Probability Essentials

Brain Fralix and Peter C. Kiessler
(Notes by Shuai Wei)

October 4, 2023

Contents

1 Sets and Events **1**
1.1 Indicator functions 1

2 Probability Spaces **5**
2.1 Basic Definitions and Properties 5
2.2 More on Closure 7
2.3 Two Constructions 8
2.4 Construction of Probability Spaces 9
2.5 Measure Constructon 9
2.5.1 Lebesgue Measure on $(0, 1]$ 9
2.5.2 Construction of a Probability Measure on $\mathbb{R}$ with Given Distribution Function $F(x)$. 9

3 Random Variables, Elements, and Measurable Maps **11**
3.1 Inverse Maps 11
3.2 Measurable Maps, Random Elements, Induced Probability Measures 11
3.3 σ-fields Generated by Maps 14

4 Independence **15**
4.1 Basic Definitions 15
4.2 Independent Random Variables 15
4.3 Two Examples of Independence 16
4.3.1 Records, Ranks, Renyi Theorem 16
4.3.2 Dyadic Expansions of Uniform Random Numbers 18
4.4 Independence, Zero-One Laws, Borel-Cantelli Lemma 19
4.4.1 Borel-Cantelli Lemma 19
4.4.2 Kolmogorov Zero-One law 20

5 Integration and Expectation **23**
5.1 Simple Functions 23
5.2 Expectation and Integration 24
5.2.1 Expectation of Simple Functions 24
5.2.2 Extension of the Definition 24
5.3 The Transformation Theorem and Densities 25
5.3.1 Expectation is Always an Integral on $\mathbb{R}$ 26

5.3.2 Densities . . . 26
5.4 Product Spaces, Independence, Fubini Theorem . . . 26
5.5 Probability Measures on Product Spaces . . . 28
5.6 Approximation theorem for measures . . . 31
5.7 Fubini's theorem . . . 32

6 Convergence Concepts **39**
6.1 Almost Sure Convergence . . . 39
6.2 Convergence in Probability . . . 39
6.3 Statistical Terminology . . . 40

7 Laws of Large Numbers and Sums of Independent Random Variables **41**
7.1 Truncation and Equivalence . . . 41
7.2 A General Weak Law of Large Numbers . . . 42
7.3 Almost Sure Convergence of Sums of Independent Random Variables . . . 42
7.4 Strong Laws of Large Numbers . . . 42
7.4.1 Record counts . . . 42
7.4.2 Explosions in the Pure Birth Process . . . 43
7.5 The Strong Law of Large Numbers for IID Sequences . . . 43

8 Convergence in Distribution **45**
8.1 Basic Definitions . . . 45
8.2 Scheffe' lemma . . . 46
8.2.1 Scheffe's lemma and Order Statistics . . . 47
8.3 Appetizer . . . 49
8.4 Left Continuous Inverse . . . 50
8.5 The Baby Skorohod Theorem . . . 52
8.6 The Delta Method . . . 54
8.7 More Relations Among Modes of Convergence . . . 56
8.8 Convergence of types theorem . . . 59
8.8.1 Maximum process . . . 75
8.8.2 Extremal process . . . 76

9 Charateristic Functions and the Central Limit Theorem **79**
9.1 Characteristic Functions . . . 79
9.2 Uniqueness and Continuity Theorem . . . 85
9.3 The Classical CLT for iid Random Variables . . . 93
9.4 Lindeberg CLT . . . 96
9.5 CLT for m-dependent random variables . . . 104

10 Point process **109**
10.1 Borel Measure . . . 109
10.2 Measure . . . 110
10.3 Random measure and Point process . . . 112
10.4 Poisson random measure . . . 116
10.5 Integration w.r.t Measure . . . 118
10.6 Kernel . . . 124

10.7 Randomization of a Point Process . . . 125
10.7.1 $M/G/\infty$ . . . 130
10.8 Transformation of random measures . . . 131
10.9 The Distribution of a Point Process . . . 134
10.10 Stationary Random measure . . . 134
10.11 Invariant Sets and the ergodic theorem for random measures . . . 136
10.12 Stochastic process . . . 138
10.13 Jointly Stationary Random Measures and Stochastic Processes . . . 139
10.13.1 Event and Time Averges . . . 139
10.13.2 Joint processes . . . 139
10.14 Palm distribution . . . 141

11 Stochastic intensity **147**
11.1 Hawkes Process . . . 150
11.1.1 Probability generating functional . . . 153

12 Martingales **159**
12.1 Signed Measure (Royden) . . . 159
12.2 Radon Nikodym Derivative . . . 161
12.3 Conditional Expectation . . . 165
12.4 Properties of conditional expectation . . . 171
12.5 Martingale . . . 176
12.5.1 Examples of Martingales . . . 179
12.6 Stopping time . . . 189
12.7 Positive Supermartingale . . . 192
12.7.1 Operations on Supermartingale . . . 193
12.7.2 Upcrossings . . . 195
12.7.3 Boundedness Properties . . . 196
12.7.4 Convergence of Positive Supermartingale . . . 197
12.7.5 Closure . . . 200
12.8 Stopping Supermartingales . . . 204
12.8.1 Gambler's Ruin . . . 205
12.9 Martingale and Submartingale Convergence . . . 206
12.9.1 Krickeberg Decomposition . . . 206
12.9.2 Doob's (Sub)martingale Convergence Theorem . . . 207
12.10 Regularity and Closure . . . 208
12.11 Regularity and Stopping . . . 209
12.12 Stopping Theorems . . . 210
12.13 Wald's identity and random walks . . . 215
12.13.1 The basic Martingales . . . 216
12.13.2 Regular stopping times . . . 217
12.13.3 Examples of integrable stopping times . . . 221
12.13.4 The simple random walk . . . 222
12.14 Reversed martingales . . . 224

Chapter 1

Sets and Events

Let Ω be a set and $\mathbb{N} = \{1, 2, \ldots, \}$.

1.1 Indicator functions

Definition 1.1. Let $A \subseteq \Omega$.

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{otherwise.} \end{cases}$$

Fact 1.2. The following hold:

(a) $\mathbb{1}_A \leqslant \mathbb{1}_B$ if and only if $A \subseteq B$.

(b) $\mathbb{1}_{A^c} = 1 - \mathbb{1}_A$.

Definition 1.3. Let $\{A_i\}_{i=1}^{\infty}$ be a sequence of subsets of Ω. Define

(a)

$$\inf_{k \geqslant n} A_k := \bigcap_{k=n}^{\infty} A_k, \qquad \sup_{k \geqslant n} A_k := \bigcup_{k=n}^{\infty} A_k,$$

and

(b)

$$\liminf_{n \to \infty} A_n = \bigcup_{k=1}^{\infty} \bigcap_{k=n}^{\infty} A_k, \qquad \limsup_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Proposition 1.4. By de Morgan's law,

$$(\liminf_{n \to \infty} A_n)^c = \limsup_{n \to \infty} A_n^c,$$

$$(\limsup_{n \to \infty} A_n)^c = \liminf_{n \to \infty} A_n^c.$$

Definition 1.5. For some sequence $\{B_n\}$ of subsets of Ω, if $\limsup_{n\to\infty} B_n = B = \liminf_{n\to\infty} B_n$, then the *limit* of $\{B_n\}$ exists, written as

$$\limsup_{n\to\infty} = B.$$

Lemma 1.6. Let $\{A_n\}$ be a sequence of subsets of Ω.

(a)

$$\limsup_{n\to\infty} A_n = \left\{\omega \,\middle|\, \sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega) = \infty\right\} = \{\omega \mid \omega \in A_{n_k}, k = 1, 2, \cdots\},$$

for some subsequence n_k depending on ω. Consequencely, we write $\limsup_{n\to\infty} A_n = \{A_n \text{ i.o.}\}$.

(b)

$$\liminf_{n\to\infty} A_n = \left\{\omega \,\middle|\, \sum_{n=1}^{\infty} \mathbb{1}_{A_n^c}(\omega) < \infty\right\} = \{\omega \mid \omega \in A_n, \forall n \geqslant n_0(\omega)\}.$$

Proof. (a) If $\omega \in \limsup_{n\to\infty} A_n$, then for any $n \in \mathbb{N}$, $\omega \in \bigcup_{k=n}^{\infty} A_k$. So for any $n \in \mathbb{N}$, there exists $k_n \geqslant n$ such that $\omega \in A_{k_n}$, and therefore $\sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega) \geqslant \sum_n \mathbb{1}_{A_{k_n}}(\omega) = \infty$, which implies $\omega \in \left\{\omega \,\middle|\, \sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega) = \infty\right\}$. Thus, $\lim_{n\to\infty} \sup A_n \subseteq \left\{\omega \,\middle|\, \sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega) = \infty\right\}$. Conversely, if $\omega \in \left\{\omega \,\middle|\, \sum_{n=1}^{\infty} \mathbb{1}(A_n)(\omega) = \infty\right\}$, then there exists $k_n \to \infty$ such that $\omega \in A_{k_n}$, and therefore for any $n \in \mathbb{N}$, $\omega \in \bigcup_{k=n}^{\infty} A_k$ so that $\omega \in \lim_{n\to\infty} \sup A_n$. Thus, $\left\{\omega \,\middle|\, \sum_{n=1}^{\infty} \mathbb{1}(A_n)(\omega) = \infty\right\} \subseteq \lim_{n\to\infty} \sup A_n$.

(b) It is similar. □

Example 1.7.

$$\liminf_{n\to\infty} \left[0, \frac{n}{n+1}\right) = [0, 1) = \limsup_{n\to\infty} \left[0, \frac{n}{n+1}\right).$$

Definition 1.8. We say $X_n \to X_0$ *almost surely* (a.s.) (a.e.) if $P\left(\lim_{n\to\infty} X_n = X_0\right) = 1$.

Theorem 1.9 (Almost surely convergence)**.** *Let $\{X_n\}$ be a set of measurable functions from Ω to $\mathbb{R}$. If $X_n \to X_0$ a.e., then*

$$\begin{aligned}
1 &= P\left(\lim_{n\to\infty} X_n = X_0\right) = P\left(\omega \,\middle|\, |X_n(\omega) - X_0(\omega)| \leqslant \epsilon, \forall n \geqslant n_0(\omega)\right) \\
&= P\left(\omega \mid \omega \in \{|X_n - X_0| \leqslant \epsilon\}, \forall n \geqslant n_0(\omega)\right) = P\left(\omega \mid \omega \in A_n, \forall n \geqslant n_0(\omega)\right) \\
&= P(\liminf_{n\to\infty} A_n) = P\left(\liminf_{n\to\infty} \{|X_n - X_0| \leqslant \epsilon\}\right), \forall \epsilon > 0,
\end{aligned}$$

and so

$$0 = P\left(\limsup_{n\to\infty} A_n^c\right) = P\left(\limsup_{n\to\infty} \{|X_n - X_0| > \epsilon\}\right), \forall \epsilon > 0.$$

Proposition 1.10. Suppose $\{A_n\}$ is a monotone sequence of subsets.

(a) If $A_n \uparrow$, then $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$.

(b) If $A_n \downarrow$, then $\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$.

Proof. (a) If follows from

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k\geqslant n} A_k \subseteq \bigcup_{k\geqslant 1} A_k = \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k\geqslant n} A_k = \liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n. \qquad \square$$

Remark. Recall that the limit of monotone increasing sequence is its supremum.

Remark. Since $\{\bigcap_{k=n}^{\infty} A_k\}_{n\geqslant 1}$ is monotone increasing,

$$\liminf_{n\to\infty} A_n := \bigcup_{k=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \lim_{n\to\infty} \bigcap_{k=n}^{\infty} A_k.$$

Since $\{\bigcup_{k=n}^{\infty} A_k\}_{n\geqslant 1}$ is monotone decreasing,

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \lim_{n\to\infty} \bigcup_{k=n}^{\infty} A_k.$$

Proposition 1.11. We have the following.

(a)

$$\mathbb{1}_{\inf_{n\geqslant k} A_n} = \inf_{n\geqslant k} \mathbb{1}_{A_n}, \qquad \mathbb{1}_{\sup_{n\geqslant k} A_n} = \sup_{n\geqslant k} \mathbb{1}_{A_n}.$$

(b)

$$\mathbb{1}_{\limsup_{n\to\infty} A_n} = \limsup_{n\to\infty} \mathbb{1}_{A_n}, \qquad \mathbb{1}_{\liminf_{n\to\infty} A_n} = \liminf_{n\to\infty} \mathbb{1}_{A_n}.$$

(c)

$$\mathbb{1}_{A\Delta B} \equiv \mathbb{1}_A + \mathbb{1}_B \pmod 2.$$

Proof. (a) $\mathbb{1}_{\inf_{n\geqslant k} A_n} = 1$ if and only if $\omega \in \inf_{n\geqslant k} A_n = \cap_{n=k}^{\infty} A_n$ if and only if $\omega \in A_n$ for all $n \geqslant k$ if and only if $\mathbb{1}_{A_n}(\omega) = 1$ for all $n \geqslant k$ if and only if $\inf_{n\geqslant k} \mathbb{1}_{A_n}(\omega) = 1$.

(b)

$$\mathbb{1}_{\limsup_{n\to\infty} A_n} = \mathbb{1}_{\inf_{n\geqslant 1}\sup_{k\geqslant n} A_k} = \inf_{n\geqslant 1} \mathbb{1}_{\sup_{k\geqslant n} A_k} = \inf_{n\geqslant 1}\sup_{k\geqslant n} \mathbb{1}_{A_k} = \limsup_{n\to\infty} \mathbb{1}_{A_n}. \qquad \square$$

Chapter 2

Probability Spaces

Definition 2.1. Suppose $\Omega = \mathbb{R}$ and let

$$\mathcal{C} = \{(a, b] \mid -\infty \leqslant a \leqslant b < \infty\}.$$

Define

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

and call $\mathcal{B}(\mathbb{R})$ the *Borel* subsets of $\mathbb{R}$.

2.1 Basic Definitions and Properties

s

Definition 2.2. A *probability space* is a triple $(\Omega, \mathcal{B}, P)$ where

(a) Ω is the sample space corresponding to outcomes of some experiment.

(b) $\mathcal{B}$ is the σ-algebra of subsets of Ω. These subsets are called events.

(c) P is a probability measure, i.e., P is a function $P : \mathcal{B} \to [0, 1]$ such that

(1) $P(A) \geqslant 0$ for all $A \in \mathcal{B}$.
(2) P is σ-additive.
(3) $P(\Omega) = 1$.

Theorem 2.3 (inclusion-exclusion formula)**.**

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \left(\sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} P\left(A_{i_1} \cap \cdots \cap A_{i_k}\right) \right) = \sum_{\emptyset \neq J \subseteq \{1,2,\ldots,n\}} (-1)^{|J|-1} P\left(\bigcap_{j \in J} A_j\right).$$

Theorem 2.4 (More continuity)**.** *We have the following.*

(a)

$$P\left(\lim_{n\to\infty} \inf A_n\right) \leqslant \lim_{n\to\infty} \inf P(A_n) \leqslant \lim_{n\to\infty} \sup P\left(A_n\right) \leqslant P\left(\lim_{n\to\infty} \sup A_n\right).$$

(b) P is continuous.

Proof. (a) The first inequality follows from Fatou's Lemma:

$$P(\liminf_{n\to\infty} A_n) = P(\lim_{n\to\infty} \bigcap_{k\geqslant n} A_k) = \lim_{n\to\infty} P(\bigcap_{k\geqslant n} A_k) = \liminf_{n\to\infty} P(\bigcap_{k\geqslant n} A_k) \leqslant \liminf_{n\to\infty} P(A_n),$$

where the second equality follows from the monotone continuity property. The last inequality follows from Proposition 1.4 and the first inequality.

(b) Let $A_n \to A$ as $n \to \infty$. Then $\lim_{n\to\infty} \sup A_n = \lim_{n\to\infty} \inf A_n$. So by (a), $P(A) = P\left(\lim_{n\to\infty} \inf A_n\right) \leqslant \lim_{n\to\infty} \inf P(A_n) \leqslant \lim_{n\to\infty} \sup P\left(A_n\right) \leqslant P\left(\lim_{n\to\infty} \sup A_n\right) = P(A)$. □

Definition 2.5. A function $F : \mathbb{R} \to [0, 1]$ satisfying

(a) F is right continuous,

(b) F is monotone nondecreasing,

(c) F has limit at $\pm\infty$: $F(\infty) := \lim_{x\uparrow\infty} F(x) = 1$ and $F(-\infty) := \lim_{x\downarrow-\infty} F(x) = 0$,

is called a *(probability) distribution* function.

Proposition 2.6. Define $F : \mathbb{R} \to [0, 1]$ by $F(x) := \mu\left((-\infty, x]\right)$, where μ is a finite Borel measure on $\mathbb{R}$. Then F is a distribution function.

Proof. (a) Let $x_n \downarrow x$, then $(-\infty, x_n] \downarrow (-\infty, x]$. By the (right) continuity of probability measure, $\mu\left((-\infty, x_n]\right) = \int_{\mathbb{R}} \mathbb{1}_{(-\infty,x_n]} d\mu \downarrow \int_{\mathbb{R}} \mathbb{1}_{(-\infty,x]} d\mu = \mu\left((-\infty, x]\right)$. So $F(x_n) \downarrow F(x)$.

(b) If $x, y \in \mathbb{R}$ and $x < y$, then $(-\infty, x] \subseteq (-\infty, y]$. So by the monotonicity of μ, $F(x) = \mu\left((-\infty, x]\right) \leqslant \mu\left((-\infty, y]\right) = F(y)$.

(c) Let $x_n \uparrow \infty$. Since μ is continuous, $F(\infty) = \lim_{x_n\uparrow\infty} F(x_n) = \lim_{x_n\uparrow\infty} \mu\left((-\infty, x_n]\right) = \mu\left(\lim_{x_n\uparrow\infty}(-\infty, x_n]\right) = \mu\left(\bigcup_{n=1}^{\infty}(-\infty, x_n]\right) = \mu\left((-\infty, \infty)\right) = \mu(\mathbb{R}) = \mu(\Omega) = 1$. Likewise, let $x_n \downarrow -\infty$, $F(-\infty) = \lim_{x_n\downarrow-\infty} F(x_n) = \lim_{x_n\downarrow-\infty} \mu\left((-\infty, x_n]\right) = \mu\left(\lim_{x_n\downarrow-\infty}(-\infty, x_n]\right) = \mu\left(\bigcap_{n=1}^{\infty}(-\infty, x_n]\right) = \mu(\emptyset) = 0$. □

Remark. In practice, we start with a known distribution function $F : \mathbb{R} \to [0, 1]$ and wish to construct a probability space $(\Omega, \mathcal{B}, P)$ such that $F(x) = P\left((-\infty, x]\right)$.

Example 2.7 (Coincidences)**.** Suppose the integers $1, 2, \ldots, n$ are randomly permuted. What is the probability that there is an integer left unchanged by the permutation?

We first construct a probability space. Let Ω be the set of all permutations of $1, 2, \ldots, n$ so that

$$\Omega = \{(x_1, \ldots, x_n) \mid x_1 \cup \cdots \cup x_n = \{1, \ldots, n\}\}.$$

Thus Ω is the set of outcomes from the experiment of sampling n times without replacement from the population $1, \ldots, n$. We let $\mathcal{B} = \mathcal{P}(\Omega)$ be the power set of Ω and define $P((x_1, \ldots, x_n)) = \frac{1}{n!}$ for $(x_1, \ldots, x_n) \in \Omega$, and $P(B) = \frac{1}{n!}|B|$ for $B \in \mathcal{B}$. For $i = 1, \ldots, n$, let A_i be the set of all elements of Ω with i in the i^{th} spot. From Theorem 2.3, we have

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1\leqslant i<j\leqslant n} P(A_i A_j) + \sum_{1\leqslant i<j<k\leqslant n} P(A_i A_j A_k) - \cdots + (-1)^{n+1} P(A_1 \cdots A_n).$$

To compute $P(A_i)$, we fix an integer i in the i^{th} spot and count the number of ways to distribute $n-1$ objects in $n-1$ spots, which is $(n-1)!$ and then divide by $n!$. To compute $P(A_iA_j)$, we fix i and j and count the number of ways to distribute $n-2$ integers into $n-2$ spots, and so on. Thus,

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &= n\frac{(n-1)!}{n!} - \binom{n}{2}\frac{(n-2)!}{n!} + \binom{n}{3}\frac{(n-3)!}{n!} - \cdots + (-1)^{n+1}\frac{1}{n!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^n\frac{1}{n!}. \end{aligned}$$

Taking into account the expansion of e^x for $x=-1$, we see that for large n, the probability of a conincidence is approximately $P\left(\bigcup_{i=1}^{n} A_i\right) \approx 1 - e^{-1} \approx 0.632$.

2.2 More on Closure

Definition 2.8. $\mathcal{P}$ is a *π-system* if it is closed under finite intersections.

