that S_n was a sum of unit exponentials, returning a $\mathcal{G}amma(n,1)$ distribution (Thm. 16.2). Here instead:

$$S_n = \sum_{k=1}^{n} \underbrace{Y_k}_{\sim \mathcal{E}xp(k)} = \sum_{k=1}^{n} \frac{1}{k} \underbrace{E_1}_{\sim \mathcal{E}xp(1)} \overset{d}{=} \max\{E_1, \dots, E_k\}$$

By a reverse time heuristic argument or a mgf argument. Eventually:

$$\begin{aligned}
\mathcal{P}(S_n < t) &= \mathcal{P}(E_1 \le t, \dots, E_n \le t) && E_k \overset{iid}{\sim} \mathcal{E}xp(1)\\
&= (\mathcal{P}(E_1 < t))^k && iid\\
&= (1-e^{-t})^k\\
&= \mathcal{P}(Z_t > n) && \text{Eqn. 21.3}\\
\implies \mathcal{P}(Z_t = n) &= e^{-t}(1-e^{-t})^k \implies Z_t \sim \mathcal{G}eom(e^{-t})
\end{aligned}$$

($\triangledown$ growth rate)wts

$$\frac{Z_t}{\mathbb{E}[Z_t]} \overset{a.s.}{\to} W \sim \mathcal{E}xp(1)$$

First of all, observe that the unnormalized rate would explode exponentially fast:

$$\mathbb{E}[Z_t] = \frac{1}{e^{-t}} = e^t \nearrow \infty$$

thus a normalized version. Let $W_t = e^{-t} Z_t = \dfrac{Z_t}{\mathbb{E}[Z_t]}$ and inspect the process $(W_t)_{t \in \mathbb{R}_+}$.
($\triangledown$♠ subpoint, W is a martingale) we show for $\mathcal{F} = \sigma(Z)$ that W is a martingale according to Definition 11.35. Adaptedness and integrability are easily verified. The martingale equality holds since:

$$\begin{aligned}
\mathbb{E}_s[W_t] &= \mathbb{E}[W_t | Z_s] && Z \text{ only determinator}\\
&= e^{-t}\mathbb{E}[Z_t | Z_s]\\
&= e^{-t}\mathbb{E}[Z_{t-s} | Z_0 = Z_s]\\
&= e^{-t} Z_s \mathbb{E}[Z_{t-s} | Z_0 = 1]\\
&= e^{-t} Z_s e^{t-s} && \text{previous results}\\
&= e^{-s} Z_s = W_s
\end{aligned}$$

($\triangledown$♣ subpoint, Gumbell limit identity) recall that for $S = (S_n)_{n\in\mathbb{N}} = \max\{E_1, \dots, E_n\}$ it holds that:

$$S_n - \log(n) \overset{d}{\to} \Psi \sim \mathcal{G}umbell \quad \mathbb{P}[\Psi = x] = e^{-e^{-x}}$$

21.2.5 Poisson Process

Poisson Process basics Let N be a counting process (Def. 11.13) such that $N_t = \sum_{k=0}^{\infty} \mathbb{1}_{[0,t]}(T_k)$ is adapted to $\mathcal{F}$ (Def. 11.7).

($\triangle$ aim) we want to show that:

$$N \sim \mathcal{P}ois(c) \quad \text{Def. 12.2} \implies M_t = exp\{-rN_t + ct - cte^{-r}\} \quad \mathcal{F}\text{-}martingale \quad \forall r \in \mathbb{R}_+ \quad \text{Def. 11.35}$$

The relation is actually $\iff$ but we only show one side.

(□ Laplace approach) recall that by Definition 6.11 for a Poisson random variable we have that:

$$X \sim Po(\lambda) \iff \widehat{\mathcal{P}}_X(r) = \mathbb{E}[e^{-rX}] = exp\{-\lambda(1 - e^{-r})\}$$

From this simple fact we could show that for any time point the martingale equality holds by noticing that from Definition 12.2 we have $N_t \sim Po(ct)$ and $N_t - N_s|\mathcal{F}_s \sim Po(c(t-s))$ by Proposition 12.3#3. This allows us to say that the Laplace transform of $N_t - N_s$ is:

$$\mathbb{E}_s\left[exp\left\{-r(N_t - N_s)\right\}\right] = exp\left\{-c(t-s)(1-e^{-r})\right\} \tag{21.4}$$

And we could instead check that:

$$\begin{aligned}
\mathbb{E}_s\left[\frac{M_t}{M_s}\right] &= \mathbb{E}_s\left[exp\{-r(N_t - N_s + c(t-s)(1-e^{-r})\}\right] \\
&= \mathbb{E}_s\left[exp\{-r(N_t - N_s\}\right] exp\{c(t-s)(1-e^{-r})\} \\
&= exp\{-c(t-s)(1-e^{-r})\}exp\{c(t-s)(1-e^{-r})\} && \text{Eqn. 21.4} \\
&= 1
\end{aligned}$$

(▽ adaptedness and integrability) we have:

$$M_t = exp\{-r\underbrace{N_t}_{\in\mathcal{F}_t} + c\underbrace{t}_{\in\mathcal{F}_t} - \underbrace{cte^{-r}}_{\in\mathcal{F}_t}\} \in \mathcal{F}_t \quad \forall t$$

Which proves adaptedness. Concerning integrability:

$$\begin{aligned}
\mathbb{E}[|M_t|] = \mathbb{E}[M_t] &= e^{ct}exp\left\{-cte^{-r}\right\}\mathbb{E}[e^{-rN_t}] \\
&= e^{ct}exp\left\{-cte^{-r}\right\}\widehat{\mathcal{P}}_X(r) \\
&= e^{ct}exp\left\{-cte^{-r}\right\}\exp\{-ct(1-e^{-r})\} && \textit{by } X_t \sim Po(ct) \\
&= 1 < \infty
\end{aligned}$$

Poisson compound process, customers in a store Consider a sequence of arrival times $(T_i)_{i\geq 1}$ from a p.r.m. $N \sim \mathcal{P}ois(cLeb)$. We can visualize a sequence of customers spending random money $Y \perp\!\!\!\perp T$ where $Y \sim \pi$ has mean a and variance b^2.
Applying Corollary 15.3 we can safely say (T, Y) is a p.r.m. such that:

$$(T, Y) \sim \mathcal{P}ois(cLeb \times \pi) \quad on \quad \mathbb{R}_+ \times \mathbb{R}_+$$

Where for a fixed time $t \geq 0$ we have that the amount of money spent is:

$$\begin{aligned}
Z_t = \sum_{T_i \leq t} Y_i = \sum_{i=1}^{\infty} Y_i \mathbb{1}_{[0,t]}(T_i) &= \sum_{i=1}^{\infty} f(T_i, Y_i) && f(x,y) := y\mathbb{1}_{[0,t]}(x) \\
&= \int_{[0,t]\times\mathbb{R}_+} \widetilde{N}(dx, dy)y \\
&= \widetilde{N}f
\end{aligned}$$

where $\widetilde{N} = (T, Y)$ is a Poisson Random measure.
We can use the previous results for p.r.m.s from Chapter 13 and 14. The new mean is $\mu = cLeb \times \pi$ with

$\mu(dx, dy) = cdx\pi(dy)$ and:

$$\begin{aligned}
\mathbb{E}[Z_t] = \mathbb{E}[\widetilde{N}f] &= \mu f && \text{Prop. 13.18\#1} \\
&= \int_{\mathbb{R}_+\times\mathbb{R}_+} f(x,y)\mu(dx,dy) \\
&= \int_{\mathbb{R}_+\times\mathbb{R}_+} y\mathbb{1}_{[0,t]}(x)\, cdx\pi(dy) \\
&= ct\int_{\mathbb{R}_+} y\pi(dy) \\
&= cta && \text{by } a = \mathbb{E}[Y]
\end{aligned}$$

Similarly the variance is:

$$\begin{aligned}
V[Z_t] = V[\widetilde{N}f] &= \mu f^2 && \text{Prop. 13.18\#2} \\
&= \int_{\mathbb{R}_+\times\mathbb{R}_+} (y\mathbb{1}_{[0,t]}(x))^2 cdx\pi(dy) \\
&= ct(a^2+b^2) && \text{by } a^2+b^2 = \mathbb{E}[Y]^2 + V[Y] = \mathbb{E}[Y^2]
\end{aligned}$$

Concluding with the Laplace transform:

$$\begin{aligned}
\widehat{\mathcal{P}}_{Z_t}(r) = \widehat{\mathcal{P}}_{\widetilde{N}}(rf) &= \mathbb{E}[e^{-r\widetilde{N}f}] \\
&= exp\left\{-\mu(1-e^{-rf})\right\} && \text{Thm. 13.19} \\
&= \exp\left\{-\int_{\mathbb{R}_+\times\mathbb{R}_+} 1 - e^{-r(y\mathbb{1}_{[0,t]}(x))} cdx\pi(dy)\right\} \\
&= \exp\left\{-ct\int_{\mathbb{R}_+} 1-e^{-ry}\pi(dy)\right\}
\end{aligned}$$

Notice that we used the random variable version with r instead of the functional version since Z_t is a random variable and not the underlying random measure.

Poisson jump structure By the Itô-Lévy decomposition [Çin11](Thm. VII.5.2) we have that $(T_a)_{a\geq 0}$ is an increasing Lévy process (Def. 15.14). For a general one we have $S = (S_t)_{t\in\mathbb{R}_+}$ with Lévy measure satisfying:

$$\int (1\wedge z)\lambda(dz) < \infty$$

which is described in the sense of Definition 15.16 as an integral wrt the underlying p.r.m. on $\mathbb{R}_+\times\mathbb{R}_+$ with mean $dx\lambda(dz)$:

$$S_t = \int_{[0,t]\times\mathbb{R}_+} zN(dx,dz) = \sum_{i:X_i\leq t} Z_i$$

In our specific case we obtain the jump structure by means of a p.r.m. $N(dx, dz)$:

$$N(B) = \sum_a \mathbb{1}_B(a, T_a - T_{a-}) \quad B\in\mathcal{B}(\mathbb{R}_+\times\mathbb{R}_+) \quad mean\ \ dx\lambda(dz) = dx\frac{1}{\sqrt{2\pi z^3}}$$

If (a, z) is an atom then the map $a \to T_a$ controls the abstract *jumps* of size a at *time* z. With:

$$T_a = \int_{[0,a]\times\mathbb{R}_+} zN(dx,dz) = \sum_{i:X_i\leq a} z_i$$

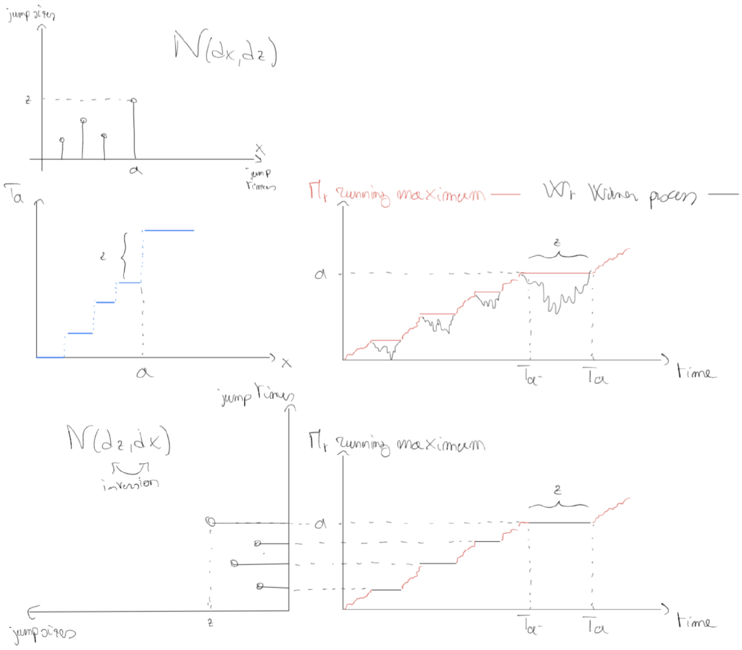

Figure 21.2: Poisson Jumps final plot

Plots of Poisson Jumps Consider Figure 21.2. Atoms of $N(dx, dz)$ are marked with little circles, corresponding to the atom (a, z) there is a jump of size z from T_{a-} to $T_{a-} + z = T_a$. The path $t \to M_t$ stays constant at level a during $[T_{a-}, T_a]$, an interval of length z. Since $N(dx, dz)$ has only countably many atoms, the situation occurs at countably many levels a only. Since there are infinitely many atoms in the strip $[a, a + b] \times \mathbb{R}_+$, the path $t \to M_t$ stays flat at infinitely many levels on its way from a to $a + b$. However, for every $\epsilon > 0$, only finitely many of those sojourns exceed ϵ in duration. The situation at fixed a is simpler. For $a > 0$ almost surely, there are no atoms on the line $\{a\} \times \mathbb{R}_+$, therefore $T_{a-} \overset{a.s.}{=} T_a = 0$.

21.2.6 Stones in a field

The "stones in a field" perspective Let $K \sim \mathcal{P}o(c)$ be Poisson distributed. Consider K to be the random number of stones in a field $E \subset \mathbb{R}^2$. This throwing process is done always with the same mechanism with no regard to total or previous positions (i.e. **independence**).

$$\mathbb{P}[K = k] = \frac{e^{-c}c^k}{k!} \mathbb{1}_{[0,1,\ldots]}(k)$$

Let X_i be the i^{th} stone position. $X_i \sim \lambda(d\vec{x})$ is a distribution over $E \subset \mathbb{R}^2$.
Assume $K \perp\!\!\!\perp \{X_i\}$, as argued before.
The random measure $M(dx)$ assigns the number of stones to the $A \subset E$ region, mathematically:

$$M(A) = \sum_{i=1}^{K} \mathbb{1}_A(X_i)$$

Is the number of stones in region A.
We will show that $M(dx)$ is atomic counting whenever λ is diffuse, i.e. no two stones are in the same position (i.e. Thm. 14.10).

Mean measure In the "Stones in a field" formalism, the mean measure is:

$$c\lambda \quad c = \mathbb{E}[K]$$

We can see this as follows.
For $f = \mathbb{1}_A$ it holds:

$$Mf = M\mathbb{1}_A = \sum_{i=1}^{K} \mathbb{1}_A(X_i)$$

Similarly for $f \in \mathcal{E}_+$:

$$Mf = \int_E f(x)M(dx) = \sum_{i=1}^{K} f(X_i) = \sum_{i=1}^{\infty} f(X_i)\mathbb{1}_{\{K\geq i\}}$$

namely, a sum of images under a random number of K atoms. The last form is for convenience. Then, applying the Definition of mean measure (Def. 13.6):

$$\begin{aligned}
\mathbb{E}[Mf] &= \mathbb{E}\left[\sum_{i=1}^{\infty} f(X_i)\mathbb{1}_{\{K\geq i\}}\right] \\
&= \sum_{i=1}^{\infty} \mathbb{E}\left[f(X_i)\mathbb{1}_{\{K\geq i\}}\right] && \text{linearity} \\
&= \sum_{i=1}^{\infty} \mathbb{E}[f(X_i)]\,\mathbb{E}\left[\mathbb{1}_{\{K\geq i\}}\right] && \text{independence} \\
&= \mathbb{E}[f(X_1)] \sum_{i=1}^{\infty} \underbrace{\mathbb{E}\left[\mathbb{1}_{\{K\geq i\}}\right]}_{=\mathbb{P}[K\geq i]} && \text{iid} \\
&= \mathbb{E}[f(X_1)]\,\mathbb{E}[K] && \text{Lem. 13.7} \\
&= \int_E f(x)\lambda(dx)\cdot c \\
&= c(\lambda f) && \text{integral notation} \\
&= (c\lambda)f
\end{aligned}$$

Eventually, the mean measure is $\nu(dx) = c\lambda(dx)$ where $c = \mathbb{E}[K]$ as claimed.

Laplace functional For $c = \mathbb{E}[K]$ it holds that:

$$\widehat{\mathcal{P}}_M(f) = exp\left\{-c(\lambda(1 - e^{-f}))\right\}$$

(□ **solution)** We perform the following long computation:

$$
\begin{aligned}
\mathbb{E}\left[e^{-Mf}\right] &= \mathbb{E}\left[\exp\left\{-\sum_{i=1}^{K} f(X_i)\right\}\right] && \text{mean measure}\\
&= \mathbb{E}\left[\prod_{i=1}^{K} \exp\left\{-f(X_i)\right\}\right]\\
&= \mathbb{E}\left[\mathbb{E}_K\left[\prod_{i=1}^{K} \exp\left\{-f(X_i)\right\}\right]\right] && \text{unconditioning}\\
&= \mathbb{E}\left[\prod_{i=1}^{K} \mathbb{E}\left[\exp\left\{-f(X_i)\right\}\right]\right] && \text{independence \& Fubini Thm. B.30}\\
&= \mathbb{E}\left[\left(\mathbb{E}\left[\exp\left\{-f(X_1)\right\}\right]\right)^K\right] && \text{iid}\\
&= \mathbb{E}\left[\left(\int_E e^{-f(x)}\lambda(dx)\right)^K\right]\\
&= \mathbb{E}\left[\left(\lambda(e^{-f})\right)^K\right]\\
&= \sum_{k=0}^{\infty} \underbrace{\left(\lambda e^{-f}\right)^k \overbrace{\frac{e^{-c}c^k}{k!}}^{\mathbb{P}[K=k]}}_{\text{pgf of } K \text{ at } t=\lambda e^{-f}} && \lambda(e^{-f}) \text{ is a number}\\
&= \exp\left\{-c\left(1-\lambda e^{-f}\right)\right\} && \text{pgf closed form } X\sim \mathcal{P}o(\lambda) \implies pgf(s)=\sum_{x\geq 0}\mathbb{P}[X=x]s^x = e^{-\lambda(1-s)}\\
&= \exp\left\{-c\left(\int_E \lambda(dx) - \lambda e^{-f}\right)\right\}\\
&= \exp\left\{-c\left(\lambda(1)-\lambda e^{-f}\right)\right\}\\
&= \exp\left\{-c\left(\lambda(1-e^{-f})\right)\right\} && \text{linearity}
\end{aligned}
$$

"Stones in a field" is a Poisson Random measure $N(dx)$ is a p.r.m. in the Definition 13.13 sense.
(□ **solution)** (△ **setup)** wts for $\{A_i\}_{i=1}^n \subset \mathcal{E}$ disjoint it holds:

$$
\begin{cases}
\mathbb{P}[N(A_1)=i_1,\dots,N(A_n)=i_n] = \dfrac{e^{-\nu(A_1)}(\nu(A_1))^{i_1}}{i_1!}\cdots\dfrac{e^{-\nu(A_n)}(\nu(A_n))^{i_n}}{i_n!}\\
\nu = c\lambda : X_i \overset{iid}{\sim} \lambda, c=\mathbb{E}[K]
\end{cases}
$$

(□ **baseline)** wlog let $n=2$ and $A_1\cap A_2=\emptyset$ with $A_3=(A_1\cup A_2)^c$. The collection $\{A_1,A_2,A_3\}$ is a partition of E and we might show △ there. Indeed:

$$
\begin{cases}
\lambda(A_1)+\lambda(A_2)+\lambda(A_3)=1\\
i_1+i_2+i_3=k
\end{cases}
$$

where we call the former $\square(1)$ and the latter $\square(2)$, with k a realization of the r.v. K.
($\bigcirc$ work) it holds that:

$$\begin{aligned}
\mathbb{P}[N(A_1)=i_1, N(A_2)=i_2, N(A_3)=i_3] &= \mathbb{P}[N(A_1)=i_1, N(A_2)=i_2, N(A_3)=i_3, K=k] \\
&[\square(2)] \\
&= \mathbb{P}[N(A_1)=i_1, N(A_2)=i_2, N(A_3)=i_3 | K=k] \\
&[\text{distribution is } \mathcal{M}ultinom\left(3, (\lambda(A_i))_{i=1}^3\right)] \\
&= \frac{e^{-c}c^k}{k!}\frac{k!}{i_1!i_2!i_3!}\left(\lambda(A_1)\right)^{i_1}\left(\lambda(A_2)\right)^{i_2}\left(\lambda(A_3)\right)^{i_3} \\
&= \frac{e^{-(\lambda(A_1)+\lambda(A_2)+\lambda(A_3))}c^{i_1+i_2+i_3}}{i_1!i_2!i_3!}\left(\lambda(A_1)\right)^{i_1}\left(\lambda(A_2)\right)^{i_2}\left(\lambda(A_3)\right)^{i_3} \\
&[\square(1), \square(2)] \\
&= \frac{e^{-c\lambda(A_1)}\left(\lambda(A_1)\right)^{i_1}}{i_1!}\frac{e^{-c\lambda(A_2)}\left(\lambda(A_2)\right)^{i_2}}{i_2!}\frac{e^{-c\lambda(A_3)}\left(\lambda(A_3)\right)^{i_3}}{i_3!}
\end{aligned}$$

21.2.7 Gamma process

The Poisson compound process we consider has form:

$$S_t = \int_{\mathbb{R}_+\times\mathbb{R}_+} f(x,z)N(dx,dz) \quad N \sim \mathcal{P}ois(Leb\times\lambda),\ f(x,z) = \mathbb{1}_{[0,t]}(x)\,z$$

Namely, the c constant in the Lebesgue measure is ignored. It is a simplified version of the candidate of Definition 15.16.

