Course Notes for *Stochastic Process*

Fei Li*

*Email: fei.li.best@gmail.com.

Contents

Part I

Probability Theory

1 Measures and Integrations

1.1 The Monotone Class Theorem

Definition 1.1. Let X be some set, and let $\mathcal{P}(X)$ be its power set. A subset $\mathcal{A} \subseteq \mathcal{P}(X)$ is called a *σ-algebra* if it satisfies the following three properties:

1. $X \in \mathcal{A}$;
2. $\mathcal{A}$ is closed under complementation: $A \in \mathcal{A} \Rightarrow X \setminus A \in \mathcal{A}$;
3. $\mathcal{A}$ is closed under countable unions: $A_1, A_2, \cdots \in X \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

It follows from the definition that a σ-algebra is also closed under countable intersections. Elements of the σ-algebra are called *measurable sets*. An ordered pair $(X, \mathcal{A})$, where X is a set and $\mathcal{A}$ is a σ-algebra over X, is called a *measurable space*. A function between two measurable spaces is called a *measurable function* if the preimage of every measurable set is measurable. We can characterize σ-algebras in terms of simpler structures, π-systems and Dynkin systems:

Theorem 1.2. $\mathcal{A} \subseteq \mathcal{P}(X)$ is a σ-algebra if and only if it is both a π-system and a λ-system.

- A π-system (or p-system) P is a collection of subsets of X that is closed under finitely many intersections;
- a Dynkin system (or λ-system) D on X is a collection of subsets of X that contains X and is closed under complement and under countable unions of *disjoint* subsets:

  1. $X \in D$;
  2. if $A \in D$, then $A^c \in D$;
  3. if $A_1, A_2, A_3, \ldots$ are such that $A_i \cap A_j = \varnothing$ for all $i \neq j$, then $\bigcup_{n=1}^{\infty} A_n \in D$.

  Equivalently, D is a Dynkin system if

  1. $X \in D$;
  2. if $A, B \in D$ and $A \subseteq B$, then $B \setminus A \in D$;
  3. for $A_1 \subset A_2 \subset A_3 \subset \cdots$ we have $\bigcup_{n=1}^{\infty} A_n \in D$.

The statement can be verified by noting that conditions 2 and 3 in the defninition of Dynkin system together with closure under finite intersections imply closure under countable unions.

Why do we care about π-systems and λ-systems? One reason is the Dynkin's π-λ theorem (or the monotone class theorem). It is an essential tool for proving many results about properties of specific σ-algebras.

Theorem 1.3 (Monotone Class Theorem / π-λ Theorem)**.** If P is a π-system and D is a λ-system such that $P \subseteq D$, then $\sigma(P) \subseteq D$.

Often, we want some σ-algebra $\mathcal{A}$ to have some desirable property. To do so, we may collect all sets satisfying some property into a collection D. We may then find some collection $P \subset D$ such that it is closed under intersection, and generates $\mathcal{A}$. If we can demonstrate that D is a λ-system, then we can use the theorem to conclude that $\mathcal{A} = \sigma(P)$ also enjoys the property. One of the most fundamental uses of the π-λ theorem is to show equivalence of separately defined measures or integrals.

Example 1.4. Let $([0,1], \mathcal{B}(\mathbb{R}), \mu_L)$ be the unit interval $[0,1]$ with the Lebesgue measure on Borel sets. Let μ' be another measure on $[0,1]$ satisfying $\mu'([a,b]) = b - a$, and let D be the family of sets S such that $\mu_L(S) = \mu'(S)$. Let $I = \{(a,b), [a,b), (a,b], [a,b] : 0 < a \leq b < 1\}$, and observe that I is closed under finite intersections, that $I \subset D$, and that $\mathcal{B}(\mathbb{R})$ is the σ-algebra generated by I. It may be shown that D is a Dynkin-system. From Dynkin's π-λ theorem it follows that $\sigma(I) = \mathcal{B}(\mathbb{R}) \subset D$, from which we conclude that the Lebesgue measure is unique on $\mathcal{B}(\mathbb{R})$.

Example 1.5. The π-λ theorem can also motivate the use of distribution functions in probability. Recall for random variable $X : \Omega \to \mathbb{R}$, its distribution function is defined as

$$F_X(x) = \mathbb{P}\{X \leq x\}, \quad x \in \mathbb{R}.$$

Recall the measure it induced on $\mathbb{R}$ is

$$\mu(B) = \mathbb{P}\{X^{-1}(B)\}, \quad B \in \mathcal{B}(\mathbb{R}),$$

so the distribution function specifies the measure on $\{(-\infty, x] : x \in \mathbb{R}\}$, which is a π-system. By the same argument as in Example 1.4, if two random variables X and Y equal in distribution ($F_X = F_Y$), then they have the same probability measure on $\mathbb{R}$. Distribution functions thus uniquely characterize random variables.

A similar result of the monotone class theorem for functions also holds.

Theorem 1.6. Let $(X, \mathcal{A})$ be a measurable space. Let P be a π-system that constains X and let $\mathcal{F} = \{f : X \to \mathbb{R}\}$ (similar to the role of a Dynkin system) be a collection of real-valued functions with the following properties:

1. $A \in P \Rightarrow \mathbb{1}_A \in \mathcal{F}$;
2. $f, g \in \mathcal{F} \Rightarrow f + g \in \mathcal{F}$ and $cf \in \mathcal{F}$ for any real number c;
3. if $\{f_n\} \subset \mathcal{F}$ is a sequence of non-negative functions that increase to a bounded function f, then $f \in \mathcal{F}$.

Then $\mathcal{F}$ contains all bounded functions that are measureable with respect to $\sigma(P)$.

Often, P is a generating class of $\mathcal{A}$ (i.e. $\sigma(P) = \mathcal{A}$), so if $\mathcal{F}$ enjoys some desired property, then so does the space of all bounded (real-valued) measurable functions on X.

Proof. The assumption that $X \in P$, together with 2 and 3 imply that $D = \{A \in X : \mathbb{1}_A \in \mathcal{F}\}$ is a λ-system. By 1, $P \subset D$. By the π-λ theorem (Theorem 1.3), $\sigma(P) \subset D$. This means $\mathcal{F}$ contains all indicator functions defined on sets in $\sigma(P)$. 2 then implies that $\mathcal{F}$ contains all simple functions defined with respect to $\sigma(P)$, and then 3 implies that $\mathcal{F}$ contains all bounded measurable functions with respect to $\sigma(P)$. □

In the course of the proof, we used an important approximation result: a function $f : X \to \mathbb{R}_+$ is measurable if and only if there exists a non-decreasing sequence $\{f_n\}_{n \geq 1}$ of simple functions such that $f_n \uparrow f$, i.e. for any $x \in X$

$$f(x) = \lim_{n \to \infty} f_n(x).$$

1.2 Measurable Spaces and Functions

We record here on some properties about measurable spaces and functions:

- *Measurability on generating class*: if $f : (E, \mathcal{E}) \to (F, \mathcal{F})$ and $\mathcal{F} = \sigma(\mathcal{C})$, then f is $\mathcal{E}/\mathcal{F}$-measurable if and only if

$$f^{-1}(B) \in \mathcal{E} \qquad \forall B \in \mathcal{C}.$$

- *Continuous functions are measurable*: if $f : E \to \mathbb{R}$ is continuous, then f is $\mathcal{E}/\mathcal{B}(\mathbb{R})$-measurable where $\mathcal{E} = \mathcal{B}(E)$ is the Borel σ-algebra on E. This is because the preimage of f is open for every open set in $\mathbb{R}$.
- *Products of measurable spaces*: Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be two measurable spaces. The product σ-algebra $\mathcal{E} \otimes \mathcal{F}$ is defined as
$$\mathcal{E} \otimes \mathcal{F} = \sigma(\{A \times B : A \in \mathcal{E}, B \in \mathcal{F}\})$$
and $(E \times F, \mathcal{E} \otimes \mathcal{F})$ is a product measurable space. A special case is when $E = F = \mathbb{R}$, where A and B take the form of open intervals, so that the generating class of $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ is the set of open rectangles. We have $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$. This is because the set of open rectangles is a basis for the topology of $\mathbb{R}^2$.
- *Other examples*
  - ◇ Since $\mathcal{C} = \{(-\infty, z] : z \in \mathbb{R}\}$ is a generating class for the Borel σ-algebra of $\mathbb{R}$, a real-valued function $f : E \to \mathbb{R}$ is $\mathcal{E}/\mathcal{B}(\mathbb{R})$-measurable if and only if $f^{-1}((-\infty, z]) = \{x \in E : f(x) \leq z\} \in \mathcal{E}$ for any $z \in \mathbb{R}$.
  - ◇ Composite functions are measurable: for
$$(E, \mathcal{E}) \xrightarrow{f} (F, \mathcal{F}) \xrightarrow{g} (G, \mathcal{G})$$
the function $g \circ f$ is $\mathcal{E}/\mathcal{G}$-measurable.
  - ◇ An indicator function $\mathbb{1}_A : (E, \mathcal{E}) \to \{0, 1\}$ is measurable if and only if $A \in \mathcal{E}$.
  - ◇ If $\{f_n\}_{n \geq 1}$ is a sequence of $\mathcal{E}/\mathcal{B}(\mathbb{R})$-measurable functions, then
$$\inf_n f_n \qquad \text{and} \qquad \sup_n f_n$$
are measurable as well. Consider the latter. It suffices to show that $\{x \in E : \sup_n f_n(x) \leq z\} \in \mathcal{E}$ for any $z \in \mathbb{R}$. But
$$\{x \in E : \sup_n f_n(x) \leq z\} = \bigcap_{n \geq 1} \{x \in E : f_n(x) \leq z\}$$
and since each set in the intersection is in $\mathcal{E}$ and $\mathcal{E}$ is closed under countable intersections, we have the desired result. In a similar fashion,
$$\limsup_n f_n = \lim_{N \to \infty} \sup_{n \geq N} f_n = \inf_{N \geq 1} \sup_{n \geq N} f_n \quad \text{and} \quad \liminf_n f_n = \lim_{N \to \infty} \inf_{n \geq N} f_n = \sup_{N \geq 1} \inf_{n \geq N} f_n$$
are all measurable.
  - ◇ As stated above, a positive real-valued function $f : E \to \mathbb{R}_+$ is measurable if and only if there exists a non-decreasing sequence $\{f_n\}_{n \geq 1}$ of simple functions such that $f_n \uparrow f$.

1.3 Measures

Definition 1.7. Let $(E, \mathcal{E})$ be a measurable space. A mapping $\mu : \mathcal{E} \to [0, +\infty]$ is called a *measure* on $(E, \mathcal{E})$ if

1. $\mu(\varnothing) = 0$;
2. for any collection of pairwise disjoint sets $A_1, A_2, \ldots \in \mathcal{E}$, i.e. $A_i \cap A_j = \varnothing \quad \forall i \neq j$, one has
$$\mu\left(\bigcup_{i \geq 1} A_i\right) = \sum_{i \geq 1} \mu(A_i).$$

When $\mu(E) < \infty$, the measure μ is said to be *finite*.

Example 1.8. Below are some basic examples.

- *Point mass/Dirac measure at x*: let $x \in E$. Define $\mu_x(A)$ for any $A \in \mathcal{E}$ as

$$\mu_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

- *Counting measure*: Let $D \subset E$ be countable and for any $A \in \mathcal{E}$ define $\mu(A)$ as

$$\mu(A) = |A \cap D| = \sum_{x \in D} \delta_x(A).$$

- *Discrete measure*: let $D \subset E$ be countable and $m : D \to \mathbb{R}_+$. Define $\nu(A)$ for any $A \in \mathcal{E}$ as

$$\nu(A) = \sum_{x \in D} m(x)\delta_x(A).$$

  Note the relationship between counting measure and discrete measure: for $A \in \mathcal{E}$, $\mu(A) = 0$ implies that $\nu(A) = 0$. We say that ν is *absolutely continuous* with respect to μ. We write $\nu \ll \mu$.

1.4 Lebesgue Integration

Let $(E, \mathcal{E}, \mu)$ be a measure space. We consider how to define integration $\mu(f)$ for measurable functions $f : E \to \mathbb{R}$. For a simple function like

$$f = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i},$$

where $a_i \geq 0$ and $\{A_1, \ldots, A_n\}$ is a partition of E into $\mathcal{E}$-sets, we can define the integration of f as

$$\mu(f) := \sum_{i=1}^{n} a_i \mu(A_i).$$

For a non-negative function $f : E \to \mathbb{R}_+$, we know there exists a sequence of simple functions $\{f_n\}_{n \geq 1}$ such that $f_n \uparrow f$. Note that

- since f_n is simple, $\mu(f_n)$ is defined;
- since $f_n \leq f_{n+1}$, we have $\mu(f_n) \leq \mu(f_{n+1})$.

Thus we can defnine $\mu(f)$ as

$$\mu(f) := \lim_{n \to \infty} \mu(f_n).$$

For a general measurable function $f : E \to \mathbb{R}$, we can split it into positive part $f^+ = \max\{f, 0\} = f \vee 0$ and negative part $f^- = -\min\{f, 0\} = -(f \wedge 0)$, so that $f = f^+ - f^-$. Since both $\mu(f^+)$ and $\mu(f^-)$ are defined, we may define

$$\mu(f) := \mu(f^+) - \mu(f^-)$$

provided at least one of the two summands on the right hand is finite. If $\mu(f^+) = \mu(f^-) = +\infty$, then $\mu(f)$ is undefined. Note that f is integrable if and only if $\mu(|f|) < +\infty$.

2 Probability Theory

2.1 Probability Spaces

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a measure space where $\mathbb{P}(\Omega) = 1$. Below are some elementary properties:

- (Monotonicity) For $A, B \in \mathcal{F}$ such that $A \subset B$, one has $\mathbb{P}(A) \leq \mathbb{P}(B)$.
- (Inclusion/exclusion formula) For any $n \geq 1$ and $A_1, \ldots, A_n \in \mathcal{F}$ one has

$$\mathbb{P}\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \ldots + (-1)^{n+1} \mathbb{P}(A_1 \cap \cdots \cap A_n).$$

  Note that for $n = 2$, the formula is $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$. If $A_1, \ldots, A_n$ are pairwise disjoint, then

$$\mathbb{P}\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

- For any collection $A_1, A_2 \ldots$ of sets in $\mathcal{F}$ one has

$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

- (Continuity of $\mathbb{P}$) If $A_n \uparrow A \in \mathcal{F}$, then $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$. If $A_n \downarrow A \in \mathcal{F}$, then $\mathbb{P}(A_n) \downarrow \mathbb{P}(A)$.

  Proof. Let $\{A_n\}_{n\geq 1}$ be non-decreasing and let $B_1 = A_1$ and $B_n = A_n \setminus A_{n-1}$ for $n \geq 2$. We have

$$A_n = \bigcup_{i=1}^{n} B_i, \quad A = \bigcup_{i=1}^{\infty} B_i, \quad \text{and} \quad B_i \cap B_j = \varnothing \quad \forall i \neq j,$$

  so that

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(B_i) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} B_i\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

  If $\{A_n\}_{n\geq 1}$ is non-increasing and $A_n \downarrow A$, then $\{A_n^c\}_{n\geq 1}$ is non-decreasing and $A_n^c \uparrow A^c$. From the previous part $\mathbb{P}(A_n^c) \uparrow \mathbb{P}(A^c)$, so that

$$\mathbb{P}(A_n) = \mathbb{P}(\Omega \setminus A_n^c) = 1 - \mathbb{P}(A_n^c) \downarrow 1 - \mathbb{P}(A^c) = \mathbb{P}(A).$$

  □

- (Continuity implies countable additivity) The above point demonstrated that countable additivity implies continuity. Here we prove a converse: if $\mathbb{P}$ is continuous along monotone sequences, i.e. $A_n \downarrow \varnothing \Rightarrow \mathbb{P}(A_n) \downarrow 0$, then $\mathbb{P}$ is countably additive.

  Proof. Let $\{B_i\}_{i=1}^{\infty}$ be a collection of pairwise disjoint sets in $\mathcal{F}$, let $B = \bigcup_{i=1}^{\infty} B_i$, and let $C_n = \bigcup_{i=n}^{\infty} B_i$. Then $C_n \supset C_{n+1}$ and $C_n \downarrow \varnothing$. We have

$$\begin{aligned}\mathbb{P}(C_n) &= \mathbb{P}\left(\bigcup_{i=n}^{\infty} B_i\right) = \mathbb{P}\left[B \setminus \left(\bigcup_{i=1}^{n-1} B_i\right)\right] = \mathbb{P}(B) - \mathbb{P}\left(\bigcup_{i=1}^{n-1} B_i\right) \\ &= \mathbb{P}(B) - \sum_{i=1}^{n-1} \mathbb{P}(B_i).\end{aligned}$$

Let $n \to \infty$ we have by continuity

$$0 = \lim_{n\to\infty} \mathbb{P}(C_n) = \mathbb{P}(B) - \sum_{i=1}^{\infty} B_i,$$

as desired. □

2.2 Random Variables

If $X : \Omega \to \mathbb{R}$ is a real-valued random variable, then

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}\{X^{-1}(A)\}, \qquad A \in \mathcal{B}(\mathbb{R})$$

defines a probability measure on $\mathbb{R}$. Since $\mathcal{C} = \{(-\infty, x] : x \in \mathbb{R}\}$ is a π-system generating $\mathcal{B}(\mathbb{R})$, $\mathbb{P}_X$ is identified by the function

$$x \mapsto F_X(x) = \mathbb{P}_X((-\infty, x]) = \mathbb{P}\{X \le x\},$$

known as the distribution function. The distribution function satisfies three properties:

1. $x < y \Rightarrow F_X(x) \le F_X(y)$, i.e. F_X is non-decreasing;
2. $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to+\infty} F_X(x) = 1$;
3. F_X is right continuous, i.e. for any $x \in \mathbb{R}$ one has $\lim_{\Delta x \downarrow 0} F_X(x + \Delta x) = F_X(x)$.

The converse is also true: if a function F satisfies the three properties, then there exists a unique probability measure $\mathbb{P}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ admitting F as its distribution function, i.e.

$$\mathbb{P}((-\infty, x]) = F(x) \quad \forall x \in \mathbb{R}.$$

Given a distribution function F, there is also a corresponding probability space and a random variable: take $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}[0, 1]$, $\mathbb{P} = \mu_L$ and $X(\omega) = \inf\{z \in \mathbb{R} : F(z) \ge \omega\}$, namely for $\omega \in [0, 1]$ on the vertical y-axis, draw a horizontal line and cross the graph of F; the x-value of the intersection is the value of $X(\omega)$. Roughly $F(x) = \omega$, so that $\mathbb{P}\{X \le x\} = \mu_L([0, \omega]) = \omega = F(x)$.

Now let's talk about densities. Recall the Radon-Nikodym theorem:

Theorem 2.1 (Radon-Nikodym theorem)**.** Let μ be a σ-finite measure on $(E, \mathcal{E})$ and let ν be another measure such that $\nu \ll \mu$. Then there exists a measurable function $f : E \to \mathbb{R}_+$ such that for any $A \in \mathcal{E}$ one has

$$\nu(A) = \int_A f d\mu. \tag{1}$$

The function f is almost-everywhere unique with respect to μ, in the sense that if g is another function that satisfies Eq. (1) then $\mu\{f \ne g\} = 0$. It is called the *Radon-Nikodym derivative* and is denoted by $\frac{d\nu}{d\mu}$.

- Let $D \subset \mathbb{R}$ be a countable set and note that the counting measure μ is σ-finite. If $\mathbb{P}_X \ll \mu$, then the probability distribution of X is discrete and the Radon-Nikodym derivative of $\mathbb{P}_X$ with respect to μ is

  $$f(x) = \frac{d\,\mathbb{P}_X}{d\mu}(x) = \mathbb{P}\{X = x\},$$

  the probability mass function of X. We have $f(x) \ge 0$ for any $x \in \mathbb{R}$ and $\sum_{x\in D} f(x) = 1$.

- The Lebesgue measure μ_L on $\mathbb{R}$ is σ-finite. If $\mathbb{P}_X \ll \mu_L$, then
$$f(x) = \frac{d\,\mathbb{P}_X}{d\mu_L}(x)$$
is the probability density function of X. It can be seen that $f(x) \geq 0$ for any $x \in \mathbb{R}$ and $\int_{\mathbb{R}} f(x)dx = 1$. So in this course, we regard the "probability density function" taught in elementary probability courses, as the Radon-Nikodym derivative of two measures. This approach gives a precise meaning of the concept of densitiy functions, and it unifies both the discrete case and the continuous case.

Example 2.2 (Examples of Densities)**.** Here are common probability densities in terms of Radon-Nikodym derivatives.

- (Degenerate distribution at x_0) For $X = x_0$, we have $\mathbb{P}_X(A) = \delta_{x_0}(A)$ for $A \in \mathcal{B}(\mathbb{R})$. Its distribution function is
$$F_X(x) = \mathbb{P}_X((-\infty, x]) = \mathbb{1}_{[x_0,+\infty)}(x).$$
- (Poisson distribution) Let $\nu = \sum_{n\in\mathbb{N}} \delta_n$ be the counting measure on $\mathbb{N}$. We say X has Poisson distribution with mean $\lambda > 0$ if $\mathbb{P}_X \ll \nu$ and
$$\mathbb{P}\{X = x\} = \frac{d\,\mathbb{P}_X}{d\nu}(x) = \frac{\lambda^x}{x!}e^{-\lambda}$$
so for any $A \in \mathcal{B}(\mathbb{R})$
$$\mathbb{P}_X(A) = \int_A \frac{d\,\mathbb{P}_X}{d\nu}d\nu = \sum_{k\in A\cap\mathbb{N}} \frac{\lambda^k}{k!}e^{-\lambda}.$$
- (Binomial distribution) Let $p \in (0, 1)$ and let $\nu_n = \sum_{i=0}^n \delta_i$ be the counting measure on $\{0, 1, \ldots, n\}$. A random variable X is said to have the binomial distribution with parameters (n, p) if $\mathbb{P}_X \ll \nu_n$ and
$$\mathbb{P}\{X = x\} = \frac{d\,\mathbb{P}_X}{d\nu_n}(x) = \binom{n}{x}p^x(1-p)^{n-x}$$
so for any $A \in \mathcal{B}(\mathbb{R})$
$$\mathbb{P}_X(A) = \int_A \frac{d\,\mathbb{P}_X}{d\nu_n}d\nu_n = \sum_{k\in A\cap\{0,\ldots,n\}} \binom{n}{k}p^k(1-p)^{n-k}.$$
- (Gamma distribution) A random variable X is said to have the gamma distribution with shape $\alpha > 0$ and rate $\beta > 0$ if $\mathbb{P}_X \ll \mu_L$ and
$$\frac{d\,\mathbb{P}_X}{d\mu_L}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}\mathbb{1}_{(0,\infty)}(x),$$
so for any $A \in \mathcal{B}(\mathbb{R})$
$$\mathbb{P}_X(A) = \int_A \frac{d\,\mathbb{P}_X}{d\mu_L}d\mu_L = \int_{A\cap(0,\infty)} \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}dx.$$
Recall the gamma function is defined as
$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$$
so that $\mathbb{P}_X((0,\infty)) = 1$.
- (Gaussian distribution) A random variable X is said to have the Gaussian distribution with mean μ and variance σ^2 if $\mathbb{P}_X \ll \mu_L$ and
$$\frac{d\,\mathbb{P}_X}{d\mu_L}(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \forall x \in \mathbb{R}$$
so for any $A \in \mathcal{B}(\mathbb{R})$
$$\mathbb{P}_X(A) = \int_A \frac{d\,\mathbb{P}_X}{d\mu_L}d\mu_L = \frac{1}{\sigma\sqrt{2\pi}}\int_A e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx.$$

- (Uniform distribution) for the uniform distribution on interval $[a, b]$, its density is

$$\frac{d\,\mathbb{P}_X}{d\mu_L}(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$$

so that

$$\mathbb{P}_X(A) = \int_A \frac{d\,\mathbb{P}_X}{d\mu_L} d\mu_L = \int_{A\cap[a,b]} \frac{1}{b-a} dx.$$

2.3 Expectations

Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 2.3. The *expectation* of X is defined as

$$\mathbb{E}X = \int_\Omega X d\,\mathbb{P}.$$

Below are several important theorems.