Definition 2.9. A class subsets $\mathcal{L}$ of Ω is called a *λ-system (σ-additive class, Dynkin class)* if it satisfies

old (a) $\Omega \in \mathcal{L}$,

(b) if $A \subseteq B$ with $A, B \in \mathcal{L}$, then $B \setminus A \in \mathcal{L}$,

(c) for increasing $\{A_n\} \subseteq \mathcal{L}$: $\bigcup_{n=1}^{\infty} A_n \in \mathcal{L}$;

new (a) $\Omega \in \mathcal{L}$,

(b) if $A \in \mathcal{L}$, then $A^c \in \mathcal{L}$,

(c) for pairwise disjoint $\{A_n\} \subseteq \mathcal{L}$: $\bigcup_{n=1}^{\infty} A_n \in \mathcal{L}$.

Definition 2.10. The *minimal structure* $\mathcal{S}$ generated by a class $\mathcal{C}$ is a non-empty structure satisfying

(a) $\mathcal{C} \subseteq \mathcal{S}$,

(b) If S' is some other structure containing $\mathcal{C}$, then $\mathcal{S} \subseteq \mathcal{S}'$.

Denote the minimal structure by $\mathcal{S}(\mathcal{C})$.

Proposition 2.11. The minimal structure $\mathcal{S}$ generated by a class $\mathcal{C}$ exists and is unique. Let $\mathcal{N} = \{\mathcal{G} \mid \mathcal{G} \text{ is a strucure}, \mathcal{C} \subseteq G\}$, then $\mathcal{S}(\mathcal{C}) = \bigcap_{\mathcal{G}\in\mathcal{N}} \mathcal{G}$.

Theorem 2.12 (Dynkin' theorem)**.** *(a) If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system such that $\mathcal{P} \subseteq \mathcal{L}$, then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.*

(b) If $\mathcal{P}$ is a π-system, then $\sigma(P) = \mathcal{L}(\mathcal{P})$, that is, the minimal σ-algebra over $\mathcal{P}$ equals the minimal λ-system over $\mathcal{P}$.

Proposition 2.13. Let P_1, P_2 be two probability measure on $(\Omega, \mathcal{B})$. $\{A \in \mathcal{B} \mid P_1(A) = P_2(A)\}$ is a λ-system.

Corollary 2.14. If P_1, P_2 are two probability measures in $(\Omega, \mathcal{B})$ and if $\mathcal{P}$ is a π-system such that for $A \in \mathcal{P} : P_1(A) = P_2(A)$, then for $B \in \sigma(\mathcal{P}) : P_1(B) = P_2(B)$.

Proof. $\mathcal{L} := \{A \in \mathcal{B} \mid P_1(A) = P_2(A)\}$ is a λ-system. But $\mathcal{P} \subseteq \mathcal{L}$, and hence by Dynkin's theorem $\sigma(\mathcal{P}) \subseteq \mathcal{L}$. □

Corollary 2.15. Let $\Omega = \mathbb{R}$. Let P_1, P_2 be two probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that their distribution functions are equal: for all $x \in \mathbb{R} : F_1(x) = P_1((-\infty, x]) = F_2(x) = P_2((-\infty, x])$. Then $P_1 \equiv P_2$ on $\mathcal{B}(\mathbb{R})$. So a probability measure on $\mathbb{R}$ is uniquely determined by its distribution function.

Proof. Let $\mathcal{P} = \{(-\infty, x] : x \in \mathbb{R}\}$. Then $\mathcal{P}$ is a π-system since $(-\infty, x] \cap (-\infty, y] = (-\infty, x \wedge y] \in \mathcal{P}$. So $F_1(x) = F_2(x), x \in \mathbb{R}$ means $P_1 \equiv P_2$ on $\mathcal{P}$. Furthermore $\sigma(\mathcal{P}) = \mathcal{B}(\mathbb{R})$. Thus, by CorollAry 2.14, $P_1 \equiv P_2$ on $\sigma(\mathcal{P}) = \mathcal{B}(\mathbb{R})$. □

Proposition 2.16. If a class $\mathcal{C}$ is both a π-system and a λ-system, then it is a σ-algebra.

2.3 Two Constructions

(a) Discrete models: Suppose $\Omega = \{\omega_1, \omega_2, \cdots\}$ is countable. For each $i \geqslant$, associate to ω_i the number p_i such that $p_i \geqslant 0$ and $\sum_{i=1}^{\infty} p_i = 1$. Define $\mathcal{B} = \mathcal{P}(\Omega)$, and for $A \in \mathcal{B}$, set $P(A) = \sum_{\omega_i \in A} p_i$. Then we have the following properties of P:

(1) $P(A) \geqslant 0, \forall A \in \mathcal{B}$.

(2) $P(\Omega) = \sum_{i=1}^{\infty} p_i = 1$.

(3) If $\{A_j\}_{j \geqslant 1}$ are mutually disjoint subsets, then

$$P\left(\bigsqcup_{j=1}^{\infty} A_j\right) = \sum_{\omega_i \in \bigsqcup_{j=1}^{\infty} A_j} p_i = \sum_{j=1}^{\infty} \sum_{\omega_i \in A_j} p_i = \sum_{j=1}^{\infty} P(A_j),$$

where the last step is justified because the series, being positive, can be added in any order.

This gives the general construction of probabilities when Ω is countable.

(b) Coin tossing N times: Set

$$\Omega = \{0,1\}^N = \{(\omega_1, \ldots, \omega_N) : \omega_i = 0 \text{ or } 1\}.$$

For $p \geqslant 0$, $q \geqslant 0$, $p + q = 1$, define $p_{(\omega_1, \ldots, \omega_N)} = p^{\sum_{j=1}^N \omega_j} q^{N - \sum_{j=1}^N \omega_j} = p^{\#1's} q^{\#0's}$. Let $\mathcal{B} = \mathcal{P}(\Omega)$ and for $A \subseteq \Omega$, define $P(A) = \sum_{\omega \in A} p_\omega$. As in (a), this gives a probability model provided $\sum_{\omega \in \Omega} p_\omega = 1$. Note the product form $p_{(\omega_1, \ldots, \omega_N)} = \prod_{i=1}^N p^{\omega_i} q^{1-\omega_i}$. So

$$\sum_i p_i = \sum_{\omega_1, \ldots, \omega_N} \prod_{i=1}^{N} p^{\omega_i} q^{1-\omega_i} = \sum_{\omega_1, \ldots, \omega_{N-1}} \prod_{i=1}^{N-1} p^{\omega_i} q^{1-\omega_i} \left(p^1 q^0 + p^0 q^1\right) = \cdots = 1.$$

2.4 Construction of Probability Spaces

Lemma 2.17 (The algebra generated by a semialgebra)**.** Suppose $\mathcal{S}$ is a semialgebra of subsets of Ω. Let $\mathcal{A}(\mathcal{S})$ be the smallEst algebra containing $\mathcal{S}$. Then

$$\mathcal{A}(\mathcal{S}) = \left\{ \bigsqcup_{i}^{\text{finite}} S_i \;\middle|\; \{S_i\} \subseteq \mathcal{S} \text{ are disjoint} \right\}.$$

Theorem 2.18 (First Extension Theorem)**.** *Suppose $\mathcal{S}$ is a semialgebra of subsets of Ω and $P : \mathcal{S} \to [0,1]$ is a σ-additive on $\mathcal{S}$ and satisfies $P(\Omega) = 1$, implying P is a probability measure on S. There is a unqiue extension P' of P to $\mathcal{A}(\mathcal{S})$, defined by $P'(\bigsqcup_{i\in I} S_i) = \sum_{i\in I} P(S_i)$, which is a probability measure on $\mathcal{A}(\mathcal{S})$; that is $P'(\Omega) = 1$ and P' is σ-additive on $\mathcal{A}(\mathcal{S})$.*

Theorem 2.19 (Second Extension Theorem)**.** *A probability measure P defined on a algebra $\mathcal{A}$ of subsets has a unique extension to a probility measure on $\sigma(\mathcal{A})$, the σ-algebra generated by $\mathcal{A}$.*

Theorem 2.20 (Combo Extension Theorem)**.** *Suppose $\mathcal{S}$ is a semialgebra of subsets of Ω and that P is a σ-addtive set function mapping $\mathcal{S}$ into $[0,1]$ such that $P(\Omega) = 1$. There is a unique probability measure on $\sigma(\mathcal{S})$ that extends P.*

2.5 Measure Constructon

2.5.1 Lebesgue Measure on $(0,1]$

Suppose $\Omega = (0,1]$, $\mathcal{B} = \mathcal{B}((0,1])$ and $\mathcal{S} = \{(a,b] \mid 0 \leqslant a \leqslant b \leqslant 1\}$. Define on $\mathcal{S}$ the function $\lambda : \mathcal{S} \longrightarrow [0,1]$ by $\lambda(\emptyset) = 0$, $\lambda(a,b] = b - a$. With a view to applying Extension Theorem, note that $\lambda(A) \geqslant 0$. To show that λ has unique extension we need to show that λ is σ-additive.

2.5.2 Construction of a Probability Measure on $\mathbb{R}$ with Given Distribution Function $F(x)$.

Given Lebesgue measure λ constructed in last section and a distribution function $F(x)$, we construct a probability measure P_F on $\mathbb{R}$ such that $P_F((-\infty, x]) = F(x)$. Define the *left continuous inverse* of F as $F^{\leftarrow}(y) = \inf\{s : F(s) \geqslant y\}, 0 < y < 1$. Now define for $A \subseteq \mathbb{R}$, $\xi_F(A) = \{x \in (0,1] : F^{\leftarrow}(x) \in A\}$. If A is a Borel subset of $\mathbb{R}$, then $\xi_F(A)$ is a Borel subset of $(0,1]$.

Lemma 2.21. If $A \in \mathcal{B}(\mathbb{R})$, then $\xi_F(A) \in \mathcal{B}((0,1])$.

Proof. Define $\mathcal{G} = \{A \subseteq R : \xi_F(A) \in \mathcal{B}((0,1])\}$. $\mathcal{G}$ contains finite intervals of the form $(a,b] \subseteq \mathbb{R}$ since

$$\begin{aligned} \xi_F((a,b]) &= \{x \in (0,1] : F^{\leftarrow}(x) \in (a,b]\} = \{x \in (0,1] : a < F^{\leftarrow}(x) \leqslant b\} \\ &= \{x \in (0,1] : F(a) < x \leqslant F(b)\} = (F(a), F(b)] \in B((0,1]). \end{aligned}$$

Next, we verify that $\mathcal{G}$ is a σ-field.

(a) Since $\xi_F(\mathbb{R}) = (0,1]$, we have $\mathbb{R} \in \mathcal{G}$.

(b) Assume $A \in \mathcal{G}$, then $A^c \subseteq \mathbb{R}$ and then

$$\xi_F(A^c) = \{x \in (0,1] : F^{\leftarrow}(x) \in A^c\} = \{x \in (0,1] : F^{\leftarrow}(x) \in A\}^c = (\xi_F(A))^c \in B\left((0,1]\right),$$

since $\xi_F(A) \in B\left((0,1]\right)$.

(c) Let $\{A_n\} \subseteq \mathcal{G}$, then $\xi_F\left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty} \xi_F(A_n) \in B\left((0,1]\right)$.

So $\mathcal{G}$ contains intervals and $\mathcal{G}$ is a σ-field and therefore $\mathcal{B}(\mathbb{R}) \supseteq \mathcal{G}$. □

Define $P_F(A) = \lambda\left(\xi_F(A)\right)$, where λ is Lebesgue measure on $(0,1]$. It is easy to check that P_F is a probability measure. Note that

$$\begin{aligned} P_F\left((-\infty, x]\right) &= \lambda\left(\xi_F\left((-\infty, x]\right)\right) = \lambda\left\{y \in (0,1] : F^{\leftarrow}(y) \leqslant x\right\} \\ &= \lambda\left\{y \in (0,1] : y \leqslant F(x)\right\} = \lambda\left((0, F(x)]\right) = F(x). \end{aligned}$$

Chapter 3

Random Variables, Elements, and Measurable Maps

3.1 Inverse Maps

Random variables are conveninet tools that allow us to focus on properties of interest about experiment being modelled. Suppose Ω and Ω' are two sets. Suppose $X : \Omega \longrightarrow \Omega'$. Then X determines a function $X^{-1} : \Omega' \longrightarrow \Omega$ defined by $X^{-1}(A') = \{\omega \in \Omega : X(\omega) \in A'\}$ for $A' \subseteq \Omega'$.

Remark. X^{-1} preserves complementation, union and intersection.

Proposition 3.1. If $\mathcal{B}'$ is a σ-field of subsets of Ω', then $X^{-1}(\mathcal{B}')$ is σ-field of subsets of Ω.

Proposition 3.2. If C' is a class of subsets of Ω', then

$$X^{-1}\left(\sigma(C')\right) = \sigma\left(X^{-1}(C')\right).$$

3.2 Measurable Maps, Random Elements, Induced Probability Measures

Definition 3.3. A pair $(\Omega, \mathcal{B})$ consisting of a set Ω and a σ-field of subsets $\mathcal{B}$ is called a measurable space.

Definition 3.4. If $(\Omega, \mathcal{B})$ and $(\Omega', \mathcal{B}')$ are two measurable spaces, then a map

$$\begin{aligned} X : \Omega &\longrightarrow \Omega' \\ \omega &\longmapsto X(\omega) \end{aligned}$$

is called *measurable* if $X^{-1}(\mathcal{B}') \subseteq \mathcal{B}$. Denoted $X \in \mathcal{B}/\mathcal{B}'$ or $X : (\Omega, \mathcal{B}) \to (\Omega', \mathcal{B}')$. When $(\Omega', \mathcal{B}') = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, X is called a *random variable*.

Definition 3.5. Let $(\Omega, \mathcal{B})$ be a measurable space. A *measure* on $(\Omega, \mathcal{B})$ is a function $\mu : \mathcal{B} \to [0, \infty]$ such that

(a)

$$\mu(\emptyset) = 0.$$

(b)

$$\mu\left(\bigsqcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i),$$

where $\{A_i\} \subseteq \mathcal{B}$ are pairwise disjoint.

We call $(\Omega, \mathcal{B}, \mu)$ a measure space.

Remark. If μ is a probability measure, we just replace condition (a) as $\mu(\Omega) = 1$.

Proposition 3.6. Let $(\Omega, \mathcal{B}, P)$ be a probability (measure) space and suppose $X : (\Omega, \mathcal{B}) \longrightarrow (\Omega', \mathcal{B}')$. Define

$$\begin{aligned} P \circ X^{-1} : \mathcal{B}' &\longrightarrow \mathbb{R}^+ \\ A' &\longmapsto (P \circ X^{-1})(A') = P(X^{-1}(A')). \end{aligned}$$

Then $P \circ X^{-1}$ is a probability measure on $(\Omega', \mathcal{B}')$, called *induced* probability measure or the *distribution* of X.

Proof. (a) $(P \circ X^{-1})(A') \in \mathbb{R}^+$ for any $A' \in \mathcal{B}'$.

(b) $(P \circ X^{-1})(\Omega') = P(X^{-1}(\Omega')) = P(\Omega) = 1$.

(c) Let $\{A_n\} \subseteq \mathcal{B}'$ be pairwise disjoint, then

$$(P \circ X^{-1})\left(\bigsqcup_{n=1}^{\infty} A'_n\right) = P\left(\bigsqcup_{n=1}^{\infty} X^{-1}(A'_n)\right) = \sum_{n=1}^{\infty} P(X^{-1}(A'_n)) = \sum_{n=1}^{\infty} (P \circ X^{-1})(A'_n). \qquad \square$$

Remark. Usually we write

$$P \circ X^{-1}(A') = P(X \in A').$$

For example, if X is random variable, then

$$(P \circ X^{-1})((-\infty, x]) = P(X \leqslant x).$$

Definition 3.7. Let $(\Omega, \mathcal{B})$ be a measurable space. $X : \Omega \to [-\infty, \infty]$ is called *measurable* if $\{f \in A\} \subseteq \mathcal{B}$ for any $A \in \mathcal{B}(\mathbb{R})$.

Remark. X is Borel measurable if $\mathcal{B}$ is a Borel σ-algebra. X is Lebesgue measurable if $\mathcal{B} = \mathcal{B}(\Omega)$ on $\mathbb{R}$ is a Lebesgue σ-algebra (All Lebesgue measurable sets on $\Omega \subseteq \mathbb{R}$).

Verification that a map is measurable is sometimes made easy by decomposing the map into the composition of two (or more) maps. If each map in the composition is measurable, then the composition is measurable.

Proposition 3.8 (Composition)**.** Let $X_1 : (\Omega_1, \mathcal{B}_1) \to (\Omega_2, \mathcal{B}_2)$ and $X_2 : (\Omega_2, \mathcal{B}_2) \to (\Omega_3, \mathcal{B}_3)$ where $(\Omega_i, \mathcal{B}_i), i = 1, 2, 3$ are measurable spaces. Define

$$\begin{aligned} X_2 \circ X_1 : \Omega_1 &\longrightarrow \Omega_3 \\ \omega_1 &\longmapsto X_2(X_1(\omega_1)). \end{aligned}$$

Then $X_2 \circ X_1 \in \mathcal{B}_1/\mathcal{B}_3$.

Proof. Since for any $B_3 \subseteq \Omega_3$:

$$\begin{aligned} (X_2 \circ X_1)^{-1}(B_3) &= \{\omega_1 : X_2 \circ X_1(\omega_1) \in B_3\} = \{\omega_1 : X_1(\omega_1) \in X_2^{-1}(B_3)\} \\ &= \{\omega_1 : \omega_1 \in X_1^{-1}\left(X_2^{-1}\right)(B_3)\} = X_1^{-1}X_2^{-1}(B_3), \end{aligned}$$

we have $(X_2 \circ X_1)^{-1} = X_1^{-1}X_2^{-1}$. Let $B_3 \in \mathcal{B}_3$, then $(X_2 \circ X_1)^{-1}(B_3) = X_1^{-1}X_2^{-1}(B_3) \in \mathcal{B}_1$ since $X_2^{-1}(B_3) \in \mathcal{B}_2$. (Or we can say $(X_2 \circ X_1)^{-1}(\mathcal{B}_3) \subseteq \mathcal{B}_1$.) □

Proposition 3.9. If $X_i : (\Omega, \mathcal{B}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a random variable for $i = 1, \dots, n$, and $\phi : U(\subseteq \mathbb{R}^n) \longrightarrow \mathbb{R}$ is continuous, then

$$\begin{aligned} \phi(X_1, \dots, X_n) : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto \phi(X_1(\omega), \dots, X_n(\omega)) \end{aligned}$$

is a random variable. (Note you have to check range of $(X_1, \dots, X_n)$ is in U.)

Proof. Let

$$V := \{\phi < t\} = \{(s_1, \dots, s_n) \in \mathbb{R}^n \mid \phi(s_1, \dots, s_n) < t\}.$$

Then V is open since ϕ is continuous. By Lindelof theorem for $\mathbb{R}^n$, there exists $\{c_k\}_{k\geqslant 1}$ such that $V = \bigsqcup_{k=1}^{\infty} c_k$, where

$$c_k = \left(a_1^{(k)}, b_1^{(k)}\right) \times \dots \times \left(a_n^{(k)}, b_n^{(k)}\right), \forall k \geqslant 1,$$

are disjoint with $a_i^{(k)} \in \mathbb{R}, \forall i = 1, \dots, n, \forall k \geqslant 1$, and such that

$$\begin{aligned} \{\phi(X_1, \dots, X_n) < t\} &= \{\omega \in \Omega \mid (X_1(\omega), \dots, X_n(\omega)) \in V\} \\ &= \left\{\omega \in \Omega \,\middle|\, (X_1(\omega), \dots, X_n(\omega)) \in \bigsqcup_{k=1}^{n} c_k\right\} \\ &= \bigsqcup_{k=1}^{\infty} \{\omega \in \Omega \mid \ (X_1(\omega), \dots, X_n(\omega)) \in c_k\} \\ &= \bigsqcup_{k=1}^{\infty} \left\{\omega \in \Omega \,\middle|\, X_1(\omega) \in \left(a_1^{(k)}, b_1^{(k)}\right), \dots, X_n(\omega) \in \left(a_n^{(k)}, b_n^{(k)}\right)\right\} \\ &= \bigsqcup_{k=1}^{\infty} \bigcap_{j=1}^{n} \left\{\omega \in \Omega \mid a_i^{(k)} < X_i(x) < b_i^{(k)}\right\} \\ &\in \mathcal{B}. \end{aligned}$$

Thus, $\phi(X_1, \dots, X_n)$ is a random variable. □

3.3 σ-fields Generated by Maps

Let $X : (\Omega, \mathcal{B}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a random variable. Define

$$\sigma(X) := X^{-1}(\mathcal{B}(\mathbb{R})).$$

This is the σ-algebra generated by information about X, which is a way of isolating that information in the probability space that pertains to X. If $\mathcal{F} \subseteq \mathcal{B}$ is a sub-σ-field of $\mathcal{B}$, we say X is measurable w.r.t to $\mathcal{F}$, written $X \in \mathcal{F}$ if $\sigma(X) \subseteq \mathcal{F}$.