Gamma process, basics Consider the (soon to be) Lévy measure:

$$\lambda(dz) = a\frac{e^{-cz}}{z}dz \quad z\in\mathbb{R}_+, a\in(0,1), c>0$$

We call the arising compound Poisson process (Def. 15.5) $S=(S_t)_{t\in\mathbb{R}_+}$ a Gamma process, and aim to show that it is also an increasing Lévy process (Def. 15.14) with the construction just explained.
($\triangle$ integrability) we want to show that $\int\lambda(dz)(z\wedge 1)<\infty$. This holds since:

- $\int_1^\infty \lambda(dz)(z\wedge 1) = \int_1^\infty \lambda(dz) = \int_1^\infty a\frac{e^{-cz}}{z}dz \to 0$ as $z\to\infty$ sufficiently fast (we take this for granted)
- $\int_0^1 \lambda(dz)(z\wedge 1) = \int_0^1 \lambda(dz)z = \int_0^1 a\frac{e^{-cz}}{z}zdz = \int_0^1 ae^{-cz}dz < \infty$

Given that the condition of Proposition 15.17 is satisfied, we conclude that S is an increasing Lévy process.
($\square$ why Gamma?) wts $(S_t)_{t\in\mathbb{R}_+}$ is such that $S_t \overset{d}{=} X_t \sim Gamma(at,c) \quad \forall t$
We do this by using the Laplace functional. We recall that a Gamma distribution is such that:

$$\widehat{\mathcal{P}}_{X_t}(r) = \left(\frac{c}{r+c}\right)^{at} \tag{21.5}$$

For the Gamma process at a fixed $t\in\mathbb{R}_+$:

$$\begin{aligned}
\mathbb{E}[e^{-rS_t}] &= exp\left\{-t\int_0^\infty(1-e^{-rz})\lambda(dz)\right\} && \text{Prop. 15.17\#2} \\
&= exp\left\{-t\int_0^\infty(1-e^{-rz})a\frac{e^{-cz}}{z}dz\right\} \\
&= exp\left\{-at\int_0^\infty\frac{e^{-cz}-e^{-(c+r)z}}{z}dz\right\}
\end{aligned}$$

(◯ **blue integral**) We focus on the highlighted part for a moment and observe that the inside can be seen as the integral in dt:

$$\begin{aligned}
\int_0^\infty \frac{e^{-cz} - e^{-(c+r)z}}{z} dz &= \int_0^\infty \frac{-e^{-tz}}{z}\Bigg|_{t=c}^{c+r} dz \\
&= \int_0^\infty \int_c^{c+r} \frac{-d}{dt}\frac{e^{-tz}}{z} dtdz \\
&= \int_0^\infty \int_c^{c+r} e^{-tz} dtdz && \text{deriving} \\
&= \int_c^{c+r} \int_0^\infty e^{-tz} dzdt && \text{Fubini Thm. B.30} \\
&= \int_c^{c+r} \frac{-e^{-tz}}{t}\Bigg|_{z=0}^{\infty} \\
&= \int_c^{c+r} \frac{1}{t} dt \\
&= \log t\Bigg|_{t=c}^{c+r} \\
&= \log\left(\frac{c+r}{c}\right)
\end{aligned}$$

(◇ **back to Laplace**) we plut the result of ◯ into □ and conclude that:

$$\begin{aligned}
\mathbb{E}[e^{-rS_t}] &= \exp\left\{-at\log\left(\frac{c+r}{c}\right)\right\} \\
&= \exp\left\{\log\left(\frac{c+r}{c}\right)^{-at}\right\} \\
&= \left(\frac{c}{c+r}\right)^{at}
\end{aligned}$$

Which is equal to the Laplace transfom of $X_t \sim Gamma(at, c)$. By Theorem 6.12 this means that the two variables are equivalent. This holds $\forall t \in \mathbb{R}_+$.

---

j **approximation for simulation** For a Lévy density as that of the gamma process the integral:

$$\lambda((\epsilon, \infty)) = \int_\epsilon^\infty a\frac{e^{-cz}}{z} dz$$

is not available in closed form. To simulate from it, we resort to the notion of incomplete Gamma function (Def. 15.36) and the result of Lemma 15.37. Indeed:

$$\Gamma(0, x) = \gamma_0(x) = \int_x^\infty u^{-1}e^{-u} du \quad \Gamma_1(x) \overset{x\to 0}{\to} \infty$$

And we can express the Lévy measure as:

$$\begin{aligned}
\lambda((\epsilon, \infty)) &= \int_\epsilon^\infty a\frac{e^{-cz}}{z} dz \\
&= \int_{c\epsilon}^\infty a\frac{e^{-x}}{xc} cdz && \text{let } x = cz, dx = cdz \\
&= \int_{c\epsilon}^\infty ae^{-x}x^{-1} dx \\
&= a\gamma_0(c\epsilon)
\end{aligned}$$

So that the following chain holds:

$$a\gamma_0(c\epsilon) = u \iff c\epsilon = \gamma_0^{-1}\left(\frac{u}{a}\right) \iff j(u) = \frac{1}{c}\gamma_0^{-1}\left(\frac{u}{a}\right)$$

And eventually:

$$\begin{aligned} S_t &= \sum_{i=1}^{\infty} j\left(\frac{1}{t}G_i\right) && \text{Obs. 15.33} \\ &= \sum_{i=1}^{\infty} \frac{1}{c}\gamma_0^{-1}\left(\frac{G_i}{at}\right) \end{aligned}$$

And we know how to approximate the inverse of the incomplete Gamma function (Lem. 15.37).

Symmetric Gamma process Recall that a Gamma process is an increasing Lévy process (Def. 15.14) with measure and distribution:

$$\lambda(dz) = \frac{ae^{-cz}}{z}dz,\ z \in \mathbb{R}_+, \quad X_t \sim \mathcal{G}amma(at, c)\ \forall t \in \mathbb{R}_+$$

Let $X_t^+ \perp\!\!\!\perp X_t^-$ be independent copies, and set $X_t = X_t^+ - X_t^-$. Then, X_t is a pure jump Lévy process according to Definition 17.7 with measure:

$$\lambda(dz) = \frac{ae^{-c|z|}}{|z|}dz$$

We aim to evaluate its characteristic function to see if it coincides with some known distribution.

$$\begin{aligned} \mathbb{E}\left[e^{irX_t}\right] &= \mathbb{E}\left[e^{irX_t^+}\right]\mathbb{E}\left[e^{i(-r)X_t^-}\right] && \text{independence} \\ &= \left(\frac{c}{c+ir}\right)^{at}\left(\frac{c}{c-ir}\right)^{at} && \text{previous result} \\ &= \left(\frac{c^2}{(c+ir)(c-ir)}\right)^{at} \\ &= \left(\frac{c^2}{c^2+r^2}\right)^{at} \end{aligned}$$

Which means that the characteristic exponent is real:

$$\psi(r) = \frac{1}{t}\log\left[\mathbb{E}[e^{irX_t}]\right] = a\log\left[\frac{c^2}{c^2+r^2}\right] \in \mathbb{R}$$

While X_t has no known distribution, it can be shown that the total variation $V = X^+ + X^-$ is such that:

$$V_t \sim Gamma(2at, c) \quad \forall t \in \mathbb{R}_+$$

21.2.8 Stable process

The Poisson compound process we consider has form:

$$S_t = \int_{\mathbb{R}_+\times\mathbb{R}_+} f(x,z)N(dx,dz) \quad N \sim \mathcal{P}ois(Leb \times \lambda),\ f(x,z) = \mathbb{1}_{[0,t]}(x)\, z$$

Namely, the c constant in the Lebesgue measure is ignored. It is a simplified version of the candidate of Definition 15.16.

Stable process of index α Consider the (soon to be) Lévy measure:

$$\lambda(dz) = \frac{1}{\Gamma(1-a)} acz^{-1-a}dz \quad z \in \mathbb{R}_+, a \in (0,1), c > 0$$

We will see that the arising process $S = (S_t)_{t\in\mathbb{R}_+}$ in the postulated form above is an increasing Lévy process (Def. 15.14) and has some nice properties.

($\triangle$ integrability) we aim to show that the integrability condition for being increasing Lévy holds. For this purpose, notice that:

$$\Gamma(x) := \int_0^\infty t^{x-1}e^{-t}dt \quad \Gamma(n) = (n-1)! \, \forall n \in \mathbb{N}$$

So that:

- $\int_1^\infty \lambda(dz)(z \wedge 1) = \int_1^\infty \lambda(dz) = \frac{1}{\Gamma(1-a)} ac \int_1^\infty z^{-1-a}dz = \frac{1}{\Gamma(1-a)} ac\frac{1}{a} - z^{-a}\Big|_{z=1}^{\infty} = \frac{1}{\Gamma(1-a)} c\frac{1}{a} < \infty$
- by $a \in (0,1)$

$$\int_0^1 \lambda(dz)(z \wedge 1) = \frac{1}{\Gamma(1-a)} ac \int_0^1 z^{-1-a}z = \frac{1}{\Gamma(1-a)} ac \int_0^1 z^{-a}$$
$$= \frac{1}{\Gamma(1-a)} ac\frac{1}{1-a} z^{1-a}\Big|_{z=0}^{1} = \frac{1}{\Gamma(1-a)} \frac{ac}{1-a} < \infty$$

Making their sum finite. By $\int_0^\infty \lambda(dz)(z \wedge 1) < \infty$ we can apply Proposition 15.17#1 and conclude that S is an increasing Lévy process.

About the stable process Notice that even though $S_t < \infty$ a.s. the process has no expectation. Infact:

$$\begin{aligned}
\mathbb{E}[S_t] &= \mathbb{E}\left[\int_{(0,t]\times\mathbb{R}_+} z \; N(dx, dz)\right] && \text{postulated form of } S_t \; b = 0 \\
&= \mathbb{E}\left[Nf\right] && f(x,z) := \mathbb{1}_{[0,t]}(x)\, z \\
&= \nu f && \text{Def. 13.6} \\
&= \int f\nu(dx) \\
&= \int \mathbb{1}_{[0,t]}(x)\, z dx\lambda(dz) \\
&= \int_0^t \int_0^\infty z\lambda(dz)dx \\
&= t\int_0^\infty z\lambda(dz) \\
&= t\int_0^\infty z\frac{1}{\Gamma(1-\alpha)} acz^{-1-a}dz \\
&= \frac{tca}{\Gamma(1-\alpha)} \int_0^\infty z^{-a}dz && \text{improper integral}
\end{aligned}$$

Where the improper integral diverges at ∞, and S_t has no expectation. For the sake of completeness, we report the calculation here below. An improper integral of this form can be calculated considering the discontinuity at zero and the divergent limit on the other side:

$$\int_0^\infty z^{-a}dz = \int_0^1 z^{-a}dz + \int_1^\infty z^{-a}dz = \lim_{b\to 0} \int_b^1 z^{-a}dz + \lim_{c\to\infty} \int_1^c z^{-a}dz$$

while the indefinite integral is easily found as $\frac{1}{1-a} z^{-a+1} + K, K \in \mathbb{R}$. Ignoring the constant which is positive since $a \in (0,1)$ by construction and $1 - a > 0$, we get:

$$\lim_{b\to 0} z^{-a+1}\Big|_{z=b}^{1} = 1 - \lim_{b\to 0} b^{1-a} = 1 < \infty$$

But

$$\lim_{c\to\infty} z^{1-a}\Big|_{z=1}^{c} = \lim_{c\to\infty} c^{1-a} - 1 = \infty$$

and the sum diverges. All the constants are positive and the claim is proved.

The stability of the stable process $S = (S_t)_{t\in\mathbb{R}_+}$ from the first paragraph is stable in the sense that:

$$S_{ut} \overset{d}{=} u^{\frac{1}{a}} S_t \quad \forall u, t \in \mathbb{R}_+ \quad i.e. \quad S_t \overset{d}{=} t^{\frac{1}{a}} S_1 \quad \forall t \in \mathbb{R}_+$$

(△ Laplace approach) use the Laplace transform from Proposition 15.17#2.

$$\begin{aligned}
\mathbb{E}\left[e^{-rS_t}\right] &= \exp\left\{-t\int_0^\infty (1-e^{-rz})\lambda(dz)\right\} \\
&= \exp\left\{-t\int_0^\infty (1-e^{-rz})\frac{ac}{\Gamma(1-a)}z^{-1-a}dz\right\} \\
&= \exp\left\{-tcr^a\right\} && \text{proved below in } \square
\end{aligned}$$

Where the last equality is $\int_0^\infty (1-e^{-rz})az^{-1-a}dz = r^a\Gamma(1-a)$.
(□ missing equality) by direct computation:

$$\begin{aligned}
\int_0^\infty (1-e^{-rz})az^{-1-a}dz &= \int_0^\infty (1-e^{-t})a\left(\frac{t}{r}\right)^{-1-a}\frac{dt}{r} && t = rz \quad dt = rdz \\
&= r^a\int_0^\infty (1-e^{-t})at^{-1-a}dt \\
&= -r^a\int_0^\infty \underbrace{(1-e^{-t})}_{g}\underbrace{(-at^{-1-a})}_{f'}dt && \text{integrate by parts} \\
&= -r^a\left(\underbrace{(1-e^{-t})t^{-a}\Big|_0^\infty}_{=0} - \int_0^\infty e^{-t}t^{-a}dt\right) && t^{-a} = t^{1-a-1} \\
&= -r^a\left(-\int_0^\infty e^{-t}t^{1-a-1}dt\right) && \text{Gamma integral at } 1-a \\
&= r^a\Gamma(1-a)
\end{aligned}$$

(♠ back to Laplace) by △ the general form at time ut is:

$$\begin{aligned}
\mathbb{E}\left[e^{-rS_{ut}}\right] &= \exp\left\{-utcr^a\right\} \\
&= \exp\left\{-ct\left(u^{\frac{1}{a}}r\right)^a\right\} \\
&= \mathbb{E}\left[e^{-u^{\frac{1}{a}}rS_t}\right] \\
&= \mathbb{E}\left[e^{-r(u^{\frac{1}{a}}S_t)}\right] \\
\implies S_{ut} &\overset{d}{=} u^{\frac{1}{a}}S_t && \forall u, t
\end{aligned}$$

Stable process j, for simulation Remember that $\lambda(dz) = \frac{ac}{\Gamma(1-a)}z^{-1-a}dz$ and:

$$\lambda((\epsilon,\infty)) = \frac{c}{\Gamma(1-a)}\epsilon^{-a}$$

Using Definition 15.31 for j we have that the solution in ϵ to the infimization is:

$$j(u)\ :\ \lambda((\epsilon,\infty)) = u \implies j(u) = \left(\frac{c}{\Gamma(1-a)}\right)^{\frac{1}{a}} u^{-\frac{1}{a}}$$

Using Observation 15.33 we can safely say that by the exponential distribution of the $G_i \sim \mathcal{E}xp(1)$ with $\frac{1}{t}\sum G_i \sim Exp(1)$ it is the case that:

$$\begin{cases} S_t = \sum_{i:T_i \le t} \widehat{c}(U_i)^{-\frac{1}{a}} = \sum_{i=1}^{\infty} \widehat{c}\left(\frac{1}{t}G_i\right)^{-\frac{1}{a}} \\ \widehat{c} = \left(\frac{c}{\Gamma(1-a)}\right)^{\frac{1}{a}} \end{cases}$$

If we take $t = 1$, (U_i) forms a p.r.m. with unit intensity. In particular the arrival times of the allied counting process (that is U_i in increasing order) are equal in distribution to:

$$G_1 = E_1,\ G_2 = E_1 + E_2, \dots, G_k = E_1 + \cdots E_k$$

Where (E_i) are exponential iid of unit rate. Hence:

$$S_1 = \sum_{i=1}^{\infty} \widehat{c} G_i^{-\frac{1}{a}}$$

while in general:

$$S_t = \sum_{i=1}^{\infty} \widehat{c}\left(\frac{1}{t}G_i\right)^{-\frac{1}{a}} \qquad \frac{1}{t}G_i \overset{iid}{\sim} \mathcal{E}xp(1)$$

Isotropic stable process Let $X_t^+ \perp\!\!\!\perp X_t^-$ be independent copies of the stable process from the second paragraph. The density of $X_t = X_t^+ - X_t^-$ is:

$$\lambda(dz) = \frac{ac}{\Gamma(1-a)}|z|^{1-a}dz, \quad z \neq 0, a \in (0,1)$$

Such a process is pure jump Lévy according to Definitions 17.1, 17.7 and has Laplace transform:

$$\begin{aligned} \mathbb{E}[e^{irX_t}] &= \exp\left\{t\int_{\mathbb{R}}(e^{irz}-1)\lambda(dz)\right\} \\ &= \exp\left\{tc\cos\left(\frac{1}{2}\pi a\right)|r|^a\right\} \end{aligned}$$

With characteristic exponent:

$$\psi(r) = \int_{\mathbb{R}}(e^{irz}-1)\lambda(dz) = -c\cos\left(\frac{1}{2}\pi a\right)|r|^a$$

Which is stable since $X_t \overset{d}{=} t^{\frac{1}{a}}X_t\ \forall t$.

21.2.9 Wiener Process

Wiener process is stable We showed that a Wiener process $W = (W_t)_{t\in\mathbb{R}_+}$ (Def. 11.55) is such that:

$$W_t = \sqrt{t}W_1, W_t \sim \mathcal{N}(0,t), W_1 \sim \mathcal{N}(0,1)$$

which is the result of Proposition 11.56#3 for $s = 0$.
This is equivalent to saying that the process is stable as that of the first paragraph.

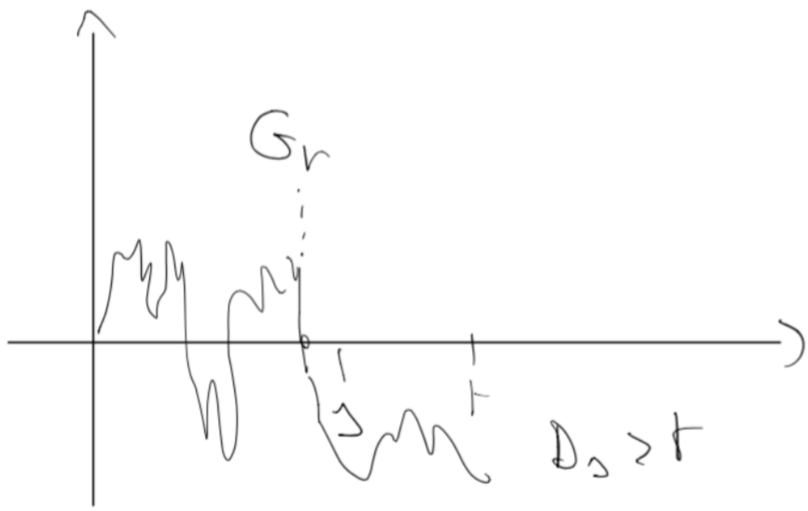

Figure 21.3: Recurrence times of a Wiener process

Recurrence times Define G_t as the last time at zero before t and D_t as the first time at zero after t. Namely:

$$G_t \coloneqq \sup\{s \in [0,t] : W_s = 0\} \quad D_t = \inf\{u \in (t, \infty) : W_u = 0\}$$

Accordingly, the forward recurrence time is $R_t = D_t - t$ and the backward recurrence time is $Q_t = t - G_t$.
By the process $(W_t)_{t\in\mathbb{R}_+}$ being such that $W_t \sim \mathcal{N}(0,t) \forall t$ we have $\mathbb{P}[W_t = 0] = 0$ a.s. by the diffusivity of a normal distribution. Then:

$$G_t < t < D_t \; a.s. \quad \& \quad s \in [0,t] \implies \{G_t < s\} = \{D_s > t\}$$

an intuition is given in Figure 21.3. So, if $W_t = a > 0 \implies R_t$ is the hitting time from above of the barrier $-a$ of the rescaled process:

$$(W_u \circ \theta_t)_{u\geq 0} = (W_{t+u} - W_t)_{u \geq 0}$$

By the markov property of Wiener processes $\widetilde{W}_u = W_{t+u} - W_t \perp\!\!\!\perp \mathcal{F}_t$ is again Wiener and we can see that:

$$R_t = \inf\{u > 0 : \widetilde{W}_u < -a\}$$

Additionally, by symmetry (Thm. 18.16#1) we have:

$$\widetilde{W}_u < -a \iff -\widetilde{W}_u > a \iff \widetilde{W}_u > a$$

So that:

$$R_t \overset{d}{=} T_a = \inf\{u > 0 : W_u \circ \theta_t = \widetilde{W}_u > a\} \quad a = W_t, \; T_a \perp\!\!\!\perp W_t$$

Which means that if T_a is known $\forall a > 0 \implies R_t$ is known and so is $D_t = R_t + t$ and G_t via $\mathbb{P}[G_t < s] = \mathbb{P}[D_s > t]$, as well as $Q_t = t - G_t$.

