Advanced Probability Theory

Rutgers STAT 680-681

Instructor: Min Xu
Scriber: Xinwei Zhang[1,2]

[1]Last updated on May 3, 2021
[2]These notes are continuously edited and may contain errors.

Contents

Chapter 0

Motivation

0.1 Curious Problems

Example 0.1.1. Let X, Y be independent continuous random variables taking value in $\mathbb{R}$. Let $f, g : \mathbb{R} \to \mathbb{R}$ be 2 arbitrary continuous functions, are $f(X)$ and $g(Y)$ independent?

Yes, but proving this without measure theoretic probability is extremely difficult: one has to use change of variables to compute the joint density of $(f(X), g(Y))$ and show that it factors. The proof is trivial with measure theory. Follow-up: what if $f, g : \mathbb{R} \to \mathbb{R}^3$? $f(X), g(Y)$ may not have densities.

Example 0.1.2. Let $p(x, y)$ be a bivariate density for $(X, Y) \in \mathbb{R}^2$. We know that $p(x \,|\, y)$ is a density of X for every $y \in \mathbb{R}$. What about the converse? Given a collection of univariate densities $\{f_y(\cdot)\}_{y\in\mathbb{R}}$, when can we "stitch" them together to form a valid joint distribution?

Example 0.1.3. There are various notions of distances between densities. For example, the Hellinger distance between densities p, q on $\mathbb{R}$ is defined as $H(p, q) := \int_{\mathbb{R}} (\sqrt{p(x)} - \sqrt{q(x)})^2 \, dx$; the KL-divergence is defined as $\mathrm{KL}(p, q) := \int_{\mathbb{R}} p(x) \log \frac{p(x)}{q(x)} \, dx$. Now consider two mixed distributions

$$P : \begin{cases} N(0,1) & \text{w.p. } \frac{1}{2} \\ 0 & \text{w.p. } \frac{1}{2} \end{cases} \quad Q : \begin{cases} N(\mu,1) & \text{w.p. } \frac{1}{2} \\ \text{Poisson}(1) & \text{w.p. } \frac{1}{2}. \end{cases}$$

Is there a notion of Hellinger distance between these? What about KL-divergence?

We will see that the answer is yes. The difference between discrete and continuous distributions is superficial; they are given an unified treatment under the Radon–Nikodym theorem.

Example 0.1.4. Let X, Y be random variables and let $f : \mathbb{R} \to \mathbb{R}$ be any function. One general formulation of the data processing inequality says that, for a very wide notions of distances between distributions $d(\cdot, \cdot)$, we have $d(X, Y) \geq d(f(X), f(Y))$. We will see that the Lebesgue decomposition theorem offers a relatively short proof of this fact.

Example 0.1.5. Let P be the uniform distribution on the surface of the 3D unit sphere $\mathbb{S}^2 := \{x \in \mathbb{R}^3 : \|X\|_2 = 1\}$. Let $(X_1, X_2, X_3) \sim P$ be a random vector, does a quantity like $\mathbb{E}[X_1 | X_2]$ have meaning?

Let us parametrize (X_1, X_2, X_3) in terms of a random longitude $\Theta \in [-\pi, \pi]$ and a random latitude $\Phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$.

By symmetry, the distribution of Θ is uniform on $[-\pi, \pi]$. What is the distribution of Φ conditioned on $\Theta = 0$ say?

Is it also uniform? No, conditioned on $\Theta = 0$, Φ is much more likely to be close to 0, near the equator.

Why? Because a "longitudinal slice" is relatively flatter around the equator than the north/south pole. This remains true if we take the limit and let the width of the slice go to 0.

Figure 1: Borel-Kolmogrov paradox

This example is known as "Borel-Kolmogrov Paradox" and shaped the development of probability theory.

> "The concept of a conditional probability with regard to an isolated hypothesis whose probability equals 0 is ambiguous. For we can obtain a probability distribution for the latitude on the meridian circle only if we regard this circle as an element of the decomposition of the entire spherical surface onto meridians circles with the given poles" – A.N. Kolmogorov

The actual conditional distribution of Φ given $\Theta = 0$ (or any other value) has the density

$$\phi \to \frac{1}{2}\cos\phi \quad \text{for } \phi \in [-\frac{\pi}{2}, \frac{\pi}{2}].$$

Example 0.1.6. Let A be a shape drawn on $\mathbb{R}^2$. Suppose we stretch X-axis by two-fold: $\widetilde{A} = \{(2x, y) : (x, y) \in A\}$. How does the "area" of A compare to the "area" of $\tilde{A}$?

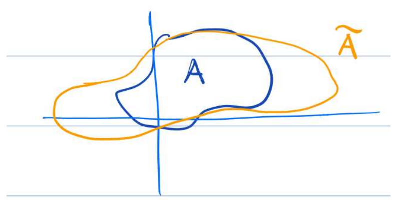


Figure 2: Stretch of a set

It always double! Why? We can approximate A arbitrarily well by smaller and smaller squares.

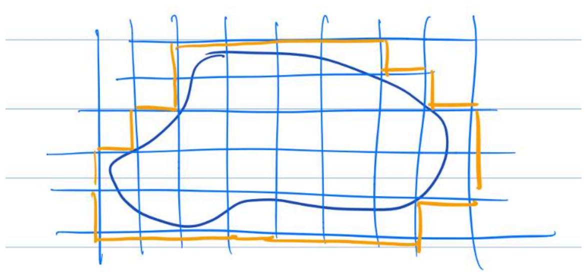


Figure 3: Area of a square doubles if we stretch X-axis by two-fold.

This is the basic intuition of measure theory: approximate complicated objects by simple ones.

Chapter 1

Construction of Measures

1.1 Algebra of Sets

Definition 1.1.1. Let Ω be a set and write 2^Ω as the set of all subsets of Ω. Let $\mathscr{F} \subset 2^\Omega$ be a family of subsets of Ω. We say that $\mathscr{F}$ is a field/algebra if

(a) $\Omega, \emptyset \in \mathscr{F}$.

(b) (Closed under complementation) For any $A \subseteq \Omega$, $A \in \mathscr{F} \iff A^c \in \mathscr{F}$.

(c) (Closed under finite union) For any $A, B \in \mathscr{F}$, $A \cup B \in \mathscr{F}$.

If, in addition, $A_1, A_1, A_3, \dots \in \mathscr{F}$ implies $\cup_{i=1}^\infty A_i \in \mathscr{F}$, (Closed under countable union), then $\mathscr{F}$ is a σ-field/σ-algebra.

Remark 1.1.1. Note that if $\mathscr{F}$ is a field and $A, B \in \mathscr{F}$, then $A \cap B = (A^c \cup B^c)^c \in \mathscr{F}$ as well. (So that $\mathscr{F}$ is closed under finite intersection). Likewise, if $\mathscr{F}$ is a σ-field, and $A_1, A_2, A_3, \dots \in \mathscr{F}$, then $\cap_{i=1}^\infty A_i = (\cup_{i=1}^\infty A_i^c)^c \in \mathscr{F}$. (Closed under countable intersection).

In fact, by the same argument, we may show that if $\mathscr{F}$ is a family of subsets of Ω that satisfies condition (a) and (b) of Definition 1.1.1 and is closed under intersection (or countable intersection), then $\mathscr{F}$ is a field (or σ-field).

Example 1.1.1. Let Ω be any set and $\mathscr{F} = 2^\Omega$, then $\mathscr{F}$ is trivially a σ-field.

Example 1.1.2. Let $\Omega = \mathbb{R}$. For $-\infty \le a \le b < \infty$, we say that $(a, b]$ is a right semi-closed interval. By convention, we let $(a, a]$ be the empty set and also define (a, ∞) to be right semi-closed. Let

$$\mathscr{F}_0 = \left\{\cup_{i=1}^n A_i : n \in \mathbb{N}, A_1, A_2, \dots, A_n \text{ disjoint right semi-closed intervals}\right\},$$

where $\cup_{i=1}^0 A_i$ is the empty set. Then $\mathscr{F}_0$ is a field.

To see this, note that $\mathbb{R} = (-\infty, \infty) \in \mathscr{F}_0$ so (a) in Definition 1.1.1 is satisfied.

For (b), let $I \in \mathscr{F}_0$ have the representation

$$I = \cup_{i=1}^n (a_i, b_i], \quad -\infty \le a_i < b_i < \infty \text{ and } b_i < a_{i+1} \quad i \in [n-1].$$

Then $I^c = (-\infty, a_1] \cup (b_1, a_2] \cup \dots \cup (b_{n-1}, a_n] \cup (b_n, \infty) \in \mathscr{F}_0$. If I contains (b, ∞) for some $b \in \mathbb{R}$, then we may obtain the same conclusion by applying the same argument.

For condition (c), it is more convenient prove that $\mathscr{F}_0$ is closed under intersection instead of union. Since $(a, b] \cap (x, y]$ is either empty or the interval $(a \vee x, b \wedge y]$, it holds that each component of $I \cap (x, y] = \cup_{i=1}^n (a_i, b_i] \cap (x, y]$ is right semi-closed.

Note that $\mathscr{F}_0$ is NOT a σ-field. Observe that $(0,1) = (0, \frac{1}{2}] \cap (\frac{1}{2}, \frac{2}{3}] \cup (\frac{2}{3}, \frac{3}{4}] \cup \cdots = \cup_{i=1}^{\infty}(0, 1 - \frac{1}{i}]$. We will show that $(0,1) \notin \mathscr{F}_0$. Let $I = \cup_{i=1}^{n}(a_i, b_i]$. If $b_n \geq 1$, then $b_n \in I$ but $b_n \notin (0,1)$. So $I \neq (0,1)$. If $b_n < 1$, then $\exists t \in (b_n, 1)$ such that $t \in (0,1)$ but $t \notin I$. So $I \neq (0,1)$.

Definition 1.1.2. Let $\mathcal{A} \subset 2^{\Omega}$ be a family of subsets of Ω. We define the field generated by $\mathcal{A}$ as

$$\mathscr{F}_0(\mathcal{A}) := \bigcap_{\{\mathscr{F} \text{ field}, \mathcal{A} \subseteq \mathscr{F}\}} \mathscr{F}.$$

That is, $\mathscr{F}_0(\mathcal{A})$ is the smallest field containing $\mathcal{A}$. Similarly, we may define a σ-field generated by $\mathcal{A}$ as

$$\sigma(\mathcal{A}) := \bigcap_{\{\mathscr{F} \ \sigma-\text{field}, \ \mathcal{A} \subseteq \mathscr{F}\}} \mathscr{F}.$$

Note that $\mathscr{F}_0(\mathscr{A})$ is nonempty since it contains $\mathcal{A}$. To verify that $\mathscr{F}_0(\mathcal{A})$ is a field, observe that $\Omega, \emptyset \in \mathscr{F}_0(\mathcal{A})$ and that if $A, B \in \mathscr{F}_0(\mathcal{A})$, then $A, B \in \mathscr{F}$ for all $\mathscr{F}$ containing $\mathcal{A}$ and thus A^c and $A \cup B \in \mathscr{F}$ for all $\mathscr{F}$ containing $\mathcal{A}$. Thus, $\mathscr{F}_0(\mathcal{A})$ is a field. Similar argument shows that $\sigma(\mathcal{A})$ is a σ-field.

Remark 1.1.2. Let $\mathcal{A} = \{\text{all right semi-closed intervals}\}$. Recall from Example 1.1.2 that the finite union of disjoint right semi-closed intervals $\mathscr{F}_0 = \{\cup_{i=1}^{n} A_i : n \in \mathbb{N}, A_i \in \mathcal{A} \text{ disjoint}\}$ is a field.

We claim that

$$\mathscr{F}_0(\mathcal{A}) = \mathscr{F}_0.$$

To see this, note that $\mathscr{F}_0(\mathcal{A}) \subseteq \mathscr{F}_0$ since $\mathscr{F}_0(\mathcal{A})$ is the smallest field containing $\mathcal{A}$. Now let $I \in \mathscr{F}_0$ be of the form $I = \cup_{i=1}^{n}(a_i, b_i]$. Since $(a_i, b_i] \in \mathscr{F}_0(\mathcal{A})$, we have $I \in \mathscr{F}_0(\mathcal{A})$ since $\mathscr{F}_0(\mathcal{A})$ is a field. Thus $\mathscr{F}_0 \subseteq \mathscr{F}_0(\mathcal{A})$.

Now let $\widetilde{\mathcal{A}} = \{\text{open intervals}\}$; we have seen that $\mathscr{F}_0(\widetilde{\mathcal{A}}) \neq \mathscr{F}_0(\mathcal{A})$. We claim that

$$\sigma(\mathcal{A}) = \sigma(\widetilde{\mathcal{A}}).$$

To show $\sigma(\mathcal{A}) \subseteq \sigma(\widetilde{\mathcal{A}})$, we will prove that $\mathcal{A} \subseteq \sigma(\widetilde{\mathcal{A}})$. Indeed, since $\sigma(\mathcal{A})$ is the smallest σ-field containing $\mathcal{A}$ and $\sigma(\widetilde{\mathcal{A}})$ is a σ-field, we have $\sigma(\mathcal{A}) \subseteq \sigma(\widetilde{\mathcal{A}})$. We may apply the same argument for the other direction.

We observe that for $-\infty < a < b < \infty$, $(a,b) = \cup_{n=1}^{\infty}(a, b - \frac{1}{n}] \in \sigma(\mathcal{A})$ and so $\widetilde{\mathcal{A}} \subseteq \sigma(\mathcal{A})$. On the other hand, $(a,b] = \cap_{n=1}^{\infty}(a, b + \frac{1}{n}) \in \sigma(\widetilde{\mathcal{A}})$ and so $\mathcal{A} \subseteq \sigma(\widetilde{\mathcal{A}})$ and the claim follows as desired.

Example 1.1.3. Let $\Omega = \mathbb{R}^d$ and let $\mathcal{A}$ be defined as in Remark 1.1.2. Let $\mathcal{A}^d = \mathcal{A} \times \mathcal{A} \times \ldots \mathcal{A} = \{A_1 \times A_2 \times \cdots \times A_d : A_i \in \mathcal{A}, \forall i \in [d]\}$ be the set of hyper-rectangles. We will be interested in the field and σ-field generated by $\mathcal{A}^d$. We refer to the latter as the Borel σ-field and denote it by $\mathscr{B}(\mathbb{R}^d)$.

Example 1.1.4. The notion of Borel σ-field is more general. Let $(\mathcal{X}, d)$ be a metric space, define an open ball around $x \in \mathcal{X}$ with radius $r > 0$ as $B(x,r) := \{y \in \mathcal{X} : d(x,y) < r\}$. Suppose also that $\mathcal{X}$ is *separable* in that there exists a countable set $x_1, x_2, \ldots, \in \mathcal{X}$ such that for all $x \in \mathcal{X}$, $\inf_{n \in \mathbb{N}} d(x, x_n) = 0$; for instance, if $\mathcal{X} = \mathbb{R}^d$, then we can take $x_1, x_2, \ldots$ to be points with rational coordinates.

Let $\mathcal{C} = \{B(x,r) : x \in \mathcal{X}, r > 0\}$ be the set of all open balls. Then the σ-field generated by $\mathcal{C}$ is known as the Borel σ-field (or Borel sets). It holds that $\mathscr{B}(\mathbb{R}^d)$ is the σ-field generated by Euclidean balls and thus rightly named Borel σ-field; remark 1.1.2 proves this for $d = 1$ and similar reasoning applies when $d > 1$.

More generally, if $(\mathcal{X}, d)$ is not necessarily separable, then the Borel σ-field is generated by the set of all open sets, where we say that $A \subset \mathcal{X}$ is open if for all $x \in A$, there exists $\epsilon > 0$ such that $B(x, \epsilon) \subset A$.

Remark 1.1.3. When $\mathcal{A}$ is the set of right semi-closed intervals, we showed that $\{\cup_{i=1}^{n} A_i : n \in \mathbb{N}, \{A_i \in \mathcal{A}\} \text{ disjoint}\}$ is the field generated by $\mathcal{A}$ (example 1.1.3, remark 1.1.2). This need not be true when $\mathcal{A}$ is a general set. When $\mathcal{A} = \{\text{right semi-closed intervals}\}$. $\mathcal{A}$ satisfies

(a) $\Omega, \emptyset \in \mathcal{A}$,

(b) $A, B \in \mathcal{A}$ implies $A \cap B \in \mathcal{A}$,

(c) $A, B \in \mathcal{A}$ implies that $\exists K \in \mathbb{N}$ and disjoint $C_1, \ldots, C_K \in \mathcal{A}$ such that $A \setminus B = \cup_{k=1}^{K} C_k$.

A set that satisfies (a), (b), and (c) is known as a semi-ring. Note that $\mathcal{A}^d$ (see example 1.1.4) is also a semi-ring. If $\mathcal{A}$ is a semi-ring, then $\{\cup_{i=1}^{n} A_i : n \in \mathbb{N}, \{A_i \in \mathcal{A}\} \text{ disjoint}\}$ is the field generated by $\mathcal{A}$.

Definition 1.1.3. Let $\mathscr{F}_0$ be a field and let $\mu : \mathscr{F}_0 \to [0, \infty]$ be a non-negative set function. We say that μ is a pre-measure if

(a) $\mu(\emptyset) = 0$,

(b) (Finite additivity) For disjoint $A, B \in \mathscr{F}_0$, $\mu(A \cup B) = \mu(A) + \mu(B)$,

(c) (Countable additivity) For any sequence of disjoint sets $A_1, A_2, \cdots \in \mathscr{F}_0$, if $\cup_{i=1}^{\infty} A_i \in \mathscr{F}_0$, then $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.

If $\mathscr{F}$ is a σ-field and $\mu : \mathscr{F} \to [0, \infty]$ is a set function that satisfies (a), (b) and (c), then μ is a measure.

If $\mu(\Omega) < \infty$, then μ is a finite measure.

If $\mu(\Omega) = 1$, then μ is a probability measure.

Remark 1.1.4. It is possible for a set function $\mu : \mathscr{F}_0 \to [0, \infty]$ to be finitly additive but not countably additive. Let $\mathscr{F}_0$ be defined as in example 1.1.2 (generated by right semi-closed intervals). Define μ such that for $A \in \mathscr{F}_0$,

$$\mu(A) = \begin{cases} 0 & \text{if } A \text{ is bounded } (\sup A - \inf A < \infty) \\ \infty & \text{else.} \end{cases}$$

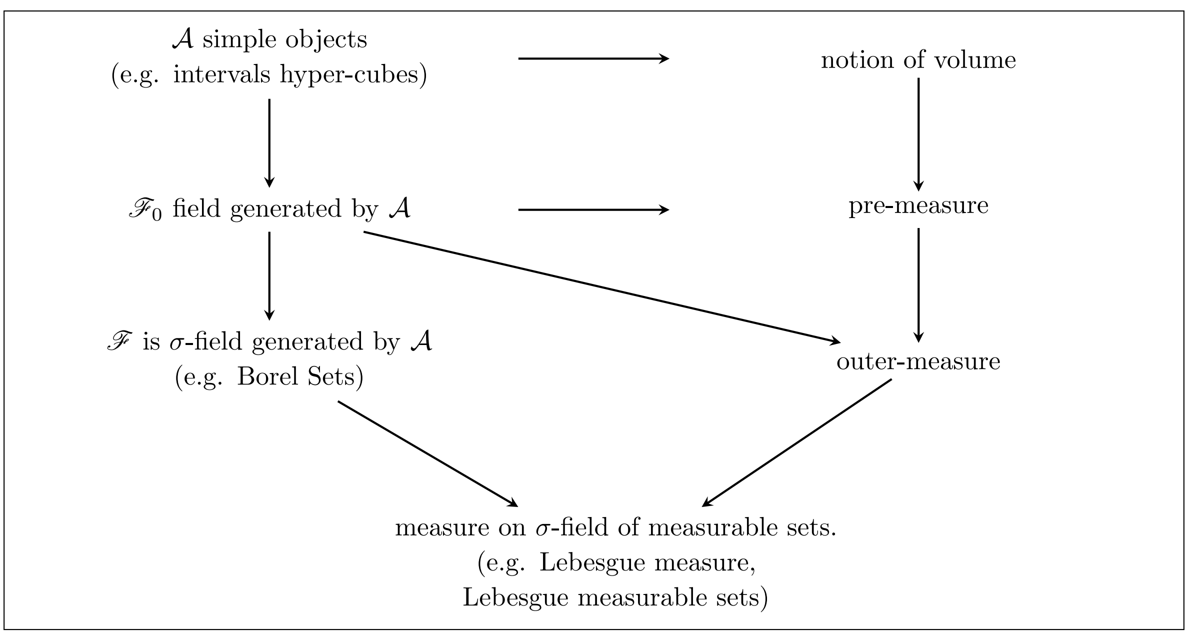


Figure 1.1: Road map of constructing measure

1.2 Extension by Outer Measure

Definition 1.2.1. Let $A_1 \subseteq A_2 \subseteq A_2 \dots$ be an increasing sequence of subsets of Ω. We write $A_i \nearrow \cup_{i=1}^{\infty} A_i$. Likewise, let $A_1 \supseteq A_2 \supseteq A_3 \dots$ be a decreasing sequence of subsets of Ω. We say $A_i \searrow \cap_{i=1}^{\infty} A_i$. In these cases, we say that $\cup_{i=1}^{\infty} A_i$ and $\cap_{i=1}^{\infty} A_i$ are the monotone set limits.

Given a field $\mathscr{F}_0$ and a pre-measure $\mu : \mathscr{F}_0 \to [0, \infty]$, we want to extend μ to the σ-field generated by $\mathscr{F}_0$. First, we will define μ on the monotone set limits of $\mathscr{F}_0$.

Lemma 1.2.1. Let $\mathscr{F}_0$ be a field on Ω and let

$$\mathscr{G} = \{A \subseteq \Omega : A_i \nearrow A \text{ where } A_1 \subseteq A_2 \subseteq \dots \mathscr{F}_0\},$$

Note that $\mathscr{F}_0 \subseteq \mathscr{G}$ since we may take $A_1 = A_2 = A_3 = \dots$. Let $\mu : \mathscr{F}_0 \to [0, \infty]$ be a pre-measure. For $A \in \mathscr{G}$ such that $A = \cup_{i=1}^{\infty} A_i$ where $A_1 \subseteq A_2 \subseteq \dots \in \mathscr{F}_0$, define

$$\mu(A) = \lim_{n\to\infty} \mu(A_n). \tag{$\star$}$$

Then, $\mu : \mathscr{G} \to [0, \infty]$ is well-defined. Moreover, $\mathscr{G}$ and μ satisfy the following:

(a) If $G_1, G_2 \in \mathscr{G}$, then $G_1 \cap G_2 \in \mathscr{G}$, $G_1 \cup G_2 \in \mathscr{G}$ and

$$\mu(G_1 \cup G_2) + \mu(G_1 \cap G_2) = \mu(G_1) + \mu(G_2).$$

(b) If $G_1, G_2 \in \mathscr{G}$ and $G_1 \subseteq G_2$, then $\mu_1(G_1) \leq \mu(G_2)$.

(c) If $G_1 \subseteq G_2 \subseteq G_3 \dots \in \mathscr{G}$, then $\cup_{n=1}^{\infty} G_n \in \mathscr{G}$ and $\lim_{n\to\infty} \mu(G_n) = \mu(\cup_{n=1}^{\infty} G_n)$. (The limit exists since $\mu(G_n)$ is nondecreasing by (b).)

Proof.
To show that μ is well-defined, we need to first show that when a monotone set limit is in $\mathscr{F}_0$, our definition $(\star)$ is consistent the original pre-measure.

To see this, note that if $A_1 \subseteq A_2 \subseteq \dots \in \mathscr{F}_0$ and $\cup_{n=1}^{\infty} A_n \in \mathscr{F}_0$ as well, then $\cup_{n=1}^{\infty} A_n = A_1 \cup (A_2 \cap A_1^c) \cup (A_3 \cap A_2^c) \cup \dots$ and

$$\begin{aligned} \mu(A_n) &= \mu(A_1) + \mu(A_2 \cap A_1^c) + \dots \mu(A_n \cap A_{n-1}^c) \\ &= \mu(A_1) + \sum_{i=2}^{n} \mu(A_i \cap A_{i-1}^c). \end{aligned}$$

Therefore,

$$\lim_{n\to\infty} \mu(A_n) = \mu(A_1) + \sum_{i=2}^{\infty} \mu(A_i \cap A_{i-1}^c) = \mu(\cup_{i=1}^{\infty} A_i). \qquad \text{(by (c) in definition 1.1.3)}$$

Thus, $(\star)$ is consistent with original $\mu : \mathscr{F}_0 \to [0, \infty)$.

Now, suppose $G \in \mathscr{G}$ has 2 representations

$$G = \cup_{m=1}^{\infty} A_m = \cup_{n=1}^{\infty} A'_n$$

for 2 sequences $A_1 \subseteq A_2 \subseteq \dots \in \mathscr{F}_0$ and $A'_1 \subseteq A'_2 \subseteq \dots \in \mathscr{F}_0$. We must show that $\mu(\cup_{m=1}^{\infty} A_m) = \mu(\cup_{n=1}^{\infty} A'_n)$.

To do this, we first claim the following: for any two monotone sequences $A_1 \subseteq A_2 \subseteq \dots \in \mathscr{F}_0$ and $A'_1 \subseteq A'_2 \subseteq \dots \in \mathscr{F}_0$ that satisfy $\cup_{m=1}^{\infty} A_m \subseteq \cup_{n=1}^{\infty} A'_n$, we claim that $\lim_{m\to\infty} \mu(A_m) \leq \lim_{n\to\infty} \mu(A'_n)$.

To see this, fix any $m \in \mathbb{N}$ and observe that $(A_m \cap A_1') \subseteq (A_m \cap A_2') \dots \in \mathscr{F}_0$ is a monotone sequence of sets and that $(A_m \cap A_n') \nearrow \cup_{n=1}^{\infty}(A_m \cap A_n') = A_m \cap (\cup_{n=1}^{\infty} A_n') = A_m \in \mathscr{F}_0$. Therefore, by (c) in Definition 1.1.3,

$$\mu(A_m) = \mu\left(\bigcup_{n=1}^{\infty}(A_m \cap A_n')\right) = \lim_{n\to\infty} \mu(A_m \cap A_n') \leq \lim_{n\to\infty} \mu(A_n'). \tag{1.1}$$

Since this is true for any m, we have $\lim_{n\to\infty} \mu(A_n') \geq \lim_{m\to\infty} \mu(A_m)$.

Now, if $\cup_{n=1}^{\infty} A_n' = \cup_{m=1}^{\infty} A_m$, then $\cup_{n=1}^{\infty} A_n' \subseteq \cup_{m=1}^{\infty} A_m$ and $\cup_{m=1}^{\infty} A_m \subseteq \cup_{n=1}^{\infty} A_n'$. So by what we just showed, it must be that $\lim_{n\to\infty} \mu(A_n') = \lim_{m\to\infty} \mu(A_m)$ and thus, $\mu : \mathscr{G} \to [0, \infty)$ is well-defined.

We now prove (a). Let $A_1 \subseteq A_2 \subseteq \dots \in \mathscr{F}_0$ and $A_1' \subseteq A_2' \subseteq \dots \in \mathscr{F}_0$ and denote $G_1 = \cup_{n=1}^{\infty} A_n$ and $G_2 = \cup_{n=1}^{\infty} A_n'$. Note that

$$\begin{aligned} G_1 \cap G_2 &= (\cup_{n=1}^{\infty} A_n) \cap (\cup_{n=1}^{\infty} A_n') = \cup_{n=1}^{\infty}(A_n \cap A_n') \\ G_2 \cup G_2 &= (\cup_{n=1}^{\infty} A_n) \cup (\cup_{n=1}^{\infty} A_n') = \cup_{n=1}^{\infty}(A_n \cup A_n'). \end{aligned}$$

For any $n \in \mathbb{N}$, we have $\mu(A_n \cap A_n') + \mu(A_n \cup A_n') = \mu(A_n) + \mu(A_n')$ by (b) of Definition 1.1.3. We may then obtain (a) by taking $n \to \infty$.

Property (b) follows from our earlier claim (see 1.1).

Lastly, we prove (c). Let $G_1 \subseteq G_2 \subseteq \dots \in \mathscr{G}$ and denote $G = \cup_{n=1}^{\infty} G_n$; we must show that $G \in \mathscr{G}$. For any $n \in \mathbb{N}$, let $G_n = \cup_{m=1}^{\infty} A_{nm}$ for $A_{n1} \subseteq A_{n2} \subseteq \dots \in \mathscr{F}_0$.

$$\begin{array}{ccccc} A_{11} \subseteq A_{12} \subseteq \dots A_{1m} \subseteq \dots \nearrow G_1 \\ A_{21} \subseteq A_{22} \subseteq \dots A_{2m} \subseteq \dots \nearrow G_2 \\ \vdots \\ A_{n1} \subseteq A_{n2} \subseteq \dots A_{nm} \subseteq \dots \nearrow G_n \\ \vdots \quad \vdots \qquad \vdots \qquad \vdots \nearrow G \end{array}$$

Define $D_m = A_{1m} \cup A_{2m} \cup \dots A_{mm}$. Note that $D_1 \subseteq D_2 \subseteq \dots \in \mathscr{F}_0$ by construction. Observe that, for any $n \leq m$,

$$A_{nm} \subseteq D_m \subseteq G_m \tag{$\star$}$$
$$\mu(A_{nm}) \leq \mu(D_m) \leq \mu(G_m) \tag{$\star\star$}$$

where the last inequality holds by (b) in this lemma. Since ($\star$) is true, $\forall m \geq n$, we get that, for any n,

$$G_n = \cup_{m=1}^{\infty} A_{nm} = \cup_{m=n}^{\infty} A_{nm} \subseteq \cup_{m=n}^{\infty} D_m \subseteq \cup_{m=1}^{\infty} D_m \subseteq \cup_{m=1}^{\infty} G_m = G.$$

Thus, we have that

$$G_n \subseteq \cup_{m=1}^{\infty} D_m \subseteq G. \tag{$\star\star\star$}$$

Since ($\star\star\star$) is true for all n, $G = \cup_{n=1}^{\infty} G_n \subseteq \cup_{m=1}^{\infty} D_m \subseteq G$ implies $\cup_{m=1}^{\infty} D_m = G$. Since $D_1 \subseteq D_2 \subseteq \dots \in \mathscr{F}_0$, we have $G \in \mathscr{G}$. Hence, by ($\star\star$),

$$\lim_{m\to\infty} \mu(G_m) \geq \lim_{m\to\infty} \mu(D_m) = \mu(G).$$

On the other hand, for any m, $G_m \subseteq G$ and so $\mu(G_m) \leq \mu(G)$ by property (b). It implies

$$\lim_{m\to\infty} \mu(G_m) \leq \mu(G).$$

Thus, $\lim_{m\to\infty} \mu(G_m) = \mu(G)$ as desired. $\square$

Definition 1.2.2. Let $\mu, \mathscr{G}$ be defined in Lemma 1.2.1. Define an outer measure $\mu^\star : 2^\Omega \to [0, \infty]$ as $\forall A \subseteq \Omega$, $\mu^\star(A) = \inf\{\mu(G) : G \in \mathscr{G}, A \subseteq G\}$. Note that the set is nonempty since it contains Ω.

Lemma 1.2.2. (a) $\forall A \in \mathscr{G}$, $\mu(A) = \mu^\star(A)$. If $\mu(\Omega) = 1$, then $\mu^\star \le 1$.

(b) $\forall A, B \in 2^\Omega$,

$$\mu^\star(A \cup B) + \mu^\star(A \cap B) \le \mu^\star(A) + \mu^\star(B).$$

In particular, if $\mu(\Omega) = 1$,

$$\mu^\star(A) + \mu^\star(A^c) \ge \mu^\star(\Omega) + \mu^\star(\emptyset) = 1.$$

(c) If $A \subseteq B \in 2^\Omega$, then $\mu^\star(A) \le \mu^\star(B)$.

(d) If $A_1 \subseteq A_2 \cdots \in 2^\Omega$, then $\lim_{n\to\infty} \mu^\star(A_n) = \mu^\star(\cup_{n=1}^\infty A_n)$. (Continuous under monotone set limit)

Proof.

(a) Let $A \in \mathscr{G}$, then $\mu^\star(A) \le \mu(A)$ by definition. If $\mu^\star(A) < \mu(A)$, then $\exists B \in \mathscr{G}$ such that $\mu(B) < \mu(A)$ and $A \subseteq B$; contradiction. Let $A \in 2^\Omega$ and suppose $\mu(\Omega) = 1$, then $\mu^\star(A) \le \mu(\Omega) \le 1$.

(b) Let $A, B \in 2^\Omega$. Fix $\varepsilon > 0$. $\exists G_1, G_2 \in \mathscr{G}$ such that $A \subseteq G_1$ and $\mu(G_1) \le \mu^\star(A) + \frac{\varepsilon}{2}$ and $B \subseteq G_2$ and $\mu(G_2) \le \mu^\star(B) + \frac{\varepsilon}{2}$. Then by Lemma 1.2.1 (b),

$$\begin{aligned} \epsilon + \mu^\star(A) + \mu^\star(B) \ge \mu(G_1) + \mu(G_2) &= \mu(G_1 \cup G_2) + \mu(G_1 \cap G_2) \\ &\ge \mu^\star(A \cup B) + \mu^\star(A \cap B). \end{aligned}$$

Since ε is arbitrary, (b) follows.

(c) Claim (c) follows since, if $G \in \mathscr{G}$ contains B, it also contains A.

(d) Let $A_1 \subseteq A_2 \subseteq \cdots \in 2^\Omega$ and let $A = \cup_{n=1}^\infty A_n$, since $A_n \subseteq A$, $\forall n$, we have $\mu^\star(A_n) \le \mu^\star(A)$, $\forall n$ by (c) and thus $\lim_{n\to\infty} \mu^\star(A_n) \le \mu^\star(A)$. If $\lim_{n\to\infty} \mu^\star(A_n) = \infty$, the claim follows trivially. We may thus assume that $\lim_{n\to\infty} \mu^\star(A_n) < \infty$. Fix $\varepsilon > 0$. There exist, for each $n \in \mathbb{N}$, $G_n \in \mathscr{G}$ such that $A_n \subseteq G_n$ and

$$\mu^\star(A_n) \le \mu(G_n) \le \mu^\star(A_n) + \varepsilon 2^{-n}. \qquad \text{(note that } \mu(G_n) < \infty \text{ as well)}$$

Since $A = \cup_{n=1}^\infty A_n \subseteq \cup_{n=1}^\infty G_n = G_1 \cup (G_1 \cup G_2) \cup \cdots \in \mathscr{G}$ by Lemma 1.2.1 (c), we have

$$(*) \qquad \mu^\star(A) \le \mu(\cup_{n=1}^\infty G_n) = \lim_{n\to\infty} \mu(\cup_{k=1}^n G_k). \qquad \text{(by Lemma 1.2.1 (c))}$$

On the other hand, we have $\mu(G_1) \le \mu^\star(A_1) + \frac{\varepsilon}{2}$. Assume inductive hypothesis, $\mu(\cup_{k=1}^n G_k) \le \mu^\star(A_n) + \varepsilon \sum_{i=1}^n 2^{-i}$. Then, since $\mu((\cup_{k=1}^n G_k) \cap G_{n+1}) < \infty$, we have

$$\begin{aligned} \mu(\cup_{k=1}^{n+1} G_k) &= \mu(\cup_{k=1}^n G_k) + \mu(G_{n+1}) - \mu((\cup_{k=1}^n G_k) \cap G_{n+1}) \qquad \text{(by Lemma 1.2.1 (a))} \\ &\le \mu^\star(A_n) + \varepsilon \sum_{i=1}^n 2^{-i} + \mu^\star(A_{n+1}) + \varepsilon 2^{-(n+1)} - \mu(G_n \cap G_{n+1}) \\ &\quad \text{(by inductive hypothesis. Note } G_n \cap G_{n+1} \subseteq A_n \cap A_{n+1} \subseteq A_n \text{ and } G_n \cap G_{n+1} \in \mathscr{G}.) \\ &\le \mu^\star(A_n) + \varepsilon \sum_{i=1}^{n+1} 2^{-i} + \mu^\star(A_{n+1}) - \mu^\star(A_n) \\ &\le \mu^\star(A_{n+1}) + \varepsilon \sum_{i=1}^{n+1} 2^{-i}. \end{aligned}$$

Thus, $\forall n \in \mathbb{N}$, $\mu(\cup_{k=1}^n G_k) \le \mu^\star(A_n) + \varepsilon \sum_{i=1}^n 2^{-i}$. Therefore, from $(*)$,

$$\begin{aligned}\mu^\star(A) &\le \lim_{n\to\infty} \mu(\cup_{k=1}^n G_k)\\ &\le \lim_{n\to\infty} \mu^\star(A_n) + \varepsilon \sum_{i=1}^n 2^{-i}\\ &= \lim_{n\to\infty} \mu^\star(A_n) + \varepsilon.\end{aligned}$$

Since ε is arbitrary, $\mu^\star(A) \le \lim_{n\to\infty} \mu^\star(A_n)$. This proves (d).

$\square$

Remark 1.2.1. An outer measure is usually defined more broadly as any set function $\lambda : 2^\Omega \to [0, \infty]$ satisfying

(a) $\lambda(\emptyset) = 0$,

(b) $A \subseteq B \in 2^\Omega$ implies $\lambda(B) \ge \lambda(A)$,

(c) $A_1, A_2, \ldots, \in 2^\Omega \implies \lambda(\cup_{n=1}^\infty A_n) \le \sum_{n=1}^\infty \lambda(A_n)$.

We see that $\mu^\star$ as defined in Lemma 1.2.2 satisfies (a) and (b). Homework 2 shows that it satisfies (c).

Suppose $\mu(\Omega) = 1$, then $1 - \mu^\star(H^c)$ is like an "inner measure". We will show that sets whose "inner" and outer measure "match", that is, $\{H \subseteq \Omega : \mu^\star(H) + \mu^\star(H^c) = 1\}$, constitute a σ-field.

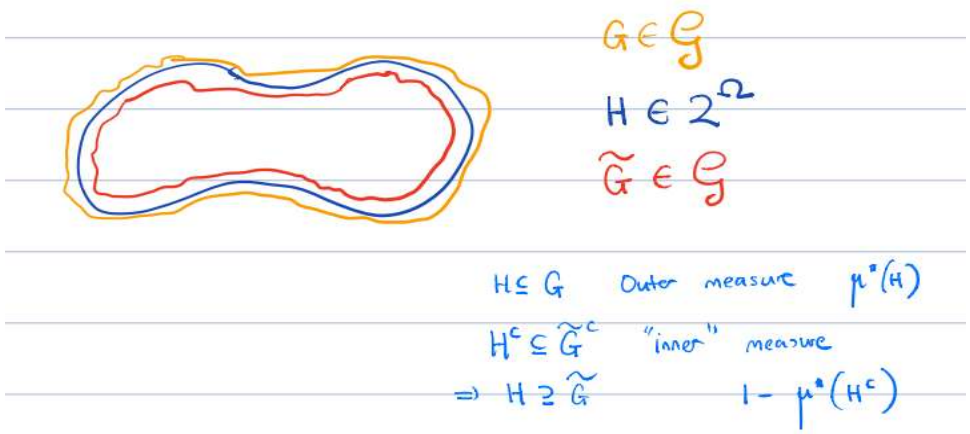


Figure 1.2: Outer and inner measures

The following lemma will be useful:

Lemma 1.2.3. (a) Let $\mathscr{F}$ be a field. If $\mathscr{F}$ is closed under monotone set limits, that is, $A_1 \subseteq A_2 \subseteq \cdots \in \mathscr{F} \implies \cup_{n=1}^\infty A_n \in \mathscr{F}$, then $\mathscr{F}$ is a σ-field.

(b) Let $\mathscr{F}$ be a σ-field and let $\mu : \mathscr{F} \to [0, \infty)$ be finitely additive (satisfying definition 1.1.3 (b)). If for any $A_1 \subseteq A_2 \subseteq \cdots \in \mathscr{F}$, we have $\lim_{n\to\infty} \mu(A_n) = \mu(\cup_{n=1}^\infty A_n)$, then μ is countably additive (definition 1.1.3 (b)).

Proof.

We prove claim (a). Let $\mathscr{F}$ be a field and let $B_1, B_2, \cdots \in \mathscr{F}$, then $\cup_{n=1}^\infty B_n = B_1 \cup (B_2 \cup B_1) \cup (B_3 \cup B_2 \cup B_1) \cup \cdots = \cup_{n=1}^\infty (\cup_{k=1}^n B_k) \in \mathscr{F})$, since $\cup_{k=1}^n B_k$ is increasing.

Now we prove (b). Let $B_1, B_2, \dots \in \mathscr{F}$ be disjoint. Define $A_n = \cup_{k=1}^n B_k$. Then

$$\begin{aligned}\mu(\cup_{n=1}^\infty B_n) = \mu(\cup_{n=1}^\infty A_n) &= \lim_{n\to\infty} \mu(A_n) \\ &= \lim_{n\to\infty} \sum_{k=1}^n \mu(B_k) \qquad \text{(by finite additivity)} \\ &= \sum_{k=1}^\infty \mu(B_k).\end{aligned}$$

□

Theorem 1.2.1. Let $\mathscr{F}_0$ be a field, let $\mu : \mathscr{F} \to [0, \infty]$ be a pre-measure and suppose $\mu(\Omega) = 1$. Let $\mu^\star$ be defined as lemma 1.2.2. Let $\mathscr{H} = \{H \subseteq \Omega : \mu^\star(H) + \mu^\star(H^c) = 1\}$, then, $\mathscr{H}$ is a σ-field and $\mu^\star : \mathscr{H} \to [0, 1]$ is a probability measure.

Note in particular that $\mathscr{F}_0 \subseteq \mathscr{H}$ by lemma 1.2.2 (a). Therefore, $\mathscr{H}$ contains $\sigma(\mathscr{F}_0)$, the σ-field generated by $\mathscr{F}_0$.

Proof.

We first show that $\mathscr{H}$ is a field. Note that $\mu^\star(\Omega) = 1$ and $\mu^\star(\emptyset) = 0$ by lemma 1.2.2. Let $A \in \mathscr{H}$, then $A^c \in \mathscr{H}$ by definition. Let $A, B \in \mathscr{H}$. Observer that by lemma 1.2.2 (b),

$$\begin{aligned}\mu^\star(A \cup B) + \mu^\star(A \cap B) &\le \mu^\star(A) + \mu^\star(B) \qquad (*)\\ \mu^\star(A^c \cup B^c) + \mu^\star(A^c \cap B^c) &\le \mu^\star(A^c) + \mu^\star(B^c)\end{aligned}$$

we obtain

$$\mu^\star(A \cup B) + \mu^\star((A \cup B)^c) + \mu^\star(A \cap B) + \mu^\star((A \cap B)^c) \le 2.$$

By lemma 1.2.2 (b), it must be that $\mu^\star(A \cup B) + \mu^\star((A \cup B)^c) = 1$, which implies $A \cup B \in \mathscr{H}$. Hence, $\mathscr{H}$ is a field. We note that also $(*)$ must hold with equality.

Now, let $A_1 \subseteq A_2 \subseteq \dots \in \mathscr{H}$. Fix $\varepsilon > 0$. By lemma 1.2.2 (d), we have $\mu^\star(\cup_{n=1}^\infty A_n) = \lim_{n\to\infty} \mu^\star(A_n)$ and hence $\exists n_0 \in \mathbb{N}$ such that

$$\mu^\star(\cup_{n=1}^\infty A_n) \le \mu^\star(A_{n_0}) + \varepsilon.$$

By lemma 1.2.2 (c), $\mu^\star((\cup_{n=1}^\infty A_n)^c) \le \mu^\star(A_{n_0}^c)$. Hence

$$\mu^\star(\cup_{n=1}^\infty A_n) + \mu^\star((\cup_{n=1}^\infty A_n)^c) \le \mu^\star(A_{n_0}) + \mu^\star(A_{n_0}^c) + \varepsilon \le 1 + \varepsilon.$$

Since $\varepsilon > 0$ is chosen arbitrarily, we have $\cup_{n=1}^\infty A_n \in \mathscr{H}$. Thus, by lemma 1.2.3 , $\mathscr{H}$ is a σ-field.

Now we show that $\mu^\star$ is finitely additive on $\mathscr{H}$. Let $A, B \in \mathscr{H}$ be disjoint. Since $(*)$ holds with equality, $\mu^\star$ is finitely additive. By lemma 1.2.2(d), $\mu^\star$ is continuous under monotone set limit, and by lemma 1.2.3 , it is thus a probability measure.

□

Remark 1.2.2. We have $\sigma(\mathscr{F}_0) \subseteq \mathscr{H}$, the reverse is generally not true. Note that $\mathscr{H}$ is defined with respect to a pre-measure μ while $\sigma(\mathscr{F}_0)$ is not. We can say more about the relationship between $\sigma(\mathscr{F}_0)$ and $\mathscr{H}$.

Definition 1.2.3. We say that $(\Omega, \mathscr{F}, \mu)$ is a <u>measure space</u> if $\mathscr{F}$ is a σ-field on Ω and $\mu : \mathscr{F} \to [0, \infty)$ is a measure.

We say that $(\Omega, \mathscr{F}, \mu)$ is <u>complete</u> if $\forall N \in 2^\Omega$, if $\exists A \in \mathscr{F}$ such that $A \supseteq N$ and $\mu(A) = 0$, then $N \in \mathscr{F}$ (and $\mu(N) = 0$ necessarily.)

Theorem 1.2.2. Let $(\Omega, \mathscr{F}, \mu)$ be a measure space, not necessarily complete. Define

$$\mathscr{F}^{(\mu)} := \{A \cup N : A \in \mathscr{F}, \exists B \in \mathscr{F} \text{ s.t. } N \subseteq B \text{ and } \mu(B) = 0\}.$$

Then $\mathscr{F}^{(\mu)}$ is a σ-field. Define $\mu : \mathscr{F}^{(\mu)} \to [0, \infty]$ with $\mu(A \cup N) = \mu(A)$ with $A \cup N \in \mathscr{F}^{(\mu)}$; then μ is well-defined and the space $(\Omega, \mathscr{F}^{(\mu)}, \mu)$ is complete.

We call $(\Omega, \mathscr{F}^{(\mu)}, \mu)$ the completion of $(\Omega, \mathscr{F}, \mu)$.

Proof.

We first prove that $\mathscr{F}^{(\mu)}$ is a σ-field. It is clear that $\emptyset, \Omega \in \mathscr{F}^{(\mu)}$. Let $A \cup N \in \mathscr{F}^{(\mu)}$ and $B \in \mathscr{F}$ such that $N \subseteq B$ and $\mu(B) = 0$. Then,

$$(A \cup N)^c = A^c \cap (B^c \cup (B \backslash N)) = (A^c \cap B^c) \cup (A^c \cap (B \backslash N)).$$

Since $A^c \cap B^c \in \mathscr{F}$ and that $(A^c \cap (B \backslash N)) \subseteq B$, we have that $(A \cup N)^c \in \mathscr{F}^{(\mu)}$.

Now let $\{A_n \cup N_n\}$ be a countable collection of elements in $\mathscr{F}^{(\mu)}$ and suppose $N_n \subset B_n \in \mathscr{F}$ and $\mu(B_n) = 0$. Then,

$$\cup_{n=1}^{\infty} (A_n \cup N_n) = \left(\cup_{n=1}^{\infty} A_n\right) \cup \left(\cup_{n=1}^{\infty} N_n\right),$$

and $\cup_{n=1}^{\infty} N_n \subset \cup_{n=1}^{\infty} B_n \in \mathscr{F}$. Since $\mu(\cup_{n=1}^{\infty} B_n) \leq \sum_{n=1}^{\infty} \mu(B_n) = 0$, we have that $\cup_{n=1}^{\infty}(A_n \cup N_n) \in \mathscr{F}^{(\mu)}$. Hence, $\mathscr{F}^{(\mu)}$ is a σ-field.

Now we verify that the extension of μ to $\mathscr{F}^{(\mu)}$ is well-defined. First suppose that for $A \in \mathscr{F}$ and $N \in 2^{\Omega}$ where $N \subseteq B \in CF$ and $\mu(B) = 0$, we have $A \cup N \in \mathscr{F}$. In this case, $\mu(A \cup N) \geq \mu(A)$ and $\mu(A \cup N) \leq \mu(A \cup B) \leq \mu(A) + \mu(B) \leq \mu(A)$ which means that $\mu(A \cup N) = \mu(A)$ and hence, the extension of μ is consistent with its original domain.

Now suppose $A \cup N \in \mathscr{F}^{(\mu)}$ has an alternative representation as $\tilde{A} \cup \tilde{N}$; we want to show that $\mu(A) = \mu(\tilde{A})$. For $R \in 2^{\Omega}$, define the outer measure $\mu^*(R) := \inf\{\mu(S) : S \in \mathscr{F}, R \subseteq S\}$. Note that $\mu^*(B) = \mu(B)$ for all $B \in \mathscr{F}$ and $\mu^*(N) = \mu^*(\tilde{N}) = 0$. Since

$$\mu(A) = \mu^*(A) \leq \mu^*(\tilde{A} \cup \tilde{N}) \leq \mu^*(\tilde{A}) + \mu^*(\tilde{N}) = \mu^*(\tilde{A}),$$

we have by symmetry that $\mu(A) = \mu(\tilde{A})$.

It is straightforward to verify that μ is countably additive on $\mathscr{F}^{(\mu)}$ and so $\mu : \mathscr{F}^{(\mu)} \to [0, \infty]$ is valid measure. Finally, it is straightforward to verify that the space $(\Omega, \mathscr{F}^{(\mu)}, \mu)$ is complete. □

Theorem 1.2.3. Let $\mathscr{F}_0$ be a field and let $\mu : \mathscr{F}_0 \to [0, \infty]$ be a pre-measure satisfying $\mu(\Omega) = 1$. Let $\mathscr{H}$ and $\mu^\star$ be defined as in lemma 1.2.2. Then $(\Omega, \mathscr{H}, \mu^\star)$ is the completion of $(\Omega, \sigma(\mathscr{F}_0), \mu^\star)$.

Proof.

Write $\mathscr{F}^{(\mu^\star)}$ as the completion of $(\Omega, \sigma(\mathscr{F}_0), \mu^\star)$. We first show that $\mathscr{H} \in \mathscr{F}^{(\mu^\star)}$.

Let $A \in \mathscr{H}$, then by definition of $\mu^\star(A)$ and of $\mu^\star(A^c)$, there exist $G_1, G_2, \cdots \in \mathscr{G}$ and $\widetilde{G}_1, \widetilde{G}_2, \cdots \in \mathscr{G}$ such that $\widetilde{G}_n \subseteq A \subseteq G_n$, $\forall n \in \mathbb{N}$ and $\mu^\star(G_n) \to \mu^\star(A)$ and $\mu^\star(\widetilde{G}_n) \to \mu^\star(A)$. Write $\widetilde{G} := \cup_{n=1}^{\infty} \widetilde{G}_n$ and write $A = \widetilde{G} \cup (A \setminus \widetilde{G})$. Note that $\widetilde{G} \in \mathscr{G} \subseteq \mathscr{H}$ and $A \setminus \widetilde{G} \subset (\cap_{n=1}^{\infty} G_n) \setminus \widetilde{G} \in \sigma(\mathscr{F}_0)$. Thus,

$$\begin{aligned} \mu^\star(\cap_{n=1}^{\infty} G_n \setminus \widetilde{G}) &\leq \mu^\star(G_m \setminus \widetilde{G}_m) \text{ for any } m \in \mathbb{N} \\ &= \mu^\star(G_m) - \mu^\star(\widetilde{G}_m) \end{aligned}$$

since $\mu^\star$ is a measure on $\mathscr{H}$. Taking $m \to \infty$ shows that LHS $= 0$, which implies $\mathscr{H} \subseteq \mathscr{F}^{(\mu^\star)}$.

Now we show $(\Omega, \mathscr{H}, \mu^\star)$ is complete. Let $B \in \mathscr{H}$ such that $\mu^\star(B) = 0$, then, for any $N \subseteq B$, $\mu^\star(N) \leq \mu^\star(B) = 0$ and $\mu^\star(N^c) \leq 1$, which implies $\mu^\star(N) + \mu^\star(N^c) = 1$ by definition 1.2.2 (b). Thus $N \in \mathscr{H}$ and $(\Omega, \mathscr{H}, \mu^\star)$ is complete.

Now let $A \cup N \in \mathscr{F}^{(\mu^\star)}$, with $A \in \sigma(\mathscr{F}_0)$ and $B \in \sigma(\mathscr{F}_0)$ such that $N \subseteq B$, $\mu^\star(B) = 0$. Since $A, B \in \mathscr{H}$ and $(\Omega, \mathscr{H}, \mu^\star)$ is complete, $N \in \mathscr{H}$ as well. So $A \cup N \in \mathscr{H}$. □

Remark 1.2.3. Although Theorem 1.2.3 assumes that $\mu(\Omega) = 1$, it may be used to show that the same conclusion when a pre-measure μ satisfies: $\exists$ disjoint $K_1, K_2, \dots \in \mathscr{F}_0$ such that $\cup_{n=1}^\infty K_n = \Omega$ and $\mu(K_n) < \infty$, $\forall n \in \mathbb{N}$. Such a pre-measure is called σ-finite.

Note that if $\mathscr{F}_0 = \{\cup_{n=1}^\infty A_n : \{A_n\}$ disjoint right semi-closed intervals$\}$ and if μ is defined as the sum of the length of the intervals, then μ is σ-finite. Let us call μ the Lebesgue pre-measure.

Corollary 1.2.1. Let $\mathscr{F}_0$ be a field and $\mu : \mathscr{F}_0 \to [0, \infty]$ be a σ-finite pre-measure. Then, there exists an extension measure $\mu^\star : \sigma(\mathscr{F}_0) \to [0, \infty]$.

Proof.

Let $K_1, K_2, \dots \in \mathscr{F}_0$ be such that $\mu(K_n) < \infty$. Define $\mu^{(n)} : \mathscr{F}_0 \to [0, \infty]$ by, for any $A \in \mathscr{F}_0$, $\mu^{(n)}(A) := \mu(A \cap K_n)$. Then, $\mu^{(n)}$ is a finite pre-measure and we may apply Theorem 1.2.3 on $\mu^{(n)}/\mu^{(n)}(\Omega)$ to obtain extensions $\mu^{\star(n)} : \sigma(\mathscr{F}_0) \to [0, \infty)$. Define $\mu^\star : \sigma(\mathscr{F}_0) \to [0, \infty]$ by $\mu^\star = \sum_{n=1}^\infty \mu^{\star(n)}$. If $A \in \mathscr{F}_0$, $\mu^\star(A) = \sum_{n=1}^\infty \mu^{(n)}(A) = \sum_{n=1}^\infty \mu(A \cap K_n) = \mu(A)$. If $A_1, A_2, \dots \in \sigma(\mathscr{F}_0)$ are disjoint, then

$$\begin{aligned}
\mu^\star(\cup_{m=1}^\infty A_m) &= \sum_{n=1}^\infty \mu^{\star(n)}(\cup_{m=1}^\infty A_m) \\
&= \sum_{n=1}^\infty \sum_{m=1}^\infty \mu^{\star(n)}(A_m) \\
&= \sum_{m=1}^\infty \sum_{n=1}^\infty \mu^{\star(n)}(A_m) = \sum_{m=1}^\infty \mu^\star(A_m). && \text{(since } \mu^{\star(n)}(A_m) \geq 0)
\end{aligned}$$

□

1.3 Uniqueness of the Extension

The key ingredient is the monotone class theorem.

Definition 1.3.1. We say that $\mathscr{C} \subseteq 2^\Omega$ is a monotone class if for any sequence $A_1 \subseteq A_2 \subseteq \dots \in \mathscr{C}$, we have that $\cup_{n=1}^\infty A_n \in \mathscr{C}$ and for any sequence $A_1 \supseteq A_2 \supseteq \dots \in \mathscr{C}$, we have that $\cap_{n=1}^\infty A_n \in \mathscr{C}$.

Let $\mathcal{A} \subseteq 2^\Omega$ and define

$$m(\mathcal{A}) := \bigcap_{\substack{\mathscr{C} \text{ monotone} \\ \mathcal{A} \subseteq \mathscr{C}}} \mathscr{C}$$

as the monotone class generated by $\mathcal{A}$. Note that the intersection of any number of monotone classes is still a monotone class.

Theorem 1.3.1 (Monotone Class Theorem)**.** Let $\mathscr{F}_0 \subseteq 2^\Omega$ be a field and let $\mathscr{C} \subseteq 2^\Omega$ be a monotone class. If $\mathscr{F}_0 \subseteq \mathscr{C}$, then $\sigma(\mathscr{F}_0) \subseteq \mathscr{C}$.

Proof.

Let $\mathscr{M}$ be the monotone class generated by $\mathscr{F}_0$. We claim that $\mathscr{M}$ is a field and thus a σ-field by lemma 1.2.3. Fix $A \in \mathscr{M}$ and $\mathscr{M}_A := \{B \in \mathscr{M} : A \cap B, A \cap B^c, A^c \cap B \in \mathscr{M}\}$. If $B_1 \subseteq B_2 \subseteq \dots \in \mathscr{M}_A$, then

$$A \cap (\cup_{n=1}^\infty B_n) = \cup_{n=1}^\infty \underbrace{(B_n \cap A)}_{\text{increasing \& in } \mathscr{M}} \in \mathscr{M}$$

and $A^c \cap (\cup_{n=1}^\infty B_n) \in \mathscr{M}$ by same logic. Also,

$$A \cap (\cup_{n=1}^\infty B_n)^c = A \cap (\cap_{n=1}^\infty B_n^c) = \cap_{n=1}^\infty \underbrace{(B_n^c \cap A)}_{\text{decreasing \& in } \mathscr{M}} \in \mathscr{M}$$

Thus, $\cup_{n=1}^{\infty} B_n \in \mathscr{M}_A$. By same reasoning, if $B_1 \supseteq B_2 \supseteq \cdots \in \mathscr{M}_A$, $\cap_{n=1}^{\infty} B_n \in \mathscr{M}_A$ implies $\mathscr{M}_A$ is a monotone class. Note that $\mathscr{M}_A \subseteq \mathscr{M}$ by definition.

Fix another $\widetilde{A} \in \mathscr{F}_0$, then $\mathscr{F}_0 \subseteq \mathscr{M}_{\widetilde{A}}$ since $\mathscr{F}_0$ is a field and since $\mathscr{M} = m(\mathscr{F}_0)$, we have that $\mathscr{M}_{\widetilde{A}} = \mathscr{M}$. This implies $\widetilde{A} \cap A$, $\widetilde{A} \cap A^c$, $\widetilde{A}^c \cap A \in \mathscr{M}$. So $\widetilde{A} \in \mathscr{M}_A$. Since $\widetilde{A}$ was an arbitrary element of $\mathscr{F}_0$, we have that $\mathscr{F}_0 \subseteq \mathscr{M}_A$ and $\mathscr{M}_A = \mathscr{M}$.

Since A was an arbitrary element of $\mathscr{M}$, for all $A \in \mathscr{M}$, for all $B \in CM$, $A \cap B$, $A^c \cap B$, $A \cap B^c \in \mathscr{M}$ and hence, $\mathscr{M}$ is a field (and σ-field by Lemma 1.2.3).

Since $\sigma(\mathscr{F}_0)$ is also a monotone class, $\sigma(\mathscr{F}_0) \supseteq \mathscr{M}$. Since $\mathscr{M}$ is a σ-field, $\sigma(\mathscr{F}_0) \subseteq \mathscr{M}$. So $\sigma(\mathscr{F}_0) = \mathscr{M} \subseteq \mathscr{C}$. □

Now, we put everything together.

Theorem 1.3.2 (Caratheodory Extension Theorem)**.** Let $\mathscr{F}_0 \subseteq 2^{\Omega}$ be a field and let $\mu : \mathscr{F}_0 \to [0, \infty]$ be a σ-finite pre-measure. Then μ has a unique extension to $\sigma(\mathscr{F}_0) \to [0, \infty]$.

Proof.

Corollary 1.2.1 has already shown existence of an extension. Let $\lambda : \sigma(\mathscr{F}_0) \to [0, \infty]$ be a measure and suppose $\lambda(A) = \mu(A)$, $\forall A \in \mathscr{F}_0$.

Write $\mathscr{C} := \{A \in \sigma(\mathscr{F}_0) : \lambda(A) = \mu(A)\}$. Note that $\mathscr{C}$ is a monotone class and that $\mathscr{F}_0 \subseteq \mathscr{C}$, so by theorem 1.3.1, $\mathscr{C} = \sigma(\mathscr{F}_0)$.

□

Remark 1.3.1. Monotone class theorem (Theorem 1.3.1) is equivalent to the so called Dynkin's $\pi - \lambda$ theorem.

We say that $\Pi \subset 2^{\Omega}$ is a π-system if $A, B \in \Pi \Rightarrow A \cap B \in \Pi$. We say that $\Lambda \subset 2^{\Omega}$ is a λ-system if (a) $\emptyset, \Omega \in \Lambda$, if (b) $A \in \Lambda \Rightarrow A^c \in \Lambda$, and if (c) $A_1 \subset A_2 \ldots \in \Lambda \Rightarrow \cup_{i=1}^{\infty} A_i \in \Lambda$. Dynkin's theorem states that if $\Pi \subseteq \Lambda$, then $\sigma(\Pi) \subseteq \Lambda$. This follows from Theorem 1.3.1 since Λ is a monotone class and contains the field generated by Π.

1.4 Lebesgue-Stieljes Measure

Definition 1.4.1. We say that a measure $\mu : \mathscr{B}(\mathbb{R}) \to [0, \infty]$ is a Lebesgue-Stieljes measure if for any bounded intervals $I \subseteq \mathbb{R}$, $\mu(I) < \infty$.

We say that $F : \mathbb{R} \to \mathbb{R}$ is a distribution function if F is non-decreasing and right-continuous. We write $F(\infty) := \lim_{x \to \infty} F(x)$ and $F(-\infty) = \lim_{x \to -\infty} F(x)$.

Theorem 1.4.1. Let F be a distribution function. There exists a unique Lebesgue-Stieljes measure $\mu : \mathscr{B}(\mathbb{R}) \to [0, \infty]$ such that $\forall -\infty \leq a < b < \infty$, $\mu((a, b]) = F(b) - F(a)$ and $\forall -\infty \leq b < \infty$, $\mu((b, \infty)) = F(\infty) - F(a)$.

Proof.

Let $\mathcal{A} :=$ {right semi-closed intervals} and $\mathscr{F}_0 := \{\cup_{i=1}^{n} A_i : A_1, A_2, \ldots, A_n \in \mathcal{A} \text{ disjoint}\}$ be the field generated by $\mathcal{A}$. Define $\mu : \mathcal{A} \to [0, \infty]$ as follows:

(1) If $\exists -\infty \leq a < b < \infty$ such that $A = (a, b]$, define $\mu(A) := F(b) - F(a)$ (note that if $a > -\infty$, then $\mu((a, b]) < \infty$), and

(2) If $A = (a, \infty)$, $\mu(A) = F(\infty) - F(a)$.

We define extension $\mu : \mathscr{F}_0 \to [0, \infty]$ as, for any $A_1, A_2, \ldots, A_n \in \mathcal{A}$ disjoint, $\mu(\cup_{i=1}^{n} A_i) := \sum_{i=1}^{n} \mu(A_i)$. This is well-defined since if $\cup_{n=1}^{N} A_n = \cup_{m=1}^{M} A'_m \in \mathscr{F}_0$ for some $N, M \in \mathbb{N}$, where $A_1, A_2, \ldots, A_N \in \mathcal{A}$ are disjoint and $A'_1, A'_2, \ldots, A'_M \in \mathcal{A}$ are disjoint, then

$$\begin{aligned}\sum_{n=1}^{N} \mu(A_n) = \sum_{n=1}^{N} \mu(A_n \cap \cup_{m=1}^{M} A'_m) &= \sum_{n=1}^{N} \mu(\cup_{m=1}^{M}(A_n \cap A'_m)) \\ &= \sum_{n=1}^{N}\sum_{m=1}^{M} \mu(A_n \cap A'_m) \qquad \text{(by $(*)$ and the fact that $A_n \cap A'_m \in \mathcal{A}$)} \\ &= \sum_{m=1}^{M} \mu(\cup_{n=1}^{N}(A_n \cap A'_m)) \qquad \text{(by $(*)$ and the fact that $A_n \cap A'_m \in \mathcal{A}$)} \\ &= \sum_{m=1}^{M} \mu(A'_m).\end{aligned}$$

It is also clear that if $B_1, B_2 \in \mathscr{F}_0$ are disjoint, then $\mu(B_1 \cup B_2) = \mu(B_1) + \mu(B_2)$. Now we show μ is countably additive on $\mathscr{F}_0$ and thus a pre-measure. Assume first that $\mu(\Omega) < \infty$.

We first show that if $B_1 \supseteq B_2 \ldots \in \mathscr{F}_0$ such that $B_n \searrow \emptyset$, then $\lim_{n\to\infty} \mu(B_n) = 0$. To see this, first note that for any r.s.c. interval $(a, b]$ and for any sequence of real numbers $a_k \searrow a$ such that $a_1 < b$, we have that

$$\lim_{k\to\infty} \mu((a_k, b]) = \lim_{k\to\infty} F(b) - F(a_k) = F(b) - F(a) = \mu((a, b])$$

by the right-continuity of F. Hence, for every B_n, we may find $C_n \in \mathscr{F}_0$ such that $\overline{C}_n \subset B_n$ and

$$\mu(B_n) \le \mu(C_n) + \epsilon 2^{-n}.$$

Since $\cap_{n=1}^{\infty} \overline{C}_n = \emptyset$ and $\overline{C}_n$ is compact, by Cantor's Intersection Theorem, there exists $n_\epsilon \in \mathbb{N}$ such that $\cap_{n=1}^{n_\epsilon} \overline{C}_n = \emptyset$.

Therefore, $\cap_{n=1}^{n_\epsilon} C_n = \emptyset$ and we have, for all $n \ge n_\epsilon$,

$$\begin{aligned}\mu(B_n) &= \mu(B_n \backslash \big(\cap_{k=1}^{n_0} C_k\big)) \\ &\le \mu\big(\cup_{k=1}^{n}(B_k \backslash C_k)\big) \\ &\le \sum_{k=1}^{n} \epsilon 2^{-k} \le \epsilon.\end{aligned}$$

Since ϵ is arbitrary, $\lim_{n\to\infty} \mu(B_n) = 0$.

Now still assume that $\mu(\Omega) < \infty$ and let $B_1, B_2, \ldots \in \mathscr{F}_0$ be disjoint and write $B := \cup_{n=1}^{\infty} B_n$. Then,

$$\mu(B) = \mu\big(B \backslash (\cup_{k=1}^{n} B_k)\big) + \mu(\cup_{k=1}^{n} B_k)$$

Since $B \backslash (\cup_{k=1}^{n} B_k) \searrow \emptyset$ as $n \to \infty$, we may take the limit on both sides and obtain countable additivity.

Now suppose $F(\infty) - F(-\infty) = \infty$. Then define, for $n \in \mathbb{N}$,

$$F_n(x) = \begin{cases} F(x) & \text{if } |x| \le n \\ F(-n) & \text{if } x < -n \\ F(n) & \text{if } x > n \end{cases}$$

We then observe that F_n is also non-decreasing and right-continuous. Let $\mu^{(n)}$ be the corresponding finite pre-measure on $\mathscr{F}_0$. We note that $\mu^{(n)} \le \mu$ and that for all $B \in \mathscr{F}_0$, $\mu^{(n)}(B) \nearrow \mu(B)$.

Now let $B_1, B_2, \ldots \in \mathscr{F}_0$ be disjoint and write $B = \cup_{n=1}^{\infty} B_n$.

We observe that

$$\begin{aligned}\mu(B) &= \lim_{n\to\infty} \mu^{(n)}(B)\\ &= \lim_{n\to\infty} \sum_{k=1}^{\infty} \mu^{(n)}(B_k)\\ &= \sum_{k=1}^{\infty} \lim_{n\to\infty} \mu^{(n)}(B_k) = \sum_{k=1}^{\infty} \mu(B_k).\end{aligned}$$

where the third equality follows from the monotone convergence theorem for infinite series (generalized later in the class).

Since $\sigma(\mathscr{F}_0) = \mathscr{B}(\mathbb{R})$, see Remark 1.1.2, the theorem follows from Caratheodory extension theorem. □

Theorem 1.4.2. Let $\mu : \mathscr{B}(\mathbb{R}) \to [0,\infty]$ be a Lebesgue-Stieljes measure. Define $F : \mathbb{R} \to \mathbb{R}$ such that $F(0) = 0$ (or any other fixed value) and for $a \in (-\infty, 0)$, $F(a) = F(0) - \mu((a, 0])$ and for $a \in [0,\infty)$, $F(a) = \mu((0, a]) + F(0)$. Then F is a distribution function.

Proof.

F is well-defined since μ is Lebesgue-Stieljes. It is non-decreasing since μ is a measure. Let $a_1 \geq a_2 \geq \cdots \in \mathbb{R}$ such that $a_n \searrow a$, then, if $a \geq 0$,

$$\begin{aligned}\lim_{n\to\infty} F(a_n) &= \lim_{n\to\infty} \mu((0, a_n])\\ &= \mu(\cap_{n=1}^{\infty}(0, a_n]) = \mu((0, a]) = F(a).\end{aligned}$$

If $a < 0$, then $\exists n_0 \in \mathbb{N}$ such that $a_n < 0$, $\forall n \geq n_0$. We assume $n_0 = 1$ WLOG. Then

$$\begin{aligned}\lim_{n\to\infty} F(a_n) &= -\lim_{n\to\infty} \mu((a_n, 0])\\ &= -\mu(\cup_{n=1}^{\infty}(a_n, 0]) = -\mu((a, 0]) = F(a).\end{aligned}$$

□

Remark 1.4.1. When, for $-\infty < a \leq b < \infty$, $F(b) - F(a) = b - a$, the resulting measure λ is the Lebesgue measure. Let $(\mathbb{R}, \mathscr{F}^{(\lambda)}, \lambda)$ be the completion of $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda)$, then $\mathscr{F}^{(\lambda)}$ is called Lebesgue measurable sets.

Some interesting mathematical facts:

- $\mathscr{B}(\mathbb{R}) \subsetneq \mathscr{F}^{(\lambda)} \subsetneq 2^{\Omega}$.
- For any $x \in \mathbb{R}$, $\lambda(\{x\}) = \lim_{n\to\infty} \lambda((x - \frac{1}{n}, x]) = 0$. If $A \subseteq 2^{\mathbb{R}}$ is countable, $\lambda(A) = 0$ by countable additivity. In particular, $\lambda(\text{rationals}) = 0$ and $\lambda(\text{irrationals}) = 1$.
- All open sets and closed sets are in $\mathscr{B}(\mathbb{R})$ and hence, for any $A \subset 2^{\mathbb{R}}$, the closure $\overline{A}$ and interior $\text{int}(A)$ are in $\mathscr{B}(\mathbb{R})$.
- $\exists A \subseteq 2^{\mathbb{R}}$ such that interior(closure(A)) $= \emptyset$ and $\lambda(A) > 0$. (Fat Cantor set).
- Let μ be a σ-finite measure on $\mathscr{B}(\mathbb{R})$. Let $B \in \mathscr{B}(\mathbb{R})$ such that $\mu(B) < \infty$. Fix $\varepsilon > 0$, $\exists N, M \in \mathbb{N}$ and $A_1, A_2, \ldots, A_N \in \mathcal{A}$ disjoint and $A'_1, A'_2, \ldots, A'_M \in \mathcal{A}$ disjoint such that $\cup_{m=1}^{M} A'_m \subseteq B \subseteq \cup_{n=1}^{N} A_n$ and $\mu(\cup_{n=1}^{N} A_n \setminus \cup_{m=1}^{M} A'_m) < \varepsilon$.

Chapter 2

Lebesgue Integration

2.1 Measurable Functions

Definition 2.1.1. Let $(\Omega, \mathscr{F})$ be a measurable space ($\mathscr{F} \subseteq 2^\Omega$ is a σ-field). We say that $f : \Omega \to \mathbb{R}$ is a Borel-measurable function if for any $B \in \mathscr{B}(\mathbb{R})$, $f^{-1}(B) \in \mathscr{F}$.

Define $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ and define the extended Borel σ-field

$$\mathscr{B}(\overline{\mathbb{R}}) := \big\{B, B \cup \{-\infty\}, B \cup \{\infty\}, B \cup \{-\infty, \infty\} \,:\, B \in \mathscr{B}(\mathbb{R})\big\}.$$

Again, we say that $f : \Omega \to \overline{\mathbb{R}}$ is a Borel-measurable function if for any $B \in \mathscr{B}(\overline{\mathbb{R}})$, $f^{-1}(B) \in \mathscr{F}$.

More generally, if $(\widetilde{\Omega}, \widetilde{\mathscr{F}})$ is a measurable space, we say $f : \Omega \to \widetilde{\Omega}$ is measurable if $\forall B \in \widetilde{\mathscr{F}}$, $f^{-1}(B) \in \mathscr{F}$ and we write f is $\mathscr{F}/\widetilde{\mathscr{F}}$–measurable.

Lemma 2.1.1. Let $f : \Omega \to \tilde{\Omega}$ and let $\widetilde{\mathcal{A}} \subset 2^{\widetilde{\Omega}}$ be a collection of subsets that generates $\widetilde{\mathscr{F}}$, i.e., $\sigma(\widetilde{\mathcal{A}}) = \widetilde{\mathscr{F}}$. If, for all $S \in \widetilde{\mathcal{A}}$, $f^{-1}(S) \in \mathscr{F}$, then f is $\mathscr{F}/\widetilde{\mathscr{F}}$–measurable.

In particular, for given a real-valued function $f : \Omega \to \mathbb{R}$ and supposing $\forall c \in \mathbb{R}$, $f^{-1}((-\infty, c]) = \{\omega \in \Omega : f(\omega) \le c\} \in \mathscr{F}$, then f is Borel-measurable.

Proof.

Define $\mathcal{C} := \{B \in \widetilde{\mathscr{F}} : f^{-1}(B) \in \mathscr{F}\}$. We claim that $\mathcal{C}$ is a σ-field. To see this, note that

(1) $f^{-1}(\emptyset) = \emptyset$, $f^{-1}(\widetilde{\Omega}) = \Omega$.

(2) $\forall B \in 2^{\widetilde{\Omega}}$, $f^{-1}(B^c) = f^{-1}(B)^c$ and thus, if $f^{-1}(B) \in \mathscr{F}$, then $f^{-1}(B^c) \in \mathscr{F}$ as well;

(3) $\forall \{B_i\}_{i \in \mathcal{I}}$, $f^{-1}(\cup_{i \in \mathcal{I}} B_i) = \cup_{i \in \mathcal{I}} f^{-1}(B_i)$ and $f^{-1}(\cap_{i \in \mathcal{I}} B_i) = \cap_{i \in \mathcal{I}} f^{-1}(B_i)$, where $\mathcal{I}$ is an arbitrary index set.

Thus, if $\widetilde{\mathcal{A}} \subset \mathcal{C}$, then $\widetilde{F} = \sigma(\widetilde{\mathcal{A}}) \subset \mathcal{C}$ and f is $\mathscr{F}/\widetilde{\mathscr{F}}$–measurable.

For the second claim, we observe that intervals of the form $(-\infty, c]$ generates $\mathscr{B}(\mathbb{R})$. □

Corollary 2.1.1. Let $f, g : \Omega \to \mathbb{R}$ be Borel-measurable. Then, $f \vee g := \max(f, g)$ and $f \wedge g := \min(f, g)$ are Borel-measurable functions. In particular, $f^+ := f \vee 0$ and $f^- := (-f) \vee 0$ are Borel-measurable.

Proof.

Since, for $c \in \mathbb{R}$, $\{\omega \in \Omega : (f \vee g)(\omega) > c\} = \{\omega \in \Omega : f(\omega) > c\} \cup \{\omega \in \Omega : g(\omega) > c\} \in \mathscr{F}$, the result follows from Lemma 2.1.1. Same reasoning for $f \wedge g$. □

Theorem 2.1.1. Let Ω and $\widetilde{\Omega}$ be metric spaces (with metrics d and $\tilde{d}$ respectively) and let $\mathscr{F}, \widetilde{\mathscr{F}}$ be Borel σ-fields. If $f : \Omega \to \widetilde{\Omega}$ is continuous, then f is measurable.

Recall continuous functions have 2 equivalent definitions:

① $\forall \omega \in \Omega$ and all $\varepsilon > 0$, $\exists \delta > 0$ such that $\forall \omega' \in \Omega$, $d(\omega', \omega) < \delta$ implies $\tilde{d}(f(\omega'), f(\omega)) < \varepsilon$;

② $\forall$ open set $B \subseteq \widetilde{\Omega}$, $f^{-1}(B)$ is open in Ω.

Proof.

Define $\mathcal{C} := \{B \in \widetilde{\mathscr{F}} : f^{-1}(B) \in \mathscr{F}\}$. Since f is continuous, $f^{-1}(B) \in 2^{\Omega}$ is an open set when $B \subseteq 2^{\widetilde{\Omega}}$ is an open ball. Thus, we have that $\widetilde{C}$ contains all open balls. Since $\mathcal{C}$ is a σ-field and since $\widetilde{\mathscr{F}}$ is the σ-field generated by open balls, $\mathcal{C} = \mathscr{F}$ is desired. $\square$

Lemma 2.1.2. Let $h_n : \Omega \to \mathbb{R}$ be Borel measurable for $n \in \mathbb{N}$. Let $h : \Omega \to \overline{\mathbb{R}}$ be a function such that $\forall \omega \in \Omega$, $\limsup_{n\to\infty} h_n(\omega) = h(\omega)$, then h is also Borel-measurable. (Same with $\liminf$). In particular, if $\exists h : \Omega \to \overline{\mathbb{R}}$ such that $\forall \omega \in \Omega$, $\lim_{n\to\infty} h_n(\omega) = h(\omega)$, then h is Borel-measurable.

Proof.

Let $h := \limsup h_n$ and let $c \in \mathbb{R}$. Note that for any $\omega \in \Omega$,

$$h(\omega) > c \iff \limsup_{n\to\infty} h_n(\omega) > c \iff \inf_{n\in\mathbb{N}} \sup_{k\geq n} h_k(\omega) > c$$
$$\iff \forall n \in \mathbb{N}, \exists k \geq n \text{ s.t. } h_k(\omega) > c.$$

Therefore, $\{\omega : h(\omega) > c\} = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} \{\omega : h_k(\omega) > c\} \in \mathscr{F}$. Applying the same argument to $-h_n$ proves the same result for $\liminf$. The lemma thus follows. $\square$

Definition 2.1.2. We say that $f : \Omega \to \mathbb{R}$ is <u>simple function</u> if $\exists n \in \mathbb{N}$, $A_1, A_2, \ldots, A_n \in \mathscr{F}$ disjoint and $c_1, c_2, \ldots, c_n \in \mathbb{R}$ such that $f = \sum_{i=1}^{n} c_i \mathbb{1}_{A_i}$. That is,

$$f(x) = \begin{cases} c_i & \text{if } x \in A_i \text{ for } i \in [n] \\ 0 & \text{else .} \end{cases}$$

Simple functions are Borel measurable.

The following approximation theorem is fundamental:

Theorem 2.1.2. A function $f : \Omega \to [0, \infty)$ is Borel-measurable if and only if $\exists$ simple functions $0 \leq s_1 \leq s_2 \leq s_3 \cdots \leq f$ such that for any $\omega \in \Omega$, $\lim_{n\to\infty} s_n(\omega) = f(\omega)$ (Pointwise convergence).

A function $f : \Omega \to \mathbb{R}$ is Borel-measurable if and only if $\exists$ simple functions $s_1, s_2, \ldots$ such that $|s_1| \leq |s_2| \leq \ldots |f|$ and $\forall \omega \in \Omega$, $\lim_{n\to\infty} s_n(\omega) = f(\omega)$.

Proof.

Since simple functions are Borel-measurable, limit of s_n is Borel-measurable by lemma 2.1.2. Now suppose $f \geq 0$ is Borel-measurable. For $n \in \mathbb{N}$, define

$$s_n(\omega) := \begin{cases} \frac{k-1}{2^n} & \text{if } f(\omega) \in [\frac{k-1}{2^n}, \frac{k}{2^n}) \text{ for some } k \in [n2^n] \\ n & \text{if } f(\omega) \geq n. \end{cases}$$

Note that $s_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbb{1}_{f^{-1}([\frac{k-1}{2^n}, \frac{k}{2^n}))} + n\mathbb{1}_{f^{-1}([n,\infty))}$. Since f is Borel-measurable, $f^{-1}([\frac{k-1}{2^n}, \frac{k}{2^n})) \in \mathscr{F}$, which implies s_n is simple.

For any $\omega \in \Omega$ and $n \in \mathbb{N}$, if $f(\omega) \geq n$, then $s_n(\omega) \leq s_{n+1}(\omega)$. Now suppose there exists $k \in [n2^n]$ such that $f(\omega) \in \left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)$. Then, we have that $f(\omega) \in \left[\frac{2k-2}{2^{n+1}}, \frac{2k}{2^{n+1}}\right)$ where $2k \in [(n+1)2^{n+1}]$ and hence,

$$\text{either } s_{n+1}(\omega) = \frac{2k-2}{2^{n+1}} \text{ or } s_{n+1}(\omega) = \frac{2k-1}{2^{n+1}}.$$

In either cases, we have $s_{n+1}(\omega) \geq s_n(\omega)$. Thus, $s_n \leq s_{n+1}$. Now fix $\omega \in \Omega$. For any $n \geq f(\omega)$,

$$0 \leq f(\omega) - s_n(\omega) \leq \frac{1}{2^n},$$

which implies $\lim_{n\to\infty} s_n(\omega) = f(\omega)$.

For the second claim, let $f = f^+ - f^-$. Since both f^+, f^- are Borel-measurable (Corrolary 2.1.1), there exists non-negative simple functions $s_1^- \leq s_2^- \leq \dots$ and $s_1^+ \leq s_2^+ \leq \dots$ such that $\forall \omega \in \Omega$, $s_n^-(\omega) \to f^-(\omega)$, $s_n^+(\omega) \to f^+(\omega)$. Define $s_n := s_n^+ - s_n^-$ and s_n satisfies the desired properties. □

Corollary 2.1.2. Let $f, g : \Omega \to \mathbb{R}$ be Borel-measurable. Then $f+g$,fg, and $\frac{1}{f}\mathbb{1}_{\{f\neq 0\}}$ are Borel measurable.

Proof.

Let $s_1, s_2, \dots$ and $t_1, t_2, \dots$ be simple functions such that $\forall \omega \in \Omega$, $s_n(\omega) \to f(\omega)$ and $t_n(\omega) \to g(\omega)$. Then, $\forall \omega \in \Omega$, $(s_n + t_n)(\omega) \to (f+g)(\omega)$, $(s_n \cdot t_n)(\omega) \to (f \cdot g)(\omega)$, $\frac{1}{s_n}(\omega)\mathbb{1}_{\{s_n(\omega)\neq 0\}} \to \frac{1}{f}(\omega)\mathbb{1}_{\{f(\omega\neq 0\}}$ □

Theorem 2.1.3. Let $(\Omega, \mathscr{F})$, $(\widetilde{\Omega}, \widetilde{\mathscr{F}})$, $(\Omega', \mathscr{F}')$ be measurable spaces. Let $f : \Omega \to \widetilde{\Omega}$ and $g : \widetilde{\Omega} \to \Omega'$ be $\mathscr{F}/\widetilde{\mathscr{F}}$–measurable and $\widetilde{\mathscr{F}}/\mathscr{F}'$–measurable respectively. Then $g \circ f : \Omega \to \Omega'$ is $\mathscr{F}/\mathscr{F}'$–measurarble.

Proof.

Let $A' \in \mathscr{F}'$, then $(g \circ f)^{-1}(A') = f^{-1}(g^{-1}(A')) \in \mathscr{F}$ since $g^{-1}(A') \in \widetilde{\mathscr{F}}$. □

2.2 Lebesgue Integral

Definition 2.2.1. Let $(\Omega, \mathscr{F}, \mu)$ be a measure space. Let $n \in \mathbb{N}$, $A_1, \dots, A_n \in \mathscr{F}$ be disjoint, $c_1, \dots, c_n \in (0, \infty)$ and $s = \sum_{i=1}^n c_i \mathbb{1}_{A_i}$ be a simple function. Then

$$\int s \mathrm{d}\mu := \sum_{i=1}^n c_i \mu(A_i) \in [0, \infty].$$

Let $f : \Omega \to [0, \infty)$ be Borel-measurable. Then,

$$\int f \mathrm{d}\mu := \sup\left\{\int s \mathrm{d}\mu : s \text{ simple}, s \leq f\right\} \in [0, \infty].$$

We say that $f : \Omega \to \mathbb{R}$ is integrable if $\int f^+ \mathrm{d}\mu < \infty$ or $\int f^- \mathrm{d}\mu < \infty$. If f is integrable,

$$\int f \mathrm{d}\mu := \int f^+ \mathrm{d}\mu - \int f^- \mathrm{d}\mu.$$

Let $A \in \mathscr{F}$ and f be integrable, then $\int_A f \mathrm{d}\mu := \int f \mathbb{1}_A \mathrm{d}\mu$.

Remark 2.2.1. Suppose a simple function $s = \sum_{i=1}^n c_i \mathbb{1}_{A_i}$ has an alternative representation $\sum_{i=1}^n c_i \mathbb{1}_{A_i} = \sum_{j=1}^m d_j \mathbb{1}_{B_j}$ for $d_1, \dots, d_m > 0$ and $B_1, \dots, B_m \in \mathscr{F}$ disjoint. Then, for any pair (i, j), either $A_i \cap B_j = \emptyset$ or $c_i = d_j$. Thus we have that

$$s = \sum_{i=1}^n \sum_{j=1}^m c_i \mathbb{1}_{A_i \cap B_j} = \sum_{i=1}^n \sum_{j=1}^m d_j \mathbb{1}_{A_i \cap B_j}.$$

Thus,

$$\sum_{i=1}^n c_i \mu(A_i) = \sum_{i=1}^n c_i \sum_{j=1}^m \mu(A_i \cap B_j) = \sum_{i=1}^n \sum_{j=1}^m d_j \mu(A_i \cap B_j) = \sum_{j=1}^m d_j \mu(B_j).$$

So, $\int s \mathrm{d}\mu$ is well defined for simple $s \geq 0$.

Remark 2.2.2. By taking $\Omega = \mathbb{N}$, $\mathscr{F} = 2^{\mathbb{N}}$, and $\mu : \mathscr{F} \to [0, \infty]$ as the counting measure, i.e., $\mu(A) = |A|$, or any discrete probability, we have that any sequence $f : \mathbb{N} \to \mathbb{R}$ is measurable and $\int f \mathrm{d}\mu = \sum_{i=1}^\infty f_i$.

Remark 2.2.3. For simple functions $s, t : \Omega \to \mathbb{R}$, it is easy to show that familiar properties of integration, e.g., $\int (s+t)\mathrm{d}\mu = \int s\mathrm{d}\mu + \int t\mathrm{d}\mu$; our goal is to derive these for general integrable functions. The steps generally goes

(1) prove property for non-negative simple functions,

(2) then for non-negative measurable functions,

(3) then for integrable functions.

Theorem 2.2.1. Let $\mathcal{I}$ be either $\mathbb{N}$ or $\{1, \ldots, n\}$ for $n \in \mathbb{N}$, and let $\{B_i \in \mathscr{F}\}_{i \in \mathcal{I}}$ be disjoint, and let $f : \Omega \to \mathbb{R}$ be integrable, then $\int_{\cup_{i\in\mathcal{I}} B_i} f\mathrm{d}\mu$ and $\sum_{i\in\mathcal{I}} \int_{B_i} f\mathrm{d}\mu$ both exist and

$$\int_{\cup_{i\in\mathcal{I}} B_i} f\mathrm{d}\mu = \sum_{i\in\mathcal{I}} \int_{B_i} f\mathrm{d}\mu.$$

In particular, if $f \geq 0$, then $B \mapsto \int_B f\mathrm{d}\mu$ on $\mathscr{F}$ is a measure.

Proof.

We first prove the claim for a simple functions $s \geq 0$. With $s = \sum_{m=1}^M c_m \mathbb{1}_{A_m}$ for $c_m \geq 0$ and $A_m \in \mathscr{F}$. Then

$$\begin{aligned}
\int_B s\mathrm{d}\mu &= \sum_{m=1}^M c_m \mu(A_m \cap B) \\
&= \sum_{m=1}^M c_m \mu(A_m \cap \cup_{i\in\mathcal{I}} B_i) \\
&= \sum_{m=1}^M \sum_{i\in\mathcal{I}} c_m \mu(A_m \cap B_i) \\
&= \sum_{i\in\mathcal{I}} \sum_{m=1}^M c_m \mu(A_m \cap B_i) = \sum_{i\in\mathcal{I}} \int_{B_I} s\mathrm{d}\mu.
\end{aligned}$$

Now suppose $f \geq 0$. Write $B := \cup_{i\in\mathcal{I}} B_i$ and let s be any simple function such that $0 \leq s \leq f$, then

$$\int_B s\mathrm{d}\mu = \sum_{i\in\mathcal{I}} \int_{B_i} s\mathrm{d}\mu \leq \sum_{i\in\mathcal{I}} \int_{B_i} f\mathrm{d}\mu,$$

since $s\mathbb{1}_{B_i}$ is simple and $s\mathbb{1}_{B_i} \leq f\mathbb{1}_{B_i}$. Since s is arbitrary, $\int_B f\mathrm{d}\mu \leq \sum_{i\in\mathcal{I}} \int_{B_i} f\mathrm{d}\mu$.

If $\exists i' \in \mathcal{I}$ such that $\int_{B_{i'}} f\mathrm{d}\mu = \infty$, then

$$\begin{aligned}
\int_B f\mathrm{d}\mu &= \sup\left\{\int t\mathrm{d}\mu : 0 \leq t \leq f\mathbb{1}_B, t \text{ simple}\right\} \\
&\geq \sup\left\{\int t\mathrm{d}\mu : 0 \leq t \leq f\mathbb{1}_{B_i'}, t \text{ simple}\right\} \qquad (*) \\
&= \int_{B_{i'}} f\mathrm{d}\mu = \infty = \sum_{i\in\mathcal{I}} \int_{B_i} f\mathrm{d}\mu.
\end{aligned}$$

Thus, assume that $\int_{B_i} f\mathrm{d}\mu < \infty$, $\forall i \in \mathcal{I}$. Fix $n \in \mathbb{N}$, $n \leq |\mathcal{I}|$, and $\varepsilon > 0$, then there exists simple functions $t_1, t_2, \ldots, t_n$ such that $\forall i \in [n]$, $0 \leq t_i \leq f\mathbb{1}_{B_i}$ and

$$\int_{B_i} t_i\mathrm{d}\mu = \int t_i\mathrm{d}\mu > \int_{B_i} f\mathrm{d}\mu - \frac{\varepsilon}{n}.$$

Thus, since $(t_1 \vee t_2 \vee \ldots \vee t_n)\mathbb{1}_B$ is simple and no greater than $f\mathbb{1}_B$, we obtain

$$\begin{aligned}\int_B f \mathrm{d}\mu &\geq \int_B t_1 \vee t_2 \ldots \vee t_n \mathrm{d}\mu \\ &= \sum_{i\in\mathcal{I}} \int_{B_i} (t_1 \vee t_2 \ldots \vee t_n) \mathrm{d}\mu \\ &\geq \sum_{i=1}^{n} \int_{B_i} t_i \mathrm{d}\mu \geq \sum_{i=1}^{n} \int_{B_i} f \mathrm{d}\mu - \varepsilon.\end{aligned}$$

Since n, ε are arbitrary, $\int_B f\mathrm{d}\mu \geq \sum_{i\in\mathcal{I}} \int_{B_i} f\mathrm{d}\mu$.

Now, suppose f is integrable and write $f = f^+ - f^-$. Suppose WLOG that $\int f^- \mathrm{d}\mu < \infty$. By reasoning similar to $(*)$, $\int_B f^-\mathrm{d}\mu \leq \int f^- \mathrm{d}\mu < \infty$ as well. Then

$$\begin{aligned}\int_B f \mathrm{d}\mu &= \int_B f^+ \mathrm{d}\mu - \int_B f^- \mathrm{d}\mu \\ &= \sum_{i\in\mathcal{I}} \int_{B_i} f^+ \mathrm{d}\mu - \sum_{i\in\mathcal{I}} \int_{B_i} f^- \mathrm{d}\mu \\ &= \sum_{i\in\mathcal{I}} \int_{B_i} f \mathrm{d}\mu.\end{aligned}$$

□

Lemma 2.2.1. Let $f : \Omega \to \mathbb{R}$ be integrable and let $s_1, s_2, \ldots$ be simple functions such that $\forall n \in \mathbb{N}$, $|s_n| \leq |f|$ and $\forall \omega \in \Omega$, $s_n(\omega) \to f(\omega)$. Then $\int f\mathrm{d}\mu = \lim_{n\to\infty} \int s_n \mathrm{d}\mu$.

Proof.

First suppose $f \geq 0$ and that $\forall n \in \mathbb{N}$, $0 \leq s_n \leq f$. By considering $s_1, s_1 \vee s_2, s_1 \vee s_2 \vee s_3, \ldots$ if necessary, we may assume without loss of generality that $0 \leq s_1 \leq s_2 \leq \ldots$.

Suppose first that $\int f\mathrm{d}\mu < \infty$. Fix $\varepsilon > 0$. By definition 2.2.1, $\exists$ a simple function $t : \Omega \to [0, \infty)$ such that $t \leq f$ and $\int f\mathrm{d}\mu > \int t\mathrm{d}\mu > \int f\mathrm{d}\mu - \frac{\varepsilon}{2}$. Define $B_n := \{\omega \in \Omega : s_n(\omega) < t(\omega)\}$. Note that $B_1 \supseteq B_2 \supseteq \ldots$ and

$$\cap_{n=1}^{\infty} B_n = \{\omega \in \Omega : \forall n \in \mathbb{N}, s_n(\omega) < t(\omega)\} \subseteq \{\omega \in \Omega : \lim_{n\to\infty} s_n(\omega) < t(\omega)\} = \emptyset$$

since $s_n(\omega) \to f(\omega)$, $\forall \omega \in \Omega$.

By Theorem 2.2.1, the set function $B \mapsto \int_B t d\mu$ is a measure and hence, $\lim_{n\to\infty} \int_{B_n} t d\mu = 0$. Thus, there exists n_ε such that for all $n \geq n_\varepsilon$, $\int_{B_n} t d\mu \leq \varepsilon/2$.

Hence, $\forall n \geq n_\epsilon$,

$$\begin{aligned}\int s_n \mathrm{d}\mu \geq \int_{B_n^c} s_n \mathrm{d}\mu &\geq \int_{B_n^c} t \mathrm{d}\mu \\ &= \int t\mathrm{d}\mu - \int_{B_n} t d\mu \geq \int t\mathrm{d}\mu - \frac{\varepsilon}{2} \\ &\geq \int f\mathrm{d}\mu - \varepsilon.\end{aligned}$$

Since s_n is simple $s_n \leq f$, we have $\int s_n \mathrm{d}\mu \leq \int f\mathrm{d}\mu \leq \int s_n \mathrm{d}\mu + \varepsilon$. Since ε is arbitrary, we have that $\lim_{n\to\infty} \int s_n \mathrm{d}\mu = \int f\mathrm{d}\mu$.

Now if $\int f\mathrm{d}\mu = \infty$, then, for any $M > 0$, we have a simple function t such that $0 \leq t \leq f$ and $\int t\mathrm{d}\mu \geq M$. Define $B_1 \supseteq B_2 \ldots$ as before, then there exists n_M such that for all $n \geq n_M$, $\int_{B_n} t d\mu \leq M/2$.

Thus, we have that, for all $n \geq n_M$,

$$\begin{aligned}\int s_n d\mu &\geq \int_{B_n^c} t d\nu = \int t d\mu - \int_{B_n} t d\mu \\ &\geq \int t d\mu - \frac{M}{2} \geq \frac{M}{2}.\end{aligned}$$

Since $M > 0$ is arbitrary, $\lim_{n\to\infty} \int s_n \mathrm{d}\mu = \infty = \int f \mathrm{d}\mu$ as desired.

Now suppose $f : \Omega \to \mathbb{R}$ is integrable. Write $f = f^+ - f^-$ and suppose WLOG that $\int f^+ \mathrm{d}\mu < \infty$. By what we showed for non-negative measurable functions, we have that $\int s_n^+ \mathrm{d}\mu \to \int f^+ \mathrm{d}\mu$ and $\int s_n^- \mathrm{d}\mu \to \int f^- \mathrm{d}\mu$. Since $\lim_{n\to\infty} \int s_n^+ \mathrm{d}\mu < \infty$, we have

$$\begin{aligned}\lim_{n\to\infty} \int s_n^+ - s_n^- \mathrm{d}\mu &= \lim_{n\to\infty} (\int s_n^+ \mathrm{d}\mu - \int s_n^- \mathrm{d}\mu) \\ &= \lim_{n\to\infty} \int s_n^+ \mathrm{d}\mu - \lim_{n\to\infty} \int s_n^- \mathrm{d}\mu \\ &= \int f^+ \mathrm{d}\mu - \int f^- \mathrm{d}\mu = \int f \mathrm{d}\mu. && \text{(by Definition 2.2.1)}\end{aligned}$$

□

Example 2.2.1. Note that lemma 2.2.1 is not true without the condition that $|s_n| \leq |f|$, $\forall n \in \mathbb{N}$. Consider $(\Omega, \mathscr{F}) = ([0,1], \mathscr{B}([0,1]))$ and μ as the Lebesgue measure. Let $f(\omega) = 0$, $\forall \omega \in [0,1]$ and

$$s_n(\omega) = \begin{cases} 2^n & \text{if } \omega \in (\frac{1}{n}, \frac{2}{n}] \\ 0 & \text{else.} \end{cases}$$

We have that for any $\omega \in [0,1]$, $s_n(\omega) \to 0 = f(\omega)$. However, $\int s_n \mathrm{d}\mu = 2^n \frac{1}{n} \to \infty$ while $\int f \mathrm{d}\mu = 0$.

Theorem 2.2.2. Let $f, g : \Omega \to \mathbb{R}$ be integrable.

(a) If $\int f \mathrm{d}\mu$ and $\int g \mathrm{d}\mu$ are not $\infty / -\infty$, then $f + g$ is integrable and $\int f + g \mathrm{d}\mu = \int f \mathrm{d}\mu + \int g \mathrm{d}\mu$.

(b) For any $\alpha \in \mathbb{R}$, αf is integrable and $\int \alpha f \mathrm{d}\mu = \alpha \int f \mathrm{d}\mu$.

(c) $\int |f| \mathrm{d}\mu \geq |\int f \mathrm{d}\mu|$.

(d) If $f \geq g$, then $\int f \mathrm{d}\mu \geq \int g \mathrm{d}\mu$.

Proof.

We prove (d) first. First suppose $f \geq g \geq 0$. Then $\{s \text{ simple } : 0 \leq s \leq g\} \subseteq \{s \text{ simple } : 0 \leq s \leq f\}$ and thus $\int f \mathrm{d}\mu \geq \int g \mathrm{d}\mu$. If $f, g : \Omega \to \mathbb{R}$ are integrable and $f \geq g$, then $f^+ \geq g^+$ and $f^- \leq g^-$. The conclusion follows again.

We now prove claim (b). Observe that (b) is true for any simple $s \geq 0$ and thus for f if $f \geq 0$. If $\alpha \geq 0$, then $\alpha f = \alpha f^+ - \alpha f-$ and if $\alpha < 0$, then $\alpha f = |\alpha| f^- - |\alpha| f^+$. Claim follows.

Lastly, we consider claim (a). Since $(f+g)^- \leq f^- + g^-$ and $(f+g)^+ \leq f^+ + g^+$, $f + g$ is integrable by (d).

Case 1: $f \geq 0$ and $g \geq 0$. Let $s_1 \leq s_2 \leq s_3 \leq \cdots \leq f$ and $t_1 \leq t_2 \leq \ldots g$ be sequences of simple functions such that $\forall \omega \in \Omega$, $s_n(\omega) \to f(\omega)$ and $t_n(\omega) \to g(\omega)$. Then, $s_n(\omega) + t_n(\omega) \to f(\omega) + g(\omega)$ and $s_1 + t_1 \leq s_2 + t_2 \cdots \leq f + g$. Thus,

$$\int f + g \mathrm{d}\mu = \lim_{n\to\infty} \int s_n + t_n \mathrm{d}\mu = \int f \mathrm{d}\mu + \int g \mathrm{d}\mu. \qquad \text{(by lemma 2.2.1)}$$

Case 2: $f \geq 0$ and $g \leq 0$ but $h := f + g \geq 0$. Then we have that $h + (-g) = f$ and thus, $\int h \mathrm{d}\mu - \int g \mathrm{d}\mu = \int f \mathrm{d}\mu$.

Now consider general f, g. Define

$$\begin{aligned}
E_1 :=&\{f \geq 0, g \geq 0\} := \{\omega \in \Omega : f(\omega) \geq 0, g(\omega) \geq 0\},\\
E_2 :=&\{f > 0, g \leq 0, f + g \geq 0\},\\
E_3 :=&\{f > 0, g \leq 0, f + g < 0\},\\
E_4 :=&\{f \leq 0, g > 0, f + g \geq 0\},\\
E_5 :=&\{f \leq 0, g > 0, f + g < 0\},\\
E_6 :=&\{f \leq 0, g \leq 0\}.
\end{aligned}$$

By case 1 and 2, $\forall i \in [6]$, $\int_{E_i} f + g \mathrm{d}\mu = \int_{E_i} f \mathrm{d}\mu + \int_{E_i} g \mathrm{d}\mu$. Since $f + g$, f,g are all integrable and since $\cup_{i=1}^6 E_i = \Omega$, the conclusion follows from Theorem 2.2.1.

For (c), observe that by (a),

$$\begin{aligned}
\int |f| \mathrm{d}\mu &= \int f^+ + f^- \mathrm{d}\mu\\
&= \int f^+ \mathrm{d}\mu + \int f^- \mathrm{d}\mu\\
&\geq \left| \int f^+ \mathrm{d}\mu - \int f^- \mathrm{d}\mu \right| = \left| \int f \mathrm{d}\mu \right|.
\end{aligned}$$

The theorem thus follows. □

Theorem 2.2.3. (a) If $E \in \mathscr{F}$ is such that $\mu(E) = 0$, then $\forall f : \Omega \to \mathbb{R}$ Borel-measurable, $\int_E f \mathrm{d}\mu = 0$.

(b) Let $f, g : \Omega \to \mathbb{R}$ be integrable and let $E := \{\omega \in \Omega : f(\omega) = g(\omega)\}$. If $\mu(E^c) = 0$ then we say $f = g$ μ-almost-everywhere and $\int f \mathrm{d}\mu = \int g \mathrm{d}\mu$.

(c) $f \geq 0$, μ-a.e. $\iff$ $\forall A \in \mathscr{F}$, $\int_A f \mathrm{d}\mu \geq 0$.

Proof.

We first consider (a). Let s be simple, then $\int_E s \mathrm{d}\mu = 0$. Thus if $f \geq 0$, $\int_E f \mathrm{d}\mu = 0$. Hence $\int_E f \mathrm{d}\mu = 0$, $\forall$ Borel-measurable f.

Now we turn to (c). Let $E := \{\omega \in \Omega : f(\omega) \geq 0\}$ and assume $\mu(E^c) = 0$. Note that $\int_{E^c} s \mathrm{d}\mu = 0$ for any simple s. Thus, $\int_{E^c} f \mathrm{d}\mu = 0$. Thus, $\int_A f \mathrm{d}\mu = \int_{A \cap E} f \mathrm{d}\mu + \int_{A \cap E^c} f \mathrm{d}\mu \geq 0$, $\forall A \in \mathscr{F}$. Now suppose $\int_A f \mathrm{d}\mu \geq 0$, $\forall A \in \mathscr{F}$. Define $B_n := \{\omega \in E^c : f(\omega) < -\frac{1}{n}\}$ and observe that $E^c = \cup_{n=1}^\infty B_n$. Then $\forall n \in \mathbb{N}$,

$$0 = \int_{B_n} -f \mathrm{d}\mu \geq \frac{1}{n}\mu(B_n) \implies \mu(B_n) = 0 \implies \mu(E^c) = 0.$$

Claim (b) follows by considering $f - g$ and $g - f$. □

Remark 2.2.4. We may append "μ-almost-everywhere" to all statements in Theorem 2.2.2. For example, if $f \geq g$ μ-a.e., then $\int f \mathrm{d}\mu \geq \int g \mathrm{d}\mu$.

2.3 Convergence of Lebesgue Integral

Theorem 2.3.1 (Monotone Convergence Theorem)**.** Let $0 \leq f_1 \leq f_2 \leq \ldots$ be a sequence of Borel-measurable functions, $f_n : \Omega \to [0, \infty)$. Suppose $\exists f : \Omega \to [0, \infty]$ such that for μ-a.e. $\omega \in \Omega$, $\lim_{n\to\infty} f_n(\omega) = f(\omega)$. Then

$$\int f_n \mathrm{d}\mu \to \int f \mathrm{d}\mu. \tag{2.1}$$

Proof.

Let $E := \{\omega \in \Omega : \lim_{n\to\infty} f_n(\omega) = f(\omega)\}$ so that $\mu(E^c) = 0$. By considering $f_n \mathbb{1}_E$ and $f\mathbb{1}_E$ and noticing that $\int_E f_n d\mu = \int f_n d\mu$ and $\int_E f d\mu = \int f d\mu$, we may assume without loss of generality that $\lim_{n\to\infty} f_n(\omega) = f(\omega)$ for all $\omega \in \Omega$.

For each $n \in \mathbb{N}$, $\exists$ a sequence of simple functions $0 \le s_1^{(n)} \le s_2^{(n)} \le \dots$ such that $\lim_{m\to\infty} s_m^{(n)}(\omega) = f_n(\omega)$, $\forall \omega \in \Omega$. (Theorem 2.1.2).

Define $t_n := \max(s_n^{(1)}, \dots, s_n^{(n)})$, then t_n is simple and $0 \le t_1 \le t_2 \le \dots$. Since $s_n^{(1)} \le f_1 \le f_n$, $s_n^{(2)} \le f_2 \le f_n$, ..., and $s_n^{(n)} \le f_n$, we have that $t_n \le f_n$. Thus,

$$\text{for all } \omega \in \Omega,\ \lim_{n\to\infty} t_n(\omega) \le f(\omega).$$

Moreover, for any fixed $n \in \mathbb{N}$ and any $m \ge n$, $t_m \ge s_m^{(n)}$. Thus, we have that

$$\lim_{m\to\infty} t_m \ge f_n.$$

Since n is arbitrary, $\forall \omega \in \Omega$, $\lim_{m\to\infty} t_m(\omega) = f(\omega)$. Thus, $\lim_{n\to\infty} \int t_n \mathrm{d}\mu = \int f \mathrm{d}\mu$ by lemma 2.2.1.

Since $\forall n \in \mathbb{N}$, $\int t_n \mathrm{d}\mu \le \int f_n \mathrm{d}\mu$, it holds that $\int f \mathrm{d}\mu \le \lim_{n\to\infty} \int f_n \mathrm{d}\mu$. But, $f \ge f_n$ for any $n \in \mathbb{N}$ which implies that $\int f \mathrm{d}\mu \ge \lim_{n\to\infty} \int f_n \mathrm{d}\mu$. So $\int f \mathrm{d}\mu = \lim_{n\to\infty} \int f_n \mathrm{d}\mu$ as desired. □

Remark 2.3.1. The monotonicity requirement is important. See Example 2.2.1. Theorem 2.3.1 also applies if $0 \ge f_1 \ge f_2 \ge f_3 \dots$ and $f_n \to f$ μ-a.e.

We will weaken the non-negativity requirement in the next Lemma but it cannot be completely removed. For instance, consider functions $f_1 \le f_2 \le \dots$ defined in $(0, \infty)$ such that

$$f_n(x) = 1 - \frac{1}{nx} \quad \text{for } x \in (0, \infty).$$

Note that for all $x \in (0, \infty)$, $\lim_{n\to\infty} f_n(x) = 1$. However, $\int_0^\infty f_n(x)\,dx = -\infty$ for all n.

Lemma 2.3.1 (Fatou's Lemma)**.** Let $h : \Omega \to \mathbb{R}$ be an integrable function such that $\int h d\mu < \infty$. Let $f_1, f_2, \dots,$ be Borel-measurable and suppose $\forall n \in \mathbb{N}$, $f_n \le h$ (f_n thus integrable). Then, $\limsup_{n\to\infty} f_n : \Omega \to \mathbb{R} \cup \{\infty\}$ is integrable and

$$\limsup_{n\to\infty} \int f_n \mathrm{d}\mu \le \int \limsup_{n\to\infty} f_n \mathrm{d}\mu. \tag{2.2}$$

Likewise, if $\int h d\mu > -\infty$ and if $\forall n \in \mathbb{N}$, $f_n \ge h$ μ-a.e., then $\liminf_{n\to\infty} f_n : \Omega \to \mathbb{R} \cup \{-\infty\}$ is integrable and

$$\liminf_{n\to\infty} \int f_n \mathrm{d}\mu \ge \int \liminf_{n\to\infty} f_n \mathrm{d}\mu. \tag{2.3}$$

Proof.

We prove the liminf case first. Suppose that $f_1, f_2, \dots \ge h$. We note that $\liminf_{n\to\infty} f_n$ is Borel–measurable by Lemma 2.1.2.

If $\int h d\mu = \infty$, then, since $f_n \ge h$ for all n and $\liminf_{n\to\infty} f_n \ge h$, we have $\int \liminf_{n\to\infty} f_n d\mu = \infty$ and $\liminf_{n\to\infty} \int f_n d\mu = \infty$. Thus, we may assume that $\int h d\mu < \infty$.

Define $g_n := \inf_{m\ge n} f_m - h$ and $g := \lim_{n\to\infty} g_n = \liminf_{n\to\infty} f_n - h$ so that we have $0 \le g_1 \le g_2 \le \dots \le g$.

For any $n \in \mathbb{N}$ and any $m \ge n$, $g_n \le f_m - h$ and thus,

$$\int g_n \mathrm{d}\mu \le \inf_{m\ge n} \int (f_m - h)\mathrm{d}\mu = \inf_{m\ge n} \int f_m \mathrm{d}\mu - \int h\mathrm{d}\mu.$$

where the equality uses the fact that $\int h d\mu \in \mathbb{R}$.

Since this is true for all $n \in \mathbb{N}$, we have by the monotone convergence theorem that

$$\int \liminf_{n\to\infty} f_n d\mu = \int \lim_{n\to\infty} g_n d\mu + \int h d\mu = \lim_{n\to\infty} \int g d\mu + \int h d\mu \le \lim_{n\to\infty} \inf_{m\ge n} \int f_m d\mu.$$

The $\liminf$ case follows by considering $-f_n$ and $-h$. $\square$

Theorem 2.3.2 (Dominated Convegence Theorem)**.** Let $h : \Omega \to [0, \infty)$ be Borel-measurable and suppose $\int h d\mu < \infty$. Suppose $f_1, f_2, \ldots$ is a sequence of Borel-measurable functions such that $|f_n| \le h$ μ-a.e. and that $\exists f : \Omega \to \mathbb{R}$ such that for μ-a.e. $\omega \in \Omega$, $f_n(\omega) = f(\omega)$. Then, $\lim_{n\to\infty} \int f_n d\mu$ exists and

$$\lim_{n\to\infty} \int f_n \mathrm{d}\mu = \int f \mathrm{d}\mu. \tag{2.4}$$

Proof.

For $n \in \mathbb{N}$, let $E_n := \{\omega \in \Omega : |f_n(\omega)| \le h(\omega)\}$ and $E_0 := \{\omega \in \Omega : f_n(\omega) \to f(\omega)\}$. Then $\mu(\cup_{n=0}^{\infty} E_n^c) \le \sum_{n=0}^{\infty} \mu(E_n^c) = 0$. Let $E := \cap_{n=0}^{\infty} E_n$.

Since $f_n \mathbb{1}_E \ge -h\mathbb{1}_E$, $\forall n \in \mathbb{N}$, by Lemma 2.3.1,

$$\limsup_{n\to\infty} \int f_n \mathrm{d}\mu = \limsup_{n\to\infty} \int_E f_n \mathrm{d}\mu \le \int_E \limsup_{n\to\infty} f_n \mathrm{d}\mu = \int f \mathrm{d}\mu.$$

Since $-f_n \mathbb{1}_E \le h\mathbb{E}$ $\forall n \in \mathbb{N}$,

$$\liminf_{n\to\infty} \int f \mathrm{d}\mu = \liminf_{n\to\infty} \int_E f_n \mathrm{d}\mu \ge \int_E \liminf_{n\to\infty} f_n \mathrm{d}\mu = \int f \mathrm{d}\mu.$$

Hence, the limit exists and is equal to $\int f d\mu$ as desired. $\square$

Remark 2.3.2. Let λ be the Lebesgue measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. If $f : \mathbb{R} \to \mathbb{R}$ is Riemann–integrable, then f is also Lebesgue–integrable and

$$\int f(x)\, dx = \int_{\mathbb{R}} f\, d\lambda.$$

In Lebesgue integral, to emphasize the variable over which we integrate, we often write

$$\int_{\mathbb{R}} f(x)\, d\lambda(x).$$

The following theorem is particularly important for statistics.

Theorem 2.3.3. Let $(\Omega, \mathscr{F}, \mu)$ be a measure space and let $\Theta \subset \mathbb{R}$ be open. Let $f : \Omega \times \Theta \to \mathbb{R}$ be a function satisfying

1. $f(\cdot, \theta)$ is integrable for every $\theta \in \Theta$.
2. $(\partial_\theta f)(\cdot, \theta)$ exists and is integrable for every $\theta \in \Theta$.

Let $\theta_0 \in \Theta$ and suppose there exists $\epsilon > 0$ such that

$$\left| \sup_{\theta \in (\theta_0 - \epsilon, \theta_0 + \epsilon)} (\partial_\theta f)(x, \theta) \right| \le h(x) \quad \text{for some } h : \Omega \to \mathbb{R} \text{ s.t.} \int h d\mu < \infty. \tag{2.5}$$

Then, we have that $\theta \mapsto \int f(x, \theta) d\mu(x)$ is differentiable at θ_0 and

$$\partial_\theta \int f(x, \theta_0) d\mu(x) = \int (\partial_\theta f)(x, \theta_0) d\mu(x)$$

Proof.

Let θ_n for $n \in \mathbb{N}$ be any sequence in $(\theta_0 - \epsilon, \theta_0 + \epsilon)$ such that $\lim_{n\to\infty} \theta_n = \theta_0$. Define $g_n(x) = \frac{f(x,\theta_n)-f(x,\theta_0)}{\theta_n - \theta_0}$ so that $\lim_{n\to\infty} g_n(\cdot) = (\partial_\theta f)(\cdot, \theta_0)$.

Note that for every $x \in \Omega$, by the mean value theorem, there exists θ_x between θ_n and θ_0 such that $|g_n(x)| = |(\partial_\theta f)(x, \theta_x)| \leq h(x)$. Thus, by the dominated convergence theorem, $\lim_{n\to\infty} \int g_n d\mu$ exists and

$$\begin{aligned}\lim_{n\to\infty} \int g_n \mathrm{d}\mu &= \int \lim_{n\to\infty} g_n d\mu \\ &= \int (\partial_\theta f)(x, \theta_0) d\mu(x).\end{aligned}$$

Since

$$\lim_{n\to\infty} \int g_n d\mu = \lim_{n\to\infty} \frac{1}{\theta_n - \theta_0} \left(\int f(x, \theta_n) d\mu(x) - \int f(x, \theta_0) d\mu(x) \right)$$

and since the sequence $\{\theta_n\}$ was an arbitrary sequence that converges to θ_0, the desired conclusion follows. $\square$

Remark 2.3.3. To check condition (2.5), we usually first check that

$$\int |(\partial_\theta f)(x, \theta_0)| d\mu(x) < \infty$$

and then find a function $h_\epsilon : \Omega \to \mathbb{R}$ such that $\int h_\epsilon d\mu < \infty$ and that

$$|(\partial_\theta f)(x, \theta) - (\partial_\theta f)(x, \theta_0)| \leq h_\epsilon(x) \text{ for all } \theta \in (\theta_0 - \epsilon, \theta_0 + \epsilon) \text{ and for all } x \in \Omega.$$

2.4 Push-forward measure and change of variables

Definition 2.4.1. Let λ, μ be 2 measures on $(\Omega, \mathscr{F})$. If there exists a Borel-measurable function $f : \Omega \to [0, \infty]$ such that $\forall A \in \mathscr{F}, \mu(A) = \int_A f \mathrm{d}\lambda$, then we say that f is the Radon-Nikodym derivative of μ w.r.t. λ and write $f = \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$. Note that $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$ is unique λ-a.e. by Theorem 2.2.3.

Remark 2.4.1.

Theorem 2.4.1. Let λ, μ be 2 measures on $(\Omega, \mathscr{F})$ and suppose $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$ exists. Then, $\forall g : \Omega \to \mathbb{R}$ integrable w.r.t. μ, we have

$$\int g \mathrm{d}\mu = \int g \cdot \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \mathrm{d}\lambda \tag{2.6}$$

in the sense that if either exists, then the other also exists and they are equal.

Proof.

First let $s : \Omega \to [0, \infty)$ be a simple function, with $s = \sum_{i=1}^n c_i \mathbb{1}_{A_i}$, for $c_1, c_2, \ldots, c_n \geq 0$. Then

$$\int s \mathrm{d}\mu = \sum_{i=1}^n c_i \mu(A_i) = \sum_{i=1}^n c_i \int_{A_i} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \mathrm{d}\lambda = \int \sum_{i=1}^n c_i \mathbb{1}_{A_i} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \mathrm{d}\lambda = \int s \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \mathrm{d}\lambda.$$

Now let $g \geq 0$ be Borel-measurable. By Theorem 2.1.2, there exist simple functions $0 \leq s_1 \leq s_2 \ldots$ such that $\forall \omega \in \Omega$, $s_n(\omega) \to g(\omega)$, and, by Lemma 2.2.1 , $\int g \mathrm{d}\mu = \lim_{n\to\infty} \int s_n \mathrm{d}\mu$. Define $h_n := s_n \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$. Then $0 \leq h_1 \leq h_2 \leq \ldots$ and $\forall \omega \in \Omega$, $h_n(\omega) \to g(\omega) \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(\omega)$. Thus, by monotone convergence theorem,

$$\int g \frac{\mathrm{d}u}{\mathrm{d}\lambda} \mathrm{d}\lambda = \lim_{n\to\infty} \int h_n \mathrm{d}\lambda = \lim_{n\to\infty} \int s_n \mathrm{d}\mu = \int g \mathrm{d}\mu.$$

Now, if $g : \Omega \to \mathbb{R}$ is integrable w.r.t. μ, then decompose $g = g^+ - g^-$ and the claim follows. If $g \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$ is integrable w.r.t. λ, decompose $g \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} = g^+ \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} - g^- \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$ and the conclusion holds as desired. $\square$

Definition 2.4.2. Let $(\Omega, \mathscr{F}, \mu)$ be a measure space, let $(\widetilde{\Omega}, \widetilde{\mathscr{F}})$ be another measurable space and let $f : \Omega \to \widetilde{\Omega}$ be $\mathscr{F}/\widetilde{\mathscr{F}}$–measurable.