Example 3.10 (Extreme example)**.** Let $X(\omega) = 17$ for $\omega \in \Omega$. Then $\sigma(X) = \sigma(\emptyset, \Omega) = \{\emptyset, \Omega\}$.

Example 3.11 (Less extreme example)**.** Suppose $X = \mathbb{1}_A$ for some $A \in \mathcal{B}$ (So mesurable). Note X has range $\{0, 1\}$. Then

$$\sigma(X) = \sigma\left(X^{-1}(\{0\}), X^{-1}(\{1\})\right) = \sigma(A, A^c) = \{\emptyset, \Omega, A, A^c\}.$$

Example 3.12 (Useful example: Simple function)**.** Let $A_i := X^{-1}(a_i) = \{X = a_i\}$. Then $\{A_i, i = 1, \ldots, k\}$ partitions Ω. We may represent X as $X = \sum_{i=1}^k a_i \mathbb{1}_{A_i}$. Then

$$\sigma(X) = \sigma(A_1, \ldots, A_k) = \left\{ \bigsqcup_{i \in I} A_i : I \subseteq \{1, \ldots, k\} \right\}.$$

Theorem 3.13. *For $t \in T$, let $X_t : (\Omega, \mathcal{B}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be measurable. Then*

$$\sigma(X_t, t \in T) = \bigvee_{t \in T} \sigma(X_t),$$

which is the smallest σ-algebra containing all $\sigma(X_t)$.

In stochastic process theory, we frequently keep track of potential information that can be revealed to us by observing the evolution of a stochastic process by an increasing famility of σ-field. If $\{X_n\}_{n \geqslant 1}$ is a (discrete time) stochastic process, we may define

$$\mathcal{B}_n := \sigma(X_1, \ldots, X_n), \forall n \geqslant 1.$$

Then $\mathcal{B}_n \subseteq B_{n+1}$ and we think of $\mathcal{B}_n$ as the information potentially available at time n. This is a way of cataloguing what information is contained in the probability model. Properties of the stochastic process are sometines expressed in terms of $\{\mathcal{B}_n\}_{n \geqslant 1}$. For instance, one formualtion of the Markov property is that the conditional distribution of X_{n+1} given $\mathcal{B}_n$ is the same as the conditional distribution of X_{n+1} given X_n.

Chapter 4

Independence

The occurrence or non-occurrence of an event has no effect on our estimate of the probability that an independent event will or will not occur.

4.1 Basic Definitions

Definition 4.1 (Independence of a finite number of events)**.** The events $A_1, \ldots, A_n (n \geqslant 2)$ are *independent* if

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i),$$

for all finite $I \subseteq \{1, \ldots, n\}$. There are $\sum_{k=2}^{n} \binom{n}{k} = 2^n - n - 1$ equations. It can be reprased as follows: The events $A_1, \ldots, A_n$ are *independent* if

$$P(B_1 \cap B_2 \cdots \cap B_n) = \prod_{i=1}^{n} P(B_i),$$

where for $i = 1, \ldots, n$, B_i equals A_i or Ω.

Definition 4.2 (Independent classes)**.** Let $\mathcal{C}_i \subseteq \mathcal{B}$ for $i = 1, \ldots, n$. The classes $\mathcal{C}_i$ are *independent*, if for any choice $A_1, \ldots, A_n$ with $A_i \in \mathcal{C}_i$ for $i = 1, \ldots, n$, we have the events $A_1, \ldots, A_n$ are independent events.

Definition 4.3 (Arbitrary number of independent classes)**.** Let T be an arbitrary index set. The classes $\mathcal{C}_t, t \in T$ are *independent families* if for each finite I with $I \subseteq T$ we have $\{\mathcal{C}_t\}_{t \in I}$ is independent.

4.2 Independent Random Variables

Definition 4.4 (Independent random variables)**.** $\{X_t, t \in T\}$ is an independent family of random variables if $\{\sigma(X_t), t \in T\}$ are independent σ-fields ($\sigma(X) = X^{-1}((\mathbb{R})))$.

Theorem 4.5 (Factorization Criterion)**.** *A family of random variables $\{X_t\}_{t\in T}$, is independent if and only if for for all finite $J \subseteq T$,*

$$F_J(x_t, t \in J) = \prod_{t\in J} P(X_t \leqslant x_t),\ \forall x \in \mathbb{R}.$$

4.3 Two Examples of Independence

4.3.1 Records, Ranks, Renyi Theorem

Let $\{X_n, n \geqslant 1\}$ be iid with common continuous distribution function $F(x)$. The continuity of F implies $P(X_i = X_j) = 0$, so that if we define $[\text{Ties}] = \bigcup_{i\neq j}[X_i = X_j]$, then $P[\text{Ties}] = 0$. Call X_n a record of the sequence if $X_n > \bigvee_{i=1}^{n-1} X_i$, and define $A_n = [X_n \text{is a record}]$. A result due to Renyi says that the events $\{A_j,\ j \geqslant 1\}$ are independent and $P(A_j) = \frac{1}{j}$ for $j \geqslant 2$. This is a special case of result about *relative ranks*. Let R_n be the relative rank of X_n among $X_1, \dots, X_n$ where $R_n = \sum_{j=1}^n \mathbb{1}(X_j \geqslant X_n)$. So

$$\begin{aligned} R_n &= 1 \text{ if and only if } X_n \text{ is a record,} \\ &= 2 \text{ if and only if } X_n \text{ is the second largest of } X_1, \dots, X_n, \end{aligned}$$

and so on.

Theorem 4.6 (Renyi Theorem)**.** *Assume $\{X_n,\ n \geqslant 1\}$ are iid with common, continuous distribution function $F(x)$.*

(a) The sequence of random variables $\{R_n,\ n \geqslant 1\}$ is independent and

$$P(R_n = k) = \frac{1}{n}, \forall k = 1, \dots, n.$$

(b) The sequence of events $\{A_n,\ n \geqslant 1\}$ is independent and

$$P(A_n) = \frac{1}{n}.$$

Proof. All possible orderings have the same probability $\frac{1}{n!}$, so for example,

$$P(\omega : X_2(\omega) < X_3(\omega) < \cdots < X_n(\omega) < X_1(\omega)) = \frac{1}{n!}.$$

Each realization of $R_1, \dots, R_n$ has the same probability as a particular ordering of $X_1, \dots, X_n$. Hence

$$P(R_1 = r_1, \dots, R_n = r_n) = \frac{1}{n!},$$

for $r_i \in \{1, \dots, i\}$, $i = 1, \dots, n$ since each realization of $R_1, \dots, R_n$ uniquely determines an ordering: For example, if $n = 3$, suppose $R_1(\omega) = 1,\ R_2(\omega) = 1$, and $R_3(\omega) = 1$. This tells us that

$$X_1(\omega) < X_2(\omega) < X_3(\omega).$$

Note that

$$P(R_n = r_n) = \sum_{r_1,\dots,r_{n-1}} P(R_1 = r_1, \dots, R_{n-1} = r_{n-1}, R_n = r_n) = \sum_{r_1,\dots,r_{n-1}} \frac{1}{n!}.$$

Since r_i ranges over i values, the number of terms in the sum is $1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-1) = (n-1)!$. Thus

$$P(R_n = r_n) = \frac{(n-1)!}{n!} = \frac{1}{n}, \ n = 1, 2, \cdots .$$

Therefore $P(R_1 = r_1, \dots, R_n = r_n) = \frac{1}{n!} = P(R_1 = r_1) \cdots P(R_n = r_n)$. □

Remark. If $\{X_n, \ n \geqslant 1\}$ is iid with common continuous distribution $F(x)$, why is the probability of ties zero? We have

$$P(\text{Ties}) = P\left(\bigcup_{i \neq j} [X_i = X_j]\right).$$

and by subadditivity, this probability is bounded above by $\sum_{i \neq j} P(X_i = X_j)$. Then it suffices to show that $P(X_1 = X_2) = 0$. Note the set containment: For every n,

$$[X_1 = X_2] \subseteq \bigcup_{k=-\infty}^{\infty} \left[\frac{k-1}{2^n} < X_1, X_2 \leqslant \frac{k}{2^n}\right].$$

By monotonicity and subadditivity

$$\begin{aligned} P(X_1 = X_2) &\leqslant \sum_{k=-\infty}^{\infty} P\left(\frac{k-1}{2^n} < X_1 \leqslant \frac{k}{2^n}, \ \frac{k-1}{2^n} < X_2 \leqslant \frac{k}{2^n}\right) \\ &= \sum_{k=-\infty}^{\infty} \left(P\left(\frac{k-1}{2^n} < X_1 \leqslant \frac{k}{2^n}\right)\right)^2. \end{aligned} \tag{4.1}$$

Write $F(a, b] = F(b) - F(a)$, then (4.1) is equal to

$$\begin{aligned} \sum_{k=-\infty}^{\infty} F(\frac{k-1}{2^n}, \frac{k}{2^n}] F(\frac{k-1}{2^n}, \frac{k}{2^n}] &= \max_{-\infty<k<\infty} F(\frac{k-1}{2^n}, \frac{k}{2^n}] \sum_{k=-\infty}^{\infty} F(\frac{k-1}{2^n}, \frac{k}{2^n}] \\ &\leqslant \max_{-\infty<k<\infty} F(\frac{k-1}{2^n}, \frac{k}{2^n}] \cdot 1 = \max_{-\infty<k<\infty} F(\frac{k-1}{2^n}, \frac{k}{2^n}]. \end{aligned}$$

Since F is continuous on $\mathbb{R}$, because F is also a probability distribution, F is uniformly continuous on $\mathbb{R}$. Thus given any $\epsilon > 0$, $\exists n_0(\epsilon)$, when $\frac{k}{2^n} - \frac{k-1}{2^n} = \frac{1}{2^n}$, as $n \geqslant n_0(\epsilon)$, we have for all k,

$$F(\frac{k-1}{2^n}, \frac{k}{2^n}] = F\left(\frac{k}{2^n}\right) - F\left(\frac{k-1}{n}\right) \leqslant \epsilon,$$

Thus for any $\epsilon > 0$, $P(X_1 = X_2) \leqslant \epsilon$.

4.3.2 Dyadic Expansions of Uniform Random Numbers

Here we consider $(\Omega, \mathcal{B}, P) = ((0,1], \mathcal{B}((0,1])), \lambda)$, where λ is Lebesgue measure. We write $\omega \in (0,1]$ using its dyadic expansion

$$\omega = \sum_{n=1}^{\infty} \frac{d_n(\omega)}{2^n} = .d_1(\omega)d_2(\omega)d_3(\omega)\cdots ,$$

where each $d_n(\omega)$ is either 0 or 1.
If a number such as $\frac{1}{2}$ has two possible expansions, we agree to use the nonterminating one.

Fact 4.7. Each d_n is a r.v.. Since d_n is discrete with possible values 0,1, it suffices to check

$$[d_n = 0] \in \mathcal{B}((0,1]),\ [d_n = 1] \in \mathcal{B}((0,1]),$$

for any $n \geqslant 1$. In fact, since $[d_n = 0] = [d_n = 1]^c$, it suffices to check $[d_n = 1] \in \mathcal{B}((0,1])$. For $n = 1$,

$$[d_1 = 1] = (.1000\cdots, .1111\cdots] = (\frac{1}{2}, 1] \in \mathcal{B}((0,1]).$$

The left endpoint is open because of the convention that we take the nonterminating expansion. Note $P(d_1 = 1) = P(d_1 = 0) = \frac{1}{2}$. For $n \geqslant 2$,

$$\begin{aligned}[d_n = 1] &= \bigcup_{(u_1,\ldots,u_{n-1}) \in [0,1]^{n-1}} (.u_1u_2\cdots u_{n-1}1000\cdots, .u_1u_2\cdots u_{n-1}1111\cdots] \\ &= \text{ disjoint union of } 2^{n-1} \text{ intervals } \in \mathcal{B}((0,1]).\end{aligned}$$

For example $[d_2 = 1] = (\frac{1}{4}, \frac{1}{2}] \cup (\frac{3}{4}, 1]$.

Fact 4.8. We have

$$\begin{aligned}P(d_n = 1) &= \sum_{(u_1,\ldots,u_{n-1}) \in [0,1]^{n-1}} P(.u_1u_2\cdots u_{n-1}1000\cdots, .u_1u_2\cdots u_{n-1}1111\cdots]) \\ &= 2^{n-1} \sum_{i=n+1}^{\infty} \frac{1}{2^i} = \frac{1}{2}.\end{aligned}$$

We thus conclude that $P(d_n = 0) = P(d_n = 1) = \frac{1}{2}$.

Fact 4.9. The sequence $\{d_n,\ n \geqslant 1\}$ is iid. It is suffices to show $\{d_n\}$ is independent. For this, it suffices to pick $n \geqslant 1$ and prove $\{d_1, \ldots, d_n\}$ is independent.
For $(u-1, \ldots, u_n) \in [0,1]^n$, we have

$$\bigcap_{i=1}^{n} [d_i = u_i] = (.u_1u_2\cdots u_n000\cdots, .u_1u_2\cdots u_n111\cdots].$$

Since the probability of an interval is its length, we get

$$P\left(\bigcap_{i=1}^{n} [d_i = u_i]\right) = \sum_{i=n+1}^{\infty} \frac{1}{2^i} = \frac{1}{2^n} = \prod_{i=1}^{n} P(d_i = u_i).$$

4.4 Independence, Zero-One Laws, Borel-Cantelli Lemma

4.4.1 Borel-Cantelli Lemma

Proposition 4.10 (Borel-Cantelli Lemma)**.** Let $\{A_n\}$ be any events. If $\sum_{n=1} P(A_n) < \infty$, then $P\left(\lim_{n\to\infty} \sup A_n\right) = 0$.

Example 4.11. Suppose $\{X_n, n \geqslant 1\}$ are BernoullI random variables with $P(X_n = 1) = p_n = 1 - P(X_n = 0)$. We assert that $P(\lim_{n\to\infty} X_n = 0) = 1$, if $\sum_{n=1}^{\infty} p_n < \infty$.

Proposition 4.12 (Borel Zro-One Law)**.** If $\{A_n\}$ is a sequence of independent events, then

$$P([A_n \ i.o.]) = \begin{cases} 0, & \text{if and only if } \sum_{n=1}^{\infty} P(A_n) < \infty, \\ 1, & \text{if and only if } \sum_{n=1}^{\infty} P(A_n) = \infty. \end{cases}$$

Example 4.13 (Behavior of exponential random variables)**.** We assume that $\{E_n,\ n \geqslant 1\}$ are iid uit exponential random variales; then

$$P\Big(\lim_{n\to\infty} \sup E_n / \log_n = 1\Big) = 1.$$

That is, every often, the sequence $\{E_n\}$ spits out a large value and the growth of these large values approximately matches that of $\{\log n,\ n \geqslant 1\}$.

Proof. To prove it, we need the following simple fact: If $\{B_k\}$ are any events satisfying $P(B_k) = 1$ for $k \in \mathbb{N}$, then $P\left(\bigcap_{k=1}^{\infty} B_k\right) = 1$. For any $\omega \in \Omega$,

$$\lim_{n\to\infty} \sup \frac{E_n(\omega)}{\log n} = 1$$

means

(a) for any $\epsilon > 0$, $\frac{E_n(\omega)}{\log n} \leqslant 1 + \epsilon$, for all large n, (Otherwise, there is infinitely many n such that $\frac{E_n(\omega)}{\log n} > 1$) and

(b) for any $\epsilon > 0$, $\frac{E_n(\omega)}{\log n} > 1 - \epsilon$, for infinitely many n.

Note (a) says that for any ϵ, there is no subsequential limit bigger than $1+\epsilon$ and part (b) says that for any ϵ, there is always some subsequential limit bounded below by $1 - \epsilon$. We have the following set equality: Let $\epsilon_k \downarrow 0$ and observe

$$\left[\lim_{n\to\infty} \sup \frac{E_n}{\log n} = 1\right] = \bigcap_{k=1}^{\infty} \left\{\lim_{n\to\infty} \inf \left[\frac{E_n}{\log n} \leqslant 1 + \epsilon_k\right]\right\} \bigcap \bigcap_{k=1}^{\infty} \left\{\left[\frac{E_n}{\log n} > 1 - \epsilon_k\right] \ i.o.\right\}.$$

Then it suffices to show every braced event on the right side has prob. 1.
For fixed k,

$$\sum_{n=1}^{\infty} P\left(\frac{E_n}{\log n} > 1 - \epsilon_k\right) = \sum_{n=1}^{\infty} P(E_n > (1 - \epsilon_k) \log n) = \sum_{n=1}^{\infty} e^{-((1-\epsilon_k)\log n))} = \sum_{n=1}^{\infty} \frac{1}{n^{1-\epsilon_k}} = \infty.$$

So by the Borel Zero-One law, $P\left(\left[\frac{E_n}{\log n} > 1-\epsilon_k\right]\ i.o.\right) = 1$. Likewise,

$$\sum_{n=1}^{\infty} P\left[\frac{E_n}{\log n} > 1+\epsilon_k\right] = \sum_{n=1}^{\infty} e^{-(1+\epsilon_k)\log n} = \sum_{n=1}^{\infty} \frac{1}{n^{1+\epsilon_k}} < \infty,$$

so

$$P\left(\lim_{n\to\infty}\sup\left[\frac{E_n}{\log n} > 1+\epsilon_k\right]\right) = 1,$$

implies

$$P\left(\lim_{n\to\infty}\inf\left[\frac{E_n}{\log n} \leqslant 1+\epsilon_k\right]\right) = 1-0 = 1.$$

□

4.4.2 Kolmogorov Zero-One law

Let X_n be a sequence of random variables and define

$$\mathcal{F}'_n = \sigma\left(X_{n+1}, X_{n+2}, \cdots\right),\ n = 1,2,\ldots,$$

where F'_n is the smallest σ-algebra on Ω consisting of all events that depends only on $X_{n+1}, X_{n+2}, \cdots$. The *tail σ-algebra* $\mathcal{T}$ is defined as

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{F}'_n = \lim_{n\to\infty} \downarrow \sigma\left(X_n, X_{n+1}, \cdots\right).$$

If $A \in \mathcal{T}$, we will call A a *tail event* and a random variable measurable with respect to $\mathcal{T}$ is called a *tail random variable.*

Example 4.14. (a) Observe that

$$\left\{\omega : \sum_{n=1}^{\infty} X_n(\omega) \text{ converges}\right\} \in \mathcal{T}.$$

To see this note that, for any $m \in \mathbb{Z}^+$, the sum $\sum_{n=1}^{\infty} X_n(\omega)$ converges if and only if $\sum_{n=m}^{\infty} X_n(\omega)$ converges. So

$$\left[\sum_{n=1}^{\infty} X_n \text{ converges}\right] = \left[\sum_{n=m+1}^{\infty} X_n \text{ converges}\right] \in \mathcal{F}'_m.$$

This holds for all m and after intersecting over m.

(b) We have $\lim_{n\to\infty}\sup X_n \in \mathcal{T}$, $\lim_{n\to\infty}\inf X_n \in \mathcal{T}$, $\{\omega : \lim_{n\to\infty} X_n(\omega) \text{ exists}\} \in \mathcal{T}$. This is true since the lim sup of the sequence $\{X_1, X_2, \ldots, \cdots\}$ is the same as the lim sup of the sequence $\{X_m, X_{m+1}, \cdots\}$ for all m.

(c) Let $S_n = X_1 + \cdots + X_n$. Then

$$\left\{\omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n}\right\} \in \mathcal{T}$$

since for any m,

$$\lim_{n\to\infty} \frac{S_n(\omega)}{n} = \lim_{n\to\infty} \frac{\sum_{i=1}^n X_i(\omega)}{n} = \lim_{n\to\infty} \frac{\sum_{i=m+1}^n X_i(\omega)}{n},$$

and so for any m, $\lim_{n\to\infty} \frac{S_n(\omega)}{n} \in \mathcal{F}'_m$.

Theorem 4.15 (Kolmogorov Zero-One Law)**.** *If $\{X_n\}$ are independent random variables with tail σ-field $\mathcal{T}$, then $\Lambda \in \mathcal{T}$ implies $P(\Lambda) = 0$ or 1 so that the tail σ-field $\mathcal{T}$ is almost trivial.*

Lemma 4.16 (Almost trivial σ-fields)**.** Let $\mathcal{G}$ be an almost trivial σ-field and let X be a random variable measurable with respect to $\mathcal{G}$. Then there exists c such that $P(X = c) = 1$.