More results from the Proposition We could also get:

1. $R_t \overset{d}{=} tC^2$ (by direct application of Lemma 19.17#4) for $C \sim \mathcal{C}auchy$
2. G_t has density $f(x) = \frac{2}{\pi} \arcsin \sqrt{\frac{s}{t}}$
3. $Q_t \overset{d}{=} G_t \overset{d}{=} tA$

Proof. **(Claim #1)** trivial.
(Claim #2) from $G_t \stackrel{d}{=} tA$ using Lemma 19.17#3 we get:

$$\mathcal{P}(G_t < s) = \mathcal{P}(tA < s) = \mathcal{P}\left(A < \frac{s}{t}\right) = \frac{2}{\pi} \arcsin \sqrt{\frac{s}{t}}$$

(Claim #3) consider $Q_t = t - G_t$ where:

$$\begin{aligned} \mathbb{P}[Q_t < s] &= \mathbb{P}[t - G_t < s] \\ &= \mathbb{P}[t - tA < s] \\ &= \mathbb{P}[t(1 - A) < s] && 1 - A \stackrel{d}{=} A \\ &= \mathbb{P}[tA < s] \\ &= \mathbb{P}[G_t < s] \end{aligned}$$

We proved $Q_t \stackrel{d}{=} G_t$, a shift of proportions. □

21.2.10 Cauchy Process

Standard Cauchy process Let the Lévy measure be:

$$\lambda(dz) = \frac{1}{\pi z^2} dz \quad z \in \mathbb{R}_+$$

It holds that (be careful with the second as it is a bit tricky):

$$\begin{aligned} \int_{-1}^{1} z^2 \lambda(dz) &= \int_{\mathbb{B}} z^2 \lambda(dz) = \frac{2}{\pi} < \infty \\ \int_{\mathbb{B}} |z| \lambda(dz) &= \int_{-1}^{1} |z| \frac{1}{\pi z^2} dz \\ &= \int_{-1}^{0} -z \frac{1}{\pi z^2} dz + \int_{0}^{1} z \frac{1}{\pi z^2} dz && \text{basically without modulus undefined} \\ &= \frac{1}{\pi} \left(\int_0^1 \frac{1}{z} dz - \int_{-1}^{0} \frac{1}{z} dz \right) \\ &= \frac{1}{\pi} \left(\ln(|z|) \Big|_0^1 - \ln(|z|) \Big|_{-1}^{0} \right) \\ &= \frac{1}{\pi} (\infty + \infty) \\ &= +\infty \end{aligned}$$

So we can apply Theorem 17.16 having infinite total variation.
Let $X_t = X_t^d + X_t^e$ where:

$$\begin{aligned} X_t^d &= \int_{[0,t] \times \mathbb{B}} z \left(N(dx, dz) - dx \lambda(dz) \right) && \text{jumps size} \leq 1 \\ X_t^e &= \int_{[0,t] \times \mathbb{B}^c} z N(dx, dz) && \text{jumps size} > 1 \end{aligned}$$

Here X_t^e is such that $\int_1^\infty \lambda(dz) < \infty$.
The characteristic exponent is:

$$\psi(r) = \int_{\mathbb{B}} (e^{irz} - 1 - irz) \lambda(dz) + \int_{\mathbb{B}^c} (e^{irz} - 1) \lambda(dz)$$

since we can split the process into the infinite variation part and the finite one. Recalling Observation 17.17 we also have that:

$$X_t^d = \lim_{\epsilon \downarrow 0} X_t^{d,\epsilon} \quad X_t^{d,\epsilon} = \int_{[0,t]\times\mathbb{B}_\epsilon} zN(dx,dz) - t \underbrace{\int_{\mathbb{B}_\epsilon} z\lambda(dx,dz)}_{=0 \text{ by symmetry}}$$

so that $X_t^{d,\epsilon}$ requires no compensation and we eventually get:

$$\begin{aligned}
X_t &= \lim_{\epsilon\downarrow 0} \int_{[0,t]\times\mathbb{B}_\epsilon} zN(dx,dz) + \int_{[0,t]\times\mathbb{B}^c} zN(dx,dz) \\
&= \lim_{\epsilon\downarrow 0} \int_{[0,t]\times\mathbb{R}_\epsilon} zN(dx,dz) + \int_{[0,t]\times\mathbb{B}^c} zN(dx,dz) && \text{Dominated conv. Thm. A.51}
\end{aligned}$$

where $\mathbb{R}_\epsilon = \{x : |x| > \epsilon\}$ dominates $\mathbb{B}_\epsilon$. Notice that this is **not a pure jump process** in the sense of Definition 17.7. Nevertheless, the Laplace transform is:

$$\begin{aligned}
\mathbb{E}[e^{irX_t}] &= \exp\left\{t \lim_{\epsilon\downarrow 0} \int_{\mathbb{R}_\epsilon} (e^{irz}-1)\frac{1}{\pi z^2}dz\right\} \\
&= \exp\left\{t \int_{\mathbb{R}} (e^{irz}-1)\frac{1}{\pi z^2}dz\right\} \\
&= \exp\left\{t \int_{\mathbb{R}} (\cos(rz)+i\sin(rz)-1)\frac{1}{\pi z^2}dz\right\} \\
&= \exp\left\{t \int_{\mathbb{R}} (\cos(rz)-1)\frac{1}{\pi z^2}dz\right\} && \frac{\sin(rz)}{\pi z^2} \text{ symm around } 0 \\
&= \exp\{t|r|\} && \text{Prop. 17.21\#2}
\end{aligned}$$

And we have that $X_t \overset{d}{=} t^{\frac{1}{1}} X_1$ (stability with index 1). Moreover by:

$$X_1 \overset{d}{=} \frac{Z_1}{Z_2}, \quad Z_1, Z_2 \sim \mathcal{N}(0,1), \quad X_1 \sim \mathcal{C}auchy(1) \text{ (Prop. 17.21\#1)}, \; f(x) = \frac{1}{\pi(1+x)^2}$$

We have that:

$$X_t \; : \; f(x) = \frac{t}{\pi(t^2+x^2)} \quad \forall x\in\mathbb{R}, \forall t \in \mathbb{R}_+$$

Concluding Example Standard Cauchy process In the context of the standard Cauchy process we add that:

1. $X_1 \sim \mathcal{C}auchy(1)$
2. $\int_{\mathbb{R}}(\cos(rz)-1)\frac{1}{\pi z^2}dz = |r|$

(Claim #1) use Lemma 17.20 and the fact that the Characteristic function of a $\mathcal{C}auchy(1)$ distribution is $\Phi(r) = e^{-|r|}$. We have:

$$\begin{aligned}
\frac{1}{2\pi}\int_{\mathbb{R}} e^{-irx}e^{-|r|}dr &= \frac{1}{\pi}\int_{\mathbb{R}} e^{ir(-x)} \underbrace{\frac{1}{2}e^{-|r|}}_{\text{density Laplace}} dr \\
&= \frac{1}{\pi}\frac{1}{1+(-x)^2} && \text{characteristic function Laplace} \\
&= \frac{1}{\pi(1+x^2)} \\
&= f(x) && X_1
\end{aligned}$$

(Claim #2) by the symmetry of the integrated function, the claim is equivalent to:

$$2\int_0^\infty (1-\cos(rz))\frac{1}{\pi(z^2)}dz = |r|$$

(△ first step) for a triangular distribution $U + U' - 1$ where $U \sim \mathcal{U}nif(0,1)$ the density is

$$f(x) = 1 - |x| \quad |x| < 1$$

with characteristic function:

$$\begin{aligned}
\Phi(r) &= \int_{-1}^{1} e^{-rx}(1-|x|)dx \\
&= \int_{-1}^{1} (\cos(rx) + i\sin(rx))(1-|x|)dx \\
&= \int_{-1}^{1} \cos(rx)(1-|x|)dx && \text{symmetry of second term} \\
&= 2\int_{0}^{1} \underbrace{\cos(rx)}_{f'}\underbrace{(1-x)}_{g}dx && \cos(-x) = \cos(x) \quad \forall x \\
&= \underbrace{2\frac{-\sin(rx)(1-x)}{r}\Big|_{x=0}^{1}}_{=0} -2\int_0^1 \frac{-\sin(rx)(-1)}{r}dx && \text{integration by parts} \\
&= -\frac{2}{r^2}\cos(rx)\Big|_{x=0}^{1} \\
&= \frac{2}{r^2}(1-\cos(r)) && \cos' = -\sin',\ \sin' = \cos
\end{aligned}$$

(□ density) using Lemma 17.20 for the triangular distribution of △ we have

$$f(x) = (1-|x|)\mathbb{1}_{[-1,1]}(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-irx}\frac{2}{r^2}(1-\cos(r))dr \implies 1 = \frac{1}{\pi}\int_{\mathbb{R}}\frac{1-\cos(r)}{r^2}dr \quad at\ x = 0$$

where, using $r = u,\ u = rx : du = |r|dx$:

$$\int_{\mathbb{R}} \frac{1-\cos(u)}{\pi u^2}du = \int_{\mathbb{R}} \frac{1-\cos(rx)}{\pi r^2 x^2}|r|dx = \int_{\mathbb{R}} \frac{1-\cos(rx)}{\pi |r| x^2}dx = 1$$

So that:

$$\int_{\mathbb{R}} \frac{1-\cos(rx)}{\pi x^2}dx = |r|$$

and we have proved the claim.

21.2.11 Miscellaneous Examples

Product of independent random variables martingale Let:

- $R_1, R_2, \ldots$ be independent and such that $\mathbb{E}[R_k] = 1$ and $V[R_k] < \infty\ \forall k$
- $M_0 = 1$ and $M_n = M_{n-1}R_n = M_0 R_1 R_2 \cdots R_n$

We check that M is a martingale with respect to its natural filtration according to Definition 11.35.
(adaptedness) Clearly $(M_n)_{n\in\mathbb{N}}$ is adapted to $\mathcal{F} = \sigma\left((M_n)_{n\in\mathbb{N}}\right)$.
(integrability) Observe that:

$$\mathbb{E}[|M_n|] = \mathbb{E}\left[|M_{n-1}R_n|\right] = \mathbb{E}\left[\left|M_0\prod_{k=1}^{n} R_k\right|\right]$$

Where by induction we can show that:

- $\mathbb{E}[|M_1|] = \mathbb{E}[|M_0 R_1|] = \mathbb{E}[|R_1|] < \infty$ by hypothesis

- one step forward

$$\begin{aligned}
\mathbb{E}[|M_2|] &= \mathbb{E}[|M_0 R_1 R_2|] && M_0 R_1 R_2 = M_1 R_2 \\
&\leq \sqrt{\mathbb{E}[|M_1^2|]\mathbb{E}[|R_2^2|]} && \text{Cauchy-Schwartz} \\
&= \sqrt{\mathbb{E}[R_1^2]\mathbb{E}[R_2^2]} && \\
&< \infty && \text{by } V[R_k] < \infty \forall k
\end{aligned}$$

- naturally iterate

So that $\mathbb{E}[M_n] < \infty \forall n$.
(martingale equality) Using Proposition 11.40 we check only for $k = 1$ increments:

$$\begin{aligned}
\mathbb{E}_n[M_{n+1}] = \mathbb{E}_n[M_n R_{n+1}] &= M_n \mathbb{E}_n[R_{n+1}] && \text{since } M_n \in \mathcal{F}_n \text{ and deterministic conditioning} \\
&= M_n \mathbb{E}[R_{n+1}] && \text{remove } n \text{ since } R_{n+1} \perp\!\!\!\perp R_n \\
&= M_n \cdot 1 && \text{hypothesis}
\end{aligned}$$

And the claim holds: $(M_n)_{n\in\mathbb{N}}$ is a martingale.

An investment strategy Let T be the random time to exit the market. Assume T is a stopping time wrt $(\mathcal{F}_n)_{n\in\mathbb{N}}$ as per Definition 11.9.
Let $F_n = \mathbb{1}_{[0,T]}(n)$ be a random indicator. Then:

$$\begin{aligned}
X_n &= \int_{[0,n]} F dM \\
&= \int_{[0,n]} \mathbb{1}_{[0,T]} dM \\
&= \int \mathbb{1}_{[0,n]} \mathbb{1}_{[0,T]} dM \\
&= \int \mathbb{1}_{[0,n\wedge T]} dM \\
&= M_{n\wedge T} \\
&= \begin{cases} M_n & n < T \\ M_T & n \geq T \end{cases}
\end{aligned}$$

Namely, the simplest strategy one can think of invest everything up to time T, and sell right after. With perfectly shared information, there is no profit or loss.
We will show in Corollary 12.10 that up to reasonable conditions this is again a martingale.

Some predictable processes and their integral martingales We present four easy examples with $S \leq T$ almost surely two stopping times and $V \in \mathcal{F}_S$.
(one extreme) let T be a stopping time. Then the process $F_n = \mathbb{1}_{[0,T]}(n)$ is such that :

$$F_{n+1} = \mathbb{1}_{[0,T]}(n+1) = \mathbb{1}_{\{n+1\leq T\}} = \mathbb{1}_{\{T\leq n\}^c} \in \mathcal{F}_n$$

Since T is a stopping time. Clearly the process $(F_n)_{n\in\mathbb{N}}$ is predictable.
(other extreme) Let $F_n = V\mathbb{1}_{(S,\infty)}(n)$ for $S \leq T$ two stopping times, and $V \in \mathcal{F}_S$. Then:

$$V \in \mathcal{F}_S \implies F_{n+1} = V\mathbb{1}_{(S,\infty)}(n+1) = V\mathbb{1}_{\{n+1\geq S\}} = V\mathbb{1}_{\{S\leq n\}^c} \in \mathcal{F}_n$$

Where we applied Theorem 11.23#1. The process is predictable.
(two extremes) for $F_n = \mathbb{1}_{(S,T]}(n) = \mathbb{1}_{[S,\infty)}(n) \cdot \mathbb{1}_{[0,T)}(n)$ the product of two predictable processes, the process

is predictable.
(two extremes + V) let $F_n = V\mathbb{1}_{(S,T]}(n)$, the result is trivial by the previous ones.
For all three cases, we have a stochastic integral:

$$X_n = \int_{[0,n]} F dM$$

Where $(X_n)_{n\in\mathbb{N}}$ is a martingale by Theorem 12.9.

Occupancy problem Consider n independent bins and m balls. We are interested in the number of empty bins, denoted as Z.

(△ setting) We set:

$$C_i \coloneqq \text{bin chosen at } i^{th} \text{ ball} \quad : \quad \mathbb{P}[C_i = j] = \frac{1}{n} \quad j = 1, \dots, n$$

Which are *iid* random variables.

(□ Azuma inequality by martingales) let $(\mathcal{F}_n)_{n\in\mathbb{N}} = \sigma(\{C_i\}_{i=1}^m)$ and $Z_t \coloneqq \mathbb{E}_t[Z]$. In this setting, Z_t is the estimate of the number of empty bins at the end having observed t throws.
Using Proposition 11.50 we have:

$$Z \in [0,n] \; bounded \implies (Z_n)_{n\in\mathbb{N}} \quad uniformly\ integrable$$

Then set $Z_0 = \mathbb{E}_0[Z] = \mu =$ by the martingale equality (Def. 11.35#3).
Notice that $Z \in \mathcal{F}_m \implies Z_t = Z \forall t \geq m$, meaning that after having thrown all the balls (m throws), Z belongs to the σ-algebra. This is rather intuitive.

(◯ Azuma inequality applied) We have that:

$$\square \implies \mathbb{P}[|Z-\mu| > \delta\mu] = \mathbb{P}[|Z_m - Z_0| > \delta\mu]$$

And using Azuma Inequality (Thm. 12.49), by $|Z_{t+1} - Z_t| \leq 1$ we set $c = 1$ and get that:

$$\implies \mathbb{P}\left[|Z_t - Z_0| > \lambda\sqrt{t}\right] \leq 2e^{-\frac{\lambda^2}{2}} \quad : \quad c = 1$$

With

$$\lambda\sqrt{t} = \delta\mu \implies \lambda = \frac{\delta\mu}{c\sqrt{m}}$$
$$\implies \mathbb{P}[|Z-\mu| > \delta\mu] \leq 2exp\left\{-\frac{-\delta^2\mu^2}{2m}\right\} \quad c = 1 \tag{21.6}$$

(◇ finding μ) letting $X_j \coloneqq \#$ balls in $j \in \{1,\dots,n\}$ we get:

$$Z = \sum_{j=1}^{n} \mathbb{1}\{X_j = 0\} \quad : \quad X_1,\dots,X_n \sim \mathcal{M}ultinom\left(m, \left(\frac{1}{n},\dots,\frac{1}{n}\right)\right)$$
$$X_j \sim \mathcal{B}inom\left(m, \frac{1}{n}\right)$$

So that:

$$\begin{aligned}
\mathbb{E}[Z] = Z_0 = \mu = \sum \mathbb{E}\left[\mathbb{1}\{X_j = 0\}\right] && \text{linearity of integral} \\
= \sum \mathbb{P}(\{X_j = 0\}) && \\
= n\mathbb{P}[X_1 = 0] && \text{iid assumption} \\
= n\left(1 - \frac{1}{n}\right)^m && \\
= np && p := \left(1 - \frac{1}{n}\right)^m
\end{aligned}$$

And Equation 21.6 in □ becomes:

$$\begin{aligned}
\mathbb{P}[|Z-\mu| > \delta\mu] \le 2exp\left\{-\frac{1}{2}\frac{\delta^2 n^2 p^2}{m}\right\} && \\
= exp\left\{-\frac{1}{2}\delta^2 np^2\right\} && \text{if } n = m
\end{aligned}$$

Notice also that as $m = n \to \infty$ we also have that $p \to e^{-1}$

(▽ informal Chernoff's bound) ignoring the dependency let $\delta \in (0,1), c > 0, \mu = np$ and derive a much more restrictive bound on the probability by Theorem 7.25:

$$\mathbb{P}[|Z-\mu| > \delta\mu] \le 2e^{-\frac{1}{2}cnp\delta^2}$$

Averages We show that the average process needs a specific condition to be a martingale, as discussed earlier. Assume a discrete process $(X_n)_{n\in\mathbb{N}}$ is adapted to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ and is integrable. Then let:

$$\overline{X_n} = \frac{1}{n}\sum^n X_i \quad assume \quad \mathbb{E}_n[X_{n+1}] = \overline{X_n}$$

(△ aim) we want to show that $\left(\overline{X}_n\right)_{n\in\mathbb{N}}$ an $\mathcal{F}$-martingale according to Definition 11.35.

(□ adaptedness) as $\overline{X}_n = \frac{1}{n}\sum^n X_i : X_i \in \mathcal{F}_n \ \forall i \le n$ adapteness is trivial.

(○ integrability) Notice that $\mathbb{E}[|X_n|] \le \frac{1}{n}\sum \mathbb{E}[|X_i|] < \infty$, by trivial application of linearity,Jensen's inequality and the hypothesis of integrability.

(◇ martingale equality) we proceed by manipulation:

$$\begin{aligned}
\mathbb{E}_n\left[\overline{X_{n+1}} - \overline{X_n}\right] &= \frac{1}{n(n+1)}\mathbb{E}_n\left[n\left(\sum^{n+1} X_i\right) - (n+1)\left(\sum^n X_i\right)\right] \\
&= \frac{1}{n(n+1)}\mathbb{E}_n\left[n\left(\sum^{n+1} X_i\right) - n\left(\sum^n X_i\right) - \left(\sum^n X_i\right)\right] \\
&= \frac{1}{n(n+1)}\mathbb{E}_n\left[nX_{n+1} - \left(\sum^n X_i\right)\right] \\
&= \frac{1}{n(n+1)}\mathbb{E}_n\left[n\frac{\sum X_i}{n} - \left(\sum^n X_i\right)\right] && \text{by hypothesis} \\
&= 0
\end{aligned}$$

And the equality holds. By $\square, \bigcirc, \diamondsuit \implies \triangle$ claim is verified.
Notice however that $\overline{X}_n \neq 0$ a.s. since $X_{n+1} \perp\!\!\!\perp \mathcal{F}_n$ so that:

$$\mathbb{E}[X_{n+1}] = \overline{X_n} = 0 \iff X_i = 0 \; \forall i$$

Homogeneous counting measure and Weibull] Let $N(dx, dy)$ be a p.r.m. on $E = \mathbb{R}^2$, with mean measure $\nu(dx, dy) = cLeb(dx, dy)$. It holds that N is invariant to translations and rotations (i.e. homogeneous). Let R be the distance of the closest atom of N from the origin $\mathbf{0} = (0, 0)$. We describe R via its probability distribution $\mathbb{P}[R > r]$. It turns out that this is equivalent to a ball having null mass:

$$B_r(\mathbf{0}) = \{(x, y) : x^2 + y^2 \leq r^2\} : N(B_r(\mathbf{0})) = 0 \quad \forall r > 0$$

This can be seen as:

$$\begin{aligned}
\mathbb{P}[R > r] &= \mathbb{P}[N(B_r(\mathbf{0})) = 0] \\
&= e^{-\nu(B_r(\mathbf{0}))} && N(B_r(\mathbf{0})) \sim \mathcal{P}o(\nu(B_r(\mathbf{0}))) \\
&= \exp\{-c \cdot Leb(B_r(\mathbf{0}))\} \\
&= \exp\{-c \cdot Area(B_r(\mathbf{0}))\} \\
&= \exp\left\{-c\pi r^2\right\}
\end{aligned}$$

Which is the well known **Weibull** distribution.

Homogeneous Poisson random measure visibility Let the atoms of N have radius $a \approx 0$. We interpret the model as a forest with density $c = \mathbb{E}[K]$ and mean measure $\nu = Leb$. For simplicity, we ignore the overlapping trees. By construction, N is homogeneous, and the horizontal direction is as good as any by rotation invariance. We refer to the distance between the origin and the closest tree as a measure of **visibility**. An atom with radius a intersects $y = 0$ if and only if the distance between y and the center is $\leq a$. Then:

$$\{V \geq x\} = \{N(D_x) = 0\} \quad D_x = [0, x] \times [-a, a]$$

is the expression in terms of sets of the visibility being greater than x. We describe the r.v. in terms of its distribution as:

$$\begin{aligned}
\mathbb{P}[V \geq x] &= \mathbb{P}[N(D_x) = 0] && N(D_x) \sim \mathcal{P}o(\nu(D_x)) \\
&= e^{-\nu(D_x)} \\
&= \exp\{-cLeb([0, x] \times [-a, a])\} \\
&= \exp\{-c(2ax)\}
\end{aligned}$$

Shot Noise, Ornstein Uhlenbeck process The following is a descriptive discussion of a famous process, which will be generalized in Example 22.28.
($\triangle$ $\mathbb{R}$ case) We aim to describe a p.r.m. N on the real line $\mathbb{R}$ with mean $\nu(dx) = cdx$. The arrival times in this case are:

$$\ldots < T_{-2} < T_{-1} < T_0 < 0 < T - 1 < \ldots$$

This could model the arrivals to an anode of electrons producing a current intensity g decreasing as a function of the elapsed time $u \geq 0$. We assume for simplicity that currents are additive.