- **Fatou's lemma.** If $\{X_n\}_{n\geq 1}$ is such that $\mathbb{P}\{X_n \geq 0\} = 1$ for all $n \geq 1$ then

$$\mathbb{E}\liminf X_n \leq \liminf \mathbb{E}X_n.$$

- **Monotone convergence theorem.** If $\{X_n\}_{n\geq 1}$ is such that $\mathbb{P}\{X_n \geq 0\} = 1$ for all $n \geq 1$ and $X_n \uparrow X$ a.s., then

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

- **Dominated convergence theorem.** Let $\{X_n\}_{n\geq 1}$ be such that $\mathbb{P}\{X_n \leq Y\} = 1$ for all $n \geq 1$, where $Y \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$, and suppose $X_n \xrightarrow{a.s.} X$. Then X_n's and X are in $\mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ and

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

- **Bounded convergence theorem** Let $\{X_n\}_{n\geq 1}$ be such that $\mathbb{P}\{X_n \leq b\} = 1$ for all $n \geq 1$ and some $b < \infty$. If $X \xrightarrow{a.s.} X$, then

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

Theorem 2.4. For any measurable function $f : \mathbb{R} \to \mathbb{R}_+$, we have

$$\mathbb{E}f(X) = \int_\Omega f(X) d\,\mathbb{P} = \int_{\mathbb{R}} f d\,\mathbb{P}_X.$$

On the other hand, if there exists a probability measure $\mathbb{P}'$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that

$$\int_{\mathbb{R}} f d\,\mathbb{P}_X = \int_{\mathbb{R}} f d\,\mathbb{P}'$$

for any measurable function $f : \mathbb{R} \to \mathbb{R}_+$, then $\mathbb{P}' = \mathbb{P}_X$.

Proof. The first statement is a change of measure from $\mathbb{P}$ to $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$. For the second statement, take $f = \mathbb{1}_A$ where $A \in \mathcal{B}(\mathbb{R})$. □

If we only know that the condition in Theorem 2.4 holds only for non-negative, bounded and continuous functions, instead of all measurable functions, then we can still get $\mathbb{P}_X = \mathbb{P}'$. This is becuase we can approximate $\mathbb{1}_{(a,b)}$ by some sequence $\{f_n\}_{n\geq 1}$ of such functions. By the monotone convergence theorem,

$$\mathbb{P}_X((a,b)) = \lim_{n\to\infty} \mathbb{P}_X(f_n) = \lim_{n\to\infty} \mathbb{P}'(f_n) = \mathbb{P}'((a,b)),$$

so $\mathbb{P}_X$ and $\mathbb{P}'$ agree on the p-system $\mathcal{C} = \{(a,b) : -\infty < a < b < \infty\}$, so $\mathbb{P}_X$ and $\mathbb{P}'$ agree on $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$.

Example 2.5. Here we show several examples of random variables with infinite or undefined expectation.

- If $X \in [1,\infty)$ has density $f(x) = 1/x^2$ on $[1,\infty)$, then the expectation is
$$\mathbb{E}X = \int_1^\infty x \cdot \frac{1}{x^2} dx = \int_1^\infty \frac{1}{x} dx = +\infty.$$
- Let X be a random variable that is equal to 2^n with probability 2^{-n} (for positive integer n). Then
$$\mathbb{E}X = \sum_{n=1}^\infty 2^{-n} \cdot 2^n = \sum_{n=1}^\infty 1 = +\infty.$$
- Cauchy distribution. The Cauchy distribution has probability density function
$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma}\left[\frac{\gamma^2}{(x-x_0)^2 + \gamma^2}\right]$$
where x_0 is the location parameter and γ is the scale parameter. The expectation of a Cauchy distribution is undefined. This is because, for an arbittrary $a \in \mathbb{R}$,
$$\int_{-\infty}^\infty x f(x) dx = \int_{-\infty}^a x f(x) dx + \int_a^\infty x f(x) dx,$$
and the left term is $-\infty$ while the right term is $+\infty$, so the mean does not exist at all. Various results in probability theory about expected values, such as the strong law of large numbers, fail to hold for the Cauchy distribution. It is like a distribution with fat tails on both sides, and it oscillates between large positive values and large negative values.

  The second raw moment $\mathbb{E}X^2$ does exist and is $+\infty$. Similarly, higher even raw moments exist and are all $+\infty$, but all odd raw moments do not exist. Since the mean does not exist, the variance — which is the second central moment — is likewise non-existent.

If X is such that $\mathbb{E}|X|^n < \infty$, then $\mathbb{E}X^n = \int_{\mathbb{R}} x^n d\,\mathbb{P}_X$ is the n-th moment of X. In case $X \geq 0$ a.s., we can obtain a formula for this:

$$X^n = \int_0^X nx^{n-1} dx = \int_0^\infty \mathbb{1}_{\{0<x<X\}} nx^{n-1} dx$$

so that

$$\mathbb{E}X^n = \int_\Omega \int_0^\infty \mathbb{1}_{\{0<x<X\}} nx^{n-1} dx\, d\,\mathbb{P} = \int_0^\infty nx^{n-1} \int_\Omega \mathbb{1}_{\{0<x<X\}} d\,\mathbb{P} dx = \int_0^\infty nx^{n-1}\,\mathbb{P}\{X > x\} dx.$$

In particular, if X takes values in $\mathbb{N}$, then

$$\mathbb{E}X = \sum_{n=1}^\infty \mathbb{P}\{X > n\}.$$

In the lecture Markov's inequality and Jensen's inequality ($\mathbb{E}f(X) \geq f(\mathbb{E}X)$ for convex f) are also mentioned. If $f : \mathbb{R} \to \mathbb{R}_+$ is increasing, then for any $b \in \mathbb{R}$, one has

$$\mathbb{P}\{X > b\} = \int_b^\infty d\,\mathbb{P}_X \leq \int_b^\infty \frac{f(x)}{f(b)} d\,\mathbb{P}_X \leq \frac{1}{f(b)} \int_{\mathbb{R}} f(x) d\,\mathbb{P}_X = \frac{1}{f(b)} \mathbb{E}f(X).$$

2.3.1 Laplace transforms and characteristic functions

For a non-negative random variable X, one can define the Laplace transform of X as

$$\hat{\mathbb{P}}_X(t) = \mathbb{E}e^{-tX} = \int_{\mathbb{R}_+} e^{-tx} d\,\mathbb{P}_X \quad \forall t \in [0,\infty). \tag{2}$$

In terms of the usual definition of Laplace transform for functions ($\mathcal{L}\{f\}(s) = \int_0^\infty e^{-st} f(t)dt$), Eq. (2) can be seen as the Laplace transform of the probability density of X. Or in terms of Laplace transform for measures ($\mathcal{L}\{\mu\}(s) = \int_{[0,\infty)} e^{-st}\mu(dt)$), Eq. (2) can be seen as the Laplace transform of the probability measure $\mathbb{P}_X$. Since $e^{-tX} \in [0,1]$, its expectations is also in $[0,1]$. Thus $\hat{\mathbb{P}}_X \in [0,1]$ just as $\mathbb{P}_X \in [0,1]$. The importance of Laplace transforms comes from the fact that they uniquely identify probability measures. Thus, for example, if we know the Laplace transform of some distribution, then to prove a random variable Y has that distribution too, we may prove that its Laplace transform has the same form.

Theorem 2.6. If X and Y are non-negative random variables, then

$$\hat{\mathbb{P}}_X(t) = \hat{\mathbb{P}}_Y(t) \quad \forall t \geq 0 \quad \iff \mathbb{P}_X = \mathbb{P}_Y.$$

Example 2.7. If X and Y are independent and non-negative random variables, then

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

This formula can help us determine the probability distribution of $X+Y$. For example, if $X \sim G(\alpha_1,\beta)$ and $Y \sim G(\alpha_2,\beta)$, then

$$\hat{\mathbb{P}}_X(t) = \left(\frac{\beta}{\beta+t}\right)^{\alpha_1} \qquad \hat{\mathbb{P}}_Y(t) = \left(\frac{\beta}{\beta+t}\right)^{\alpha_2}.$$

Using this, the Laplace transform of $X+Y$ is

$$\hat{\mathbb{P}}_{X+Y}(t) = \hat{\mathbb{P}}_X(t)\hat{\mathbb{P}}_Y(t) = \left(\frac{\beta}{\beta+t}\right)^{\alpha_1+\alpha_2},$$

so using Theorem 2.6 we can conclude that $X+Y$ is distributed as $G(\alpha_1+\alpha_2,\beta)$.

Example 2.8. From Eq. (2), the first derivative of $\hat{\mathbb{P}}_X(t)$ is

$$\frac{d}{dt}\hat{\mathbb{P}}_X(t) = -\mathbb{E}Xe^{-tX}$$

so that

$$\frac{d}{dt}\hat{\mathbb{P}}_X(t)\bigg|_{t=0} = -\mathbb{E}X.$$

One can show that

$$(-1)^n \frac{d^n}{dt^n}\hat{\mathbb{P}}_X(t)\bigg|_{t=0} = \mathbb{E}X^n.$$

If X is real-valued instead of non-negative valued, it is convenient to use the characteristic function or Fourier transform instead of the Laplace transform:

$$\varphi_X(t) = \mathbb{E}e^{itX} = \int_{\mathbb{R}} e^{itx} d\,\mathbb{P}_X = \int_{\mathbb{R}} \cos(tx) d\,\mathbb{P}_X + i\int_{\mathbb{R}} \sin(tx) d\,\mathbb{P}_X$$

for any $t \in \mathbb{R}$. Since $|e^{itx}| = 1$ we have $|\varphi_X(t)| \leq 1$. It can be shown that

$$\varphi_X(t) = \varphi_Y(t) \quad \forall t \in \mathbb{R} \quad \iff \mathbb{P}_X = \mathbb{P}_Y.$$

Just as the distribution function

$$F_X(x) = \mathbb{E}\left[\mathbb{1}_{\{X \le x\}}\right]$$

completely determines the behavior and properties of the probability distribution of the random variable X, the characteristic function

$$\varphi_X(t) = \mathbb{E}\left[e^{itX}\right]$$

also completely determines the behavior and properties of the probability distribution of the random variable X. The two approaches are equivalent in the sense that knowing one of the functions it is always possible to find the other, yet they provide different insights for understanding the features of the random variable. However, in particular cases, there can be differences in whether these functions can be represented as expressions involving simple standard functions.

If a random variable admits a density function, then the characteristic function is its dual, in the sense that each of them is a Fourier transform of the other. The characteristic function approach is particularly useful in analysis of linear combinations of independent random variables: a classical proof of the Central Limit Theorem uses characteristic functions and Lévy's continuity theorem. Another important application is to the theory of the decomposability of random variables.

2.4 Uniform Integrability

Proposition 2.9. A real-valued random variable X is in $\mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ if and only if

$$\lim_{K\to\infty} \mathbb{E}|X|\mathbb{1}_{(K,+\infty)}(|X|) = \lim_{K\to\infty} \int_{x:|x|>K} |x| d\,\mathbb{P}_X = 0. \tag{3}$$

Proof. First note that $|X|\mathbb{1}_{(K,+\infty)}(|X|) \le |X|$ and $\{|X| > K\} \downarrow \varnothing$ as $K \to \infty$. If $X \in \mathcal{L}_1$, then one can apply the dominated convergence theorem to obtain

$$\lim_{K\to\infty} \mathbb{E}|X|\mathbb{1}_{(K,+\infty)}(|X|) = \mathbb{E} \lim_{K\to\infty} \left[|X|\mathbb{1}_{(K,+\infty)}(|X|)\right] = 0.$$

On the other hand, if the above condition holds true, then one can use the bound

$$|X| = |X|\mathbb{1}_{[0,K]}(|X|) + |X|\mathbb{1}_{(K,+\infty)}(|X|) \le K + |X|\mathbb{1}_{(K,+\infty)}(|X|)$$

to deduce that

$$\mathbb{E}|X| \le K + \mathbb{E}|X|\mathbb{1}_{(K,+\infty)}(|X|) < \infty,$$

thus proving that X is in $\mathcal{L}_1$. □

The above proposition motivates the definition of uniform integerability.

Definition 2.10. A collection of real-valued random variable $\mathcal{C}$ is *uniformly integrable* if

$$\lim_{K\to\infty} \sup_{X\in\mathcal{C}} \mathbb{E}|X|\mathbb{1}_{(K,+\infty)}(|X|) = 0 \tag{4}$$

- From Proposition 2.9, $X \in \mathcal{L}_1$ if and only if for any $\varepsilon > 0$ there exists K_0 such that for any $K > K_0$

  $$\mathbb{E}|X|\mathbb{1}_{(K,+\infty)}(|X|) < \varepsilon.$$

  So another definition of uniform integrability would be: a collection of real-valued random variable $\mathcal{C}$ is *uniformly integrable* if for any $\varepsilon > 0$ there exists $K > 0$ such that

  $$\sup_{X\in\mathcal{C}} \mathbb{E}|X|\mathbb{1}_{(K,+\infty)}(|X|) < \varepsilon. \tag{5}$$

- From Eq. (5) we can see that if $\mathcal{C}$ is uniformly integrable, then it is also $\mathcal{L}_1$-bounded. However, uniform integrability is a more stringent concept than $\mathcal{L}_1$-boundedness. A classical example to illustrate this is the collection $\mathcal{C} = \{X_n\}_{n=1}^{\infty}$ of random variables where $X_n = n\mathbb{1}_{(0,1/n)}$ on $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}[0,1], \mu_L)$. It is $\mathcal{L}_1$-bounded since $\mathbb{E}|X_n| = n \cdot (1/n) = 1$ for any X_n, but for any $K > 0$ and $n > K$ one has $\mathbb{E}|X_n|\mathbb{1}_{(K,+\infty)}(|X_n|) = n \cdot (1/n) = 1$ so that $\mathcal{C}$ is not uniformly integrable.
- Here are two sufficient conditions for uniform integrability:
  - If $\mathcal{C}$ is such that $|X| \leq Z$ for some integrable random variable Z, then $\mathcal{C}$ is uniformly integrable.
  - If $\mathcal{C}$ is a collection of $\mathcal{L}_p$-bounded random variables for some $p > 1$, then $\mathcal{C}$ is uniformly integrable.

  To prove the first statement, note $\mathbb{E}|X|\mathbb{1}_{(K,+\infty)}(|X|) \leq \mathbb{E}|Z|\mathbb{1}_{(K,+\infty)}(|Z|) < \varepsilon$ for some suitably chosen $K > 0$, since Z is integrable. Hence $\mathcal{C}$ is uniformly integrable. The second condition means that there exists $M > 0$ such that for any $X \in \mathcal{C}$ one has $\mathbb{E}|X|^p < M$. Take $v \geq K > 0$. Using $p > 1 \Rightarrow v^{1-p} \leq K^{1-p} \Rightarrow v \leq K^{1-p}v^p$, we have

$$\mathbb{E}|X|\mathbb{1}_{(K,+\infty)}(|X|) \leq K^{1-p}\mathbb{E}|X|^p\mathbb{1}_{(K,+\infty)}(|X|) \leq K^{1-p}M < \varepsilon$$

  where the last inequality follows from taking K large enough.

Theorem 2.11. A collection $\mathcal{C}$ of random variables is uniformly integrable if and only if it is $\mathcal{L}_1$ bounded and for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for any $H \in \mathcal{F}$ with $\mathbb{P}(H) < \delta$, one has

$$\sup_{X \in \mathcal{C}} \mathbb{E}|X|\mathbb{1}_H < \varepsilon.$$

2.5 Information and Determinability

The σ-algebra generated by X, $\sigma(X)$, is the smallest σ-algebra with respect to which X is measurable. It is exactly $\sigma(X) = \{X^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\}$.

Theorem 2.12. Y is measurable with respect to $\sigma(X)$ if and only if there is a deterministic function $f : \mathbb{R} \to \mathbb{R}$ such that

$$Y = f(X).$$

2.6 Independence

Definition 2.13. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The sub-σ-algebras $\mathcal{G}_1, \ldots, \mathcal{G}_n$ of $\mathcal{F}$ are independent if for any $A_i \in \mathcal{G}_i$,

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \prod_{i=1}^{n} \mathbb{P}(A_i).$$

We can characterize independence in terms of p-systems.

Theorem 2.14. Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be sub-σ-algebras of $\mathcal{F}$ and suppose $\mathcal{C}_1$ and $\mathcal{C}_2$ are p-systems that generate $\mathcal{G}_1$ and $\mathcal{G}_2$, namely

$$\sigma(\mathcal{C}_i) = \mathcal{G}_i \qquad i = 1, 2.$$

Then $\mathcal{G}_1$ and $\mathcal{G}_2$ are independent if and only if $\mathcal{C}_1$ and $\mathcal{C}_2$ are, i.e.

$$\mathbb{P}(C_1 \cap C_2) = \mathbb{P}(C_1)\mathbb{P}(C_2) \quad \forall C_1 \in \mathcal{C}_1, C_2 \in \mathcal{C}_2.$$

Two random variables X_1 and X_2 are independent if $\sigma(X_1)$ and $\sigma(X_2)$ are independent. They are independent if and only if $\mathbb{E} f_1(X_1) f_2(X_2) = \mathbb{E} f_1(X_1) \mathbb{E} f_2(X_2)$ for all positive measurable functions f_1 and f_2, if and only if their joint distribution is the product of their marginal distributions, i.e. $F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1) F_{X_2}(x_2)$. We can also extend these (equivalent) definitions to more than two random variables.

If X_1 and X_2 are independent, and f_1 and f_2 are measurable functions, then $Y_1 = f_1(X_1)$ and $Y_2 = f_2(X_2)$ are independent. This follows from the fact that $\sigma(Y_i) \subset \sigma(X_i)$ for $i = 1, 2$, so independence of $\sigma(X_1)$ and $\sigma(X_2)$ implies independence of $\sigma(Y_1)$ and $\sigma(Y_2)$.

2.7 Convolutions

In probability theory, the probability distribution of the sum of two or more independent random variables is the *convolution* of their individual distributions. The term is motivated by the fact that the probability density function of a sum of independent random variables is the convolution of their corresponding probability density functions.

Let X and Y be independent random variables taking values in $\mathbb{R}$, with $\mathbb{P}_X$ and $\mathbb{P}_Y$ denoting their respective probability distributions. Moreover, with $H \in \mathcal{B}(\mathbb{R})$ define $B = \{(x, y) : x + y \in H\} \subset \mathbb{R}^2$, one has

$$\begin{aligned}\mathbb{P}_{X+Y}(H) = \mathbb{P}\{X + Y \in H\} &= \int_B d\,\mathbb{P}_{X,Y} = \int_B d\,\mathbb{P}_X\, d\,\mathbb{P}_Y \\ &= \int_{\mathbb{R}} \int_{\{y : x+y \in H\}} d\,\mathbb{P}_Y\, d\,\mathbb{P}_X \\ &= \int_{\mathbb{R}} \mathbb{P}_Y\{H - x\} d\,\mathbb{P}_X.\end{aligned}$$

Given two probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ on $\mathbb{R}$, the convolution of $\mathbb{P}_1$ and $\mathbb{P}_2$ is a new probability measure defined as

$$\mathbb{P}_1 * \mathbb{P}_2(H) = \int_{\mathbb{R}} \mathbb{P}_1\{H - x\} d\,\mathbb{P}_2 \quad \forall H \in \mathcal{B}(\mathbb{R}).$$

Hence if X and Y are independent then

$$\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y = \mathbb{P}_Y * \mathbb{P}_X.$$

- If $\mathbb{P}_X \ll \lambda$ and $\mathbb{P}_Y \ll \mu$ with $p = d\,\mathbb{P}_X / d\lambda$, $q = d\,\mathbb{P}_Y / d\mu$, then

$$\mathbb{P}_{X+Y}(H) = \int_{\mathbb{R}} p(x) \int_H q(y - x) d\mu d\lambda.$$

- If $H = (-\infty, z]$ then

$$F_{X+Y}(z) = \mathbb{P}\{X + Y \le z\} = \int_{\mathbb{R}} F_Y(z - x) d\,\mathbb{P}_X.$$

- If $\mathbb{P}_X \ll \mu_L$ and $\mathbb{P}_Y \ll \mu_L$ with respective densities $p = d\,\mathbb{P}_X / d\mu_L$, $q = d\,\mathbb{P}_Y / d\mu_L$, then

$$\begin{aligned}F_{X+Y}(z) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} q(y) dy p(x) dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{z} q(y - x) dy p(x) dx \\ &= \int_{-\infty}^{z} \int_{-\infty}^{+\infty} q(y - x) p(x) dx dy.\end{aligned}$$

  This implies that $\mathbb{P}_{X+Y} \ll \mu_L$ and

$$p^*(z) = \frac{d\,\mathbb{P}_{X+Y}}{d\mu_L}(z) = F'_{X+Y}(z) = \int_{-\infty}^{+\infty} q(z - x) p(x) dx.$$

Example 2.15. Using the convolution formula, we can determine distributions of sum of common independent random variables. For example, calculations show that

(1) The sum of two uniformly distributed random variables on $[0, 1]$ with density $\mathbb{1}_{[0,1]}$ has density $z\mathbb{1}_{[0,1]}(z) + (2 - z)\mathbb{1}_{(1,2]}(z)$, the so-called triangular distribution;

(2) The sum of two independent Poisson random variables with parameter λ_1 and λ_2 is another Poisson random variable with parameter $\lambda_1 + \lambda_2$;

(3) The sum of two independent Gamma random variables with parameters (α_1, β) and (α_2, β) is another Gamma random variable with parameter $(\alpha_1 + \alpha_2, \beta)$;

(4) See more examples at: https://en.wikipedia.org/wiki/List_of_convolutions_of_probability_distributions.

2.8 Borel-Cantelli Lemmas

Kolmogorov's 0-1 law states that an tail event corresponding to a sequence of *independent* variables will either almost surely happen or almost surely not happen; that is, the probability of such an event occurring is zero or one.