Proof. Let $F(x) = P(X \leqslant x)$. Since $\{X \leqslant x\} \in \sigma(X) \subseteq \mathcal{G}$, we have $F(x) = 0$ or $1, \forall x \in \mathbb{R}$. Let $c = \sup\{x : F(x) = 0\}$. Since F is non-decreasing, the distribution function must have a jump of size 1 at c and thus $P(X = c) = 1$. □

Corollary 4.17 (CorollAries of the Kolmogorov Zero-One Law)**.** Let $\{X_n\}$ be independent random variables. Then the following are true.

(a) The event $[\sum_{n=1}^\infty X_n$ converges] has probability 0 or 1.

(b) The random variable $\limsup_{n\to\infty} X_n$ and $\liminf_{n\to\infty} X_n$ are constant with prob. 1.

(c) The event $\{\omega : S_n(\omega)/n \to 0\}$ has probability 0 or 1.

Chapter 5

Integration and Expectation

5.1 Simple Functions

A function $X : \Omega \to \mathbb{R}$ on the probability space $(\Omega, \mathcal{B}, P)$ is *simple* if it has a finite range. Henceforth, assume that a simple function is $\mathbb{R}/\mathbb{R}(\mathbb{R})$ measurable. Such a function can always be written in the form

$$X(\omega) = \sum_{i=1}^{k} a_i \mathbb{1}_{A_i}(\omega),$$

where $a_i \in \mathbb{R}$ and $A_i \in \mathcal{B}$ and $A_1, \ldots, A_k$ are disjoint and $\bigsqcup_{i=1}^k A_i = \Omega$.
Recall

$$\mathcal{B}(X) = \mathcal{B}(A_i, i = 1, \ldots, k) = \left\{ \bigsqcup_{i \in I} A_i : I \subseteq \{1, \ldots, k\} \right\}.$$

Let $\mathcal{E}$ be the set of all simple functions on Ω. We have the following important properties of $\mathcal{E}$.

(a) $\mathcal{E}$ is vector space. This means the following two properties hold.

(1) If $X = \sum_{i=1}^k a_i \mathbb{1}_{A_i} \in \mathcal{E}$, then $\alpha X = \sum_{i=1}^k \alpha a_i \mathbb{1}_{A_i} \in \mathcal{E}$.

(2) If $X = \sum_{i=1}^k a_i \mathbb{1}_{A_i}$, and $Y = \sum_{i=1}^m b_j \mathbb{1}_{B_j}$ and $X, Y \in \mathcal{E}$, then $X + Y = \sum_{i,j} (a_i + b_j) \mathbb{1}_{A_i \cap B_j}$ and $\{A_i \cap B_j, 1 \leqslant i \leqslant k, 1 \leqslant j \leqslant m\}$ is a partition of Ω. So $X + Y \in \mathcal{E}$.

(b) If $X, Y \in \mathcal{E}$, then $XY \in \mathcal{E}$ since $XY = \sum_{ij} a_i b_j \mathbb{1}_{A_i \cap B_j}$.

(c) If $X, Y \in \mathcal{E}$, then $X \vee Y$, $X \wedge Y \in \mathcal{E}$, since, for instance, $X \vee Y = \sum_{i,j} a_i \vee b_j \mathbb{1}_{A_i B_j}$.

Theorem 5.1 (Measurability Theorem)**.** *Suppose $X(\omega) \geqslant 0, \forall \omega$. Then $X \in \mathcal{B}/\mathcal{B}(\mathbb{R})$ if and only if there exists simple functions $X_n \in \mathcal{E}$ and $0 \leqslant X_n \uparrow X$.*

Proof. $\Longleftarrow$ If $X_n \to \mathcal{E}$, then $X_n \in \mathcal{B}/\mathcal{B}(\mathbb{R})$, and if $X = \lim_{n\to\infty} \uparrow X_n$, then $X \in \mathcal{B}/\mathcal{B}(\mathbb{R})$ since taking limits preserves measurability.

$\Longrightarrow$ Suppose $0 \leqslant X \in \mathcal{B}/\mathcal{B}(\mathbb{R})$. Define

$$X_n := \sum_{k=1}^{n2^n} \left(\frac{k-1}{2^n} \right) \mathbb{1}_{\{\frac{k-1}{2^n} \leqslant X < \frac{k}{2^n}\}} + n \mathbb{1}_{\{X \geqslant n\}}.$$

Because $X \in \mathcal{B}/\mathcal{B}(\mathbb{R})$, it follows that $X_n \in \mathcal{E}$. Also $X_n \leqslant X_{n+1}$ and if $X(\omega) < \infty$, then for all large enough n

$$|X(\omega) - X_n(\omega)| \leqslant \frac{1}{2^n} \to 0.$$

If $X(\omega) = \infty$, then $X_n(\omega) = n \to \infty$. (If $M : \sup_{\omega\in\Omega}|X(\omega)| < \infty$, $\sup_{\omega\in\Omega}|X(\omega) - X_n(\omega)| \to 0$.) □

5.2 Expectation and Integration

Suppose $(\Omega, \mathcal{B}, P)$ is a probability space and $X : (\Omega, \mathcal{B}) \to \overline{\mathbb{R}},\ \mathcal{B}\left(\overline{\mathbb{R}}\right)$, where $\overline{\mathbb{R}} = [-\infty, \infty]$. We will define the expectation of X, written $E(X)$ or $\int_\Omega X dP$ or $\int_\Omega X(\omega)P(d\omega)$, as the Lebesgue-Stieltjes integral of X w.r.t. P.

5.2.1 Expectation of Simple Functions

5.2.2 Extension of the Definition

In stochastic modeling, for instance, we often deal with waiting times for an event to happen or return times to a state or set. If the event never occurs, it is natural to say the waiting time is infinite. If the process never returns to a state or set, it is natural to say the return time is infinite. Let $\mathcal{E}_+$ be the non-negative valued simple functions, and define

$$\overline{\mathcal{E}_+} = \left\{X \geqslant 0 : X : (\Omega, \mathcal{B}) \to \left(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})\right)\right\}$$

to be nonnegative, measurable functions with domain Ω. Let $X \in \overline{\mathcal{E}_+}$, if $P(X = \infty) > 0$, define $E(X) = \infty$, otherwise by Theorem 5.1, we may find $X_n \in \overline{\mathcal{E}_+}$, such that $0 \leqslant X_n \uparrow X$. We call $\{X_n\}$ the approximating sequence to X. The sequence $\{E(X_n)\}$ is nondecreasing by the monotonicity of expectations applied to $\mathcal{E}_+$. Since limits of monotone sequences always exists, we conclude that $\lim_{n\to\infty} E(X_n)$ exists. We define $E(X) := \lim_{n\to\infty} E(X_n)$. This extends expectation from $\mathcal{E}$ to $\overline{\mathcal{E}_+}$.

Proposition 5.2. E is well defined on $\overline{\mathcal{E}_+}$, since if $X_n \in \mathcal{E}_+$ and $Y_m \in \mathcal{E}_+$ and $X_n \uparrow X$, $Y_m \uparrow X$, then $\lim_{n\to\infty} E(X_n) = \lim_{n\to\infty} E(Y_m)$.

Theorem 5.3 (MCT). *If $X_n \geqslant 0$ for any $n \in \mathbb{N}$, then $E\left(\lim_{n\to\infty} \uparrow X_n\right) = \lim_{n\to\infty} \uparrow E(X_n)$.*

If $E(X^+)$ and $E(X^-)$ are both finite, call X integrable. This is the case if and only if $E\left(|X|\right) < \infty$.

If $E(X^+) < \infty$ but $E(X^+) = \infty$, then $E(X) = -\infty$.

If $E(X^+) = \infty$ but $E(X^-) < \infty$, then $E(X) = \infty$.

If $E(X^+) = \infty$ and $E(X^-) = \infty$, then $E(X)$ does not exists.

Example 5.4. Assume the pdf $f(x)$ of the r.v. X exists.
If

$$f(x) = \begin{cases} x^{-2}, & \text{if } x > 1, \\ 0, & \text{otherwise,} \end{cases}$$

then $E(X) = \infty$. On the other hand, if

$$f(x) = \begin{cases} \frac{1}{2}|x|^{-2}, & \text{if } |x| > 1, \\ 0, & \text{otherwise,} \end{cases}$$

then $E(X^+) = E(X^-) = \infty$, and $E(X)$ does not exist. The same conclusion would hold if f were the Cauchy density $f(x) = \frac{1}{\pi(1+x^2)}$ for $x \in \mathbb{R}$.

Properties of the expectation operator E.

(a) If X is integrable, then $P(X = \pm\infty) = 0$.

(b) (WLLN) Let $\{X_n,\ n \geqslant 1\}$ be iid with finite mean and variance. Suppos $E(X_n) = \mu < \infty$ and $\mathrm{Var}(X_n) = \sigma^2 < \infty$. Then for any $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \epsilon\right) = 0.$$

5.3 The Transformation Theorem and Densities

Suppose we are given two measurable spaces $(\Omega, \mathcal{B})$ and $(\Omega', \mathcal{B}')$, and $T : (\Omega, \mathcal{B}) \to (\Omega', \mathcal{B}')$ is a measurable map. P is a probability measure on $\mathcal{B}$.
Define $P' = P \circ T^{-1}$ to be the probability measure on $\mathcal{B}'$ given by $P'(A') = P\left(T^{-1}(A')\right),\ A' \in \mathcal{B}'$.

Theorem 5.5. *Suppose $X' : (\Omega', \mathcal{B}') \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a random variable with domain Ω'. (Then $X' \circ T : \Omega \to \mathbb{R}$ is also a random variable by composition.)*

(a) If $X' > 0$, then

$$\int_\Omega X'(T(\omega))P(d\omega) = \int_{\Omega'} X'(\omega')P'(d\omega'),$$

where $P' = P \circ T^{-1}$. It can also be expressed as $E(X' \circ T) = E'(X')$.

(b) We have $X' \in L^1(P')$ if and only if $X' \circ T \in L^1(P)$, in which case

$$\int_{T^{-1}(A')} X'(T(\omega))P(d\omega) = \int_{A'} X'(\omega')P'(d\omega'), \forall A' \in \mathcal{B}'.$$

Proof. Typical of many integration proofs, we proceed in a series of steps, starting with X as an indicator function, proceeding to X as a simple function and concluding with X being general.

(a) Suppose $X' = \mathbb{1}_{A'},\ A' \in \mathcal{B}'$. If $T(\omega) \in A'$, then $\omega \in T^{-1}T(\omega) \subseteq T^{-1}(A)$. If $\omega \in T^{-1}A'$, then $T(\omega) \in TT^{-1}A' \subseteq A'$. So $T(\omega) \in A'$ if and only if $\omega \in T^{-1}A'$. Thus, $X'(T(\omega)) = \mathbb{1}_{A'}(T(\omega)) = \mathbb{1}_{T^{-1}A'}(\omega)$. Then

$$\begin{aligned}\int_\Omega X'(T(\omega))P(d\omega) &= \int_\Omega \mathbb{1}'_A(T(\omega))P(d\omega) = \int_\Omega \mathbb{1}_{T^{-1}(A')}(\omega)P(d\omega)\\ &= P(T^{-1}(A')) = P'(A') = \int_{\Omega'} \mathbb{1}_{A'}(\omega)P'(d\omega').\end{aligned}$$

(b) Let X' be simple: $X' = \sum_{i=1}^k a_i'\mathbb{1}_{A_i'}$. Then

$$\begin{aligned}\int_\Omega X'(T\omega)P(d\omega) &= \int_\Omega \sum_{i=1}^k a_i'\mathbb{1}_{A_i'}(T(\omega))P(d\omega) = \sum_{i=1}^k a_i' \int_\Omega \mathbb{1}_{T^{-1}(A_i')}(\omega)P(d\omega)\\ &= \sum_{i=1}^k a_i'P(T^{-1}(A_i')) = \sum_{i=1}^k a_i'P'(A_i') = \int_\Omega \sum_{i=1}^k a_i'\mathbb{1}_{A_i'}(\omega')P'(d\omega').\end{aligned}$$

(c) Let $X' \geqslant 0$ be measurable. There exists a sequence of simple functions $\{X'_n\}$ such that $X'_n \uparrow X'$. Then it is also true that $X'_n \circ T \uparrow X' \circ T$. Then

$$\int_\Omega X'(T\omega)P(d\omega) \overset{MCT}{=} \lim_{n\to\infty} \uparrow \int_\Omega X'_n(T(\omega))P(d\omega) = \lim_{n\to\infty} \uparrow \int_{\Omega'} X'_n(\omega')P'(d\omega')$$
$$\overset{MCT}{=} \int_{\Omega'} X'_n(\omega')P'(d\omega').$$

Proof (ii) is similar and for the second we replace X' in (i) by $X'\mathbb{1}_{A'}$. □

5.3.1 Expectation is Always an Integral on $\mathbb{R}$

Recall the the distribution of X is the measure $F := P \circ X^{-1}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined by $(A \in \mathcal{B}(\mathbb{R}))$: $F(A) = P \circ X^{-1}(A) = P(X \in A)$. The d.f. of X is

$$F(x) := F((-\infty, x]) = P(X \leqslant x).$$

Corollary 5.6. (a) If X is an integrable random variable with distribution F, then

$$E(X) = \int_\Omega X dP = \int_\mathbb{R} xF(dx).$$

(b) Suppose $X : (\Omega, \mathcal{B}) \to (\mathbb{E}, \mathcal{E})$ is a random element of $\mathbb{E}$ with distribution $F = P \circ X^{-1}$ and suppose $g : (\mathbb{E}, \mathcal{E}) \to (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ is a non-negative measurable function. The expectation of $g(X)$ is

$$E[g(X)] = \int_\Omega g(X(\omega))P(d\omega) = \int_{x\in\mathbb{E}} g(x)F(dx).$$

Proof. (a) $X : (\Omega, \mathcal{B}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $X' : (\mathbb{R}, , \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $X'(x) = x$, (identity map), $T = X$, $P' = P \circ X^{-1} =: F$, and then apply the Transformation Theorem (a).

(b) Let $X' = g$. $X' : (\mathbb{E}, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. $T = X$. $P' = P \circ X^{-1} = F$. □

5.3.2 Densities

Let $\vec{X} : (\Omega, \mathcal{B}) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ be a random vector on $(\Omega, \mathcal{B}, P)$ with distribution F. We say X or F is absolutely continuous if there exists a nonnegative function $f : (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k)) \to (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ such that $F(A) = \int_A f(\vec{x})d\vec{x}$, where $d\vec{x}$ stands for Lebesgue measure.

5.4 Product Spaces, Independence, Fubini Theorem

Definition 5.7. If $A \subseteq \Omega_1 \times \Omega_2$, define

$$A_{\omega_1} = \{\omega_2 : (\omega_1, \omega_2) \in A\} \subseteq \Omega_2,$$

$$A_{\omega_2} = \{\omega_1 : (\omega_1, \omega_2) \in A\} \subseteq \Omega_1,$$

where A_{ω_1} is called the *section of* A at ω_1.

Here are some properties of set sections.

(a) If $A \subseteq \Omega_1 \times \Omega_2$, then
$$(A^c)_{\omega_1} = (A_{\omega_1})^c.$$

(b) For an index set T, let $A_\alpha \subseteq \Omega_1 \times \Omega_2, \forall \alpha \in T$. Then
$$\left(\bigcup_{\alpha \in T} A_\alpha\right)_{\omega_1} = \bigcup_{\alpha \in T} (A_\alpha)_{\omega_1},$$
$$\left(\bigcap_{\alpha \in T} A_\alpha\right)_{\omega_1} = \bigcap_{\alpha \in T} (A_\alpha)_{\omega_1}.$$

Definition 5.8. Define the *section of* the function $X : \Omega_1 \times \Omega_2 \to S$ as $X_{\omega_1}(\omega_2) = X(\omega_1, \omega_2)$. So $X_{\omega_1} : \Omega_2 \to S$. We think of ω_1 as fixed and the section is a function of varying ω_2. Call X_{ω_1} the section of X at ω_1.

Basic properties of sections of functions mapping from $\Omega_1 \times \Omega_2$ to S are the following:

(a) $(1_A)_{\omega_1} = 1_{A\omega_1}$.

(b) If $S = \mathbb{R}^k$ for some $k \in \mathbb{N}$ and if for $i = 1, 2$, we have $X_i : \Omega_1 \times \Omega_2 \to S$, then $(X_1 + X_2)_{\omega_1} = (X_1)_{\omega_1} + (X_2)_{\omega_1}$.

(c) Suppose S is a metric space for $n \in \mathbb{N}$, $X_n : \Omega_1 \times \Omega_2 \to S$ and $\lim_{n\to\infty} X_n$ exists. Then
$$\left(\lim_{n\to\infty} X_n\right)_{\omega_1} = \lim_{n\to\infty} (X_n)_{\omega_1}.$$

Definition 5.9. A rectangle is called *measurable* if it is of the form $A_1 \times A_2$ where $A_i \in \mathcal{B}_i$ for $i = 1, 2$.

Definition 5.10. The class of measurable rectangles is a semi-algebra which we call RECT.

Proof. (a) $\emptyset, \Omega \in$ RECT.

(b) RECT is a π-class: If $A_1 \times A_2$, $A_1' \times A_2' \in$ RECT, then
$$(A_1 \times A_2) \cap (A_1' \times A_2') = (A_1 \cap A_1') \times (A_2 \cap A_2') \in \text{RECT}.$$

(c) RECT is closed under complementation. Suppose $A_1 \times A_2 \in$ RECT. Then
$$(\Omega_1 \times \Omega_2) \smallsetminus (A_1 \times A_2) = ((\Omega_1 \smallsetminus A_1) \times A_2) \bigsqcup (A_1 \times (\Omega_2 \smallsetminus A_2)) \bigsqcup (A_1^c \times A_2^c).$$

□

Definition 5.11 (product σ-field)**.**
$$\mathcal{B}_1 \times \mathcal{B}_2 := \sigma(\text{RECT}).$$

Remark. Another way to generate the product σ-field on $\mathbb{R}^2$ is

$$\mathcal{B}_1 \times \mathcal{B}_2 = \sigma\left(\{I_1 \times I_2, I_j \in I, j = 1, 2\}\right).$$

Lemma 5.12 (Sectioning Sets)**.** Sections of measurable sets are measurable. If $A \in \mathcal{B}_1 \times \mathcal{B}_2$, then for all $\omega_1 \in \Omega_1$, $A_{\omega_1} \in \mathcal{B}_2$.

Proof. Define

$$\mathcal{C}_{\omega_1} := \{A \subseteq \Omega_1 \times \Omega_2 : A_{\omega_1} \in \mathcal{B}_2\}.$$

If $A \in$ RECT and $A = A_1 \times A_2$ where $A_1 \in \mathcal{B}_1$ and $A_2 \in \mathcal{B}_2$, then

$$A_{\omega_1} = \{\omega_2 : (\omega_1 \times \omega_2) \in A_1 \times A_2\} = \begin{cases} A_2 \in \mathcal{B}_2, & \text{if } \omega_1 \in A_1, \\ \emptyset \in \mathcal{B}_2, & \text{if } \omega_1 \notin A_1. \end{cases}$$

So RECT $\subseteq \mathcal{C}_{\omega_1}$. Check $\mathcal{C}_{\omega_1}$ is a λ-system. By Dynkin's theorem, $\mathcal{B}_1 \times \mathcal{B}_2 = \sigma(\text{RECT}) \subseteq \mathcal{C}_{\omega_1}$. Thus, $A_{\omega_1} \in \mathcal{B}_2$. □

Corollary 5.13. Sections of measurable function are measurable. That is, if

$$X : (\Omega \times \Omega_2, \mathcal{B}_1 \times \mathcal{B}_2) \to (S, \mathcal{S})$$

is $\mathcal{B}_1 \times \mathcal{B}_2/\mathcal{S}$-measurable, then $X_{\omega_1} \in \mathcal{B}_2$. Also, $X_{\omega_2} \in \mathcal{B}_1$.

Proof. Let $\Lambda \in \mathcal{S}$, then $X^{-1}(\lambda) \in \mathcal{B}_1 \times \mathcal{B}_2$. Since sections of measurable sets are measurable, $(X^{-1}(\Lambda))_{\omega_1} \in \mathcal{B}_2$. Note $X_{\omega_1} : \Omega_2 \to S$. Then

$$\begin{aligned} \mathcal{B}_2 \ni (X_{\omega_1})^{-1}(\Lambda) &= \{\omega_2 : X_{\omega_1}(\omega_2) \in \Lambda\} = \{\omega_2 : X(\omega_1, \omega_2) \in \Lambda\} \\ &= \{\omega_2 : (\omega_1, \omega_2) \in X^{-1}(\Lambda)\} = (X^{-1}(\Lambda))_{\omega_1}, \end{aligned}$$

by the def of the section of the set $X^{-1}(\Lambda)$. So $X_{\omega_1} \in \mathcal{B}_2$. □

5.5 Probability Measures on Product Spaces

Definition 5.14. Call a funtion

$$K(\omega_1, A_2) : \Omega_1 \times \mathcal{B}_2 \to [0, 1]$$

a *transition function* if

(a) for each ω_1, $K(\omega_1, \cdot)$ is a probability measure on $\mathcal{B}_2$, and

(b) for each $A_2 \in \mathcal{B}_2$, $K(\cdot, A_2)$ is $\mathcal{B}_1/B([0, 1])$ measurable.