The total current at time t is then modelled as:

$$\begin{aligned}
X_t &= \sum_{n:T_n \le t} g(t-T_n) \\
&= \sum_{n=\infty}^{\infty} g(t-T_n)\mathbb{1}_{(-\infty,t]}(T_n) \\
&= \int g(t-x)\mathbb{1}_{(-\infty,t]}(x)\, N(dx) \\
&= Nf && f(x) \coloneqq g(t-x)\mathbb{1}_{[0,t]}(x)
\end{aligned}$$

We wish to describe the moments and the Laplace functional of this process on $\mathbb{R}$:

$$\begin{aligned}
\mathbb{E}[X_t] = \mathbb{E}[Nf] &= \nu f && \text{Prop. 13.18\#1} \\
&= \int_{\mathbb{R}} g(t-x)\mathbb{1}_{(-\infty,t]}(x)\, c dx \\
&= c\int_{-\infty}^{t} g(t-x)dx && \text{let } u=t-x, du=-dx \\
&= c\int_{\infty}^{0} -g(u)du \\
&= c\int_{0}^{\infty} g(u)du
\end{aligned}$$

Which is $\perp\!\!\!\perp t$ once we integrate. Moving on to the variance:

$$\begin{aligned}
V[X_t] = V[Nf] &= \nu f^2 && \text{Prop. 13.18\#2} \\
&= \int_{\mathbb{R}} \left(g(t-x)\mathbb{1}_{(-\infty,t]}(x)\right)^2 c dx \\
&= c\int_{-\infty}^{t} [g(t-x)]^2 dx && \text{let } u=t-x, du=-dx \\
&= c\int_{\infty}^{0} -[g(u)]^2 du \\
&= c\int_{0}^{\infty} [g(u)]^2 du
\end{aligned}$$

Again $\perp\!\!\!\perp t$. Lastly, the Laplace transform is:

$$\begin{aligned}
\mathbb{E}[e^{-rX_t}] = \mathbb{E}[e^{-rNf}] &= \exp\left\{-\nu(1-e^{-rf})\right\} && \text{Thm. 13.19} \\
&= \exp\left\{-\int_{\mathbb{R}} (1-e^{-rg(t-x)\mathbb{1}_{(\infty,t]}(x)}) c dx\right\} \\
&= \exp\left\{-c\int_{\mathbb{R}} (1-e^{-rg(t-x)})\mathbb{1}_{(-\infty,t]}(x)\, dx\right\} && \text{move indicator out} \\
&= \exp\left\{-\int_{-\infty}^{t} (1-e^{-rg(t-x)})dx\right\} && \text{let } u=t-x, du=-dx \\
&= \exp\left\{-c\int_{0}^{\infty} (1-e^{-rg(u)})du\right\} && (21.7)
\end{aligned}$$

Since the independence from t carries over to the Laplace functional, we can safely say that by Theorem 13.19 we have that:

$$X_t \overset{d}{=} \widetilde{X}_0 \;\forall t \quad \widetilde{X}_0 \; \textit{with transform as above}$$

(□ $(0,\infty)$ **case)** consider now a more realistic p.r.m. on $\mathbb{R}_+$, which would allow a researcher to simulate the phenomenon[1]. We consider as intensity function

$$g(u) = ae^{-bu} \quad a>0, b>0$$

[1] indeed, a physicist has to start somewhere, but the process in reality does not have a starting point itself.

and set a starting current to our X_t amount:

$$\begin{cases} X_t = e^{-bt}X_0 + \sum_{n=1}^{\infty} \underbrace{ae^{-b(t-T_n)}}_{=g(t-T_n)} \mathbb{1}_{[0,t]}(T_n) \\ X_0 \perp\!\!\!\perp T_1 < T_2 < \dots \end{cases}$$

In this context, we want to show that $X_t \overset{d}{\to} \widetilde{X}_0$ as before. Proceeding in the same way, we inspect moments and Laplace functional:

$$\begin{aligned} \mathbb{E}[X_t] &= \mathbb{E}[e^{-bt}X_0 + Nf] && f(x) := g(t-x)\mathbb{1}_{[0,t]}(x) \\ &= e^{-bt}\mathbb{E}[X_0] + \int_0^t g(t-x)c\,dx && \text{again by Prop. 13.18\#1} \\ &= e^{-bt}\mathbb{E}[X_0] + c\int_0^t g(u)du && \text{same ch. variable} \end{aligned}$$

which is **dependent** on t. For what concerns the variance

$$V[X_t] = \mathbb{E}\left[(e^{-bt}X_0 + Nf)^2\right] - \mathbb{E}\left[e^{-bt}X_0 + Nf\right]^2$$

the result is the same. Moving to the Laplace transform:

$$\begin{aligned} \mathbb{E}[e^{-rX_t}] &= \mathbb{E}\left[e^{-re^{-bt}X_0}\right]\mathbb{E}\left[e^{-rNf}\right] && \text{by } X_0 \perp\!\!\!\perp t \\ &= \mathbb{E}\left[e^{-re^{-bt}X_0}\right]\exp\left\{-c\int_0^t \left(1-e^{-rg(t-x)}\right)dx\right\} && \text{let } u = t-x, du = -dx \\ &= \underbrace{\mathbb{E}\left[e^{-re^{-bt}X_0}\right]}_{\to 1 \; as \; t\to\infty}\exp\left\{c\int_0^t \left(1-e^{-rg(u)}\right)du\right\} \\ &\overset{t\to\infty}{\to} \exp\left\{c\int_0^t \left(1-e^{-rae^{-bu}}\right)du\right\} \end{aligned}$$

In the limit, the laplace transform converges pointwise to that of $\widetilde{X}_0$ of Equation 21.7. Laplace pointwise convergence ensures that $X_t \overset{d}{\to} \widetilde{X}_0$ (Thm. 9.39).

(◯ **stationarity**) we aim to show $X_0 \overset{d}{=} \widetilde{X}_0 \implies X_t \overset{d}{=} \widetilde{X}_0 \; \forall t \in \mathbb{R}_+$. This would mean that the distribution is stationary around the realistic one over $\mathbb{R}$, but feasible for experimentation as argued in □. We start with a split:

$$\mathbb{E}[e^{-rX_t}] = \underbrace{\mathbb{E}[e^{-re^{-bt}\widetilde{X}_0}]}_{=\mathbf{A}} \underbrace{\exp\left\{-c\int_0^t e^{-rg(u)}du\right\}}_{=\mathbf{B}} \tag{$\star$}$$

A is the Laplace transform of $\widetilde{X}_0$ at $r' = re^{-bt} > 0$. Using Theorem 13.19 together with the explicit form in Equation 21.7 we get that:

$$\begin{aligned} \mathbf{A} &= \exp\left\{-c\int_0^\infty \left(1-e^{-re^{-bt}g(u)}\right)du\right\} \\ &= \exp\left\{-c\int_0^\infty \left(1-e^{-re^{-bt}ae^{-bu}}\right)du\right\} \\ &= \exp\left\{-c\int_0^\infty \left(1-e^{-rae^{-b(t+u)}}\right)du\right\} && \text{let } x = t+u, dx = du \\ &= \exp\left\{-c\int_t^\infty \left(1-e^{-rae^{-bx}}\right)dx\right\} && \text{for clearness, let } x = u \\ &= \exp\left\{-c\int_t^\infty \left(1-e^{-rae^{-bu}}\right)du\right\} \end{aligned}$$

So that $(\star)$ becomes:

$$\begin{aligned}\mathbb{E}[e^{-rX_t}] &= \exp\left\{-c\int_t^\infty \left(1-e^{-rae^{-bu}}\right)du\right\}\exp\left\{-c\int_0^t e^{-rg(u)}du\right\}\\ &= \exp\left\{-c\int_0^\infty \left(1-e^{-rg(u)}\right)du\right\} \perp\!\!\!\perp t\\ \implies X_t &\overset{d}{=} X_0 \overset{d}{=} \widetilde{X}_0 \quad \forall t\in\mathbb{R}_+\end{aligned}$$

($\diamondsuit$ stochastic differential equation) we want to show that such a process satisfies the SDE:

$$X_t = X_0 - b\int_0^t X_s ds + aN([0,t])$$

Also written as $dX_t = -bX_t dt + aN(dt)$. This is equivalent to:

$$\iff X_t = e^{-bt}X_0 + \int_0^t g(t-x)N(dx) = X_0 - b\int_0^t X_s ds + N([0,t])$$

Where the form we have is the LHS and the form we want is the RHS. Inspecting the integral in the RHS with the result of the LHS [2]:

$$\begin{aligned}\int_0^t X_s ds &= \int_0^t e^{-bs}X_0 ds + \int_0^t\int_0^s g(s-x)N(dx)ds && \text{where } s\le x\le t\\ &= \frac{1}{b}X_0(1-e^{-bt}) + \int_0^t\int_x^t g(s-x)dsN(dx) && \text{order change in accordance with } s\le x\le t\end{aligned}$$

Where the blue integral is precisely

$$\int_0^t\int_x^t g(s-x)dsN(dx) = \int_0^t -\frac{a}{b}e^{-b(s-x)}\Big|_{s=x}^t N(dx) = \int_0^t \frac{a}{b}\left(1-e^{-b(t-x)}\right)N(dx)$$

Eventually substituting in the RHS one gets:

$$\begin{aligned}X_0 - b\int_0^t X_s ds + aN([0,t]) &= X_0 + e^{-bt}X_0 - X_0 - a\int_0^t N(dx) + a\int_0^t e^{-b(t-x)}N(dx)\\ &= e^{-bt}X_0 + \int_0^t g(t-x)N(dx)\end{aligned}$$

Which is the LHS.

[2] this is slightly informal to say

Chapter 22

Itô Integration

The content of this Section is based on the last two lectures, not included in the exam, and presented for the sake of completeness only. For this reason, it might be less informal, especially at the end. A reference is [BZ99].

Disclaimer: due to changes in the layout of the digital document it may refer to the wrong statements when cross referencing other Chapters.

◇ **Observation 22.1** (Setting). *We want to integrate wrt a Wiener process $(W_t)_{t\in\mathbb{R}_+}$ adapted to a filtration $\mathcal{F}$ (Defs. 11.2, 11.7). Just like in the didactic derivation of the Riemann sum we will proceed step by step. The ingredients of the first part are:*

- *trying to characterize an integral in terms of $\mathcal{L}^2$ convergent sequences exploiting the result of Theorem 20.4*
- *circumavigating the difficulty of W being nowhere differentiable, a result of Proposition 20.6*

♡ **Example 22.2** (Differences with Riemann integral). *Recall the classic definition of Riemann sum for a function $f : \mathbb{R}_+ \to \mathbb{R}$:*

$$\sum_{j=0}^{n-1} f(s_j)(t_{j+1} - t_j)$$

*where n is the number of sub-intervals and a **very useful property** is that the choice of s_j does not influence the final result.*
Contrarily, a stochastic integral in terms of a Wiener process of the form:

$$\sum_{j=0}^{n-1} f(s_j)(W_{t_{j+1}} - W_{t_j})$$

behaves differently. We will show that the choice of $s_j \in [t_j, t_{j+1}]$ for a subdivision $t_j^n = (\frac{jT}{n})_j$ partitioning the interval $[0, T]$ is dependent on the choice. To do so, we choose as test function $f(t) = W_t$ and we assign the symbol $\triangle$ to the choice $s_j = t_j$ and $\square$ to the choice $s_j = t_{j+1}$. It suffices to show that the two are different.
***($\triangle$ case)** we will make use of the trivial identity:*

$$a(b - a) = \frac{b^2 - a^2}{2} - \frac{(b - a)^2}{2}$$

Computations show that:

$$
\begin{aligned}
\sum_{j=0}^{n-1} f(s_j)(W_{t_{j+1}} - W_{t_j}) &= W_{t_j}(W_{t_{j+1}} - W_{t_j}) \\
&= \frac{1}{2}\sum_{j=0}^{n-1} W_{t_{j+1}}^2 - W_{t_j}^2 - \frac{1}{2}\sum_{j=0}^{n-1}(W_{t_{j+1}} - W_{t_j})^2 && \text{trivial equality} \\
&= \frac{1}{2}(W_T^2 - W_0^2) - \frac{1}{2}\sum_{j=0}^{n-1}(W_{t_{j+1}} - W_{t_j})^2 && \text{telescopic sum} \\
&\overset{n\to\infty}{\to} \frac{1}{2}W_T^2 - \frac{1}{2}T && W_0 = 0 \text{ a.s. by Def. and Thm. 20.4}
\end{aligned}
$$

($\square$ case) *similarly, we use another trivial identity*

$$
b(b-a) = \frac{b^2 - a^2}{2} + \frac{(b-a)^2}{2}
$$

and get:

$$
\begin{aligned}
\sum_{j=0}^{n-1} f(s_j)(W_{t_{j+1}} - W_{t_j}) &= W_{t_{j+1}}(W_{t_{j+1}} - W_{t_j}) \\
&= \frac{1}{2}\sum_{j=0}^{n-1} W_{t_{j+1}}^2 - W_{t_j}^2 + \frac{1}{2}\sum_{j=0}^{n-1}(W_{t_{j+1}} - W_{t_j})^2 && \text{trivial equality} \\
&= \frac{1}{2}(W_T^2 - W_0^2) + \frac{1}{2}\sum_{j=0}^{n-1}(W_{t_{j+1}} - W_{t_j})^2 && \text{telescopic sum} \\
&\overset{n\to\infty}{\to} \frac{1}{2}W_T^2 + \frac{1}{2}T && W_0 = 0 \text{ a.s. by Def. and Thm. 20.4}
\end{aligned}
$$

For $T > 0$ we have $\triangle \neq \square$ and the Stochastic integral depends on the choice of the point for each subinterval.

◇ **Observation 22.3** (Intuition)**.** *By Theorem 20.4 we have that the quadratic variation for over an interval $[a,b]$ is such that $V_n \overset{\mathcal{L}^2}{\to} b - a$. This means that $\mathbb{E}[V_n] < \infty$ and so $V_n < \infty$ a.s.*
For this reason, we might as well use $s_j = t_j$ in a subdivision (t_j^n) by the adaptedness of $f(t_j) \in \mathcal{F}_{t_j}$ and work with square integrable r.v.s.

22.1 Constructive Definition of the Itô Integral

♠ **Definition 22.4** (Random step processes M^2_{step})**.** *M^2_{step} is a collection of functions of the form $f : \mathbb{R}_+ \to E$ such that:*

$$
\exists 0 < t_0 < t_1 < \ldots < t_n \ : \ f(t) = \sum_{j=0}^{n-1} \eta_j \mathbb{1}_{t_j, t_{j+1}}(t)
$$

with:

- *(measurability)* $\eta_j \in \mathcal{F}_{t_j} \forall j$
- *(square integrability)* $\eta_j \in \mathcal{L}^2 \forall j$

By construction, we get for free that $f(t) \in \mathcal{F}_t \forall t$ and $f \in \mathcal{L}^2$.

♠ **Definition 22.5** (Stochastic integral of step process)**.** *For $f \in M^2_{step}$ its integral is:*

$$
I(f) = \int_{\mathbb{R}_+} f(t) dW_t = \sum_{j=0}^{n-1} \eta_j (W_{t_{j+1}} - W_{t_j})
$$

♣ **Proposition 22.6** (Isometry of step process integral). *The integral of a step process is square integrable,* $f \in M^2_{step} \implies I(f) \in \mathcal{L}^2$ *and:*

$$\mathbb{E}[|I(f)|^2] = \mathbb{E}\left[\int_0^\infty |f(t)|^2 dt\right]$$

♠ **Definition 22.7** (Random process class M^2). *Making use of* M^2_{step} *we define the broader class of random step approximable functions as:*

$$M^2 := \left\{ f : \mathbb{R}_+ \to E \ : \ f \in \mathcal{L}^2, \exists (f_n) \subset M^2_{step} \ : \ \lim_{n\to\infty} \mathbb{E}\left[\int_0^\infty |f(t) - f_n(t)|^2 dt\right] = 0 \right\}$$

♠ **Definition 22.8** (Itô stochastic integral). *For* $f \in M^2$ *the integral* $\int_0^\infty f(t)dW_t$ *is defined via:*

$$\lim_{n\to\infty} \mathbb{E}\left[|I(f) - I(f_n)|^2\right] = 0 \quad (f_n) \subset M^2_{step}$$

♣ **Proposition 22.9** (Itô isometry of stochastic integral). *For* $f \in M^2 \exists! I(f) \in \mathcal{L}^2$ *a.s. (the integral is square integrable and uniquely defined a.s.), with form:*

$$\mathbb{E}[|I(f)|^2] = \mathbb{E}\left[\int_0^\infty |f(t)|^2 dt\right]$$

♠ **Definition 22.10** (Extension to finite times). $\forall T > 0$ *we define*

$$M^2_T := \{ f : \mathbb{R}_+ \to E \ : \ \mathbb{1}_{[0,T)} f \in M^2 \}$$

with integral

$$I_T(f) = I(\mathbb{1}_{[0,T)} f) = \int_0^T f(t) dW_t$$

♢ **Observation 22.11** (What is missing and specifications). *we used the linearity of the integral* I *in* $I(f_n) - I(f_m) = I(f_n - f_m)$ *in the proof of Proposition 22.9. This is rather easy to prove for* $f_n, f_m \in M^2_{step}$ *but holds also for* $f \in M^2_T$*, we will do this in Theorem 22.15#1.*
*Additionally, we are missing **existance conditions**, a result shown in the next Theorem.*

22.2 Properties of the Itô Integral

♣ **Theorem 22.12** (Itô integral existance). *Let* $f : \mathbb{R}_+ \to E$ *be such that:*

- *(continuity)* $t \to f(t)$ *is a.s. continuous*
- *(adaptedness)* $f(t) \in \mathcal{F}_t \forall t$

Then:

1. *the Itô integral exists*
$$\mathbb{E}\left[\int_0^\infty |f(t)|^2 dt\right] < \infty \implies f \in M^2 \text{ i.e. } \exists I(f)$$
2. *the Itô integral exists at finite times*
$$\mathbb{E}\left[\int_0^T |f(t)|^2 dt\right] < \infty \implies f \in M^2_T \text{ i.e. } \exists I_T(f)$$

♡ **Example 22.13** (Back to Example 22.2). *Recall that for a Wiener process* W*:*

- *the map* $t \to W_t$ *is continuous a.s. (Thm. 18.22)*
- *the integrability condition holds*
$$\mathbb{E}[\int_0^t |W_s|^2 ds] < \infty \quad \forall t$$

Then, by Theorem 22.12 we can use a sequence $(f_n) \subset M^2_{step}$ of the form:

$$f_n(t) = \sum_{k=0}^{n-1} W_{t_k} \mathbb{1}_{[t_k, t_{k+1}]}(t)$$

With (t^n_k) made of equally spaced intervals and make an approximation of the form:

$$\int f_n(s)ds \underset{n\to\infty}{\overset{\mathcal{L}^2}{\to}} \frac{1}{2}W_T^2 - \frac{1}{2}T$$

A second example could be setting $f(t) = W_t^2$. In this case the integrability condition is:

$$\mathbb{E}[\int_0^\infty |W_s^2|^2 ds] < \infty$$

which is easily checked. To ease out the computation, we use the trivial equality:

$$a^2(b-a) = \frac{b^3 - a^3}{3} - a(b-a)^2 - \frac{1}{3}(b-a)^3$$

To get as an approximation:

$$\begin{aligned}
\sum_{j=0}^{n-1} W_{t_j}^2 (W_{t_{j+1}} - W_{t_j}) &= \sum_{j=0}^{n-1} \frac{W_{t_{j+1}}^3 - W_{t_j}^3}{3} - \sum_{j=0}^{n-1} W_{t_j}(W_{t_{j+1}} - W_{t_j})^2 - \frac{1}{3}\sum_{j=0}^{n-1}(W_{t_{j+1}} - W_{t_j})^3 \\
&= \frac{1}{3}(W_{t_n}^3 - W_{t_0}^3) - \sum_{j=0}^{n-1} W_{t_j}(t_{j+1} - t_j) - \frac{1}{3}\sum_{j=0}^{n-1}(W_{t_{j+1}} - W_{t_j})^3 \\
&\quad [\textit{telescopic sum \& Thm. 20.4}] \\
&= \frac{1}{3}W_{t_n}^3 - \sum_{j=0}^{n-1} W_{t_j}(t_{j+1} - t_j) - \frac{1}{3}\sum_{j=0}^{n-1}(t_{j+1} - t_j)^{\frac{3}{2}} Z_k^3 \\
&\quad [W_{t_0} = 0 \textit{ and scaling Thm. 18.16\#1}] \\
&= \frac{1}{3}W_T^3 - \sum_{j=0}^{n-1} W_{t_j}(t_{j+1} - t_j) \qquad T = t_n
\end{aligned}$$

Where in the last passage we also use $\mathbb{E}[Z^3] = 0$, i.e. null kurtosis of a normal distribution. Eventually in the $n \to \infty$ limit the result becomes:

$$\int_0^T W_s^2 dW_s = \frac{1}{3}W_T^3 - \int_0^T W_s ds$$

Where the second term is a Riemann integral, which can be computed with classical methods.