Define a measure $\tilde{\mu}$ on $\widetilde{\mathscr{F}}$ by

$$\tilde{\mu}(B) := \mu(f^{-1}(B)), \ \forall B \in \widetilde{\mathscr{F}}.$$

We call $\tilde{\mu}$ the push-forward measure induced by f and write it as $\mu^{(f)}$. Note: if $\mu(\Omega) = 1$, then $\mu^{(f)}(\widetilde{\Omega}) = 1$ as well.

Example 2.4.1. (a) Let $f : \mathbb{R} \to \mathbb{R}$ be defined as $f(x) = \alpha x$ for some $\alpha > 0$. Let λ_1 be the Lebesgue measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, then for any $-\infty \leq a < b < \infty$,

$$\begin{aligned}\lambda_1^{(f)}((a,b]) &= \lambda_1(f^{-1}(a,b]) \\ &= \lambda_1(\{\omega \in \mathbb{R} : f(\omega) \in (a,b]\}) = \lambda((\frac{a}{\alpha}, \frac{b}{\alpha}]) \\ &= (b-a)\frac{1}{\alpha} = \int_b^a \frac{1}{\alpha}\mathrm{d}x = \int_{[a,b]} \frac{1}{\alpha}\mathrm{d}\lambda_1.\end{aligned}$$

By Caratheodory extension theorem, $\lambda_1^{(f)}(B) = \int_B \frac{1}{\alpha}\mathrm{d}\lambda_1$, $\forall B \in \mathscr{B}(\mathbb{R})$. By Theorem 2.4.1, $\frac{\mathrm{d}\lambda_1^{(f)}}{\mathrm{d}\lambda_1} = \frac{1}{\alpha}$. So, for any $g : \mathbb{R} \to \mathbb{R}$ Lebesgue–integrable, we have that $\int g \mathrm{d}\lambda_1^{(f)} = \int \frac{g}{\alpha}\mathrm{d}\lambda$.

(b) Let $f : \mathbb{R}^p \to [0, \infty)$ be defined as, $\forall x \in \mathbb{R}^p$, $f(x) := \|x\|_\infty$. Let λ_p be the Lebesgue measure on $(\mathbb{R}^p, \mathscr{B}(\mathbb{R}^p))$. Then, $\forall a \geq 0$,

$$\begin{aligned}\lambda_p^{(f)}([0,a]) &= \lambda_p(\{x \in \mathbb{R}^p : \|x\|_\infty \leq a\}) \\ &= a^p \lambda_p(\{x \in \mathbb{R}^p : \|x\|_\infty \leq 1\}) = 2^p a^p \\ &= \int_0^a p2^p r^{p-1}\mathrm{d}r \\ &= \int_{[0,a]} p2^p r^{p-1}\mathrm{d}\lambda_1(r).\end{aligned}$$

By Theorem 2.2.1, we have that $B \mapsto \int_B 2^p r^{p-1}\mathrm{d}r$ is a measure on $([0,\infty), \mathscr{B}([0,\infty)))$ and by Caratheodory extension theorem,

$$\forall B \in \mathscr{B}([0,\infty)), \quad \lambda^{(f)}(B) = \int_B p2^p r^{p-1}\mathrm{d}\lambda_1(r).$$

By Theorem 2.4.1, $\frac{\mathrm{d}\lambda^{(f)}}{\mathrm{d}\lambda_1} = r \mapsto p2^p r^{p-1}$. So, for any $g : \mathbb{R} \to \mathbb{R}$ Lebesgue-integral,

$$\int g \mathrm{d}\lambda^{(f)} = \int g p 2^p r^{p-1}\mathrm{d}\lambda_1(r).$$

Remark 2.4.2. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable (Borel-measurbale function), then we write $\mathbb{P}^{(X)}$ as the push–forward probability measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ induced by X and call it the distribution of X.

Theorem 2.4.2 (Generalized Change of Variable). Let $(\Omega, \mathscr{F}, \mu)$ be a measurable space and $(\widetilde{\Omega}, \widetilde{\mathscr{F}})$ be a measurable space, let $f : \Omega \to \widetilde{\Omega}$ be a measurable function and $\mu^{(f)} : \widetilde{\mathscr{F}} \to [0, \infty]$ be push–forward measure on $\widetilde{\Omega}$ induced by f.

Then, $\forall g : \widetilde{\Omega} \to \mathbb{R}$ that is integrable w.r.t. $\mu^{(f)}$,

$$\int_\Omega g \circ f \mathrm{d}\mu = \int_{\widetilde{\Omega}} g \mathrm{d}\mu^{(f)}$$

in the sense if one exists, both exist and are equal.

Let us see some examples before we prove this.

Example 2.4.2. (a) Let $f : \mathbb{R} \to \mathbb{R}$ be such that $f(x) = \alpha x$, $\forall x \in \mathbb{R}$. Then, $\forall g : \mathbb{R} \to \mathbb{R}$ integrable, we have

$$\int g \circ f \mathrm{d}\lambda_1 = \int g(f(x))\mathrm{d}x = \int g(\alpha x)\mathrm{d}x$$

and also

$$\begin{aligned} \int g \circ f \mathrm{d}\lambda_1 &= \int g \mathrm{d}\lambda^{(f)} && \text{(By Theorem 2.4.2)} \\ &= \int f \, \frac{\mathrm{d}\lambda_1^{(f)}}{\mathrm{d}\lambda_1} \mathrm{d}\lambda_1 && \text{(By Theorem 2.4.1)} \\ &= \int \frac{g}{\alpha} \mathrm{d}\lambda_1 && \text{(Example 2.4.1 (a))} \\ &= \frac{1}{\alpha} \int g(x)\mathrm{d}x. \\ \implies \int g(\alpha x)\mathrm{d}x &= \frac{1}{\alpha} \int g(x)\mathrm{d}x. \end{aligned}$$

(b) Let $f : \mathbb{R}^p \to [0, \infty)$ be such that $f(x) = \|x\|_\infty$, $\forall x \in \mathbb{R}^p$. Then, $\forall g : [0, \infty) \to \mathbb{R}$,

$$\int_{\mathbb{R}^p} g \circ f \mathrm{d}\lambda_p = \int_{\mathbb{R}^p} g(\|x\|_\infty)\mathrm{d}x$$

but also

$$\int_{\mathbb{R}^p} g \circ f \mathrm{d}\lambda_p = \int_{[0,\infty)} g \mathrm{d}\lambda_p^{(f)} = \int_{[0,\infty)} g \frac{\mathrm{d}\lambda_p^{(f)}}{\mathrm{d}\lambda_1} \mathrm{d}\lambda_1 = \int_0^\infty g(r) p 2^p r^{p-1} \mathrm{d}r.$$

For example,

$$\begin{aligned} \int_{\mathbb{R}^p} \|x\|_\infty^2 \mathbb{1}_{\{\|x\|_\infty \le 1\}} \mathrm{d}x &= \int_0^\infty r^2 \mathbb{1}_{\{r \le 1\}} p 2^p r^{p-1} \mathrm{d}r \\ &= \int_0^1 r^{p+1} p 2^p \mathrm{d}r = p 2^p \frac{r^{p+2}}{p+2}\Big|_0^1 = \frac{p}{p+2} 2^p. \end{aligned}$$

As another example,

$$\begin{aligned} \int_{\mathbb{R}^p} e^{-\|x\|_\infty} \mathrm{d}x &= \int_0^\infty e^{-r} p 2^p r^{p-1} \mathrm{d}r \\ &= p 2^p \Gamma(p) = p! 2^p. \end{aligned}$$

So $x \mapsto \frac{1}{p!2^p} e^{-\|x\|_\infty}$ is a density on $\mathbb{R}^p$.

Remark 2.4.3. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Let $X : \Omega \to \mathbb{R}$ be a random variable (Borel-measure function), then Theorem 2.4.2 implies that, for any Borel-measurable $g : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[g(X)] := \int_\Omega g(X(\omega))\mathrm{d}\mathbb{P}(\omega) = \int_{\mathbb{R}} g(x) \mathrm{d}\mathbb{P}^{(X)}(x)$$

and, writing $Y = g \circ X : \Omega \to \mathbb{R}$ and $\mathrm{Id} : \mathbb{R} \to \mathbb{R}$ as the identity, $\mathrm{Id}(x) = x$, $\mathbb{E}[g(X)] = \mathbb{E}[Y] = \int_\Omega \mathrm{Id} \circ Y \mathrm{d}\mathbb{P} = \int_{\mathbb{R}} y \mathrm{d}\mathbb{P}^{(Y)}(y)$.

Proof of Theorem 2.4.2.

If $s = \sum_{i=1}^n c_i \mathbb{1}_{\widetilde{A}_i} : \widetilde{\Omega} \to [0, \infty)$ is simple, for $c_1, c_2, \ldots, c_n > 0$ and disjoint $\widetilde{A}_1, \widetilde{A}_2, \ldots, \widetilde{A}_n \in \widetilde{\mathscr{F}}$, then $(s \circ f)(\omega) = s(f(\omega)) = c_i$ if $\exists i \in [n]$ s.t. $\omega \in f^{-1}(\widetilde{A}_i) \in \mathscr{F}$ and 0 else. Thus, $s \circ f = \sum_{i=1}^n c_i \mathbb{1}_{f^{-1}(\widetilde{A}_i)}$, and

$$\int s \circ f \mathrm{d}\mu = \sum_{i=1}^n c_i \mu(f^{-1}(\widetilde{A}_i)) = \sum_{i=1}^n c_i \mu^{(f)}(\widetilde{A}_i) = \int s \mathrm{d}\mu^{(f)}.$$

Now let $g : \widetilde{\Omega} \to [0, \infty)$ be Borel–measure, then $\exists$ a sequence of simple functions $0 \leq s_1 \leq s_2 \leq \cdots \leq g$ of simple functions such that $\forall \tilde{\omega} \in \widetilde{\Omega}$, $s_n(\tilde{\omega}) \to g(\tilde{\omega})$. Since $0 \leq s_1 \circ f \leq s_2 \circ f \leq \cdots \leq g \circ f$ is also a sequence of simple functions and $\forall \omega \in \Omega$, $s_n \circ f(\omega) = s_n(f(\omega)) \to (g \circ f)(\omega)$.

$$\int g \circ f \mathrm{d}\mu = \lim_{n \to \infty} \int s_n \circ f \mathrm{d}\mu = \lim_{n \to \infty} \int s_n \mathrm{d}\mu^{(f)} = \int g \mathrm{d}\mu^{(f)}.$$

Now, let $g : \widetilde{\Omega} \to \mathbb{R}$ be Borel measurable. Decompose $g = g^+ - g^-$ and the conclusion follows from previous analysis. $\square$

Chapter 3

Radon-Nikodym

3.1 Signed measure

Definition 3.1.1. Let $(\Omega, \mathscr{F})$ be a measurable space, let $\lambda : \mathscr{F} \to [-\infty, \infty]$ be a set function. We say that λ is a signed-measure if it is countably additive, i.e. $\forall B_1, B_2, \dots \in \mathscr{F}$ disjoint,

$$\lambda(\cup_{i=1}^n B_i) = \sum_{i=1}^{\infty} \lambda(B_i).$$

Note that $\lambda(\emptyset) = 0$ since $\emptyset = \cup_{i=1}^{\infty} \emptyset$.

Remark 3.1.1. Suppose $\exists A \in \mathscr{F}$ such that $\lambda(A) = \infty$. Then $\forall B \in \mathscr{F}$ disjoint from A, $\lambda(B) > -\infty$ or else $\lambda(A) + \lambda(B)$ is undefined. Also note that if $A \subseteq B$, it may be that $\lambda(A) > \lambda(B)$. The key result of this section is to show that $\exists$ measures λ^+ and $\lambda^- : \mathscr{F} \to [0, \infty]$ such that $\lambda = \lambda^+ - \lambda^-$ and at least one of which is finite.

Lemma 3.1.1. Let $\lambda : \mathscr{F} \to [-\infty, \infty]$ be a signed measure. Then, $\exists C, D \in \mathscr{F}$ such that $\lambda(C) = \sup\{\lambda(A) : A \in \mathscr{F}\}$ and $\lambda(D) := \inf\{\lambda(A) : A \in \mathscr{F}\}$.

Proof.

First consider the supremum. If $\exists A \in \mathscr{F}$ such that $\lambda(A) = \infty$, then we may take $C = A$. We thus assume $\lambda(A) < \infty \; \forall A \in \mathscr{F}$. Then, $\exists A_1, A_2 \dots \in \mathscr{F}$ such that $\lambda(A_n) \to \sup\{\lambda(A) : A \in \mathscr{F}\}$ and write $A := \cup_{n=1}^{\infty} A_n$. Note $A_1, A_2, \dots$ is not monotone since λ is a signed measure. And $\lambda(A) \neq \lim_{n\to\infty} \lambda(A_n)$ since λ is signed measure. We cannot set $C = A$.

For each $n \in \mathbb{N}$, let $\alpha^{(n)} \in \{0,1\}^n$ and write

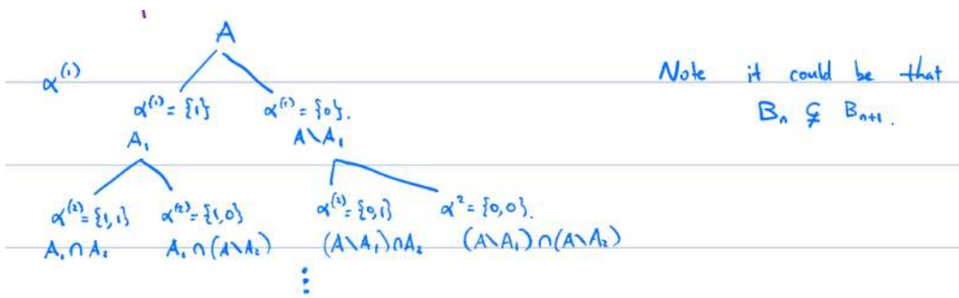
$$A_{\alpha^{(n)}} := \{\omega \in A : \forall i \in [n], \omega \in A_i \text{ iff } \alpha^{(n)}(i) = 1\}.$$

In other words, $A_{\alpha^{(n)}} = \cap_{i=1}^n A_i^\star$ where

$$A_i^\star = \begin{cases} A_i & \text{if } \alpha^{(n)}(i) = 1 \\ A \setminus A_i & \text{if } \alpha^{(n)}(i) = 0. \end{cases}$$

Since $A_n \subseteq A$, we have that $\cup_{\alpha^{(n)} \in \{0,1\}^n} A_{\alpha^{(n)}} = A$. Note that $\{A_{\alpha^{(n)}}\}_{\alpha^{(n)} \in \{0,1\}^n}$ are disjoint. Now define, for $n \in \mathbb{N}$, $B_n = \cup_{\alpha^{(n)} \in \{0,1\}^n} \{A_{\alpha^{(n)}} : \lambda(A_{\alpha^{(n)}}) \geq 0\}$ and $B_n = \emptyset$ if $\lambda(A_{\alpha^{(n)}}) < 0 \; \forall \alpha^{(n)} \in \{0,1\}^n$. Since $A_n = \cup_{\alpha^{(n)} \in \{0,1\}^n} \{A_{\alpha^{(n)}} : \alpha^{(n)}(n) = 1\}$, we have that $\lambda(A_n) \leq \lambda(B_n)$ $(\star)$.

On the other hand, for any $m \leq n$, for any $\beta^{(m)} \in \{0,1\}^m$, $\alpha^{(n)} \in \{0,1\}^n$, if $\alpha^{(n)}(i) = \beta^{(m)}(i) \; \forall i \in [m]$, $A_{\alpha^{(n)}} \subseteq A_{\beta^{(m)}}$, if $\exists i \in [m]$ such that $\alpha^{(n)}(i) \neq \beta^{(m)}(i)$, $A_{\alpha^{(n)}} \cap A_{\beta^{(m)}} = \emptyset$. For example, if $\alpha^{(n)}(i) = 1$



and $\beta^{(m)}(i) = 0$, then $A_{\alpha^{(n)}} \subseteq A_i$ and $A_{\beta^{(m)}} \subseteq A \setminus A_i$. Therefore, $\{A_{\alpha^{(1)}}\}_{\alpha^{(1)} \in \{0,1\}}$, $\{A_{\alpha^{(2)}}\}_{\alpha^{(2)} \in \{0,1\}^2}$, $\{A_{\alpha^{(n)}}\}_{\alpha^{(n)} \in \{0,1\}^n}$ form a binary hierarchical partition of A:

Thus, for $r \geq n$, $\cup_{k=n}^{r} B_k = B_n \cup E$ where $E \in \mathscr{F}$ is disjoint from B_n and $\lambda(E) \geq 0$, Thus, $\lambda(\cup_{k=n}^{r} B_k) \geq \lambda(B_n) + \lambda(E) \geq \lambda(B_n)$. Since r is arbitrary, define $C_n := \cup_{k=n}^{\infty} B_k$ and we have $\lambda(C_n) \geq \lambda(B_n)$).

Define $C := \limsup B_n := \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} B_k = \cap_{n=1}^{\infty} C_n$. Since $\lambda(C_1) < \infty$ by assumption and $C_1 \supseteq C_2 \supseteq C_3 \cdots \supseteq C$,

$$\begin{aligned}
\lambda(C) = \lambda(C_1) - \lambda(C_1 \setminus C) &= \lambda(C_1) - \lambda\{\underbrace{\cup_{n=1}^{\infty}(C_n \setminus C_{n+1})}_{\text{disjoint}}\} \\
&= \lambda(C_1) - \sum_{n=1}^{\infty}\{\lambda(C_n) - \lambda(C_{n+1})\} \\
&= \lim_{n\to\infty} \lambda(C_n).
\end{aligned}$$

Since $\lambda(C_n) \geq \lambda(B_n) \geq \lambda(A_n)$, we have that $\lambda(C) \geq \lim_{n\to\infty} \lambda(A_n)$.

On the other hand, $\lambda(C) \leq \sup\{\lambda(A') : A' \in \mathscr{F}\} = \lim_{n\to\infty} \lambda(A_n)$ implies $\lambda(C) = \sup\{\lambda(A') : A' \in \mathscr{F}\}$. To show that inf is attained, we apply previous analysis to $-\lambda$.

□

Remark 3.1.2. If λ is a measure, then $C = \Omega$ and $D = \emptyset$. If $\exists$ a measure $\mu : \mathscr{F} \to [0, \infty]$ and an integrable $f : \Omega \to \mathbb{R}$ such that $\forall A \in \mathscr{F}$, $\lambda(A) = \int_A f \mathrm{d}\mu$, then $C := \{\omega \in \Omega : f(\omega) \geq 0\}$ and $D := \{\omega \in \Omega : f(\omega) < 0\}$. In general, there is no simple expression for C, D. We will see in Theorem 3.1.1 that we can take $D = C^c$.

Theorem 3.1.1 (Jordan-Hahn Decomposition Theorem). Let $\lambda : \mathscr{F} \to [-\infty, \infty]$ be a signed measure. Define, $\forall A \in \mathscr{F}$, $\lambda^+(A) := \sup\{\lambda(B) : B \subseteq A, B \in \mathscr{F}\}$ and $\lambda^-(A) := \inf\{\lambda(B) : B \subseteq A, B \in \mathscr{F}\}$. Then, λ^+, λ^- are both measures on $(\Omega, \mathscr{F})$ and $\lambda = \lambda^+ - \lambda^-$. We write ($|\lambda| := \lambda^+ + \lambda^-$).

Proof.

Let $C = \arg\sup \lambda$ and $D = \arg\inf \lambda$. (Lemma 3.1.1) If $\lambda(C) = \infty$ and $\lambda(D) = -\infty$, then $\lambda(C \cup D) = \lambda(C) + \lambda(D)$ is undefined. Thus, assume WLOG that $\lambda(C) < \infty$ and thus $\forall A \in \mathscr{F}$, $\lambda(A) < \infty$. It implies $\forall A, A' \in \mathscr{F}$, $\lambda(A) - \lambda(A')$ is well-defined.

We note that $\forall A \in \mathscr{F}$, $\lambda(A \cap C) = \lambda(C) - \lambda(C \setminus A) \geq 0$ and $\lambda(A \cap C^c) = \lambda(C \cup (A \cap C^c)) - \lambda(C) \leq 0$. $(\star)$

We now claim that $\forall A \in \mathscr{F}$, $\lambda^+(A) = \lambda(A \cap C)$ and $\lambda^-(A) = -\lambda(A \cap C^c)$. To see this, let $B \in \mathscr{F}$ such that $B \subseteq A$, then

$$\begin{aligned}
\lambda(B) &= \lambda(B \cap C) + \lambda(B \cap C^c) \\
&= \lambda(A \cap C) - \underbrace{\lambda((A \setminus B) \cap C)}_{\geq 0 \text{ by } (\star)} + \underbrace{\lambda(B \cap C^c)}_{\leq 0 \text{ by } (\star)} \\
&\leq \lambda(A \cap C).
\end{aligned}$$

Likewise, we have

$$\begin{aligned}\lambda(B) &= \lambda(A\cap C^c) - \underbrace{\lambda((A\setminus B)\cap C^c)}_{\leq 0 \text{ by } (\star)} + \underbrace{\lambda(B\cap C)}_{\geq 0 \text{ by } (\star)}\\ &\geq \lambda(A\cap C^c).\end{aligned}$$

It is clear that $\lambda = \lambda^+ - \lambda^-$. It is then straightforward to verify that λ^+ and λ^- are measures. $\square$

Remark 3.1.3. Note that in the proof of Theorem 3.1.1, for any set $\widetilde{C}\in\mathscr{F}$ such that $\forall A\in\mathscr{F}$, $\lambda(A\cap\widetilde{C})\geq 0$ and $\lambda(A\cap\widetilde{C}^c)\leq 0$, then we may define, $\forall A\in\mathscr{F}$, $\lambda^+(A)=\lambda(A\cap\widetilde{C})$ and $\lambda^-(A)=\lambda(A\cap\widetilde{C}^c)$. So $\forall A\in\mathscr{F}, \lambda(\widetilde{C})\geq\lambda(A)$ and $\lambda(\widetilde{C}^c)\leq\lambda(A)$.

3.2 Radon-Nikodym Theorem

Definition 3.2.1. Let μ be a measure and λ be a signed measure on $(\Omega,\mathscr{F})$. We say that λ is absolutely continuous w.r.t. μ if

$$\forall E\in\mathscr{F}, \mu(E)=0 \implies \lambda(E)=0.$$

We write $\lambda\ll\mu$ and also say that μ dominates λ.

Note that if λ,μ are both measures, we may have $\lambda\ll\mu$ and $\mu\ll\lambda$.

Remark 3.2.1. Let μ be a measure and λ be a signed measure on $(\Omega,\mathscr{F})$. Recall, if $\exists$ Borel-measure $f:\Omega\to\mathbb{R}\cup\{\infty,-\infty\}$ such that $\forall A\in\mathscr{F}$, $\lambda(A)=\int_A f\mathrm{d}\mu$, then $\frac{\mathrm{d}\lambda}{\mathrm{d}\mu} := f$ is the Radon-Nikodym derivative of λ w.r.t. μ. By Theorem 2.2.3 (a), $\forall E\in\mathscr{F}$, $\mu(E)=0$ implies $\int_E f\mathrm{d}\mu=\lambda(E)=0$. Thus, if $\frac{\mathrm{d}\lambda}{\mathrm{d}\mu}$ exist, then $\mu\gg\lambda$. The Radon-Nikodym theorem say that, when μ is σ-finite, this condition is also sufficient,

$$\mu\gg\lambda \iff \frac{\mathrm{d}\lambda}{\mathrm{d}\mu} \text{ exists.}$$

Theorem 3.2.1 (Radon-Nikodym)**.** Let μ be a σ-finite measure and λ be a signed measure on $(\Omega,\mathscr{F})$. If $\mu\gg\lambda$, then $\exists f:\Omega\to\mathbb{R}\cup\{\infty,-\infty\}$ Borel-measurable such that

$$\forall A\in\mathscr{F}, \lambda(A)=\int_A f\mathrm{d}\mu.$$

If $g:\Omega\to\mathbb{R}\cup\{\infty,-\infty\}$ also satisfy this property, then $f=g$ μ-a.e. If $|\lambda|$ is σ-finite, then $|f|<\infty$ μ-a.e., if λ is a measure, then $f\geq 0$ μ-a.e.

Proof.

Uniqueness follows from Theorem 2.2.3. To show existence, we proceed in several steps:

Step 1: Suppose μ,λ are both finite measures, i.e., $\lambda,\mu\geq 0$ and $\lambda(\Omega),\mu(\Omega)<\infty$. Define

$$\mathscr{G} := \{f:\Omega\to[0,\infty] \text{ Borel-meas. s.t. } \forall A\in\mathscr{F}, \lambda(A)\geq\int_A f\mathrm{d}\mu\},$$

and $s:=\sup\{\int f\mathrm{d}\mu : f\in\mathscr{G}\}$. We claim $\exists g\in\mathscr{G}$ such that $\int g\mathrm{d}\mu = s$. Observe that if $f,\tilde{f}\in\mathscr{G}$, then, $\forall A\in\mathscr{F}$,

$$\begin{aligned}\int_A f\vee\tilde{f}\mathrm{d}\mu &= \int_{A\cap\{f\geq\tilde{f}\}} f\mathrm{d}\mu + \int_{A\cap\{f<\tilde{f}\}}\tilde{f}\mathrm{d}\mu\\ &\leq \lambda(A\cup\{f\geq\tilde{f}\}) + \lambda(A\cap\{f<\tilde{f}\})\\ &= \lambda(A) \implies f\vee\tilde{f}\in\mathscr{G}.\end{aligned}$$

Now let $f_1, f_2, \dots \in \mathscr{G}$ be such that $\int f_n \mathrm{d}\mu \to s$. Define $g_n : f_1 \vee f_2 \vee \ldots \vee f_n$; note that $g_n \in \mathscr{G}$ and $0 \le g_1 \le g_2 \le \ldots$. Define $g : \Omega \to [0, \infty]$ such that $\forall \omega \in \Omega$, $g(\omega) := \lim_{n\to\infty} g_n(\omega)$. Then,

$$\int f_n \mathrm{d}\mu \le \int g_n \mathrm{d}\mu \underset{\text{(by MCT)}}{\to} \int g \mathrm{d}\mu \implies \int g \mathrm{d}\mu \ge s.$$

Let $A \in \mathscr{F}$, then $\omega \in \Omega$,

$$\begin{aligned} &(g_n \mathbb{1}_A)(\omega) \to (g\mathbb{1}_A)(\omega) \\ \implies &\int_A g \mathrm{d}\mu \underset{\text{(by MCT)}}{=} \lim_{n\to\infty} \int_A g_n \mathrm{d}\mu \le \lambda(A) \end{aligned}$$

since $\lambda(A) \ge \int_A g_n \mathrm{d}\mu$ as $g_n \in \mathscr{G}$. So $g \in \mathscr{G}$ and $\int g\mathrm{d}\mu = s$.

Now fix $k \in (0, \infty)$ and define, $\forall A \in \mathscr{F}$, $\lambda_1(A) := \lambda(A) - \int_A g\mathrm{d}\mu$. Note that λ_1 is a finite measure and $\lambda_1 \ll \mu$. Since μ is a finite measure, $\mu - k\lambda_1$ is a signed measure. By Lemma 3.1.1 and Theorem 3.1.1, $\exists C \in \mathscr{F}$ such that

$$\forall A \in \mathscr{F}, \mu(A \cap C) - k\lambda_1(A \cap C) \ge 0 \text{ and } \mu(A \cap C^c) - k\lambda_1(A \cap C^c) \le 0. \tag{$\star$}$$

Thus, $\forall A \in \mathscr{F}$,

$$\begin{aligned} \int_A \frac{1}{k}\mathbb{1}_{C^c} + g\mathrm{d}\mu &= \frac{1}{k}\mu(A \cap C^c) + \int_A g\mathrm{d}\mu \\ &\le \lambda_1(A \cap C^c) + \int_A g\mathrm{d}\mu && (\text{ by } (\star)) \\ &\le \lambda_1(A) + \int_A g\mathrm{d}\mu = \lambda(A). && (\ \lambda_1 \text{ is a measure}) \end{aligned}$$

Since $\int g\mathrm{d}\mu = \sup\{\int f\mathrm{d}\mu : f \in \mathscr{G}\}$ and $\frac{1}{k}\mathbb{1}_{C^c} + g \in \mathscr{G}$, it must be that $\int \frac{1}{k}\mathbb{1}_{C^c}\mathrm{d}\mu = \frac{1}{k}\mu(C^c) = 0 \implies \mu(C^c) = 0 \implies \lambda_1(C^c) = 0$ since $\lambda_1 \ll \mu$. Thus,

$$\begin{aligned} \mu(\Omega) - k\lambda_1(\Omega) &= \mu(C) - k\lambda_1(C) + \mu(C^c) - k\lambda_1(C^c) \\ &= \mu(C) - k\lambda_1(C) \ge 0 && (\text{by } (\star)) \end{aligned}$$

Since $k \in (0, \infty)$ is arbitrary, it must be that $\lambda_1(\Omega) = 0$. So $\lambda_1(A) = \lambda(A) - \int_A g\mathrm{d}\mu = 0\ \forall A \in \mathscr{F}$ as desired. We may take $\frac{\mathrm{d}\lambda}{\mathrm{d}\mu} = g$.

Step 2: μ is a finite measure and λ is a σ-finite measure. Since λ is σ-finite, $\exists A_1, A_2, \dots \in \mathscr{F}$ disjoint such that $\lambda(A_n) < \infty\ \forall n \in \mathbb{N}$ and $\Omega = \cup_{n=1}^\infty A_n$. Define $\forall B \in \mathscr{F}$, $\lambda_n(B) := \lambda(B \cap A_n)$. By step 1, $\forall n \in \mathbb{N}$, $\exists g_n : \Omega \to [0, \infty]$ such that $\forall B \in \mathscr{F}$, $\lambda_n(B) = \int_B g_n\mathrm{d}\mu$. Then, $\forall B \in \mathscr{F}$,

$$\begin{aligned} \lambda(B) = \sum_{n=1}^\infty \lambda_n(B) &= \sum_{n=1}^\infty \int_B g_n \mathrm{d}\mu \\ &= \int_B \sum_{n=1}^\infty g_n\mathrm{d}\mu && (\text{By MCT applied to } \textstyle\sum_{m=1}^n g_m) \end{aligned}$$

So we may take $\frac{\mathrm{d}\lambda}{\mathrm{d}\mu} = \sum_{n=1}^\infty g_n$.

Step 3: μ is a finite measure and λ is an arbitrary measure. Let

$$\mathcal{C} := \{C \in \mathscr{F} : \exists C^{(1)}, C^{(2)}, \dots \in \mathscr{F} \text{ disjoint s.t. } \lambda(C^{(n)}) < \infty, \forall n \in \mathbb{N} \text{ and } C = \cup_{n=1}^\infty C^{(n)}\}.$$

Note that $\emptyset \in \mathcal{C}$ so that $\mathcal{C}$ is non-empty. Let $s := \sup\{\mu(A) : A \in \mathcal{C}\}$ and let $C_1, C_2, \dots \in \mathcal{C}$ be such that $\mu(C_n) \to s$. Define $C := \cup_{n=1}^\infty C_n$. Note that

$$C = (\cup_{n=1}^\infty C_1^{(n)}) \cup (\cup_{n=1}^\infty C_2^{(n)} \setminus C_1) \cup (\cup_{n=1}^\infty C_3^{(n)} \setminus (C_1 \cup C_2)) \cup \ldots,$$

so $C \in \mathcal{C}$, which implies $s \geq \mu(C)$. But $\mu(C) \geq \mu(C_n) \implies \mu(C) \geq \lim_{n\to\infty} \mu(C_n) = s \implies \mu(C) = s$. Define $\mathscr{F}_C := \{A \cap C : A \in \mathscr{F}\}$. Note that $\mathscr{F}_C$ is a σ-field and λ is σ-finite on $(C, \mathscr{F}_C)$. Hence, by Step 2, $\exists \tilde{g} : C \to [0,\infty]$ measurable w.r.t. $(C, \mathscr{F}_C)$ and $([0,\infty], \mathscr{B}([0,\infty]))$ such that

$$\forall A \in \mathscr{F}, \lambda(A \cap C) = \int_{A\cap C} \tilde{g} \mathrm{d}\mu.$$

Define $g : \Omega \to [0,\infty]$ such that $\forall \omega \in \Omega$,

$$g(\omega) = \begin{cases} \tilde{g}(\omega) & \text{if } \omega \in C \\ \infty & \text{if } \omega \in C^c. \end{cases}$$

Let $A \in \mathscr{F}$. If $\lambda(A \cap C^c) = \infty$, then $\mu(A \cap C^c) > 0$, $\lambda(A) = \infty$ and

$$\int_A g \mathrm{d}\mu = \int_{A\cap C} g \mathrm{d}\mu + \int_{A\cap C^c} g \mathrm{d}\mu = \infty = \lambda(A).$$

Now suppose $\lambda(A \cap C^c) < \infty$, then $C \cup (A \cap C^c) \in \mathcal{C}$. Then $\mu(A \cap C^c) = \mu(C \cup (A \cap C^c)) - \mu(C) \leq 0$ since $\mu(C) = \sup\{\mu(A) : A \in \mathcal{C}\} \implies \mu(A \cap C^c) = 0 \implies \lambda(A \cap C^c) = 0$ since $\mu \gg \lambda$. Thus,

$$\int_A g \mathrm{d}\mu = \int_{A\cap C} g \mathrm{d}\mu + \int_{A\cap C^c} g \mathrm{d}\mu = \int_{A\cap C} \tilde{g} \mathrm{d}\mu = \lambda(A \cap C) = \lambda(A).$$

We may then take $\frac{\mathrm{d}\lambda}{\mathrm{d}\mu} = g$.

Step 4: μ is σ-finite measure, λ is an arbitrary measure. Let $A_1, A_2, \dots \in \mathscr{F}$ disjoint such that $\cup_{n=1}^{\infty} A_n = \Omega$ and $\mu(A_n) < \infty$. Define, for any $A \in \mathscr{F}$,

$$\lambda_n(A) = \lambda(A \cap A_n), \ \mu_n(A) = \mu(A \cap A_n).$$

Since $\lambda_n \ll \mu_n$, we have by Step 3 that there exists $g_n := \frac{d\lambda_n}{d\mu_n} : \Omega \to [0,\infty]$. Moreover, note that $\mu_n \ll \mu$ and that $\mathbb{1}_{A_n} = \frac{d\mu_n}{d\mu}$.

Thus,

$$\lambda(A) = \sum_{n=1}^{\infty} \lambda_n(A) = \sum_{n=1}^{\infty} \int_A g_n \mathrm{d}\mu_n = \int_A \sum_{n=1}^{\infty} g_n \mathbb{1}_{A_n} \mathrm{d}\mu.$$

So we may take $\frac{\mathrm{d}\lambda}{\mathrm{d}\mu} = \sum_{n=1}^{\infty} g_n \mathbb{1}_{A_n}$. Now, note that if λ is σ-finite, then $\exists$ disjoint $B_1, B_2, \dots \in \mathscr{F}$ such that $\lambda(B_n) < \infty$ and $\cup_{n=1}^{\infty} B_n = \Omega$. Define $E := \{\omega \in \Omega : \frac{\mathrm{d}\lambda}{\mathrm{d}\mu}(\omega) = \infty\}$, then,

$$\begin{aligned} \int_{E\cap B_n} \frac{\mathrm{d}\lambda}{\mathrm{d}\mu} \mathrm{d}\mu = \underbrace{\lambda(E \cap B_n)}_{\text{finite}} \ \ \forall n \in \mathbb{N} \\ \implies \mu(E \cap B_n) = 0 \quad \forall n \in \mathbb{N} \\ \implies \mu(E) = 0. \end{aligned}$$

Thus, we may take $\frac{\mathrm{d}\lambda}{\mathrm{d}\mu} : \Omega \to [0,\infty)$.

Step 5: μ is σ-finite measure and λ is arbitrary signed measure. By Theorem 3.1.1, we can write $\lambda = \lambda^+ - \lambda^-$ for 2 measures λ^+, λ^- one of which is finite. Assume λ^- is finite WLOG. By Step 4, $\exists g_1, g_2 : \Omega \to [0,\infty]$ such that $\forall A \in \mathscr{F}$, $\lambda^+(A) = \int_A g_1 \mathrm{d}\mu$ and $\lambda^-(A) = \int_A g_2 \mathrm{d}\mu$. Since λ^- is finite, $g_2 < \infty$ μ-a.e. and thus, $g_1 - g_2$ is well defined, and

$$\lambda(A) = \int_A g_1 \mathrm{d}\mu - \int_A g_2 \mathrm{d}\mu = \int_A g_1 - g_2 \mathrm{d}\mu \qquad \left(\text{since } \int_A g_2 \mathrm{d}\mu < \infty\right)$$

So we may take $\frac{\mathrm{d}\lambda}{\mathrm{d}\mu} = g_1 - g_2$ as desired. $\square$

Theorem 3.2.2. Let μ, η be σ-finite measure and λ be a signed measure.

(a) If $\mu \gg \lambda$ and $\eta \gg \mu$, then $\eta \gg \lambda$ and $\frac{\mathrm{d}\lambda}{\mathrm{d}\eta} = \frac{\mathrm{d}\lambda}{\mathrm{d}\mu} \cdot \frac{\mathrm{d}\mu}{\mathrm{d}\eta}$. In particular, if $\mu \gg \eta$, and $\eta \ll \mu$, then $\frac{\mathrm{d}\mu}{\mathrm{d}\eta} = (\frac{\mathrm{d}\eta}{\mathrm{d}\mu})^{-1}$.

(b) If ν is signed measure, and ν^-, λ^- are both finite, say, then $\nu + \lambda$ is a signed measure and $\mu \gg \nu$ and $\mu \gg \lambda$, then $\frac{\mathrm{d}(\nu+\lambda)}{\mathrm{d}\mu} = \frac{\mathrm{d}\nu}{\mathrm{d}\mu} + \frac{\mathrm{d}\lambda}{\mathrm{d}\mu}$.

3.3 Lebesgue Decomposition

Definition 3.3.1. Let μ_1, μ_2 be measures on $(\Omega, \mathscr{F})$. We say that μ_1, μ_2 are <u>mutually singular</u> if $\exists A \in \mathscr{F}$ such that $\mu_1(A) = 0$ and $\mu_2(A^c) = 0$. We write $\mu_1 \perp \mu_2$. We say that 2 signed measures λ_1, λ_2 are singular if $|\lambda_1| \perp |\lambda_2|$. (Equivalently, $\lambda_1^+ \perp \lambda_2^+, \lambda_1^- \perp \lambda_2^+, \lambda_1^+ \perp \lambda_2^-, \lambda_1^- \perp \lambda_2^-$).

Lemma 3.3.1 (Borel-Cantelli)**.** Let μ be a measure and let $A_1, A_2, \dots \in \mathscr{F}$ such that $\sum_{n=1}^\infty \mu(A_n) < \infty$, then

$$\mu\left(\bigcap_{n=1}^\infty \bigcup_{k=n}^\infty A_k\right) = 0. \tag{3.1}$$

Recall $\omega \in \cap_{n=1}^\infty \cup_{k=n}^\infty A_k$ iff $\omega \in A_n$ for infinitely many $n \in \mathbb{N}$.

Proof.

Observe that for any $n \in \mathbb{N}$, $\mu(\cup_{k=n}^\infty A_k) \leq \sum_{k=n}^\infty \mu(A_k)$. Since $\sum_{n=1}^\infty \mu(A_n) < \infty$,

$$0 = \lim_{n\to\infty} \sum_{k=n}^\infty \mu(A_k) = \lim_{n\to\infty} (\cup_{k=n}^\infty A_k) = \mu(\cap_{n=1}^\infty \cup_{k=n}^\infty A_k).$$

□

Lemma 3.3.2. Let μ be a measure and λ_1, λ_2 be signed measures on $(\Omega, \mathscr{F})$.

(a) If $\lambda_1 \perp \mu$ and $\lambda_2 \perp \mu$, then $\lambda_1 + \lambda_2 \perp \mu$ so long as $\lambda_1 + \lambda_2$ is well-defined. The same applies for countable sums. (Note that if $\lambda_1, \lambda_2 \ll \mu$, then $\lambda_1 + \lambda_1 \ll \mu$ as well.)

(b) $\lambda_1 \ll \mu$ iff $|\lambda_1| \ll \mu$.

(c) If $\lambda_1 \ll \mu$, $\lambda_2 \perp \mu$, then $\lambda_1 \perp \lambda_2$.

(d) If $\lambda_1 \ll \mu$ and $\lambda_1 \perp \mu$, then $\lambda_1 = 0$.

(e) If λ_1 is finite, then $\lambda_1 \ll \mu$ iff for any $A_1, A_2, \dots \in \mathscr{F}$, $\lim_{n\to\infty} \mu(A_n) = 0$ implies $\lim_{n\to\infty} \lambda_1(A_n) = 0$.

Proof.

(a) $\exists A, B \in \mathscr{F}$ such that $|\lambda_1|(A) = 0$ and $\mu(A^c) = 0$ and $|\lambda_2|(B) = 0$ and $\mu(B^c) = 0$. Thus,

$$|\lambda_1 + \lambda_2|(A \cap B) \leq |\lambda_1|(A \cap B) + |\lambda_2|(A \cap B) = 0$$

and $\mu(A^c \cup B^c) = 0$. In the case of countable sums, the argument is identical.

(b) Suppose $\lambda_1 \ll \mu$. Let $E \in \mathscr{F}$ be such that $\mu(E) = 0$, then, by Remark 3.1.3, $\lambda_1^+(E) = \sup\{\lambda_1(B) : B \subseteq E, B \in \mathscr{F}\} = 0$ since $\mu(B) = 0$ and thus $\lambda_1(B) = 0$ for any $B \in \mathscr{F}$ that is a subset of E. Likewise, $\lambda_1^-(E) = 0$ and we have that $|\lambda_1| = \lambda_1^+ + \lambda_1^- \ll \mu$ by claim (a). Other direction is obvious.

(c) $\exists A \in \mathscr{F}$ such that $|\lambda_2|(A) = 0$ and $\mu(A^c) = 0 \implies |\lambda_1|(A^c) = 0$.

(d) $\exists A \in \mathscr{F}$ such that $|\lambda_1|(A) = 0$, $\mu(A^c) = 0 \implies |\lambda_1|(A^c) = 0 \implies |\lambda_1| = 0 \implies \lambda_1 = 0$.

(e) Suppose $\lambda_1 \ll \mu$ and let $A_1, A_2, \dots \in \mathscr{F}$ be such that $\lim_{n\to\infty} \mu(A_n) = 0$. Let $\{\widetilde{A}_1, \widetilde{A}_2, \dots\} \subseteq \{A_1, A_2, \dots\}$ be a subsequence such that

$$\lim_{n\to\infty} |\lambda_1|(\widetilde{A}_n) = \limsup_{n\to\infty} |\lambda_1|(A_n).$$

Since $\lim_{n\to\infty} \mu(\widetilde{A}_n) = 0$, there exists subsequence $1 \le n_1 \le n_2 \le \dots \in \mathbb{N}$ such that $\forall m \in \mathbb{N}$, $\mu(\widetilde{A}_{n_m}) \le 2^{-m}$ and that $\lim_{m\to\infty} |\lambda_1|(\tilde{A}_{n_m}) = \lim_{n\to\infty} |\lambda_1|(\tilde{A}_n)$. Let $\tilde{A} := \cap_{m=1}^\infty \cup_{k=m}^\infty \widetilde{A}_{n_k}$. Then $\mu(\widetilde{A}) = 0$ by Lemma 3.3.1 implies $|\lambda_1|(\widetilde{A}) = 0$. Since $|\lambda_1|$ is finite,

$$0 = |\lambda_1|(\tilde{A}) = \lim_{m\to\infty} |\lambda_1|(\cup_{k=m}^\infty \widetilde{A}_{n_k}) \ge \lim_{m\to\infty} |\lambda_1|(\widetilde{A}_{n_m}) \implies \lim_{n\to\infty} |\lambda_1|(\widetilde{A}_n) = 0.$$

□

Theorem 3.3.1 (Lebesgue Decomposition). Let μ be a measure and λ be a σ-finite signed measure on $(\Omega, \mathscr{F})$. Then $\exists$ signed measure $\lambda_1 \ll \mu$, $\lambda_2 \perp \mu$ such that $\lambda = \lambda_1 + \lambda_2$. Moreover, this decomposition is unique.

Proof.
Let us first prove uniqueness. Assume $|\lambda|$ is finite. If $\exists \lambda_1, \tilde{\lambda} \ll \mu$ and $\lambda_2, \tilde{\lambda}_2 \perp \mu$ such that $\lambda = \lambda_1 + \lambda_2 = \tilde{\lambda}_1 + \tilde{\lambda}_2$, then $\lambda_1 - \tilde{\lambda}_1 = \lambda_2 - \tilde{\lambda}_2$ is both absolutely continuous with respect to μ and singular to μ. So $\lambda_1 - \tilde{\lambda}_1 = \lambda_2 - \tilde{\lambda}_2 = 0$ by lemma 3.3.2 (d). If $|\lambda|$ is σ-finite, then $\exists A_1, A_2, \dots \in \mathscr{F}$ disjoint such that $|\lambda|$ is finite on A_n $\forall n \in \mathbb{N}$. We may then apply the same argument for each $n \in \mathbb{N}$.

Now we prove existence. First assume that λ is a finite measure. Define $\mathcal{C} := \{A \in \mathscr{F} : \mu(A) = 0\}$ and $s := \sup\{\lambda(A) : A \in \mathcal{C}\} \le \lambda(\Omega) < \infty$. Let $A_1, A_2, \dots \in \mathcal{C}$ such that $\lambda(A_n) \to s$, then $A^\star := \cup_{n=1}^\infty A_n \in \mathcal{C}$ and $\lambda(A^\star) = s$. Note that for every $B \in \mathcal{C}$,

$$\lambda(B \cap A^{\star c}) = \lambda(B \cup A^\star) - \lambda(A^\star) \le 0 \implies \lambda(B \cap A^{\star c}) = 0 \qquad (\star)$$

Define, $\forall A \in \mathscr{F}$, $\lambda_1(A) := \lambda(A \cap A^{\star c})$ and $\lambda_2 := \lambda(A \cap A^\star)$. We have that λ_1, λ_2 are measures and $\lambda = \lambda_1 + \lambda_2$. If $E \in \mathscr{F}$ such that $\mu(E) = 0$, then $\lambda_1(E) = 0$ by $(\star)$ and so $\lambda_1 \ll \mu$. Also, $\lambda_2 \perp \mu$ since $\lambda_2(A^{\star c}) = 0$ and $\mu(A^\star) = 0$. If λ is a signed measure such that $|\lambda|$ is finite, then we may apply the same argument to λ^+ and λ^- and obtain the same conclusion.

Now, if λ is a σ-finite signed measure, then $\exists A_1, A_2, \dots, \in \mathscr{F}$ disjoint such that $|\lambda|$ is finite on A_n $\forall n \in \mathbb{N}$. Thus, $\forall n \in \mathbb{N}$, $\exists \lambda_1^{(n)}, \lambda_2^{(n)}$ such that $\lambda_1^{(n)} \ll \mu$ and $\lambda_2^{(n)} \perp \mu$ and $\lambda = \sum_{n=1}^\infty \lambda_1^{(n)} + \sum_{n=1}^\infty \lambda_2^{(n)}$. Setting $\lambda_1 := \sum_{n=1}^\infty \lambda_1^{(n)}$ and $\lambda_2 := \sum_{n=1}^\infty \lambda_2^{(n)}$ finishes the proof. □

3.4 Extended Example

Definition 3.4.1. We say that $\eta : [0,1] \to \{0,1\}$ is finite point pattern if $|\{t \in [0,1] : \eta(t) = 1\}| < \infty$. Let $\mathcal{X}$ denote the set of all finite point patterns on $[0,1]$. For any $S \subseteq [0,1]$, define $\eta(S) := |\{t \in S : \eta(t) = 1\}|$.

For any $B \in \mathscr{B}([0,1])$ and $k \in \mathbb{N}_0$, define

$$A_{B,k} := \{\eta \in \mathcal{X} : \eta(B) = k\}.$$

Let $\mathscr{G} := \sigma(\{A_{[0,b],k} : b \in [0,1], k \in \mathbb{N}_0\})$ so that $(\mathcal{X}, \mathscr{G})$ is a measurable space.

Let μ be the Lebesgue measure on $[0,1]$. For a Borel-measurable function $\lambda : [0,1] \to [0,\infty)$ such that $\int_{[0,1]} \lambda d\mu < \infty$, we say that PP_λ is a Poisson point process if it is a probability measure on $(\mathcal{X}, \mathscr{G})$ such that for any $n \in \mathbb{N}$, any $B_1, B_2, \dots, B_n \in \mathscr{B}([0,1])$ disjoint, any $k_1, \dots, k_n \in \mathbb{N}_0$,

$$PP_\lambda(\cap_{i=1}^n A_{B_i,k_i}) = \prod_{i=1}^n PP_\lambda(A_{B_i,k_i}) = \prod_{i=1}^n \frac{\left(\int_{B_i} \lambda d\mu\right)^{k_i} e^{-\int_{B_i} \lambda d\mu}}{k_i!}.$$

We will formally show the existence of Poisson process later on. We say that λ is the intensity function of PP_λ.

Proposition 3.4.1. Let PP_1 denote the Poisson point process with a constant intensity function of 1. Let $\lambda : [0,1] \to [0,\infty)$ be finitely integrable with respect to the Lebesgue measure μ. Then, $PP_\lambda \ll PP_1$ and

$$\log \frac{dPP_\lambda}{dPP_1}(\eta) = \sum_{i=1}^{N_\eta} \log \lambda(\eta_i) + 1 - \int \lambda d\mu,$$

where $N_\eta := \eta([0,1])$ and $\eta_1, \ldots, \eta_{N_\eta} \in [0,1]$ are the locations of the points.

Proof.

We prove the most basic case where λ is a constant, say $\lambda(t) = \alpha \geq 0$ for all $t \in [0,1]$. In this case, we want to show that for any $A \in \mathscr{G}$,

$$PP_\alpha(A) = \int_A \alpha^{N_\eta} e^{\alpha-1} dPP_1(\eta). \tag{3.2}$$

Note that $N_\eta : \mathcal{X} \to [0,\infty)$ is $\mathscr{G}/\mathscr{B}([0,\infty))$-measurable and thus the RHS is well-defined.

Define $\mathcal{A} := \{\cap_{i=1}^n A_{B_i,k_i} : n \in \mathbb{N}, B_1, \ldots, B_n \text{ disjoint intervals}, k_, \ldots, k_n \in \mathbb{N}_0\}$. We may verify that $\mathcal{A} \cup \{\emptyset\}$ is a Π-system (closed under intersection). Since $\mathscr{G} = \sigma(\mathcal{A})$, it suffices by the $\Pi - \Lambda$ theorem (see Remark 1.3.1) to show that (3.2) holds for sets in $\mathscr{A}$. By the definition of Poisson point process, it suffices to further limit our attention to $A_{[0,b],k}$ for some $b \in [0,1]$ and $k \in \mathbb{N}_0$.

To that end, observe that

$$\begin{aligned}
&\int_{A_{[0,b],k}} \alpha^{N_\eta} e^{1-\alpha} dPP_1(\eta) \\
&= e^{1-\alpha} \sum_{m=0}^{\infty} \int_{A_{[0,b],k} \cap A_{(b,1],m}} \alpha^{N_\eta} dPP_1(\eta) \\
&= e^{1-\alpha} \sum_{m=0}^{\infty} \alpha^{k+m} PP_1(A_{[0,b],k} \cap A_{(b,1],m}) \\
&= e^{1-\alpha} \sum_{m=0}^{\infty} \alpha^{k+m} \frac{b^k e^{-b}}{k!} \frac{(1-b)^m e^{-(1-b)}}{m!} \\
&= \frac{(\alpha b)^k e^{-\alpha}}{k!} \sum_{m=0}^{\infty} \frac{\alpha^m (1-b)^m}{m!} = \frac{(\alpha b)^k e^{-\alpha b}}{k!} = PP_\alpha(A_{[0,b],k}),
\end{aligned}$$

as desired.

We may then prove same result for a simple λ and obtain the general conclusion through an application of MCT.

□

Lemma 3.4.1. Let ν, μ, λ be σ-finite measures on $(\Omega, \mathscr{F})$. If $\nu \ll \mu$ and $\nu \ll \lambda$, then, writing $\lambda = \lambda_1 + \lambda_2$ where $\lambda_1 \ll \mu$ and $\lambda_2 \perp \mu$, we have that $\nu \ll \lambda_1$ and

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\lambda_1} \mathbb{1}_A$$

where $A \in \mathscr{F}$ is a set where $\mu(A^c) = 0$ and $\lambda_2(A) = 0$.

Remark 3.4.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Let Θ be a set and let $\{P_\theta\}_{\theta \in \Theta}$ be a family of probability measures on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that, for some σ-finite measure μ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, $P_\theta \ll \mu$ for all $\theta \in \Theta$.

For $\theta_0 \in \Theta$, let $X : \Omega \to \mathbb{R}$ be random variable (Borel-measurable) such that $\mathbb{P}^{(X)} = P_{\theta_0}$.

Given independent samples (formalized later) $X_1, \ldots, X_n$, the maximum likelihood estimator for θ_0 is then

$$\hat{\theta} := \arg\max_{\theta} \sum_{i=1}^{n} \log \frac{dP_\theta}{d\mu}(X_i).$$

Suppose there exists another measure λ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that for all $\theta \in \Theta$, $P_\theta \ll \lambda$. Let $\lambda = \lambda_1 + \lambda_2$ where $\lambda_1 \ll \mu$ and $\lambda_2 \perp \mu$, then $P_\theta \ll \lambda_1$ for all $\theta \in \Theta$ and

$$\frac{dP_\theta}{d\lambda} = \frac{dP_\theta}{d\lambda_1}.$$

$$\log \frac{dP_\theta}{d\lambda} = \log \frac{dP_\theta}{d\lambda_1} = \log \frac{dP_\theta}{d\lambda_1}\frac{d\lambda_1}{d\mu} - \log \frac{d\lambda_1}{d\mu} = \log \frac{dP_\theta}{d\mu} - \log \frac{d\lambda_1}{d\mu}.$$

Thus we have that

$$\arg\max_{\theta} \sum_{i=1}^{n} \log \frac{dP_\theta}{d\mu}(X_i) = \arg\max_{\theta} \sum_{i=1}^{n} \log \frac{dP_\theta}{d\lambda}(X_i).$$

Therefore, given just one sample η, we may estimate the intensity function of an imhomogeneous Poisson process by

$$\arg\max_{\lambda \in \Lambda} \sum_{i=1}^{N_\eta} \log \lambda(\eta_i) - \int \lambda d\mu$$

where Λ is some subset of all finitely integrable functions from $[0,1]$ to $[0,\infty)$.

3.5 Data Processing Inequality

Lemma 3.5.1. Let $(\mathcal{X}, \mathscr{F}, P)$ be a probability space and let $(\mathcal{Y}, \mathscr{G})$ be a measurable space. Let $f : \mathcal{X} \to \mathbb{R}$ be $\mathscr{F}/\mathscr{B}(\mathbb{R})$-measurable and $\phi : \mathcal{X} \to \mathcal{Y}$ be $\mathscr{F}/\mathscr{G}$-measurable. Recall that $\sigma(\phi) := \{\phi^{-1}(B) : B \in \mathscr{G}\}$ is the σ-field generated by ϕ and is a sub-σ-field of $\mathscr{F}$.

The following are true:

1. If f is $\sigma(\phi)/\mathscr{B}(\mathbb{R})$-measurable, then there exists $g : \mathcal{Y} \to \mathbb{R}$ Borel-measurable such that $f(x) = g(\phi(x))$ for all $x \in \mathcal{X}$.

2. For any f that is $\mathscr{F}/\mathscr{B}(\mathbb{R})$-measurable and integrable, there exists a P-a.e. unique function $E_P[f|\phi] : \mathcal{X} \to \mathbb{R}$ that is $\sigma(\phi)/\mathscr{B}(\mathbb{R})$-measurable such that $\int_A f dP = \int_A E_P[f|\phi] dP$ for all $A \in \sigma(\phi)$.

3. If $f \geq 0$ P-a.e, then $E_P[f|\phi] \geq 0$. Also, for any $a, b \in \mathbb{R}$, $E_P[af + b|\phi] = aE_P[f|\phi] + b$.

4. Let $r : \mathbb{R} \to \mathbb{R}$ be a convex function. Then, $r \circ E_P[f|\phi] \leq E_P[r \circ f|\phi]$.

Proof.

Claim 1 is proved in Lemma 6.1.1. Claim 2 is proved in Theorem 6.1.1 but we repeat it here. We define $\lambda : \sigma(\phi) \to \bar{\mathbb{R}}$ by

$$\lambda(A) = \int_A f dP.$$

Then, λ is a signed measure on $(\mathcal{X}, \sigma(\phi))$ and $\lambda \ll P$. Hence, we may take $E_P[f|\phi] = \frac{d\lambda}{dP}$.

For the first part of the third, claim, note that λ is measure if $f \geq 0$. Hence, $E_P[f|\phi] \geq 0$. The second part follows by noting that for any $A \in \sigma(\phi)$,

$$\int_A (af + b)\, dP = a \int_A f dP + bP(A) = \int_A (aE_P[f|\phi] + b)\, dP.$$

For the last claim, we note that for any $z \in \mathbb{R}$,

$$r(z) = \sup_{h\,:\,\mathbb{R}\to\mathbb{R}\text{ linear},\, h\le z} h(z).$$

Hence, by claim 3, we have that, for any linear function $h \le r$,

$$E_P[r \circ f|\phi] \ge E_P[h \circ f|\phi] = h \circ E_P[f|\phi].$$

Since this is true for any linear $h \le r$, we have that $E_P[r \circ f|\phi] \ge r \circ E_P[f|\phi]$ as desired. □

Example 3.5.1. Let $(\mathcal{X}, \mathscr{F}) = ([0,1], \mathscr{B}([0,1]))$ with P as the Lebesgue measure and let $f(t) = t$.

Let $\phi : \mathcal{X} \to \{0,1\}$ be such that $\phi(t) = 0$ if $t < 1/2$ and $\phi(t) = 1$ if $t \ge 1/2$.

Then, we have that

$$\sigma(\phi) = \{\emptyset, [0,1], [0,1/2), [1/2,1]\}.$$

Note that f is not $\sigma(\phi)/\mathscr{B}(\mathbb{R})$-measurable. We claim that

$$E_P[f|\phi](t) = \begin{cases} 1/4 & \text{if } t < 1/2 \\ 3/4 & \text{if } t \ge 1/2 \end{cases}$$

Indeed, we have that

$$\int_{[0,1/2)} f\, dP = \int_{[0,1/2)} E_P[f|\phi]\, dP = 1/8$$
$$\int_{[1/2,1]} f\, dP = \int_{[1/2,1]} E_P[f|\phi]\, dP = 3/8.$$

Proposition 3.5.1. Let P, Q be probability measures on $(\mathcal{X}, \mathscr{F})$. Let $(\mathcal{Y}, \mathscr{G})$ be another measurable space and let $\phi : \mathcal{X} \to \mathcal{Y}$ be $\mathscr{F}/\mathscr{G}$-measurable.

If $P \ll Q$, then $P^{(\phi)} \ll Q^{(\phi)}$ and

$$\frac{dP^{(\phi)}}{dQ^{(\phi)}} \circ \phi = E_Q\left[\frac{dP}{dQ}\middle|\phi\right].$$

Proof.

By Lemma 3.5.1 and the fact that $E_Q\left[\frac{dP}{dQ}\middle|\phi\right]$ is $\sigma(\phi)/\mathscr{B}(\mathbb{R})$-measurable, there exists $g : \mathcal{Y} \to [0,\infty]$ such that

$$g \circ \phi = E_Q\left[\frac{dP}{dQ}\middle|\phi\right].$$

We will show that $g = \frac{dP^{(\phi)}}{dQ^{(\phi)}}$. To that end, let $B \in \mathscr{G}$, then

$$\begin{aligned} P^{(\phi)}(B) = P(\phi^{-1}(B)) &= \int_{\phi^{-1}(B)} \frac{dP}{dQ} dQ \\ &= \int_{\phi^{-1}(B)} E_Q\left[\frac{dP}{dQ}\middle|\phi\right] dQ \\ &= \int_{\phi^{-1}(B)} g \circ \phi\, dQ = \int_B g\, dQ^{(\phi)}. \end{aligned}$$

The desired conclusion follows. □

Definition 3.5.1. Let $f : (0,\infty) \to \mathbb{R}$ be convex. Then, for any $s > 0$, the function $t \mapsto \frac{f(t)-f(s)}{t-s}$ is increasing in t and its limit $\lim_{t\to\infty} \frac{f(t)-f(s)}{t-s}$ does not depend on s. To see this, note that for $t' > t$,

$$\frac{f(t') - f(s)}{t' - s} = \frac{t - s}{t' - s}\frac{f(t) - f(s)}{t - s} + \frac{t' - t}{t' - s}\frac{f(t') - f(t)}{t' - t}.$$

Thus, we may define the maximal slope of f by

$$M_f := \lim_{t\to\infty} \frac{f(t)-f(0)}{t} = \lim_{t\to\infty} \frac{f(t)}{t} \in \mathbb{R} \cup \{\infty\},$$

where $f(0) = \lim_{x\searrow 0} f(x)$ is defined by continuity.

Note that for all $s, t \in [0,\infty)$, we have that

$$f(s+t) \leq f(s) + tM_f.$$

Definition 3.5.2. Let P, Q be two probability measures on $(\mathcal{X}, \mathscr{F})$. Write $P = P_1 + P_2$ where $P_1 \ll Q$ and $P_2 \perp Q$. Given a convex function $f : (0,\infty) \to \mathbb{R}$ with $f(1) = 0$, we define the f-divergence

$$D_f(P,Q) = \int_{\mathcal{X}} f\left(\frac{dP_1}{dQ}(x)\right) dQ(x) + P_2(\mathcal{X})M_f.$$

Note that, by Jensen's inequality,

$$D_f(P,Q) \geq f(P_1(\mathcal{X})) + P_2(\mathcal{X})M_f \geq f(P_1(\mathcal{X}) + P_2(\mathcal{X})) \geq 0.$$

Note also that, for any $t \in [0,1]$,

$$D_f(tP + (1-t)\tilde{P}, tQ + (1-t)\tilde{Q}) \leq tD_f(P,Q) + (1-t)D_f(\tilde{P},\tilde{Q}).$$

Example 3.5.2.

1. Let $f(x) = x\log x$ so that $M_f = \infty$. Then,

$$D_f(P,Q) = \begin{cases} \int_{\mathcal{X}} \log\left(\frac{dP}{dQ}\right) dP & \text{if } P \ll Q \\ \infty & \text{else .} \end{cases}$$

   This is known as the KL–divergence.

2. Let $f(x) = x^2 - 1$ so that $M_f = \infty$. Then,

$$D_f(P,Q) = \begin{cases} \int_{\mathcal{X}} \left(\frac{dP}{dQ}\right)^2 dQ & \text{if } P \ll Q \\ \infty & \text{else .} \end{cases}$$

   This is known as the χ^2–divergence.

3. Let $f(x) = (x^{1/2} - 1)^2$ so that $M_f = 1$. Then,

$$D_f(P,Q) = \int_{\mathcal{X}} \left(\sqrt{\frac{dP_1}{dQ}} - 1\right)^2 dQ + P_2(\mathcal{X}) = \int_{\mathcal{X}} \left(\sqrt{\frac{dP}{d\nu}} - \sqrt{\frac{dQ}{d\nu}}\right)^2 d\nu,$$

   where ν is any measure that dominates both P and Q, for example $P+Q$. This is the square of the Hellinger distance.

4. Let $f(x) = |x-1|/2$ so that $M_f = 1/2$. Then,

$$D_f(P,Q) = \int_{\mathcal{X}} \frac{1}{2}\left|\frac{dP_1}{dQ} - 1\right| dQ + \frac{1}{2}P_2(\mathcal{X}) = \int_{\mathcal{X}} \frac{1}{2}\left|\frac{dP}{d\nu} - \frac{dQ}{d\nu}\right| d\nu.$$

   where ν is any measure that dominates both P and Q, for example $P+Q$. This is known as the total variation distance.

Theorem 3.5.1. (Data Processing Inequality)
Let $(\mathcal{X}, \mathscr{F})$ and $(\mathcal{Y}, \mathscr{G})$ be measurable spaces and let $\phi : \mathcal{X} \to \mathcal{Y}$ be $\mathscr{F}/\mathscr{G}$-measurable. Let P, Q be two probability measures on $(\mathcal{X}, \mathscr{F})$. Let $f : (0,\infty) \to \mathbb{R}$ be convex. Then,

$$D_f(P,Q) \geq D_f(P^{(\phi)}, Q^{(\phi)}).$$

Proof.

Write $P = P_1 + P_2$ where $P_1 \ll Q$ and $P_2 \perp Q$. Note that by ..., $P_1^{(\phi)} \ll Q^{(\phi)}$. We write $P_2^{(\phi)} = [P_2^{(\phi)}]_1 + [P_2^{(\phi)}]_2$ where $[P_2^{(\phi)}]_1 \ll Q^{(\phi)}$ and $[P_2^{(\phi)}]_2 \perp Q^{(\phi)}$. Thus, we have that $P_1^{(\phi)} + [P_2^{(\phi)}]_1$ is the part of $P^{(\phi)}$ that is absolutely continuous with respect to Q and that $[P_2^{(\phi)}]_2$ is the part that is singular.

Then, we have

$$\begin{aligned}
D_f(P,Q) &= \int_{\mathcal{X}} f \circ \frac{dP_1}{dQ}\, dQ + P_2(\mathcal{X})M_f \\
&= \int_{\mathcal{X}} E\left[f \circ \frac{dP_1}{dQ} \middle| \phi \right] dQ + P_2(\mathcal{X})M_f \\
&\geq \int_{\mathcal{X}} f \circ E\left[\frac{dP_1}{dQ} \middle| \phi \right] dQ + P_2(\mathcal{X})M_f \\
&= \int_{\mathcal{X}} f \circ \frac{dP_1^{(\phi)}}{dQ^{(\phi)}} \circ \phi \, dQ + P_2^{(\phi)}(\mathcal{Y})M_f \\
&= \int_{\mathcal{Y}} f \circ \frac{dP_1^{(\phi)}}{dQ^{(\phi)}}\, dQ^{(\phi)} + [P_2^{(\phi)}]_1(\mathcal{Y})M_f + [P_2^{(\phi)}]_2(\mathcal{Y})M_f \\
&= \int_{\mathcal{Y}} f \circ \frac{dP_1^{(\phi)}}{dQ^{(\phi)}} + M_f \frac{d[P_2^{(\phi)}]_1}{dQ^{(\phi)}}\, dQ^{(\phi)} + [P_2^{(\phi)}]_2(\mathcal{Y})M_f \\
&\geq \int_{\mathcal{Y}} f \circ \left(\frac{dP_1^{(\phi)}}{dQ^{(\phi)}} + \frac{d[P_2^{(\phi)}]_1}{dQ^{(\phi)}} \right) dQ^{(\phi)} + [P_2^{(\phi)}]_2(\mathcal{Y})M_f \\
&= \int_{\mathcal{Y}} f \circ \frac{dP^{(\phi)}}{dQ^{(\phi)}}\, dQ^{(\phi)} + [P_2^{(\phi)}]_2(\mathcal{Y})M_f = D_f(P^{(\phi)}, Q^{(\phi)}),
\end{aligned}$$

as desired. □

Theorem 3.5.2. (Fano's Inequality)
Let $\mathbb{P}, \mathbb{Q}$ be probability measures on $(\Omega, \mathscr{F})$. For any $A \in \mathscr{F}$, we have

$$\mathbb{P}(A) \leq \frac{\mathrm{KL}(\mathbb{P}, \mathbb{Q}) + \log 2}{-\log \mathbb{Q}(A)}.$$

Let $M \in \mathbb{N}$ and let $I : \Omega \to [M]$ be $\mathscr{F}/2^{[M]}$-measurable. For each $i \in [M]$, define $A_i = \{\omega : I(\omega) = i\} \in \mathscr{F}$ and assume that $\mathbb{P}(A_i) = \mathbb{Q}(A_i) = 1/M$. Define probability measure $\mathbb{P}_i$ on $(\Omega, \mathscr{F})$ by $A \mapsto \frac{\mathbb{P}(A \cap A_i)}{\mathbb{P}(A_i)}$ for all $A \in \mathscr{F}$. Define $\mathbb{Q}_i$ similarly.

Let $(\mathcal{X}, \mathscr{G})$ be a measure space and let $X : \Omega \to \mathcal{X}$ be $\mathscr{F}/\mathscr{G}$-measurable. Let $\phi : \mathcal{X} \to [M]$ be $\mathscr{G}/2^{[M]}$-measurable. We have that

$$\mathbb{P}(\phi(X) = I) \leq \frac{\frac{1}{M}\sum_{i=1}^M \mathrm{KL}(\mathbb{P}_i^{(X)}, \mathbb{Q}_i^{(X)}) + \log 2}{-\log \mathbb{Q}(\phi(X) = I)}.$$

In particular, if $\mathbb{Q}_i^{(X)} = \mathbb{Q}^{(X)}$ for all i, then

$$\mathbb{P}(\phi(X) = I) \leq \frac{\frac{1}{M}\sum_{i=1}^M \mathrm{KL}(\mathbb{P}_i^{(X)}, \mathbb{Q}^{(X)}) + \log 2}{\log M}.$$

Proof.

First, we observe that for any $p, q \in [0,1]$,

$$p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} \geq p \log \frac{1}{q} + p \log p + (1-p)\log(1-p) \geq p \log \frac{1}{q} - \log 2.$$

Then, we have that

$$\begin{aligned}&\mathbb{P}(A)\log\frac{1}{\mathbb{Q}(A)}-\log 2\\&\leq \mathrm{KL}(\mathrm{Ber}(\mathbb{P}(A)),\mathrm{Ber}(\mathbb{Q}(A)))\leq \mathrm{KL}(\mathbb{P},\mathbb{Q}).\end{aligned}$$

where we applied the data processing inequality with the transformation $\mathbb{1}_A$.

For the second claim, we note that

$$\mathbb{P}(\phi(X)=I)=\sum_{i=1}^{M}\mathbb{P}(\{\phi(X)=i\}\cap A_i)=\frac{1}{M}\sum_{i=1}^{M}\mathbb{P}_i(\phi(X)=i).$$

Thus, we have

$$\begin{aligned}\mathbb{P}(\phi(X)=I)\log\frac{1}{\mathbb{Q}(\phi(X)=I)}-\log 2&\leq \mathrm{KL}\big(\mathrm{Ber}(\mathbb{P}(\phi(X)=I)),\mathrm{Ber}(\mathbb{Q}(\phi(X)=I))\big)\\&\leq \frac{1}{M}\sum_{i=1}^{M}\mathrm{KL}\big(\mathrm{Ber}(\mathbb{P}_i(\phi(X)=i)),\mathrm{Ber}(\mathbb{Q}_i(\phi(X)=i))\big)\\&=\frac{1}{M}\sum_{i=1}^{M}\mathrm{KL}\big(\mathrm{Ber}(\mathbb{P}_i^{(X)}(\phi=i)),\mathrm{Ber}(\mathbb{Q}_i^{(X)}(\phi=i))\big)\\&\leq \frac{1}{M}\sum_{i=1}^{M}\mathrm{KL}(\mathbb{P}_i^{(X)},\mathbb{Q}_i^{(X)}),\end{aligned}$$

where we applied the data processing inequality with $\mathbb{1}_{\phi=i}$.

For the last claim, we note that

$$\mathbb{Q}(\phi(X)=I)=\frac{1}{M}\sum_{i=1}^{M}\mathbb{Q}_i(\phi(X)=i)=\frac{1}{M}\sum_{i=1}^{M}\mathbb{Q}_i^{(X)}(\phi=i)=\frac{1}{M}.$$

□

Chapter 4

Measures on Product Spaces

4.1 Product σ-field

Definition 4.1.1. Let $(\Omega_1, \mathscr{F}_1)$ and $(\Omega_2, \mathscr{F}_2)$ be measurable spaces. If $A_1 \in \mathscr{F}_1$, $A_2 \in \mathscr{F}_2$, we call $A_1 \times A_2 := \{(\omega_1, \omega_2) : \omega_1 \in A_1, \omega_2 \in A_2\}$ a measurable rectangles. We call $\mathscr{F} := \sigma(\{A_1 \times A_2 : A_1 \in \mathscr{F}, A_2 \in \mathscr{F}\})$ the product σ-field (σ-field generated by measurable rectangles), denoted by $\mathscr{F} = \mathscr{F}_1 \otimes \mathscr{F}_2$.

If we have $(\Omega_1, \mathscr{F}_1), \ldots, (\Omega_n, \mathscr{F}_n)$, then we define $\mathscr{F} := \mathscr{F}_1 \otimes \cdots \otimes \mathscr{F}_n = \sigma(\{A_1 \times \cdots \times A_n : A_i \in \mathscr{F}_i, \forall i \in [n]\})$.

A common notation for product σ-field is $\mathscr{F}_1 \times \mathscr{F}_2$. This can be misleading.

Lemma 4.1.1. The set of measurable rectangles $\mathcal{A} := \{A_1 \times A_2 : A_1 \in \mathscr{F}_1, A_2 \in \mathscr{F}_2\}$ is a semi-ring. ($\emptyset \in \mathcal{A}$; $\Omega \in \mathcal{A}$; $A \in \mathcal{A}, B \in \mathcal{A} \implies A \cap B \in \mathcal{A}$; $A \setminus B = \cup_{k=1}^K C_k \in \mathcal{A}$ disjoint). Likewise for products of n σ-fields. As a consequence, $\{\cup_{n=1}^N R_n : N \in \mathbb{N}, R_1, \ldots, R_n \in \mathcal{A} \text{ disjoint}\}$ is a field.

Proof.

Let $A_1 \times A_2$, $\widetilde{A}_1 \times \widetilde{A}_2 \in \mathcal{A}$, then $(A_1 \times A_2) \cap (\widetilde{A}_1 \times \widetilde{A}_2) = (A_1 \cap \widetilde{A}_1) \times (A_2 \cap \widetilde{A}_2) \in \mathcal{A}$. Also,

$$(A_1 \times A_2) \setminus (\widetilde{A}_1 \times \widetilde{A}_2) = \{(A_1 \cap \widetilde{A}_1^c) \times (A_2 \cap \widetilde{A}_2^c)\} \cup \{(A_1 \cap \widetilde{A}_1) \times (A_2 \cap \widetilde{A}_2^c)\} \cup \{(A_1 \cap \widetilde{A}_1^c) \times (A_2 \cap \widetilde{A}_2)\}.$$

It is clear that $\emptyset = \emptyset \times \emptyset \in \mathcal{A}$ and $\Omega_1 \times \Omega_2 \in \mathcal{A}$. The n-product case follows similarly or through induction, by first proving that $\mathscr{F}_1 \otimes \mathscr{F}_2 \otimes \mathscr{F}_3 = \mathscr{F}_1 \otimes (\mathscr{F}_2 \otimes \mathscr{F}_3)$. $\square$

Lemma 4.1.2. Define $\pi_1 : \Omega_1 \times \Omega_2 \to \Omega_1$ as, $\forall (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$, $\pi_1(\omega_1, \omega_2) = \omega_1$ and $\pi_2 : \Omega_1 \times \Omega_2 \to \Omega_2$ similarly. Define

$$\sigma(\pi_1, \pi_2) := \bigcap \left\{\mathscr{G} : \mathscr{G} \ \sigma\text{-field }, \forall A_1 \in \mathscr{F}_1, \pi_1^{-1}(A_1) \in \mathscr{G}, \forall A_2 \in \mathscr{F}_2, \pi_2^{-1}(A_2) \in \mathscr{G}\right\}$$

as the σ-field generated by π_1, π_2. Then, we have that

$$\sigma(\pi_1, \pi_2) = \mathscr{F}_1 \otimes \mathscr{F}_2.$$

Proof.

For any $A_1 \in \mathscr{F}_1$ and $A_2 \in \mathscr{F}_2$, we have that

$$\pi_1^{-1}(A_1) = A_1 \times \Omega_2 \in \mathscr{F}_1 \otimes \mathscr{F}_2$$
$$\pi_2^{-1}(A_1) = \Omega_1 \times A_2 \in \mathscr{F}_1 \otimes \mathscr{F}_2.$$

Hence, π_1 is $\mathscr{F}_1/\mathscr{F}_1 \otimes \mathscr{F}_2$-measurable and π_2 is $\mathscr{F}_2/\mathscr{F}_1 \otimes \mathscr{F}_2$-measurable and we have that $\sigma(\pi_1, \pi_2) \subset \mathscr{F}_1 \otimes \mathscr{F}_2$.

On the other hand, for any $A_1 \in \mathscr{F}_1$ and $A_2 \in \mathscr{F}_2$, we have that

$$\pi_1^{-1}(A_1) \cap \pi_2^{-1}(A_2) = A_1 \times A_2 \in \sigma(\pi_1, \pi_2).$$

Thus, $\mathscr{F}_1 \otimes \mathscr{F}_2 \subset \sigma(\pi_1, \pi_2)$. □

Lemma 4.1.3. Let $C \in \mathscr{F}_1 \otimes \mathscr{F}_2$ and let $\omega_1 \in \Omega_1$. Define the section $C(\omega_1) := \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in C\} \subset \Omega_2$. Then $C(\omega_1) \in \mathscr{F}_2$.

Proof.
Fix $\omega_1 \in \mathscr{F}_1$ and define $\mathscr{G} = \{C \in \mathscr{F}_1 \otimes \mathscr{F}_2 : C(\omega_1) \in \mathscr{F}_2\}$. Note that $\emptyset$, $\Omega_1 \times \Omega_2 \in \mathscr{G}$. Let $C_1, C_2 \dots \in \mathscr{G}$. Then,

$$\begin{aligned}(\cup_{n=1}^{\infty} C_n)(\omega_1) &:= \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in \cup_{n=1}^{\infty} C_n\} \\ &= \cup_{n=1}^{\infty} \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in C_n\} \\ &= \cup_{n=1}^{\infty} C_n(\omega_1) \in \mathscr{F}_2 \implies \cup_{n=1}^{\infty} C_n \in \mathcal{C}.\end{aligned}$$

Finally,

$$\begin{aligned}(C_1 \setminus C_2)(\omega_1) &= \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in C_1 \text{ and } (\omega_1, \omega_2) \notin C_2\} \\ &= C_1(\omega_1) \setminus C_2(\omega_1) \in \mathscr{F}_2 \implies C_1 \setminus C_2 \in \mathcal{C}.\end{aligned}$$

Thus, $\mathcal{C}$ is a σ-field. Since all sets of the form $A \times B \in \mathcal{C}$ for $A \in \mathscr{F}_1$ and $B \in \mathscr{F}_2$, we have that $\mathcal{C} = \mathscr{F}_1 \otimes \mathscr{F}_2$ as desired. □

4.2 Markov Kernels

Definition 4.2.1. Let $(\Omega_1, \mathscr{F}_1)$ and $(\Omega_2, \mathscr{F}_2)$ be measurable spaces. We say $\mathscr{K} : \Omega_1 \times \mathscr{F}_2 \to [0, 1]$ is a <u>Markov kernel</u> if

(a) $\forall \omega_1 \in \Omega_1$, $\mathscr{K}(\omega_1, \bullet) : \mathscr{F}_2 \to [0, 1]$ is probability a measure.

(b) $\forall B \in \mathscr{F}_2$, $\mathscr{K}(\bullet, B) : \Omega_1 \to [0, 1]$ is $\mathscr{F}_1/\mathscr{B}([0, 1])$-measurable.

We say that $\mathscr{K} : \Omega_1 \times \mathscr{F}_2 \to [0, \infty]$ is a σ-finite kernel if there exists disjoint $K_1, K_2, \ldots \in \mathscr{F}_2$ such that

(a) $\forall \omega_1 \in \Omega_1$, $\mathscr{K}(\omega_1, \bullet \cap K_n) : \mathscr{F}_2 \to [0, 1]$ is a probability measure for all $n \in \mathbb{N}$.

(b) $\forall B \in \mathscr{F}_2$, $\mathscr{K}(\bullet, B) : \Omega_1 \to [0, \infty]$ is $\mathscr{F}_1/\mathscr{B}([0, \infty])$-measurable.

Remark 4.2.1. As a trivial example, any probability measure μ_2 on $(\Omega_2, \mathscr{F}_2)$ induces a Markov kernel through $\mathscr{K}(\omega_1, B) := \mu_2(B)$, $\forall \omega_1 \in \Omega_1$, $B \in \mathscr{F}_2$. Markov kernels are also called regular conditional probability.

Theorem 4.2.1 (Fubini-Tonelli)**.** Let $(\Omega_1, \mathscr{F}_1, \mu_1)$ be a measure space with μ_1 being a σ-finite measure, let $(\Omega_2, \mathscr{F}_2)$ be a measurable space, and let $\mathscr{K} : \Omega_1 \times \mathscr{F}_2 \to [0, \infty]$ be a σ-finite kernel, then $\exists$ a unique σ-finite measure $\mu : \mathscr{F}_1 \otimes \mathscr{F}_2 \to [0, \infty]$ such that $\forall A \in \mathscr{F}_1$, $B \in \mathscr{F}_2$,

$$\mu(A \times B) = \int_A \mathscr{K}(\omega_1, B) \mathrm{d}\mu_1(\omega_1). \tag{4.1}$$

Moreover, for any $f : \Omega_1 \times \Omega_2 \to [-\infty, \infty]$, Borel-measurable, we have

$$\int_{\Omega_1 \times \Omega_2} f \mathrm{d}\mu = \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2) \mathrm{d}\mathscr{K}(\omega_1, \bullet)(\omega_2) \mathrm{d}\mu_1(\omega_1). \tag{4.2}$$

Proof.

Note that (4.1) is well defined since $\mathscr{K}(\bullet, B) : \Omega_1 \to [0,\infty]$ is Borel-measurable. If we can show that (4.1) extends to a σ-finite measure on $(\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2)$, then uniqueness follows then from Lemma 4.1.1 and Caratheodory extension theorem.

We first assume μ_1 is a probability measure and that $\mathscr{K}$ is a Markov kernel.

For $C \in \mathscr{F}_1 \otimes \mathscr{F}_2$, we write

$$\tilde{\mu}(C) := \int_{\Omega_1} \mathscr{K}(\omega_1, C(\omega_1)) \mathrm{d}\mu_1(\omega_1),$$

and claim that $\tilde{\mu}$ is a well-defined probability measure. Note that $C(\omega_1) \in \mathscr{F}_2$ by Lemma 4.1.3 so that $\mathscr{K}(\omega_1, C(\omega_1))$ is well-defined.

Step A: We claim that $\omega_1 \mapsto \mathscr{K}(\omega_1, C(\omega_1))$ as a function $\Omega_1 \to [0,1]$ is Borel-measurable so that $\tilde{\mu}$ is well-defined. To see this in general, define

$$\mathcal{C} := \{C \in \mathscr{F}_1 \otimes \mathscr{F}_2 : \omega_1 \mapsto \mathscr{K}(\omega_1, C(\omega_1)) \text{ is Borel-measurable}\}$$

Note that $\mathcal{A} \subseteq \mathcal{C}$ ($\mathcal{A}$ is the set of measurable rectangles) since $\forall A \in \mathscr{F}_1, B \in \mathscr{F}_2$, we have $\omega_1 \mapsto \mathscr{K}(\omega_1, (A \times B)(\omega_1)) = \mathscr{K}(\omega_1, B)\mathbb{1}_A(\omega_1)$, which is Borel-measurable.

If $C_1, C_2, \dots, C_n \in \mathcal{C}$ are disjoint,

$$\omega_1 \mapsto \mathscr{K}(\omega_1, (\cup_{i=1}^n C_i)(\omega_1)) = \mathscr{K}(\omega_1, \cup_{i=1}^n C_i(\omega_1)) = \sum_{i=1}^n \underbrace{\mathscr{K}(\omega_1, C_i(\omega_1))}_{\substack{\forall i \in [n], \omega_1 \mapsto \mu(\omega_1, C_i(\omega_1)) \\ \text{is Borel Meas.}}}$$

is Borel-measurable. Thus, $\cup_{i=1}^n C_i \in \mathcal{C}$. Likewise, if $C \in \mathcal{C}$, then for any $\omega_1 \in \Omega_1$, $(C^c)(\omega_1) = \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \notin C\} = C(\omega_1)^c$. Thus,

$$\omega_1 \mapsto \mathscr{K}(\omega_1, (C^c)(\omega_1)) = \mathscr{K}(\omega_1, C(\omega_1)^c) = 1 - \mathscr{K}(\omega_1, C(\omega_1))$$

is Borel-measurable. So $C_1^c \in \mathcal{C} \implies \mathcal{C}$ contains field generated by $\mathcal{A}$. Now let $C_1 \subseteq C_2 \subseteq \dots \in \mathcal{C}$, then

$$\omega_1 \mapsto \mathscr{K}(\omega_1, (\cup_{n=1}^\infty C_n)(\omega_1)) = \mathscr{K}(\omega_1, \cup_{n=1}^\infty C_n(\omega_1)) = \lim_{n\to\infty} \mathscr{K}(\omega_1, C_n(\omega_1)) \quad (\text{since } C_1(\omega_1) \subseteq C_2(\omega_2) \subseteq \dots)$$

is Borel-measurable by lemma 2.1.2. So $\cup_{n=1}^\infty C_n \in \mathcal{C}$. Likewise, if $C_1 \supseteq C_2 \supseteq \dots \in \mathcal{C}$, we have $\cap_{n=1}^\infty C_n \in \mathcal{C}$.

Thus, $\mathcal{C}$ is a monotone class and by monotone class theorem (Theorem 1.3.1), $C = \mathscr{F}_1 \otimes \mathscr{F}_2$. Thus, $\omega_1 \mapsto \kappa(\omega_1, C(\omega_1))$ is Borel-measurable.

Step B: We prove that $\tilde{\mu}$ is a probability measure. Let $C_1, C_2 \dots \in \mathscr{F}_1 \otimes \mathscr{F}_2$ be disjoint, then

$$\begin{aligned}
\tilde{\mu}(\cup_{n=1}^\infty C_n) &= \int_{\Omega_1} \mathscr{K}(\omega_1, (\cup_{n=1}^\infty C_n)(\omega_1)) \mathrm{d}\mu_1(\omega_1) \\
&= \int_{\Omega_1} \mathscr{K}(\omega_1, \cup_{n=1}^\infty C_n(\omega_1)) \mathrm{d}\mu_1(\omega_1) \\
&= \int_{\Omega_1} \sum_{n=1}^\infty \mathscr{K}(\omega_1, C_n(\omega_1)) \mathrm{d}\mu_1(\omega_1) && (\text{note that } C_1(\omega_1), C_2(\omega_1), \dots \text{ are disjoint}) \\
&= \sum_{n=1}^\infty \int_{\Omega_1} \mathscr{K}(\omega_1, C_n(\omega_1)) \mathrm{d}\mu_1(\omega_1) && (\text{by MCT}) \\
&= \sum_{n=1}^\infty \tilde{\mu}(C_n).
\end{aligned}$$

Since $\tilde{\mu}(\emptyset) = 0$ and $\tilde{\mu}(\Omega_1 \times \Omega_2) = 1$, $\tilde{\mu}$ is a probability measure. Since $\tilde{\mu}$ agrees with (4.1) on measurable rectangles, we see that $\tilde{\mu}$ is the unique extension of μ.

Now let μ_1 be a σ-finite measure on $(\Omega_1, \mathscr{F}_1)$ and let $\mathcal{K}$ be a σ-finite kernel. Let $A_1, A_2, \ldots \in \mathscr{F}_1$ be a partition of Ω_1 such that $\mu_1^{(n)}(\bullet) := \mu_1(\bullet \cap A_n)$ is a probability measure for all $n \in \mathbb{N}$ and let $K_1, K_2, \ldots \in \mathscr{F}_2$ be a partition of Ω_2 such that for all $\omega_1 \in \Omega_1$, $\mathcal{K}(\omega_1, \bullet \cap K_n)$ is a probability measure.

Then, for any pair (m, n), there exists a probability measure $\mu^{(m,n)}$ on $\Omega_1 \times \Omega_2$ such that $\mu^{(m,n)}(A_m \times K_n) = 1$, such that, for any $A \in \mathscr{F}_1$ and $B \in \mathscr{F}_2$, $\mu^{(m,n)}(A \times B) = \int_A \mathcal{K}(\omega_1, B \cap K_n) d\mu_1^{(m)}(\omega_1)$. We define $\mu = \sum_{m,n} \mu^{(m,n)}$ and observe that for any $A \in \mathscr{F}_1, B \in \mathscr{F}_2$,

$$\begin{aligned}
\mu(A \times B) &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \int_A \mathcal{K}(\omega_1, B \cap K_n) d\mu_1^{(m)}(\omega_1) \\
&= \sum_m \int_A \mathcal{K}(\omega_1, B) \underbrace{\frac{d\mu_1^{(m)}(\omega_1)}{d\mu_1(\omega_1)}}_{\mathbb{1}_{A_m}(\omega_1)} d\mu_1(\omega_1) \\
&= \int_A \mathcal{K}(\omega_1, B) d\mu_1(\omega_1),
\end{aligned}$$

as desired.

Now we prove (4.2). Let $s = \sum_{i=1}^n a_i \mathbb{1}_{C_i}$ be a simple function with $a_1, a_2, \ldots, a_n > 0$ and $C_1, C_2 \ldots, C_n \in \mathscr{F}_1 \otimes \mathscr{F}_2$ disjoint. Then

$$\begin{aligned}
\int_{\Omega_1 \times \Omega_2} s \mathrm{d}\mu = \sum_{i=1}^n a_i \mu(C_i) &= \sum_{i=1}^n a_i \int_{\Omega_1} \mathcal{K}(\omega_1, C_i(\omega_1)) \mathrm{d}\mu_1(\omega_1) \\
&= \sum_{i=1}^n a_i \int_{\Omega_1} \int_{\Omega_2} \mathbb{1}_{C_i(\omega_1)}(\omega_2) \mathrm{d}\mathcal{K}(\omega_1, \bullet)(\omega_2) \mathrm{d}\mu_1(\omega_1) \\
&= \int_{\Omega_1} \int_{\Omega_2} \sum_{i=1}^n a_i \mathbb{1}_{C_i(\omega)}(\omega_2) \mathrm{d}\mathcal{K}(\omega_1, \bullet)(\omega_2) \mathrm{d}\mu_1(\omega_1) \\
&= \int_{\Omega_1} \int_{\Omega_2} s(\omega_1, \omega_2) \mathrm{d}\mathcal{K}(\omega_1, \bullet)(\omega_2) \mathrm{d}\mu_1(\omega_1).
\end{aligned}$$

Now let $f : \Omega_1 \times \Omega_2 \to [0, \infty]$ be Borel-measurable, then $\exists\ 0 \leq s_1 \leq s_2 \leq \ldots$ simple functions such that $\forall (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$, $s_n(\omega_1, \omega_2) \to f(\omega_1, \omega_2)$. Then

$$\begin{aligned}
\int_{\Omega_1 \times \Omega_2} f \mathrm{d}\mu &= \lim_{n \to \infty} \int_{\Omega_1 \times \Omega_2} s_n \mathrm{d}\mu \\
&= \lim_{n \to \infty} \int_{\Omega_1} \int_{\Omega_2} s_n(\omega_1, \omega_2) \mathrm{d}\mathcal{K}(\omega_1, \bullet)(\omega_2) \mathrm{d}\mu_1(\omega_1) \\
&= \int_{\Omega_1} \int_{\Omega_2} \left(\lim_{n \to \infty} s_n(\omega_1, \omega_2) \right) \mathrm{d}\mathcal{K}(\omega_1, \bullet)(\omega_2) \mathrm{d}\mu_1(\omega_1) \qquad \text{(by 2 applications of MCT)} \\
&= \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2) \mathrm{d}\mathcal{K}(\omega_1, \bullet)(\omega_2) \mathrm{d}\mu_1(\omega_1).
\end{aligned}$$

Finally, suppose $f : \Omega_1 \times \Omega_2 \to [-\infty, \infty]$ is Borel-measurable. Let $f = f^+ - f^-$.

$$\begin{aligned}
\int_{\Omega_1 \times \Omega_2} f \mathrm{d}\mu &= \int_{\Omega_1 \times \Omega_2} f^+ \mathrm{d}\mu - \int_{\Omega_1 \times \Omega_2} f^- \mathrm{d}\mu \qquad \left(\text{Assume } \int_{\Omega_1 \times \Omega_2} f^- \mathrm{d}\mu < \infty \text{ say}\right) \\
&= \int_{\Omega_1} \int_{\Omega_2} f^+(\omega_1, \omega_2) \mathrm{d}\mathcal{K}(\omega_1, \bullet)(\omega_2) \mathrm{d}\mu_1(\omega_1) - \int_{\Omega_1} \underbrace{\int_{\Omega_2} f^-(\omega_1, \omega_2) \mathrm{d}\mathcal{K}(\omega_1, \bullet)(\omega_2)}_{\substack{\omega_1 \mapsto \int_{\Omega_2} f^-(\omega_1, \omega_2) \mathrm{d}\mathcal{K}(\omega_1, \bullet)(\omega_2) \\ \text{must be finite } \mu_1\text{-a.e.}}} \mathrm{d}\mu_1(\omega_1) \\
&= \int_{\Omega_1} \int_{\Omega_2} f^+(\omega_1, \omega_2) - f^-(\omega_1, \omega_2) \mathrm{d}\mathcal{K}(\omega_1, \bullet)(\omega_2) \mathrm{d}\mu_1(\omega_1).
\end{aligned}$$

$\square$

Remark 4.2.2. If $\forall \omega_1 \in \Omega_1$, $\mathscr{K}(\omega_1, \bullet) = \mu_2(\bullet)$ for some probability measures μ_2 on $(\Omega_2, \mathscr{F}_2)$, then we call μ the product measure on $(\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2)$.

The classic Fubini Theorem is a special case of Theorem 4.2.1: Let μ_1, μ_2 be probability (can be extended to σ-finite) measures on $(\Omega_1, \mathscr{F}_1)$ and $(\Omega_2, \mathscr{F}_2)$. Then, for any $f : \Omega_1 \times \Omega_2 \to [-\infty, \infty]$ Borel-measurable,

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} f \mathrm{d}\mu &= \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2) \mathrm{d}\mu_2(\omega_2) \mathrm{d}\mu_1(\omega_1) \\ &= \int_{\Omega_2} \int_{\Omega_1} f(\omega_1, \omega_2) \mathrm{d}\mu_1(\omega_1) \mathrm{d}\mu_2(\omega_2).\end{aligned}$$

Example 4.2.1. Fubini's Theorem, in addition to being fundamental in defining probabilistic dependence, also has applications in simplifying calculations.

In this example, we show that if $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space and $X : \Omega \to [0, \infty]$ is a random variable, then $\int X d\mathbb{P} = \int_0^\infty \mathbb{P}(X \geq t)\, dt$. Observe that

$$\begin{aligned}\int X d\mathbb{P} &= \int_0^\infty x \, d\mathbb{P}^{(X)}(x) \\ &= \int_0^\infty \left(\int_0^\infty \mathbb{1}(t \leq x)\, dt \right) d\mathbb{P}^{(X)}(x) \\ &= \int_0^\infty \left(\int_t^\infty d\mathbb{P}^{(X)}(x) \right) dt \\ &= \int_0^\infty \mathbb{P}(X \geq t)\, dt.\end{aligned}$$

Example 4.2.2. We will use Fubini's theorem to derive the following bound: $\mathrm{TV}(N(0, I_p), N(\mu, I_p)) \leq 1 \vee \frac{1}{\sqrt{2\pi}} \|\mu\|_2$. Let $\lambda_p(\bullet)$ denote Lebesgue-measure.

First we show the following: if $\mathbb{P}, \mathbb{Q}$ are probability measure on $\mathbb{R}^p$ with densities p, q, then

$$\begin{aligned}\mathrm{TV}(\mathbb{P}, \mathbb{Q}) &:= \sup_{A \in \mathscr{B}(\mathbb{R}^p)} |\mathbb{P}(A) - \mathbb{Q}(A)| \\ &= \sup_{A \in \mathscr{B}(\mathbb{R})} \left| \int_A p - q \mathrm{d}\lambda_p \right| \\ &= \max \left\{ \int_{\{p > q\}} p - q \mathrm{d}\lambda_p, \int_{\{q > o\}} q - p \mathrm{d}\lambda_p \right\}.\end{aligned}$$

But,

$$\int_{\{p>q\}} p - q \mathrm{d}\lambda_p = \int_{\{p>q\}} p - q \mathrm{d}\lambda_p - \underbrace{\int p - q \mathrm{d}\lambda_p}_{=0} = \int_{\{q>p\}} q - p \mathrm{d}\lambda_p.$$

So $\mathrm{TV}(\mathbb{P},\mathbb{Q}) = \frac{1}{2}\int |p-q| \mathrm{d}\lambda_p$.

$$\begin{aligned}
\mathrm{TV}(N(0,I_p),N(\mu,I_p)) &= \frac{1}{2}\int_{\mathbb{R}^p} |p(x:0,I_p) - p(x:\mu,I_p)|\,\mathrm{d}x \\
&= \frac{1}{2}(2\pi)^{-\frac{p}{2}}\int_{\mathbb{R}^p}\left|e^{-\frac{1}{2}\|x\|^2} - e^{-\frac{1}{2}\|x-\mu\|_2^2}\right|\mathrm{d}x \\
&= \frac{1}{2}(2\pi)^{-\frac{p}{2}}\int_{\mathbb{R}^p}\int_0^\infty e^{-t}\mathbb{1}_{\{t\in[\frac{1}{2}\|x\|_2^2,\frac{1}{2}\|x-\mu\|_2^2]\}}\mathrm{d}t\mathrm{d}x \\
&= \frac{1}{2}(2\pi)^{-\frac{p}{2}}\int_0^\infty e^{-t}\int_{\mathbb{R}^p}\mathbb{1}_{\{t\in[\frac{1}{2}\|x\|_2^2,\frac{1}{2}\|x-\mu\|_2^2]\}}\mathrm{d}x\mathrm{d}t && \text{(Fubini)} \\
&= \frac{(2\pi)^{-\frac{\beta}{2}}}{2}\int_0^\infty e^{-t}\underbrace{\lambda_p\left\{B(0,\sqrt{2t})\Delta B(\mu,\sqrt{2t})\right\}}_{B(0,\sqrt{2t})\cup B(\mu,\sqrt{2t})\setminus\{B(0,\sqrt{2t})\cap B(\mu,\sqrt{2t})\}}\mathrm{d}t \\
&\leq \frac{(2\pi)^{-\frac{\beta}{2}}}{2}\int_0^\infty e^{-t}2\|\mu\|_2\,\lambda_{p-1}\left\{B_{p-1}(0,\sqrt{2t})\right\}\mathrm{d}t \\
&= (2\pi)^{-\frac{\beta}{2}}\|\mu\|_2\,\lambda_{p-1}(B_{p-1}(0,1))\int_0^\infty e^{-t}(2t)^{\frac{p-1}{2}}\mathrm{d}t \\
&= (2\pi)^{-\frac{p}{2}}\frac{\pi^{\frac{p-1}{2}}}{\Gamma(\frac{p-1}{2}+1)}2^{\frac{p-1}{2}}\Gamma(\frac{p+1}{2})\|\mu\|_2 \\
&= \sqrt{\frac{2}{\pi}}\|\mu\|_2\,.
\end{aligned}$$

Theorem 4.2.2. Let $(\Omega_1,\mathscr{F}_1),\ldots,(\Omega_n,\mathscr{F}_n)$ be measurable spaces and suppose, for every $j\in[n-1]$, $\{(\omega_1,\ldots,\omega_j),B_{j+1}\}\mapsto P(\omega_1,\ldots,\omega_j,B_{j+1}):(\Omega_1\times\cdots\times\Omega_j)\times\mathscr{F}_{j+1}\to[0,1]$ is Markov kernel w.r.t. $\mathscr{F}_1\otimes\mathscr{F}_2\otimes\cdots\otimes\mathscr{F}_j$. Let μ_1 be a probability measure on $(\Omega_1,\mathscr{F}_1)$. Then $\exists$ unique probability measure P on $(\Omega_1\times\cdots\times\Omega_n,\mathscr{F}_1\otimes\cdots\otimes\mathscr{F}_n)$ such that $\forall A_1\in\mathscr{F}_1,\ldots,A_n\in\mathscr{F}_n$,

$$P(A_1\times\cdots\times A_n) = \int_{A_1}\cdots\int_{A_{n-1}}\left(\int_{A_n}\mathrm{d}P(\omega_1,\ldots,\omega_{n-1},\bullet)\right)\mathrm{d}P(\omega_1,\ldots,\omega_{n-2},\bullet)(\omega_{n-1})\ldots\mathrm{d}\mu_1(\omega_1).$$

Moreover, for any Borel-measurable $f:\Omega_1\times\cdots\times\Omega_n\to[-\infty,\infty]$,

$$\int_{\Omega_1\times\ldots\Omega_n} f\mathrm{d}\mu = \int_{\Omega_1}\cdots\int_{\Omega_n} f(\omega_1,\ldots,\omega_n)\mathrm{d}P(\omega_1,\ldots,\omega_{n-1},\bullet)(\omega_n)\mathrm{d}P(\omega_1,\ldots,\omega_{n-2},\bullet)(\omega_{n-1})\ldots\mathrm{d}\mu_1(\omega_1).$$

Proof.

We use induction. The $n=1$ case is trivial. Assume then what $\exists$ unique probability measure $\widetilde{P}$ on $(\Omega_1\times\cdots\times\Omega_{n-1},\mathscr{F}_1\otimes\cdots\otimes\mathscr{F}_{n-1})$. Then we may use Theorem 4.2.1 to obtain a unique a probability measure P on $(\Omega_1\times\cdots\times\Omega_{n-1})\times\Omega_n = \Omega_1\times\cdots\times\Omega_n$ and $(\mathscr{F}_1\otimes\cdots\otimes\mathscr{F}_{n-1})\otimes\mathscr{F}_n = \mathscr{F}_1\otimes\cdots\otimes\mathscr{F}_n$ satisfying the desired properties. □

4.3 Countable Product Space

Definition 4.3.1. Let $(\Omega_n,\mathscr{F}_n)$ for $n\in\mathbb{N}$ be measurable space. For any subset $S\subseteq\mathbb{N}$, we define project $\pi_S:\prod_{n\in\mathbb{N}}\Omega_n\to\prod_{k\in S}\Omega_k$ by

$$\pi_S(\omega_1,\omega_2,\ldots) = \omega_S$$

where $\omega_S = (\omega_k)_{k\in S}$ is the subsequence of $(\omega_1,\omega_2,\ldots)$ indexed by S.

We say that $C\subseteq\prod_{n=1}^\infty\Omega_n$ is a measurable cylinder if there exists a finite $S\subseteq\mathbb{N}$ and $B_S\in\otimes_{n\in S}\mathscr{F}_n$ such that $C=\pi_S^{-1}(B_S) = \{(\omega_1,\omega_2,\ldots)\in\prod_{n=1}^\infty\Omega_n : \omega_S\in B_s\}$. We call (S,B_S) the base of C.

For $S \subseteq \mathbb{N}$, $B_S \subseteq \otimes_{n\in S}\mathscr{F}_n$, we also write $C(S, B_S) \subseteq \prod_{n=1}^{\infty}\Omega_n$ as the measurable cylinder generated by base (S, B_S). If $B_S = \prod_{n\in S} A_n$ for $A_n \in \mathscr{F}_n$, then we call C a rectangular cylinder. Note that if C is a rectangular cylinder, then

$$C = \pi_S^{-1}\left(\prod_{k\in S} A_k\right) = \bigcap_{k\in S}\pi_k^{-1}(A_k),$$

for some finite $S \subset \mathbb{N}$ and $A_k \in \mathscr{F}_k$ for all $k \in S$.

Let $\mathcal{A} := \{\pi_S^{-1}(\prod_{k\in S} A_k) : S \subseteq \mathbb{N} \text{ finite}, A_k \in \mathscr{F}_k, k \in S\}$. Then we define $\otimes_{n=1}^{\infty}\mathscr{F}_n := \sigma(\mathcal{A})$.

Remark 4.3.1. Let $S_1 \subseteq S_2 \subseteq \mathbb{N}$ be finite. For any $B_{S_1} \in \otimes_{n\in S_1}\mathscr{F}_n$, define $\tau_{S_2}(B_{S_1}) := \{\tilde{\omega} \in \prod_{n\in S_2}\Omega_n : \tilde{\omega}_{S_1} \in B_{S_1}\} \in \otimes_{n\in S_2}\mathscr{F}_n$ as the embedding of $B_{S_1} \subseteq \prod_{n\in S_1}\Omega_n$ into $\prod_{n\in S_2}\Omega_n$.

For example, if $S_1 = \{2\}$, $S_2 = \{1,2,3\}$, then $\tau_{S_2}(B_{S_1}) = \Omega_1 \times B_{S_1} \times \Omega_3$. We note then that $\pi_{S_1}^{-1}(B_{S_1}) = \pi_{S_2}^{-1}(\tau_{S_2}(B_{S_1}))$.

Lemma 4.3.1. Define $\mathscr{B} := \{C(S, B_S) : S \subseteq \mathbb{N} \text{ finite}, B_S \in \otimes_{n\in S}\mathscr{F}_n\}$. We have $\mathcal{A}$ is a semi-ring and $\mathscr{B}$ is a field (not necessarily generated by $\mathcal{A}$). Moreover,

$$\otimes_{k\in\mathbb{N}}\mathscr{F}_k = \sigma(\mathcal{A}) = \sigma(\mathscr{B}) = \sigma\big(\{\pi_S\}_{\text{ finite } S\subset\mathbb{N}}\big) = \sigma(\{\pi_k\}_{k\in\mathbb{N}}).$$

Proof.

We first show $\mathcal{A}$ is a semi-ring. Note that $\prod_{n=1}^{\infty}\Omega_n = C(\{1\}, \Omega_1) \in \mathcal{A}$ and that $\emptyset = \emptyset \times \prod_{n=2}^{\infty}\Omega_n \in \mathcal{A}$. Let $\pi_{S_1}^{-1}(B_{S_1}), \pi_{S_2}^{-1}(B_{S_2}) \in \mathcal{A}$ for $S_1, S_2 \subseteq \mathbb{N}$ finite and where $B_{S_1} \in \otimes_{n\in S_1}\mathscr{F}_n$ and $B_{S_2} \in \otimes_{n\in S_2}\mathscr{F}_n$ are rectangles.

Let $S = S_1 \cup S_2$, then $\pi_{S_1}^{-1}(B_{S_1}) = \pi_S^{-1}(\tau_S(B_{S_1}))$ and $\pi_{S_2}^{-1}(B_{S_2}) = \pi_S^{-1}(\tau_S(B_{S_2}))$. Since $\tau_S(B_{S_1}))$ and $\tau_S(B_{S_2})$ are rectangles in $\prod_{n\in S}\Omega_n$, we have that $\tau_S(B_{S_1}) \cap \tau_S(B_{S_2})$ is a rectangle and thus, $\pi_S^{-1}(\tau_S(B_{S_1}) \cap \tau_S(B_{S_2})) \in \mathcal{A}$. Likewise, $\tau_S(B_{S_1}) \setminus \tau_S(B_{S_2})$ is a disjoint union of rectangles by Lemma 4.1.1 and hence,

$$\pi_S^{-1}(\tau_S(B_{S_1})) \setminus \pi_S^{-1}(\tau_S(B_{S_2})) = \pi_S^{-1}(\tau_S(B_{S_1}) \setminus \tau_S(B_{S_2}))$$

is a disjoint union of sets in $\mathscr{A}$. So $\mathcal{A}$ is a semi-ring.

Now, let $\pi_{S_1}^{-1}(B_{S_1})$, $\pi_{S_2}^{-1}(B_{S_2}) \in \mathscr{B}$ for $S_1, S_2 \subseteq \mathbb{N}$ finite and $B_{S_1} \in \otimes_{n\in S_1}\mathscr{F}_n$, $B_{S_2} \in \otimes_{n\in S_2}\mathscr{F}_n$. Then, $\pi_{S_1}^{-1}(B_{S_1}) \cup \pi_{S_2}^{-1}(B_{S_2}) = \pi_S^{-1}(\tau_S(B_{S_1}) \cup \tau_S(B_{S_2})) \in \mathscr{B}$ where $S = S_1 \cup S_2$. Moreover, $\prod_{n=1}^{\infty}\Omega_n \setminus \pi_{S_1}^{-1}(B_{S_1}) = \pi_{S_1}^{-1}((\prod_{n\in S_1}\Omega_n) \setminus B_{S_1}) \in \mathscr{B}$.

Now, it is clear that $\sigma(\mathcal{A}) \subset \sigma(\mathscr{B})$. For any $S \subset \mathbb{N}$ be finite and any $B_S \in \otimes_{k\in S}\mathscr{F}_k$, we have that $\pi_S^{-1}(B_S) \in \sigma(\{\pi_S\}_{\text{finite } S\subset\mathbb{N}})$. Hence, $\sigma(\mathscr{B}) \subset \sigma(\{\pi_S\}_{\text{finite } S\subset\mathbb{N}})$.

Now, let $S \subset \mathbb{N}$ be finite and define $\mathscr{G} := \{B \in \otimes_{k\in S}\mathscr{F}_k : \pi_S^{-1}(B) \in \sigma(\{\pi_k\}_{k\in S})\}$. It is straightforward to verify that $\mathscr{G}$ is a sub-σ-field of $\otimes_{k\in S}\mathscr{F}_k$ and that for any $A_k \in \mathscr{F}_k$, we have that $\pi_S^{-1}(\prod_{k\in S} A_k) = \cap_{k\in S}\pi_k^{-1}(A_k)$ and thus $\prod_{k\in S} A_k \in \mathscr{G}$. Therefore, $\mathscr{G} = \otimes_{k\in S}\mathscr{F}_k$ and hence, π_S is $\sigma(\{\pi_k\}_{k\in\mathbb{N}})/\otimes_{k\in S}\mathscr{F}_k$-measurable. Thus, $\sigma(\{\pi_S\}_{\text{finite } S\subset\mathbb{N}} \subseteq \sigma(\{\pi_k\}_{k\in\mathbb{N}})$.

Finally, since, for any $k \in \mathbb{N}$, the projection π_k is $\sigma(\mathcal{A})/\mathscr{F}_k$-measurable, we have that $\sigma(\{\pi_k\}_{k\in\mathbb{N}}) \subseteq \sigma(\mathcal{A})$, which completes the proof. $\square$

Theorem 4.3.1. Let $\Omega = \prod_{n=1}^{\infty}\Omega_n$ and $\mathscr{F} = \otimes_{n=1}^{\infty}\mathscr{F}_n$. Suppose, for every $n \in \mathbb{N}$, $\exists$ Markov kernel $\{(\omega_1,\ldots,\omega_n), \mathscr{B}_{n+1}\} \mapsto P(\omega_1,\ldots,\omega_n, B_{n+1}) \in [0,1]$ for $(\omega_1,\ldots,\omega_n) \in \prod_{i=1}^{n}\Omega_i$ and $B_{n+1} \in \mathscr{F}_{n+1}$.

Let μ_1 be a probability measure on $(\Omega_1, \mathscr{F}_1)$. Then, $\exists$ unique probability measure P on $(\Omega, \mathscr{F})$ such that for any $S \subseteq \mathbb{N}$ finite and $B_S \in \otimes_{n\in S}\mathscr{F}_n$, with $m = \max S$,

$$\begin{aligned} P(\pi_S^{-1}(B_S)) &= P(\pi_{[m]}^{-1}(\tau_{[m]}(B_S)) \\ &= \int_{\Omega_1}\cdots\int_{\Omega_m}\mathbb{1}_{\tau_{[m]}(B_S)}(\omega_1,\ldots,\omega_m)\mathrm{d}P(\omega_1,\ldots,\omega_{m-1},\bullet)(\omega_m)\ldots\mathrm{d}\mu_1(\omega_1). \end{aligned} \tag{4.3}$$

Proof.

First we show that (4.3) is well-defined. Note that if we let $\tilde{m} > m := \max S$, then

$$\mathbb{1}_{\tau_{[\tilde{m}]}(B_S)}(\omega_1, \dots, \omega_m, \omega_{m+1}, \dots, \omega_{\tilde{m}}) = \mathbb{1}_{\tau_{[m]}(B_S)}(\omega_1, \dots, \omega_m).$$

Thus,

$$\begin{aligned}
&\int_{\Omega_1} \cdots \int_{\Omega_m} \mathbb{1}_{\tau_{[m]}(B_S)}(\omega_1, \dots, \omega_m) \mathrm{d}P(\omega_1, \dots, \omega_{m-1}, \bullet)(\omega_m) \dots \mathrm{d}\mu_1(\omega_1) \\
&= \int_{\Omega_1} \cdots \int_{\Omega_m} \mathbb{1}_{\tau_{[m]}(B_S)}(\omega_1, \dots, \omega_m) \underbrace{\left\{ \int_{\Omega_{m+1}} \cdots \int_{\Omega_{\tilde{m}}} \mathrm{d}P(\omega_1, \dots, \omega_{\tilde{m}-1}) \dots \mathrm{d}P(\omega_1, \dots, \omega_m, \bullet)(\omega_{m+1}) \right\}}_{=1} \\
&\qquad \mathrm{d}P(\omega_1, \dots, \omega_{m-1}, \bullet)(\omega_m) \dots \mathrm{d}\mu_1(\omega_1).
\end{aligned}$$

Let $C = C(S, B_S) = C(\widetilde{S}, B_{\widetilde{S}})$ for $S, \widetilde{S} \subseteq \mathbb{N}$ finite and $B_S \in \otimes_{n \in S} \mathscr{F}_n$ and $B_{\widetilde{S}} \in \otimes_{n \in \widetilde{S}} \mathscr{F}_n$. Then, writing $m = \max S \cup \widetilde{S}$, it must be that $\tau_{[m]}(B_S) = \tau_{[m]}(B_{\widetilde{S}})$. Thus, $P(C(S, B_S)) = P(C(\widetilde{S}, B_{\widetilde{S}})) = P([m], \tau_{[m]}(B_S))$.

By Lemma 4.3.1, we need only show that $P : \mathscr{B} \to [0, 1]$ is a pre-measure, i.e., countably additive. Let $C_1, C_2, \dots, C_n \in \mathscr{B}$ be disjoint with $C_i = C(S_i, B_{S_i})$ $\forall i \in [n]$. By letting $m := \max \cup_{i=1}^n S_i$, and embedding B_{S_i} into $[m]$ as $\tau_{[m]}(B_{S_i})$, we have that $P(\cup_{i=1}^n C_i) = \sum_{i=1}^n P(C_i)$.

Let $C_1, C_2, \dots \in \mathscr{B}$ be disjoint. Write $C = \cup_{n=1}^\infty C_n$ and define $\widetilde{C}_1 = C \setminus C_1$, $\widetilde{C}_2 := C \setminus (C_1 \cup C_2)$, $\dots$. We claim that $\lim_{n \to \infty} P(\widetilde{C}_n) = 0$; Since

$$P(C) = P(\widetilde{C}_n) + P(C_1 \cup C_2 \dots \cup C_n) = P(\widetilde{C}_n) + \sum_{i=1}^n P(C_i) \quad \forall n \in \mathbb{N},$$

this claim implies $P(C) = \lim_{n \to \infty} P(\widetilde{C}_n) + \sum_{i=1}^n P(C_i) = \sum_{i=1}^\infty P(C_i)$. To see the claim, note that $\widetilde{C}_1 \supseteq \widetilde{C}_2 \supseteq \widetilde{C}_2 \dots$ and $\cap_{n=1}^\infty \widetilde{C}_n = \emptyset$. Let us write $\widetilde{C}_n = C(S_n, B_{S_n})$ for $S_n \subseteq \mathbb{N}$ finite and $B_{S_n} \in \otimes_{i \in S_n} \mathscr{F}_i$. By taking embeddings if necessary, we may assume $S_1 \subseteq S_2 \subseteq \dots$ and define $m_n := \max S_n$.

Now, define, for $n \in \mathbb{N}$, $g_n^{(1)} : \Omega_1 \to [0, 1]$ as

$$g_n^{(1)}(\omega_1) = \int_{\Omega_2} \cdots \int_{\Omega_{m_n}} \mathbb{1}_{\tau_{[m_n]}(B_{S_n})}(\omega_1, \dots, \omega_{m_n}, \bullet) \mathrm{d}P(\omega_1, \dots, \omega_{m_n - 1})(\omega_{m_n}) \dots \mathrm{d}P(\omega_1, \bullet)(\omega_2),$$

so that $\int_{\Omega_1} g_n^{(1)}(\omega_1) \mathrm{d}\mu_1(\omega_1) = P(C(S_n, B_{S_n}))$. Since $\widetilde{C}_n \supseteq \widetilde{C}_{n+1}$, we have that

$$\mathbb{1}_{\tau_{[m_n]}(B_S)}(\omega_1, \dots, \omega_{m_n}) \geq \mathbb{1}_{\tau_{[m_{n+1}]}(B_{S_{n+1}})}(\omega_1, \dots, \omega_{m_n}, \dots, \omega_{m_{n+1}}).$$

So $g_1^{(1)} \geq g_2^{(1)} \geq \dots$. Thus, $\exists h : \Omega_1 \to [0, 1]$ such that $\forall \omega_1 \in \Omega_1$, $g_n^{(1)}(\omega_1) \to h^{(1)}(\omega_1)$ and by Dominated Convergence Theorem,

$$\lim_{n \to \infty} P(\widetilde{C}_n) = \lim_{n \to \infty} \int_{\Omega_1} g_n^{(1)} \mathrm{d}\mu_1 = \int_{\Omega_1} h^{(1)} \mathrm{d}\mu_1.$$

Suppose for sake of contradiction that $\lim_{n \to \infty} P(\widetilde{C}_n) > 0$ then let $E_1 := \{h_1 > 0\} \in \mathscr{F}_1$ and we have $\mu_1(E_1) > 0$ and hence $E_1 \neq \emptyset$. Note that $\forall \omega_1 \in E_1$, $\{\omega_1\} \times \Omega_2 \times \Omega_3 \dots \cap \widetilde{C}_1 \neq \emptyset$, or else $g_1^{(1)}(\omega_1) = 0 \implies h_1(\omega_1) = 0$.