Transition functions are used to define discrete time Markov processes where $K(\omega_1, A_2)$ represents the conditional probability that starting from ω_1, the next movement of the system results in a state in A_2.

Theorem 5.15. *Let P_1 be a probability measure on $\mathcal{B}_1$, and suppose $K : \Omega_1 \times \mathcal{B}_2 \to [0,1]$ is a transition function. Then K and P_1, uniquely determine a probability on $\mathcal{B}_1 \times \mathcal{B}_2$ via the formula*

$$P(A_1 \times A_2) = \int_{A_1} K(\omega_1, A_2) P_1(d\omega_1),$$

for all $A_2 \times A_2 \in RECT$. Let $A := A_1 \times A_2$ (genreal case since we can not always find a rectangle set $A_1 \times A_2$ and then A_2 depend on ω_1), then

$$P(A) = \int_{A_1} K(\omega_1, A_{\omega_1}) P_1(d\omega_1),$$

Proof. The measure P given above is specified on the semealgebra RECT and we need to verify that the conditions of the Combo Extension Theorem are applicibale so that P can be extended to $\sigma(\text{RECT}) = \mathcal{B}_1 \times \mathcal{B}_2$. We verify that P is σ-additive on RECT. Suppose $A_1 \times A_2 = \bigsqcup_{n=1}^{\infty} A_1^{(n)} \times A_2^{(n)}$, where $A_1^{(n)} \times A_2^{(n)} \in \text{RECT}$ for $n \in \mathbb{N}$. Then

$$A_1 = \bigsqcup_{n=1}^{\infty} A_1^{(n)} \in \mathcal{B}_1 \text{ and } A_2 = \bigsqcup_{n=1}^{\infty} A_2^{(n)} \in \mathcal{B}_2.$$

Then $A_1 \times A_2 \in \text{RECT}$. Need to show

$$P\left(\bigsqcup_{n=1}^{\infty} A_1^{(n)} \times A_2^{(n)}\right) = \sum_{n=1}^{\infty} P\left(A_1^{(n)} \times A_2^{(n)}\right).$$

Note that

$$\begin{aligned}\mathbb{1}_{A_1}(\omega_1)\mathbb{1}_{A_2}(\omega_2) = \mathbb{1}_{A_1 \times A_2}(\omega_1, \omega_2) &= \mathbb{1}_{\bigsqcup_{n=1}^{\infty} A_1^n \times A_2^{(n)}}(\omega_1, \omega_2) \\ &= \sum_{n=1}^{\infty} \mathbb{1}_{A_1^{(n)} \times A_2^{(n)}}(\omega_1, \omega_2) = \sum_{n=1}^{\infty} \mathbb{1}_{A_1^{(n)}}(\omega_1)\mathbb{1}_{A_2^{(n)}}(\omega).\end{aligned}$$

Then

$$\begin{aligned}
P(A_1 \times A_2) &= \int_{A_1} K(\omega_1, A_2) P_1(d\omega_1) = \int_{\Omega_1} \mathbb{1}_{A_1}(\omega_1) K(\omega_1, A_2) P_1(d\omega_1) \\
&= \int_{\Omega_1} \mathbb{1}_{A_1}(\omega_1) \int_{\Omega_2} \mathbb{1}_{A_2}(\omega_2) K(\omega_1, d\omega_2) P_1(d\omega_1) \\
&= \int_{\Omega_1} \int_{\Omega_2} \mathbb{1}_{A_1}(\omega_1) \mathbb{1}_{A_2}(\omega_2) K(\omega_1, d\omega_2) P_1(d\omega_1) \text{ (Riemman)} \\
&= \int_{\Omega_1} \int_{\Omega_2} \sum_{n=1}^{\infty} \mathbb{1}_{A_1^{(n)}}(\omega_1) \mathbb{1}_{A_2^{(n)}}(\omega_2) K(\omega_1, d\omega_2) P_1(d\omega_1) \\
&= \int_{\Omega_1} \sum_{i=1}^{\infty} \int_{\Omega_2} \mathbb{1}_{A_1^{(n)}}(\omega_1) \mathbb{1}_{A_2^{(n)}}(\omega_2) K(\omega_1, d\omega_2) P_1(d\omega_1) \\
&= \sum_{i=1}^{\infty} \int_{\Omega_1} \mathbb{1}_{A_1^{(n)}}(\omega_1) \left[\int_{\Omega_2} \mathbb{1}_{A_2^{(n)}}(\omega_2) K(\omega_1, d\omega_2) \right] P_1(d\omega_1) \\
&= \sum_{i=1}^{\infty} \int_{\Omega_1} \mathbb{1}_{A_1^{(n)}}(\omega_1) K(\omega_1, A_2^{(n)}) P_1(d\omega_1) \\
&= \sum_{n=1}^{\infty} P(A_1^{(n)} \times A_2^{(n)}).
\end{aligned}$$

□

Remark. Special case. Suppose for some probability measure P_2 on $\mathcal{B}_2$ that $K(\omega_1, A_2) = P_2(A_2)$. Then $P(A_1 \times A_2) = P_1(A_1)P_2(A_2)$. We denote this P by $P_1 \times P_2$ and call P *product measure*. Define σ-fields in $\Omega_1 \times \Omega_2$ by

$$\mathcal{B}_1^{\#} = \{A_1 \times \Omega_2 : A_1 \in \mathcal{B}_1\},$$

$$\mathcal{B}_2^{\#} = \{\Omega_1 \times A_2 : A_2 \in \mathcal{B}_2\}.$$

With respect to the product measure $P = P_1 \times P_2$, we have $\mathcal{B}_1^{\#} \perp\!\!\!\perp \mathcal{B}_2^{\#}$ since

$$\begin{aligned}
P(A_1 \times \Omega_2 \cap \Omega_1 \times A_2) &= P(A_1 \times A_2) = P_1(A_1)P_2(A_2) \\
&= P_1(A_1)P_2(\Omega_2)P_1(\Omega_1)P_2(A_2) \\
&= P(A_1 \times \Omega_2)P(\Omega_1 \times A_2).
\end{aligned}$$

Suppose $X_i : (\Omega_i, \mathcal{B}_i) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is random variable on Ω_i for $i = 1, 2$. Define on $\Omega_1 \times \Omega_2$ the new functions

$$X_1^{\#}(\omega_1, \omega_2) = X_1(\omega_1), \; X_2^{\#}(\omega_1, \omega_2) = X_2(\omega_2).$$

Then for any $B_1 \in \mathcal{B}(\mathbb{R})$,

$$\{X_1^{\#} \in B_1\} = \left\{(\omega_1, \omega_2) : X_1^{\#}(\omega_1, \omega_2) \in B_1\right\} = \{(\omega_1, \omega_2) : X_1(\omega_1) \in B_1\} = X_1^{-1}(B_1) \times \Omega_2.$$

Likewise,$\forall\, B_2 \in \mathcal{B}(\mathbb{R})$,

$$\left\{X_2^{\#} \in B_2\right\} = \Omega_1 \times X_2^{-1}(B_2).$$

With respect to $P = P_1 \times P_2$, the variables $X_1^{\#}$ and $X_2^{\#}$ are independent since$\forall B_1, B_2 \in \mathcal{B}(\mathbb{R})$, we have

$$\begin{aligned} P(X_1^{\#} \in B_1, X_2^{\#} \in B_2) &= P\left(X_1^{-1}(B_1) \times \Omega_2, \Omega_1 \times X_2^{-1}(B_2)\right) \\ &= P\left(\left(X_1^{-1}(B_1) \cap \Omega_1\right) \times \left(\Omega_2 \cap X_2^{-1}(B_2)\right)\right) \\ &= P\left(X_1^{-1}(B_1) \times X_2^{-1}(B_2)\right) \\ &= P_1\left(X_1^{-1}(B_1)\right) P_2\left(X_2^{-1}(B_2)\right) \\ &= P\left(X_1^{-1}(B_1) \times \Omega_2\right) P\left(\Omega_1 \times X_2^{-1}(B_2)\right) \\ &= P\left(X_1^{\#} \in B_1\right) P\left(X_2^{\#} \in B_2\right) \end{aligned}$$

The point of the remark is that independence is automatically built into the model by construction when using product measure.

5.6 Approximation theorem for measures

By the second extension theorem, the probability defined on RECT has a unique extension to a probability measure on σ(RECT).

Theorem 5.16. *Let RECT be the semialgebra on $\Omega_1 \times \Omega_1$ and let P be a probability measure on σ(RECT). Then for any $A \in \sigma$(RECT) and for any $\epsilon > 0$, there exists $n \in \mathbb{N}$ and mutually disjoint sets $A_1, \ldots, A_n \in$RECT such that*

$$P\left(A \Delta \bigsqcup_{k=1}^{n} A_k\right) < \epsilon.$$

Proof. By the definition

$$P^*(A) = \inf\left\{\sum_{E \in \mathcal{E}} P(E) : \mathcal{E} \in \mathcal{U}(A)\right\}, \forall A \subseteq \Omega_1 \times \Omega_2.$$

where the covering of A

$$\mathcal{U}(A) = \left\{\mathcal{E} \subseteq \text{RECT} \mid \mathcal{E} \text{ is at most countable and } A \subseteq \bigcup_{E \in \mathcal{E}} E\right\}.$$

Let $A \in \sigma$(RECT), $\exists A_1, A_2 \in$ RECT such that $A = \bigsqcup_{i=1}^{\infty} A_i$. Then by the defintion of P^* and since P is a measure on σ(RECT), P is countable additive: $P^*(A) = \sum_{i=1}^{\infty} P(A_i) = P(A)$. So P and P^* coincide on σ(RECT) ($\star$). Also, there exists a covering $\mathcal{E} = \bigcup_{i=1}^{\infty} B_i \subseteq$ RECT such that

$$P(A) = P^*(A) \geqslant \sum_{i=1}^{\infty} P(B_i) - \epsilon/2, \forall A \in \sigma(\text{RECT}).$$

Since P is bounded, there exists $n \in \mathbb{N}$ such that $\sum_{i=n+1}^{\infty} P(B_i) < \frac{\epsilon}{2}$. For any three sets C, D, E, we have

$$C \Delta D = (D \setminus C) U (C \setminus D) \subseteq (D \setminus C) \cup (C \setminus (DUE)) UE \subseteq (C \setminus \Delta(D \cup E)) \cup E.$$

Choose $C = A$, $D = \bigcup_{i=1}^{n} B_i$ and $E = \bigcup_{i=n+1}^{\infty} B_i$, since $A \setminus \bigcup_{i=1}^{\infty} B_i = \emptyset$,

$$P\left(A\Delta \bigcup_{i=1}^{n} B_i\right) \leqslant P\left(A\Delta \bigcup_{i=1}^{\infty} B_i\right) + P\left(\bigcup_{i=n+1}^{\infty} B_i\right) \leqslant P\left(\bigcup_{i=1}^{\infty} B_i\right) - \mu(A) + \frac{\epsilon}{2} \leqslant \epsilon.$$

Since RECT is a semialgebra, there exist $B_1', \ldots, B_n' \in$RECT such that

$$\bigsqcup_{i=1}^{n} B_i' := B_1 \sqcup \bigsqcup_{i=2}^{n} \bigcap_{j=1}^{i-1} (B_i \setminus B_j) = \bigcup_{i=1}^{n} B_i.$$ □

Theorem 5.17. *Let $A \subseteq \Omega$ and $\{A_n\}_{n=1}^{\infty} \subseteq \Omega$. Then $A_n \to A \Longleftrightarrow \mathbb{1}_{A_n} \to \mathbb{1}_A$.*

Proof. It is clear. □

5.7 Fubini's theorem

We conitnue to work on the product space $(\Omega \times \Omega_2, \mathcal{B}_1 \times \mathcal{B}_2)$.

Theorem 5.18. *Let P_1 be a probability measure on $(\Omega_1, \mathcal{B}_1)$ and suppose $K : \Omega_1 \times \mathcal{B}_2 \to [0, 1]$ is a transition kernel. Define P on $(\Omega_1 \times \Omega_2, \mathcal{B}_1 \times \mathcal{B}_2)$ by*

$$P(A_1 \times A_2) = \int_{A_1} K(\omega_1, A_2) P_1(d\omega_1), \forall A_1 \times A_2 \in \mathcal{B}_1 \times \mathcal{B}_2.$$

Assume $X : (\Omega_1 \times \Omega_2, \mathcal{B}_1 \times \mathcal{B}_2) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is $\mathcal{B}_1 \times \mathcal{B}_2/\mathcal{B}(\mathbb{R})$-measurable. Furthermore, suppose $X \geqslant 0$ (X is integrable). Then

$$Y(\omega_1) = \int_{\Omega_2} K(\omega_1, d\omega_2) X_{\omega_1}(\omega_2)$$

has the properties:

(a) Y is well defined,

(b) $Y \in \mathcal{B}_1$,

(c) $Y \geqslant 0$ ($Y \in L^1(P_1)$), and furthermore

$$\int_{\Omega_1 \times \Omega_2} X dP = \int_{\Omega_1} Y(\omega_1) P_1(d\omega_1) = \int_{\Omega_1} \left[\int_{\Omega_2} K(\omega_1, d\omega_2) X_{\omega_1}(\omega_2)\right] P_1(d\omega_1)$$

Proof. (a) Since section of random variable (measurable function) is still a random variable, we have $X_{\omega_1}(\omega_2) \in \mathcal{B}_2, \forall \omega_1 \in \Omega_1$. Also, $K(\omega_1, \cdot)$ is a probability measure on $(\Omega_2, \mathcal{B}_2)$.
So Y is well-defined.

(b) (1) Assume $X = \mathbb{1}_A$ for $A = A_1 \times A_2 \in$ RECT. Then

$$X(\omega_1, \omega_2) = \mathbb{1}_{A_1 \times A_2}(\omega_1, \omega_2) = \mathbb{1}_{A_1}(\omega_1)\mathbb{1}_{A_2}(\omega_2), \forall (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2.$$

Then

$$Y(\omega_1) = \int_{\Omega_2} \mathbb{1}_{A_1}(\omega_1)\mathbb{1}_{A_2}(\omega_2)K(\omega_1, d\omega_2) = \int_{A_2} \mathbb{1}_{A_1}(\omega_1)\ K(\omega_1, d\omega_2) = \mathbb{1}_{A_1}(\omega_1)K(\omega_1, A_2).$$

So $Y = \mathbb{1}_{A_1} \cdot K(\cdot, A_2) \in \mathcal{B}_1$.

(2) Assume $X = \mathbb{1}_A$ for $A = A_1 \times A_2 \in \mathcal{Z}_*^R$, where $\mathcal{Z}_*^R$ is the algebra of finite unions of disjoint **rectangles**. Then there exists $n \in \mathbb{N}$ and (disjoint) $A_{1,i} \times A_{2,i} \in \text{RECT}$ for $i = 1, \ldots, n$ such that

$$A = \bigsqcup_{i=1}^{n} A_{1,i} \times A_{2,i}.$$

Then since $A_{1,i} \times A_{2,i}$'s are disjoint,

$$\begin{aligned} X(\omega_1, \omega_2) &= \mathbb{1}_{\bigsqcup_{i=1}^n A_{1,i} \times A_{2,i}}(\omega_1, \omega_2) = \sum_{i=1}^{n} \mathbb{1}_{A_{1,i} \times A_{2,i}}(\omega_1, \omega_2) \\ &= \sum_{i=1}^{n} \mathbb{1}_{A_{i,1}}(\omega_1)\mathbb{1}_{A_{2,i}}(\omega_2), \forall (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2. \end{aligned}$$

Then

$$\begin{aligned} Y(\omega_1) &= \int_{\Omega_2} \sum_{i=1}^{n} \mathbb{1}_{A_1}(\omega_1)\mathbb{1}_{A_2}(\omega_2)K(\omega_1, d\omega_2) = \sum_{i=1}^{n} \int_{A_{2,i}} \mathbb{1}_{A_{1,i}}(\omega_1)K(\omega_1, d\omega_2) \\ &= \sum_{i=1}^{n} \mathbb{1}_{A_{1,i}}(\omega_1)K(\omega_1, A_{2,i}). \end{aligned}$$

So $Y = \sum_{i=1}^{n} \mathbb{1}_{A_{1,i}} K(\cdot, A_{2,i}) \in \mathcal{B}_1$.

(3) Let $X = \mathbb{1}_A$ for $A \in \mathcal{B}_1 \times \mathcal{B}_2$. By the theorem 5.16, there exists a sequence of sets

$$\{A_{1,n} \times A_{2,n}\}_{n \in \mathbb{N}} \subseteq \mathcal{Z}_*^R.$$

such that $A_{1,n} \times A_{2,n} \to A$ in P-measure. By the theorem 5.1,

$$\mathbb{1}_{A_{1,n} \times A_{2,n}} \to \mathbb{1}_A = X.$$

Then

$$X(\omega_1, \omega_2) = \lim_{n \to \infty} \mathbb{1}_{A_{1,n} \times A_{2,n}}(\omega_1, \omega_2) = \lim_{n \to \infty} \mathbb{1}_{A_{1,n}}(\omega_1)\mathbb{1}_{A_{2,n}}(\omega_2), \forall (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2.$$

Then

$$\begin{aligned} Y(\omega_1) &= \int_{\Omega_2} \lim_{n \to \infty} \mathbb{1}_{A_{1,n}}(\omega_1)\mathbb{1}_{A_{2,n}}(\omega_2)K(\omega_1, d\omega_2) = \lim_{n \to \infty} \int_{A_{2,n}} \mathbb{1}_{A_{1,n}}(\omega_1)K(\omega_1, d\omega_2) \\ &= \lim_{n \to \infty} \mathbb{1}_{A_{1,n}}(\omega_1)K(\omega_1, A_{2,n}). \end{aligned}$$

Since limits of $\mathcal{B}_1$-measureable functions are $\mathcal{B}_1$-measurable, $Y \in \mathcal{B}_1$.

(4) Since $X \geqslant 0$, there exists a sequence of (positive) simple functions $\{Z_n\}_{n\in\mathbb{N}}$ mapping from $\Omega_1 \times \Omega_2$ to $\mathbb{R}$ such that $Z_n \uparrow X$, where

$$Z_n(\omega_1, \omega_2) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbb{1}_{A_k}(s) + n\mathbb{1}_{A_{n2^n+1}}(\omega_1, \omega_2),$$

and

$$A_k = \left\{ (\omega_1, \omega_2) \in \Omega_1 \times \Omega_1 : \frac{k-1}{2^n} < X(\omega_1, \omega_2) \leqslant \frac{k}{2^n} \right\} \in \mathcal{B}_1 \times \mathcal{B}_2,$$

$$A_{n2^n+1} = \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : X(\omega_1, \omega_2) > n\} \in \mathcal{B}_1 \times \mathcal{B}_2.$$

Then by MCT,$\forall \omega_1 \in \Omega_1$,

$$\begin{aligned} Y(\omega_1) &= \int_{\Omega_2} X_{\omega_1}(\omega_2) K(\omega_1, d\omega_2) = \int_{\Omega_2} \lim_{n\to\infty} Z_n(\omega_1, \omega_2) K(\omega_1, d\omega_2) \\ &= \lim_{n\to\infty} \int_{\Omega_2} Z_n(\omega_1, \omega_2) K(\omega_1, d\omega_2) = \lim_{n\to\infty} \sum_{k=1}^{n2^n} \frac{k-1}{2^n} K(\omega_1, A_k) + nK(\omega_1, A_{n2^n+1}). \end{aligned}$$

Since

$$\frac{k-1}{2^n} K(\cdot, A_k) \in \mathcal{B}_1, \forall k = 1, \ldots, n2^n, \text{ and } nK(\cdot, A_{n2^n+1}) \in \mathcal{B}_1,$$

we have $Y \in \mathcal{B}_1$.

Alternative: First show it holds for simple functions, then by MCT, we have the conclusion for measurable functions.

(c) Define LHS $:= \int_{\Omega_1 \times \Omega_2} X dP$ and RHS $:= \int_{\Omega_1} Y(\omega_1) P(d\omega_1)$. We begin by supposing $X = \mathbb{1}_{A_1 \times A_2}$ where $A_1 \times A_2 \in$ RECT. Then LHS $:= \int_{A_1 \times A_2} dP = P(A_1 \times A_2)$ and

$$\text{RHS} = \int_{\Omega_1} \left[\int_{\Omega_2} K(\omega_1, d\omega_2) \mathbb{1}_{A_1}(\omega_1) \mathbb{1}_{A_2}(\omega_2) \right] P_1(d\omega) = \int_{A_1} K(\omega_1, A_2) P_1(d\omega_1) = P(A_1 \times A_2).$$

So LHS = RHS for any $A \in$ RECT. Define $\mathcal{C} := \{A \in \mathcal{B}_1 \times \mathcal{B}_2 : \text{LHS} = \text{RHS} \forall X = \mathbb{1}_A\}$. Note RECT $\subseteq \mathcal{C}$. We claim $\mathcal{C}$ is a λ-system.