◇ **Observation 22.14** (The example in the context of stochastic differential equations: Itô correction)**.** *Stochastic differential equations (SDEs) are an interesting topic in which Itô integrals are instrumental. It is worth noticing that we have just established:*

$$\int_0^t W_s dW_s = \frac{1}{2}W_t^2 - \frac{1}{2}t \quad \forall t \implies W_t^2 = 2\int_0^t W_s dW_s + t, \quad dW_t^2 = 2W_t dW_t + dt$$

which looks like the usual differential equation:

$$X^2(t) = 2\int_0^t X(s)dX(s) \quad dX^2(t) = 2X(t)dX(t)$$

*except for an added red term. We call this **Itô correction**.*
Similarly for the second part of the example:

$$\int_0^t W_s^2 dW_s = \frac{1}{3}W_t^3 - \int_0^t W_s ds \quad \forall t \implies W_t^3 = 3\int_0^t W_s^2 dW_s + 3\int_0^t W_s ds, \quad dW_t^3 = 3W_t^2 dW_t + 3W_t dt$$

as well corrected wrt the classic integral formulation.

♣ **Theorem 22.15** (Properties of Itô integral). *Let $f, g \in M_t^2$ and $\alpha, \beta \in \mathbb{R}$. Then $\forall 0 \leq s < t$ it holds that:*

1. *(linearity)* $\int_0^t (\alpha f(r) + \beta g(r)) dW_r = \alpha \int_0^t f(r) dW_r + \beta \int_0^t g(r) dW_r$
2. *(general isometry)* $\mathbb{E}\left[| \int_0^t f(r) dW_r|^2 \right] = \mathbb{E}\left[\int_0^t |f(r)|^2 dr \right]$
3. *(martingale property)* $\mathbb{E}_s \left[\int_0^t f(r) dW_r \right] = \int_0^s f(r) dW_r$

Lemma 22.16 (Square integrable $\mathcal{L}^2$ convergence in conditional expectation). *Let $(\xi_n), \xi \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ then:*

$$\xi_n \xrightarrow{\mathcal{L}^2} \xi \implies \mathbb{E}[\xi_n | \mathcal{G}] \xrightarrow{\mathcal{L}^2} \mathbb{E}[\xi | \mathcal{G}] \quad \forall \mathcal{G} \subset \mathcal{F}$$

♠ **Definition 22.17** (Versions of stochastic process). *For two stochastic processes $\xi(t), \zeta(t)$ on $t \in \mathbb{T} \subset \mathbb{R}$ seen as functions, we say that they are a version of each other when:*

$$\mathbb{P}[\{\xi(t) = \zeta(t)\}] = 1 \quad \forall t \in \mathbb{T}$$

Namely, they are a.s. equal.

♣ **Theorem 22.18** (Almost sure continuous version of stochastically integrable processes). *Let $f \in M_t^2$, then the integral process:*

$$\xi(t) = \int_0^t f(s) dW_s$$

is such that we can always find an almost continuous version:

$$\exists \zeta(t) \overset{a.s.}{=} \xi(t) \quad t \to \zeta(t) \quad a.s. \text{ continuous}$$

Corollary 22.19 (Direct consequence of almost continuous version existance). *We can safely say that the modification of $\xi(t)$ is in M_2^T*

Assumption 22.20 (Almost sure continuity). *From now on, we work with the always existing (Thm. 22.18) almost continuous version of the stochastic integral process we consider.*

♠ **Definition 22.21** (Itô process). *A process $\xi(t)$ on $\mathbb{R}_+$ is an Itô process when:*

1. *the path $t \to \xi(t)$ is almost surely continuous*
2. $\xi(T) = \xi(0) + \int_0^T a(t) dt + \int_0^T b(t) dW_t \quad a.s.$
3. $b(t) \in M_T^2 \; \forall T > 0$
4. *$a(t)$ is adapeted i.e. $a(t) \in \mathcal{F}_t \; \forall$ and such that* $\int_0^T |a(t)| dt < \infty \quad a.s. \; \forall T > 0$

We write this result in compact differential form as:

$$d\xi(t) = a(t) dt + b(t) dW_t$$

♡ **Example 22.22** (Observation 22.14 as an Itô process). *Recognize that:*

$$dW_t^2 = 2W_t dW_t + dt \quad a(t) \equiv 1, \; b(t) = 2W_t : W_t \in M_T^2 \forall T > 0$$

and also

$$dW_t^3 = 3W_t^2 dW_t + W_t dt \quad a(t) = W_t : \mathbb{E}\left[\int_0^t |W_s| ds \right] < \infty, \; b(t) = 3W_t^2 \in M_T^2 \forall T > 0$$

♠ **Definition 22.23** ($\mathcal{L}_T^1$ space for the function $a(t)$). *The function $a(\cdot)$ from Definition 22.21 is enclosed in a space denoted as:*

$$\mathcal{L}_T^1 := \left\{ a(t) \text{ adapted } \mathcal{F}, \; \int_0^T |a(t)| dt < \infty \text{ a.s. } \forall T > 0 \right\}$$

Where the second requirement is equivalent to having finite expectation a.s.

♣ **Theorem 22.24** (Itô formula, simplified). *For $F(t, x)$ a real valued function such that:*

$$F_t(t, x) = \frac{\partial F(t, x)}{\partial t}, \; F_x(t, x), F_{xx}(t, x) \in C(\mathbb{R}) \quad \forall t \geq 0, \forall x \in \mathbb{R}$$

where additionally $F_x(t, W_t) \in M_T^2 \forall T > 0$ is such that:

$$F(T, W_T) - F(0, W_0) = \int_0^T F_t(t, W_t) + \frac{1}{2} F_{xx}(t, W_t)\, dt + \int_0^T F_x(t, W_t)\, dW_t \quad a.s. \qquad \textit{integral form}$$

$$dF(t, W_t) = \left(F_t(t, W_t) + \frac{1}{2} F_{xx}(t, W_t) \right) dt + F_x(t, W_t) dW_t \quad a.s. \qquad \textit{differential form}$$

♡ **Example 22.25** (Itô formula for Example 22.22)**.** *Let $\xi_t = F(t, W_t) = W_t^2$. Using Theorem 22.24, the differential form is:*

$$\begin{aligned} d\xi_t &= \left(\frac{\partial W_t^2}{\partial t} + \frac{1}{2} \frac{\partial^2 W_t^2}{\partial W_t^2} \right) dt + \frac{\partial W_t^2}{\partial W_t} dW_t \\ &= \left(0 + \frac{1}{2} 2 \right) dW_t + 2W_t dW_t \\ &= dt + 2W_t dW_t \end{aligned}$$

As in the form of Example 22.22. Similarly for $\xi_t = F(t, W_t) = W_t^3$:

$$\begin{aligned} d\xi_t &= \left(\frac{\partial W_t^3}{\partial t} + \frac{1}{2} \frac{\partial^2 W_t^3}{\partial W_t^2} \right) dt + \frac{\partial W_t^3}{\partial W_t} dW_t \\ &= \left(0 + \frac{1}{2} 6W_t \right) dt + 3W_t^2 dW_t \\ &= 3W_t dt + 3W_t^2 dW_t \end{aligned}$$

again as in Example 22.22.

♡ **Example 22.26** (Brownian bridge)**.** *Consider a Wiener process $W = (W_t)_{t \in \mathbb{R}_+}$ and a second process $X_t = W_t - tW_1$ for $t \in [0, 1]$. This construction is such that:*

$$\begin{cases} X_t = W_0 = 0 \; a.s. & t = 0 \\ X_t = W_1 - W_1 = 0 \; a.s. & t = 1 \end{cases}$$

We explore its properties.
*(△ **mean**) we have $\mathbb{E}[X_t] = 0$ being a sum of Wiener processes.*
*(◯ **distribution**) X_t is Gaussian $\forall t$ since it is a sum of Gaussians*
*(□ **variance**) after some computations:*

$$\begin{aligned} V[X - t] &= V[W_t - tW_1] \\ &= V[W_t] + t^2 V[W_1] - 2tCoV[W_1, W1] \\ &= t + t^2 \cdot 1 - 2tCoV[W_t, W_1] && \textit{Wiener-Gauss Thm. 18.11} \\ &= t + t^2 - 2t(t \wedge 1) && \textit{Wiener as Brownian Def. 18.7} \\ &= t + t^2 - 2t^2 && t < 1 \\ &= t - t^2 \end{aligned}$$

*(▽ **covariance**) for $s \neq t$:*

$$\begin{aligned} CoV[X_s, X_t] &= CoV(W_s - sW_1, W_t - tW_1) \\ &= CoV(W_s, W_t) - sCoV(W_1, W_t) - tCoV(W_s, W_1) + stCoV(W_1, W_1) \\ &= (s \wedge t) - t(s \wedge 1) - s(t \wedge 1) + st \cdot 1 \\ &[\textit{as above Thm. 18.11, Def. 18.7}] \\ &= s \wedge t - st - stm + st && s, t < 1 \\ &= s \wedge t - st \\ &= \begin{cases} s(1-t) \; s < t \\ t(1-s) \; s \geq t \end{cases} \end{aligned}$$

(♢ *conditionals*) *wts for* $t_1 < t_2$ *it holds:*

$$\mathbb{P}[W_{t_1} \in dx_2, W_{t_2} \in dx_2 | W_1 = 0] = \mathbb{P}[X_{t_1} \in dx_1, X_{t_2} \in dx_2]$$

namely, the two points in the path of X *are equivalent to two the same times of a Wiener process conditioned on the future* $W_1 = 0$*. This is rather difficult to work with, so we will use another property, that of time inversion, Theorem 18.16#3 and the Markov property Thm. 19.5. Namely:*

$$tW_{\frac{1}{t}} \overset{d}{=} W_t \qquad \begin{cases} W_t | W_s = x \text{ Brownian} & s < t \\ W_t - W_s \overset{d}{=} W_{t-s} \end{cases}$$

While we known the LHS, we do not known how the RHS behaves. In this context, for $s \geq t$*:*

$$\begin{aligned} \mathbb{P}[W_t \leq y | W_s = x] &= \mathbb{P}[tW_{\frac{1}{t}} \leq y | sW_{\frac{1}{s}} = x] \\ &= \mathbb{P}\left[W_{\frac{1}{t}} \leq \frac{y}{t} \middle| W_{\frac{1}{s}} = \frac{x}{s}\right] \\ &= \mathbb{P}\left[W_{\frac{1}{t}} + \frac{x}{s} - \frac{x}{s} \leq \frac{y}{t} \middle| W_{\frac{1}{s}} = \frac{x}{s}\right] && \text{notice } \frac{x}{s} = W_{\frac{1}{s}} \text{ and } \pm\frac{x}{s} \\ &= \mathbb{P}\left[W_{\frac{1}{t}} - W_{\frac{1}{s}} \leq \frac{y}{t} - \frac{x}{s} \middle| W_{\frac{1}{s}} = \frac{x}{s}\right] \\ &= \mathbb{P}\left[t(W_{\frac{1}{t}} - W_{\frac{1}{s}}) + \frac{xt}{s} \leq y\right] && \text{Markov property} \end{aligned}$$

Where $W_t | W_s \overset{d}{=} t(W_{\frac{1}{t}} - W_{\frac{1}{s}}) + \frac{tx}{s} \overset{d}{=} t(W_{\frac{1}{t}-\frac{1}{s}}) + \frac{tx}{s}$ *so that:*

$$\implies W_t | W_1 = 0 \overset{d}{=} tW_{\frac{1}{t}} - 1 + 0 \quad \text{setting } s = 0$$

Which has mean 0 *and variance* $V[tW_{\frac{1}{t}} - 1] = t^2 \left(\frac{1}{t} - 1\right) = t(1-t)$*. The joint distribution then becomes:*

$$W_s, W_t | W_1 = 0 \overset{d}{=} \left(sW_{\frac{1}{s}-1}, tW_{\frac{1}{t}-1}\right)$$

with covariance:

$$\begin{aligned} CoV[W_s, W_t] &= CoV\left(sW_{\frac{1}{s}-1}, tW_{\frac{1}{t}-1}\right) \\ &= st\left[\left(\frac{1}{s} - 1\right) \wedge \left(\frac{1}{t} - 1\right)\right] \\ &= st\left[\left(\frac{1}{s} \wedge \frac{1}{t}\right) - 1\right] \\ &= st\left(\frac{1}{s} \wedge \frac{1}{t}\right) - st \\ &= \begin{cases} s - st & t \geq s \\ t - st & t < s \end{cases} \end{aligned}$$

Just like ▽*, proving the claim.*

♡ **Example 22.27** (Exponential martingale)**.** *For* $t \geq 0$ *consider the process* $X_t = e^{W_t} e^{-\frac{t}{2}}$*.*
(△ *aim*) *wts* X_t *is an Itô process according to Definition 22.21, namely, an Itô stochastic integral (Def. 22.8. The form should be that of:*

$$dX_t = X_t dW_t, \quad X_t = X_0 + \int_0^t X_s dW_s, \quad X_0 = 1$$

(□ *recap of idea*) *An Itô process* ξ_t *is such that:*

$$d\xi_t = a(t)dt + b(t)dW_t, \quad \xi_t = \xi(0) + \int_0^T a(s)ds + \int_0^t b(s)dW_s, \quad \begin{cases} \mathbb{E}[\int_0^t |a(s)|ds] < \infty \; \forall t & \text{i.e. } a \in \mathcal{L}_t^1 \\ \mathbb{E}[\int_0^t |b(s)|^2 ds] < \infty \; \forall t & \text{i.e. } b \in M_t^2 \end{cases}$$

We use the simplified Itô formula (Thm. 22.24) setting $\xi_t = F(t,x) = e^x e^{-\frac{t}{2}}$ which is differentiable:

$$\begin{aligned} dF(t,W_t) &= \left(\frac{\partial}{\partial t}F + \frac{\partial^2}{\partial W_t^2}F\right)dt + \left(\frac{\partial}{\partial W_t}F\right)dW_t \\ &= \left(-\frac{1}{2}e^{W_t}e^{-\frac{t}{2}} + \frac{1}{2}e^{W_t}e^{-\frac{t}{2}}\right)dt + \left(e^{W_t}e^{-\frac{t}{2}}\right)dW_t \\ &= \left(e^{W_t}e^{-\frac{t}{2}}\right)dW_t \\ &= \xi_t dW_t \end{aligned}$$

*(○ **conditions check**) we want to check the conditions on the functions a and b. Clearly, $a(t) \equiv 0$ is in $\mathcal{L}_t^1 \forall t \in \mathbb{R}_+$. For $b(t) = X_t$ instead:*

$$\begin{aligned} \mathbb{E}\left[\int_0^t |X(s)|^2\right] ds &= \int_0^t \mathbb{E}\left[|X(s)|^2 ds\right] && \textit{Fubini Thm. B.30 provided that it exists} \\ &= \int_0^t \mathbb{E}\left[e^{2W_s}e^{-s}\right] ds \\ &= \int_0^t e^{-s}\mathbb{E}\left[e^{2W_s}\right] ds \\ &= \int_0^t e^{-s}\exp\left\{\frac{1}{2}2V[W_s]\right\} ds && \textit{mgf} \\ &= \int_0^t e^{-s}e^{s} ds && \textit{Wiener variance} \\ &= \int_0^t 1 < \infty \ \forall t \in \mathbb{R}_+ \end{aligned}$$

By the arguments of $\square, \bigcirc \implies \triangle$ and the process is an Itô process.

♡ **Example 22.28** (Continuous Ornstein-Uhlenbeck process)**.** *We inspect the traditional version in continuous time of the process from Example 16.9.*
*(△ **symmetris and definition**) the process analyzed is:*

$$X_t = X_0 e^{-bt} + \int_0^t a e^{-b(t-s)} dW_s, \quad t \in \mathbb{R}_+$$

Focusing on the integral, in Example 16.9 it was the integral of a p.r.m. on $\mathbb{R}$, shown to be equivalent to the experimentally feasible case of a p.r.m. on $\mathbb{R}_+$ subject to a specific value of the starting point. We then showed that the process satisfied a SDE of the form:

$$dX_t = -bX_t + a dW_t, \quad X_t = X_0 - b\int_0^t X_t dt + aW_t$$

Where a was the number of arrivals at time t.
*(□ **gaussianity of integral**) wts:*

$$\int_0^t a e^{-b(t-s)} dW_s \overset{d}{=} a e^{-bt} W_{\frac{e^{-2bt}-1}{2b}} \iff \int_0^t e^{bs} dW_s \overset{d}{=} W_{\frac{e^{-2bt}-1}{2b}} \sim \mathcal{N}\left(0, \frac{e^{2bt}-1}{2b}\right)$$

*(○ **two facts**) we take for granted without proof that:*

- *any integral of the form $\int_0^t f(s)dW_s$ is a martingale by Theorem 22.15#3, and thus has constant mean*
- *$\int_0^1 f(s)dW_s$ is such that $f(s) \in \mathcal{L}_1$ is Gaussian itself.*

For this reason, to prove the claim in □ we just need to inspect the variance.
*(▽ **variance**) continuing the discussion:*

$$\begin{aligned} V\left[\int_0^t e^{bs} dW_s\right] &= \mathbb{E}\left[\left(\int_0^t e^{bs} dW_s\right)^2\right] && \textit{mean zero martingale} \\ &= \mathbb{E}\left[\int_0^t (e^{bs})^2 ds\right] && \textit{Itô isometry Thm. 22.15\#2} \\ &= \int_0^t e^{2bs} ds && \textit{deterministic value} \\ &= \left.\frac{e^{2bs}}{2b}\right|_0^t \\ &= \frac{e^{2bt}-1}{2b} \end{aligned}$$

which is the variance needed to prove the claim in □.
*(◊ **limiting distribution**) wts*

$$X_t \xrightarrow{d} \sqrt{\frac{a^2}{2b}} Z \sim \mathcal{N}\left(0, \frac{a^2}{2b}\right)$$

Where

$$X_t = \underbrace{X_0 e^{-bt}}_{\to 0 \text{ as } t\to\infty} + a e^{-bt} W_{\frac{e^{2bt}-1}{2b}} \to a e^{-bt} W_{\frac{e^{2bt}-1}{2b}} = a e^{-bt} W_{\eta_t - \eta_0}$$

Where we let $\eta_t = \frac{e^{2bt}}{2b}$ *which* $\to \infty$ *as* $t \to \infty$*. At the limit, the second term is:*

$$\begin{aligned} a e^{-bt} W_{\eta_t-\eta_0} &= \frac{a}{\sqrt{2b}} \frac{\sqrt{2b}}{e^{bt}} W_{\eta_t-\eta_0} \\ &= \frac{a}{\sqrt{2b}} \sqrt{\frac{1}{\frac{e^{2bt}}{2b}}} W_{\eta_t-\eta_0} \\ &= \frac{a}{\sqrt{2b}} \frac{1}{(\eta_t)^{\frac{1}{2}}} W_{\eta_t-\eta_0} \end{aligned}$$

Where the Mgf is that of a gaussian, since ignoring the coefficient in front:

$$\begin{aligned} \mathbb{E}\left[\exp\left\{r \frac{1}{(\eta_t)^{\frac{1}{2}}} W_{\eta_t-\eta_0}\right\}\right] &= \exp\left\{\frac{1}{2}\frac{r^2}{\eta_t}[\eta_t - \eta_0]\right\} && \eta_t - \eta_0 \textit{ is the variance} \\ &= \exp\left\{\frac{r^2}{2}\frac{\eta_t - \eta_0}{\eta_t}\right\} \\ &\xrightarrow{t\to\infty} \exp\left\{\frac{r^2}{2}\right\} && \textit{mgf of } Z \sim \mathcal{N}(0,1) \end{aligned}$$

Concluding the claim that:

$$\begin{aligned} X_t &= X_0 e^{-bt} + a e^{-bt} W_{\frac{e^{2bt}-1}{2b}} \\ &\xrightarrow{d} o(1) + \frac{a}{\sqrt{2b}} Z \sim \mathcal{N}\left(0, \frac{a^2}{2b}\right) \end{aligned}$$

*(⋆ **stationarity**) assume* $X_0 \sim \mathcal{N}\left(0, \frac{a^2}{2b}\right)$*. Then the process is such that:*

- $\mathbb{E}[X_t] = 0 \forall t$ *easily by sum of Gaussians*
- X_t *is Gaussian by sum of Gaussians*
- $V[X_t] = V[X_0 e^{-bt} + a e^{-bt} W_{\eta_t-\eta_0}] = e^{-2bt}\frac{a^2}{2b} + a^2 e^{-2bt}\frac{e^{2bt}-1}{2b} = \frac{a^2}{2b}$

Which is the same variance of X_0. This means that the process is stationary at $\mathcal{N}\left(0, \frac{a^2}{2b}\right)$.