Definition 2.16. Let $\{\mathcal{G}_n\}_{n\geq 1}$ be a collection of sub-σ-algebras of $\mathcal{F}$ and $\mathcal{T}_n = \sigma\left(\bigcup_{m\geq n} \mathcal{G}_m\right)$. Then

$$\mathcal{T} = \bigcap_{n\geq 1} \mathcal{T}_n$$

is called the *tail* σ-algebra.

A typical case is where $\{X_n\}_{n\geq 1}$ is a sequence of random variables:

$$\mathcal{G}_n = \sigma(X_n) \Rightarrow \mathcal{T}_n = \sigma(X_n, X_{n+1}, \ldots) \Rightarrow \mathcal{T} = \bigcap_{n\geq 1} \sigma(X_n, X_{n+1}, \ldots).$$

Tail events are precisely those events whose occurrence can still be determined if an arbitrarily large but finite initial segment of the $\{X_n\}_{n\geq 1}$ are removed. Examples are

$$\left\{\lim_{n\to\infty} S_n/n = 0\right\}, \left\{\sum_{n=1}^{\infty} X_n \text{ converges}\right\}, \left\{\lim_{n\to\infty} X_n \text{ exists}\right\}, \{X_n \in B \text{ i.o.}\}.$$

On the other hand, events like $\{S_n \in B \text{ i.o.}\}$ and $\{\limsup_n S_n > b\}$ are not in $\mathcal{T}$ since they depend on all the $X_1, X_2, \ldots$.

Theorem 2.17 (Kolmogorov's 0-1 law)**.** Let $\{\mathcal{G}_n\}_{n\geq 1}$ be a sequence of independent sub-σ-algebras of $\mathcal{F}$ and $\mathcal{T}$ the corresponding tail σ-algebra. Then

$$H \in \mathcal{T} \quad \Rightarrow \quad \mathbb{P}(H) \in \{0, 1\}.$$

The Borel-Cantelli lemmas provide sufficient conditions for almost sure convergence. Recall the following definitions:

Definition 2.18. Let $\{A_n\}_{n\geq 1}$ be a sequence of events in $\mathcal{F}$. We define

$$\limsup_{n\to\infty} A_n = \bigcap_{N\geq 1} \bigcup_{n\geq N} A_n = \{A_n \text{ occur i.o.}\},$$
$$\liminf_{n\to\infty} A_n = \bigcup_{N\geq 1} \bigcap_{n\geq N} A_n = \{A_n \text{ occur for all but finitely many } n\text{'s}\}.$$

It is easy to check that

$$\limsup_{n\to\infty} A_n = \left\{ \omega \in \Omega : \sum_{n\geq 1} \mathbb{1}_{A_n}(\omega) = +\infty \right\}. \tag{6}$$

The Borel-Cantelli lemmas provide us sufficient conditions to evaluate $\mathbb{P}$ on such limits of events. From this perepective, the lemmas can be seen as providing conditions to determine if $\mathbb{P}(H) = 0$ or $\mathbb{P}(H) = 1$ in Kolmogorov's 0-1 law.

Lemma 2.19 (First Borel–Cantelli lemma)**.** Let $\{A_n\}_{n\geq 1}$ be a sequence of events in $\mathcal{F}$ such that $\sum_{n\geq 1} \mathbb{P}(A_n) < \infty$. Then

$$\mathbb{P}\left[\limsup_{n\to\infty} A_n\right] = 0.$$

Proof. Let $G_N = \bigcup_{n\geq N} A_n$, so that

$$\lim_{N\to\infty} G_N = \bigcap_{N\geq 1} \bigcup_{n\geq N} A_n = \limsup_{n\to\infty} A_n.$$

Since $\limsup A_n \subset G_N$, we have

$$0 \leq \mathbb{P}\left[\limsup_{n\to\infty} A_n\right] \leq \mathbb{P}[G_N] = \mathbb{P}\left[\bigcup_{n\geq N} A_n\right] \leq \sum_{n\geq N} \mathbb{P}[A_n].$$

As $N \to \infty$, the last term goes to zero, so we are able to conclude the proof. □

In view of Eq. (6), the first Borel-Cantelli lemma says

$$\sum_{n\geq 1} \mathbb{P}(A_n) < \infty \quad \Rightarrow \quad \sum_{n\geq 1} \mathbb{1}_{A_n} < \infty \qquad \text{almost surely,}$$

which is straightforward since this amounts to saying that $\mathbb{E}N < \infty \Rightarrow N < \infty$ almost surely, where $N = \sum_{n\geq 1} \mathbb{1}_{A_n}$.

Theorem 2.20 (Second Borel–Cantelli lemma)**.** If $\{A_n\}_{n\geq 1}$ is a sequence of independent events in $\mathcal{F}$ such that $\sum_{n\geq 1} \mathbb{P}(A_n) = \infty$, then

$$\mathbb{P}\left[\limsup_{n\to\infty} A_n\right] = 1.$$

Proof. We prove the equivalent statement that $\mathbb{P}\left[(\limsup A_n)^c\right] = \mathbb{P}\left[\liminf A_n^c\right] = 0$. Using the fact that $1-x \leq e^{-x}$ for any $x > 0$, we obtain

$$\mathbb{P}\left[\bigcap_{n=N}^{N+j} A_n^c\right] = \prod_{n=N}^{N+j} \mathbb{P}(A_n^c) = \prod_{n=N}^{N+j} \{1 - \mathbb{P}(A_n)\} \leq e^{-\sum_{n=N}^{N+j} \mathbb{P}(A_n)}$$

for any $j \geq 1$. Since $\sum_n \mathbb{P}(A_n)$ diverges, we have

$$\mathbb{P}\left[\bigcap_{n\geq N} A_n^c\right] = \lim_{j\to\infty} \mathbb{P}\left[\bigcap_{n=N}^{N+j} A_n^c\right] \leq \lim_{j\to\infty} e^{-\sum_{n=N}^{N+j} \mathbb{P}(A_n)} = 0.$$

Hence

$$\mathbb{P}\left[\liminf A_n^c\right] \leq \sum_{N\geq 1} \left[\bigcap_{n=N}^{N+j} A_n^c\right] = 0.$$

□

Example 2.21. Here we record the example of Riemann zeta function. See the lecture notes for the coin tossing example. Let $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ for $s > 1$. Let X be a random variable taking valued in $\mathbb{N} = \{1, 2, \ldots\}$ such that

$$\mathbb{P}\{X = x\} = \frac{x^{-s}}{\zeta(s)} \mathbb{1}_{\mathbb{N}}(x).$$

Let Π denote the set of primes in $(1, +\infty)$ and for any $p \in \Pi$ define $E_p = \bigcup_{n\geq 1}\{X = np\}$. Then

$$\mathbb{P}(E_p) = p^{-s} \frac{\sum_{n\geq 1} n^{-s}}{\zeta(s)} = p^{-s}.$$

Note that $\{E_p\}_{p\in\Pi}$ is a collection of independent events, since for any $k \geq 2$ and choice of $p_1 \neq \cdots \neq p_k$ we have

$$\mathbb{P}(E_{p_1} \cap \cdots \cap E_{p_k}) = \sum_{n\geq 1} \mathbb{P}\{X = np_1 \cdots p_k\} = (p_1 \cdots p_k)^{-s} = \mathbb{P}(E_{p_1}) \cdots \mathbb{P}(E_{p_k}).$$

We can then derive

$$\mathbb{P}\left[\bigcap_{p\in\Pi} E_p^c\right] = \prod_{p\in\Pi} \mathbb{P}\left[E_p^c\right] = \prod_{p\in\Pi}(1 - p^{-s}).$$

Thus

$$\mathbb{P}\{X = 1\} = \frac{1}{\zeta(s)} = \mathbb{P}\left[\bigcap_{p\in\Pi} E_p^c\right] = \prod_{p\in\Pi}(1 - p^{-s}).$$

This leads to the Euler formula:

$$\zeta(s) = \frac{1}{\prod_{p\in\Pi}(1 - p^{-s})}.$$

Finally, since

$$\sum_{p\in\Pi} \mathbb{P}[E_p] = \sum_{p\in\Pi} p^{-s} < \zeta(s) < +\infty,$$

the first Borel-Cantelli lemma implies that $\mathbb{P}\{E_p \text{ i.o.}\} = 0$. On the other hand, since events in $\{\mathbb{E}_p^c\}$ are independent and

$$\sum_{p\in\Pi} \mathbb{P}[E_p^c] = \sum_{p\in\Pi}(1 - p^{-s}) = +\infty,$$

the second Borel-Cantelli lemma implies that $\mathbb{P}\{E_p^c \text{ i.o.}\} = 1$.

2.9 Convergence

Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of real-valued random variable. Recall the four definitions of convergence:

- *Almost sure convergence*:

  $$\mathbb{P}\left\{\lim_{n\to\infty} X_n = X\right\} = 1.$$

  Using the notion of the limit inferior of a sequence of sets, almost sure convergence can also be defined as follows:

  $$\mathbb{P}\Big(\liminf_{n\to\infty}\{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \varepsilon\}\Big) = 1 \quad \text{for all} \quad \varepsilon > 0.$$

  Almost sure convergence implies convergence in probability (by Fatou's lemma), and hence implies convergence in distribution. It is the notion of convergence used in the strong law of large numbers.

  There is no topology on the space of random variables such that the almost surely convergent sequences are exactly the converging sequences with respect to that topology. In particular, there is no metric of almost sure convergence.

- *Convergence in probability*: for any $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\{|X_n - X| > \varepsilon\} = 0.$$

Convergence in probability implies convergence in distribution. In the opposite direction, convergence in distribution implies convergence in probability when the limiting random variable X is a constant.

Convergence in probability defines a topology on the space of random variables over a fixed probability space.

Convergence in probability to X implies there exists a sub-sequence $\{X_{n_k}\}$ which almost surely converges to X.

- *Convergence in $\mathcal{L}_p$*:

$$\lim_{n\to\infty} \mathbb{E}|X_n - X|^p = 0.$$

Convergence in $\mathcal{L}_p$, for $p \geq 1$, implies convergence in probability (by Markov's inequality). Furthermore, if $p \geq q \geq 1$, convergence in $\mathcal{L}_p$ implies convergence in $\mathcal{L}_q$.

- *Convergence in distribution*:

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for all $x \in \mathbb{R}$ at which F is continuous. Namely,

  - $\mathbb{P}\{X_n \leq x\} \to \mathbb{P}\{X \leq x\}$ for all continuous points of $x \mapsto \mathbb{P}\{X \leq x\}$.

According to the portmanteau lemma, this is equivalent to:

  - $\mathbb{E} f(X_n) \to \mathbb{E} f(X)$ for all bounded, continuous functions f;
  - $\mathbb{E} f(X_n) \to \mathbb{E} f(X)$ for all bounded, Lipschitz functions f;
  - $\liminf \mathbb{E} f(X_n) \geq \mathbb{E} f(X)$ for all nonnegative, continuous functions f;
  - $\liminf \mathbb{P}\{X_n \in G\} \geq \mathbb{P}\{X \in G\}$ for every open set G;
  - $\limsup \mathbb{P}\{X_n \in F\} \leq \mathbb{P}\{X \in F\}$ for every open set F;
  - .....

For this reason, convergence in distribution is also referred to as **weak convergence**.

The Lévy's continuity theorem established that $X_n \xrightarrow{d} X$ if and only if the characteristic functions $\{\varphi_n\}$ converges pointwise to φ of X.

Note that convergence in distribution of $\{X_n\}$ to X and $\{Y_n\}$ to Y does in general not imply convergence in distribution of $\{X_n + Y_n\}$ to $X + Y$ or of $\{X_n Y_n\}$ to XY.

Theorem 2.22 (Continous mapping theorem)**.** Let g be a continuous function. Then

$$X_n \xrightarrow{\text{d}} X \quad \Rightarrow \quad g(X_n) \xrightarrow{\text{d}} g(X); \tag{7}$$

$$X_n \xrightarrow{\text{p}} X \quad \Rightarrow \quad g(X_n) \xrightarrow{\text{p}} g(X); \tag{8}$$

$$X_n \xrightarrow{\text{a.s.}} X \quad \Rightarrow \quad g(X_n) \xrightarrow{\text{a.s.}} g(X). \tag{9}$$

The chain of implications between the various notions of convergence is summarized as:

$$\boxed{\xrightarrow{\mathcal{L}_p}} \underset{p>q\geq 1}{\Longrightarrow} \boxed{\xrightarrow{\mathcal{L}_q}}$$

$$\Downarrow$$

$$\boxed{\xrightarrow{\text{a.s.}}} \Longrightarrow \boxed{\xrightarrow{p}} \Longrightarrow \boxed{\xrightarrow{d}}$$

Let's return back to the Lévy's continuity theorem. It can transfer the proof of convergence in distribution of something to convergence of their characteristic functions. We should immediately come up with the idea that we can use the theorem to prove the Central Limit Theorem, which is a theorem about convergence in distribution.

Theorem 2.23 (Central Limit Theorem)**.** If $\{X_n\}_{n=1}^{\infty}$ is a sequence of i.i.d. random variables with $\mathbb{E}X_n = \mu$ and $\mathrm{Var}(X_n) = \sigma^2 < \infty$, then

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} Z$$

where $Z \sim N(0,1)$.

Proof. For a characteristic function φ_X, we have the property (you can derive this informally by deferentiating the formula $\varphi_X(t) = \mathbb{E}e^{itX}$)

$$\varphi_X^{(k)}(0) = i^k \mathbb{E}X^k.$$

In particular, if X has mean μ and variance σ^2, then

$$\varphi_X'(0) = i\,\mathbb{E}X = \mu, \qquad \varphi_X''(0) = i^2\mathbb{E}X^2 = -\sigma^2 - \mu^2.$$

Let φ denote the characteristic function of the random variable $(X_n - \mu)/\sigma$, which has mean 0 and variance 1, so that

$$\varphi'(0) = 0, \qquad \varphi''(0) = -1.$$

A second order Taylor expansion of φ yields

$$\varphi(t) = \varphi(0) + \varphi'(0)t + \frac{\varphi''(0)}{2}t^2 + \varepsilon_t = 1 - \frac{t^2}{2} + \varepsilon_t$$

with $\lim_{t\to\infty}(|\varepsilon_t|/t^2) = 0$. Using the independence of X_n's, one has

$$\mathbb{E}e^{itZ_n} = \left(\mathbb{E}e^{it\frac{X_n-\mu}{\sigma\sqrt{n}}}\right)^n = \left(\varphi(t/\sqrt{n})\right)^n.$$

For n large enough, $t/\sqrt{n} \approx 0$ and so

$$\mathbb{E}e^{itZ_n} = \left(1 - \frac{t^2}{2n} + \varepsilon_{t/\sqrt{n}}\right)^n \to e^{-t^2/2} \quad \text{as } n \to +\infty.$$

□

Theorem 2.24 (Law of Large Numbers)**.** If $\{X_n\}_{n=1}^{\infty}$ is a sequence of pairwise independent and identically distributed random variables with $\mathbb{E}X_n = \mu$ and $\mathrm{Var}(X_n) = \sigma^2 < \infty$, then

$$\bar{X}_n \xrightarrow{\mathcal{L}_2} \mu, \quad \bar{X}_n \xrightarrow{p} \mu, \quad \bar{X}_n \xrightarrow{a.s.} \mu.$$

Proof. Since $\mathbb{E}\bar{X}_n = \mu$ and $\mathrm{Var}(\bar{X}_n) = \sigma^2/n$, we have

$$\|\bar{X}_n - \mu\|_2^2 = \mathbb{E}|\bar{X}_n - \mu|^2 = \mathrm{Var}(\bar{X}_n) \to 0 \quad \text{as } n \to \infty,$$

which entails $\bar{X}_n \xrightarrow{\mathcal{L}_2} \mu$ and *a fortiori* $\bar{X}_n \xrightarrow{p} \mu$. To show almost sure convergence, we assume without loss of generality that $X_n \geq 0$. Let $n_k = k^2$ and by Chebychev's inequality

$$\sum_{k=1}^{\infty} \mathbb{P}\left\{|\bar{X}_{n_k} - \mu| > \varepsilon\right\} \leq \frac{\sigma^2}{\varepsilon^2}\sum_{k=1}^{\infty}\frac{1}{k^2} < +\infty,$$

so by the first Borel-Cantelli lemma

$$\mathbb{P}\left\{|\bar{X}_{n_k} - \mu| > \varepsilon \text{ i.o.}\right\} = 0.$$

This means $\bar{X}_{n_k} \xrightarrow{a.s.} \mu$ as $k \to \infty$. Hence, there exists $\Omega_0 \in \mathcal{F}$ with $\mathbb{P}(\Omega_0) = 1$ such that

$$\bar{X}_{n_k}(\omega) \to \mu \quad \text{for all } \omega \in \Omega_0.$$

There is a lemma in analysis that says if for a subsequence $\{x_{n_k}\}$ in $\{x_n\}$ such that $n_{k+1}/n_k \to r \geq 1$ as $k \to \infty$ and $\lim_{k\to\infty} \bar{x}_{n_k} = x$, then $x/r \leq \liminf_n \bar{x}_n \leq \limsup_n \bar{x}_n \leq rx$. Using the lemma, we have

$$\mu \leq \liminf_n \bar{X}_n(\omega) \leq \limsup_n \bar{X}_n(\omega) \leq \mu \quad \forall \omega \in \Omega_0,$$

which implies that $\lim_{n\to\infty} \bar{X}_n = \mu$ on Ω_0. □

See the lecture notes for application of Law of Large Numbers in proving the Weierstrass approximation theorem.

2.10 Conditional Expectations

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{H}$ be a sub-σ-algebra of $\mathcal{F}$ and let X be a random variable with finite expectation. We would like to obtain an representation (approximation) of X in terms of $\mathcal{H}$, but X may not be $\mathcal{H}$-measurable. A *conditional expectation* of X given $\mathcal{H}$, denoted as $\mathbb{E}(X|\mathcal{H})$, is any $\mathcal{H}$-measurable random variable that satisfies

$$\int_H \mathbb{E}(X|\mathcal{H}) d\mathbb{P} = \int_H X d\mathbb{P} \quad \forall H \in \mathcal{H}. \tag{10}$$

Namely, the values of $\mathbb{E}(X|\mathcal{H})$ are "defined" as averages of X on various elements in $\mathcal{H}$. Note that Eq. (10) is equivalent to requiring that $\mathbb{E}V\mathbb{E}(X|\mathcal{H}) = \mathbb{E}VX$ for any $\mathcal{H}$-measurable random variable V. The existence of $\mathbb{E}(X|\mathcal{H})$ can be established by the following. First assume $X \geq 0$ (for generalization work on $X = X^+ - X^-$). Note that $\mu_X : A \mapsto \int_A X d\mathbb{P}$ for $A \in \mathcal{F}$ defines a finite measure on $(\Omega, \mathcal{F})$. Let h be the natural injection from $\mathcal{H}$ to $\mathcal{F}$, so that $\mu_X \circ h = \mu_X|_{\mathcal{H}}$ is the restriction of μ_X to $\mathcal{H}$ and $\mathbb{P} \circ h = \mathbb{P}|_{\mathcal{H}}$ is the restriction of $\mathbb{P}$ to $\mathcal{H}$. Furthermore, $\mu_X|_{\mathcal{H}}$ is absolutely continuous with respect to $\mathbb{P}|_{\mathcal{H}}$ because $\mathbb{P}|_{\mathcal{H}}(H) = \mathbb{P} \circ h(H) = 0$ implies that $\mu_X(h(H)) = \mu_X|_{\mathcal{H}}(H) = 0$. Thus, by the Radon-Nikodym theorem, there exists a function $d\mu_X|_{\mathcal{H}}/d\mathbb{P}_{\mathcal{H}} : \Omega \to \mathbb{R}_+$ such that

$$\mu_X|_{\mathcal{H}}(H) = \int_H X d\mathbb{P} = \int_H \left[\frac{d\mu_X|_{\mathcal{H}}}{d\mathbb{P}_{\mathcal{H}}}\right] d\mathbb{P}|_{\mathcal{H}} \quad \forall H \in \mathcal{H}.$$

We thus see that the conditional expectation is exactly this Radon-Nikodym derivative.

If $X \in \mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P})$, we can define the subspace $\mathcal{K} = \mathcal{L}_2(\Omega, \mathcal{H}, \mathbb{P})$ and use the theorem on orthogonal projection to conclude that there exists $Y \in \mathcal{K}$ such that

$$\|X - Y\|_2 = \inf_{W\in\mathcal{K}} \|X - W\|_2 \quad \text{and} \quad \langle X - Y, W\rangle = 0 \quad \forall W \in \mathcal{K}.$$

Taking $\mathbb{1}_H \in \mathcal{K}$ for any $H \in \mathcal{H}$, we get

$$0 = \langle X - Y, \mathbb{1}_H\rangle = \int_H (X - Y) d\mathbb{P}$$

so that

$$\int_H Y d\mathbb{P} = \int_H X d\mathbb{P} \qquad \forall H \in \mathcal{H}.$$

We see that this Y is a version of $\mathbb{E}(X|\mathcal{H})$. In this case, the conditional expectation of X given $\mathcal{H}$ is the orthogonal projection of X onto $\mathcal{K} = \mathcal{L}_2(\Omega, \mathcal{H}, \mathbb{P})$.

Here are several properties of conditional expectation:

- For $\mathcal{H} = \{\varnothing, \Omega\}$, the conditional expectation is the constant $\mathbb{E}X$, and for $\mathcal{H} = \mathcal{F}$, the conditional expectation is X itself.
- Since $\Omega \in \mathcal{H}$, we have
$$\int_\Omega \mathbb{E}(X|\mathcal{H}) = \int_\Omega X d\,\mathbb{P} \quad \Rightarrow \quad \mathbb{E}\mathbb{E}(X|\mathcal{H}) = \mathbb{E}X,$$
which is the law of iterated expectations.
- If X is $\mathcal{H}$-measurable, then $\mathbb{E}(X|\mathcal{H}) = X$ almost surely.
- Linearity: $\mathbb{E}(aX + bY|\mathcal{H}) = a\mathbb{E}(X|\mathcal{H}) + b\mathbb{E}(Y|\mathcal{H})$.
- Monotonicity: if $X \leq Y$ almost surely then $\mathbb{E}(X|\mathcal{H}) \leq \mathbb{E}(Y|\mathcal{H})$ almost surely.
- Monotone convergence: if $0 \leq X_n \uparrow X$, then $\mathbb{E}(X_n|\mathcal{H}) \uparrow \mathbb{E}(X|\mathcal{H})$ almost surely.
- Dominated convergence: If $X_n \xrightarrow{a.s.} X$ and $|X_n| \leq Y$ for some Y with $\mathbb{E}|Y| < +\infty$, then $\mathbb{E}(X_n|\mathcal{H}) \xrightarrow{a.s.} \mathbb{E}(X|\mathcal{H})$.
- Towering property: if $\mathcal{H} \subset \mathcal{G}$, then $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$.