(1) $\Omega_1 \times \Omega_2 \in \mathcal{C}$ since $\Omega_1 \times \Omega_2 \in$ RECT $\subseteq \mathcal{C}$.

(2) If $A \in \mathcal{C}$, then for $X = \mathbb{1}_{A^c}$, we have LHS $= P(A^c) = 1 - P(A)$. Since by definition,

$\int_{\Omega_1} K(\omega_1, \Omega_2) = P(\Omega_1 \times \Omega_2) = 1.$

$$\begin{aligned}
\text{LHS} &= 1 - \int_{\Omega_1}\int_{\Omega_2} K(\omega_1, d\omega_2)\mathbb{1}_{A_{\omega_1}(\omega_2)}P_1(d\omega_1) \\
&= \int_{\Omega_1}\int_{\Omega_2} K(\omega_1, d\omega_2)P_1(d\omega_1) - \int_{\Omega_1}\int_{\Omega_2} K(\omega_1, d\omega_2)\mathbb{1}_{A_{\omega_1}(\omega_2)}P_1(d\omega_1) \\
&= \int_{\Omega_1}\int_{\Omega_2} K(\omega_1, d\omega_2)\left(1 - \mathbb{1}_{A_{\omega_1}}(\omega_2)\right)P_1(d\omega_1) \\
&= \int_{\Omega_1}\int_{\Omega_2} K(\omega_1, d\omega_2)\mathbb{1}_{(A_{\omega_1})^c}(\omega_2)P_1(d\omega_2) \\
&= \int_{\Omega_1}\int_{\Omega_2} K(\omega_1, d\omega_2)\mathbb{1}_{(A^c)_{\omega_1}}(\omega_2)P_1(d\omega_2) \\
&= \text{RHS}.
\end{aligned}$$

So $A^c \in \mathcal{C}$.

(3) If $\{A_n\}_{n=1}^\infty \subseteq \mathcal{C}$ are disjoint, then

$$\begin{aligned}
\text{LHS} &= \int_{\Omega_1\times\Omega_2} \mathbb{1}_{\bigsqcup_{n=1}^\infty A_n} dP = P\left(\bigsqcup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty P(A_n) \\
&= \sum_{n=1}^\infty \int_{\Omega_1}\int_{\Omega_2} K(\omega_1, d\omega_2)\mathbb{1}_{(A_n)_{\omega_1}}(\omega_2)P_1(d\omega_1),
\end{aligned}$$

since

$$\begin{aligned}
\int_{\Omega_1}\int_{\Omega_2} K(\omega_1, d\omega_2)\mathbb{1}_{(A_n)_{\omega_1}}(\omega_2)P_1(d\omega_1) &= \int_{\Omega_1}\int_{(A_n)_{\omega_1}} K(\omega_1, d\omega_2)P_1(d\omega_1) \\
&= \int_{\Omega_1} K(\omega_1, (A_n)_{\omega_1})P_1(d\omega_1) \\
&= P(A_n).
\end{aligned}$$

Then by MCT

$$\begin{aligned}
\int_{\Omega_1\times\Omega_2} \mathbb{1}_{\bigsqcup_{n=1}^\infty A_n} dP &= \int_{\Omega_1}\int_{\Omega_2} K(\omega_1, d\omega_2)\sum_{n=1}^\infty \mathbb{1}_{(A_n)_{\omega_1}}(\omega_2)P_1(d\omega_1) \\
&= \int_{\Omega_1}\int_{\Omega_2} K(\omega_1, d\omega_2)\mathbb{1}_{(\bigsqcup_{n=1}^\infty A_n)_{\omega_1}}(\omega_2)P_1(d\omega_1) \\
&= \int_{\Omega_1}\int_{\Omega_2} K(\omega_1, d\omega_2)\mathbb{1}_{(\bigsqcup_{n=1}^\infty A_n)_{\omega_1}}(\omega_2)P_1(d\omega_1) \ (\star\ \textit{not for sets}) \\
&= \text{RHS}.
\end{aligned}$$

So $\bigsqcup_{n=1}^\infty A_n \in \mathcal{C}$.

Then $\mathcal{B}_1 \times \mathcal{B}_2 = \sigma(\text{RECT}) \subseteq \mathcal{C}$. We may conclude that for any $A \in \mathcal{B}_1 \times \mathcal{B}_2$, if $X = \mathbb{1}_A$, then LHS = RHS. Now consider the simple functions of the form

$$X = \sum_{i=1}^{k} a_i \mathbb{1}_{A_i}, \;\; A_i \in \mathcal{B}_1 \times \mathcal{B}_2, \forall i = 1, \ldots, k.$$

Then

$$\begin{aligned}
\text{LHS} &= \int_{\Omega_1 \times \Omega_2} \sum_{i=1}^{k} a_i \mathbb{1}_{A_i} dP = \sum_{i=1}^{k} a_i \int_{\Omega_1 \times \Omega_2} \mathbb{1}_{A_i} dP = \sum_{i=1}^{k} a_i P(A_i) \\
&= \sum_{i=1}^{k} a_i \int_{\Omega_1} \int_{\Omega_2} K(\omega_1, d\omega_2) \mathbb{1}_{(A_i)_{\omega_1}}(\omega_2) P_1(d\omega_1) \\
&= \int_{\Omega_1} \int_{\Omega_2} K(\omega_1, d\omega_2) \sum_{i=1}^{k} a_i \mathbb{1}_{(A_i)_{\omega_1}}(\omega_2) P_1(d\omega_1) \\
&= \int_{\Omega_1} \int_{\Omega_2} K(\omega_1, d\omega_2) \sum_{i=1}^{k} a_i (\mathbb{1}_{A_i})_{\omega_1}(\omega_2) P_1(d\omega_1) \\
&= \int_{\Omega_1} \int_{\Omega_2} K(\omega_1, d\omega_2) \left(\sum_{i=1}^{k} a_i \mathbb{1}_{A_i} \right)_{\omega_1} (\omega_2) P_1(d\omega_1) \\
&= \int_{\Omega_1} \int_{\Omega_2} K(\omega_1, d\omega_2) \left(\sum_{i=1}^{k} a_i \mathbb{1}_{A_i} \right) (\omega_1, \omega_2) P_1(d\omega_1) \\
&= \text{RHS}.
\end{aligned}$$

So LHS = RHS holds for simple functions. For arbitrary $X \geqslant 0$, there exists a sequence of simple $\{X_n\}_{n \in \mathbb{N}}$ such that $X_n \uparrow X$. Note $\text{LHS}(X_n) = \text{RHS}(X_n)$ for $n \in \mathbb{N}$. By MCT, $\text{LHS}(X_n) \uparrow \text{LHS}(X)$. Also, we get for RHS, by applying monotone convergence twice, that

$$\begin{aligned}
\text{LHS} &= \lim_{n\to\infty} \text{LHS}(X_n) = \lim_{n\to\infty} \text{RHS}(X_n) = \lim_{n\to\infty} \int_{\Omega_1} \int_{\Omega_2} K(\omega_1, d\omega_2)(X_n)_{\omega_1}(\omega_2) P_1(d\omega_1) \\
&= \int_{\Omega_1} \lim_{n\to\infty} \int_{\Omega_2} K(\omega_1, d\omega_2)(X_n)_{\omega_1}(\omega_2) P_1(d\omega_1) = \int_{\Omega_1} \int_{\Omega_2} \lim_{n\to\infty} K(\omega_1, d\omega_2)(X_n)_{\omega_1}(\omega_2) P_1(d\omega_1) \\
&= \int_{\Omega_1} \int_{\Omega_2} K(\omega_1, d\omega_2) X_{\omega_1}(\omega_2) P_1(d\omega_1) = \text{RHS}(X).
\end{aligned}$$

□

We can now give the result, called Fubino's theorem, which justifies interchange of the order of integration.

Theorem 5.19 (Fubini Theorem). *Let $P = P_1 \times P_2$ be product measure. If X is $\mathcal{B}_1 \times \mathcal{B}_2$ measurable and is either nonnegative or integrable w.r.t. P, then*

$$\int_{\Omega_1 \times \Omega_2} X dP = \int_{\Omega_1} \left[\int_{\Omega_2} X_{\omega_1}(\omega_2) P_2(d\omega_2) \right] P_1(d\omega_1) = \int_{\Omega_2} \left[\int_{\Omega_1} X_{\omega_2}(\omega_1) P_1(d\omega_1) \right] P_2(d\omega_2).$$

Proof. Let $K(\omega_1, A_2) = P_2(A_2)$. Then

$$\int_{\Omega_1\times\Omega_2} X dP = \int_{\Omega_1}\int_{\Omega_2} K(\omega_1, d\omega_2)X_{\omega_1}(\omega_2)P_1(d\omega_1) = \int_{\Omega_1}\left[\int_{\Omega_2} X_{\omega_1}(\omega_2)P_2(d\omega_2)\right] P_1(d\omega_1).$$

Similarly, let $\widetilde{K}(\omega_2, A_1) = P_1(A_1)$ be a transition function with $\widetilde{K} : \Omega_1 \times \mathcal{B}_1 \to [0, 1]$. Then

$$\begin{aligned}\int_{\Omega_1\times\Omega_2} X dP &= \int_{\Omega_2}\int_{\Omega_1} \widetilde{K}(\omega_2, d\omega_1)X_{\omega_2}(\omega_1)P_2(d\omega_2)\\ &= \int_{\Omega_2}\left[\int_{\Omega_1} X_{\omega_2}(\omega_1)P_1(d\omega_1)\right] P_2(d\omega_2).\end{aligned}$$

□

Example 5.20. Let $X_i \geqslant 0, i = 1, 2$ be two independent random variables. Then $E[X_1X_2] = E[X_1]E[X_2]$.

Proof. Define the random vector $X := (X_1, X_2)$ as

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R}\\ \omega &\longmapsto (X_1(\omega), X_2(\omega))\end{aligned}$$

$$\begin{aligned} g : \mathbb{R}_+ \times \mathbb{R}_+ &\longrightarrow \mathbb{R}_+\\ (x_1, x_2) &\longmapsto x_1 + x_2\end{aligned}$$

Note $P \circ X^{-1} = F_1 \times F_2$, where F_i is the distribution of X_i. This follows since

$$\begin{aligned} P\circ X^{-1}(A_1\times A_2) &= P\left((X_1, X_2) \in A_1 \times A_2\right) = P(X_1 \in A_1, X_2 \in A_2) = P(X_1 \in A_1)P(X_2 \in A_2)\\ &= F_1(A_1)F_2(A_2) = F_1 \times F_2(A_1 \times A_2).\end{aligned}$$

So $P \circ X^{-1}$ and $F_1 \times F_2$ agree on RECT and hence on $\mathcal{B}(\text{RECT}) = \mathcal{B}_1 \times \mathcal{B}_2$. Then by CorollAry 5.6,

$$\begin{aligned} E[X_1X_2] = E[g(X)] &= \int_\Omega g(X(\omega))P(d\omega) = \int_{\mathbb{R}^+\times\mathbb{R}^+} g(x)P\circ X^{-1}(dx) = \int_{\mathbb{R}_+\times\mathbb{R}_+} gd(F_1\times F_2)\\ &= \int_{\mathbb{R}^+} x_2 \int_{\mathbb{R}^+} x_1F_1(dx_1)F_2(d_2) \text{ (Fubini)} = E[X_1]\int_{\mathbb{R}^+} x_2F_2(dx_2) = E[X_1]E[X_2].\end{aligned}$$

□

Example 5.21 (Convolution)**.** Suppose X_1, X_2 are two independent random variables with distributions F_1, F_2. The distribution function of the random variable X_1+X_2 is given by the convolution $F_1 * F_2$ of the distribution functions. For $x \in \mathbb{R}$,

$$P(X_1 + X_2 \leqslant x) =: F_1 * F_2(x) = \int_{\mathbb{R}} F_1(x-u)F_2(du) = \int_{\mathbb{R}} F_2(x-u)F_1(du).$$

Proof. To see this, proceed as in the previous example. Let $X = (X_1, X_2)$ which has a distribution $F_1 \times F_2$ and set

$$g(x_1, x_2) = \mathbb{1}_{\{(u,v)\in\mathbb{R}^2 : u+v\leqslant x\}}(x_1, x_2),\ (x_1, x_2) \in \mathbb{R}^2.$$

From Corollary 5.6,

$$\begin{aligned}
P(X_1 + X_2 \leqslant x) &= E[\mathbb{1}_{\{X_1+X_2\leqslant x\}}] = E\left[\mathbb{1}_{\{(u,v)\in\mathbb{R}^2: u+v\leqslant x\}}(X_1, X_2)\right] = E[g(X)] = \int_{\mathbb{R}^2} g d(F_1 \times F_2) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{\{(u,v)\in\mathbb{R}^2: u+v\leqslant x\}}(x_1, x_2) F_1(dx_1) F_2(dx_2) \text{ (independence)} \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{\{v\in\mathbb{R}: v\leqslant x-x_2\}}(x_1) F_1(dx_1) F_2(dx_2) \\
&= \int_{\mathbb{R}} F_1(x - x_2) F_2(dx_2). \qquad \square
\end{aligned}$$

Chapter 6

Convergence Concepts

Much of classical probability theory and its applications to statistics concerns limit theorem; that is, the asymptotic behavior of a sequence of random variables. The sequence could consist of sample averges, cumulative sums, extremes, sample quantiles, sample correlations, and so on. Whereas probability theory disscusses limit theorem, the theory of statistics is concerned with large sample properties of statistics, where a statistics is just a function of the sample.

6.1 Almost Sure Convergence

Proposition 6.1. Let $\{X_n\}$ be iid random variables with common distribution function $F(x)$. Asume that $F(x) < 1$, for all x. Set

$$M_n = \bigvee_{i=1}^{n} X_i \to \infty \text{ a.s..}$$

Proof. Since $F(j) < 1$, we have $\sum_{n=1}^{\infty} P(M_n \leqslant j) = \sum_{n=1}^{\infty} F^n(j) < \infty$. By the Borel-Cantelli Lemma, $P\left(\lim_{n\to\infty} \sup\{M_n \leqslant j\}\right) = 0$. Let $N_j = \lim_{n\to\infty} \sup\{M_n \leqslant j\}$ for $j \in \mathbb{N}$. Then $P(N_j) = 0$ for $j \in \mathbb{N}$. Note $N_j^c = \lim_{n\to\infty} \inf\{M_n \geqslant j\}$. So for $\omega \in N_j^c$, we get $M_n(\omega) > j$ for all large n. Let $N = \bigcup_{j=1}^{\infty} N_j$, so $P(N) \leqslant \sum_{j=1}^{\infty} P(N_j) = 0$. Note

$$N^c = \bigcap_{j=1}^{\infty} N_j^c = \bigcap_{j=1}^{\infty} \lim_{n\to\infty} \inf\{M_n > j\},$$

and $P(N^c) = 1$. If $\omega \in N^c$, we have the property that for any j, $M_n(\omega) > j$ for all sufficiently large n. Thus, $P(M_n \to \infty) = 1$. □

6.2 Convergence in Probability

Theorem 6.2 (Convergence a.s implies convergence i.p.). *Suppose that* $\{X_n, n \geqslant 1, X\}$ *are r.v.'s on a probability space* $(\Omega, \mathcal{B}, P)$. *If* $X_n \to X$, *a.s., then* $X_n \xrightarrow{p} X$.

Proof. If $X_n \to X$ a.s., then for any $\epsilon > 0$

$$0 = P\left(|X_n - X| > \epsilon \text{ i.o.}\right) = P(\lim_{n\to\infty}\sup\{|X_n - X| > \epsilon\}) = P\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\{|X_n - X| > \epsilon\}\right)$$
$$= \lim_{n\to\infty} P\left(\bigcup_{k=n}^{\infty}\{|X_n - X| > \epsilon\}\right) \geqslant \lim_{n\to\infty} P(|X_n - X| > \epsilon). \qquad \square$$

Remark. The definition of convergence i.p. and convergence a.s. can be readily extended to random elements of metric spaces.

6.3 Statistical Terminology

In statistical estimation theory, almost sure and in probability convergence have analogues as strong or weak consistency. Given a family of probability models $((\Omega, \mathcal{B}, P_\theta))$. Suppose the statistician ges to observe random variables $X_1 \cdots, X_n$ defined on Ω and based on these observations must decide which is the correct model; that is, which is the correct value of θ. Statistical estimation means: select the correct model. For example, suppose $\Omega = \mathbb{R}^\infty$, $\mathcal{B} = \mathcal{B}(\mathbb{R}^\infty)$. Let $\omega = (x_1, x_2, \cdots)$ and define $X_n(\omega) = x_n, \forall n \in \mathbb{Z}^+$. For each $\theta \in \mathbb{R}$, let P_θ be product measure on $\mathbb{R}^\infty$ which makes $\{X_n,\ n \geqslant 1\}$ iid with common $N(\theta, 1)$ distribution. Based on observing $X_1, \ldots, X_n$, one estimate θ with an appropriate function of the observations $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$. $\hat{\theta}_n(X_1, \ldots, X_n)$ is called a *statistic* and is also an *estimator*. When one actually does the experiment and observes, $X_1 = x_1, \ldots, X_n = x_n$, then $\hat{\theta}_n(x_1, \ldots, x_n)$ is called the estimate. So the estimator is a random element while the estimate is a number or maybe a vector if θ is multidimentional. In this example, the usual choice of estimator is $\hat{\theta}_n = \sum_{i=1}^n X_i/n$. The estimator $\hat{\theta}_n$ is weakly consistent if for all $\theta \in \Theta$,

$$P_\theta\left(\left|\hat{\theta}_n - \theta\right| > \epsilon\right) \to 0,$$

that is, $\hat{\theta}_n \stackrel{P_\theta}{\to} \theta$. This indicates that no matter what the true parameter is or to put it another way, no matter what the true (but unknown) state of nature is, $\hat{\theta}$ does a good job estimating the true parameter. $\hat{\theta}_n$ is strongly consistent if for all $\theta \in \Theta$, $\hat{\theta}_n \to \theta$, P_θ-a.s..

Chapter 7

Laws of Large Numbers and Sums of Independent Random Variables

7.1 Truncation and Equivalence

We will see that is is easier to deal with random variables that are uniformly bounded or that have moments. Many technique rely on these desirable properties being present. If these properties are not present, a technique called truncation can induce their presence but then a comparison must be made between the original random variables and the truncated ones. For instance, we often want to compare $\{X_j\}$ with $\left\{X_j \mathbb{1}_{\{|X_j| \leqslant n\}}\right\}$. The following is a useful concept, expecially for problems needing almost sure convergence.

Definition 7.1. Two sequences $\{X_n\}$ and $\{X'_n\}$ are *tail equivalent* if $\sum_{n=1}^{\infty} P(X_n \neq X'_n) < \infty$.

When two sequences are tail equivalent, their sums behave asympotically the same as shown next.

Proposition 7.2 (Equivalence)**.** Suppose the two sequences $\{X_n\}$ and $\{X'_n\}$ are tail equivalent. Then

(a) $\sum_{n=1}^{\infty}(X_n - X'_n)$ converges a.s.

(b) Two two series $\sum_{n=1}^{\infty} X_n$ and $\sum_{n=1}^{\infty} X'_n$ converges a.s. together or diverge a.s. together; that is $\sum_{n=1}^{\infty} X_n$ converges a.s. if and only if $\sum_{n=1}^{\infty} X'_n$ converges a.s..

(c) If there exists a sequence $\{a_n\}$ such that $a_n \uparrow \infty$ and if there exists a random variable X such that $\frac{1}{a_n}\sum_{j=1}^{n} X_j \xrightarrow{\text{a.s.}} X$, then also $\frac{1}{a_n}\sum_{j=1}^{n} X'_j \xrightarrow{\text{a.s.}} X$.

Proof. (a) By the Borel-CantellI Lemma, since they are tail equivalent, $\lim_{n\to\infty}\sup\{X_n \neq X'_n\} = 0$, or equivalently $P\left(\lim_{n\to\infty}\inf\{X_n = X'_n\}\right) = 1$. So for $\omega \in \lim_{n\to\infty}\inf\{X_n = X'_n\}$, we have that $X_n(\omega) = X'_n(\omega)$ from some index onwars, say for $n \geqslant N(\omega)$.