(♠ SDE and general Itô formula) *wts:*

$$dX_t = -bX_t dt + adW_t, \quad X_t = X_0 - b\int_0^t X_s ds + aW_t$$

to do so, we will resort to a constructive derivation of the general Itô formula.

$$\begin{aligned} X_t &= X_0 e^{-bt} + ae^{-bt}\int_0^t e^{bs} dW_s \\ &= e^{-bt}\left(X_0 + a\int_0^t e^{bs} dW_s\right) \\ &= e^{-bt}\xi_t \\ \text{with} \quad \xi_t &= X_0 + \int_0^t ae^{bs} dW_s \\ d\xi_t &= ae^{bt} dW_t = a(t)dt + b(t)dW_t & a(t) \equiv 0,\ b(t) = ae^{bt} \end{aligned}$$

Here we set $X_t = F(t, \xi_t)$ so that $F(t, x) = e^{-bt}x$ in general. In this context, we aim to do a 2^{nd} order Taylor expansion, and corrected according with a procedure similar to the result of Obs. 22.14. A normal expansion of:

$$dF(t,x) = \frac{\partial F}{\partial t}dt + \frac{\partial F}{\partial x}dx + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}dx^2 + \frac{1}{2}\frac{\partial^F}{\partial t^2}dt^2 + \frac{1}{2}\frac{\partial^2 F}{\partial x \partial t}dxdt + \frac{1}{2}\frac{\partial^2 F}{\partial t \partial x}dtdx$$

in the context of SDEs becomes:

$$\begin{aligned} dF(t,\xi_t) &= \left(\frac{\partial F}{\partial t} + \frac{\partial F}{\partial \xi_t}a(t) + \frac{1}{2}\frac{\partial^2 F}{\partial \xi_t^2}b(t)^2\right)dt + \frac{\partial F}{\partial \xi_t}b(t)dW_t \\ &= \left(-be^{-bt}\xi_t + 0 + 0\right)dt + e^{-bt}b(t)dW_t \\ &= -bX_t dt + e^{-bt}ae^{bt}dW_t & b(t) = ae^{-bt} \\ &= -bX_t dt + adW_t \end{aligned}$$

Which is equivalent to the claim. We have also "formally" proved the SDE of Example 16.9.

While some results in previous sections were not part of the course, the content of the Appendix is 100% not study material. I decided to add it to better understand arguments presented in class, and make the notes *almost* self contained. For this reason, I most likely will not copy over the handwritten proofs, typing only the claims on LaTeX.

Appendix A

Sets, Measures, Probability

A.1 The p-system extension

♠ **Definition A.1** (d-system). *A collection $\mathcal{D} = \{A_1, \ldots\}$ where $A_i \subset E\ \forall i$ is a d-system when:*

1. $E \in \mathcal{D}$
2. $A, B \in \mathcal{D}, B \subset A \implies A \setminus B \in \mathcal{D}$
3. *(closedness for increasing sequences)* $(A_n) \subset \mathcal{D}, A_n \nearrow A \implies A = \bigcup A_n \in \mathcal{D}$

♣ **Proposition A.2** (σ-algebras, p&d systems equivalence). *Recalling Definitions 1.6 and 1.8:*

$$\{A_1, \ldots\} : A_i \subset E\ \forall i \quad \sigma\text{-algebra} \iff p, d\text{-system on } E$$

Lemma A.3 (Nested d-system form). *Let $\mathcal{D}$ be a d-system on E. Then, for a given $D \in \mathcal{D}$:*

$$\widehat{\mathcal{D}} = \{A \in \mathcal{D} : A \cap D \in \mathcal{D}\} \quad \text{is a d-system}$$

♣ **Theorem A.4** (Monotone Class Theorem I). *For $\mathcal{C}$ a p-system (Def. 1.8) and $\mathcal{D}$ a d-system:*

$$\mathcal{C} \subset \mathcal{D} \implies \sigma(\mathcal{C}) \subset \mathcal{D}$$

♠ **Definition A.5** (Product of measurable spaces). *Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces (Def. 2.1). Consider $A \subset E$, $B \subset F$.*

- $A \in \mathcal{E}, B \in \mathcal{F} \implies A \times B$ *is a measurable rectangle*
- $\mathcal{E} \otimes \mathcal{F}$ *is the product σ-algebra on $E \times F$*

Accordingly, we define the product measurable space as:

$$(E \times F, \mathcal{E} \otimes \mathcal{F}) = (E, \mathcal{E}) \times (F, \mathcal{F})$$

Lemma A.6 (Mapping properties). *Let $f : E \to F$, where $f^{-1}(B) = \{x \in E : f(x) \in B\} \subset \mathcal{E}\ \forall B \in \mathcal{F}$. Then for all $B, C, \{B_i\}$:*

1. $f^{-1}(\emptyset) = \emptyset$
2. $f^{-1}(F) = E$
3. $f^{-1}(B \setminus C) = f^{-1}(B) \setminus f^{-1}(C)$
4. $f^{-1}\left(\bigcup_i B_i\right) = \bigcup_i f^{-1}(B_i)$
5. $f^{-1}\left(\bigcap_i B_i\right) = \bigcap_i f^{-1}(B_i)$

♠ **Definition A.7** (Measurable function). *We enclose the results of Lemma A.6 in a Definition. For measurable spaces $(E, \mathcal{E}), (F, \mathcal{F})$ a function $f : E \to F$ is measurable when:*

$$f^{-1}(B) \in \mathcal{E} \quad \forall B \in \mathcal{F}$$

♣ **Proposition A.8** (Generating algebra measurability). *We characterize measurability as:*

$$f : E \to F \quad \text{measurable wrt } \mathcal{E}, \mathcal{F} \iff f^{-1}(B) \in \mathcal{B}\ \forall B \in \mathcal{F}_0 : \sigma(\mathcal{F}_0) = \mathcal{F}$$

A.2 More about measures

♣ **Proposition A.9** (Composition of measurable functions measurability). *For f measurable wrt $\mathcal{E},\mathcal{F}$ and g measurable wrt $\mathcal{F},\mathcal{G}$ their composition is measurable as well:*

$$f \circ g \quad \textit{measurable wrt } \mathcal{E},\mathcal{G}$$

♠ **Definition A.10** (Numerical functions, measurable). *For a measurable space $(E,\mathcal{E})$ a numerical function is a mapping to $\mathbb{R}$. Namely:*

$$f : E \to \overline{\mathbb{R}} = [-\infty,\infty]$$

The measurability condition is wrt $\mathcal{E},\mathcal{B}(\mathbb{R})$ and using the results of Propositions 1.21 and A.8:

$$f \quad \textit{measurable} \iff \forall r \in \mathbb{R} \quad f^{-1}([-\infty,r]) \in \mathcal{E}$$

Since $\{[-\infty,r] : r \in \mathbb{R}\}$ is the generating Borel set of $\mathbb{R}$.
For numerical functions, we omit saying that they are measurable wrt $\mathcal{E},\mathcal{B}(\mathbb{R})$ and simply write wrt $\mathcal{E}$.

♢ **Observation A.11** (About numerical functions). *Consider $f : E \to F$ where $F \subset \overline{\mathbb{R}}$ is* ***countable****. Then:*

$$f \quad \mathcal{E}\textit{-measurable} \iff f^{-1}(\{a\}) = \{x \in E : f(x) = a\} \in \mathcal{E}\ \forall a \in F$$

♣ **Proposition A.12** (Positive-negative decomposition measurability). *Any measurable function can be decomposed into measurable functions since:*

$$f \quad \textit{measurable} \iff f^{+} = f \vee 0, f^{-} = f \wedge 0 \quad \textit{measurable}$$

♠ **Definition A.13** (Canonical form of simple function). *A numerical function can always be reduced to its canonical form when it is simple. We say it is simple when:*

$$f = \sum^{n} a_i \mathbb{1}_{A_i} \quad a_i \in \mathbb{R}, A_i \in \mathcal{E}$$

And in this case we can safely say that:

$$\exists m \in \mathbb{N}^*,\ \textit{distinct}\ \{b_i\}_{i=1}^{m}\ \&\ \{B_i\}\ \textit{measurable partition}$$

which compose the canonical form of the simple function:

$$f = \sum_{i=1}^{m} b_i \mathbb{1}_{B_i}$$

♣ **Proposition A.14** (Properties of simple functions). *Let f, g be simple. Then:*

1. *f is always $\mathcal{E}$-measurable*
2. *$f+g$, $f-g$, fg, $\frac{f}{g}$ (provided that $g \neq 0 \forall x$), $f \vee g$, $f \wedge g$ are $\mathcal{E}$-measurable*
3. *for f $\mathcal{E}$-measurable taking finite values in $\mathbb{R}$ $\implies$ f is simple*

♠ **Definition A.15** (Limits of sequences of functions). *According to previous discussions, for a numerical function f_n (Def. A.10) it holds that:*

- $\inf f_n, \sup f_n, \liminf f_n, \limsup f_n$ *are pointwise extremizations*
- $\liminf f_n = \limsup f_n = f \implies \lim f_n = f$ *as usual*

♣ **Theorem A.16** (Limits of sequences measurability). *For f numerical all the functions of Definition A.15 are $\mathcal{E}$-measurable. $\lim f_n$ is so if it exists.*

Lemma A.17 (Dyadic functions properties). *For $n \in \mathbb{N}^*$ dyadic functions:*

$$d_n(r) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbb{1}_{[\frac{k-1}{2^n},\frac{k}{2^n}]}(r) + n\mathbb{1}_{[n,\infty]}(r) \qquad r \in \overline{R}_+$$

are such that:

1. $\forall n \; d_n(r)$ *is increasing, right continuous and simple*
2. $\lim_{n\to\infty} d_n(r) = r \quad \forall r \in \overline{\mathbb{R}}_+$

♣ **Theorem A.18** (Positive function measurability with respect to simple functions).

$$f : E \to \overline{\mathbb{R}}_+ \quad \mathcal{E}\text{-measurable} \iff (f_n) \nearrow f \quad : \quad (f_n) \text{ positive simple}$$

♠ **Definition A.19** (Monotone class of functions $\mathcal{M}$). *A collection of numerical functions $\mathcal{M} = \{f : E \to \overline{\mathbb{R}}\}$ with subsets $\mathcal{M}_+, \mathcal{M}_b$ of positive/bounded functions such that:*

1. $1 \in \mathcal{M}$
2. $f, g \in \mathcal{M}_b,\; a, b \in \mathbb{R} \implies af + bg \in \mathcal{M}$
3. $(f_n) \subset \mathcal{M}_+, (f_n) \nearrow f \implies f \in \mathcal{M}$

♣ **Theorem A.20** (Monotone Class Theorem II). *This is a revisited version of Theorem A.4. For $\mathcal{M}$ monotone on E, $\mathcal{C}$ a p-system (Def. 1.8). Then:*

$$\mathbb{1}_A \in \mathcal{M}\; \forall A \in \mathcal{C}, \quad \sigma(\mathcal{C}) = \mathcal{E} \implies M \supset \left\{\text{positive or bounded } \mathcal{E}\text{-measurable } f : E \to \overline{\mathbb{R}}\right\}$$

♠ **Definition A.21** (Isomorphism of measurbale spaces $\cong$). *A function f between measurable spaces $(E, \mathcal{E}), (F, \mathcal{F})$ that is a bijection induces an isomorphism. The inverse map: $\widehat{f}(y) = x \iff f(x) = y$ is measurable wrt $\mathcal{F}, \mathcal{E}$.*

♠ **Definition A.22** (Standard measurable space). *A measurable space $(E, \mathcal{E})$ is standard if:*

$$(E, \mathcal{E}) \cong (F, \mathcal{B}(F)), \quad F \subset \mathbb{R}$$

♣ **Theorem A.23** (Some standard measurable spaces). *Recognize that:*

1. $\mathbb{R}, \mathbb{R}^d, \mathbb{R}^\infty$ *and their respective Borel σ-algebras are standard*
2. *for E complete and separable* $\implies (E, \mathcal{B}(E))$ *is standard*
3. $([0,1], \mathcal{B}([0,1]))$ *is standard*
4. $(\mathbb{N}^*, \sigma(\mathbb{N}^*)), (\mathbb{N}, \sigma(\mathbb{N}))$ *are standard*

♢ **Observation A.24** (Confusing notation). *Notice that we may use the following symbol:*

$$\mathcal{E} \setminus \mathcal{F}$$

To denote functions that are measurable wrt the two σ-algebras.

♣ **Proposition A.25** (Arithmetic of measures). *Let $(E, \mathcal{E})$ be measurable and μ be a measure on it (Def. 2.2). Then:*

1. $\forall c \in \mathbb{R}_+ \; c\mu$ *is a measure*
2. *if ν a measure on $(E, \mathcal{E}) \implies \nu + \mu$ is a measure*
3. *for countably many measures $\mu_1, \mu_2, \ldots \implies \sum_n \mu_n$ is a measure*

♠ **Definition A.26** (σ-finite measure). *A measure μ on $(E, \mathcal{E})$ such that $\exists\{E_n\}$ measurable partitioning E with:*

$$\mu(E_n) < \infty \forall n$$

♠ **Definition A.27** (Σ-finite measure). *A measure μ on $(E, \mathcal{E})$ such that $\exists\{\mu_n\}$ **finite** measures with:*

$$\sum_n \mu_n = \mu$$

♣ **Proposition A.28** (Order of measure types).

$$\mu \quad finite \implies \mu \quad \sigma\text{-}finite \implies \mu \quad \Sigma\text{-}finite$$

And properties propagate in the opposite direction as always.

♣ **Proposition A.29** (Measure specification in p-systems, finite measures case). *Let $(E, \mathcal{E})$ be measurable, and μ, ν measures on it, both finite, and $\mathcal{C}$ be a p-system generating $\mathcal{E}$, i.e. $\sigma(\mathcal{C}) = \mathcal{E}$. Then:*

$$\mu(A) = \nu(A) \quad \forall A \in \mathcal{C} \implies \mu = \nu$$

◇ **Observation A.30** (Extensions extended). *We can prove the result of the Theorem for σ-finite measures as well. Notice that this holds also for probability measures on $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, and their Borel sets of intervals.*

♠ **Definition A.31** (Atom of a measure). *For a measurable space $(E, \mathcal{E}, \mu)$ we say:*

$$x \in E \quad atom \quad : \quad \mu(\{x\}) = 0$$

♠ **Definition A.32** (Diffuse & purely atomic measures). *Measures can be distinguished in terms of presence of atoms, but can also be of mixed types. We call:*

- *μ diffuse when $\nexists x \in E$ atom*
- *μ purely atomic when $\forall x \in E$ x is an atom*

♡ **Example A.33** (Some diffuse and purely atomic measures). *in some cases we will work with*

- *Leb (Def. 2.8), which is diffuse*
- *δ Dirac (Def. 2.3) which is purely atomic*
- *any discrete random variable (Def. 5.11) which has a purely atomic probability measure.*

Lemma A.34 (Finite & Σ-finite countability of atoms). *Establish that:*

1. *$\mu(E) < \infty \implies x$ atoms are countably many*
2. *μ Σ-finite $\implies x$ atoms are countable*

♣ **Proposition A.35** (Atomic Diffuse decomposition of measures). *For μ on $(E, \mathcal{E})$ Σ-finite it holds that:*

$$\mu = \lambda + \nu \quad \lambda\, diffuse, \quad \nu \text{ purely atomic}$$

♠ **Definition A.36** (Negligible sets in measurable spaces). *For a measure space $(E, \mathcal{E}, \mu)$ we say:*

- *B is measurable negligible when $\mu(B) = 0$*
- *an arbitrary subset (**not** necessarily measurable) $E_i \subset E$ is negligible if $E_i \subset B$, where B measurable negligible*

♠ **Definition A.37** (Complete measurable space). *A measure space $(E, \mathcal{E}, \mu)$ such that $\forall B \subset E : \mu(B) = 0$ the set B is measurable.*

♠ **Definition A.38** (Completion $(E, \overline{\mathcal{E}}, \overline{\mu})$). *Let $\mathcal{N}$ be a collection of negligible subsets. Define:*

- *$\overline{\mathcal{E}} = \sigma(\mathcal{E} \cup \mathcal{N})$*
- *$\overline{\mu}(A \cup N)$ for $A \in \mathcal{E}$ and $N \in \mathcal{N}$*

♣ **Proposition A.39** (Completion properties). *For a completion $(E, \overline{\mathcal{E}}, \overline{\mu})$ it holds that:*

1. *$\forall B \in \overline{\mathcal{E}} \quad B = A \cup N, A \in \mathcal{E}, N \in \mathcal{N}$*
2. *$\overline{\mu}(A \cup N) = \mu(A)$ is the unique measure on $\overline{\mathcal{E}}$ such that $\overline{\mu}(A) = \mu(A) \forall A \in \mathcal{E}$*
3. *the completion is complete according to Definition A.37*

♠ **Definition A.40** (Almost everywhere (almost every x)). *When a property is true $\forall x \in E$ but negligible sets we say the property holds almost everywhere (a.e.). When considering multiple measures, we say μ-almost everywhere.*

♡ **Example A.41** (Equivalence almost everywhere on functions). *See handwritten notes*

A.3 More about integrals of measures

Lemma A.42 (Integral properties for simple positive functions). *Recall Definition 4.1 and 4.4, the following is a similar result for general measures (Def. 2.2):*

$$f = \sum a_i \mathbb{1}_{A_i} \quad : \quad \mu(f) = \sum a_i \mu(A_i) = \int_E \mu(dx) f(x)$$

Satisfies:

- *linearity*
- *positivity*
- *monotone convergence*

♠ **Definition A.43** (Integral over a set). *Let $f \in \mathcal{E}$ be a measurable function, and A be measurable. Then $f\mathbb{1}_A \in \mathcal{E}$ and:*

$$\mu(f\mathbb{1}_A) = \int_E \mu(dx) f\mathbb{1}_A \int_A \mu(dx) f(x) = \int_A f d\mu$$

Lemma A.44 (Finite additivity of positive measurable functions). *Let $f \in \mathcal{E}_+$, $A, B \in \mathcal{E}$ be such that $A \cap B = \emptyset$ and $A \cup B = C$. Then:*

$$\mu(f\mathbb{1}_A) + \mu(f\mathbb{1}_B) = \mu(f\mathbb{1}_C)$$

Lemma A.45 (Positivity and monotonicity of integral). *For $f \in \mathcal{E}_+$ it holds that:*

1. $\mu f \geq 0$
2. $g \in \mathcal{E}_+ : f \leq g \implies \mu f \leq \mu g$

♣ **Theorem A.46** (General monotone convergence theorem). *This is the expanded result of Theorem 4.21, with proof.*
Let $(f_n)_{n\geq 1} \in \mathcal{E}_+$ be increasing. Then:

$$\implies \mu\left(\lim_{n\to\infty} f_n\right) = \lim_{n\to\infty} \mu(f_n)$$

♣ **Proposition A.47** (Linearity of integration). *For $f, g \in \mathcal{E}_+$ it holds:*

$$\mu(af + bg) = a\mu f + b\mu g \quad \forall a, b \in \mathbb{R}_+$$

♣ **Proposition A.48** (Insensitivity of the integral). *The following conclusions can be drawn when dealing with measures.*

1. $A \in \mathcal{E} : \mu(A) = 0 \implies \mu(f\mathbb{1}_A) = 0 \quad \forall f \in \mathcal{E}$
2. $f, g \in \mathcal{E}_+,\ f = g$ *a.e.* $x \implies \mu f = \mu g$
3. $f \in \mathcal{E}_+,\ \mu f = 0 \implies f = 0$ *a.e.* x

Lemma A.49 (Fatou's Lemma). *This is a very important result.*

$$(f_n) \subset \mathcal{E}_+ \implies \mu\left(\liminf_n f_n\right) \leq \liminf_n \mu(f_n)$$

Corollary A.50 (lim sup bounds for integrable bounded functions). *For $(f_n) \subset \mathcal{E}$ and g integrable we have:*

1. $f_n \geq g \quad \forall n \implies \mu(\liminf_n f_n) \leq \liminf \mu(f_n)$
2. $f_n \leq g \quad \forall n \implies \mu(\limsup_n f_n) \geq \limsup_n \mu(f_n)$

♣ **Theorem A.51** (General dominated convergence theorem). *This is the result with proof for the general case of Theorem 4.24.*
Let $(f_n)_{n\geq 1} \subset \mathcal{E}$ where $|f_n| \leq g\ \forall n$ and g is integrable according to Definition 4.5. Then:

$$\exists \lim_{n\to\infty} f_n \implies f \in \mathcal{L}_1, \quad \mu\left(\lim_{n\to\infty} f_n\right) = \lim_{n\to\infty} \mu(f_n)$$

Corollary A.52 (General bounded convergence Theorem). *This is the result of Corollary 4.26 for finite measures.*

$$(f_n) \subset \mathcal{E}, \quad (f_n) \text{ bounded}, \quad \mu \text{ finite}, \quad \exists \lim_{n\to\infty} f_n \implies f \in \mathcal{L}_1 \text{ bounded}, \quad \mu\left(\lim_{n\to\infty} f_n\right) = \lim_{n\to\infty} \mu(f_n)$$

♢ **Observation A.53** (About sequential continuity and monotone convergence). *Sequential continuity comes from sets assumed in the definition of measure (Def. 2.2).*
Monotone convergence (Thm. A.46) holds for measurable functions.