Now, define, $\forall n \in \mathbb{N}$, $g_n^{(2)} : \Omega_1 \times \Omega_2 \to [0, 1]$ as

$$g_n^{(2)}(\omega_1, \omega_2) = \int_{\Omega_3} \cdots \int_{\Omega_m} \mathbb{1}_{\tau_{[m_n]}(B_{S_n})}(\omega_1, \dots, \omega_{m_n}) \mathrm{d}P(\omega_1, \dots, \omega_{m_n - 1}, \bullet)(\omega_{m_n}) \dots \mathrm{d}P(\omega_1, \omega_2, \bullet)(\omega_3),$$

so that $g_n^{(1)}(\omega_1) = \int_{\Omega_2} g_n^{(2)}(\omega_1, \omega_2) \mathrm{d}P(\omega_1, \bullet)(\omega_2)$. $(\star)$

By same argument, we have that $g_1^{(2)} \geq g_2^{(2)} \geq \dots$ and $g_n^{(2)} \to h^{(2)}$ for some $h^{(2)} : \Omega_1 \times \Omega_2 \to [0,1]$. Define $E_2 = \{h^{(2)} > 0\} \in \mathscr{F}_1 \otimes \mathscr{F}_2$. By $(\star)$, $\forall \omega \in \Omega_1$,

$$\begin{aligned} h^{(1)}(\omega_1) &= \lim_{n\to\infty} g_n^{(1)}(\omega_1) \\ &= \lim_{n\to\infty} \int_{\Omega_2} g_n^{(2)}(\omega_1,\omega_2) \mathrm{d}P(\omega_1,\bullet)(\omega_2) \\ &= \int_{\Omega_2} h^{(2)}(\omega_1,\omega_2)\mathrm{d}P(\omega_1,\bullet)(\omega_2). && \text{(DCT since } g_n^{(2)} \leq 1) \end{aligned}$$

Note $\forall \omega_1 \in \Omega_1$, $\mathbb{E}_2(\omega_1) = \{\omega_2 : (\omega_1,\omega_2) \in E_2\} \neq \emptyset$. So $\pi_1(E_2) := \{\omega \in \Omega_1 : \exists (\omega_1,\omega_2) \in E_2\} = E_1 \neq \emptyset$ and $\pi_2(E_2) \neq \emptyset$. Also, $\forall (\omega_1,\omega_2) \in E_2$, $((\omega_1,\omega_2) \times \Omega_3 \times \Omega_4 \times \dots) \cap \widetilde{C}_2) \neq \emptyset$ or else $g_2^{(2)}(\omega_1,\omega_2) = 0$. So $h^{(2)}(\omega_1,\omega_2) = 0$.

By repeating this argument, we have $E_1 \times \Omega_2 \times \Omega_3 \times \dots \supseteq E_2 \times \Omega_3 \times \Omega_4 \dots \supseteq \dots$, and

- $\forall n \in \mathbb{N}$, $(E_n \times \Omega_{n+1} \times \Omega_{n+2} \dots) \cap \widetilde{C}_n \notin \emptyset$,
- $\forall n \in \mathbb{N}$, $\forall m \leq n$, $\pi_{[m]}(E_n) = E_m$ and $\pi_m(E_n) \neq \emptyset$.

Define $E := \cap_{n=1}^{\infty}(E_n \times \Omega_{n+1} \times \dots) \subseteq \prod_{i=1}^{\infty} \Omega_i$. Since, $\forall m \in \mathbb{N}$, $\pi_m(E) = \cap_{n=1}^{\infty}\pi_m(E_n \times \Omega_{n+1} \times \dots) \neq \emptyset$, we have that $E \neq \emptyset$ by the axiom of choice. Observe

$$\omega \in E \implies \{\omega\} \cap (\cap_{i=1}^{n} \widetilde{C}_i) \neq \emptyset, \forall n \in \mathbb{N} \implies \omega \in \cap_{i=1}^{\infty} \widetilde{C}_i.$$

But, $\cap_{n=1}^{\infty} \widetilde{C}_n = \emptyset$ leads to contradiction. It must be that $P(\widetilde{C}_n) \to 0$. □

4.4 Probability Measure on Metric Spaces

Definition 4.4.1. Let $\mathcal{X}$ be a set. We say that $d : \mathcal{X} \times \mathcal{X} \to [0,\infty)$ is a metric if $\forall x,y,z \in X$,

(a) $d(x,z) \leq d(x,y) + d(y,z)$,

(b) $d(x,y) = d(y,x)$,

(c) $d(x,y) = 0 \iff x = y$.

We call $\mathcal{X} := \mathcal{X}_d$ a metric space. Write, for $x \in \mathcal{X}$, $r > 0$, $B(x,r) := \{x' \in \mathcal{X} : d(x,x') < r\}$ as the open ball around x and $\bar{B}(x,r) := \{x' \in \mathcal{X} : d(x,x') \leq r\}$ as the closed ball around x.

We say that $A \subseteq \mathcal{X}$ is open if $\forall x \in A$, $\exists \varepsilon > 0$ such that $B(x,\varepsilon) \in A$. We say $A \subseteq \mathcal{X}$ is closed if A^c is open, e.g. if $x \in \mathcal{X}$ satisfy $\forall \varepsilon > 0$, $B(x,\varepsilon) \cap A \neq \emptyset$, then $x \in A$.

We say that $A \subseteq \mathcal{X}$ is compact if for any collection of open sets $\{A_\gamma\}_{\gamma\in\Gamma}$ such that $A \subseteq \cup_{\gamma\in\Gamma} A_\gamma$, $\exists$ a finite subset $\{\gamma_1,\dots,\gamma_n\} \subseteq \Gamma$ such that $A \subseteq \cup_{t=1}^{n} A_{\gamma_t}$ as well.

Given a metric space $\mathcal{X}_d$, we say $\mathscr{F}$ is the Borel σ-field if

$$\mathscr{F} = \sigma(\{\text{open sets}\}) \quad \text{Borel } \sigma\text{-field}.$$

Example 4.4.1. Examples of metric spaces:

- $\forall p$, $\mathbb{R}^p$ with $d(x,y) := \|x-y\|_2$, $\forall x,y \in \mathbb{R}^p$ is a metric space. We can also take $d(x,y) := \|x-y\|_q$, $\forall q \in [0,\infty)$ and $d(x,y) = 1 \wedge \|x-y\|_2$ for a bounded metric.
- Let $\mathcal{X} = C[0,1]$ be the set of continuous functions $[0,1] \to \mathbb{R}$. Let $d(f,g) = \sup_{x\in[0,1]} |f(x) - g(x)|$, $\forall f,g \in \mathcal{X}$, then $\mathcal{X}_d$ is a metric space.

- Let $\mathcal{P}$ be the set of all probability measures on some measurable space $(\Omega, \mathscr{F})$, and let $d(P,Q) = \mathrm{TV}(P,Q) := \sup_{A\in\mathscr{F}} |P(A) - Q(A)|$, $\forall P, Q \in \mathcal{P}$, then $\mathcal{P}_d$ is a metric space (although TV is often not a useful metric).

- Let $PCX = \mathbb{R}^{\mathbb{N}}$ and for $x = (x_1, x_2, \dots)$, $y = (y_1, y_2, \dots)$, let $d_{\text{unif}}(x,y) := \sup_{k\in\mathbb{N}} |x_k - y_k| \wedge 1$ and $d_{\text{point}}(x,y) := \sum_{k=1}^{\infty} 2^{-k}|x_k - y_k|$, then PCX_d is a metric space for either d_{unif} or d_{point}. Let $x, x^{(1)}, x^{(2)}, \dots$ be a sequence of functions in $\mathbb{R}^{\mathbb{N}}$. Then,

$$\forall\, k \in \mathbb{N},\ \lim_{n\to\infty} x_k^{(n)} = x_k \Leftrightarrow \lim_{n\to\infty} d_{\text{point}}(x^{(n)}, x) = 0.$$

  In this case, we say that d_{point} metrizes pointwise convergence. Notice that $x^{(n)} \to x$ pointwise does not imply that $d_{\text{unif}}(x^{(n)}, x) \to 0$.

Lemma 4.4.1. Let $\mathcal{X}$ be a metric spaces.

(a) Open balls are open; closed balls are closed.

(b) If $\{A_\gamma\}_{\gamma\in\Gamma}$ are open, $\cup_{\gamma\in\Gamma} A_\gamma$ is open and if Γ is finite, $\cap_{\gamma\in\Gamma} A_\gamma$ is open.

(c) If $\{A_\gamma\}_{\gamma\in\Gamma}$ are closed, then $\cap_{\gamma\in\Gamma} A_\gamma$ is closed and if Γ is finite, $\cup_{\gamma\in\Gamma} A_\gamma$ is closed.

Partial Proof.

(a) let $y \in B(x,r)$ and let $\widetilde{r} = d(x,y) < r$. Then, we have that $\forall z \in B(y, r - \tilde{r})$, $d(z,x) \le d(x,y) + d(y,z) < r$. So $z \in B(x,r)$ and hence $B(x,r)$ is open.

   Now suppose $y \notin \bar{B}(x,r)$, then $r_y := d(x,y) > r$. $\forall y' \in \mathcal{X}$ such that $d(y,y') < r_y - r$, we have $d(x,y') \ge d(x,y) - d(y,y') > r$. So $B(y, r_y - r) \subseteq \bar{B}(x,r)^c$ and hence $\bar{B}(x,r)$ is closed.

(b) Let $x \in \cup_{\gamma\in\Gamma} A_\gamma$, then $\exists \gamma \in \Gamma$ such that $x \in A_\gamma$. So $\exists \varepsilon > 0$ such that $B(x,\varepsilon) \subseteq A_\gamma \subseteq \cup_{\gamma\in\Gamma} A_\gamma$. Now suppose Γ is finite. Let $x \in \cap_{\gamma\in\Gamma} A_\gamma$, then $\exists r_\gamma > 0$ such that $B(x, r_\gamma) \subseteq A_\gamma$. Take $r = \min_{\gamma\in\Gamma} r_\gamma$ and we have $B(x,r) \subseteq \cap_{\gamma\in\Gamma} A_\gamma$. Hence, $\cap_{\gamma\in\Gamma} A_\gamma$ is open.

(c) Follows from (b).

□

Lemma 4.4.2. Again, let $\mathcal{X}$ be a metric space.

(a) Compact sets are closed.

(b) If $A \subseteq \mathcal{X}$ is compact and $B \subseteq A$ is closed, then B is compact.

(c) The following are equivalent:

   (i) $A \subseteq \mathcal{X}$ is compact,

   (ii) For any $x_1, x_2, x_3, \dots \in A$, $\exists$ subsequence $x_{n_1}, x_{n_2}, \dots$ convergent, i.e. $\exists x \in A$ such that $\lim_{m\to\infty} d(x, x_{n_m}) = 0$.

   (iii) A is complete (Cauchy sequence converges) and totally bounded (For every $\varepsilon > 0$, $\exists n \in \mathbb{N}$ and $x_1, x_2, \dots x_n \in A$ such that $A \subseteq \cup_{i=1}^n B(x_i, \varepsilon)$).

(d) If $\mathcal{X}$ is complete and $A \subseteq \mathcal{X}$ is closed and totally bounded, then A is compact.

(e) Let $C_1, C_2, \dots \subseteq \mathcal{X}$ be compact and suppose $\forall n \in \mathbb{N}$, $\cap_{i=1}^n C_i \neq \emptyset$. Then $\cap_{i=1}^\infty C_i \neq \emptyset$.

Partial Proof.

(a) Let $\tilde{x} \in A^c$ and for all $x \in A$, define $\varepsilon_x := d(x, \tilde{x})$. Then $A \subseteq \cup_{x \in A} B(x, \frac{\varepsilon_x}{2})$, which means we can find a finite subcover, $n \in \mathbb{N}$, $x_1, x_2, \dots, x_n \in A$ such that $A \subseteq \cup_{i=1}^n B(x_i, \frac{\varepsilon_{x_i}}{2})$. Define $\varepsilon := \min_{i \in [n]} \varepsilon_{x_i}$, then $d(\tilde{x}, x_i) = \varepsilon_{x_i} \geq \frac{\varepsilon}{2} + \frac{\varepsilon_{x_i}}{2}$. Note $\forall i \in [n], B(x_i, \frac{\varepsilon_{x_i}}{2}) \cap B(\tilde{x}, \frac{\varepsilon}{2}) = \emptyset$. So $B(\tilde{x}, \frac{\varepsilon}{2}) \cap A = \emptyset$ and hence A^c is open.

(b) Let A be compact and $C \subseteq A$ closed. Since C^c is open, for all $x \notin C$, $\exists \varepsilon_x > 0$ such that $B(x, \varepsilon_x) \subseteq C^c$. Let $\{B_\gamma\}_{\gamma \in \Gamma}$ be any collection of opens sets that cover C, i.e. $C \subseteq \cup_{\gamma \in \Gamma} B_\gamma$, then $A \subseteq (\cup_{\gamma \in \Gamma} B_\gamma) \cup (\cup_{x \in A \setminus C} B(x, \varepsilon_x))$ where the $\cup_{x \in A \setminus C} B(x, \varepsilon_x)) \subseteq C^c$. Since A is compact, $\exists$ finite $\widetilde{\Gamma} \subseteq \Gamma$ and $\{x_1 \dots, x_n\} \subseteq A \setminus C$ such that $A \subseteq (\cup_{\gamma \in \widetilde{\Gamma}} B_\gamma) \cup (\cup_{i=1}^n B(x_i, \varepsilon_{x_i}))$. As $\cup_{i=1}^n B(x_i, \varepsilon_{x_i}) \subseteq C^c$ by construction, $\widetilde{\Gamma}$ must be non-empty and $C \subseteq \cup_{\gamma \in \widetilde{\Gamma}} B_\gamma$. So C is compact.

(c) We will show that (i) $\implies$ (ii) and (i), (ii) $\implies$ (iii). Assume $A \subseteq \mathcal{X}$ is compact, let $\{x_1, x_2, \dots\} \subseteq A$ be a sequence. WLOG, we may assume that $\{x_1, x_2, \dots\}$ are all distinct. For each $x \in A$, define $\varepsilon_x := \inf_{\tilde{x} \in \{x_1, x_2, \dots\} \setminus x} d(x, \tilde{x})$ and define $A_1 := \{x \in A : \varepsilon_x > 0\}$ and $A_2 := \{x \in A : \varepsilon_x = 0\}$. Fix arbitrary $\varepsilon > 0$ and note that $A \subseteq \cup_{x \in A_1} B(x, \frac{\varepsilon_x}{2}) \cup \cup_{x \in A_2} B(x, \varepsilon)$. Since A is compact and since, $\forall x \in A_1$, $B(x, \frac{\varepsilon_x}{2})$ contains at most one of $\{x_1, x_2, \dots\}$. Hence A_2 is non-empty. So $\exists x \in \mathcal{X}$ such that for some subsequence $\{x_{n_1}, x_{n_2}, \dots\}$, $\lim_{m \to \infty} d(x_{n_m}, x) = 0$.

Now assume (i), (ii). Let $\{x_1, x_2, \dots\}$ be a Cauchy sequence. Let $\{x_{n_1}, x_{n_2}, \dots\}$ be a subsequence converging to $x \in A$. Then $\forall n, m \in \mathbb{N}$, $d(x_n, x) \leq d(x_n, x_{n_m}) + d(x, x_{n_m})$. Taking $n, m \to \infty$ yields that $d(x_n, x) \to 0$. The totally bounded condition follows immediately from the definition of compactness.

(d) We need only show that A is complete. Let $x_1, x_2, \dots \in A$ be a Cauchy sequence, then $\exists x \in X$ (as X is complete) such that $\lim_{n \to \infty} d(x, x_n) = 0$. So $\forall \varepsilon > 0$, $B(x, \varepsilon) \cap A \neq \emptyset$. Therefore, $x \in A$ since A is closed.

(e) Let $x_1 \in C_1$, $x_2 \in C_1 \cap C_2$, $\dots$. Since $\{x_1, x_2, \dots\} \subseteq C_1$, $\exists$ subsequence $\{x_{n_1}, x_{n_2}, \dots\}$ convergent to $x \in C_1$. But $\{x_2, x_3, \dots\} \subseteq C_2 \implies x \in C_2$ as well. In this way, we have that $x \in C_n$, $\forall n$. So $x \in \cap_{n=1}^\infty C_n$.

$\square$

Theorem 4.4.1. Let $(\mathcal{X}, \mathscr{F})$ be a metric space with $\mathscr{F}$ as the Borel σ-field. Let $(\mathcal{X}, \mathscr{F}, P)$ be a probability measure. Then, $\forall A \in \mathscr{F}$,

$$P(A) = \inf\{P(U) : A \subseteq U \text{ open}\} = \sup\{P(V) : A \supseteq V \text{ closed}\}. \tag{4.4}$$

Proof.
Define

$$\mathcal{G} := \big\{A \in \mathscr{F} : P(A) = \inf\{P(U) : A \subseteq U \text{ open}\}\,(\star) \text{ and } P(A) = \sup\{P(A) : A \supseteq V \text{ closed}\}\,(\star\star)\big\}$$

If $A \subseteq \mathcal{X}$ be open, then $(\star)$ holds. Define $C_n = \cup_{x \notin A} B(x, \frac{1}{n})$ so that C_n^c is closed and $C_1^c \subseteq C_2^c \subseteq \cdots \subseteq A$. If $\tilde{x} \in A$, then $\exists r > 0$ such that $B(\tilde{x}, r) \subseteq A \implies \tilde{x} \notin C_n$ for any n satisfying $\frac{1}{n} < r \implies \cup_{n=1}^\infty C_n^c = A$. Hence, $P(A) = \lim_{n \to \infty} P(C_n^c)$.

Since $P(C_n^c) \leq \sup\{P(V) : A \supseteq V \text{ closed}\}$, $\forall n \in \mathbb{N}$, we have $P(A) \leq \sup\{P(V) : A \supseteq V \text{ closed}\}$. But, it is clear that $P(A) \geq P(V)$ for any $V \subseteq A$ and thus $(\star\star)$ holds. This means that $\mathscr{G}$ contains all the open sets. Thus, we need only show that $\mathscr{G}$ is a σ-field.

Let $A_1, A_2, \dots \in \mathscr{G}$ and fix $\varepsilon > 0$. For $i \in \mathbb{N}$, let U_i be open and V_i be closed such that $V_i \subseteq A_i \subseteq U_i$ and

$P(V_i) + 2^{-i-1}\varepsilon \geq P(A_i) \geq P(U_i) - 2^{-i}\varepsilon$. Write $A = \cup_{i=1}^{\infty} A_i$, we have

$$\begin{aligned} P(A) &\geq P\left(\bigcup_{i=1}^{\infty} U_i\right) - P\left(\bigcup_{i=1}^{\infty} U_i \setminus A\right) \\ &\geq P\left(\bigcup_{i=1}^{\infty} U_i\right) - P\left(\bigcup_{i=1}^{\infty} (U_i \backslash A_i)\right) \\ &\geq P\left(\bigcup_{i=1}^{\infty} U_i\right) - \sum_{i=1}^{\infty} \left(P\left(U_i\right) - P\left(A_i\right)\right) \\ &\geq P\left(\bigcup_{i=1}^{\infty} U_i\right) - \sum_{i=1}^{\infty} 2^{-i}\varepsilon = P\left(\bigcup_{i=1}^{\infty} U_i\right) - \varepsilon. \end{aligned}$$

Since $\cup_{i=1}^{\infty} U_i$ is open by Lemma 4.4.1, we have that $P(A) \geq \inf\{P(U) : A \subseteq U \text{ open}\} - \varepsilon$.

Now, since $P(\cup_{i=1}^{\infty} V_i) = \lim_{n\to\infty} P(\cup_{i=1}^{n} V_i)$, $\exists m \in \mathbb{N}$ such that $P(\cup_{i=1}^{m} V_i) \geq P(\cup_{i=1}^{\infty} V_i) - \frac{\varepsilon}{2}$. Then,

$$\begin{aligned} P(A) &\leq P\left(\bigcup_{i=1}^{\infty} V_i\right) + P\left(A \setminus \bigcup_{i=1}^{\infty} V_i\right) \\ &\leq P\left(\bigcup_{i=1}^{\infty} V_i\right) + P\left(\bigcup_{i=1}^{\infty} (A \setminus V_i)\right) \\ &\leq P\left(\bigcup_{i=1}^{\infty} V_i\right) + \sum_{i=1}^{\infty} \left(P\left(A_i\right) - P\left(V_i\right)\right) \\ &\leq P\left(\bigcup_{i=1}^{m} V_i\right) + \frac{\varepsilon}{2} + \frac{1}{2}\sum_{i=1}^{\infty} 2^{-i}\varepsilon = P\left(\bigcup_{i=1}^{m} V_i\right) + \varepsilon. \end{aligned}$$

Since $\cup_{i=1}^{m} V_i$ is closed, $P(A) \leq \sup\{P(V) : A \supseteq V \text{ closed}\} + \varepsilon$. Since ε is arbitrary, $A \in \mathcal{G}$. Let $A \in \mathcal{G}$, fix $\varepsilon > 0$, and let $V \subseteq A \subseteq U$ be such that V is closed, U is open, and $P(A) \leq P(V) + \frac{\varepsilon}{2}$, $P(A) \geq P(U) - \frac{\varepsilon}{2}$. Then, $V^c \supseteq A^c \supseteq U^c$ where V^c is open, U^c is closed, and $P(A^c) = 1 - P(A) \geq 1 - P(V) - \frac{\varepsilon}{2} = P(V^c) - \frac{\varepsilon}{2}$, and $P(A^c) \leq 1 - P(U) + \frac{\varepsilon}{2} = P(U^c) + \frac{\varepsilon}{2}$. Therefore $A^c \in \mathcal{G}$. Thus, $\mathcal{G}$ is a σ-field and $\mathcal{G} = \mathscr{F}$. □

Definition 4.4.2. Let $\mathcal{X}$ be a metric space and $(\mathcal{X}, \mathscr{F}, P)$ be a probability space with Borel σ-field. We say that P is <u>tight</u> if $\forall\varepsilon > 0$, $\exists$ compact $K \subseteq \mathcal{X}$ such that $P(\mathcal{X} \setminus K) < \varepsilon$.

Lemma 4.4.3. If P is <u>tight</u> probability measure on $(\mathcal{X}, \mathscr{F})$, then $\forall A \in \mathscr{F}$, $P(A) = \sup\{P(K) : A \supseteq K \text{ compact}\}$.

Proof.

Let $\forall\varepsilon > 0$, and let $K_0 \subseteq X$ be compact such that $P(X \setminus K_0) < \varepsilon$. Then, $\forall A \in \mathscr{F}$,

$$\begin{aligned} P(A) &= P(A \setminus K_0) + P(A \cap K_0) \\ &\leq \varepsilon + P(A \cap K_0) \\ &= \varepsilon + \sup\{P(V) : A \cap K_0 \supseteq V \text{ closed}\} \\ &\leq \varepsilon + \sup\{P(K) : A \supseteq K \text{ compact}\} \end{aligned}$$

because $V \subseteq A \cap K_0$ and closed. So V is compact by Lemma 4.4.2 (b), and $V \subseteq A$. Since $\varepsilon > 0$ is arbitrary, the conclusion follows as desired. □

Theorem 4.4.2. Let $\mathcal{X}$ be a metric space such that

- $\mathcal{X}$ is complete and

- $\exists \tilde{x}_1, \tilde{x}_2, \cdots \in \mathcal{X}$ such that $\forall x \in \mathcal{X}$, $\inf_{n\in\mathbb{N}} d(x, \tilde{x}_n) = 0$ (separable)

Then every probability measure on $(\mathcal{X}, \mathscr{F})$, with $\mathscr{F}$ as the Borel σ-field, is tight. (Complete separable metric space is called a Polish space.)
Example:

- $(\mathbb{R}^d, \ell_2)$ is complete and separable; consider $\mathbb{Q}^d$.
- $(C[0,1], L_\infty)$ is complete and separable by the Stone-Weierstrass theorem.

Proof.
Fix $\varepsilon > 0$ and let P be any probability measure on $(\mathcal{X}, \mathscr{F})$. Note that for any $\delta > 0$, $\mathcal{X} = \cup_{i=1}^\infty B(\tilde{x}_i, \delta)$ and thus, $P(\mathcal{X}) = \lim_{n\to\infty} P(\cup_{i=1}^n B(\tilde{x}_i, \delta))$. Hence, for any $m \in \mathbb{N}$, $\exists n_m \in \mathbb{N}$ such that

$$P(\mathcal{X}) \leq P\left(\bigcup_{i=1}^{n_m} B(\tilde{x}_i, \frac{1}{m})\right) + 2^{-m}\varepsilon.$$

Define $K = \cap_{m=1}^\infty \cup_{i=1}^{n_m} \bar{B}(\tilde{x}_i, \frac{1}{m})$, then K is closed. Also $\forall \delta > 0$, we may take $m > \frac{1}{\delta}$ to obtain $K \subseteq \cup_{i=1}^{n_m} \bar{B}(\tilde{x}_i, \frac{1}{m}) \subseteq \cup_{i=1}^{n_m} \bar{B}(\tilde{x}_i, \delta)$. So K is totally bounded and thus compact by Lemma 4.4.2 (d). Then,

$$\begin{aligned}
P(\mathcal{X} \setminus K) &= P\left(\bigcup_{m=1}^\infty \left(\mathcal{X} \setminus \bigcup_{i=1}^{n_m} \bar{B}(\tilde{x}_i, \frac{1}{m})\right)\right) \\
&\leq \sum_{m=1}^\infty P\left(\mathcal{X} \setminus \bigcup_{i=1}^{n_m} \bar{B}(\tilde{x}_i, \frac{1}{m})\right) \\
&= \sum_{m=1}^\infty \left\{P(\mathcal{X}) - P\left(\bigcup_{i=1}^{n_m} \bar{B}(\tilde{x}_i, \frac{1}{m})\right)\right\} \leq \sum_{m=1}^\infty 2^{-m}\varepsilon \leq \varepsilon.
\end{aligned}$$

So P is tight. □

4.5 Kolmogorov Extension Theorem

Definition 4.5.1. Let T be any index set and for $t \in T$, let $(\Omega_t, \mathscr{F}_t)$ be a measurable space. For any finite $S \subseteq T$, let P_S be a probability measure on $(\prod_{s\in S} \Omega_s, \otimes_{s\in S} \mathscr{F}_s)$.

We say that the family $\{P_S : S \subseteq T \text{ finite}\}$ is consistent if, for any $S_0 \subseteq S_1 \subseteq T$ finite, we have that $P_{S_0} = P_{S_1}^{(\pi_{S_1,S_0})}$ where $P_{S_1}^{(\pi_{S_1,S_0})}$ is the pushforward measure on $(\prod_{s\in S_0} \Omega_s, \otimes_{s\in S_0} \mathscr{F}_s)$ induced by the projection $\pi_{S_1,S_0} : \prod_{s\in S_1} \Omega_s \to \prod_{s\in S_0} \Omega_s$, i.e. $\forall B \in \otimes_{s\in S_0} \mathscr{F}_s$,

$$P_{S_0}(B) = P_{S_1}(\pi_{S_1,S_0}^{-1}(B)) = P_{S_1}\left(\left\{\omega \in \prod_{s\in S_1} \Omega_s : \omega_{S_0} \in B\right\}\right).$$

Recall that for $S \subseteq T$ finite and $B_S \in \otimes_{s\in S} \mathscr{F}_s$, we have that $\pi_S^{-1}(B) := \pi_{T,S}^{-1}(B) := C(S, B_S) = \{\omega \in \prod_{t\in T} \Omega_t : \omega_S \in B_S\}$ is a measurable cylinder. We define the product σ-field $\otimes_{t\in T} \mathscr{F}_t := \sigma(\{C(S, B_S) : S \subseteq T \text{ finite}, B_S \in \otimes_{s\in S} \mathscr{F}_s\})$. Recall by Lemma 4.3.1 that

$$\otimes_{t\in T} \mathscr{F}_t = \sigma(\{\pi_s\}_{s\in T}).$$

Think of $\omega \in \prod_{t\in T} \Omega_t$ as a function $T \to \cup_{t\in T} \Omega_t$.

Remark 4.5.1. Let T be an arbitrary index set. For simplicity, suppose $\Omega_t = \mathcal{X}$ and $\mathscr{F}_t = \mathscr{G}$ for all $t \in T$ so that $\prod_{t\in T} \Omega_t = \mathcal{X}^T$. Let $(\Omega, \mathscr{F})$ be a measure space. Let $Z : Omega \to \mathcal{X}^T$ be a function-valued function.

For each $t \in T$, define $Z_t : \Omega \to \mathcal{X}$ by $Z_t = \pi_t \circ Z$ so that for every $\omega \in \Omega$, $Z_t(\omega) = Z(\omega)(t)$. Then,

$$Z \text{ is } \mathscr{F}/\mathscr{G}^{\otimes T}\text{-measurable iff } Z_t \text{ is } \mathscr{F}/\mathscr{G}\text{-measurable for all } t \in T.$$

Theorem 4.5.1 (Kolmogorov Consistency/Extension Theorem). Using the setting of Definition 4.5.1, suppose, for any $t \in T$, (Ω_t, d_t) is a complete and separable metric space. Then, for any consistent family $\{P_S : S \subseteq T \text{ finite}\}$, there exists a unique probability measure P on $(\prod_{t\in T} \Omega_t, \otimes_{t\in T}\mathscr{F}_t)$ such that for any $S \subseteq T$ finite and $B_S \in \otimes_{k\in S}\mathscr{F}_k$, $P(\pi_S^{-1}(B_S)) = P_S(B_S)$.

Example 4.5.1. Let $T = \mathbb{R}^d$ and $\forall t \in T$, let $\Omega_t = \mathbb{R}$, with $\mathscr{F}_t := \mathscr{B}(\mathbb{R})$. Let $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a covariance function, i.e., for any $n \in \mathbb{N}$, $t_1, t_2, \dots, t_n \in \mathbb{R}^d$,

$$\begin{pmatrix} K(t_1,t_2) & \dots & K(t_1,t_n) \\ \vdots & \ddots & \vdots \\ K(t_n,t_1) & \dots & K(t_n,t_n) \end{pmatrix} \in \mathbb{R}^{n\times n}$$

is positive semi-definite. E.g.

- $K(s,t) = \sigma^2 \mathbb{1}_{\{s=t\}}$ for $\sigma > 0$ (white noise).
- $K(s,t) = \sigma^2 \exp\{-\frac{\|s-t\|_2^2}{\ell^2}\}$ for $\sigma > 0$, $\ell > 0$ (Square exponential)
- Matern kernel

Let $\mu : T \to \mathbb{R}$ be a mean function. For $S = \{t_1, t_2, \dots, t_n\} \subseteq T$ finite, let P_S be $N(\mu, \Sigma)$ where $\mu = (\mu(t_1), \dots, \mu(t_n)) \in \mathbb{R}^n$ and

$$\Sigma := \begin{pmatrix} K(t_1,t_2) & \dots & K(t_1,t_n) \\ \vdots & \ddots & \vdots \\ K(t_n,t_1) & \dots & K(t_n,t_n) \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

Note that P_S is a valid probability measure on $\prod_{s\in S}\Omega_s = \mathbb{R}^n$ and $\otimes_{s\in S}\mathscr{F}_s = \mathscr{B}(\mathbb{R}^n)$. Also, $\{P_S : S \subseteq T \text{ finite}\}$ is consistent implies $\exists$ unique probability measure on $(\mathbb{R}^T, \mathscr{B}(\mathbb{R})^{\otimes T})$; this is know as a Gaussian process.

Example 4.5.2. Let $T = [0,\infty)$, $\forall t \in T$, let $\Omega_t = \mathbb{R}$ and $\mathscr{F}_t = \mathscr{B}(\mathbb{R})$. For $n \in \mathbb{N}$ and $S = \{t_1, t_2, \dots, t_n\} \in T$ where $0 \le t_1 \le t_2 \le \dots t_n$, let

$$X_1 \sim N(0, t_1), X_2 \sim N(X_1, t_2 - t_1), X_3 \sim N(X_1 + X_2, t_3 - t_2), \dots, X_n \sim N(\sum_{i=1}^{n-1} X_i, t_n - t_{n-1}),$$

and let P_S on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ be the jointly probability measure induced by $(X_1, X_2, \dots, X_n)$, i.e. $\forall B \in \mathscr{B}(\mathbb{R}^n)$, $P_S(B) = \mathbb{P}((X_1, \dots, X_n) \in B)$.

Then, $\{P_S : S \subseteq T \text{ finite}\}$ is consistent and the resulting probability measure on $(\mathbb{R}^T, \mathscr{B}(\mathbb{R})^{\otimes T})$ is called one-sided Brownian motion or Wiener process.

Remark 4.5.2. Additional work is required to show that a realization from a Gaussian process (with appropriate convariance function) or Wiener process is continuous. Formally, one can show that $\exists C \in \mathscr{B}(\mathbb{R})^{\otimes T}$ such that $\forall f \in C, f : T \to \mathbb{R}$ is continuous and $\forall A \in \mathscr{B}(\mathbb{R})^{\otimes T}$, $P(A \cap C) = P(A)$.

Stated another way, let $(\Omega, \mathscr{F}, \mu)$ be some probability space and let $X : \Omega \to \mathbb{R}^T$ be measurable with respect to $(\mathbb{R}^T, \mathscr{B}(\mathbb{R})^{\otimes T}))$, such that the pushforward measure induced by X is a Gaussian process, then $\exists Y : \Omega \to \mathbb{R}^T$ such that $\mathbb{P}(Y \text{ is continuous}) = 1$ and $\forall t \in T$, $\mathbb{P}(X_t = Y_t) = 1$.

Lemma 4.5.1. Let $\mathcal{C} = \{C_1, C_2, C_3, \dots\}$ be a family of subsets of some set X. We say that $\mathcal{C}$ is a compact class if $\forall n \in \mathbb{N}$, $\cap_{i=1}^n C_i \ne \emptyset \implies \cap_{i=1}^\infty C_i \ne \emptyset$.

Let $\mathscr{F}_0 \subseteq 2^X$ be a field and let $P : \mathscr{F}_0 \to [0,1]$ be a finitely additive function. Let $\mathcal{C} \subseteq \mathscr{F}_0$ be a compact class. If $\forall A \in \mathscr{F}_0$, $P(A) = \sup\{P(K) : K \subseteq A, K \in \mathcal{C}\}$, then P is a pre-measure.

By Lemma 4.4.2, any collection of compact sets is a compact class.

Proof.

Let $A_1, A_2, \ldots, \in \mathscr{F}_0$ be disjoint and let $A = \cup_{i=1}^{\infty} A_i$. Define $\widetilde{A}_1 = A \setminus A_1$, $\widetilde{A}_2 = A \setminus (A_1 \cup A_2)$, $\ldots$ so that $\widetilde{A}_1 \supseteq \widetilde{A}_2 \supseteq \ldots$ and that $\cap_{i=1}^{\infty} \widetilde{A}_i = \emptyset$. We need only show that $P(\widetilde{A}_n) \to 0$. Fix $\varepsilon > 0$ and for each $i \in \mathbb{N}$, let $K_i \in \mathcal{C}$ such that $K_i \subseteq \widetilde{A}_i$ and $P(\widetilde{A}_i) \leq P(K_i) + 2^{-i}\varepsilon$.

Since $\cap_{i=1}^{\infty} K_i \subseteq \cap_{i=1}^{\infty} \widetilde{A}_i = \emptyset$, $\exists m \in \mathbb{N}$ such that $\cap_{i=1}^{m} K_i = \emptyset$. Hence,

$$\begin{aligned} P(\widetilde{A}_m) = P(\widetilde{A}_m \setminus \cap_{i=1}^{m} K_i) &\leq P(\cup_{i=1}^{m} (\widetilde{A}_m \setminus K_i)) \\ &\leq P(\cup_{i=1}^{m} (\widetilde{A}_i \setminus K_i)) \\ &\leq \sum_{i=1}^{m} ((P(\widetilde{A}_i) - P(K_i)) \leq \varepsilon. \end{aligned}$$

Since ε is arbitrary, $\lim_{n\to\infty} P(\widetilde{A}_n) = 0$ as desired. □

Lemma 4.5.2. Let X be a set and let $\mathcal{C} := \{C_1, C_2, C_3, \ldots\} \in 2^X$ be a compact class. Let $\mathscr{D} = \{\cup_{i=1}^{n} C_i : n \in \mathbb{N}, C_1, \ldots, C_n \in \mathcal{C} \text{ disjoint}\}$, then $\mathscr{D}$ is also a compact class.

Proof.

Let $D_1, D_2, \cdots \in \mathscr{D}$, and suppose, $\forall n \in \mathbb{N}$, $\cap_{i=1}^{n} D_i \neq \emptyset$. Write, for any $i \in \mathbb{N}$, $D_i = \cup_{j=1}^{m_i} C_j^{(i)}$, where $C_1^{(i)}, \ldots, C_{m_i}^{(i)} \in \mathcal{C}$ are disjoint. For $n \in \mathbb{N}$, define $S_n := \{(s_1, \ldots, s_n) := s_1 \in [m_1], s_2 \in [m_2], \ldots, s_n \in [m_n]\}$ and $S_\infty := \{(s_1, s_2, \ldots) : s_1 \in [m_1], s_2 \in [m_2], \ldots\}$.

For $n \in \mathbb{N}$, $\emptyset \neq \cap_{i=1}^{n} D_i = \cap_{i=1}^{n} \cup_{j=1}^{m_i} C_j^{(i)} = \cup_{(s_1,\ldots,s_n)\in S_n} \cap_{k=1}^{n} C_{s_k}^{(k)} \implies \exists (s_1, \ldots, s_n) \in S_n$ such that $\cap_{k=1}^{n} C_{s_k}^{(k)} \neq \emptyset$. Since n is arbitrary, $\exists (s_1, s_2, \ldots) \in S_\infty$ such that $\cap_{k=1}^{\infty} C_{s_k}^{(k)} \neq \emptyset$ since $\forall k \in \mathbb{N}$, $C_{s_k}^{(k)} \in \mathcal{C}$ and $\mathcal{C}$ is a compact class.

Thus, $\cap_{i=1}^{\infty} D_i = \cap_{i=1}^{\infty} \cup_{j=1}^{m_i} C_j^{(i)} = \cup_{(s_1,s_2,\ldots)\in S_\infty} \cap_{k=1}^{\infty} C_{s_k}^{(k)} \neq \emptyset$. So $\mathscr{D}$ is a compact class as well. □

Proof of Theorem 4.5.1.

Let $\mathcal{A} := \{C(S, B_S) : S \subseteq T \text{ finite}, B_S = \prod_{s\in S} A_s, A_s \in \mathscr{F}_s, \forall s \in S\}$, then $\mathcal{A}$ is a semi-ring by Lemma 4.3.1. Define, $\mathscr{F}_0 = \{\cup_{i=1}^{n} C_i : n \in \mathbb{N}, C_1, C_2, \ldots, C_n \in \mathcal{A} \text{ disjoint}\}$, then $\mathscr{F}_0$ is a field. Define $\widetilde{P} : \mathscr{F}_0 \to [0, 1]$ such that, for $C_1, \ldots, C_n \in \mathcal{A}$ disjoint,

$$\widetilde{P}(\cup_{i=1}^{n} C_i) = \sum_{i=1}^{n} \widetilde{P}(C_i) = \sum_{i=1}^{n} P_{S_i}(B_{S_i})$$

where (S_i, B_{S_i}) is the base of C_i. This is well-defined since $\{P_S : S \subseteq T \text{ finite}\}$ is consistent. P is clearly finitely additive. Since $\forall t \in T$, Ω_t is a complete and separable with $\mathscr{F}_t$ as Borel σ-field, $\forall A \in \mathscr{F}_t$,

$$P_t(A) = \sup\{P(K) : K \subseteq A, K \text{ compact}\}.$$

<u>Step 1</u>: We claim that

$$\mathcal{C} := \left\{ C(S, B_S) \subseteq \prod_{t\in T} \Omega_t : S \subseteq T \text{ finite}, B_S = \prod_{s\in S} K_s, K_s \in \mathscr{F}_s \text{ compact} \right\} \subseteq \mathcal{A}$$

is a compact class. Let $\widetilde{C}_1, \widetilde{C}_2, \cdots \in \mathcal{C}$ and suppose $\forall n \in \mathbb{N}$, $\cap_{i=1}^{n} \widetilde{C}_i = \emptyset$. Let $S \subseteq T$ countable, be the union of the base coordinates of $\widetilde{C}_1, \widetilde{C}_2, \ldots$

Fix $n \in \mathbb{N}$ and write $\widetilde{C}_i = C(s, B_i)$ with $S \subseteq T$ countable and $B_i = \prod_{s\in S} K_s^{(i)}$, where $\forall s \in S$, $K_s^{(i)}$ is either Ω_s or compact. Then, we have that, for all $n \in \mathbb{N}$,

$$\begin{aligned} \emptyset \neq \bigcap_{i=1}^{n} \widetilde{C}_i = \bigcap_{i=1}^{n} C(S, B_i) &= C\left(S, \bigcap_{i=1}^{n} B_i\right) \\ &= C\left(S, \prod_{i=1}^{n} \prod_{s\in S} K_s^{(i)}\right) = C\left(S, \prod_{s\in S}\left(\bigcap_{i=1}^{n} K_s^{(i)}\right)\right) \end{aligned}$$

Hence $\prod_{s\in S}\left(\cap_{i=1}^n K_s^{(i)}\right)\neq\emptyset$. So $\forall s\in S$, $\cap_{i=1}^n K_s^{(i)}\neq\emptyset$. Since $\forall s\in S$, $\forall i\in\mathbb{N}$, $K_s^{(i)}$ is either compact or Ω_s, we have that $\cap_{i=1}^\infty K_s^{(i)}\neq\emptyset$. Hence $\prod_{s\in S}(\cap_{i=1}^\infty K_s^{(i)})\neq\emptyset$ by Axiom of choice. It implies

$$\begin{aligned}\emptyset\neq C\left(S,\prod_{s\in S}\left(\bigcap_{i=1}^\infty K_s^{(i)}\right)\right)&=C\left(S,\bigcap_{i=1}^\infty\prod_{s\in S}K_s^{(i)}\right)\\&=\bigcap_{i=1}^\infty C\left(S,\prod_{s\in S}K_s^{(i)}\right)=\bigcap_{i=1}^\infty\widetilde{C}_i.\end{aligned}$$

So $\mathcal{C}$ is a compact class. Thus, $\mathscr{D}:=\{\cup_{i=1}^n\widetilde{C}_i:n\in\mathbb{N},\widetilde{C}_1,\widetilde{C}_2,\dots,\widetilde{C}_n\in\mathcal{C}\text{ disjoint}\}\subseteq\mathscr{F}_0$ is a compact class.
Step 2: Let $\cup_{i=1}^n C_i\in\mathscr{F}_0$ where $C_1,\dots,C_n\in\mathcal{A}$ are disjoin. We claim that

$$\widetilde{P}(\cup_{i=1}^n C_i)=\sup\{\widetilde{P}(\cup_{i=1}^n\widetilde{C}_i):n\in\mathbb{N},\widetilde{C}_i\in\mathcal{C}\}=\sup\{\widetilde{P}(D):D\subseteq\cup_{i=1}^n C_i,D\in\mathscr{D}\}.$$

To see this, write $S\subseteq T$ finite as the union of the base coordinates of $C_1,C_2,\dots,C_n$, and write $C_i=C(S,B_i)$ where $B_i=\prod_{s\in S}A_s^{(i)}$, with $A_s^{(i)}\in\mathscr{F}_s$, $\forall s\in S$. Since, for all $s\in T$, $(\Omega_s,\mathscr{F}_s)$ is a complete and separable metric space, for any $i\in[n]$ and $s\in S$, $\exists K_s^{(i)}\in\mathscr{F}_s$ compact such that $K_s^{(i)}\subseteq A_s^{(i)}$ and $P_s(A_s^{(i)})\leq P_s(K_s^{(i)})+\frac{\varepsilon}{n|S|}$. Then,

$$\begin{aligned}\widetilde{P}(\cup_{i=1}^n C_i)=\sum_{i=1}^n\widetilde{P}(C_i)&=\sum_{i=1}^n P_S\left(\prod_{s\in S}A_s^{(i)}\right)\\&=\sum_{i=1}^n P_S\left((\prod_{s\in S}A_s^{(i)})\setminus(\prod_{s\in S}K_s^{(i)})\right)+\sum_{i=1}^n P_S\left(\prod_{s\in S}K_s^{(i)}\right)\\&=\sum_{i=1}^n P_S\left(\bigcup_{s\in S}\tau_S\left(A_s^{(i)}\setminus K_s^{(i)}\right)\right)+\sum_{i=1}^n\widetilde{P}\left(C\left(S,\prod_{s\in S}K_s^{(i)}\right)\right)\\&\leq\sum_{i=1}^n\sum_{s\in S}P_S\left(\tau_S\left(A_s^{(i)}\setminus K_s^{(i)}\right)\right)+\widetilde{P}\left(\bigcup_{i=1}^n C\left(S,\prod_{s\in S}K_s^{(i)}\right)\right)\\&\leq\sum_{i=1}^n\sum_{s\in S}P_s\left(A_s^{(i)}\setminus K_s^{(i)}\right)+\widetilde{P}\left(\bigcup_{i=1}^n C\left(S,\prod_{s\in S}K_s^{(i)}\right)\right)\\&\leq\widetilde{P}\left(\bigcup_{i=1}^n C\left(S,\prod_{s\in S}K_s^{(i)}\right)\right)+\varepsilon.\end{aligned}$$

Thus, $\widetilde{P}:\mathscr{F}_0\to[0,1]$ is countably additive. So there exists unique extension $P:\sigma(\mathscr{F}_0)\to[0,1]$. ☐

Example 4.5.3. Let $T=[0,\infty)$, $\forall t\in T$, let $\Omega_t=\mathbb{N}$ and $\mathscr{F}_t=2^{\mathbb{N}}$. Let $\lambda:T\to[0,\infty)$ be Lebesgue integrable. Let $n\in\mathbb{N}$ and $S=\{t_1,t_2,\dots,t_n\}\subseteq T$, define

$$\begin{aligned}&N_1\sim\text{Poisson}\left(\int_0^{t_1}\lambda(t)\mathrm{d}t\right),\\&N_2=N_1+M_2,\ \text{where } M_2\sim\text{Poisson}\left(\int_{t_1}^{t_2}\lambda(t)\mathrm{d}t\right),\ \text{indep of } N_1\\&\quad\vdots\\&N_n=N_{n-1}+M_n,\ \text{where } M_n\sim\text{Poisson}\left(\int_{t_{n-1}}^{t_n}\lambda(t)\mathrm{d}t\right),\ \text{indep of } N_1,N_2,\dots,N_{n-1}.\end{aligned}$$

Let P_S on $(\mathbb{N}^n,(2^{\mathbb{N}})^n)$ be the joint probability measure induced by $(N_1,N_2,\dots,N_n)$. Then $\{P_S:S\subseteq T\text{ finite}\}$ is consistent and the resulting probability measure on $(\mathbb{N}^T,(2^{\mathbb{N}})^{\otimes T})$ is called Poisson Point Process. (counting process representation).

Example 4.5.4. Recall that, for $\alpha_1, \dots, \alpha_n \geq 0$, if $(Z_1, Z_2, \dots, Z_n) \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_n)$, then $\sum_{i=1}^n Z_i = 1$, and $\mathbb{P}(Z_i > 0) = 1 \iff \alpha_i > 0$.

Let $(\mathcal{X}, \mathscr{G})$ be a measurable space. Let $\alpha > 0$ and let H be a probability measure on $(\mathcal{X}, \mathscr{G})$. Let $T = \mathscr{G}$ and, for any $A \in T$, let $\Omega_A = [0,1]$, $\mathscr{F}_A = \mathscr{B}([0,1])$.

For $n \in \mathbb{N}$, $S = \{A_1, A_2, \dots, A_n\} \subseteq T$, i.e., $A_1, A_2, \dots, A_n \in \mathscr{G}$, let $\Gamma := \{\gamma : [n] \to \{0,1\}\}$ and for $\gamma \in \Gamma$, let $\widetilde{A}_\gamma := \{x \in \mathcal{X} : x \in A_i \text{ iff } \gamma(i) = 1\}$ so that $\{\widetilde{A}_\gamma\}_{\gamma \in \Gamma}$ from a disjoint partition of $\mathcal{X}$. Let $\{Z_\gamma\}_{\gamma \in \Gamma} \sim \text{Dirichlet}(\alpha \{H(\widetilde{A}_\gamma)\}_{\gamma \in \Gamma})$ and for $i = 2, 3, \dots, n$, let $X_i = \sum_{\gamma \in \Gamma} Z_\gamma \mathbb{1}_{\{\gamma_i = 1\}}$. Let P_S be the probability measure on $([0,1]^n, \otimes_{i=1}^n \mathscr{B}([0,1]))$ induced by $(X_1, \dots, X_n)$.

Then $\{P_S : S \subseteq T \text{ finite}\}$ is consistent and the resulting probability measure on $([0,1]^{\mathscr{G}}, \mathscr{B}([0,1])^{\otimes \mathscr{G}})$ is called the Dirichlet process.

Note that if $A_1, \dots, A_n$ is already a disjoint partition of $\mathcal{X}$, then $(X_1, \dots, X_n) \sim \text{Dirichlet}(\alpha \{H(A_i)\}_{i=1}^n)$.

Example 4.5.5. Let $(\Omega, \mathscr{F})$ be a background probability space and let T be any set. Let $\mathcal{X}$ be a sub-interval of $\mathbb{R}$ and let $\mathscr{B}(\mathcal{X})$ be the Borel σ-field on $\mathcal{X}$.

For a function $X : \Omega \to \mathcal{X}^T$, for $t \in T$, define $X_t : \Omega \to \mathcal{X}$ by $X_t(\omega) = X(\omega)(t)$. Equivalently, $X_t = \pi_t \circ X$ where π_t is the projection on the t-th coordinate. Then, X is $\mathscr{F}/\mathscr{B}(\mathcal{X})^{\otimes T}$–measurable if and only if X_t is $\mathscr{F}/\mathscr{B}(\mathcal{X})$–measurable for all $t \in T$.

Let $Z_1, \dots, Z_n : \Omega \to \mathbb{R}$ be random variables ($\mathscr{F}/\mathscr{G}$–measurable). Let $T = \mathscr{B}(\mathbb{R})$. Define $P_n : \Omega \to [0,1]^{\mathscr{B}(\mathbb{R})}$ by

$$P_n(\omega)(A) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{Z_i(\omega) \in A\}, \quad \text{for } A \in \mathscr{B}(\mathbb{R}).$$

Therefore, P_n is $\mathscr{F}/\mathscr{B}([0,1])^{\mathscr{B}(\mathbb{R})}$–measurable and a valid random measure. We call P_n is the empirical measure of $Z_1, \dots, Z_n$.

Chapter 5

Convergence of Probability Measures

5.1 Weak Convergence

Definition 5.1.1. Let $(\mathcal{X}_d, \mathscr{F})$ be metric space with Borel σ-field $\mathscr{F}$. Let $P_1, P_2, \dots$ be probability measure on $(\mathcal{X}, \mathscr{F})$. We say P_n converges to a probability measure P weakly if $\forall f : \mathcal{X} \to \mathbb{R}$ continuous and bounded,

$$\lim_{n\to\infty} \int f \mathrm{d}P_n = \int f \mathrm{d}P. \tag{5.1}$$

We can also write $\lim_{n\to\infty} \mathbb{E}_{P_n} f = \mathbb{E}_P f$. Recall that $f : \mathcal{X} \to \mathbb{R}$ is continuous if $\forall x \in \mathcal{X}$, $\forall \varepsilon > 0$, $\exists \delta > 0$ such that $d(x, y) < \delta$ implies $|f(x) - f(y)| < \varepsilon$. f is bounded if $\sup_{X\in\mathcal{X}} |f(x)| < \infty$.

Remark 5.1.1. Let $(\Omega, \mathscr{F})$ be a general measure space and let $P_1, P_2, \dots$ be probability measures. We say P_n converges to P in total variation if $\lim_{n\to\infty} \mathrm{TV}(P_n, P) = 0$ (recall that $\mathrm{TV}(P_n, P) := \sup_{A\in\mathscr{F}} |P_n(A) - P(A)|$). Note then that

$$\begin{aligned}
\lim_{n\to\infty} \mathrm{TV}(P_n, P) = 0 &\iff \lim_{n\to\infty} \sup\{|P_n(A) - P(A)| : A \in \mathscr{F}\} = 0 \\
&\iff \lim_{n\to\infty} \sup\{|\mathbb{E}_{P_n} f - \mathbb{E}_P f| : f \in \Omega \to \mathbb{R} \text{ meas.}, \sup_{\omega\in\Omega} |f(\omega)| \le 1\} \to 0 \\
&\implies \text{weak convergence.}
\end{aligned}$$

To see this, first show that

$$\sup\{|P_n(A) - P(A)| : A \in \mathscr{F}\} = \sup\{|\mathbb{E}_{P_n} s - \mathbb{E}_p s| : s : \Omega \to \mathbb{R} \text{ simple}, 0 \le s \le 1\}.$$

We say $P_n \to P$ strongly if for any $A \in \mathscr{F}$, $\lim_{n\to\infty} P_n(A) = P(A)$, or equivalently, for any $f : \Omega \to \mathbb{R}$ Borel-measurable, bounded, $\lim_{n\to\infty} \mathbb{E}_{P_n} f = \mathbb{E}_P f$. Note total variation convergence implies strong convergence which implies weak convergence.

Convergence in total variation is too strict. For example, for $n \in \mathbb{N}$, let $X_1, X_2, \dots, X_n \overset{iid}{\sim} \mu = \mathrm{N}(0, 1)$ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Note that $\exists(\Omega, \mathscr{F}, P)$ measurable space such that $\forall i \in \mathbb{N}$, $X_i : \Omega \to \mathbb{R}$ (Borel-measurable) and the pushforward probability measures on $\mathbb{R}^{\mathbb{N}}$ induced by $(X_1, X_2, \dots)$ is a product of Normal distributions.

Let $\mathbb{P}_n$ denote the (random) empirical measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ induced by $\{X_1, \dots, X_n\}$, i.e., $\forall \omega \in \Omega$, $\forall A \in \mathscr{B}(\mathbb{R})$, $\mathbb{P}_n(\omega)(A) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i(\omega)\in A\}}$. Then $\forall A \in \mathscr{B}(\mathbb{R})$,

$$P\left(\lim_{n\to\infty} \mathbb{P}_n(A) = \mu(A)\right) = P\left(\left\{\omega \in \Omega : \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i(\omega)\in A\}} = \mu(A)\right\}\right) = 1.$$

by strong LLN so that $\mathbb{P}_n \to \mu$ strongly P-almost surely.

However, $\forall \omega \in \Omega$, define $A_\omega := \{X_1(\omega), \ldots, X_n(\omega)\} \in \mathscr{B}(\mathbb{R})$, then

$$\sup_{A \in \mathscr{B}(\mathbb{R})} |\mathbb{P}_n(\omega)(A) - \mu(A)| \geq |\mathbb{P}_n(\omega)(A_\omega) - \mu(A_\omega)| = 1.$$

Thus, $\forall n \in \mathbb{N}$, $\forall \omega \in \Omega$, $\mathrm{TV}(\mathbb{P}_n(\omega), \mu) = 1$ implies $P(\mathrm{TV}(\mathbb{P}_n, \mu) = 1) = P(\omega) = 1$.

But, strong convergence is also too strict. For $n \in \mathbb{N}$, let $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Ber}(\frac{1}{2})$. Again $\exists (\Omega, \mathscr{F}, P)$ such that $\forall i \in \mathbb{N}$, $X_i : \Omega \to \mathbb{R}$ (Borel-measurable). Let μ_n be the pushforward probability measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ induced by $\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \frac{1}{2}) : \Omega \to \mathbb{R}$. We will see that $\mu_n \to N(0, \frac{1}{4})$ weakly through CLT. But, let $A = \mathbb{Q}$. Then, whenever $\sqrt{n} \in \mathbb{N}$, we have that $\forall \omega \in \Omega$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i(\omega) - \frac{1}{2}) \in \mathbb{Q} \implies \mu_n(\mathbb{Q}) = 1.$$

However, $N(0, \frac{1}{4})(\mathbb{Q}) = 0$ since $\mathrm{Lebesgue}(\mathbb{Q}) = 0$ and $N(0, \frac{1}{4}) \ll \mathrm{Lebesgue}$. Thus, μ_n does not converge strongly to $N(0, \frac{1}{4})$.

As a final example, for $n \in \mathbb{N}$, let μ_n be a probability measure on $(\mathbb{R}^2, \mathscr{B}(\mathbb{R}^2))$ that is uniform on $[0,1] \times [0, \frac{1}{n}]$. Let μ be the probability measure on $(\mathbb{R}^2, \mathscr{B}(\mathbb{R}^2))$ that is uniform on $[0,1] \times \{0\}$. Then we will see that $\mu_n \to \mu$ weakly. But, μ_n does not converge to μ strongly: Let $A = [0,1] \times \{0\} \in \mathscr{B}(\mathbb{R}^2)$, then $\forall n \in \mathbb{N}$, $\mu_n(A) = 0$ whereas $\mu(A) = 1$.

Theorem 5.1.1 (Portmanteau). Let $(\mathcal{X}, \mathscr{F})$ be a metric space with Borel σ-Field $\mathscr{F}$. Let $P_1, \ldots, P_n$ and P be probability measure on $(\mathcal{X}, \mathscr{F})$. The following are equivalent:

(i) $\forall f : \mathcal{X} \to \mathbb{R}$ continuous and bounded, $\mathbb{E}_{P_n} f \to \mathbb{E}_P f$. (Weak convergence)

(ii) $\forall f : \mathcal{X} \to \mathbb{R}$ Lipschitz and bounded, $\mathbb{E}_{P_n} f \to \mathbb{E}_P f$.

(iii) $\forall U \subseteq \mathcal{X}$ open, $\liminf_{n\to\infty} P_n(U) \geq P(U)$.

(iv) $\forall V \subseteq \mathcal{X}$ closed, $\limsup_{n\to\infty} P_n(V) \leq P(V)$.

(v) $\forall C \subseteq \mathcal{X}$ such that $P(\partial C) = 0$, $\lim_{n\to\infty} P_n(C) = P(C)$.

- Recall f is Lipschitz if $\exists L > 0$ s.t. $\forall x, y \in \mathcal{X}$, $|f(x) - f(y)| \leq L d(x,y)$.
- Recall $\partial C := \bar{C} \backslash \mathrm{int}(C)$ is the boundary of C, where $\bar{C}$ is the closure of C, i.e. $\cap\{V : V \text{ closed}, V \supseteq A\}$, and $\mathrm{int}(C)$ is the interior of C, i.e. $\cup\{U : U \text{ open}, U \subseteq A\}$.

  We have $x \in \partial C$ iff $\forall \varepsilon > 0$, $B(x, \varepsilon) \cap C \neq \emptyset$ and $B(x, \varepsilon) \cap C^c \neq \emptyset$.

Proof.

(i) $\implies$ (ii) is obvious.

(ii) $\implies$ (iv): Let $V \subseteq \mathcal{X}$ be closed. Fix $m \in \mathbb{N}$, let $U_m := \{x \in \mathcal{X} : d(x, V) < \frac{1}{m}\}$ where $d(x,v) := \inf\{d(x,y) : y \in V\}$. Then $U_m = \cup_{x \in V} B(x, \frac{1}{m})$ is open and U_m^c is closed. Define $f : \mathcal{X} \to \mathbb{R}$ as $\forall x \in \mathcal{X}$, $f(x) := \min\{1, m \cdot d(x, U_m^c)\}$.

Note that $x \in V \implies d(x, U_m^c) \geq \frac{1}{m} \implies f(x) = 1$, $x \in U_m^c \implies f(x) = 0$, and that $\forall x, y \in \mathcal{X}$,

$$|f(x) - f(y)| \leq m |d(x, U_m^c) - d(y, U_m^c)| \leq m \cdot d(x,y)$$

is m-Lipschitz. Thus, we have $\forall n \in \mathbb{N}$, $P_n(V) = \int_{\mathcal{X}} \mathbb{1}_C \mathrm{d}P_n \leq \int_{\mathcal{X}} f \mathrm{d}P_n,$, and thus,

$$\begin{aligned}\limsup_{n\to\infty} P_n(V) \leq \limsup_{n\to\infty} \int_X f \mathrm{d}P_n &= \lim_{n\to\infty} \int_X f \mathrm{d}P_n \\ &= \int_{\mathcal{X}} f \mathrm{d}P \leq \int_{\mathcal{X}} 1 - \mathbb{1}_{U_m^c} \mathrm{d}P \leq P(U_m). \qquad (*)\end{aligned}$$

However, $\cap_{m=1}^{\infty} U_m = \cap_{m=1}^{\infty}\{x \in \mathcal{X} : d(x, V) < \frac{1}{m}\} = V$ as V is closed. Since $(*)$ is true for an arbitrary $m \in \mathbb{N}$, we have that $P(V) = \lim_{m\to\infty} P(U_m) \geq \limsup_{n\to\infty} P_n(V)$.

(iii) $\iff$ (iv): Obvious since U open implies U^c closed and $P(U) = 1 - P(U^c)$.

(iii + iv) $\implies$ (v): Let $C \in \mathcal{X}$ be such that $P(\partial C) = 0$.

$$\begin{aligned}
\limsup_{n\to\infty} P_n(C) \leq \limsup_{n\to\infty} P_n(\bar{C}) \leq P(\bar{C}) &= P(\bar{C}\setminus C) + P(C) \\
&\leq P(C) + P(\bar{C}\setminus \mathrm{int}(C)) = P(C), \\
\liminf_{n\to\infty} P_n(C) \geq \liminf_{n\to\infty} P_n(\mathrm{int}C) \geq P(\mathrm{int}(C)) &= P(C) - P(C\setminus \mathrm{int}(C)) \\
&\geq P(C) - P(\bar{C}\setminus \mathrm{int}(C)) = P(C), \\
\implies \lim_{n\to\infty} P_n(C) = P(C) \text{ as desired.}
\end{aligned}$$

(v) $\implies$ (i): Recall that if $f : \mathcal{X} \to \mathbb{R}$ is continuous, then $\forall a < b \in \mathbb{R}$, $f^{-1}((a, b)) \in \mathscr{F}$ is open. To see this, let $x_0 \in f^{-1}((a, b))$, then $\exists \varepsilon > 0$ such that $\forall x \in B(x_0, \varepsilon)$,

$$|f(x) - f(x_0)| \leq \min\{f(x_0) - a, b - f(x_0)\} \implies B(x_0, \varepsilon) \subseteq f^{-1}((a, b)).$$

Likewise, $f^{-1}([a, b])$ is closed.

Now, fix $f : \mathcal{X} \to \mathbb{R}$ continuous and bounded and let $a < -\inf_{x\in\mathcal{X}} f(x)$ and $b > \sup_{x\in X} f(x)$. Let μ be the pushforward measure on $([a, b], \mathscr{B}([a, b]))$ induced by P and f, i.e., $\forall B \in \mathscr{B}([a, b])$, $\mu(B) = P(f^{-1}(B))$. Since $\mu([a, b]) = 1$, then

$$\{t \in [a, b] : \mu(\{t\}) > 0\} = \bigcup_{k=1}^{\infty} \underbrace{\left\{t \in [a, b] : \mu(\{t\}) > \frac{1}{k}\right\}}_{\substack{\text{must be finite since} \\ \mu([a,b])=1}}$$

must be countable.

Fix $\varepsilon > 0$, we may thus find $m \in \mathbb{N}$ and $a = t_0 < t_1 < t_2 < \cdots < t_m = b$ such that

(1) $\forall i = 0, \ldots, m$, $\mu(\{t_i\}) = 0$,

(2) $\forall i = 1, \ldots, m$, $t_i - t_{i-1} \leq \varepsilon$.

Define, for $i \in [m]$, $A_i := \{x \in \mathcal{X} : t_{i-1} \leq f(x) < t_i\} = f^{-1}([t_{i-1}, t_i)) \in \mathscr{F}$. Then, $\cup_{i=1}^{m} A_i = \mathcal{X}$ and $\forall i \in [m]$, $\bar{A}_i \subseteq f^{-1}([t_{i-1}, t_i])$ and $\mathrm{int}(A) \supseteq f^{-1}((t_{i-1}, t_i))$. Note that, $\forall i \in [m]$,

$$\begin{aligned}
P(\partial A_i) &= P(\bar{A}_i \setminus \mathrm{int}(A_i)) \\
&\leq P(f^{-1}([t_{i-1}, t_i]) \setminus f^{-1}((t_{i-1}, t_i))) \\
&= P(\{x \in \mathcal{X} : f(x) = t_{i-1}\}) + P(\{x \in \mathcal{X} : f(x) = t_i\}) = 0. \qquad (**)
\end{aligned}$$

Define $g := \sum_{i=1}^{m} t_{i-1}\mathbb{1}_{A_i}$, note then that $\forall x \in \mathcal{X}$, $g(x) \leq f(x) \leq g(x) + \varepsilon$. Thus, $\forall n \in \mathbb{N}$,

$$\begin{aligned}
\left|\int f\mathrm{d}P - \int f\mathrm{d}P_n\right| &\leq \left|\int f\mathrm{d}P - \int g\mathrm{d}P\right| + \left|\int g\mathrm{d}P - \int g\mathrm{d}P_n\right| + \left|\int g\mathrm{d}P_n - \int f\mathrm{d}P_n\right| \\
&= 2\varepsilon P(X) + \left|\sum_{i=1}^{m} t_{i-1}(P(A_i) - P_n(A_i))\right| \\
&= 2\varepsilon + \sum_{i=1}^{m} |t_{i-1}||P(A_i) - P_n(A_i)|.
\end{aligned}$$

By $(**)$, we have $\lim_{n\to\infty} |\int f\mathrm{d}P - \int f\mathrm{d}P_n| \leq 2\varepsilon$. Since ε is arbitrary, we have $\lim_{n\to\infty} \int f\mathrm{d}P_n = \int f\mathrm{d}P$ as desired. $\square$

Remark 5.1.2. Weak convergence is also equivalent to

(a) $\forall f : \mathcal{X} \to \mathbb{R}$ upper semi-continuuous and $\sup\{f(x) : x \in \mathcal{X}\} < \infty$, $\limsup_{n\to\infty} \mathbb{E}_{P_n} f \leq \mathbb{E}_P f$.

(b) $\forall f : \mathcal{X} \to \mathbb{R}$ lower semi-continuous and $\inf\{f(x) : x \in \mathcal{X}\} > -\infty$, $\liminf_{n\to\infty} \mathbb{E}_{P_n} f \geq \mathbb{E}_P f$.

Recall f is upper semi-continuous if $\forall x_0 \in \mathcal{X}$ $\forall \varepsilon > 0$, $\exists \delta > 0$ s.t. $d(x_0, x) < \delta \implies f(x_0) - f(x) \geq -\varepsilon$, and lower semi-continuous if $\forall x_0 \in \mathcal{X}$ $\forall \varepsilon > 0$, $\exists \delta > 0$ s.t. $d(x_0, x) < \delta \implies f(x_0) - f(x) \leq \varepsilon$.

Theorem 5.1.2 (Continuous Mapping Theorem). Let $(\mathcal{X}, \mathscr{F})$, $(\mathcal{Y}, \mathscr{G})$ be measurable metric space and let $P_1, P_2, \ldots$ be probability measures on $(\mathcal{X}, \mathscr{F})$ such that $P_n \to P$ weakly for some probability measure P. Let $f : \mathcal{X} \to \mathcal{Y}$ be continuous, an let $P_n^{(f)}$ be the pushforward probability measure on $(\mathcal{Y}, \mathscr{G})$. Then $P_n^{(f)} \to P^{(f)}$ weakly.

Proof.

Let $g : \mathcal{Y} \to \mathbb{R}$ be continuous an bounded, then $g \circ f : \mathcal{X} \to \mathbb{R}$ is also continuous and bounded. Thus $\lim_{n\to\infty} \int_{\mathcal{X}} g \circ f \mathrm{d}P_n = \int_{\mathcal{X}} g \circ f \mathrm{d}P$. Since, $\forall n \in \mathbb{N}$, $\int_{\mathcal{Y}} g \mathrm{d}P_n^{(f)} = \int_{\mathcal{X}} g \circ f \mathrm{d}P_n$ and $\int_{\mathcal{Y}} g \mathrm{d}P^{(f)} = \int_{\mathcal{X}} g \circ f \mathrm{d}P_n$, the claim follows. □

5.2 Convergence of Random Variables

Remark 5.2.1. Given a background probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and measurable space $(\mathcal{X}, \mathscr{G})$, recall $X : \Omega \to \mathcal{X}$ is a random object (r.o.) if it is $\mathscr{F}/CG$–measurable, e.g., $\forall B \in \mathscr{G}$, $X^{-1}(B) \in \mathscr{F}$. For a random object X, we ~~say~~ refer to $\mathbb{P}^{(X)}$ as the distribution of X. Where $\mathbb{P}^{(X)} : \mathscr{G} \to [0, 1]$ is the pushforward measure induced by X if $\forall B \in \mathscr{G}$, $\mathbb{P}^{(X)}(B) = \mathbb{P}(X^{-1}(B))$.

For any countable sequence of random variables $X_1, X_2, \ldots$ with any infinite dimensional joint distribution P on $(\mathcal{X}^{\mathbb{N}}, \mathscr{G}^{\otimes\mathbb{N}})$ (specified completely by a consistent family of finite dimensional distributions), there always exists a background probability space $(\Omega, \mathscr{F}, \mathbb{P})$ such that

$$(X_1, X_2, \ldots) : \Omega \to \mathcal{X}^{\mathbb{N}}$$

is $\mathscr{F}/\mathscr{G}^{\otimes\mathbb{N}}$–measurable and that $P = \mathbb{P}^{(X_1, X_2, \ldots)}$.

To show that the marginal distributions $\mathbb{P}^{(X_1)}, \mathbb{P}^{(X_2)}, \ldots$ converge weakly, we may specify any joint distribution $\mathbb{P}^{(X_1, X_2, \ldots)}$ so long as the marginal distributions are fixed. Choosing a convenient joint distribution is called coupling.

Example 5.2.1. This is from Pollard, 2001, section 10.1. We will prove the following: $\forall n \in \mathbb{N}$, $\alpha \in [0, 1]$,

$$\mathrm{TV}(\mathrm{Bin}(n, \alpha), \mathrm{Poisson}(n\alpha)) \leq n\alpha^2. \tag{5.2}$$

We view $\mathrm{Bin}(n, \alpha)$ as probability measure on $(\mathbb{N}, 2^{\mathbb{N}})$.

We first observe the following: Let $\widetilde{P}, \widetilde{Q}$ be probability measure on measurable space $(\mathcal{X}, \mathscr{G})$. If there exists a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and $X, Y : \Omega \to \mathcal{X}$ such that $\widetilde{P}, \widetilde{Q}$ are the pushforward measure induced by X, Y, then, writing $B = \{\omega \in \Omega : X(\omega) \neq Y(\omega)\}$, we have $\forall \widetilde{A} \in \mathscr{G}$,

$$\begin{aligned}
|\widetilde{P}(\widetilde{A}) - \widetilde{Q}(\widetilde{A})| &= |\mathbb{P}(X^{-1}(\widetilde{A})) - \mathbb{P}(Y^{-1}(\widetilde{A}))| \\
&= |P(X^{-1}(\widetilde{A}) \cap B) - P(X^{-1}(\widetilde{A}) \cap B^c) - P(Y^{-1}(\widetilde{A}) \cap B) + P(Y^{-1}(\widetilde{A}) \cap B^c)| \\
&= |\mathbb{P}(X^{-1}(\widetilde{A}) \cap B) - \mathbb{P}(Y^{-1}(\widetilde{A}) \cap B)| \qquad (*) \\
&\leq \mathbb{P}(B) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \neq Y(\omega)\}) \\
\implies \mathrm{TV}(\widetilde{P}, \widetilde{Q}) &\leq \mathbb{P}(X \neq Y).
\end{aligned}$$

The $(*)$ holds since

$$\begin{aligned}X^{-1}(\widetilde{A}) \cap B^c &= \{\omega \in \Omega : X(\omega) \in \widetilde{A}, X(\omega) = Y(\omega)\}\\ &= \{\omega \in \Omega : X(\omega) \in \widetilde{A}, Y(\omega) \in \widetilde{A}, X(\omega) = Y(\omega)\}\\ &= \{\omega \in \Omega : Y(\omega) \in \widetilde{A}, X(\omega) = Y(\omega)\} = Y^{-1}(\widetilde{A}) \cap B^c.\end{aligned}$$

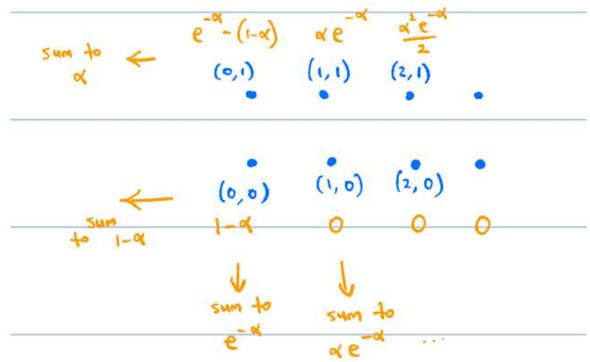


Figure 5.1: Coupling

We first couple Ber(α) and Poisson(α). Let $\Omega_1 = \mathbb{N} \times \mathbb{N}$, $\mathscr{F}_1 = 2^{\Omega_1}$, and let

$$\mathbb{P}_1((0,0)) = \min(1-\alpha, e^{-\alpha}) = 1-\alpha, \qquad \mathbb{P}_1((1,1)) = \min(\alpha, \alpha e^{-\alpha}) = \alpha e^{-\alpha},$$
$$\mathbb{P}_1((0,1)) = e^{-\alpha} - (1-\alpha), \qquad \mathbb{P}_1((1,0)) = 0,$$

$\forall k \in \mathbb{N}$, $k \geq 2$, let $\mathbb{P}_1((k,1)) = \frac{\alpha^k e^{-\alpha}}{k!}$ and $\mathbb{P}_1((k,0)) = 0$, and let $\mathbb{P}_1((k,\ell)) = 0$, $\forall \ell \geq 2$. Let $X_1, Y_1 : \Omega_1 \to \mathbb{N}$ be such that $\forall (n,m) \in \Omega_1$, $X_1(n,m) = n$, $Y_1(n,m) = m$. Then $X_1 \sim$ Poisson(α) and $Y_1 \sim$ Ber(α), and

$$\mathbb{P}_1(X_1 \neq Y_1) = 1 - \mathbb{P}_1((0,0)) - \mathbb{P}_1((1,1)) = \alpha - \alpha e^{-\alpha} = \alpha(1-e^{-\alpha}) \leq \alpha^2.$$

Now, for $i \in [n]$, let (X_i, Y_i) be iid with same distribution as (X_1, Y_1). Specifically, let $\Omega = (\mathbb{N} \times \mathbb{N})^n$, $\mathscr{F} = 2^\Omega$ and $\mathbb{P}$ be n products of $\mathbb{P}_1$.

$$\mathrm{TV}(\mathrm{Poisson}(n\alpha), \mathrm{Bin}(n,\alpha)) \leq \mathbb{P}\left(\sum_{i=1}^n X_i \neq \sum_{i=1}^n Y_i\right) \leq \sum_{i=1}^n \mathbb{P}(X_i \neq Y_i) = n\mathbb{P}_1(X_1 \neq Y_1) = n\alpha^2.$$

Thus, e.g., letting $P_n :=$ Bin$(n, \frac{1}{n})$, we have $P_n \to$ Poisson(1) in total variation.

Definition 5.2.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let $(\mathcal{X}, \mathscr{G})$ be a metric space with Borel σ-field. Let $X_1, X_2, X_3, \ldots X : \Omega \to \mathcal{X}$ be random objects.

(1) We say that $X_n \to X$ almost surely if $\mathbb{P}\big(\{\omega \in \Omega : \lim_{n\to\infty} d(X_n(\omega), X(\omega)) = 0\}\big) = 1$.

(2) We say that $X_n \to X$ in probability if $\forall \varepsilon > 0$, $\mathbb{P}\big(\{\omega \in \Omega : d(X_n(\omega), X(\omega)) > \varepsilon\}\big) \to 0$.

(3) We say that, for $p \geq 0$, $X_n \to X$ in L_p if $\int_\Omega d(X_n(\omega), X(\omega))^p \mathrm{d}\mathbb{P}(\omega) \to 0$.

(4) We say that $X_n \to X$ in distribution (in law) if $\mathbb{P}^{(X_n)} \to \mathbb{P}^{(X)}$ weakly, where $\mathbb{P}^{(X_n)}$, $\mathbb{P}^{(X)}$ on $(\mathcal{X}, \mathscr{G})$ are the pushforward measures. Equivalently, $X_n \to X$ in law if for any $f : \mathcal{X} \to \mathbb{R}$ that is continuous and bounded,
$$\mathbb{E}f(X_n) \to \mathbb{E}f(X).$$

Theorem 5.2.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let $(\mathcal{X}, \mathscr{G})$ be a metric space with Borel σ-field. Let $X_1, X_2, X_3, \dots X : \Omega \to \mathcal{X}$ be random objects.

(a) $X_n \to X$ a.s. $\implies$ $X_n \to X$ in prob. $\implies$ $X_n \to X$ in law.

(b) For any $p \geq 1$, $X_n \to X$ in L_p implies $X_n \to X$ in probability.

(c) $X_n \to x_0$ in law for some $x_0 \in \mathcal{X}$ implies $X_n \to x_0$ in probability.

(d) Suppose $\mathcal{X}$ and $\mathcal{Y}$ are separable metric spaces (recall that $\mathcal{X}$ is separable if there exists a countable subset $A \subseteq \mathcal{X}$ such that $\bar{A} = \mathcal{X}$). Let $Y_1, Y_2, \dots : \Omega \to \mathcal{Y}$, and suppose $Y_n \to \{y_0\}$ in law for some $y_0 \in \mathcal{Y}$; suppose also $X_n \to X$ in law. Then, $f(X_n, Y_n) \to f(X, y_0)$ in law for any $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ continuous.

A function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ is continuous if $\forall (x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$, $\forall \varepsilon > 0$, $\exists \delta > 0$ such that $\forall (x, y) \in \mathcal{X} \times \mathcal{Y}$, $d_{\mathcal{X}}(x, x_0) \vee d_{\mathcal{Y}}(y, y_0) \leq \delta$ implies $d_{\mathcal{X}}(f(x, y), f(x_0, y_0)) \leq \varepsilon$. Note that we can define metric on $\mathcal{X} \times \mathcal{Y}$ by $\tilde{d}((x, y), (x', y')) := d_{\mathcal{X}}(x, x') \vee d_{\mathcal{Y}}(y, y')$.

If $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ and are equipped with the Borel σ-field, then $+ : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $\times : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ are continuous. Thus, (d) implies that if $Y_n \to y_0 \in \mathbb{R}$ in law and $X_n \to X$ in law, $X_n + Y_n \to X_n + y_0$ in law and $X_n \cdot Y_n \to X_n y_0$ in law. This is known as Slutsky's theorem.

Proof.

(a) Suppose $X_n \to X$ a.s. Fix $\varepsilon > 0$, define $A_n = \{\omega \in \Omega : d(X_n(\omega), X(\omega)) > \varepsilon\}$ and define

$$\tilde{A}_n := \cup_{m=n}^{\infty} A_m = \left\{ \sup_{m \geq n} d(X_m, X) > \varepsilon \right\}$$
$$A := \{\omega \in \Omega : \limsup_{n \to \infty} d(X_n(\omega), X(\omega)) > \varepsilon\} = \cap_{n=1}^{\infty} \tilde{A}_n = \cap_{n=1}^{\infty} \cup_{m=n}^{\infty} A_m.$$

Then, we have that $X_n \to X$ a.s. if and only if for any $\varepsilon > 0$, $0 = \mathbb{P}(A) = \lim_{n \to \infty} \mathbb{P}(\tilde{A}_n)$. Since $\mathbb{P}(\tilde{A}_n) \geq \mathbb{P}(A_n)$, we have that $X_n \to X$ a.s. implies that $X_n \to X$ in probability.

Now suppose $X_n \to X$ in probability. Fix $f : \mathcal{X} \to \mathbb{R}$ such that $\exists L, B > 0$ such that $\forall x, y \in \mathcal{X}$, $|f(x) - f(y)| \leq L d(x, y)$ and $\sup_{x \in \mathcal{X}} |f(x)| \leq B$.

Fix $\varepsilon > 0$ and define $A_n = \{\omega \in \Omega : d(X_n(\omega), X(\omega)) > \varepsilon\}$ as before, then

$$\begin{aligned}
\left| \int_{\mathcal{X}} f \mathrm{d}\mathbb{P}^{(X_n)} - \int_{\mathcal{X}} f \mathrm{d}\mathbb{P}^{(X)} \right| &= \left| \int_{\Omega} f(X_n(\omega)) \mathrm{d}\mathbb{P}(\omega) - \int_{\Omega} f(X(\omega)) \mathrm{d}\mathbb{P}(\omega) \right| \\
&\leq \int_{\Omega} |f(X_n(\omega)) - f(X(\omega))| \, \mathrm{d}\mathbb{P}(\omega) \\
&\leq \int_{A_n} |f(X_n(\omega)) - f(X(\omega))| \, \mathrm{d}\mathbb{P}(\omega) + \int_{A_n^c} |f(X_n(\omega)) - f(X(\omega))| \, \mathrm{d}\mathbb{P}(\omega) \\
&\leq L\varepsilon \mathbb{P}(A_n^c) + 2B\mathbb{P}(A_n).
\end{aligned}$$

Thus $\lim_{n \to \infty} |\int_{\mathcal{X}} f \mathrm{d}\mathbb{P}^{(X_n)} - \int_{\mathcal{X}} f \mathrm{d}\mathbb{P}^{(X)}| \leq L\varepsilon$. Since ε is arbitrary, we have that $X_n \to X$ in law.

Remark 5.2.2. Note that $X_n \to X$ in law does not imply $X_n \to X$ in probability. Let $X : \Omega \to \mathbb{R}$ such that $\mathbb{P}^{(X)}$ is $N(0, 1)$. Let $X_n = (-1)^n X$. Then $X_n \to X$ in law but when n is odd,

$$\mathbb{P}(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon) = \mathbb{P}(\{\omega \in \Omega : 2|X(\omega)| > \varepsilon\})$$

and so $\limsup_{n \to \infty} \mathbb{P}(|X_n - X| > \varepsilon) > 0$. Intuitively, this is because convergence in probability depends on the joint distribution of (X_n, X) whereas convergence in law depends only on marginal distributions

$\mathbb{P}^{(X_n)}, \mathbb{P}^{(X)}$. Also $X_n \to X$ in probability does not imply $X_n \to X$ a.s. Let $\Omega = [0,1]$, $\mathscr{F} = \mathscr{B}([0,1])$, and $\mathbb{P}$ be uniform. Let $X : \Omega \to \{0,1\}$ be $\forall \omega \in \Omega$, $X(\omega) = 0$. For $n \in \mathbb{N}$, define, for $\omega \in \Omega$,

$$X_n(\omega) = \begin{cases} 1 & \text{if } \exists k \in \mathbb{N} \text{ s.t. } |\omega - \frac{k}{2^n}| \leq \frac{1}{4^n} \\ 0 & \text{else.} \end{cases}$$

Then, $\forall \varepsilon > 0$,

$$\mathbb{P}(\{\omega \in \Omega : |X(\omega) - X_n(\omega)| > \varepsilon\}) = \mathbb{P}(X_n \neq 0) \leq 2 \cdot \frac{2^n}{4^n} \to 0.$$

Thus $X_n \to X$ in probability. However, $\forall \omega \in \Omega$, $\limsup_{n\to\infty} X_n(\omega) = 1 \neq X(\omega)$. Thus,

$$\mathbb{P}(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 0.$$

Thus, X_n does not converge to X a.s. Intuitively, this is because almost sure convergence depends on the joint distribution $(X_n, X_{n+1}, \ldots, X)$ whereas convergence in probability depends only on the joint distribution (X_n, X).

(b) Continuing with the proof, suppose $X_n \to X$ in L_p. Fix $\varepsilon > 0$, then $A := \{\omega \in \Omega : d(X_n(\omega), X(\omega)) > \varepsilon\}$.

$$\begin{aligned}\mathbb{P}(A) \leq \int_A \frac{d(X_n(\omega), X(\omega))^p}{\varepsilon^p} \mathrm{d}\mathbb{P}(\omega) &\leq \frac{1}{\varepsilon^p} \int_\Omega d(X_n(\omega), X(\omega))^p \mathrm{d}\mathbb{P}(\omega) \\ &= \frac{\mathbb{E}_{\mathbb{P}}[d(X_n, X)^p]}{\varepsilon^p} \to 0.\end{aligned}$$

Thus $X_n \to X$ in probability.

(c) Suppose $X_n \to x_0$ in law. Fix $\varepsilon > 0$ and define $A_{n,\varepsilon} := \{\omega \in \Omega : d(x_0, X_n(\omega)) > \varepsilon\}$. Define $g : \mathcal{X} \to \mathbb{R}$ as $g(x) = \frac{d(x,x_0)}{\varepsilon} \wedge 1$, so that g is Lipschitz and bounded. Then,

$$\begin{aligned}\mathbb{P}(A_{n,\varepsilon}) &\leq \int_{A_{n,\varepsilon}} \frac{d(X_n(\omega), x_0)}{\varepsilon} \wedge 1 \mathrm{d}\mathbb{P}(\omega) \\ &\leq \int_\Omega g(X_n(\omega)) - g(x_0) \mathrm{d}\mathbb{P}(\omega) = \int_{\mathcal{X}} g \mathrm{d}\mathbb{P}^{(X_n))} - \int_{\mathcal{X}} g \mathrm{d}\mathbb{P}^{(x_0)} \to 0.\end{aligned}$$

(d) Consider, $\forall n \in \mathbb{N}$, $(X_n, Y_n) : \Omega \to \mathcal{X} \times \mathcal{Y}$. Let $\mathscr{G}, \tilde{\mathscr{G}}$ be the Borel σ-field of $\mathcal{X}$ and $\mathcal{Y}$ respectively. Recall that we may define metric $\tilde{d}((x,y),(x',y')) = d_{\mathcal{X}}(x,x') \vee d_{\mathcal{Y}}(y,y')$ on $\mathcal{X} \times \mathcal{Y}$. Let $\mathscr{H}$ be the Borel σ-field on $\mathcal{X} \times \mathcal{Y}$ with respect to this metric. We claim that $\mathscr{G} \otimes \tilde{\mathscr{G}} = \mathscr{H}$ so that (X_n, Y_n) is $\mathscr{F}/\mathscr{G}$–measurable and $\mathbb{P}^{(X_n,Y_n)}$ is defined in $\mathscr{G}$. (This claim requires the separability of $\mathcal{X}, \mathcal{Y}$)

Let us first assume this and prove that $(X_n, Y_n) \to (X, y_0)$ in law. Let $g : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be L-Lipschitz and bounded by $B > 0$. Then

$$\begin{aligned}\left|\int_{\mathcal{X}\times\mathcal{Y}} g \mathrm{d}\mathbb{P}^{(X_n,Y_n)} - \int_{\mathcal{X}\times\mathcal{Y}} g \mathrm{d}\mathbb{P}^{(X,y_0)}\right| &= \left|\int_\Omega g(X_n, Y_n) \mathrm{d}\mathbb{P} - \int_\Omega g(X, y_0) \mathrm{d}\mathbb{P}\right| \\ &\leq \underbrace{\left|\int_\Omega g(X_n, Y_n) - g(X_n, y_0) \mathrm{d}\mathbb{P}\right|}_{\text{term 1}} + \underbrace{\left|\int_\Omega g(X_n, y_0) - g(X, y_0) \mathrm{d}\mathbb{P}\right|}_{\text{term 2}}.\end{aligned}$$

Now, since g is Lipschitz, $\exists L > 0$ such that $\forall x, y \in \mathcal{X}$,

$$|g(x,y) - g(x,y_0)| \leq L\tilde{d}((x,y),(x,y_0)) = Ld_{\mathcal{Y}}(y, y_0).$$

Moreover, $|g(x,y) - g(x,y_0)| \leq 2B$. Hence,

$$\text{term } 1 = \left| \int_\Omega g(X_n, Y_n) - g(X_n, y_0) \mathrm{d}\mathbb{P} \right| \leq \int_\Omega (L d_\mathcal{Y}(Y_n, y_0)) \wedge 2B \, \mathrm{d}\mathbb{P} \to 0,$$

by the fact that $Y_n \xrightarrow{d} y_0$. By a similar argument, we may show that term $2 \to 0$. Claim then follows by the continuous mapping theorem.

Now let us prove the claim that $\mathscr{G} \otimes \tilde{\mathscr{G}} = \mathscr{H}$. We first note that since $\mathcal{X}, \mathcal{Y}$ are separable, we may show that $\mathscr{G}, \tilde{\mathscr{G}}$ are the σ-field generated by the set of open balls (see Proposition 10.1.1 for a proof). Now, consider an open ball in $\mathcal{X} \times \mathcal{Y}$ with metric $\tilde{d}$: for $(x,y) \in \mathcal{X} \times \mathcal{Y}$ and $r > 0$, we have

$$\begin{aligned} B((x,y),r) &= \{(x',y') : d_\mathcal{X}(x,x') \vee d_\mathcal{Y}(y,y') \leq r\} \\ &= B(x,r) \times B(y,r). \end{aligned}$$

Therefore, $B((x,y),r) \in \mathscr{G} \otimes \tilde{\mathscr{G}}$. Since $\mathcal{X}, \mathcal{Y}$ is also separable (with separating set as $A \times B$ where A, B are the countable separating set for $\mathcal{X}, \mathcal{Y}$ respectively), we have that $\mathscr{H}$ is generated by the open balls and hence $\mathscr{H} \subset \mathscr{G} \otimes \tilde{\mathscr{G}}$. Since projections are continuous with respect to the $\tilde{d}$ metric, they are also $\mathscr{F}/\mathscr{H}$–measurable. Thus, we have that $\mathscr{H} = \mathscr{G} \otimes \tilde{\mathscr{G}}$.

□

If $Y_n \to Y$ in law and $Y \neq y_0$, then (d) may not hold, depending on the coupling (X_n, Y_n), (X, Y). For example, let $X_n = Y_n$ and suppose $\mathbb{P}^{(X_n)} = N(0,1)$ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Let $X = -Y$ and suppose $\mathbb{P}^{(X)} = N(0,1)$. Then $\mathbb{P}^{(X_n+Y_n)} = N(0,2)$ and $\mathbb{P}^{(X+Y)} = \{0\}$.

Theorem 5.2.2 (Skorohod's Theorem). Let $(\mathcal{X}, \mathscr{G})$ be a separable measurable space. Let $P_1, P_2, \ldots, P$ be probability measures on $(\mathcal{X}, \mathscr{G})$ such that $P_n \to P$ weakly. Then $\exists(\Omega, \mathscr{F}, \mathbb{P})$ and $X_1, X_2, \ldots, X : \Omega \to \mathcal{X}$ s.t. $X_n \to X$ almost surely.

Proof.

See Pollard, 2001, section 10.2. □

Definition 5.2.2. Let $(\mathcal{X}, \mathscr{G})$ be a measurable metric space and let P, Q be probability measure on $(\mathcal{X}, \mathscr{G})$. For $A \subseteq \mathcal{X}$, $\varepsilon > 0$, define $A_\varepsilon := \{x \in \mathcal{X} : d(x, A) < \varepsilon\}$. Define Levy-Prokhorov distance

$$d_{\mathrm{LP}}(P,Q) := \inf\{\varepsilon > 0 : \forall A \in \mathscr{G}, P(A) \leq Q(A_\varepsilon) + \varepsilon, Q(A) \leq P(A_\varepsilon) + \varepsilon\}.$$

If $P(A) \leq Q(A_\varepsilon) + \varepsilon$ and $Q(A_\varepsilon) \leq P'(A_{\varepsilon+\varepsilon'}) + \varepsilon'$, then $P(A) \leq P'(A_{\varepsilon+\varepsilon'}) + (\varepsilon + \varepsilon')$ implies $d_{\mathrm{LP}}(P, P') \leq d_{\mathrm{LP}}(P,Q) + d_{\mathrm{LP}}(Q, P')$.

Corollary 5.2.1. Let $(\mathcal{X}, \mathscr{G})$ be a separable metric measurable space. Let $P_1, P_2, P_3, \ldots, P$ be probability measures on $(\mathcal{X}, \mathscr{G})$. Then $P_n \to P$ weakly iff $d_{\mathrm{LP}}(P_n, P) \to 0$. We say that d_{LP} metrizes weak convergence on separable metric spaces.

Proof.

Suppose $d_{\mathrm{LP}}(P_n, P) \to 0$ and let $A \subseteq \mathcal{X}$ be such that A is closed. For any $m \in \mathbb{N}$, $\exists n_m \in \mathbb{N}$ such that $\forall n \geq n_m$,

$$P_n(A) \leq P(A_{\frac{1}{m}}) + \frac{1}{m} \implies \limsup_{n\to\infty} P_n(A) \leq P(A_{\frac{1}{m}}) + \frac{1}{m}.$$

Since m is arbitrary and $\lim_{m\to\infty} P(A_{\frac{1}{m}}) = P(A)$, we have that $P_n \to P$ weakly.

Now suppose $P_n \to P$ weakly. By Theorem 5.2.2, $\exists(\Omega, \mathscr{F}, \mathbb{P})$ and $X_1, X_2, \ldots, X : \Omega \to \mathcal{X}$ such that $X_n \to X$ a.s. Hence, $X_n \to X$ in probability. Fix $\varepsilon > 0$, then $\exists n_\varepsilon \in \mathbb{N}$ such that $\forall n \geq n_\varepsilon$, $\mathbb{P}(d(X_n, X) \geq \varepsilon) < \varepsilon$.

Thus, for any $B \in \mathscr{G}$, we have

$$\begin{aligned} P_n(B) = \mathbb{P}(X_n \in B) &\leq \mathbb{P}(X_n \in B, d(X_n, X) < \varepsilon) + \mathbb{P}(d(X_n, X) \geq \varepsilon) \\ &\leq \mathbb{P}(X \in B_\varepsilon) + \varepsilon = P(B_\varepsilon) + \varepsilon. \end{aligned}$$

We may use the same argument to show that $P(B) \leq P_n(B_\varepsilon) + \varepsilon$. This implies that for all $n \geq n_\epsilon$, $d_{LP}(P_n, P) \leq \epsilon$. Hence, $d_{\mathrm{LP}}(P, P_n) \to 0$. □

Lemma 5.2.1 (Egorov's Theorem). Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $(\mathcal{X}, \mathscr{G})$ be metric measurable space. Suppose $X_1, X_2, \dots, X : \Omega \to \mathcal{X}$ are r.o. and $X_n \to X$ a.s., then $\forall \varepsilon > 0$, $\exists B \in \mathscr{F}$ such that $P(B^c) < \varepsilon$ and $\lim_{n\to\infty} \sup_{\omega \in B} d(X_n(\omega), X(\omega)) = 0$.

Proof.

For $n, k \in \mathbb{N}$, define $E_n^{(k)} := \cup_{m \geq n}\{\omega \in \Omega : d(X_m(\omega), X(\omega)) \geq \frac{1}{k}\}$, note that $E_1^{(k)} \supseteq E_2^{(k)} \supseteq \dots$. Then, since $X_n \to X$ a.s. ,

$$0 = \mathbb{P}(\cap_{n=1}^\infty E_n^{(k)}) = \lim_{n\to\infty} \mathbb{P}(E_n^{(k)}).$$

Fix $\varepsilon > 0$, then there exists $n_k \in \mathbb{N}$ such that $\mathbb{P}(E_{n_k}^{(k)}) \leq \frac{\varepsilon}{2^k}$. Let $B = \cap_{k=1}^\infty E_{n_k}^{(k)c}$, then

$$\mathbb{P}(B^c) \leq \sum_{k=1}^\infty \mathbb{P}(E_{n_k}^{(k)}) \leq \sum_{k=1}^\infty \frac{\varepsilon}{2^k} \leq \varepsilon.$$

For any $k \in \mathbb{N}$, $\forall n \geq n_k$,

$$\sup\left\{d(X_n(\omega), X(\omega)) : \omega \in B \subseteq {E_{n_k}^{(k)}}^c\right\} < \frac{1}{k}.$$

Thus, $\lim_{n\to\infty} \sup\{d(X_n(\omega), X(\omega)) : \omega \in B\} = 0$. □

Lemma 5.2.2. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $(\mathcal{X}, \mathscr{G})$ be metric measurable space. Suppose $X_1, X_2, \dots, X : \Omega \to \mathcal{X}$ are r.o. and $X_n \to X$ in probability, then there exists subsequence $\{n_1, n_2, \dots\}$ s.t. $X_{n_k} \to X$ a.s.

Proof.

For all $k \in \mathbb{N}$, $\exists n_k \in \mathbb{N}$ such that

$$\mathbb{P}(d(X_{n_k}, X) > 2^{-k}) \leq 2^{-k}.$$

Fix $\varepsilon > 0$. For $k \in \mathbb{N}$, define $E_k := \{\omega \in \Omega : d(X_{n_k}(\omega), X(\omega)) > \varepsilon\}$. Let k_0 be such that $2^{-k_0} < \varepsilon$. Then, $\forall k > k_0$, $\mathbb{P}(\mathbb{E}_k) \leq 2^{-k}$. Then, $\mathbb{P}(\cap_{m=1}^\infty \cup_{k=m}^\infty E_k) \leq \mathbb{P}(\cap_{m=k_0}^\infty \cup_{k=m}^\infty E_k) = 0$ by Borel-Cantelli lemma, which implies

$$\mathbb{P}(\limsup_{k\to\infty} d(X_{n_k}, X) > \varepsilon) = 0.$$

Since $\varepsilon > 0$ is arbitrary,

$$\mathbb{P}(\limsup_{k\to\infty} d(X_{n_k}, X) > 0) = \mathbb{P}\left(\bigcup_{m=1}^\infty \left\{\limsup_{k\to\infty} d(X_{n_k}, X) > \frac{1}{m}\right\}\right) = 0.$$

So $X_{n_k} \to X$ a.s. □

Definition 5.2.3 (Uniform Integrability). Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $X_1, X_2, \dots : \Omega \to \mathbb{R}$ be r.v. We say $\{X_n\}_{n=1}^\infty$ is <u>uniformly integrable</u> (UI) if

(a) $\sup_{n\in\mathbb{N}} \mathbb{E}|X_n| < \infty$ and

(b) $\forall \varepsilon > 0$, $\exists \delta > 0$ such that for all $n \in \mathbb{N}$, $\forall E \in \mathscr{F}$ where $\mathbb{P}(E) < \delta$, we have $\int_E |X_n| \mathrm{d}\mathbb{P} < \varepsilon$. Equivalently,

$$\lim_{\delta\to 0} \sup_{B\,:\,\mathbb{P}(B)<\delta} \sup_{n\in\mathbb{N}} \int_B |X_n| d\mathbb{P} = 0.$$

Lemma 5.2.3. $\{X_n\}_{n=1}^{\infty}$ is UI iff

$$\lim_{K\to\infty} \sup_{n\in\mathbb{N}} \int_{\{|X_n|>K\}} |X_n| \mathrm{d}\mathbb{P} = 0, \tag{5.3}$$

i.e.,

$$\lim_{K\to\infty} \sup_{n\in\mathbb{N}} \left\{ \mathbb{E}|X_n| - \int_{-K}^{K} |x| \mathrm{d}\mathbb{P}^{(X_n)}(x) \right\} = 0.$$

Proof.

Suppose $\{X_n\}_{n=1}^{\infty}$ is uniform integrable. Fix $\varepsilon > 0$ and let $\delta > 0$ be such that for all $A \in \mathscr{F}$ satisfying $\mathbb{P}(A) \leq \delta$ and for all $n \in \mathbb{N}$, we have $\int_A |X_n| d\mathbb{P} \leq \epsilon$. Since $\exists M > 0$ such that $\sup_{n\in\mathbb{N}} \mathbb{E}|X_n| < M$, we have $\forall K \geq \frac{M}{\delta}$, $\forall n \in \mathbb{N}$, $\mathbb{P}(\{\omega \in \Omega : |X_n(\omega)| > K\}) < \frac{M}{K} \leq \delta$ which implies

$$\sup_{n\in\mathbb{N}} \int_{\{|X_n|>K\}} |X_n| \mathrm{d}\mathbb{P} < \varepsilon.$$

Since ε is arbitrary, $\lim_{K\to\infty} \sup_{n\in\mathbb{N}} \int_{\{|X_n|>K\}} |X_n| \mathrm{d}\mathbb{P} = 0$. Now assume (5.3) and fix $\varepsilon > 0$. $\exists K \in \mathbb{N}$ such that $\sup_{n\in\mathbb{N}} \int_{\{|X_n|>K\}} |X_n| \mathrm{d}\mathbb{P} < \frac{\varepsilon}{2}$. Choose $\delta \leq \frac{1}{K}\frac{\varepsilon}{2}$, then $\forall E \in \mathscr{F}$ such that $\mathbb{P}(E) < \delta$ and for any $n \in \mathbb{N}$, we have

$$\begin{aligned}
\int_E |X_n| \mathrm{d}\mathbb{P} \leq & \int_{|X_n|>K} |X_n| \mathrm{d}\mathbb{P} + \int_{E\cap\{|X_n|<K\}} |X_n| \mathrm{d}\mathbb{P} \\
\leq & \frac{\varepsilon}{2} + K \cdot P(E) \\
< & \frac{\varepsilon}{2} + K\delta \leq \varepsilon.
\end{aligned}$$

Now, fix $\varepsilon = 1$, then $\exists K \in \mathbb{N}$ such that $\sup_{n\in\mathbb{N}} \int_{\{|X_n|>K\}} |X_n| \mathrm{d}\mathbb{P} < 1$. Then $\forall n \in \mathbb{N}$,

$$\int_\Omega |X_n| \mathrm{d}\mathbb{P} = \int_{\{|X_n|>K\}} |X_n| \mathrm{d}\mathbb{P} + K \cdot \mathbb{P}(\{|X_n| \leq K\}) \leq 1 + K.$$

Therefore, $\{X_n\}_{n\in\mathbb{N}}$ is uniformly integrable. □

Remark 5.2.3. If $n \in \mathbb{N}$ and $X_1, \ldots, X_n$ satisfy $\mathbb{E}|X_i| < \infty$, $\forall i \in [n]$, then $\{X_1, \ldots, X_n\}$ is UI. To see this, note that $\forall i \in [n]$,

$$\lim_{K\to\infty} \int_{\{|X_i|>K\}} |X_i| \mathrm{d}\mathbb{P} = \mathbb{E}|X_i| - \lim_{K\to\infty} \int_{\{|X_i|\leq K\}} |X_i| \mathrm{d}\mathbb{P} = 0.$$

(Note $B \mapsto \int_B |X_i| \mathrm{d}\mathbb{P}$ is a finite measure since $\mathbb{E}|X_i| < \infty$. Thus, $\lim_{K\to\infty} \max_{i\in[n]} \int_{\{|X_i|>K\}} |X_i| \mathrm{d}\mathbb{P} = 0$.

Now let $X_1, X_2, \cdots : \Omega \to \mathbb{R}$ be random variables. If $\exists Y : \Omega \to \mathbb{R}$ such that $\mathbb{E}|Y| < \infty$ and $\forall \omega \in \Omega$, $|X_n(\omega)| \leq |Y(\omega)|$, then

$$\lim_{K\to\infty} \sup_{n\in\mathbb{N}} \int_{\{|X_n|>K\}} |X_n| \mathrm{d}\mathbb{P} = \lim_{K\to\infty} \int_{\{|Y|>K\}} |Y| \mathrm{d}\mathbb{P} = 0.$$

Thus, domination implies U.I. We will see that U.I. is necessary and sufficient for convergence of integrals.

Theorem 5.2.3. Let $p \geq 1$. Let $X_1, X_2, \ldots, X : \Omega \to \mathbb{R}$ be random variables. Suppose $X_n \to X$ a.s. and suppose $\mathbb{E}|X_n|^p < \infty$ $\forall n \in \mathbb{N}$. Then the following are equivalent:

(i) $\{|X_n|^p\}_{n=1}^\infty$ is U.I.

(ii) $X_n \to X$ in L_p.

(iii) $\mathbb{E}|X_n|^p \to \mathbb{E}|X|^p < \infty$.

Moreover, if $X_n \to X$ in law and $\{|X_n|^p\}_{n=1}^\infty$ is U.I., then $\mathbb{E}|X_n|^p \to \mathbb{E}|X|^p$.

Proof.

Suppose $X_n \to X$ a.s. and suppose $\mathbb{E}|X_n|^p < \infty$ $\forall n \in \mathbb{N}$.

(i) $\implies$ (ii) Assume $\{|X_n|^p\}_{n=1}^\infty$ is U.I. Fix $\varepsilon > 0$, then $\exists \delta_\varepsilon > 0$ such that $\forall E \in \mathscr{F}$ where $\mathbb{P}(E) < \delta_\varepsilon$ and for all $n \in \mathbb{N}$, $\int_E |X_n|^p \mathrm{d}\mathbb{P} < \frac{\varepsilon}{3 \cdot 2^p}$.

By Lemma 5.2.1, $\exists B \in \mathscr{F}$, $\mathbb{P}(B^c) < \delta_\varepsilon$ and $n_\varepsilon \in \mathbb{N}$ such that $\forall n \in n_\varepsilon$, $\sup_{\omega \in B} |X_n(\omega) - X(\omega)|^p < \frac{\varepsilon}{3}$.

Using U.I., we have that $\sup_{n \in \mathbb{N}} \int_{B^c} |X_n|^p \mathrm{d}\mathbb{P} < \frac{\varepsilon}{3 \cdot 2^p}$ and that, by Fatou's lemma,

$$\int_{B^c} |X|^p \mathrm{d}\mathbb{P} \leq \liminf_{n \to \infty} \int_{B^c} |X_n|^p \mathrm{d}\mathbb{P} \leq \frac{\varepsilon}{3 \cdot 2^p}.$$

Thus, $\forall n \geq n_\varepsilon$,

$$\begin{aligned}
\int_\Omega |X_n - X|^p \mathrm{d}\mathbb{P} &= \int_{B^c} |X_n - X|^p \mathrm{d}\mathbb{P} + \int_B |X_n - X|^p \mathrm{d}\mathbb{P} \\
&\leq \int_{B^c} (|X_n| + |X|)^p \mathrm{d}\mathbb{P} + \frac{\varepsilon}{3} \cdot \mathbb{P}(B) \\
&\leq 2^p \int_{B^c} |X_n|^p \mathrm{d}\mathbb{P} + 2^p \int_{B^c} |X|^p \mathrm{d}\mathbb{P} + \frac{\varepsilon}{3} \leq \varepsilon.
\end{aligned}$$

So $X_n \to X$ in L_p.

(ii) $\implies$ (iii) Note when $p = 1$, $\mathbb{E}|X_n| - \mathbb{E}|X| \leq \mathbb{E}|X_n - X|$. For a general $p \geq 1$, note that $(\mathbb{E}|X_n|^p)^{1/p} \leq (\mathbb{E}|X|^p)^{1/p} + (\mathbb{E}|X - X_n|^p)^{1/p}$ (Minkowski inequality). Thus,

$$\begin{aligned}
&|(\mathbb{E}|X_n|^p)^{\frac{1}{p}} - (\mathbb{E}|X|^p)^{\frac{1}{p}}| \leq (\mathbb{E}|X - X_n|^p)^{\frac{1}{p}} \to 0 \\
\implies &|\mathbb{E}|X_n|^p - \mathbb{E}|X|^p| \to 0.
\end{aligned}$$

(iii) $\implies$ (i): Fix $\varepsilon > 0$. Then $\exists n_0 \in \mathbb{N}$ such that $\forall n \geq n_0$, $|\mathbb{E}|X_n|^p - \mathbb{E}|X|^p| < \delta/3$. For $K \in \mathbb{N}$, define $\phi_K : [0, \infty) \to [0, \infty)$ as

$$\phi_K(x) = \begin{cases} x & \text{if } x \leq K - 1, \\ 0 & \text{if } x > K, \\ (K-1)(K-x) & \text{else}, \end{cases}$$

so that $\forall x > 0$, $\phi_K(x) \leq x \mathbb{1}_{\{x \leq K\}} \leq x$. Note that $\forall \omega \in \Omega$, $\lim_{K \to \infty} \phi_K(|X(\omega)|^p) = |X(\omega)|^p$, and that $\forall K \in \mathbb{N}$, $\phi_K(|X|^p) \leq |X|^p$, and $\mathbb{E}|X|^p < \infty$. Thus, by Dominated Convergence Theorem, $\exists K_0 \in \mathbb{N}$ such that $\forall K \geq K_0$,

$$\left|\mathbb{E}_p |X|^p - \mathbb{E}\phi_K(|X|^p)\right| \leq \frac{\varepsilon}{3}.$$

Now observe that ϕ_{K_0} is continuous and bounded, thus, since $X_n \to X$ a.s., $|X_n|^p \to |X|^p$ a.s. and $|X_n|^p \to |X|^p$ in law, $\exists n_1 \in \mathbb{N}$ such that $\forall n \geq n_1$,

$$\left|\mathbb{E}\phi_{K_0}(|X_n|^p) - \mathbb{E}\phi_{K_0}(|X|^p)\right| < \frac{\varepsilon}{3}.$$

Thus, we have, $\forall n \geq n_0 \vee n_1$,

$$\begin{aligned}\int_{\{|X_n|^p > K_0\}} |X_n|^p \mathrm{d}\mathbb{P} &= \int_\Omega |X_n|^p \mathrm{d}\mathbb{P} - \int_{\{|X_n|^p \leq K_0\}} |X_n|^p \mathrm{d}\mathbb{P} \\ &= \mathbb{E}|X_n|^p - \mathbb{E}\phi_{K_0}(|X_n|^p) \\ &\leq \frac{\varepsilon}{3} + \mathbb{E}|X|^p - \mathbb{E}\phi_{K_0}(|X|^p) + \frac{\varepsilon}{3} < \varepsilon.\end{aligned}$$

Thus, $\forall K \geq K_0$, we have

$$\sup_{n \geq n_1 \vee n_0} \int_{\{|X_n|^p > K\}} |X_n|^p \mathrm{d}\mathbb{P} < \sup_{n \geq n_1 \vee n_0} \int_{\{|X_n|^p > K_0\}} |X_n|^p \mathrm{d}\mathbb{P} < \varepsilon.$$

By Remark 5.2.3, $\exists K_1 \in \mathbb{N}$ such that $\forall K \geq K_1$, $\sup_{n < n_1 \vee n_0} \int_{\{|X_n|^p > K\}} |X_n|^p \mathrm{d}\mathbb{P} < \varepsilon$.

Hence, $\forall K \geq K_0 \vee K_1$, $\sup_{n \in \mathbb{N}} \int_{\{|X_n|^p > K\}} |X_n|^p \mathrm{d}\mathbb{P} < \varepsilon$. Since $\varepsilon > 0$ is arbitrary,

$$\lim_{K \to \infty} \sup_{n \in \mathbb{N}} \int_{\{|X_n|^p > K\}} |X_n|^p \mathrm{d}\mathbb{P} = 0.$$

Now, suppose $X_n \to X$ in law and $\{|X_n|^p\}_{n=1}^\infty$ is U.I. By Theorem 5.2.2, there exists $(\Omega', \mathscr{F}', \mathbb{Q})$ and $Y_1, Y_2, \ldots, Y : \Omega' \to \mathcal{X}$ such that $\mathbb{Q}^{(Y_n)} = \mathbb{P}^{(X_n)}$, $\forall n \in \mathbb{N}$, $\mathbb{Q}^{(Y)} = \mathbb{P}^{(X)}$, and $Y_n \to Y$ a.s. By lemma 5.2.3, $\{|Y_n|\}_{n=1}^\infty$ is U.I. and thus $\forall n \in \mathbb{N}$,

$$\mathbb{E}_\mathbb{P}|X_n|^p = \int_\mathbb{R} |x|^p \mathrm{d}\mathbb{P}^{(X_n)}(x) = \int_\mathbb{R} |x|^p \mathrm{d}\mathbb{Q}^{(Y_n)}(x) = \mathbb{E}_\mathbb{Q}|Y_n|^p \to \mathbb{E}_\mathbb{Q}|Y|^p = \mathbb{E}_\mathbb{P}|X|^p.$$

$\square$

Remark 5.2.4. Let $(\mathcal{X}, \mathscr{G})$ be a metric space and let $X_1, X_2, \ldots, X : \Omega \to \mathcal{X}$. Suppose $X_n \to X$ a.s. Then, $\forall p \geq 1$, $\mathbb{E}_\mathbb{P} d(X_n, X)^p \to$ if and only if $\exists x_0 \in \mathcal{X}$ such that $\{d(X_n, x_0)^p\}_{n=1}^\infty$ is U.I. Note that if $\exists x_0 \in \mathcal{X}$ such that

$$\lim_{K \to \infty} \sup_{n \in \mathbb{N}} \int_{B(x_0, K)^c} d(x, x_0)^p \mathrm{d}\mathbb{P}^{(X)}(x) = 0.$$

Let $\widetilde{X} \in \mathcal{X}$ and let $D = d(x, x_0)$. Then, $\forall n \in \mathbb{N}$, $\forall K \geq 2D$,

$$\begin{aligned}\int_{B(\tilde{x}, K)^c} d(x, \tilde{x})^p \mathrm{d}\mathbb{P}^{(X_n)}(x) &\leq \int_{B(x_0, K-D)^c} (d(x, x_0) + D)^p \mathrm{d}\mathbb{P}^{(X_n)}(x) \\ &\leq 2^p \int_{B(x_0, K-D)^c} d(x, x_0)^p \mathrm{d}\mathbb{P}^{(X_n)}(x).\end{aligned}$$

So $\lim_{K \to \infty} \sup_{n \in \mathbb{N}} \int_{B(\tilde{x}, K)^c} d(x, \tilde{x})^p \mathrm{d}\mathbb{P}^{(X_n)}(x) \to 0$ as well. Thus, the choice of $x_0 \in \mathcal{X}$ is unimportant.

Definition 5.2.4. Let P, Q be probability measure on $(\mathcal{X}, \mathscr{G})$. Let $\Omega = \mathcal{X} \times \mathcal{X}$ and $\mathscr{F} = \mathscr{G} \otimes \mathscr{G}$. Define the set of couplings as

$$\mathcal{C}(P, Q) := \{\mathbb{P} \text{ on } (\Omega, \mathscr{F}) : \text{ projections } X = \pi_1, Y = \pi_2 \text{ satiisfy } P = \mathbb{P}^{(X)} \text{ and } Q = \mathbb{P}^{(Y)}\}.$$

Let $p \geq 1$, define Wasserstein/Earth-mover/Monge-Kantorovich/Optimal Transport distance as

$$d_{W_p}(P, Q) = \inf_{\mathbb{P} \in \mathcal{C}(P, Q)} \mathbb{E}_\mathbb{P}[d(X, Y)^p]^{\frac{1}{p}},$$

which is a metric on $\{P \text{ on } (\mathcal{X}, \mathscr{G}) : \exists x_0 \in \mathcal{X}, \int_\mathcal{X} d(x, x_0)^p \mathrm{d}P(x) < \infty\}$. Let $P_1, P_2, \ldots, P$ be probability measure on $(\mathcal{X}, \mathscr{G})$. We say that, for $p \geq 1$, $P_n \to P$ in p-Wasserstein if $d_{W_p}(P_n, P) \to 0$.

Theorem 5.2.4. Let $p \geq 1$.

(a) If $\exists(\Omega, \mathscr{F}, \mathbb{P})$ and $X_1, X_2, \ldots, X : \Omega \to \mathcal{X}$ such that $\mathbb{P}^{(X_n)} = P_n$ and $\mathbb{P}^{(X)} = P$. Then $X_n \to X$ in L_p implies $P_n \to P$ in p-Wasserstein.

(b) If $(\mathcal{X}, d)$ is a complete and separable metric space, then for any P, Q probability measure on $(\mathcal{X}, d)$,

$$d_{W_1}(P, Q) = \sup \left\{ \left| \int_{\mathcal{X}} f \mathrm{d}P - \int_{\mathcal{X}} f \mathrm{d}Q \right| : f : \mathcal{X} \to \mathbb{R} \text{ is 1-Lipschitz} \right\}.$$

(c) Let $P_1, P_2, \ldots, P$ be probability measure on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ such that $\int_{\mathbb{R}^d} \|x\|^d \,\mathrm{d}P_n(x) < \infty$. Then $P_n \to P$ in p-Wasserstein if and only if $\int_{\mathbb{R}^d} \|X\|_2^p \,\mathrm{d}P_n(x) \to \int_{\mathbb{R}^d} \|x\|_2^p \,\mathrm{d}P(x)$ and $P_n \to P$ weakly.

Proof.

(a) It follows since $(\mathbb{E}_{\mathbb{P}} d(X_n, X)^p)^{1/p} \geq d_{W_p}(P_n, P)$.

(b) This one is difficult; see Edwards, 2011 and Kellerer, 1985 or note by Basso.

(c) Suppose $P_n \to P$ in p-Wasserstein. Fix $\varepsilon > 0$, then $\exists(\Omega, \mathscr{F}, \mathbb{P})$ and $X_1, X_2, \ldots, X : \Omega \to \mathbb{R}^d$ such that $\forall n \in \mathbb{N}$,

$$\mathbb{E}_{\mathbb{P}} \|X_n - X\|^p \leq d_{W_p}(P_n, P) + \frac{\varepsilon}{n} \to 0.$$

E.g. let $\Omega = (\mathbb{R}^d)^{\mathbb{N}}$, $\mathscr{F} = (\mathscr{B}(\mathbb{R}^d))^{\otimes \mathbb{N}}$, and let $Q_n \in \mathcal{C}(P_n, P)$ such thhat

$$\mathbb{E}_{Q_n} \|X_n - X\|^p \leq d_{W_p}(P_n, P) = \frac{\varepsilon}{n}.$$

Let $\mathbb{P}(A_1 \times A_2 \times \mathbb{R}^d \times \mathbb{R}^d \ldots) = P_1(A_1) P_2(A_2)$. Let

$$\mathbb{P}(A_1 \times A_2 \times A_3 \times \mathbb{R}^d \times \mathbb{R}^d \times \cdots \times A) = \begin{cases} \frac{Q_1(A_1, A) Q_2(A_2, A) Q_3(A_3, A)}{P(A)^2} & \text{if } P(A) > 0 \\ 0 & \text{if } P(A) = 0. \end{cases}$$

This definition on rectangular cylinders is consistent. Use Kolmogrov extension theory.