□

7.2 A General Weak Law of Large Numbers

Theorem 7.3 (General weak law of large numbers)**.** *Suppose $\{X_n, n \geqslant 1\}$ are independent random variables and define $S_n = \sum_{j=1}^n X_j$. If*

(a) $\sum_{j=1}^n P(|X_j| > n) \to 0$,

(b) $\frac{1}{n^2}\sum_{j=1}^n EX_j^2 \mathbb{1}_{\{|X_j| \leqslant n\}} \to 0$,

then if we define $a_n = \sum_{j=1}^n E\left(X_j \mathbb{1}_{\{|X_j| \leqslant n\}}\right)$, we get

$$\frac{S_n - a_n}{n} \xrightarrow{P} 0. \tag{7.1}$$

One of the virtues of this result is that no assumptions about moments need to be made. Also, although this result is presented as conditions which are sufficient for (7.1)*.*

7.3 Almost Sure Convergence of Sums of Independent Random Variables

Reminder: If $\{X_n\}$ is a monotone sequence of random variables, then $X_n \xrightarrow{P} X$ implies $X_n \xrightarrow{\text{a.s.}} X$.

Theorem 7.4 (Levy's theorem)**.** *If $\{X_n, n \geqslant 1\}$ is an independent sequence of random variables, then $\sum_n^\infty X_n$ converges in probability if and only if $\sum_n^\infty X_n$ converges a.s..*

Theorem 7.5 (Kolmogorov Convergence Criterion)**.** *Suppose $\{X_n,\ n \geqslant 1\}$ is a sequence of independent random variables. If $\sum_{j=1}^\infty \mathrm{Var}(X_j) < \infty$, then $\sum_{j=1}^\infty (X_j - E(X_j))$ converges almost surely.*

7.4 Strong Laws of Large Numbers

Lemma 7.6 (Kronecker's lemma)**.** Suppose we have two sequence $\{x_k\}$ and $\{a_n\}$ such that $x_k \in \mathbb{R}$ and $0 < a_n \uparrow \infty$. If $\sum_{k=1}^\infty \frac{x_k}{a_k}$ converges, then $\lim_{n\to\infty} a_n^{-1} \sum_{k=1}^n x_k = 0$.

7.4.1 Record counts

Suppose $\{X_n,\ n \geqslant 1\}$ is an iid sequence with common continuous distribution function F. Define

$$u_N = \sum_{j=1}^N \mathbb{1}_{\{X_j \text{ is a record}\}} = \sum_{j=1}^N \mathbb{1}_j,$$

where $\mathbb{1}_j = \mathbb{1}_{\{X_j \text{ is a record}\}}$. So u_N is the number of records in the first N observations.

Proposition 7.7 (Logarithmic growth rate)**.** The number of records in an iid sequence grows logarithmically and we have the almost sure limit

$$\lim_{n\to\infty} \frac{\mu_N}{\log N} \to 1.$$

7.4.2 Explosions in the Pure Birth Process

Let $\{X_j,\ j \geqslant 1\}$ be nonnegative independent random variables and suppose

$$P(X_n > x) = e^{-\lambda_n x},\ x > 0,$$

where $\lambda_n \in \mathbb{R}^+$ for $n \in \mathbb{N}$ are called the *birth parameters*. Define the birth time process $S_n = \sum_{t=1}^n X_t$ and the population size process $\{X(t),\ t \geqslant 0\}$ of the pure birth process by

$$X(t) = \begin{cases} 1, & \text{if } 0 \leqslant t < S_1, \\ 2, & \text{if } S_1 \leqslant t < S_2, \\ 3, & \text{if } S_2 \leqslant t < S_3, \\ \vdots & \end{cases}$$

Next define the event explosion by

$$\{\text{explosion}\} = \left[\sum_{n=1}^{\infty} X_n < \infty\right] = [X(t) = \infty \text{ for some finite } t]\,.$$

Proposition 7.8. For the probability of explosion, we have

$$P[\text{explosion}] = \begin{cases} 1, & \text{if } \sum_{n=1}^{\infty} \lambda_n^{-1} < 0, \\ 0 & \text{if } \sum_{n=1}^{\infty} \lambda_n^{-1} = \infty. \end{cases}$$

Recall that we know that $P(\sum_{n=1}^{\infty} X_n < \infty) = 0$ or 1 by Kologorov Zero-One Law.

7.5 The Strong Law of Large Numbers for IID Sequences

Theorem 7.9 (Kolmogorov' SLLN)**.** *Let $\{X_n,\ n \geqslant 1\}$ be an iid sequence of random variables and set $S_n = \sum_{i=1}^n X_i$. There exists $c \in \mathbb{R}$ such that $\overline{X}_n = S_n/n \xrightarrow{a.s.} c$ if and only if $E(|X_1|) < \infty$ in which case $c = E(X_1)$.*

Corollary 7.10. If $\{X_n\}$ is iid, then

$$E(|X_1|) < \infty \text{ implies } \overline{X}_n \xrightarrow{\text{a.s.}} \mu = E(X_1),$$

and

$$EX_1^2 < \infty \text{ implies } S_n := \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \xrightarrow{\text{a.s.}} \sigma^2 =: \text{Var}(X_1).$$

Chapter 8

Convergence in Distribution

8.1 Basic Definitions

Definition 8.1. F is a *distribution function* if

(a) $0 \leqslant F(x) \leqslant 1$;

(b) F is non-decreasing;

(c) $F(x^+) = F(x)$ for $x \in \mathbb{R}$, where $F(x^+) = \lim_{\epsilon>0, \epsilon\downarrow 0} F(x+\epsilon)$.

Also, remember the notation

$$F(\infty) := \lim_{y\uparrow\infty} F(y),$$

$$F(-\infty) := \lim_{y\downarrow-\infty} F(y).$$

F is a *probability distribution function* if $F(-\infty) = 0$ and $F(+\infty) = 1$. In this case, F is *proper* or *non-defective.*

Lemma 8.2. A distribution function $F(x)$ is determined on a dense set. Let D be dense in $\mathbb{R}$. Suppose $F_D(\cdot)$ is defined on D and satisfies the following:

(a) $F_D(\cdot)$ is non-decreasing on D.

(b) $0 \leqslant F_D(x) \leqslant 1$ for $x \in D$.

(c) $\lim_{x\in D, x\to+\infty} F_D(x) = 1$, $\lim_{x\in D, x\to-\infty} F_D(x) = 0$.

Define for $x \in \mathbb{R}$,

$$F(x) := \inf_{y>x, y\in D} F_D(y) = \lim_{y\downarrow x, y\in D} F_D(y).$$

Then F is a right continuous probability distribution function. Thus, any two right continuous df's agreeing on a dense set will agree everywhere.

Proof. Fix $x \in \mathbb{R}$. Given $\epsilon > 0$, there exists $x' \in D, x' > x$ such that $F(x) + \epsilon \geqslant F_D(x')$. By the definition of F for any $y \in (x, x')$, $F_D(x') \geqslant F(y)$. Then $F(x) + \epsilon \geqslant F(y)$ for any $y \in (x, x')$. Now F is monotone ($F(x^+) = \lim_{y \downarrow x} F(y)$ exists), so let $y \downarrow x$ to get $F(x) + \epsilon \geqslant F(x^+)$. This is true for all small $\epsilon > 0$, so let $\epsilon \downarrow 0$ and we get $F(x) \geqslant F(x^+)$. Since monotonicity of F implies $F(x^+) \geqslant F(x)$, we get $F(x) = F(x^+)$ as desired. □

Remark. If $g : \mathbb{R} \to \mathbb{R}$ have the property that for any $x \in \mathbb{R}$, $g(x^+) = \lim_{y \downarrow x} g(y)$ exists. Set $h(x) = g(x^+)$. Then h is right continuous.

Definition 8.3. *Weak convergence.* The sequence $\{F_n\}$ converges weakly to F, written $F_n \Rightarrow F$, if $F_n(x) \to F(x)$ for all $x \in C(F)$.

Remark. In the definition of weak convergence, F may not be proper.

Proposition 8.4. $C(F)^c$ is a at most countable set.

Proof.

$$C(F)^c = \{x \in \mathbb{R}, F(x) - F(x^-) > 0\} = \bigcup_{n=1}^{\infty} \left\{x \in \mathbb{R}, F(x) - F(x^-) \geqslant \frac{1}{n}\right\} = \bigcup_{n=1}^{\infty} B_n.$$

Claim. B_n has at most n distinct elements. Suppose not, then we can find $x_1, x_2, \ldots, x_{n+1} \in B_n$ and $x_1 < x_2 \cdots x_n < x_{n+1}$. Then $F(x_{n+1}^-) \geqslant F(x_n)$ and

$$\begin{aligned} F(x_{n+1}) &= F(x_{n+1}) - F(x_{n+1}^-) + F(x_{n+1}^-) \geqslant \frac{1}{n} + F(x_n) \geqslant \frac{2}{n} + F(x_{n-1}) \\ &\geqslant \frac{n}{n} + F(x_1) \geqslant \frac{n+1}{n} + F(x_1^-) > 1, \end{aligned}$$

which is a contradiction. □

Example 8.5. Let N be an $N(0,1)$ random variable so that the distribution function is symmetric. Define for $n \geqslant 1$, $X_n = (-1)^n N$. Then $X_n \overset{d}{=} N$, so automatically $X_n \Rightarrow N$. But of course $\{X_n\}$ neither converges almost surely nor in probability.

Remark. Weak limits are unique. If $F_n \to F$, and also $F_n \to G$, then $F = G$. There is a simple reason for this. The set $(C(F))^c \cup (C(G))^c$ is countable so $\text{INT} = \mathcal{C}(F) \cap \mathcal{C}(G) = \mathbb{R} \setminus$ a countable set and hence is dense. For $x \in \text{INT}$, $F_n(x) \to F(x)$, $F_n(x) \to G(x)$. So $F(x) = G(x)$ for $x \in \text{INT}$, and hence by Lemma 8.2, we have $F = G$.

Example 8.6. Suppose F_n puts mass $\frac{1}{n}$ at points $\{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\}$. If $F(x) = x$, $0 \leqslant x \leqslant 1$ is the uniform distribution on $[0,1]$, then for $x \in (0,1)$, $F_n(x) = \frac{\lfloor nx \rfloor}{n} \to x = F(x)$. Thus we have weak convergence $F_n \Rightarrow F$. However if $\mathbb{Q}$ is the set of rationals in $[0,1]$, $F_n(\mathbb{Q}) = 1$, $F(\mathbb{Q}) = 0$. So $F_n(A) \not\to F(A)$ for $A \in \mathcal{B}(\mathbb{R})$.

8.2 Scheffe' lemma

Lemma 8.7 (Scheffe's lemma)**.** Suppose F and $\{F_n\}_{n \geqslant 1}$ are probability distributions with densities $\{f, f_n \ n \geqslant 1\}$. Then

$$\sup_{B \in \mathcal{B}(\mathbb{R})} |F_n(B) - F(B)| = \frac{1}{2} \int |f_n(x) - f(x)| dx.$$

If $f_(x) \to f(x)$ almost everywhere, then $\int |f_n(x) - f(x)|dx \to 0$, and thus $F_n \to F$ in total variation (and hence weakly).

8.2.1 Scheffe's lemma and Order Statistics

Proposition 8.8. Suppose $\{U_n\}_{n\geqslant 1}$ are iid $U(0,1)$ random variables so that

$$P(U_j \leqslant x) = x,\ 0 \leqslant x \leqslant 1, \forall j \in \mathbb{N}$$

and suppose $U_{(1,n)} \leqslant U_{(2,n)} \leqslant \cdots \leqslant U_{(n,n)}$ are the order statistics. Assume $k = k(n)$ is a function of n satisfying $k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$. Let

$$\xi_n = \frac{U_{(k,n)} - \frac{k}{n}}{\sqrt{\frac{k}{n}\left(1 - \frac{k}{n}\right)\frac{1}{n}}}.$$

Then the density of ξ_n converges to a standard normal density and hence by Scheffe' lemma

$$\sup_{B\in\mathcal{B}(\mathbb{R})}\left|P(\xi_n \in B) - \int_B \sqrt{\frac{1}{2\pi}} e^{-u^2/2} du\right| \to 0.$$

Proof. For $0 < x < 1$, $P(U_{(k,n)} \leqslant x)$ is the binomial probability of at least k successes in n trials when the success probability is x. So

$$P(U_{(k,n)} \leqslant x) = \sum_{i=k}^{n} \binom{n}{i} x^i (1-x)^{n-i}.$$

Dif and only iferentiating, we get the density $f_{(k,n)}(x)$ of $U_{(k,n)}$ to be

$$\begin{aligned}
f_n(x) &= \sum_{i=k}^{n} \binom{n}{i} i x^{i-1}(1-x)^{n-i} - \sum_{i=k}^{n} \binom{n}{i} x^i (n-i)(1-x)^{n-i-1} \\
&= \sum_{i=k-1}^{n-1} \binom{n}{i+1} (i+1) x^i (1-x)^{n-i-1} - \sum_{i=k}^{n-1} \binom{n}{i} x^i (n-i)(1-x)^{n-i-1} \\
&= \sum_{i=k-1}^{n-1} \frac{n!}{i!(n-i-1)!} x^i (1-x)^{n-i-1} - \sum_{i=k}^{n-1} \frac{n!}{i!(n-i-1)!} x^i (1-x)^{n-i-1} \\
&= \binom{n}{k} k x^{k-1} (1-x)^{n-k} \\
&= \frac{n!}{(k-1)!(n-k)!} x^{k-1} (1-x)^{n-k},\ 0 < x < 1.
\end{aligned}$$

Since $\sqrt{\frac{k}{n}\left(1 - \frac{k}{n}\right)\frac{1}{n}} \sim \frac{\sqrt{k}}{n}$ as $n \to \infty$, by convergence of types theorem discussed below assures us we can replace the square root in the expression for ξ_n by $\sqrt{k}/n$ and by transformation theorem, the pdf of ξ_n is

$$g_n(x) = \frac{\sqrt{k}}{n} f_n\left(\frac{\sqrt{k}}{n} x + \frac{k}{n}\right).$$

By Stirling's formula, as $n \to \infty$, ($\frac{k}{n} \to 0$ as $n \to \infty$.)

$$\frac{n!}{(k-1)!(n-k)!} \sim \frac{\sqrt{n}}{\sqrt{2\pi}\left(\frac{k}{n}\right)^{k-1/2}\left(1-\frac{k}{n}\right)^{n-k}}.$$

So as $n \to \infty$,

$$\begin{aligned}
g_n(x) &\sim \frac{\sqrt{k}}{n}\frac{\sqrt{n}}{\sqrt{2\pi}\left(\frac{k}{n}\right)^{k-1/2}\left(1-\frac{k}{n}\right)^{n-k}}\left(\frac{\sqrt{k}}{n}x+\frac{k}{n}\right)^{k-1}\left(1-\frac{\sqrt{k}}{n}x-\frac{k}{n}\right)^{n-k} \\
&= \frac{1}{\sqrt{2\pi}}\frac{\left(\frac{\sqrt{k}}{n}x+\frac{k}{n}\right)^{k-1}}{\left(\frac{k}{n}\right)^{k-1}}\frac{\left(1-\frac{k}{n}-\frac{\sqrt{k}}{n}x\right)^{n-k}}{\left(1-\frac{k}{n}\right)^{n-k}} \\
&= \frac{1}{\sqrt{2\pi}}\left(1+\frac{x}{\sqrt{k}}\right)^{k-1}\left(1-\frac{x}{(n-k)/\sqrt{k}}\right)^{n-k}.
\end{aligned}$$

It suffices to prove that

$$\left(1+\frac{x}{\sqrt{k}}\right)^{k-1}\left(1-\frac{x}{(n-k)/\sqrt{k}}\right)^{n-k},$$

or equivalently,

$$(k-1)\log\left(1+\frac{x}{\sqrt{k}}\right)+(n-k)\log\left(1-\frac{x}{(n-k)/\sqrt{k}}\right) \to -\frac{x^2}{2}.$$

Observe that, for $|t|<1$, $-\log(1-t)=\sum_{n=1}^{\infty}\frac{t^n}{n}$, and therefore

$$\delta(t) := \left|-\log(1-t)-(t+\frac{t^2}{2})\right| \leqslant \sum_{n=3}^{\infty}|t|^n = \frac{|t|^3}{1-|t|} \leqslant 2|t|^3, \text{ if } |t|<\frac{1}{2}.$$

Then

$$\begin{aligned}
&(k-1)\log\left(1+\frac{x}{\sqrt{k}}\right)+(n-k)\log\left(1-\frac{x}{(n-k)/\sqrt{k}}\right) \\
&= (k-1)\left(\frac{x}{\sqrt{k}}-\frac{x^2}{2k}\right)-(n-k)\left(\frac{x}{(n-k)/\sqrt{k}}+\frac{x^2}{2(n-k)^2/k}\right)+0(1) \\
&= -\frac{x^2}{2}-\frac{x}{\sqrt{k}}+\frac{x^2}{2k}-\frac{x^2}{2(n-k)/k}+0(1). \\
&= -\frac{x}{\sqrt{k}}-\frac{x^2}{2}\left(1-\frac{1}{k}-\frac{1}{n/k-1}\right)+0(1) \\
&\to -\frac{x^2}{2}
\end{aligned}$$

since

$$0(1) = (k-1)\delta\left(\frac{x}{\sqrt{k}}\right)+(n-k)\delta\left(\frac{x}{(n-k)/\sqrt{k}}\right) \to 0,$$

and $k(n) \to \infty$ and $\frac{k}{n} \to 0$ as $n \to \infty$.

□

8.3 Appetizer

Definition 8.9. Suppose $X_1, \ldots, X_n$ are random variables. Set $M_n = \max\{X_1, \ldots, X_n\}$.

Example 8.10. Let Y_i be the low temperature in year i. Set $X_i = -Y_i$ so that

$$-M_n = -\max\{X_1, \ldots, X_n\} = \min\{-X_1, \ldots, -X_n\} = \min\{Y_1, \ldots, Y_n\}$$

is the lowest temperature in year $1, \ldots, n$.

Example 8.11. Suppose $X_1, \ldots, X_n$ are iid with d.f. F. Since $M_n \leqslant x$ if and only if $X_1 \leqslant x, \ldots, X_n \leqslant x$,

$$P(M_n \leqslant x) = P(X_1 \leqslant x, \ldots, X_n \leqslant x) = \prod_{i=1}^n P(X_i \leqslant x) = \prod_{i=1}^n F(x) = F(x)^n.$$

Definition 8.12. Let $X_r = \inf\{x : F(x) = 1\}$ with $\inf\{\emptyset\} = \infty$.

Example 8.13. Uniform$[0, 1]$.

$$F(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leqslant x < 1 \\ 1, & x = 1 \end{cases}.$$

Then $X_r = 1$.

Example 8.14. $\exp(\lambda)$.

$$F(x) = \begin{cases} 0, & x \leqslant 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$$

Then $X_r = \infty$.

Corollary 8.15. When $x < X_r$, we have $F(x) < 1$ and then $P(M_n \leqslant x) = (F(x))^n \to 0$. If $x > X_r$, then $F(x) = 1$ and $P(M_n \leqslant x) = [F(x)]^n = 1 \to 1$. Thus, $M_n \xrightarrow{d} M$, where M is degenerate and

$$F_M(x) = \begin{cases} 0, & x < X_r \\ 1, & x > X_r \end{cases}$$

Remark. If $X_r = \infty$, then $F_M(x) = 0$ for $x \in \mathbb{R}$ since $F(x) < \infty$ for $x \in \mathbb{R}$.

$X_1, X_2, \cdots$ are iid, $E(X_1) = \mu$, $\mathrm{Var}(X_1) = \sigma^2$.

Theorem 8.16. *Central Limit Theorem.*

First, make a single variable center 0 and variance 1,

$$\frac{X_i - \mu}{\sigma}.$$

Then make the sum center 0 and variance 1,

$$\sum_{i=1}^n \frac{X_i - \mu}{\sigma\sqrt{n}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} = \frac{S_n - b_n}{a_n} \xrightarrow{d} N(0, 1).$$

Then

$$P\left(\frac{S_n - b_n}{a_n} \leqslant x\right) = P(S_n \leqslant a_n x + b_n) = F_{S_n}(a_n x + b_n) \longrightarrow N(0,1).$$

Can we find constants and b_n so that

$$P\left(\frac{M_n - b_n}{a_n} \leqslant x\right) = P(M_n \leqslant a_n x + b_n) \longrightarrow G(x)?$$

where G is not degenerate.