Part II

Stochastic Process

3 Martingales

Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a fixed probability space in the background, and let $\mathbb{T}$ be some time index set.

3.1 Filtrations and Stopping Times

A filtration $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ on $\mathbb{T}$ is an incresing family of sub-σ-algebras of $\mathcal{H}$, i.e. $\mathcal{F}_t \subset \mathcal{F}_{t'}$ whenever $t < t'$. For a stochastic process $X = (X_t)_{t\in\mathbb{T}}$, the filtration generated by X is $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ with $\mathcal{F}_t = \sigma\{X_s : s \le t\}$. X is said to be adapted to $\mathcal{F}$ if X_t is $\mathcal{F}_t$-measurable for each t.

Definition 3.1. Let $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ be a filtration on $\mathbb{T}$. A random variable $T : \Omega \to \mathbb{T} \cup \{+\infty\}$ is called a *stopping time* of $\mathcal{F}$ if

$$\{T \le t\} \in \mathcal{F}_t \qquad \forall t \in \mathbb{T}. \tag{11}$$

Be aware that both of the sets $\{T \le t\}$ and $\mathcal{F}_t$ increase in their size as t increases. Eq. (11) is equivalent to requiring that the process $Z = (Z_t)_{t\in\mathbb{T}}$ with $Z_t = \mathbb{1}_{\{T\le t\}}$ be adapted to $\mathcal{F}$, and for $\mathbb{T} = \mathbb{N}$ this is also equivalent to requiring that $Z = (Z_t)_{t\in\mathbb{T}}$ with $Z_t = \mathbb{1}_{\{T=n\}}$ be adapted to $\mathcal{F}$.

Example 3.2. Let $\mathbb{T} = \mathbb{N}$, let $X = (X_n)_{n\in\mathbb{N}}$ be a process and consider

$$T = \inf\{n \in \mathbb{N} : X_n \in A\}.$$

T is called the first entrance to A. It is a stopping time because

$$\{T \le n\} = \bigcup_{k=0}^{n} \{X_k \in A\} \in \mathcal{F}_n$$

since each set in the union is in $\mathcal{F}_n$. The information $\{T \le n\}$ at each n can be determined by information of $X = (X_n)_{n\in\mathbb{N}}$ up to n, instead of having to resort to future information.

Example 3.3 (Counting Process)**.** Let $0 < T_1 < T_2 < \cdots$ be some random times taking values in $\mathbb{R}_+$ and assume $T_n \to +\infty$. Define

$$N_t = \sum_{n=1}^{\infty} \mathbb{1}_{[0,t]} \circ T_n, \qquad t \in \mathbb{R}_+.$$

We imagine that for each $\omega \in \Omega$, a sequence of time is determined and N_t "expands" the positive real axis to infinity and counts the occurance time it encountered along the way. If $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is the filtration generated by $N = (N_t)_{t\in\mathbb{R}_+}$, then it is obvious that every occurence time T_n is a stopping time of $\mathcal{F}$. Indeed,

$$\{T_n \le t\} = \{N_t \ge n\} \in \mathcal{F}_t \quad \forall t \in \mathbb{R}_+.$$

Another stopping time is $T = \inf\{t \ge a : N_t = N_{t-a}\}$, namely the first time that an interval of length a passed without an arrival.

Let $\mathcal{F}$ be a filtration on $\mathbb{T}$ and let T be a stopping time of it. We define[1]

$$\mathcal{F}_T = \{H \in \mathcal{H} : H \cap \{T \le t\} \in \mathcal{F}_t \quad \forall t \in \bar{\mathbb{T}}\} \tag{12}$$

[1] To tackle the case that $T(\omega)$ may be $+\infty$ for some ω, we define $\mathcal{F}_\infty$ as $\sigma(\bigcup_{t\in\mathbb{T}} \mathcal{F}_t)$, the σ-algebra generated by the union of all the $\mathcal{F}_t$.

as the *past until* T. It is a sub-σ-algebra of $\mathcal{H}$ on Ω. Note that if $T \equiv t_0$ is a fixed constant, then $\{T \le t\} = \{t_0 \le t\}$ is $\varnothing$ when $t < t_0$ and it is Ω when $t \ge t_0$. Remembering $F_{t_0} \subset \mathcal{F}_t$ for all $t \ge t_0$, we see that $\mathcal{F}_T = \mathcal{F}_{t_0}$ in this case. Also note that $\{T \le s\} \in \mathcal{F}_T$ for each $s \ge 0$, because $\{T \le s\} \cap \{T \le t\} = \{T \le s \wedge t\} \in \mathcal{F}_t$ for any t. This shows that T is $\mathcal{F}_T$-measurable.

If a positive random variable V is $\mathcal{F}_T$-measurable, then $\{V > s\} \in \mathcal{F}_T$ for any $s \ge 0$. By Eq. (12), this means $\{V > s\} \cap \{T \le t\} = \{V \mathbb{1}_{\{T \le t\}} > s\} \in \mathcal{F}_t$ for all $t \in \mathbb{T}$. In other words, the random variable $V \mathbb{1}_{\{T \le t\}}$ should be $\mathcal{F}_t$-measurable for any $t \in \mathbb{T}$. This is the content of the following theorem in the book.

Theorem 3.4. A random variable V belongs to $\mathcal{F}_T$ if and only if

$$V \mathbb{1}_{\{T \le t\}} \in \mathcal{F}_t$$

for every $t \in \bar{\mathbb{T}}$. In particular, if $\mathbb{T} = \mathbb{N}$, then the condition is equivalent to requiring that

$$V \mathbb{1}_{\{T = n\}} \in \mathcal{F}_n \qquad \forall n \in \bar{\mathbb{N}}.$$

In the book, the author identifies a filtration $\mathcal{F} = (\mathcal{F}_t)$ with the collection of all stochastic processes X such that

1. $X = (X_t)$ is adapted to $\mathcal{F} = (\mathcal{F}_t)$;
2. $t \mapsto X_t(\omega)$ is right-continuous for each $\omega \in \Omega$.

Then we can identify $\mathcal{F}_T$ as the set of values (random variables) of all processes X in $\mathcal{F}$ at time T, i.e. $\mathcal{F}_T = \{X_T : X \in \mathcal{F}\}$.

Theorem 3.5. Let S and T be stopping times of $\mathcal{F}$. Then

1. $S \wedge T$ and $S \vee T$ are stopping times of $\mathcal{F}$;
2. if $S \le T$ then $\mathcal{F}_S \subset \mathcal{F}_T$;
3. in gengeral, $\mathcal{F}_{S \wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$;
4. if $V \in \mathcal{F}_S$ then the following are in $\mathcal{F}_{S \wedge T}$:

$$V \mathbb{1}_{\{S \le T\}}, \qquad V \mathbb{1}_{\{S = T\}}, \qquad V \mathbb{1}_{\{S < T\}}.$$

Proof. You shoule be able to work out the proof yourself, following items 1, 2, 4 and finally 3. Or see page 177 in the book. □

Definition 3.6. We define $\mathbb{E}_T$ as the conditional expectation operator based on the σ-algebra $\mathcal{F}_T$, i.e. $\mathbb{E}_T := \mathbb{E}(\cdot|\mathcal{F}_T)$.

Borrowing this notation, we use $\mathbb{E}_t$ to mean $\mathbb{E}(\cdot|\mathcal{F}_t)$, the conditional expectation given the σ-algebra $\mathcal{F}_t$.

Theorem 3.7. The following hold for all positive random variables X, Y, Z and for all stopping times S and T of $\mathcal{F}$:

1. $Y = \mathbb{E}_T X$ if and only if $Y \in \mathcal{F}_T$ and $\mathbb{E} VX = \mathbb{E} VY$ for every positive $V \in \mathcal{F}_T$.
2. $\mathbb{E}\mathbb{E}_T X = \mathbb{E} X$.
3. $\mathbb{E}_S \mathbb{E}_T X = \mathbb{E}_{S \wedge T} X$.
4. $\mathbb{E}_T(X + YZ) = X + Y \mathbb{E}_T Z$ if $X, Y \in \mathcal{F}_T$.

3.2 Martingales

Definition 3.8. A real-valued stochastic process $X = (X_t)_{t\in\mathbb{T}}$ is called an *$\mathcal{F}$-submartingale* if X is adapted to $\mathcal{F}$, $\mathbb{E}|X_t| < +\infty$ for each t, and

$$\mathbb{E}_s(X_t - X_s) \geq 0 \qquad \forall t > s.$$

It is an *$\mathcal{F}$-supermartingale* if

$$\mathbb{E}_s(X_t - X_s) \leq 0 \qquad \forall t > s.$$

It is an *$\mathcal{F}$-martingale* if

$$\mathbb{E}_s(X_t - X_s) = 0 \qquad \forall t > s.$$

Several remarks:

- Let X be an $\mathcal{F}$-submartingale. For $u > t > s$,

$$\mathbb{E}_s(X_u - X_t) = \mathbb{E}_s\mathbb{E}_t(X_u - X_t) \geq \mathbb{E}_s 0 = 0,$$

  so that any remote future increment is also positive.

- When $\mathbb{T} = \mathbb{N}$, the condition for martingale is equivalent to

$$\mathbb{E}_n(X_{n+1} - X_n) = 0, \qquad \forall n \in \mathbb{N}.$$

- In fact, X is a martingale if $\mathbb{E}X_t = \mathbb{E}X_0$ for all times t.

Example 3.9. Here are two basic examples of martingales.

1. Let $X_1, X_2, \ldots$ be independent random variables with mean 0, and put $S_0 = 0$. Then $S_n = S_0 + X_1 + \cdots + X_n$ is a martingale adapted to the filtration generated by itself, since $\mathbb{E}_n(S_{n+1} - S_n) = \mathbb{E}_n X_{n+1} = \mathbb{E}X_{n+1} = 0$.
2. Similarly, if $R_1, R_2, \ldots$ are independent random variables all with mean 1 and finite variance, then $M_n = M_0 R_1 R_2 \cdots R_n$ with $M_0 = 1$ is a martingale adapted to the filtration generated by itself, since $\mathbb{E}_n M_{n+1} = \mathbb{E}_n M_n R_{n+1} = M_n \mathbb{E}_n R_{n+1} = M_n \mathbb{E}R_{n+1} = M_n$.

Theorem 3.10. Let Z be an integrable random variable. Define

$$X_t = \mathbb{E}_t Z$$

for $t \in \mathbb{T}$. Then $X = (X_t)_{t\in\mathbb{T}}$ is an $\mathcal{F}$-martingale and is uniformly integrable.

Proof. Adaptedness is immediate, each X_t is integrable, and the martingale property is $\mathbb{E}_s X_t = \mathbb{E}_s\mathbb{E}_t Z = \mathbb{E}_s Z = X_s$ for $s < t$. For uniform integrability, the proof in the book used a proposition that says $\mathcal{C}$ is uniformly integrable if and only if $\sup_{X\in\mathcal{C}} \mathbb{E}\, f(|X|) < \infty$ for some increasing convex function such that $\lim_{x\to\infty} f(x)/x = \infty$ (page 74). But it should be intuitively clear that $\{X_t\}_{t\in\mathbb{T}}$ is uniformly integrable. □

Wiener process and Poisson process are two processes that have stationary and independent increments.

Definition 3.11. Put $W_0 = 0$. The continuous process $W = (W_t)_{t\in\mathbb{R}_+}$ is called a *Wiener process* with respect to $\mathcal{F}$ if it is adapted to $\mathcal{F}$ and

$$\mathbb{E}_s f(W_{s+t} - W_s) = \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} f(x) e^{-x^2/2t} dx$$

for all $s, t \in \mathbb{R}_+$ and all positive Borel functions on $\mathbb{R}$.

Theorem 3.12. W is Wiener with respect to $\mathcal{F}$ if and only if it is continuous and

1. W is an $\mathcal{F}$-martingale, and
2. $Y = (W_t^2 - t)_{t\in\mathbb{R}_+}$ is an $\mathcal{F}$-martingale.

Proof of neccessity. If W is Wiener, then each W_t is normal with mean 0 and variance t. For $s < t$, the increment $W_t - W_s$ is independent of $\mathcal{F}_s$ and so $\mathbb{E}_s(W_t - W_s) = \mathbb{E}(W_t - W_s) = 0$. To show the second process is a martingale[2], we note that

$$Y_t - Y_s = (W_t - W_s)^2 + 2W_s(W_t - W_s) - (t - s)$$

so that

$$\mathbb{E}_s(Y_t - Y_s) = \mathbb{E}(W_t - W_s)^2 + 2W_s\mathbb{E}(W_t - W_s) - (t - s) = (t - s) - (t - s) = 0.$$

□

Theorem 3.12 signifies ubiquity of normal distribution: if a collection of continuous i.i.d. random variables $\{X_t\}_{t\in\mathbb{R}_+}$ is such that they have mean $\mathbb{E}X_t = 0$ and variance $\mathbb{E}X_t^2 = t$ for all t, then it must have a normal distribution, i.e. it is a Wiener process.

Proposition 3.13. W is a Wiener process with respect to $\mathcal{F}$ if and only if, for any $r \in \mathbb{R}$,

$$M_t = \exp\left\{rW_t - \frac{1}{2}r^2t\right\}, \quad t \in \mathbb{R}_+$$

is an $\mathcal{F}$-martingale[3].

Proof. First recall that, for a normal random variable $X \sim N(\mu, \sigma^2)$, its moment-generating function is $m(r) = \mathbb{E}e^{rX} = e^{r\mu+\frac{1}{2}\sigma^2r^2}$, so for W_t it is $m(r) = e^{\frac{1}{2}r^2t}$. If W is Wiener, then for $s < t$

$$\mathbb{E}_s\left[\frac{M_t}{M_s}\right] = \mathbb{E}_s \exp\left\{r(W_t - W_s) - \frac{1}{2}r^2(t - s)\right\} = 1 \tag{13}$$

so that $\mathbb{E}_sM_t = \mathbb{E}_s\left[\frac{M_t}{M_s} \cdot M_s\right] = M_s\mathbb{E}_s\left[\frac{M_t}{M_s}\right] = M_s \cdot 1 = M_s$. This proves $M = (M_t)_{t\in\mathbb{R}_+}$ is an $\mathcal{F}$-martingale. Conversely, if M is an $\mathcal{F}$-martingale, then $\mathbb{E}_s(M_t/M_s) = 1$, which means Eq. (13) holds, or equivalently

$$\mathbb{E}_s \exp\{r(W_{s+t} - W_s)\} = e^{\frac{1}{2}r^2t}.$$

This proves that $W = (W_t)_{t\in\mathbb{R}_+}$ is Wiener. □

A *counting process* $N = (N_t)_{t\in\mathbb{R}_+}$ is a process with state space $(\mathbb{N}, 2^{\mathbb{N}})$ whose every path $t \mapsto N_t(\omega)$ starts from $N_0(\omega) = 0$, is increasing and right-continuous, and increases by jumps of size one only. Therefore N_t is the number of jumps in the interval $(0, t]$.

Definition 3.14. The counting process N is said to be a Poisson process with rate c with respect to $\mathcal{F}$ if it is adapted to $\mathcal{F}$ and

$$\mathbb{E}_s f(N_{s+t} - N_s) = \sum_{k=0}^{\infty}\left[\frac{(ct)^k}{k!}e^{-ct}\right] \cdot f(k) \tag{14}$$

for all $s, t \in \mathbb{R}_+$ and all positive functions f on $\mathbb{N}$.

[2]The proof is from the book. However, it is easy to see that $\mathbb{E}_sW_t^2 = \mathbb{E}W_t^2 = t$, so if we subtract the variance t we get a martingale. $\mathbb{E}(W_t^2 - t) = 0 = W_0$ for all $t \in \mathbb{R}_+$.

[3]M_t is the moment generating function of W_t, e^{rW_t}, times its reciprocal.

With $f(k) = k$, we get the mean of $N_{t+s} - N_s$ from Eq. (14): it is $\mathbb{E}(N_{t+s} - N_s) = \mathbb{E}_s(N_{t+s} - N_s) = ct$. Each N_t is Poisson distributed with mean ct. It turns out that

Theorem 3.15. A counting process N is a Poisson process with rate c if and only if $(N_t - ct)_{t\in\mathbb{R}_+}$ is an $\mathcal{F}$-martingale.

The theorem says, if a (discrete) counting process $\{N_t\}_{t\in\mathbb{R}_+}$ is such that they are independent, and the expectation $\mathbb{E}N_t = ct$ grows linearly in t with a single rate c, then only Poisson distribution can be used to describe their probability laws.

In fact, if N is a Poisson process, then we can substitute s with stopping time S:

$$\mathbb{E}_S f(N_{S+t} - N_S)\mathbb{1}_{\{S<\infty\}} = \sum_{k=0}^{\infty}\left[\frac{(ct)^k}{k!}e^{-ct}\right]\cdot f(k)\mathbb{1}_{\{S<\infty\}}. \tag{15}$$

Eq. (15) is called *strong Markov property* for the Poisson process.

Example 3.16. Let S be the first time of an interval of length a passes without a jump, that is

$$S = \inf\{t \geq a : N_t = N_{t-a}\}.$$

Let T be the time of first jump after S. Note that the interval that includes S has length $a + (T - S)$, and for a large, the raw intuition expects $T - S$ to be small. Instead, noting that $\{T - S > t\} = \{N_{S+t} - N_S = 0\}$, we see that $T - S$ is independent of $\mathcal{F}_S$ and has the same exponential distribution as if S is a jump time.

3.3 Martingale Transformation and Maxima

Definition 3.17. A process $X = (X_n)_{n\in\mathbb{N}}$ is said to be $\mathcal{F}$-predictable if $X_{n+1} \in \mathcal{F}_n$ for every $n \in \mathbb{N}$.

Theorem 3.18. Let X be adapted and integrable. Then it can be decomposed as

$$X_n = X_0 + M_n + A_n, \qquad n \in \mathbb{N}, \tag{16}$$

where M is a martingale with $M_0 = 0$, and A is predictable with $A_0 = 0$. This decomposition is unique up to equivalence. In particular, if X is a submartingale, then A is increasing, and if X is a supermartingale, A is decreasing.

Proof. The way we achieve Eq. (16) is to define $M_0 = A_0 = 0$ and define M and A through their increments:

$$A_{n+1} - A_n = \mathbb{E}_n(X_{n+1} - X_n), \qquad M_{n+1} - M_n = (X_{n+1} - X_n) - (A_{n+1} - A_n)$$

for each $n \in \mathbb{N}$, so that $X_{n+1} - X_n = (M_{n+1} - M_n) + (A_{n+1} - A_n) = (M_{n+1} + A_{n+1}) - (M_n + A_n) \Rightarrow X_n = X_0 + M_n + A_n$. □

Let $M = (M_n)$ and $F = (F_n)$ be real-valued stochastic processes and define

$$X_n = F_0 \cdot M_0 + F_1 \cdot (M_1 - M_0) + \cdots + F_n \cdot (M_n - M_{n-1}), \quad n \in \mathbb{N}.$$

The $X = (X_n)$ is called the *integral* of F with respect to M, or the *transform* of M by F, and we write

$$X = \int F dM.$$

If $F \equiv 1$, then X_n is just M_n. From another perspective, F is like a (random) function on $\mathbb{N}$, and M is like a (random) (signed) measure on $\mathbb{N}$, with mass $(M_n - M_{n-1})$ at n. Thus we can also write

$$X_n = \int_{[0,n]} F dM.$$

Theorem 3.19. Let F be a bounded predictable process and let $X = \int F dM$. If M is a martingale, then so is X. If M is a submartingale and F is positive, then X is a submartingale.

Proof. X is adapted to $\mathcal{F}$ because F and M are. Since F is bounded, say by $b > 0$, $|X_n| \leq b \cdot (|M_0| + \cdots + |M_n - M_{n-1}|)$, which is integrable, so X is integrable. Finally,

$$E_n(X_{n+1} - X_n) = F_{n+1} \cdot \mathbb{E}_n(M_{n+1} - M_n) = F_{n+1} \cdot 0 = 0$$

so that $X = (X_n)$ is a martingale. □

Example 3.20. Let S and T be stopping times of $\mathcal{F}$ with $S \leq T$ (they take values in $\bar{\mathbb{N}}$). Let $V \in \mathcal{F}_S$. Then the following processes

$$V\mathbb{1}_{(S,T]}, \quad V\mathbb{1}_{(S,\infty]}, \quad \mathbb{1}_{(S,T]}, \quad \mathbb{1}_{[0,T]}$$

are all **predictable processes**. Start with the second one:

- Let $X_n = V \cdot \mathbb{1}_{(S,\infty]}(n) = V \cdot \mathbb{1}_{\{S<n\}}(\omega)$. If we know the value of X_n, plus information of the stopping time (recall $\{S \leq n\} \in \mathcal{F}_n$) at time n, then can know the value of $X_{n+1} = V \cdot \mathbb{1}_{\{S<n+1\}}(\omega) = V \cdot \mathbb{1}_{\{S\leq n\}}(\omega)$, i.e. $X_{n+1} \in \mathcal{F}_n$. Note that $\{S < n+1\} = \{S \leq n\}$ because S can only take integer values. Thus, $V\mathbb{1}_{(S,\infty]}$ is predictable.
- From $S \leq T$, we have $\mathcal{F}_S \subset \mathcal{F}_T$, so that $V \in \mathcal{F}_T$ also. Changing S to T, we see $V\mathbb{1}_{(T,\infty]}$ is also predictable.
- Their difference $V\mathbb{1}_{(S,\infty]} - V\mathbb{1}_{(T,\infty]} = V\mathbb{1}_{(S,T]}$ is also predictable.
- Taking $V = 1$ shows that $\mathbb{1}_{(S,T]}$ is predictable.
- Taking $T = \infty$ shows $\mathbb{1}_{(S,\infty]}$ is predictable.
- Finally $\mathbb{1}_{[0,S]} = 1 - \mathbb{1}_{(S,\infty]}$ is predictable. Alterntatively, $\mathbb{1}_{[0,S]}(n+1) = \mathbb{1}_{\{S\geq n+1\}} = \mathbb{1}_{\{S<n+1\}^c} = \mathbb{1}_{\{S\leq n\}^c}$ and since $\{S \leq n\} \in \mathcal{F}_n$ we have $\{S \leq n\}^c \in \mathcal{F}_n$ also.

Definition 3.21. Let $M = (M_n)$ be a process. Let $T \in \bar{\mathbb{N}}$ be a random time. Then the process X defined by

$$X_n(\omega) = M_{n\wedge T(\omega)}(\omega) = \begin{cases} M_n(\omega) & \text{if } n \leq T(\omega), \\ M_{T(\omega)}(\omega) & \text{if } n > T(\omega) \end{cases}$$

is called the process M stopped at T.

Observe that X is a "transform" of M, and indeed we can express X as the integral

$$X = \int \mathbb{1}_{[0,T]} dM = \int_{[0,T]} dM$$

so that

$$X_n = \int_{[0,n]\cap[0,T]} dM = \int_{[0,n\wedge T]} dM = M_{n\wedge T}.$$

Thus, from Theorem 3.19, if M is a martingale, then so is the martingale stopped at T.