Then $X_n \to X$ in probability implies $X_n \to X$ in law. Also, by Minkowski inequality,

$$\left| (\mathbb{E}_{\mathbb{P}} \|X_n\|^p)^{\frac{1}{o}} - (\mathbb{E}_{\mathbb{P}} \|X\|^p)^{\frac{1}{p}} \right| \leq (\mathbb{E}_{\mathbb{P}} \|X_n - X\|^p)^{\frac{1}{p}} \to 0.$$

Now suppose $P_n \to P$ weakly and $\int_{\mathbb{R}^d} \|x\|_2^p \,\mathrm{d}P_n(x) \to \int_{\mathbb{R}^d} \|x\|_2^p \,\mathrm{d}P(x)$. Then by Theorem 5.2.2, $\exists(\Omega, \mathscr{F}, \mathbb{P})$ and $X_1, X_2, \ldots, X : \Omega \to \mathbb{R}^d$ such that $\mathbb{P}^{(X_n)} = P_n$ and $\mathbb{P}^{(X)} = P$. Then, combining the assumption $\mathbb{E}_{\mathbb{P}} |X_n|^p < \infty$, $\forall n \in \mathbb{N}$ and Theorem 5.2.3 yields $\mathbb{E}_{\mathbb{P}} \|X - X\|^p \to 0$. So $\mathbb{P}^{(X_n)} \to \mathbb{P}^{(X)}$ in p-Wasserstein by (a).

□

5.3 Central Limit Theorem

Definition 5.3.1. Let $d \in \mathbb{N}$. Define the bump function $\psi : \mathbb{R}^d \to [0, \infty)$ by

$$\psi(x) = \begin{cases} C_d \exp\left(-\frac{1}{1 - \|x\|^2}\right) & \text{if } \|x\|_2 \leq 1 \\ 0 & \text{else}, \end{cases}$$

where $C_d > 0$ is the normalization so that ψ is a density on $\mathbb{R}^d$. We note that ψ is infinitely differentiable on $\mathbb{R}^d$ (but not analytic). Since ψ is compactly supported, all its derivatives are continuous, compactly supported, and bounded.

Lemma 5.3.1. Let $d \in \mathbb{N}$. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and suppose $X_1, X_2, \dots, X : \Omega \to \mathbb{R}^d$. If, for any $f : \mathbb{R}^d \to \mathbb{R}$ that is bounded and whose third derivatives exist and are bounded, $\lim_{n\to\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X)$, then for any $g : \mathbb{R}^d \to \mathbb{R}$ bounded and Lipschitz, $\lim_{n\to\infty} \mathbb{E}g(X_n) = \mathbb{E}g(X)$, that is, $X_n \to X$ in law.

Proof.

Let $g : \mathbb{R}^d \to \mathbb{R}$ be L-Lipschitz and bounded by B. For $\sigma > 0$, define

$$g_\sigma(x) = \mathbb{E}_{Z\sim P_\psi} g(x + \sigma Z),$$

where P_ψ denotes the probability measure whose density is the bump function. We observe that

$$\begin{aligned} g_\sigma(x) &= \int_{\mathbb{R}^d} g(x + \sigma z)\psi(z)dz \\ &= \int_{\mathbb{R}^d} g(y)\psi\big(\frac{y-x}{\sigma}\big)\frac{1}{\sigma}dy. \end{aligned}$$

Since the derivatives of ψ are bounded and compactly supported,

$$\begin{aligned} g'_\sigma(x) &= -\int_{\mathbb{R}^d} g(y)\psi'\big(\frac{y-x}{\sigma}\big)\frac{1}{\sigma^2}dy \\ &\cdots \\ g^{(3)}_\sigma(x) &= -\int_{\mathbb{R}^d} g(y)\psi^{(3)}\big(\frac{y-x}{\sigma}\big)\frac{1}{\sigma^4}dy. \end{aligned}$$

Since g and $\psi^{(3)}$ are both bounded, we have that $g^{(3)}_\sigma$ is bounded and continuous. Thus, we have that $\mathbb{E}g_\sigma(X_n) \to \mathbb{E}g_\sigma(X)$.

Moreover, since g is L-Lipschitz, we have that, for any $x \in \mathbb{R}^d$,

$$|g(x) - g_\sigma(x)| \le \mathbb{E}_{Z\sim P_\psi}|g(x) - g(x + \sigma Z)| \le \sigma L \mathbb{E}_{Z\sim P_\psi}\|Z\|_2 \le \sigma L.$$

Fix $\varepsilon > 0$ and let $\sigma = \frac{\varepsilon}{2L}$. Then, for any $n \in \mathbb{N}$,

$$\begin{aligned} |\mathbb{E}g(X_n) - \mathbb{E}g(X)| &\le \mathbb{E}|g(X_n) - g_\sigma(X_n)| + \mathbb{E}|g(X) - g_\sigma(X)| + |\mathbb{E}g_\sigma(X_n) - \mathbb{E}g_\sigma(X)| \\ &\le \varepsilon + |\mathbb{E}g_\sigma(X_n) - \mathbb{E}g_\sigma(X)|. \end{aligned}$$

Taking the limit $n \to \infty$ on both sides, we have that $\lim_{n\to\infty} |\mathbb{E}g(X_n) - \mathbb{E}g(X)| \le \varepsilon$. Since ε is arbitrary, the claim follows.

□

Theorem 5.3.1 (CLT under Finite Third Moment Condition). Suppose $X_1, X_2, X_3, \dots : \Omega \to \mathbb{R}^d$ are independent random vectors. Suppose, $\forall i \in \mathbb{N}$, $\mu_i := \mathbb{E}X_i \in \mathbb{R}^d$, $\Sigma_i := \mathbb{E}(X_i - \mu_i)(X_i - \mu_i)^\intercal \in \mathbb{R}^{d\times d}$ and $\sigma_{ij} := (\mathbb{E}|X_{ij} - \mu_{ij}|^2)^{\frac{1}{2}} \ge 0$. Suppose $\Sigma = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \Sigma_i \in \mathbb{R}^{d\times d}$. Suppose also $\exists C > 0$ such that $\sup_{i\in\mathbb{N}, j\in[d]} \mathbb{E}|X_{ij} - \mu_{ij}|^3 \le C$. Then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - \mu_i) \to N(0, \Sigma) \text{ in law.}$$

Proof.

Let $n \in \mathbb{N}$. Define, for $i \in [n]$, $Z_i : \Omega \to \mathbb{R}^d$ such that $\mathbb{P}^{(Z_i)} = N(0, \Sigma_i)$. Observe that $\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i$ has pushforward measure $N(0, \frac{1}{n}\sum_{i=1}^n Z_i) \to N(0, \Sigma)$ weakly.

Let $f : \mathbb{R}^d \to \mathbb{R}$ be such that $\exists B_1, B_2 > 0$,

$$\sup_{x\in\mathbb{R}^d} |f(x)| \le B_1, \max_{i,j,k\in[d]} \sup_{x\in\mathbb{R}^d} \left|\frac{\partial^3 f(x)}{\partial x_i \partial x_j \partial x_k}\right| \le B_2.$$

For $i \in [n]$, define $\widetilde{X}_i = \frac{1}{\sqrt{n}}(X_i - \mu_i)$ and $\widetilde{Z}_i = \frac{1}{\sqrt{n}} Z_i$. Write $Y_1 = \widetilde{X}_2 + \widetilde{X}_3 + \cdots + \widetilde{X}_n$, write

$$f(\widetilde{X}_1 + \widetilde{X}_2 + \cdots + \widetilde{X}_n) = f(\widetilde{X}_1 + Y_1) = f(Y_1) + \nabla f(Y_1)^\top \widetilde{X}_1 + \frac{1}{2}\widetilde{X}_1^\top (Hf)(Y_1)\widetilde{X}_1 + R(\widetilde{X}_1, Y_1)$$

and

$$f(\widetilde{Z}_1 + \widetilde{X}_2 + \cdots + \widetilde{X}_n) = f(\widetilde{Z}_1 + Y_1) = f(Y_1) + \nabla f(Y_1)^\top \widetilde{Z}_1 + \frac{1}{2}\widetilde{Z}_1^\top (Hf)(Y_1)\widetilde{Z}_1 + R(\widetilde{Z}_1, Y_1).$$

Since Y_1 is independent of $\widetilde{X}_1$, $\widetilde{Z}_1$ and $\mathbb{E}\widetilde{X}_1 = \mathbb{E}\widetilde{Z}_1$ and $\mathbb{E}\widetilde{X}_1\widetilde{X}_1^\top = \mathbb{E}\widetilde{Z}_1\widetilde{Z}_1^\top$, we have

$$|\mathbb{E}f(\widetilde{X}_1 + \widetilde{X}_2 + \cdots + \widetilde{X}_n) - \mathbb{E}f(\widetilde{Z}_1 + \widetilde{X}_2 + \cdots + \widetilde{X}_n)| \leq |\mathbb{E}R(\widetilde{X}_1, Y_1)| + |\mathbb{E}R(\widetilde{Z}_1, Y_1)| \tag{$*$}$$

By letting, for $i = 2, 3, \ldots, n$, $Y_i = \widetilde{Z}_1 + \cdots + \widetilde{Z}_{i-1} + \widetilde{X}_{i+1} + \cdots + \widetilde{X}_n$, we have, by similar derivation, that

$$\begin{aligned}&|\mathbb{E}f(\widetilde{Z}_1 + \cdots + \widetilde{Z}_{i-1} + \widetilde{X}_i + \widetilde{X}_{i+1} + \cdots + \widetilde{X}_n) - \mathbb{E}f(\widetilde{Z}_1 + \cdots + \widetilde{Z}_{i-1} + \widetilde{Z}_i + \widetilde{X}_{i+1} + \cdots + \widetilde{X}_n)| \\ \leq& |\mathbb{E}R(\widetilde{X}_i, Y_i)| + |\mathbb{E}R(\widetilde{Z}_i, Y_i)| \end{aligned} \tag{$**$}$$

Combining $(*)$ and $(**)$, we have

$$|\mathbb{E}f(\widetilde{Z}_1 + \cdots + \widetilde{Z}_n) - \mathbb{E}f(\widetilde{X}_1 + \cdots + \widetilde{X}_n)| \leq \sum_{i=1}^n |\mathbb{E}R(\widetilde{X}_i, Y_i)| + |\mathbb{E}R(\widetilde{Z}_i, Y_i)|.$$

By Taylor's theorem, $\forall i \in [n]$,

$$\begin{aligned}|\mathbb{E}R(\widetilde{X}_i, Y_i)| &\leq \mathbb{E}\left\{\sum_{i,k,l\in[d]} \sup_{y\in\mathbb{R}^d}\left|\frac{\partial^3 f}{\partial y_j \partial y_k \partial y_l}(y)\right| |\tilde{X}_{ij}\tilde{X}_{ik}\tilde{X}_{il}|\right\} \\ &\leq B_2\mathbb{E}\left(\sum_{j=1}^d |\widetilde{X}_{ij}|\right)^3 = B_2 d^3 \mathbb{E}\left(\frac{1}{d}\sum_{j=1}^d |\widetilde{X}_{ij}|\right)^3 \leq B_2 d^3\mathbb{E}\left(\frac{1}{d}\sum_{j=1}^d |\widetilde{X}_{ij}|^3\right) \leq B_2\frac{d^3}{n^{\frac{3}{2}}}C, \\ |\mathbb{E}R(\widetilde{Z}_i, Y_i)| &\leq B_2 d^3 \mathbb{E}\left(\frac{1}{d}\sum_{j=1}^d |\widetilde{Z}_{ij}|^3\right) = B_2 d^3\mathbb{E}\left(\frac{1}{d}\sum_{j=1}^d \sigma_{ij}^3\left|\frac{\widetilde{Z}_{ij}}{\sigma_{ij}}\right|^3\right) \leq B_2\frac{d^3}{n^{3/2}}2C_1^3,\end{aligned}$$

where $C_1 := \sup_{i\in\mathbb{N}, j\in[d]}\sigma_{ij} < \infty$. Therefore, $|\mathbb{E}f(\widetilde{Z}_1 + \ldots \widetilde{Z}_n) - \mathbb{E}f(\widetilde{X}_1 + \cdots + \widetilde{X}_n)| \leq B_2(2C_1^3 + C)\frac{d^3}{\sqrt{n}}$. So

$$\lim_{n\to\infty}\mathbb{E}f\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - \mu_i)\Big) = \lim_{n\to\infty}\mathbb{E}f\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i\Big).$$

The claim follows from Lemma 5.3.1.

□

5.4 Convergence through Moments

Definition 5.4.1. Let $(\mathcal{X}, \mathscr{G})$ be a measure metric space. We say that a collection $\{P_n\}_{n=1}^\infty$ of probability measures on $(\mathcal{X}, \mathscr{G})$ is <u>tight</u> if $\forall \varepsilon > 0$, $\exists K_\varepsilon \in \mathscr{G}$ compact such that $\forall n \in \mathbb{N}$, $P_n(K_\varepsilon) \geq 1 - \varepsilon$.

We showed in Theorem 4.4.2 that if $\mathcal{X}$ is complete and separable, then any finite $\{P_1, \ldots, P_n\}$ is tight.

Theorem 5.4.1 (**Prokhorov's Theorem**)**.** Let $(\mathcal{X}, \mathscr{G})$ be separable. Then a collection $\{P_n\}_{n=1}^\infty$ is <u>tight</u> iff $\exists$ a subsequence $\{P_{n_1}, P_{n_2}, \ldots\}$ and a probability measure P such that $P_{n_k} \to P$ weakly.

Proof.

See note by Van Gaans, Theorem 5.2. □

Lemma 5.4.1. Let $(\mathcal{X}, \mathscr{G})$ be a measurable metric space an let $\{P_n\}_{n=1}^{\infty}$ be a tight collection of probability measures. Suppose $\exists$ probability measure P such that for any weakly convergent subsequence $\{P_{n_k}\}_{k=1}^{\infty}$, $P_{n_k} \to P$ weakly. Then, we have that $P_n \to P$ weakly.

Proof.

Suppose $\{P_n\}_{n=1}^{\infty}$ does not converge to P weakly. Then, $\exists f : \mathcal{X} \to \mathbb{R}$ bounded and continuous and $c > 0$ such that $\limsup_{n\to\infty} |\mathbb{E}_{P_n} f - \mathbb{E}_P f| \geq c$. Thus, $\forall k \in \mathbb{N}$, $\exists n_k \geq k$ such that

$$|\mathbb{E}_{P_{n_k}} f - \mathbb{E}_P f| \geq c.$$

Since $\{P_n\}_{n=1}^{\infty}$ is tight, $\{P_{n_k}\}_{k=1}^{\infty}$ is tight as well by Theorem 5.4.1. There exists a further subsequence $\{m_1, m_2, \dots\} \subseteq \{n_1, n_2, \dots\}$ such that $\{P_{m_k}\}_{k=1}^{\infty}$ is convergent and thus converge to P, that is, $\lim_{k\to\infty} |\mathbb{E}_{P_{m_k}} f - \mathbb{E}_P f| = 0$. This is a contradiction. Thus, $P_n \to P$ weakly as desired. ☐

Definition 5.4.2. Let $d \in \mathbb{N}$ and let $X : \Omega \to \mathbb{R}$ have distribution (pushforward measure) P on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. The moment generating function $M_p : \mathbb{R} \to [0, \infty]$ is $\forall t \in \mathbb{R}$,

$$M_P(t) = \mathbb{E}[e^{tX}]. \tag{5.4}$$

Remark 5.4.1. If $\exists \varepsilon > 0$ such that $|t| < \varepsilon$ implies $M_p(t) < \infty$, then

(i) $\mathbb{E}[e^{tX}] = 1 + \sum_{k=1}^{\infty} \frac{t^k \mathbb{E}[X^k]}{k!}$, $\forall t \in (-\varepsilon, \varepsilon)$,

(ii) Let Q be another distribution on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, if $M_P(t) = M_Q(t)$, $\forall t \in (-\varepsilon, \varepsilon)$, then $P = Q$. That is, if $X, Y : \Omega \to \mathbb{R}$ and $P = \mathbb{P}^{(X)}$ and $Q = \mathbb{P}^{(Y)}$, then $P = Q$ iff $\forall k \in \mathbb{N}$, $\mathbb{E}X^k = \mathbb{E}Y^k$.

(iii) There exists $P \neq Q$ on $\mathbb{R}$ satisfying $\mathbb{E}e^{tX} = \infty$, $\forall t \neq 0$ such that $\forall k \in \mathbb{N}$, $\mathbb{E}X^k = \mathbb{E}Y^k$ (with X, Y defined as in (ii)).

Theorem 5.4.2. Let $\{P_n\}_{n=1}^{\infty}$ be a sequence of probability measures on $\mathbb{R}$. Let P be a probability measure on $\mathbb{R}$ such that $M_P(t) < \infty$, $\forall t \in (-c, c)$ for some $c > 0$. Let $X_1, X_2, \dots, X : \Omega \to \mathbb{R}$ such that $\mathbb{P}^{(X_n)} = P_n$, $\mathbb{P}^{(X)} = P$. If $\mathbb{E}X_n^k \to \mathbb{E}X^k$, $\forall k \in \mathbb{N}$, then $P_n \to P$ weakly.

Proof.

First we show that $\{P_n\}$ is tight. Since $M_P(t) < \infty$ for some $t \neq 0$, $\mathbb{E}X^2 < \infty$, and since $\mathbb{E}X_n^2 \to \mathbb{E}X^2$, there exists $M > 0$ such that $\sup_{n\in\mathbb{N}} \mathbb{E}X_n^2 \leq M$. Hence, for any $\varepsilon > 0$, we may take $K_\varepsilon := [-(\frac{M}{\varepsilon})^{\frac{1}{2}}, (\frac{M}{\varepsilon})^{\frac{1}{2}}]$ to obtain $\forall n \in \mathbb{N}$,

$$1 - P_n(K_\varepsilon) = \mathbb{P}(X_n^2 > \frac{M}{\varepsilon}) < \frac{\mathbb{E}X_n^2}{M}\varepsilon < \varepsilon.$$

Now, let $\{P_{n_1}, P_{n_2}, \dots\}$ be a subsequence that converges weakly to some probability measure Q on $\mathbb{R}$. Then, let $Y : \Omega \to \mathbb{R}$ such that $\mathbb{P}^{(Y)} = Q$, we have $\mathbb{E}X^k = \mathbb{E}Y^k$, $\forall k \in \mathbb{N}$. So $Q = P$ by Remark 5.4.1 (ii). Thus, by Lemma 5.4.1, we have that $P_n \to P$ weakly. ☐

Definition 5.4.3. Let $X : \Omega \to \mathbb{R}^d$ be a random vector with distribution P and $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. Define the characteristic function $\phi : \mathbb{R}^d \to \mathbb{C}$ as

$$\phi_P(t) := \mathbb{E}e^{it^\top X} = \mathbb{E}\cos(t^\top X) + i\mathbb{E}\sin(t^\top X), \text{ for } t \in \mathbb{R}^d.$$

Remark 5.4.2. (i) Since $\forall x \in \mathbb{R}$, $|e^{ix}| = |\cos x + i \sin x| = 1$, for any probability measure P on $\mathbb{R}^d$, we have that $|\phi_P(t)| \leq 1$, $\forall t \in \mathbb{R}^d$.

(ii) By Levy inversion formula, we have that for any pair P, Q probability measures on $\mathbb{R}^d$, $P = Q$ iff $\phi_P(t) = \phi_Q(t)$, $\forall t \in \mathbb{R}^d$.

Theorem 5.4.3 (Levy Continuity Theorem)**.** Let $\{P_n\}_{n=1}^{\infty}$ be probability measure on $\mathbb{R}^d$ with characteristic function $\{\phi_{P_n}\}_{n=1}^{\infty}$. If $P_n \to P$ weakly for some probability measure P on $\mathbb{R}^d$, then $\forall t \in \mathbb{R}^d$, $\phi_{P_n}(t) \to \phi_P(t)$.

If $\forall t \in \mathbb{R}^d$, $\phi_{P_n}(t) \to \phi(t)$ for some $\phi : \mathbb{R}^d \to \mathbb{C}$ and ϕ is continuous at 0, then $\exists$ probability measure P such that $\phi = \phi_P$ and $P_n \to P$ weakly.

Proof.

Similar structure to the proof of theorem 5.4.2, see proof of Theorem 7.2.9 in Ash and Dolean-Dade. □

Theorem 5.4.4 (CLT under Linderberg Condition)**.** Let $X_1, X_2, X_3, \cdots : \Omega \to \mathbb{R}$ be independent r.v. such that $\forall n \in \mathbb{N}$, $\mu_n := \mathbb{E}X_n < \infty$ and $\sigma_n^2 := \mathrm{Var}(X_n) < \infty$. Write $c_n^2 = \sum_{i=1}^n \sigma_i^2$ and suppose $c_n^2 > 0$, $\forall n \in \mathbb{N}$. If $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{c_n^2} \sum_{i=1}^n \int_{\{x:|x-\mu_i|>\varepsilon c_n\}} (x-\mu_i)^2 \mathrm{d}\mathbb{P}^{(X_i)}(x) = 0,$$

then,

$$\frac{1}{c_n}(\sum_{i=1}^n (X_i - \mu_i)) \to N(0,1) \text{ in law.}$$

Proof.

Through Levy continuity theorem. See theorem 7.3.1 in Ash and Dolean-Dade. □

Corollary 5.4.1. If $X_1, X_2, \cdots : \Omega \to \mathbb{R}$ are iid with $\mu := \mathbb{E}X_1 < \infty$ and $\sigma^2 := \mathrm{Var}(X_1) < \infty$, then

$$\frac{1}{\sigma\sqrt{n}}(\sum_{i=1}^n (X_i - \mu)) \to N(0,1) \text{ in law.}$$

Proof.

Using notation in Theorem 5.4.4, we have $c_n^2 = \sum_{i=1}^n \mathrm{Var}(X_i) = n\sigma^2$. Thus, $\forall \varepsilon > 0$, $n \in \mathbb{N}$,

$$\frac{1}{c_n^2} \sum_{i=1}^n \int_{\{x:|x-\mu_i|>\varepsilon c_n\}} (x-\mu_i)^2 \mathrm{d}\mathbb{P}^{(X_i)}(x) = \frac{1}{\sigma^2} \int_{\{x:|x-\mu|>\varepsilon\sqrt{n}\sigma\}} (x-\mu)^2 \mathrm{d}\mathbb{P}^{(X_1)}(x) \to 0$$

as $n \to \infty$. The conclusion follows. □

Chapter 6

Conditional Expectation and Probability

6.1 Conditional Expectation

Theorem 6.1.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let $X : \Omega \to \mathbb{R}$ be measurable with respect to $(\Omega, \mathscr{F})/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ and integrable, and let $\mathscr{G} \subseteq \mathscr{F}$ be a sub-σ-filed. Then, there exists $g : \Omega \to \mathbb{R}$, measurable w.r.t $(\Omega, \mathscr{G})/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that $\forall B \in \mathscr{G}$,

$$\int_B g \mathrm{d}\mathbb{P} = \int_B X \mathrm{d}\mathbb{P}. \tag{6.1}$$

Moreover, g is unique $\mathbb{P}$-a.e. We write $g(\omega) = \mathbb{E}[X|\mathscr{G}](\omega)$, $\forall \omega \in \Omega$.

Proof.

Define $\lambda : \mathscr{G} \to \mathbb{R}$ by $\lambda(B) := \int_B X \mathrm{d}\mathbb{P}$, $\forall B \in \mathscr{G}$. Then λ is a signed measure by Theorem 2.2.1. Note that $\lambda \ll \mathbb{P}$ since if $\mathbb{P}(E) = 0$ for $E \in \mathscr{G}$, then $\lambda(E) = \int_E X \mathrm{d}\mathbb{P}$ as well. Thus, by Radon-Nikodym theorem, $\frac{\mathrm{d}\lambda}{\mathrm{d}\mathbb{P}} : \Omega \to \mathbb{R}$ exists and $\forall B \in \mathscr{G}$,

$$\int_B \frac{\mathrm{d}\lambda}{\mathrm{d}\mathbb{P}} \mathrm{d}\mathbb{P} = \lambda(B) = \int_B X \mathrm{d}\mathbb{P}.$$

Setting $g = \frac{\mathrm{d}\lambda}{\mathrm{d}\mathbb{P}}$ yields desired result. □

Definition 6.1.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $Y : \Omega \to \widetilde{\Omega}$ be a r.o. where $(\widetilde{\Omega}, \mathscr{H})$ is measurable space. Define $\sigma(Y) := \{Y^{-1}(\widetilde{A}) : \widetilde{A} \in \mathscr{H}\} \subseteq \mathscr{F}$ as the sub-σ-algebra generated by Y. Let $X : \Omega \to \mathbb{R}$ be r.v. Define $\mathbb{E}[X|Y] : \Omega \to \mathbb{R}$ as $\forall \omega \in \Omega$,

$$\mathbb{E}[X|Y](\omega) := \mathbb{E}[X|\sigma(Y)](\omega). \tag{6.2}$$

Thus, $\mathbb{E}[X|Y]$ is $\mathbb{P}$ a.e. uniquely defined by

(a) $\forall \widetilde{A} \in \mathscr{H}$,

$$\int_{Y^{-1}(\widetilde{A})} \mathbb{E}[X|Y] \mathrm{d}\mathbb{P} = \int_{Y^{-1}(\widetilde{A})} X \mathrm{d}\mathbb{P} = \mathbb{E}[X \mathbb{1}_{\{Y \in \widetilde{A}\}}]. \tag{6.3}$$

(b) $\forall S \in \mathscr{B}(\mathbb{R})$, $\mathbb{E}[X|Y]^{-1}(S) \in \sigma(Y)$, or equivalently, $\exists \widetilde{A} \in \mathscr{H}$, such that

$$\forall \omega \in \Omega, \mathbb{E}[X|Y](\omega) \in S \iff Y(\omega) \in \widetilde{A}. \tag{6.4}$$

Example 6.1.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be probability space, let $X : \Omega \to \mathbb{R}$ be r.v. and let $Y : \Omega \to \{0, 1\}$ so that $\sigma(Y) := \{\emptyset, \Omega, Y^{-1}(0), Y^{-1}(1) = Y^{-1}(0)^c\}$.

Write $g := \mathbb{E}[X|Y]$, we have that g must satisfy $\int_\emptyset g\mathrm{d}\mathbb{P} = 0$, $\int_\Omega g\mathrm{d}\mathbb{P} = \int_\Omega X\mathrm{d}\mathbb{P} = \mathbb{E}X$,

$$\int_{Y^{-1}(0)} g\mathrm{d}\mathbb{P} = \int_{Y^{-1}(0)} X\mathrm{d}\mathbb{P} = \mathbb{E}[X\mathbb{1}_{\{Y=0\}}] \quad \text{and} \quad \int_{Y^{-1}(1)} g\mathrm{d}\mathbb{P} = \mathbb{E}[X\mathbb{1}_{\{Y=1\}}].$$

Since g must be $(\Omega, \sigma(Y))/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ measurable, it must be that $\exists a, b \in \mathbb{R}$ such that $g(\omega) = a$, $\forall\omega$ such that $Y(\omega) = 1$ and $g(\omega) = b$, $\forall\omega$ such that $Y(\omega) = 0$.

Thus, $\mathbb{E}[X\mathbb{1}_{\{Y=1\}}] = \int_{Y^{-1}(1)} g\mathrm{d}\mathbb{P} = a \int_{Y^{-1}(1)} \mathrm{d}\mathbb{P}$ implies $a = \frac{\mathbb{E}[X\mathbb{1}_{\{Y=1\}}]}{\mathbb{P}(Y=1)}$. Likewise, $b = \frac{\mathbb{E}[X\mathbb{1}_{\{Y=0\}}]}{\mathbb{P}(Y=0)}$.

Lemma 6.1.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$, $(\widetilde{\Omega}, \mathscr{H})$ be probability spaces. Let $Y : \Omega \to \widetilde{\Omega}$ be r.o. measurable w.r.t. $(\Omega, \mathscr{F})/(\widetilde{\Omega}, \mathscr{H})$ and $Y(\Omega) \in \mathscr{H}$. Let $X : \Omega \to \mathbb{R}$ be measurable w.r.t $(\Omega, \mathscr{F})/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Then, X is measurable w.r.t. $(\Omega, \sigma(Y))/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ if and only if $\exists\phi : \widetilde{\Omega} \to \mathbb{R}$ measurable w.r.t. $(\widetilde{\Omega}, \mathscr{H})/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that $X = \phi(Y)$, i.e. $\forall\omega \in \Omega, X(\omega) = \phi(Y(\omega))$.

$$\begin{tikzcd} \Omega \arrow[r, "Y"] \arrow[rr, bend right, "X"'] & \widetilde{\Omega} \arrow[r, "\phi"] & \mathbb{R} \end{tikzcd}$$

Figure 6.1: lemma illustration

Note that X is $(\Omega, \sigma(Y))/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ measurable if and only if $\forall s \in \mathscr{B}(\mathbb{R})$, $X^{-1}(s) \in \sigma(Y)$ iff $\sigma(X) \subseteq \sigma(Y)$.

Proof.

First suppose $X \geq 0$ and is simple, i.e. $\exists c_1, c_2, \ldots, c_m > 0$ and $A_1, A_2, \ldots, A_m \in \mathscr{F}$ such that $X = \sum_{i=1}^m c_i \mathbb{1}_{A_i}$. If X is measurable w.r.t. $(\Omega, \sigma(Y))/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, then $\forall i \in [m]$, $X^{-1}(c_i) = A_i \in \sigma(Y)$. It implies $\exists\widetilde{A}_i \in \mathscr{H}$ such that $A_i = Y^{-1}(\widetilde{A}_i)$. Thus, we may write $\phi = \sum_{i=1}^m c_i \mathbb{1}_{\widetilde{A}_i}$, which is $(\widetilde{\Omega}, \mathscr{H})/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable, and we have $\forall\omega \in \Omega$,

$$\phi(Y(\omega)) = \sum_{i=1}^m c_i \mathbb{1}_{\{Y(\omega) \in \widetilde{A}_i\}} = \sum_{i=1}^m c_i \mathbb{1}_{A_i}(\omega) = X(\omega).$$

Now suppose $X \geq 0$. If X is measurable w.r.t. $(\Omega, \sigma(Y))/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, then $\exists$ sequence $0 \leq X_1 \leq X_2 \leq \ldots$ simple, measurable w.r.t $(\Omega, \sigma(Y))/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, such that $\forall\omega \in \Omega$, $X_n(\omega) \to X(\omega)$. By previous analysis, $\exists\phi_1, \phi_2, \cdots : \widetilde{\Omega} \to \mathbb{R}$, all measurable w.r.t. $(\widetilde{\Omega}, \mathscr{H})/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, such that $X_n = \phi_n(Y)$. We then have, $\forall\omega \in \Omega$, $X(\omega) = \lim_{n\to\infty} \phi_n(Y(\omega))$. Define $\phi : \widetilde{\Omega} \to \mathbb{R}$ by $\phi(\omega) = \lim_{n\to\infty} \phi_n(\omega)$, $\forall\omega \in Y(\Omega)$ and $\phi(\omega) = 0$ if $\omega \notin Y(\Omega)$. Then ϕ is $(\widetilde{\Omega}, \mathscr{H})/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ measurable and $X = \phi(Y)$.

If $X : \Omega \to \mathbb{R}$, then we may apply previous reasoning on X^+, X^- to obtain $\phi : \widetilde{\Omega} \to \mathbb{R}$ such that $X = \phi(Y)$.

If, on the other hand, suppose $\exists\phi : \widetilde{\Omega} \to \mathbb{R}$, $(\widetilde{\Omega}, \mathscr{H})/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable such that $X = \phi(Y)$, then, $\forall S \in \mathscr{B}(\mathbb{R})$, $X^{-1}(S) = Y^{-1}(\phi^{-1}(S)) \in \sigma(Y)$. It implies X is $(\Omega, \sigma(Y))/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable. $\square$

Theorem 6.1.2. Same setting as Definition 6.1.1. Then, there exists a function $\phi : \widetilde{\Omega} \to \mathbb{R}$ measurable w.r.t $(\widetilde{\Omega}, \mathscr{H})/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that $\mathbb{E}[X|Y] = \phi(Y)$, i.e.

$$\forall\omega \in \Omega, \mathbb{E}[X|Y](\omega) = \phi(Y(\omega)).$$

Moreover, for all $\widetilde{A} \in \mathscr{H}$,

$$\int_{\widetilde{A}} \phi\mathrm{d}\mathbb{P}^{(Y)} = \int_{Y^{-1}(\widetilde{A})} X\mathrm{d}\mathbb{P}. \tag{6.5}$$

and if $\psi : \widetilde{\Omega} \to \mathbb{R}$ satisfy (6.5), then $\psi = \phi$, $\mathbb{P}^{(Y)}$-a.e.

We write, $\forall y \in \widetilde{\Omega}$, $\mathbb{E}[X|Y = y] := \phi(y)$.

Proof.

First assume $Y(\Omega) \in \mathscr{H}$. Since $\mathbb{E}[X|Y] : \Omega \to \mathbb{R}$ is $(\Omega, \sigma(Y))/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable, by Lemma 6.1.1, $\exists \phi : \widetilde{\Omega} \to \mathbb{R}$ such that $\mathbb{E}[X|Y] = \phi(Y)$. For $\widetilde{A} \in \mathscr{H}$, we have that

$$\int_{\widetilde{A}} \phi(y) \mathrm{d}\mathbb{P}^{(Y)}(y) = \int_{Y^{-1}(\widetilde{A})} \phi(Y) \mathrm{d}\mathbb{P} = \int_{Y^{-1}(\widetilde{A})} X \mathrm{d}\mathbb{P}$$

by Theorem 6.1.1. Uniqueness follows from Theorem 6.1.1.

In general, defined signed measure $\lambda : \mathscr{H} \to [-\infty, \infty]$ by

$$\lambda(\widetilde{A}) = \int_{Y^{-1}(\widetilde{A})} X \mathrm{d}\mathbb{P}, \quad \forall \widetilde{A} \in \mathscr{H}.$$

We have that $\lambda \ll \mathbb{P}^{(Y)}$ and thus may let $\phi = \frac{\mathrm{d}\lambda}{\mathrm{d}\mathbb{P}^{(Y)}}$. Since

$$\int_{Y^{-1}(\widetilde{A})} X \mathrm{d}\mathbb{P} = \int_{\widetilde{A}} \phi \mathrm{d}\mathbb{P}^{(Y)} = \int_{Y^{-1}(\widetilde{A})} \phi(Y) \mathrm{d}\mathbb{P}, \quad \forall \widetilde{A} \in \mathscr{H},$$

we have that $\phi(Y) = \mathbb{E}[X|Y]$. $\square$

Theorem 6.1.3. Let $(\Omega, \mathscr{F}, \mathbb{P})$, $(\widetilde{\Omega}, \mathscr{H})$ be probability space and let $Y : \Omega \to \widetilde{\Omega}$ be $(\Omega, \mathscr{F})/(\widetilde{\Omega}, \mathscr{H})$-measurable and let $X : \Omega \to \mathbb{R}$ be $(\Omega, \mathscr{F})/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable. Suppose $\mathbb{E}|X| < \infty$. Then, for all $\psi : \widetilde{\Omega} \to \mathbb{R}$ such that $\mathbb{E}|\psi(Y)| < \infty$,

$$\int_{\Omega} (X - \mathbb{E}[X|Y]) \psi(Y) \mathrm{d}\mathbb{P} = 0. \tag{6.6}$$

Moreover, we have that

$$\int (\mathbb{E} - \mathbb{E}[X|Y])^2 \mathrm{d}\mathbb{P} \leq \int (X - \psi(Y))^2 \mathrm{d}\mathbb{P}. \tag{6.7}$$

Proof.

For the first claim, first suppose $\psi \geq 0$ and is simple function, i.e. for some $c_1, \dots, c_m \geq 0$ and $\widetilde{A}_1, \dots, \widetilde{A}_m \in \mathscr{H}$. $\psi = \sum_{i=1}^m c_i \mathbb{1}_{\widetilde{A}_i}$. Then, we have

$$\int_{\Omega} (X - \mathbb{E}[X|Y]) \psi(Y) \mathrm{d}\mathbb{P} = \sum_{i=1}^m c_i \int_{Y^{-1}(\widetilde{A}_i)} X - \mathbb{E}[X|Y] \mathrm{d}\mathbb{P} = 0.$$

Now, let $\psi \geq 0$ and let $0 \leq \psi_1 \leq \psi_2 \dots$ be simple functions such that $\lim_{n\to\infty} \psi_n(y) = \psi(y)$, $\forall y \in \widetilde{\Omega}$. Observe that since $\mathbb{E}\big(\mathbb{E}[X|Y]\big) = \mathbb{E}X$ is finite, $\mathbb{E}|\mathbb{E}[X|Y]| < \infty$.

We thus have

$$\begin{aligned} &\left| \int_{\Omega} (X - \mathbb{E}[X|Y]) \psi(Y) \mathrm{d}\mathbb{P} - \int_{\Omega} (X - \mathbb{E}[X|Y]) \psi_n(Y) \mathrm{d}\mathbb{P} \right| \\ &\leq \int_{\Omega} |X - \mathbb{E}[X|Y]| (\psi(Y) - \psi_n(Y)) d\mathbb{P} \to 0 \end{aligned}$$

where the final limit follows from the monotone convergence theorem. The general case immediately follows.

For the second claim, note that

$$\begin{aligned}
&\int_\Omega (X-\psi(Y))^2 \mathrm{d}\mathbb{P} \\
&= \int_\Omega (X-\mathbb{E}[X|Y]+\mathbb{E}[X|Y]-\psi(Y))^2 \mathrm{d}\mathbb{P} \\
&= \int_\Omega (X-\mathbb{E}[X|Y])^2 \mathrm{d}\mathbb{P} + \underbrace{\int_\Omega 2(X-\mathbb{E}[X|Y])(\mathbb{E}[X|Y]-\psi(Y))\mathrm{d}\mathbb{P}}_{=0 \text{ by Thm 6.1.2}} + \int_\Omega (\mathbb{E}[X|Y]-\psi(Y))^2 \mathrm{d}\mathbb{P} \\
&\geq \int_\Omega (X-\mathbb{E}[X|Y])^2 \mathrm{d}\mathbb{P}.
\end{aligned}$$

□

Remark 6.1.1. Let $\mathbb{P}^{(X,Y)}$ be a probability measure on $(\mathbb{R}\times\widetilde{\Omega}, \mathscr{B}(\mathbb{R})\otimes\mathscr{H})$ induced by (X,Y). Let us assume there exists Markov Kernel $\mathcal{K}:\widetilde{\Omega}\times\mathbb{R}\to[0,1]$ such that $\forall s\in\mathbb{R}$, $\widetilde{A}\in\mathscr{H}$,

$$\mathbb{P}(\{X\in S\}\cap\{Y\in\widetilde{A}\}) = \mathbb{P}^{(X,Y)}(S,\widetilde{A}) = \int_{\widetilde{A}} \mathcal{K}(y,S)\mathrm{d}\mathbb{P}^{(Y)}(y).$$

We claim then that for $\mathbb{P}^{(Y)}$-a.e. $y\in\widetilde{\Omega}$,

$$\mathbb{E}[X|Y=y] = \int_{\mathbb{R}} x\mathrm{d}\mathcal{K}(y,\cdot)(x).$$

To see this, define $\Psi:\widetilde{\Omega}\to\mathbb{R}$ by $\Psi(y)=\int_{\mathbb{R}} x\mathrm{d}\mathcal{K}(y,\cdot)(x)$, $\forall y\in\widetilde{\Omega}$. We note that, any $\widetilde{A}\in\mathscr{H}$,

$$\begin{aligned}
\int_{\widetilde{A}} \Psi(y)\mathrm{d}\mathbb{P}^{(Y)}(y) &= \int_{\widetilde{A}}\int_{\mathbb{R}} x\mathrm{d}\mathcal{K}(y,\cdot)(x)\mathrm{d}\mathbb{P}^{(Y)}(y) \\
&= \int_{\mathbb{R}\times\widetilde{A}} x\mathrm{d}\mathbb{P}^{(X,Y)}(x,y) \\
&= \int_{(X,Y)^{-1}(\mathbb{R}\times\widetilde{A})} X\mathrm{d}\mathbb{P} \\
&= \int_{Y^{-1}(\widetilde{A})} X\mathrm{d}\mathbb{P}.
\end{aligned}$$

Thus, $\Psi=\mathbb{E}[X|Y=\bullet]$, $\mathbb{P}^{(Y)}$-a.e. by Theorem 6.1.2.

6.2 Properties of Conditional Expectation

Theorem 6.2.1. Let $(\Omega,\mathscr{F},\mathbb{P})$ be a probability space and let $\mathscr{G}\subseteq\mathscr{F}$ be a sub-σ-field. Let $X_1,X_2:\Omega\to\mathbb{R}$ be $(\Omega,\mathscr{F})/(\mathbb{R},\mathscr{B}(\mathbb{R}))$-measurable and integrable.

(a) If $X_1\leq X_2$ $\mathbb{P}$-a.e, then $\mathbb{E}[X_1|\mathscr{G}]\leq\mathbb{E}[X_2|\mathscr{G}]$ $\mathbb{P}$-a.e. as well.

(b) If $a,b\in\mathbb{R}$, and if $a\mathbb{E}X_1+b\mathbb{E}X_2$ is well-defined, then $\mathbb{E}[aX_1+bX_2|\mathscr{G}]=a\mathbb{E}[X_1|\mathscr{G}]+b\mathbb{E}[X_1|\mathscr{G}]$.

(c) $\big|\mathbb{E}[X_1|\mathscr{G}]\big|\leq\mathbb{E}\big[|X_1|\big|\mathscr{G}\big]$.

(d) For any convex function $r:\mathbb{R}\to\mathbb{R}$, $\mathbb{E}[r(X)|\mathscr{G}]\geq r(\mathbb{E}[X|\mathscr{G}])$.

Let $(\widetilde{\Omega},\mathscr{H})$ be a measurable space, let $Y:\Omega\to\widetilde{\Omega}$ be $(\Omega,\mathscr{F})/\widetilde{\Omega},\mathscr{H})$-measurable.

(a') If $X_1\leq X_2$ $\mathbb{P}$-a.e., then $\mathbb{E}[X_1|Y=\bullet]\leq\mathbb{E}[X_2|Y=\bullet]$ $\mathbb{P}^{(Y)}$-a.e.

(b') If $a, b \in \mathbb{R}$, and if $a\mathbb{E}X_1+b\mathbb{E}X_2$ is well defined, then $\mathbb{E}[aX_1+bX_2|Y=\bullet] = a\mathbb{E}[X_1|Y=\bullet]+b\mathbb{E}[X_2|Y=\bullet]$.

(c') $|\mathbb{E}[X_1|Y=\bullet]| \le \mathbb{E}[|X_1|\big|Y=\bullet]$

(d') For any convex function $r : \mathbb{R} \to \mathbb{R}$, $\mathbb{E}[r(X)|Y=\bullet] \ge r(\mathbb{E}[X|Y=\bullet])$.

Proof.

(a) Suppose $X_1 \le X_2$ $\mathbb{P}$-a.e. Then, for all $B \in \mathscr{G}$,

$$\int_B \mathbb{E}[X_1|Y]\mathrm{d}\mathbb{P} = \int_B X_1\mathrm{d}\mathbb{P} \le \int_B X_2\mathrm{d}\mathbb{P} = \int_B \mathbb{E}[X_2|Y]\mathrm{d}\mathbb{P}.$$

Thus, $\mathbb{E}[X_1|Y] \le \mathbb{E}[X_2|Y]$ $\mathbb{P}$-a.e. by Theorem 2.2.3.

(a') For any $\widetilde{B} \in \mathscr{H}$, we have

$$\int_{\widetilde{B}} \mathbb{E}[X_1|Y=\tilde{\omega}]\mathrm{d}\mathbb{P}^{(Y)}(\tilde{\omega}) = \int_{Y^{-1}(\widetilde{B})} X_1\mathrm{d}\mathbb{P} \le \int_{Y^{-1}(\widetilde{B})} X_2\mathrm{d}\mathbb{P} = \int_{\widetilde{B}} \mathbb{E}[X_2|Y=\tilde{\omega}]\mathrm{d}\mathbb{P}^{(Y)}(\tilde{\omega}).$$

(b) For any $B \in \mathscr{G}$, we have

$$\begin{aligned}\int_B \mathbb{E}[aX_1+bX_2|Y]\mathrm{d}\mathbb{P} &= \int_B aX_1+bX_2\mathrm{d}\mathbb{P}\\ &= a\int_B X_1\mathrm{d}\mathbb{P} + b\int_B X_2\mathrm{d}\mathbb{P}\\ &= a\int_B \mathbb{E}[X_1|Y]\mathrm{d}\mathbb{P} + b\int_B \mathbb{E}[X_2|Y]\mathrm{d}\mathbb{P}.\end{aligned}$$

(b') Similar to (a').

(c) Note that $|X_1| \ge X_1$ and $|X_1| \ge -X_1$. Thus,

$$\mathbb{E}\big[|X_1|\big||Y\big] \ge \max\{\mathbb{E}[X_1|Y], -\mathbb{E}[X_2|Y]\} = \big|\mathbb{E}[X_1|Y]\big|.$$

(c') Similar to (a').

(d) We note that for any $z \in \mathbb{R}$,

$$r(z) = \sup_{h\,:\,\mathbb{R}\to\mathbb{R}\text{ linear},\, h\le r} h(z).$$

Hence, by (a) and (b), we have that, for any linear function $h \le r$,

$$\mathbb{E}[r(X)|\mathscr{G}] \ge \mathbb{E}[h(X)|\mathscr{G}] = h\big(\mathbb{E}[X|\mathscr{G}]\big).$$

Since this is true for any linear $h \le r$, we have that $\mathbb{E}[r(X)|\mathscr{G}] \ge r\big(\mathbb{E}[X|\mathscr{G}]\big)$ as desired.

(d') Same as (d).

□

Theorem 6.2.2. Let $X_1, X_2, \ldots, X : \Omega \to \mathbb{R}$.

(a) If $0 \le X_1 \le X_2 \ldots$ and $X_n \to X$ $\mathbb{P}$ a.e., then

$$\mathbb{E}[X_n|\mathscr{G}] \to \mathbb{E}[X|\mathscr{G}] \text{ }\mathbb{P}\text{-a.e.} \quad \text{and} \quad \mathbb{E}[X_n|Y=\bullet] \to \mathbb{E}[X|Y=\bullet] \text{ }\mathbb{P}^{(Y)}\text{-a.e.}$$

(b) If $\exists Z : \Omega \to \mathbb{R}$ such that $|X_n| \leq Z$ $\mathbb{P}$-a.e. and $\mathbb{E}Z < \infty$ and if $X_n \to X$ $\mathbb{P}$-a.e., then

$$\mathbb{E}[X_n|\mathscr{G}] \to \mathbb{E}[X|\mathscr{G}] \ \mathbb{P}\text{-a.e.} \quad \text{and} \quad \mathbb{E}[X_n|Y = \bullet] \to \mathbb{E}[X|Y = \bullet] \ \mathbb{P}^{(Y)}\text{-a.e.}$$

Proof.

(a) By Theorem 6.2.1 (a), $0 \leq \mathbb{E}[X_1|\mathscr{G}] \leq \mathbb{E}[X_2|\mathscr{G}] \ldots$. Define $h : \Omega \to \overline{\mathbb{R}}$ as $h(\omega) = \lim_{n\to\infty} \mathbb{E}[X_n|\mathscr{G}](\omega)$, $\forall \omega \in \Omega$, then h is $(\Omega, \mathscr{G})/(\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$-measurable. For $B \in \mathscr{G}$,

$$\begin{aligned}
\int_B h \mathrm{d}\mathbb{P} &= \int_B \lim_{n\to\infty} \mathbb{E}[X_n|\mathscr{G}] \mathrm{d}\mathbb{P} \\
&= \lim_{n\to\infty} \int_B \mathbb{E}[X_n|\mathscr{G}] \mathrm{d}\mathbb{P} && \text{(MCT)} \\
&= \lim_{n\to\infty} \int_B X_n \mathrm{d}\mathbb{P} \\
&= \int_B X \mathrm{d}\mathbb{P}. && \text{(MCT)}
\end{aligned}$$

The proof of the second claim is identical.

(b) Define $W_n := \sup_{k\geq n} |X_n - X|$ so that $W_{n+1} \leq W_n$, $\forall n \in \mathbb{N}$ and $\forall \omega \in \Omega$, $W_n(\omega) \to 0$. Since $\mathbb{E}|X| \leq \mathbb{E}Z < \infty$, we have

$$\left| \int_\Omega \mathbb{E}[X|\mathscr{G}] \mathrm{d}\mathbb{P} \right| = |\mathbb{E}X| < \infty$$

and thus, $|\mathbb{E}[X|\mathscr{G}]| < \infty$ $\mathbb{P}$-a.e. Hence, $\mathbb{E}[X|\mathscr{G}] - \mathbb{E}[X_n|\mathscr{G}]$ is well-defined and

$$\mathbb{E}\big[|X - X_n|\mathscr{G}\big] \geq \big|\mathbb{E}[X|\mathscr{G}] - \mathbb{E}[X_n|\mathscr{G}]\big|.$$

Since $\mathbb{E}\big[|X - X_n|\big|\mathscr{G}\big] \leq \mathbb{E}[W_n|\mathscr{G}]$, we need only show that $\lim_{n\to\infty} \mathbb{E}[W_n|\mathscr{G}] = 0$. Define $h : \Omega \to [0, \infty)$ such that $\forall \omega \in \Omega$, $h(\omega) = \lim_{n\to\infty} \mathbb{E}[W_n|\mathscr{G}](\omega)$. Then

$$\int_\Omega h \mathrm{d}\mathbb{P} = \lim_{n\to\infty} \int_\Omega \mathbb{E}[W_n|\mathscr{G}] \mathrm{d}\mathbb{P} = \lim_{n\to\infty} \int_\Omega W_n \mathrm{d}\mathbb{P} = 0.$$

Thus, $h = 0$ $\mathbb{P}$-a.e.

□

Theorem 6.2.3. Let $X : \Omega \to \mathbb{R}$ be $(\Omega, \mathscr{F})/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable.

(a) $\mathbb{E}[X|\{\emptyset, \Omega\}] = \mathbb{E}[X]$.

(b) $\mathbb{E}[X|\mathscr{F}] = X$.

(c) If $\mathscr{G}_1 \subseteq \mathscr{G}_2 \subseteq \mathscr{F}$, then $\mathbb{E}[\mathbb{E}[X|\mathscr{G}_1]|\mathscr{G}_2] = \mathbb{E}[X|\mathscr{G}_1]$ and $\mathbb{E}[\mathbb{E}[X|\mathscr{G}_2]|\mathscr{G}_1] = \mathbb{E}[X|\mathscr{G}_1]$.

(d) If $Z : \Omega \to \mathbb{R}$ is $(\Omega, \mathscr{G})/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable and if $\mathbb{E}|X| < \infty$ and $\mathbb{E}|ZX| < \infty$, then

$$\mathbb{E}[ZX|\mathscr{G}] = Z \cdot \mathbb{E}[X|\mathscr{G}].$$

Proof.

(a) Since $\int_\Omega \mathbb{E}X \mathrm{d}\mathbb{P} = \int_\Omega X \mathrm{d}\mathbb{P}$ and $\int_\emptyset \mathbb{E}X \mathrm{d}\mathbb{P} = 0 = \int_\emptyset X \mathrm{d}\mathbb{P}$, the claim follows.

(b) Follows from definition.

(c) Since $\mathbb{E}[X|\mathscr{G}_1] : \Omega \to \mathbb{R}$ is also $(\Omega, \mathscr{G}_2)/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable, we have from (b) that $\mathbb{E}[\mathbb{E}[X|\mathscr{G}_1]\mathscr{G}_2] = \mathbb{E}[X|\mathscr{G}_1]$.

For the second claim, let $B \in \mathscr{G}_1$, then $B \in \mathscr{G}_2$ as well, hence

$$\int_B \mathbb{E}[\mathbb{E}[X|\mathscr{G}_2]|\mathscr{G}_1] \mathrm{d}\mathbb{P} = \int_B \mathbb{E}[X|\mathscr{G}_2] \mathrm{d}\mathbb{P} = \int_B X \mathrm{d}\mathbb{P}.$$

Since this is true $\forall B \in \mathscr{G}_1$ and since $\mathbb{E}[\mathbb{E}[X|\mathscr{G}_2]|\mathscr{G}_1] : \Omega \to \mathbb{R}$ is $(\Omega, \mathscr{G}_1)/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable, $\mathbb{E}[\mathbb{E}[X|\mathscr{G}_2]|\mathscr{G}_1] = \mathbb{E}[X|\mathscr{G}_1]$.

(d) By Theorem 6.1.3 and the fact that Z is $\mathscr{G}/\mathscr{B}(\mathbb{R})$–measurable, we have that, for any $A \in \mathscr{G}$,

$$\int_A Z\mathbb{E}[X|\mathscr{G}] d\mathbb{P} = \int_A ZX \, d\mathbb{P}.$$

Hence, it must be that $Z\mathbb{E}[X|\mathscr{G}] = \mathbb{E}[ZX|\mathscr{G}]$.

□

Example 6.2.1. (a) Let $X, Y : \Omega \to \mathbb{R}$. Then $\mathbb{E}[Y|X] = \mathbb{E}[Y|\alpha X]$, $\forall \alpha \neq 0$ since $\sigma(X) = \sigma(\alpha X)$. In fact, let $\phi : \mathbb{R} \to \mathbb{R}$ be invertible such that ϕ, ϕ^{-1} are both $(\mathbb{R}, \mathscr{B}(\mathbb{R}))/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable, then $\mathbb{E}[Y|X] = \mathbb{E}[Y|\phi(X)]$.

Note, however, for $x \neq 0$, $\mathbb{E}[Y|X = x] \neq \mathbb{E}[Y|2X = x]$, we have

$$\mathbb{E}[Y|X = \bullet](x) = \mathbb{E}[Y|X] = \mathbb{E}[Y|2X] = \mathbb{E}[Y|2X = \bullet](2x).$$

(b) Let $X, Y, Z : \Omega \to \mathbb{R}$ and let Z be independent of Y, X. Then $\mathbb{E}[Y|X, Z] = \mathbb{E}[Y|X]$.

To see this, note that

$$\sigma(X, Z) = \{(X, Z)^{-1}(C) : C \subseteq \mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R})\} = \sigma(\{X^{-1}(Z) \cap Z^{-1}(Y) : S, T \subseteq \mathscr{B}(\mathbb{R})\}.$$

For any $S, T \in \mathscr{B}(\mathbb{R})$, we have

$$\begin{aligned} \int_{X^{-1}(S) \cap Z^{-1}(T)} \mathbb{E}[Y|X] \mathrm{d}\mathbb{P} &= \int_{X^{-1}(S)} \mathbb{E}[Y|X] \mathbb{1}_{\{Z \in T\}} \mathrm{d}\mathbb{P} \\ &= \int_{X^{-1}(S)} \mathbb{E}[Y|X] \mathrm{d}\mathbb{P} \cdot \mathbb{P}(Z \in T) \\ &= \int_{X^{-1}(S)} Y \mathrm{d}\mathbb{P} \cdot \mathbb{P}(Z \in T) = \int_{X^{-1}(S) \cap Z^{-1}(T)} Y \mathrm{d}\mathbb{P}. \end{aligned}$$

Thus, $\mathbb{E}[Y|X, Z] = \mathbb{E}[Y|X]$.

(c) Recall that if $X : \Omega \to \mathbb{R}^p$ and $X \sim N(0, I_p)$, then for $v, w \in \mathbb{R}^p$ such that $v^{\mathrm{T}} w = 0$, $v^{\mathrm{T}} x$ is independent of $w^{\mathrm{T}} x$. Thus, for any $\beta, \gamma \in \mathbb{R}^p$, $\beta \neq 0$, we have

$$\gamma = \frac{\gamma^{\mathrm{T}} \beta}{\|\beta\|^2} \beta + \gamma'$$

where $\beta^{\mathrm{T}} \gamma' = 0$ and

$$\mathbb{E}[\gamma^{\mathrm{T}} X | \beta^{\mathrm{T}} X] = \mathbb{E}[\frac{\gamma^{\mathrm{T}} \beta}{\|\beta\|^2} \beta^{\mathrm{T}} X + (\gamma')^{\mathrm{T}} X | \beta^{\mathrm{T}} X] = \frac{\gamma^{\mathrm{T}} \beta}{\|\beta\|^2} \beta^{\mathrm{T}} X.$$

Note that $\frac{\gamma^{\mathrm{T}} \beta}{\|\beta\|^2} \beta^{\mathrm{T}} X = \gamma^{\mathrm{T}} \frac{\beta}{\|\beta\|} \cdot \frac{\beta^{\mathrm{T}}}{\|\beta\|} X$ does not depend on the norm of β.

(d) Suppose $X, Y, Z : \Omega \to \mathbb{R}$ and Z is independent of (X, Y). We claim that

$$\underbrace{\mathbb{E}(Y - \mathbb{E}[Y|X+Z])^2}_{\text{residual variance of error-in-variable}} \geq \mathbb{E}(Y - \mathbb{E}[Y|X])^2.$$

To see this, note that

$$\mathbb{E}(Y - \mathbb{E}[Y|X+Z])^2 \geq \mathbb{E}(Y - \mathbb{E}[Y|X,Z])^2 = \mathbb{E}(Y - \mathbb{E}[Y|X])^2.$$

To understand this more, observe that

$$\begin{aligned}
\mathbb{E}[Y|X+Z] &= \mathbb{E}[\mathbb{E}[Y|X+Z,Z]|X+Z] && (\text{since } \sigma((X+Z,Z)) \supseteq \sigma(X+Z)) \\
&= \mathbb{E}[\mathbb{E}[Y|X,Z]|X+Z] && (\text{since } \sigma((X+Z),Z)) = \sigma(X,Z)) \\
&= \mathbb{E}[\mathbb{E}[Y|X]|X+Z].
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}(Y - \mathbb{E}[Y|X])^2 &= \mathbb{E}\big\{Y - \mathbb{E}[Y|X+Z] + \mathbb{E}[Y|X+Z] - \mathbb{E}[Y|X]\big\}^2 \\
&= \mathbb{E}(Y - \mathbb{E}[Y|X+Z])^2 + \mathbb{E}\big\{\mathbb{E}[Y|X+Z] - \mathbb{E}[Y|X]\big\}^2 \\
&\qquad + 2\mathbb{E}\big[(Y - \mathbb{E}[Y|X+Z])(\mathbb{E}[Y|X+Z] - \mathbb{E}[Y|X])\big] \\
&= \mathbb{E}(Y - \mathbb{E}[Y|X+Z])^2 - \mathbb{E}\big\{\mathbb{E}[Y|X+Z] - \mathbb{E}[Y|X]\big\}^2.
\end{aligned}$$

where

$$\begin{aligned}
&2\mathbb{E}\big[(Y - \mathbb{E}[Y|X+Z])(\mathbb{E}[Y|X+Z] - \mathbb{E}[Y|X])\big] \\
=&2\mathbb{E}\big[\mathbb{E}\big\{(Y - \mathbb{E}[Y|X+Z])(\mathbb{E}[Y|X+Z] - \mathbb{E}[Y|X])|X,Z\big\}\big] \\
=&-2\mathbb{E}(\mathbb{E}[Y|X] - \mathbb{E}[Y|X+Z])^2.
\end{aligned}$$

Consider a concrete case: $X, Z \sim N(0, I_p)$ and $\mathbb{E}[Y|X] = \alpha X$ for some $\alpha \neq 0$. Then

$$\begin{aligned}
\mathbb{E}[Y|X+Z] = \mathbb{E}[\mathbb{E}[Y|X]|X+Z] &= \alpha \mathbb{E}[X|X+Z] \\
&= \alpha \frac{(1,0)^\top (1,1)}{2}(X+Z) = \frac{1}{2}\alpha(X+Z).
\end{aligned}$$

More generally, for $\lambda > 0$,

$$\mathbb{E}[Y|X+\lambda Z] = \alpha \frac{(1,0)^\top (1,\lambda)}{1+\lambda^2}(X+Z) = \frac{\alpha}{1+\lambda^2}(X+Z).$$

6.3 Regular Conditional Probability

Remark 6.3.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ and $(\widetilde{\Omega}, \mathscr{H})$ be probability spaces and let $Y : \Omega \to \widetilde{\Omega}$ be $(\Omega, \mathscr{F})/(\widetilde{\Omega}, \mathscr{H})$-measurable.

Our goal is to use conditional expectation to construct a Markov Kernel $\mathcal{K} : \widetilde{\Omega} \to \mathscr{F} \to [0,1]$, i.e.

- $\{\mathcal{K}(y, \bullet)\}_{y \in \widetilde{\Omega}}$ are probability measure on $(\Omega, \mathscr{F})$
- For any $B \in \mathscr{F}$, $y \mapsto \mathcal{K}(y, B)$ is $(\widetilde{\Omega}, \mathscr{H})/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable.
- For any $B \in \mathscr{F}$, $\widetilde{A} \in \mathscr{H}$,

$$\int_{\widetilde{A}} \mathcal{K}(y, B) \mathrm{d}\mathbb{P}^{(Y)}(y) = \mathbb{P}(B \cap Y^{-1}(\widetilde{A})).$$

We interpret $\{\mathscr{K}(y,\bullet)\}_{y\in\tilde{\Omega}}$ as the condition probability on $(\Omega,\mathscr{F})$ conditional on $Y=y$. Note that $\{\mathscr{K}(y,\bullet)\}_{y\in\tilde{\Omega}}$ is only unique for $\mathbb{P}^{(Y)}$-a.e. $y\in\tilde{\Omega}$.

One attempt to construct the Markov kernel is to note that, for any $B\in\mathscr{F}$, the indicator $\mathbb{1}_B:\Omega\to\mathbb{R}$ is $(\Omega,\mathscr{F})/(\mathbb{R},\mathscr{B}(\mathbb{R}))$-measurable and we may define

$$\forall y\in\widetilde{\Omega},\quad \mathbb{P}(B|Y=y):=\mathbb{E}[\mathbb{1}_B|Y=y].$$

Thus, we may consider defining $\widetilde{\mathscr{K}}:\widetilde{\Omega}\to\widetilde{\mathscr{F}}\to[0,1]$ by

$$\widetilde{\mathscr{K}}(y,B):=\mathbb{E}[\mathbb{1}_B|Y=y].$$

But this may not be well-defined. Recall that, for a fixed $B\in\mathscr{F}$, $\mathbb{E}[\mathbb{1}_B|Y=\bullet]$ is only unique $\mathbb{P}^{(Y)}$-a.e. In other words, $\exists\widetilde{E}_B\in\mathscr{H}$ such that $\mathbb{P}^{(Y)}(\widetilde{E}_B)=0$ and $\mathbb{E}[\mathbb{1}_B|Y=\bullet]$ is unique for $y\in\widetilde{E}_B^c$. Therefore, $B\mapsto\widetilde{\mathscr{K}}(y,B)$ is only well-defined for all $y\in\cup_{B\in\mathscr{F}}\widetilde{E}_B$. However, it may be that $\mathbb{P}^{(Y)}(\cup_{B\in\mathscr{F}}\widetilde{E}_B)>0$. Luckily, this is not the case so long as Ω is a complete separable metric space.

Theorem 6.3.1. Let $(\Omega,\mathscr{F})$ be a measure space and let $(\mathcal{X},\mathscr{G})$ be a complete and separable metric space with Borel σ-field $\mathscr{G}$. Let $X:\Omega\to\mathcal{X}$ be $\mathscr{G}/\mathscr{F}$–measurable.

For any probability measure $\mathbb{P}$ on $(\Omega,\mathscr{F})$, there exists a Markov kernel $\mathscr{K}:\Omega\times\mathscr{G}\to[0,1]$ such that

1. $\{\mathscr{K}(\omega,\bullet)\}_{\omega\in\Omega}$ is a family of probability measures on $(\mathcal{X},\mathscr{G})$,
2. for any $S\in\mathscr{G}$, $\omega\mapsto\mathscr{K}(\omega,S)$ is $\mathscr{F}/\mathscr{B}([0,1])$–measurable, for any $A\in\mathscr{F}$ and $S\in\mathscr{G}$,

$$\mathbb{P}(A\cap\{X\in S\})=\int_A\mathscr{K}(\omega,S)\,d\mathbb{P}(\omega).$$

Note that we may define probability measure $\mathbb{Q}$ on the product space $(\mathcal{X}\times\Omega,\ \mathscr{H}\otimes\mathscr{F})$ by $\mathbb{Q}(S,A)=\mathbb{P}(A\cap\{X\in S\})$ for any $S\in\mathscr{H}$ and $A\in\mathscr{F}$.

Proof.
See Theorem 3.6 in the note by P. Orbanz. □

We also call $\mathscr{K}$ regular conditional probability measures or regular conditional distribution.
The following is an equivalent formulation of the same theorem.

Theorem 6.3.2. (Version 2)
Let $(\Omega,\mathscr{F})$ and $(\mathcal{Y},\mathscr{H})$ be measurable space. Let $(\mathcal{X},\mathscr{G})$ be a complete and separable metric space with the Borel σ-field $\mathscr{G}$. Let $X:\Omega\to\mathcal{X}$ and $Y:\Omega\to\mathcal{Y}$ be r.o. (measurable function). Then, for any probability measure $\mathbb{P}$ on $(\Omega,\mathscr{F})$, there exists Markov kernel $\mathbb{P}^{(X)}(\bullet|Y=\bullet):\mathcal{Y}\times\mathscr{G}\to[0,1]$ such that

(1) $\{\mathbb{P}^{(X)}(\bullet|Y=y)\}_{y\in\mathcal{Y}}$ is a family of probability measure on $(\mathcal{X},\mathscr{G})$.

(2) $\forall S\in\mathscr{G}$, $y\mapsto\mathbb{P}^{(X)}(S\,|\,Y=y)$ is $(\mathcal{Y},\mathscr{H})/(\mathbb{R},\mathscr{B}(\mathbb{R}))$-measurable.

(3) $\forall S\in\mathscr{G}$, $\forall T\in\mathscr{H}$,

$$\mathbb{P}(\{X\in S\}\cap\{Y\in T\})=\mathbb{P}^{(X,Y)}(S\times T)=\int_T\mathbb{P}^{(X)}(S\,|\,Y=y)\mathrm{d}\mathbb{P}^{(Y)}(y).$$

Remark 6.3.2. We can consider a third equivalent formulation. If $(\Omega,\mathscr{F})$ is complete and separable with the Borel σ-field, then we may take $X:\Omega\to\Omega$ as the identity so that $\mathbb{P}^{(X)}=\mathbb{P}$. Then there exists a Markov kernel $\mathbb{P}(\bullet|Y=\bullet):\mathcal{Y}\times\mathscr{F}\to[0,1]$

(1) $\{\mathbb{P}(\bullet|Y=y)\}_{y\in\mathcal{Y}}$ is a family of probability measure on $(\Omega,\mathscr{F})$.

(2) $\forall B \in \mathscr{F}$, $y \mapsto \mathbb{P}(B \,|\, Y = y)$ is $(\mathcal{Y}, \mathscr{H})/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable.

(3) $\forall B \in \mathscr{F}, \forall T \in \mathscr{H}$,

$$\mathbb{P}(B \cap \{Y \in T\}) = \int_T \mathbb{P}(B \,|\, Y = y) \mathrm{d}\mathbb{P}^{(Y)}(y).$$

Remark 6.3.3. Note that for $X : \Omega \to \mathbb{R}$ r.v., we have that $\mathbb{E}[X|Y = y] = \int_\Omega X \mathrm{d}\mathbb{P}(\,\bullet\,|Y = y)$. To see this, first note that if X is simple and non-negative, i.e. $X = \sum_{i=1}^m c_i \mathbb{1}_{A_i}$ for $c_1, \ldots, c_m > 0$ and $A_1, \ldots, A_m \in \mathscr{F}$, then for any $\widetilde{A} \in \mathscr{H}$,

$$\begin{aligned}
\int_{\widetilde{A}} \int_\Omega X \mathrm{d}\mathbb{P}(\,\bullet\,|Y = y) \mathrm{d}\mathbb{P}^{(Y)}(y) &= \int_{\widetilde{A}} \int_\Omega \sum_{i=1}^m c_i \mathbb{1}_{A_i} \mathrm{d}\mathbb{P}(\,\bullet\,|Y = y) \mathrm{d}\mathbb{P}^{(Y)}(y) \\
&= \sum_{i=1}^m c_i \int_{\widetilde{A}} \mathbb{P}(A_i|Y = y) \mathrm{d}\mathbb{P}^{(Y)}(y) \\
&= \sum_{i=1}^m c_i \mathbb{P}(A_i \cap Y^{-1}(\widetilde{A})) \\
&= \int_{Y^{-1}(\widetilde{A})} \sum_{i=1}^m c_i \mathbb{1}_{A_i} \mathrm{d}\mathbb{P} = \int_{Y^{-1}(\widetilde{A})} X \mathrm{d}\mathbb{P}.
\end{aligned}$$

Chapter 7

Martingales

7.1 Introduction

Definition 7.1.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \cdots \subseteq \mathscr{F}$ be an increasing sequence of σ-fields (filtration). Let $X_1, X_2, \cdots : \Omega \to \mathbb{R}$ be such that $\forall n \in \mathbb{N}$, X_n is $(\Omega, \mathscr{F}_n)/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable and $\mathbb{E}|X_n| < \infty$. We say that $X_1, X_2, \ldots,$ is a

- martingale if $\forall n \in \mathbb{N}$, $\mathbb{E}[X_{n+1}|\mathscr{F}_n] = X_n$,
- sub-martingale if $\forall n \in \mathbb{N}$, $\mathbb{E}[X_{n+1}|\mathscr{F}_n] \geq X_n$,
- super-martingale if $\forall n \in \mathbb{N}$, $\mathbb{E}[X_{n+1}|\mathscr{F}_n] \leq X_n$.

Let $\mathscr{F} \supseteq \mathscr{F}_1 \supseteq \mathscr{F}_2 \supseteq \mathscr{F}_3 \ldots$ be a decreasing sequence of σ-filed. Suppose $X_n : \Omega \to \mathbb{R}$ is $(\Omega, \mathscr{F}_n)/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable. We say that $X_1, X_2, \ldots$ is

- reverse martingale if $\forall n \in \mathbb{N}$, $\mathbb{E}[X_n|\mathscr{F}_{n+1}] = X_{n+1}$,
- reverse sub-martingale if $\forall n \in \mathbb{N}$, $\mathbb{E}[X_n|\mathscr{F}_{n+1}] \geq X_{n+1}$,
- reverse super-martingale if $\forall n \in \mathbb{N}$, $\mathbb{E}[X_n|\mathscr{F}_{n+1}] \leq X_{n+1}$.

Remark 7.1.1. (a) If $\{X_, \mathscr{F}_n\}_{n+1}^{\infty}$ is a martingale, then $\forall n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E}[X_{n+2}|\mathscr{F}_n] =& \mathbb{E}[\mathbb{E}[X_{n+2}|\mathscr{F}_{n+1}]|\mathscr{F}_n] \\ =& \mathbb{E}[X_{n+1}|\mathscr{F}_n] = X_n. \qquad \text{(Similar for super/sub-martingales)} \end{aligned}$$

(b) We say that $\{X_n\}_{n=1}^{\infty}$ is a martingale if it satisfies

$$\mathbb{E}[X_{n+1}|X_1, \ldots, X_n] = \mathbb{E}[X_{n+1}\,|\,\sigma(X_1, \ldots, X_n)] = X_n.$$

That is, writing $\mathscr{F}_n := \sigma(X_1, \ldots, X_n)$, $\{X_n, \mathscr{F}_n\}_n$ is a martingale.

Example 7.1.1. (a) Let $Y_1, Y_2, \cdots : \Omega \to \mathbb{R}$ be independent random variables with mean zero. For $n \in \mathbb{N}$, let $\mathscr{F}_n := \sigma(Y_1, \ldots, Y_n)$ and $X_n = \sum_{i=1}^n Y_i$. Then we have

$$\mathbb{E}[X_{n+1}|\mathscr{F}_n] = \mathbb{E}\left[\sum_{i=1}^n Y_i + Y_{n+1}|Y_1, \ldots, Y_n\right] = \sum_{i=1}^n Y_i = X_n$$

which implies $\{X_n, \mathscr{F}_n\}$ is a martingale.

(b) Let $Y_1, Y_2, \dots : \Omega \to \mathbb{R}$ be independent and write, for $j \in \mathbb{N}$, $\mu_j = \mathbb{E}Y_j \neq 0$. For $n \in \mathbb{N}$, let $\mathscr{F}_n = \sigma(Y_1, \dots, Y_n)$, $X_n = \prod_{j=1}^n \frac{Y_j}{\mu_j}$. Then

$$\mathbb{E}[X_{n+1}|\mathscr{F}_n] = \mathbb{E}\left[\frac{Y_{n+1}}{\mu_{n+1}} \prod_{j=1}^n \frac{Y_j}{\mu_j} | Y_1, \dots, Y_n\right] = \prod_{j=1}^n \frac{Y_j}{\mu_j} = X_n.$$

So $\{X_n, \mathscr{F}_n\}$ is a martingale.

(c) Let $Y : \Omega \to \mathbb{R}$ be such that $\mathbb{E}|Y| < \infty$ and let $\mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \dots \subseteq \mathscr{F}$. For $n \in \mathbb{N}$, let $X_n = \mathbb{E}[Y|\mathscr{F}_n]$. Then

$$\mathbb{E}[X_{n+1}|\mathscr{F}_n] = \mathbb{E}[\mathbb{E}[Y|\mathscr{F}_{n+1}|\mathscr{F}_n] = \mathbb{E}[Y|\mathscr{F}_n] = X_n.$$

So $\{X_n, \mathscr{F}_n\}$ is a martingale. Now let $\mathscr{F} \supseteq \mathscr{F}_1 \supseteq \mathscr{F}_2 \supseteq \dots$ be decreasing. Let $X_n = \mathbb{E}[Y|\mathscr{F}_n]$, then

$$\mathbb{E}[X_n|\mathscr{F}_{n+1}] = \mathbb{E}[\mathbb{E}[Y|\mathscr{F}_n]|\mathscr{F}_{n+1}] = \mathbb{E}[Y|\mathscr{F}_{n+1}] = X_{n+1}.$$

So $\{X_n, \mathscr{F}_n\}$ is a reverse martingale.

In particular, let $Z_1, Z_2, \dots : \Omega \to \mathbb{R}$ be iid. For $n \in \mathbb{N}$, let $S_n = \sum_{i=1}^n Z - i$. Then, we claim $\forall n \in \mathbb{N}$, $\frac{S_n}{n} = \mathbb{E}[Z_1|S_n] = \mathbb{E}[Z_1|S_n, Z_{n+1}, \dots]$.

To see this, let $B \in \mathscr{B}(\mathbb{R})$, then

$$\begin{aligned}
\int_{S_n \in B} Z_1 \mathrm{d}\mathbb{P} &= \int_{\{z_1 + \dots + z_n \in B\}} z_1 \mathrm{d}\mathbb{P}^{(Z_1, \dots, Z_n)}(z_1, \dots, z_n) \\
&= \frac{1}{n} \sum_{i=1}^n \int_{\{z_1 + \dots + z_n \in B\}} z_i \mathrm{d}\mathbb{P}^{(Z_1, \dots, Z_n)}(z_1, \dots, z_n) \\
&= \frac{1}{n} \int_{S_n \in B} \sum_{i=1}^n Z_i \mathrm{d}\mathbb{P} = \int_{S_n \in B} \frac{S_n}{n} \mathrm{d}\mathbb{P}.
\end{aligned}$$

Thus, for $n \in \mathbb{N}$, letting $\mathscr{F}_n = \sigma(S_n, Z_{n+1}, Z_{n+2}, \dots)$, we have that $\mathscr{F} \supseteq \mathscr{F}_1 \supseteq \mathscr{F}_2 \dots$ and $\{\frac{S_n}{n}, \mathscr{F}_n\}$ is a reverse martingale.

(d) (Polya's Urn). Let $n^{(0)}, n_R^{(0)} \in \mathbb{N}$ where $n^{(0)}$ is the total number of balls and $n_R^{(0)}$ is the number of red balls. For $t = 1, 2, \dots, n^{(0)}$, we have that $n^{(t)} = n^{(t-1)} + 1$ and

$$n_R^{(t)} = \begin{cases} n_R^{(t-1)} + 1 & \text{with probability } \frac{n_R^{(t-1)}}{n^{(t-1)}} \\ n_R^{(t-1)} & \text{with probability } 1 - \frac{n_R^{(t-1)}}{n^{(t-1)}} \end{cases}.$$

Let $t \in [n^{(0)}]$, let $\mathscr{F}_t := \sigma(n_R^{(1)}, \dots, n_R^{(t)})$, and let $X_t := \frac{n_R^{(t)}}{n^{(t)}}$. Then

$$\begin{aligned}
\mathbb{E}[X_{t-1}|\mathscr{F}_t] = \mathbb{E}\left[\frac{n_R^{(t+1)}}{n^{(t+1)}} | \mathscr{F}_t\right] &= \frac{n_R^{(t)} + 1}{n^{(t+1)}} \cdot \frac{n_R^{(t)}}{n^{(t)}} + \frac{n_R^{(t)}}{n^{(t+1)}} \cdot \frac{n^{(t)} - n_R^{(t)}}{n^{(t)}} \\
&= \frac{n_R^{(t)} n^{(t)} + n_R^{(t)}}{n^{(t+1)} n^{(t)}} = \frac{n_R^{(t)}}{n^{(t)}} = X_t.
\end{aligned}$$

So $\{X_n, \mathscr{F}_n\}$ is a martingale.

(e) (Stochastic gradient descent) Let $X : \Omega \to \mathbb{R}^p$ be a random vector and $Y : \Omega \to \mathbb{R}$ be a random variable. For simplicity, let us also suppose $\mathbb{E}[XX^\top] = I_p$.

We wish to minimize $F : \mathbb{R}^p \to [0, \infty)$, $F(\beta) = \mathbb{E}[(Y - X^\top \beta)^2]$. Write

$$\beta^* := \arg\min_{\beta \in \mathbb{R}^p} F(\beta) = \mathbb{E}[XY],$$

and note that β^* is the point where the gradient

$$\nabla F(\beta) = \mathbb{E}XX^\top \beta - \mathbb{E}XY$$

is equal to 0. If we know ∇F, we could perform gradient descent; but ∇F is unobserved. In stochastic gradient descent (SGD), we assume that we can observe a noisy version of the gradient.

More precisely, let $\beta^{(0)} \in \mathbb{R}^p$. For $t = 1, 2, 3, \ldots$, let $W_t : \Omega \to \mathbb{R}^p$ be independent, $\mathbb{E}W_t = 0$ and assume $\mathbb{E}\left\|W_t\right\|^2 \leq p$. We suppose that we observe the stochastic gradient $\nabla F(\beta^{(t-1)}) + W_t$.

For example, if (X_t, Y_t) is a single observation, we may have

$$\nabla F(\beta^{(t-1)}) + W_t := X_t X_t^\top \beta^{(t-1)} - X_t Y_t.$$

If X_t, Y_t have finite fourth moment and if the entries of $\beta^{(t-1)}$ are bounded, then we may verify that $\mathbb{E}\|W_t\|^2 \leq p$.

The SGD update, with step-size $1/t$, is then

$$\begin{aligned}
\beta^{(t)} &= \beta^{(t-1)} + \frac{1}{t}(\nabla F(\beta^{(t-1)}) + W_t) \\
&= \beta^{(t-1)} + \frac{1}{t}\big(\mathbb{E}[XX^\top]\beta^{(t-1)} - \mathbb{E}[XY] + W_t\big) \\
&= \beta^{(t-1)} + \frac{1}{t}(\beta^{(t-1)} - \beta^\star + W_t).
\end{aligned}$$

For $t \in \mathbb{N}$, let $\mathscr{F}_t := \sigma(W_1, \ldots, W_t)$ and let $Z_t = \left\|\beta^\star - \beta^{(t)}\right\|^2 + \frac{p}{t}$. Then,

$$\begin{aligned}
\mathbb{E}[Z_t|\mathscr{F}_{t-1}] &= \mathbb{E}_{\cdot|\mathscr{F}_{t-1}} \left\|\beta^\star - \beta^{(t-1)} - \frac{1}{t}(\beta^{(t-1)} - \beta^\star) - \frac{1}{t}W_t\right\|^2 + \frac{p}{t} \\
&= (1 - \frac{1}{t})^2 \left\|\beta^\star - \beta^{(t-1)}\right\|^2 + \frac{1}{t^2}\mathbb{E}\left\|W_t\right\|^2 + \frac{p}{t} \\
&\leq \left\|\beta^\star - \beta^{(t-1)}\right\|^2 + \frac{p}{t}(1 + \frac{1}{t}) \leq Z_{t-1}.
\end{aligned}$$

Thus, $\{Z_t, \mathscr{F}_t\}$ is a super martingale.

Lemma 7.1.1. $\{X_n, \mathscr{F}_n\}$ is a martingale if and only if $\forall n \in \mathbb{N}$, $\forall A \in \mathscr{F}_n$, $\int_A X_n \mathrm{d}\mathbb{P} = \int_A X_{n+1} \mathrm{d}\mathbb{P}$. (sub-martingale if and only if $\int_A X_n \mathrm{d}\mathbb{P} \leq \int_A X_{n+1}\mathrm{d}\mathbb{P}$.)

In particular, for martingale, $\mathbb{E}[X_1] = \mathbb{E}[X_2] = \ldots$.

Proof.

Recall that $X_{n+1} = \mathbb{E}[X_n|\mathscr{F}_n]$ iff $\forall A \in \mathscr{F}_n$, $\int_A X_{n+1}\mathrm{d}\mathbb{P} = \int_A X_n \mathrm{d}\mathbb{P}$. □

Lemma 7.1.2. (a) If $\{X_n, \mathscr{F}_n\}$ and $\{Y_n, \mathscr{F}_n\}$ are sub-martingales, then $\{\max(X_n, Y_n), \mathscr{F}_n\}$ is also a sub-martingale.

(b) Let $\{X_n, \mathscr{F}_n\}$ be a sub-martingale. Let $g : \mathbb{R} \to \mathbb{R}$ be convex and increasing and suppose $\mathbb{E}|g(X_n)| < \infty$. Then $\{g(X_n), \mathscr{F}_n\}$ is a sub-martingale. e.g. $\{X_n^+, \mathscr{F}_n\}$ is a sub-martingale.

If $\{X_n, \mathscr{F}_n\}$ is a martingale and $g : \mathbb{R} \to \mathbb{R}$ is convex and $\mathbb{E}|g(X_n)| < \infty$, then $\{g(X_n), \mathscr{F}_n\}$ is a sub-martingale.

Proof.

(a) Note that $\mathbb{E}[\max(X_{n+1}, Y_{n+1})|\mathscr{F}_n] \geq \max(\mathbb{E}[X_{n+1}|\mathscr{F}_n], \mathbb{E}[Y_{n+1}|\mathscr{F}_n]) \geq \max(X_n, Y_n)$.

(b) By Jensen's inequality, we have $\mathbb{E}[g(X_{n+1})|\mathscr{F}_n] \geq g(\mathbb{E}[X_{n+1}|\mathscr{F}_n]) \geq g(X_n)$ if either g is increasing or if $\mathbb{E}[X_{n+1}|\mathscr{F}_n] = X_n$.

$\square$

7.2 Martingale Convergence

Definition 7.2.1. Let $\mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \dots$ be a filtration. We say that a sequence of random variables $A_2, A_3, \dots$ is a **predictable** process if for all $k \geq 2$, A_k is $\mathscr{F}_{k-1}/\mathscr{B}(\mathbb{R})$–measurable.

Suppose $\{X_n, \mathscr{F}_n\}_{n\in\mathbb{N}}$ is a sub-martingale (or martingale), then we may define a sequence of random variables $\{(X \cdot A)_n\}_{n\in\mathbb{N}}$ by

$$\begin{aligned}
(X \cdot A)_1 &= X_1 \\
(X \cdot A)_2 &= X_1 + A_2(X_2 - X_1) \\
&\dots \\
(X \cdot A)_n &= X_1 + \sum_{k=2}^{n} A_k(X_k - X_{k-1}).
\end{aligned}$$

We call $(X \cdot A)$ the **martingale transform** of $\{A_n\}$.

Lemma 7.2.1 (Optional skipping theorem)**.** Let $\{X_n, \mathscr{F}_n\}$ be as sub-martingale and let $\{A_k\}_{k\geq 2}$ be a predictable process. Suppose $\mathbb{E}|A_k(X_k - X_{k-1})| < \infty$ for all $k \geq 2$. (We can use a stronger assumption that $\sup_k |A_k| < M < \infty$ almost surely. This is easier to verify.)

1. If $A_k \geq 0$, then $(X \cdot A)$ is a sub-martingale.

2. If $\{X_n\}_n$ is a martingale, then $(X \cdot A)$ is a martingale and $\mathbb{E}(X \cdot A)_n = \mathbb{E}X_n = \mathbb{E}X_1$.

Moreover, if $\{X_n\}$ is a sub-martingale and $A_n \in [0, 1]$, then $\mathbb{E}(X \cdot A)_n \leq \mathbb{E}X_n$.

Proof.

For $n \in \mathbb{N}$, let us write $Y_n := (X \cdot A)_n = X_1 + \sum_{k=2}^{n} A_k(X_k - X_{k-1})$. It is clear that Y_n is $\mathscr{F}_n/\mathscr{B}(\mathbb{R})$–measurable and $\mathbb{E}|Y_n| < \infty$. Since $\{A_n\}_{n\geq 2}$ is a predictable process, we have, for any $n \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{E}[Y_{n+1}|\mathscr{F}_n] &= \mathbb{E}[Y_n + A_n(X_n - X_{n-1}) \,|\, \mathscr{F}_n] \\
&= Y_n + A_n\mathbb{E}[X_{n+1} - X_n \,|\, \mathscr{F}_n] \begin{cases} = Y_n & \text{if } \{X_n, \mathscr{F}_n\} \text{ martingale} \\ \geq Y_n & \text{if } \{X_n, \mathscr{F}_n\} \text{ sub-martingale and } A_n \geq 0 \end{cases}
\end{aligned}$$

If $\{X_n, \mathscr{F}_n\}$ is a martingale, then it is clear that $\mathbb{E}[Y_n] = \mathbb{E}[Y_1] = \mathbb{E}[X_1]$.

Assume now that $\{X_n, \mathscr{F}_n\}$ is a sub-martingale and that $A_n \in [0, 1]$. Since $Y_1 = X_1$, $\mathbb{E}[Y_1] = \mathbb{E}[X_1]$. For $k \in \mathbb{N}$, assume as inductive hypothesis that $\mathbb{E}[X_k - Y_k] \geq 0$. Then

$$X_{k+1} - Y_{k+1} = X_{k+1} - Y_k - A_{k+1}(X_{k+1} - X_k) = (1 - A_k)(X_{k+1} - X_k) + (X_k - Y_k).$$

Thus,

$$\begin{aligned}
\mathbb{E}[X_{k+1} - Y_{k+1}|\mathscr{F}_k] &= (1 - A_k)\mathbb{E}[X_{k+1} - X_k|\mathscr{F}_k] + (X_k - Y_k) \\
&\geq X_k - Y_k
\end{aligned}$$

Thus, $\mathbb{E}[X_{k+1} - Y_{k+1}] \geq \mathbb{E}[X_k - Y_k] \geq 0$. $\square$

Let $(X_1, \mathscr{F}_1), \dots, (X_n, \mathscr{F}_n)$ be a sub-martingale. Fix $a < b \in \mathbb{R}$. Define

$$
\begin{aligned}
&T_1 : \Omega \to \mathbb{N} \text{ s.t. } \forall \omega \in \Omega, T_1(\omega) = \min\{i \in [n] : X_i(\omega) \le a\}, \text{ or } \infty \text{ if } X_i(\omega) > a, \forall i \in [n] \\
&T_2 : \Omega \to \mathbb{N} \text{ s.t. } \forall \omega \in \Omega, T_1(\omega) = \min\{i \ge T_1(\omega) : X_i(\omega) \ge b\}, \dots\dots \\
&T_3 : \Omega \to \mathbb{N} \text{ s.t. } \forall \omega \in \Omega, T_1(\omega) = \min\{i \ge T_2(\omega) : X_i(\omega) \le a\}, \dots\dots \\
&T_4 : \Omega \to \mathbb{N} \text{ s.t. } \forall \omega \in \Omega, T_1(\omega) = \min\{i \ge T_3(\omega) : X_i(\omega) \ge b\}, \dots\dots \\
&\qquad\qquad\qquad\vdots
\end{aligned}
$$

For $\omega \in \Omega$, define $N(\omega) := |\{T_1(\omega), T_2(\omega), \dots : \text{finite}\}|$ and define

$$
U_{ab}(\omega) = \begin{cases} \dfrac{N(\omega)}{2} & \text{if } N(\omega) \text{ even,} \\ \dfrac{N(\omega) - 1}{2} & \text{if } N(\omega) \text{ odd.} \end{cases}
$$

So U_{ab} is the number of up-crossing.

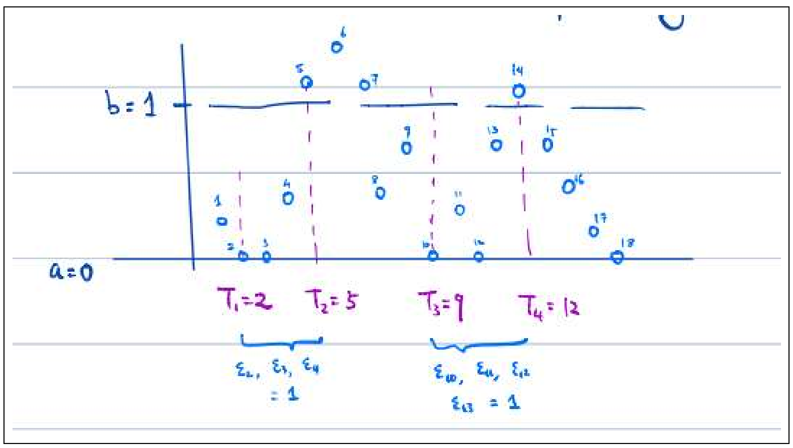


Figure 7.1: Upcrossing

Theorem 7.2.1 (Doob's Upcrossing Theorem). Let $\{X_i, \mathscr{F}_i\}_{i=1}^n$ be a sub-martingale. Then

$$
\mathbb{E}U_{ab} \le \frac{1}{b-a}\mathbb{E}[(X_n - a)_+].
$$

Proof.

First assume that $a = 0$ and that $X_j \ge 0$, $\forall j \in [n]$. For $j \in [n-1]$, define $\varepsilon_j : \Omega \to \{0, 1\}$ by

$$
\varepsilon_j(\omega) = \begin{cases} 1 & \text{if } T_1(\omega) \le j < T_2(\omega), \text{ or } T_3(\omega) \le j < T_4(\omega), \text{ or } T_5(\omega) \le j < T_6(\omega), \dots \\ 0 & \text{else.} \end{cases}
$$

Note that ε_j is a binary function of $X_1, X_2, \dots, X_j$ and hence ε_j is $(\Omega, \mathscr{F}_j)/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable. Note also if $\varepsilon_{j-1}(\omega) = 0$, $\varepsilon_j(\omega) = 1$, $\varepsilon_{j+1}(\omega) = 1$, $\dots$, $\varepsilon_{j+k}(\omega) = 1$, $\varepsilon_{j+k+1}(\omega) = 0$, then $X_{j-1}(\omega) > 0$, $X_j(\omega) = 0$, and $X_{j+1}(\omega)$, $\dots$, $X_{j+k}(\omega) \in [a, b)$ and $X_{j+k+1}(\omega) \ge b$.

Define $Y_n = X_1 + \varepsilon_1(X_2 - X_1) + \varepsilon_2(X_3 - X_2) + \cdots + \varepsilon_{n-1}(X_n - X_{n-1})$. As an example, in Figure 7.1, we have that

$$
\begin{aligned}
Y_n(\omega) = {} & X_1(\omega) + (X_3(\omega) - X_2(\omega)) + (X_4(\omega) - X_3(\omega)) + (X_5(\omega) - X_4(\omega)) \\
& + (X_{11}(\omega) - X_{10}(\omega)) + (X_{12}(\omega) - X_{11}(\omega)) + (X_{13}(\omega) - X_{12}(\omega)) + (X_{14}(\omega) - X_{13}(\omega)) \\
= {} & X_1(\omega) + (X_5(\omega) - X_3(\omega)) + (X_{14}(\omega) - X_{11}(\omega)) \ge 2b.
\end{aligned}
$$

Thus, $Y_n \geq bU_{ab}$. Since ε_j is $(\Omega, \mathscr{F}_j)/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable and takes value in $\{0, 1\}$, we have by lemme 7.2.1, that $\mathbb{E}U_{ab} \leq \frac{1}{b}\mathbb{E}Y_n \leq \frac{1}{b}\mathbb{E}X_n$ as desired.

Now consider the general case. Then $\{(X_j - a)_+, \mathscr{F}_j\}_{j=1}^n$ is also a sub-martingale by Lemma 7.1.2. Define $\widetilde{U}_{b-a}$ as the number of $(0, b-a)$ up-crossing of $\{(X_j - a)_+, \mathscr{F}_j\}_{j=1}^n$, then $U_{ab} = \widetilde{U}_{b-a}$ and thus, $\mathbb{E}U_{ab} = \mathbb{E}\widetilde{U}_{b-a} \leq \frac{1}{b-a}\mathbb{E}(X_n - a)_+$, as desired. □

Theorem 7.2.2. Let $\{X_n, \mathscr{F}_n\}_{n+1}^\infty$ be a sub-martingale and define $\mathscr{F}_\infty := \sigma(\cup_{n=1}^\infty \mathscr{F}_n)$. If $\sup_{n\in\mathbb{N}} \mathbb{E}[(X_n)_+] < \infty$, then there exists a random variable $X_\infty : \Omega \to \mathbb{R}$, $(\Omega, \mathscr{F}_\infty)/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable such that $\mathbb{E}|X_\infty| < \infty$ and $X_n \to X_\infty$ a.s.

Proof.

Let $E = \{\omega \in \Omega : \nexists x_0 \in \mathbb{R} \text{ s.t. } X_n(\omega) \to x_0\}$. Then,

$$E = \bigcup_{\substack{a<b \\ \text{rational}}} \{\omega \in \Omega : \liminf_{n\to\infty} X_n(\omega) < a < b < \limsup_{n\to\infty} X_n(\omega)\}.$$

Suppose for sake of contradiction that $\mathbb{P}(E) > 0$. Then, $\exists a < b \in \mathbb{Q}$ such that

$$\mathbb{P}(\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n) > 0.$$

Let U_{ab} be the number of up-crossing, then $U_{ab}(\omega) = \infty$, $\forall \omega \in \Omega$ such that

$$\liminf_{n\to\infty} X_n(\omega) < a < b < \limsup_{n\to\infty} X_n(\omega) \implies \mathbb{E}U_{ab} = \infty.$$

For $n \in \mathbb{N}$, let $U_{ab,n}$ be the number of up-crossing among $\{X_1, \ldots, X_n\}$. We have that $\lim_{n\to\infty} U_{ab,n} = U_{ab}$. Since $0 \leq U_{ab,1} \leq U_{ab,2} \leq \ldots$, we have that $\lim_{n\to\infty} \mathbb{E}U_{ab,n} = \infty$ by MCT. However,

$$\begin{aligned}
\sup_{n\in\mathbb{N}} \mathbb{E}U_{ab,n} &\leq \frac{1}{b-a} \sup_{n\in\mathbb{N}} \mathbb{E}(X_n - a)_+ \\
&\leq \frac{1}{b-a} \sup_{n\in\mathbb{N}} \mathbb{E}(X_n)_+ < \infty,
\end{aligned}$$

which is a contradiction. Thus, $\mathbb{P}(E) = 0$. Hence, for $\mathbb{P}$-a.e. $\omega \in \Omega$, $X_n(\omega) \to x_0(\omega)$ and we may then define $X_\infty(\omega) = x_0(\omega)$. ($X_\infty$ may be defined arbitrarily on the probability 0 set that the X_n's do not converge.) Since X_∞ is the almost sure pointwise limit of X_n's, it is $\mathscr{F}_n/\mathscr{B}(\mathbb{R})$–measurable for any n. Therefore, it is $\mathscr{F}_\infty/\mathscr{B}(\mathbb{R})$–measurable.

Now,

$$\begin{aligned}
\mathbb{E}|X_n| = \mathbb{E}(X_n)_+ + \mathbb{E}(X_n)_- &= 2\mathbb{E}(X_n)_+ - \mathbb{E}X_n \\
&\leq 2\mathbb{E}(X_n)_+ - \mathbb{E}X_1 && (\text{since } \mathbb{E}X_n \geq \mathbb{E}X_1) \\
&\leq 2\sup_{n\in\mathbb{N}} \mathbb{E}(X_n)_+ - \mathbb{E}X_1 < \infty.
\end{aligned}$$

Thus, $\mathbb{E}|X_\infty| \leq \liminf_{n\to\infty} \mathbb{E}|X_n| < \infty$ by Fatou's lemma.

□

Corollary 7.2.1. Let $\{X_n, \mathscr{F}_n\}$ be a reverse sub-martingale. There exists a random variable $X_\infty : \Omega \to \mathbb{R}$ such that $X_n \to X_\infty$ a.s. Moreover, if $\inf_{n\in\mathbb{N}} \mathbb{E}X_n > -\infty$, then $\mathbb{E}|X_\infty| < \infty$.

Proof.

For $a < b \in \mathbb{Q}$, $n \in \mathbb{N}$, let $U_{ab,n}$ be the number of up-crossing of $\{X_n, X_{n-1}, \ldots, X_1\}$ (note that the order is reversed). Note that $\{X_n, X_{n-1}, \ldots, X_1\}$ is a sub-martingale with respect to $\mathscr{F}_n \subset \mathscr{F}_{n-1} \subset \ldots \mathscr{F}_1$. Thus,

$$\mathbb{E}U_{ab,n} \leq \frac{1}{b-a}\mathbb{E}(X_1 - a)^+ < \infty,$$

which implies $\exists X_\infty : \Omega \to \mathbb{R}$ s.t. $X_n \to X_\infty$.

To show that $\mathbb{E}|X_\infty| < \infty$, note that $|X_n| = 2(X_n)_+ - X_n$ and $\mathbb{E}X_n \geq \inf_n \mathbb{E}X_n > -\infty$. Note also that $\{(X_n)_+, (X_{n-1})_+, \ldots, (X_1)_+\}$ is a sub-martingale and hence, $\mathbb{E}(X_n)_+ \leq \mathbb{E}(X_1)_+$. Thus,

$$\sup_n \mathbb{E}|X_n| \leq 2\mathbb{E}(X_1)_+ - \inf_n \mathbb{E}X_n < \infty.$$

Hence, the result follows by Fatou's lemma again. □

Remark 7.2.1. Recall that if $Z_1, Z_2, \ldots$ are iid random variables such that $\mathbb{E}|Z_1| < \infty$, then $\frac{S_n}{n} = \mathbb{E}[Z_1 \mid \mathscr{F}_n]$ where $\mathscr{F}_n := \sigma(S_n, Z_{n+1}, \ldots)$ and $S_n := \sum_{i=1}^n Z_i$. Assume without the loss of generality that $\mathbb{E}Z_1 = 0$.

Since $\{S_n/n, \mathscr{F}_n\}$ is a reverse martingale, we have that $S_n/n \to Z_\infty$ almost surely for some random variable Z_∞ satisfying $\mathbb{E}|Z_\infty| < \infty$. Thus, to prove the strong law of large numbers, we need only show that $Z_\infty = 0$. We can do this right away if Z_1 has finite second moment. For the general case, we need to wait until Kolmogorov's zero-one law.

Remark 7.2.2. Recall that random variables $\{X_n\}$ is uniformly integrable if

$$\lim_{K\to\infty} \sup_{n\in\mathbb{N}} \int_{|X_n|>K} |X_n| \mathrm{d}\mathbb{P} = 0.$$

Note that if $\{X_n\}$ is U.I., then $\forall n \in \mathbb{N}$,

$$\sup_{n\in\mathbb{N}} \mathbb{E}|X_n| = \sup_{n\in\mathbb{N}} \int_{|X_n|\leq K} |X_n| \mathrm{d}\mathbb{P} + \sup_{n\in\mathbb{N}} \int_{|X_n|>K} |X_n| \mathrm{d}\mathbb{P} \leq K + \sup_{n\in\mathbb{N}} \int_{|X_n|>K} |X_n| \mathrm{d}\mathbb{P} < \infty.$$

Lemma 7.2.2. Let $Y : \Omega \to \mathbb{R}$ be $(\Omega, \mathscr{F})/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable with $\mathbb{E}|Y| < \infty$ and let $\mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \cdots \subseteq \mathscr{F}$ be a filtration. For $n \in \mathbb{N}$, let $X_n := \mathbb{E}[Y|\mathscr{F}_n]$. Then $\{X_n\}$ is uniformly integrable.

Proof.

We need to show that $\forall K > 0$, $\lim_{K\to\infty} \sup_{n\in\mathbb{N}} \int_{|X_n|>K} |X_n| \mathrm{d}\mathbb{P} = 0$. For $n \in \mathbb{N}$, $|X_n| = |\mathbb{E}[Y|\mathscr{F}_n]| \leq \mathbb{E}[|Y||\mathscr{F}_n]$. Thus, for any $K > 0$,

$$\int_{|X_n|>K} |X_n| \mathrm{d}\mathbb{P} \leq \int_{|X_n|>K} \mathbb{E}[|Y||\mathscr{F}_n] \mathrm{d}\mathbb{P} \leq \int_{|X_n|>K} |Y| \mathrm{d}\mathbb{P}.$$

Since $\cup_{K=0}^\infty \{|X_n| > K\} = \Omega$, we have

$$\lim_{K\to\infty} \int_{|X_n|>K} |Y| \mathrm{d}\mathbb{P} = \mathbb{E}|Y| - \lim_{K\to\infty} \int_{|X_n|\leq K} |Y| \mathrm{d}\mathbb{P} = 0.$$

□

Theorem 7.2.3. Let $\mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \ldots$ be a filtration, and let $\mathscr{F}_\infty = \sigma(\cup_{n=1}^\infty \mathscr{F}_n) \subset \mathscr{F}$. Let $Y : \Omega \to \mathbb{R}$ be $\mathscr{F}/\mathscr{B}(\mathbb{R})$–measurable such that $\mathbb{E}|Y| < \infty$ and let $X_n := \mathbb{E}[Y|\mathscr{F}_n]$. Then, $X_n \to \mathbb{E}[Y|\mathscr{F}_\infty]$ a.s. and in L_1.

Let $\mathscr{F} \supseteq \mathscr{F}_1 \supseteq \mathscr{F}_2 \supseteq \ldots$ be a reverse filtration, let $\mathscr{F}_\infty = \cap_{n=1}^\infty \mathscr{F}_n$ and let $X_n := \mathbb{E}[Y|\mathscr{F}_n]$. Then $X_n \to \mathbb{E}[Y|\mathscr{F}_\infty]$ a.s. and in L_1.

Proof.

For the first claim, we have that $\{X_n, \mathscr{F}_n\}$ is a martingale and uniformly integrable. By Remark 7.2.2, $\sup_{n\in\mathbb{N}} \mathbb{E}|X_n| < \infty$. Thus, $\exists$ random variable $X_\infty : \Omega \to \mathbb{R}$ such that $X_n \to X_\infty$ a.s. and in L_1 by Theorem 7.2.2 and Theorem 5.2.3.

To show that $X_\infty = \mathbb{E}[Y|\mathscr{F}_\infty]$, let $A \in \mathscr{F}_n$ for some $n \in \mathbb{N}$, then

$$\int_A Y \mathrm{d}\mathbb{P} = \int_A X_n \mathrm{d}\mathbb{P} \to \int_A X_\infty \mathrm{d}\mathbb{P}$$

since $\int_A |X_n - X_\infty| \mathrm{d}\mathbb{P} \to 0$. Thus $\int_A Y \mathrm{d}\mathbb{P} = \int_A X_\infty \mathrm{d}\mathbb{P}, \forall A \in \cup_{n=1}^\infty \mathscr{F}_n$. Then $\int_A Y \mathrm{d}\mathbb{P} = \int_A X_\infty \mathrm{d}\mathbb{P}, \forall A \in \mathscr{F}_\infty$, by monotone class theorem.

To see the second claim, note that $\exists X_\infty : \Omega \to \mathbb{R}$ such that $X_n \to X_\infty$ a.s. and in L_1 by Corollary 7.2.1, Theorem 5.2.3. To show that $X_\infty = \mathbb{E}[Y|\mathscr{F}_\infty]$, let $A \in \mathscr{F}_\infty \subseteq \mathscr{F}_n$, $\forall n \in \mathbb{N}$, then

$$\int_A Y \mathrm{d}\mathbb{P} = \int_A X_n \mathrm{d}\mathbb{P} \to \int_A X_\infty \mathrm{d}\mathbb{P}.$$

□

Theorem 7.2.4. Let $\{X_n, \mathscr{F}_n\}_{n=1}^\infty$ be a sub-martingale and write $\mathscr{F}_\infty = \sigma(\cup_{n=1}^\infty \mathscr{F}_n)$. If $\{X_n\}$ is uniform integrable, then $\sup_{n\in\mathbb{N}} \mathbb{E}X_n^+ < \infty$ and $\exists X_\infty : \Omega \to \mathbb{R}$ random variable, $\mathscr{F}_\infty/\mathscr{B}(\mathbb{R})$–measurable such that $X_n \to X_\infty$ a.s. and in L_1. Moreover, if $\{X_n, \mathscr{F}_n\}_{n=1}^\infty$ is a martingale, then $X_n = \mathbb{E}[X_\infty|\mathscr{F}_n]$.

Proof.

First suppose that $\{X_n\}$ is a sub-martingale and uniformly integrable. We have from Remark 7.2.2 that $\sup_{n\in\mathbb{N}} \mathbb{E}(X_n)_+ < \infty$ and hence, by Theorem 7.2.2, there exists $X_\infty : \Omega \to \mathbb{R}$, $\mathscr{F}_\infty/\mathscr{B}(\mathbb{R})$–measurable, such that $X_n \to X_\infty$ almost surely. By Theorem 5.2.3, we have that $X_n \to X_\infty$ in L_1 as well.

Now assume that $\{X_n\}$ is a martingale. Let $A \in \mathscr{F}_n$ and let $\epsilon > 0$ be arbitrary. Let $m > n$ be such that $\int |X_\infty - X_m| d\mathbb{P} < \epsilon$. Then,

$$|\int_A X_\infty - X_n d\mathbb{P}| = |\int_A X_\infty - X_m + X_m - X_n d\mathbb{P}| \leq \epsilon$$

since $\mathbb{E}[X_m \,|\, \mathscr{F}_n] = X_n$. Since $\epsilon > 0$ is arbitrary, the claim follows.

□

In summary, we have the following:

Corollary 7.2.2. Let $\mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \ldots$ be a filtration, write $\mathscr{F}_\infty = \sigma(\cup_{n=1}^\infty \mathscr{F}_n)$, and let $\{X_n\}_{n=1}^\infty$ be a sequence such that X_n is $\mathscr{F}_n/\mathscr{B}(\mathbb{R})$–measurable.

1. Suppose $\{X_n\}$ is a sub-martingale. Then, $\{X_n\}$ is uniformly integrable if and only if there exists $X_\infty : \Omega \to \mathbb{R}$, $\mathscr{F}_\infty/\mathscr{B}(\mathbb{R})$–measurable, such that $X_n \to X_\infty$ almost surely and in L_1.
2. $\{X_n\}$ is a uniformly integrable martingale if and only if there exists $X_\infty : \Omega \to \mathbb{R}$, $\mathscr{F}_\infty/\mathscr{B}(\mathbb{R})$–measurable, such that $X_n \to X_\infty$ almost surely and in L_1, and $X_n = \mathbb{E}[X_\infty \,|\, \mathscr{F}_n]$.

7.3 Optional Sampling Theorem

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space.

Definition 7.3.1. Let $\mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \ldots$ be a filtration. Let $T : \Omega \to \mathbb{N} \cup \{\infty\}$ such that $\forall n \in \mathbb{N}$, $\{\omega : T(\omega) \leq n\} \in \mathscr{F}_n$, i.e. $T \wedge n$ is measurable w.r.t $\mathscr{F}_n$, $\forall n \in \mathbb{N}$. We say that T is a stopping time.

Remark 7.3.1. (a) T is a stopping time iff $\{\omega : T(\omega) = n\} \in \mathscr{F}_n$, $\forall n \in \mathbb{N}$. The "if" direction follows because $\{\omega : T(\omega) \leq n\} = \cup_{i=1}^n \{\omega : T(\omega) := i\} \in \mathscr{F}_n$.

(b) If S, T are stopping time, then $S \vee T$ and $S \wedge T$ are also stopping times. $T = k$ for any fixed k is also a stopping time. To see this, observe that for any $n \in \mathbb{N}$,

$$\{T \wedge S \leq n\} = \{T \leq n\} \cap \{S \leq n\} \in \mathscr{F}_n.$$

The same conclusion holds for $T \vee S$.

Example 7.3.1. Let $X_1, X_2, \cdots : \Omega \to \mathbb{R}$ such that X_n is $(\Omega, \mathscr{F}_n)/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable, not necessarily a martingale. Let $B \in \mathscr{B}(\mathbb{R})$ and define $T : \Omega \to \mathbb{N}$ as

$$T(\omega) = \begin{cases} \infty & \text{if } X_n(\omega) \notin B \text{ for } n \in \mathbb{N}, \\ \min\{n \in \mathbb{N} : X_n(\omega) \in B\}, & \text{else.} \end{cases} \tag{7.1}$$

Then T is a stopping time since $\{T \leq n\} = \cup_{k=1}^n \{X_k \in B\} \in \mathscr{F}_n$. We refer to (7.1) as hitting time.

Theorem 7.3.1. Let $T : \Omega \to \mathbb{N}$ be a stopping time. Define

$$\mathscr{F}_T := \{A \in \mathscr{F} : A \cap \{T \leq n\} \in \mathscr{F}_n, \forall n \in \mathbb{N}\}$$

and define $X_T : \Omega \to \mathbb{R}$ as

$$X_T(\omega) := X_{T(\omega)}(\omega).$$

Then, $\mathscr{F}_T \subseteq \mathscr{F}$ is a σ-field, T and X_T are $(\Omega, \mathscr{F}_T)/(\mathbb{R}, \mathscr{B}(\mathbb{R}))$-measurable. We call $\mathscr{F}_T$ as events "prior" to T.

Proof.

Let $A \in \mathscr{F}_T$, then $A^c \cap \{T \leq n\} = \{T \leq n\} \cap \{A \cap \{T \leq n\}\}^c \in \mathscr{F}_T$. Countable union clearly hold and $\{\emptyset, \Omega\} \in \mathscr{F}_T$. So $\mathscr{F}_T$ is a σ-field.

To see that T is $\mathscr{F}_T/\mathscr{B}(\mathbb{R})$–measurable, note that for any $k \in \mathbb{N}$, we have that

$$\forall n \in \mathbb{N}, \{T \leq k\} \cap \{T \leq n\} = \{T \leq \min(k, n)\} \in \mathscr{F}_n.$$

Before showing that X_T is $\mathscr{F}_T/\mathscr{B}(\mathbb{R})$–measurable, we note that if $A \cap \{T = n\} \in \mathscr{F}_n$ for all $n \in \mathbb{N}$, then $A \cap \{T \leq n\} = \cup_{k=1}^n A \cap \{T = k\} \in \mathscr{F}_n$, and hence, $A \in \mathscr{F}_T$.

Let $B \in \mathscr{B}(\mathbb{R})$. Then, for any $n \in \mathbb{N}$,

$$\begin{aligned} \{T = n\} \cap \{\omega : X_T(\omega) \in B\} &= \{T = n\} \cap \{\omega : X_{T(\omega)}(\omega) \in B\} \\ &= \{X_n \in B\} \in \mathscr{F}_n, \\ &\implies \{X_T \in B\} \in \mathscr{F}_T. \end{aligned}$$

□

Theorem 7.3.2. Let $\{X_k, \mathscr{F}_k\}_{k \in \mathbb{N}}$ be a martingale, let $T : \Omega \to \mathbb{N}$ be a stopping time, and suppose there exists $n \in \mathbb{N}$ such that $T \leq n$ almost surely. Then, $\mathbb{E}X_T = \mathbb{E}X_1$. If $\{X_k, \mathscr{F}_k\}$ is a sub-martingale, then $\mathbb{E}X_1 \leq \mathbb{E}X_T$.

Proof.

First, consider the martingale case. We have that

$$\begin{aligned} \int_\Omega X_T \mathrm{d}\mathbb{P} = \sum_{k=1}^n \int_{\{T=k\}} X_T \mathrm{d}\mathbb{P} &= \sum_{k=1}^n \int_{\{T=k\}} X_k \mathrm{d}\mathbb{P} \\ &= \int_{\{T=n\}} X_n \mathrm{d}\mathbb{P} + \int_{\{T=n-1\}} X_{n-1} \mathrm{d}\mathbb{P} + \sum_{k=1}^{n-2} \int_{\{T=k\}} X_k \mathrm{d}\mathbb{P}. \end{aligned}$$

Since $\{T = n\} = \{T \leq n-1\}^c \in \mathscr{F}_{n-1}$, we have that

$$\int_{\{T=n\}} X_n \mathrm{d}\mathbb{P} = \int_{\{T=n\}} \mathbb{E}[X_n | \mathscr{F}_{n-1}] \mathrm{d}\mathbb{P} = \int_{\{T=n\}} X_{n-1} \mathrm{d}\mathbb{P}.$$

Thus,

$$\int_\Omega X_T \mathrm{d}\mathbb{P} = \int_{\{T\geq n-1\}} X_{n-1}\mathrm{d}\mathbb{P} + \int_{\{T=n-2\}} X_{n-2}\mathrm{d}\mathbb{P} + \sum_{k=1}^{n-3}\int_{\{T=k\}} X_k\mathrm{d}\mathbb{P}.$$

By noting that $\{T \geq n-1\} = \{T \leq n-2\}^c \in \mathscr{F}_{n-2}$ and repeating the same argument many times, we have

$$\begin{aligned}\int_\Omega X_T \mathrm{d}\mathbb{P} &= \int_{\{T\geq 2\}} X_2\mathrm{d}\mathbb{P} + \int_{\{T=1\}} X_1\mathrm{d}\mathbb{P}\\ &= \int_{\{T\geq 2\}} X_1\mathrm{d}\mathbb{P} + \int_{\{T=1\}} X_1\mathrm{d}\mathbb{P} = \int_\Omega X_1\mathrm{d}\mathbb{P}.\end{aligned}$$

Same analysis follows if we have a sub-martingale. □

Theorem 7.3.3 (**Optional Sampling Theorem**)**.** Let $\{X_n, \mathscr{F}_n\}_{n=1}^\infty$ be a sub-martingale and let $T_1 \leq T_2 \leq \cdots : \Omega \to \mathbb{N}$, sequence of stopping times. Let $Y_n := X_{T_n}$, $\forall n \in \mathbb{N}$. If $n \in \mathbb{N}$, $\mathbb{E}|Y_n| < \infty$ and T_n is bounded a.s., then $\{Y_n, \mathscr{F}_{T_n}\}$ is also a sub-martingale.

Same conclusion for Martingale.

Proof.

First, we note that since $T_1 \leq T_2 \leq \ldots$, we have that $\mathscr{F}_{T_1} \subseteq \mathscr{F}_{T_2} \subseteq \ldots$. We want to show that $\forall A \in \mathscr{F}_{T_n}$, $\int_A Y_{n+1}\mathrm{d}\mathbb{P} \geq \int_A Y_n\mathrm{d}\mathbb{P}$. Write $A = \cup_{j=1}^n A \cap \{T_n = j\}$ and let $D_j := A \cap \{T_n = j\} \in \mathscr{F}_j$. Since T_{n+1} is a.s. bounded, $\exists m \in \mathbb{N}$ such that $\mathbb{P}(\{\omega : T_{n+1}(\omega) \leq m\}) = 1$. Thus,

$$\begin{aligned}\int_{D_j} Y_{n+1}\mathrm{d}\mathbb{P} &= \sum_{k=j}^m \int_{D_j\cap\{T_{n+1}=j\}} Y_{n+1}\mathrm{d}\mathbb{P}\\ &= \int_{D_j\cap\{T_{n+1}=m\}} X_m\mathrm{d}\mathbb{P} + \int_{D_j\cap\{T_{n+1}=m-1\}} X_{m-1}\mathrm{d}\mathbb{P} + \sum_{k=j}^m \int_{D_j\cap\{T_{n+1}=j\}} X_k\mathrm{d}\mathbb{P}\\ &= \int_{D_j\cap\{T_{n+1}\geq m-1\}} X_{m-1}\mathrm{d}\mathbb{P} + \sum_{k=j}^m \int_{D_j\cap\{T_{n+1}=j\}} X_k\mathrm{d}\mathbb{P}\\ &= \int_{D_j\cap\{T_{n+1}\geq j\}} X_j\mathrm{d}\mathbb{P}\\ &\geq \int_{D_j} Y_n\mathrm{d}\mathbb{P},\end{aligned}$$

since $D_j \cap \{T_{n+1} \geq j\} \subseteq D_j$ and $Y_n = X_j$ on D_j. □

Theorem 7.3.4. (Doob's Maximal Inequality)
Let $\{X_i, \mathscr{F}_i\}_{i=1}^n$ be a finite length sub-martingale. For any $\lambda > 0$, we have

$$\mathbb{P}\left(\max_{j\in[n]} X_j > \lambda\right) \leq \frac{1}{\lambda}\int X_n \mathbb{1}\left\{\max_{j\in[n]} X_j > \lambda\right\} d\mathbb{P} \leq \frac{1}{\lambda}\mathbb{E}|X_n|.$$

Proof.