♢ **Observation A.54** (Almost everywhere version)**.** *For the results of this Section, one can derive an almost everywhere version and work with completions using Proposition A.39.*

♣ **Theorem A.55** (Characterization of integral)**.** *For a measurable space* $(E, \mathcal{E})$ *and a function* $L : \mathcal{E}_+ \to \overline{\mathbb{R}}_+$ *we state that:*

$$\exists! \mu \text{ on } (E, \mathcal{E}) : L(f) = \mu f \quad \forall f \in \mathcal{E}_+$$

satisfying the properties of an integral over a measure:

1. $f = 0 \implies L(f) = 0$
2. *L is linear*
3. *L is monotone convergent in* $\mathcal{E}_+$

Namely, once we have a general function satisfying properties $1, 2, 3$ *we are working with a valid measure induced by the integral representation. This characterization is unique on both directions.*

Appendix B

Transforms, Kernels, Product Spaces

B.1 Combining measures through kernels

♠ **Definition B.1** (Image measures of ν under h, $\nu \circ h^{-1}$)**.** *For measurable spaces $(F, \mathcal{F}), (E, \mathcal{E})$ and ν on $(F, \mathcal{F})$ consider:*

$$h : F \to E \quad \text{measurable wrt } \mathcal{F}, \mathcal{E}$$

Then we define:

$$\nu \circ h^{-1} : \mathcal{E} \to \mathbb{R}_+ \quad \nu \circ h^{-1}(B) = \nu(h^{-1}(B)) \ \forall B \in \mathcal{E}$$

Which is well defined since $h^{-1}(B) \in \mathcal{F}$ by the measurability of h (refer to Eqn. 3.1). This makes $\nu \circ h^{-1}$ a measure.

♢ **Observation B.2** (Other notions for $\nu \circ h^{-1}$)**.** *It often happens that the notation changes to $h \circ \nu, h(\nu), \nu \circ h, \nu_h$.*

Lemma B.3 (Finiteness relations)**.** *Recalling Definitions A.26, and A.27 it trivially holds that:*

1. *ν finite $\implies$ $\nu \circ h^{-1}$ finite*
2. *ν Σ-finite $\implies$ $\nu \circ h^{-1}$ Σ-finite*
3. *ν σ-finite $\not\Longrightarrow$ $\nu \circ h^{-1}$ σ-finite*
   but $\implies$ $\nu \circ h^{-1}$ Σ-finite

♣ **Theorem B.4** (Integral of image measure, integral change of variable)**.**

$$\forall f \in \mathcal{E}_+ \quad (\nu \circ h^{-1})(f) = \nu(f \circ h)$$

♢ **Observation B.5** (Extending the Theorem)**.** *The limitation $f \in \mathcal{E}_+$ can be removed upon noting that we just require that both parts of an arbitrary $f \in \mathcal{E}$ need to be well defined.*

♢ **Observation B.6** (Relationship to change of variable)**.** *Let $\mu = \nu \circ h^{-1}$ then:*

$$\int_F \nu(dx) f(h(x)) = \int_E \mu(dy) f(y) \tag{B.1}$$

In calculus we refer to this formula in the euclidean case for $E = \mathbb{R}^d$ and $F = \mathbb{R}^{d'}$ with ν, μ expressed in terms of the Lebesgue measure and the Jacobian of h.
In probability we often define ν through Equation B.1 in terms of μ, h.

♣ **Theorem B.7** (Lebesgue measure and Σ-finite h map link)**.** *Let $(E, \mathcal{E})$ be standard measurable (Def. A.22), μ be Σ-finite on $(E, \mathcal{E})$ and $b = \mu(E) \in \overline{\mathbb{R}}_+$. Then*

$$\implies \exists h : [0, b) \to E \quad \text{measurable wrt } \mathcal{B}([0, b)), \mathcal{E}$$

Such that:

$$\mu = \lambda \circ h^{-1} \quad \lambda = Leb[0, b) \to \overline{\mathbb{R}}_+$$

♠ **Definition B.8** (Indefinite integral of a function). *Let* $(E, \mathcal{E}, \mu)$ *be a measurable space,* $p \in \mathcal{E}_+$. *We can define:*

$$\nu(A) = \mu(p\mathbb{1}_A) = \int_A \mu(dx)p(x) \quad \forall A \in \mathcal{E}$$

as the indefinite integral of p *wrt* μ.

♣ **Proposition B.9** (Properties of $\nu(A)$). *For* $\nu(A)$ *as in Definition B.8 we have that:*

1. $\nu(A)$ *is a measure according to Definition 2.2*
2. $\forall f \in \mathcal{E}_+$ *it holds* $\nu(f) = \mu(pf)$

♢ **Observation B.10** (About the indefinite integral definition). *We can create new measures by simply using the fact that:*

$$\int_E \nu(dx)f(x) = \int_E \mu(dx)p(x)f(x)$$

Informally we may write $\nu(dx) = \mu(dx)p(x)$ *for* $x \in E$, ν, μ *measurable,* $p \in \mathcal{E}_+$.
We call p *the mass density of* ν *wrt* μ.

♠ **Definition B.11** (Density function p). *We write* $p = \frac{d\nu}{d\mu}$ *or* $p(x) = \frac{d\nu(x)}{d\mu(x)}$ *for* $x \in E$.

The next result is equivalent to Theorem 5.7, we state it here for the purpose of keeping the discourse flow.

♣ **Theorem B.12** (Radon-Nykodym Theorem). *Let* μ *be* σ*-finite, and* $\nu \ll \mu$.

$$\implies \exists p \in \mathcal{E}_+ : \int_E \nu(dx)f(x) = \int_E \mu(dx)p(x)f(x)$$

Where p *is* ***unique up to equivalences*** *meaning that if* $\exists p, p'$ *satisfying the requirement then* $p(x) = p'(x)$ *for a.e.* $x \in E$.

♠ **Definition B.13** (Transition Kernel K).

A transition kernel is a function between measurable space $(E, \mathcal{E}), (F, \mathcal{F})$, of the form:

$$K : (E \times \mathcal{F}) \to \overline{\mathbb{R}}_+$$

such that:

1. $x \to K(x, B)$ is $\mathcal{E}$ measurable $\forall B \in \mathcal{F}$
2. $B \to K(x, B)$ is a measure on $(F, \mathcal{F})$ $\forall x \in E$

♡ **Example B.14** (A transition kernel). *we will later show that for* ν *finite on* $(F, \mathcal{F})$ *a kernel* $K \in (\mathcal{E} \otimes \mathcal{F})_+$ *is:*

$$K(x, B) = \int_B \nu(dy)K(x, y) \quad x \in E, B \in \mathcal{F}$$

♣ **Theorem B.15** (Measure kernel function). *Let* K *be as in Definition B.13, conclude that:*

1. *For* $x \in E$ *we have* $Kf(x) = \int_F K(x, dy)f(y) \in \mathcal{E}_+ \quad \forall f \in \mathcal{F}_+$
2. *for* $B \in \mathcal{F}$ *sets* $\mu(K(B)) = \int_E \mu(dx)K(x, B)$ *is a measure on* $(F, \mathcal{F})$ *for any measure* μ *on* $(E, \mathcal{E})$
3. *For* $f \in \mathcal{F}_+$ *functions* $(\mu K)f = \mu(Kf) = \int_E \mu(dx) \int_F K(x, dy)f(y)$ *is a measure on* $(F, \mathcal{F})$ *for any measure* μ *on* $(E, \mathcal{E})$

♢ **Observation B.16** (About the Theorem). *The notation* $Kf, \mu K$ *can be seen as* $f \equiv$ *columns and* $\mu \equiv$ *rows generalized.*
To specify K *it is enough to find a form of* Kf $\forall f \in \mathcal{F}_+$.

Corollary B.17 (K specification, a "converse" of Theorem B.15). *A mapping* $K : \mathcal{F}_+ \to \mathcal{E}_+$ *such that* $f \to Kf$ *is a transition kernel where* $K(x, B) = K\mathbb{1}_B(x)$ *if and only if:*

$$\iff \begin{cases} K0 = 0 & \#1 \\ K(af + bg) = aKf + bKg \quad \forall f, g \in \mathcal{F}_+, a, b \in \mathbb{R} & \#2 \\ Kf_n \nearrow Kf \quad \forall (f_n) \nearrow f \in \mathcal{F}_+ & \#3 \end{cases}$$

♠ **Definition B.18** (Product Kernel). *For a kernel K from $(E,\mathcal{E})$ to $(F,\mathcal{F})$ and a kernel L form $(F,\mathcal{F})$ to $(G,\mathcal{G})$ we define the product as the function KL from $(E,\mathcal{E})$ to $(G,\mathcal{G})$.*

♣ **Proposition B.19** (Product is a transition kernel). *It holds that:*

1. *KL is a transition kernel*
2. *$KL(x,B) = \int_F K(x,dy)L(y,B)$ for $x \in E$ and $B \in \mathcal{G}$ is the representation*

♠ **Definition B.20** (Markov, subMarkov kernel). *A kernel K is **Markov** when $K(x,E) = 1\ \forall x$, **subMarkov** if $K(x,E) \leq 1\ \forall x$.*

♠ **Definition B.21** (Identity Kernel I). *The identity Kernel is a kernel from $(E,\mathcal{E})$ to $(E,\mathcal{E})$ defined as:*

$$I(x,A) = \delta_x(A) = \mathbb{1}_A(x) \quad x \in E, \quad A \in \mathcal{E}$$

♣ **Proposition B.22** (Kernels relations). *We establish the following results:*

1. *For a Kernel $K \implies K^0 = I, K^2, \ldots$ are kernels*
2. *for μ a measure and f a function: $If = f, \mu I = \mu, \mu If = \mu f, IK = KI = K$*
3. *for K a Markov kernel $\Longrightarrow K^n$ is Markov $\forall n \geq 1$*

♠ **Definition B.23** (Finite Kernels, Bounded Kernels). *Using the notions already encountered in Definitions A.26, A.27 and other discussions accordingly define:*

- *K finite such that $K(x,F) < \infty)\forall x$*
- *K σ-finite such that $B \to K(x,B)$ is σ-finite $\forall x$*
- *K bounded: $x \to K(x,F)$ bounded*
- *σ-bounded: $\exists\{F_n\}$ a measurable partition such that $x \to K(x,F_n)$ is bounded $\forall n$*
- *Σ-finite: K_n kernels are such that $K = \sum K_n$*
- *Σ-bounded: K_n are bounded in a partition $\forall n$*

♠ **Definition B.24** (Transition probability kernel). *A kernel is a probability kernel when it has mass 1 for each of its measures, i.e. $K(x,F) = 1\forall x$.*

♢ **Observation B.25.** *Markov Kernels (Def. B.20) are as in Definition B.24 and on $(E,\mathcal{E})$ to $(E,\mathcal{E})$.*

♣ **Proposition B.26** (Measurable functions of products). *Conclude that for $f \in \mathcal{E} \otimes \mathcal{F}$ measurable:*

1.
$$\begin{cases} x \to f(x,y) \in \mathcal{E} & \forall y \in F \\ y \to f(x,y) \in \mathcal{F} & \forall x \in E \end{cases}$$

2. *the opposite in general is not true!*

♣ **Proposition B.27** (A Σ-finite kernel makes measurable functions). *Let K be Σ-finite from $(E,\mathcal{E})$ to $(F,\mathcal{F})$. Then:*
$\forall f \in (\mathcal{E} \otimes \mathcal{F})_+$ the map:

$$Tf(x) = \int_F K(x,dy)f(x,y) \quad x \in E$$

is such that:

1. *$Tf \in \mathcal{E}_+$*
2. *$T : (\mathcal{E} \otimes \mathcal{F})_+ \to \mathcal{E}_+$ is linear an continuous under increasing limits and such that:*
   (a) $T(af + bg) = aTf + bTg \quad \forall f, g, \forall a, b \in \mathbb{R}_+$
   (b) $Tf_n \nearrow Tf \quad \forall f_n, f : f + n \nearrow f$

B.2 Noteworthy results at divergent sizes & Fubini's Theorem

♣ **Theorem B.28** (Measures on product space). *Let μ be a measure on $(E,\mathcal{E})$, K a Σ-finite kernel on $(E,\mathcal{E})$ to $(F,\mathcal{F})$. Then:*

1. $\pi f = \int_E \mu(dx) \int_F K(x, dy) f(x, y)$ *for* $f \in (\mathcal{E} \otimes \mathcal{F})_+$ *is a measure on the measurable space* $(E \times F, \mathcal{E} \otimes \mathcal{F})$
2. *if* μ *is* σ*-finite and* K *is* σ*-bounded additionally:*

$$\exists! \mu : \pi(A \times B) = \int_A \mu(dx) K(x, B) \quad A \in \mathcal{E}, B \in \mathcal{F}$$

♢ **Observation B.29** (Special case of the Theorem). *If* $K(x, B) = \nu(B)$ *a* Σ*-finite measure on* $(F, \mathcal{F})$ *then* π *is said to be a* ***product measure*** *and we write* $\pi = \mu \times \nu$.

♣ **Theorem B.30** (Fubini's Theorem). *Let* μ, ν *be* Σ*-finite measures respectively on* $(E, \mathcal{E})$ *and* $(F, \mathcal{F})$*. Then:*

1. $\exists! \Sigma$*-finite* $\pi = \mu \times \nu : \forall f \in (\mathcal{E} \otimes \mathcal{F})_+$ *it holds:*

$$\pi f = \int_E \mu(dx) \int_f \nu(dy) f(x, y) = \int_F \nu(dy) \int_E \mu(dx) f(x, y)$$

2. *if* $f \in (\mathcal{E} \otimes \mathcal{F})$ *is* π*-integrable then:*

$$\begin{cases} y \to f(x, y) & \nu\text{-integrable} \quad \mu\text{-a.e. } x \\ x \to f(x, y) & \mu\text{-integrable} \quad \nu\text{-a.e. } y \\ \#1 \text{ holds again} \end{cases}$$

Corollary B.31 (Extending results to finite products). *For* $\bigotimes_{i=1}^n (E_i, \mathcal{E}_i)$ *such that the sigma algebra is generated by the rectangles, namely:* $\mathcal{E}_1 \otimes \ldots \otimes \mathcal{E}_n = \sigma(A_1 \times \cdots \times A_n)$ *and for* $\mu_1, \ldots, \mu_n$ *finite measures:*

1. $\pi = \bigtimes_{i=1}^n \mu_i$ *is such that:*

$$\pi f = \int_{E_1} \mu_1(dx_1) \cdot \int_{E_n} \mu_n(dx_n) f(x_1, \ldots, x_n)$$

2. $\bigotimes_{i=1}^n (E_i, \mathcal{E}_i, \mu_i)$ *is a measure space*
3. *for* f *positive or* π*-integrable Theorem B.30 hods with any order of the integrals*

Corollary B.32 (More general measures for finite products). *For* μ_1 *on* $(E_1, \mathcal{E}_1)$ *and* $\{K_i\}_{i=2}^n$ *kernels on* $(E_i, \mathcal{E}_i)$ *to* $(E_{i+1}, \mathcal{E}_{i+1}) \forall i$ *we can define:*

$$\pi f = \int_{E_1} \mu_1(dx_1) \int_{E_2} K_2(x_1, dx_2) \int_{E_3} \cdots \int_{E_n} K_n(x_{n-1}, dx_n)$$

Such that π *is a measure on* $\left(\bigtimes_{i=1}^n E_i, \bigotimes_{i=1}^n \mathcal{E}_i\right)$ *and* $\pi = \mu_1 \times K_2 \times \cdot \times K_n$.

♠ **Definition B.33** (Infinite products measurability setting). *For* $\mathbb{T}$ *an arbitrary index set where* $(E_t, \mathcal{E}_t)$ *is measurable* $\forall t \in \mathbb{T}$ *we can let:*

$$\bigtimes_{t \in \mathbb{T}} A_t = \{X = (X_t)_{t \in \mathbb{T}} \, product \ space : \ X_t \in A_t \quad \forall t \in \mathbb{T}\} \quad where \quad A_t \neq E_t \ for \ \{t_i\}_{i=1}^n \ finite$$

So that the measurable space is $\bigotimes_{t \in \mathbb{T}} (E_t, \mathcal{E}_t)$

♣ **Proposition B.34** (Infinite products measurability condition). *Let:*

- $(\Omega, \mathcal{H})$ *be measurable and* $(F, \mathcal{F}) = \bigotimes_{t \in \mathbb{T}} (E_t, \mathcal{E}_t)$
- $\forall t \in \mathbb{T} \ f_t : \Omega \to E_t$
- $\forall \omega \in \Omega \ (f_t(\omega))_{t \in \mathbb{T}} \in F$

Then:

$$f : \Omega \to F \quad measurable \ wrt \quad \mathcal{H}, \mathcal{F} \iff f_t \quad measurable \ wrt \quad \mathcal{H}, \mathcal{E}_t \ \forall t \in \mathbb{T}$$

Appendix C

Miscellaneous Results

C.1 Martingales & Stopping times

♣ **Theorem C.1** (Uniform integrability & convex function). *This is the supporting proof for Lemma 11.51 extended.*
The following are equal (TFAE):

1. $\mathcal{C}$ *u.i.*
2. $h(b) = \sup_{\mathcal{C}} \int_b^\infty dy\, \mathbb{P}[|X| > y] \to 0 \quad$ *as* $b \to \infty$
3. *for some increasing positive convex function* $f : \mathbb{R} \to \mathbb{R}_+$ *with* $\lim_{x\to\infty} \frac{f(x)}{x} \to \infty$ *we have*

$$\sup_{\mathcal{C}} \mathbb{E}\left[f \circ |X|\right] < \infty$$

C.2 Random Variables

What follows is a treatment of the theoretical results needed to reach Carathéodory's Theorem and eventually prove Theorem 3.21.

♠ **Definition C.2** (Outer measure μ^* for an algebra). *For* μ *a measure (Def. 2.2) on a* σ*-algebra* $\mathcal{C}$ *we define:*

$$\mu^*(A) := \inf \sum_n \mu(A_n) \quad \forall A \in \Omega : \begin{cases} (A_n) \subset \mathcal{C} \\ A \subset \bigcup_n A_n \quad (A_n) \textit{ is a cover} \end{cases}$$

Where the sequence covers the set.

♠ **Definition C.3** (μ^* measurability). $A \in \Omega$ *is* μ^**-measurable when:*

$$\mu^*(A \cup B) + \mu^*(A^c \cup B) = \mu^*(B) \quad \forall B \subset \Omega$$

♠ **Definition C.4** (μ^*-measurable sets $\mathcal{M}$). *We denote the collection of* μ^**-measurable sets as* $\mathcal{M}$.

♣ **Proposition C.5** (Trivial properties of μ^*). *For an outer measure* μ^* *on a* σ*-algebra* $\mathcal{F}$*:*

1. *(nullity)* $\mu^*(\emptyset) = 0$
2. *(non negativity)* $\mu^*(A) \geq 0 \; \forall A \subset \Omega$
3. *(monotonicity)* $A \subset B \implies \mu^*(A) \leq \mu^*(B)$

♣ **Proposition C.6** (Countable subadditivity of μ^*). *For a collection* $(A_n) \subset \mathcal{M}$*:*

$$\mu^*\left(\bigcup_n A_n\right) \leq \sum_n \mu^*(A_n)$$

Corollary C.7 (Measurability in μ^* characterization). *It holds:*

$$A \in \mathcal{M} \iff \forall B \subset \Omega \quad \mu^*(A \cap B) + \mu^*(A^c \cap B) \geq \mu^*(B)$$

♣ **Proposition C.8** ($\mathcal{M}$ is an Algebra). *$\mathcal{M}$ from Definition C.4 is an algebra in the sense of Definition 1.4.*

♣ **Proposition C.9** (Intersection union rule of disjoint sets). *For a collection of disjoint countable sets $(A_n) \subset \mathcal{M}$:*

$$\mu^*\left(B \cap \left(\bigcup_n A_n\right)\right) = \sum_n \mu^*(B \cap A_n) \quad \forall B \in \Omega$$

♣ **Proposition C.10** ($\mathcal{M}$ is a σ-algebra and the restriction of μ^* additivity). *As the title preludes:*

1. *$\mathcal{M}$ is a σ-algebra (Def. 1.6)*
2. *$\mu^*_{\mathcal{M}}$ is countably additive*

♣ **Proposition C.11** (Agreement of measure and outer measure + order). *Consider a measure μ, a measurable σ-algebra $\mathcal{C}$ an the outer measure μ^*. The collectio $\mathcal{M}$ with the conditions of Definition C.4 is such that:*

1. *$\forall A \in \mathcal{C} \quad \mu^*(A) = \mu(A)$*
2. *$\mathcal{C} \subset \mathcal{M}$*

♣ **Theorem C.12** (Carathéodory extension Theorem).

$$\mu \quad finite\ on \quad \mathcal{C}\ \sigma\text{-}algebra \implies \exists!\ extension\ \mu^*\ on\ \sigma(\mathcal{C})$$

♢ **Observation C.13** (What we know so far). *We briefly recollect some facts already in hand:*

- *(Prop. A.29 and the subsequent Obs.) for a p-system (Def. 1.8) $\mathcal{C}$ and ν, μ both σ-finite (Def. A.26) on $\sigma(\mathcal{C})$ equality $\forall A \in \mathcal{C}$ implies equality $\forall A \in \sigma(\mathcal{C})$*
- *(Thm. C.12) μ finite on an algebra $\mathscr{A} \implies \exists! \mu^*$ on $\sigma(\mathscr{A})$*
- *(Prop. A.2) $\mathscr{A}$ algebra $\iff$ $\mathscr{A}$ is a p-d-system (Defs. 1.8, A.1)*

Are such that ν, μ on $\mathcal{E} = \sigma(\mathscr{A})$ are finite on $\mathscr{A}$ and agree $\forall A \in \mathscr{A}$ then they agree $\forall A \in \mathcal{E}$.
However, in practical cases, we can do more by requiring less that equality over an algebra $\mathscr{A}$, and we can also extend Theorem C.12 for σ-finite measures while doing so.