Theorem 3.22 (Doob's stopping theorem)**.** Let M be adapted to $\mathcal{F}$. The following are equivalent:

1. M is a submartingale;
2. For every pair of bounded stopping time $S \leq T$, M_S and M_T are integrable and

$$\mathbb{E}_S(M_T - M_S) \geq 0.$$

3. For every pair of bounded stopping time $S \leq T$, M_S and M_T are integrable and

$$\mathbb{E}(M_T - M_S) \geq 0.$$

We now want to define the notion of crossings of an interval $(a, b) \in \mathbb{R}$ by some process M. Put $T_0 = -1$ for convenience and for each $k \geq 1$ define

$$S_k = \inf\{n > T_{k-1} : M_n \leq a\}, \qquad T_k = \inf\{n > S_k : M_n \geq b\},$$

the downcrossing times and upcrossing times. Then

$$U_n(a, b) = \sum_{k=1}^{\infty} \mathbb{1}_{(0,n]} \circ T_k = \mathbb{1}_{(0,n]}(T_1) + \mathbb{1}_{(0,n]}(T_2) + \cdots$$

is the number of upcrossings of (a, b) during $[0, n]$. The following inequality will be used to prove the martingale convergence theorem.

Proposition 3.23. If M is a submartingale, then

$$(b - a)\mathbb{E}U_n(a, b) \leq \mathbb{E}[(M_n - a)^+ - (M_0 - a)^+]$$

Proof. An upcrossing of (a, b) by M is the same as an upcrossing of $(0, b - a)$ by $(M - a)^+$, and the later us again a submartingale, so we may assume $a = 0$ and $M \geq 0$. Let $F_n = \sum_{k=1}^{\infty} \mathbb{1}_{(S_k, T_k]}(n)$ for $n \geq 1$ and put $F_0 = 0$. Note $F_n \in \{0, 1\}$ with $F_n = 1$ if $n \in (S_k, T_k]$ for some particular k and 0 otherwise. Let $X = \int F dM$. Since F is predictable, we have $F_{k+1} \in \mathcal{F}_k$, so

$$\mathbb{E}_k(X_{k+1} - X_k) = \mathbb{E}_k F_{k+1} \cdot (M_{k+1} - M_k) = F_{k+1}\mathbb{E}_k(M_{k+1} - M_k) \leq \mathbb{E}_k(M_{k+1} - M_k).$$

Taking expectations on bothe side and summing over k:

$$\mathbb{E}(X_n - X_0) \leq \mathbb{E}(M_n - M_0).$$

On the other hand, $bU_n(0, b) \leq X_n - X_0$. The justification is to view $M = (M_n)$ as stock price, and so X_n is like the total profit if you buy one share when M hits 0 and sell the share when it goes above b during $[0, n]$, which is larger than b times total number of upcrossings in the interval $[0, n]$. Thus

$$b\mathbb{E}U_n(0, b) \leq \mathbb{E}(X_n - X_0) \leq \mathbb{E}(M_n - M_0).$$

□

The following inequalities on maxima and minima should be more or less obvious. Let $S = \{\max_{k \leq n} M_k \geq b\} \subset \Omega$. Note $\mathbb{P}(S) = \mathbb{E}\mathbb{1}_S$ and the inequality says $b \cdot \mathbb{P}(S) = \mathbb{E}\, b \cdot \mathbb{1}_S \leq \mathbb{E}M_n \cdot \mathbb{1}_S$. The submartingale $\{M_n\}$ has a tendency to increase, so on the domain S, M_n should be close to or larger than b, i.e. $b \leq \mathbb{E}M_n$ over that domain, and this is exactly what the inequality says.

Theorem 3.24. Let $M = (M_n)$ be a process adapted to $\mathcal{F}$. Suppose M is a submartingale, then for $b > 0$,

$$b\mathbb{P}\{\max_{k \leq n} M_k \geq b\} \leq \mathbb{E}M_n \mathbb{1}_{\{\max_{k \leq n} M_k \geq b\}} \leq \mathbb{E}M_n^+,$$

$$b\mathbb{P}\{\min_{k \leq n} M_k \leq -b\} \leq \mathbb{E}M_n \mathbb{1}_{\{\min_{k \leq n} M_k > -b\}} - \mathbb{E}M_0 \leq \mathbb{E}M_n^+ - \mathbb{E}M_0.$$

Proof. Define the stopping times

$$T = \inf\{n \geq 0 : M_n \geq b\}, \quad S = \inf\{n \geq 0 : M_n \leq -b\},$$

so that

$$\left\{\max_{k\leq n} M_k \geq b\right\} = \{T \leq n\}, \quad \left\{\min_{k\leq n} M_k \leq -b\right\} = \{S \leq n\}.$$

Note that on $\{T \leq n\}$, we have $b \leq M_T = M_{T\wedge n}$, so

$$b\mathbb{1}_{\{T\leq n\}} \leq M_{T\wedge n}\mathbb{1}_{\{T\leq n\}} \leq (\mathbb{E}_{T\wedge n} M_n)\mathbb{1}_{\{T\leq n\}} = \mathbb{E}_{T\wedge n} M_n \mathbb{1}_{\{T\leq n\}},$$

where the second inequality is Doob's submartingale inequality and the last equality is because $\{T \leq n\} \in \mathcal{F}_{T\wedge n}$. Taking expectations on both sides yields the desired inequality. Note that since $M_n \leq M_n^+$ and $\mathbb{1}_{\{\max_{k\leq n} M_k \geq b\}} \leq 1$, the second inequality $\mathbb{E} M_n \mathbb{1}_{\{\max_{k\leq n} M_k \geq b\}} \leq \mathbb{E} M_n^+$ is obvious.

On $\{S \leq n\}$, we have $M_S \leq -b$, so that

$$M_{S\wedge n} = M_S \mathbb{1}_{\{S\leq n\}} + M_n \mathbb{1}_{\{S>n\}} \leq -b\mathbb{1}_{\{S\leq n\}} + M_n \mathbb{1}_{\{S>n\}}.$$

Taking expectations on both side and noting that $\mathbb{E} M_0 \leq \mathbb{E} M_{S\wedge n}$ by Doob's martingale inequality, we get the desired result. □

If M is a martingale, then$|M|^p$ is a submartingale for $p \in [1, \infty)$. Apply the above theorem to $|M|^p$ we get a generalization of Kolmogorov's inequality:

Corollary 3.25. Let M be a martingale in $\mathcal{L}_p$ for some $p \in [1, \infty)$. Then, for $b > 0$,

$$b^p \mathbb{P}\left\{\max_{k\leq n} |M_k| > b\right\} \leq \mathbb{E}|M_n|^p.$$

Theorem 3.26. Let M be a martingale in $\mathcal{L}_p$ for some $p > 1$, with $1/p + 1/q = 1$. Then

$$\mathbb{E} \max_{k\leq n} |M_k|^p \leq q^p \mathbb{E}|M_n|^p.$$

Proof. Fix n and introduce $Z = \max_{k\leq n} |M_k|$. We want to show

$$\mathbb{E} Z^p \leq q^p \mathbb{E}|M_n|^p.$$

We have

$$Z^p = \int_0^Z px^{p-1}dx = \int_0^\infty px^{p-2} x \mathbb{1}_{\{Z\geq x\}} dx$$

and by Theorem 3.24,

$$\mathbb{E}\, x\mathbb{1}_{\{Z\geq x\}} \leq \mathbb{E}|M_n| \cdot \mathbb{1}_{\{Z\geq x\}},$$

so

$$\mathbb{E} Z^p \leq \mathbb{E}|M_n| \int_0^\infty px^{p-2}\mathbb{1}_{\{Z\geq x\}} dx.$$

Note that

$$\int_0^\infty px^{p-2}\mathbb{1}_{\{Z\geq x\}}dx = \int_0^Z \frac{p-1}{p-1} px^{p-2}dx = \frac{p}{p-1}\int_0^Z (p-1)x^{p-2}dx = qZ^{p-1},$$

so with $p - 1 = p/q$,

$$\mathbb{E} Z^p \leq \mathbb{E}|M_n| qZ^{p-1} \leq q\left(\mathbb{E}|M_n|^p\right)^{1/p}\left(\mathbb{E} Z^p\right)^{1/q}$$

from Hölder's inequality. Solving for $\mathbb{E} Z^p$ we have

$$\left(\mathbb{E} Z^p\right)^{1-1/q} = \left(\mathbb{E} Z^p\right)^{1/p} \leq q\left(\mathbb{E}|M_n|^p\right)^{1/p} \quad \Rightarrow \quad \mathbb{E} Z^p \leq q^p \mathbb{E}|M_n|^p.$$

□

3.4 Martingale Convergence

Theorem 3.27 (Martingale convergence theorem)**.** Let $X = \{X_n\}_{n=0}^{\infty}$ be a submartingale. If

$$\sup_n \mathbb{E} X_n^+ < \infty, \tag{17}$$

then $X = \{X_n\}_{n=0}^{\infty}$ converges almost surely to an integrable random variable.

Proof. First note that

$$\mathbb{E} X_n^+ \leq \mathbb{E}|X_n| = 2\mathbb{E} X_n^+ - \mathbb{E} X_n \leq 2\mathbb{E} X_n^+ - \mathbb{E} X_0 \tag{18}$$

so Eq. (17) is equivalent to requiring that $X = \{X_n\}$ be $\mathcal{L}_1$ bounded, i.e. $\sup_n \mathbb{E}|X_n| < \infty$.

If, for an outcome ω, the sequence $\{X_n(\omega)\}$ does not have a limit, then we can pick two rationals $\liminf X_n(\omega) \leq a < b \leq \limsup X_n(\omega)$ such that the sequence upcross (a, b) infinitely often, i.e. $U(a, b) = \infty$ where $U(a, b) = \lim_{n\to\infty} U_n(a, b)$.Thus, to show $\lim X_n$ exists almost surely, we can show for any pair of rationals with $a < b$, one has $U(a, b) < \infty$ almost surely.

Fix $a < b$. By Proposition 3.23,

$$(b-a)\mathbb{E}U(a,b) = (b-a)\lim_{n\to\infty} \mathbb{E}U_n(a,b) \leq \sup \mathbb{E}(X_n - a)^+ \leq \sup \mathbb{E} X_n^+ + |a| < \infty$$

where we used the monotone convergence theorem in the first equality. Thus, $U(a, b) < \infty$ almost surely. It follows that $X_\infty = \lim_{n\to\infty} X_n$ exists almost surely. By Fatou's lemma

$$\mathbb{E}|X_\infty| = \mathbb{E} \liminf |X_n| \leq \liminf \mathbb{E}|X_n| \leq 2 \sup_n \mathbb{E} X_n^+ - \mathbb{E} X_0 < \infty$$

so X_∞ is integrable. □

Theorem 3.28. Let X be a submartingale. Then X converges almost surely and in $\mathcal{L}_1$ if and only if it is uniformly integrable. Moreover, setting $X_\infty = \lim X_n$ extends X to a submartingale $\bar{X} = (X_n)_{n\in\bar{\mathbb{N}}}$.

Proof. Theorem III.4.6. (p106) in the book says a sequence $\{X_n\}$ of random variables converges in $\mathcal{L}_1$ if and only if it converges in probability and is uniformly integrable. So if X converges almost surely and in $\mathcal{L}_1$, then it is uniformly integrable. On the other hand, if it is uniformly integrable, then it is $\mathcal{L}_1$-bounded, so we can use Theorem 3.27 to conclude that it converges almost surely, and also in $\mathcal{L}_1$ by Theorem III.4.6. again. □

Theorem 3.29. A process $M = (M_n)_{n\in\mathbb{N}}$ is a uniformly integrable martingale if and only if

$$M_n = \mathbb{E}_n Z, \qquad n \in \mathbb{N} \tag{19}$$

for some integrable random variable Z. If so, it converges almost surely and in $\mathcal{L}_1$ to the integrable random variable

$$M_\infty = \mathbb{E}_\infty Z \tag{20}$$

and $\bar{M} = (M_n)_{n\in\bar{\mathbb{N}}}$ is again a uniformly integrable martingale.

Proof. If M has the form Eq. (19), then it is uniformly integrable by Theorem 3.10.

If M is uniformly integrable, then Theorem 3.28 shows it converges almost surely and in $\mathcal{L}_1$ to some integrable random variable M_∞ and that $\bar{M} = (M_n)_{n\in\bar{\mathbb{N}}}$ is again a martingale. Define $Z = M_\infty$ so that by the martingale property for $\bar{M}$ one has $M_n = \mathbb{E}_n M_\infty$. □

Corollary 3.30. For every integrable random variable Z,

$$\mathbb{E}_n Z \to \mathbb{E}_\infty Z$$

almost surely and in $\mathcal{L}_1$.

We can use the above corollary to give a proof to Kolmogorov's 0-1 law.

Theorem 3.31 (Kolmogorov's 0-1 law)**.** Let $\{\mathcal{G}_n\}_{n\geq 1}$ be a sequence of independent sub-σ-algebras of $\mathcal{F}$ and $\mathcal{T}$ the corresponding tail σ-algebra. Then

$$H \in \mathcal{T} \quad \Rightarrow \quad \mathbb{P}(H) \in \{0, 1\}.$$

Proof. By Corollary 3.30, for every event H,

$$\mathbb{E}_n \mathbb{1}_H \to \mathbb{E}_\infty \mathbb{1}_H$$

almost surely. When $H \in \mathcal{T}$, since $\mathcal{T}$ is independent of $\mathcal{G}_n$, we have $\mathbb{E}_n \mathbb{1}_H = \mathbb{E}\mathbb{1}_H = \mathbb{P}(H)$. On the other hand, since $\mathcal{G}_n \subset \mathcal{G}_\infty$ for every n, we have $\mathcal{T} \subset \mathcal{G}_\infty$ (see Definition 2.16), which implies that $\mathbb{E}_\infty \mathbb{1}_H = \mathbb{1}_H$. This implies that $\mathbb{P}(H)$ is either 0 or 1. □

4 Poisson Random Measures

Throughout, $(E, \mathcal{E})$ is a measurable space that is often $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. $(\Omega, \mathcal{H}, \mathbb{P})$ is the probability space in the background.

4.1 Random Measures

Definition 4.1. A *random measure* is a measure-valued random variable. Specifically, $M : \Omega \times \mathcal{E} \to \bar{\mathbb{R}}_+$ is a random measure if for every left coordinate it is a measure, and for every right coordinate it is a mearuable function (i.e. a real-valued random variable).

We can also view a random measure as a function $M : \Omega \to \{\mu : \mu : \mathcal{E} \to \bar{\mathbb{R}}_+\}$, from Ω to the space of all measures on $(E, \mathcal{E})$. It is a *random counting measure* if, for almost every $\omega \in \Omega$, M_ω is purely atomic and its every atom has weight one. Often, we also work with random probability measures, for example in Bayesian statistics.

Here are some notations regarding random measures. For $f \in \mathcal{E}_+$,

$$M_\omega f := \int_E f(x) dM_\omega$$

is a (positive) random variable as a function of $\omega \in \Omega$, and

$$\mu(A) = \mathbb{E} M_\omega(A) = \int_\Omega M_\omega(A) d\,\mathbb{P}$$

defines a measure on $(E, \mathcal{E})$, called the mean of M, i.e. $\mu = \mathbb{E}M$. By Fubini's theorem one has $\mu f = \mathbb{E}Mf, \forall f \in \mathcal{E}_+$.

Proposition 4.2. The probability law of a random measure M is completely determined by the Laplace functional $\varphi : \mathcal{E}_+ \to [0, 1]$ defined by

$$\varphi(f) = \mathbb{E}e^{-Mf}, \quad f \in \mathcal{E}_+.$$

Proposition 4.3. If $(f_n) \subset \mathcal{E}_+$ is increasing to f, then

$$\lim_{n\to\infty} \mathbb{E}e^{-Mf_n} = \mathbb{E}e^{-Mf}.$$

Proof. If $f_n \uparrow f$, then $M_\omega f_n \uparrow M_\omega f$ for each $\omega \in \Omega$ by the monotone convergence theorem, and so $e^{-M_\omega f_n} \uparrow e^{-M_\omega f}$ for each $\omega \in \Omega$. The desired conclusion follows from the bounded convergence theorem. □

Proposition 4.4. Two random measures M and N are independent if and only if

$$\mathbb{E}e^{-(Mf+Ng)} = \left(\mathbb{E}e^{-Mf}\right)\left(\mathbb{E}e^{-Ng}\right), \quad f, g \in \mathcal{E}_+.$$

Example 4.5. Let $X = \{X_i\}_{i=1}^\infty$ be an independency of random variables taking values in $\mathbb{R}^n$ according to some common distribution λ. Let K be independent of X and have Poisson distribution with mean c. Then

$$M(A) = \sum_{i=1}^K \mathbb{1}_A(X_i), \quad A \in \mathcal{B}(\mathbb{R}^n)$$

defines a random measure, i.e. the measure of A is the count of how much X_i's fall in A. The integral of an $f : \mathbb{R}^n \to \mathbb{R}_+$ with respect to M is

$$Mf = \sum_{i=1}^K f(X_i) = \sum_{i=1}^\infty f(X_i)\mathbb{1}_{\{K\geq i\}}.$$

The mean of the random variable Mf is

$$\mathbb{E}Mf = \sum_{i=1}^{\infty} \mathbb{E}f(X_i)\mathbb{E}\mathbb{1}_{\{K\geq i\}} = (\lambda f)\mathbb{E}K = c\cdot(\lambda f),$$

namely, c times the value $\mathbb{E}f(X_i)$. To compute the Laplace functional of M, note

$$e^{-Mf} = e^{-\sum_{i=1}^{K} f(X_i)} = \prod_{i=1}^{K} e^{-f(X_i)}.$$

Because K and $\{X_i\}$ are independent, we can first take the expectation of the terms in the product, to get

$$\left(\mathbb{E}e^{-f(X_1)}\right)\left(\mathbb{E}e^{-f(X_2)}\right)\cdots\left(\mathbb{E}e^{-f(X_K)}\right) = (\lambda e^{-f})(\lambda e^{-f})\cdots(\lambda e^{-f}) = (\lambda e^{-f})^K.$$

Then we take the expectation with respect to K, to get

$$\begin{aligned}\mathbb{E}e^{-Mf} &= \mathbb{E}(\lambda e^{-f})^K = \sum_{k=0}^{\infty}\frac{c^k}{k!}e^{-c}\cdot(\lambda e^{-f})^k \\ &= \sum_{k=0}^{\infty}\frac{(c\cdot\lambda e^{-f})^k}{k!}e^{-c} = e^{c\lambda e^{-f}}\cdot e^{-c} \\ &= e^{-c\lambda(1-e^{-f})},\end{aligned}$$

where we recall $\lambda(1) = \int_{\mathbb{R}} d\lambda = 1$.

4.2 Poisson Random Measures

First a small note: we extend the definition of Poisson distribution to mean $c = 0$ and $c = \infty$. For $c = 0$, $X = 0$ almost surely and for $c = +\infty$, $X = +\infty$ almost surely. Recall if X_1 and X_2 are independent Poisson wih mean c_1 and c_2, then $X_1 + X_2$ is Poisson with mean $c_1 + c_2$.

Definition 4.6. Let $(E, \mathcal{E})$ be a measureable space and let ν be a measure on it. A random measure N on $(E, \mathcal{E})$ is said to be *Poisson* with mean ν if

1. $N(A)$ is Poisson distributed with mean $\nu(A)$ for every $A \in \mathcal{E}$;
2. if $A_1, \ldots, A_n$ are disjoint, then $N(A_1), \ldots, N(A_n)$ are independent.

As we can see, the random measure in Example 4.5 is exactly a Poisson random measure with mean $c\lambda$. For a discrete space like $(\mathbb{N}, 2^{\mathbb{N}})$, we can define a Poisson random measure as followes. Let ν be some measure on $\mathbb{N}$ and for each $n \in \mathbb{N}$ let W_n be a Poisson distributed random variable with mean $\nu(\{n\})$. Assume $\{W_n\}$ is an independency and define

$$N(A) = \sum_{n\in A} W_n, \quad A \subset \mathbb{N}.$$

This is a Poisson random measure on $\mathbb{N}$ with mean ν.

Let's now do some small computations for a Poisson random measure N on $\mathbb{R}^2$ with mean $\nu = c\cdot\mu_L$.