Define, $\forall \omega \in \Omega$,

$$T(\omega) := \begin{cases}\min\{j \in [n] : X_j(\omega) > \lambda\} & \text{if } \max_{j\in[n]} X_j > \lambda\\ n & \text{else}\end{cases}$$

Note that for any $j \in [n]$, $\{T \leq j\} = \cup_{k=1}^j \{X_k > \lambda\} \in \mathscr{F}_j$ implies T is a stopping time.

We have $\{\max_{j\in[n]} X_j > \lambda\} = \cup_{j=1}^{n-1}\{T=j\}\cup\{T=n, X_n>\lambda\}$. Therefore,

$$\begin{aligned}
\mathbb{P}(\{\max_{j\in[n]} X_j > \lambda\}) &= \sum_{j=1}^{n-1}\mathbb{P}(T=j) + \mathbb{P}(T=n, X_n>\lambda)\\
&\le \frac{1}{\lambda}\sum_{j=1}^{n-1}\underbrace{\mathbb{E}[X_j \mathbb{1}_{\{T=j\}}]}_{\le \mathbb{E}[\mathbb{E}[X_n|\mathscr{F}_j]\mathbb{1}_{\{T=j\}}]} + \frac{1}{\lambda}\mathbb{E}[X_n \mathbb{1}_{\{T=n, X_n>\lambda\}}]\\
&\le \frac{1}{\lambda}\sum_{j=1}^{n-1}\underbrace{\mathbb{E}[\mathbb{E}[X_n \mathbb{1}_{\{T=j\}}|\mathscr{F}_j]]}_{=\mathbb{E}[X_n \mathbb{1}_{\{T=j\}}]} + \frac{1}{\lambda}\mathbb{E}[X_n \mathbb{1}_{\{T=n, X_n>\lambda\}}]\\
&\le \frac{1}{\lambda}\sum_{j=1}^{n-1}\mathbb{E}[X_n \mathbb{1}_{\{T=j\}}] + \frac{1}{\lambda}\mathbb{E}[X_n \mathbb{1}_{\{T=n, X_n>\lambda\}}]\\
&= \frac{1}{\lambda}\mathbb{E}[X_n \mathbb{1}_{\{\max_{j\in[n]} X_n>\lambda\}}].
\end{aligned}$$

□

7.4 Tail σ-field

Definition 7.4.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a background probability space and let $X_1, X_2, \ldots$ be a sequence of random variables. Define $\mathscr{F}_n := \sigma(X_n, X_{n+1}, \ldots)$ and define $\mathscr{F}_\infty = \cap_{n=1}^\infty \mathscr{F}_n$ as the tail σ-field.

Remark 7.4.1. We note that $\limsup_{n\to\infty} X_n$ is $\mathscr{F}_\infty/\mathscr{F}(\mathbb{R})$–measurable. Indeed, for any $c\in\mathbb{R}$, for any $m\in\mathbb{N}$,

$$\{\limsup_{n\to\infty} X_n \le c\} = \left\{\lim_{n\ge m,\, n\to\infty}\sup_{k\ge n} X_k \le c\right\} \in \mathscr{F}_m.$$

Likewise, if $\sum_{n=1}^\infty X_n$ converges absolutely almost surely, then $\lim_{n\to\infty}\sum_{k=n}^\infty X_n$ is $\mathscr{F}_\infty/\mathscr{B}(\mathbb{R})$–measurable.

Theorem 7.4.1 (Kolmogorov Zero–One Law). Suppose $X_1, X_2, \ldots$ are independent random variables. If $A\in\mathscr{F}_\infty$, then either $\mathbb{P}(A)=0$ or $\mathbb{P}(A)=1$. If a function $f:\Omega\to\mathbb{R}$ is $\mathscr{F}_\infty/\mathscr{B}(\mathbb{R})$–measurable, then it is a constant almost surely.

Proof.
Let $A\in\mathscr{F}_\infty$. We will show that $\mathbb{P}(A\cap A)=\mathbb{P}(A)\mathbb{P}(A)$, that is, A is independent of itself. The theorem follows from this immediately.

Since $\mathscr{F}_\infty\subset\mathscr{F}_1$, there exists $\tilde{A}\in\mathscr{B}(\mathbb{R})^{\otimes\mathbb{N}}$ such that $A=\{(X_1,X_2,\ldots)\in\tilde{A}\}$. Define

$$\mathscr{C} := \big\{\tilde{C}\in\mathscr{B}(\mathbb{R})^{\otimes\mathbb{N}} : \mathbb{P}(A\cap\{(X_1,X_2,\ldots)\in\tilde{C}\}) = \mathbb{P}(A)\mathbb{P}((X_1,X_2,\ldots)\in\tilde{C})\big\}.$$

We now claim that all measurable cylinders are in $\mathscr{C}$. Let $B_n\in\mathscr{B}(\mathbb{R}^n)$ for some $n\in\mathbb{N}$. Since $A\in\mathscr{F}_{n+1}$, there exists $\tilde{A}_{n+1}\in\mathscr{B}(\mathbb{R})^{\otimes\{n+1,n+2,\ldots\}}$ such that $A=\{(X_{n+1},X_{n+2},\ldots)\in\tilde{A}_{n+1}\}$. Thus,

$$\begin{aligned}
\mathbb{P}\big(A\cap\{(X_1,X_2,\ldots X_n)\in B_n\}\big) &= \mathbb{P}\big(\{(X_{n+1},X_{n+2},\ldots)\in\tilde{A}_{n+1}\}\cap\{(X_1,X_2,\ldots X_n)\in B_n\}\big)\\
&= \mathbb{P}((X_{n+1},X_{n+2},\ldots)\in\tilde{A}_{n+1})\mathbb{P}((X_1,X_2,\ldots X_n)\in B_n)\\
&= \mathbb{P}(A)\mathbb{P}((X_1,X_2,\ldots X_n)\in B_n).
\end{aligned}$$

Now, let $\tilde{C}_1\subset\tilde{C}_2\subset\ldots$ be an increasing sequence of sets in $\mathscr{C}$ and write $\tilde{C}=\cup_{n=1}^\infty\tilde{C}_n$. Then, $\mathbb{P}(A\cap\{(X_1,X_2,\ldots)\in\tilde{C}_n\})\to\mathbb{P}(A\cap\{(X_1,X_2,\ldots)\in\tilde{C})$ and $\mathbb{P}(X_1,X_2,\ldots\in\tilde{C}_n)\to\mathbb{P}(X_1,X_2,\ldots\in\tilde{C})$ and hence $\tilde{C}\in\mathscr{C}$. By applying the same reasoning with decreasing sequences of sets, we have that $\mathscr{C}$ is a monotone class. Therefore, by the Π-Λ theorem or the monotone class theorem (Theorem 1.3.1), $\mathscr{C}$ contains the σ-field generated by measurable cylinders, which implies that $\mathscr{C}=\mathscr{B}(\mathbb{R})^{\otimes\mathbb{N}}$. Hence, $\tilde{A}\in\mathscr{C}$ which proves the result as desired. □

Theorem 7.4.2 (Hewitt–Savage zero–one law). We say that a set $A \in \mathscr{B}(\mathbb{R})^{\otimes \mathbb{N}}$ is symmetric if, for any $n \in \mathbb{N}$, for any finite permutation $\tau : [n] \to [n]$, $(x_1, x_2, \ldots) \in A$ iff $(x_{\tau(1)}, x_{\tau(2)}, \ldots) \in A$. For example, the set $\{x \in \mathbb{R}^{\mathbb{N}} : \sum_{n=1}^{\infty} x_n < \infty\}$ is symmetric.

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables and let A be symmetric. Then, $\mathbb{P}(X_1, X_2, \ldots \in A) = 0$ or 1.

Chapter 8

Brownian Motion

8.1 Gaussian Process

Definition 8.1.1 (Brownian Motion). Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and, for each $t \in [0, \infty)$, let $W_t : \Omega \to \mathbb{R}$ be $\mathscr{F}/\mathscr{B}(\mathbb{R})$-measurable. We say that $\{W_t\}_{t\in[0,\infty)}$ is a standard Brownian motion or Wiener process if the following are true:

(1) $W_0 = 0$,

(2) There exists $B \in \mathscr{F}$ such that $\mathbb{P}(B) = 1$ and for all $\omega \in B$, the function $t \to W_t(\omega)$ is continuous.

(3) (Independent increment property) For any $n \in \mathbb{N}$, for any sequence $0 \le s_1 < t_1 \le s_2 \le t_2 \le \cdots \le s_n < t_n < \infty$, $W_{t_1} - W_{s_1}$, $W_{t_2} - W_{s_2}$, $\ldots$, $W_{t_n} - W_{s_n}$ are jointly independent.

(Stationary increment) Moreover, for any $0 < s \le t < \infty$, $W_t - W_s$ has the same distribution as $W_{t-s} - W_0$,

(4) For any $0 < s \le t < \infty$, $W_t - W_s \sim N(0, t-s)$.

Remark 8.1.1. (a) It is not clear that such $(\Omega, \mathscr{F}, \mathbb{P})$ and $\{W_t\}_{t\in[0,\infty)}$ exist. We will prove this with Kolmogorov extension theorem and continuity theorem later. For now, we assume that Brownian motion exists.

(b) We note that if time is discrete, $t \in \{0, \Delta, 2\Delta, 3\Delta, \ldots\}$ for some $\Delta > 0$, we can simulate Brownian motion by generating i.i.d. $Z_1, Z_2, \cdots \sim N(0, \Delta)$ and defining $W_t = \sum_{i=1}^{t/\Delta} Z_i = W_{t-\Delta} + Z_{t/\Delta}$ for $t \in \{\Delta, 2\Delta, \ldots\}$ and $W_0 = 0$.

(c) Properties (2) and (3) in Definition 8.1.1 in fact imply that

$$W_t - W_s \sim N(\mu(t-s), \sigma^2(t-s)) \text{ for } \mu \in \mathbb{R}, \sigma^2 > 0.$$

To see this informally, suppose $s = 0$, $t = 1$, and let $0 = t_0 < t_1 < \ldots < t_n = 1$ be an evenly spaced partition of $[0, 1]$. We then have

$$W_1 = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sqrt{n}(W_{t_j} - W_{t_{j-1}}),$$

which is a sum of iid random variables by the independent and stationary increment properties. Using the continuity condition, one may show that $\sqrt{n}(W_{t_j} - W_{t_{j-1}})$ has mean $O(1/\sqrt{n})$ and variance $O(1)$. CLT then shows that W_1 must be Gaussian.

(d) In stating the continuity property, it is not valid to say $\mathbb{P}(\{\omega \in \Omega : t \mapsto W_t(\omega) \text{ is continuous}\}) = 1$ because the set of continuous functions $\{f \in \mathbb{R}^{[0,\infty)} : f \text{ is continuous}\}$ is not in the product σ-field $\mathscr{B}(\mathbb{R})^{\otimes[0,\infty)}$. Therefore, viewing $W : \Omega \to \mathbb{R}^{[0,\infty)}$ as a random function ($\mathscr{F}/\mathscr{B}(\mathbb{R})^{\otimes[0,\infty)}$–measurable), it need not be that $W^{-1}(\text{ continuous functions }) \in \mathscr{F}$.

Proposition 8.1.1. Let $\{W_t\}_{t\in[0,\infty)}$ be a standard Brownian motion. Then each of the following is also a standard Brownian motion:

(a) $\{-W_t\}_{t\geq 0}$.

(b) For any $s_0 \geq 0$, $\{W_{t+s_0} - W_{s_0}\}_{t\geq 0}$

(c) For any $a > 0$, $\{a \cdot W_{t/a^2}\}_{t\geq 0}$

Proof.

(a) For $0 \leq s \leq t < \infty$, $-W_t - (-W_s) = -(W_t - W_s) \sim N(0, t-s)$. Property (4) thus follows. We may show property (3) in the same way. Properties (1) and (2) are obvious.

(b) For any $0 \leq s \leq t < \infty$, $(W_{t+s_0} - W_{s_0}) - (W_{s+s_0} - W_{s_0}) = W_{t+s_0} - W_{s+s_0} \sim N(0, t-s)$.

(c) For any $0 \leq s \leq t < \infty$, $aW_{t/a^2} - aW_{s/a^2} = a(W_{t/a^2} - W_{s/a^2}) \sim aN(0, \frac{t-s}{a^2}) = N(0, t-s)$.

□

Remark 8.1.2. (1) These simple properties immediately imply properties such as $\max\{W_t : t \in [0,1]\} \stackrel{d}{=} -\min\{W_t : t \in [0,1]\}$.

(2) Note that $M(\omega) := \max\{W_t(\omega) : t \in [0,1]\}$ for $\omega \in \Omega$ is attained almost surely since $t \to W_t(\omega)$ is continuous a.s.

(3) Moreover, $M : \Omega \to \mathbb{R}$ is $\mathscr{F}/\mathscr{B}(\mathbb{R})$-measurable because $M = \sup\{W_t : t \in \mathbb{Q} \cap [0,1]\}$ by continuity. Recall that if $f_1, f_2, \cdots : \Omega \to \mathbb{R}$ are $\mathscr{F}/\mathscr{B}(\mathbb{R})$-measurable, then for any $t \in \mathbb{R}$,

$$(\sup_n f_n)^{-1}((-\infty, t]) = \{\omega : f_n(\omega) \leq t, \forall n \in \mathbb{N}\} = \cap_{n=1}^{\infty} f_n^{-1}(-\infty, t] \in \mathscr{B}(\mathbb{R}).$$

Remark 8.1.3. We can view Brownian motion as $\{W_t\}_{t\in[0,\infty)}$ or as $W : \Omega \times [0,\infty) \to \mathbb{R}$ with $W(t,\omega) = W_t(\omega)$ or as $W : \Omega \to \mathbb{R}^{[0,\infty)}$ where $\mathbb{R}^{[0,\infty)} := \{f : [0,\infty) \to \mathbb{R}\}$, where $W(\omega) = \{t \mapsto W_t(\omega)\}$. All these views are useful. Recall that W is $\mathscr{F}/\mathscr{B}(\mathbb{R})^{\otimes[0,\infty)}$–measurable by the definition of $\mathscr{B}(\mathbb{R})^{\otimes[0,\infty)}$ as the smallest σ-field such that all the marginals W_t's are $\mathscr{F}/\mathscr{B}(\mathbb{R})$–measurable.

Definition 8.1.2 (Product σ-Field). Let $(\mathcal{X}, \mathscr{G})$ be a measurable space and let T be any set (can be finite). Recall that $\mathcal{X}^T := \{f : T \to \mathcal{X}\}$. For each $t \in T$, define the projection operator $\pi_t : \mathcal{X}^T \to \mathcal{X}$ by $\forall f \in \mathcal{X}^T$, $\pi_t(f) = f(t)$. Also, for $S \subseteq T$, define $\pi_S : \mathcal{X}^T \to \mathcal{X}^S$ by $\pi_S(f) = \{f(t)\}_{t\in S}$, $\forall f \in \mathcal{X}^T$. We define the product σ-field $\mathscr{G}^{\otimes T} \subseteq 2^{\mathcal{X}^T}$ as the smallest σ-field s.t. $\forall t \in T$, π_t is $\mathscr{G}^{\otimes T}/\mathscr{G}$-measurable. More precisely,

$$\mathscr{G}^{\otimes T} := \cap \left\{ \mathscr{H} \subseteq 2^{\mathcal{X}^T} : \mathscr{H} \text{ is } \sigma\text{-field and } \forall t \in T, \forall A \in \mathscr{G}, \pi_t^{-1}(A) = \{f \in \mathcal{X}^T : f(t) \in A\} \in \mathscr{H} \right\}$$

Equivalently, we say that $\mathscr{G}^{\otimes T}$ is σ-field generated by sets of the form $\pi_t^{-1}(A)$ for some $t \in T, A \in \mathscr{G}$:

$$\mathscr{G}^{\otimes T} = \sigma\{\pi_t^{-1}(A) : t \in T, A \in \mathscr{G}\}.$$

As an example, if $\mathcal{X} = \mathbb{R}$, $\mathscr{G} = \mathscr{B}(\mathbb{R})$, $T = \{1,2,3\}$, then $\mathcal{X}^T = \mathbb{R}^3$, and for $A \in \mathscr{B}(\mathbb{R})$, $\pi_1^{-1}(A) = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 \in A\}$. Thus, we have $\mathscr{B}(\mathbb{R})^{\otimes\{1,2,3\}} = \mathscr{B}(\mathbb{R}^3)$.

Definition 8.1.3 (Stochastic Process). Let T be any set, we say that $X : \Omega \to \mathbb{R}^T$ is a stochastic process if it is $\mathscr{F}/\mathscr{B}(\mathbb{R})^{\otimes T}$-measurable, or equivalently, if X_t is a well-defined random variable for every $t \in T$.

Definition 8.1.4 (Gaussian Process). Given $\mu : T \to \mathbb{R}$ and positive definite kernel $K : T \times T \to \mathbb{R}$, we say that a stochastic process X is a Gaussian process, denoted as,

$$X \sim \mathrm{GP}(\mu, K),$$

if for all finite collection $\{t_1, \dots, t_n\}$, writing $X_{t_i} := \pi_{t_i} \circ X$ where $\pi_{t_i} : \mathbb{R}^T \to \mathbb{R}$ is the evaluation/projection ($\pi_{t_i}(f) = f(t_i)$), we have that

$$(X_{t_1}, \dots, X_{t_n}) \sim N\left(\begin{pmatrix}\mu(t_1)\\ \vdots \\ \mu(t_n)\end{pmatrix}, \begin{pmatrix}K(t_1,t_1) & K(t_1,t_2) & \dots \\ K(t_2,t_1) & \ddots & \vdots \\ \dots & \dots & K(t_n,t_n)\end{pmatrix}\right)$$

Remark 8.1.4. (a) Key is that $\{X_t\}_{t\in T}$ is a Gaussian process iff every finite-dimensional random vector $(X_{t_1}, \dots, X_{t_n})$ for $n \in \mathbb{N}$, $t_1, \dots, t_n \in T$ is Gaussian in a "consistent" manner.

(b) This definition may be extended to the case where X_t, $\forall t \in T$, is $\mathbb{R}^d$-valued by taking $\mu : T \to \mathbb{R}^d$ and $K : T \times T \to \mathbb{R}^{d\times d}$.

Example 8.1.1. (a) We take $T = [0, \infty)$, $\mu = 0$, and $K(s,t) = \mathrm{Cov}(W_s, W_t) = \min(s,t)$ for $s, t \ge 0$, to obtain standard Brownian motion. To verify covariance kernel, note that $\forall 0 \le s \le t < \infty$, $\mathrm{Cov}(W_s, W_t) = \mathrm{Cov}(W_s, W_s + W_t - W_s) = \mathrm{Var}(W_s) = s$. We may also take $T = \mathbb{R}$, $W_0 = 0$, and define 2-sided Brownian motion.

(b) We take $T = [0, \infty)$, $\mu = 0$, and $K(s,t) = \beta e^{-\alpha|t-s|}$ for $s, t \ge 0$, $\alpha, \beta > 0$ and denote resulting process by $\{Y_t\}_{t\ge 0}$. Note $Y_0 \sim N(0, \beta)$. This yields the Uhlenbeck-Ornstein process. It is stationary, mean-reverting, and the continuous analogue of AR(1).

To see that $K(\cdot,\cdot)$ is a positive definite kernel, assume $\alpha = \beta = 1$ and note that $\{e^{-t}W_{e^{2t}}\}_{t\ge 0}$ is a Gaussian process.

For $0 \le s \le t < \infty$, $\mathrm{Cov}(e^{-s}W_{e^{2s}}, e^{-t}W_{e^{2t}}) = \mathbb{E}e^{-s}W_{e^{2s}}W_{e^{2t}}e^{-t} = e^{-s-t} \cdot e^{2s} = e^{s-t} = e^{-|t-s|}$. Since multivariate Gaussian distribution is uniquely specified by mean and covariance, $(Y_{t_1}, \dots, Y_{t_n}) \stackrel{d}{=} (e^{-t_1}W_{e^{2t_1}}, \dots, e^{-t_n}W_{e^{2t_n}})$, $\forall n \in \mathbb{N}$, $\{t_1, \dots, t_n\} \subseteq [0, \infty)$.

(c) We take $T = [0,1]$, $\mu = 0$, and $K(s,t) = \min(s,t) - st$ for $s, t \in [0,1]$, and denote the process by $\{\widetilde{W}_t\}_{t\in[0,1]}$. Note that $\mathrm{Var}(\widetilde{W}_0) = \mathrm{Var}(\widetilde{W}_1) = 0$. For $t \in [0,1]$, $\mathrm{Var}(\widetilde{W}_t) = t(1-t)$, maximized at $t = 1/2$. This gives us the Brownian bridge: $\widetilde{W}$ has the same distribution as standard Brownian motion W conditioned on $W_1 = 0$. To see this intuitively, note that, for any $0 \le s \le t < \infty$,

$$\mathrm{Var}(W_s, W_t, W_1) = \begin{pmatrix} s & s & s \\ s & t & t \\ s & t & 1 \end{pmatrix}$$

and therefore,

$$\mathrm{Var}(W_s, W_t | W_1) = \begin{pmatrix} s & s \\ s & t \end{pmatrix} - \begin{pmatrix} s \\ t \end{pmatrix}\begin{pmatrix} s \\ t \end{pmatrix}^T 1^{-1} = \begin{pmatrix} s - s^2 & s - st \\ s - st & t - t^2 \end{pmatrix}$$
$$\implies \mathrm{Cov}(W_s, W_t | W_1 = 0) = s - st.$$

We also see that $\forall t \in [0,1]$, $\widetilde{W}_t \stackrel{d}{=} W_t - tW_1$. In general, for $T = [a,b]$, for $a < b \in \mathbb{R}$, $\mu(t) = \alpha + \frac{t-a}{b-a}\beta$ for $\alpha, \beta \in \mathbb{R}$, $K(s,t) = (b-t)(s-1)/(b-a)$, the resulting process $\{\widetilde{W}_t\}_{t\in[a,b]}$ has the distribution of $\{W_t\}_{t\in[a,b]}$ conditioned on $W_a = \alpha$, $W_b = \beta$.

(d) We take $T = \mathbb{R}$, $\mu = 0$, $K(s,t) = e^{-|s-t|^2}$ for $s, t \in \mathbb{R}$ and denote the process by X. Then the sample path $X(\omega)(\cdot) : \mathbb{R} \to \mathbb{R}$ is differentiable for a.e. $\omega \in \Omega$.

(e) We take $T = \mathbb{R}$, $\mu = 0$, $K(s,t) = \mathbf{1}_{\{s=t\}}$ for $s, t \in \mathbb{R}$, and denote the resulting process by X. Then, the sample path of X is discontinuous.

Definition 8.1.5. For any finite subset $S = \{t_1, \dots, t_n\} \subseteq T$, let P_S be a probability measure on $\left(\mathcal{X}^S, \mathscr{G}^{\otimes S}\right)$. We say that the family $\{P_S : S \subseteq T \text{ finite}\}$ is Kolmogorov consistent (or projective) if for any $S, S' \subseteq T$ where $S \subseteq S'$, we have that $P_S = P_{S'}^{(\pi_S)}$ where $P_{S'}^{(\pi_S)}$ is a probability measure defined on $(X^S, \mathscr{G}^{\otimes T})$ by

$$\forall B \in \mathscr{G}^{\otimes S}, P_{S'}^{(\pi_S)}(B) = P_{S'}(\pi_S^{-1}(B)) = P_{S'}(B \times \mathcal{X}^{S \setminus S'}) = P_{S'}(\{f : S' \to \mathcal{X} : (f(t))_{t \in S} \in B\}).$$

As an example, let $T = \{a,b,c,d\}$, $S = \{a,b\}$, $S' = \{a,b,c\}$, $\mathcal{X} = \mathbb{R}$, $\mathcal{G} = \mathscr{B}(\mathbb{R})$. For $B \in \mathscr{B}(\mathbb{R})^{\otimes a,b} = \mathscr{B}(\mathbb{R}^2)$, $P_{S'}^{(\pi_S)}(B) = P_{S'}(\{(x_a, x_b, x_c) \in \mathbb{R}^3 : (x_a, x_b) \in B)$.

Recall that for any probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and measurable space $(\mathcal{X}, \mathcal{G})$, for any $g : \Omega \to \mathcal{X}$, $\mathscr{F}/\mathcal{G}$-measurable, we have $\mathbb{P}^{(g)}$ is a probability measure on $(\mathcal{X}, \mathcal{G})$ such that $\forall B \in \mathcal{G}$, $\mathbb{P}^{(g)}(B) = \mathbb{P}(\{\omega \in \Omega : g(\omega) \in B\}) = \mathbb{P}(\{g \in B\})$. We call $\mathbb{P}^{(g)}$ the pushforward measure induced by g or the distribution of g. (See Remark 2.4.1, Thm 2.4.2).

Note that any stochastic process $\{X_t\}_{t \in T}$ has associated with it a consistent family $\{P_S : S \subset T \text{ finite }\}$ where P_S is the distribution of the random vector X_S.

Theorem 8.1.1 (Kolmogorov Extension Theorem). Let T be any set, let $(\mathcal{X}, \mathcal{G}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Let $\{P_S : S \subseteq T \text{ finite}\}$ be a consistent family of probability measure. Then, there exists a unique probability measure P on $(\mathbb{R}^T, \mathscr{B}(\mathbb{R})^{\otimes T})$ s.t.

$$P^{(\pi_S)} = P_S \quad \forall S \subseteq T \text{ finite},$$

where again, $P^{(\pi_S)}$ is a probability measure on $(\mathbb{R}^S, \mathscr{B}(\mathbb{R})^{\otimes S})$, s.t. for $B \in \mathscr{B}(\mathbb{R})^{\otimes S}$, $P^{(\pi_S)}(B) = P(\{f : T \to \mathbb{R} : (f(t))_{t \in S} \in B\})$. In particular, if there exists background probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and stochastic process $X : \Omega \to \mathbb{R}^T$, then there exists a distribution of X which we write $\mathbb{P}^{(X)}$ over $(\mathbb{R}^T, \mathscr{B}^{\otimes T})$. This distribution is uniquely specified by the finite dimensional marginal distributions $\{\mathbb{P}^{(X_S)} : S \subset T \text{ finite }\}$.

Remark 8.1.5. It is now easy to show existence of Gaussian process:

Fix a set T, $\mu : T \to \mathbb{R}$, positive semidefinite kernel $K : T \times T \to \mathbb{R}$ (giving a possibly degenerate GP). Let $(\Omega, \mathscr{F}, \mathbb{P}) = (\mathbb{R}^T, \mathscr{B}(\mathbb{R})^{\otimes T}, P)$ where P is the unique probability measure on $(\mathbb{R}^T, \mathscr{B}(\mathbb{R})^{\otimes T})$ such that for any $\{t_1, \dots, t_n\} \subseteq T$, the marginal distribution of P on $\mathbb{R}^{\{t_1, \dots, t_n\}}$ is

$$P^{(\pi_{t_1}, \dots, \pi_{t_n})} = P_{\{t_1, \dots, t_n\}} := N\left(\begin{pmatrix}\mu(t_1) \\ \vdots \\ \mu(t_n)\end{pmatrix}, \begin{pmatrix} K(t_1,t_1) & K(t_1,t_2) & \dots \\ K(t_2,t_1) & \ddots & \vdots \\ \dots & \dots & K(t_n,t_n)\end{pmatrix}\right).$$

Note that $\{P_{\{t_1, \dots, t_n\}} : n \in \mathbb{N}, \{t_1, \dots, t_n\} \subseteq T\}$ is a consistent family of finite dimensional probability distributions, thus Theorem 8.1.1 applies.

Example 8.1.2. As a concrete example, if $T = [0, \infty)$ and P is the process distribution of standard Brownian motion, then for any $t \in [0, \infty)$, $a, b \in \mathbb{R}$, writing $A_{t,a,b} = \{f : [0, \infty) \to \mathbb{R} : f(t) \in (a,b]\} = \pi_t^{-1}((a,b]) \in \mathscr{B}(\mathbb{R})^{\otimes [0,\infty)}$, we have $P(A_{t,a,b}) = \mathbb{P}_{N(0,t)}((a,b])$. Now, let $X : \Omega \to \mathbb{R}^T$ be the identity function so that the distribution of X, $\mathbb{P}^{(X)}$, is equal to P. Then, for any $n \in \mathbb{N}$, $\{t_1, \dots, t_n\} \subseteq T\}$, writing $X_{t_1} : \Omega \to \mathbb{R}$ as $\pi_{t_1} \circ X$, that is $X_{t_1}(\omega) = \pi_{t_1}(X(\omega)) = X(\omega)(t_1)$, $\forall \omega \in \Omega$, we have

$$\mathbb{P}^{(X_{t_1}, \dots, X_{t_n})} = \mathbb{P}^{(X)(\pi_{t_1}, \dots, \pi_{t_n})} = P^{(\pi_{t_1}, \dots, \pi_{t_n})} = N\left(\begin{pmatrix}\mu(t_1) \\ \vdots \\ \mu(t_n)\end{pmatrix}, \begin{pmatrix} K(t_1,t_1) & K(t_1,t_2) & \dots \\ K(t_2,t_1) & \ddots & \vdots \\ \dots & \dots & K(t_n,t_n)\end{pmatrix}\right).$$

Corollary 8.1.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let T be any set and let $X, Y : \Omega \to \mathbb{R}^T$ be $\mathscr{F}/\mathscr{B}(\mathbb{R})^{\otimes T}$-measurable such that

$$\forall t \in T, \quad \mathbb{P}(X_t = Y_t) = \mathbb{P}(\{\omega \in \Omega : X_t(\omega) = Y_t(\omega)\}) = 1, \tag{$*$}$$

then we have that $\mathbb{P}^{(X)} = \mathbb{P}^{(Y)}$, i.e., $X \overset{d}{=} Y$. We call Y a modification of X if $(*)$ holds.

Note that $(*)$ is weaker than $\mathbb{P}(\{\omega : \forall t \in T, X_t(\omega) = Y_t(\omega)\})$, where the set $\{\omega : \forall t \in T, X_t(\omega) = Y_t(\omega)\}$ may not be in $\mathscr{B}(\mathbb{R})^{\otimes T}$.

Proof.
Let $n \in \mathbb{N}$, $\{t_1, \dots, t_n\} \subseteq T$. Then

$$\mathbb{P}(\{\omega : (X_{t_1}(\omega), \dots, X_{t_n}(\omega)) \neq (Y_{t_1}(\omega), \dots, Y_{t_n}(\omega))\}) \leq \sum_{i=1}^{n} \mathbb{P}(X_{t_i} \neq Y_{t_i}) = 0,$$

which implies $\mathbb{P}^{(X)(\pi_{t_1}, \dots, \pi_{t_n})} = \mathbb{P}^{(Y)(\pi_{t_1}, \dots, \pi_{t_n})}$. Then $\mathbb{P}^{(X)} = \mathbb{P}^{(Y)}$ by Kolmogorov extension theorem. $\square$

Corollary 8.1.2. Let $X : \Omega \to \mathbb{R}^T$ be $\mathscr{F}/\mathscr{B}(\mathbb{R})^{\otimes T}$-measurable (stochastic process). Then X is a Gaussian process (possibly degenerate) iff for all $n \in \mathbb{N}$, $\{t_1, \dots, t_n\} \subseteq T$, $(X_{t_1}, \dots, X_{t_n})$ is jointly Gaussian.

Proof.
Suppose $\forall n \in \mathbb{N}$, $\{t_1, \dots, t_n\} \subseteq T$, $(X_{t_1}, \dots, X_{t_n})$ is Gaussian. Define, $\forall t \in T$, $\mu(t) = \mathbb{E}X_t$ and $\forall s, t \in T$, $K(s,t) = \operatorname{Cov}(X_s, X_t)$. Then μ, K are well-defined and K is p.s.d. kernel since $\{\mathbb{P}^{(X_{t_1}, \dots, X_{t_n})} : n \in \mathbb{N}, \{t_1, \dots, t_n\} \subseteq T\}$ is a consistent family. $\square$

Example 8.1.3. Fix $m \in \mathbb{N}$ and functions $\phi_1, \dots, \phi_m : [0,1] \to \mathbb{R}$. Let $(Z_1, \dots, Z_m) : \Omega \to \mathbb{R}^m$ be a Gaussian random vector and let $X = \sum_{i=1}^m Z_i \phi_i$, then X is a degenerate Gaussian process.

8.2 Kolmogorov–Chentsov Continuity

Definition 8.2.1. For a function $f : \mathbb{R}^d \to \mathbb{R}$, we say that, for $\gamma \in (0,1]$, it is γ-Hölder-continuous if $\exists C_\gamma$ such that $\forall x, y \in \mathbb{R}^d$,

$$|f(x) - f(y)| \leq C_\gamma \|x - y\|_2^\gamma. \tag{8.1}$$

Remark 8.2.1.
- If f is γ-Hölder for any $\gamma > 0$, f is uniformly continuous, i.e., $\forall \varepsilon > 0$, $\exists \delta > 0$ such that $\forall x, y \in \mathbb{R}^d$, $\|x - y\|_2 < \delta \implies |f(x) - f(y)| < \varepsilon$.
- The case when $\gamma = 1$ is Lipschitz continuity.
- If (8.1) holds for $\gamma > 1$, then f is a constant. If f is differentiable, the notion of Hölder-continuity can be extended for $\gamma > 1$.
- For $\gamma \in (0,1]$, if $\exists C_\gamma$, $\delta > 0$ such that $\forall x, y \in \mathbb{R}^d$ where $\|x - y\| < \delta$, we have $|f(x) - f(y)| \leq C_\gamma \|x - y\|^\gamma$, then f is γ-Hölder-continuous.

Theorem 8.2.1 (**Kolmogorov-Chentsov Continuity Theorem).** Let $X : \Omega \to \mathbb{R}^{\mathbb{R}^d}$ be a stochastic process such that, for some $\alpha, \beta, C > 0$, for all $s, t \in \mathbb{R}^d$, $\mathbb{E}|X_s - X_t|^\alpha \leq C\|s - t\|_2^{d+\beta}$. Then X has a modification $\widetilde{X} : \Omega \to \mathbb{R}^{\mathbb{R}^d}$ such that $\exists B \in \mathscr{F}$ with $\mathbb{P}(B) = 1$ such that, for all ω in B,

(1) $\widetilde{X}(\omega)$ is continuous,

(2) For any compact set $D \subseteq \mathbb{R}^d$, $\widetilde{X}(\omega)$ restricted to D is γ-Hölder-continuous for any $0 < \gamma < \frac{\beta}{\alpha}$.

(Recall $\widetilde{X}$ is a modification of X if $\forall t \in \mathbb{R}^d$, $\mathbb{P}(\widetilde{X}_t = X_t) = 1$.)

Example 8.2.1. For a standard Brownian motion $W : \Omega \to \mathbb{R}^{[0,\infty)}$, we have $\forall k \in \mathbb{N}$, $\exists C_k > 0$ such that $\forall s,t \in [0,\infty)$, $\mathbb{E}|W_s - W_t|^{2k} \le C_k|s-t|^k$. Thus, W has a continuous modification that is γ-Hölder-continuous on any bounded interval for any $\gamma < \sup_{k\in\mathbb{N}} \frac{k-1}{2k} = \frac{1}{2}$.

Proof of Theorem 8.2.1.

We first consider process over $[0,1]^d$. Suppose $X : \Omega \to \mathbb{R}^{[0,1]^d}$ satisfy the hypothesis for $\alpha, \beta, C > 0$. Fix any $\gamma \in (0, \frac{\alpha}{\beta})$.

Step 1: Define, for $n \in \mathbb{N}$, $D_n := \{(x_1, \dots, x_d)2^{-n} : x_1, \dots, x_d \in [2^n]\}$, so that D_n is a grid over $[0,1]^d$. Define $D := \cup_{n=1}^{\infty} D_n$. For $n \in \mathbb{N}$, let $\xi_n : \Omega \to [0,\infty)$ be a random variable such that, $\forall \omega \in \Omega$,

$$\xi_n(\omega) := \max\{|X_s(\omega) - X_t(\omega)| : s,t \in D_n, \|s-t\|_2 = 2^{-n}\}.$$

(We call such s,t neighbouring pairings). We note that

$$\begin{aligned}
\mathbb{E}\xi_n^\alpha &= \mathbb{E}\left\{\max_{(s,t)\in D_n^2, \|s-t\|_2 = 2^{-n}} |X_s - X_t|^\alpha\right\} \\
&\le \mathbb{E}\left\{\sum_{(s,t)\in D_n^2, \|s-t\|_2 = 2^{-n}} |X_s - X_t|^\alpha\right\} \\
&= \sum_{(s,t)\in D_n^2, \|s-t\|_2 = 2^{-n}} \mathbb{E}\{|X_s - X_t|^\alpha\} \\
&\le \left|\{(s,t) \in D_n^2 : \|s-t\|_2 = 2^{-n}\}\right| \cdot \max_{(s,t)\in D_n^2, \|s-t\|_2 = 2^{-n}} \mathbb{E}|X_s - X_t|^\alpha \\
&\le 2d2^{dn}C2^{-(d+\beta)n} \le 2dC \cdot 2^{-\beta n}.
\end{aligned}$$

Therefore,

$$\mathbb{E}\left[\sum_{n=1}^{\infty}(2^{\gamma n}\xi_n)^\alpha\right] = \sum_{n=1}^{\infty} 2^{\alpha\gamma n}\mathbb{E}[\xi_n^\alpha] \le 2dC\sum_{n=1}^{\infty} 2^{(\alpha\gamma-\beta)n} < \infty,$$

where the first equality by monotone convergence theorem and $\xi_n^\alpha \ge 0$ and the last inequality holds since $\alpha\gamma - \beta < 0$,

Therefore, since $\sum_{n=1}^{\infty}(2^{\gamma n}\xi_n)^\alpha \ge 0$, it is almost surely finite. Hence, $\exists B \in \mathscr{F}$, $\mathbb{P}(B) = 1$, such that $\forall \omega \in B$, $\sum_{n=1}^{\infty}(2^{\gamma n}\xi_n(\omega))^\alpha < \infty$.

So, $\forall \omega \in B$, $\exists N_\omega \in \mathbb{N}$ such that $\forall n \ge N_\omega$, $\xi_n(\omega)2^{\gamma n} \le 1 \iff \xi_n(\omega) \le 2^{-\gamma n}$.

Step 2: We will show that $\forall \omega \in B$, $X(w)$ restricted to D is γ-Hölder. Fix $\omega \in B$. Let $s,t \in D$, such that $2^{-m-1} \le \|s-t\|_2 \le 2^{-m}$ for some $m \ge N_\omega$. For every $n \in \mathbb{N}$, define $g_n(x) := \arg\min\{\|\widetilde{s} - x\|_2 : \tilde{s} \in D_n\}$, $\forall x \in [0,1]^d$.

FACT A: $\forall x \in [0,1]^d$, $\|g_n(x) - x\|_2 \le d^{1/2}2^{-n}$, and thus,

$$\|g_n(x) - g_{n+1}(x)\|_2 \le \|g_n(x) - x\|_2 + \|g_{n+1}(x) - x\|_2 \le 2d^{1/2} \cdot 2^{-n} = 4d^{1/2}2^{-(n+1)}$$

FACT B: $\forall x \in [0,1]^d$, since $g_n(x), g_{n+1}(x) \in D_{n+1}$ and by FACT A, we have that they are connected by at most $4d$ neighbouring pairs in D_{n+1}, because each neighbouring pairs in D_{n+1} has distance $2^{-(n+1)}$. In addition, for each of these connecting neighbouring pairs $\{(s',t')\}$, by the conclusion of Step 1, we have, when $n \ge N_\omega$,

$$|X_{s'}(\omega) - X_{t'}(\omega)| \le 2^{-\gamma(n+1)}.$$

So, $\forall n \ge N_\omega$, by adding and subtracting these pairs, we can upper-bounded

$$|X_{g_n(x)}(\omega) - X_{g_{n+1}(x)}(\omega)| \le 4d \cdot 2^{-\gamma(n+1)}.$$

FACT C: we have that

$$\|g_m(s) - g_m(t)\|_2 \leq \|g_m(s) - s\|_2 + \|g_m(t) - t\|_2 + \|s - t\|_2 \leq 4d^{1/2}2^{-m} + 2^{-m} \leq 5d^{1/2}2^{-m}.$$

Hence, $|X_{g_m(s)}(\omega) - X_{g_m(t)}(\omega)| \leq 5d2^{-\gamma m}$ by same logic as FACT B.

Let $\tilde{m} \in \mathbb{N}$ be such that $s, t \in D_{\tilde{m}}$ and note $\tilde{m} \geq m$. We thus have that

$$\begin{aligned}
|X_s(\omega) - X_t(\omega)| &\leq \left|X_{g_m(s)}(\omega) - X_{g_m(t)}(\omega)\right| \\
&+ \left|X_{g_{m+1}(s)}(\omega) - X_{g_m(s)}(\omega)\right| + \left|X_{g_{m+2}(s)}(\omega) - X_{g_{m+1}(s)}(\omega)\right| + \cdots + \left|X_s(\omega) - X_{g_{\tilde{m}-1}(s)}(\omega)\right| \\
&+ \left|X_{g_m(t)}(\omega) - X_{g_{m+1}(t)}(\omega)\right| + \cdots + \left|X_{g_{\tilde{m}-1}(t)}(\omega) - X_t(\omega)\right| \\
&\leq 5d2^{-\gamma m} + 2\sum_{k=1}^{\infty} 4d \cdot 2^{-\gamma(m+k)} \qquad (*) \\
&\leq 8d2^{-\gamma m}(1 + \sum_{k=1}^{\infty} 2^{-\gamma k}) \\
&\leq 8d(2^{-m})^{\gamma} \frac{1}{1 - 2^{-\gamma}} \leq C_{\gamma,d} \|s - t\|_2^{\gamma}.
\end{aligned}$$

Step 3: Define $\widetilde{X} : \Omega \to \mathbb{R}^{[0,1]^d}$ such that $\forall \omega \in B$,

$$\widetilde{X}_s(\omega) := \begin{cases} X_s(\omega) & \text{if } s \in D \\ \lim_{n\to\infty} X_{g_n(s)}(\omega) & \text{if } s \notin D \end{cases}.$$

For $\omega \notin B$, define $\widetilde{X}$ arbitrarily. Note that $\forall \omega \in B$, $s \in [0,1]^d$, $\lim_{n\to\infty} X_{g_n(s)}(\omega)$ exists since, $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ such that $\sup_{n,n' \geq N} |X_{g_n(s)}(\omega) - X_{g'_n(s)}(\omega)| < \varepsilon$, and so $\{X_{g_n(s)}(\omega)\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$. By the same logic as $(*)$, we have that, $\forall s, t \in [0,1]^d$,

$$|\widetilde{X}_s(\omega) - \widetilde{X}_t(\omega)| \leq C_{\gamma,d} \|s - t\|_2^{\gamma}.$$

We claim that $\widetilde{X}$ is a modification of X. Fix $t \in [0,1]^d$ and let $\{t_n\}_{n=1}^{\infty} \in D$ be a sequence such that $t_n \to t$. Then we have that $\forall \varepsilon > 0$, $\mathbb{P}(|X_{t_n} - X_t| > \varepsilon) \leq \frac{C\|t_n - t\|_2^{d+p}}{\varepsilon^{\alpha}} \to 0$ as $n \to \infty$. Thus $X_{t_n} \to X_t$ in probability, which implies there exists a subsequence $\{n_1, n_2, \dots\}$ such that $X_{t_{n_k}} \to X_t$ a.s. as $k \to \infty$ (by lemma 5.2.2). Since $X_{t_n} \to \widetilde{X}_t$ a.s., we have $X_t = \widetilde{X}_t$ a.s.

To show that X has a continuous modification over $\mathbb{R}^d$, note that $\exists B_1 \supseteq B_2 \supseteq \cdots \in \mathscr{F}$ such that for every $k \in \mathbb{N}$, $\mathbb{P}(B_k) = 1$ and $\exists \widetilde{X} : \Omega \to \mathbb{R}^{[-k,k]^d}$ such that $\forall \omega \in B_k$, $\widetilde{X}(\omega)(s) = X(\omega)(s)$ for all $s \in (D - (\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2})) \cdot 2k$ and $\widetilde{X}(\omega)$ is continuous. Thus, on $B = \cap_{k=1}^{\infty} B_k$, we have that $\widetilde{X} : \Omega \to \mathbb{R}^{\mathbb{R}^d}$ is continuous. Since $\mathbb{P}(B^c) \leq \sum_{k=1}^{\infty} \mathbb{P}(B_k^c) = 0$, the conclusion follows as desired. $\square$

Remark 8.2.2. It is necessary to work with modifications. Let $X : \Omega \to \mathbb{R}^{[0,\infty)}$ be a continuous process, i.e., $\forall \omega \in \Omega$, $X(\omega) : [0,\infty) \to \mathbb{R}$ is continuous. Let $h : \mathbb{R} \to \{0,1\}$ be such that $h(s) = 1$ if $s \in \mathbb{Q}$ and $h(s) = 0$ otherwise. Let $Z : \Omega \to \mathbb{R}$ be a random variable that is $N(0,1)$ distributed. Define $Y : \Omega \to \mathbb{R}^{[0,\infty)}$ such that $\forall \omega \in \Omega$, $\forall s \in [0,\infty)$,

$$Y(\omega)(s) := X(\omega)(s) + h(Z(\omega) + s).$$

Then, $Y(\omega)$ is nowhere continuous on $[0,\infty)$. And, $\forall s \in [0,\infty)$,

$$\mathbb{P}(\{\omega \in \Omega : Y(\omega)(s) = X(\omega)(s)\}) = \mathbb{P}(\{\omega \in \Omega : Z(\omega) + s \notin \mathbb{Q}\}) = 1,$$

which implies Y is a modification of X.

Remark 8.2.3. Since the standard Brownian motion $\{W_t\}_{[0,\infty)}$ almost surely has a continuous modification and since this modification is unique, we take $\{W_t\}_{t\in[0,\infty)}$ to be this *canonical modification.*

Define $C[0,1] := \{f \in \mathbb{R}^{[0,1]} : f \text{ is continuous } \}$ and for $f \in C[0,1]$, let $\|f\|_\infty := \sup_{x\in[0,1]} |f(x)|$ so that $\|\cdot\|_\infty$ is a norm on $C[0,1]$. Then, viewing $W : \Omega \to \mathbb{R}^{[0,1]}$, we see that the image of W lies in $C[0,1]$. Let $\mathscr{B}(C[0,1])$ be the Borel σ-field associated with $\|\cdot\|_\infty$ (generated by open sets where openness is defined with respect to $\|\cdot\|_\infty$). We will show later that $\mathscr{B}(\mathbb{R})^{\otimes[0,1]} \subset \mathscr{B}(C[0,1])$ and that W is $\mathscr{F}/\mathscr{B}(C[0,1])$–measurable as well. Therefore, we may view the distribution of Brownian motion $\mathbb{P}^{(W)}$ as a probability measure over $(C[0,1], \mathscr{B}(C[0,1]))$. This characterization is crucial for defining weak convergence of stochastic processes. Recall that weak convergence is only defined with respect to Borel probability measures over some metric space.

8.3 Path of Brownian Motion

Theorem 8.3.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $W : \Omega \to \mathbb{R}^{[0,\infty)}$ be a standard Brownian motion. Then, $\exists B \in \mathscr{F}$ with $\mathbb{P}(B) = 1$ such that $\forall \omega \in B$, $W(\omega)$ nowhere differentiable on $[0,\infty)$.

Proof.

Step 1: Fix $K \in \mathbb{N}$ arbitrarily and let $f : [0,K) \to \mathbb{R}$ be a function. Suppose f is differentiable at $t_0 \in [0,K)$ with $|f'(t_0)| \le M$ for some $M \in \mathbb{N}$. Since $|f'(t_0)| = \lim_{t\to t_0} \frac{|f(t)-f(t_0)|}{|t-t_0|} \le M$, $\exists n_0 \in \mathbb{N}$ such that for all $t \in [0,K)$ such that $t + \frac{3}{n_0} < K$ and

$$\sup_{t\in[0,K),|t-t_0|\le\frac{3}{n_0}} \frac{|f(t)-f(t_0)|}{|t-t_0|} \le 2M.$$

Let $n \ge n_0$, and let $k \in \{0,1,2,\ldots,Kn-1\}$ such that $\frac{k}{n} \le t_0 < \frac{k+1}{n}$. Then, $|\frac{k}{n} - t_0|$, $|\frac{k+1}{n} - t_0|$, $|\frac{k+2}{n} - t_0|$, $|\frac{k+3}{n} - t_0| \le \frac{3}{n_0}$ and thus,

$$\left|f(\frac{k}{n}) - f(\frac{k+1}{n})\right| \le \left|f(\frac{k}{n}) - f(t_0)\right| + \left|f(t_0) - f(\frac{k+1}{n})\right| \le \frac{2}{n}\cdot 2M$$

and

$$\left|f(\frac{k+1}{n}) - f(\frac{k+2}{n})\right| \le \frac{3}{n}\cdot 2M, \quad \left|f(\frac{k+3}{n}) - f(\frac{k+2}{n})\right| \le \frac{5}{n}\cdot 2M.$$

To summarize, if $f'(t_0)$ exists for some $t_0 \in [0,K)$, then there exists $M \in \mathbb{N}$ and $n_0 \in \mathbb{N}$ such that $\forall n \ge n_0$, $\exists k \in \{0,1,2,\ldots,Kn-1\}$ such that

$$\left|f(\frac{k}{n}) - f(\frac{k+1}{n})\right| \vee \left|f(\frac{k+2}{n}) - f(\frac{k+1}{n})\right| \vee \left|f(\frac{k+3}{n}) - f(\frac{k+2}{n})\right| \le \frac{10M}{n}.$$

Step 2: For $n \in \mathbb{N}$, $k \in \mathbb{N}\cup\{0\}$, define $X_{n,k} := |W_{\frac{k}{n}} - W_{\frac{k+1}{n}}| \vee |W_{\frac{k+2}{n}} - W_{\frac{k+1}{n}}| \vee |W_{\frac{k+3}{n}} - W_{\frac{k+2}{n}}|$. For $M, K, n \in \mathbb{N}$, define the event

$$A_n^{(M,K)} := \bigcup_{k=0}^{Kn-1} \left\{\omega \in \Omega : X_{n,k}(\omega) \le \frac{10M}{n}\right\} \in \mathscr{F}.$$

Then, we have that, for $n \geq 10M$,

$$\begin{aligned}
\mathbb{P}(A_n^{(M,K)}) &\leq \sum_{k=0}^{Kn-1} \mathbb{P}\left(X_{n,k} \leq \frac{10M}{n}\right) \\
&\leq \sum_{k=0}^{Kn-1} \mathbb{P}\left(|W_{\frac{1}{n}}| \leq \frac{10M}{n}\right)^3 \\
&= Kn \cdot \mathbb{P}\left(|W_{\frac{1}{n}}|\sqrt{n} \leq \frac{10M}{\sqrt{n}}\right)^3 = Kn \cdot \left\{\int_{-\frac{10M}{\sqrt{n}}}^{\frac{10M}{\sqrt{n}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} \mathrm{d}s\right\}^3 \\
&\leq Kn \left(\frac{20M}{\sqrt{2\pi n}}\right)^3 \to 0 \text{ as } n \to \infty.
\end{aligned}$$

Therefore, for any $M, K \in \mathbb{N}$, writing $A^{(M,K)} := \cup_{n_0=1}^{\infty} \cap_{n=n_0}^{\infty} A_n^{(M,K)}$, we have

$$\mathbb{P}(A^{(M,K)}) = \lim_{n_0 \to \infty} \mathbb{P}(\cap_{n=n_0}^{\infty} A_n^{(M,K)}) \leq \lim_{n_0 \to \infty} \mathbb{P}(A_{n_0}^{(M,K)}) = 0.$$

Since

$$\{\omega \in \Omega : W(\omega)'(t_0) \text{ exists for some } t_0 \in [0, K) \text{ and } |W(\omega)'(t_0)| \leq M\} \subseteq A^{(M,K)}$$

by step 1, and that $\mathbb{P}(\cup_{m \in \mathbb{N}} \cup_{K \in \mathbb{N}} A^{(M,K)}) = 0$, the conclusion holds as desired.

$\square$

Remark 8.3.1. Let $\mathscr{P} := \{0 = t_1 < t_2 < \cdots < t_K = 1 : K \in \mathbb{N}\}$ be the set of all partitions of $[0,1]$. We say that $f : [0,1] \to \mathbb{R}$ is of bounded total variation if $\sup_{(t_1,\ldots,t_K) \in \mathscr{P}} \sum_{k=1}^{K-1} |f(t_k) - f_{t_{k+1}}| < \infty$.

It is known that $f : [0,1] \to \mathbb{R}$ is bounded total variation if and only if there exists monotone $g, h : [0,1] \to \mathbb{R}$ such that $f = h - g$, which implies f is Lebesgue-a.e. differentiable on $[0,1]$. Thus, the path of Brownian motion is of unbounded total variation almost surely.

Theorem 8.3.2. Let $W : \Omega \to \mathbb{R}^{[0,\infty)}$ be standard Brownian motion. Let $\{(t_1^{(n)}, \ldots, t_{K_n}^{(n)}) \in \mathscr{P}\}_{n \in \mathbb{N}}$ be a sequence of partitions of $[0,1]$. If $\lim_{n\to\infty} \max_{i \in [K_n - 1]} |t_{i+1}^{(n)} - t_i^{(n)}| = 0$, then

$$\sum_{i=1}^{K_n - 1} |W_{t_{i+1}^{(n)}} - W_{t_i^{(n)}}|^2 \to 1 \quad \text{in } L_2(\mathbb{P}).$$

Proof.

Define

$$\mathscr{Q}_n := \sum_{i=1}^{K_n - 1} |W_{t_{i+1}^{(n)}} - W_{t_i^{(n)}}|^2,$$

then

$$
\begin{aligned}
\mathbb{E}(\mathscr{Q}_n - 1)^2 &= \mathbb{E}\left\{\sum_{i=1}^{K_n-1}\left|W_{t_{i+1}^{(n)}} - W_{t_i^{(n)}}\right|^2 - (t_{i+1}^{(n)} - t_i^{(n)})\right\}^2 \\
=&\mathbb{E}\sum_{i=1}^{K_n-1}\sum_{j=1}^{K_n-1}\left\{\left|W_{t_{i+1}^{(n)}} - W_{t_i^{(n)}}\right|^2 - (t_{i+1}^{(n)} - t_i^{(n)})\right\}^2\left\{\left|W_{t_{j+1}^{(n)}} - W_{t_j^{(n)}}\right|^2 - (t_{j+1}^{(n)} - t_j^{(n)})\right\}^2 \\
=&\mathbb{E}\sum_{i=1}^{K_n-1}\left\{\left|W_{t_{i+1}^{(n)}} - W_{t_i^{(n)}}\right|^2 - (t_{i+1}^{(n)} - t_i^{(n)})\right\}^2 \qquad (*) \\
=&\mathbb{E}\sum_{i=1}^{K_n-1}(Z_i^2 - 1)^2(t_{i+1}^{(n)} - t_i^{(n)})^2 \text{ where } Z_1, \ldots, Z_{K_n-1} \sim N(0,1) \\
\leq&\mathbb{E}(Z_1^2 - 1)^2 \cdot \max_{i\in[K_n-1]}|t_{i+1}^{(n)} - t_i^{(n)}| \cdot \underbrace{\sum_{i=1}^{K_n-1}|t_{i+1}^{(n)} - t_i^{(n)}|}_{=1} \to 0.
\end{aligned}
$$

The simplification in the $(*)$ line is because the two terms of the product would be independent for $i \neq j$ as $W_{t_{i+1}^{(n)}} - W_{t_i^{(n)}} \sim N(0, t_{i+1}^{(n)} - t_i^{(n)}) \perp\!\!\!\perp W_{t_{j+1}^{(n)}} - W_{t_j^{(n)}} \sim N(0, t_{j+1}^{(n)} - t_j^{(n)})$. □

Remark 8.3.2. We say W be quadratic variation 1 on $[0,1]$. More generally, we can show that W has quadratic variation $b-a$ on $[a,b]$.

Proposition 8.3.1. Let $W : \Omega \to \mathbb{R}^{[0,\infty)}$ be standard Brownian motion. Then,

$$\mathbb{P}\left(\sup_{t\geq 0} W_t = \infty,\ \inf_{t\geq 0} W_t = -\infty\right) = 1.$$

Proof.

By proposition 8.1.1 (c), for any $a > 0$, $\{aW_{t/a^2}\}_{t\geq 0} \overset{d}{=} \{W_t\}$. Define $Z := \sup_{t\geq 0} W_t \geq 0$, then, for any $a > 0$,

$$Z = \sup_{t\geq 0} W_t \overset{d}{=} \sup_{t\geq 0} a \cdot W_{t/a^2} = a\left(\sup_{t\geq 0} W_t\right) = aZ.$$

Therefore,

$$
\begin{aligned}
\mathbb{P}(Z=0) = \lim_{n\to\infty}\mathbb{P}\left(Z \in [0, n^{-1})\right) = \lim_{n\to\infty}\mathbb{P}\left(\frac{1}{n^2}Z \in [0, n^{-1})\right) = \lim_{n\to\infty}\mathbb{P}(Z \in [0, n)) = \mathbb{P}(Z \in [0,\infty)) \\
\implies \mathbb{P}(Z \in \{0, \infty\}) = 1.
\end{aligned}
$$

We need only show that $\mathbb{P}(Z = 0) = 0$. By Proposition 8.1.1 (b), we have

$$Z = \sup_{t\in[0,1]} W_t + W_1 + \sup_{t\geq 0}(W_{t+1} - W_1) \overset{d}{=} \underbrace{\sup_{t\in[0,1]} W_t}_{\geq 0} + W_1 + \widetilde{Z}$$

where $\widetilde{Z}$ is a random variable that is identically distributed as Z and independent of W_1 and $\sup_{t\in[0,1]} W_t$. We thus have

$$
\begin{aligned}
\mathbb{P}(Z=\infty) = 1 - \mathbb{P}(Z=0) = \mathbb{P}(Z>0) &\geq \mathbb{P}(\widetilde{Z} + W_1 > 0) \\
&\geq \mathbb{P}(\widetilde{Z} + W_1 > 0|\widetilde{Z} = \infty)\mathbb{P}(\widetilde{Z} = \infty) + \mathbb{P}(W_1 > 0|\widetilde{Z} = 0)\mathbb{P}(\widetilde{Z} = 0) \\
&= \mathbb{P}(\widetilde{Z} = \infty) + \frac{1}{2}\mathbb{P}(\widetilde{Z} = 0) = \mathbb{P}(Z = \infty) + \frac{1}{2}\mathbb{P}(Z = 0),
\end{aligned}
$$

which implies $\mathbb{P}(Z = 0) = 0$. Thus $\mathbb{P}(\sup_{t\geq 0} W_t = \infty) = 1$ and conclusion follows by symmetry. □

8.4 Strong Markov Property

Definition 8.4.1. (1) We say, for $T \subseteq [0, \infty)$, that a family of σ-field $\{\mathscr{F}_t\}_{t \in T}$ is a filtration if $\forall s \leq t \in T$, $\mathscr{F}_s \subseteq \mathscr{F}_t$.

(2) A random variable $\tau : \Omega \to T \cup \{\infty\}$ is a stopping time w.r.t. $\{\mathscr{F}_t\}$ if, $\forall t \in T$, $\{\tau \leq t\} \in \mathscr{F}_t$.

(3) We say that τ is a weakly-optional time w.r.t. $\{\mathscr{F}_t\}$ if $\forall t \in T$, $\{\tau < t\} \in \mathscr{F}_t$.

(4) A process $X : \Omega \to \mathbb{R}^T$ is adapted w.r.t. $\{\mathscr{F}_t\}$ if, $\forall t \in T$, $X_t : \Omega \to \mathbb{R}$ is $\mathscr{F}_t/\mathscr{B}(\mathbb{R})$-measurable.

Remark 8.4.1. Fix filtration $\{\mathscr{F}_t\}_{t \geq 0}$. If τ is a stopping time, then it is a weakly-optional time. To see this, note that $\forall t > 0$, $\{\tau < t\} = \cup_{n=1}^{\infty} \{\tau \leq t - \frac{1}{n}\}$.

Example 8.4.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $W : \Omega \to \mathbb{R}^{[0,\infty)}$ be standard Brownian motion. Let $T = [0, \infty)$ and define, for $t \geq 0$, $\mathscr{B}_t := \sigma(\{W_s : s \leq t\})$, i.e., let $\mathscr{B}_t$ be smallest σ-field such that W_s is $\mathscr{B}_t/\mathscr{B}(\mathbb{R})$-measurable $\forall s \in [0, t]$. Then $\{\mathscr{B}_t\}$ is a filtration and W is adapted on $\{\mathscr{B}_t\}_{t \geq 0}$ by definition.

Proposition 8.4.1. Let $X : \Omega \to \mathbb{R}^{[0,\infty)}$ be an adapted continuous process w.r.t. filtration $\{\mathscr{F}_t\}_{t \geq 0}$. For $A \subseteq \mathbb{R}$, define the hitting time

$$\tau_A := \inf\{t \geq 0 : X_t \in A\},$$

where $\inf \emptyset := \infty$. If A is closed, then τ_A is a stopping time and if A is open, then τ_A is a weakly-optional time.

Proof.

Since $X(\omega) : [0, \infty) \to \mathbb{R}$ is continuous, $\forall \omega \in \Omega$, for any $t_0 \geq 0$, $X(\omega)([0, t_0]) \subseteq \mathbb{R}$ is a closed interval.

Let $A \subseteq$ be closed. For any $t_0 \in [0, \infty)$, $X(\omega)([0, t_0]) \cap A \neq \emptyset$ implies that $\tau_A(\omega) \leq t_0$. Conversely, $X(\omega)([0, t_0]) \cap A = \emptyset$ implies $\exists \varepsilon > 0$ such that $X(\omega)([0, t_0 + \varepsilon]) \cap A = \emptyset$, which further implies $\tau_A(\omega) > t_0$.

Thus,

$$\begin{aligned}
\{\omega : \tau_A(\omega) \leq t_0\} &= \{\omega : X(\omega)([0, t_0]) \cap A \neq \emptyset\} \\
&= \{\omega : \overline{X(\omega)(\mathbb{Q} \cap [0, t_0])} \cap A \neq \emptyset\} \\
&= \left\{\omega : \forall n \in \mathbb{N}, \exists s \in \mathbb{Q} \cap [0, t_0] \text{ s.t. } \left(X_s(\omega) - \frac{1}{n}, X_s(\omega) + \frac{1}{n}\right) \cap A \neq \emptyset\right\} \\
&= \bigcap_{n \in \mathbb{N}} \bigcup_{s \in \mathscr{Q} \cap [0, t_0]} \left\{\omega : X_s(\omega) \in A + \left(-\frac{1}{n}, \frac{1}{n}\right)\right\} \in \mathscr{F}_{t_0}.
\end{aligned}$$

Now, let A be open, then $\forall t_0 \geq 0$, $X(\omega)([0, t_0]) \cap A \neq \emptyset$ implies that $\exists \varepsilon > 0$ such that $X(\omega)([0, t_0 - \varepsilon]) \cap A \neq \emptyset$, which further implies $\tau_A(\omega) < t_0$. Conversely, $X(\omega)([0, t_0]) \cap A = \emptyset \implies \tau_A(\omega) \geq t_0$. By same logic as before, the claim follows.

$\square$

Remark 8.4.2. (1) If A is open, τ_A may not be a stopping time.

For example, for $a > 0$, let $\tau = \inf\{t \geq 0 : W_t > a\}$. Suppose $\omega \in \Omega$ is such that $W_t(\omega) = a$, $W_s(\omega) \leq a$, $\forall s \leq t$. If, for some $\varepsilon > 0$, we have $W(\omega)([t, t + \varepsilon]) \cap (a, \infty) \neq \emptyset$, then $\tau(\omega) = t$. On the other hand, if, for some $\varepsilon > 0$, we have that $W(\omega)([t, t + \varepsilon]) \subseteq (-\infty, a]$, then $\tau(\omega) > t$. Thus, $\mathscr{B}_t$ may not have enough information to determine whether $\tau = t$.

(2) If a filtration $\{\mathscr{F}_t\}$ satisfies the condition that $\forall t \in T$, $\cap_{\{s \in T, s \leq t\}} \mathscr{F}_s = \mathscr{F}_t$, then any weakly-optional time is also a stopping time, and we say that $\{\mathscr{F}_t\}$ is right-continuous.

Definition 8.4.2. Fix $T \subseteq [0,\infty)$ and filtration $\{\mathscr{F}_t\}_{t\in T}$. For a stopping time $\tau : \Omega \to T \cup \{\infty\}$, define the "events prior to τ" as $\mathscr{F}_\tau := \{A \in \mathscr{F} : A \cap \{\tau \le t\} \in \mathscr{F}_t, \forall t \in T\}$.

Proposition 8.4.2. (a) $\mathscr{F}_\tau$ is a σ-field and τ is $\mathscr{F}_\tau/\mathscr{B}(T)$-measurable.

(b) If τ, π are stopping time and $\tau \le \pi$, then $\mathscr{F}_\tau \subseteq \mathscr{F}_\pi$.

(c) If τ, π are stopping time, $\tau + \pi$,$\tau \vee \pi$, $\tau \wedge \pi$ are stopping times.

(d) If $U : \Omega \to T$ is $\mathscr{F}_\tau/\mathscr{B}(T)$-measurable, and $\tau \le U$, then U is a stopping time.

Proof.

(a) Let $A \in \mathscr{F}_\tau$, then $\forall t \in T$, $A^c \cap \{\tau \le t\} = \{\tau \le t\} \setminus (A \cap \{\tau \le t\}) \in \mathscr{F}_t$.

Closure under countable intersection follows similarly.

For any $s \in T$, since $\{\tau \le s\} \in \mathscr{F}_s$, it holds that $\{\tau \le s\} \cap \{\tau \in t\} \in \mathscr{F}_{s\wedge t} \subseteq \mathscr{F}_t$. Hence, for any $s \in T$, $\{\tau \le s\} \in \mathscr{F}_\tau$ and τ is $\mathscr{F}_\tau/\mathscr{B}(T)$-measurable.

(b) Let $t \in T$, then $A \in \mathscr{F}_\tau$ implies

$$A \cap \{\pi \le t\} = \underbrace{A \cap \{\tau \le t\}}_{\in \mathscr{F}_t} \cap \{\pi \le t\} \in \mathscr{F}_t \implies A \in \mathscr{F}_\pi$$

where the first inequality holds because $\{\omega : \pi(\omega) \le t\} \subseteq \{\omega : \tau(\omega) \le t\}$ as $\tau \le \pi$.

(c) $\{\tau \vee \pi \le t\} = \{\tau \le t\} \cup \{\pi \le t\} \in \mathscr{F}_t$, $\{\tau \wedge \pi \le t\} = \{\tau \le t\} \cap \{\pi \le t\} \in \mathscr{F}_t$.

Now, $\tau + \pi = \tau \vee \pi + \tau \wedge \pi$. Since $\tau \wedge \pi$ is $\mathscr{F}_{\tau\vee\pi}/\mathscr{B}(T)$-measurable by (a) and (b), we have that $\tau + \pi$ is $\mathscr{F}_{\tau\vee\pi}/\mathscr{B}([0,\infty])$-measurable. Thus, $\forall t \ge 0$,

$$\{\tau + \pi \le t\} = \{\tau + \pi \le t\} \cap \{\tau \vee \pi \le t\} \in \mathscr{F}_t.$$

(d) Let $t \in T$, $\{U \le t\} = \{U \le t\} \cap \{\tau \le t\} \in \mathscr{F}_t$ since $\{U \le t\} \in \mathscr{F}_\tau$.

□

Remark 8.4.3. Note that for any $s \ge 0$, we have $W_{s+t} - W_s \perp\!\!\!\perp W_{s'}$, $\forall t \ge 0$ and $s' \in [0,s]$ by the independent increment property. Using Kolmogorov Extension Theorem, we may show that $\{W_{s+t} - W_s\}_{t\in[0,\infty)} \perp\!\!\!\perp \{W_{s'}\}_{s'\in[0,s]}$.

Thus, we say $W_{s+t} - W_s$ is independent of σ-field $\mathscr{B}_s = \sigma(\{W_{s'} : s' \in [0,s]\}) \subseteq \mathscr{F}$ in the sense that $\forall A \in \mathscr{B}(\mathbb{R})^{[0,\infty)}$, $\widetilde{A} \in \mathscr{B}_s$ (any $\widetilde{A} \in \mathscr{B}_s$ can be written as $\{W_{[0,s]} \in A'\}$ for some $A' \in \mathscr{B}(\mathbb{R}^{\otimes[0,s]})$),

$$\mathbb{P}(\{W_{s+t} - W_s\}_{t\in[0,\infty)} \in A, \widetilde{A}) = \mathbb{P}(\{W_{s+t} - W_s\}_{t\in[0,\infty)} \in A)\mathbb{P}(\widetilde{A}). \qquad (*)$$

If $\{\mathscr{F}_t\}_{t\ge 0}$ is a filtration such that $(*)$ holds, we say $\{W_t\}$ is Markovian w.r.t. $\{\mathscr{F}_t\}$. If s is a random stopping time, then $(*)$ in fact also holds.

Before stating the strong Markov property, we will state a lemma that is used in its proof.

Lemma 8.4.1. Let $X, X_1, X_2, \ldots$ be random variables such that $X_n \to X$ almost surely. Let $A \in \mathscr{B}(\mathbb{R})$. Then,

1. if A is open, then $\liminf_{n\to\infty} \mathbb{P}(X_n \in A) \ge \mathbb{P}(X \in A)$ and
2. if A is closed, then $\limsup_{n\to\infty} \mathbb{P}(X_n \in A) \le \mathbb{P}(X \in A)$.

Proof.

Suppose A is open. Then, $x \mapsto \mathbb{1}\{x \in A\}$ is lower semi-continuous and thus, $\liminf_{n\to\infty} \mathbb{1}\{X_n \in A\} \geq \mathbb{1}\{X \in A\}$. The conclusion thus follows by Fatou's lemma. If A is closed, we may obtain the desired conclusion by analyzing A^c. $\square$

Theorem 8.4.1 (Strong Markov Property). Let $W : \Omega \to \mathbb{R}^{[0,\infty)}$ be a standard Brownian motion and let $\{\mathscr{F}_t\}_{t\geq 0}$ be any filtration for which W is adapted and Markovian. Let $\tau : \Omega \to [0, \infty]$ be an almost surely finite stopping time.

(1) Let $W_\tau : \Omega \to \mathbb{R}$ where $W_\tau(\omega) = W_{\tau(\omega)}(\omega)$, $\forall \omega \in \Omega$ (where we define $W_\infty(\omega) := 0$). Then W_τ is $\mathscr{F}_\tau/\mathscr{B}(\mathbb{R})$-measurable.

(2) $\{\widetilde{W}_t\}_{t\geq 0} := \{W_{\tau+t} - W_\tau\}_{t\geq 0}$ is a standard Brownian motion and independent of $\mathscr{F}_\tau$ in the sense $\forall A \in \mathscr{B}(\mathbb{R})^{\otimes[0,\infty)}$, $\forall \widetilde{A} \in \mathscr{F}_\tau$,

$$\mathbb{P}(\{\widetilde{W} \in A\} \cap \widetilde{A}) = \mathbb{P}(\widetilde{W} \in A)\mathbb{P}(\widetilde{A}).$$

In particular, $\widetilde{W}$ is independent of τ, W_τ, or any $\mathscr{F}_\tau/\mathscr{B}(\mathbb{R})$-measurable random variable.

Note that $\forall a > 0$, with $\tau_a := \inf\{t \geq 0 : W_t = a\}$, we have $\mathbb{P}(\tau_A < \infty) = 1$ by Proposition 8.3.1.

Proof.

Let $B \in \mathscr{F}$ such that $\mathbb{P}(B) = 1$ and $\tau(\omega) < \infty$, $\forall \omega \in B$.

To prove the first claim, fix $n \in \mathbb{N}$ arbitrarily. We have that $B = \cup_{k=0}^\infty \{\frac{k}{n} \leq \tau < \frac{k+1}{n}\}$. Define $\tau_n : \Omega \to [0, \infty)$ by the following: for $\omega \in B$, define $\tau_n(\omega) = \frac{k}{n}$ where $k \in \mathbb{N}$ satisfies $\frac{k}{n} \leq \tau(\omega) < \frac{k+1}{n}$; for $\omega \notin B$, define $\tau_n(\omega) = \infty$ (this value does not matter). Note that τ_n is not necessarily a stopping time.

We claim that W_{τ_n} is $\mathscr{F}_\tau/\mathscr{B}(\mathbb{R})$–measurable. To see this, observe that for all $a \in \mathbb{R}$, for all $t \geq 0$,

$$\begin{aligned}
\{W_{\tau_n} \leq a\} \cap \{\tau \leq t\} &= \bigcup_{k=0}^{\infty} \{W_{\tau_n} \leq a\} \cap \{\tau \leq t\} \cap \{\tau_n = \frac{k}{n}\} \\
&= \bigcup_{k=0}^{\infty} \underbrace{\{W_{k/n} \leq a\}}_{\in \mathscr{F}_t \text{ since } W \text{ is adapted}} \cap \underbrace{\{\tau \leq t\} \cap \{\frac{k}{n} \leq \tau < \frac{k+1}{n}\}}_{\in \mathscr{F}_t \text{ since } \tau \text{ is } \mathscr{F}_t/\mathscr{B}([0,\infty))\text{–meas.}}
\end{aligned}$$

Therefore, $\{W_{\tau_n} \leq a\} \cap \{\tau \leq t\} \in \mathscr{F}_t$ and W_{τ_n} is $\mathscr{F}_\tau/\mathscr{B}([0, \infty))$–measurable. Since for all $\omega \in B$, we have $\lim_{n\to\infty} \tau_n(\omega) = \tau(\omega)$ and thus, By continuity of W, it holds that $\lim_{n\to\infty} W_{\tau_n}(\omega) = W_\tau(\omega)$. Thus, W_τ is $\mathscr{F}_\tau/\mathscr{B}(\mathbb{R})$-measurable.

For the second claim, we first show that for any $t \geq 0$, the random variable $W_{t+\tau} - W_\tau$ is independent of $\mathscr{F}_\tau$. For $n \in \mathbb{N}$, redefine $\tau_n : \Omega \to [0, \infty)$ by

$$\forall \omega \in B, \tau_n(\omega) = \frac{k}{n} \text{ if } \frac{k-1}{n} \leq \tau(\omega) < \frac{k}{n} \text{ for } k \in \mathbb{N}, \quad \forall \omega \notin B, \tau_n(\omega) := \infty.$$

Then, τ_n is $\mathscr{F}_\tau/\mathscr{B}([0, \infty))$-measurable and a stopping time by Proposition 8.4.2(d). Hence, if $\widetilde{A} \in \mathscr{F}_\tau$, then

$$\widetilde{A} \cap \left\{\tau_n = \frac{k}{n}\right\} = \underbrace{\widetilde{A} \cap \left\{\tau \leq \frac{k}{n}\right\}}_{\in \mathscr{F}_{\frac{k}{n}} \text{ since } \widetilde{A} \in \mathscr{F}_\tau} \cap \underbrace{\left\{\tau_n = \frac{k}{n}\right\}}_{\substack{\in \mathscr{F}_{\frac{k}{n}} \text{ b/c } \tau_n \\ \text{is stopping time}}} \in \mathscr{F}_{\frac{k}{n}}.$$

Since W is Markovian on $\{\mathscr{F}_t\}_{t\geq 0}$, we have $\forall A \in \mathscr{B}(\mathbb{R})$, $\widetilde{A} \in \mathscr{F}_\tau$, $\forall t \geq 0$,

$$\begin{aligned}
\mathbb{P}(\{W_{t+\tau_n} - W_{\tau_n} \in A\} \bigcap \widetilde{A}) &= \mathbb{P}\Big(\{W_{t+\tau_n} - W_{\tau_n} \in A\} \cap \widetilde{A} \bigcap \Big\{\cup_{k=1}^{\infty} \{\tau_n = \frac{k}{n}\Big\}\Big) \\
&= \sum_{k=0}^{\infty} \mathbb{P}\Big(\{W_{t+\frac{k}{n}} - W_{\frac{k}{n}} \in A\} \cap \underbrace{\widetilde{A} \cap \{\tau_n = \frac{k}{n}\}}_{\in \mathscr{F}_{\frac{k}{n}}}\Big) \\
&= \sum_{k=0}^{\infty} \mathbb{P}\left(W_{t+\frac{k}{n}} - W_{\frac{k}{n}} \in A\right) \mathbb{P}(\widetilde{A} \cap \{\tau_n = \frac{k}{n}\}) \\
&= \mathbb{P}(W_t \in A) \cdot \sum_{k=0}^{\infty} \mathbb{P}\left(\widetilde{A} \cap \left\{\tau_n = \frac{k}{n}\right\}\right) \\
&= \mathbb{P}(W_t \in A)\mathbb{P}(\widetilde{A}). \qquad (\star)
\end{aligned}$$

Now take $A = (a, b)$ for $a \leq b \in \mathbb{R}$. Since for almost all $\omega \in \Omega$, $\tau_n(\omega) \to \tau(\omega)$ and thus, by continuity of W, $W_{t+\tau_n}(\omega) - W_{\tau_n}(\omega) \to W_{t+\tau}(\omega) - W_\tau(\omega)$. By Lemma 8.4.1, we have that , $\forall \widetilde{A} \in \mathscr{F}_\tau$,

$$\begin{aligned}
\mathbb{P}(\{a < W_{t+\tau} - W_\tau < b\} \cap \widetilde{A}) &\leq \liminf_{n\to\infty} \mathbb{P}(\{a < W_{t+\tau_n} - W_{\tau_n} < b\} \cap \widetilde{A}) \\
&\leq \limsup_{n\to\infty} \mathbb{P}(\{a \leq W_{t+\tau_n} - W_{\tau_n} \leq b\} \cap \widetilde{A}) \\
&\leq \mathbb{P}(\{a \leq W_{t+\tau} - W_\tau \leq b\} \cap \widetilde{A}).
\end{aligned}$$

Note by $(\star)$,

$$\liminf_{n\to\infty} \mathbb{P}(\{a < W_{t+\tau_n} - W_{\tau_n} < b\} \cap \widetilde{A}) = P(\widetilde{A}) \cdot \liminf_{n\to\infty} \mathbb{P}(a < W_t < b) = \mathbb{P}(\widetilde{A})\mathbb{P}(a < W_t < b),$$

so we obtain

$$\mathbb{P}(\{a < W_{t+\tau} - W_\tau\} \cap \widetilde{A}) \leq \mathbb{P}(\widetilde{A})\mathbb{P}(a < W_t < b) \leq \mathbb{P}(\{a \leq W_{t+\tau} - W_\tau \leq b\} \cap \widetilde{A}).$$

For any $a \in \mathbb{R}$, for $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}(W_{t+\tau} - W_\tau = a) &\leq \liminf_{n\to\infty} \mathbb{P}(a - \varepsilon < W_{t+\tau_n} - W_{\tau_n} < a + \varepsilon) \\
&\leq \mathbb{P}(a - \varepsilon \leq W_t \leq a + \varepsilon) \qquad (\text{by } (*) \text{ with } \widetilde{A} = \Omega)
\end{aligned}$$

Letting $\varepsilon \to 0$ shows that $\mathbb{P}(W_{t+\tau} - W_\tau = a) = 0$. Thus, we have that $\mathbb{P}(\{a < W_{t+\tau} - W_\tau < b\} \cap \widetilde{A}) = \mathbb{P}(a < W_t < b)\mathbb{P}(\widetilde{A})$.

Similarly, one may show that for any finite $\{t_1, \ldots, t_n\} \subseteq [0, \infty)$, $(W_{t_1+\tau} - W_\tau, \ldots, W_{t_n+\tau} - W_\tau)$ is independent of $\mathscr{F}_\tau$. By Kolmogorov extension theorem, we have that $\{W_{t+\tau} - W_\tau\}_{t\geq 0} \perp\!\!\!\perp \mathscr{F}_\tau$ as desired. ☐

Theorem 8.4.2 (Reflection Principle). Let W be a standard Brownian Motion. For any $t \geq 0$, $a > 0$,

$$\mathbb{P}(\sup_{s\in[0,t]} W_s \geq a) = 2\mathbb{P}(W_t \geq a).$$

As an immediate consequence, $\sup_{s\in[0,t]} W_s \overset{d}{=} |W_t|$.

Proof.

Define $\tau := \inf\{t \geq 0 : W_t \geq a\}$. We know $\tau < \infty$ a.s. since $\mathbb{P}(\sup_{s\geq 0} W_s = \infty) = 1$. The claim is trivially true for $t = 0$. We offer two different proofs.

First proof. Define the stochastic process $\tilde{W}$ such that $\tilde{W}_s = W_{\tau+s} - W_\tau$. Then, by the Strong Markov property, $\tilde{W}$ is a standard Brownian motion independent of τ. Hence, we have that $(-\tilde{W}, \tau) \overset{d}{=} (\tilde{W}, \tau)$.

Therefore, we obtain

$$
\begin{aligned}
\mathbb{P}(W_t \geq a) &= \mathbb{P}(W_t \geq a, \tau \leq t) = \mathbb{P}(W_{\tau-(\tau-t)} - W_\tau \geq 0, \tau \leq t) \\
&= \mathbb{P}(\tilde{W}_{\tau-t} \geq 0, \tau \leq t) \\
&= \frac{1}{2}\mathbb{P}(\tilde{W}_{\tau-t} \geq 0, \tau \leq t) + \frac{1}{2}\mathbb{P}(-\tilde{W}_{\tau-t} \geq 0, \tau \leq t) \\
&= \frac{1}{2}\mathbb{P}(\tau \leq t) = \frac{1}{2}\mathbb{P}(\sup_{s\in[0,t]} W_s \geq a).
\end{aligned}
$$

Second proof. Fix $t > 0$, we have

$$
\mathbb{P}(W_t \geq a) = \mathbb{P}(W_t \geq a, \tau \leq t) = \int_{[0,t]} \mathbb{P}(W_t \geq a|\tau = s)\mathrm{d}\mathbb{P}^{(\tau)}(s) = \int_{[0,t)} \mathbb{P}(W_t \geq a|\tau = s)\mathrm{d}\mathbb{P}^{(\tau)}(s)
$$

since $\mathbb{P}^{(\tau)}(\{t\}) = \mathbb{P}(\tau = t) = \mathbb{P}(W_t = a) = 0$.

By property of regular conditional distribution, we have that for $\mathbb{P}^{(\tau)}$-a.e. $s \in [0,t)$,

$$
\begin{aligned}
\mathbb{P}(W_t \geq a|\tau = s) &= \mathbb{P}(W_{\tau+(t-s)} - a \geq 0|\tau = s) \\
&= \mathbb{P}(W_{\tau+(t-s)} - W_\tau|\tau = s) && (W_\tau = a) \\
&= \mathbb{P}(W_{t-s} \geq 0) = \frac{1}{2}. && \text{(Strong Markov Property)}
\end{aligned}
$$

The first equality for the reason as follows. For fixed $s < t$, define $f(W,\tau) = \mathbb{1}_{\{W_{\tau+(t-s)}-a\geq 0\}}$, so that $f(w,s) = \mathbb{1}_{\{W_t - a \geq 0\}}$ for $w \in \mathbb{R}^T$. We have then

$$
\mathbb{E}[f(W,\tau)|\tau = s] = \int_C f(\omega, s) dP(\cdot|\tau = s)(\omega) = \mathbb{P}(W_t \geq a|\tau = s).
$$

Hence, $2\mathbb{P}(W_t \geq a) = \mathbb{P}(\tau < t) = \mathbb{P}(\tau \leq t) = \mathbb{P}(\sup_{s\in[0,t]} W_s \geq a)$. □

Corollary 8.4.1. For any $a > 0$, $\tau_a := \inf\{t \geq 0 : W_t \geq a\}$ has the density

$$
\frac{\mathrm{d}\mathbb{P}^{(\tau_a)}}{\mathrm{dLeb}}(t) = \frac{a}{\sqrt{2\pi}} t^{-\frac{3}{2}} \exp\left\{-\frac{a^2}{2t}\right\} \quad \text{for } t \geq 0.
$$

In particular, $\mathbb{E}\tau_a = \infty$.

Proof.

For any $t > 0$, we have

$$
\begin{aligned}
\mathbb{P}(\tau_a \leq t) &= 2\mathbb{P}(W_t \geq a) = 2\mathbb{P}\left(\frac{1}{\sqrt{t}}W_t \geq \frac{a}{\sqrt{t}}t\right) && \text{(Reflection Principle)} \\
&= 2\int_{\frac{a}{\sqrt{t}}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx = 2\left(1 - \int_{-\infty}^{\frac{a}{\sqrt{t}}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx\right).
\end{aligned}
$$

Thus, $\frac{\mathrm{d}\mathbb{P}(\tau\leq t)}{\mathrm{d}t} = (2\pi)^{-1/2}\exp\{-\frac{a^2}{2t}\}\frac{a}{t^{3/2}}$.

□

Remark 8.4.4. To intuitively see that, for any $a > 0$, $\mathbb{E}\tau = \infty$, we define $\pi := \inf\{t \geq 0 : W_t \leq -a\}$. Then $\pi \overset{d}{=} \tau$ since $-W \overset{d}{=} W$. Thus,

$$
\mathbb{E}\tau = \frac{1}{2}(\mathbb{E}\tau + \mathbb{E}\pi) = \frac{1}{2}\mathbb{E}\tau \wedge \pi + \frac{1}{2}\mathbb{E}\tau \vee \pi.
$$

Now,

$$\begin{aligned}
\tau \vee \pi &= \inf\left\{t \geq 0 : \sup_{s \leq t} W_s \geq a,\ \inf_{s \leq t} W_s \leq -a\right\} \\
&\overset{d}{=} \inf\left\{t \geq 0 : \sup_{s \leq t} W_s \geq 3a\right\} && \text{(Reflection Principle)} \\
&\overset{d}{=} 3\tau, && \text{(Strong Markov Property)}
\end{aligned}$$

which implies $\mathbb{E}\tau = \frac{1}{2}\mathbb{E}\tau \wedge \pi + \frac{3}{2}\mathbb{E}\tau$. So $\mathbb{E}\tau \geq \frac{3}{2}\mathbb{E}\tau$ implies $\mathbb{E}\tau = \infty$.

Remark 8.4.5. Define $\tau_1 := \inf\{t \geq 0 : |W_t| \geq 1\}$, $\tau_2 := \inf\{t \geq 0 : |W_{t+\tau_1} - W_{\tau_1}| \geq 1\}$, ... By strong Markov property, $\{\tau_1, \tau_2, \dots\}$ are i.i.d. One may show, through an argument similar to Proposition 8.4.1, that

$$\tau_1 + \tau_2 = \inf\{t \geq \tau_1 : |W_t - W_{\tau_1}| \geq 1\}$$

is a stopping time.

Let $\{X_1, X_2, \dots\}$ be iid binary random variables such that $X_i = 1$ with probability $1/2$ and $X_i = -1$ with probability $1/2$. Let $S_n = \sum_{i=1}^n X_i$. Then, by Strong Markov property and reflection principle, $W_{\tau_1}, W_{\tau_2}, \dots) \overset{d}{=} (S_1, S_2, \dots)$.

We observe that for any $n \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{P}(\tau_1 \geq n) &\leq \mathbb{P}(|W_1| \leq 2, |W_1 - W_2| \leq 2, |W_2 - W_3| \leq 2, \dots, |W_n - W_{n-1}| \leq 2) \\
&= \mathbb{P}(|W_1| \leq 2)^n = e^{-cn}
\end{aligned}$$

for some $c > 0$.

We will later show that $\mathbb{E}\tau_i = 1$, which implies $(\sum_{i=1}^n \tau_i)/n \xrightarrow{\text{a.s.}} 1$. This is known as the Skorokhod embedding of S_n.

Theorem 8.4.3 (Skorokhod Embedding). Let $W : \Omega \to \mathbb{R}^{[0,\infty)}$ be a standard Brownian motion and let P be a distribution on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with mean 0 and variance $\sigma^2 < \infty$. There exists a stopping time $\tau : \Omega \to [0, \infty]$ such that

(a) W_τ has distribution P, i.e., $\mathbb{P}^{(W_\tau)} = P$.

(b) $\mathbb{E}\tau = \sigma^2$, $\mathbb{E}\tau^2 \leq \int_{\mathbb{R}} x^4 dP(x)$.

Proof.

See, e.g. Billingsley Theorem 37.6. □

Corollary 8.4.2. Let $X_1, X_2, \dots$ be independent random variables with mean 0 and finite variance. Let $S_n := \sum_{i=1}^n X_i$ for $n \in \mathbb{N}$. There exist stopping times $\pi_1, \pi_2, \dots$ such that

(a) $\{S_n\}_{n \in \mathbb{N}} \overset{d}{=} \{W_{\pi_n}\}_{n \in \mathbb{N}}$

(b) $\{\pi_1, \pi_2 - \pi_1, \pi_3 - \pi_2, \dots\}$ are mutually independent.

(c) $\forall n \in \mathbb{N}$, $\mathbb{E}(\pi_n - \pi_{n-1}) = \mathbb{E}X_n^2$, $\mathbb{E}(\pi_n - \pi_{n-1})^2 \leq \mathbb{E}X_n^4$. (with $\pi_0 = 0$).

8.5 Law of Iterated Logarithm

Lemma 8.5.1 (Mill's Ratio). Let $X \sim N(0, 1)$. For all $x \geq 2$, it holds that

$$\frac{1}{2\pi x} e^{-\frac{1}{2}x^2} \leq \mathbb{P}(X > x) \leq e^{-\frac{1}{2}x^2}.$$

Theorem 8.5.1 (Law of Iterated Logarithm). Let W be a standard Brownian motion. We have

$$\limsup_{t\to\infty} \frac{W_t}{\sqrt{2t\log\log t}} = 1 \text{ and } \liminf_{t\to\infty} \frac{W_t}{\sqrt{2t\log\log t}} = -1 \text{ a.s.}$$

Moreover, we have

$$\limsup_{t\to 0} \frac{W_t}{\sqrt{2t\log\log \frac{1}{t}}} = 1 \text{ and } \liminf_{t\to 0} \frac{W_t}{\sqrt{2t\log\log \frac{1}{t}}} = -1 \text{ a.s.}$$

An equivalent way to interpret the lim sup part of the theorem is that, almost surely, two things occur: (1) there exists $\eta(t) \to 0$ as $t \to \infty$ such that $W_t \le (1+\eta(t))\sqrt{2t\log\log t}$ and (2) there exists points $t_1, t_2, \ldots$ such that $W_{t_i} \ge \sqrt{2t_i \log\log t_i}$.

Proof.

To obtain the latter claim, observe that $\widetilde{W}_t := tW_{\frac{1}{t}}$ is also a standard Brownian motion. To see this, note that for $t \ge s \ge 0$,

$$\begin{aligned}
\mathbb{E}(\widetilde{W}_t - \widetilde{W}_s)^2 &= \mathbb{E}(tW_{\frac{1}{t}} - sW_{\frac{1}{t}} + sW_{\frac{1}{t}} - sW_{\frac{1}{s}})^2 \\
&= (t-s)^2\mathbb{E}(W_{\frac{1}{t}})^2 + s^2\mathbb{E}(W_{\frac{1}{t}} - W_{\frac{1}{s}})^2 \\
&= \frac{(t-s)^2}{t} + s^2(\frac{1}{s} - \frac{1}{t}) = t - s.
\end{aligned}$$

Hence,

$$\limsup_{t\to\infty} \frac{tW_{\frac{1}{t}}}{\sqrt{2t\log\log t}} = \limsup_{s\to 0} \frac{W_s}{\sqrt{2s\log\log\frac{1}{s}}} \text{ where we set } s := \frac{1}{t}.$$

We thus need only prove the first claim.

Since $W \overset{d}{=} -W$, we need only show that $\limsup_{t\to\infty} \frac{W_t}{\sqrt{2t\log\log t}} = 1$ a.s.

Define $\psi(t) = \sqrt{2t\log\log t}$. Let us first prove $\lim_{n\to\infty}\sup_{t\ge n} \frac{W_t}{\psi(t)} \le 1$ a.s. To that end, fix $\varepsilon > 0$ and $q > 1$ arbitrarily. We claim that $\mathbb{P}(\lim_{n\to\infty}\sup_{t\ge n}\frac{W_t}{\psi(t)} > (1+\varepsilon)q) = 0$. To show this, define, for $n \in \mathbb{N}$,

$$A_n^{(\varepsilon,q)} := \left\{\sup_{0\le t\le q^n} W_t \ge (1+\varepsilon)\psi(q^n)\right\} \text{ and } A^{(\varepsilon,q)} := \cap_{m=1}^\infty \cup_{n=m}^\infty A_n^{(\varepsilon.q)}.$$

Then,

$$\begin{aligned}
\mathbb{P}(A_n^{(\varepsilon,q)}) &= 2\mathbb{P}\left(W_{q^n} \ge (1+\varepsilon)\psi(q^n)\right) && \text{(by Reflection principle)} \\
&= 2\mathbb{P}\left(\frac{1}{\sqrt{q^n}}W_{q^n} \ge (1+\varepsilon)\sqrt{2\log(n\log q)}\right) \\
&\le 2e^{-(1+\varepsilon)^2\log(n\log q)} = 2(n\log q)^{-(1+\varepsilon)^2},
\end{aligned}$$

so that $\sum_{n=1}^\infty \mathbb{P}(A_n^{(\varepsilon,q)}) < \infty$. Therefore,

$$\mathbb{P}(A^{(\varepsilon,q)}) \le \lim_{n\to\infty}\sum_{m=n}^\infty \mathbb{P}(A_m^{(\varepsilon,q)}) = 0.$$

Let $\omega \notin A^{(\varepsilon,q)}$, then there exists $n_0 \in \mathbb{N}$ such that for all $n \ge n_0$,

$$\sup_{t\in[0,q^n]} W_t \le (1+\epsilon)\psi(q^n).$$

Therefore, for any $t \in [q^{n-1}, q^n]$ for any n large enough such that $n \geq n_0 + 1$ and $\frac{\log n}{\log(n-1)} \leq q$, we have that

$$\begin{aligned} \frac{W_t(\omega)}{\psi(t)} = \frac{W_t(\omega)}{\psi(q^n)} \frac{\psi(q^n)}{\psi(t)} &\leq (1+\varepsilon) \frac{\psi(q^n)}{\psi(q^{n-1})} \\ &= (1+\varepsilon) \left(q \frac{\log n + \log\log q}{\log(n-1) + \log\log q} \right)^{\frac{1}{2}} \leq (1+\varepsilon) q. \end{aligned}$$

This implies that

$$\limsup_{t\to\infty} \frac{W_t(\omega)}{\psi(t)} \leq (1+\varepsilon) q.$$

Since $\omega \notin A^{(\epsilon,q)}$ was arbitrary, we further conclude that

$$\left\{ \limsup_{t\to\infty} \frac{W_t}{\psi(t)} > (1+\varepsilon) q \right\} \subseteq A^{(\varepsilon,q)}.$$

Noting that $\mathbb{P}(\cup_{n=1}^\infty \cup_{m=1}^\infty A^{(\frac{1}{n}, 1+\frac{1}{m})}) = 0$ finishes proof of the first part.

For the second part, fix $q \geq 2$. We claim that $\mathbb{P}(\lim_{n\to\infty} \sup_{t\geq n} \frac{W_t}{\psi(t)} \geq 1 - \frac{3}{\sqrt{q}}) = 1$. To that end, define $B := \{\liminf_{n\to\infty} \frac{-W_{q^n}}{\psi(q^n)} \geq 1\}$ and note $\mathbb{P}(B) = 1$ by the first part.

Define, for $n \in \mathbb{N}$, $D_n := \{W_{q^n} - W_{q^{n-1}} \geq \psi(q^n - q^{n-1})\}$, and $D := \cap_{m=1}^\infty \cup_{n=m}^\infty D_n$. Since

$$\begin{aligned} \mathbb{P}(D_n) = \mathbb{P}\Bigg(\frac{W_{q^n} - W_{q^{n-1}}}{\sqrt{q^n - q^{n-1}}} &\geq \underbrace{\frac{\psi(q^n - q^{n-1})}{\sqrt{q^n - q^{n-1}}}}_{\leq \sqrt{2\log\{n\log q\}}} \Bigg) \\ &\geq \mathbb{P}\big(N(0,1) \geq \sqrt{2\log(n\log q)}\big) \\ &\geq \frac{1}{2\pi(2\log\{n\log q\})^{\frac{1}{2}}} \underbrace{e^{-\log\{n\log q\}}}_{\{n\log q\}^{-1}}. \end{aligned}$$

Thus, $\sum_{n=1}^\infty \mathbb{P}(D_n) = \infty$. Since $\{D_1, D_2, \dots\}$ are mutually independent,

$$\begin{aligned} \mathbb{P}(D^c) = \mathbb{P}(\cup_{m=1}^\infty \cap_{n=m}^\infty D_m^c) &= \lim_{m\to\infty} \prod_{n=m}^\infty (1 - \mathbb{P}(D_n)) \\ &\leq \lim_{n\to\infty} \prod_{n=m}^\infty e^{-\mathbb{P}(D_n)} = \lim_{m\to\infty} e^{-\sum_{n=m}^\infty \mathbb{P}(D_n)} = 0 \\ \implies \mathbb{P}(D) = 1. \end{aligned}$$

Now, let $\omega \in D \cap B$, then for infinitely many $n \in \mathbb{N}$, we have that

$$-W_{q^n} \geq \psi(q^n) \qquad W_{q^n} - W_{q^{n-1}} \geq \psi(q^n - q^{n-1}).$$

Therefore, for any $n_0 \in \mathbb{N}$, there exists $n \geq n_0$ such that

$$\begin{aligned} \frac{W_{q^n}(\omega)}{\psi(q^n)} &\geq \frac{W_{q^n}(\omega) - W_{q^{n-1}}(\omega) - (-W_{q^{n-1}}(\omega))}{\psi(q^n)} \\ &\geq \frac{\psi(q^n - q^{n-1})}{\psi(q^n)} - \frac{\psi(q^{n-1})}{\psi(q^n)} \\ &= \left(\frac{(q-1)\log\{(n-1)\log q + \log(q-1)\}}{q\log\{n\log q\}} \right)^{\frac{1}{2}} - \left(\frac{\log\{(n-1)\log q\}}{q\log\{n\log q\}} \right)^{\frac{1}{2}} \\ &\geq \left(\frac{q-1}{q} \left(1 - \frac{1}{q}\right) \right)^{1/2} - (1/q)^{1/2} \qquad \text{(by choosing } n \text{ large enough)} \\ &\geq 1 - \frac{3}{\sqrt{q}}, \end{aligned}$$

where the last inequality follows by the sub-additivity of the square root function. Hence, we have that $\limsup_{n\to\infty} \frac{W_{q^n}(\omega)}{\psi(q^n)} \geq 1 - \frac{s}{\sqrt{q}}$. Therefore,

$$A^{(q)} := \left\{ \limsup_{n\to\infty} \frac{W_{q^n}}{\psi(q^n)} \geq 1 - \frac{3}{\sqrt{q}} \right\} \supseteq D \cap B \implies \mathbb{P}(A^{(q)}) = 1.$$

Hence, $\mathbb{P}\{\limsup_{t\to\infty} \frac{W_t}{\psi(t)} \geq 1\} = \mathbb{P}(\cap_{q=2}^{\infty} A^{(q)}) = 1.$ □

Lemma 8.5.2. Let $\tau_1, \tau_2, \tau_3, \cdots : \Omega \to [0, \infty]$ be random variable such that $\tau_n \to \infty$ and $\frac{\tau_n}{n} = 1$ a.s. Then,

$$\limsup_{n\to\infty} \frac{W_{\tau_n}}{\sqrt{2n \log\log n}} = 1 \text{ a.s.}$$

Proof.
See note by Choi.

□

The following is a direct corollary of the above lemma and Skorokhod embedding.

Theorem 8.5.2 (Hartman-Wintner L.I.L.). For $n \in \mathbb{N}$, let $S_n : \Omega \to \mathbb{R}$ be random variable such that $\{S_1, S_2 - S_1, S_3 - S_2, \dots\}$ are independent and $\mathbb{E}(S_n - S_{n-1}) = 0$ and $\mathbb{E}(S_n - S_{n-1})^2 = 1$. We have

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = 1.$$

Theorem 8.5.3 (Levy's Arcsine Law). Let $Z := \sin^2(2\pi U)$ for $U \sim \text{Unif}[0,1]$ so that $\mathbb{P}(Z \leq t) = \frac{2}{\pi} \arcsin \sqrt{t}$ for $t \in [0,1]$, with density $z \mapsto \frac{1}{\pi\sqrt{z(1-z)}}$. Let W be standard B.M., we have

$$\text{Leb}(t \in [0,1] : W_t > 0) \overset{d}{=} \inf \left\{ t \in [0,1] : W_t = \sup_{s\in[0,1]} W_s \right\} \overset{d}{=} \sup\{t \in [0,1] : W_t = 0\} \overset{d}{=} Z.$$

Let W^0 be a Brownian bridge on $[0,1]$. Then

$$\text{Leb}(t \in [0,1] : W_t^0 > 0) \overset{d}{=} \inf \left\{ t \in [0,1] : W_t^0 = \sup_{s\in[0,1]} W_s \right\} \overset{d}{=} U.$$

Proof.
See Kallenberg Thm 11.16. □

Theorem 8.5.4 (Erdö and Kac Arcsine Law). For $n \in \mathbb{N}$, let $S_n : \Omega \to \mathbb{R}$ be random variables such that $\{S_1, S_2 - S_1, S_3 - S_2, \dots\}$ are independent and $\mathbb{E}(S_n - S_{n-1}) = 0$ and $\mathbb{E}(S_n - S_{n-1})^2 = 1$. Let

$$\tau_n^1 := \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{S_k > 0\}}$$
$$\tau_n^2 := \frac{1}{n} \min\{k \geq 0 : S_k = \max_{j\in[n]} S_j\}$$
$$\tau_n^3 := \frac{1}{n} \max\{k \geq 0 : S_k S_n \leq 0\}.$$

We have that $\tau_n^1, \tau_n^2, \tau_n^3 \xrightarrow{d} Z$ where $Z := (\sin 2\pi U)^2$ for $U \sim \text{Unif}[0,1]$.

Proof.
See Kallenberg Thm 12.11. □

8.6 Continuous Martingales

Definition 8.6.1. Let $\tau \subseteq [0, \infty)$ and let $\{\mathscr{F}_t\}_{t \in T}$ be a filtration. Let $X : \Omega \to \mathbb{R}^T$ be an adapted process such that $\mathbb{E}|X_t| < \infty$, $\forall t \in T$. We say

(1) X is a martingale if $\mathbb{E}[X_t|\mathscr{F}_s] = X_s$, $\forall s \leq t \in T$.

(2) X is a sub-martingale if $\mathbb{E}[X_t|\mathscr{F}_s] \geq X_s$, $\forall s \leq t \in T$.

(3) X is a super-martingale if $-X$ is a sub-martingale.

Note that if $T = \mathbb{N}$, this reverts to discrete time martingale.

Remark 8.6.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let $X : \Omega \to \mathbb{R}$ be $\mathscr{F}/\mathscr{B}(\mathbb{R})$-measurable. Let $\mathcal{G} \subseteq \mathscr{F}$ be a σ-field. Recall that

$$\mathbb{E}[X|\mathcal{G}] : \Omega \to \mathbb{R}$$

is the $\mathbb{P}$-a.e. unique $\mathcal{G}/\mathscr{B}(\mathbb{R})$-measurable such that $\forall B \in \mathcal{G}$, $\int_B \mathbb{E}[X|\mathcal{G}] \mathrm{d}\mathbb{P} = \int_B X \mathrm{d}\mathbb{P}$.

If $\mathbb{E}|X|^2 < \infty$, then $\mathbb{E}(X - \mathbb{E}[X|\mathcal{G}])^2 \leq \mathbb{E}(X - Z)^2$ for all $Z : \Omega \to \mathbb{R}$ that is $\mathcal{G}/\mathscr{B}(\mathbb{R})$-measurable.

If $Y : \Omega \to \mathbb{R}$ is random variable, then $\mathbb{E}[X|Y] := \mathbb{E}[X|\sigma(Y)]$ where $\sigma(Y) := \{Y^{-1}(\widetilde{B}) : \widetilde{B} \in \mathscr{B}(\mathbb{R})\}$.

Note that, if $\phi : \mathbb{R} \to \mathbb{R}$ is convex (and hence $\mathscr{B}(\mathbb{R})/\mathscr{B}(\mathbb{R})$-measurable), then $\forall \omega \in \Omega$,

$$\mathbb{E}[\phi(X)|\mathcal{G}](\omega) - \phi(\mathbb{E}[X|\mathcal{G}](\omega)) \geq 0 \text{ for } \mathbb{P} - a.e., \omega \in \Omega.$$

In particular, $\mathbb{E}[|X|\big|\mathcal{G}] \geq |\mathbb{E}[X|\mathcal{G}]|$.

Lemma 8.6.1. If $\{X_t, \mathscr{F}_t\}_{t \in T}$ is a martingale and $\phi : \mathbb{R} \to \mathbb{R}$ is convex, then $\{\phi(X_t), \mathscr{F}_t\}_{t \in T}$ is a sub-martingale.

Proof.
Apply the conclusion in the previous remark to obtain, for $s \leq t \in T$,

$$\mathbb{E}[\phi(X_t)|\mathscr{F}_s] \geq \phi(\mathbb{E}[X_t|\mathscr{F}_s]) = \phi(X_s)$$

where the last equality holds since $\{X_t\}$ is a martingale. □

Remark 8.6.2. Recall, when $T = \mathbb{N}$, the martingale convergence theorems: let $\{X_n\}_{n \in N}$ be a sub-martingale. If $\sup_{n \in \mathbb{N}} \mathbb{E}(X_n)_+ < \infty$, then there exists random variable $X_\infty : \Omega \to \mathbb{R}$, $\mathbb{E}|X_\infty| < \infty$, and $X_n \to X_\infty$ a.s.

Moreover, recall that $\{X_n\}$ is uniformly integrable (U.I.) if

$$\lim_{K \to \infty} \sup_{n \in \mathbb{N}} \int_{|X_n| > K} |X_n| \mathrm{d}\mathbb{P} = 0.$$

Uniform integrable (U.I.) is an important condition. Let $Z_1, Z_2, \ldots, Z : \Omega \to \mathbb{R}$ be random variable such that $Z_n \to Z$ a.s. Then, $\mathbb{E}|Z_n - Z| \to 0$ if and only if $\{Z_n\}$ is uniformly integrable. Hence, if sub-martingale $X_n \to X_\infty$ a.s., then $\mathbb{E}|X_n - X_\infty| \to 0$ if and only if $\{X_n\}$ is U.I.

Moreover, if $\{X_n\}$ is a martingale and $\{X_n\}$ is U.I., then $X_n = \mathbb{E}[X_\infty|\mathscr{F}_n]$. Conversely, if we take some random variable Y and define $X_n = \mathbb{E}[Y \mid \mathscr{F}_n]$, then $\{X_n\}$ is U.I. martingale.

Lemma 8.6.2. If a set of random variable $\{X_\gamma\}_{\gamma\in\Gamma}$ is U.I., then $\sup_{\gamma\in\Gamma}\mathbb{E}|X_\gamma|<\infty$. Let $\phi:[0,\infty)\to[0,\infty)$ be a measurable function such that $\lim_{x\to\infty}\frac{\phi(x)}{x}=\infty$. If $\sup_{\gamma\in\Gamma}\mathbb{E}\phi(|X_\gamma|)<\infty$, then $\{X_\gamma\}$ is U.I.

Proof.

For the first claim, $\exists K_0\in\mathbb{N}$ such that

$$\sup_{\gamma\in\Gamma}\int_{|X_\gamma|>K_0}|X_\gamma|\mathrm{d}\mathbb{P}\le 1.$$

Thus, for any $\gamma\in\Gamma$, we have

$$\mathbb{E}|X_\gamma|=\int_{|X_\gamma|\le K_0}|X_\gamma|\mathrm{d}\mathbb{P}+\int_{|X_\gamma|>K_0}|X_\gamma|\mathrm{d}\mathbb{P}\le K_0+1<\infty.$$

Fort the second claim, fix $\varepsilon>0$. There exists $K\in\mathbb{N}$ s.t. $\frac{x}{\phi(x)}\le\frac{\varepsilon}{\{\sup_{\gamma\in\Gamma}\mathbb{E}\phi(|X_\gamma|)\}}$, $\forall x>K$. Then, for any $\gamma\in\Gamma$,

$$\int_{|X_\gamma|>K}|X_\gamma|\mathrm{d}\mathbb{P}=\int_{|X_\gamma|>K}\phi(|X_\gamma|)\cdot\frac{|X_\gamma|}{\phi(|X_\gamma|)}\mathrm{d}\mathbb{P}\le\frac{\mathbb{E}\phi(|X_\gamma|)}{\sup_{\gamma\in\Gamma}\phi(|X_\gamma|)}\varepsilon\le\varepsilon.$$

$\square$

Example 8.6.1. Let $\widetilde{Z}$ be a random variable satisfying $\mathbb{E}|\widetilde{Z}|<\infty$, then one may always find a convex $\phi:[0,\infty)\to[0,\infty)$ such that $\mathbb{E}\phi(|\widetilde{Z}|)<\infty$ and $\lim_{x\to\infty}\frac{\phi(x)}{x}=\infty$. Thus, if $\{Z_n\}_{n\in\mathbb{N}}$ is a collection of random variables such that $|Z_n|\le|\widetilde{Z}|$ a.s. $\forall n\in\mathbb{N}$, then $\{Z_n\}_{n\in\mathbb{N}}$ is U.I.

To construct such a ϕ, we first note that

$$\int_\Omega\phi(|\tilde{Z}|)d\mathbb{P}=\int_\Omega\int_0^\infty\phi'(t)\mathbb{1}\{|\tilde{Z}|\ge t\}\,dt\,d\mathbb{P}=\int_0^\infty\phi'(t)\mathbb{P}(|\tilde{Z}|\ge t)dt.$$

Thus, the problem reduces to the following: given a function $H:[0,\infty)\to[0,\infty)$ such that $A:=\int_0^\infty H(t)dt<\infty$, construct an increasing function $g:[0,\infty)\to[0,\infty)$ such that $g(t)\to\infty$ as $t\to\infty$ and $\int_0^\infty g(t)H(t)dt<\infty$. To this end, fix $\alpha\in(0,1)$ and define $g(t)=\left\{\int_t^\infty H(t)dt\right\}^{-\alpha}$. Since $H\ge0$, the function g is increasing. Since $\int_0^\infty H(t)dt<\infty$, we have that $\lim_{t\to\infty}g(t)=\infty$. Moreover, defining, for any $k\in\mathbb{N}$, $t_k=\inf\{t\ge0\,:\,\int_t^\infty H(t)dt\le A2^{-k}\}$, we have

$$\begin{aligned}\int_0^\infty g(t)H(t)dt&=\sum_{k=0}^\infty\int_{t_k}^{t_{k+1}}g(t)H(t)\,dt\\&\le\sum_{k=0}^\infty g(t_{k+1})\int_{t_k}^{t_{k+1}}H(t)dt\\&\le A^{1+\alpha}\sum_{k=0}^\infty 2^{\alpha(k+1)-k}<\infty.\end{aligned}$$

Remark 8.6.3. Recall the discrete time optional sampling theorem. Let $\{X_n,\mathscr{F}_n\}_{n\in\mathbb{N}}$ be a sub-martingale and let $\tau,\pi:\Omega\to\mathbb{N}$ be stopping times such that $\tau\le\pi$. If π is a.s. bounded, i.e., there exists $N_0\in\mathbb{N}$ such that $\mathbb{P}(\pi\le N_0)=1$, then

$$\mathbb{E}\left[X_\pi|\mathscr{F}_\tau\right]\ge X_\tau.$$

Note X_τ is always $\mathscr{F}_\tau/\mathscr{B}(\mathbb{R})$-measurable in discrete time.

As a consequence, for any stopping time τ,

$$\{X_{n\wedge\tau},\mathscr{F}_n\}\text{ is a sub-martingale.}\tag{$*$}$$

Equivalent statement hold for Martingales.

The statement $(*)$ may be directly proved. Note that

$$X_{n\wedge\tau} = X_1 + \sum_{k=2}^{n-1} \mathbb{1}_{\{\tau>k-1\}}(X_k - X_{k-1}).$$

To see this, if $\omega \in \Omega$ is such that $\tau(\omega) = m \leq n$, then RHS becomes

$$(X_m(\omega) - X_{m-1}(\omega)) + (X_{m-1}(\omega) - X_{m-2}(\omega)) + \cdots = X_m(\omega).$$

This is an example of *martingale transform*: let $\{(A\cdot X)_n\}$ be a process where $(A\cdot X)_m := X_0 + \sum_{k=1}^{n} A_k(X_k - X_{k-1})$ and where $\{A_k\}$ is a *predictable process* in that A_k is $\mathscr{F}_{k-1}/\mathscr{B}(\mathbb{R})$–measurable. We may show that if X is a sub-martingale and if $\{A_k\}$ is non-negative and bounded, then $(A\cdot X)$ is a sub-martingale.

To prove $(*)$, first note that if $\omega \in \Omega$ is such that $\tau(\omega) > n$, then RHS is $X_n(\omega)$. Now, because $\mathbb{1}_{\{\tau\leq n-1\}}$ is $\mathscr{F}_{n-1}$-measurable and hence $\mathbb{1}_{\{\tau>n-1\}}$ is also $\mathscr{F}_{n-1}$-measurable, we have

$$\begin{aligned}\mathbb{E}[X_{n\wedge\tau}|\mathscr{F}_{n-1}] &= X_0 + \sum_{k=0}^{n-2} \mathbb{1}_{\{\tau>k\}}\,(X_{k+1} - X_k) + \mathbb{1}_{\{\tau>n-1\}}(\mathbb{E}[X_n|\mathscr{F}_{n-1}] - X_{n-1})\\ &\geq X_{(n-1)\wedge\tau} \quad \text{ since } \mathbb{E}[X_n|\mathscr{F}_{n-1}] \geq X_{n-1}.\end{aligned}$$

Theorem 8.6.1 (Doob Stopping Theorem)**.** Let $\{X_t, \mathscr{F}_t\}_{t\geq 0}$ be a continuous martingale. Let $\tau : \Omega \to [0,\infty]$ be a stopping time, let $\widetilde{X}_t := X_{t\wedge\tau}$ (Stopped process). Then $\{\widetilde{X}_t, F_t\}_{t\geq 0}$ is also a continuous martingale. Note $X_{t\wedge\tau}$ is $\mathscr{F}_{t\wedge\tau}/\mathscr{B}(\mathbb{R})$-measurable and hence $\mathscr{F}_t/\mathscr{B}(\mathbb{R})$-measurable.

Proof.

Step 1: first we verify that $\mathbb{E}|X_{t\wedge\tau}| < \infty$, $\forall t \in [0,\infty)$.

Fix $t \in [0,\infty)$ and define $S_n := \{\frac{k}{2^n} : k = 0, 1, 2, \ldots\} \cup \{t\}$. Define $\tau_n : \Omega \to S_n$ as $\tau_n(\omega) := \inf\{u \in S_n : u \geq \tau(\omega)\}$, $\forall \omega$ such that $\tau(\omega) < \infty$. Define $\tau_n(\omega) = \infty$ else. Note then that $\forall \omega \in \Omega$, $\lim_{n\to\infty} \tau_n(\omega) = \tau(\omega)$ and thus, by continuity,

$$\lim_{n\to\infty} X_{t\wedge\tau_n(\omega)}(\omega) = X_{t\wedge\tau(\omega)}(\omega).$$

Note also $\{X_u, \mathscr{F}_u\}_{u\in S_n}$ is a discrete martingale and τ_n is a stopping time (Proposition 8.4.2) taking value in S_n. We have by Remark 8.6.3 (with $\pi = t$ and $\tau = t\wedge\tau_n$) that $\mathbb{E}[X_t|\mathscr{F}_{t\wedge\tau_n}] = X_{t\wedge\tau_n}$ and hence,

$$\mathbb{E}[|X_t|\big|\mathscr{F}_{t\wedge\tau_n}] \geq |X_{t\wedge\tau_n}| \implies \infty > \mathbb{E}|X_t| \geq \mathbb{E}|X_{t\wedge\tau_n}|.$$

By Fatou's lemma, $\mathbb{E}|X_{t\wedge\tau}| \leq \liminf_{n\to\infty} \mathbb{E}|X_{t\wedge\tau_n}| \leq \mathbb{E}|X_t| < \infty$.

Step2: Let $s \leq t \in [0,\infty)$, we claim that $\mathbb{E}[X_{t\wedge\tau}|\mathscr{F}_s] = X_{s\wedge\tau}$. Redefine

$$S_n := \left\{\frac{k}{2^n} : k = 0, 1, 2, \ldots\right\} \cup \{s, t\}$$

and define τ_n as before. Then

$$\mathbb{E}[X_{t\wedge\tau_n}|\mathscr{F}_s] = X_{s\wedge\tau_n}. \tag{8.2}$$

On the RHS of (8.2), we have $\forall \omega \in \Omega$,

$$\lim_{n\to\infty} X_{s\wedge\tau_n(\omega)}(\omega) = X_{s\wedge\tau(\omega)}(\omega).$$

To analyze the LHS of (8.2), we first show that $\{X_{t\wedge\tau_n}\}_{n\in\mathbb{N}}$ is U.I. To see this, note by Example 8.6.1 that there exists $\phi : [0,\infty) \to [0,\infty)$ convex and satisfying $\lim_{x\to\infty} \frac{\phi(x)}{x} = \infty$ such that $\mathbb{E}\phi(|X_t|) < \infty$. Fix $n \in \mathbb{N}$ arbitrarily. Since t and $t\wedge\tau_n$ are stopping times $\Omega \to S_n$,

$$\mathbb{E}[\phi(|X_t|)\big|\mathscr{F}_{t\wedge\tau_n}] \geq \phi(|X_{t\wedge\tau_n}|) \implies \mathbb{E}\phi(|X_t|) \geq \mathbb{E}\phi(|X_{t\wedge\tau_n}|),$$

which implies $\{X_{t\wedge\tau_n}\}_{n\in\mathbb{N}}$ is U.I. since n is arbitrary.