♠ **Definition C.14** (Semi algebra $\mathscr{B}$). *Let $\mathscr{B} \neq \emptyset$ be a collection of subsets of the sample space of arrival:*

$$\mathscr{B} = \{B_i : B_i \subset E\}$$

Additionally let $\mathscr{A} := \{\bigcup_{i=1}^n B_i, B_i \cap B_j = \emptyset, B_i \in \mathscr{B} \forall i\}$ where:

1. *$B_1, B_2 \in \mathscr{B} \implies B_1 \cap B_2 \in \mathscr{B}$*
2. *$B \in \mathscr{B} \implies B^c \in \mathscr{A}$*
3. *we refer to $\mathscr{A}$ as the algebra (semi)generated by $\mathscr{B}$ and we write $\mathscr{A} = \widetilde{\sigma}(\mathscr{B})$*

♢ **Observation C.15** (Setting). *Suppose μ is a map on the semialgebra $\mathscr{B}$ such that $\mu(B) \in [0, \infty] \quad \forall B \in \mathscr{B}$. Additionally $\mathscr{A} = \widetilde{\sigma}(\mathscr{B})$ and $\mathcal{E} = \sigma(\mathscr{A}) = \sigma(\widetilde{\sigma}(\mathscr{B}))$. By the additivity of measures:*

$$A = \bigcup_{i=1}^n B_i,\ disjoint, \quad A \in \mathscr{A}, \quad \mu(A) = \sum_{i=1}^n \mu(B_i)$$

However, how do we assign $\mu(B)$ for $B \in \mathscr{B}$?

♣ **Proposition C.16** (Semialgebra to Algebra extension). *Let $\mathscr{B}$ be a semialgebra on E and $\mathscr{A} = \widetilde{\sigma}(\mathscr{B})$. Then, the map:*

$$\mu : \mathscr{B} \to [0, \infty]$$

*is **uniquely extended** to $\mathscr{A}$ if and only if:*

1. *(nullity) $\emptyset \in \mathscr{B} \implies \mu(\emptyset) = 0$*
2. *(finite additivity) $\{B_i\}_{i=1}^n \subset \mathscr{B}$ finite disjoint $\implies \mu(A) = \mu(\bigcup_{i=1}^n B_i) = \sum_{i=1}^n \mu(B_i)$*

3. *(countable subadditivity)* $(B_n) \subset \mathscr{B}$ *countable disjoint and* $B \subset \bigcup_n B_n, \quad B \in \mathscr{B} \implies \mu(B) \leq \sum_n \mu(B_n)$

Corollary C.17 (Double extension)**.** *This is a corollary of Carathéodory's extension Thm. C.12. For a semialgebra* $\mathscr{B}$ *where* $\mathscr{A} = \widetilde{\sigma}(\mathscr{B})$ *and* μ *a measure on* $\mathscr{A}$*,* σ*-finite (Def. A.26), then:*

$$\exists!\ extension\ \mu\ on \quad \mathcal{E} = \sigma(\mathscr{A}) = \sigma(\widetilde{\sigma}(\mathscr{B}))$$

satisfying the requirements of Proposition C.16.

♡ **Example C.18** (Proving Theorem 3.21)**.** *Recall that* $\mathcal{B}(\mathbb{R}) = \sigma(\{(-\infty, a],\ a \in \mathbb{R}\})$ *(Ex. 1.20), also:*

$$\forall b \geq a \quad Leb([a,b]) = Leb([a,b)) = Leb((a,b]) = Leb((a,b)) = b - a$$

Notice that $\{(-\infty, a],\ a \in \mathbb{R}\}$ *is trivially a semialgebra (Def. C.14) by being a p-system (Def. 1.8). Then:*

$$\widetilde{\sigma}(\{(-\infty, a],\ a \in \mathbb{R}\}) = \mathscr{A} \quad Algebra$$

And by Proposition C.16 since μ *satisfies #1, #2, #3 we can extend it to a measure* $\mu = Leb$ *on* $\mathscr{A}$ *to* $\overline{\mu}$ *where* $\overline{\mu}(I) := Leb(I)\ \forall I \in \mathscr{B}$.
Now $\overline{\mu}$ *is on* $\mathscr{A} = \widetilde{\sigma}(\mathscr{B})$*, thus on finite disjoint unions. Additionally,* $\overline{\mu}$ *is* σ*-finite since:*

$$\mathbb{R} = \bigcup_{a \in \mathbb{Q}} [-a, a]$$

by $\mathbb{Q}$ *being dense in* $\mathbb{R}$ *(Prop. 18.15) and we could decompose the measure to a countable set of measures. Then, by Carathéodory's Extension Thm. C.12 we identify a further unique extension of* $\overline{\mu}$ *into* μ^* *on the* σ*-algebra generated by* $\mathscr{A}$ *which is:*

$$\mathcal{E} = \sigma(\mathscr{A}) = \sigma(\{(-\infty, a],\ a \in \mathbb{R}\}) = \mathcal{B}(\mathbb{R})$$

And we can safely say that:

$$F_X(x) = F_Y(x) \quad \forall x \in \mathbb{R} \iff \mathcal{P}_X(A) = \mathcal{P}_Y(A) \quad \forall A \in \mathcal{B}(\mathbb{R})$$

a result which can be adapted to sample spaces E *different than* $\mathbb{R}$*.*

C.3 Laplace, More filtration types, Poisson vs Martingale

In this Section we give more context for the treatment of Sections 18, 19, 22 and provide a proof of the missing piece of Theorem 12.4. The very first result is a useful property.

♣ **Proposition C.19** (Laplace transforms and finiteness)**.** *This result is used in Lemma 14.3. Recall the definition of Laplace transform (Def. 6.11). Then:*

$$X \geq 0 \quad a.s. \implies \lim_{r \downarrow 0} \widehat{\mathcal{P}}_X(r) = \mathbb{P}[\{X < \infty\}]$$

♠ **Definition C.20** ($\mathcal{F}$-predictable σ-algebra $\mathcal{F}^p$)**.** *For:*

$$\mathcal{F}^{pp} = \{H \times (s,t] : 0 \leq s < t < \infty, H \in \mathcal{F}_s\} \cup \{H \times \{0\} : H \in \mathcal{F}_0\}$$

We set $\mathscr{F}^p = \sigma(\mathcal{F}^{pp})$ *on* $\Omega \times \mathbb{R}_+$*.*
A process $F = (F_t)_{t \in \mathbb{R}_+}$ *is said to be* $\mathcal{F}$*-predictable if its path is measurable wrt* $\mathcal{F}^p$*. Namely,* $(\omega, t) \to F_t(\omega) \in \mathcal{F}^p$*.*
We call elements of $\mathcal{F}^{pp}$ ***primitive sets*** *and indicators* ***primitives****.*

♣ **Proposition C.21** (Predictability, left continuity and adaptedness)**.** *Say* $\mathcal{F}^p = \sigma(\mathcal{G})$ *then:*

$$\mathcal{G} = \left\{G = (G_t)_{t \in \mathbb{R}_+} : adapted,\ left\ continuous\ on\ \Omega \times \mathbb{R}_+\right\}$$

♠ **Definition C.22** (Setting)**.** *For the results of this Section, we consider:*

- $N = (N_t)_{t \in \mathbb{R}_+}$ *an increasing, right continuous process adapted to* $\mathcal{F}$

- $\nu_t = \mathbb{E}[N_t] < \infty \ \forall t$ *and* $\mathbb{E}_s[N_t - N_s] = \nu_t - \nu_s \ \forall s \leq t \in \mathbb{R}_+$ *which means that:*

$$\widetilde{N}_t = N_t - \nu_t \quad \mathcal{F}\text{-martingale}$$

♢ **Observation C.23** (Feasibility of the Definition requirements). *Observe that for a process N with $\perp\!\!\!\perp$ increments and finite expectation at all times $\mathbb{E}[N_t] < \infty \forall t \in \mathbb{R}_+$ we are in the setting of Definition C.22.*

♣ **Theorem C.24** (Stochastic integrals reduce to the mean in expectation). *Consider a process $F = (F_t)_{t \in \mathbb{R}_+}$ positive and predictable (Def. 12.5). Then:*

$$\mathbb{E} \int_{\mathbb{R}_+} F_t dN_t = \mathbb{E} \int_{\mathbb{R}_+} F_t d\nu_t$$

Corollary C.25 (Stopping times version). *For a positive predictable process $F = (F_t)_{t \in \mathbb{R}_+} \in \mathcal{F}^p$ and stopping times S, T (Def. 11.9) such that $S \leq T$ it holds that:*

$$\mathbb{E}_S \int_{(S,T]} F_t dN_t = \mathbb{E}_S \int_{(S,T]} F_t d\nu_t$$

♣ **Theorem C.26** (Martingale Creation by predictability). *Predictability and boundedness are sufficient to form a martingale once integrating.*

$$F = (F_t)_{t \in \mathbb{R}_+} \in \mathcal{F}^p : \forall t \ F_t \leq b \in \mathbb{R}_+ \implies M_t = \int_{[0,t]} F_s d\widetilde{N}_s \quad t \in \mathbb{R}_+ \quad \mathcal{F}\text{-martingale}$$

Where $\widetilde{N}_s$ was constructed in Definition C.22.

Lemma C.27 (Counting process decomposition of bounded functions). *For f a bounded function and N a counting process (Def. 11.13) we can decompose f into:*

$$f(N_t) = f(0) + \int_{[0,t]} f(N_{s-} + 1) + f(N_{s-}) dN_s \quad N_{s-} = \lim_{r \uparrow s} N_r$$

♣ **Proposition C.28** (Martingale is Poisson counterdirection). *This is the opposite direction required for the proof of Theorem 12.4.*
For a counting process $N = (N_t)_{t \in \mathbb{R}_+}$ (Def. 11.13) adapted to $\mathcal{F}$:

$$\widetilde{N} = (N_t - ct)_{t \in \mathbb{R}_+} \quad \mathcal{F}\text{-martingale} \ (Def.11.35) \implies N \sim Pois(c) \quad (Def.12.2)$$

C.4 More about characteristic functions

The following is a quick explanation of the result of Lemma 17.20.

Lemma C.29 (Characteristic function "symmetry"). *Recall Definition 6.17 for $\Phi_X(t)$.*
For densities f, g of X, X' it holds that:

$$\int_{\mathbb{R}} \Phi_X(t) g(t) dt = \int_{\mathbb{R}} f(t) \Phi_{X'}(t) dt$$

Lemma C.30 (A strange consequence of the CLT). *For X a r.v. and $\epsilon > 0$ it holds that:*

$$X \perp\!\!\!\perp N_\epsilon \sim \mathcal{N}(0, \epsilon) \implies \int_{\mathbb{R}} g(y) \rho_\epsilon(x - y) dy \overset{\epsilon \to 0}{\to} g(x)$$

For ρ_ϵ the Radon Nykodym density (Def. 5.9) of $\mathcal{N}(0, \epsilon)$.

♢ **Observation C.31** (Idea for inverse of $\Phi_X(t)$). *By the above Lemma $g(x) = \lim_{\epsilon \to 0} g(y) \rho_\epsilon(x - y) dy$. Then, given x, ϵ we can find a map:*

$$h_{x,\epsilon}(\cdot) \quad s.t. \quad y \to \Phi_h(y) \ is \ y \to \rho_\epsilon(x - y)$$

so that applying Lemma C.29:

$$g(x) = \lim_{\epsilon \to 0} g(y) \rho_\epsilon(x - y) dy = \lim_{\epsilon \to 0} \int_{\mathbb{R}} \Phi_g(t) h_{x,\epsilon}(t) dt$$

which, together with the continuity and monotonicity of the characteristic function (a form of Theorem 6.18) ensures uniqueness of the relationship $x \iff h_{x,\epsilon}$.

Lemma C.32 (An identity for distribution & characteristic function)**.** *For X a r.v. in $\mathbb{R} = E$ with density $f(x)$:*

$$\Phi_X(r) = \int_{\mathbb{R}} e^{irx} f(x) dx \iff f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{irx} \Phi_X(r) dr$$

Appendix D

Completing Chapter 18 and 19, Brownian Motion

♣ **Theorem D.1** (Lévy characterization as Wiener). *This is the proof of Theorem 18.4#2.*

$$X_t = at + bW_t \ : \ (X_t)_{t\in\mathbb{R}_+} \ Lévy \implies (W_t)_{t\in\mathbb{R}_+} \ Wiener$$

Lemma D.2 (Kolmogorov Maximal inequality). *This is the proof of Lemma 18.13.*
For an independency of r.v.s $\{X_i\}_{i=1}^n$ *with* $\mathbb{E}[X_i] = 0 \forall i$ *with sum* $S_n = \sum^n X_i$ *it holds:*

$$a^2\mathbb{P}[\max_{k\leq n} |S_k| > a] \leq V[S_n] \quad \forall a \in (0,\infty)$$

♣ **Proposition D.3** (Dyadics are dense in $\mathbb{R}$). *This is the proof of Proposition 18.15.*

$$\forall r \in \mathbb{R}, \forall \epsilon > 0, \ \exists k, m \in \mathbb{N} : t \in (k2^{-m}, (k+1)2^{-m}), \ t - k2^{-m} < \epsilon$$

D.1 More about augmented filtrations

This Section is devoted to providing the additional details needed to understand well the results arising from the augmented filtration construction, such as Blumenthal's law or relations with stopping times. The starting point for this context is Definition 19.1, here we start from a parent of it.

♠ **Definition D.4** (Augmented filtration). *A filtration* $\mathcal{F}$ *on a complete (Def. A.37) space* $(\Omega, \mathcal{H}, \mathbb{P})$ *is a filtration such that:*

$$\forall N \in \mathcal{H} : \mathbb{P}[H] = 0 \quad N \in \mathcal{F}_0$$

Namely, the filtration always contains (from the start onwards) negligible sets of a given probability space.

♢ **Observation D.5** (About augmented filtrations). *For a complete probability space, a filtration* $\mathcal{F}$ *and the set of negligible sets* $\mathcal{N} \subset \mathcal{H}$ *we can say that the filtration:*

$$\overline{\mathcal{F}} = \sigma\left(\{\mathcal{F}_t \cup \mathcal{N} \forall t\}\right)$$

is augmented and that the original fitration $\mathcal{F}$ *is augmented if and only if* $\mathcal{F} = \overline{\mathcal{F}}$.

♠ **Definition D.6** ("Right continuous" filtration $\mathcal{F}_+$). *we define the filtration:*

$$\mathcal{F}_{t+} := \bigcap_{\epsilon>0} \mathcal{F}_{t+\epsilon} \quad \forall t \in \mathbb{R}_+, \quad \mathcal{F}_+ = (\mathcal{F}_{t+})_{t\in\mathbb{R}_+}$$

Notice that $\mathcal{F}_+$ *is finer (Def. 11.6) than* $\mathcal{F}$. *Additionally, it is right continuous if the sense that* $\mathcal{F}_t = \mathcal{F}_{t+} \ \forall t \in \mathbb{R}_+$.

♡ **Example D.7** (Interpreting the new filtrations). *For a smooth motion $t \to X_t$ we could see $\mathcal{F}_t$ as the σ-algebra containing all the information about past and present positions, while $\mathcal{F}_{t+}$ contains also the velocity at time t, which formally is:*

$$v = \lim_{\epsilon \to 0} \frac{X_{t+\epsilon} - X_t}{\epsilon}$$

as well as the acceleration etc.

♣ **Theorem D.8** (Augmentation and stopping times). *For a filtration $\mathcal{F}$ and a random time $T : \Omega \to \mathbb{R}_+$:*

$$T \text{ stopping } \mathcal{F}_+ \quad (Def.11.9) \iff \{T < t\} \in \mathcal{F}_t \ \forall t \in \mathbb{R}_+$$

Namely, T is a stopping time of the right continuous filtration if it is so for the filtration itself.

Corollary D.9 (Right continuous variant). *$\mathcal{F}$ right continuous s.t. $\mathcal{F}_t = \mathcal{F}_{t+} \forall t \iff \{t < T\} \in \mathcal{F}_t \forall t$ which is in accordance with Definition 11.9.*

♠ **Definition D.10** (Past until T). *For a filtration $\mathcal{F}$ and a stopping time T for $\mathcal{F}_+$ the corresponding past until T is:*

$$\mathcal{F}_{T+} := \{H \in \mathcal{H} : H \cap \{T \le t\} \in \mathcal{F}_{t+} \ \forall t \in \mathbb{R}_+\}$$

Which in case of right continuity can be declined to the classical Definition of stopped filtration (Def. 11.19).

♣ **Proposition D.11** (Sequences of stopping times). *For a sequence of random times $(T_n)_{n \in \mathbb{N}}$ stopping either $\mathcal{F}$ or $\mathcal{F}_+$ it holds:*

1. *$T = \inf T_n$ is stopping $\mathcal{F}_+$*
2. *$\mathcal{F}_{T+} = \bigcap_n \mathcal{F}_{T_n+}$*

♢ **Observation D.12** (Building limits with the result and right continuity). *If $\mathcal{F}$ is right continuous and $(T_n)_{n \in \mathbb{N}}$ is a sequence of stopping times then $\liminf, \limsup, \sup, \inf$ are all stopping times.*

♣ **Proposition D.13** (Partitions of independencies). *Given an independency $\{\mathcal{F}_t, \ t \in \mathbb{T}\}$ and a sequence $\{T_1, \dots\}$ partitioning T we can say that:*

$$\implies \mathcal{F}_T := \{\mathcal{F}_t \ : \ t \in T_i\} \ i \in \mathbb{N}^* \ : \ \{\mathcal{F}_{T_1}, \dots, \mathcal{F}_{T_n}\} \text{ independency}$$

♠ **Definition D.14** (Future σ-algebras and tail σ-algebra $\mathcal{T}_n, \mathcal{T}$). *define:*

- *$\mathcal{T}_n = \bigvee_{m > n} \mathcal{G}_m$ for $(\mathcal{G}_n) \subset \mathcal{H}$*
- *$\mathcal{T} = \bigcap_n \mathcal{T}_n$ the remote future tail σ-algebra*

♣ **Theorem D.15** (Kolmogorov's 0-1 law). *For an independency the evens in the tail σ-algebra are either null or certain:*

$$(\mathcal{G}_n) \subset \mathcal{H} \text{ independency} \implies \forall H \in \mathcal{T} \quad \mathbb{P}[H] \in \{0, 1\}$$

♢ **Observation D.16** (Setting for next Theorem). *The probability space is $(\Omega, \mathcal{H}, \mathbb{P})$, $X : \Omega \to \mathbb{R}^d$ is a stochastic process, and $\mathcal{G} = \sigma(\{X\})$ is the filtration it generates.*
The completion (Def. A.38) is denoted as $(\Omega, \overline{\mathcal{H}}, \overline{\mathbb{P}})$ with sigma algebra generated by the negligible sets denoted as $\mathcal{N} = \sigma(\{N \in \overline{\mathcal{H}} \ : \ \overline{\mathbb{P}}[N] = 0\})$.
Here the augmented filtration is $\overline{\mathcal{G}}$ built as $\overline{\mathcal{G}}_t \bigvee \mathcal{N} \ \forall t$.
Recall that right continuity is defined as:

$$\bigcap_{\epsilon > 0} \overline{\mathcal{G}}_{t+\epsilon} = \overline{\mathcal{G}}_t \quad \forall t \in \mathbb{R}_+$$

This fact will hold in the specific setting of the next Theorem.

♣ **Theorem D.17** (Lévy process in augmentation is Lévy). *Let $X = (X_t)_{t \in \mathbb{R}_+}$ be a Lévy process (Def. 17.1) wrt $\mathcal{G}$ over $(\Omega, \mathcal{H}, \mathbb{P})$. Then:*

1. *X is Lévy wrt $\overline{\mathcal{G}}$ over $(\Omega, \overline{\mathcal{H}}, \overline{\mathbb{P}})$*
2. *$\mathcal{G}$ the augmentation is (augmented and) right continuous*

Corollary D.18 (Blumenthal's 0-1 law). *The results of this Section eventually allow us to prove Corollary 19.9. With the setting of the last Theorem:*

$$\forall G \in \overline{\mathcal{G}}_0 \quad \mathbb{P}[G] \in \{0, 1\}$$

Bibliography

[Çin11] Erhan Çinlar. *Probability and Stochastics*. Vol. 261. Graduate Texts in Mathematics. New York, NY: Springer, 2011. ISBN: 978-0-387-87858-4 978-0-387-87859-1. DOI: 10.1007/978-0-387-87859-1. URL: http://link.springer.com/10.1007/978-0-387-87859-1 (visited on 11/07/2022).

[Ver18] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. 1st ed. Cambridge University Press, Sept. 27, 2018. ISBN: 978-1-108-23159-6 978-1-108-41519-4. DOI: 10.1017/9781108231596. URL: https://www.cambridge.org/core/product/identifier/9781108231596/type/book (visited on 12/12/2022).

[BZ99] Zdzisław Brzeźniak and Tomasz Zastawniak. *Basic Stochastic Processes*. Springer Undergraduate Mathematics Series. London: Springer London, 1999. ISBN: 978-3-540-76175-4 978-1-4471-0533-6. DOI: 10.1007/978-1-4471-0533-6. URL: http://link.springer.com/10.1007/978-1-4471-0533-6 (visited on 11/07/2022).

[Dur19] Richard Durrett. *Probability: Theory and Examples*. Fifth edition. Cambridge Series in Statistical and Probabilistic Mathematics 49. Cambridge ; New York, NY: Cambridge University Press, 2019. ISBN: 978-1-108-47368-2.