1. What is the probability distribution of the distance R from the origin to the nearest atom? $R > r$ if and only if $N(B_r) = 0$, where B_r is the closed disk centered at the origin. Thus

$$\mathbb{P}\{R > r\} = \mathbb{P}\{N(B_r) = 0\} = e^{-\nu(B_r)} = e^{-c\cdot\pi r^2}, \quad r \in \mathbb{R}_+.$$

2. Imagine atoms as centers of small disks of radius a. What is the distribution of the distance V from the origin to the nearest disk along the positive x-axis? We have $V > x$ if and only if $N(D_x) = 0$ where $D_x = [0, x] \times [-a, a]$. Thus
$$\mathbb{P}\{V > x\} = \mathbb{P}\{N(D_x) = 0\} = e^{-\nu(D_x)} = e^{-c\cdot 2ax}.$$

It can be calculated that for $f \in \mathcal{E}_+$ one has $\mathbb{E}Nf = \nu f$ and $\mathrm{Var}Nf = \nu(f^2)$. The first one on mean is clear, and for the one on variance, we prove as follows:

1. For $f = a\mathbb{1}_A$, we have $Nf = aN(A) \Rightarrow \mathrm{Var}Nf = \mathrm{Var}[aN(A)] = a^2\mathrm{Var}[N(A)] = a^2\nu(A) = \nu(a\mathbb{1}_A)^2$.
2. For simple function $f = \sum_{i=1}^n a_i\mathbb{1}_{A_i} = \sum_{i=1}^n f_i$ where $A_1, \ldots, A_n$ are disjoint, we have $Nf = a_1N(A_1) + \cdots + a_nN(A_n)$ so
$$\begin{aligned}\mathrm{Var}Nf &= \mathrm{Var}\left[a_1N(A_1) + \cdots + a_nN(A_n)\right]\\ &= a_1^2\mathrm{Var}[N(A_1)] + \cdots + a_n^2\mathrm{Var}[N(A_n)]\\ &= a_1^2\nu(A_1) + \cdots + a_n^2\nu(A_n)\\ &= \nu(a_1^2\mathbb{1}_{A_1} + \cdots + a_n^2\mathbb{1}_{A_n}).\end{aligned}$$
The last term is equal to f^2, because for $i \neq j$, $A_i \cap A_j = \varnothing$ so $f_if_j = 0 \Rightarrow f^2 = f_1^2 + \cdots + f_n^2$.
3. For $f \in \mathcal{E}_+$, there exists a sequence of simple functions $(f_n) \subset \mathcal{E}_+$ such that $f_n \uparrow f$. Using continuity of $f \mapsto \mathrm{Var}Nf$,
$$\mathrm{Var}Nf = \mathrm{Var}N\left(\lim_{n\to\infty} f_n\right) = \lim_{n\to\infty}\mathrm{Var}Nf_n = \lim_{n\to\infty}\nu(f_n^2) = \nu\left(\lim_{n\to\infty} f_n^2\right) = \nu(f^2).$$

Theorem 4.7. *N is Poisson with mean ν if and only if*
$$\varphi(f) = \mathbb{E}e^{-Nf} = e^{-\nu(1-e^{-f})}, \quad f \in \mathcal{E}_+. \tag{21}$$

Proof. Suppose N is Poisson with mean ν. For $f = a\mathbb{1}_A$ with $\nu(A) < \infty$, we have $Nf = a \cdot N(A)$ where $N(A)$ has Poisson distribution with mean $\nu(A)$, so
$$\begin{aligned}\mathbb{E}e^{-Nf} &= \sum_{k=0}^{\infty}\frac{\nu(A)^k}{k!}e^{-\nu(A)}\cdot e^{-ak}\\ &= \sum_{k=0}^{\infty}\frac{[\nu(A)\cdot e^{-a}]^k}{k!}e^{-\nu(A)}\\ &= e^{\nu(A)e^{-a}}e^{-\nu(A)}\\ &= e^{-\nu(A)(1-e^{-a})}\\ &= \exp\{-\nu\mathbb{1}_A(1-e^{-a})\} \leftarrow \text{integral of } \mathbb{1}_A \text{ times a constant}\\ &= \exp\left\{-\nu(1-e^{-a\mathbb{1}_A})\right\} \leftarrow \text{integral of the function } (1-e^{-a\mathbb{1}_A})\\ &= e^{-\nu(1-e^{-f})}.\end{aligned}$$

The result remains true even when $\nu(A) = +\infty$. Next, if $f \in \mathcal{E}_+$ is simple, say $f = \sum_{i=1}^n a_i\mathbb{1}_{A_i} = \sum_{i=1}^n f_i$ where

$A_1, \ldots, A_n$ are disjoint, then $Nf_1, \ldots, Nf_n$ are independent by definition of Poisson random measures, so

$$\begin{aligned}\mathbb{E}e^{-Nf} &= \prod_{i=1}^{n} \mathbb{E}e^{-Nf_i} = \exp\left\{-\nu(1-e^{-f_1})\right\} \cdots \exp\left\{-\nu(1-e^{-f_n})\right\} \\ &= \exp\left\{-\nu\left(n - \sum_{i=1}^{n} e^{-f_i}\right)\right\} \\ &= \exp\left\{-\nu\left(1 - e^{-\sum_{i=1}^{n} f_i}\right)\right\} \\ &= \exp\left\{-\nu\left(1-e^{-f}\right)\right\}.\end{aligned}$$

The two terms in red are equal because, when $x \notin A_1 \cup \cdots \cup A_n$, we have $f_i = 0 \quad \forall i = 1, \ldots, n$ so $e^{-f_i} = 1 \quad \forall i = 1, \ldots, n \Rightarrow \sum_{i=1}^{n} e^{-f_i} = n \Rightarrow n - \sum_{i=1}^{n} e^{-f_i} = 0$ and for the second term it is $1 - e^0 = 1 - 1 = 0$ also. If $x \in A_j$ for some j, then $f_j = 1$ and $f_i = 0$ for $i \neq j$, so $\sum_{i=1}^{n} e^{-f_i} = (n-1) + f_j \Rightarrow n - [(n-1) + e^{-f_j}] = 1 - e^{-f_j}$, which agrees with the second term in this case.

Finally, let $f \in \mathcal{E}_+$ be arbitrary. Let $(f_n) \subset \mathcal{E}_+$ be a sequence of simple functions such that $f_n \uparrow f$. By Proposition 4.3,

$$\mathbb{E}e^{-Nf} = \mathbb{E}e^{-N \lim_{n\to\infty} f_n} = \lim_{n\to\infty} \mathbb{E}e^{-Nf_n} = \lim_{n\to\infty} e^{-\nu(1-e^{-f_n})}.$$

As $n \to \infty$, $1 - e^{-f_n}$ increases to $1 - e^{-f}$, and $\nu(1 - e^{-f_n})$ increases to $\nu(1 - e^{-f})$ due to monotone convergence theorem. Thus $\lim_{n\to\infty} e^{-\nu(1-e^{-f_n})} = e^{-\nu(1-e^{-f})}$. □

The following proposition is quite obvious. Since N has mean ν, if $\nu(\mathbb{1}_E) = \nu(E) = +\infty$ then $N(\mathbb{1}_E) = N(E) = +\infty$ also. If $\nu(E) < \infty$ then $N(E) < \infty$ also.

Proposition 4.8. Let N be a Poisson random measure on $(E, \mathcal{E})$ with mean ν. Let $f \in \mathcal{E}_+$.

1. If $\nu(f \wedge 1) = +\infty$, then $Nf = +\infty$ almost surely.
2. If $\nu(f \wedge 1) < \infty$, then $Nf < \infty$ almost surely.

Theorem 4.9. Let N be a Poisson random measure on $(E, \mathcal{E})$ with mean ν and suppose ν is Σ-finite. Then N is a random counting measure if and only if ν is diffuse.

4.3 Transformations

We remind the reader what is a transition kernel: let $(S, \mathcal{S}), (T, \mathcal{T})$ be two measurable spaces. A function

$$\kappa : S \times \mathcal{T} \to [0, +\infty]$$

is called a transition kernel if

1. fix the left coordinate, you get a measure on $(T, \mathcal{T})$;
2. fix the right coordinate, you get a measurable function.

A *transition probability kernel* is a transition kernel such that $\kappa(s, \cdot) \in [0, 1]$ is a probability measure on T for every $s \in S$.

Theorem 4.10. Let $X = \{X_i : i \in I\}$ be a collection of random variables that form a Poisson random measure N with mean ν on $(E, \mathcal{E})$, and let $Y = \{Y_i : i \in I\}$ be a collection of random variables taking values in $(F, \mathcal{F})$. Let $Q : E \times \mathcal{F} \to [0, 1]$ be a transition probability kernel from $(E, \mathcal{E})$ to $(F, \mathcal{F})$. If given X, the variables $\{Y_i\}$ are conditionally independent and have the respective distribution $Q(X_i, \cdot)$, then

1. Y forms a Poisson random measure on $(F, \mathcal{F})$ with mean νQ;
2. (X, Y) forms a Poisson random measure on $(E \times F, \mathcal{E} \otimes \mathcal{F})$ with mean $\nu \times Q$.

We remind the reader that $\nu \times Q$ is the product measure on $E \times F$, i.e. $(\nu \times Q)(dx, dy) = \nu(dx) \cdot Q(x, dy)$, and νQ is the marginal of $\nu \times Q$ on F, i.e. $(\nu Q)(dy) = \int_E \nu(dx) Q(x, dy)$.

Proof. Let M be the random measure formed by (X, Y) on $E \times F$. The random measure formed by Y on F is the image of M under $h(x, y) = y$, so we shall only prove 2. We prove by showing that the Laplace functional of M has the form of Eq. (21). For positive f in $\mathcal{E} \otimes \mathcal{F}$, $Mf = \sum_{i \in I} f(X_i, Y_i)$ so

$$e^{-Mf} = \prod_{i \in I} e^{-f(X_i, Y_i)}.$$

Since Y_i's are conditionally independent given X, the conditional expectation of e^{-Mf} given $X = \{X_i\}$ is

$$\mathbb{E}\left[e^{-Mf} | X\right] = \prod_{i \in I} \int_F e^{-f(X_i, y)} Q(X_i, dy) = \prod_{i \in I} e^{-g(X_i)} = e^{-Ng},$$

where $g(x)$ is defined by

$$e^{-g(x)} = \int_F e^{-f(x,y)} Q(x, dy).$$

Now take the expectation to get

$$\mathbb{E} e^{-Mf} = \mathbb{E} e^{-Ng} = e^{-\nu(1 - e^{-g})}$$

where we used Theorem 4.7 for the Poisson random measure N on E. Now

$$\begin{aligned}
\nu(1 - e^{-g}) &= \int_E \nu(dx) \left(1 - \int_F Q(x, dy) e^{-f(x,y)}\right) \\
&= \int_E \nu(dx) \left(\int_F Q(x, dy) - \int_F Q(x, dy) e^{-f(x,y)}\right) \\
&= \int_E \nu(dx) \int_F Q(x, dy) \left(1 - e^{-f(x,y)}\right) \\
&= (\nu \times Q)(1 - e^{-f}).
\end{aligned}$$

So we see that

$$\mathbb{E} e^{-Mf} = e^{-(\nu \times Q)(1 - e^{-f})}.$$

This completes the proof that M is a Poisson random measure on $E \times F$ with mean $\nu \times Q$. □

Corollary 4.11. If X forms a Poisson random measure on E with mean ν and Y is independent of X and is an independency of variables with distribution π on $(F, \mathcal{F})$, then (X, Y) forms a Poisson random measure on $(E \times F, \mathcal{E} \otimes \mathcal{F})$ with mean $\nu \times \pi$.

Proof. Take $Q(x, \cdot)$ to be $\pi(\cdot)$ in Theorem 4.10. □

Example 4.12 (Compound Poisson Process)**.** Let's do some computation of a model of customer arrival. Suppose the arrival time $\{T_i\}_{i=1}^{\infty}$ of customers at a store form a Poisson random measure N on $\mathbb{R}_+$ with intencity c, i.e. the mean measure is $\nu = c \cdot \mu_L$. Each customer spends, independently of each other, a random amount of money Y_i at the store which is distributed as π, with mean m and variance σ^2. In other words, $\{Y_i\}$ and $\{T_i\}$ are independent and furthermore $\{Y_i\}$ is an independency. We are interested in knowing the distribution of

$$Z_t = \sum_{i=1}^{\infty} Y_i \mathbb{1}_{(0,t]}(T_i),$$

the total amount of purchase before time t. From the preceding corollary, $\{(T_i, Y_i)\}_{i=1}^{\infty}$ forms a Poisson random measure M on $\mathbb{R}_+ \times \mathbb{R}_+$ with mean $\nu \times \pi$. The random measure puts weight 1 on each of the point (T_i, Y_i) in the space $\mathbb{R}_+^2$. This means, for $A \subset \mathbb{R}_+^2$ measurable, one has

$$M(A) = \sum_{i=1}^{\infty} \mathbb{1}_A(T_i, Y_i),$$

the count of those (T_i, Y_i) falling into A, and $M(A)$ has Poisson distribution with mean $(\nu \times \pi)(A)$. In particular, if $A = [0, t] \times B$, then $M(A)$ is the count of those in $\{(T_i, Y_i)\}_{i=1}^{\infty}$ who fall before time t and at the same time have sizes in B. It is Poisson distributed and its mean is $ct \cdot \pi(B)$. For $f : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ measurable,

$$Mf = \sum_{i=1}^{\infty} f(T_i, Y_i).$$

For $f(x, y) = \mathbb{1}_{[0,t]} \cdot y$, we have $f(T_i, Y_i) = \mathbb{1}_{[0,t]}(T_i) \cdot Y_i$, so that

$$Mf = M\left(\mathbb{1}_{[0,t]} \cdot y\right) = \sum_{i=1}^{\infty} Y_i \mathbb{1}_{(0,t]}(T_i),$$

which is exactly Z_t. Thus, we can represent Z_t as

$$Z_t = \int_{\mathbb{R}_+ \times \mathbb{R}_+} \mathbb{1}_{[0,t]} \cdot y M(dt, dy) = Mf$$

with $f = \mathbb{1}_{[0,t]} \cdot y$. Some information on distribution of Z_t:

$$\begin{aligned}
\mathbb{E}Z_t &= (\nu \times \pi)(f) = \nu\left(\mathbb{1}_{[0,t]}\right) \cdot \pi(y) = ct \cdot m, \\
\mathrm{Var}Z_t &= (\nu \times \pi)(f^2) = \nu\left(\mathbb{1}_{[0,t]}\right) \cdot \pi(y^2) = ct \cdot (m^2 + \sigma^2), \\
\mathbb{E}e^{-rZ_t} &= \mathbb{E}e^{-M(rf)} = \exp\left\{-ct \int_{\mathbb{R}_+} (1 - e^{-ry})\pi(dy)\right\}.
\end{aligned}$$

The process $Z = (Z_t)_{t \in \mathbb{R}_+}$ is an example of a *compound Poisson Process*.

4.4 Additive Random Measures and Lévy Processes

Definition 4.13. Let M be a random measure on E. It is said to be *additive* if $M(A_1), \ldots, M(A_n)$ are independent for all choices of finitely many disjoint sets $A_1, \ldots, A_n$ in $\mathcal{E}$.

Deterministic measures and Poisson random measures are of course additive. Below are two more.

1. (Fixed atoms, random weights) If $D \subset E$ is countable and $\{W_x : x \in D\}$ is an independency of positive random variables, then the random measure defined by

$$K(A) = \sum_{x \in D} W_x \mathbb{1}_A(x)$$

is an additive random measure.

2. (Random atoms, fixed weights) Let N be a Poisson random measure on $E \times \mathbb{R}_+$ with mean ν. Define

$$L(A) = \int_{A \times \mathbb{R}_+} y N(dx, dy) = N(\mathbb{1}_A \cdot y).$$

Then L is an additive random measure on E. The Laplace transform for $L(A)$ is

$$\mathbb{E}e^{-rL(A)} = \exp\left\{-\int_{A\times\mathbb{R}_+}(1-e^{-ry})\nu(dx,dy)\right\}$$

by using Theorem 4.7.

Let α be a deterministic measure and suppose K and L are independent. Then $M = \alpha + K + L$ is an additive measure. Conversely, it can be shown that the preceding is, basically, the general form of an additive random measure.

Definition 4.14. Let $S = (S_t)_{t\in\mathbb{R}_+}$ be an increasing and right-continuous stochastic process with state space $\mathbb{R}_+$ and $S_0 = 0$. It is said to be an *increasing Lévy process* (or *subordinator*) if it has stationary and independent increments.

Given an additive random measure M on $\mathbb{R}_+$, putting $S_t = M([0,t])$ yields an increasing right-continuous process. Once we assure that $S_t < \infty$ almost surely for all t, independence of increments follows from the additivity of M. Stationarity of increments is achieved by making sure that the mean measure is chosen appropriately and there be no fixed atoms and the deterministic measure α be a constant multiple of the Lebesgue measure. In other words, the following proposition is in fact a complete characterization of increasing Lévy processes, Here we state the sufficiency part.

Proposition 4.15. Let $b \in \mathbb{R}_+$ be a constant and let N be a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}_+$ with mean measure $\nu = \mu_L \times \lambda$, where λ satisfies

$$\int_{\mathbb{R}_+}(y\wedge 1)\lambda(dy) < \infty. \tag{22}$$

Define

$$S_t = bt + \int_{[0,t]\times\mathbb{R}_+} yN(dt,dy) = bt + N\left(\mathbb{1}_{[0,t]}\cdot y\right).$$

Then $S = (S_t)_{t\in\mathbb{R}_+}$ is an increasing Lévy process on $\mathbb{R}_+$, and

$$\mathbb{E}e^{-rS_t} = \exp\left\{-t\left[br + \int_{\mathbb{R}_+}(1-e^{-ry})\lambda(dy)\right]\right\}.$$

Note $S_t = M([0,t])$ where $M = \alpha + L$.

Example 4.16 (Gamma process)**.** Let S be as in Proposition 4.15 with $b = 0$ and

$$\lambda(dy) = a\frac{e^{-cy}}{y}dy, \quad y > 0$$

for some constant $a, c \in (0, +\infty)$. Then λ satisfies Eq. (22) so S is an increasing Lévy process and

$$\begin{aligned}\mathbb{E}e^{-rS_t} &= \exp\left\{-t\int_0^\infty (1-e^{-ry})\cdot a\frac{e^{-cy}}{y}dy\right\}\\ &= \exp\left\{-at\cdot\log\frac{c+r}{c}\right\}\\ &= \left(\frac{c}{c+r}\right)^{at}.\end{aligned}$$

The integral can be worked out as follows:

$$\begin{aligned}\int_0^\infty \left(\frac{e^{-cy}}{y} - \frac{e^{-(r+c)y}}{y}\right) dy &= -\int_0^\infty \frac{e^{-zy}}{y}\Big|_r^{c+r} dy \\ &= \int_0^\infty \left(\int_c^{c+r} e^{-zy} dz\right) dy \\ &= \int_c^{c+r} \left(\int_0^\infty e^{-zy} dy\right) dz \\ &= \int_c^{c+r} \frac{1}{z} dz \\ &= \log \frac{c+r}{c}.\end{aligned}$$

Thus, S_t has a gamma distribution with shape index a and scale parameter c. For this reason S is called a *gamma process*.

Example 4.17 (Increasing stable process)**.** Let S be as in Proposition 4.15 with $b = 0$ and

$$\lambda(dy) = \frac{ac}{\Gamma(1-a)} z^{-1-a}, \quad z > 0,$$

where $a \in (0,1)$ and $c \in (0, +\infty)$. Again λ satisfies Eq. (22), so S is increasing Lévy. Even though $S_t < \infty$ almost surely,

$$\mathbb{E}S_t = t \int_{\mathbb{R}_+} y\lambda(dy) = t \cdot (+\infty) = +\infty$$

for every $t > 0$. The process S is called a *stable process with index* $a \in (0,1)$, because (S_{ut}) has the same distribution as $(u^{1/a} S_t)$ for every $u > 0$, as can be demonstrated from the Laplace transform

$$\mathbb{E}e^{-rS_t} = e^{-c \cdot t \cdot r^a}.$$

The distribution of S_t does not have an explicit form in general. However for $a = 1/2$ one has

$$\mathbb{P}\{S_t \in dy\} = \frac{ct}{\sqrt{4\pi y^3}} e^{-c^2 t^2/4y} dy.$$

4.5 Poisson Processes

Let $N = (N_t)_{t \in \mathbb{R}_+}$ be a counting process on $\mathbb{R}_+$. Then there is an increasing sequence of random variables (T_k) taking values in $\mathbb{R}_+$ such that

$$N_t = \sum_{k=1}^\infty \mathbb{1}_{[0,t]}(T_k), \quad t \in \mathbb{R}_+.$$

The sequence (T_k) forms a random counting measure M on $\mathbb{R}_+$, and

$$Mf = \sum_{k=1}^\infty f(T_k)$$

for positive Borel functions. Indeed $N_t = M([0,t])$. Finally, let $\mathcal{F}$ be the filtration generated by N.

Theorem 4.18. Let $c \in (0, +\infty)$. The following are equivalent:

1. N is a Poisson process with rate c.

2. M is a Poisson random measure with mean $\nu = c \cdot \mu_L$.
3. $(N_t - ct)_{t \in \mathbb{R}_+}$ is an $\mathcal{F}$-martingale.
4. (T_k) is an increasing sequence of $\mathcal{F}$-stopping times, and the differences $T_1, T_2 - T_1, T_3 - T_2, \ldots$ are independent and exponentially distributed with parameter c.

The next theorem's characterization is often used as a definition: A Poisson process is a counting process with stationary and independent increments.

Theorem 4.19. The counting process N is a Poisson process if and only if it is a Lévy process.

5 Lévy Processes

5.1 Itô-Lévy Decomposition

Definition 5.1. X is a *Lévy process* in $\mathbb{R}^d$ with respect to $\mathcal{F}$ if it is adapted to $\mathcal{F}$, $X_0 = 0$, and

1. for almost every ω, the path $t \mapsto X_t(\omega)$ is right-continuous and left-limited;
2. the increment $X_{t+u} - X_t$ is independent of $\mathcal{F}_t$ and has the same distribution as X_u for every $t, u \in \mathbb{R}_+$.

Example 5.2.

- The simplest and trivial Lévy process is the drift $X_t = bt$ for some $b \in \mathbb{R}^d$.
- A Wiener process W is a Lévy process in $\mathbb{R}$ that has continuous paths and has Gaussian distribution with mean 0 and variance u for the increment $W_{t+u} - W_t$. The most general continuous Lévy process in $\mathbb{R}$ has the form
$$X_t = bt + cW_t, \quad t \in \mathbb{R}_+.$$
A similar result holds for processes in $\mathbb{R}^d$.
- A Poisson process N with rate c is a Lévy process that is a counting process having the Poisson distribution with mean cu for the increments $N_{t+u} - N_t$.
- A compound Poisson process is a Lévy process. Let N be a Poisson process and (Y_n) an independency of identically distributed random variables. Then the process defined by
$$X_t = \sum_{n=1}^{\infty} Y_n \mathbb{1}_{\{n \le N_t\}}$$
is a compound Poisson process, in agreement with Example 4.12. Its every path is a step function; its jumps occur at jump times of N and the size of successive jumps are $Y_1, Y_2, \ldots$. It can be shown that, conversely, every Lévy process whose paths are step functions is a compound Poisson process.
- Increasing Lévy processes are Lévy processes with state space $\mathbb{R}_+$, because the positivity of S_u and stationarity of $X_{t+u} - X_t$ implies that every increment is positive. Increasing Lévy processes include Poisson process, compound Poisson process with positive jumps, gamma process and stable process with indices in $(0, 1)$.

Recall that a random variable is said to be *infinitely divisible* if, for every integer n, it can be written as the sum of n independent and identically distributed random variables. If X is a Lévy process, then for fixed $t > 0$ and $n \ge 1$, we can write X_t as the sum of the increments over $(0, \delta], (\delta, 2\delta], \ldots, ((n-1)\delta, n\delta]$ where $\delta = t/n$, and those increments are independent and identically distributed. Thus X_t is infinitely divisible for every t, and so is every increment $X_{t+u} - X_t$. It follows that the characteristic function of a Lévy process X has the form

$$\varphi(r) = \mathbb{E}e^{irX_t} = e^{t\phi(r)},$$

so that for $t = t_1 + \cdots + t_n$ and $X_t = X_{t_1} + \cdots + X_{t_n}$, we have

$$\begin{aligned}
\varphi(r) &= \mathbb{E}e^{ir(X_{t_1} + \cdots + X_{t_n})} = \mathbb{E}e^{irX_{t_1}} \cdots e^{irX_{t_n}} \\
&= \left(\mathbb{E}e^{irX_{t_1}}\right) \cdots \left(\mathbb{E}e^{irX_{t_n}}\right) \\
&= \varphi_{X_{t_1}}(r) \cdots \varphi_{X_{t_n}}(r) \\
&= e^{t_1\phi(r)} \cdots e^{t_n\phi(r)} \\
&= e^{(t_1 + \cdots + t_n)\phi(r)} \\
&= e^{t\phi(r)}.
\end{aligned}$$

$\phi(r)$ is called the *characteristic component* of X.

If X is a Lévy process, then it is possible that $\mathbb{E}X_t$ does not exist. This is the case, for instance, if X is a compound Poisson process and the Y_n do not have expected values. Or it is possible that $\mathbb{E}X_t$ is well defined but is equal to infinity. However, it means and variances of X are well-defined, then they must be linear in t, i.e.

$$\mathbb{E}X_t = at, \qquad \mathrm{Var}X_t = vt, \quad t \in \mathbb{R}_+,$$

which is a consequence of the stationarity and independence of the increments. For example, let $f(t) = \mathbb{E}X_t$ and since $X_t + X_t = X_{2t}$ we have $\mathbb{E}(X_t + X_t) = 2\mathbb{E}X_t = \mathbb{E}X_{2t} \Rightarrow 2f(t) = f(2t)$. Here a is a vector and v is a symmetric and positive definite $d \times d$ matrix.