Since $X_{t\wedge\tau_n} \overset{a.s.}{\to} X_{t\wedge\tau}$, by Theorem 5.2.3, we have

$$\begin{aligned}\mathbb{E}|X_{t\wedge\tau} - X_{t\wedge\tau_n}| \to 0 &\implies \mathbb{E}|\mathbb{E}[X_{t\wedge\tau}|\mathscr{F}_s] - \mathbb{E}[X_{t\wedge\tau_n}|\mathscr{F}_s]| \to 0,\\ &\implies \mathbb{E}[X_{t\wedge\tau_n}|\mathscr{F}_s] \overset{\text{in prob.}}{\longrightarrow} \mathbb{E}[X_{t\wedge\tau}|\mathscr{F}_s] \text{ by Markov inequality,}\end{aligned}$$

which further implies there exists a subsequence $\{n_1, n_2, \dots\}$ such that

$$\mathbb{E}[X_{t\wedge\tau_{n_k}}|\mathscr{F}_s] \xrightarrow{a.s.} \mathbb{E}[X_{t\wedge\tau}|\mathscr{F}_s].$$

Thus, (8.2) implies that $\mathbb{E}[X_{t\wedge\tau}|\mathscr{F}_s] = X_{s\wedge\tau}$. □

Example 8.6.2. Let W be a standard B.M. with induced filtration $\{\mathscr{F}_t\}$. We have that $\{W_t^2 - t\}_{t\geq 0}$ is a martingale. To see this, note that for $s \leq t \in [0,\infty)$,

$$\begin{aligned}\mathbb{E}[W_t^2 - t|\mathscr{F}_s] &= \mathbb{E}[(W_t - W_s + W_s)^2 - t|\mathscr{F}_s]\\ &= \mathbb{E}[(W_t - W_s)^2 + 2(W_t - W_s)W_s + W_s^2 - t|\mathscr{F}_s]\\ &= t - s + 0 - \mathbb{E}[W_s^2|\mathscr{F}_s] - t = W_s^2 - s.\end{aligned}$$

Define $\tau := \inf\{t \geq 0 : |W_t| \geq 1\}$, then, for all $t \geq 0$, $\mathbb{E}[W_{t\wedge\tau}^2 - t\wedge\tau] = 0$ by Theorem 8.6.1. Since $\lim_{n\to\infty} n\wedge\tau = \tau$ everywhere, we have $\lim_{n\to\infty}\mathbb{E}n\wedge\tau = \mathbb{E}\tau$ by monotone convergence theorem. Similarly, $\lim_{n\to\infty} W_{n\wedge\tau}^2 = W_\tau^2 = 1$ by continuity and since $W_{n\wedge\tau}^2 \leq 1$ by definition of τ, we have $\lim_{n\to\infty}\mathbb{E}W_{n\wedge\tau}^2 = \mathbb{E}W_\tau^2 = 1$ by dominated convergenc theorem. We may conclude then that $\mathbb{E}\tau = 1$.

Theorem 8.6.2 (Doob's Maximal Inequality). Let $\{X_t, \mathscr{F}_t\}_{t\geq 0}$ be a continuous sub-martingale. For any $t \geq 0$, $\lambda > 0$,

$$\mathbb{P}\left(\sup_{s\in[0,t]} X_s > \lambda\right) \leq \frac{1}{\lambda}\int_{\{\sup_{s\in[0,t]} X_s > \lambda\}} X_t \mathrm{d}\mathbb{P} \leq \frac{1}{\lambda}\mathbb{E}|X_t|.$$

Proof.

First consider finite discrete time sub-martingale $\{X_u, \mathscr{F}_u\}_{u=1}^m$. Define the event

$$A := \{\max_{u\in[m]} X_u > \lambda\}.$$

Define, for all $\omega \in \Omega$,

$$\tau(\omega) := \begin{cases}\min\{u \in [m] : X_u(\omega) > \lambda\} & \text{if } \omega \in A\\ m & \text{else}\end{cases}$$

Note that for any $u \in [m-1]$, we have $\{\tau \leq u\} = \cup_{k=1}^u \{X_k > \lambda\} \in \mathscr{F}_u$, which implies τ is a stopping time.

Then, we have that

$$\begin{aligned}\mathbb{P}(\max_{k\in[m]} X_k > \lambda) &\leq \frac{1}{\lambda}\mathbb{E}[X_\tau \mathbb{1}_A] = \frac{1}{\lambda}\mathbb{E}X_\tau - \frac{1}{\lambda}\mathbb{E}[X_\tau \mathbb{1}_{A^c}]\\ &\overset{(a)}{\leq} \frac{1}{\lambda}\mathbb{E}X_m - \frac{1}{\lambda}\mathbb{E}[X_m \mathbb{1}_{A^c}] = \frac{1}{\lambda}\mathbb{E}[X_m \mathbb{1}_A],\end{aligned}$$

where inequality (a) follows by optional stopping theorem and also because $X_\tau = X_m$ on the event A^c.

For the continuous time setting, fix $t \in [0,\infty)$ and define $\mathbb{Q}_t = \mathbb{Q} \cap [0,t] \cup \{t\}$. Let $I_1 \subseteq I_2 \subseteq \cdots \subseteq \mathbb{Q}_t$ be finite sets such that $\cup_{k=1}^\infty I_k = \mathbb{Q}_t$. It holds that $\cup_{k=1}^\infty \{\max_{u\in I_k} X_u > \lambda\} = \{\max_{u\in\mathbb{Q}_t} X_u > \lambda\}$. Thus, $\mathbb{P}(\sup_{u\in\mathbb{Q}_t} X_u > \lambda) = \lim_{k\to\infty}\mathbb{P}(\max_{u\in I_k} X_u > \lambda) \leq \lim_{k\to\infty}\frac{1}{\lambda}\int_{\{\max_{u\in I_k} X_u > \lambda\}} X_t \mathrm{d}\mathbb{P}$. And, since $\mathbb{E}|X_t| < \infty$, the set function $B \mapsto \int_B X_t \mathrm{d}\mathbb{P}$ on $\mathscr{F}_t$ is a finite signed measure. Therefore,

$$\lim_{k\to\infty}\frac{1}{\lambda}\int_{\{\max_{u\in I_k} X_u > \lambda\}} X_t \mathrm{d}\mathbb{P} = \frac{1}{\lambda}\int_{\{\sup_{u\in\mathbb{Q}_t} X_u > \lambda\}} X_t \mathrm{d}\mathbb{P}.$$

Since X is continuous, we have $\sup_{u\in\mathbb{Q}_t} X_u = \sup_{u\in[0,t]} X_u$ and the desired conclusion follows. □

Theorem 8.6.3 (Doob's maximal norm inequality). Let $\{X_t, \mathscr{F}_t\}_{t\geq 0}$ be a continuous martingale. If $\mathbb{E}X_t^2 < \infty$, $\forall t \geq 0$, we have

$$\mathbb{E}X_t^2 \leq \mathbb{E}\left(\sup_{s\in[0,t]} X_s\right)^2 \leq 4\mathbb{E}X_t^2.$$

Proof.

Let $Y_t := \sup_{s\leq t} |X_s|$ for $t \in [0,\infty)$. Since $\{|X_t|\}$ is a sub-martingale, by Theorem 8.6.1, we have $\forall\lambda > 0$,

$$\mathbb{P}(Y_t > \lambda) \leq \frac{1}{\lambda}\int_{Y_t>\lambda} |X_t| \mathrm{d}\mathbb{P}.$$

Since we do not know whether $\mathbb{E}Y_t^2 < \infty$, define $Y_{t,n} := n \wedge Y_t$ for $n \in \mathbb{N}$. Note that $\{Y_{t,n} > \lambda\} = \{Y_t > \lambda\}$ if $\lambda < n$ and $\{Y_{n,t} > \lambda\} = \emptyset$ else. Thus,

$$\mathbb{P}(Y_{n,t} > \lambda) \leq \frac{1}{\lambda}\int_{Y_{n,t}>\lambda} |X_t| \mathrm{d}\mathbb{P}.$$

From the fact that $s^2 = \int_0^s 2\lambda \mathrm{d}\lambda$ for $s \geq 0$, we have

$$\begin{aligned}
\mathbb{E}Y_{t,n}^2 &= \int_0^\infty s^2 \mathrm{d}\mathbb{P}^{(Y_{t,n})}(s) = \int_0^\infty \int_0^s 2\lambda \mathrm{d}\lambda \mathrm{d}\mathbb{P}^{(Y_{t,n})}(s) \\
&= \int_{\{s,\lambda:s\geq\lambda\}} 2\lambda \mathrm{d}\mathbb{P}^{(Y_t,s)}(s)\mathrm{d}\lambda \\
&= \int_0^\infty 2\lambda \int_\lambda^\infty \mathrm{d}\mathbb{P}^{(Y_{t,n})}(s)\mathrm{d}\lambda \\
&= \int_0^\infty 2\lambda \mathbb{P}(Y_{t,n} > \lambda)\mathrm{d}\lambda \\
&\leq 2\int_0^\infty \int_\Omega |X_t(\omega)| \mathbb{1}_{\{Y_{t,n}(\omega)>\lambda\}} \mathrm{d}\mathbb{P}(\omega)\mathrm{d}\lambda \\
&= 2\int_\Omega |X_t(\omega)| \int_0^\infty \mathbb{1}_{\{Y_{t,n}(\omega)>\lambda\}} \mathrm{d}\lambda \mathrm{d}\mathbb{P}(\omega) \\
&= 2\int_\Omega |X_t(\omega)| Y_{t,n}(\omega) \mathrm{d}\mathbb{P} \leq 2\sqrt{\mathbb{E}X_t^2}\sqrt{\mathbb{E}Y_{t,n}^2},
\end{aligned}$$

which implies $\sqrt{\mathbb{E}Y_{t,n}^2} \leq 2\sqrt{\mathbb{E}X_t^2}$. Since n is arbitrary, the desired result follows. $\square$

Definition 8.6.2. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Let $\{\mathscr{F}_t\}_{t\geq 0}$ be a filtration where $\mathscr{F}_t \subseteq \mathscr{F}$, $\forall t \geq 0$. Let $\mathscr{C} = \{A \in \mathscr{F} : \mathbb{P}(A) = 0\}$ and let $\mathcal{N} := \{B \subseteq A : A \in \mathscr{C}\}$. Define $\widetilde{\mathscr{F}_t} := \sigma(\mathscr{F}_t, \mathcal{N})$ and $\widetilde{\mathscr{F}} := \sigma(\mathscr{F}, \mathcal{N})$ and extend $\mathbb{P}$ to $\widetilde{\mathscr{F}}$ by defining $\mathbb{P}(A) = 0$, $\forall A \in \mathcal{N}$. Recall that this gives the completion of $\mathbb{P}$.

We call $\{\widetilde{\mathscr{F}_t}\}_{t\geq 0}$ the complete filtration. One may show that if $\{\mathscr{B}_t\}_{t\geq 0}$ is the filtration induced by standard Brownian motion, then $\{\widetilde{\mathscr{B}}_t\}_{t\geq 0}$ is also right-continuous. A filtration that is complete and right-continuous is said to satisfy the usual conditions.

A consequence is that if $X : \Omega \to \mathbb{R}^{[0,\infty)}$ is a process adapted to a complete filtration $\{\mathscr{F}_t\}_{t\geq 0}$ and $Y = X$ a.s., then Y is also adapted.

Theorem 8.6.4. Let $\{\mathscr{F}_t\}_{t\geq 0}$ be a complete filtration and, for $a \geq 0$, define $\mathcal{M}_a$ as the set of all continuous martingale $\{X_t\}_{t\in[0,a]}$ adapted w.r.t. $\{\mathscr{F}_t\}_{t\in[0,a]}$ such that $X_0 = 0$ and $\mathbb{E}X_a^2 < \infty$. It holds that $\mathcal{M}_a$ is a Hilbert space.

Remark 8.6.4. Technically, $\mathcal{M}_a$ is a set of equivalence classes where 2 processes X, Y are equivalent if they equal almost surely.

Proof.

For $X, Y \in \mathcal{M}_a$, define $\langle X, Y\rangle := \mathbb{E}X_aY_a$. It is easy to verify that $\mathcal{M}_a$ is a vector space and that $\langle\cdot,\cdot\rangle$ is bilinear.

Suppose $\mathbb{E}X_a^2 = 0$, then $\mathbb{E}(\sup_{t\le a} X_t)^2 = 0 \implies \sup_{t\le a} X_t^2 = 0$ a.s. $\implies X = 0$ a.s. since X is continuous. Hence, $\|X\| := \langle x, x\rangle^{\frac{1}{2}}$ is a norm.

Now we show that $\mathcal{M}_a$ is complete. Let $\{X^{(n)}\}_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathcal{M}_a$ so that

$$\lim_{n,m\to\infty} \mathbb{E}\{X_a^{(n)} - X_a^{(m)}\}^2 = 0.$$

By taking subsequence if necessary, we may assume that $\forall n \in \mathbb{N}$, $\mathbb{E}\{X_a^{(n)} - X_a^{(n+1)}\}^2 \le 2^{-3n}$.

Define $A^{(n)} = \{\sup_{t\le a} |X_t^{(n)} - X_t^{(n+1)}| > 2^{-n}\}$. Since $\{|X_t^{(n)} - X_t^{(n+1)}|^2\}_{t\in[0,a]}$ is a sub-martingale, we have by Theorem 8.6.2 that

$$\mathbb{P}(A^{(n)}) = \mathbb{P}\{\sup_{t\le a} |X_t^{(n)} - X_t^{(n+1)}|^2 > 2^{-2n}\} \le \frac{2^{-3n}}{2^{-2n}} = 2^{-n}.$$

Write $A = \cap_{n=1}^\infty \cup_{m=n}^\infty A^{(m)}$, then $\mathbb{P}(A) = 0$. Suppose $\omega \in A^c = \cup_{n=1}^\infty \cap_{m=n}^\infty {A^{(m)}}^c$, then $\exists n_\omega \in \mathbb{N}$ such that $\forall n \ge n_\omega$, $\sup_{t\le a} |X_t^{(n)}(\omega) - X_t^{(n+1)}(\omega)| \le 2^{-n}$. It implies $\forall n \ge n_\omega$, $\forall k \in \mathbb{N}$,

$$\sup_{t\le a} |X_t^{(n)}(\omega) - X_t^{(n+k)}(\omega)| \le \sum_{j=0}^{k-1} 2^{-(n+j)} \le 2^{-(n-1)}.$$

Thus, $\lim_{n,m\to\infty} \left\|X^{(n)}(\omega) - X^{(m)}(\omega)\right\|_\infty = 0$. Since $(C[0,a], \|\cdot\|_\infty)$ is complete, $\exists X(\omega) : [0,a] \to \mathbb{R}$ that is continuous and that $\lim_{n\to\infty} \left\|X^{(n)}(\omega) - X(\omega)\right\|_\infty = 0$. If $\omega \in A$, define $X(\omega) = 0$.

We claim that (a) X is adapted, (b) X is a martingale, and (c) $\mathbb{E}(X_a^{(n)} - X_a)^2 \to 0$.

To see (a), note that for any $t \in [0,a]$, $\forall \omega \in \Omega$,

$$X_t(\omega) = \lim_{n\to\infty} X_t^{(n)}(\omega)\mathbb{1}_{\{\omega\in A^c\}}$$

where $X_t^{(n)}$ is $\mathscr{F}_t/\mathscr{B}(\mathbb{R})$-measurable since $X^{(n)}$ is adapted and $\mathbb{1}_{\{\omega\in A^c\}}$ is also $\mathscr{F}_t/\mathscr{B}(\mathbb{R})$-measurable since $A \in \mathscr{N}$ and $\{\mathscr{F}_t\}$ is complete. Thus X is adapted.

For claim (b), fix $s < t \in [0,a]$. For any $n \in \mathbb{N}$, we have

$$\mathbb{E}[X_t^{(n)}|\mathscr{F}_s] = X_s^{(n)}. \tag{$\star$}$$

On the RHS of ($\star$), we have that $\lim_{n\to\infty} X_s^{(n)} = X_s$ a.s. For the LHS of ($\star$), we first show that $\{X_t^{(n)}\}_{n\in\mathbb{N}}$ is U.I. Since, for all $n \in \mathbb{N}$, $\{{X_t^{(n)}}^2\}_{t\in[0,a]}$ is a sub-martingale, $\mathbb{E}{X_t^{(n)}}^2 \le \mathbb{E}[\mathbb{E}[{X_a^{(n)}}^2|\mathscr{F}_s]] = \mathbb{E}{X_a^{(n)}}^2$. Thus,

$$\sup_{n\in\mathbb{N}} \mathbb{E}{X_t^{(n)}}^2 \le \sup_{n\in\mathbb{N}} \mathbb{E}{X_a^{(n)}}^2 < \infty$$

where the latter inequality follows because $\lim_{n,m\to\infty} \mathbb{E}(X_a^{(n)} - X_a^{(m)})^2 = 0$ and, since $L_2(\mathbb{P})$ is complete, there exists $Z \in L_2(\mathbb{P})$ such that

$$\lim_{n\to\infty} \mathbb{E}(X_a^{(n)} - Z)^2 = 0.$$

So $\lim_{n\to\infty} \mathbb{E}{X_a^{(n)}}^2 = \mathbb{E}Z^2$. Thus, $\{X_t^{(n)}\}$ is U.I. and by an argument very similar to that of Theorem 8.6.1, we have that $\mathbb{E}[X_t|\mathscr{F}_s] = X_s$.

For claim (c), note that since $\lim_{n\to\infty} \mathbb{E}(X_a^{(n)} - Z)^2 = 0$, we have that $X_a^{(n)} \to Z$ in probability and hence, there is a subsequence $\{n_1, n_2, \dots\}$ such that $X_a^{(n_k)} \to Z$ a.s. as $k \to \infty$. However, $X_a^{(n)} \to X_a$ a.s. by definition. So $X_a = Z$ a.s. and $\lim_{n\to\infty} \mathbb{E}(X_a^{(n)} - X_a)^2 = 0$. □

Chapter 9

Stochastic Integral

9.1 Stochastic Integral Definition

Remark 9.1.1. Let $f : [0,1] \to \mathbb{R}$ be a continuous function. For $n \in \mathbb{N}$, let $0 = t_1^{(n)} < t_2^{(n)} < \cdots < t_n^{(n)} = 1$ be a sequence of partitions of $[0,1]$ such that $\max_{i\in[n-1]} |t_{i+1}^{(n)} - t_i^{(n)}| \to 0$ as $n \to \infty$. We omit superscript index to mate the notation simpler. It hold that $\int_0^1 f(t)\mathrm{d}t = \lim_{n\to\infty} \sum_{i=1}^{n-1} f(t_i)(t_{i+1} - t_i)$ for any sequence of partitions.

Similarly, if $G : [0,1] \to \mathbb{R}$ is a continuous non-decreasing function, then G defines a measure by $\mathbb{P}^G((a,b]) = G(b) - G(a)$. Then, we have the Lebesgue-Stieljes integral

$$\int_0^1 f(t)\mathrm{d}G(t) = \int_{[0,1]} f(t)\mathrm{d}\mathbb{P}^G(t) = \lim_{n\to\infty} \sum_{i=1}^{n-1} f(t_i)(G(t_{i+1}) - G(t_i))$$

for any sequence of partitions. If $H : [0,1] \to \mathbb{R}$ is a continuous function of bounded total variation, we may write $H = G_1 - G_2$ for non-decreasing G_1, G_2 and define $\int_0^1 f(t)\mathrm{d}H(t)$ analogously.

Let W be standard Brownian motion on $[0,1]$. For a.e. $\omega \in \Omega$, $W(\omega)$ is continuous but has unbounded total variation and thus, the limit

$$\lim_{n\to\infty} \sum_{i=1}^{n-1} f(t_i)\left(W_{t_{i+1}}(\omega) - W_{t_i}(\omega)\right)$$

may not exists. Thus, we need a new notation of integration.

Definition 9.1.1. Let $a > 0$ and let $\{\mathscr{F}_t\}_{t\in[0,a]}$ be the standard Brownian motion filtration. We say $F : \Omega \to \mathbb{R}^{[0,a]}$ is a <u>simple process</u> if $\exists n \in \mathbb{N}$ and

(1) $0 = t_1 < t_2 < \cdots < t_n = a$,

(2) $\phi_0 : \Omega \to \mathbb{R}$ which is $\mathscr{F}_0/\mathscr{B}(\mathbb{R})$-measurable and $\phi_1, \phi_2, \ldots \phi_{n-1} : \Omega \to \mathbb{R}$ where ϕ_i is $\mathscr{F}_{t_i}/\mathscr{B}(\mathbb{R})$-measurable and $\mathbb{E}\phi_i^2 < \infty$ $\forall i = 0, 1, \ldots, n-1$,

such that

$$F(\omega, t) = \phi_0(\omega)\mathbb{1}_{\{t=0\}} + \sum_{i=1}^{n-1} \phi_i(\omega)\mathbb{1}_{\{t\in(t_i,t_{i+1}]\}}. \tag{9.1}$$

Let $\mathcal{I}_a$ denote the set of all simple process on $[0,a]$. Note that $F \in \mathcal{I}_a$ is adapted w.r.t. $\{\mathscr{F}_t\}_{t\in[0,a]}$.

Definition 9.1.2. For $F \in \mathcal{I}_a$, define $F \cdot W : \Omega \to \mathbb{R}^{[0,a]}$ as a process such that

$$(F \cdot W)(\omega, t) = \sum_{i=1}^{n-1} \phi_i(\omega) \left(W_{t_{i+1} \wedge t}(\omega) - W_{t_i \wedge t}(\omega) \right). \tag{9.2}$$

Equivalently, let $i(t) := \max\{i \in [n-1] : t_i < t\}$,

$$(F \cdot W)(\omega, t) = \sum_{i=1}^{i(t)-1} \phi_i(\omega) \left(W_{t_i+1}(\omega) - W_{t_i}(\omega) \right) + \phi_{i(t)}(\omega) \left(W_t(\omega) - W_{t_{i(t)}}(\omega) \right).$$

Theorem 9.1.1. Let $F \in \mathcal{I}_a$ be a simple process.

(1) $F \cdot W$ is well-defined (does not depend on representation of F).

(2) For $F, G \in \mathcal{I}_a$, $\alpha, \beta \in \mathbb{R}$,

$$(\alpha F + \beta G) \cdot W = \alpha F \cdot W + \beta F \cdot W.$$

(3) $F \cdot W \in \mathcal{M}_a$

(4) $\mathbb{E}\{(F \cdot W)_a\}^2 = \mathbb{E} \int_0^a F_t^2 \mathrm{d}t$. (Ito Isometry).

Proof.

(1) Suppose $F \in \mathcal{I}_a$ has 2 representations, i.e., $\forall t \in [0, a]$, $\omega \in \Omega$

$$\begin{aligned} F(\omega, t) &= \phi_0(\omega) \mathbb{1}_{\{t=0\}} + \sum_{i=1}^{n-1} \phi_i(\omega) \mathbb{1}_{\{t \in (t_i, t_{i+1}]\}} \\ &= \psi_0(\omega) \mathbb{1}_{\{t=0\}} + \sum_{i=1}^{m-1} \psi_i(\omega) \mathbb{1}_{\{t \in (s_i, s_{i+1}]\}}. \end{aligned}$$

By definition $0 = r_1 < r_2 < \cdots < r_M = a$ such that $\{t_1, \ldots, t_n\}, \{s_1, \ldots, s_m\} \subseteq \{r_1, \ldots, r_M\}$ and writing F as a simple process over $\{r_1, \ldots, r_M\}$, we may verify that $F \cdot W$ is well-defined.

(2) Let $F, G \in \mathcal{I}_a$. Write $F(\omega, t) = \phi_0(\omega) \mathbb{1}_{\{t=0\}} + \sum_{i=1}^{n-1} \phi_i(\omega) \mathbb{1}_{\{t \in (t_i, t_{i+1}]\}}$ and $F(\omega, t) = \psi_0(\omega) \mathbb{1}_{\{t=0\}} + \sum_{i=1}^{m-1} \psi_i(\omega) \mathbb{1}_{\{t \in (s_i, s_{i+1}]\}}$. We may again find $0 = r_1 < r_2 < \cdots < r_m = a$ such that $\{t_1, \ldots, t_n\}$, $\{s_1, \ldots, s_m\} \subseteq \{r_1, \ldots, r_m\}$. The claim is straightforward.

(3) Suppose $F \in \mathcal{I}_a$ is of the form

$$F(\omega, t) = \phi(\omega) \mathbb{1}_{\{t \in (u, v]\}} \text{ for } u, v \in [0, a] \text{ where } \phi \text{ is } \mathscr{F}_u / \mathscr{B}(\mathbb{R})\text{-measurable.}$$

It is clear then that $(F \cdot W)(\omega, t) = \phi(\omega) \{W_{v \wedge t}(\omega) - W_{u \wedge t}(\omega)\}$ is continuous. We will show that $\mathbb{E}(F \cdot W)_a^2 < \infty$ in claim (4).

Fix $s < t \in [0, a]$:

- If $s \leq u$, then

$$\begin{aligned} \mathbb{E}_{\bullet|\mathscr{F}_s} \phi \left(W_{v \wedge t} - W_{u \wedge t} \right) &= \mathbb{E}_{\bullet|\mathscr{F}_s} \left[\phi \left\{ \mathbb{E}_{\bullet|\mathscr{F}_u} \left(W_{v \wedge t} - W_{u \wedge t} \right) \right\} \right] \\ &= 0 = \phi \left(W_{v \wedge s} - W_{u \wedge s} \right). \end{aligned}$$

- If $s \in (u, v]$, then ϕ is $\mathscr{F}_s / \mathscr{B}(\mathbb{R})$–measurable as well and

$$\begin{aligned} \mathbb{E}_{\bullet|\mathscr{F}_s} \phi \left(W_{v \wedge t} - W_{u \wedge t} \right) &= \phi \mathbb{E}_{\cdot|\mathscr{F}_s} \left(W_{v \wedge t} - W_{u \wedge t} \right) \\ &= \phi \left(W_{v \wedge s} - W_{u \wedge s} \right). \end{aligned}$$

Thus, $F \cdot W$ is a martingale. The general case follows from claim (2) since $\mathcal{M}_a$ is a vector space.

(4) Let $F \in \mathcal{I}_a$ be of the form $F(\omega,t) = \phi_0(\omega)\mathbb{1}_{\{t=0\}} + \sum_{i=1}^{n-1} \phi_i(\omega)\mathbb{1}_{\{t\in(t_i,t_{i+1}]\}}$,

$$\begin{aligned}
\mathbb{E}\{(F\cdot W)_a\}^2 &= \mathbb{E}\left\{\sum_{i=1}^{n-1} \phi_i(W_{t_{i+1}} - W_{t_i})\right\}^2 \\
&= \sum_{i=1}^{n-1} \mathbb{E}\phi_i^2(W_{t_{i+1}} - W_{t_i})^2 \\
&= \sum_{i=1}^{n-1} \mathbb{E}\phi_i^2 \mathbb{E}_{\cdot|\mathscr{F}_{t_i}}(W_{t_{i+1}} - W_{t_i})^2 \\
&= \sum_{i=1}^{n-1} \mathbb{E}\phi_i^2 \cdot (t_{i+1} - t_i) = \int_\Omega \sum_{i=1}^{n-1} \phi_i^2(\omega) \int_{t_i}^{t_{i+1}} 1\mathrm{d}t \\
&= \int_\Omega \int_0^a \sum_{i=1}^{n-1} \phi_i^2(\omega)\mathbb{1}_{\{t\in(t_i,t_{i+1}]\}}\mathrm{d}t\mathrm{d}\mathbb{P}(\omega).
\end{aligned}$$

□

Definition 9.1.3. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let $a > 0$ and let $\{\mathscr{F}_t\}_{t\in[0,a]}$ be the standard Brownian motion filtration. Let $F : \Omega \to \mathbb{R}^{[0,a]}$ be a process. We say F is <u>measurable</u> if $(\omega, t) \mapsto F(\omega, t)$ as a function from $\Omega \times [0, a] \to \mathbb{R}$ is $(\mathscr{F} \otimes \mathscr{B}([0, a]))/\mathscr{B}(\mathbb{R})$-measurable. Note that if F is measurable, then

$$\mathbb{E}\int_0^a F_t^2\mathrm{d}t = \int_\Omega \int_0^a F(\omega,t)^2\mathrm{d}t\mathrm{d}\omega = \int_0^a \mathbb{E}F_t^2\mathrm{d}t.$$

We write

$$\mathcal{H}_a := \left\{F \text{ process on } [0,a]\ : F \text{ is measurable, adapted, } \mathbb{E}\int_0^a F_t^2\mathrm{d}t < \infty\right\}. \tag{9.3}$$

Note that if $F \in \mathcal{I}_a$, then F is of the form

$$F(\omega,t) = \phi_0(\omega)\mathbb{1}_{\{t=0\}} + \sum_{i=1}^{n-1} \phi_i(\omega)\mathbb{1}_{\{t\in(t_i,t_{i+1}]\}} \in \mathcal{H}_a$$

where $\phi_i(\omega)$ is $\mathscr{F}$-measurable and $\mathbb{1}_{\{t\in(t_i,t_{i+1})\}}$ is $\mathscr{B}([0,a])$-measurable, and that $\mathcal{H}_a$ is an inner product space with

$$\langle F, G\rangle_{L_2(\mathbb{P}\times\mathrm{Leb}[0,a])} = \int_\Omega \int_{[0,a]} F(\omega,t)G(\omega,t)\mathrm{d}\mathbb{P}(\omega)\mathrm{d}t \quad \text{for } F, G \in \mathcal{H}_a.$$

Proposition 9.1.1. If a process $X : \Omega \to \mathbb{R}^{[0,a]}$ is right-continuous, then it is measurable.

Proof.

Without loss of generality, assume $a = 1$. Define, for $n \in \mathbb{N}$, $\omega \in \Omega$,

$$X^{(n)}(\omega,t) = X(\omega, \frac{k}{n}) \text{ where } \frac{k-1}{n} < t \le \frac{k}{n} \text{ for } k \in [n].$$

Note that

$$X^{(n)}(\omega,t) = X_0(\omega)\mathbb{1}_{\{t=0\}} + \sum_{k=1}^{\infty} X_{\frac{k}{n}}(\omega)\mathbb{1}_{\{t\in(\frac{k-1}{n},\frac{k}{n}]\}}$$

is measurable. Since $\lim_{n\to\infty} X^{(n)}(\omega,t) = X(\omega,t)$ by right-continuous, X is also measurable. □

Theorem 9.1.2. For any $H \in \mathcal{H}_a$, $\exists F_1, F_2, \cdots \in \mathcal{I}_a$ s.t.

$$\lim_{n\to\infty} \mathbb{E} \int_0^a |H(\cdot,t) - F_n(\cdot,t)|^2 \mathrm{d}t = 0.$$

In other words, $\bar{\mathcal{I}}_a = \mathcal{H}_a$ where the closure is taken w.r.t. $L_2(\mathbb{P} \times \mathrm{Leb}[0,a])$.

Proof.

See Lemma 2.4 in Shreve and Karatzas. □

Remark 9.1.2. Another commonly occurring definition is that of progressive measurability. We say that a process $X : \Omega \times [0,a] \to \mathbb{R}$ is progressively measurable if for all $t \in [0,T]$, we have that X restricted to $\Omega \times [0,t]$ is $\mathscr{F}_t \otimes \mathscr{B}([0,t])$–measurable. It holds that every measurable adapted process X has a progressively measurable modification (Lemma 1.1.12 in Shreve and Karatzas). Hence, any continuous measurable adapted process is also progressively measurable.

Definition 9.1.4 (Ito Integral I). Let $H \in \mathcal{H}_a$ and let $F_1, F_2, \cdots \in \mathcal{I}_a$ such that

$$\lim_{n\to\infty} \mathbb{E} \int_0^a |H(\cdot,t) - F_n(\cdot,t)|^2 \mathrm{d}t = 0.$$

Thus, $\{F_n\}_{n=1}^\infty$ is a Cauchy sequence with respect to $L_2(\mathbb{P}\times\mathrm{Leb}[0,a])$. By Ito isometry, $\{M^{(n)} = F_n \cdot W\}_{n=1}^\infty$ is a Cauchy sequence in $\mathcal{M}_a$ with respect to $\|M^{(n)}\| = \sqrt{\mathbb{E}(M_a^{(n)})^2}$. Since $\mathcal{M}_a$ is complete (Theorem 8.6.4), $\exists M \in \mathcal{M}_a$ s.t. $\mathbb{E}\{M_a^{(n)} - M_a\}^2 \to 0$. We define

$$\int_0^t H_s \mathrm{d}W_s = M_t, \quad \forall t \in [0,a]. \tag{9.4}$$

Remark 9.1.3. We need to check that the definition does not depend on the limiting sequence. Indeed, suppose $F_1, F_2, \dots$ and $\widetilde{F}_1, \widetilde{F}_2, \dots \in \mathcal{I}_a$ are simple process sequences such that for $H \in \mathcal{H}_a$, $\mathbb{E}\int_0^a |H(\cdot,t) - F_n(\cdot,t)|^2 dt \to 0$ and $\mathbb{E}\int_0^a |H(\cdot,t) - \widetilde{F}_n(\cdot,t)|^2 dt \to 0$. Then $\mathbb{E}\int_0^a |F_n(\cdot,t) - \widetilde{F}_n(\cdot,t)|^2 dt \to 0$ by triangle inequality. Hence, $\mathbb{E}\{(F_n \cdot W)_a - (\widetilde{F}_n \cdot W)_a\}^2 \to 0$ by linearity and Ito isometry, which implies they converge to the same $M \in \mathcal{M}_a$.

Example 9.1.1. As an example, we claim that, for a.e. $\omega \in \Omega$,

$$M(\omega,t) := \left(\int W_s \mathrm{d}W_s\right)(\omega,t) = \frac{1}{2}W^2(\omega,t) - \frac{1}{2}t \quad \forall t \geq 0.$$

Note the Ito integral differs from Lebesgue Integral in the extra $\frac{1}{2}t$ factor. As a sanity check, we have $\mathbb{E}M_t = 0$ since $M \in \mathcal{M}_a$, and $\mathbb{E}(\frac{1}{2}W_t^2 - \frac{1}{2}) = 0$. Moreover, by Ito isometry, $\mathrm{Var}(M_t) = \int_0^t \mathbb{E}W_s^2 \mathrm{d}s = \frac{1}{2}t^2$.

$$\mathbb{E}(\frac{1}{2}W_t^2 - \frac{1}{2})^2 = \frac{1}{4}\mathbb{E}[W_t^4 - 2W_t^2 \cdot t + t^2] = \frac{1}{4} \cdot (3t^2 - 2t^2 + t^2) = \frac{1}{2}t^2.$$

Fix $a > 0$. For $n \in \mathbb{N}$, define $t_i^{(n)} = \frac{i}{n}a$ for $i \in [n] \cup \{0\}$. We omit superscript for simplicity. Define $F^{(n)}(\omega,t) = \sum_{i=0}^{n-1} W_{t_i}(\omega)\mathbb{1}_{\{t\in(t_i,t_{i+1}]\}}$ so that $F^{(n)} \in \mathcal{I}_a$. First, note that

$$\begin{aligned}
\mathbb{E}\int_0^a |F^{(n)}(\cdot,t) - W(\cdot,t)|^2 \mathrm{d}t &= \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}|F^{(n)}(\cdot,t) - W(\cdot,t)|^2 \mathrm{d}t \\
&= \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E}(W_{t_i} - W_t)^2 \mathrm{d}t \\
&= \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} t - t_i dt = \sum_{i=0}^{n-1} \frac{1}{2}(t_{i+1} - t_i)^2 = \frac{1}{2}\frac{a^2}{n} \to 0.
\end{aligned}$$

Now, define $M^{(n)} = (F^{(n)} \cdot W)$ so that $M_t^{(n)} = \sum_{i=0}^{n-1} W_{t_i}(W_{t_{i+1}\wedge t} - W_{t_i \wedge t})$.

We will show that $\mathbb{E}\{M_a^{(n)} - (\frac{1}{2}W_a^2 - \frac{1}{2}a)^2\} \to 0$. To see this, note that

$$\begin{aligned} M_a^{(n)} = \sum_{i=0}^{n-1} W_{t_i}(W_{t_{i+1}} - W_{t_i}) &= \sum_{i=0}^{n-1} \left\{ \frac{1}{2}(W_{t_{i+1}}^2 - W_{t_i}^2) - \frac{1}{2}(W_{t_{i+1}} - W_{t_i})^2 \right\} \\ &= \frac{1}{2}(W_a^2 - W_0^2) - \frac{1}{2}\sum_{i=0}^{n-1}(W_{t_{i+1}} - W_{t_i})^2 \\ &= \frac{1}{2}W_a^2 - \frac{1}{2}\sum_{i=0}^{n-1}(W_{t_{i+1}} - W_{t_i})^2. \end{aligned}$$

Hence, by Theorem 8.3.2,

$$\mathbb{E}\left(M_a^{(n)2} - \frac{1}{2}W_a^2 + \frac{1}{2}a \right)^2 = \frac{1}{4}\mathbb{E}\left\{ \sum_{i=0}^{n-1}(W_{t_{i+1}} - W_{t_i})^2 - a \right\}^2 \to 0.$$

Proposition 9.1.2. Let $a > 0$. For $H \in \mathcal{H}_a$, let $\int H_s dW_s \in \mathcal{M}_a$ be the Ito integral.

(a) For $\alpha, \beta \in \mathbb{R}$, $G, H \in \mathcal{H}_a$, $\alpha \int G_s \mathrm{d}W_s + \beta \int H_s \mathrm{d}W_s = \int (\alpha G_s + \beta H_s)\mathrm{d}W_s$.

(b) $\mathbb{E}\int_0^a H(\cdot,t)^2 \mathrm{d}t = \mathbb{E}(\int_0^a H_s \mathrm{d}W_s)^2$.

Proof.

Immediate since $\bar{\mathcal{I}}_a = \mathcal{H}_a$. □

Remark 9.1.4. Let $\tilde{a} > a > 0$ so that $\mathcal{H}_{\tilde{a}} \subseteq \mathcal{H}_a$. We claim that, for $H \in \mathcal{H}_{\tilde{a}}$, the definition of $\int_0^t H_s \mathrm{d}W_s$, for $t \in [0, a]$, does not depend on whether we view H as an element of $\mathcal{H}_a$ or of $\mathcal{H}_{\tilde{a}}$.

To see this, let $F_1, F_2, \cdots \in \mathcal{I}_{\tilde{a}}$ and that $\lim_{n\to\infty} \mathbb{E}\int_0^{\tilde{a}}(F_n(\cdot,t) - H(\cdot,t))^2 \mathrm{d}t = 0$. Then, $F_1, F_2, \cdots \in \mathcal{I}_a$ and $\mathbb{E}\int_0^a (F_n(\cdot,t) - H(\cdot,t))^2 \mathrm{d}t \to 0$ as well. By definition, $\forall \omega \in \Omega$, $(F_n \cdot W)(\omega, \cdot)$ defined over $[0, a]$ is the restriction of $(F_n \cdot W)(\omega, \cdot)$ defined over $[0, \tilde{a}]$. Let $\widetilde{M} \in \mathcal{M}_{\tilde{a}}$ and $M \in \mathcal{M}_a$ such that

$$\mathbb{E}\{(F_n \cdot W)_{\tilde{a}} - \widetilde{M}_{\tilde{a}}\}^2 \to 0 \text{ and } \mathbb{E}\{(F_n \cdot W)_a - M_a\}^2 \to 0.$$

Then,

$$\mathbb{E}\{(F_n \cdot W)_a - \widetilde{M}_a\}^2 \le \mathbb{E}\sup_{s \le \tilde{a}} |(F_n \cdot W)_s - \widetilde{M}_s|^2 \le 4\mathbb{E}\{(F_n \cdot W)_{\tilde{a}} - \widetilde{M}_{\tilde{a}}\}^2 \to 0.$$

It implies $\mathbb{E}(M_a - \widetilde{M}_a)^2 = 0$. Thus, there exists $B \in \mathscr{F}$, $\mathbb{P}(B) = 1$, such that $\forall \omega \in B$, $M(\omega, t) = \widetilde{M}(\omega, t)$ for all $t \in [0, a]$. So $M(\omega, \cdot)$ is the restriction of $\widetilde{M}(\omega, \cdot)$ to $[0, a]$. In particular, $M_a = \mathbb{E}[\widetilde{M}_{\tilde{a}} | \mathscr{F}_a]$ a.s.

Definition 9.1.5 (Ito Integral II). Define

$$\mathcal{H} := \left\{ F \text{ process on } [0, \infty) : \text{ measurable, adapted, and } \mathbb{E}\int_0^a F_t^2 \mathrm{d}t < \infty, \forall a \ge 0 \right\} \tag{9.5}$$

so that $\mathcal{H} \subseteq \mathcal{H}_a$, $\forall a \ge 0$. For $H \in \mathcal{H}$, let $M^{(n)} \in \mathcal{M}_n$ for $n \in \mathbb{N}$ be the integral of H viewed as an element of $\mathcal{H}_n$.

Define a process $M : \Omega \times [0, \infty) \to \mathbb{R}$ by $M(\omega, t) = M^{(n)}(\omega, t)$ for $\omega \in \Omega$ and $t \in [n-1, n)$. For $n < m \in \mathbb{N}$, let $B_{n,m} \in \mathscr{F}$ be the event such that $\mathbb{P}(B_{n,m}) = 1$ and $\forall \omega \in B_{n,m}$, we have $M^{(n)}(\omega, t) = M^{(m)}(\omega, t)$ for all $t \in [0, n]$. Let $B = \cap_{n=1}^\infty \cap_{m \ge n} B_{n,m}$, then $\mathbb{P}(B) = 1$ and $M(\omega, \cdot)$ is continuous on $[0, \infty)$, $\forall \omega \in B$.

For $s < t \in [0, \infty)$, suppose $s \in [m-1, m)$ and $t \in [n-1, n)$ for $m \le n \in \mathbb{N}$, then $M_s = M_s^{(m)} = \mathbb{E}[M_t^{(n)} | \mathscr{F}_s] = \mathbb{E}[M_t | \mathscr{F}_s]$. So M is a martingale. We then define

$$\int_0^t H_s \mathrm{d}W_s = M_t, \forall t \ge 0. \tag{9.6}$$

9.2 Localization

Remark 9.2.1. In the definition of $\mathcal{H}_a$, the restriction that $\mathbb{E}\int_0^a F_t^2 \mathrm{d}t < \infty$ is restrictive and often difficult to verify. For instance, let $f(x) = e^{x^4}$ for $x \in \mathbb{R}$ and let $F(\omega,t) = f(W(\omega,t))$ for $(\omega,t) \in \Omega \times [0,a]$, then $\int_0^1 \mathbb{E}F_t^2 dt = \int_0^1 \mathbb{E}e^{2W_t^4}\mathrm{d}t = \infty$.

A much easier class of integrand is

$$\mathscr{L}_{\mathrm{LOC}} := \left\{ F \text{ process on } [0,a]:\ \text{measurable, adapted, } \mathbb{P}\left(\int_0^a F_t^2 \mathrm{d}t < \infty\right) = 1 \right\}. \tag{9.7}$$

Observe that if $f : \mathbb{R} \to \mathbb{R}$ is continuous, then $\forall \omega \in \Omega$, $t \to f(W(\omega,t))$ is bounded on $[0,a]$ since $W(\omega,t)$ is bounded for $t \in [0,a]$. Hence $f \circ W \in \mathscr{L}_{\mathrm{LOC}}[0,a]$.

Theorem 9.2.1. Let $a > 0$ and let $H \in \mathcal{H}_a$. Let $\tau : \Omega \to [0,\infty]$ be a stopping time. It holds that $H\mathbb{1}_{[0,\tau]} \in \mathcal{H}_a$ and writing $M = \int H_s \mathrm{d}W_s$ and $\widetilde{M} = \int H_s \mathbb{1}_{\{s\le\tau\}}\mathrm{d}W_s \in \mathcal{M}_a$, we have that for a.e. $\omega \in \Omega$, $M(\omega, t\wedge\tau(\omega)) = \widetilde{M}(\omega,t)$, $\forall t \in [0,a]$.

Proof.

Note that, $\forall t \ge 0$, $\omega \mapsto \mathbb{1}_{\{t\le\tau(\omega)\}} = (1 - \mathbb{1}_{\{\tau(\omega)<t\}})$ is $\mathscr{F}_t/\mathscr{B}(\mathbb{R})$-measurable and thus $H\mathbb{1}_{[0,\tau]}$ is adapted. Also, $(\omega,t) \mapsto \mathbb{1}_{\{t\le\tau(\omega)\}} = \mathbb{1}_{\{t-\tau(\omega)\le 0\}}$ is $\mathscr{F}\otimes\mathscr{B}([0,a])/\mathscr{B}(\mathbb{R})$-measurable and hence $H\mathbb{1}_{[0,\tau]}$ is measurable. Since $(H\mathbb{1}_{[0,\tau]})^2 \le H^2$, we have that $H\mathbb{1}_{[0,\tau]} \in \mathcal{H}_a$.

Step 1: First suppose $\tau : \Omega \to [0,\infty]$ takes a finite set of values. Let $H \in \mathcal{H}_a$ and $F_1, F_2, \dots \in \mathcal{I}_a$ such that $\mathbb{E}\int_0^a |F_n(\cdot,t) - H(\cdot,t)|^2 dt \to 0$.

Fix arbitrary $n \in \mathbb{N}$ and suppose F_n has the form

$$F_n(\omega,t) = \sum_{i=1}^{m-1} \phi_i(\omega)\mathbb{1}_{\{t\in(t_i,t_{i+1}]\}}.$$

By increasing m if necessary, we may assume range$(\tau) \subseteq \{t_1,\dots,t_m\}$, so that for any $\omega \in \Omega$, $\mathbb{1}_{\{t\le\tau(\omega)\}}$ is constant on intervals $(t_i,t_{i+1}]$'s. Thus,

$$F_n\mathbb{1}_{[0,\tau]}(\omega,t) = F_n(\omega,t)\mathbb{1}_{\{t\le\tau(\omega)\}} = \sum_{i=1}^{m-1}\phi_i(\omega)\underbrace{\mathbb{1}_{\{t_i<\tau(\omega)\}}}_{=\mathbb{1}_{\{t_{i+1}\le\tau(\omega)\}}}\mathbb{1}_{\{t\in(t_i,t_{i+1}]\}}.$$

Note that

$$\int_\Omega\int_0^a |F_n(\omega,t)\mathbb{1}_{\{t\le\tau(\omega)\}} - H(\omega,t)\mathbb{1}_{\{t\le\tau(\omega)\}}|^2\mathrm{d}t\mathrm{d}\mathbb{P}(\omega) = \int_\Omega\int_0^{\tau(\omega)} |F_n(\omega,t) - H(\omega,t)|^2\mathrm{d}t\mathrm{d}\mathbb{P}(\omega) \to 0,$$

and that

$$\begin{aligned}(F_n\mathbb{1}_{[0,\tau]}\cdot W)(\omega,t) &= \sum_{i=1}^{m-1}\phi_i(\omega)\mathbb{1}_{\{t_{i+1}\le\tau(\omega)\}}(W(\omega,t_{i+1}\wedge t) - W(\omega,t_i\wedge t)) \\ &= \sum_{i=1}^{m-1}\phi_i(\omega)\{W(\omega,t_{i+1}\wedge t\wedge\tau(\omega)) - W(\omega,t_i\wedge t\wedge\tau(\omega))\} \\ &= (F_n\cdot W)(\omega,t\wedge\tau(\omega))\end{aligned}$$

Denote $M^{(n)} = F_n\cdot W$, $\widetilde{M}^{(n)} = F_n\mathbb{1}_{[0,\tau]}\cdot W$, $M = \int H_s\mathrm{d}W_s$, and $\widetilde{M} = \int H_s\mathbb{1}_{\{s\le\tau\}}\mathrm{d}W_s$. Recall that $\{M_{t\wedge\tau}\}_{t\in[0,a]} \in \mathcal{M}_a$ by Doob's optional stopping theorem. We have that

$$\begin{aligned}\mathbb{E}(M_{a\wedge\tau} - \widetilde{M}_a)^2 &\le \mathbb{E}(M_{a\wedge\tau} - M_{a\wedge\tau}^{(n)})^2 + \mathbb{E}(M_{a\wedge\tau}^{(n)} - \widetilde{M}_a)^2 \\ &\le \mathbb{E}\sup_{s\le a}(M_s - M_s^{(n)})^2 + \mathbb{E}(\widetilde{M}_a^{(n)} - \widetilde{M}_a)^2 \\ &\le 4\mathbb{E}(M_a - M_a^{(n)})^2 + \mathbb{E}(\widetilde{M}_a^{(n)} - \widetilde{M}_a)^2.\end{aligned}$$

By taking limit $n \to \infty$, we have that for a.e. $\omega \in \Omega$, $M(\omega, t \wedge \tau(\omega)) = \widetilde{M}(\omega, t)$, $\forall t \in [0, a]$

Step 2: Let $\tau : \Omega \to [0, \infty]$ be a general stopping time. For $n \in \mathbb{N}$, define $\tau_n(\omega) = \frac{k}{n}$ for $k \in \mathbb{N}$ such that $\frac{k-1}{n} \leq \tau(\omega) \leq \frac{k}{n}$, so that $\tau_n(\omega) \to \tau(\omega)$, $\forall \omega \in \Omega$.

Observe that

$$\int_\Omega \int_0^a |H(\omega,t)\mathbb{1}_{\{t \leq \tau_n(\omega)\}} - H(\omega,t)\mathbb{1}_{\{t \leq \tau(\omega\}}|^2 \mathrm{d}t\mathrm{d}\mathbb{P}(\omega) \leq \int_\omega \int_{\tau(\omega)}^{\tau_n(\omega)\wedge a} H(\omega,t)^2 \mathrm{d}t\mathrm{d}\mathbb{P}(\omega) \to 0.$$

Thus, $\int H_s \mathbb{1}_{\{s \leq \tau_n\}} \mathrm{d}W_s \to \int H_s \mathbb{1}_{\{s \leq \tau\}} \mathrm{d}W_s$ in $\mathcal{M}_a$-norm by Ito isometry.

Define $M = \int H_s \mathrm{d}W_s \in \mathcal{M}_a$ and define $\widetilde{M} = \int H_s \mathbb{1}_{\{s \leq \tau\}} \mathrm{d}W_s$. Then, for any $n \in \mathbb{N}$,

$$\mathbb{E}(M_{a\wedge\tau} - \widetilde{M}_a)^2 \leq \underbrace{\mathbb{E}(M_{a\wedge\tau} - M_{a\wedge\tau_n})^2}_{\text{term 1}} + \underbrace{\mathbb{E}(M_{a\wedge\tau_n} - \widetilde{M}_a)^2}_{\text{term 2}}.$$

First, since $M_{a\wedge\tau_n} = \int_0^a H_s \mathbb{1}_{\{s \leq \tau_n\}} \mathrm{d}W_s$, term 2 $\to 0$ as $n \to \infty$. In addition, for term 1, note that the sequence of functions

$$\omega \mapsto M_{a\wedge\tau(\omega)}(\omega) - M_{a\wedge\tau_n(\omega)}(\omega), \text{ indexed by } n \in \mathbb{N},$$

converges to 0 a.s. by continuity of M. Since $\mathbb{E}(M_{a\wedge\tau_n} - M_{a\wedge\tau})^2 \leq 2\mathbb{E}M^2_{a\wedge\tau_n} + 2\mathbb{E}M^2_{a\wedge\tau} \leq 4\mathbb{E}M_a^2 < \infty$, we have that term $1 \to 0$ as $n \to \infty$ by dominated convergence theorem.

□

Corollary 9.2.1. Let $G, H \in \mathcal{H}_a$ and let $\tau : \Omega \to [0, \infty]$ be a stopping time. Suppose, for a.e. $\omega \in \Omega$, $G(\omega, t) = H(\omega, t)$, $\forall t \in [0, a \wedge \tau(\omega)]$, and let $X, Y \in \mathcal{M}_a$ such that $X_t = \int_0^t G_s \mathrm{d}W_s$ and $Y_t = \int_0^t H_s \mathrm{d}W_s$. Then, for a.e. $\omega \in \Omega$, $X(\omega, t) = Y(\omega, t)$, $\forall t \in [0, a \wedge \tau(\omega)]$.

Proof.

By Theorem 9.2.1, for a.e. $\omega \in \Omega$, $\forall t \in [0, a]$,

$$\begin{aligned} X(\omega, t \wedge \tau(\omega)) &= \left(\int G_s \mathbb{1}_{\{s \leq \tau\}} \mathrm{d}W_s\right)(\omega, t) \\ &= \left(\int H_s \mathbb{1}_{\{s \leq \tau\}} \mathrm{d}W_s\right)(\omega, t) = Y(\omega, t \wedge \tau(\omega)). \end{aligned}$$

Thus, for a.e. $\omega \in \Omega$, for $t \in [0, a \wedge \tau(\omega)]$, we have that

$$X(\omega, t) = X(\omega, t \wedge \tau(\omega)) = Y(\omega, t \wedge \tau(\omega)) = Y(\omega, t).$$

□

Definition 9.2.1. Let $H \in \mathscr{L}_{\text{LOC}}[0, a]$, we say that a sequence of stopping times $\tau_1 \leq \tau_2 \leq \cdots : \Omega \to [0, \infty]$ is a localizing sequence for H if

(a) $H\mathbb{1}_{[0,\tau_n]} \in \mathcal{H}_a$, $\forall n \in \mathbb{N}$,

(b) $\mathbb{P}(\cup_{n=1}^\infty \{\tau_n \geq a\}) = 1$.

For example, we may define $\tau_n(\omega) = \inf\{s \geq 0 : \int_0^s H(\omega, t)^2 dt \geq n\}$. Note that $(\omega, s) \mapsto \int_0^s H(\omega, t)^2 \mathrm{d}t$ is a continuous process (use the fact that $\exists F_1, F_2, \dots \in \mathcal{I}_a$ such that $F_n \mapsto H$ in $L_2(\mathbb{P} \times \text{Leb}[0, a])$ and hence, τ_n is a stopping time.

Define $M^{(n)} = \int H_s \mathbb{1}_{\{s \leq \tau_n\}} \mathrm{d}W_s \in \mathcal{M}_a$ and let $B_{n,m} \in \mathscr{F}$ for $n \leq m \in \mathbb{N}$ be defined as

$$B_{n,m} := \{\omega \in \Omega : M^{(n)}(\omega, t) = M^{(m)}(\omega, t) \text{ for } t \in [0, \tau_n(\omega)]\}.$$

By Corollary 9.2.1, $P(B_{n,m}) = 1$ since $H_t \mathbb{1}_{\{t \leq \tau_n\}} = H_t \mathbb{1}_{\{t \leq \tau_m\}}$ for $t \in [0, \tau_n(\omega)]$. In addition, writing $B = \cap_{n=1}^\infty \cap_{m \geq n} B_{n,m}$, $\mathbb{P}(B) = 1$ by union bound. For $\omega \in B \cap (\cup_{n=1}^\infty \{\tau_n \geq a\})$, for $t \in (0, a]$, define $M(\omega, t) = M^{(n)}(\omega, t)$ where $n \in \mathbb{N}$ is such that $\tau_{n-1}(\omega) < t \leq \tau_n(\omega)$. (Define $\tau_0 = 0$). For other ω, define $M(\omega, t) = 0$. We then define $\left(\int H_s \mathrm{d}W_s\right)(\omega, t) = M(\omega, t)$.

Remark 9.2.2. (a) For $H \in \mathscr{L}_{\text{LOC}}[0,a]$, $\int H_s \mathrm{d}W_s$ is not necessarily a martingale, but if $\tau_1 \le \tau_2 \le \dots$ is a localizing sequence for H, then $\{M_{t\wedge\tau_n}\}_{t\ge 0} \in \mathcal{M}_a$ for all $n \in \mathbb{N}$ since $\{M_{t\wedge\tau_n}\}_{t\ge 0} = \int H_s \mathbb{1}_{\{s\le\tau_n\}}\mathrm{d}W_s$ and $H_s\mathbb{1}_{\{s\le\tau_n\}} \in \mathcal{H}_a$. Thus, M is called a local martingale.

(b) If $\tau_1 \le \tau_2 \le \dots$ and $\tau_1' \le \tau_2' \le \dots$ are 2 localizing sequences for $H \in \mathscr{L}_{\text{LOC}}[0,a]$, then so is $\{\pi_n := \tau_n \wedge \tau_n'\}_{n=1}^{\infty}$.

Writing $M^{(n)} := \int H_s \mathbb{1}_{\{s\le\tau_n\}}\mathrm{d}W_s$ and $\widetilde{M}^{(n)} := \int H_s \mathbb{1}_{\{s\le\pi_n\}}\mathrm{d}W_s$, we have that for a.e. $\omega \in \Omega$,

$$M^{(n)}(\omega,t) = \widetilde{M}^{(n)}(\omega,t) \quad \forall t \in [0,\pi_n(\omega)].$$

Therefore, writing $M = \int H_s\mathrm{d}W_s$, we have that $M(\omega,t) = \widetilde{M}^{(n)}(\omega,t)$, $\forall t \in [0,\pi_n(\omega)]$. It implies M does not depend on the choice of localizing sequence.

(c) We may define

$$\mathscr{L}_{\text{LOC}} = \left\{F \text{ process on } [0,\infty) : \text{ measurable, adapted } \forall a > 0, \mathbb{P}\left(\int_0^a F_t^2\mathrm{d}t < \infty\right) = 1\right\}$$

and extend the definition of Ito integral to $\mathscr{L}_{\text{LOC}}$.

9.3 Ito's Lemma

Remark 9.3.1. Let W be the standard Brownian motion and let $f : \mathbb{R} \to \mathbb{R}$ be twice continuously differentiable. Let $t > 0$ and let $0 = t_1 < t_2 < \dots < t_n = t$ be a partition. Then

$$\begin{aligned} f(W_t) - f(W_0) &= \sum_{i=1}^{n-1} f(W_{t_{i+1}}) - f(W_{t_i}) \\ &= \underbrace{\sum_{i=1}^{n-1} f'(W_{t_i})(W_{t_{i+1}} - W_{t_i})}_{\text{term 1}} + \underbrace{\frac{1}{2}\sum_{i=1}^{n-1} f''(W_{s_i})(W_{t_{i+1}} - W_{t_i})^2}_{\text{term 2}}, \end{aligned}$$

where s_i is a random variable taking value in $[t_i, t_{i+1}]$.

Term 1 "converges" to $\int_0^t f'(W_s)\mathrm{d}W_s$ as $n \to \infty$. Term 1 resembles the Ito integral of a simple function but we do not know if, for all $i \in [n-1]$, $\mathbb{E}f'(W_{t_i})^2 < \infty$. Since $\sum_{i=1}^{n-1}(W_{t_{i+1}} - W_{t_i})^2 \to t$ in $L_2(\mathbb{P})$, we expect term 2 to "converge" to $\frac{1}{2}\int_0^t f''(W_s)\mathrm{d}s$. This is the intuition for Ito's lemma.

Theorem 9.3.1 (Ito's lemma). Let W be a standard Brownian Motion and let $f : \mathbb{R} \to \mathbb{R}$ be twice continuously differentiable. Then, for a.e. $\omega \in \Omega$, $\forall t \ge 0$,

$$f(W(\omega,t)) - f(W(\omega,0)) = \left(\int f'(W_s)\mathrm{d}W_s\right)(\omega,t) + \frac{1}{2}\int_0^t f''(W(\omega,s))\mathrm{d}s. \tag{9.8}$$

This is informally expressed as $\forall t \ge 0$,

$$f(W_t) - f(W_0) = \int_0^t f'(W_s)\mathrm{d}W_s + \frac{1}{2}\int_0^t f''(W_s)\mathrm{d}s.$$

Example 9.3.1. (a) If $f(x) = x$ is the identity, then $f'(x) = 1$ and $f''(x) = 0$. We have, for a.e. $\omega \in \Omega$, $\forall t \ge 0$,

$$f(W(\omega,t)) - f(W(\omega,0)) = \left(\int 1dW\right)(\omega,t) = W(\omega,t) \text{ as expected.}$$

(b) If $f(x) = x^2$, then $f'(x) = 2x$ and $f''(x) = 2$. We have, for a.e. $\omega \in \Omega$, $\forall t \geq 0$,

$$f(W(\omega,t)) - f(W(\omega,0)) = \left(\int 2W_s \mathrm{d}W_s\right)(\omega,t) + \frac{1}{2}\int_0^t 2\mathrm{d}s.$$

which implies $2(\int W_s \mathrm{d}W_s)(\omega,t) = W(\omega,t)^2 - t$ as expected from Example 9.1.1.

Proof of Theorem 9.3.1 (Ito's lemma).

For simplicity, we write $fW(\omega,t) := f(W(\omega,t))$.

First, we note that $\{f'W_t\}_{t\geq 0}$ is a continuous and adapted process. Moreover, $\forall a \geq 0$, $\forall \omega \in \Omega$, we have that $\int_0^a (f'W(\omega,t))^2 \mathrm{d}t < \infty$ since $t \mapsto W(\omega,t)$ is continuous and thus maps $[0,a]$ to some interval. Thus, $f'W \in \mathscr{L}_{\mathrm{LOC}}$ and $\int f'W_s \mathrm{d}W_s$ is well-defined.

Now, $\forall \omega \in \Omega$, $t \mapsto f''W(\omega,t)$ is continuous and hence $\forall \omega \in \Omega$, $\int_0^t f''W(\omega,s)\mathrm{d}s$ is well-defined as a Lebesgue-Riemann integral.

Now, fix any $a > 0$ and let $0 = t_1^{(n)} < t_2^{(n)} < \cdots < t_n^{(n)} = a$ be a sequence of partitions such that $\lim_{n\to\infty} \max_{i\in[n-1]} |t_{i+1}^{(n)} - t_i^{(n)}| = 0$. We omit superscript for simplicity.

For $m \in \mathbb{N}$, define stopping times $\tau_m(\omega) = \inf\{t \geq 0 : |W(\omega,t)| \geq m\}$. Note that $\mathbb{P}(\cup_{m\in\mathbb{N}}\{\tau_m \geq a\}) = 1$ and that $\tau_1 \leq \tau_2 \leq \ldots$. We will prove, for any $m \in \mathbb{N}$, for a.e. $\omega \in \Omega$, $\forall t \in [0,a]$,

$$fW(\omega, t\wedge\tau_m(\omega)) - fW(\omega,0) = \left(\int f'W_s \mathrm{d}W_s\right)(\omega, t\wedge\tau_m(\omega)) + \frac{1}{2}\int_0^{t\wedge\tau_m(\omega)} f''W(\omega,s)\mathrm{d}s.$$

Since, for a.e. $\omega \in \Omega$, $\exists m \in \mathbb{N}$ such that $\tau_m(\omega) \geq a$, this proves the theorem.

Fix $m \in \mathbb{N}$, $\forall \omega \in \Omega$, $\forall t \in [0,a]$, $\exists s_1(\omega) \in [t_1,t_2], \ldots, s_{n-1}(\omega) \in [t_{n-1},t_n]$ such that

$$\begin{aligned}
&fW(\omega, t\wedge\tau_m(\omega)) - fW(\omega,0) \\
=&\sum_{i=1}^{n-1} fW(\omega, t_{i+1}\wedge t\wedge\tau_m(\omega)) - fW(\omega, t_i\wedge t\wedge\tau_m(\omega)) \\
=&\sum_{i=1}^{n-1} f'W(\omega,t_i)\{W(\omega, t_{i+1}\wedge t\wedge\tau_m(\omega)) - W(\omega, t_i\wedge t\wedge\tau_m(\omega))\} \to \text{Term1} \\
&\quad + \frac{1}{2}\sum_{i=1}^{n-1} f''W(\omega, s_i(\omega))\{W(\omega, t_{i+1}\wedge t\wedge\tau_m(\omega)) - W(\omega, t_i\wedge t\wedge\tau_m(\omega))\}^2. \to \text{Term2}
\end{aligned}$$

We will show that $\text{Term1}(\omega,\cdot) \to \left(\int f'W_s \mathrm{d}W_s\right)(\omega,\cdot)$ and $\text{Term2}(\omega,\cdot) \to \frac{1}{2}\int_0^{\cdot\wedge\tau_m(\omega)} f''W(\omega,s)\mathrm{d}s$ in $\|\cdot\|_\infty$ as $n \to \infty$ for a.e. $\omega \in \Omega$. Since the LHS does not depend on n, the theorem follows.

We first analyse Term1. We claim that $\forall n \in \mathbb{N}$, the processes

$$\begin{aligned}
U^{(n)}(\omega,t) &:= \sum_{i=1}^{n-1} f'W(\omega,t_i)\{W(\omega, t_{i+1}\wedge t\wedge\tau_m(\omega)) - W(\omega, t_i\wedge t\wedge\tau_m(\omega))\} \\
V(\omega,t) &:= \left(\int f'W_s \mathrm{d}W_s\right)(\omega, t\wedge\tau_m(\omega))
\end{aligned}$$

are in $\mathcal{M}_a$ and that $\lim_{n\to\infty} \mathbb{E}(U_a^{(n)} - V_a)^2 = 0$ $(\star)$. This means that $\lim_{n\to\infty} \mathbb{E}\sup_{t\in[0,a]}(U_t^{(n)} - V_t)^2 = 0$.

Define, $\forall n \in \mathbb{N}$, the process $F^{(n)} : \Omega \times [0,a] \to \mathbb{R}$ such that $\forall \omega \in \Omega$, $t \in [0,a]$,

$$F^{(n)}(\omega,t) = \sum_{i=1}^{n-1} f'W(\omega,t_i)\mathbb{1}_{\{t_i \leq \tau_m(\omega)\}}\mathbb{1}_{\{t\in(t_i,t_i+1]\}}$$

Note that $F^{(n)} \in \mathcal{I}_a$ since for all $i \in [n-1]$,

$$\mathbb{E}[(f'W_{t_i})^2\mathbb{1}_{\{t_i\leq\tau_m(\omega)\}}] \leq \sup_{\omega\in\Omega}\sup_{t\leq\tau_m(\omega)} |f'W(\omega,t)|^2 \leq \sup_{u\in[-m,m]} |f'(u)| := C_m < \infty.$$

Likewise, $f'W\mathbb{1}_{[0,\tau_m]} \in \mathcal{H}_a$. Note that

$$\begin{aligned}
U^{(n)}(\omega,t) &= \sum_{i=1}^{n-1} f'W(\omega,t_i)\mathbb{1}_{\{t_i\le\tau_m(\omega)\}}\{W(\omega,t_{i+1}\wedge t\wedge\tau_m(\omega)) - W(\omega,t_i\wedge t\wedge\tau_m(\omega))\} \\
&= \left(\int F_s^{(n)}\mathrm{d}W_s\right)(\omega, t\wedge\tau_m(\omega)) \\
&= \left(\int F_s^{(n)}\mathbb{1}_{\{s\le\tau_m(\omega)\}}\mathrm{d}W_s\right)(\omega,t).
\end{aligned}$$

We will show that $F^{(n)}\mathbb{1}_{[0,\tau_m]} \to f'W\mathbb{1}_{[0,\tau_m]}$ in $L_2(\mathbb{P}\times \mathrm{Leb}[0,a])$. First, observe that

$$\begin{aligned}
\sup_{\omega\in\Omega}\sup_{t\in[t_i\wedge\tau_m(\omega),t_{i+1}\wedge\tau_m(\omega)]}|f'W(\omega,t_i) - f'W(\omega,t)|^2 &\le \sup_{\omega\in\Omega}\sup_{t\le\tau_m(\omega)}|f''W(\omega,t)|^2(t_{i+1}-t_i)^2 \\
&\le (t_{i+1}-t_i)^2\cdot\sup_{u\in[-m,m]}f''(u)^2 := (t_{i+1}-t_i)^2\widetilde{C}_m.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\int_\Omega\int_0^a |F^{(n)}(\omega,t)\mathbb{1}_{\{t\le\tau_m(\omega)\}} - f'W(\omega,t)\mathbb{1}_{\{t\le\tau_m(\omega)\}}|^2\mathrm{d}t\mathrm{d}\mathbb{P}(\omega) \\
=&\int_\Omega\int_0^{\tau_m(\omega)} |F^{(n)}(\omega,t) - f'W(\omega,t)|^2\mathrm{d}t\mathrm{d}\mathbb{P}(\omega) \\
=&\int_\Omega\sum_{i=1}^{n-1}\int_{t_i\wedge\tau_m(\omega)}^{t_{i+1}\wedge\tau_m(\omega)} |f'W(\omega,t_i) - f'W(\omega,t)|^2\mathrm{d}t\mathrm{d}\mathbb{P}(\omega) \\
\le&\sum_{i=1}^{n-1}\int_\Omega\int_{t_i\wedge\tau_m(\omega)}^{t_{i+1}\wedge\tau_m(\omega)} \widetilde{C}_m(t_{i+1}-t_i)^2\mathrm{d}t\mathrm{d}\mathbb{P}(\omega) \\
\le&\widetilde{C}_m\sum_{i=1}^{n-1}(t_{i+1}-t_i)^2 \\
\le&\widetilde{C}_m\max_{i\in[n-1]}(t_{i+1}-t_i)^2\sum_{i=1}^{n-1}(t_{i+1}-t_i) \to 0 \text{ as } n\to\infty,
\end{aligned}$$

which implies $F^{(n)}\mathbb{1}_{[0,\tau_m]} \to f'W\mathbb{1}_{[0,\tau_m]}$ in $L_2(\mathbb{P}\times\mathrm{Leb}[0,a])$. Our claim $(\star)$ thus follows by Ito isometry.

For Term2, we write

$$\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{n-1} f''W(\omega,s_i(\omega))\{W(\omega,t_{i+1}\wedge t\wedge\tau_m(\omega)) - W(\omega,t_i\wedge t\wedge\tau_m(\omega))\}^2 \\
=&\frac{1}{2}\mathrm{Term2A}(\omega,t) + \frac{1}{2}\mathrm{Term2B}(\omega,t) + \underbrace{\frac{1}{2}\sum_{i=1}^{n-1} f''W(\omega,t_i)\{t_{i+1}\wedge t\wedge\tau_m(\omega) - t_i\wedge t\wedge\tau_m(\omega)\}}_{\mathrm{Term2C}(\omega,t)},
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{Term2A}(\omega,t) := &\sum_{i=1}^{n-1} f''W(\omega,t_i)\Big[\big\{W(\omega,t_{i+1}\wedge t\wedge\tau_m(\omega)) - W(\omega,t_i\wedge t\wedge\tau_m(\omega)\big\}^2 \\
&\qquad - \big\{t_{i+1}\wedge t\wedge\tau_m(\omega) - t_i\wedge t\wedge\tau_m(\omega)\big\}\Big] \\
=&\sum_{i=1}^{n-1} f''W(\omega,t_i)\mathbb{1}_{\{t_i\le\tau_m(\omega)\}}\Big[\big\{W(\omega,t_{i+1}\wedge t\wedge\tau_m(\omega)) - W(\omega,t_i\wedge t\wedge\tau_m(\omega))\big\}^2 \\
&\qquad - \big\{t_{i+1}\wedge t\wedge\tau_m(\omega) - t_i\wedge t\wedge\tau_m(\omega)\big\}\Big]
\end{aligned}$$

and

$$\text{Term2B}(\omega,t) := \sum_{i=1}^{n-1}(f''W(\omega,s_i(\omega)) - f''W(\omega,t_i))\big\{W(\omega,t_{i+1}\wedge t\wedge\tau_m(\omega)) - W(\omega,t_i\wedge t\wedge\tau_m(\omega))\big\}^2.$$

Consider Term2C first. We note that, for any $\omega\in\Omega$, the function $t\mapsto f''W(\omega,t)$ is continuous on $[0,a]$ and therefore is uniformly continuous.

> Recall that $\psi:[0,a]\to\mathbb{R}$ is continuous if $\forall x_0$, the modulus of continuity
>
> $$W_{\psi,x_0}(\delta) := \sup\{|\psi(x)-\psi(x_0)| : x\in[0,a], |x-x_0|<\delta\}$$
>
> satisfy $\lim_{\delta\to 0} m_{\psi,x_0}(\delta) = 0$.
>
> The function ψ is uniformly continuous if
>
> $$m_\psi(\delta) := \sup_{x_0\in[0,a]} m_{\psi,x_0}(\delta) = \sup\{|\psi(x)-\psi(x_0)| : x,x_0\in[0,a], |x-x_0|<\delta\}$$
>
> satisfy $\lim_{\delta\to 0} m_\psi(\delta) = 0$.

Therefore, for any fixed $\omega\in\Omega$,

$$\begin{aligned}
&\sup_{t\in[0,a]}\left|\sum_{i=1}^{n-1} f''W(\omega,t_i)(t_{i+1}\wedge t\wedge\tau_m(\omega) - t_i\wedge t\wedge\tau_m(\omega)) - \int_0^{t\wedge\tau_m(\omega)} f''W(\omega,s)\mathrm{d}s\right| \\
\le& \sup_{t\in[0,a]}\sum_{i=1}^{n-1}\int_{t_i\wedge t\wedge t_m(\omega)}^{t_{i+1}\wedge t\wedge\tau_m(\omega)} |f''W(\omega,t_i) - f''W(\omega,s)|\mathrm{d}s \\
\le& \sum_{i=1}^{n-1}(t_{i+1}-t_i)\sup_{s\in[t_{i+1},t_i]}|f''W(\omega,t_i) - f''W(\omega,s)| \\
\le& a\cdot\max_{i\in[n-1]}\sup_{s\in[t_i,t_{i+1}]}|f''W(\omega,t_i) - f''W(\omega,s)| \\
\le& a\cdot\sup_{s,s':|s-s'|\le\delta_n}|f''W(\omega,s) - f''W(\omega,s')|
\end{aligned}$$

where $\delta_n := \max_{i\in[n-1]}|t_{i+1}-t_i|$. Since $\delta_n\to 0$ as $n\to\infty$, we have that, $\forall\omega\in\Omega$,

$$\sup_{t\in[0,a]}\left|\text{Term2C}(\omega,t) - \int_0^{t\wedge\tau_m(\omega)} f''W(\omega,s)\mathrm{d}s\right|\to 0.$$

We will show that Term2A and Term2B both converge to 0 in the sense that

$$\text{for a.e. } \omega\in\Omega,\ \sup_{t\in[0,a]}|\text{Term2A}(\omega,t)|\to 0 \text{ (for a subsequence)}$$

and likewise for Term2B.

Consider Term2A, we claim Term2A $\in\mathcal{M}_a$ because:

- $(\omega,t)\mapsto\{W_{t_{i+1}\wedge t} - W_{t_i\wedge t}\}^2 - (t_{i+1}\wedge t - t_i\wedge t)\in\mathcal{M}_a$,
- τ_m is a stopping time,
- $\omega\mapsto f''W(\omega,t_i)\mathbb{1}_{\{t_i\le\tau(\omega)\}}$ is $\mathscr{F}_{t_i}/\mathscr{B}(\mathbb{R})$-measurable,
- we will show that $\mathbb{E}\text{Term2A}(\cdot,a)^2<\infty$.

Thus, we need only show that $\lim_{n\to\infty}\mathbb{E}\mathrm{Term2A}(\cdot,t)^2=0$. Then Doob's Maximal norm inequality (Theorem 8.6.3) yields that $\lim_{n\to\infty}\mathbb{E}\sup_{t\in[0,a]}\mathrm{Term2A}(\cdot,t)^2=0$, and thus $\sup_{t\in[0,a]}\mathrm{Term2A}(\cdot,t)^2\to 0$ in probability and a subsequence converge a.s.

Now, note that

$$\begin{aligned}
&\mathbb{E}\mathrm{Term2A}(\cdot,a)^2=\int_\Omega \mathrm{Term2A}(\omega,a)^2\mathrm{d}\mathbb{P}(\omega)\\
\le&\sup_{\omega\in\Omega}\max_{i\in[n-1]}|f''W(\omega,t_i)|^2\cdot\mathbb{E}\Big(\sum_{i=1}^{n-1}\Big[\{W(\omega,t_{i+1}\wedge\tau_m(\omega))-W(\omega,t_i\wedge\tau_m(\omega))\}^2\\
&\qquad\qquad\qquad\qquad\qquad\qquad-\{t_{i+1}\wedge\tau_m(\omega)-t_i\wedge\tau_m(\omega)\}\Big]\Big)^2\\
\le&\widetilde{C}_m\left(\mathbb{E}\sum_{i=1}^{n-1}V_i^2+2\mathbb{E}\sum_{i=1}^{n-2}\sum_{l=i+1}^{n-1}V_iV_l\right), \qquad (\star\star)
\end{aligned}$$

where $V_i(\omega):=\{W(\omega,t_{i+1}\wedge\tau_m(\omega))-W(\omega,t_i\wedge\tau_m(\omega))\}^2-(t_{i+1}\wedge\tau_m(\omega)-t_i\wedge\tau_m(\omega)$. For each $i\in[n-1]$,

$$\begin{aligned}
\mathbb{E}\{(W_{t_{i+1}\wedge\tau_m}-W_{t_i\wedge\tau_m})^2-(t_{i+1}\wedge\tau_m-t_i\wedge\tau_m)\}^2&\le\mathbb{E}\sup_{s\in[t_i,t_{i+1}]}\big\{(W_s-W_{t_i})^2-(s-t_i)\big\}^2\\
&\le 4\mathbb{E}\left\{(W_{t_{i+1}}-W_{t_i})^2-(t_{i+1}-t_i)\right\}^2\\
&\le 4(t_{i+1}-t_i)^2\mathbb{E}\{Z^2-1\}\\
&\le 8(t_{i+1}-t_i)^2,
\end{aligned}$$

where $Z\sim N(0,1)$ is an independent standard normal. Therefore

$$\mathbb{E}\sum_{i=1}^{n-1}V_i^2\le 8\sum_{i=1}^{n-1}(t_{i+1}-t_i)^2\le 8\max_{i\in[n-1]}|t_{i+1}-t_i|\underbrace{\left(\sum_{i=1}^{n-1}t_{i+1}-t_i\right)}_{=a}\to 0\text{ as }n\to\infty.$$

For the cross terms, we have, for $i<l\in[n-1]$,

$$\mathbb{E}V_iV_l=\mathbb{E}[V_i\underbrace{\mathbb{E}[V_l|\mathscr{F}_{t_{i+1}\wedge\tau_m}]}_{=0}]=0.$$

Thus, $\mathbb{E}\mathrm{Term2A}(\cdot,a)^2\to 0$ as desired.

Consider Term 2B. For $\omega\in\Omega,t\in[0,a]$,

$$\begin{aligned}
&|\mathrm{Term2B}(\omega,t)|\\
=&\sum_{i=1}^{n-1}(f''W(\omega,s_i(\omega))-f''W(\omega,t_i))\{W(\omega,t_{i+1}\wedge t\wedge\tau_m(\omega))-W(\omega,t_i\wedge t\wedge\tau_m(\omega))\}^2\\
\le&\sup_{i\in[n-1]}|f''W(\omega,s_i(\omega))-f''W(\omega,t_i)|\cdot\left(\sum_{i=1}^{n-1}\Big[\Big\{W(\omega,t_{i+1}\wedge t\wedge\tau_m(\omega))-W(\omega,t_i\wedge t\wedge\tau_m(\omega))\Big\}^2\right.\\
&\qquad\left.-(t_{i+1}\wedge t\wedge\tau_m(\omega)-t_i\wedge t\wedge\tau_m(\omega))\Big]+\underbrace{\sum_{i=1}^{n-1}(t_{i+1}\wedge t\wedge\tau_m(\omega)-t_i\wedge t\wedge\tau_m(\omega)}_{t\wedge\tau_m(\omega)}\right).
\end{aligned}$$

Therefore, we have that

$$\sup_{t\in[0,a]}|\mathrm{Term2B}(\omega,t)|\le\sup_{i\in[n-1]}\sup_{s\in[t_i,t_{i+1}]}|f''W(\omega,s)-f''W(\omega,t)|\cdot(\sup_{t\in[0,a]}M^{(n)}(\omega,t)+a) \qquad (\star\star\star)$$

where $M^{(n)}(\omega,t) := \sum_{i=1}^{n-1}\{W(\omega,t_{i+1}\wedge t) - W(\omega,t_i\wedge t)\}^2 - (t_{i+1}\wedge t - t_i \wedge t)$.

For every $\omega\in\Omega$, $t\mapsto f''W(\omega,t)$ is uniformly continuous on $[0,a]$. By similar logic that we applied to analyse Term2C, we have that

$$\lim_{n\to\infty}\sup_{i\in[n-1]}\sup_{s\in[t_i,t_{i+1}]}|f''W(\omega,s)-f''W(\omega,t_i)| = 0.$$

Moreover, in the proof of Theorem 8.3.2, we showed that $\lim_{n\to\infty}\mathbb{E}(M_a^{(n)})^2 = 0$. Since $M^{(n)}\in\mathcal{M}_a$, we have that $\lim_{n\to\infty}\mathbb{E}\sup_{t\in[0,a]}(M_t^{(n)})^2 = 0$ and that for a subsequence $n_1, n_2, \ldots$,

$$\lim_{n_k\to\infty}\sup_{t\in[0,a]} M^{(n_k)}(\omega,t) = 0 \text{ for a.e. } \omega\in\Omega.$$

Hence, for a.e. $\omega\in\Omega$, $(\star\star\star)\to 0$ as $n_k\to\infty$. Thus,

$$\lim_{n_k\to\infty}\sup_{t\in[0,a]}\left|\text{Term2}(\omega,t) - \frac{1}{2}\int_0^{t\wedge\tau_m(\omega)} f''W(\omega,s)\mathrm{d}s\right| = 0 \text{ for a.e. } \omega\in\Omega.$$

□

The following enhancement of Ito's lemma will be useful:

Theorem 9.3.2 (Ito's lemma II). Let $f:[0,\infty)\to\mathbb{R}$ be a bivariate function and suppose $f_t(t,x) := \frac{\partial f}{\partial t}(t,x)$, $f_x(t,x) := \frac{\partial f}{\partial x}(t,x)$, and $f_{xx}(t,x) := \frac{\partial^2 f}{\partial x^2}(t,x)$ exists $\forall t\geq 0$ and $x\in\mathbb{R}$. Let $\{W_t\}_{t>0}$ be the standard Brownian motion. We have that, for a.e. $\omega\in\Omega$,

$$f(t,W(\omega,t)) - f(0,W(\omega,0)) = \left(\int f_x(s,W_s)\mathrm{d}W_s\right)(\omega,t) + \int_0^t f_t(s,W(\omega,s))\mathrm{d}s + \frac{1}{2}\int_0^t f_{xx}(s,W(\omega,s))\mathrm{d}s. \tag{9.9}$$

As shorthand, we write

$$f(t,W_t) - f(0,0) = \int_0^t f_x(s,W_s)\mathrm{d}W_s + \int_0^t f_t(s,W_s)\mathrm{d}t + \frac{1}{2}\int_0^t f_{xx}(s,W_s)\mathrm{d}s.$$

Proof Sketch.

The proof closely follows that of Theorem 9.3.1. Letting $a>0$ and writing $0=t_1<t_2<\cdots<t_n=a$ as a partition of $[0,a]$, we have that

$$\begin{aligned}
f(a,W(\omega,a)) - f(0,W(\omega,0)) &= \sum_{i=1}^{n-1} f(t_{i+1},W(\omega,t_{i+1})) - f(t_i,W(\omega,t_i))\\
&= \sum_{i=1}^{n-1} f_x(t_i,W(\omega,t_i))\{W(\omega,t_{i+1})-W(\omega,t_i)\}\text{①}\\
&\qquad + f_t(t_i,W(\omega,t_i))(t_{i+1}-t_i)\text{②}\\
&\qquad + \frac{1}{2}f_{xx}(r_i(\omega),W(\omega,s_i(\omega)))\{W(\omega,t_{i+1})-W(\omega,t_i)\}^2\text{③}\\
&\qquad + f_{xt}(r_i(\omega),W(\omega,s_i(\omega)))\{W(\omega,t_{i+1})-W(\omega,t_i)\}(t_{i+1}-t_i)\text{④}\\
&\qquad + \frac{1}{2}f_{tt}(r_i(\omega),W(\omega,s_i(\omega)))(t_{i+1}-t_i)^2\text{⑤},
\end{aligned}$$

where $r_i(\omega)$, $s_i(\omega)\in[t_i,t_{i+1}]$, and where $f_{xt}(t,x) := \frac{\partial^2 f}{\partial x\partial t}(t,x)$, $f_{tt}(t,x) = \frac{\partial^2 f}{\partial t^2}(t,x)$.

Term ①, ②, ③ can be analyzed in a way similar to the proof of Theorem 9.3.1. Term ④, ⑤ are both of lower order and "converge" to 0. □

Corollary 9.3.1. Let $f : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ be a function that satisfy

$$\frac{\partial f}{\partial t}(t,x) = -\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t,x) \quad \forall t \geq 0, x \in \mathbb{R}. \tag{9.10}$$

In addition, suppose that, for some $a > 0$, $\mathbb{E}\int_0^a \frac{\partial f}{\partial x}(t, W_t)^2 \mathrm{d}t < \infty$. Then, we have that $\{f(t, W_t)\}_{t\in[0,a]}$ is a martingale.

Proof.
By Theorem 9.3.2, for a.e. $\omega \in \Omega$, $\forall t \geq 0$,

$$f(t, W(\omega, t)) - f(0,0) = \int \frac{\partial f}{\partial x}(s, W_s)\mathrm{d}W_s.$$

If, for some $a > 0$, $\mathbb{E}\int_0^a \frac{\partial f}{\partial x}(s, W_s)^2\mathrm{d}s < \infty$, then the function $(\omega, t) \mapsto \frac{\partial f}{\partial x}(t, W(\omega,t)) \in \mathcal{H}_a$ implies $\int \frac{\partial f}{\partial x}(s, W_s)\mathrm{d}W_s \in \mathcal{M}_a$ is a martingale. □

Example 9.3.2 (Gambler's Ruin). Define $X_t = \mu t + \sigma W_t$ for $t \geq 0$ as Brownian motion with drift. For $\alpha, \beta > 0$, define stopping time $\tau := \inf\{t \geq 0 : X_t = \alpha \text{ or } X_t = -\beta\}$. We want to compute $\mathbb{P}(X_\tau = \alpha)$.

Suppose there exists continuous $h : \mathbb{R} \to [0,1]$ such that $h(\alpha) = 1$, $h(-\beta) = 0$, and $\{h(X_t)\}_{t\geq 0}$ is a martingale, then

$$\begin{aligned}
\mathbb{P}(X_\tau = \alpha) &= \mathbb{E}h(X_\tau) \\
&= \lim_{n\to\infty} \mathbb{E}h(X_{\tau\wedge n}) && \text{(By DCT and continuity of } \{h(X_t)\}_{t\geq 0}) \\
&= \mathbb{E}h(X_0) = h(0).
\end{aligned}$$

To find a function h, write $f(t,x) := h(\mu t + \sigma x)$. We require $\frac{\partial f}{\partial t}(x,t) = -\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t,x)$, $\forall t \geq 0, x \in \mathbb{R}$ by Corollary 9.3.1. It implies

$$\begin{aligned}
\mu h'(\mu t + \sigma x) &= -\frac{1}{2}\sigma^2 h''(\mu t + \sigma x) \\
\implies -\frac{2\mu}{\sigma^2}h'(\cdot) &= h''(\cdot).
\end{aligned}$$

We also have boundary condition $h(\alpha) = 1$, $h(-\beta) = 0$. Solving the ODE, we get

$$h(z) = \frac{\exp(-\frac{2\mu z}{\sigma^2}) - \exp(\frac{2\mu\beta}{\sigma^2})}{\exp(-\frac{2\mu\alpha}{\sigma^2}) - \exp(\frac{2\mu\beta}{\sigma^2})}.$$

We may check that $\mathbb{E}\int_0^a \frac{\partial f}{\partial x}(t, W_t)^2\mathrm{d}t = \mathbb{E}\int_0^a \sigma^2 h'(X_t)^2\mathrm{d}t < \infty$, $\forall a > 0$, so that $\{h(X_t)\}_{t\geq 0}$ is a martingale. So we have

$$\mathbb{P}(X_\tau = \alpha) = h(0) = \frac{1 - \exp(\frac{2\mu\beta}{\sigma^2})}{\exp(-\frac{2\mu\alpha}{\sigma^2}) - \exp(\frac{2\mu\beta}{\sigma^2})} = \frac{\exp(-\frac{2\mu\beta}{\sigma^2}) - 1}{\exp(-\frac{2\mu(\alpha+\beta)}{\sigma^2}) - 1}.$$

Note, if we take limit $\mu \to 0$, we get

$$\lim_{\mu\to 0} \frac{\exp(-\frac{2\mu\beta}{\sigma^2}) - 1}{\exp(-\frac{2\mu(\alpha+\beta)}{\sigma^2}) - 1} = \lim_{\mu\to 0} \frac{-\frac{2\mu\beta}{\sigma^2}}{-\frac{2\mu(\alpha+\beta)}{\sigma^2}} = \frac{\beta}{\alpha+\beta}.$$

Now suppose $\mu < 0$ and define $\tau_n := \inf\{t \geq 0 : X_t = \alpha \text{ or } X_t = -n\}$.

Note that $\cup_{n\in\mathbb{N}}\{\omega : X_{\tau_n}(\omega) = \alpha\} = \{\omega : \sup_{t\geq 0} X(\omega, t) \geq \alpha\}$. Also $\{\omega : X_{\tau_n}(\omega) = \alpha\} \subseteq \{\omega : X_{\tau_{n+1}(\omega)} = \alpha\}$. Hence

$$\begin{aligned}
\mathbb{P}(\sup_{t\geq 0} X_t \geq \alpha) &= \lim_{n\to\infty} \mathbb{P}(X_{\tau_n} = \alpha) \\
&= \lim_{n\to\infty} \frac{\exp(-\frac{2\mu n}{\sigma^2}) - 1}{\exp(-\frac{2\mu(n+\alpha)}{\sigma^2}) - 1} \\
&= \exp(-\frac{2|\mu|\alpha}{\sigma^2}).
\end{aligned}$$

In other word, $Y := \sup_{t\geq 0} X_t \sim \mathrm{Exp}(-\frac{2|\mu|}{\sigma^2})$.

9.4 Stochastic Differential Equations

Definition 9.4.1 (Stochastic Differential Equation). Let $\{W_t\}_{t\geq 0}$ be standard Brownian motion. We say that a process $\{X_t\}_{t\geq 0}$ satisfy the Stochastic Differential Equation (SDE)

$$\mathrm{d}X_t = \phi(t, X_t)\mathrm{d}t + \psi(t, X_t)\mathrm{d}W_t \tag{9.11}$$

for some $\phi, \psi : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ if $\{X_t\}_{t\geq 0}$ satisfies the integral equation for a.e. $\omega \in \Omega$,

$$X(\omega, t) - X(\omega, 0) = \int_0^t \phi(s, X(\omega, s))\mathrm{d}s + \left(\int \psi(s, X_s)\mathrm{d}W_s\right)(\omega, t) \quad \forall t \geq 0. \tag{9.12}$$

Shorthand: $X_t - X_0 = \int_0^t \phi(s, X_s)\mathrm{d}s + \int_0^t \psi(s, X_s)\mathrm{d}W_s$. Note that X_0 can be arbitrary.

Note that we require $(\omega, s) \to \psi(s, X(\omega, s)) \in \mathscr{L}_{\mathrm{LOC}}$ and $s \mapsto \phi(s, X(\omega, s))$ to be integrable a.e. $\omega \in \Omega$. (That is, for a.e. $\omega \in \Omega$, $\forall t \geq 0$, $\int_0^t \psi(s, X(\omega, s))^2 ds < \infty$ and $\int_0^t |\phi(s, X(\omega, s))|\mathrm{d}s < \infty$.)

Example 9.4.1. (a) Consider the SDE

$$\mathrm{d}X_t = \mu\mathrm{d}t + \sigma\mathrm{d}W_t.$$

This implies

$$X_t - X_0 = \int_0^t \mu\mathrm{d}s + \int_0^t \sigma\mathrm{d}W_s = t\mu + \sigma(W_t - W_0) = t\mu + \sigma W_t$$

is Brownian motion with drift. X_0 is the random start location.

(b) Consider the SDE

$$\mathrm{d}X_t = \mu X_t\mathrm{d}t + \sigma X_t\mathrm{d}W_t.$$

This implies

$$X_t - X_0 = \int_0^t \mu X_s\mathrm{d}s + \int_0^t \sigma X_s\mathrm{d}W_s.$$

To solve the SDE, we guess that $X_t = X_0 f(t, W_t)$ for some $f : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ twice differentiable such that $f(0, 0) = 1$. By Ito's lemma, we have

$$\begin{aligned}
&X_t - X_0 = X_0 \int_0^t f_x(s, W_s)\mathrm{d}W_s + X_0 \int_0^t f_t(s, W_s)\mathrm{d}s + \frac{x_0}{2} \int_0^t f_{xx}(s, W_s)\mathrm{d}s \\
&\implies \mathrm{d}X_t = X_0 f_x(t, W_t)\mathrm{d}W_t + X_0 f_t(t, W_t)\mathrm{d}t + \frac{X_0}{2} f_{xx}(t, W_t)\mathrm{d}t.
\end{aligned}$$

By matching terms, we have

$$X_0 f_x(t, W_t) = \sigma X_t = \sigma X_0 f(t, W_t)$$

and

$$X_0 f_t(t, W_t) + \frac{X_0}{2} f_{xx}(t, W_t) = \mu X_t = \mu X_0 f(t, W_t).$$

Thus, we solve the deterministic PDE

$$f_x(t, x) = \sigma f(t, x) \quad \text{and} \quad f_t(t, x) + \frac{1}{2} f_{xx}(t, x) = \mu f(t, x) \quad \forall t \geq 0, x \in \mathbb{R}$$

with boundary condition $f(0, 0) = 1$, and we obtain $f(t, x) = \exp((\mu - \frac{1}{2}\sigma^2)t + \sigma x)$. Hence,

$$X_t = X_0 \exp\{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t\}.$$

We will see later that the solution is unique.

Definition 9.4.2 (Geometric Bronian Motion). Let the process $\{X_t\}_{t \geq 0}$ be

$$X_t = X_0 \exp\{(\mu - \frac{\sigma^2}{2})t + \sigma W_t\} \tag{9.13}$$

for some random variable X_0 and scalars $\mu \in \mathbb{R}$ and $\sigma > 0$ so that $\{X_t\}$ satisfy SDE

$$\mathrm{d}X_t = \mu X_t \mathrm{d}t + \sigma X_t \mathrm{d}W_t. \tag{9.14}$$

The process $\{X_t\}$ is known as geometric Brownian motion.

Remark 9.4.1. We call (9.13) geometric Brownian motion because for $s < t \in [0, \infty)$, we have

$$\log \frac{X_t}{X_s} = (\mu - \frac{\sigma^2}{2})(t - s) + \sigma(W_t - W_s)$$

so that $\{X_t\}$ exhibits geometric/exponential growth with variance.

Suppose $\mu > 0$ but $\mu < \frac{\sigma^2}{2}$, then we have that

$$\mathbb{E}X_t = \mathbb{E}X_0 \cdot \mathbb{E}_{\cdot|\mathscr{F}_0} e^{(\mu - \frac{\sigma^2}{2})t + \sigma W_t} = (\mathbb{E}X_0) \cdot e^{\mu t} \to \infty \text{ as } t \to \infty$$

but for a.e. $\omega \in \Omega$,

$$X_t(\omega) = X_0(\omega) \exp\left\{ t \left((\mu - \frac{\sigma^2}{2}) + \frac{\sigma W_t(\omega)}{t} \right) \right\} \to 0 \quad \text{as } t \to \infty$$

since $\frac{W_t(\omega)}{t} \to 0$ as $t \to \infty$ by Law of iterated logarithm.

Thus, if the growth rate μ is too small compared to volatility $\frac{\sigma^2}{2}$, there is no possibility of long term gain. This is called risk without possibility of reward.

Remark 9.4.2. Suppose process $\{X_t\}_{t \geq 0}$ satisfies $\mathrm{d}X_t = \mu X_t \mathrm{d}t + \sigma X_t \mathrm{d}W_t$, then for a.e. $\omega \in \Omega$, for all $t \geq 0$, $\Delta t \geq 0$,

$$X(\omega, t + \Delta t) - X(\omega, t) = \int_t^{t+\Delta t} \mu X(\omega, s)\mathrm{d}s + \left(\int_t^{t+\Delta t} \sigma X_s \mathrm{d}W_s \right)(\omega).$$

Note we regard $\left(\int_t^{t+\Delta t} \sigma X_s \mathrm{d}W_s\right)(\omega)$ as a shorthand of

$$\left(\int \sigma X_s \mathrm{d}W_s\right)(\omega, t+\Delta t) - \left(\int \sigma X_s \mathrm{d}W_s\right)(\omega, t) = \left(\int \sigma X_s \mathbb{1}_{\{s\geq t\}} \mathrm{d}W_s\right)(\omega, t+\Delta t).$$

Fix $\omega \in \Omega$, $t \geq 0$. If $s \mapsto X(\omega, s)$ is continuous on $[0,\infty)$, then

$$\begin{aligned}\left|\int_t^{t+\Delta t} \mu X(\omega,s)\mathrm{d}s - \mu X(\omega,t)\Delta t\right| &\leq \mu \int_t^{t+\Delta t} |X(\omega,s) - X(\omega,t)|\mathrm{d}s \\ &\leq \mu \cdot \Delta t \cdot \sup_{s\in[t,t+\Delta t]} |X(\omega,s) - X(\omega,t)| \to 0 \text{ as } \Delta t \to 0.\end{aligned}$$

Additionally, if $\mathbb{E}\int_t^{t+\Delta t} |X_s - X_t|^2 \mathrm{d}s \to 0$ as $\Delta t \to 0$, then by Ito isometry,

$$\begin{aligned}\mathbb{E}\left\{\int_t^{t+\Delta t} \sigma X_s \mathrm{d}W_s - \sigma X_t (W_{t+\Delta t} - W_t)\right\}^2 &= \mathbb{E}\left\{\int_t^{t+\Delta t} \sigma(X_s - X_t)\mathrm{d}W_s\right\}^2 \\ &= \mathbb{E}\int_t^{t+\Delta t} \sigma|X_s - X_t|^2 \mathrm{d}s \to 0 \text{ as } \Delta t \to 0.\end{aligned}$$

Also, by Jensen's inequality or Cauchy-Schwarz inequality,

$$\mathbb{E}\left(\int_t^{t+\Delta t} \mu X_s \mathrm{d}s - \mu X_t \Delta t\right)^2 \leq \mathbb{E}\mu\Delta t \cdot \int_t^{t+\Delta t} |X_s - X_t|^2 \mathrm{d}s \to 0 \quad \text{ as } \Delta t \to 0.$$

Therefore, for any $t \geq 0$, $\exists$ sequence $\Delta t \to 0$ such that for a.e. $\omega \in \Omega$,

$$|X(\omega, t+\Delta t) - X(\omega, t) - \{\mu \cdot \Delta t \cdot X(\omega, t) + \sigma \cdot X(\omega,t)\left(W(\omega, t+\Delta t) - W(\omega, t)\right)\}| \to 0.$$

and

$$\mathbb{E}\left|X_{t+\Delta t} - X_t - \{\mu X_t \Delta_t + \sigma X_t (W_{t+\Delta t} - W_t)\}\right|^2 \to 0.$$

Intuitively, we have that for small Δt,

$$X_{t+\Delta t} - X_t \approx X_t \cdot \{\mu \cdot \Delta t + \sigma(W_{t+\Delta t} - W_t)\}$$

Example 9.4.2 (Uhlenbeck-Ornstein Process). Consider SDE

$$\mathrm{d}X_t = -\alpha X_t \mathrm{d}t + \sigma \mathrm{d}W_t \text{ for } \alpha, \sigma > 0.$$

We see that $\{X_t\}_{t\geq 0}$ is mean-reverting in that $-\alpha X_t \geq 0$ if $X_t < 0$ and $-\alpha X_t \leq 0$ if $X_t \geq 0$.

We will see that the solution to this SDE is

$$X(\omega, t) = X(\omega, 0)e^{-\alpha t} + \sigma\left(\int e^{-\alpha(t-s)}\mathrm{d}W_s\right)(\omega, t).$$

Note:

- If X_0 is Gaussian or a constant, then $\{X_t\}_{t\geq 0}$ is a Gaussian process. Note that for a deterministic $f : [0,\infty) \to \mathbb{R}$ such that $\int_0^a f(s)^2 \mathrm{d}s < \infty$, $\forall a \geq 0$, the process $(\omega, t) \mapsto \int_0^t f_s \mathrm{d}W_s$ is a Gaussian process.
- This process $\{X_t\}_{t\geq 0}$ is not of the form $f(t, W_t)$ for some function f. Thus, we need another extension of Ito's lemma to handle a larger class of processes.

9.5 Ito Process

Definition 9.5.1 (Ito Process). Let process $\{F_t\}_{t\geq 0}$, $\{G_t\}_{t\geq 0}$ satisfy for a.e. $\omega\in\Omega$, $\forall t\geq 0$,

$$\int_0^t |F(\omega,s)|\mathrm{d}s<\infty \text{ and } \int_0^t |G(\omega,s)|^2\mathrm{d}s<\infty \quad (G\in\mathscr{L}_{\mathrm{LOC}}).$$

We say a process $\{X_t\}_{t\geq 0}$ is an <u>Ito process</u> if

$$X(\omega,t)-X(\omega,0)=\int_0^t F(\omega,s)\mathrm{d}s+\left(\int G_s\mathrm{d}W_s\right)(\omega,t). \tag{9.15}$$

(Note that if $f:[0,\infty)\times\mathbb{R}\to\mathbb{R}$ is continuously twice differentiable, then $\{f(t,W_t)-f(0,0)\}_{t\geq 0}$ is an Ito process by Theorem 9.3.2.)

For a process $\{H_t\}_{t\geq 0}$ satisfying for a.e. $\omega\in\Omega$, $\forall t\geq 0$,

$$\int_0^t |F(\omega,s)H(\omega,s)|\mathrm{d}s<\infty \text{ and } \int_0^t G(\omega,s)^2H(\omega,s)^2\mathrm{d}s<\infty,$$

we define $\int H_s\mathrm{d}X_s$ as a process such that

$$\left(\int H_s\mathrm{d}X_s\right)(\omega,t):=\int_0^t F(\omega,s)H(\omega,s)\mathrm{d}s+\left(\int G_sH_s\mathrm{d}W_s\right)(\omega,t).$$

Theorem 9.5.1 (Ito's lemma III). Let $\{X_t\}_{t\geq 0}$ be an Ito process with representation

$$X(\omega,t)=\int_0^t F(\omega,s)\mathrm{d}s+\left(\int G_s\mathrm{d}W_s\right)(\omega,t),$$

and let $f:[0,\infty)\times\mathbb{R}\to\mathbb{R}$ be continuously twice differentiable. Then, for a.e. $\omega\in\Omega$, $\forall t\geq 0$,

$$\begin{aligned} f(t,X(\omega,t))-f(0,0)=\int_0^t\frac{\partial f}{\partial t}(s,X(\omega,s))\mathrm{d}s+\frac{1}{2}\int_0^t\frac{\partial^2 f}{\partial x^2}(s,X(\omega,s))G(\omega,s)^2\mathrm{d}s\\ +\left(\int\frac{\partial f}{\partial x}(s,X_s)\mathrm{d}X_s\right)(\omega,t).\end{aligned} \tag{9.16}$$

Note the term $\left(\int\frac{\partial f}{\partial x}(s,X_s)\mathrm{d}X_s\right)(\omega,t)$ is in fact

$$\int_0^t\frac{\partial f}{\partial x}(s,X(\omega,s))F(\omega,s)\mathrm{d}s+\left(\frac{\partial f}{\partial x}(s,X_s)G_s\mathrm{d}W_s\right)(\omega,t).$$

Note that if $F=0$ and $G=1$, then $X(\omega,t)=W(\omega,t)$ and we obtain Theorem 9.3.2.

Proof Sketch.

We take similar approach as before. Let $a>0$ and let $0=t_1<t_2<\cdots<t_n=a$ be as sequence of partitions. Then,

$$\begin{aligned} f(a,X(\omega,a))-f(0,0)&=\sum_{i=1}^{n-1}f(t_{i+1},X(\omega,t_{i+1}))-f(t_i,X(\omega,t_i))\\ &=\sum_{i=1}^{n}\frac{\partial f}{\partial x}(t_i,X(\omega,t_i))\{X(\omega,t_{i+1})-X(\omega,t_i)\} \text{ ①}\\ &\quad+\frac{\partial f}{\partial t}(t_i,X(\omega,t_i))(t_{i+1}-t_i) \text{ ②}\\ &\quad+\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(r_i(\omega),X(\omega,s_i(\omega)))\{X(\omega,t_{i+1})-X(\omega,t_i)\}^2 \text{ ③}\\ &\quad+\text{ lower order terms for } s_i(\omega),r_i(\omega)\in[t_i,t_{i+1}].\end{aligned}$$

For term ①, note that

$$\begin{aligned}X(\omega,t_{i+1})-X(\omega,t_i) &= \int_{t_i}^{t_{i+1}} F(\omega,s)\mathrm{d}s+\left(\int_{t_i}^{t_{i+1}} G_s\mathrm{d}W_s\right)(\omega)\\ &\approx F(\omega,t_i)(t_{i+1}-t_i)+G(\omega,t_i)(W(\omega,t_{i+1})-W(\omega,t_i)).\end{aligned}$$

For term ③, we have the similarly that

$$\begin{aligned}(X(\omega,t_{i+1})-X(\omega,t_i))^2 \approx F(\omega,t_i)(t_{i+1}-t_i)^2 &+ 2F(\omega,t_i)G(\omega,t_i)(t_{i+1}-t_i)(W(\omega,t_{i+1})-W(\omega,t_i))\\ &+G(\omega,t_i)^2(W(\omega,t_{i+1})-W(\omega,t_i))^2.\end{aligned}$$

□

Example 9.5.1. Consider again the SDE

$$\mathrm{d}U_t=-\alpha U_t\mathrm{d}t+\sigma\mathrm{d}W_t.$$

Suppose $U_0=x_0\in\mathbb{R}$. We assume that $\{U_t\}$ is of the form $\{f(t,X_t)\}$ for some continuously twice differentiable $f:[0,\infty)\times\mathbb{R}\to\mathbb{R}$ and Ito process $\{X_t\}$. In addition, we will in particular guess that $f(t,x)=a(t)\{x_0+x\}$ and $X_t=\int_0^t b(s)\mathrm{d}W_s$ for some $a,b:[0,\infty)\to 0$. We require $a(0)=1$, a be continuously twice differentiable and b be square integrable.

By Theorem 9.5.1, we have

$$\begin{aligned}U_t-U_0=f(t,X_t)-f(0,X_0) &= \int_0^t \frac{\partial f}{\partial t}(s,X_s)\mathrm{d}s+\frac{1}{2}\int_0^t\frac{\partial^2 f}{\partial x^2}(s,X_s)b(s)^2\mathrm{d}s+\int_0^t\frac{\partial f}{\partial x}(s,X_s)\mathrm{d}X_s\\ &= \int_0^t a'(s)(x_0+X_s)\mathrm{d}s+\int_0^t\frac{\partial f}{\partial x}(s,X_s)b(s)\mathrm{d}W_s\\ &= \int_0^t a'(s)(x_0+X_s)\mathrm{d}s+\int_0^t a(s)b(s)\mathrm{d}W_s\end{aligned}$$

since

$$\frac{\partial f}{\partial t}(s,x)=a'(s)(x_0+x),\quad \frac{\partial f}{\partial x}(s,x)=a(s),\quad \frac{\partial^2 f}{\partial x^2}=0.$$

By coefficient matching, with SDE $\mathrm{d}U_t=-\alpha U_t\mathrm{d}t+\sigma\mathrm{d}W_t$, we have

$$\begin{aligned}a'(s)(x_0+X_s)=-\alpha U_s&=-\alpha a(s)(x_0+X_s),\\ a(s)b(s)&=\sigma\quad s\geq 0.\end{aligned}$$

By first equation, we get $a'(s)=-\alpha a(s)$, $\forall s\geq 0$ and using boundary condition $a(0)=1$, we get unique solution $a(s)=e^{-\alpha s}$. Thus, $b(s)=\sigma a(s)^{-1}=\sigma e^{\alpha s}$, which implies

$$U_t=e^{-\alpha t}\left(x_0+\int_0^t\sigma e^{\alpha s}\mathrm{d}W_s\right)=e^{-\alpha t}x_0+\sigma\int_0^t e^{-\alpha(t-s)}\mathrm{d}W_s.$$

Note:

(1) U_t is a Gaussian process (assuming U_0 is fixed or Gaussian).

(2) $\mathbb{E}U_t=e^{-\alpha t}x_0$ since $\int_0^t e^{-2\alpha(t-s)}\mathrm{d}s<\infty$ and so stochastic integral is a martingale and has mean 0 ($\mathbb{E}M_0=M_0=\int_0^0 e^{-\alpha(0-s)}\mathrm{d}W_s=0$).

(3) Assume $t > t' \in [0, \infty)$,

$$\begin{aligned}
\mathrm{Cov}(U_t, U_{t'}) &= \sigma^2 \mathbb{E} \int_0^t e^{-\alpha(t-s)} \mathrm{d}W_s \cdot \int_0^{t'} e^{-\alpha(t'-s)} \mathrm{d}W_s \\
&= \int_0^{t'} e^{-\alpha(t-s)} e^{-\alpha(t'-s)} \mathrm{d}s \qquad (*) \\
&= e^{-\alpha(t+t')} \int_0^{t'} e^{2\alpha s} \mathrm{d}s \\
&= \frac{1}{2\alpha} \left(e^{-\alpha(t-t')} + e^{-\alpha(t+t')} \right).
\end{aligned}$$

The $(*)$ holds by Ito Isometry, that is, for $H, G \in \mathcal{H}_a$, we have

$$\mathbb{E} \int H_s \mathrm{d}W_s \int G_s \mathrm{d}W_s = \langle H, G \rangle_{\mathcal{M}_a} = \langle H, G \rangle_{L_2(\mathbb{P} \times \mathrm{Leb})} = \mathbb{E} \int_0^a H_s G_s \mathrm{d}s.$$

Thus, UO process is not stationary but becomes stationary as $t \to \infty$. (Stationary here means the covariance $\{U_{t+s_1}, U_{t+s_2}, U_{t+s_3}\}$ does not depend on t.

Remark 9.5.1. Let $a > 0$ and $x_0 \in \mathbb{R}$ and consider the SDE

$$\mathrm{d}X_t = \phi(t, X_t)\mathrm{d}t + \psi(t, X_t)\mathrm{d}W_t \text{ for } t \in [0, a]$$

with initial condition $X_0 = x_0$.

Suppose $\exists K > 0$ such that $\forall t \in [0, a]$ and $x, y \in \mathbb{R}$,

$$|\phi(t, x) - \phi(t, y)|^2 + |\psi(t, x) - \psi(t, y)|^2 \leq K|x - y|^2 \text{ and } |\phi(t, x)|^2 + |\psi(t, x)|^2 \leq K(1 + |x|^2).$$

Then, $\exists$ solution $\{X_t\}_{t \geq 0}$ continuous, adapted, and $\sup_{t \in [0,a]} \mathbb{E} X_t^2 < \infty$ $(\star)$.

Moreover, if $\{Y_t\}_{t \geq 0}$ is another solution satisfy $(\star)$, then for a.e. $\omega \in \Omega$, $\sup_{t \in [0,a]} |X(\omega, t) - Y(\omega, t)| = 0$, See Steele (Theorem 9.1).

Thus, the solution of SDE exists and is unique.

9.6 Black-Scholes-Merton Theory

Definition 9.6.1. Let $\{\mathscr{F}_t\}_{t \geq 0}$ be a filtration and let the prices of commodities be represented as an adapted multivariate process State : $\Omega : \times [0, \infty) \to \mathbb{R}^d$.

For simplicity, we assume $d = 2$ and write

$$\mathrm{State}_t = (\mathrm{Stock}_t, \mathrm{Bond}_t) = (S_t, \beta_t).$$

(The names don't matter yet; we can also have $\mathrm{State}_t = (\mathrm{Bitcoin}_t, \mathrm{Gold}_t)$.)

For $T \geq 0$, define a contract/option with maturity time T as $h_T : \mathbb{R}^2 \to \mathbb{R}$ where $h_T(S_T, \beta_T)$ is the payout contract h_T.

For adapted process $a, b : \Omega \times [0, \infty) \to \mathbb{R}$, we define portfolio

$$V(\omega, t) = a(\omega, t)S(\omega, t) + b(\omega, t)\beta(\omega, t).$$

$a(\omega, t)$ and $b(\omega, t)$ are interpreted as amounts of stocks and bonds. We allow $a(\omega, t), b(\omega, t)$ to be negative to reflect short-selling.

Example 9.6.1. Suppose $t \in \{0, 1\}$.

At time $t = 0$, $S_0 = \$2$ and $\beta_0 = \$1$.

At time $t = 1$, $S_1 = \$4$ with probability p and $S_1 = \$1$ otherwise. And $\beta_1 = \$1$.

Define contract with maturity time $T = 1$ as

$$h(4, 1) = \$3, \quad h(1, 1) = \$1.$$

At time $t = 0$, how much should we pay to purchase this contract? Before 1970s, we would estimate $p := \mathbb{P}(S_1 = \$4)$ and pay expected value $4\hat{p} + 1(1 - \hat{p})$. This turns out to be the wrong approach.

Consider a portfolio with $a(\omega, t) = 1$ and $b(\omega, t) = -1$ for all $t \in \{0, 1\}$, $\omega \in \Omega$. The value of this portfolio at time $t = 1$ is, $\forall \omega \in \Omega$,

$$V(\omega, 1) = \begin{cases} \$3 & \text{if } S(\omega, 1) = 4 \\ \$0 & \text{if } S(\omega, 1) = 1 \end{cases}.$$

It implies $\forall \omega \in \Omega$, $V_1 = h(S_1, p_1)$. This is known as replicating portfolio. At time $t = 0$ it takes only $V_0 = \$1$ to acquire this portfolio, which implies the arbitrage-free price of contract h should be \$1, regardless of $\mathbb{P}(S_1 = 4)$.

If someone is willing to buy contract h at price higher than V_0, then one can get risk-free profit (arbitrage) by buying portfolio V and selling contract h. Likewise, if someone is selling contract h at a price lower than V_0, than there is arbitrage opportunity by buying h and selling V.

Consider example contract h:

$$h(4, 1) = -3, \quad h(1, 1) = 1.$$

We compute $a, b \in \mathbb{R}$ such that $aS_1 + b\beta_1 = h(S_1, \beta_1)$ a.s. It implies $4a + b = -3\text{W}$, $a + b = 1$. And therefore $a = -\frac{4}{3}$, $b = \frac{7}{3}$. Thus, the arbitrage-free price of h is $V_0 = aS_0 + b\beta_0 = \frac{1}{2}$.

Example 9.6.2. Now suppose $t \in [0, \infty)$ and assume that $\mathrm{d}S_t = \mu S_t \mathrm{d}t + \sigma S_t \mathrm{d}W_t$ (Geometric Brownian Motion) and $\mathrm{d}\beta_t = r\beta_t \mathrm{d}t$ (deterministic: $\beta_t = \beta_0 e^{rt}$) where $S_0, \beta_0 > 0$ and $\sigma, r > 0$.

Let h be a contract with maturity time $T > 0$ depending only on S_t. For example, we may take $h(S_T) := \max(S_T - K, 0)$ for some $K \geq 0$. (European call option).

How should we price h?

We construct a portfolio $V(\omega, t) = a(\omega, t)S(\omega, t) + b(\omega, t)\beta_t$ with 2 conditions:

(1) (Replicating) For a.e. $\omega \in \Omega$, $V(\omega, t) = h(S(\omega, t), \beta_t)$, $\forall t \in [0, T]$

(2) (Self-financing) $\mathrm{d}V_t = a_t \mathrm{d}S_t + b_t \mathrm{d}\beta_t$ (Note increase in stock share a_t must be offset by decrease in bond share b_t.)

Using the fact that $\{S_t\}$ is an Ito process with the representation,

$$S_t - S_0 = \int_0^t \mu S_s \mathrm{d}s + \int_0^t \sigma S_s \mathrm{d}W_s,$$

we have that

$$\begin{aligned} (2) \implies V_t - V_0 &= \int_0^t a_s \mathrm{d}S_t + \int_0^t b_t \mathrm{d}\beta_t \\ &= \int_0^t a_s \mu S_s \mathrm{d}s + \int_0^t a_s \sigma S_s \mathrm{d}W_s + \int_0^t b_s r \beta_s \mathrm{d}s \\ \implies \mathrm{d}V_t &= (\mu a_t S_t + r b_t \beta_t)\mathrm{d}t + \sigma a_t S_t \mathrm{d}W_t. \end{aligned}$$

To solve this SDE we guess $\{V_t\}$ is of the form $\{f(t, S_t)\}$ for $f : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ twice continuously-differentiable and obtain by Theorem 9.5.1 that

$$\begin{aligned} \mathrm{d}V_t &= \left(\frac{\partial f}{\partial t}(t, S_t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, S_t)\sigma^2 S_t^2 \right) \mathrm{d}t + \frac{\partial f}{\partial x}(t, S_t)\mathrm{d}S_t \\ &= \left(\frac{\partial f}{\partial t}(t, S_t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, S_t)\sigma^2 S_t^2 + \frac{\partial f}{\partial x}(t, S_t)\mu S_t \right) \mathrm{d}t + \frac{\partial f}{\partial x}(t, S_t)\sigma S_t \mathrm{d}W_t. \end{aligned}$$

By coefficient matching, we get $a_t = \frac{\partial f}{\partial x}(t, S_t)$ and hence

$$\mu \frac{\partial f}{\partial x}(t, S_t) S_t + r b_t \beta_t = \frac{\partial f}{\partial t}(t, S_t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, S_t)\sigma^2 S_t^2 + \frac{\partial f}{\partial x}(t, S_t)\mu S_t$$
$$\implies b_t = \frac{1}{r\beta_t}\left\{\frac{\partial f}{\partial t}(t, S_t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, S_t)\sigma^2 S_t^2\right\}$$

Since, by (1), we have that almost surely, $\forall t \in [0, T]$, $V_t = h(S_t, \beta_t) = f(t, S_t)$, it implies

$$f(t, S_t) = a_t S_t + b_t \beta_t = \frac{\partial f}{\partial x}(t, S_t) S_t + \frac{1}{r\beta_t}\left(\frac{\partial f}{\partial t}(t, S_t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, S_t)\sigma^2 S_t^2\right)\beta_t \text{ a.s.}$$

Thus, f must satisfy PDE, $\forall x \in \mathbb{R}, t \in [0, T]$,

$$f(t, x) = \frac{\partial f}{\partial x}(t, x)x + \frac{1}{r}\left(\frac{\partial f}{\partial t}(t, x) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, x)\sigma^2 x^2\right) \qquad \text{(Black-Scholes equation)}$$

with boundary condition $f(T, x) = h(x)$, $\forall x \in \mathbb{R}$.

Once we find a solution, we get the arbitrage free price of contract h at time $t = 0$ in $V_0 = f(0, S_0)$.

A few notes:

- For $h(S_T) = \max(S_T - K, 0)$, we obtain

$$V_t = S_t \Phi\left(\frac{\log(S_t/K) + (r + \frac{1}{2}\sigma^2)(T - t)}{\sigma\sqrt{T - t}}\right) + K e^{-r(T-t)} \cdot \Phi\left(\frac{\log(S_t/K) + (r - \frac{1}{2}\sigma^2)(T - t)}{\sigma\sqrt{T - t}}\right)$$
$$V_0 = S_0 \Phi\left(\frac{\log(S_0/K) + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) + K e^{-rT} \cdot \Phi\left(\frac{\log(S_0/K) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right)$$

  where Φ is the err function.

- Black-Scholes equation does not depend on μ.
- For other contract/options, Black-Scholes may be solved by numerical techniques such as Feynman-Kac theorem.

Note (Informal Overview). Ito's lemma II (Theorem 9.3.2):

$$f(t, W_t) - f(0, W_0) = \int_0^t \frac{\partial f}{\partial t}(s, W_s)\mathrm{d}s + \frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}(s, W_s)\mathrm{d}s + \int_0^t \frac{\partial s}{\partial x}(s, W_s)\mathrm{d}W_s$$
$$\implies \mathrm{d}f(t, W_t) = \left(\frac{\partial f}{\partial t}(t, W_t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, W_t)\right)\mathrm{d}t + \frac{\partial f}{\partial x}(t, W_t)\mathrm{d}W_t.$$

Now, X_t is Ito process if $\mathrm{d}X_t = F_t \mathrm{d}t + G_t \mathrm{d}W_t$. If

$$Y_t = \int_0^t H_s \mathrm{d}X_s := \int_0^t H_s F_s \mathrm{d}s + \int_0^t H_s G_s \mathrm{d}W_s,$$

then

$$\mathrm{d}Y_t = H_t \mathrm{d}X_t = H_t F_t \mathrm{d}t + H_t G_t \mathrm{d}W_t.$$

Ito's lemma III (Theorem 9.5.1): suppose $\mathrm{d}X_t = F_t \mathrm{d}t + G_t \mathrm{d}W_t$,

$$\mathrm{d}f(t, X_t) = \left(\frac{\partial f}{\partial t}(t, X_t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, X_t)G_t^2\right)\mathrm{d}t + \frac{\partial f}{\partial x}(t, X_t)\mathrm{d}X_t$$
$$= \left\{\frac{\partial f}{\partial t}(t, X_t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, X_t)G_t^2 + \frac{\partial f}{\partial x}(t, X_t)F_t\right\}\mathrm{d}t + \frac{\partial f}{\partial x}(t, X_t)G_t \mathrm{d}W_t.$$

Chapter 10

Uniform Central Limit Theorem

10.1 Donsker's Theorem

Definition 10.1.1. Let $C := C[0,1] = \{f \in \mathbb{R}^{[0,1]} : f \text{ continuous}\}$ and let $(C, \|\cdot\|)$ be the resulting complete and separable metric space. Define $\mathscr{C} := \mathscr{B}(C[0,1]) \subseteq 2^C$ as the Borel σ-field w.r.t $\|\cdot\|$. For $f \in C$ and $r > 0$, define the open ball $B(f,r) := \{g \in C : \|f-g\|_\infty < r\}$.

Proposition 10.1.1. Define $\mathscr{C}_{\text{ball}} := \sigma\{B(f,r) : f \in C, r > 0\}$ as the ball σ-field. Then we have $\mathscr{C} = \mathscr{C}_{\text{ball}}$.

Proof.

It is clear that $\mathscr{C}_{\text{ball}} \subseteq \mathscr{C}$. To show the other direction, we claim that if $A \subseteq \mathscr{C}$ is open, then A is countable union of open balls.

Since $(C, \|\cdot\|)$ is separable, there exists countable Γ and $\{f_\gamma\}_{\gamma\in\Gamma} \in C$ such that $\overline{\{f_\gamma\}_{\gamma\in\Gamma}} = C$. Let $A \subseteq C$ be open and for each $f_\gamma \in A$, define $r_\gamma := \sup\{r > 0 : B(f_\gamma, r) \subseteq A\}$. (Note $B(f_\gamma, r_\gamma) \subseteq A$).

We claim that $A = \cup_{f_\gamma\in A} B(f_\gamma, r_\gamma)$. To see this, let $g \in A$ and $r > 0$ such that $B(g,r) \subseteq A$. There exists $f_\gamma \in B(g, \frac{r}{2})$ and thus, $B(f_\gamma, \frac{r}{2}) \subseteq B(g,r) \subseteq A \implies r_\gamma \geq \frac{r}{2}$. Hence, $g \in B(f_\gamma, \frac{r}{2}) \subseteq B(f_\gamma, r_\gamma)$. Since g is arbitrary, the claim follows. $\square$

Corollary 10.1.1. $C \cap \mathscr{B}(\mathbb{R})^{\otimes[0,1]} = \mathscr{C}$.

Proof.

Since evaluation functional is continuous in $(C, \|\cdot\|_\infty)$, it is $\mathscr{C}/\mathscr{B}(\mathbb{R})$–measurable and hence, $C \cap \mathscr{B}(\mathbb{R})^{\otimes[0,1]} \subseteq \mathscr{C}$.

To show the other direction, observe that $\forall f \in C$, $\gamma > 0$,

$$\begin{aligned}
B(f,\gamma) &= \{g \in C : \sup_{x\in\mathbb{Q}\cap[0,1]} |g(x) - f(x)| < \gamma\} \\
&= \cap_{x\in\mathbb{Q}\cap[0,1]} \{g \in C : g(x) \in (f(x)-\gamma, f(x)+\gamma)\} \\
&= \cap_{x\in\mathbb{Q}\cap[0,1]} \underbrace{e_x^{-1}((f(x)-\gamma, f(x)+\gamma))}_{\in C\cap\mathscr{B}(\mathbb{R})^{\otimes[0,1]}}.
\end{aligned}$$

Combining the result of the proposition, it implies $\mathscr{C} \subseteq C \cap \mathscr{B}(\mathbb{R})^{\otimes[0,1]}$. $\square$

Definition 10.1.2. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. For $n \in \mathbb{N}$, let $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ be iid random variables. Define the <u>scaled random walk</u> $F_n : \Omega \to C$ as, $\forall \omega \in \Omega$, $\forall t \in [0,1]$,

$$F_n(\omega, t) := \frac{1}{\sqrt{n}} \sum_{i=1}^{k-1} X_i(\omega) + \left(t - \frac{k-1}{n}\right) X_k(\omega)\sqrt{n} \tag{10.1}$$

where $k = \lceil tn \rceil$ so that $\frac{k-1}{n} \le t \le \frac{k}{n}$.

Note that $F_n(\omega, 0) = 0$, $F_n(\omega, \frac{k}{n}) = \frac{1}{\sqrt{n}} \sum_{i=1}^k X_i(\omega)$, $\forall k \in [n]$. F_n is continuous, and we show that it is $\mathscr{F}/\mathscr{C}$-measurable.

Note that for any $\{t_1, t_2, \ldots, t_m\} \subseteq [0,1]$, $(F_n(\cdot, t_1), \ldots, F_n(\cdot, t_n)) : \Omega \to \mathbb{R}^m$ is a random vector and thus, F_n is $\mathscr{F}/C \cap \mathscr{B}(\mathbb{R})^{\otimes[0,1]}$-measurable, which implies $\mathscr{F}/\mathscr{C}$-measurable.

Remark 10.1.1. For a matrix space $(\mathcal{X}, d)$ with Borel σ-field $\mathscr{B}(\mathcal{X})$, define

$$\mathcal{P}(\mathcal{X}) := \{\text{probability measures on } (\mathcal{X}, \mathscr{B}(\mathcal{X}))\}.$$

We say a sequence $\{P_n\}_{n\in\mathbb{N}}$ in $\mathcal{P}(\mathcal{X})$ converges weakly to $P \in \mathcal{P}(\mathcal{X})$ if

$$\int f \mathrm{d}P_n \to \int f \mathrm{d}P, \quad \forall f : \mathcal{X} \to \mathbb{R} \text{ continuous and bounded. (Equivalently, Lipschitz and bounded)}$$

We say a sequence of random objects $X_n : \Omega \to \mathcal{X}$ converges in distribution to random object $X : \Omega \to \mathcal{X}$ if $\mathbb{P}^{(X_n)} \xrightarrow{w} \mathbb{P}^{(X)}$, i.e.

$$\mathbb{E}f(X_n) \to \mathbb{E}f(X) \quad \forall f : \mathcal{X} \to \mathbb{R} \text{ continuous and bounded.}$$

Donsker's Theorem states that with $(\mathcal{X}, \mathscr{B}(\mathcal{X})) = (C, \mathscr{C})$, with $X_1, \ldots, X_n$ as mean-zero, unit-variance iid random variables, we have $F_n \xrightarrow{d} W$ where $\{W_t\}_{t\in[0,1]}$ is the standard Brownian Motion. In particular, since evaluation functional is 1-Lipschitz,

$$(F_n(\cdot, t_1), \ldots, F(\cdot, t_m)) \xrightarrow{d} (W_{t_1}, \ldots, W_{t_m}) \quad \forall \{t_1, \ldots, t_m\} \subseteq [0,1]$$

by continuous mapping Theorem (Theorem 5.1.2).

Thus, by CLT, we see that if $\mathbb{P}^{(F_n)}$ has a weak limit, the limit has to be standard Brownian motion. The challenge is to establish the limit. As another example, the max-functional $\phi(f) = \sup_{x\in[0,1]} |f(x)|$ is continuous, which implies

$$\sup_{t\in[0,1]} |F_n(\cdot, t)| \xrightarrow{d} \sup_{t\in[0,1]} W_t \stackrel{d}{=} |W_1|.$$

Example 10.1.1. Let $X_1, X_2, \ldots, X_n$ be iid $N(0,1)$ random variable. Define process $\widetilde{F}_n : \Omega \to C[0,1]$ such that

$$\widetilde{F}_n(\omega, t) = n\left(\frac{k}{n} - t\right) X_{k-1}(\omega) + n\left(t - \frac{k-1}{n}\right) X_k(\omega)$$

where $k = \lceil nt \rceil$ and where we define $X_0 = 0$.

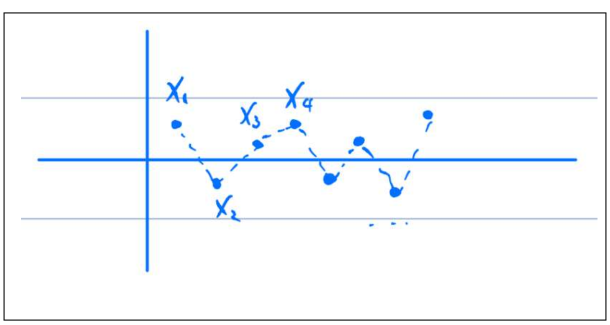


Figure 10.1: Scaled-random walk with iid Normals

Then $\{\mathbb{P}^{(\widetilde{F}_n)}\}_{n\in\mathbb{N}}$ does not have a weak limit.

In particular, for $s, t \in [0,1]$, $\mathrm{Cov}(\widetilde{F}_n(\cdot, s), \widetilde{F}_n(\cdot, t)) \to \mathbb{1}_{\{s=t\}}$. One may think of $\widetilde{F}_n$ as "tending" to Gaussian white noise, but Gaussian white noise is not continuous ($\notin \mathcal{P}(C)$).

Definition 10.1.3. Let $(\mathcal{X}, d)$ be a metric space with Borel σ-field $\mathscr{B}(\mathcal{X})$ We say $K \subseteq \mathcal{X}$ is <u>compact</u> if for any sequence $\{x_n\}_{n\in\mathbb{N}} \subseteq K$, there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ and a $x \in K$ such that $d(x_{n_k}, x) \to 0$ as $k \to \infty$.

Equivalently, K is compact if and only if K is complete and totally bounded (i.e. $\forall \varepsilon > 0\ \exists x_1, \dots, x_n \in \mathcal{X}$ s.t. $K \subseteq \cup_{i=1}^n B(x_i, \varepsilon)$).

Let $\{P_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}(\mathcal{X})$. We say that $\{P_n\}_{n\in\mathbb{N}}$ is <u>tight</u> if $\forall \varepsilon > 0$, $\exists K \subseteq \mathcal{X}$ compact such that $\inf_{n\in\mathbb{N}} P_n(K) > 1 - \varepsilon$. Note by Theorem 4.4.2 that any finite subset $\mathcal{Q} \subseteq \mathcal{P}(\mathcal{X})$ is tight if $\mathcal{X}$ is complete and separable.

Theorem 10.1.1 (Prokhorov). Let $(\mathcal{X}, d)$ be a separable space. If a sequence $\{P_n\}_{n\in\mathbb{N}} \subseteq \mathcal{P}(\mathcal{X})$ is tight, then $\exists$ a subsequence $\{n_1, n_2, \dots\}$ and $P \in \mathcal{P}(\mathcal{X})$ such that $P_{n_k} \overset{w}{\to} P$ as $k \to \infty$.

Suppose $(\mathcal{X}, d)$ be complete and separable metric space. Define the Levy-Prokhorov metric

$$d_{\mathrm{LP}}(P, Q) := \inf\{\varepsilon > 0 : P(A) \le Q(A^\varepsilon) + \varepsilon \text{ and } Q(A) \le P(A^\varepsilon) + \varepsilon, \forall A \in \mathscr{B}(\mathcal{X})\},$$

where $A^\varepsilon := \{x \in \mathcal{X} : d(x, A) \le \varepsilon\}$. Then we have that

(1) $(\mathcal{P}(\mathcal{X}), d_{\mathrm{LP}})$ is a metric space,

(2) $P_n \overset{w}{\to} P$ iff $d_{\mathrm{LP}}(P_n, P) \to 0$,

(3) $\mathcal{Q} \subseteq \mathcal{P}(\mathcal{X})$ is compact iff it is tight.

Proof.

The proof of the Theorem 10.1.1 requires more advanced functional analysis. See note by Van Gaans. $\square$

Theorem 10.1.2 (Arzela-Ascoli). Let $A \subseteq C[0,1]$. Then $\bar{A}$ is compact with respect to $\|\cdot\|_\infty$ metric if and only if

(a) $\sup_{f\in A} |f(0)| < \infty$ and,

(b) (Uniform equicontinuity) $\lim_{\delta\to 0} \sup_{f\in A} w(f, \delta) = 0$, where $w(f, \delta) := \sup\{|f(s) - f(t)| : s, t \in [0,1], |s - t| < \delta\}$ is the modulus of continuity.

Proof.

We will prove one direction by showing that, for any sequence $\{f_n\}$ in A, there exists $\{n_1, n_2, \dots\}$ and $f \in \bar{A}$ such that $\|f_{n_k} - f\|_\infty \to 0$ as $k \to \infty$.

Step 0.5: Assume that $A \subseteq C[0,1]$ satisfy (a) + (b). Choose $k \in \mathbb{N}$ such that $\sup_{f\in A} w(f, \frac{1}{k}) := C < \infty$. Then, for any $f \in A$, for any $t \in [0,1]$,

$$\begin{aligned} |f(t)| \le& |f(t) - f(0)| + |f(0)| \\ \le& \sum_{i=1}^k \left| f(\frac{i-1}{k}t) - f(\frac{i}{k}t) \right| + f(0) \\ \le& kC + |f(0)|. \end{aligned}$$

Thus, $\sup_{f\in A} \sup_{t\in[0,1]} |f(t)| \le kC + \sup_{f\in A} |f(0)| < \infty$.

Step 1: Let $\{x_m\}_{m\in\mathbb{N}}$ be an enumeration of $\mathbb{Q} \cap [0,1]$. The sequence $\{f_1(x_1), f_2(x_1), \dots\}$ is bounded by Step 0.5 and thus, there exists subsequence $\{n_1^{(1)}, n_2^{(1)}, \dots,\}$ and $y_1 \in \mathbb{R}$ such that $f_{n_k^{(1)}}(x_1) \to y_1$. Since $\{f_{n_1^{(1)}}(x_2), f_{n_2^{(1)}}(x_2), \dots\}$ is bounded, there exists subsequence $\{n_1^{(2)}, n_2^{(2)}, \dots\} \subseteq \{n_1^{(1)}, n_2^{(1)}, \dots\}$ and $y_2 \in \mathbb{R}$ such that $f_{n_k^{(2)}}(x_2) \to y_2$ and $f_{n_k^{(2)}}(x_1) \to y_1$. Thus, for any $m \in \mathbb{N}$, there exists subsequence $\{n_1^{(m)}, n_2^{(m)}, \dots\}$ and $y_m \in \mathbb{R}$ such that $f_{n_k^{(m)}}(x_1) \to y_1, \dots, f_{n_k^{(m)}}(x_m) \to y_m$. Define $g_k := f_{n_k^{(k)}}$. Then, $\forall m \in \mathbb{N}$, $\lim_{k\to\infty} g_k(x_m) = y_m$.

Step 2: We will show that $\{g_k\}_{k\in\mathbb{N}}$ is Cauchy w.r.t. $\|\cdot\|_\infty$, i.e., $\lim_{n,m\to\infty}\|g_n - g_m\|_\infty = 0$. Theorem follows since C is complete. To this end, let $\varepsilon > 0$, let $\delta > 0$ be such that $\sup_{f\in A} W(f,\delta) \le \frac{\varepsilon}{3}$. Let $0 \le t_0 < t_1 < \cdots < t_m \le 1$ be such that $t_i \in \mathbb{Q}\cap[0,1]$ and $\max_{i\in[m]}|t_{i+1}-t_i| \le \delta$.

There exists $n_0 \in \mathbb{N}$ such that $\forall n, m \ge n_0$, we have

$$\max_{i=0,\dots,m} |g_n(t_i) - g_m(t_i)| \le \frac{\varepsilon}{3}.$$

Then, for any $x \in [0,1]$, $\forall n, m \ge n_0$,

$$\begin{aligned} |g_n(x) - g_m(x)| &\le |g_n(t_i) - g_n(x)| + |g_m(t_i) - g_m(x)| + |g_n(t_i) - g_m(t_i)| && \text{(where } x \in [t_i, t_{i+1}]) \\ &\le \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. && \text{(by } \textstyle\sup_{f\in A} w(f,\delta) \le \frac{\varepsilon}{3}) \end{aligned}$$

So $\|g_n - g_m\|_\infty \to 0$ as $n, m \to \infty$. Since $(C[0,1], \|\cdot\|_\infty)$ is complete, the claim follows.

For the other direction, suppose $A \subseteq C$ such that $\bar{A}$ is compact. Since $e_0 := f \mapsto f(0)$ is continuous, $\sup_{f\in\bar{A}}|e_0(f)| = \sup_{f\in\bar{A}}|f(0)| < \infty$. Now, let $\{\delta_1, \delta_2, \dots\}$ be such that $\delta_k \to 0$. Define $\phi_k := f \mapsto w(f, \delta_k)$. Then, the facts that

(i) $\forall f \in \bar{A}$, $\phi_k(f) \to 0$,

(ii) $\phi_1 \ge \phi_2 \ge \dots$,

(iii) $\phi_k : C[0,1] \to [0,\infty)$ is a continuous functional,

implies $\lim_{k\to\infty}\sup_{f\in\bar{A}}\phi_k(f) = 0$ by Dini's Theorem. □

Corollary 10.1.2. Let $G_1, G_2, \dots$ be random processes taking value in $C[0,1]$ (i.e. $G_n : \Omega \to C[0,1]$), $\{\mathbb{P}^{(G_n)}\}_{n\in\mathbb{N}}$ is tight iff the following two conditions (i) and (ii) hold:

(1) $\forall \eta > 0$, $\exists a > 0, n_0 \in \mathbb{N}$ such that

$$\sup_{n\ge n_0} \mathbb{P}(|G_n(\cdot,0)| > a) \le \eta,$$

(2) $\forall \eta > 0, \varepsilon > 0$, $\exists \delta > 0, n_0 \in \mathbb{N}$ such that

$$\sup_{n\ge n_0} \mathbb{P}(w(G_n,\delta) > \varepsilon) \le \eta.$$

Proof.

$\Longrightarrow$: Suppose $\{\mathbb{P}^{(G_n)}\}_{n\in\mathbb{N}}$ is tight. Fix $\eta > 0$ and let $K \subseteq C[0,1]$ be a compact set such that $\inf_{n\in\mathbb{N}}\mathbb{P}(G_n \in K) \ge 1-\eta$. By Theorem 10.1.2, $\exists a > 0$ such that $|f(0)| \le a$, $\forall f \in K$. So $\forall n \in \mathbb{N}$,

$$\mathbb{P}(|G_n(\cdot,0)| > a) \le \mathbb{P}(G_n \notin K) \le \eta.$$

Similarly, fix $\varepsilon > 0$, $\exists \delta > 0$ such that $w(f,\delta) \le \varepsilon$, $\forall f \in K$ by Theorem 10.1.2. So $\forall n \in \mathbb{N}$,

$$\mathbb{P}(w(G_n,\delta) > \varepsilon) \le \mathbb{P}(G_n \notin K) \le \eta.$$

$\Longleftarrow$: Now suppose $\{\mathbb{P}^{(G_n)}\}_{n\in\mathbb{N}}$ satisfy (1) and (2).

Since any finite set of probability measures is tight, we may increase a and decrease δ if necessary to assume that $n_0 = 1$ in conditions (1) and (2). Fix $\eta > 0$, then there exists $a > 0$ such that writing $B := \{f \in C : |f(0)| \le a\}$, it holds that $\inf_{n\in\mathbb{N}}\mathbb{P}(G_n \in B) \ge 1-\eta/2$ by (1).

Also, for any $k \in \mathbb{N}$, there exists $\delta_k > 0$ such that, writing $B_k := \{f \in C[0,1] : w(f,\delta_k) < \frac{1}{k}\}$, we have $\inf_{n\in\mathbb{N}}\mathbb{P}(G_n \in B_k) \ge 1 - \frac{\eta}{2}2^{-k}$ by (2).

Define $A := B \cap (\cap_{k\in\mathbb{N}} B_k)$. Since A satisfy (a), (b) in Theorem 10.1.2, $\bar{A}$ is compact. Note that $\sup_{n\in\mathbb{N}}\mathbb{P}(G_n \notin \bar{A}) \le \frac{\eta}{2} + \sum_{k=1}^{\infty}\frac{\eta}{2}2^{-k} \le \eta$. It implies $\{\mathbb{P}^{(G_n)}\}$ is tight. □

Lemma 10.1.1. Let $0 = t_0 < t_1 < t_2 < \dots t_L = 1$ such that $\min_{i\in[L]} |t_i - t_{i-1}| \geq \delta$. Then, $\forall f \in C$,

$$w(f,\delta) \leq 3 \max_{i\in[L]} \sup_{s\in(t_{i-1},t_i]} |f(s) - f(t_{i-1})|.$$

Proof.

Let $s,t \in [0,1]$ be such that $|s-t| < \delta$. Then either s,t are in the same bin or in adjacent bins. If $\exists i \in \lceil L \rceil$ such that $s,t \in (t_{i-1}, t_i]$, then

$$|f(s) - f(t)| \leq |f(s) - f(t_{i-1})| + |f(t) - f(t_{i-1})| \leq 2 \sup_{s\in(t_{i-1},t_i]} |f(s) - f(t_{i-1})|.$$

The other cases follows similarly. □

Theorem 10.1.3 (Donsker). Let $X_1, X_2, \dots$ be iid random variables with $\mathbb{E}X_1 = 0$ and $\mathbb{E}X_1^2 = 1$. Define, for $n \in \mathbb{N}$, F_n be the scaled random walk, that is, define $F_n : \Omega \to C$ as, $\forall \omega \in \Omega$, $\forall t \in [0,1]$,

$$F_n(\omega,t) := \frac{1}{\sqrt{n}} \sum_{i=1}^{k-1} X_i(\omega) + \left(t - \frac{k-1}{n}\right) X_k(\omega)\sqrt{n}$$

where $k = \lceil tn \rceil$ so that $\frac{k-1}{n} \leq t \leq \frac{k}{n}$. Then $F_n \xrightarrow{d} W$, (i.e., $\mathbb{P}^{(F_n)} \xrightarrow{w} \mathbb{P}^{(W)}$).

Proof.

Step1: Define, for $k \in \mathbb{N}$, such that $S_k = \sum_{i=1}^k X_i$. Assume

$$\lim_{\lambda\to\infty} \limsup_{m\to\infty} \lambda^2 \mathbb{P}(\max_{k\in[m]} |S_k| \geq \lambda\sqrt{m}) = 0. \qquad (\star)$$

We will show that $\{\mathbb{P}^{(F_n)}\}_{n\in\mathbb{N}}$ is tight. Since $F_n(\cdot,0) = 0$, we need only show that, $\forall \varepsilon > 0$,

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}(w(F_n,\delta) > \varepsilon) = 0. \qquad \text{(Corollary 10.1.2)}$$

To that end, fix $\varepsilon > 0$. For any $\delta > 0$, integer $n \geq \frac{1}{2\delta}$, let $m \in \mathbb{N}$ be such that

$$\frac{1}{2\delta} \leq \frac{n}{m} := L \leq \frac{1}{\delta}$$

so that $\frac{m}{n} \geq \delta$. (For any fixed δ, $n \to \infty \iff m \to \infty$).

Define $t_\ell := \ell \cdot \frac{m}{n}$ for $\ell = 0, 1, \dots, L$. By Lemma 10.1.1,

$$\begin{aligned}
\mathbb{P}(w(F_n,\delta) > \varepsilon) &\leq \mathbb{P}\left(\max_{\ell\in[L]} \sup_{s\in(t_{\ell-1},t_\ell]} |F_n(\cdot,s) - F_n(\cdot,t_{\ell-1})| > \frac{\varepsilon}{3}\right)\\
&\leq \sum_{\ell=1}^{L} \mathbb{P}\left(\sup_{s\in(t_{\ell-1},t_\ell]} |F_n(\cdot,s) - \frac{1}{\sqrt{n}} S_{(\ell-1)m}| > \frac{\varepsilon}{3}\right)\\
&\leq L\mathbb{P}\left(\max_{k\in[m]} |S_k| \geq \frac{\varepsilon}{3}\sqrt{n}\right)\\
&\leq \frac{1}{\delta}\mathbb{P}\left(\max_{k\in[m]} |S_k| \geq \frac{\varepsilon}{3}\sqrt{\frac{m}{2\delta}}\right) \qquad \text{(write } \lambda := \tfrac{\varepsilon}{3\sqrt{2\delta}}\text{)}\\
&\leq \frac{18}{\varepsilon^2} \lambda^2 \mathbb{P}(\max_{k\in[m]} |S_k| \geq \lambda\sqrt{m}).
\end{aligned}$$

Thus,

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}(w(F_n,\delta) > \varepsilon) \leq \lim_{\lambda\to\infty} \limsup_{m\to\infty} \frac{18}{\varepsilon^2} \lambda^2 \mathbb{P}(\max_{k\in[m]} |S_k| \geq \lambda\sqrt{m}) = 0.$$

So $\{\mathbb{P}^{(F_n)}\}_{n\in\mathbb{N}}$ is tight.

Step 2: we prove $(\star)$. By Etemadi's inequality (Lemma 10.1.2), for any $m \in \mathbb{N}$, $\lambda > 0$, we have

$$\mathbb{P}(\max_{k\in[m]} |S_k| \geq \lambda\sqrt{m}) \leq 3 \max_{k\in[m]} \mathbb{P}(|S_k| \geq \frac{\lambda}{3}\sqrt{m}). \tag{$\star\star$}$$

For each $k \in [m]$, we have $\mathbb{P}(|S_k| \geq \frac{\lambda}{3}\sqrt{m}) \leq \frac{k}{\lambda^2 m}$. Also, since $\frac{1}{\sqrt{k}} S_k \xrightarrow{d} Z \sim N(0,1)$ as $k \to \infty$, we have Portmanteau theorem that

$$\limsup_{k\to\infty} \mathbb{P}(|S_k| \geq \frac{\lambda}{3}\sqrt{k}) \leq \mathbb{P}(|Z| \geq \frac{\lambda}{3}) \leq e^{-\frac{1}{18}\lambda^2}.$$

It implies for any $\lambda > 0$, there exists $k_\lambda \in \mathbb{N}$ such that $\forall k \geq k_\lambda$,

$$\mathbb{P}(|S_k| \geq \frac{\lambda}{3}\sqrt{m}) \leq \mathbb{P}(|S_k| \geq \frac{\lambda}{3}\sqrt{k}) \leq 2e^{-\frac{1}{18}\lambda^2}.$$

We then have that

$$(\star\star) \leq \underbrace{3 \max_{k\in[m]} \left(\frac{k}{\lambda^2 m} \mathbb{1}_{\{k<k_\lambda\}} + 2e^{-\frac{1}{18}\lambda^2} \mathbb{1}_{\{k\geq k_\lambda\}} \right)}_{\text{Bound}(\lambda,m)}.$$

$(\star)$ follows, because

$$\lim_{\lambda\to\infty} \limsup_{m\to\infty} \lambda^2 \cdot \text{Bound}(\lambda, m) = 0.$$

Step 3: Suppose $\mathbb{P}^{(F_n)} \stackrel{w}{\not\to} \mathbb{P}^{(W)}$. Then $\exists \phi : C[0,1] \to \mathbb{R}$ Lipschitz such that $\mathbb{E}\phi(F_n) \not\to \mathbb{E}\phi(W)$. Then $\exists$ subsequence $\{n_1, n_2, \dots\}$ and $\varepsilon > 0$ such that $|\mathbb{E}\phi(F_{n_k}) - \mathbb{E}\phi(W)| > \varepsilon$, $\forall k \in \mathbb{N}$. But $\{\mathbb{P}^{(F_{n_k})}\}_{k\in\mathbb{N}}$ is tight implies $\exists$ a further subsequence $\{n'_1, n'_2, \dots\}$ and $P \in \mathcal{P}(C)$ such that $\mathbb{P}^{(F_{n'_k})} \xrightarrow{w} P$. P must be $\mathbb{P}^{(W)}$. Contradiction. It must be $\mathbb{P}^{(F_n)} \xrightarrow{w} \mathbb{P}^{(W)}$. ☐

Lemma 10.1.2. (Etemadi's Inequality)
Let $X_1, \dots, X_n$ be independent random variables. Write $S_k := \sum_{i=1}^k X_i$. Then, for any $t > 0$, we have that

$$\mathbb{P}\left(\max_{k\in[n]} |S_k| \geq t\right) \leq 3 \max_{k\in[n]} \mathbb{P}(|S_k| \geq t/3).$$

Proof.

For any $k \in [n]$, define the event

$$A_k := \left\{ k = \min\{j \in [n] : |S_j| \geq t\} \right\},$$

where the minimum over an empty set is defined to be infinity. It is clear that $A_1, \dots A_n$ are disjoint and that $A := \cup_{k=1}^n A_k = \left\{ \max_{k\in[n]} |S_k| \geq t \right\}$.

Therefore, we have that

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap \{|S_n| \geq t/3\}) + \mathbb{P}(A \cap \{|S_n| < t/3\}) \\ &= \mathbb{P}(|S_n| \geq t/3) + \sum_{k=1}^n \mathbb{P}(A_k \cap \{|S_n| < t/3\}) \end{aligned} \tag{10.2}$$

Let us focus on the second term:

$$\begin{aligned}\sum_{k=1}^{n}\mathbb{P}(A_k\cap\{|S_n|<t/3\}) &\le \sum_{k=1}^{n}\mathbb{P}(A_k\cap\{|S_n-S_k|>2t/3\})\\ &=\sum_{k=1}^{n}\mathbb{P}(A_k)\mathbb{P}(\{|S_n-S_k|>2t/3\})\\ &\le\sum_{k=1}^{n}\mathbb{P}(A_k)\big\{\mathbb{P}(|S_k|>t/3)+\mathbb{P}(|S_n|>t/3)\big\}\\ &\le \max_{k\in[n]}\mathbb{P}(|S_k|>t/3)+\mathbb{P}(|S_n|>t/3),\end{aligned}$$

where the first equality follows because A_k and $S_n - S_k$ are independent. Combining this with (10.2), we obtain the desired conclusion. □

10.2 Generalizing Donsker's Theorem

Definition 10.2.1. We say a function $f:[0,1]\to\mathbb{R}$ is RCLL (Right-Continuous Left-Limit) or cadlag (continu à drolte, limit à gauche) if for all $t\in[0,1]$,

$$\lim_{t_n\searrow t} f(t_n)=f(t)\text{ (Right-continuous) and }\lim_{t_n\nearrow t} f(t_n)=f(t^-)\in\mathbb{R}\text{ (Left limit).}$$

Define $D=D[0,1]=\{f\in\mathbb{R}^{[0,1]}:f\text{ cadlag}\}$. One can show that $\|f\|_\infty<\infty, \forall f\in D$ and that $(D,\|\cdot\|_\infty)$ is a non-separable metric space.

Example 10.2.1. Let $X_1,\dots,X_n$ be iid random variable taking value on $[0,1]$ with $F(t):=\mathbb{P}(X_1\le t),\forall t\in[0,1]$. Define $F_n,G_n:\Omega\to D$ by

$$F_n(\omega,t)=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{X_i(\omega)\le t\}},\ \text{and}\ G_n(\omega,t)=\sqrt{n}(F_n(\omega,t)-F(t)).$$

Note that $\forall t\in[0,1]$, $\mathbb{E}G_n(\cdot,t)=0$. Also, for all $s<t\in[0,1]$,

$$\begin{aligned}\mathrm{Cov}(G_n(\cdot,s),G_n(\cdot,t))&=\mathbb{E}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbb{1}_{\{X_i\le t\}}-F(t))\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbb{1}_{\{X_i\le s\}}-F(s))\right)\\ &=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\mathbb{1}_{\{X_i\le t\}}-F(t))\cdot(\mathbb{1}_{\{X_i\le s\}}-F(s)\right) \qquad (X_i\text{ are iid})\\ &=F(s)(1-F(t)).\end{aligned}$$

Note that if $X_1\sim\mathrm{Unif}[0,1]$, then $\mathrm{Cov}(G_n(\cdot,s),G_n(\cdot,t))=s(1-t)=\mathrm{Cov}(W_s^\circ,W_t^\circ)$ for standard Brownian bridge $\{W_t^\circ\}_{t\in[0,1]}$. By CLT,

$$(G_n(\cdot,t_1),\dots,G_n(\cdot,t_m))\xrightarrow{d}(W_{t_1}^\circ,\dots,W_{t_m}^\circ).$$

A natural guess is that G_n converge to W° as $n\to\infty$. For a general random variable X_1, we guess that G_n converges to $\{W_{F(t)}^\circ\}_{t\in[0,1]}$ where $F(\cdot):=\mathbb{P}(X_1\le\cdot)$. But, can we get weak convergence in functional space?

Remark 10.2.1. $(D,\|\cdot\|_\infty)$ is not separable. Letting $\mathscr{G}\subset 2^D$ be the Borel σ-field w.r.t. $\|\cdot\|_\infty$, we find that G_n is not $\mathscr{F}/\mathscr{G}$-measurable. Also, Prokhorov's theorem does not apply.

One solution by A.V. Skorohod is to define a new metric on D. Let

$$\Lambda:=\{\lambda:[0,1]\to[0,1]:\lambda\text{ Strictly increasing, continuous bijection}\}.\text{ “wiggle in time”}$$

Define $d_s(f,g)=\inf_{\lambda\in\Lambda}\|\lambda-Id\|_\infty\vee\|f-g\circ\lambda\|_\infty$. Then:

- (D, d_s) is a separable metric space. If we define $\mathscr{D} \subseteq 2^D$ as the Borel σ-field with respect to d_s, then G_n is $\mathscr{F}/\mathscr{D}$-measurable.
- If $\{f_n\}$, $f \in C$, then $\|f_n - f\|_\infty \to 0$ iff $d_s(f_n, f) \to 0$.
- We may extend the Arzela-Ascoli theorem to (D, d_s).
- $G_n \xrightarrow{d} W^\circ$ on $(D, \mathscr{D})$.

Another solution by R. Dudley is to abandon Borel σ-field and use ball σ-field, generated by open balls.

We will study the solution by Hoffman-Jorgensen, summarized in Van der Vaart & Wellner 2000, (Vdv&W).

Main idea: define a notation $G_n \rightsquigarrow W^\circ$ without requiring G_n to be measurable. "Convergence in law without laws being defined."

10.3 Outer Integral

Definition 10.3.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $Y : \Omega \to \overline{\mathbb{R}}$ be arbitrary. Define outer integral

$$\mathbb{E}^\star Y := \inf\{\mathbb{E}X : X \geq Y, X \text{ is } \mathscr{F}/\mathscr{B}(\mathbb{R})\text{-measurable}\}. \tag{10.3}$$

For $B \subseteq 2^\Omega$ (not necessarily in $\mathscr{F}$), define outer measure

$$\mathbb{P}^\star(B) := \inf\{\mathbb{P}(A) : A \in \mathscr{F}, A \supseteq B\}.$$

One may similarly define inner integral $\mathbb{E}_\star Y$ and inner probability $\mathbb{P}_\star(B)$.

Note that Y is $\mathscr{F}/\mathscr{B}(\mathbb{R})$-measurable and $\mathbb{E}Y$ exists iff $\mathbb{E}^\star Y = \mathbb{E}_\star Y = \mathbb{E}Y$.

Remark 10.3.1. (a) We have $\mathbb{E}^\star(Y_1 + Y_2) \leq \mathbb{E}^\star Y_1 + \mathbb{E}^\star Y_2$ since $X_1 \geq Y_1$, $X_2 \geq Y_2$ implies $X_1 + X_2 \geq Y_1 + Y_2$. Reverse is not true in general.

(b) One still have, $\forall t \geq 0$,

$$\mathbb{P}^\star(|Y| \geq t) \leq \frac{\mathbb{E}^\star|Y|}{t} \text{ from regular Markov inequality.}$$

(c) Lemma 1.2.1 in VdV & W:

For any $Y : \Omega \to \overline{\mathbb{R}}$, $\exists Y^\star : \Omega \to \overline{\mathbb{R}}$, $\mathscr{F}/\mathscr{B}(\mathbb{R})$-measurable such that

- $Y \leq Y^\star$ and $U \geq Y^\star$, for any $U : \Omega \to \overline{\mathbb{R}}$, $\mathscr{F}/\mathscr{B}(\mathbb{R})$-measurable with $U \geq Y$;
- If $\mathbb{E}^\star Y < \infty$, then $\mathbb{E}Y^\star = \mathbb{E}^\star Y$.

We say $Y^\star$ is the minimal measurable majorant of Y.

Definition 10.3.2. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and (D, d) be a metric space with Borel σ-field $\mathscr{D}$. Write $C_b(D) := \{f \in \mathbb{R}^D : f \text{ continuous and bounded}\}$. Let $G_n : \Omega \to D$ be maps and let $P \in \mathcal{P}(D)$ (probability measure on $(D, \mathscr{D})$).

We say G_n converges weakly to P, written $G_n \rightsquigarrow P$ if

$$\mathbb{E}^\star\phi(G_n) \to \int_D \phi \mathrm{d}P \quad \forall \phi \in C_b(D). \tag{10.4}$$

If $G : \Omega \to D$ is $\mathscr{F}/\mathscr{D}$-measurable and has law P, then we say $G_n \rightsquigarrow G$.

Note: "$\mathbb{P}^{(G_n)}$" is not defined since G_n is not $\mathscr{F}/\mathscr{D}$-measurable.

Theorem 10.3.1 (Continuous Mapping Theorem [Thm. 1.3.6. of VdV&W]). Let E be a metric space and let $g : D \to E$ be continuous. If $G_n \rightsquigarrow G$, then $g(G_n) \rightsquigarrow g(G)$. (Note $g(G) : \Omega \to E$ is $\mathscr{F}/\mathscr{B}(E)$-measurable.)

Definition 10.3.3. Let (D, d) be a metric space. We say that a sequence of mappings $G_n : \Omega \to D$ is asymptotically measurable if

$$\mathbb{E}^\star \phi(G_n) - \mathbb{E}_\star \phi(G_n) \to 0 \qquad \forall \phi \in C_b(D). \tag{10.5}$$

We say that $\{G_n\}_{n\in\mathbb{N}}$ is asymptotically tight if $\forall \varepsilon > 0$, $\exists$ compact set $K \subseteq D$ such that $\forall \delta > 0$,

$$\liminf_{n\to\infty} \mathbb{P}^\star(G_n \in K^\delta) \geq 1 - \varepsilon \tag{10.6}$$

where $K^\delta := \{x \in D : d(x, K) < \delta\}$.

Note: If D is complete and separable (Polish) and if G_n is $\mathscr{F}/\mathscr{D}$-measurable, then (10.6) is equivalent to tightness.

Remark 10.3.2 (Lemma 1.3.8 in VdV & W)**.** Let $G : \Omega \to D$ be $\mathscr{F}/\mathscr{D}$-measurable.

(1) If $G_n \rightsquigarrow G$, then $\forall \phi \in C_b(D)$, $\mathbb{E}^\star \phi(G_n) \to \mathbb{E}\phi(G)$ and $\mathbb{E}_\star \phi(G_n) = -\mathbb{E}^\star[-\phi(G_n)] \to \mathbb{E}\phi(G_n)$, which implies $\{G_n\}_{n\in\mathbb{N}}$ is asymptotic measurable.

(2) If $G_n \rightsquigarrow G$, then $\{G_n\}_{n\in\mathbb{N}}$ is asymptotic tight iff G is tight. (Note G is not guaranteed to be tight since D is not polish)

Theorem 10.3.2 (Prokhorov; Thm 1.3.9. in VdV& W). If $\{G_n\}_{n\in\mathbb{N}}$ is asymptotic tight and asymptotic measurable, then $\exists$ subsequence $\{n_1, n_2, \dots\}$ and tight $P \in \mathcal{P}(D)$ such that $G_n \rightsquigarrow P$.

Note: VdV&W often use the concept of a "net". A "net" is a generalization of a sequence. Let A be a partially ordered set, let $x_\alpha \in D$, $\forall \alpha \in A$. We say $x_d \to x$ iff for any $\varepsilon > 0$, $\exists \alpha_0 \in A$ such that $d(x_\alpha, x) < \varepsilon$, $\forall \alpha \succcurlyeq \alpha_0$.

10.4 Space $\ell^\infty(T)$

Definition 10.4.1. Let T be a set and define

$$\ell^\infty(T) := \{f \in \mathbb{R}^T : \|f\|_\infty < \infty\} \text{ where } \|f\|_\infty := \sup_{t\in T} |f(t)|.$$

Note that $(\ell^\infty(T), \|\cdot\|_\infty)$ is a metric space; let $\mathscr{B}(\ell^\infty(T))$ be the Borel σ-field. Note also $\ell^\infty(T) \cap \mathscr{B}(\mathbb{R})^{\otimes T} \subseteq \mathscr{B}(\ell^\infty(T))$ since the evaluation functionals $e_t : \ell^\infty(T) \to \mathbb{R}$ are still continuous. In particular, if $G : \Omega \to \ell^\infty(T)$ is $\mathscr{F}/\mathscr{B}(\ell^\infty(T))$-measurable, then $\forall t \in T$, $G(\cdot, t) : \Omega \to \mathbb{R}$ is a random variable. Because $G(\cdot, t) = e_t \circ G$, then $\forall S \in \mathscr{B}(\mathbb{R})$, we obtain $G(\cdot, t)^{-1}(S) = G^{-1}(e_t^{-1}(S)) \in \mathscr{F}$ since $e_t^{-1}(S) \in \mathscr{B}(\ell^\infty(T))$.

Lemma 10.4.1 (Lem 1.5.2 in VdV&W)**.** Let $G_n : \Omega \to \ell^\infty(T)$ be a sequence of maps. If $\{G_n\}_{n\in\mathbb{N}}$ is asymptotic tight, then it is asymptotic measurable iff $\forall t \in T$, $\{G_n(\cdot, t)\}_{n\in\mathbb{N}}$ is asymptotic measurable.

Lemma 10.4.2 (Lem 1.5.3 in VdV&W)**.** Let $G, F : \Omega \to \ell^\infty(T)$ be $\mathscr{F}/\mathscr{B}(\ell^\infty(T))$-measurable and suppose $\mathbb{P}^{(G)}$, $\mathbb{P}^{(F)}$ are tight. Then $\mathbb{P}^{(G)} = \mathbb{P}^{(F)}$ iff $\forall \{t_1, \dots, t_m\} \subseteq T$,

$$(G(\cdot, t_1), \dots, G(\cdot, t_m)) \overset{d}{=} (F(\cdot, t_1), \dots, F(\cdot, t_m)).$$

In other words, tight Borel probability measures on $\ell^\infty(T)$ is uniquely specified by finite-dimensional marginals. Needed because $\mathscr{B}(\ell^\infty(T)) \neq \ell^\infty(T) \cap \mathscr{B}(\mathbb{R})^{\otimes T}$.

Theorem 10.4.1 (Thm. 1.5.4 in VdV&W)**.** Let $G_n : \Omega \to \ell^\infty(T)$ be a sequence of maps. There exists a tight $P \in \mathcal{P}(\ell^\infty(T))$ such that $G_n \rightsquigarrow P$ iff

(a) $\{G_n\}$ is asymptotic tight and

(b) $\forall\{t_1, \dots, t_m\} \subseteq T$, $\exists P_{t_1,\dots,t_m} \in \mathcal{P}(\mathbb{R}^m)$ such that

$$(G_n(\cdot, t_1), \dots, G_n(\cdot, t_m)) \rightsquigarrow P_{t_1,\dots,t_m}.$$

Proof.

$\Longrightarrow$: If $\exists P \in \mathcal{P}(\ell^\infty(T))$ tight such that $G_n \rightsquigarrow P$, then (a) and (b) follow from Remark 10.3.2b and continuous mapping theorem.

$\Longleftarrow$: Now, suppose (a) and (b) hold. From (b) and Remark 10.3.2a, we have that $\forall t \in T$, $\{G_n(\cdot, t)\}$ is asymptotic measurable and thus, $\{G_n\}$ is asymptotic measurable. Thus, by (a) and Prokhorov's theorem, there $\exists$ subsequence $\{n_1, n_2, \dots\}$ and tight $P \in \mathcal{P}(\ell^\infty(T))$ such that $G_{n_k} \rightsquigarrow P$. By continuous mapping theorem and (b), it must be that for any $\{t_1, \dots, t_m\} \subseteq T$, $P^{(e_{t_1},\dots,e_{t_m})} = P_{t_1,\dots,t_m}$.

Now suppose $G_n \not\rightsquigarrow P$, then $\exists \phi \in C_b(\ell^\infty(T))$, subsequence $\{n'_1, n'_2, \dots\}$, and $\varepsilon > 0$ such that

$$|\mathbb{E}^\star \phi(G_{n'_k}) - \int_D \phi \mathrm{d}P| > \varepsilon, \forall k \in \mathbb{N}.$$

But since $\{G_{n'_k}\}_{k\in\mathbb{N}}$ is asymptotic tight and measurable, $\exists$ tight $Q \in \mathcal{P}(\ell^\infty(T))$ and a further subsequence $\{n''_1, n''_2, \dots\}$ such that $G_{n''_k} \rightsquigarrow Q$. Since $Q^{(e_{t_1},\dots,e_{t_m})} = P^{(e_{t_1},\dots,e_{t_m})}$, we have $P = Q$ by Lemma 10.4.2. Contradiction.

Thus, $G_n \rightsquigarrow P$ as desired. $\square$

Remark 10.4.1 (Example 1.5.10 in VdV&W)**.** Recall that any collection $\{P_{t_1,\dots,t_m} \in \mathcal{P}(\mathbb{R}^m)\}_{t_1,\dots,t_m}$ that are Kolmogorov-consistent, there is a unique probability measure on $(\mathbb{R}^T, \mathscr{B}(\mathbb{R})^{\otimes T})$ such that $P^{(e_{t_1},\dots e_{t_m})} = P_{t_1,\dots,t_m}$. If each $P_{t_1,\dots,t_m}$ is Gaussian, then P is called Gaussian.

If there $\exists$ maps $\{G_n : \Omega \to \ell^\infty(T)\}_{n\in\mathbb{N}}$ asymptotic tight and

$$(G_n(\cdot, t_1), \dots, G_n(\cdot, t_m)) \rightsquigarrow P_{t_1,\dots,t_m},$$

then P is also a tight probability measure on $(\ell^\infty(T), \mathscr{B}(\ell^\infty(T)))$, i.e., $P(A)$ is defined for all $A \in \mathscr{B}(\ell^\infty(T))$ and $P(\ell^\infty(T)) = 1$.

Note: Gaussian white noise for example, is on $(\mathbb{R}^T, \mathscr{B}(\mathbb{R})^{\otimes T})$ but not on $(\ell^\infty(T), \mathscr{B}(\ell^\infty(T)))$ and definitely not tight.

Let $G : \Omega \to \mathbb{R}^T$ be a Gaussian process. Define semi-metric, for $p > 0$,

$$\rho_p(s, t) = (\mathbb{E}|G(\cdot, s) - G(\cdot, t)|^p)^{\frac{1}{p\vee 1}} \text{ for } s, t \in T.$$

Then $\mathbb{P}^{(G)}$ is tight probability in $(\ell^\infty(T), \mathscr{B}(\ell^\infty(T)))$ iff $\forall p > 0$, for a.e. $\omega \in \Omega$, $\lim_{\delta\to 0} w_{\rho_p}(G(\omega, \cdot), \delta) = 0$ (uniform ρ_p-continuity a.e.) where

$$w_{\rho_p}(G, S) := \sup\{|G(\cdot, s) - G(\cdot, t)| : s, t \in T, \rho_p(s, t) < \delta\}.$$

Theorem 10.4.2. [Theorem 1.5.7 in VdV&W] First, we say that $\rho : T \times T \to [0, \infty)$ is a semi-metric on T if it satisfies that $\rho(s, t) = 0$ iff $s = t$ and $\rho(s, t) = \rho(t, s)$.

Let $\{G_n\}_{n\in\mathbb{N}}$ be a sequence of maps. $\{G_n\}$ is asymptotic tight iff $\forall t \in T$, $\{G_n(\cdot, t)\}_{n\in\mathbb{N}}$ is asymptotic tight and there exists semi-metric ρ such that

(1) (T, ρ) is totally bounded ($\forall\varepsilon > 0$, $\exists\{t_1, \dots, t_m\} \subseteq T$ such that $T \subseteq \cup_{i=1}^m B(t_i, \varepsilon)$).

(2) $\forall\epsilon > 0$, $\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}^*(w_\rho(G_n, \delta) > \varepsilon) = 0$ (Asymptotically uniformly ρ-equicontinuity, where, for $f : T \to \mathbb{R}$, $w_\rho(f, \delta) := \sup\{|f(s) - f(t)| : s, t \in T, \rho(s, t) < \delta\}$ is the modulus of continuity.)

Moreover, if $G_n \rightsquigarrow G$, then for a.e. $\omega \in \Omega$, $\lim_{\delta\to 0} w_\rho(G(\omega,\cdot),\delta) = 0$. $(\star)$ (Uniform ρ-continuity).

Also, if $G_n \rightsquigarrow G$ and ρ is a semi-metric such that for a.e. $\omega \in \Omega$, $\lim_{\delta\to 0} w(G(\omega,\cdot),\delta) = 0$, then $\{G_n\}$ is asymptotic uniform ρ-equicontinuity.

Proof (partial).

We prove only $(\star)$. First, note that for any semi-metric $\rho : T \times T \to [0,\infty)$, for $f, g \in \mathbb{R}^T$, $\forall \delta > 0$, $w_\rho(f+g,\delta) \le w_\rho(f,\delta) + w_\rho(g,\delta)$. It implies

$$|w_\rho(f,\delta) - w_\rho(g,\delta)| \le w_\rho(f-g,\delta) \le 2\left\|f-g\right\|_\infty,$$

So $w_\rho(\cdot,\delta)$ is 2-Lipschitz.

Assume now $G_n \rightsquigarrow G$, then $w_\rho(G_n,\delta) \rightsquigarrow w_\rho(G,\delta)$. Note that for any closed $F \subseteq \ell^\infty(T)$,

$$\limsup_{n\to\infty} \mathbb{P}^\star(G_n \in F) \le \mathbb{P}(G \in F). \qquad \text{(Thm 1.3.4(iii) in VdV\&W)}$$

Therefore, $\forall \varepsilon > 0$, $\delta > 0$,

$$\begin{aligned}\liminf_{n\to\infty} \mathbb{P}^\star(w_\rho(G_n,\delta) > \varepsilon) &\ge 1 - \limsup_{n\to\infty} \mathbb{P}^\star(w_\rho(G_n,\delta) \le \varepsilon)\\ &\ge 1 - \mathbb{P}(w_\rho(G,\delta) \le \varepsilon) = \mathbb{P}(w_\rho(G,\delta) > \varepsilon),\end{aligned}$$

where the first inequality holds since $\mathbb{P}^\star(A) + \mathbb{P}^\star(A^c) \ge 1$. So $\lim_{\delta\to 0} \mathbb{P}(w_\rho(G,\delta) > \varepsilon) = 0$. ☐

10.5 Empirical Process

Definition 10.5.1. Let $(\mathcal{X},\mathcal{A})$ be a measurable space and let $X_1,\ldots,X_n : \Omega \to \mathcal{X}$ be iid random objects. Let $P := \mathbb{P}^{(X_1)}$ and let $P_n : \Omega \to \mathcal{P}(\mathcal{X},\mathcal{A})$ be the empirical measurable, i.e. $P_n(\omega, A) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i(\omega)\in\mathcal{A}\}}$.

Let $\mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$ be a set of $\mathcal{A}/\mathscr{B}(\mathbb{R})$-measurable functions and suppose that

$$F(x) := \sup_{f\in\mathcal{F}} |f(x) - \mathbb{E}_P f(x)| < \infty \quad \forall x \in \mathcal{X}. \qquad (\star) \text{ (Envelope condtion)}$$

Define empirical process $G_n : \Omega \to \ell^\infty(\mathcal{F})$, $\ell^\infty(\mathcal{F}) = \{\phi : \mathcal{F} \to \mathbb{R} : \sup_{f\in\mathcal{F}} |\phi(f)| < \infty\}$, by

$$G_n(\omega, f) := \sqrt{n}(\mathbb{E}_{P_n(\omega)} f(X) - \mathbb{E}_P f(X)) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (f(X_i(\omega)) - \mathbb{E}f(X)).$$

Note that G_n is a map into $\ell^\infty(\mathcal{F})$ by $(\star)$. We have that $\mathbb{E}G_n(\cdot,f) = 0$ and that $\operatorname{Cov}(G_n(\cdot,f), G_n(\cdot,g)) = \mathbb{E}f(X)g(X) - \mathbb{E}f(X)\mathbb{E}g(X)$.

Example 10.5.1. Let $(\mathcal{X},\mathcal{A})$ be $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ and let $\mathcal{F} := \{x \mapsto \mathbb{1}_{\{x\le t\}} : t \in [0,1]\}$. Then we have that

$$G_n(\omega,t) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (\mathbb{1}_{\{X_i(\omega)\le t\}} - \mathbb{P}(X_1 \le t)).$$

Note that the envelope condition is satisfied since $|f| \le 1$, $\forall f \in \mathcal{F}$.

For $(\mathcal{X},\mathcal{A}) = (\mathbb{R}^d,\mathscr{B}(\mathbb{R}^d))$, we may replace $\mathbb{1}_{\{X_i(\omega)\le t\}} = \mathbb{1}_{\{X_i(\omega)\in[0,t]\}}$ by $\mathbb{1}_{\{X_i(\omega)\in R\}}$ for $R \in \mathcal{R}$ where $\mathcal{R}$ is a set of geometric sets such as hyper-rectangles, spheres, etc.

Now let $(\mathcal{X},\mathcal{A})$ be $(\mathbb{R}^d,\mathscr{B}(\mathbb{R}^d))$, let $\{p_\theta\}_{\theta\in\Theta}$ be a class of densities on $\mathbb{R}^d$ parametrized by Θ and let $\mathcal{F} := \{x \mapsto \log p_\theta(x) : \theta \in \Theta\}$ and assume that envelope condition holds.

Then we have $G_n(\omega,f) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (\log p_\theta(X_i) - \mathbb{E}_p \log p_\theta(X))$.

Remark 10.5.1. Since $f : \mathcal{X} \to \mathbb{R}$ is $\mathcal{A}/\mathscr{B}(\mathbb{R})$-measurable $\forall f \in \mathcal{F}$, we have $G_n(\cdot,f) : \Omega \to \mathbb{R}$ is a random variable. Since $G_n(\cdot,f) \xrightarrow{d} N(0,\operatorname{Var} f(X))$, $\{G_n(\cdot,f)\}_{n\in\mathbb{N}}$ is tight. Thus, to verify that $\{G_n\}_{n\in\mathbb{N}}$ is asymptotic tight, by Theorem 10.4.2 , we must show that for a semi-metric ρ,

(a) $(\mathcal{F}, \rho)$ is totally bounded and

(b) $\forall \varepsilon > 0$, $\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}^\star(w_\rho(G_n, \delta) > \varepsilon) = 0$.

First, let $G : \Omega \to \mathbb{R}^{\mathbb{R}^{\mathcal{X}}}$ be the Gaussian process with mean-zero and

$$\mathrm{Cov}(G(\cdot, f), G(\cdot, g)) = \mathbb{E}f(X)g(X) - \mathbb{E}f(X)\mathbb{E}g(X).$$

Then we may take

$$\begin{aligned}
\rho(f, g) &:= (\mathbb{E}|G(\cdot, f) - G(\cdot, g)|^2)^{\frac{1}{2}} \\
&= (\mathbb{E}G(\cdot, f)^2 - 2\mathbb{E}G(\cdot, f)G(\cdot, g) + \mathbb{E}G(\cdot, g)^2)^{\frac{1}{2}} \\
&= \{\mathbb{E}(f(X) - \mu_f - g(X) - \mu_g)^2\}^{\frac{1}{2}} && \text{(where } \mu_f := \mathbb{E}f(X), \mu_g := \mathbb{E}g(X)) \\
&\le \{\mathbb{E}(f(X) - g(X))^2\}^{\frac{1}{2}} := \|f - g\|_{L_2(P)}.
\end{aligned}$$

Now, for $\delta > 0$,

$$\begin{aligned}
w_\rho(G,\delta) &= \sup\{|G_n(\cdot, f) - G_n(\cdot, g)| : f, g \in \mathcal{F}, \|f - g\|_{L_2(P)} \le \delta\} \\
&= \sup\{|G_n(\cdot, f - g)| : f, g \in \mathcal{F}, \|f - g\|_{L_2(P)} \le \delta\} \\
&= \sup\{|G_n(\cdot, h)| : h \in \mathcal{F}_\delta\}
\end{aligned}$$

where $\mathcal{F}_\delta := \{h \in \mathbb{R}^{\mathcal{X}} : h = f - g \text{ for } f, g \in \mathcal{F} \text{ and } \|h\|_{L_2(P)} < \delta\}$. Therefore, (b) reduces to $\forall \varepsilon > 0$

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}^\star(\sup_{h\in\mathcal{F}_\delta} |G_n(\cdot, h)| > \varepsilon) = 0.$$

Definition 10.5.2. For a set $\mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$ of $\mathcal{A}/\mathscr{B}(\mathbb{R})$-measurable functions, for $\varepsilon > 0$, for a metric $d : \mathcal{F} \times \mathcal{F} \to [0, \infty)$, define an ε-covering as any finite set $\{f_1, \dots, f_m\} \subseteq \mathcal{F}$ such that $\mathcal{F} \subseteq \cup_{i=1}^m B_d(f_i, \varepsilon)$. Further define $N(\varepsilon, \mathcal{F}, d) := \min\{m \in \mathbb{N} : \exists f_1, \dots, f_m \in \mathcal{F} \text{ s.t. } \mathcal{F} \subseteq \cup_{i=1}^m B_d(f_i, \varepsilon)\}$ as the covering number of $(\mathcal{F}, d)$.

Theorem 10.5.1. (Donsker's theorem; Thm 2.5.2 in VdV&W) Let $X_1, \dots, X_n : \Omega \to \mathcal{X}$ be iid random objects with $P := \mathbb{P}^{(X_1)}$. Let $\mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$ be a set of $\mathcal{A}/\mathscr{B}(\mathbb{R})$-measurable function such that

(1) $F(x) := \sup_{f\in\mathcal{F}} |f(X) - \mu_f| < \infty$ and $\mathbb{E}^\star F(X)^2 < \infty$,

(2) and that $\int_0^1 M^{\frac{1}{2}}(\varepsilon)\mathrm{d}\varepsilon < \infty$, where we define

$$M(\varepsilon) := \sup\{\log N(\varepsilon \|F\|_{L_2(Q)}, \mathcal{F}, L_2(Q)) : Q \text{ discrete prob. meas. on } (\mathcal{X}, \mathcal{A})\}.$$

Then, we have that

$$G_n \rightsquigarrow G$$

where $G : \Omega \to \ell^\infty(\mathcal{F})$ is $\mathscr{F}/\mathscr{B}(\ell^\infty(\mathcal{F}))$-measurable and a tight Gaussian process such that $\mathbb{E}G(\cdot, f) = 0$ and $K(f, g) = \mathbb{E}G(\cdot, f)G(\cdot, g) = \mathbb{E}f(X)g(X) - \mu_f\mu_g$.

Example 10.5.2. Let $(\mathcal{X}, \mathcal{A}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ and let $\mathcal{F} := \{x \mapsto \mathbb{1}_{\{x\le t\}} : t \in \mathbb{R}\}$. Then, $F(x) = 1$ for all $x \in \mathbb{R}$. We claim that for any probability measure Q on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, there exists a universal $C > 0$ such that, for any $\varepsilon > 0$,

$$N(\varepsilon, \mathcal{F}, L_2(Q)) \le \frac{C}{\varepsilon^2}. \tag{10.7}$$

To see this, define $t_1 < t_2 < \dots t_L \in \mathbb{R}$ such that

$$\begin{aligned}
t_1 &:= \inf\{t \in \mathbb{R} : Q((-\infty, t]) \ge \varepsilon^2\} \\
t_2 &:= \inf\{t > t_1 : Q((t_1, t_2]) \ge \varepsilon^2\} \\
&\dots
\end{aligned}$$

Then, we have that if $Q(\{t\}) \geq \varepsilon^2$, then $t \in \{t_1, \ldots, t_L\}$. Moreover, we have that $Q((t_\ell, t_{\ell+1})) \leq \varepsilon^2$ and that $L \leq \varepsilon^{-2}$.

Define $f_\ell = x \mapsto \mathbb{1}_{\{x \leq t_\ell\}}$. We claim that $\{0, f_1, f_2, \ldots, f_L\}$ is an ϵ-covering of $\mathcal{F}$. Indeed, let $g := x \mapsto \mathbb{1}_{\{x \leq s\}}$. If $s < t_1$, then

$$\int_{-\infty}^{\infty} (g(x) - 0)^2 dQ(x) \leq Q((-\infty, s]) \leq \varepsilon^2.$$

If $s \in \{t_1, t_2, \ldots, t_L\}$, then g is obviously covered. Let t_ℓ be the largest point such that $t_\ell < s$. Then,

$$\int_{-\infty}^{\infty} (g(x) - f_\ell(x))^2 dQ(x) \leq Q((t_\ell, s]) \leq \varepsilon^2.$$

Hence, (10.7) follows. We then have that $M(\varepsilon) = 2 \log \frac{1}{\varepsilon}$ and $\int_0^1 M^{1/2}(\varepsilon) d\varepsilon < \infty$.

Proof Sketch of Theorem 10.5.1.

Assume $\mathcal{F}$ satisfies (1) and (2). By Remark 10.5.1, we need only to show that $(\mathcal{F}, L_2(P))$ is totally bounded and that

$$\forall \varepsilon > 0, \lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}^\star(\sup_{f \in \mathcal{F}_\delta} |G_n(\cdot, f)| > \varepsilon) = 0,$$

where $\mathcal{F}_\delta := \{f \in \mathcal{F} - \mathcal{F} : \mathbb{E} f(X)^2 < \delta^2\}$. Fix $\varepsilon > 0$, and let δ_n be any sequence going to 0, then

$$\mathbb{P}^\star \left(\sup_{f \in \mathcal{F}_{\delta_n}} |G_n(\cdot, f)| > \varepsilon \right) \leq \frac{1}{\varepsilon} \mathbb{E}^\star \sup_{f \in \mathcal{F}_{\delta_n}} |G_n(\cdot, f)|.$$

Let $\widetilde{X}_1 . \widetilde{X}_2, \ldots \widetilde{X}_n$ be iid where $\widetilde{X}_i \overset{d}{=} X_i$ and $X_i \perp\!\!\!\perp \widetilde{X}_i$. Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent Rademacher random variables. Informally, we have

$$\begin{aligned}
\mathbb{E}^\star \sup_{f \in \mathcal{F}_{\delta_n}} |G_n(\cdot, f)| &= \mathbb{E}^\star \sup_{f \in \mathcal{F}_{\delta_n}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n f(X_i) - \mathbb{E} f(X) \right| \\
&\leq \mathbb{E}_X^\star \mathbb{E}_{\widetilde{X}}^\star \sup_{f \in \mathcal{F}_{\delta_n}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n f(X_i) - f(\widetilde{X}_i) \right| \\
&= \mathbb{E}_X^\star \mathbb{E}_{\widetilde{X}}^\star \mathbb{E}_\varepsilon^\star \sup_{f \in \mathcal{F}_{\delta_n}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i (f(X_i) - f(\widetilde{X}_i)) \right| \\
&\leq \mathbb{E}_X^\star \mathbb{E}_\varepsilon^\star \sup_{f \in \mathcal{F}_{\delta_n}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i f(X_i) \right| + \mathbb{E}_{\widetilde{X}}^\star \mathbb{E}_\varepsilon^\star \sup_{f \in \mathcal{F}_{\delta_n}} \left| \frac{1}{\sqrt{n}} \varepsilon_i f(\widetilde{X}_i) \right| \\
&= 2 \mathbb{E}_X^\star \mathbb{E}_\varepsilon^\star \sup_{f \in \mathcal{F}_{\delta_n}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i f(X_i) \right|.
\end{aligned}$$

Now, fix (condition on) $X_1, \ldots, X_n$. Let $\mathbb{F} := \{(f(X_1), \ldots, f(X_n)) \in \mathbb{R}^n : f \in \mathcal{F}_{\delta_n}\}$

Define process $\Psi : \widetilde{\Omega} \to \mathbb{R}^{\mathbb{F}}$ by $\Psi(\tilde{\omega}, f) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i(\tilde{\omega}) f(X_i)$. Note, Ψ is continuous on $\mathbb{F}$ w.r.t $\frac{1}{\sqrt{n}} \|\cdot\|_2$ where $\frac{1}{\sqrt{n}} \|f\| := \sqrt{\frac{1}{n} \sum_{i=1}^n f(X_i)^2}$. Then, for every $f, g \in \mathbb{F}$,

$$\mathbb{P}(|\Psi(\cdot, f) - \Psi(\cdot, g)| > x) \leq 2 \exp\left(-\frac{1}{2} \frac{x^2}{\frac{1}{n} \|f - g\|_2^2} \right).$$

So

$$\begin{aligned}
\mathbb{E}_\varepsilon \sup_{f\in\mathbb{F}} \Psi(\cdot, f) &\le \int_0^\infty \log^{\frac{1}{2}} N(\varepsilon, \mathbb{F}, \frac{1}{\sqrt{n}} \|\cdot\|_2) \mathrm{d}\varepsilon \\
&= \int_0^\infty \log^{\frac{1}{2}} N(\varepsilon, \mathcal{F}_{\delta_n}, L_2(P_n)) \mathrm{d}\varepsilon \\
&= \int_0^{\theta_n} \log^{\frac{1}{2}} N(\varepsilon, \mathcal{F}_{\delta_n}, L_2(P_n)) \mathrm{d}\varepsilon \qquad (\text{where } \theta_n := \sup_{f\in\mathcal{F}_{\delta_n}} \|f\|_{L_2(P_n)}) \\
&= \int_0^{\theta_n/\|F\|_{L_2(P_n)}} \log^{\frac{1}{2}} N(\varepsilon \|F\|_{L_2(P)}, \mathcal{F}-\mathcal{F}, L_2(P_n)) \mathrm{d}\varepsilon \|F\|_{L_2(P_n)} \\
&= \int_0^{\theta_n/\|F_\star\|_{L_2(P_n)}} \underbrace{\sup_{Q \text{ discrete}} \sqrt{2} \log^{\frac{1}{2}} N\left(\frac{\varepsilon}{2} \|F\|_{L_2(Q)}, \mathcal{F}, L_2(Q)\right)}_{M(\varepsilon)^{\frac{1}{2}}} \mathrm{d}\varepsilon \|F\|_{L_2(P_n)}.
\end{aligned}$$

We finish the proof by showing $\theta_n := \sup_{f\in\mathcal{F}_{\delta_n}} \left\{\frac{1}{n}\sum_{i=1}^n f(X_i)^2\right\}^{\frac{1}{2}} \to 0$ in probability as $n\to\infty$. We may similarly show that $(\mathcal{F}, L_2(P))$ is totally bounded. □

Remark 10.5.2. The following argument is due to Van de Geer 2000: Let $\Psi : \Omega \to \mathbb{R}^T$ be $\mathscr{F}/\mathscr{B}(\mathbb{R})^{\otimes T}$-measurable and assume

(a) d is a metric on T such that $D = \mathrm{diam}T = \sup_{s,t\in T} d(s,t) < \infty$,

(b) $\int_0^D \log^{\frac{1}{2}} N(\varepsilon, T, d)\mathrm{d}\varepsilon < \infty$,

(c) $\exists t_0 \in T$ such that $\Psi(\cdot, t_0) = 0$ a.s. and $t \mapsto \Psi(\cdot, t)$ is continuous a.s.,

(d) $\forall s, t \in T$, $\mathbb{P}(|\Psi(\cdot, s) - \Psi(\cdot, t)| \ge t) \le 2e^{-\frac{1}{2}\frac{x^2}{d(s,t)^2}}$.

Then, $\forall x \ge \max\left(\sqrt{2\log 2}, D\int_0^D \log^{\frac{1}{2}} N(\varepsilon, T, d)\mathrm{d}\varepsilon\right)$,

$$\mathbb{P}(\sup_{t\in T} |\Psi(\cdot, t)| > x) \le 4e^{-\frac{1}{2}x^2}.$$

Proof.

For $j \in \mathbb{N}$, let $T_j := D2^{-j}$-covering of T so that $|T_j| \le N(D2^{-j}, T, d)$. Set $T_0 = \{t_0\}$. For any $\omega \in \Omega$, for any $t \in T$, $\exists$ a chain $t_0, t_1, t_2, \cdots \in T$ such that $t_j \in T_j$, and $d(t_j, t_{j+1}) \le D2^{-(j+1)}$, and $\lim_{j\to\infty} d(t_j, t) = 0$.

We have, for a.e. $\omega \in \Omega$. for any $J \in \mathbb{N}$,

$$|\Psi(\omega, t) - \Psi(\omega, t_J)| + |\Psi(\omega, t)| \le \left|\sum_{j=1}^J \Psi(\omega, t_j) - \Psi(\omega, t_{j-1})\right|.$$

Taking $J \to \infty$, we have

$$|\Psi(\omega, t)| \le \sum_{j=1}^\infty |\Psi(\omega, t_j) - \Psi(\omega, t_{j-1})|$$

by a.s. continuity. We note also that

$$d(t_j, t_{j-1}) \le d(t_j, t) + d(t_{j-1}, t) \le 4D2^{-j}.$$

Let $x_1, x_2, \cdots > 0$ be such that $\sum_{j=1}^{\infty} x_j \le x$, then $|\Psi(\omega, t)| > x$ implies there exists $j \in \mathbb{N}$ such that $|\Psi(\omega, t_j) - \Psi(\omega, t_{j-1})| > x_j$. Thus

$$\begin{aligned}\mathbb{P}(\sup_{t\in T}|\Psi(\cdot,t)| > x) &\le \sum_{j=1}^{\infty}\mathbb{P}\Big(\sup\big\{|\Psi(\omega,t_j)-\Psi(\omega,t_{j-1})| : t_j \in T_j, t_{j-1}\in T_{j-1}, d(t_j,t_{j-1}) \le 4D2^{-j}\big\} > x_j\Big)\\ &\le 2\sum_{j=1}^{\infty}\exp\left\{-\frac{1}{2}x_j^2(4D2^{-j})^{-2} + 2\log N(D2^{-j},T,d)\right\}. \qquad (\star)\end{aligned}$$

Set x_j such that $-8x_j^2 D^{-2}2^{-2j} + \log 2N(D2^{-j}, T, d) = -\frac{j}{2}x^2$ and it implies

$$x_j = \frac{1}{8}D2^{-j}\left\{2\log N(D2^{-j},T,d) + \frac{j}{2}x^2\right\}^{\frac{1}{2}}$$

and, using the fact that $\epsilon \mapsto \log N(\epsilon, T, d)$ is a decreasing function,

$$\sum_{j=1}^{\infty} x_j \le D\int_0^D \log^{\frac{1}{2}} N(\varepsilon,T,d)\mathrm{d}\varepsilon + \frac{x}{2}\int_0^1 \log^{\frac{1}{2}}\frac{1}{\varepsilon}\mathrm{d}\varepsilon \le x$$

as required. So

$$(\star) \le 2\sum_{j=1}^{\infty} e^{-\frac{j}{2}x^2} = 2e^{-\frac{j}{2}x^2} = 2e^{-\frac{1}{2}x^2}(1-e^{-\frac{1}{2}x^2})^{-1} \le 4e^{-\frac{1}{2}x^2}.$$

□

10.6 Kolmogorov-Smirnov Test

Recall Donsker's theorem for scaled empirical d.f. Let $X_1, \ldots, X_n$ be iid random variables. Define $\mathcal{F} := \{x \mapsto \mathbb{1}_{\{x\le t\}} : t \in \mathbb{R}\}$, define $G_n : \Omega \to \ell^\infty(\mathcal{F})$ as

$$G_n(\cdot, f) = \frac{1}{\sqrt{n}}\sum_{i=1}^n f(X_i) - \mathbb{E}f(X) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{1}_{\{X_i\le t\}} - F(t)$$

where $F(t) = \mathbb{P}(X_1 \le t)$. Then $G_n \rightsquigarrow G$ where $G : \Omega \to \ell^\infty(\mathcal{F})$ is a Gaussian Process with $\mathbb{E}G(\cdot, f) = 0$ and

$$\begin{aligned}\mathbb{E}G(\cdot,f)G(\cdot,g) &= \mathbb{E}f(X)g(X) - \mathbb{E}f(X)\mathbb{E}g(X)\\ &= F(s) - F(s)F(t) = F(s)(1-F(t)) \qquad \text{(for step functions with } s < t \in \mathbb{R})\end{aligned}$$

For the class $\mathcal{F}$, we write $G(\cdot, t) = \widetilde{W}(\cdot, F(t))$ where $\widetilde{W}$ is the standard Brownian bridge.

Example 10.6.1 (Goodness of Fit Testing). Now suppose we have data $X_1, \ldots, X_n$ iid random variables. We wish to test $H_0 : X_1 \sim F$ where F is a known probability measure on $\mathbb{R}$. We use F to also denote the distribution function.

The KS test statistics is

$$\phi(X_1,\ldots,X_n) = \|F_n - F\|_\infty = \sup_{t\in\mathbb{R}}\left|\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i\le t\}} - F(t)\right| = \sup_{f\in\mathcal{F}}|\mathbb{E}_{P_n}f(X) - \mathbb{E}f(X)|.$$

Let F_0 be the true d.f. of X_1. Then

$$\begin{aligned}\sqrt{n}\,\|F_n - F_0\|_\infty &\rightsquigarrow \sup_{t\in\mathbb{R}}|\widetilde{W}(\cdot,F_0(t))|\\ &= \underbrace{\sup_{t\in[0,1]}|\widetilde{W}(\cdot,t)|}_{\text{does not depend on } F_0}. \qquad \text{(assume } F_0 \text{ is continuous)}\end{aligned}$$

So $\phi = \|F_n - F_0 - (F - F_0)\|_\infty \rightsquigarrow 0$ if and only if $F_0 = F$. Also, under the null, $\sqrt{n}(F_n - F) \rightsquigarrow \{\widetilde{W}(\cdot, F(t))\}_{t\in\mathbb{R}}$ which implies $\sqrt{n}\phi \rightsquigarrow \sup_{t\in[0,1]} |\widetilde{W}(\cdot,t)|$, assuming F is continuous. The distribution of $\sup_{t\in[0,1]} |\widetilde{W}(\cdot,t)|$ is known as the Kolmogorov distribution. It is then straightforward to derive the critical values by using the density of the Kolmogorov distribution.

Example 10.6.2 (Two-sample Test). Now consider iid r.v. $\boldsymbol{X} = (X_1, \dots, X_n) \sim P$ and $\boldsymbol{Y} = (Y_1, \dots, Y_m) \sim Q$. We consider $H_0 : P = Q$. The KS test statistics is

$$\phi(\boldsymbol{X}, \boldsymbol{Y}) = \|F_{n,P} - F_{m,Q}\|_\infty = \sup_{t\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i \le t\}} - \frac{1}{m}\sum_{i=1}^m \mathbb{1}_{\{Y_i \le t\}} \right|$$

Let F_P, F_Q denote the cdf of P, Q respectively, then we have $\|F_{n,P} - F_P - (F_{m,Q} - F_Q) + F_P - F_Q\|_\infty \to 0$ as $n, m \to \infty$ iff $F_P = F_Q$. Under the null, we have

$$\sqrt{\frac{nm}{n+m}}\|F_{n,P} - F_{m,Q}\|_\infty = \left\| \sqrt{\frac{m}{n+m}}\sqrt{n}(F_{n,P} - F) - \sqrt{\frac{n}{n+m}}\sqrt{m}(F_{m,Q} - F) \right\|_\infty.$$

Note that as if $n, m \to \infty$ and $\frac{n}{n+m} \to \lambda \in (0,1)$, then, writing $\widetilde{W}$ and W° as two independent standard Brownian bridge, we have

$$\begin{aligned}\sqrt{\frac{m}{n+m}}\sqrt{n}(F_{n,P} - F) - \sqrt{\frac{n}{n+m}}\sqrt{m}(F_{m,Q} - F) &\rightsquigarrow \sqrt{1-\lambda}\{\widetilde{W}(\cdot, F(t))\}_{t\in\mathbb{R}} - \sqrt{\lambda}\{W^\circ(\cdot, F(t))\}_{t\in\mathbb{R}} \\ &\overset{d}{=} \{\widetilde{W}(\cdot, F(t))\}_{t\in\mathbb{R}} \overset{d}{=} \{\widetilde{W}_t\}_{t\in[0,1]}.\end{aligned}$$

Thus, $\sqrt{\frac{nm}{n+m}}\phi(\boldsymbol{X}, \boldsymbol{Y}) \rightsquigarrow \sup_{t\in[0,1]} |\widetilde{W}(\cdot,t)|$. One may also obtain critical values by permutation test.

Remark 10.6.1. The KS test can be inverted to form an asymptotically valid confidence band around the empirical CDF F_n. To obtain a non-asymptotic confidence band, one may use the Dvoretzky–Kiefer–Wolfowitz (DKW) inequality, which states that for iid random variables $X_1, \dots, X_n$ from any distribution, we have that

$$\mathbb{P}\left(\sup_{t\in\mathbb{R}} \sqrt{n}|F_n(t) - F(t)| \ge \epsilon\right) \le 2e^{-2\epsilon^2},$$

where the leading constant 2 is sharp due to work by P. Massart.

Example 10.6.3 (Multivariate KS test). Now let $X_1, \dots, X_n$ be iid random vectors taking values in $\mathbb{R}^2$. Let $\mathcal{F} = \{x \mapsto \mathbb{1}_{\{x\in(-\infty,a]\times(-\infty,b]\}} : a, b \in \mathbb{R}\}$. Then, $G_n : \Omega \to \ell^\infty(\mathcal{F})$ becomes

$$G_n(\cdot, (a,b)) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \big(\mathbb{1}_{\{X_i\in(-\infty,a]\times(-\infty,b]\}} - F(a,b)\big)$$

where $F(a,b) := \mathbb{P}(X \in (-\infty,a] \times (-\infty,b])$ and $G_n \rightsquigarrow G$ where $G : \Omega \to \ell^\infty(\mathcal{F})$ is a Gaussian process with mean-zero and covariance $K((a,b),(c,d)) = F(a \wedge c, b \wedge d) - F(a,b)F(c,d)$. G has no easy expression and critical values must be determined by simulation.

Remark 10.6.2. An alternative approach toward determining critical values is to use the bootstrap. Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be iid random vectors in $\mathscr{X} \subseteq \mathbb{R}^d$ with distribution $P = \mathbb{P}^{(X_1)}$. Let $P_n := P_n(\boldsymbol{X})$ be the empirical measure of $\boldsymbol{X}$ and let $X_1^*, \dots, X_n^*$ be iid samples from P_n.

For clarity, we let $X_1, \dots, X_n : \Omega \to \mathbb{R}^d$ and $X_1^*, \dots, X_n^* : \tilde{\Omega} \times \Omega \to \mathbb{R}^d$ where $X_1^*(\cdot, \omega)$ is the random value of X_1^* conditioned on $X_1(\omega), \dots, X_n(\omega)$.

Let $\mathscr{F} \subset \mathbb{R}^{\mathscr{X}}$ be a class of bounded measurable functions and define the bootstrap process $G_n^b : \tilde{\Omega} \times \Omega \to \ell^\infty(\mathscr{F})$ as

$$G_n^b(\tilde{\omega}, \omega, f) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(f(X_i^*(\tilde{\omega}, \omega)) - \frac{1}{n}\sum_{i=1}^n f(X_i(\omega)) \right).$$

For example, in the case of empirical CDF, we have that, for $t \in \mathbb{R}$,

$$G_n^b(\cdot, t) = \sqrt{n}(F_n^b(t) - F_n(t))$$

where F_n^b is the empirical CDF computed from the bootstrapped sample $X_1^*, \ldots, X_n^*$.

We have the following remarkable theorem from E. Gine and J. Zinn (1990): let $G : \Omega \to \ell^\infty(\mathscr{F})$ be the centered Gaussian process with covariance $K(f, g) = \mathbb{E}f(X)g(X) - (\mathbb{E}f(X))(\mathbb{E}g(X))$. Let $\tilde{G} : \tilde{\Omega} \to \ell^\infty(\mathscr{F})$ be a centered Gaussian process with the same covariance as G but independent of $X_1, X_2, \ldots$ and G. Then, when $\mathscr{F}$ has envelope F that satisfies $\mathbb{E}^* F(X)^2 < \infty$, we have

$$G_n^b(\cdot, \omega) \rightsquigarrow \tilde{G} \text{ on } \ell^\infty(\mathscr{F}) \text{ for } \mathbb{P}\text{-a.e. } \omega \text{ if and only if } G_n \rightsquigarrow G \text{ on } \ell^\infty(\mathscr{F}).$$

Hence, we may use many bootstrap samples to approximate G.

Chapter 11

Concentration of Measure

11.1 Outline

Remark 11.1.1. In its generic form, concentration inequalities say that, with random objects $(X_1, \ldots, X_n)$ taking value in $\mathcal{X}^n$ and with $Z := f(X_1, \ldots, X_n)$ in $\mathbb{R}$, under assumptions on f and $\mathbb{P}^{(X_1,\ldots,X_n)}$, there exist $C > 0$ and $\nu_n > 0$ such that $\forall t > 0$,

$$\mathbb{P}(Z - \mathbb{E}Z > t) \leq Ce^{-t^2\nu_n} (\text{ or } Ce^{-t\nu_n}). \tag{11.1}$$

The simplest example is if $X_1, \ldots, X_n \overset{iid}{\sim} N(0, \sigma^2)$ and $f(X_1, \ldots, X_n) = \sum_{i=1}^n X_i$, then $\mathbb{P}^{(Z)} = N(0, n\sigma^2)$ and we have by Mill's inequality,

$$\mathbb{P}(Z - \mathbb{E}Z > t) \leq \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{t} e^{-\frac{1}{2}\frac{t^2}{n\sigma^2}}.$$

Hoeffding inequality extends this to the setting where $X_1, X_2, \ldots, X_n$ are independent random variables taking value in a bounded interval $[a, b]$. Then, with $Z = \sum_{i=1}^n X_i$, we have

$$\mathbb{P}(Z - \mathbb{E}Z > t) \leq e^{-\frac{2t^2}{n(b-a)^2}}.$$

What is remarkable is that if $X_1, \ldots X_n$ are independent (not always necessary) then concentration holds for a large class of "smooth" f.

Example 11.1.1. Let $\mathcal{X}^n = \mathbb{R}^n$ and let $\mathbb{P}^{(X_1,\ldots,X_n)}$ be uniform distribution on $\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n : \|x\| = 1\}$. We will show that (11.1) holds if f is Lipschitz. Let $\mu : \mathscr{B}(\mathbb{R}^n) \cap \mathbb{S}^{n-1} \to [0, 1]$ denote the uniform probability measure on $\mathbb{S}^{n-1}$. (Recall that $\mu = \mathbb{P}^{(\frac{Y}{\|Y\|})}$ for $Y \sim N(0, I_n)$.)

We have the following facts:

(A) (Levy's isoperimetric inequality). Define a half-cap $A = \{x \in \mathbb{S}^{n-1} : x^{\mathrm{T}}u \leq 0\}$ for an arbitrary $u \in \mathbb{S}^{n-1}$, and for $t > 0$, (note $\mu(A) = \frac{1}{2}$) define $A_t := \{x \in \mathbb{S}^{n-1} : d_{\text{geo}}(x, A) < t\}$ where $d_{\text{geo}} := \arccos x^{\mathrm{T}}y$ (geodesic distance) for $x, y \in \mathbb{S}^{n-1}$. Then, if $B \in \mathscr{B}(\mathbb{R}^n) \cap \mathbb{S}^{n-1}$ and $\mu(B) \geq \frac{1}{2}$, then, $\forall t \geq 0$, $\mu(B_t) \geq \mu(A_t)$. If $\mu(B) = \frac{1}{2}$, then we have $\mu(B_t \setminus B) \geq \mu(A_t \setminus A)$.

(B) We have that $\mu(A_t^c) \leq e^{-(n-1)\frac{t^2}{2}}$. Intuitively, since $\cos(\frac{\pi}{2} - t) \geq \frac{2}{\pi}t$, $\forall t \in [0, \frac{\pi}{2}]$,

$$\begin{aligned} A_t^c &= \left\{x \in \mathbb{S}^{n-1} : d_{\text{geo}}(x, u) \leq \frac{\pi}{2} - t\right\} \\ &= \left\{x \in \mathbb{S}^{n-1} : x^{\mathrm{T}}u \geq \cos(\frac{\pi}{2} - t)\right\} \\ &\subseteq \left\{x \in \mathbb{S}^{n-1} : x^{\mathrm{T}}u \geq \frac{2}{\pi}t\right\}. \end{aligned}$$

Taking $u = \frac{1}{\sqrt{n}}(1,1,\ldots,1)$, and using the fact that if $Y \sim N(0, I_n)$, then $\|Y\| := (\sum_{i=1}^n Y_i^2)^{1/2}$ "concentrates" around $\sqrt{n}$, we have, for $n \geq 2$,

$$\mu(A_t^c) = \mathbb{P}\left(\frac{Y}{\|Y\|} \in A_t^c\right) \leq \mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Y_i}{\|Y\|} \geq \frac{2}{\pi}t\right) \leq e^{-Cnt^2}$$

for some universal $C > 0$.

Now, let $f : \mathbb{R}^n \to \mathbb{R}$ be L-Lipschitz w.r.t. $\|\cdot\|_2$. Since $\|x-y\|_2 \leq d_{\text{geo}}(x,y), \forall x,y \in \mathbb{S}^{n-1}$, f is L-Lipschitz w.r.t. d_{geo}.

Let $M_f :=$ median $f(X)$ where $X = (X_1,\ldots,X_n)$ is uniform on $\mathbb{S}^{n-1}$ and let $B := \{x \in \mathbb{S}^{n-1} : f(x) - M_f \leq 0\}$. Then, $\mu(B) = \mathbb{P}(f(X) - M_f \leq 0) \geq \frac{1}{2}$ and $\forall t > 0$, $\forall x \in \mathbb{S}^{n-1}$, $d_{\text{geo}}(x, B) < t$ implies $\exists y \in B$ such that $d_{\text{geo}}(x,y) < t$. Hence $f(x) - M_f = f(x) - f(y) + f(y) - M_f < L_t$. So $B_{\frac{t}{L}} \subseteq \{x \in \mathbb{S}^{n-1} : f(x) - M_f < t\}$.

Hence, $\forall t > 0$,

$$\begin{aligned}\mathbb{P}(f(X) - M_f \geq t) &= \mu\left\{x \in \mathbb{S}^{n-1} : f(x) - M_f \geq t\right\}\\ &\leq \mu\{B^c_{\frac{t}{L}}\} \leq \mu\{A^c_{\frac{t}{L}}\} \leq e^{-\frac{n-1}{2}\frac{t^2}{L^2}}\end{aligned}$$

By taking $B := \{x \in \mathbb{S}^{n-1} : f(x) - M_f \geq 0\}$, we also obtain

$$\mathbb{P}(f(X) - M_f \leq -t) \leq e^{-\frac{n-1}{2}\frac{t^2}{L^2}}. \tag{11.2}$$

Remark 11.1.2. Let $(\mathcal{X}^n, d)$ be a metric space and let $f : \mathcal{X} \to \mathbb{R}$ be 1-Lipschitz w.r.t $d(\cdot,\cdot)$. Write $X := (X_1,\ldots,X_n)$ and define $\forall t > 0$,

$$\begin{aligned}\alpha(t) :=& \sup\left\{\mathbb{P}(d(X,A) > t) := A \subseteq \mathcal{X}^n, \mathbb{P}(X \in A) \geq \frac{1}{2}\right\}\\ =& \sup\left\{\mathbb{P}^{(X)}(A_t) : A \subseteq \mathcal{X}^n, \mathbb{P}^{(X)}(A) = \frac{1}{2}\right\}.\end{aligned}$$

Then $\mathbb{P}(|f(X) - M_f| \geq t) \leq 2\alpha(t)$. The $\alpha(\cdot)$ is called concentration function.

As another example, take $\mathcal{X}^n = \mathbb{R}^n$, $\|\cdot\|_2$-metric, and $\mathbb{P}^{(X)} = N(0, I_n)$. Then, by Gaussian isoperimetric inequality, $\forall t > 0$, the supremum in $\alpha(t)$ is attained for half-space $A = \{x \in \mathbb{R}^n : x^{\mathrm{T}}u \leq 0\}$ for any $u \in \mathbb{R}^n$, $u \neq 0$.

Since $A_t^c \subseteq \{x \in \mathbb{R}^n : x^{\mathrm{T}}\frac{u}{\|u\|} > t\}$, we have that

$$\alpha(t) \leq \mathbb{P}(X^{\mathrm{T}}e_1 > t) \leq \frac{1}{\sqrt{2\pi}}\frac{1}{t}e^{-\frac{t^2}{2}}.$$

Another important example: let $\mathcal{X}^n$ be arbitrary and for $x, y \in \mathcal{X}^n$, define the convex distance

$$d^\star(x,y) := \sup\left\{\sum_{i=1}^n \alpha_i \mathbb{1}_{\{x_i \neq y_i\}} : \alpha \in [0,\infty)^n, \|\alpha\|_2 \leq 1\right\}.$$

Then, for any product measure $\mathbb{P}^{(X)}$, we have

$$\alpha(t) \leq 2e^{-\frac{t^2}{4}}. \qquad \text{(Talagrand's convex distance inequality)}$$

To see why this is useful, let $\mathcal{X}^n = \{-1,+1\}^n$ and let $X_1,\ldots,X_n$ be iid Rademacher. Let $\mathcal{A} \subseteq \mathbb{R}^n$ be countable and define

$$f(x) = \sup\left\{\sum_{i=1}^n a_i x_i : a \in \mathcal{A}\right\} \quad \forall x \in \{-1,+1\}^n.$$

Then, for any $x, y \in \{-1,+1\}^n$, writing $\nu := \sup_{a\in\mathcal{A}} \|a\|_2$,

$$|f(x) - f(y)| \le \sup\left\{\sum_{i=1}^n a_i(x_i - y_i) : a \in \mathcal{A}\right\} \le \nu d^\star(x, y).$$

So $\mathbb{P}(|f(x) - M_f| > t) \le 4e^{-\frac{t^2}{4\nu^2}}$.

Remark 11.1.3. Let $\mathscr{F}_{ab} \subseteq \{g \in [a,b]^{\mathcal{X}}\}$ and let $f(x) = \sup_{g\in\mathscr{F}_{ab}} \sum_{i=1}^n g(x_i)$ for $x \in \mathcal{X}^n$. We will show that if $(X_1, \dots, X_n)$ are independent, then write $Z := \sup_{g\in\mathscr{F}_{ab}} \sum_{i=1}^n g(x_i)$,

$$\mathbb{P}(Z - \mathbb{E}Z > t) \le e^{-\frac{2t^2}{n(b-a)^2}}.$$

The magnitude of $\mathbb{E}Z = \mathbb{E}\sup_{g\in\mathscr{F}_{ab}} \sum_{i=1}^n g(X_i)$ depends on the complexity of $\mathscr{F}_{ab}$. This concentration (weak version of Talagrand's Empirical Process Inequality) requires a different proof technique from isoperimetry. It uses the so-called entropy method: log-Sobolev inequality and Herbst argument.

11.2 Chernoff Bounds

Definition 11.2.1. For a random variable Z, define $\psi_Z : \mathbb{R} \to (-\infty, \infty]$ as

$$\psi_Z(\lambda) = \log \mathbb{E}e^{\lambda Z}.$$

Note that $\psi_Z(0) = 0$. Define also $\psi_Z^\star : \mathbb{R} \to [0, \infty]$ by

$$\psi_Z^\star(t) := \sup_{\lambda\ge 0} \lambda t - \psi_Z(\lambda).$$

Proposition 11.2.1. (a) ψ_Z is convex on $\mathbb{R}$. In particular, if $\exists\lambda > 0$ such that $\psi_Z(\lambda) < \infty$, then $\psi_Z < \infty$ on $[0, b)$ where $b := \sup\{\lambda \ge 0 : \psi_Z(\lambda) < \infty\}$.

(b) If $\exists\lambda > 0$ such that $\psi_Z(\lambda) < \infty$, then ψ_Z is continuous differentiable on $[0, b)$ and $\psi_Z'(\lambda) = \frac{\mathbb{E}Ze^{\lambda Z}}{\mathbb{E}e^{\lambda Z}}$ for any $\lambda \in \mathbb{R}$. In particular, if $\mathbb{E}Z = 0$, then $\psi_Z'(0+) = 0$.

(c) If $\mathbb{E}|Z| < \infty$, then $\psi_Z(\lambda) = \psi_{Z-\mathbb{E}Z}(\lambda) + \lambda\mathbb{E}Z$ for all $\lambda \in \mathbb{R}$.

(d) We have $\psi_Z^\star(t) \ge 0$ for all $t \in \mathbb{R}$. For all $t \le \mathbb{E}Z$, $\psi_Z^\star(t) = 0$.

(e) $\forall t \in \mathbb{R}$, $\mathbb{P}(Z \ge t) \le e^{-\psi_Z^\star(t)}$.

(f) If $Z_1, \dots, Z_n$ are independent, then $\psi_{Z_1+\cdots+Z_n}(\lambda) = \sum_{k=1}^n \psi_{Z_k}(\lambda)$, $\forall\lambda \in \mathbb{R}$. Also, if $Z_1, \dots, Z_n$ are i.i.d., then $\forall t \in \mathbb{R}$, $\psi^\star_{Z_1+\cdots+Z_n}(t) = n\psi^\star_{Z_1}(\frac{t}{n})$. In particular, for any $t \in \mathbb{R}$, we have that $\mathbb{P}(Z_1 + \dots Z_n \ge nt) \le e^{-n\psi^\star_{Z_1}(t)}$.

Proof.

(a) Let $\rho \in [0, 1]$, and let $\lambda_1, \lambda_2 \in \mathbb{R}$, then

$$\begin{aligned}\psi_Z(\rho\lambda_1 + (1-\rho)\lambda_2) &= \log \mathbb{E}e^{\rho\lambda_1 Z}e^{(1-\rho)\lambda_2 Z} \\ &\le \log\left\{\mathbb{E}(e^{\rho\lambda_1 Z})^{\frac{1}{\rho}}\right\}^\rho \left\{\mathbb{E}(e^{(1-\rho)\lambda_2 Z})^{\frac{1}{1-\rho}}\right\}^{1-\rho} \\ &= \rho\psi_Z(\lambda_1) + (1-\rho)\psi_Z(\lambda_2)\end{aligned}$$

So ψ_Z is convex.

(b) Let $\lambda \in (0,b)$. There exists $\varepsilon > 0$ such that $U := [\lambda-\varepsilon, \lambda+\varepsilon] \subset (0,b)$. Note that

$$\mathbb{E}\sup_{\tilde{\lambda}\in U} \frac{\partial e^{\lambda Z}}{\partial \lambda}(\tilde{\lambda}) = \mathbb{E}\sup_{\tilde{\lambda}\in U} Ze^{\tilde{\lambda}Z} \leq \mathbb{E}|Z|e^{(\lambda+\varepsilon)Z}\mathbb{1}_{\{Z\geq 0\}} + \mathbb{E}|Z|\mathbb{1}_{\{Z\leq 0\}} < \infty,$$

where the last inequality follows because $\lambda+\varepsilon < b$ and hence, there exists $b' \in (\lambda+\varepsilon, b)$ and $C > 0$ such that $|z|e^{(\lambda+\varepsilon)z} \leq e^{b'z}$ for all $z \geq C$.

Therefore, by dominated convergence theorem,

$$\frac{\partial}{\partial\lambda}\mathbb{E}e^{\lambda Z} = \mathbb{E}\frac{\partial}{\partial\lambda}e^{\lambda Z} = \mathbb{E}Ze^{\lambda Z}.$$

Hence, $\psi_Z'(\lambda) = \frac{\mathbb{E}Ze^{\lambda Z}}{\mathbb{E}e^{\lambda Z}}$, which is continuous on $(0,b)$.

If $\lambda = 0$, then we proceed in the same way except we only take right derivatives.

(c) Follows by algebra.

(d) Note that $\psi_Z^\star(t) \geq -\psi_Z(0) = 0$ for all $t \in \mathbb{R}$. By Jensen's inequality, we have $\psi_Z(\lambda) = \log \mathbb{E}e^{\lambda Z} \geq \lambda\mathbb{E}Z$. Therefore, if $t \leq \mathbb{E}Z$, then $\lambda t - \psi_Z(\lambda) \leq \lambda(t-\mathbb{E}Z) \leq 0$.

(e) $\forall t \in \mathbb{R}$, $\forall \lambda \geq 0$,

$$\mathbb{P}(Z > t) = \mathbb{P}(e^{\lambda Z} > e^{\lambda t}) \leq \frac{\mathbb{E}e^{\lambda Z}}{e^{\lambda t}} = \exp(-\lambda t + \psi_Z(\lambda)).$$

Taking infimum over λ yields desire conclusion.

(f) $\psi_{Z_1+\cdots+Z_n}(\lambda) = \log\prod_{k=1}^n \mathbb{E}e^{\lambda Z_k} = \sum_{k=1}^n \psi_{Z_k}(\lambda)$. Now, for $t \in \mathbb{R}$,

$$\psi^\star_{Z_1+\cdots+Z_n}(t) = \sup_{\lambda\geq 0} \lambda t - n\psi_Z(\lambda) = n\sup_{\lambda\geq 0}\lambda\frac{t}{n} - \psi_Z(\lambda) = n\psi_Z^\star(\frac{t}{n}).$$

□

Remark 11.2.1. Let Z be a r.v. such that $\psi_Z(\lambda), \psi_{-Z}(\lambda) < \infty$ for some $\lambda > 0$. Then $\mathbb{E}e^{\lambda|Z|} < \infty$ and hence

$$\mathbb{E}e^{\lambda|Z|} = 1 + \sum_{k=1}^\infty \frac{\lambda^k\mathbb{E}|Z|^k}{k!}.$$

For $t \geq 0$, define $M^\star(t) := \inf_{k\in\mathbb{N}} \frac{\mathbb{E}|Z|^k}{t^k}$ so that $\mathbb{P}(|Z| > t) \leq \inf_{k\in\mathbb{N}} \mathbb{P}(|Z|^k > t^k) \leq M^\star(t)$. Then, for any $t > 0$, for any $\lambda \geq 0$.

$$\frac{\mathbb{E}e^{\lambda|Z|}}{e^{\lambda t}} = \frac{1+\sum_{k=1}^\infty \frac{\lambda^k\mathbb{E}|Z|^k}{k!}}{1+\sum_{k=1}^\infty \frac{\lambda^k t^k}{k!}} \geq M^\star(t)$$

since $\forall k \in \mathbb{N}$, $\frac{\mathbb{E}|Z|^k}{t^k} \geq M^\star(t)$. Thus, $\inf_{\lambda\geq 0}\mathbb{E}e^{\lambda|Z|-\lambda t} \geq M^\star(t)$, which implies that Chernoff bound is no better than the best moment bound. We prefer Chernoff bound for convenience however.

Remark 11.2.2. Let $Z_1, Z_2, \ldots, Z_n$ be iid random variables, supported on $\mathbb{R}$, such that $\psi_{Z_1}(\lambda) < \infty$ for all $\lambda \geq 0$. Then, Chernoff's bound is optimal for $S_n := \sum_{i=1}^n Z_i$. We will later show Cramér's theorem, which implies that for all $t \geq \mathbb{E}Z_1$,

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}(S_n \geq nt) = -\psi^\star_{Z_1}(t).$$

Equivalently,

$$\mathbb{P}(S_n \geq nt) = \exp\{-n(\psi^\star_{Z_1}(t) + o(1))\},$$

where the $o(1)$ sequence depends on t.

Example 11.2.1. (a) Let $Z \sim N(0, \sigma^2)$, then $\psi_Z(\lambda) = \frac{\lambda^2\sigma^2}{2}$, $\forall \lambda \in \mathbb{R}$. To compute $\psi_Z^\star(t)$ for $t \geq 0$, we write $\tilde{\lambda} = \arg\max_{\lambda \geq 0} t\lambda - \psi_Z(\lambda)$ and observe that $t = \tilde{\lambda}\sigma^2$ and hence $\tilde{\lambda} = \frac{t}{\sigma^2}$. So $\psi_Z^\star(t) = \frac{t^2}{\sigma^2} - \frac{t^2}{2\sigma^2} = \frac{1}{2}\frac{t^2}{\sigma^2}$.

(b) Let $Y \sim \text{Poisson}(\nu)$ for $\nu > 0$. Define $Z := Y - \nu$. Then, $\forall \lambda \in \mathbb{R}$,

$$\mathbb{E}e^{\lambda Z} = \sum_{k=0}^{\infty} \frac{\nu^k e^{-\nu}}{k!} e^{\lambda(k-\nu)} = e^{-\nu\lambda-\nu} \sum_{k=0}^{\infty} \frac{(\nu e^\lambda)^k}{k!} = e^{-\lambda\nu-\nu} e^{\nu e^\lambda}$$

So $\psi(\lambda) = -\lambda\nu - \nu + \nu e^\lambda$. Let $t \geq 0$, solving $t = -\nu + \nu e^{\tilde{\lambda}}$, we get $\tilde{\lambda} = \log(1 + \frac{t}{\nu})$. So

$$\begin{aligned}
\psi_Z^\star(t) &= t\log(1+\frac{t}{\nu}) + \nu\log(1+\frac{t}{\nu}) + \nu - \nu(1+\frac{t}{\nu}) \\
&= \nu\left\{(1+\frac{t}{\nu})\log(1+\frac{t}{\nu}) - \frac{t}{\nu}\right\} := \nu h(\frac{t}{\nu}),
\end{aligned}$$

where $h(x) = (1+x)\log(1+x) - x$, $\forall x \geq 0$. Note that $h(x) = \frac{x^2}{2} - \frac{x^3}{6} + \frac{x^4}{12} - \frac{x^5}{20} + O(x^6)$.

(c) Let $Y \sim \text{Ber}(p)$ for $p \in (0,1)$, and let $Z = Y - p$. $\forall \lambda \in \mathbb{R}$,

$$\begin{aligned}
&\mathbb{E}e^{\lambda Z} = pe^{\lambda(1-p)} + (1-p)e^{-\lambda p} = e^{-\lambda p}(pe^\lambda + (1-p)) \\
\implies &\psi_Z(\lambda) = -\lambda p + \log(pe^\lambda + (1-p)).
\end{aligned}$$

Let $t \in [0, 1-p]$ so that $1-(p+t) \geq 0$, setting $t = -p + \frac{pe^\lambda}{pe^\lambda + (1-p)}$, we obtain $\tilde{\lambda} = \log\frac{(1-p)(t+p)}{p(1-p-t)}$. Write $a := p + t$, we have

$$\begin{aligned}
\psi_Z^\star(t) &= (a-p)\log\frac{1-p}{p}\frac{a}{1-a} - \log\left(\frac{(1-p)a}{1-a} + (1-p)\right) + p\log\frac{1-p}{p}\frac{a}{1-a} \\
&= a\log\frac{a}{p} + (1-a)\log\frac{1-a}{1-p} \\
&= \text{KL}(\text{Ber}(p+t)\|\text{Ber}(p)).
\end{aligned}$$

(d) Recall that for $a, b \geq 0$, X is Gamma(a,b) if X has density

$$f(x) = \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-\frac{x}{b}} \mathbb{1}_{\{x \geq 0\}},$$

where $\Gamma(a) := \int_0^\infty x^{a-1}e^{-x}\mathrm{d}x$. Note that Gamma$(1,b)$ is Exponential(b). Note also that $\mathbb{E}X = ab$ and $\text{Var}(X) = ab^2$. Define $Z := X - ab$. For all $\lambda \in [0, \frac{1}{b})$, we have

$$\begin{aligned}
\mathbb{E}e^{\lambda Z} &= \frac{1}{\Gamma(a)b^a}\int_0^\infty e^{\lambda(x-ab)} x^{a-1} e^{-\frac{x}{b}} \mathrm{d}x \\
&= \frac{e^{-\lambda ab}}{\Gamma(a)b^a}\Gamma(a)\frac{1}{(b^{-1}-\lambda)^a} = e^{-\lambda ab}\frac{1}{(1-b\lambda)^a}.
\end{aligned}$$

Thus, $\psi_Z(\lambda) = -a\log(1-b\lambda) - \lambda ab$. To get a more tractable upper bound on the tail probability, we will take an upper bound of $\psi_Z(\lambda)$. Note that if $\tilde{\psi}_Z(\lambda) \geq \psi_Z(\lambda)$ for all $\lambda \geq 0$, then for all $t \in \mathbb{R}$,

$$t\lambda - \psi_Z(\lambda) \geq t\lambda - \tilde{\psi}_Z(\lambda)$$

and hence, $\psi_Z^\star(t) \geq \tilde{\psi}_Z^\star(t)$.

Note that $\forall u \in [0,1)$, by Taylor expansion,

$$\begin{aligned} &-(1-u)\log(1-u) - u \le -\frac{u^2}{2} \\ \implies &-(1-u)\log(1-u) - u + u^2 \le \frac{u^2}{2} \\ \implies &-\log(1-u) - u \le \frac{u^2}{2(1-u)}. \end{aligned}$$

Thus, for $\lambda \in [0, \frac{1}{b})$ define $\widetilde{\psi}(\lambda) = \frac{ab^2\lambda^2}{2(1-b\lambda)} := \frac{\nu\lambda^2}{2(1-c\lambda)}$ where $\nu := ab^2$, $c := b$; note that $\tilde{\psi}(\lambda) \ge \psi(\lambda)$. By differentiation, one may check that $\widetilde{\psi}$ is convex on $[0, \frac{1}{c})$. We may thus solve the $\sup_{\lambda \ge 0} \lambda t - \widetilde{\psi}(\lambda)$. For $t \ge 0$, one may set

$$t = \frac{1}{4(1-c\lambda)^2}\{4\nu\lambda(1-c\lambda) + 2\nu c\lambda^2\} = \frac{2\lambda\nu(2-\lambda c)}{4(1-c\lambda)^2}$$

to obtain $\lambda = \frac{1}{c}(1 - (1 + 2\frac{c}{\nu}t)^{-\frac{1}{2}})$ (Hint: set $s = 1 - c\lambda$ to get $t = \frac{1}{4s^2}\frac{2\nu}{c}(1 - s^2)$) and that

$$\tilde{\psi}_Z^\star(t) = t\lambda - \frac{\lambda^2\nu}{2(1-c\lambda)} = \frac{\nu}{c^2}\left\{1 + \frac{c}{\nu}t - \sqrt{1 + 2\frac{c}{\nu}t}\right\} := \frac{\nu}{c^2}h_1(\frac{c}{\nu}t)$$

where $h_1(u) := 1 + u - \sqrt{1+2u}$, $\forall u \ge 0$. So $\forall t \ge 0$,

$$\mathbb{P}(X > t) \le \exp\left(-\frac{\nu}{c^2}h_1(\frac{c}{\nu}t)\right).$$

Note further that

(1) $h_1(u) = \frac{u^2}{2} - \frac{u^3}{2} + O(u^4)$

(2) $h_1(u) \ge \frac{u^2}{2(1+u)}$, $\forall u \ge 0$ (Hint: Consider $u \mapsto h_1(u) - \frac{u^2}{2(1+u)}$ and differentiate) and thus

$$\mathbb{P}(X > t) \le \exp\left\{-\frac{\nu}{c^2}\frac{(\frac{c}{\nu}t)^2}{2(1+\frac{c}{\nu}t)}\right\} = \exp\left\{-\frac{t^2}{2(\nu + ct)}\right\} \le \begin{cases} e^{-\frac{t^2}{4\nu}} & \text{if } t \le \frac{\nu}{c} \\ e^{-\frac{t}{4c}} & \text{if } t > \frac{\nu}{c} \end{cases}$$

(3) h_1 is invertible on $[0,\infty) \to [0,\infty)$ and $h_1^{-1}(w) = w + \sqrt{2w}, \forall w \in [0,\infty)$ (Hint: set $s = \sqrt{1+2u}$ and solve $s^2 - s - (\frac{s^2-1}{2}) = w$).

$$\mathbb{P}(X \ge \sqrt{2\nu w} + cw) \le e^{-w} \quad \forall w \ge 0.$$

Definition 11.2.2. Let X be a random variable. We say that

(a) X is sub-Gasussian with variance-factor $\nu > 0$ if

$$\psi_X(\lambda) = \log \mathbb{E}e^{\lambda X} \le \frac{\nu\lambda^2}{2} \quad \forall \lambda \in \mathbb{R}.$$

Denote $sG(\nu) := \{\text{all centered sub-Gaussian r.v.}\}$.

(b) X is right-sub-Gamma with variance-factor $\nu > 0$ and scale $c \ge 0$ if

$$\psi_X(\lambda) \le \frac{\lambda^2\nu}{2(1-c\lambda)} \quad \forall \lambda \in [0, \frac{1}{c})$$

and left-sub-Gamma if $-X$ is right-sub-Gamma, or, equivalently,

$$\psi_X(\lambda) \le \frac{\lambda^2\nu}{2(1+c\lambda)} \quad \forall \lambda \in (-\frac{1}{c}, 0].$$

Denote $s\Gamma_+(\nu, c)$ and $s\Gamma_-(\nu, c)$ as the set of all right and left sub-Gamma random variables. Write $s\Gamma(\nu, c) := s\Gamma_+(\nu, c) \cap s\Gamma_-(\nu, c)$.

(c) We say X is sub-Exponential with variance-factor $\nu > 0$ if $X \in s\Gamma(\nu, \sqrt{\nu})$. Note that $s\Gamma(\nu, 0) = sG(\nu)$.

Remark 11.2.3. We do not require a random variable X to be centered in order to characterize it as sub-Gaussian or sub-Exponential. In most applications where we have to apply tail bounds, we consider only centered random variables.

Proposition 11.2.2. Let X be a random variable.
Tail Characterization:

(1a) Let $\nu > 0$. If $X \in sG(\nu)$, then

$$\mathbb{P}(X > t) \vee \mathbb{P}(X < -t) \leq e^{-\frac{t^2}{2\nu}} \quad \forall t \geq 0 \tag{11.3}$$

If on the other hand (11.3) holds, then $X \in sG(C\nu)$ for a universal $C > 0$.

(1b) Let $\nu > 0, c \geq 0$. If $X \in s\Gamma(\nu, c)$, then

$$\mathbb{P}(X > t) \vee \mathbb{P}(X < -t) \leq \exp\left(-\frac{t^2}{2(\nu + ct)}\right) \quad \forall t \geq 0 \tag{11.4}$$

$$\mathbb{P}(X > t) \vee \mathbb{P}(X < -t) \leq \exp\left(-\frac{t^2}{4\nu} \wedge \frac{t}{4c}\right) \quad \forall t \geq 0 \tag{11.5}$$

$$\mathbb{P}(X > \sqrt{2\nu t} + ct) \vee \mathbb{P}(X < -\sqrt{2\nu t} - ct) \leq e^{-t} \quad \forall t \geq 0. \tag{11.6}$$

If on the other hand any one of (11.4), (11.5), (11.6) holds, then $X \in s\Gamma(C\nu, C'c)$ for universal $C, C' > 0$.

Moment Characterization:

(2a) If $X \in sG(\nu)$, then $\forall q \in \mathbb{N}$, we have $\mathbb{E}X^{2q} \leq q!(4\nu)^q$, which implies by Stirling's formula that

$$\|X\|_{L_{2q}(\mathbb{P})} \leq C\sqrt{\nu}\sqrt{q}$$

for an universal $C > 0$. (Note that $\mathbb{E}|X|^{2q-1} \leq (\mathbb{E}X^{2q})^{\frac{2q-1}{2q}}$.) Conversely, if $\exists A > 0$ such that $\forall q \in \mathbb{N}$, $\mathbb{E}X^{2q} \leq q!A^q$, then $X \in sG(4A)$.

(2b) If $X \in s\Gamma(\nu, c)$, then $\forall q \in \mathbb{N}$, we have $\mathbb{E}X^{2q} \leq q!(8\nu)^q + (2q)!(4c)^{2q}$, which implies

$$\|X\|_{L_{2q}(\mathbb{P})} \leq C(\sqrt{\nu q} + cq)$$

for an universal $C > 0$. Conversely, if $\exists A, B \geq 0$ such that $\forall q \in \mathbb{N}$, $\mathbb{E}X^{2q} \leq q!A^q + (2q)!B^{2q}$, then $X \in s\Gamma(4(A + B^2), 2B)$.

Proof.

We prove the results for sub-Gaussian.

(1a) Since $\psi_X(\lambda) \leq \frac{\lambda^2\nu}{2}$, for all $t \in \mathbb{R}$, $\psi_X^\star(t) \geq \frac{t^2}{2\nu}$. The desired conclusion follows from Proposition 11.2.1.
Conversely, let us assume (11.3). Then, for any $\lambda > 0$, we have

$$\begin{aligned}
\mathbb{E}e^{\lambda X} &= \int_0^\infty \mathbb{P}(e^{\lambda X} \geq t)\,dt \\
&= \int_0^\infty \mathbb{P}(X \geq \lambda^{-1}\log t)\,dt \\
&= \int_0^\infty \mathbb{P}(X \geq s)\lambda e^{\lambda s} ds \\
&\leq \int_0^\infty e^{-\frac{s^2}{2\nu} + \lambda s}\lambda\,ds \\
&\leq \sqrt{2\pi\nu}\lambda e^{\frac{\lambda^2\nu}{2}}.
\end{aligned}$$

The conclusion then follows from the fact that for all $xe^{x^2} \leq e^{2x^2}$ for all $x \geq 1$.

(2a) Assume $X \in sG(\nu)$ and is centered. Then $\forall t \geq 0$, $\mathbb{P}(|X| > t) \leq 2e^{-\frac{t^2}{2\nu}}$. Thus, for any $q \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E}X^{2q} = \int_0^\infty \mathbb{P}(X^{2q} > t)\mathrm{d}t &= \int_0^\infty \mathbb{P}(|X| > t^{\frac{1}{2q}})\mathrm{d}t \\ &= 2q\int_0^\infty \mathbb{P}(|X| > s)s^{2q-1}\mathrm{d}s \\ &\leq 4q\int_0^\infty e^{-\frac{s^2}{2\nu}}s^{2q-1}\mathrm{d}s \\ &= 2q\int_0^\infty e^{-\frac{x}{2\nu}}x^{q-1}\mathrm{d}x \\ &= 2q\Gamma(q)(2\nu)^q \leq q!(4\nu)^q. \end{aligned}$$

Conversely, assume $\forall q \in \mathbb{N}$, $\mathbb{E}X^{2q} \leq q!C^q$. Let X, X' be iid. Note that $\forall \lambda \geq 0$, $\mathbb{E}e^{\lambda|X|} < \infty$ implies

$$\begin{aligned} \mathbb{E}e^{\lambda X} &\leq \mathbb{E}e^{\lambda X}\mathbb{E}e^{-\lambda X'} && \text{(b/c } \mathbb{E}e^{-\lambda X'} \geq e^{-\lambda\mathbb{E}X'} = 1 \text{ since } \mathbb{E}X' = 0) \\ &= \mathbb{E}e^{\lambda(X-X')} \\ &= \sum_{q=0}^\infty \frac{\lambda^{2q}\mathbb{E}(X-X')^{2q}}{(2q)!} \\ &= \sum_{q=0}^\infty \frac{\lambda^{2q}}{(2q)!}\mathbb{E}\left[\frac{(2X)^{2q}}{2} + \frac{(2X')^{2q}}{2}\right] \\ &= \sum_{q=0}^\infty \frac{\lambda^{2q}}{(2q)!}2^{2q}q!C^q && \text{(Note } \tfrac{q!}{(2q)!} \leq \tfrac{1}{2^q q!}) \\ &\leq \sum_{q=0}^\infty \frac{(2C\lambda^2)^q}{q!} \leq e^{2C\lambda^2}, \end{aligned}$$

as desired.

$\square$

Definition 11.2.3. For a random variable Z and $\psi : [0, \infty) \to [0, \infty)$ convex, non-decreasing, and $\psi(0) = 0$, define the Orlicz norm

$$\|Z\|_\psi := \inf\left\{t \geq 0 : \mathbb{E}\psi\left(\frac{|Z|}{t}\right) \leq 1\right\}.$$

We define $\inf\emptyset = \infty$ by convention.