Example 5.3 (Pure jump processes)**.** Let $\Delta X_t(\omega) = X_t(\omega) - X_{t-}(\omega)$ be the size of a jump at time t. Let D_ω be the discontinuity set for the path $X(\omega)$, that is,

$$D_\omega = \{t > 0 : \Delta X_t(\omega) \neq 0\}.$$

If X is continuous then D_ω is empty for almost every ω. If X is Poisson or compound Poisson, then D_ω is an infinite countable set, but $D_\omega \cap (s, u)$ is finite for all $0 \leq s < u < \infty$. For all other processes, the set D_ω is still infinite but $D_\omega \cap (s, u)$ is infinite for all $0 \leq s < u < \infty$.

A *pure-jump process* is a process X such that X_t is equal to the size of its jumps over $[0, t]$, i.e. for almost every ω,

$$X_t(\omega) = \sum_{s \in D_\omega \cap [0,t]} \Delta X_s(\omega), \quad t \in \mathbb{R}_+,$$

where $V_t(\omega) = \sum_{s \in D_\omega \cap [0,t]} |\Delta X_s(\omega)| < \infty$. $V_t(\omega)$ is called the total variation of the path $X(\omega)$ over $[0, t]$.

Every increasing Lévy process without drift is a pure-jump Lévy process, so is the difference of two such independent processes. The following constructs such processes in general. Conversely, every pure-jump Lévy process in $\mathbb{R}^d$ has the form given in this theorem.

Theorem 5.4 (Pure Jump Processes)**.** Let M be a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d$ with mean measure $\mu_L \times \lambda$, where the measure λ on $\mathbb{R}^d$ satisfies $\lambda\{0\} = 0$ and

$$\int_{\mathbb{R}^d} (|y| \wedge 1)\lambda(dy) < +\infty. \tag{23}$$

Then almost surely,

$$X_t = \int_{[0,t]\times\mathbb{R}^d} yM(dt, dy) = M\left(\mathbb{1}_{[0,t]} \cdot y\right)$$

convergences absolutely for every t, and the path X has bounded variation over $[0, t]$ for every $t \in \mathbb{R}_+$. The process X is a pure-jump process Lévy process in $\mathbb{R}^d$, and its characteristic exponent is

$$\phi(r) = \int_{\mathbb{R}^d} (e^{ir\cdot y} - 1)\lambda(dy), \quad r \in \mathbb{R}^d.$$

Several remarks:

- The condition that $\lambda\{0\} = 0$ is for reasons of convenience: to prevent linguistic faults like "jumps of size 0", and also to ensure that X and M uniquely determine each other.
- The measure λ determines the probability laws of M and X. It is called the *Lévy measure* of X. It regulates the jumps: for every Borel subset A of $\mathbb{R}^d$ with $\lambda(A) < \infty$, the jump times of X with corresponding sizes belonging to A form the counting process $t \mapsto M([0, t] \times A) = M(\mathbb{1}_{[0,t]} \cdot \mathbb{1}_A)$, and the latter is a Poisson process with rate

$\lambda(A)$ (because the mean is $\mathbb{E}M(\mathbb{1}_{[0,t]} \cdot \mathbb{1}_A) = \mu_L([0,t]) \cdot \lambda(A) = t \cdot \lambda(A)$). If we imagine some $\{(T_i, Y_i)\}$ forms M, then

$$M\left(\mathbb{1}_{[0,t]} \cdot \mathbb{1}_A\right) = \sum_{i=1}^{\infty} \mathbb{1}_{[0,t]}(T_i) \cdot \mathbb{1}_A(Y_i) = \#\{(T_i, Y_i) : 0 \le T_i \le t \text{ and } Y_i \in A\}.$$

If $\lambda(\mathbb{R}^d) = +\infty$, then this implies that X_t has infinitely many jumps all the time, as the following remark says.

- The condition Eq. (23) is essential. It is satisfied by every finite measure. More interesting are infinite measures that satisfy it; to such measures there correspond pure-jump processes that have infinitely many jumps during every interval (s,t) with $s < t$; but, of those jumps, only finitely many may exceed ε in magnitude no matter how small $\varepsilon > 0$ is. An example is the gamma process: for $\lambda(dy) = y^{-1}e^{-y}dy$, $y \in (0, +\infty)$ we have

$$\int_0^\infty (y \wedge 1)\frac{e^{-y}}{y}dy = \int_0^1 e^{-y}dy + \int_1^\infty \frac{e^{-y}}{y}dy < +\infty,$$

but

$$\lambda(\mathbb{R}_+) = \int_0^\infty \frac{e^{-y}}{y}dy \ge \int_0^1 \frac{e^{-y}}{y}dy \ge e^{-1}\int_0^1 \frac{1}{y}dy = +\infty.$$

Example 5.5 (Gamma processes)**.** Recall that a gamma process is a stochastic process with independent gamma-distributed increments. It is a pure-jump increasing Lévy process with intensity measure $\lambda(dx) = ax^{-1}\exp(-cx)dx$ for $x \in (0, +\infty)$. Those jumps whose size lies in the interval $[x, x + dx)$ occur as a Poisson process with intensity $\lambda(dx)$.

Let X^+ and X^- be independent gamma processes. Then

$$X = X^+ - X^-$$

is a pure-jump Lévy process in $\mathbb{R}$. It is called a two-sided gamma process. If they have the same shape rate a and scale parameter c, then the Lévy measure of X is given by

$$\lambda(dx) = a\frac{e^{-c|x|}}{|x|}dx, \quad x \in \mathbb{R} \setminus \{0\},$$

with $\lambda\{0\} = 0$. In this case it is called a symmetric gamma process. The distribution of X_t is not gamma and cannot be expressed explicitely. However, its characteristic function is

$$\mathbb{E}e^{irX_t} = \left(\frac{c}{c - ir}\right)^{at}\left(\frac{c}{c + ir}\right)^{at} = \left(\frac{c^2}{c^2 + r^2}\right)^{at}, \quad r \in \mathbb{R}.$$

The total variation process $V = X^+ + X^-$ is a gamma process with shape rate $2a$ and scale parameter c.

Theorem 5.6 (Compensated Sum of Jumps)**.** Let B denote the closed unit ball in $\mathbb{R}^d$. Let M be a Poisson random measure on $\mathbb{R}_+ \times B$ with mean $\mu_L \times \lambda$, where the measure λ on B satisfies $\lambda\{0\} = 0$ and

$$\int_B |y|^2 \lambda(dy) < \infty. \tag{24}$$

For $\varepsilon \in (0,1)$, define

$$X_t^\varepsilon = \int_{[0,t]\times(B\setminus B_\varepsilon)} yM(dt, dy) - t\int_{B\setminus B_\varepsilon} y\lambda(dy), \quad t \in \mathbb{R}_+.$$

Then there exists a Lévy process X, such that

$$X_t = \lim_{\varepsilon \downarrow 0} X_t^\varepsilon$$

almost surely, the convergence being uniform in t over bounded intervals. The characteristic exponent of X is

$$\phi(r) = \int_B (e^{ir\cdot y} - 1 - ir \cdot y)\lambda(dy), \quad r \in \mathbb{R}^d.$$

We denote the limiting process X by

$$X_t = \int_{[0,t]\times B} yM(dt,dy) - t\int_B y\lambda(dy).$$

We see $X_t^\varepsilon = Y_t^\varepsilon - a_\varepsilon t$. The theorem says, if λ fails to satisfy Eq. (23), but satisfies Eq. (24), then as $\varepsilon \to 0$, Y_t^ε fails to converge and a_ε fails to converge, but their difference $X_t^\varepsilon = Y_t^\varepsilon - a_\varepsilon t$ converges to $X_t = Y_t - at$. The process X is called a compensated sum of jumps. In the simplest setting, where $y > 0$ is one-dimensional, failure of Eq. (23) amounts to

$$\int_0^1 y\lambda(dy) = +\infty,$$

which is the case if for example $\lambda(dy) = 1/y^2 dy$, but

$$\int_0^1 y^2\lambda(dy) = \int_0^1 y^2 \cdot \frac{1}{y^2}dy = \int_0^1 1dy < +\infty.$$

So for this kind of Lévy process X, it has infinite variation (i.e. $\int_B y\lambda(dy) = +\infty$) over every time interval (s,t).

The following theorem shows the construction of the most general Lévy process. It is the Itô-Lévy decomposition.

Theorem 5.7 (Itô-Lévy decomposition)**.** Let $b \in \mathbb{R}^d$, let c be a $(d \times d')$ matrix, and let λ be a measure on $\mathbb{R}^d$ with $\lambda\{0\} = 0$ and

$$\int_{\mathbb{R}^d} (|y|^2 \wedge 1)\lambda(dy) < \infty.$$

Let W be a d'-dimensional Wiener process, and independent of it, let M be a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d$ with mean $\mu_L \times \lambda$. Then

$$X_t = bt + cW_t + \left(\int_{[0,t]\times B} yM(dt,dy) - t\int_B y\lambda(dy)\right) + \int_{[0,t]\times\mathbb{R}^d\setminus B} yM(dt,dy)$$

defines a Lévy process in $\mathbb{R}^d$, and the characteristic exponent of X is, with $v = cc^T$,

$$\phi(r) = ib\cdot r - \frac{1}{2}r\cdot vr + \int_B (e^{ir\cdot y} - 1 - ir\cdot y)\lambda(dy) + \int_{\mathbb{R}^d\setminus B}(e^{ir\cdot y}-1)\lambda(dy), \quad r\in\mathbb{R}^d. \tag{25}$$

Eq. (25) is called the Lévy–Khinchine formula. Note that if λ satisfies Eq. (23), then the integral

$$a = \int_B y\lambda(dy)$$

converges absolutely and the process X becomes

$$X_t = (b-a)t + cW_t + \int_{[0,t]\times\mathbb{R}^d} yM(dt,dy), \quad t\in\mathbb{R}_+,$$

a drift plus a (continuous) Wiener process plus a pure-jump process. Accordingly, the characteristic exponent becomes

$$\phi(r) = ir\cdot(b-a) - \frac{1}{2}r\cdot vr + \int_{\mathbb{R}^d}(e^{ir\cdot y}-1)\lambda(dy), \quad r\in\mathbb{R}^d.$$

5.2 Stable Processes

Let $a \in \mathbb{R}_+$ and let $X = (X_t)_{t\in\mathbb{R}_+}$ be a Lévy process in $\mathbb{R}^d$. Then X is said to be a-stable, or stable with index a, or self-similar with index a, if the process $\hat{X} = (s^{-1/a}X_{st})_{t\in\mathbb{R}_+}$ has the same probability law as X for every $s \in (0,\infty)$. The condition is also equivalent to that $s^{-1/a}X_s$ having the same distribution as X_1, or that X_t and $t^{1/a}X_1$ having the same distribution.

It can be shown that the index a cannot exceed 2. If $X = W$ or $X = cW$ then X is stable with index 2, namely X_t has the same distribution as $\sqrt{t}X_1$ for every t. If $a \in (0,1)$, then the process is necessarily a pure-jump Lévy process whose Lévy measure is infinite and has a specific form. If $a \in (1,2)$, then the Lévy measure is again infinite and has a specific form, and the paths have infinite variation over every time interval and cannot be pure-jump type. If $a = 1$, there are three possibilities: the process can be pure drift and thus deterministic; or it can be a Cauchy process, the paths having the same qualitative features as in the case of indices in $(1,2)$, but each increment having a Cauchy distribution; or it can be a Cauchy process plus some drift.

Example 5.8 (Standard Cauchy process on $\mathbb{R}$)**.** A Cauchy process $X = (X_t)$ taking values in $\mathbb{R}$ is a Lévy, symmetric and stable process with index 1. The distribution is

$$\mathbb{P}\{X_t \in dx\} = \frac{t}{\pi(t^2 + y^2)}dy, \quad y \in \mathbb{R}.$$

The Lévy measure is

$$\lambda(dy) = \frac{1}{\pi y^2}dy, \quad y \in \mathbb{R}.$$

We have

$$\int_{B\setminus B_\varepsilon} y\lambda(dy) = \int_{B\setminus B_\varepsilon} \frac{1}{\pi y}dy = \int_{-1}^{-\varepsilon} \frac{1}{\pi y}dy + \int_{\varepsilon}^{1} \frac{1}{\pi y}dy = 0,$$

so according to Theorem 5.6, almost surely

$$X_t = \lim_{\varepsilon\downarrow 0} \int_{[0,t]\times\mathbb{R}\setminus(-\varepsilon,\varepsilon)} yM(dt,dy) := \int_{[0,t]\times\mathbb{R}} yM(dt,dy).$$

It follows that

$$\begin{aligned}
\mathbb{E}e^{irX_t} &= \exp\left\{t\lim_{\varepsilon\downarrow 0}\int_{\mathbb{R}\setminus(-\varepsilon,\varepsilon)} (e^{iry}-1)\lambda(dy)\right\} \\
&= \exp\left\{2t\int_{\mathbb{R}_+} \frac{(1-\cos ry)}{\pi y^2}dy\right\} \\
&= e^{-t|r|}.
\end{aligned}$$

The Cauchy process is not a pure-jump process, because

$$\int_{(0,1)} y\lambda(dy) = \int_{(-1,0)} (-y)\lambda(dy) = +\infty$$

and it follows that

$$\int_{(s,t)\times(0,1)} yM(dt,dy) = \int_{(s,t)\times(-1,0)} (-y)M(dt,dy) = +\infty$$

for every $s < t$. In other words, over every time interval (s,t), the path X has infinitely many upword jumps whose sizes sum up tp $+\infty$, and infinitely many downward jumps whose sizes sum up tp $-\infty$. In particular, the total variation over (s,t) is equal to $+\infty$ always.

The process X is not a martingale for the simple reason that $\mathbb{E}X_t$ does not exist.

5.3 Lévy Processes on Standard Settings

Recall the following technical definitions:

- A probability space $(\Omega, \mathcal{H}, \mathbb{P})$ is *complete* if for all $A \in \mathcal{H}$ with $\mathbb{P}(A) = 0$, one has $A' \in \mathcal{H}$ for all $A' \subset A$.

- An *extended* filtration $\mathcal{F}$ over $\mathbb{R}_+$ is one that includes $\mathcal{F}_\infty = \lim \mathcal{F}_t = \vee_t \mathcal{F}_t$, the σ-algebra generated by unions of all $\mathcal{F}_t$. We say $\mathcal{F}$ is *right-continuous* if

$$\mathcal{F}_t = \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}, \quad t \in \mathbb{R}_+.$$

  Heuristically, this means that $\mathcal{F}_t$ includes all events that can be told by an "infinitesimal peek" into the future. $\mathcal{F}$ is *augmented* if $(\Omega, \mathcal{H}, \mathbb{P})$ is complete and that $\mathcal{F}_0$ (and therefore all $\mathcal{F}_t$) contains the collection of all negligible events in $\mathcal{H}$.

Definition 5.9. A stochastic base is a collection

$$\mathcal{B} = (\Omega, \mathcal{H}, \mathcal{F}, \theta, \mathbb{P})$$

where $(\Omega, \mathcal{H}, \mathbb{P})$ is a complete probability space, $\mathcal{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$ is an augmented right-continuous filtration on it, and $0 = \{\theta_t\}_{t \in \mathbb{R}_+}$ is a semigroup of operators on Ω such that

$$\theta_0 \omega = \omega, \quad \theta_u \circ \theta_t \omega = \theta_{u+t} \omega, \quad t, u \in \mathbb{R}_+.$$

$\{\theta_t\}_{t \in \mathbb{R}_+}$ are called *time-shifts*.

Definition 5.10. Let $X = (X_t)_{t \in \mathbb{R}_+}$ be a stochastic process with state space $\mathbb{R}^d$. It is called a Lévy process over $\mathcal{B}$ if it is adapted to $\mathcal{F}$ and the following hold:

1. Regularity. X is right-continuous and left-limited, and $X_0 = 0$.
2. Additivity. $X_{t+u} = X_t + X_u \circ \theta_t$ for every $t, u \in \mathbb{R}_+$.
3. Lévy property. For every $t, u \in \mathbb{R}_+$, the increment $X_u \circ \theta_t$ is independent of $\mathcal{F}_t$ and has the same distribution as X_u.

Note that if a process X is additive, and $Z_t = Z_0 + X_t$, then $\boxed{Z_{t+u} = Z_u \circ \theta_t}$. This is because $Z_{t+u} = Z_0 + X_{t+u} = Z_0 + X_t + X_u \circ \theta_t = Z_t + X_u \circ \theta_t$, while $Z_u \circ \theta_t = (Z_0 + X_u) \circ \theta_t = Z_0 \circ \theta_t + X_u \circ \theta_t = Z_t + X_u \circ \theta_t$.

Theorem 5.11 (Markov property)**.** Suppose X is a Lévy process over $\mathcal{B}$, and let $\mathcal{G}$ denote the filtration generated by X. Then for every time t, the process $X \circ \theta_t$ is independent of $\mathcal{F}_t$ and has the same law as X. Equivalently, for every bounded random variable $V \in \mathcal{G}_\infty$,

$$\mathbb{E}_t V \circ \theta_t = \mathbb{E} V, \quad t \in \mathbb{R}_+.$$

Theorem 5.12 (Strong Markov property)**.** Let $\bar{\mathcal{G}}_\infty = \mathcal{G}_\infty \vee \mathcal{N}$ where $\mathcal{N}$ is the σ-algebra generated by the collection of negligible events in $\mathcal{H}$. If X is a Lévy process over $\mathcal{B}$, T is a stopping time of $\mathcal{F}$, then for every bounded random variable $V \in \bar{\mathcal{G}}_\infty$,

$$\mathbb{E}_T V \circ \theta_T \mathbb{1}_{\{T < \infty\}} = (\mathbb{E} V) \mathbb{1}_{\{T < \infty\}}.$$

6 Brownian Motion

6.1 Introduction

Definition 6.1. A stochastic process $X = (X_t)_{t\in\mathbb{R}_+}$ with state space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is calleed a *Brownian motion* if it is continuous and has stationary and independent increments. A process $W = (W_t)_{t\in\mathbb{R}_+}$ is called a *Wiener process* if it is a Brownian motion with

$$W_0 = 0, \quad \mathbb{E}W_t = 0, \quad \mathrm{Var}W_t = t, \quad t \in \mathbb{R}_+.$$

If X is a Brownian motion, then it follows from the remarks in Example 5.2 that X has the form

$$X_t = X_0 + at + bW_t, \quad t \in \mathbb{R}_+.$$

Theorem 6.2. $W = (W_t)_{t\in\mathbb{R}_+}$ is a Wiener process if and only if it is continuous and is a Gaussian process with mean 0 and

$$\mathrm{Cov}(W_s, W_t) = s \wedge t, \quad s, t \in \mathbb{R}_+.$$

Theorem 6.3. Let W be a Wiener process. Then the following hold:

1. Symmetry. The process $(-W_t)_{t\in\mathbb{R}_+}$ is again a Wiener process.
2. Scaling. $\hat{W} = (c^{-1/2}W_{ct})_{t\in\mathbb{R}_+}$ is a Wiener process for each $c \in (0, +\infty)$, i.e. W is stable with index 2.
3. Time inversion. Putting $\tilde{W}_0 = 0$ and $\tilde{W}_t = tW_{1/t}$ for $t > 0$ yields a Wiener process $(\tilde{W}_t)_{t\in\mathbb{R}_+}$.

Proof. Symmetry and scaling properties are immediate from the definition for Wiener processes. To show 3, we start by noting that $\{\tilde{W}_t : t > 0\}$ is a continuous Gaussian process with mean 0 and $\mathrm{Cov}(\tilde{W}_s, \tilde{W}_t) = s \wedge t$ for $s, t > 0$. For example, for $s < t$, we have $1/s > 1/t$, so

$$\mathrm{Cov}(\tilde{W}_s, \tilde{W}_t) = (st) \cdot \mathrm{Cov}(W_{1/s}, W_{1/t}) = st \cdot (1/t) = s = s \wedge t.$$

For $s > t$,

$$\mathrm{Cov}(\tilde{W}_s, \tilde{W}_t) = (st) \cdot \mathrm{Cov}(W_{1/s}, W_{1/t}) = st \cdot (1/s) = t = s \wedge t.$$

Thus, 3 will follow from Theorem 6.2 once we show that $\tilde{W}$ is continuous at time 0, that is, almost surely,

$$\lim_{t\downarrow 0} tW_{1/t} = 0.$$

We show the equivalent condition that $W_t/t \to 0$ almost surely as $t \to \infty$. To this end, we start by noting that if $n \geq 0$ is an integer and $t \in (n, n+1]$, then

$$\left|\frac{1}{t}W_t\right| \leq \frac{1}{n}|W_n + (W_t - W_n)| \leq \left|\frac{1}{n}W_n\right| + \frac{1}{n}\sup_{0\leq s\leq 1}|W_{n+s} - W_n|. \tag{26}$$

By strong law of large numbers, $W_n/n \to 0$ almost surely, since $W_n = W_1 + \cdots + W_1$ and each of the n copies of W_1 has mean $\mathbb{E}W_1 = 0$. Next, by Kolmogorov's inequality (continuous time version),

$$\mathbb{P}\left\{\frac{1}{n}\sup_{0\leq s\leq 1}|W_{n+s} - W_n| > \varepsilon\right\} \leq \frac{1}{n^2\varepsilon^2}\mathbb{E}|W_{n+1} - W_n|^2 = \frac{1}{n^2\varepsilon^2}.$$

Since $\sum 1/n^2$ is finite, the Borel-Cantelli lemma shows that the last term in Eq. (26) goes to 0 almost surely as $n \to \infty$. Hence $W_t/t \to 0$ almost surely as $t \to \infty$ and the proof is complete. □

Theorem 6.4 (Strong Markov property)**.** Let W be a Wiener process and let $\mathcal{G}$ be the filtration generated by W. Let T be a stopping time of $\mathcal{F}$. Then for every bounded random variable $V \in \mathcal{G}_\infty$,

$$\mathbb{E}_T V \circ \theta_T \mathbb{1}_{\{T<\infty\}} = (\mathbb{E}V)\mathbb{1}_{\{T<\infty\}}.$$

In particular, if $T < \infty$, then the process $W \circ \theta_T = (W_{T+u} - W_T)_{u\in\mathbb{R}_+}$ is independent of $\mathcal{F}_T$ and is again a Wiener process.

Theorem 6.5. Let T be an $\mathcal{F}$-stopping time, and let $U \in \mathcal{F}_T$ be a positive real-valued random variable. Then

$$\mathbb{E}_T f(W_{T+U} - W_T)\mathbb{1}_{\{T<\infty\}} = [\mathbb{E}f(W_U)]\,\mathbb{1}_{\{T<\infty\}}, \quad \forall f \in \mathcal{B}(\mathbb{R}) \text{ bounded.}$$

6.2 Hitting Times and Recurrence Times

Let W be a Wiener process over a stochastic base $\mathcal{B}$. We are interested in hitting times

$$T_a = \inf\{t > 0 : W_t > a\}, \quad a \in \mathbb{R}_+, \tag{27}$$

the first time that $W_t \in (a, +\infty)$. Let $\mathcal{G}$ be the (augmented) filtration generated by W.

Proposition 6.6. Almost surely, $T_0 = 0$.

Proof. According to Blumenthal's zero–one law, each event in $\mathcal{G}_0$ has probability zero or one. We have $\{T_0 = 0\} \in \mathcal{G}_0$. Note that $\{W_t > 0\}$ has probability $1/2$, and $\omega \in \{W_t > 0\} \Rightarrow \omega \in \{T_0 < t\}$ for every $t > 0$ implies that $\{W_t > 0\} \subset \{T_0 < t\}$, so $\mathbb{P}\{T_0 < t\} \geq 1/2$ for every $t > 0$. Letting $t \to 0$ concludes the proof. □

According to the above proposition, for almost every ω, and for every $\varepsilon > 0$, there is $u < \varepsilon$ such that $W_u > 0$; there is also $0 < s < \varepsilon$ such that $W_s(\omega) < 0$, this being by symmetry. Taking ε of the second phrase to be the time u of the preceding one, and recalling the continuity of the paths, we conclude that for every $\varepsilon > 0$ there are $0 < s < t < u < \varepsilon$ such that

$$W_s(\omega) < 0, \quad W_t(\omega) = 0, \quad W_u(\omega) > 0.$$

Interating the argument with s replacing ε yields

Corollary 6.7. For almost every ω, there are times $u_1 > t_1 > s_1 > u_2 > t_2 > s_2 > \cdots$ with limit 0 such that, for each n,

$$W_{s_n}(\omega) < 0, \quad W_{t_n}(\omega) = 0, \quad W_{u_n}(\omega) > 0.$$

Thus, the Wiener path $W(\omega)$ is highly oscillatory. Starting with $W_0(\omega) = 0$, the path spends no time at 0; it crosses under and over 0 infinitely many times during $(0, \varepsilon)$, no matter how small $\varepsilon > 0$ may be. By applying the corollary to the time inversion process in 3, we get

Corollary 6.8. For almost every ω there exist times $u_1 < t_1 < s_1 < u_2 < t_2 < s_2 < \cdots$ with limit $+\infty$ such that

$$\lim_{n\to\infty} W_{s_n}(\omega) = -\infty, \quad W_{t_n} = 0 \quad \forall n, \quad \lim_{n\to\infty} W_{u_n}(\omega) = +\infty.$$

In particular, the set $\{t \in \mathbb{R}_+ : W_t(\omega) = 0\}$ is unbounded.

We next explore the distribution of hitting times T_a.

Lemma 6.9. For $a > 0$, we have

$$\mathbb{P}\{T_a \leq t, W_t > a\} = \mathbb{P}\{T_a \leq t\}\mathbb{P}\{W_{t-T_a} > 0\} = \frac{1}{2}\mathbb{P}\{T_a \leq t\}.$$

Proof. Take $T = T_a$ and $U = (t - T_a)\mathbb{1}_{\{T_a \le t\}} \in \mathcal{F}_{T_a}$, and $f = \mathbb{1}_{(0,+\infty)}$ in Theorem 6.5, we get

$$\begin{aligned}
\mathbb{E}_{T_a}\left[\mathbb{1}_{(0,+\infty)}(W_{T_a+U} - W_{T_a})\,\mathbb{1}_{\{T_a \le t\}}\right] &= \mathbb{E}_{T_a}\left[\mathbb{1}_{(0,+\infty)}(W_t - a)\,\mathbb{1}_{\{T_a \le t\}}\right] \\
&= \mathbb{E}_{T_a}\left[\mathbb{1}_{(a,+\infty)}(W_t)\,\mathbb{1}_{\{T_a \le t\}}\right] \\
&= \mathbb{E}_{T_a}\left[\mathbb{1}_{\{T_a \le t\}}\mathbb{1}_{\{W_t > a\}}\right] \\
&= \mathbb{E}\left[f(W_U)\right]\mathbb{1}_{\{T_a \le t\}} \\
&= \mathbb{P}\left\{W_{t-T_a} > 0\right\}\mathbb{1}_{\{T_a \le t\}}.
\end{aligned}$$

Take expectations on both side of

$$\mathbb{E}_{T_a}\left[\mathbb{1}_{(a,+\infty)}(W_t)\,\mathbb{1}_{\{T_a \le t\}}\right] = \mathbb{P}\left\{W_{t-T_a} > 0\right\}\mathbb{1}_{\{T_a \le t\}}$$

we get

$$\mathbb{E}\,\mathbb{1}_{(a,+\infty)}(W_t)\,\mathbb{1}_{\{T_a \le t\}} = \mathbb{P}\{T_a \le t, W_t > a\} = \mathbb{P}\left\{W_{t-T_a} > 0\right\}\mathbb{P}\{T_a \le t\} = \frac{1}{2}\mathbb{P}\{T_a \le t\}.$$

□

In particular, since $\{W_t > a\} \subset \{T_a \le t\}$, the intersection of the two events is $\{W_t > a\} \cap \{T_a \le t\} = \{W_t > a\}$, so $\mathbb{P}\{T_a \le t\} = 2\mathbb{P}\{T_a \le t, W_t > a\} = 2\mathbb{P}\{W_t > a\} = \mathbb{P}\{|W_t| > a\}$. Thus

$$\begin{aligned}
\mathbb{P}\{T_a \le t\} = \mathbb{P}\{|W_t|^2 > a^2\} &= \mathbb{P}\left\{\left(\sqrt{t}W_1\right)^2 > a^2\right\} \\
&= \mathbb{P}\left\{tZ^2 > a^2\right\} \qquad \text{with } Z \sim N(0,1) \\
&= \mathbb{P}\left\{\frac{a^2}{Z^2} < t\right\}.
\end{aligned}$$

Thus T_a has the same distribution as a^2/Z^2. Since $Z \in \mathbb{R} \setminus \{0\}$ almost surely, $T_a \in (0, +\infty)$ almost surely. The density can be calculated as

$$\mathbb{P}\{T_a \in dt\} = \frac{ae^{-a^2/2t}}{\sqrt{2\pi t^3}}dt. \quad t > 0. \tag{28}$$

From the formula we can see that $\mathbb{E}T_a = +\infty$.

Proposition 6.10. Fix $a \in (0, +\infty)$ and define

$$T_{a-} = \inf\{t > 0 : W_t \ge a\} = \inf\{t > 0 : W_t = a\}.$$

Then T_{a-} is a stopping time of $\mathcal{G}$ and $T_{a-} = T_a$ almost surely.

Proof. The statement that T_{a-} is a stopping time of $\mathcal{G}$ is obvious. Clearly $T_{a-} \le T_a$ so $T_{a-} < +\infty$ almost surely, and $W \circ \theta_{T_{a-}}$ is again Wiener by the strong Markov property at T_{a-}. Thus by Proposition 6.6 we have $T_0 \circ \theta_{T-a} = 0$ almost surely, and so

$$T_a = T_{a-} + T_0 \circ \theta_{T-a} = T_{a-}$$

almost surely. □

We recall some facts from probability theory. Let X and Y be independent standard Gaussian. Then X^2 and Y^2 both have gamma distribution with shape index $1/2$ and scale index $1/2$. The random variable $A = X^2/(X^2 + Y^2)$ has beta distribution with index pair $(1/2, 1/2)$. This beta distribution is also called *arcsine distribution*, because

$$\mathbb{P}\{A \in du\} = \frac{\Gamma(1/2 + 1/2)}{\Gamma(1/2)\Gamma(1/2)}u^{\frac{1}{2}-1}(1-u)^{\frac{1}{2}-1}du = \frac{1}{\pi\sqrt{u(1-u)}}du$$

because $\Gamma(1/2) = \sqrt{\pi}$, and

$$\mathbb{P}\{A \le u\} = \int_0^u \frac{1}{\pi\sqrt{x(1-x)}} dx = \frac{2}{\pi} \arcsin\sqrt{u}, \quad 0 \le u \le 1.$$

Note that $A = X^2/(X^2+Y^2) \in [0,1]$, so $tA \in [0,t]$ and $t/A \in [t, +\infty]$.

Let

1. $G_t = \sup\{s \le t : W_s = 0\}$, the last time before t such that $W = 0$;
2. $D_t = \inf\{s > t : W_s = 0\}$, the first time after t such that $W = 0$.

Note $G_t \le t$ and $D_t > t$ for each $t \in \mathbb{R}_+$. Also note that, if $G_t < s < t$, then D_s cannot fall into the interval (s,t), otherwise the definition of G_t would be violated; thus $D_s > t$. Conversely, if $D_s > t$, then $W_{t'} \neq 0$ for all $t' \in (s,t)$, so in particular $G_t < s$ at least. Thus $\{G_t < s\} = \{D_s > t\}$.

Proposition 6.11. For each $t \in \mathbb{R}_+$, $G_t \sim tA$ and $D_t \sim t/A$.

Proof. Let X and Y be independent standard Gaussian variables. Recall $T_a \sim a^2/Y^2$. Consider $R_t = D_t - t$. If $W_t = x$, then R_t is the hitting time of the point $-x = -W_t$ by the path $W \circ \theta_t$. Since $W \circ \theta_t$ is independent of $\mathcal{F}_t$, so is $(-W) \circ \theta_t$, so we conclude from here that $R_t \sim (-W_t)^2/Y^2 = W_t^2/Y^2$. We may replace W_t by $\sqrt{t}X$ to obtain $R_t \sim tX^2/Y^2$. Thus,

$$D_t = t + R_t \sim t(X^2+Y^2)/Y^2 \sim t/A.$$

Finally,

$$\mathbb{P}\{G_t < s\} = \mathbb{P}\{D_s > t\} = \mathbb{P}\left\{\frac{s}{A} > t\right\} = \mathbb{P}\{tA < s\}$$

so that $G_t \sim tA$. □

For example, the probability of the event that $W_{t'} \neq 0$ during an interval $[s,t]$ can be calculated as

$$\mathbb{P}\{W_{t'} \neq 0 \quad \forall t' \in [s,t]\} = \mathbb{P}\{G_t < s\} = \mathbb{P}\left\{A < \frac{s}{t}\right\} = \frac{2}{\pi} \arcsin\sqrt{\frac{s}{t}}.$$

The variable $\boxed{R_t = D_t - t}$ is called *forward recurrence time*. It is the time needed for the process W to return to value 0 given that it is in W_t. $\boxed{Q_t = t - G_t}$ is called *backward rcurrence time*; it is the time needed to reach the point 0 if we go backward in time. It is shown in the proof that

$$R_t \sim tX^2/Y^2 = tC^2,$$

where C is standard Cauchy. Because A and $1-A$ have the same distribution, we have

$$\begin{aligned}\mathbb{P}\{Q_t \le s\} = \mathbb{P}\{t - tA \le s\} &= \mathbb{P}\left\{1 - A \le \frac{s}{t}\right\} = \mathbb{P}\left\{A \le \frac{s}{t}\right\} \\ &= \frac{2}{\pi} \arcsin\sqrt{\frac{s}{t}}, \quad 0 \le s \le t.\end{aligned}$$

Theorem 6.12. Let

$$A_t = \int_{[0,t]} \mathbb{1}_{\mathbb{R}_+} \circ W_s ds.$$

Then A_t has the same distribution as that of tA, where A has the arcsine distribution.

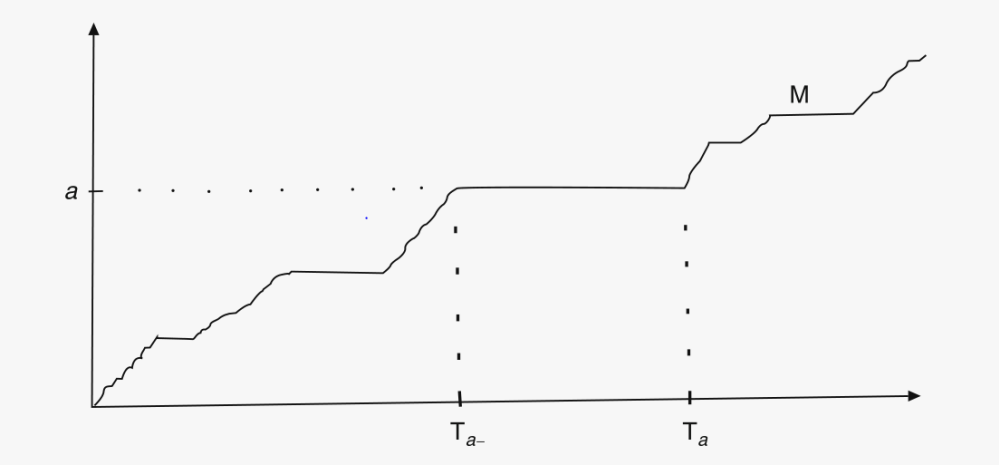

Figure 1: Relationship between T_a and M_t from the textbook (Çınlar 2011).

6.3 Hitting Times and Running Maximum

We are interested in the process $T = (T_a)_{a\in\mathbb{R}_+}$ of hitting times and its relationship with the process $M = (M_t)_{t\in\mathbb{R}_+}$ of running maximum, defined as

$$M_t = \max_{0\le s\le t} W_s, \quad t \in \mathbb{R}_+.$$

Note that the definition of hitting times Eq. (27) remains true when we replace W_t by M_t. Indeed, the paths $a \mapsto T_a(\omega)$ and $t \mapsto M_t(\omega)$ are functional inverses of each other:

$$T_a(\omega) = \inf\{t > 0 : M_t(\omega) > a\}, \quad M_t = \inf\{a > 0 : T_a(\omega) > t\}.$$

See Fig. 1 for an illustration from the textbook.

Proposition 6.13. For almost every ω, the path $a \mapsto T_a(\omega)$ is right-continuous, strictly increasing, real-valued, and with $T_0(\omega) = 0$ and $\lim_{a\to\infty} T_a(\omega) = +\infty$. For almost every ω, the path $t \mapsto M_t(\omega)$ is increasing, continuous, real-valued, and with $M_0(\omega) = 0$ and $\lim_{t\to\infty} M_t(\omega) = +\infty$.

Proposition 6.14. For every a and t in $\mathbb{R}_+$,

$$\mathbb{P}\{T_a < t\} = \mathbb{P}\{M_t > a\} = \mathbb{P}\{|W_t| > a\}. \tag{29}$$

The above proposition implies that M_t has the same distribution as $|W_t|$ for each t. Thus, in particular, $\mathbb{E}M_t = \sqrt{2t/\pi}$ and $\mathbb{E}M_t^2 = t$. However, this does not imply that the probability law of the whole M is the same with $|W|$.

We mention that Eq. (29) is also called the *reflection principle*. It says if you reflect the path around W_{t_0} at any t_0, then the reflected path (namely $(2W_{t_0} - W_t)_{t\in[t_0,\infty)}$) has the same distribution as $(W_t)_{t\in[0,\infty)}$.

Theorem 6.15. The process $T = (T_a)_{a\in\mathbb{R}_+}$ is a strictly increasing pure-jump Lévy process. It is stable with index $1/2$, and its Lévy measure is

$$\lambda(dt) = \frac{1}{\sqrt{2\pi t^3}}dt, \quad t > 0. \tag{30}$$

Proof. Fix $a, b \in (0,\infty)$. In order for W to hit $(a+b,\infty)$, it must hit (a,∞) first, and then $W \circ \theta_{T_a}$ must hit (b,∞). Thus,

$$T_{a+b} = T_a + T_b \circ \theta_{T_a}.$$

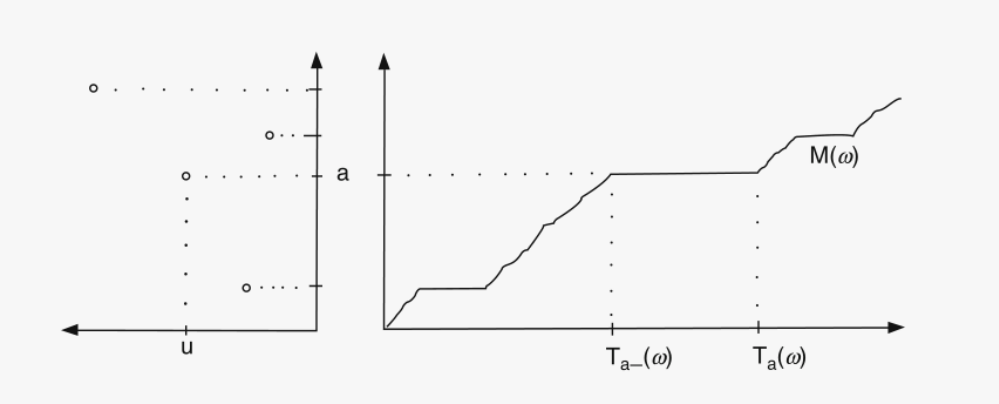

Figure 2: Visualization of the Poisson random measure N of hitting times $T = (T_a)_{a\in\mathbb{R}_+}$ from the textbook (Çınlar 2011).

Since $T_a < \infty$ almost surely, the process $W \circ \theta_{T_a}$ is independent of $\mathcal{F}_{T_a}$ and is again a Wiener process, by strong Markov property at T_a. Thus $T_{a+b} - T_a = T_b \circ \theta_{T_a}$ is independent of $\mathcal{F}_{T_a}$ and has the same distribution as T_b. Together with Proposition 6.13, this shows that the process T is a strictly increasing Lévy process over the stochastic base $(\Omega, \mathcal{H}, \hat{\mathcal{F}}, \hat{\theta}, \mathbb{P})$, where $\hat{\mathcal{F}}_a = \mathcal{F}_{T_a}$ and $\hat{\theta}_a = \theta_{T_a}$.

The distribution of T_a is the same as that of $a^2 T_1$, by Eq. (28). Thus, the Lévy process is stable with index $1/2$. Every such process is of the pure-jump type, and its Lévy measure has the form $\lambda(dt) = c/t^{3/2}dt$. The constant $c = 1/\sqrt{2\pi}$ because the following equation holds:

$$\mathbb{E}e^{-rT_a} = \exp\left\{-a\int_{\mathbb{R}_+}(1-e^{-rt})\lambda(dt)\right\} = e^{-a\sqrt{2r}}.$$

□

Theorem 6.16. Let N be the random measure on $\mathbb{R}_+ \times \mathbb{R}_+$ defined by

$$N(B) = \sum_{a:T_a>T_{a-}} \mathbb{1}_B(a, T_a - T_{a-}), \quad B \in \mathcal{B}(\mathbb{R}_+^2).$$

Then N is Poisson with mean $\mu_L \times \lambda$, where λ is given by Eq. (30). We have

$$T_a = \int_{(0,a]\times\mathbb{R}_+} uN(dx, du), \quad a \in \mathbb{R}_+.$$

See Fig. 2 for a visualization of N from the textbook. Note that, for *fixed* a, almost surely there are no atoms on the line $\{a\} \times \mathbb{R}_+$, so $T_a = T_{a-}$ almost surely, in agree with Proposition 6.10. Since N has infintely many atoms in any $(a, b) \times (0, \infty)$, the path M stays flat at infinitely many levels on its way from a to b, but only finitely many of those exceed ε in duration no matter how small $\varepsilon > 0$ is.

6.4 Other Properties

We put the following definitions or notations for reference.

- A *perfect set* is a closed set with no isolated point. The simplest example is a union of finitely many disjoint closed intervals. Another example is the Cantor set. Every perfect set C has the power of the continuum, that is, there exists an injection of $\mathbb{R}_+$ into C.

Zeros

We are interested in the qualitative features of the set

$$C_\omega = \{t \in \mathbb{R}_+ : W_t(\omega) = 0\}, \quad \omega \in \Omega,$$

the set of zeros of W. For fixed ω, it is the inverse image of the closed set $\{0\}$ under the continuous mapping $t \mapsto W_t(\omega)$, so it is closed, and its complement is the union of a countable collection of disjoint open intervals, called *contiguous intervals*.

Theorem 6.17. For almost every ω, the set C_ω is perfect and unbounded, its interior is empty, its Lebesgue measure is zero, and it has the power of the continuum.

Proof. We already showed that C_ω is closed. It is unbounded for almost every ω from Corollary 6.8. Its Lebesgue measure is zero for almost every ω since

$$\mathbb{E}\,\mu_L(C) = \mathbb{E}\int_{\mathbb{R}_+} \mathbb{1}_{\{0\}}(W_t)\,dt = \int_{\mathbb{R}_+} \mathbb{P}\{W_t = 0\}dt = 0.$$

This implies that the interior of C_ω is empty for almost every ω, because no set of Lebesgue measure zero can contain an open interval. To complete the proof, there remains to show that, for almost every ω, the set C_ω has no isolated point. Let

$$\mathbb{R}_+ \setminus C_\omega = \bigcup_{i\in\mathbb{N}} (G_i(\omega), D_i(\omega)). \tag{31}$$

According to the analysis after Proposition 6.6, there is an almost sure set Ω_{00} such that, for every $\omega \in \Omega_{00}$, there is a strictly decreasing sequence $\{t_k\} \subset C_\omega$ with limit 0, i.e. $0 \in C_\omega$ is a limit point of C_ω. Similarly, there is an almost sure event Ω_i such that $D_i(\omega)$ is a limit point of C_ω for every $\omega \in \Omega_i$. Consider the intersection $\Omega' = \Omega_{00} \cap \Omega_0 \cap \Omega_1 \cap \cdots$. For $\omega \in \Omega'$, neither 0 nor $D_i(\omega)$ is isolated. In view of Eq. (31), C_ω is perfect for every $\omega \in \Omega'$. □

Total Variation and Quadratic Variation

The (probabilistic) quadratic variation of W is $[W, W]_t = t$.

For almost every ω, the path $W(\omega)$ has infinite total variation over every interval $[a, b]$ with $a < b$.

Hölder Continuity, Nowhere Differentiability

Let $\alpha \in \mathbb{R}_+$, $B \subset \mathbb{R}_+$, and $f : \mathbb{R}_+ \to \mathbb{R}$. The function f is said to be Hölder continuous of order α on B if there is a constant k such that

$$|f(t) - f(s)| \le k \cdot |t - s|^\alpha, \quad s, t \in B.$$

It is said to be locally Hölder continuous of order α if it is such on $[0, b]$ for every $b < \infty$. Note that if f is differentiable at some point, then it is Hölder continuous of order 1 at some neighborhood of that point.

Proposition 6.18. For almost every ω, the Wiener path $W(\omega)$ is not Hölder continuous of order α on any interval for $\alpha > 1/2$. In particular, for almost every ω, the path is nowhere differentiable.

Proposition 6.19. For almost every ω, the path $W(\omega)$ is locally Hölder continuous of order α for every $\alpha < 1/2$.

References

Çınlar, E. (2011). *Probability and Stochastics*. Graduate Texts in Mathematics. Springer New York (cit. on pp. 54, 55).