Remark 11.2.4. (a) $\|\cdot\|_\psi$ is a norm on $\{Z \text{ random variable} : \|Z\|_\psi < \infty\}$.

To verify triangle inequality, let Z, Y be random variables, possibly dependent. We note that, for any $s > \|Z\|_\psi$ and $t > \|Y\|_\psi$ (so that $\mathbb{E}\psi(|Z|/s) \leq 1$ and $\mathbb{E}\psi(|Y|/t) \leq 1$ by monotonicity of ψ),

$$\begin{aligned} \mathbb{E}\psi\left(\frac{|Z+Y|}{s+t}\right) &= \mathbb{E}\psi\left(\frac{|Z|}{s}\frac{s}{s+t} + \frac{|Y|}{t}\frac{t}{s+t}\right) \\ &\leq \frac{s}{s+t}\mathbb{E}\psi(\frac{|Z|}{s}) + \frac{t}{s+t}\mathbb{E}\psi(\frac{|Y|}{t}) \leq 1. \end{aligned}$$

Therefore, $\|Z+Y\|_\psi \leq \inf\{s+t : s \geq \|Z\|_\psi, t \geq \|Y\|_\psi\} = \|Z\|_\psi + \|Y\|_\psi$.

(b) With $\psi(z) = z^p$ for $p \geq 1$, $\|Z\|_\psi = \|Z\|_{L_p(\mathbb{P})}$.

(c) If $\|X\|_\psi < \infty$, then the inf is attained, i.e., $\mathbb{E}\psi(\frac{|X|}{\|X\|_\psi}) = 1$. (Hint: let $C_n \searrow \|X\|_\psi$ and use the monotone convergence theorem.)

(d) if $\mathbb{E}\psi(\frac{|X|}{A}) \le b$ for $A > 0$, $b \ge 1$, then $\|X\|_\psi \le Ab$. (Since $\psi(\frac{|X|}{bA}) \le \frac{1}{b}\psi(\frac{|X|}{A})$ by convexity and the fact $\psi(0) = 0$.)

(e) For any $t \ge 0$,

$$\begin{aligned}
\mathbb{P}(|Z| \ge t) &= \mathbb{P}\left(\psi(\frac{|Z|}{\|Z\|_\psi}) \ge \psi(\frac{t}{\|Z\|_\psi})\right) \\
&\le \frac{\mathbb{E}\psi(\frac{|Z|}{\|Z\|_\psi})}{\psi(\frac{t}{\|Z\|_\psi})} \\
&\le \frac{1}{\psi(\frac{t}{\|Z\|_\psi})}.
\end{aligned}$$

(f) Define $\psi_1(t) = e^t - 1$ and $\psi_2(t) = e^{t^2} - 1$. Then, for random variables Z, Y, possibly dependent,

$$\|ZY\|_{\psi_1} \le \|Z\|_{\psi_2} \|Y\|_{\psi_2}.$$

(Hint: Use the fact that $a^2 + b^2 \ge 2ab$, $\forall a, b \in \mathbb{R}$.)

Proposition 11.2.3 (Orlicz Norm Characterization). Let X be a random variable.

(a) Let $\psi_2(z) = e^{z^2} - 1$. $X \in sG(\nu)$ implies $\|X\|_{\psi_2} \le 2\sqrt{\nu}$. If $\|X\|_{\psi_2} \le A < \infty$, then $X \in sG(CA^2)$ for universal $C > 0$.

(b) Let $\psi_1(z) = e^z - 1$. $X \in sE(\nu)$ implies $\|X\|_{\psi_1} \le 2\sqrt{\nu}$. If $\|X\|_{\psi_1} \le A < \infty$, then $X \in sE(CA^2)$ for universal $C > 0$.

Proof.

We prove (a); the proof for (b) is similar.

Assume that $X \in sG(\nu)$ for $\nu > 0$, then for any $a > 4\nu$,

$$\begin{aligned}
\mathbb{E}e^{\frac{X^2}{a}} &= \int_1^\infty \mathbb{P}(e^{\frac{X^2}{a}} \ge t)\mathrm{d}t \\
&= \int_1^\infty \mathbb{P}(|X| \ge \sqrt{a\log t})\mathrm{d}t \le 2\int_1^\infty e^{-\frac{a}{2\nu}\log t}\mathrm{d}t \\
&\le 2\int_1^\infty t^{-2}\mathrm{d}t \le 2.
\end{aligned}$$

This implies that $\|X\|_{\psi_2} \le \sqrt{a}$ for any $a > 4\nu$. Hence, $\|X\|_{\psi_2} \le 2\sqrt{\nu}$.

Now assume $\|X\|_{\psi_2} \le A < \infty$. Then $\mathbb{E}e^{\frac{X^2}{A^2}} \le 2$. Therefore, by Markov's inequality, we have $\mathbb{P}(|X| > t) \le 2e^{-\frac{t^2}{A^2}}$ which implies $X \in sG(CA^2)$ for universal $C > 0$. $\square$

Proposition 11.2.4. Suppose $\psi : [0, \infty) \to [0, \infty)$ is convex, non-decreasing, $\psi(0) = 0$ and $\psi(1) \le 2$ and $\exists c > 0$ such that $\forall x, y \ge 1$, $\psi(x)\psi(y) \le \psi(cxy)$. Then, for any collection of random variables $X_1, \dots, X_m$, possibly dependent,

$$\left\|\max_{i\in[m]} X_i\right\|_\psi \le 3c\psi^{-1}(2m)\max_{i\in[m]}\|X_i\|_\psi.$$

Note that $\psi_2(z) = e^{z^2} - 1$ and $\psi_1(z) = e^z - 1$ satisfy the conditions with $c \le 1$ and $\psi_2^{-1}(2m) = \sqrt{\log(2m+1)}$ and $\psi_1^{-1}(2m) = \log(2m+1)$.

Proof.

Write $A := \max_{i\in[m]} \|X_i\|_\psi$ and $Y := \max_{i\in[m]} |X_i|$. Since $\psi(\frac{x}{y})\psi(y) \le \psi(cx)$ for all $y \ge 1$ and $\frac{x}{y} \ge 1$, and since $\psi^{-1}(2m) \ge 1$, we have that

$$\psi\left(\frac{Y}{\psi^{-1}(2m)cA}\right) \le \psi(\frac{Y}{A})\frac{1}{2m} + \psi(1)\mathbb{1}_{\{\frac{Y}{\psi^{-1}(2m)cA}\le 1\}} \le \frac{1}{2m}\sum_{i=1}^m \psi(\frac{|X_i|}{A}) + 2.$$

So $\mathbb{E}\psi(\frac{Y}{\psi^{-1}(2m)cA}) \le \frac{1}{2} + 2 \le 3$. It implies $\|Y\|_\psi \le 3c\psi^{-1}(2m)\max_{i\in[m]}\|X_i\|_\psi$. □

Remark 11.2.5. The different characterizations are useful in different contexts. For example, we may easily use the Orlicz norm characterization to show that for two random variables X, Y, possibly dependent, there exists a universal $C > 0$ such that

$$X \in sG(\sigma_1^2),\, Y \in sG(\sigma_2^2) \Rightarrow X + Y \in sG(C(\sigma_1+\sigma_2)^2).$$

On the other hand, if X, Y are independent, then the moment generating function characterization easily shows that

$$X \in sG(\sigma_1^2),\, Y \in sG(\sigma_2^2) \Rightarrow X + Y \in sG(\sigma_1^2+\sigma_2^2).$$

11.3 Basic Inequalities

Theorem 11.3.1 (Hoeffding Inequality). (a) Let Z be a centered r.v. such that $Z \in [a,b] \subseteq \mathbb{R}$. Then, $\forall\lambda \in \mathbb{R}$, $\psi_Z(\lambda) := \log \mathbb{E}e^{\lambda Z} \le \frac{\lambda^2(b^2-a^2)}{8}$. So $Z \in sG(\frac{(b-a)^2}{4})$.

(b) Let $X_1, \dots, X_n$ be independent centered r.v. where $X_i \in [a_i, b_i]$ and write $S_n = \sum_{i=1}^n X_i$, then $S_n \in sG(\frac{1}{4}\sum_{i=1}^n (b_i-a_i)^2)$ and $\mathbb{P}(S_n > t) \le \exp(-\frac{2t^2}{\sum_{i=1}^n (b_i-a_i)^2})$.

Proof.

First note that $\forall\lambda \in \mathbb{R}$, $\mathbb{E}|Z|e^{\lambda|Z|} \le (b-a)e^{\lambda(b-a)} < \infty$ and likewise for $\mathbb{E}|Z|^2e^{\lambda|Z|}$. Also, for any r.v. W on $[a,b]$, note that

$$\mathrm{Var}(W) \le \mathbb{E}(W - \frac{a+b}{2})^2 \le (\frac{b-a}{2})^2.$$

Now fix $\lambda \in \mathbb{R}$, let P_λ be the distribution on $[a,b]$ such that $\forall x \in [a,b]$,

$$\frac{\mathrm{d}P_\lambda}{\mathrm{d}\mathbb{P}^{(Z)}}(x) = (\mathbb{E}e^{\lambda Z})e^{\lambda x} = e^{-\psi_Z(\lambda)}e^{\lambda x}.$$

Then

$$\psi_Z'(\lambda) = \frac{\mathbb{E}Ze^{\lambda Z}}{\mathbb{E}e^{\lambda Z}} = e^{-\psi_Z(\lambda)}\mathbb{E}Ze^{\lambda Z}$$

and

$$\begin{aligned}\psi_Z''(\lambda) &= e^{-\psi_Z(\lambda)}\mathbb{E}Z^2e^{\lambda Z} - e^{-2\psi_Z(\lambda)}(\mathbb{E}Ze^{\lambda Z})^2 \\ &= \int_a^b e^{-\psi_Z(\lambda)}x^2e^{\lambda x}\mathrm{d}\mathbb{P}^{(Z)}(x) - \left\{\int_a^b e^{-\psi_Z(\lambda)}xe^{\lambda x}\mathrm{d}\mathbb{P}^{(Z)}(x)\right\}^2 \\ &\le (\frac{b-a}{2})^2.\end{aligned}$$

Since $\psi_Z'(0) = 0$, we have that $\forall\lambda \in \mathbb{R}$, $\exists\bar\lambda$ such that $|\bar\lambda| \le |\lambda|$ such that

$$\psi_Z(\lambda) = \psi_Z(0) + \psi_Z'(0)\lambda + \frac{1}{2}\psi_Z''(\bar\lambda)\lambda^2 \le \frac{\lambda^2}{2}\frac{(b-a)^2}{4}.$$

Claim (b) follows since $\psi_{S_n}(\lambda) = \sum_{i=1}^n \psi_{X_i}(\lambda)$, $\forall\lambda \in \mathbb{R}$. □

Remark 11.3.1. Let $X_1, \dots, X_n$ be r.v. such that $X_i \in [a_i, b_i]$ and that defining $Y_0 = 0$, $Y_1 = X_1$, $Y_2 = X_1 + X_2$, $\dots, Y_n = \sum_{i=1}^n X_i$, $\{Y_t\}_{t=0}^n$ is a martingale. We call $\{X_1, \dots, X_n\}$ a martingale difference sequence.

Define $\mathscr{F}_t = \sigma(X_1, \dots, X_t)$ for $t \in [n]$. Then, $\forall \lambda \in \mathbb{R}$,

$$\begin{aligned}
\psi_{Y_n}(\lambda) = \log \mathbb{E} e^{\lambda Y_n} &= \log \mathbb{E}\mathbb{E}_{\cdot|\mathscr{F}_{n-1}} e^{\lambda(X_n + Y_{n-1})} \\
&= \log \mathbb{E} e^{\lambda Y_{n-1}} \mathbb{E}_{\cdot|\mathscr{F}_{n-1}} e^{\lambda X_n} \\
&= \psi_{Y_{n-1}}(\lambda) + \log \mathbb{E}_{\cdot|\mathscr{F}_{n-1}} e^{\lambda X_n} \\
&\le \psi_{Y_{n-1}}(\lambda) + \frac{(b_n - a_n)^2}{4} \frac{\lambda^2}{2} \\
&\le \frac{\lambda^2}{2} \sum_{i=1}^n \frac{(b_i - a_i)^2}{4} \\
\implies Y_n = S_n &\in sG\left(\frac{1}{4} \sum_{i=1}^n (b_i - a_i)^2\right).
\end{aligned}$$

This gives the Azuma-Hoeffding inequality.

Theorem 11.3.2 (Bennett). Let $X_1, \dots, X_n$ be independent random variables such that $X_i \le b < \infty$ for some $b > 0$. Let $S_n := \sum_{i=1}^n X_i - \mathbb{E}X_i$ and $\nu := \sum_{i=1}^n \mathbb{E}X_i^2 < \infty$. We have that $\forall \lambda \ge 0$, $\psi_{S_n}(\lambda) \le \frac{\nu}{b^2}(e^{b\lambda} - b\lambda - 1)$ and for $t \ge 0$,

$$\mathbb{P}(S_n > t) \le \exp\left(-\frac{\nu}{b^2} h(\frac{bt}{\nu})\right)$$

where $h(z) = (1+z)\log(1+z) - z$.

Remark 11.3.2. Recall from Example 11.2.1 (b) that if $Z \sim \text{Poisson}(\nu)$, then $\psi_{Z - \mathbb{E}Z}(\lambda) \le \nu\phi(\lambda)$. Hence, in Bennett's inequality, if ν does not increase with n, then the right-tail of S_n behaves like Poisson.

If $X_i \sim \text{Ber}(1/n)$, then $\nu = 1$ and $b = 1$. Since $\text{TV}(S_n, \text{Poisson}(\nu)) \to 0$ as $n \to \infty$, we see that Bennett's inequality cannot be improved upon in general.

Proof.

We first assume $b = 1$. In general, we may define $\tilde{X}_i = X_i/b$ and note that $\sum_{i=1}^n \mathbb{E}\tilde{X}_i^2 \le \nu/b^2$ and that $\psi_{S_n}(\lambda) = \psi_{\sum_{i=1}^n \tilde{X}_i - \mathbb{E}\tilde{X}_i}(b\lambda)$.

First, note that $u \mapsto \frac{e^u - u - 1}{u^2}$ is non-decreasing $\forall u \in \mathbb{R}$. Since $X_i \le 1$, we have, $\forall \lambda \ge 0$,

$$\frac{e^{\lambda X_i} - \lambda X_i - 1}{X_i^2} \le e^\lambda - \lambda - 1,$$

which implies $\mathbb{E}e^{\lambda X_i} - \lambda \mathbb{E}X_i - 1 \le (\mathbb{E}X_i^2)(e^\lambda - \lambda - 1)$. Hence, we have

$$\log \mathbb{E}e^{\lambda X_i} \le \log\{(\mathbb{E}X_i^2)(e^\lambda - \lambda - 1) + \lambda\mathbb{E}X_i + 1\} \le \mathbb{E}X_i^2(e^\lambda - \lambda - 1) + \lambda\mathbb{E}X_i.$$

where the last inequality holds by using $\log(1+x) \le x$, $\forall x \ge 1$. Thus,

$$\psi_{S_n}(\lambda) \le \sum_{i=1}^n \log \mathbb{E}e^{\lambda X_i} - \lambda\mathbb{E}X_i \le \nu e^\lambda - \lambda - 1.$$

Defining $\widetilde{\psi}(\lambda) := \nu(e^\lambda - \lambda - 1)$, we have that $\forall t \ge 0$,

$$\widetilde{\psi}^\star(t) = \sup_{\lambda \ge 0} \lambda t - \widetilde{\psi}(\lambda) = \nu h(\frac{t}{\nu})$$

by Example 11.2.1 (b). $\square$

Remark 11.3.3. Using the fact that $(1+z)\log(1+z) - z \geq \frac{z^2}{2(1+z/3)}$, $\forall z \geq 0$, we have

$$\mathbb{P}(S_n > t) \leq \exp\left(-\frac{t^2}{2(\nu + bt/3)}\right) \implies S_n \in s\Gamma_+(C\nu, C'b).$$

This gives us Bernstein inequality for bounded variables. This holds more generally.

Theorem 11.3.3 (Bernstein). Let $X_1, \ldots, X_n$ be independent random variable such that, for some $\nu, c > 0$, $\sum_{i=1}^n \mathbb{E}X_i^2 \leq \nu$ and $\forall$ integer $q \geq 3$, $\sum_{i=1}^n \mathbb{E}(X_i)_+^q \leq \frac{q!}{2}\nu c^{q-2}$. Then, writing $S_n = \sum_{i=1}^n (X_i - \mathbb{E}X_i)$, we have

$$\psi_{S_n}(\lambda) = \log \mathbb{E}e^{\lambda S_n} \leq \frac{\nu\lambda^2}{2(1-c\lambda)}, \quad \forall \lambda \in [0, \frac{1}{c}),$$

which implies $S_n \in s\Gamma_+(\nu, c)$ and that for all $t \in \mathbb{R}$, $\mathbb{P}(S_n \geq t) \leq \exp\left(-\frac{t^2}{2(\nu+ct)}\right)$.

Proof.

Note that $u \mapsto \frac{e^u - u - 1}{u^2}$ is non-decreasing on $\mathbb{R}$ and $\lim_{u\to 0} \frac{e^u-u-1}{u^2} = \frac{1}{2}$. Therefore, we have that

$$e^u - u - 1 \leq \frac{u^2}{2} \quad \forall u \leq 0. \qquad (\star)$$

Now, using the fact that $\log(u) \leq u - 1$ for all $u \geq 0$, we have, for $\lambda > 0$,

$$\begin{aligned}
\psi_{S_n}(\lambda) &= \sum_{i=1}^n \log \mathbb{E}e^{\lambda X_i} - \mathbb{E}\lambda X_i \\
&\leq \sum_{i=1}^n \mathbb{E}\big\{e^{\lambda X_i} - \lambda X_i - 1\big\} \\
&\leq \sum_{i=1}^n \mathbb{E}\left\{\frac{\lambda^2 X_i^2}{2} + \sum_{q=3}^\infty \frac{\lambda^q (X_i)_+^q}{q!}\right\} \\
&\leq \frac{\lambda^2\nu}{2} + \sum_{q=3}^\infty \frac{\lambda^q q! \nu c^{q-2}}{2q!} \\
&= \frac{\lambda^2\nu}{2}\left(1 + \sum_{q=3}^\infty (\lambda c)^{q-2}\right) = \frac{\lambda^2\nu}{2(1-c\lambda)}.
\end{aligned}$$

□

11.4 Connections to Large Deviation Theory

Remark 11.4.1. Suppose $X_1, X_2, \ldots, X_n$ are independent random variables such that $\mathbb{E}X_i = 0$ and $X_i \in sE(\sigma^2)$ for $\sigma > 0$; again write $S_n = \sum_{i=1}^n X_i$. Then, using Bernstein's inequality with $\nu \leftarrow n\sigma^2$ and $c \leftarrow \sigma$, we have that, for any $t \in \mathbb{R}$,

$$\mathbb{P}(S_n \geq nt) \leq \exp\left(-\left(\frac{nt^2}{\sigma^2} \wedge \frac{nt}{\sigma}\right)\right).$$

What can we say about heavier tailed random variables? For example, suppose $Z_1, \ldots Z_n \sim N(0, \sigma^2)$. Then, we can analyze the moments to show that Z_i^3 is not sub-exponential. What can we say about $\sum_{i=1}^n Z_i^3$?

Definition 11.4.1. For $\alpha > 0$, define $\psi_\alpha(z) = e^{z^\alpha} - 1$ for $z \geq 0$. We say that a centered random variable Z is sub-Weibull with exponent α and variance factor $\nu > 0$ if

$$\|Z\|_{\psi_\alpha} := \inf\left\{t \geq 0 \,:\, \mathbb{E}\psi_\alpha\left(\frac{|Z|}{t}\right) \leq 1\right\} \leq 2\sqrt{\nu}.$$

We write $Z \in sW_\alpha(\nu)$.

Remark 11.4.2. It is important to note that $\|\cdot\|_{\psi_\alpha}$ does NOT satisfy triangle inequality when $\alpha < 1$ because $\psi_\alpha(\cdot)$ is not convex when $\alpha < 1$. However, $\psi_\alpha(\cdot)$ is still non-decreasing and hence, $\|aZ\|_{\psi_\alpha} = |a|\|Z\|_{\psi_\alpha}$. We also have that if $\|Z\|_{\psi_\alpha} < \infty$, then for all $t \geq 0$,

$$\mathbb{P}(Z \geq t) = \mathbb{P}\bigg(e^{\left(\frac{|Z|}{\|Z\|_{\psi_\alpha}}\right)^\alpha} \geq e^{\left(\frac{t}{\|Z\|_{\psi_\alpha}}\right)^\alpha}\bigg) \tag{11.7}$$

$$\leq 2e^{-\left(\frac{t}{\|Z\|_{\psi_\alpha}}\right)^\alpha}. \tag{11.8}$$

It is also clear from the definition that if $X \in sE(\nu)$, then $X^k \in sW_{1/k}(\nu^k)$. If $X \in sG(\nu)$, then $X^k \in sW_{2/k}(\nu^k)$.

Theorem 11.4.1. (Consequence of Theorem 3.1 in Kuchibotla and Chakrabortty)
Let $\alpha \in (0,1)$ and let $X_1, \ldots, X_n$ be independent random variables such that $X_i \in sW_\alpha(\sigma^2)$. Write $S_n = \sum_{i=1}^n X_i$, we then have that, for all $t \geq 0$,

$$\mathbb{P}(|S_n| \geq t) \leq 2\exp\bigg(-C_\alpha\bigg\{\frac{t^2}{n\sigma^2} \wedge \big(\frac{t}{\sigma}\big)^\alpha\bigg\}\bigg),$$

where $C_\alpha > 0$ is a constant that depends only on α.

Equivalently, we have that, for any $w \in \mathbb{R}$,

$$\mathbb{P}\big(|S_n| \geq C'_\alpha(\sqrt{nw} + w^{1/\alpha})\big) \leq e^{-w},$$

where $C'_\alpha > 0$ is some other constant that depends only on α.

Remark 11.4.3. We can intuitively understand why there is a $(nt/\sigma)^\alpha$ term in the tail probability using large deviation theory (LDT). In LDT, we say that a random variable X has a *subexponential distribution* (NOT the same as $sE(\nu)$) if, for all $n \geq 2$, for iid $X_1, \ldots, X_n$ with $S_n = \sum_{i=1}^n X_i$, we have

$$\lim_{x\to\infty} \frac{\mathbb{P}(S_n \geq x)}{n\mathbb{P}(X_1 \geq x)} = 1. \tag{11.9}$$

Write $M_n = \max\{X_1, \ldots, X_n\}$. Since $\mathbb{P}(M_n \geq x) = 1 - (1 - \mathbb{P}(X_1 \geq x))^n$ and $\lim_{x\to\infty} \mathbb{P}(X_1 \geq x) = 0$, (11.9) implies that $\lim_{x\to\infty} \frac{\mathbb{P}(S_n \geq x)}{\mathbb{P}(M_n \geq x)} = 1$.

In LDT, subexponential distribution is a heavy-tail condition. Intuitively, it means that if the sum is large, then the main contribution to the sum comes from the extreme values. Note that if the X_i's are centered with unit variance, then, for any fixed $x \in \mathbb{R}$, we also have by CLT that for any fixed $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \frac{\mathbb{P}(S_n \geq \sqrt{n}t)}{\bar{\Phi}(t)} = 1$$

where Φ is the CDF of a standard normal and $\bar{\Phi} = 1 - \Phi$. There is no contradiction because in (11.9), we fix n and let x go to infinity.

Let $t > 0$ and $0 < \alpha < 1$ be fixed. If $\mathbb{P}(X_1 \geq t) = C\exp(-(t/\sigma)^\alpha)$, we expect $\mathbb{P}(|S_n| \geq t)$ to have Weibull tail $\exp(-C_\alpha(t/\sigma)^\alpha)$ if t is large. On the other hand, if $t = s/\sqrt{n}$ for some fixed $s > 0$, then we expect $\mathbb{P}(|S_n| \geq s\sqrt{n})$ to have Gaussian tails $\exp(-C_\alpha(s/\sigma)^2)$ because of central limit theorem.

Remark 11.4.4. Given a random variable X, recall that $\psi_X(\lambda) = \log \mathbb{E}e^{\lambda X}$ for $\lambda \in \mathbb{R}$. For $t \in \mathbb{R}$, we define $\psi^\star_{X_1}(t) = \sup_{\lambda\in\mathbb{R}} \lambda t - \psi_{X_1}(\lambda)$. Note that we have the following: suppose $\mathbb{E}X$ exists and is finite,

1. If $t \geq \mathbb{E}X$, then $\psi^\star_X(t) = \sup_{\lambda\geq 0} \lambda t - \psi_X(\lambda)$.

2. If $t \leq \mathbb{E}X$, then $\psi^\star_X(t) = \sup_{\lambda\leq 0} \lambda t - \psi_X(\lambda)$.

To see this, note that $\psi_X(\lambda) \geq \lambda \mathbb{E}X$ by Jensen's inequality. Hence, if $\lambda \geq 0$ and $t \leq EX$, then $t\lambda - \psi_X(\lambda) \leq \lambda(t - \mathbb{E}X) \leq 0$. Since $\lambda = 0$ yields $t\lambda - \psi_X(\lambda) = 0$, the conclusion follows.

Theorem 11.4.2. (Cramér's Theorem)
Let $X_1, X_2, \ldots$ be iid sequence such that $\psi_{X_1}(\lambda) < \infty$ for all $\lambda \in \mathbb{R}$ and define $S_n = \sum_{i=1}^n X_i$. Suppose also that $\mathbb{P}(X_1 \geq K) > 0$ for all $K > 0$. Then, we have that, for any $t \geq \mathbb{E}X_1$,

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(S_n \geq nt) = -\psi^\star_{X_1}(t).$$

For any $t \leq \mathbb{E}X_1$,

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(S_n \leq nt) = -\psi^\star_{X_1}(t).$$

Proof.
Note that $\psi_{X - \mathbb{E}X}(\lambda) = \psi_X(\lambda) - \lambda \mathbb{E}X$ for all $\lambda \in \mathbb{R}$ and thus, for all $t \in \mathbb{R}$, $\psi^\star_{X - \mathbb{E}X}(t) = \sup_{\lambda \in \mathbb{R}} t\lambda - \psi_X(\lambda) + \lambda \mathbb{E}X = \sup_{\lambda \in \mathbb{R}} \lambda(t - \mathbb{E}X) - \psi_X(\lambda) = \psi^\star_X(t - \mathbb{E}X)$. Therefore, we may assume without loss of generality that $\mathbb{E}X_1 = 0$.

We fix $t \geq 0$ and prove the first statement; the second statement follows by analyzing $-X_1$. Since $\psi_{S_n}(\lambda) = n\psi_{X_1}(\lambda)$ for all $\lambda \in \mathbb{R}$ and $\psi^\star_{S_n}(t) = n\psi^\star_{X_1}(t/n)$, we have that, for all $n \geq 1$,

$$\mathbb{P}(S_n \geq nt) \leq e^{-n\psi^\star_{X_1}(t)}.$$

Thus, we have that $\limsup_{n \to \infty} \frac{1}{n} \mathbb{P}(S_n \geq nt) \leq -\psi^\star_{X_1}(t)$.

Therefore, we need only show that $\liminf_{n \to \infty} \frac{1}{n} \mathbb{P}(S_n \geq nt) \geq -\psi^\star_{X_1}(t)$. Before proving this, we first prove a few properties of $\psi^\star_{X_1}(\cdot)$:

Claim 1 $\psi^\star_{X_1}(\cdot)$ is convex on $[0, \infty)$. As a consequence, since $\psi^\star_{X_1}(\cdot)$ is non-negative and 0 at 0, it must be non-decreasing on $[0, \infty)$. Morever, since $\psi'_{X_1}(\cdot)$ is 0 at 0, $\psi^\star_{X_1}(t) > 0$ for any $t > 0$ and thus, $\psi^\star_{X_1}(\cdot)$ must be strictly increasing on $[0, \infty)$.

Claim 2 For any $t \geq 0$, the supremum $\sup_{\lambda \geq 0} t\lambda - \psi_X(\lambda)$ is attained at some λ_t satisfying $t = \frac{\mathbb{E}Xe^{\lambda_t X}}{\mathbb{E}e^{\lambda_t X}}$.

To see claim 1, we observe that $\psi^\star_{X_1}(t)$ is the pointwise supremum of a collection of linear functions and hence convex. To see claim 2, we will show that $\lim_{\lambda \to \infty} \frac{\psi_{X_1}(\lambda)}{\lambda} = \infty$. Let $K > \mathbb{E}X_1$ be arbitrary, then, for any $\lambda \geq 0$,

$$\frac{\psi_{X_1}(\lambda)}{\lambda} \geq \frac{1}{\lambda} \log \int_K^\infty e^{\lambda X} d\mathbb{P} \leq \frac{1}{\lambda} \log\bigl(e^{\lambda K} \mathbb{P}(X \geq K)\bigr) \geq K + \frac{1}{\lambda} \log \mathbb{P}(X \geq K).$$

Hence, $\lim_{\lambda \to \infty} \frac{\psi_{X_1}(\lambda)}{\lambda} \geq K$. Since K is arbitrary, claim 2 follows.

Now, let $\epsilon > 0$ be arbitrary and choose $s > t$ such that $\psi^\star_{X_1}(s) \leq \psi^\star_{X_1}(t) + \epsilon$; this is possible since $\psi^\star_{X_1}(\cdot)$ is strictly increasing. Since $\psi^\star_{X_1}(\cdot)$ is also convex, there exists $\delta > 0$ such that $s - \delta > t$.

Let $\lambda_s \geq 0$ satisfy $s = \frac{\mathbb{E}X_1 e^{\lambda_s X_1}}{\mathbb{E}e^{\lambda_s X_1}}$ and define a probability measure Q_s on $\mathbb{R}$ such that

$$\frac{dQ_s}{d\mathbb{P}^{(X_1)}}(x) = \frac{1}{\mathbb{E}e^{\lambda_s X_1}} e^{\lambda_s x}, \qquad x \in \mathbb{R}.$$

Observe that Q_s has mean $\frac{1}{\mathbb{E}e^{\lambda_s X}} \int_{-\infty}^\infty x e^{\lambda_s x} d\mathbb{P}^{(X_1)}(x) = s$.

Thus, we have that

$$\begin{aligned}
\mathbb{P}(S_n \geq nt) &\geq \mathbb{P}(S_n \in n(s-\delta, s+\delta)) \\
&= \int_{\{x:|\sum x_i - ns|<n\delta\}} d\mathbb{P}^{(X_1)}(x_1)...d\mathbb{P}^{(X_n)}(x_n) \\
&= \int_{\{x:|\sum x_i - ns|<n\delta\}} \left(\mathbb{E}e^{\lambda_s X_1}\right)^n e^{-\lambda_s \sum_{i=1}^n x_i} dQ_s(x_1)...dQ_s(x_n) \\
&\leq \left(\mathbb{E}e^{\lambda_s X_1}\right)^n e^{-\lambda_s(ns-n\delta)} \int_{\{x:|\sum x_i - ns|<n\delta\}} dQ_s(x_1)...dQ_s(x_n).
\end{aligned}$$

By law of large numbers, $\lim_{n\to\infty} \int_{\{x:|\sum x_i - ns|<n\delta\}} dQ_s(x_1)...dQ_s(x_n) = 1$. Thus, we have that

$$\frac{1}{n}\log \mathbb{P}(S_n \geq nt) \geq -\lambda_s(s-\delta) + \psi_{X_1}(\lambda_s) + o(1) = \psi^\star_{X_1}(s) + \lambda_s\delta \geq \psi^\star_{X_1}(t) + \epsilon + \lambda_s\delta + o(1).$$

Since ϵ, δ can be chosen arbitrarily close to 0, the conclusion follows.

□

11.5 Coordinate-wise Differences and Efron–Stein Inequality

Definition 11.5.1. Let $(\mathcal{X}, \mathscr{G})$ be a measurable space and let $f : \mathcal{X}^n \to \mathbb{R}$ be $\mathscr{G}^{\otimes n}/\mathscr{B}(\mathbb{R})$-measurable.

(a) We say that f satisfies the bounded difference condition (BDC_a) with coefficients $c_1, c_2, \ldots, c_n > 0$ if $\forall i \in [n]$

$$\sup_{x\in\mathcal{X}^n, x_i'\in\mathcal{X}} |f(x_i, x_{-i}) - f(x_i', x_{-i})| \leq c_i,$$

(b) We say that f satisfies the bounded difference condition (BDC_b) with coefficients $V > 0$ if

$$\sup_{x\in\mathcal{X}^n} \sum_{i=1}^n \sup_{x_i'\in\mathcal{X}} (f(x_i, x_{-i}) - f(x_i', x_{-i}))_+^2 \leq V.$$

Note that $\mathrm{BDC}_a(c_1, \ldots, c_n) \implies \mathrm{BDC}_b(c_1^2 + c_2^2 \cdots + c_n^2)$. To see this, define $c_i(x) = \sup_{x_i'\in\mathscr{X}}(f(x_i, x_{-i}) - f(x_i', x_{-i}))_+$ and we have that f satisfies BDC_b if $\sup_{x\in\mathscr{X}^n} \sum_{i=1}^n c_i(x)^2 \leq V$.

Theorem 11.5.1 (McDiarmid/BC Ineq)**.** Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Let $X_1, \ldots, X_n : \Omega \to \mathcal{X}$ be independent random objects and let $f : \mathcal{X}^n \to \mathbb{R}$ satisfy $\mathrm{BDC}_a(c_1, \ldots, c_n)$. Then, writing $Z := f(X_1, \ldots, X_n)$, we have that $Z - \mathbb{E}Z \in sG(\frac{1}{4}\sum_{i=1}^n c_i^2)$. In particular, $\forall t > 0$,

$$\mathbb{P}(Z - \mathbb{E}Z > t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right).$$

Lemma 11.5.1. Let $X = (X_1, \ldots, X_n) : \Omega \to \mathcal{X}^n$ be independent and let $f : \mathcal{X}^n \to \mathbb{R}$ be measurable. Let $T \subseteq [n]$ and write $X_T = \{X_t\}_{t\in T}$ and $X_{-T} = \{X_t\}_{t\notin T}$.

$$\mathbb{E}[f(X)|X_T] = \int_{\mathcal{X}^{n-|T|}} f(X_T, x_{-T}) \mathrm{d}\mathbb{P}^{(X_{-T})}(x_{-T}). \tag{11.10}$$

Proof.

By definition, $\mathbb{E}[f(X)|X_T]$ is the $\mathbb{P}$-a.e. unique measurable function of X_T such that

$$\int_\Omega (f(X) - \mathbb{E}[f(X)|X_T])g(X_T)\mathrm{d}\mathbb{P} = 0 \quad \forall \text{meas. } g : \mathcal{X}^{|T|} \to \mathbb{R}.$$

Since

$$\begin{aligned}
&\int_{\Omega}(f(X)-\mathbb{E}[f(X)|X_T])g(X_T)\mathrm{d}\mathbb{P}\\
&=\int_{\mathcal{X}^n}(f(x)-\mathbb{E}[f(X)|X_T=x_t])g(x_T)\mathrm{d}\mathbb{P}^{(X_T)}(x_T)\mathrm{d}\mathbb{P}^{(X_{-T})}(x_{-T})\\
&=\int_{\mathcal{X}^{|T|}}\int_{\mathcal{X}^{n-|T|}} f(x)-\mathbb{E}[f(X)|X_T=x_T]\mathrm{d}\mathbb{P}^{(X_{-T})}(x_{-T})g(x_T)\mathrm{d}\mathbb{P}^{(X_T)}(x_T),
\end{aligned}$$

the claim follows. □

Proof of Theorem 11.5.1.
For $t\in[n]$, define $\mathscr{F}_t=\sigma(X_1,\dots,X_t)$ and define $\mathscr{F}_0=\{\emptyset,\Omega\}$. Let

$$\Delta_t:=\mathbb{E}[Z|\mathscr{F}_t]-\mathbb{E}[Z|\mathscr{F}_{t-1}]=\mathbb{E}[Z-\mathbb{E}_{\bullet|\mathscr{F}_{t-1}}Z|\mathscr{F}_t],$$

(recall that $\mathbb{E}_{\bullet|\mathscr{F}_t}\mathbb{E}_{\bullet|\mathscr{F}_{t-1}}Z=\mathbb{E}_{\bullet|\mathscr{F}_{t-1}}Z$) so that $Z-\mathbb{E}Z=\sum_{t=1}^n\Delta_t$.

Note that $\forall t\in[n-1]$, $\mathbb{E}_{\bullet|\mathscr{F}_t}\Delta_{t+1}=0$. So $\{\sum_{i=1}^t\Delta_i\}_{t\in[n]}$ is a martingale. Moreover, we have by Lemma 11.5.1 that

$$\begin{aligned}
\mathbb{E}[Z|\mathscr{F}_t]&=\mathbb{E}[f(X_1,\dots,X_n)|X_1,\dots,X_t]\\
&=\int_{\mathcal{X}^{n-t}}f(X_1,\dots,X_t,x_{t+1},\dots,x_n)\mathrm{d}\mathbb{P}^{(X_{t+1})}(x_{t+1})\dots\mathrm{d}\mathbb{P}^{(X_n)}(x_n).
\end{aligned}$$

Hence, $\forall t\in\{2,\dots,n\}$,

$$\begin{aligned}
\Delta_t&=\int_{\mathcal{X}^{n-1}}f(X_1,\dots,X_t,x_{t+1},\dots,x_n)\mathrm{d}\mathbb{P}^{(X_{t+1})}(x_{t+1})\dots\mathrm{d}\mathbb{P}^{(X_n)}(x_n)\\
&\qquad-\int_{\mathcal{X}^{n-t+1}}f(X_1,\dots,X_{t-1},x_t,x_{t+1},\dots,x_n)\mathrm{d}\mathbb{P}^{(X_t)}(x_t)\dots\mathrm{d}\mathbb{P}^{(X_n)}(x_n)\\
&\le\int_{\mathcal{X}^{n-t}}\underbrace{\left\{f(X_1,\dots,X_{t-1},X_t,x_{t+1},\dots,x_n)-\int_{\mathcal{X}}f(X_1,\dots,X_{t-1},x_t,x_{t+1},\dots,x_n)\mathrm{d}\mathbb{P}^{(X_t)}(x_t)\right\}}_{(\star)}\\
&\qquad\qquad\mathrm{d}\mathbb{P}^{(X_{t+1})}(x_{t+1})\dots\mathrm{d}\mathbb{P}^{(X_n)}(x_n).
\end{aligned}$$

We have that $\operatorname{ess\,sup}(\star)-\operatorname{ess\,inf}(\star)\le c_t$ by BDC and thus Δ_t takes value in an interval of width at most c_t. The theorem follows from Azuma-Hoeffding (see Remark 11.3.1). □

Example 11.5.1. (a) Let $\mathcal{H}\subseteq[-1,1]^{\mathcal{X}}$ be a countable set of $\mathscr{G}/\mathscr{B}(\mathbb{R})$-measurable functions. Define $\mathcal{X}^n\to\mathbb{R}$ by $f(x_1,\dots,x_n)=\sup_{h\in\mathcal{H}}\frac{1}{n}\sum_{i=1}^n h(x_i)$. Then, for any $i\in[n]$, $x\in\mathcal{X}^n$, $x_i'\in\mathcal{X}$, we have that

$$\begin{aligned}
|f(x_i,x_{-i})-f(x_i',x_{-i})|&=\left|\sup_{h\in\mathcal{H}}\frac{1}{n}\sum_{j\in[n],j\neq i}h(x_j)+\frac{1}{n}h(x_i)-\left\{\sup_{h\in\mathcal{H}}\frac{1}{n}\sum_{j\in[n]}h(x_j)+\frac{1}{n}h(x_i')\right\}\right|\\
&\le\frac{1}{n}\sup_{h\in\mathcal{H}}|h(x_i)-h(x_i')|\le\frac{2}{n}
\end{aligned}$$

Thus, f satisfy BDC_a with $c_1=c_2=\dots=c_n=\frac{2}{n}$.

(b) (Longest common subsequence) Let $\mathcal{X}=\{0,1\}$ and let $X_1,\dots,X_n$ and $Y_1,\dots,Y_n$ be iid Bernoulli random variable. Define $f:\mathcal{X}^{2n}\to\mathbb{N}$ by

$$\begin{aligned}
f(x_1,\dots,x_n,y_1,\dots,y_n)=\max\{k\in[n]:x_{i_1}=y_{j_1},\dots,x_{i_k}=y_{j_k}\text{ for }i_1<i_2<\dots<i_k\subseteq[n]\\
\text{and }j_1<j_2<\dots<j_k\subseteq[n]\}.
\end{aligned}$$

For example, if $X = (0,0,1,0,1)$ and $Y = (1,0,0,1,1)$, then the longest common subsequence is $(0,0,1,1)$.

It is clear that f satisfy BCD_a with $c_1 = c_2 = \cdots = c_n = 1$. Thus, $f(X_1,\dots,X_n,Y_1,\dots,Y_n) \in sG(\frac{n}{4})$.

(c) Let $A \in [-1,1]^{n\times n}$ be symmetric and define $f(A) = \lambda_{\max}(A) = \sup\{v^\top Av : \|v\|_2 = 1\}$. Fix $k,\ell \in [n]$, $a \in [-1,1]$, and define $\widetilde{A}^{(k,l)} \in \mathbb{R}^{n\times n}$ such that $\forall i,j \in [n]$

$$\widetilde{A}^{(k,\ell)}_{ij} = \begin{cases} A_{ij} & \text{if } (i,j) \neq (k,\ell) \text{ or } (\ell,k) \\ a & \text{else.} \end{cases}$$

Let $w \in \mathbb{R}^n$ be the maximum eigenvector of A, then,

$$\begin{aligned}\max(f(A) - f(\widetilde{A}^{(k,\ell)}),0) &\le \max(\sum_{i,j} A_{ij}w_iw_j - \sum_{i,j}\widetilde{A}^{(k,\ell)}_{ij}w_iw_j, 0)\\ &\le 4|w_kw_\ell|.\end{aligned}$$

Thus, $\sum_{k,\ell}(f(A) - f(\widetilde{A}^{(k,\ell)}))_+^2 \le 16\sum_{k,\ell} w_k^2w_\ell^2 \le 16(\sum_{i=1}^n w_i^2)^2 = 16$. So f satisfy BDC_b with $V \le 16$.

Theorem 11.5.2 (Efron-Stein). Let $X = (X_1,\dots,X_n) : \Omega \to \mathcal{X}^n$ be independent random object and let $f : \mathcal{X}^n \to \mathbb{R}$ be measurable. Define

$$V_{\mathrm{ES}} := \sum_{i=1}^n \mathbb{E}\mathrm{Var}_{\bullet|X_{-i}} f(X_i,X_{-i}) = \sum_{i=1}^n \mathbb{E}\left(f(X_i,X_{-i}) - \mathbb{E}_{\bullet|X_{-i}} f(X_i,X_{-i})\right)^2.$$

Then

$$\mathrm{Var} f(X) \le V_{\mathrm{ES}}$$

and moreover, with $X_1',\dots,X_n'$ as independent copies of $X_1,\dots,X_n$,

$$V_{\mathrm{ES}} = \frac{1}{2}\sum_{i=1}^n \mathbb{E}\left(f(X_i,X_{-i}) - f(X_i',X_{-i})\right)^2 = \sum_{i=1}^n \mathbb{E}\left(f(X_i,X_{-i}) - f(X_i',X_{-i})\right)_+^2.$$

Additionally, $V_{\mathrm{ES}} \le \sum_{i=1}^n \mathbb{E}\left(f(X_i,X_{-i}) - g(X_{-i})\right)^2$, $\forall g : \mathcal{X}^{n-1} \to \mathbb{R}$ square integrable.

Note that if f satisfy $\mathrm{BDC}_b(V)$, then we immediately obtain $\mathrm{Var} f(X) \le V$.

Proof.

Write $\mathscr{F}_t := \sigma(X_1,\dots,X_t)$ for $t \in [n]$. By Lemma 11.5.1, we have that $\forall t \in [n]$,

$$\mathbb{E}_{\bullet|\mathscr{F}_t}\mathbb{E}_{\bullet|X_{-t}} f(X) = \mathbb{E}_{\bullet|\mathscr{F}_{t-1}} f(X).$$

Thus,

$$\Delta_t := \mathbb{E}[f(X) - \mathbb{E}_{\bullet|\mathscr{F}_{t-1}} f(X)|\mathscr{F}_t] = \mathbb{E}[f(X) - \mathbb{E}_{\bullet|X_{-t}} f(X)|\mathscr{F}_t].$$

By Jensen's inequality, $\Delta_t^2 \le \mathbb{E}_{\bullet|\mathscr{F}_t}(f(X) - \mathbb{E}_{\bullet|X_{-t}} f(X))^2$. Since $f(X) - \mathbb{E}f(X) = \sum_{t=1}^n \Delta_t$ and since, for $s < t \in [n]$, $\mathbb{E}[\Delta_s\Delta_t] = \mathbb{E}\Delta_s\mathbb{E}_{\bullet|\mathscr{F}_s}\Delta_t = 0$, we have

$$\mathrm{Var} f(X) = \sum_{t=1}^n \mathbb{E}\Delta_t^2 \le \sum_{t=1}^n \mathbb{E}\left(f(X) - \mathbb{E}_{\bullet|X_{-t}} f(X)\right)^2.$$

This establishes the first claim.

For the second claim, observe that $\forall i \in [n]$,

$$\frac{1}{2}\mathbb{E}\mathbb{E}_{\cdot|X_{-i}}\left(f(X_i,X_{-i})-f(X_i,X'_{-i})\right)^2 = \mathbb{E}\left\{\mathbb{E}_{\cdot|X_{-i}}f(X_i,X_{-i})^2-\mathbb{E}_{\cdot|X_{-i}}f(X_i,X_{-i})f(X'_i,X_{-i})\right\}$$
$$= \mathbb{E}\mathrm{Var}_{\cdot|X_{-i}}f(X_i,X_{-i}).$$

Now, writing $Y_i := f(X_i,X_{-i})-f(X'_i,X_{-i})$, we have that $Y_i \overset{d}{=} -Y_i$. So

$$\frac{1}{2}\mathbb{E}Y_i^2 = \frac{1}{2}\mathbb{E}Y_{i+}^2+\frac{1}{2}\mathbb{E}Y_{i-}^2 = \mathbb{E}Y_{i+}^2.$$

□

Example 11.5.2. We say that $f:[0,1]^n \to \mathbb{R}$ is separately convex if $\forall i \in [n]$, $\forall x_{-i} \in [0,1]^{n-1}$, $x_i \mapsto f(x_i,x_{-i})$ is convex. Let $X=(X_1,\dots,X_n)$ be independent random variables taking value on $[0,1]$ and let $X'=(X'_1,\dots,X'_n)$ be an independent copy. Let $\frac{\partial}{\partial x_i}f(X)$ be a (random) subgradient of $x_i \mapsto f(x_i,x_{-i})$ at X. Then, since $f(X_i,X_{-i})-f(X'_i,X_{-i}) \le -\frac{\partial}{\partial x_i}f(X_i,X_{-i})(X'_i-X_i)$, we have that

$$\sum_{i=1}^n \left(f(X_i,X_{-i})-f(X'_i,X_{-i})\right)_+^2 \le \sum_{i=1}^n\left(\frac{\partial}{\partial x_i}f(X_i,X_{-i})\right)^2(X'_i-X_i)^2$$
$$\le \sum_{i=1}^n\left(\frac{\partial}{\partial x_i}f(X_i,X_{-i})^2\right)^2 = \|\nabla f(X)\|^2$$

where $\nabla f(X)=(\frac{\partial}{\partial x_1}f(X),\dots,\frac{\partial}{\partial x_n}f(X))$. So $\mathrm{Var}f(X)\le \mathbb{E}\,\|\nabla f(X)\|_2^2$. (Convex Poincare Inequality).

Theorem 11.5.3 (Gaussian Poincare Inequality). Let $X=(X_1,\dots,X_n)\sim N(0,\Sigma)$ for some $\Sigma \succeq 0$ and let $f:\mathbb{R}^n\to\mathbb{R}$ be continuously differentiable ($\nabla f:\mathbb{R}^n\to\mathbb{R}^n$ exists and is continuous everywhere). Then,

$$\mathrm{Var}f(X) \le \mathbb{E}\nabla f(X)^\top\Sigma\nabla f(X). \tag{11.11}$$

Note: if $\Sigma=\mathrm{Id}$ and f is 1-Lipschitz, then $\mathrm{Var}f(X)\le 1$. Equality is attained by $f(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$.

Proof.

It suffices to assume that $\Sigma=\mathrm{Id}$; to see this, note that $\widetilde{X}=\Sigma^{-\frac{1}{2}}X$ is $N(0,I_n)$ and that, writing $g(x):=f(\Sigma^{\frac{1}{2}}x)$, $\forall x\in\mathbb{R}^n$, we have that $g(\widetilde{X})=f(X)$. Since $\nabla g(x)=\Sigma^{1/2}\nabla f(\Sigma^{\frac{1}{2}}X)$ by chain rule and thus, $\nabla g(\widetilde{X})=\Sigma^{\frac{1}{2}}\nabla f(X)$.

Now it also suffices to treat only $n=1$ case. To see this, suppose (11.11) holds for univariate functions, then

$$\mathrm{Var}f(X) \le \sum_{i=1}^n \mathbb{E}\mathrm{Var}_{\cdot|X_{-i}}f(X_i,X_{-i}) \le \sum_{i=1}^n \mathbb{E}\mathbb{E}_{X_{-i}}\left(\frac{\partial}{\partial x_i}f(X_i,X_{-i})\right)^2 = \mathbb{E}\,\|\nabla f(X)\|^2.$$

Thus, we need only consider $f:\mathbb{R}\to\mathbb{R}$ and show that $\mathrm{Var}f(X)\le\mathbb{E}f'(X)^2$ where $f':\mathbb{R}\to\mathbb{R}$ is continuous.

For $b>0$, define $f_b(x):=f(x)\mathbb{1}\{|x|\le b\}$. Recall also the definition of the bump distribution (Definition 5.3.1)

$$q(x)=Ce^{-\frac{1}{1-|x|}}\mathbb{1}\{|x|\le 1\}.$$

For $\sigma>0$, define $f_{b,\sigma}(x):=\mathbb{E}_{Z\sim q}f(x+\sigma Z)$. We then have that $f_{b,\sigma}$ is continuous, continuously differentiable, and supported on $[-b-\sigma,b+\sigma]$. Since $\lim_{\sigma\to0}f_{b,\sigma}=f_b$ pointwise and $\sup_{\sigma\le1}|f_{b,\sigma}(x)|\le M_b\mathbb{1}\{|x|\le b+1\}$ where $M_b:=\sup_{\sigma\in(0,1)}\sup_{x\in[-b-1,b+1]}|f_{b,\sigma}(x)|<\infty$, we have by the dominated convergence theorem that $\lim_{\sigma\to0}\mathbb{E}f_{b,\sigma}(X)^2=\mathbb{E}f_b(X)^2$ and $\lim_{\sigma\to0}\mathbb{E}f_{b,\sigma}(X)=\mathbb{E}f_b(X)$, which implies that

$$\lim_{\sigma\to0}\mathrm{Var}f_{b,\sigma}(X)=\mathrm{Var}f_b(X).$$

By the same reasoning, it holds that

$$\lim_{\sigma\to 0} \mathbb{E} f'_{b,\sigma}(X)^2 = \mathbb{E} f'_b(X)^2.$$

By dominated convergence theorem again, we have that

$$\lim_{b\to\infty} \mathrm{Var} f_b(X) = \mathrm{Var} f(X),$$
$$\lim_{b\to\infty} \mathbb{E} f'_b(X)^2 = \mathbb{E} f'(X)^2,$$

Thus, it suffices to show that $\forall b \geq 1$ and $\sigma > 0$, $\mathrm{Var} f_{b,\sigma}(X) \leq \mathbb{E} f'_{b,\sigma}(X)^2$.

Define $\varepsilon_1, \ldots, \varepsilon_n$ as iid Rademacher random variables and define $S_n := \sum_{i=1}^n \varepsilon_i$. Since $\frac{1}{\sqrt{n}} S_n \xrightarrow{d} X$ and $f_{b,\sigma}, f'_{b,\sigma}$ are bounded, we have that $\lim_{n\to\infty} \mathrm{Var} f_{b,\sigma}(S_n/\sqrt{n}) = \mathrm{Var} f_{b,\sigma}(X)$, $\lim_{n\to\infty} \mathbb{E} f'_{b,\sigma}(\frac{S_n}{\sqrt{n}})^2 = \mathbb{E} f'_{b,\sigma}(X)^2$.

By Efron–Stein inequality again,

$$\begin{aligned}
\mathrm{Var} f_{b,\sigma}(\frac{S_n}{\sqrt{n}}) &\leq \sum_{i=1}^n \mathbb{E}\mathrm{Var}_{\cdot|\varepsilon_{-i}} f_{b,\sigma}(\frac{S_n}{\sqrt{n}}) \\
&= \sum_{i=1}^n \mathbb{E}\frac{1}{4}\left\{ f_{b,\sigma}\left(\frac{S_n + 1 - \varepsilon_i}{\sqrt{n}}\right) - f_{b,\sigma}\left(\frac{S_n - (1+\varepsilon_i)}{\sqrt{n}}\right)\right\}^2.
\end{aligned}$$

Now, writing $K_b := \sup_{x\in\mathbb{R}} |f''_{b,\sigma}(x)| < \infty$, we have

$$\begin{aligned}
&\left| f_{b,\sigma}\left(\frac{S_n + (1 - \varepsilon_i)}{\sqrt{n}}\right) - f_{b,\sigma}\left(\frac{S_n - (1+\varepsilon_i)}{\sqrt{n}}\right)\right| \leq \frac{2}{\sqrt{n}}\left| f'_{b,\sigma}(\frac{S_n}{\sqrt{n}})\right| + \frac{2K_b}{n} \\
\implies & \mathrm{Var} f_{b,\sigma}(\frac{S_n}{\sqrt{n}}) \leq \mathbb{E}\left| f'_{b,\sigma}(\frac{S_n}{\sqrt{n}})\right|^2 + \frac{K_b}{\sqrt{n}}\mathbb{E}\left| f'_{b,\sigma}(\frac{S_n}{\sqrt{n}})\right| + \frac{K_b^2}{n} \\
\implies & \lim_{n\to\infty} \mathrm{Var} f_{b,\sigma}(\frac{S_n}{\sqrt{n}}) - \mathbb{E} f'_{b,\sigma}(\frac{S_n}{\sqrt{n}})^2 \leq 0 \\
\implies & \mathrm{Var} f_{b,\sigma}(X) - \mathbb{E} f'_{b,\sigma}(X)^2 \leq 0.
\end{aligned}$$

□

Example 11.5.3. Let $A \in \mathbb{R}^{m\times n}$ and let $S_{\max}(A) = \sup\{\|Au\|_2 : \|u\|_2 = 1, u \in \mathbb{R}^n\}$ be the maximum singular value of A. Note that $S_{\max}(A)^2 = \lambda_{\max}(A^\top A)$. Since, for any fixed $u \in \mathbb{R}^n$, $A \mapsto \|Au\|_2$ is convex, and since pointwise sup of convex function is convex, we have that $S_{\max} : \mathbb{R}^{m\times n} \to [0,\infty)$ is convex. Moreover, since $S_{\max}(A)$ is the operator norm of A, we have

$$|S_{max}(A) - S_{\max}(B)| \leq \|A - B\|_F,$$

so $S_{\max}$ is 1-Lipschitz w.r.t. Frobenius norm on $\mathbb{R}^{m\times n}$.

As a direct consequence:

(a) If $\{A_{ij}\}_{i\in[m],j\in[n]}$ are independent and in $[0,1]$, then $\mathrm{Var}(S_{\max}(A)) \leq 1$. (Example 11.5.2)

(b) If $\{A_{ij}\}_{i\in[m],j\in[n]}$ are independent and $N(0,1)$, then $\mathrm{Var}(S_{\max}(A)) = \mathrm{Var}(\lambda_{\max}(A^\top A)) \leq 1$. (Theorem 11.5.3)

The same argument applies to all singular values.

Proposition 11.5.1. Let $X = (X_1, \ldots, X_n) : \Omega \to [0,1]^n$ be a random vector such that $\mathbb{P}^{(X)}$ has density $p : [0,1]^n \to [0,\infty)$. Let $p_{\max} := \operatorname{ess\,sup} p(x)$ and $p_{\min} := \operatorname{ess\,inf} p(x)$. If $p_{\max} < \infty$ and $p_{\min} > 0$, then $\forall f : \mathbb{R}^n \to \mathbb{R}$ continuous differentiable,

$$\mathrm{Var} f(X) \leq \frac{p_{\max}^2}{p_{\min}} \mathbb{E}\|\nabla f(X)\|^2.$$

Proof.

Let $\tilde{X}$ be a iid copy of X. Since, for any $f : \mathbb{R}^n \to \mathbb{R}$ that is continuous differentiable,

$$\begin{aligned}\mathrm{Var} f(X) &= \frac{1}{2}\mathbb{E}(f(X) - f(\widetilde{X}))^2 \\ &= \frac{1}{2}\int\int (f(x) - f(\tilde{x}))^2 p(x)p(\tilde{x})\mathrm{d}x \leq p_{\max}^2 \frac{1}{2}\int\int (f(x) - f(\tilde{x}))^2 \mathrm{d}x\mathrm{d}\tilde{x}\end{aligned}$$

and $\mathbb{E}\left\|\nabla f(X)\right\|^2 = \int \nabla f(x)^2 p(x)\mathrm{d}x \geq p_{\min}\int \nabla f(x)^2 \mathrm{d}x$, it suffices to prove that if $X \sim \mathrm{Unif}[0,1]^n$, then

$$\mathrm{Var} f(X) \leq \mathbb{E}\left\|\nabla f(X)\right\|^2.$$

Again, by Efron-Stein inequality, it suffices to show that if $X \sim \mathrm{Unif}[0,1]$ and $f : \mathbb{R} \to \mathbb{R}$ is continuous differentiable, then $\mathrm{Var} f(X) \leq \mathbb{E}f'(X)^2$.

To see this, note that

$$\begin{aligned}\mathrm{Var} f(X) &= \frac{1}{2}\mathbb{E}(f(X) - f(\widetilde{X}))^2 \\ &= \frac{1}{2}\int_0^1\int_0^1 (f(s) - f(t))^2 \mathrm{d}s\mathrm{d}t \\ &= \int_0^1\int_0^1 \left\{\int_s^t f'(u)\mathrm{d}u\right\}^2 \mathbb{1}_{\{s\leq t\}}\mathrm{d}s\mathrm{d}t \\ &\leq \int_0^1\int_0^1 (t-s)\int_s^t f'(u)^2 \mathbb{1}_{\{s\leq t\}}\mathrm{d}u\mathrm{d}s\mathrm{d}t \\ &= \int_0^1 f'(u)^2 \underbrace{\left[\int_0^1\int_0^1 (t-s)\mathbb{1}_{\{s\leq u\leq t\}}\mathrm{d}s\mathrm{d}t\right]}_{\int_u^1 t\mathrm{d}t - \int_0^u s\mathrm{d}s = \frac{1-u^2}{2} - \frac{u^2}{2} = 1-u^2 \leq 1}\mathrm{d}u \\ &\leq \int_0^1 f'(u)^2\mathrm{d}u.\end{aligned}$$

□

11.6 Sub-Exponential Concentration

Theorem 11.6.1. (Section 2.1 in Berestycki & Nickl)

Let $X = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ be a random vector, not necessarily independent. Suppose $\exists\gamma > 0$ where, for all $f : \mathbb{R}^n \to \mathbb{R}$ that are continuous differentiable, we have $\mathrm{Var} f(X) = \gamma\mathbb{E}\left\|\nabla f(X)\right\|^2$. Then, for any open set $A \in \mathscr{B}(\mathbb{R}^n)$ such that $\mathbb{P}^{(X)}(A) \geq \frac{1}{2}$, $\forall\varepsilon > 0$,

$$\mathbb{P}^{(X)}(A_r^c) \leq e^{-\frac{r}{3\sqrt{\gamma}}},$$

where $A_r := \{x \in \mathbb{R}^n : d(x, A) \leq r\}$.

Moreover, for any 1-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$, we have that $f(X) - \mathbb{E}f(X) \in sE(C\gamma)$ for some universal $C > 0$.

Proof.

(Somewhat Informal)

Let $A \in \mathscr{B}(\mathbb{R}^n)$ be an open set such that $a := \mathbb{P}^{(X)}(A) \geq \frac{1}{2}$. Let $\varepsilon > 0$ be arbitrary and let $b := \mathbb{P}^{(X)}(A_\varepsilon^c)$. Define $f : \mathbb{R}^n \to \mathbb{R}$ such that, $\forall x \in \mathbb{R}^n$

$$f(x) = \frac{1}{a} - \frac{\min(\varepsilon, d(x, A))}{\varepsilon}\left(\frac{1}{a} + \frac{1}{b}\right)$$

so that $f = \frac{1}{a}$ on A, $f = -\frac{1}{b}$ on A_ε^c, and for $x \in \mathrm{int}(A^c \cap A_\varepsilon)$, if $\nabla f(x)$ exists, then $\|\nabla f(x)\|_2 \leq \frac{1}{\varepsilon}(\frac{1}{a} + \frac{1}{b})$. To see this, note that for any unit vector h and scalar $t > 0$, if $x \mapsto d(x, A)$ is differentiable at x, then we have that $d(x + th, A) = d(x, A) + t(\nabla d(x, A))^\top h + O(t^2) \leq d(x, A) + t$ by triangle inequality.

Let $q(\cdot)$ be the bump function (Definition 5.3.1) on $\mathbb{R}^n$, let $\sigma \searrow 0$ and define $f_\sigma := x \mapsto \mathbb{E}_{Z\sim \mathrm{q}} f(x + \sigma Z)$. Then f_σ is continuously-differentiable, $f_\sigma \to f$ pointwise as $\sigma \to 0$, and $\lim_{\sigma\to 0} \mathrm{Var} f_\sigma(X) = \mathrm{Var} f(X)$. Morever,

$$\begin{aligned}\lim_{\sigma\to 0} \mathbb{E}\,\|\nabla f_\sigma(X)\|^2 &\leq \frac{1}{\varepsilon^2}\left(\frac{1}{a} + \frac{1}{b}\right)^2 \mathbb{P}^{(X)}(A^c \cap A_\varepsilon) \\ &= \frac{1}{\varepsilon^2}\left(\frac{1}{a} + \frac{1}{b}\right)^2 (1 - a - b).\end{aligned}$$

Thus, we have $\mathrm{Var} f(X) \leq \frac{\gamma}{\varepsilon^2}\left(\frac{1}{a} + \frac{1}{b}\right)^2 (1 - a - b)$ $(\star)$. On the other hand, we have that

$$\begin{aligned}\mathrm{Var} f(X) &= \int_{\mathbb{R}^n} (f - \mathbb{E} f)^2 \mathrm{d}\mathbb{P}^{(X)} \\ &\geq \int_A (f - \mathbb{E} f)^2 \mathrm{d}\mathbb{P}^{(X)} + \int_{A_\epsilon^c} (f - \mathbb{E} f) \mathrm{d}\mathbb{P}^{(X)} \\ &\geq a\left(\frac{1}{a} - \mathbb{E} f\right)^2 + b\left(-\frac{1}{b} - \mathbb{E} f\right)^2 \\ &\geq \min_{\mu\in\mathbb{R}} a\left(\frac{1}{a} - \mu\right)^2 + b\left(-\frac{1}{b} - \mu\right)^2 \\ &\geq \frac{1}{a} + \frac{1}{b}. \end{aligned} \tag{$\star\star$}$$

Putting $(\star)$ and $(\star\star)$ together, we have that

$$\frac{\varepsilon^2}{\gamma} \leq \left(\frac{1}{a} + \frac{1}{b}\right)(1 - a - b) \leq \frac{1 - a - b}{ab} = \frac{1 - a}{ab} - \frac{1}{a},$$

which implies

$$b \leq \frac{1 - a}{a} \frac{1}{\frac{1}{a} + \frac{\varepsilon^2}{\gamma}} \leq \frac{1 - a}{1 + \frac{\varepsilon^2}{2\gamma}}$$

since $a \geq \frac{1}{2}$. Hence, for $\varepsilon = \sqrt{2\gamma}$, we have $b \leq \frac{1-a}{2}$. So

$$\mathbb{P}^{(X)}(A_\varepsilon^c) \leq \frac{1}{2}\mathbb{P}^{(X)}(A^c) \leq \frac{1}{4}.$$

Iterating this argument, we have that $\forall k \in \mathbb{N}$,

$$\mathbb{P}^{(X)}(A_{k\varepsilon}^c) \leq 2^{-(k+1)}.$$

Fix $r > 0$ and let k be an integer such that $k\varepsilon < r \leq (k + 1)\varepsilon$; we then have that

$$\begin{aligned}\mathbb{P}^{(X)}(A_r^c) \leq \mathbb{P}^{(X)}(A_{k\varepsilon}^c) \leq 2^{-(k+1)} &\leq e^{-(k+1)\log 2} \\ &\leq e^{-\frac{r}{\varepsilon}(\log 2)} \leq \exp(-\frac{\log 2}{\sqrt{2}} \frac{r}{\sqrt{\gamma}}) \leq e^{-\frac{r}{3\sqrt{\gamma}}}.\end{aligned}$$

For the second claim, let $f : \mathbb{R}^n \to \mathbb{R}$ be 1-Lipschitz and first suppose $\mathbb{P}(f(X) = Mf) = 0$. Let $A := \{x \in \mathbb{R}^n : f(x) < Mf\}$ so that $\mathbb{P}(A) = 1/2$ and A is open. Let $t > 0$ be arbitrary. For any point $x \in \mathbb{R}^n$ such that $f(x) > Mf + t$, it holds that for any $y \in A$, $\|x - y\|_2 \leq |f(x) - f(y)| \leq t$. Thus, we have $\{x \in \mathbb{R}^n : f(x) > Mf + t\} \subseteq A_t^c$ and so

$$\mathbb{P}(f(X) - Mf(X) > t) \leq e^{-\frac{t}{3\sqrt{\gamma}}}.$$

Since $-f$ is also 1-Lipschitz, we have that $\mathbb{P}(|f(X) - Mf(X)| > t) \leq 2e^{-\frac{t}{3\sqrt{\gamma}}}$. Now, write $Z := f(X)$; we have

$$|\mathbb{E}Z - MZ| \leq \mathbb{E}|Z - MZ| = \int_0^\infty \mathbb{P}(|Z - MZ| \geq t)\mathrm{d}t \leq \int_0^\infty 2e^{-\frac{t}{3\sqrt{\gamma}}}\mathrm{d}t = 6\sqrt{\gamma}.$$

Thus,

$$\begin{aligned}\mathbb{P}(|Z - \mathbb{E}Z| > t) &= \mathbb{P}(|Z - MZ + MZ - \mathbb{E}Z| > t) \\ &\leq \mathbb{P}(|Z - MZ| > t - 6\sqrt{\gamma}) \leq 2e^{-\frac{t-6\sqrt{\gamma}}{3\sqrt{\gamma}}} \\ &\leq 2e^2 e^{-\frac{t}{3\sqrt{\gamma}}}.\end{aligned}$$

Now, if $\mathbb{P}(f(X) = Mf) > 0$, then let $\delta > 0$ and define $A = \{x : f(x) < Mf + \delta\}$. We may use the same reasoning to obtain that $\mathbb{P}(f(X) - Mf(X) > t - \delta) \leq e^{-\frac{t}{3\sqrt{\gamma}}}$. Since δ is arbitrary, the conclusion follows. $\square$

Theorem 11.6.2. Let $(\mathcal{X}, \mathscr{G})$ be some measurable space and let $X = (X_1, \ldots, X_n) : \Omega \to \mathcal{X}^n$ be independent random objects. Suppose $f : \mathcal{X}^n \to \mathbb{R}$ satisfy BDC$_b$ with $V > 0$ (see definition 11.5.1) (b)). Then, for all $t > 0$,

$$\mathbb{P}(f(X) - Mf(X) > t) \leq \exp\left(-\frac{t}{3\sqrt{V}}\right).$$

Moreover, if $-f$ also satisfies BDC$_b$ with $V > 0$, then $f(X) - \mathbb{E}f(X) \in sE(CV)$ for some universal $C > 0$.

Proof.

Define $F : \mathbb{R} \to [0,1]$ by $F(z) := \mathbb{P}(f(X) \leq z)$ and define $Q : [0,1] \to \mathbb{R}$ by $Q(\alpha) = \inf\{z \in \mathbb{R} : F(z) \geq \alpha\}$ as the quantile function of $f(X)$.

Note that $\forall \alpha \in [0,1]$, we have $F(Q(\alpha)) \geq \alpha$ and $F(Q(\alpha)^-) \leq \alpha$ where $F(Q(\alpha)^-) = \lim_{z_n \nearrow Q(\alpha)} F(z_n) = \mathbb{P}(f(X) < Q(\alpha))$. For $k \in \mathbb{N}$, define $a := Q(1-2^{-k}) \geq Mf(X)$ and $b := Q(1-2^{-(k+1)})$. Define $g_{a,b} : \mathcal{X}^n \to \mathbb{R}$ by

$$g_{a,b}(x) = \begin{cases} b & \text{if } f(x) \geq b \\ f(x) & \text{if } f(x) \in (a,b) \\ a & \text{if } f(x) \leq a \end{cases}.$$

Since $a \geq Mf(X)$, $\mathbb{E}g_{a,b}(X) \leq a\mathbb{P}(f(X) < a) + b\mathbb{P}(f(X) \geq a) \leq \frac{b+a}{2}$. Thus.

$$\begin{aligned}\mathrm{Var} g_{a,b}(X) &\geq \mathbb{P}(f(X) \geq b)(b - \mathbb{E}g_{a,b}(X))^2 \\ &\geq \frac{\mathbb{P}(f(X) \geq b)}{4}(b-a)^2 = \frac{1 - F(b^-)}{4}(b-a)^2 \geq \frac{2^{-(k+1)}}{4}(b-a)^2 \qquad (\star)\end{aligned}$$

On the other hand, note that for any $x \in \mathcal{X}^n$ and $x' \in \mathcal{X}^n$,

$$\sum_{i=1}^n (g_{a,b}(x_i, x_{-i}) - g_{a,b}(x'_i, x_{-i}))_+^2 \leq \sum_{i=1}^n (f(x_i, x_{-i}) - f(x'_i, x_{-i}))_+^2 \leq V$$

which implies $g_{a,b}$ satisfy BDC$_b$ with $V > 0$. Hence, by Efron-Stein inequality,

$$\begin{aligned}\mathrm{Var} g_{a,b}(X) &\leq \mathbb{E}\left[\sum_{i=1}^n \left(g_{a,b}(X_i, X_{-i}) - g_{a,b}(\widetilde{X}_i, X_{-i})\right)_+^2\right] \\ &= \mathbb{E}\left[\mathbb{1}_{\{f(X)>a\}} \sum_{i=1}^n \left(g_{a,b}(X_i, X_{-i}) - g_{a,b}(\widetilde{X}_i, X_{-i})\right)_+\right] \\ &\leq V\mathbb{P}(f(X) > a) \leq V(1 - F(a)) \leq V2^{-k}. \qquad (\star\star)\end{aligned}$$

Combining ($\star$) and ($\star\star$), we have

$$(b-a)^2 \leq 4V \frac{2^{-k}}{2^{-(k+1)}} \leq 8V,$$

which implies $b - a \leq 3\sqrt{V}$.

Applying the argument repeatedly, we have that $\forall k \in \mathbb{N}$,

$$\begin{aligned} & Q(1-2^{-(k+1)}) - Q(\frac{1}{2}) \leq 3\sqrt{V}k \\ \implies & Q(\frac{1}{2}) + 3\sqrt{V}k \geq Q(1-2^{-(k+1)}) \\ \implies & \underbrace{1 - F(Q(\frac{1}{2}) + 3\sqrt{V}k)}_{\mathbb{P}(f(X) > Q(\frac{1}{2}) + 3\sqrt{V}k)} \leq 1 - F(Q(1-2^{-(k+1)})) \leq 2^{-(k+1)}. \end{aligned}$$

Thus, for any $t > 0$, letting $k \in \mathbb{N}$ be such that $k3\sqrt{V} \leq t \leq (k+1)3\sqrt{V}$, we have

$$\mathbb{P}(f(X) > Mf(X) + t) \leq 2^{-(k+1)} \leq e^{-\frac{\log 2}{3}\frac{t}{\sqrt{V}}} \leq e^{-\frac{t}{3\sqrt{V}}}.$$

If $-f$ satisfy BDC$_b$ with $V > 0$ as well, then $\mathbb{P}(|f(X) - Mf(X)| > t) \leq 2e^{-\frac{t}{3\sqrt{V}}}$. We may then bound the deviation from the expectation in the same way as that of Theorem 11.6.1. $\square$

11.7 Entropy Functional

Definition 11.7.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $Z : \Omega \to [0, \infty)$ be non-negative random variable. Let $\Phi : [0, \infty) \to [0, \infty)$ be convex. Define the Φ-entropy of Z by

$$H_\Phi(Z) := \mathbb{E}\Phi(Z) - \Phi(\mathbb{E}Z). \tag{11.12}$$

Note that $H_\Phi(Z) \geq 0$. If $\Phi(x) = x^2$, then $H_\Phi(Z) = \mathrm{Var}(Z)$. In the case where $\Phi(x) = x \log x$ ($0 \log 0 = 0$), we write

$$\mathrm{Ent}(Z) := H_\Phi(Z) = \mathbb{E}(Z \log Z) - (\mathbb{E}Z)\log(\mathbb{E}Z).$$

We write $\Phi(x) := x \log x$ from this point on.

Remark 11.7.1. (a) By the equality condition of Jensen's inequality, we have that $\mathrm{Ent}(Z) = 0$ if and only if Z is constant a.s.

(b) Let $\mathbb{Q} \ll \mathbb{P}$ be another probability measure on $(\Omega, \mathscr{F})$ and let $Z := \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$. Then, $Z \geq 0$ and $\mathbb{E}Z = \int_\Omega Z \mathrm{d}\mathbb{P} = \int_\Omega \mathrm{d}\mathbb{Q} = 1$. Then,

$$\mathrm{Ent}(Z) = \int \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \log \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \mathrm{d}\mathbb{P} = \mathrm{KL}(\mathbb{Q}\|\mathbb{P}).$$

(c) Suppose $\mathbb{P}$ is uniform on $\{\omega_1, \ldots, \omega_n\}$ and $\mathbb{Q} \ll \mathbb{P}$, then

$$\begin{aligned} \mathrm{KL}(\mathbb{Q}\|\mathbb{P}) &= \sum_{i=1}^n \mathbb{Q}(\omega_i) \log \mathbb{Q}(\omega_i) + \log n \\ &= \log n - \underbrace{\left(-\sum_{i=1}^n \mathbb{Q}(\omega_i) \log \mathbb{Q}(\omega_i) \right)}_{\text{Shannon entropy of } \mathbb{Q}} \end{aligned}$$

Lemma 11.7.1. Let Z be non-negative random variable such that $\mathbb{E}\Phi(Z) < \infty$. Then,

$$\begin{aligned}\mathrm{Ent}(Z) &= \sup\left\{\mathbb{E}[UZ] : U \text{ r.v. } \mathbb{E}e^U = 1\right\}\\ &= \sup\left\{\mathbb{E}[Z(\log T - \log(\mathbb{E}T))] : T \geq 0 \text{ integrable r.v.}\right\}\end{aligned}$$

Note T need not be independent of Z.

Proof.

First, we claim that for any random variable U such that $\mathbb{E}e^U = 1$, $\mathbb{E}[UZ]$ is well-defined, that is, either $\mathbb{E}(UZ)_+ < \infty$ or $\mathbb{E}(UZ)_- < \infty$.

To see this, note that for any $u \in \mathbb{R}$,

$$\sup\{zu - z\log z : z \geq 0\} = e^{u-1}.$$

Hence, for any random variable U such that $\mathbb{E}e^U = 1$, we have

$$ZU - Z\log Z \leq e^{U-1} \implies \mathbb{E}(ZU)_+ \leq \frac{1}{e} + \mathbb{E}(Z\log Z)_+ < \infty.$$

Now, let U be any random variable such that $\mathbb{E}e^U = 1$. Define $\mathbb{Q} \ll \mathbb{P}$ such that $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = e^U$. Then,

$$\mathbb{E}_{\mathbb{Q}} Ze^{-U} = \int Ze^{-U}\mathrm{d}\mathbb{Q} = \int Z\mathrm{d}\mathbb{P} = \mathbb{E}Z$$

and that

$$\begin{aligned}\mathrm{Ent}_{\mathbb{Q}}(Ze^{-U}) &:= \int Ze^{-U}\log(Ze^{-U})\mathrm{d}\mathbb{Q} - (\mathbb{E}Z)\log(\mathbb{E}Z)\\ &= \int Z\log Z\mathrm{d}\mathbb{P} - \int ZU\mathrm{d}\mathbb{P} - (\mathbb{E}Z)\log(\mathbb{E}Z)\\ &= \mathrm{Ent}(Z) - \mathbb{E}ZU\\ \implies &\mathrm{Ent}(Z) \geq \mathbb{E}ZU.\end{aligned}$$

Taking $U = \log\frac{Z}{\mathbb{E}Z}$ yields that $\mathbb{E}e^U = 1$ and that $\mathbb{E}ZU = \mathrm{Ent}(Z)$. □

Theorem 11.7.1. Let $(\mathcal{X}, \mathcal{G})$ be a Polish space and let $X_1, \ldots, X_n : \Omega \to \mathcal{X}$ be independent random objects. Let $f : \mathcal{X}^n \to [0, \infty)$ and write $Y := f(X)$. Then we have,

$$\begin{aligned}\mathrm{Ent} f(X) :=& \mathbb{E}(Y\log Y) - (\mathbb{E}Y)\log(\mathbb{E}Y)\\ \leq& \mathbb{E}\sum_{i=1}^n \underbrace{\left\{\mathbb{E}_{\bullet|X_{-i}}(Y\log Y) - (\mathbb{E}_{\bullet|X_{-i}}Y)\log(\mathbb{E}_{\bullet|X_{-i}}Y)\right\}}_{\mathrm{Ent}_{\bullet|X_{-i}}Y}.\end{aligned}$$

Proof.

For $i \in [n]$, define $\mathscr{F}_i = \sigma(X_1, X_2, \ldots, X_i)$. Recall, by Lemma 11.5.1, that $\forall i \in [n]$, $\mathbb{E}_{\bullet|\mathscr{F}_{i-1}}Y = \mathbb{E}_{\bullet|X_{-i}}\mathbb{E}_{\bullet|\mathscr{F}_i}Y$. We observe that

$$Y(\log Y - \log\mathbb{E}Y) = \sum_{i=1}^n Y\left(\log\mathbb{E}_{\bullet|\mathscr{F}_i}Y - \log\mathbb{E}_{\bullet|\mathscr{F}_{i-1}}Y\right).$$

Thus,

$$\begin{aligned}\mathbb{E}(Y\log Y) - (\mathbb{E}Y)\log(\mathbb{E}Y) &= \mathbb{E}\sum_{i=1}^n \mathbb{E}_{\bullet|X_{-i}}\left\{Y\left(\log\mathbb{E}_{\bullet|\mathscr{F}_i}Y - \log\mathbb{E}_{\bullet|\mathscr{F}_{i-1}}Y\right)\right\}\\ &\leq \mathbb{E}\sum_{i=1}^n\left\{\mathbb{E}_{\bullet|X_{-i}}(Y\log Y) - (\mathbb{E}_{\bullet|X_{-i}}Y)\log(\mathbb{E}_{\bullet|X_{-i}}Y)\right\}\\ &= \mathbb{E}\sum_{i=1}^n \mathrm{Ent}_{\bullet|X_{-i}}Y,\end{aligned}$$

where the inequality holds by Lemma 11.7.1 using $T := \mathbb{E}_{\bullet|\mathscr{F}_i} Y$. A technical point is that $\mathbb{E}_{\bullet|X_{-i}} Y = \int Y \mathrm{d}\mathbb{P}_{\bullet|X_{-i}}$, where regular conditional distribution $\mathbb{P}_{\bullet|X_{-i}} : \mathscr{F} \times \mathcal{X}^{n-1} \to [0,1]$ exists because $\mathcal{X}^{n-1}$ is Polish. ☐

11.8 Subgaussian Concentration and Log-Sobolev Inequalities

Definition 11.8.1. For $x \in \{\pm 1\}^n$, define $x^{(i)} = \{x_1, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_n\} \in \{\pm 1\}^n$. For $f : \{\pm 1\}^n \to \mathbb{R}$, we define

$$\nabla f(x) := \frac{1}{2}(f(x) - f(x^{(1)}), \ldots, f(x) - f(x^{(i)}), \ldots, f(x) - f(x^{(n)}))$$

as the discrete gradient.

Proposition 11.8.1 (Log-Sobolev inequality for Rademacher)**.** Let $X : \Omega \to \{\pm 1\}^n$ be uniform and let $f : \{\pm 1\}^n \to \mathbb{R}$. We have that

$$\mathrm{Ent} f(X)^2 \leq \frac{1}{2}\sum_{i=1}^n \mathbb{E}\left(f(X) - f(X^{(i)})\right)^2 = 2\mathbb{E}\left\|\nabla f(X)\right\|^2.$$

Proof.

By Theorem 11.7.1, we have that

$$\mathrm{Ent} f(X)^2 \leq \mathbb{E}\sum_{i=1}^n \left\{\mathbb{E}_{\bullet|X_{-i}} f(X)^2 \log f(X)^2 - \left(\mathbb{E}_{\bullet|X_{-i}} f(X)^2\right) \log\left(\mathbb{E}_{\bullet|X_{-i}} f(X)^2\right)\right\}.$$

Thus, we need only show that for any univariate $f : \{\pm 1\} \to \mathbb{R}$,

$$\mathrm{Ent} f(X)^2 \leq \frac{1}{2}\mathbb{E}\left(f(X) - f(-X)\right)^2.$$

Write $a := f(1)$, $b = f(-1)$. We need to show that

$$\frac{a^2}{2}\log a^2 + \frac{b^2}{2}\log b^2 - \frac{a^2+b^2}{2}\log\frac{a^2+b^2}{2} \leq \frac{1}{2}(a-b)^2.$$

Since $(|a| - |b|)^2 \leq (a-b)^2$ and since the LHS contains only terms involving a^2 and b^2, we may assume that $a, b \geq 0$. We may also assume without loss of generality that $a \geq b$. Define

$$h(a) := \frac{a^2}{2}\log a^2 + \frac{b^2}{2}\log b^2 - \frac{a^2+b^2}{2}\log\frac{a^2+b^2}{2} - \frac{1}{2}(a-b)^2 \quad \text{for } a \geq b.$$

We check that (a) $h(b) = 0$, (b) $h'(b) = 0$, (c) h is concave. Therefore, $h(a) \leq 0$ as desired. ☐

Remark 11.8.1. More generally, if

$$X_i \sim \begin{cases} +1 & \text{w.p. } p \\ -1 & \text{w.p. } 1-p, \end{cases}$$

and $X = (X_1, \ldots, X_n)$ are independent, then, $\mathrm{Ent} f(X)^2 \leq c(p)\mathbb{E}\left\|\nabla f(X)\right\|^2$, where $c(p) = \frac{1}{1-2p}\log\frac{1-p}{p}$ for $p \neq 1/2$ and $c(1/2) = \lim_{p\to 1/2} c(p) = 2$.

Theorem 11.8.1. Let $X : \Omega \to \{\pm 1\}^n$ be uniform and suppose $f : \{\pm 1\}^n \to \mathbb{R}$ satisfy BCD_b with $V > 0$, i.e.,

$$\sup_{x\in\{\pm 1\}^n}\sum_{i=1}^n \left(f(x) - f(x^{(i)})\right)_+^2 \leq V.$$

Then, $f(X) - \mathbb{E}f(X) \in sG(\frac{v}{2})$.

Proof.

Let $\lambda \in \mathbb{R}$ and define, for $x \in \{\pm 1\}^n$, $g(x) = e^{\frac{\lambda f(x)}{2}}$. Write also $F(\lambda) := \mathbb{E}e^{\lambda f(X)} = \mathbb{E}g(X)^2$ and note that $F'(\lambda) = \mathbb{E}f(X)e^{\lambda f(X)}$.

Our goal is to show that $\forall \lambda \in \mathbb{R}$, $F(\lambda) \le e^{\frac{\lambda^2 V}{4} + \lambda \mathbb{E}f(X)}$. Assume $\lambda \neq 0$. Note,

$$\begin{aligned} \mathrm{Ent} g(X)^2 &= \mathbb{E}\lambda f(X)e^{\lambda f(X)} - F(\lambda)\log F(\lambda) \\ &= \lambda F'(\lambda) - F(\lambda)\log F(\lambda). \end{aligned} \tag{$\star$}$$

At the same time, by Proposition 11.8.1,

$$\begin{aligned} \mathrm{Ent} g(X)^2 &\le \frac{1}{2}\sum_{i=1}^n \mathbb{E}\left(e^{\frac{\lambda f(X)}{2}} - e^{\frac{\lambda f(X^{(i)})}{2}}\right)^2 \\ &= \sum_{i=1}^n \mathbb{E}\left(e^{\frac{\lambda f(X)}{2}} - e^{\frac{\lambda f(X^{(i)})}{2}}\right)_+^2 \end{aligned}$$

since $X \overset{d}{=} X^{(i)}$. By convexity of $z \mapsto e^z$, we have that $\forall z > y$, $e^{\frac{y}{2}} \ge e^{\frac{z}{2}} + \frac{1}{2}e^{\frac{z}{2}}(y-z) \implies \frac{1}{2}e^{\frac{z}{2}}(z-y) \ge e^{\frac{z}{2}} - e^{\frac{y}{2}}$. So

$$\begin{aligned} \mathrm{Ent} g(X)^2 &\le \sum_{i=1}^n \mathbb{E}\left(\frac{\lambda^2}{4}\left(f(X) - f(X^{(i)})\right)_+^2 e^{\lambda f(X)}\right) \\ &\le \frac{\lambda^2}{4}\mathbb{E}\left(e^{\lambda f(X)}\sum_{i=1}^n \left(f(X) - f(X^{(i)})_+\right)_+^2\right) \\ &\le \frac{\lambda^2}{4} V F(\lambda). \end{aligned} \tag{$\star\star$}$$

Combining ($\star$) and ($\star\star$), we have that

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \le \frac{\lambda^2}{4} V F(\lambda).$$

Since $F(\lambda) > 0$, $\forall \lambda \in \mathbb{R}$, we divide both sides by $\lambda^2 F(\lambda)$ to obtain

$$\frac{1}{\lambda}\frac{F'(\lambda)}{F(\lambda)} - \frac{1}{\lambda^2}\log F(\lambda) \le \frac{V}{4}.$$

Define $G(\lambda) = \frac{\log F(\lambda)}{\lambda}$ for $\lambda \neq 0$, then we have $G'(\lambda) \le \frac{V}{4}$. Define

$$G(0) = \lim_{\lambda \to 0} \frac{\log F(\lambda)}{\lambda} = \lim_{\lambda \to 0} \frac{F'(\lambda)}{F(\lambda)} = \frac{F'(0)}{F(0)} = \mathbb{E}f(X).$$

Then, for $\lambda > 0$,

$$G(\lambda) - G(0) = \int_0^\lambda G'(u)\mathrm{d}u \le \int_0^\lambda \frac{V}{4}\mathrm{d}u = \frac{\lambda V}{4}$$

So $\log F(\lambda) \le \frac{\lambda^2 V}{4} + \lambda \mathbb{E}f(X)$. Analogous analysis applies for $\lambda < 0$. $\square$

Proposition 11.8.2 (Gaussian Log-Sobolev Inequality)**.** Suppose $X \sim N(0, I_n)$ and suppose $f : \mathbb{R}^n \to \mathbb{R}$ is continuous differentiable. Then, $\mathrm{Ent} f(X)^2 \le 2\mathbb{E}\left\|\nabla f(X)\right\|^2$.

Proof.

By the same reasoning as Proposition 11.8.1, we need only consider the $n = 1$ case and show that for $f : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}f(X)^2 \log f(X)^2 - (\mathbb{E}f(X)^2)\log(\mathbb{E}f(X)^2) \le 2\mathbb{E}f'(X)^2.$$

We assume additionally that $\exists b > 0$ such that f is supported on $[-b, b]$ and that f'' exists and is continuous. Note then that $\sup_{x\in[-b,b]} |f''(X)| < \infty$. See proof of Theorem 11.5.3 on how the general case reduce to this.

Define $\varepsilon_1, \dots, \varepsilon_n$ as independent Rademacher random variable and define $S_n = \sum_{i=1}^n \varepsilon_i$. Since $\frac{1}{\sqrt{n}} S_n \xrightarrow{d} X$ and since f^2, $f^2 \log f^2$ are continuous and bounded, we have that

$$\lim_{n\to\infty} \mathrm{Ent} f(\frac{1}{\sqrt{n}} S_n)^2 = \mathrm{Ent} f(X) \tag{$\star$}$$

Moreover, since f' is continuous and bounded, we have that

$$\mathbb{E} \sum_{j=1}^n \left(f(\frac{1}{\sqrt{n}} S_n) - f(\frac{1}{\sqrt{n}} S_n - \frac{2\varepsilon_j}{\sqrt{n}}) \right)^2 = \mathbb{E} \sum_{j=1}^n \left(f'(\frac{1}{\sqrt{n}} S_n) \frac{2\varepsilon_j}{\sqrt{n}} + f''(\widetilde{Z}_j) \frac{4}{n} \right)^2$$

for some r.v. $\{\widetilde{Z}_j\}_{j=1}^n$, which implies

$$\lim_{n\to\infty} \mathbb{E} \sum_{j=1}^n \left(f(\frac{1}{\sqrt{n}} S_n) - f(\frac{1}{\sqrt{n}} S_n - \frac{2\varepsilon_j}{\sqrt{n}}) \right)^2 = 4\mathbb{E} f'(X)^2. \tag{$\star\star$}$$

Lastly, by Proposition 11.8.1,

$$\mathrm{Ent} f(\frac{1}{\sqrt{n}} S_n)^2 \le \frac{1}{2} \mathbb{E} \sum_{j=1}^n \left(f(\frac{1}{\sqrt{n}} S_n) - f(\frac{1}{\sqrt{n}} S_n - \frac{2\varepsilon_j}{\sqrt{n}}) \right)^2.$$

So the claim follows by ($\star$) and ($\star\star$).

$\square$

Theorem 11.8.2 (Tsirelson-Ibragimov-Sudakov Inequality)**.** Let $X \sim N(0, I_n)$ and $f : \mathbb{R}^n \to \mathbb{R}$ be L-Lipschitz. Then

$$f(X) - \mathbb{E} f(X) \in sG(L^2).$$

Proof.

We assume additionally that f is continuous differentiable. Note that $\|\nabla f\|_\infty \le L$. Assume also that $\mathbb{E} f(X) = 0$.

Let $\lambda \in \mathbb{R}$ and define, for $x \in \mathbb{R}^n$, $g(x) = e^{\frac{\lambda f(x)}{2}}$. We have that g is continuous differentiable and that $\nabla g(x) = \frac{\lambda}{2} \nabla f(x) e^{\frac{\lambda}{2} f(x)}$.

Write, for $\lambda \in \mathbb{R}$, $F(\lambda) = \mathbb{E} e^{\lambda f(x)} = \mathbb{E} g(X)^2$. Then,

$$\mathrm{Ent} g(X)^2 = \lambda F'(\lambda) - F(\lambda) \log F(\lambda).$$

At the same time, by Proposition 11.8.1,

$$\begin{aligned} &\mathrm{Ent} g(X)^2 \le 2\mathbb{E} \|\nabla g(X)\|^2 \le \frac{\lambda^2}{4} \mathbb{E}\big(\|\nabla f(X)\|_2^2 e^{\lambda f(X)} \big) \le \frac{\lambda^2}{2} L^2 F(\lambda) \\ \implies &\lambda F'(\lambda) - F(\lambda) \log F(\lambda) \le \frac{\lambda^2}{2} L^2 F(\lambda). \end{aligned}$$

Th rest of the proof proceeds in the same way as that of Theorem 11.8.1. $\square$

Example 11.8.1. (a) Let A be random matrix taking value in $\mathbb{R}^{n\times m}$ and suppose $A_{ij} \sim N(0,1)$, $\forall i, j \in [n]$ iid. Then, the largest singular value $s_{\max} : \mathbb{R}^{n\times m} \to [0, \infty)$ is 1-Lipschitz (w.r.t. Frobenius norm) and

$$\mathbb{P}(|s_{\max}(A) - \mathbb{E} s_{\max}(A)| \ge t) \le 2e^{-\frac{t^2}{2}}.$$

(b) Let $X \sim N(0,\Sigma)$ take value on $\mathbb{R}^n$, then $X = \Sigma^{\frac{1}{2}} Z$ where $Z \sim N(0, I_n)$. Define, for $p \in [1,\infty]$ and $z \in \mathbb{R}^n$, $f(z) := \|\Sigma^{\frac{1}{2}} z\|_p$. Define the $\ell_2 \to \ell_p$ operator norm $L := \|\Sigma^{\frac{1}{2}}\|_{\ell_2 \to \ell_p} := \sup\{\|\Sigma^{\frac{1}{2}} z\|_p : \|z\|_2 = 1\}$. Then, $\forall z, \tilde{z} \in \mathbb{R}^n$,

$$|f(z) - f(\tilde{z})| \leq \|\Sigma^{\frac{1}{2}}(z-\tilde{z})\|_p \leq L\,\|z-\tilde{z}\|_2 .$$

Thus, $\|X\|_p - \mathbb{E}\,\|X\|_p \in sG(L^2)$. Note that $\|\Sigma^{\frac{1}{2}}\|_{\ell_2\to\ell_\infty} = \max_{i\in[n]} \|\Sigma_{i\cdot}^{\frac{1}{2}}\|_2$ where $\Sigma_{i\cdot}$ is the i-th row of $\Sigma^{\frac{1}{2}}$.

(c) Same set-up as (b). Note that $\forall i \in [n]$, $\|\Sigma_{i\cdot}^{\frac{1}{2}}\|_2^2 = \mathbb{E}X_i^2$. Hence, writing $\tilde{\sigma}^2 := \max_{i\in[n]} \mathbb{E}X_i^2$, we have that

$$\max_{i\in[n]} X_i - \mathbb{E}[\max_{i\in[n]} X_i] \in sG(\tilde{\sigma}^2).$$

Note that this is true $\forall n \in \mathbb{N}$. Thus, let T be a countable set and let $X : \Omega \to \mathbb{R}^T$ be a centered Gaussian process. Assume without loss of generality that $T = \mathbb{N}$. We have,

$$\sup_{t\in T} X_t(\omega) = \lim_{n\to\infty} \max_{i\in[n]} X_i(\omega) \quad \forall \omega \in \Omega,$$

that $\mathbb{E}\sup_{t\in T} X_t = \lim_{n\to\infty} \mathbb{E}|\max_{i\in[n]} X_i|$ by monotone convergence. And we have that $\forall\lambda > 0$,

$$\begin{aligned}\mathbb{E}e^{\lambda(\sup_{t\in T} X_t - \mathbb{E}\sup_{t\in T} X_t)} &= \lim_{n\to\infty} \mathbb{E}e^{\lambda(\max_{i\in[n]} X_i - \mathbb{E}[\max_{i\in[n]} X_i])} && \text{(by MCT again)}\\ &\leq \lim_{n\to\infty} e^{\frac{\lambda^2}{2}\max_{i\in[n]}\mathbb{E}X_i^2} = e^{\frac{\tilde{\sigma}^2\lambda^2}{2}}\end{aligned}$$

where $\tilde{\sigma}^2 = \sup_{t\in T}\mathbb{E}X_t^2$ and same analysis applies for $\lambda < 0$. So

$$\sup_{t\in T} X_t - \mathbb{E}\sup_{t\in T} X_t \in sG(\tilde{\sigma}^2).$$

11.9 Entropy Method

Lemma 11.9.1 (Herbst)**.** Let Z be an integrable r.v. such that, for some $V > 0$, $\forall\lambda > 0$,

$$\begin{aligned}\frac{\mathrm{Ent}(e^{\lambda z})}{\mathbb{E}e^{\lambda Z}} &= \frac{\lambda\mathbb{E}Ze^{\lambda Z} - (\mathbb{E}e^{\lambda Z})\log(\mathbb{E}e^{\lambda Z})}{\mathbb{E}e^{\lambda Z}}\\ &= \frac{\lambda F'(\lambda)}{F(\lambda)} - \log F(\lambda) \leq \frac{\lambda^2 V}{2}\end{aligned}$$

where $F(\lambda) := \mathbb{E}e^{\lambda Z}$. Then, $\forall\lambda > 0$, $\log\mathbb{E}e^{\lambda(Z-\mathbb{E}Z)} \leq \frac{\lambda^2 V}{2}$.

Same holds for $\lambda < 0$.

Proof.

See proof of Theorem 11.8.1 □

Remark 11.9.1. Recall Hoeffding's lemma (see proof of Theorem 11.3.1), which states that if r.v. $Z \in [a,b]$, then, writing $\psi(\lambda) = \log\mathbb{E}e^{\lambda(Z-\mathbb{E}Z)}$, we have that $\psi''(\lambda) \leq \frac{(b-a)^2}{4}$, $\forall\lambda \in \mathbb{R}$. Since $\frac{\mathrm{d}}{\mathrm{d}\lambda}\{\lambda\psi'(\lambda) - \psi(\lambda)\} = \lambda\psi''(\lambda)$ and $\psi(0) = 0$, we have that, $\forall\lambda \in \mathbb{R}$,

$$\begin{aligned}\lambda\psi'(\lambda) - \psi(\lambda) &= \int_0^\lambda t\psi''(t)\mathrm{d}t \leq \frac{(b-a)^2}{8}\lambda^2,\\ \implies \frac{\mathrm{Ent}(e^{\lambda Z})}{\mathbb{E}e^{\lambda Z}} &\leq \frac{(b-a)^2\lambda^2}{8}.\end{aligned}$$

Theorem 11.9.1. Let $(\mathcal{X}, \mathscr{G})$ be a measurable space and let $X = (X_1, \ldots, X_n) : \Omega \to \mathcal{X}^n$ be independent r.o. Let $f : \mathcal{X}^n \to \mathbb{R}$ and suppose $\exists V > 0$ such that

$$\sup_{x,x' \in \mathcal{X}^n} \sum_{i=1}^{n} \left(f(x_i, x_{-i} - f(x_i', x_{-i})\right)^2 \leq V.$$

Then, $f(X) - \mathbb{E}f(X) \in sG(2V)$.

Proof.
Note that $\forall x \in \mathcal{X}^n, x' \in \mathcal{X}^n, x'' \in \mathcal{X}^n$,

$$\sum_{i=1}^{n} \left(f(x_i', x_{-i}) - f(x_i'', x_{-i})\right)^2 = \sum_{i=1}^{n} \left(f(x_i', x_{-i}) - f(x_i, x_{-i}) + f(x_i, x_{-i}) - f(x_i'', x_{-i})\right)^2 \leq 4V.$$

Thus, define, for $i \in [n]$,

$$c_i(x_{-i}) = \sup_{x_i' \in \mathcal{X}, x_i'' \in \mathcal{X}} |f(x_i', x_{-i}) - f(x_i'', x_{-i})|,$$

we have that $\forall x \in \mathcal{X}^n$, $\sum_{i=1}^{n} c_i(x_{-i})^2 \leq 4V$. Then, $\forall \lambda \in \mathbb{R}$,

$$\begin{aligned}
\operatorname{Ent}(e^{\lambda f(X)}) &\leq \mathbb{E} \sum_{i=1}^{n} \operatorname{Ent}_{\cdot|X_{-i}} e^{\lambda f(X)} \\
&\leq \mathbb{E} \sum_{i=1}^{n} \frac{c_i(X_{-i})^2 \lambda^2}{8} \mathbb{E}_{\cdot|X_{-i}} e^{\lambda f(X)} && \text{(By Remark 11.9.1)} \\
&= \mathbb{E} \left\{ e^{\lambda f(X)} \frac{\lambda^2}{8} \sum_{i=1}^{n} c_i(X_{-i})^2 \right\} \\
&\leq \frac{\lambda^2}{2} V \mathbb{E} e^{\lambda f(X)}.
\end{aligned}$$

The claim follows by Lemma 11.9.1. □

Lemma 11.9.2. Let $I \subseteq \mathbb{R}$ be an open interval and let $f : I \to \mathbb{R}$ be convex and differentiable. Let $X : \Omega \to I$ be a r.v. such that $\mathbb{E}X \in I$. Then,

$$\mathbb{E}[f(X) - \mathbb{E}f(X)] = \inf_{a \in I} \mathbb{E}[f(X) - f(a) - f'(a)(X - a)].$$

Note: for $x, y \in I$, the Bregman divergence between x, y w.r.t. f is $f(y) - f(x) - f'(x)(y - x)$.

Proof.
Let $a \in I$, then

$$\mathbb{E}f(X) - f(a) - f'(a)(\mathbb{E}X - a) \geq \mathbb{E}f(X) - f(\mathbb{E}X)$$

by convexity of f. Equality is attained at $a = \mathbb{E}X$. □

Corollary 11.9.1. For any r.v. $Z \geq 0$,

$$\operatorname{Ent} Z = \inf_{u > 0} \mathbb{E}[Z(\log Z - \log u) - (Z - u)].$$

Proof.
Apply Lemma 11.9.2 with $f(x) = x \log x$ and $I = (0, \infty)$. □

Theorem 11.9.2. Let $X = (X_1, \dots, X_n) : \Omega \to \mathcal{X}^n$ be independent r.o. and let $f : \mathcal{X}^n \to \mathbb{R}$ satisfy BDC_b with $V > 0$, i.e.

$$\sup_{x,x' \in \mathcal{X}^n} \sum_{i=1}^n (f(x_i, x_{-i}) - f(x_i', x_{-i}))_+^2 \le V.$$

Then, $\forall \lambda > 0$, $\log \mathbb{E} e^{\lambda(f(X) - \mathbb{E} f(X))} \le \frac{V\lambda^2}{2}$ and $\forall t > 0$,

$$\mathbb{P}(f(X) - \mathbb{E} f(X) \ge t) \le e^{-\frac{t^2}{2V}}.$$

Proof.

First, we note that $\forall x \in \mathcal{X}^n$,

$$\sum_{i=1}^n \left(f(x_i, x_{-i}) - \inf_{\tilde{x} \in \mathcal{X}} f(\tilde{x}, x_{-i}) \right)^2 = \sup_{x_i \in \mathcal{X}^n} \sum_{i=1}^n (f(x_i, x_{-i}) - f(x_i', x_{-i}))_+^2 \le V.$$

We write $Z := f(X)$ and $Z^{(i)} := \inf_{\tilde{x} \in \mathcal{X}} f(\tilde{x}, x_{-i})$. (If $x_{-i} \mapsto \inf_{\tilde{x} \in \mathcal{X}} f(\tilde{x}, x_{-i})$ is not measurable function $\mathcal{X}^n \to \mathbb{R}$, then take the smallest measurable majorant.)

Then, we have that, for $\lambda > 0$,

$$\begin{aligned}
\mathrm{Ent} e^{\lambda Z} &\le \mathbb{E} \sum_{i=1}^n \mathrm{Ent}_{\cdot | X_{-i}} e^{\lambda Z} \\
&\le \mathbb{E} \sum_{i=1}^n \mathbb{E}_{\cdot | X_{-i}} \left\{ e^{\lambda Z} (\lambda Z - \lambda Z^{(i)}) - (e^{\lambda Z} - e^{\lambda Z^{(i)}}) \right\} \qquad \text{(by Corollary 11.9.1 with } u = e^{\lambda Z^{(i)}}) \\
&= \mathbb{E} \sum_{i=1}^n \mathbb{E}_{\cdot | X_{-i}} \left\{ e^{\lambda Z} \underbrace{(e^{-(\lambda Z - \lambda Z^{(i)})} + (\lambda Z - \lambda Z^{(i)}) - 1)}_{\phi(-\lambda(z - z^{(i)})) \text{ where } \phi(x) = e^x - x - 1 \, \forall x \in \mathbb{R}} \right\}
\end{aligned}$$

Since $\phi(x) \le \frac{x^2}{2}$, $\forall x \le 0$ and since $-\lambda(Z - Z^{(i)}) \le 0$, we further have

$$\begin{aligned}
\mathrm{Ent} e^{\lambda Z} &\le \mathbb{E} \sum_{i=1}^n \mathbb{E}_{\cdot | X_{-i}} \left[e^{\lambda Z} \frac{\lambda^2 (Z - Z^{(i)})^2}{2} \right] \\
&= \mathbb{E} \left[e^{\lambda Z} \frac{\lambda^2}{2} \sum_{i=1}^n (Z - Z^{(i)})^2 \right] \le \frac{\lambda^2}{2} V \mathbb{E} e^{\lambda Z}.
\end{aligned}$$

The claim follows by Lemma 11.9.1. ☐

Remark 11.9.2. In the proof of Theorem 11.9.2 , we cannot show that $\forall t \ge 0$,

$$\mathbb{P}(Z - \mathbb{E} Z \le -t) \le e^{-\frac{t^2}{2V}}.$$

We do have that $Z \in sE(V)$ from Theorem 11.6.2.

Corollary 11.9.2. Let $X_1, \dots, X_n$ be independent random variables in $[0, 1]$. Let $f : [0, 1]^n \to \mathbb{R}$ be coordinate-wise convex and 1-Lipschitz. Then

$$\forall \, t > 0, \mathbb{P}(f(X) - \mathbb{E} f(X) \ge t) \le e^{-\frac{t^2}{2}}.$$

Proof.

Since f is coordinate-wise convex, $\forall x, x' \in [0, 1]^n$, $\forall i \in [n]$,

$$f(x_i, x_{-i}) - f(x_i', x_{-i}) \le \partial_{x_i} f(x)(x_i' - x_i) \le \partial_{x_i} f(x)$$

where $\partial_{x_i} f(x)$ is any sub-gradient of $f(\cdot, x_{-i})$ at x_i. So

$$\sum_{i=1}^n (f(x_i, x_{-i}) - f(x_i', x_{-i}))_+^2 \le \sum_{i=1}^n (\partial_{x_i} f(x))^2 \le 1$$

since f is 1-Lipschitz. Thus, f satisfies BDC_b with $V = 1$. ☐

11.10 Johnson Lindenstrauss Lemma

We will use the Lipschitz concentration inequality for uniform distribution on the sphere and for Gaussian distribution to derive two versions of the Johnson Lindenstrauss Lemma.

First, we formally state the Lipschitz concentration inequality on the unit sphere.

Theorem 11.10.1. Let $X \in \mathbb{S}^{n-1}$ be a random unit vector uniformly distributed on surface of the n-dimensional unit ball. Let $f : \mathbb{S}^{n-1} \to \mathbb{R}$ be an L-Lipschitz function (with respect to Euclidean norm). Then, we have that $f(X) - \mathbb{E}f(X) \in sG\big(2\frac{L^2}{n-1}\big)$.

Proof.
(Uses informal arguments)

From example 11.1.1, in particular displayed equation (11.2), we have, for all $t \geq 0$,

$$\mathbb{P}(|f(X) - Mf(X)| > t) \leq 2e^{-\frac{n-1}{2}\frac{t^2}{L^2}}.$$

From this, we have that,

$$\begin{aligned}|\mathbb{E}f(X) - Mf(X)| \leq \mathbb{E}|f(X) - Mf(X)| &\leq \int \mathbb{P}(|f(X) - Mf(X)| > t)\,dt \\ &\leq 2\int_0^\infty e^{-\frac{n-1}{2}\frac{t^2}{L^2}}\,dt = 2\sqrt{2\pi}\frac{L}{\sqrt{n-1}}.\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}\mathbb{P}(|f(X) - \mathbb{E}f(X)| > t) &= \mathbb{P}(|f(X) - Mf(X) + Mf(X) - \mathbb{E}f(X)| > t) \\ &\leq \mathbb{P}(|f(X) - \mathbb{E}f(X)| > t - |Mf(X) - \mathbb{E}f(X)|) \\ &\leq 2\exp\Big(-\frac{n-1}{2L^2}(t - 2\sqrt{2\pi}\frac{L}{\sqrt{n-1}})^2\Big) \leq 2\exp\Big(-\frac{n-1}{2L^2}\big\{\frac{t^2}{2} - \frac{4\pi L^2}{n-1}\big\}\Big) \\ &\leq 2e^{2\pi}e^{-\frac{n-1}{4}\frac{t^2}{L^2}}.\end{aligned}$$

□

Remark 11.10.1. Before stating the Johnson Lindenstrauss lemma, we need to define the notion of a uniformly random projection. For $p \in \mathbb{N}$, define $O(p) := \{U \in \mathbb{R}^{p\times p} : U^\top U = I_p\}$ as the set of rotation matrices. We say that a random rotation matrix U is *uniformly distributed on* $O(p)$ if for any **fixed** $\tilde{U} \in O(p)$, we have that $U \stackrel{d}{=} \tilde{U}U$. An important result called Haar's Theorem shows that this invariance condition is enough to guarantee the existence and uniqueness of a distribution. Haar's theorem applies broadly to any locally compact Hausdorff topological group; we will not discuss its details here.

There are many ways to generate a uniformly random rotation matrix $U \in O(p)$. One simple example is to generate a random matrix $X \in \mathbb{R}^{p\times p}$ such that all the entries and iid $N(0, 1)$. Then, we compute the SVD $X = U\Sigma V^\top$ and take the left singular vector matrix U as the desired sample. To see that this is a valid sampling algorithm, observe that for any fixed rotation $\tilde{U} \in O(p)$, we have that $\tilde{U}X \stackrel{d}{=} X$ and that $\tilde{U}U$ is the left singular vector matrix of $\tilde{U}X$. Therefore, $\tilde{U}U \stackrel{d}{=} U$.

Let q be an integer such that $q \leq p$. We define $\Pi : \mathbb{R}^p \to \mathbb{R}^q$ as a random projection if there exists a uniformly random $U \in O(p)$ such that Π comprise the first q rows of U.

Theorem 11.10.2. Let $x_1, \dots, x_n \in \mathbb{R}^p$ be an arbitrary collection of points and let $\epsilon > 0$. There exists universal constants $c > 0$ such that, for any $q \geq \frac{4}{c\epsilon^2} \log n$, if $\Pi \in \mathbb{R}^{q \times p}$ is a random projection (described in the previous remark), then, with probability at least $1 - 2e^{-\frac{c}{2}\epsilon^2 q} \geq 1 - 2n^{-4}$,

$$\text{for all } i, j \in [n]^2,\ \ (1-\epsilon)\|x_i - x_j\|_2 \leq \sqrt{p/q}\|\Pi x_i - \Pi x_j\|_2 \leq (1+\epsilon)\|x_i - x_j\|_2.$$

Or, equivalently,

$$\sup_{(i,j)\in[n]^2} \left| \frac{\sqrt{p/q}\|\Pi x_i - \Pi x_j\|_2}{\|x_i - x_j\|_2} - 1 \right| \leq \epsilon.$$

Proof.

Assume $q < p$ or else there is nothing to prove. Let $D \in \mathbb{R}^{q\times p}$ denote a matrix of the form $D = [I_q; 0]$. Let $z \in \mathbb{R}^p$ and suppose $\|z\|_2 = 1$. Since there exists a uniformly random $U \in O(p)$ such that $\Pi = DU$, we have that

$$\Pi x = D(Uz) = DZ,$$

where Z is a random vector uniformly distributed on $\mathbb{S}^{p-1}$.

Therefore, by symmetry, we have that

$$\mathbb{E}\|\Pi x\|_2^2 = \frac{q}{p}. \tag{11.13}$$

Moreover, since $z \mapsto \|Dz\|_2$ is 1-Lipschitz, we have that $\|\Pi x\|_2 - \mathbb{E}\|\Pi x\|_2 \in sG(2/(p-1))$.

By Jensen's inequality, we have that $\mathbb{E}\|\Pi x\|_2 \leq (\mathbb{E}\|\Pi x\|_2^2)^{1/2} \leq \sqrt{q/p}$. We may also get a lower bound:

$$\begin{aligned}
\mathbb{E}\|\Pi x\|_2 &\geq \left\{\mathbb{E}\|\Pi x\|_2^2 - \mathrm{Var}(\|\Pi x\|_2)\right\}^{1/2} \\
&\geq \left\{\mathbb{E}\|\Pi x\|_2^2 - \frac{8}{p-1}\right\}^{1/2} = \sqrt{\frac{q}{p}}\left(1 - \frac{8p}{q(p-1)}\right)^{1/2} \\
&\geq \sqrt{\frac{q}{p}} - \frac{c_2}{q}\sqrt{\frac{q}{p}},
\end{aligned}$$

where, in the last inequality, we used the fact that $(1-z)^{1/2} \geq 1 - 2z$ for all $z \leq 1/2$ and where $c_2 > 0$ is a universal constant.

Therefore, by Theorem 11.10.1, we have that

$$\begin{aligned}
\mathbb{P}\left(\left|\|\Pi x\|_2 - \sqrt{\frac{q}{p}}\right| > \epsilon\sqrt{\frac{q}{p}}\right) &\leq \mathbb{P}\left(\left|\|\Pi x\|_2 - \mathbb{E}\|\Pi x\|_2\right| > (\epsilon - c_2/q)\sqrt{\frac{q}{p}}\right) \\
&\leq 2\exp\left\{-\frac{q}{8}(\epsilon - c_2/q)^2\right\} \leq 2\exp\left\{-\frac{q}{16}\epsilon^2 + \frac{c_2^2}{8q}\right\} \leq 2e^{-cq\epsilon^2},
\end{aligned}$$

where $c > 0$ is a universal constant.

Assume that $\epsilon^2 \geq \frac{4}{cq}\log n$, then, by a union bound,

$$\begin{aligned}
&\mathbb{P}\left(\text{for all } (i,j) \in [n]^2,\ \left|\|\Pi(x_i - x_j)\|_2 - \sqrt{\frac{q}{p}}\right| > \epsilon\sqrt{\frac{q}{p}}\|x_i - x_j\|_2^2\right) \\
&\leq 2\exp\left(-cq\epsilon^2 + 2\log n\right) \leq 2e^{-\frac{c}{2}q\epsilon^2}.
\end{aligned}$$

$\square$

Theorem 11.10.3. Let $x_1, \dots, x_n \in \mathbb{R}^p$ be an arbitrary collection of points and let $\epsilon > 0$. Let Π be a random matrix taking value in $\mathbb{R}^{q\times p}$ such that every entry is iid $N(0, 1/p)$.

There exists universal constants $c > 0$ such that, for any $q \geq \frac{4}{c\epsilon^2}\log n$, with probability at least $1 - 2e^{-\frac{c}{2}\epsilon^2 q} \geq 1 - 2n^{-4}$,

$$\text{for all } i, j \in [n]^2,\ \ (1-\epsilon)\|x_i - x_j\|_2 \leq \sqrt{p/q}\|\Pi x_i - \Pi x_j\|_2 \leq (1+\epsilon)\|x_i - x_j\|_2.$$

Proof.

Same as that of Theorem 11.10.2 but we use TIS inequality. □

11.11 Empirical Process

Definition 11.11.1. Let $(\mathcal{X}, \mathscr{G})$ be a measurable space and let $(X_1, \dots, X_n) : \Omega \to \mathcal{X}^n$ be independent. Let $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}}$ be countable (to avoid measurability issues) and assume $\exists a, b \in \mathbb{R}$ such that $\forall x \in \mathcal{X}$, $a \le h(x) \le b$.

We define empirical process $G_n : \Omega \to \mathbb{R}^{\mathcal{H}}$ where

$$G_n(\omega, h) = \sum_{i=1}^{n} h(X_i(\omega)).$$

Define $f : \mathcal{X}^n \to \mathbb{R}$ as $f(x_1, \dots, x_n) = \sup_{h \in \mathcal{H}} \sum_{i=1}^{n} h(x_i)$ and the supremum of empirical process as

$$f(X_1, \dots, X_n) = \sup_{h \in \mathcal{H}} \sum_{i=1}^{n} h(X_i) = \sup_{h \in \mathcal{H}} G_n(\cdot, h).$$

Theorem 11.11.1. Let $(X_1, \dots, X_n) : \Omega \to \mathcal{X}^n$ be independent and let $\mathcal{H} \subseteq \{h \in \mathbb{R}^{\mathcal{X}} : h \in [a, b]\}$ for $a < b \in \mathbb{R}$ be countable. Write $f(X) = \sup_{h \in \mathcal{H}} \sum_{i=1}^{n} h(X_i)$. Then,

$$f(X) - \mathbb{E} f(X) \in sG(n(b-a)^2).$$

Proof.

Let $x, x' \in \mathcal{X}^n$ and fix $i \in [n]$. Note that

$$\begin{aligned}(f(x_i, x_{-i}) - f(x_i', x_{-i}))_+^2 &\le \left\{ \sup_{h \in \mathcal{H}} h(x_i) + \sum_{j \ne i} h(x_j) - h(x_i') - \sum_{j \ne i} h(x_j) \right\}^2 \\ &\le \sup_{h \in \mathcal{H}} (h(x_i) - h(x_i'))^2 \le (b-a)^2.\end{aligned}$$

Likewise, we have $(f(x_i, x_{-i}) - f(x_i', x_{-i}))_-^2 \le (b-a)^2$. The claim follows from Theorem 11.9.1. □

Theorem 11.11.2. Let $(X_1, \dots, X_n) : \Omega \to \mathcal{X}^n$ be independent and let $\mathcal{H} \subseteq \{h \in \mathbb{R}^{\mathcal{X}} : |h| \le 1\}$ be countable. Assume that $\forall h \in \mathcal{H}$,

$$\mathbb{E} h(X_1) = \mathbb{E} h(X_2) = \cdots = \mathbb{E} h(X_n) = 0.$$

Define $\sigma^2 := \sup_{h \in \mathcal{H}} \sum_{i=1}^{n} \mathbb{E} h(X_i)^2$ and $\tilde{\sigma}^2 := \mathbb{E} \sup_{h \in \mathcal{H}} \sum_{i=1}^{n} h(X_i)$. Write $f(X) = \sup_{h \in \mathcal{H}} \sum_{i=1}^{n} h(X_i)$. Then,

$$f(X) - \mathbb{E} f(X_i) \in s\Gamma(2(\tilde{\sigma}^2 + \sigma^2), 2),$$

i.e., $\forall t \ge 0$,

$$\mathbb{P}(|f(X) - \mathbb{E} f(X)| \ge t) \le \exp\left(-\frac{t^2}{4(\tilde{\sigma}^2 + \sigma^2) + 2t} \right).$$