Example 8.17. $X_1, X_2, \cdots$ are iid and $X_i \sim$exp(1). Consider $M_n = \max(X_1, \ldots, X_n)$. Let $X_{(0)} = 0$, then for all $i \in [n]$, $X_{(i)} - X_{(i-1)} \sim \exp(n+1-i)$, then

$$M_n = X_{(1)} + (X_{(2)} - X_{(1)}) + \cdots + (X_{(n)} - X_{(n-1)}).$$

So

$$E[M_n] = \sum_{i=1}^{n} \frac{1}{n+i-1} \sim \log n = b_n.$$

Set $a_n = 1$. For $x \in \mathbb{R}$, as soon as $x + \log n \geqslant 0$,

$$P(M_n - \log n \leqslant x) = P(M_n \leqslant x + \log n) = \left(1 - e^{-(x+\log n)}\right)^n = \left(1 - \frac{1}{n}e^{-x}\right)^n \to e^{-e^{-x}}.$$

Thus, $M_n - b_n \xrightarrow{d} G$, where $G(x) = e^{-e^{-x}}$ for $x > 0$. Choose b_n so that $1 - e^{-b_n} = F(b_n) = 1 - \frac{1}{n}$, where b_n is the last n-th tile of d.f F. Then we get $b_n = \log n$ as before.

Example 8.18. Pareto

$$F(x) = \begin{cases} 0, & x < 1 \\ 1 - \frac{1}{x}, & x \geqslant 1 \end{cases}$$

$$E[X] = \int_1^\infty (1 - F(x))dx = \int_1^\infty \frac{1}{x} \to \infty.$$

X has a **heavey tail**. Take $b_n = 0$ and choose a_n so that $1 - \frac{1}{a_n} = F(a_n) = 1 - \frac{1}{n}$. Then we get $a_n = n$. For $x > 0$, when $a_n x > 1$,

$$P\left(\frac{M_n}{a_n} \leqslant x\right) = (M_n \leqslant a_n x) = (F(nx))^n = \left(1 - \frac{1}{nx}\right)^n \to e^{-x^{-1}}.$$

Thus, $\frac{M_n}{a_n} \xrightarrow{d} G$, where

$$G(x) = \begin{cases} 0, & x \leqslant 0 \\ e^{-x^{-1}}, & x > 0 \end{cases}$$

8.4 Left Continuous Inverse

Proposition 8.19. Any c.d.f. $F : A \to \mathbb{R}$ is right continuous, where $A \subseteq \mathbb{R}$.

Proof. Fix $x \in \mathbb{R}$. Assume $\{x_n\} \subseteq \mathbb{R}$ and $x_n \downarrow x$. Then $\{X \leqslant x_1\} \supseteq \{X \leqslant x_2\} \supseteq \cdots$, and $\bigcap_{n=1}^{\infty}\{X \leqslant x_n\} = \{X \leqslant x\}$. By the continuity of probability measure, we have $P(X \leqslant x) = \lim_{n\to\infty} P(X \leqslant x_n)$. Then $F(x) = \lim_{n\to\infty} F(x_n)$. By the equavalent definition of continuity, we have F is right continuous. □

Definition 8.20. Let H be a nondecreasing function defined on $\mathbb{R}$. Define the the *continuous inverse*

$$H^{\leftarrow}(y) = \inf\{s : H(s) \geqslant y\}.$$

Remark. Given H nondecreasing, consider the graphs.
If H is strictly increasing, then $H^{\leftarrow}(y)$ is strictly increasing.
If H is flat, then there is a jump in $H^{\leftarrow}(y)$.
If there is a jump in H, then $H^{\leftarrow}(y)$ is flat.
Thus, $H^{\leftarrow}$ is also nondecreasing.

Proposition 8.21. Let $A(y) = \{x : H(x) \geqslant y\}$, then $A(y) = (H^{\leftarrow}(y), \infty)$, or $A(y) = (H^{\leftarrow}(y), \infty)$.

Proof. Suppose $s \in A(y)$, then $s \geqslant H^{\leftarrow}(y)$. So $A(y) \subseteq [H^{\leftarrow}(y), \infty)$. Since $H^{\leftarrow}(y) = \inf A(y)$, we have for any $\delta > 0$, there exists $s \in \mathbb{R}$ such that $H^{\leftarrow}(y) \leqslant s < H^{\leftarrow}(y) + \delta$. (Otherwise, for any $s \in \mathbb{R}$, $H^{\leftarrow}(y) > s$ or $H^{\leftarrow}(y) + \delta \leqslant s$, a contradiction.) Thus, $[H^{\leftarrow}(y) + \delta, \infty) \subseteq [H^{\leftarrow}(y), \infty)$. Since $\delta > 0$ is arbitrary, $(H^{\leftarrow}(y), \infty) \subseteq A(y)$. Hence, $s \geqslant H^{\leftarrow}(y)$. So $A(y) = (H^{\leftarrow}(y), \infty)$, or $A(y) = [H^{\leftarrow}(y), \infty)$. □

Proposition 8.22. Assume H is right continuous. Then

(a) $A(y)$ is closed, and then $A(y) = [H^{\leftarrow}(y), \infty)$.

(b) $H^{\leftarrow}(y) \in A(y)$ and then $H(H^{\leftarrow}(y)) \geqslant y$.

(c) $H^{\leftarrow}(y) \leqslant t$ if and only if $y \leqslant H(t)$.

(d) $H^{\leftarrow}(y) > t$ if and only if $y > H(t)$.

Proof. (a) If $s_n \in A(y)$ and $s_n \downarrow s$, then by right continuity $y \leqslant H(s_n) \downarrow H(s)$. So $H(s) \geqslant y$ and then $s \in A(y)$. If $s_n \in A(y)$ and $s_n \uparrow s$, then since H is nondecreasing, $y \leqslant H(s_n) \uparrow H(s^-) \leqslant H(s)$. So $y \leqslant H(s)$ and then $s \in A(y)$.

(b) By (a).

(c) Since $A(y) = [H^{\leftarrow}(y), \infty)$, $H^{\leftarrow}(y) \leqslant t$ if and only if $t \in A(y)$ if and only if $y \leqslant H(t)$.

(d) Similar to (c). □

Proposition 8.23. $H^{\leftarrow}$ is left continuous.

Proof. It equivalent to show if $\{y_n\} \subseteq \mathbb{R}$ and $y_n \uparrow y \in \mathbb{R}$, then $H^{\leftarrow}(y_n) \uparrow H^{\leftarrow}(y)$. Since $H^{\leftarrow}$ is nondecreasing, it sufficies to show for any $t \in \mathbb{R}$ for which

$$y_n \uparrow y \text{ and} \forall n \in \mathbb{N}^{>0}, H^{\leftarrow}(y_n) < t, \text{ then } H^{\leftarrow}(y) \leqslant t.$$

(Suppose not, then there exists $\epsilon > 0$ such that $H^{\leftarrow}(y) = t + \epsilon$, then $|H^{\leftarrow}(y_n) - H^{\leftarrow}(y)| > \epsilon$, a contradiction.) Then for $n \in \mathbb{N}$, $t \in \{x : H(x) \geqslant y_n\}$. Since $y_n \uparrow y$, $t \in \{s : H(s) \geqslant y\}$. Thus, $H^{\leftarrow}(y) \leqslant t$. □

Proposition 8.24. Let X have a df. F and $U \sim$ Uni(0,1), then $F^{\leftarrow}(U)$ also has a df. F.

Proof. $P(F^{\leftarrow}(U) \leqslant x) = P(U \leqslant F(x)) = F(x)$. □

Remark. Let $H : \mathbb{R} \to (0,1]$, then $H^{\leftarrow} : (0,1] \to \mathbb{R}$. Let $\xi_H(A) = \{x \in (0,1] : H^{\leftarrow}(x) \in A\}$ for $A \subseteq \mathbb{R}$. If $A \in \mathcal{B}(\mathbb{R})$, then $\xi_F(A) \in \mathcal{B}((0,1])$.

Example 8.25. Let X be a ramdon variable with cdf $F(x) = 1 - e^{-x}, x \geqslant 0$. Set $R(x) = -\log(1 - F(x))$. Then $R^{\leftarrow}(X)$ has has cdf F.

Proof.

$$P(R^{\leftarrow}(X) \leqslant x) = P(X \leqslant R(x)) = 1 - e^{-R(x)} = 1 - e^{\log(1-F(x))} = 1 - (1 - F(x)) = F(x). \qquad \square$$

Let $H, H_1, H_2, \cdots$ be nondecreasing and right continuous. Let $\mathcal{C}(H)$ be continuity points of H. We say $H_n \to H$ if for any $t \in \mathcal{C}(H)$, $H_n(t) \to H(t)$.

Theorem 8.26. *$H_n \to H$ implies $H_n^{\leftarrow} \to H^{\leftarrow}$.*

Proof. Let $t \in\ll C(H^{\leftarrow})$. It is equivalent to show $A : \lim_{n\to\infty} \inf H_n^{\leftarrow}(t) \geqslant H^{\leftarrow}(t)$ and $B : \lim_{n\to\infty} \sup H_n^{\leftarrow}(t) \leqslant H^{\leftarrow}(t)$. Let $\epsilon > 0$. Since the discontinuous point is countable, $\overline{\mathcal{C}(H)} = H$. Then there exists $x \in \mathcal{C}(H)$ such that $H^{\leftarrow}(t) - \epsilon < x < H^{\leftarrow}(t)$. By Proposition 8.22(4), $H(x) < t$. Since $x \in \mathcal{C}(H)$, $H_n(x) \to H(x)$. Then there exists $n_0 \in \mathbb{N}^{>0}$ such that for any $n \geqslant n_0$, $H_n(x) < t$. (Let $H(x) = t - \delta$ for some $\delta > 0$, then there exists some N such that $H_n(x) \leqslant H(x) < t$ as $n \geqslant N$.) By Proposition 8.22(4), for any $n \geqslant n_0$, $x < H_n^{\leftarrow}(t)$. So $\inf_{n\geqslant n_0} H_n^{\leftarrow}(t) \geqslant x > H^{\leftarrow}(t) - \epsilon$. Thus, $\lim_{n\to\infty} \inf H_n^{\leftarrow}(t) \geqslant \inf_{n\geqslant n_0} H_n^{\leftarrow}(t) \geqslant H^{\leftarrow}(t) - \epsilon$. Since ϵ is arbitrary, A follows. Let $\epsilon > 0$. Select $t' > t$ and choose $y \in \{x : H^{\leftarrow}(t') < x < H^{\leftarrow}(t') + \epsilon\} \cap \mathcal{C}(H)$. Since $y > H^{\leftarrow}(t')$, $H(y) \geqslant t' > t$. Since $y \in \mathcal{H}$, $H_n(y) \to H(y)$, and there exists $n_0 \in \mathbb{N}^{>0}$ such that for any $n \geqslant n_0$, $H_n(y) > t$. Then $y \in A(t) = \{s : H_n(s) \geqslant t\} = [H_n^{\leftarrow}(t), \infty)$ for any $n \geqslant n_0$. So for any $n \geqslant n_0$, $y \geqslant H_n^{\leftarrow}(t)$. Therefore, for any $n \geqslant n_0$, $H^{\leftarrow}(t') + \epsilon > y \geqslant H_n^{\leftarrow}(t)$. So $\sup_{n\geqslant n_0} H_n^{\leftarrow}(t) \leqslant H^{\leftarrow}(t') + \epsilon$. Then

$$\lim_{n\to\infty} \sup H_n^{\leftarrow}(t) \leqslant \sup_{n\geqslant n_0} H_n^{\leftarrow}(t) \leqslant H^{\leftarrow}(t') + \epsilon.$$

Since ϵ is arbitrary, $\lim_{n\to\infty} \sup H_n^{\leftarrow}(t) \leqslant H^{\leftarrow}(t')$. Since $t \in \mathcal{C}(H^{\leftarrow})$, let $t' \downarrow t$,

$$\lim_{n\to\infty} \sup H_n^{\leftarrow}(t) \leqslant H^{\leftarrow}(t). \qquad \square$$

8.5 The Baby Skorohod Theorem

Proposition 8.27. Suppose X, $\{X_n\}_{n\geqslant 1}$ are random variables. If $X_n \overset{a.s.}{=} X$, then $X_n \Rightarrow X$.

Proof. Consider the same probability space $(\Omega, \mathcal{A}, P)$. Suppose $X_n \xrightarrow{\text{a.s.}} X$ and let F and $\{F_n\}_{n\geqslant 1}$ be the distribution functions of X and $\{X_n\}_{n\geqslant 1}$, respectively. Then there exists $N \in \mathcal{A}$ such that $P(N) = 0$, and for $\omega \in N^c$, $\lim_{n\to\infty} X_n(\omega) = X(\omega)$. Fix $x \in \mathbb{R}$.

******************** *crucial* ********************

Claim I. for any $h > 0$ we have $\{X \leqslant x - h\} \cap N^c \subseteq \lim_{n\to\infty} \inf\{X_n \leqslant x\} \cap N^c$. Assuming $\{X \leqslant x - h\} \cap N^c \neq \emptyset$, fix $\omega \in \{X \leqslant x - h\} \cap N^c$. Since $\omega \in \{X \leqslant x - h\}$, $X(\omega) \leqslant x - h$. Since

$\omega \in N^c$, $X_n(\omega) \to X(\omega)$. Then $\exists N \in \mathbb{Z}^{>0}$ such that$\forall n \geqslant N$, $|X_n(\omega) - X(\omega)| < \frac{h}{2}$. Then for any $n \geqslant N$,

$$* * * * * * * * * * Trick * * * * * * * * * *$$

$$X_n(\omega) \leqslant X(\omega) + |X_n(\omega) - X(\omega)| < x - \frac{h}{2} < x.$$

Therefore, $X_n(\omega) < x$ for all but finitly many n. Thus, $\omega \in \lim_{n\to\infty} \inf\{X_n \leqslant x\}$. Also $\omega \in N^c$. Hence $\omega \in \lim_{n\to\infty} \inf\{X_n \leqslant x\} \cap N^c$.

Claim II. $\lim_{n\to\infty} \sup\{X_n \leqslant x\} \cap N^c \subseteq \{X \leqslant x\}$. Assuming $\lim_{n\to\infty} \sup\{X_n \leqslant x\} \neq \emptyset$, fix $\omega \in \lim_{n\to\infty} \sup\{X_n \leqslant x\} \cap N^c$. Since $\omega \in \lim_{n\to\infty} \sup\{X_n \leqslant x\}$, $X_n(\omega) \leqslant x$ for infinitely many n. We claim $X(\omega) \leqslant x$. Suppose not, then $\exists \epsilon > 0$ such that $X(\omega) = x + \epsilon$. Then $|X_n(\omega) - X(\omega)| \geqslant \epsilon$ for infinitely many n. Since $\omega \in N^c$, $X_n(\omega) \to X(\omega)$. Then for the same ϵ, $|X_n(\omega) - X(\omega)| < \epsilon$ for all but finitly many n, a contradiction. Hence $X(\omega) \leqslant x$. Thus, $\lim_{n\to\infty} \sup\{X_n \leqslant x\} \cap N^c \subseteq \{X \leqslant x\}$. As a result,

$$\{X \leqslant x - h\} \cap N^c \subseteq \lim_{n\to\infty} \inf\{X_n \leqslant x\} \cap N^c \subseteq \lim_{n\to\infty} \sup\{X_n \leqslant x\} \cap N^c \subseteq \{X \leqslant x\}.$$

Taking probabilites,

$$\begin{aligned} F(x-h) = P(X \leqslant x-h) = P(X \leqslant x - h \cap N^c) \leqslant P\big(\lim_{n\to\infty} \inf\{X_n \leqslant x\}\big) \leqslant \lim_{n\to\infty} \inf P(\{X_n \leqslant x\}) \\ \leqslant \lim_{n\to\infty} \sup P(\{X_n \leqslant x\}) \leqslant P\left(\lim_{n\to\infty} \sup\{X_n \leqslant x\}\right) \leqslant P(X \leqslant x). \end{aligned}$$

Since $x \in C(F)$, let $h \downarrow 0$ to get $F(x) \leqslant \lim_{n\to\infty} \inf F_n(x) \leqslant \lim_{n\to\infty} \sup F_n(x) \leqslant F(x)$. Thus, $\lim_{n\to\infty} F_n(x) = F(x)$. $\square$

Theorem 8.28 (Baby Skorohod Theorem)**.** *Suppose $X_n \Rightarrow X$. Then one can take the prob. space $([0,1], \mathcal{B}([0,1]), m)$, where m is Lebesgue measure. Construct r.v's $\widetilde{X}_n$ and $\widetilde{X}$ such that*

(a)

$$\widetilde{X} \stackrel{d}{=} X, \ \widetilde{X}_n \stackrel{d}{=} X_n, \forall n \in \mathbb{N}^{>0}.$$

(b)

$$\widetilde{X}_n \to \widetilde{X}, \ \textit{with prob 1 (a.s.)}.$$

Proof. Define

$$\begin{aligned} U : [0,1] &\to \mathbb{R} \\ \omega &\mapsto \omega. \end{aligned}$$

For $0 \leqslant x \leqslant 1$, $m(\{\omega : U(\omega) \leqslant x\}) = m([0,x]) = x$. Thus, U has a uniform distribution. Set

$$\widetilde{X} = F^{\leftarrow}(U), \text{ and } \widetilde{X}_n = F_n^{\leftarrow}(U), \forall n \in \mathbb{N}^{>0}.$$

By the Proposition 8.24,

$$\widetilde{X} \stackrel{d}{=} X, \text{ and } \widetilde{X}_n \stackrel{d}{=} X_n, \forall n \in \mathbb{N}^{>0}.$$

By the Theorem 8.26,

$$\widetilde{X}_n(\omega) \to \widetilde{X}(\omega), \forall \omega \in \mathcal{C}(\widetilde{X}).$$

The only thing left to do is to show $m\big(\mathcal{C}(\widetilde{X})\big) = 1$. If $\omega_1, \omega_2 \in [0,1]$ and $\omega_1 < \omega_2$,

$$\widetilde{X}(\omega_1) = F^{\leftarrow}(U(\omega_1)) = F^{\leftarrow}(\omega_1) \leqslant F^{\leftarrow}(\omega_2) = F^{\leftarrow}(U(\omega_2)) = \widetilde{X}(\omega_2).$$

So $\widetilde{X}$ is nondecreasing on $[0,1]$. Thus, $\widetilde{X}$ has at most a countable number of discontinuity. Hence $m\big(\mathcal{C}(\widetilde{X})\big) = 1$. $\square$

Theorem 8.29 (Continuous Mapping Theorem)**.** *Suppose $X_n \Rightarrow X$. Let $f : \mathbb{R} \to \mathbb{R}$ satisfy $P(X \in Disc(f)) = 0$. Then $f(X_n) \Rightarrow f(X)$, and if h is bounded, dominated convergence implies $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$, since assuming $|h| \leqslant M$, we have $\mathbb{E}[M] = M$, implying M is integrable.*

Proof. We take the probalility space $([0,1], \mathcal{B}([0,1], m))$, where m is Lebesgue measure. By the Baby Skorohod Theorem, we can construct $\widetilde{X}_n, \forall n \in \mathbb{Z}^{>0}$ and $\widetilde{X}$ such that $\widetilde{X}_n \stackrel{d}{=} X_n, \forall n \in \mathbb{Z}^{>0}$, and $\widetilde{X}_n \xrightarrow{\text{a.s.}} \widetilde{X}$. If $\widetilde{X}(\omega) \in C(f) = (\mathrm{Disc}(f))^c$, then $f\big(\widetilde{X}_n(\omega)\big) \to f\big(\widetilde{X}(\omega)\big)$.

$$\begin{aligned} m\big(\omega \in [0,1] : f(\widetilde{X}_n(\omega)) \to f(\widetilde{X}(\omega))\big) &\geqslant m\big(\omega \in [0,1] : \widetilde{X}(\omega) \in (\mathrm{Disc}(f))^c\big) \\ &= P\left(\{X \in \mathrm{Disc}(f)\}^c\right) = 1. \end{aligned}$$

So $f(\widetilde{X}_n) \xrightarrow{\text{a.s.}} f(\widetilde{X})$ w.r.t m. By Proposition 8.27, we have $f(\widetilde{X}_n) \Rightarrow f(\widetilde{X})$. Since $\widetilde{X}_n \stackrel{d}{=} X_n$, and $\widetilde{X} \stackrel{d}{=} X$, we have

$$f(X_n) \stackrel{d}{=} f(\widetilde{X}_n) \text{ and } f(\widetilde{X}) \stackrel{d}{=} f(X).$$

Thus,

$$f(X_n) \stackrel{d}{=} f(\widetilde{X}_n) \Longrightarrow f(\widetilde{X}) \stackrel{d}{=} f(X).$$ $\square$

8.6 The Delta Method

The delta method allows us to take a basic convergence, for instance to a limiting normal distribution, and apply smmoth functions and conclude that the functions are asymptotically normal as well. In statistical estimation we try to estimate a parameter θ from a parameter set Θ based on a random sample size n with a statistic $T_n = T_n(X_1, \ldots, X_n)$. This means we have a family of probability models $\{(\Omega, \mathcal{B}, P_\theta), \theta \in \Theta\}$, and we are trying to choose the correct model. The estimator T_n is *consistent* if $T_n \xrightarrow{P_\theta} \theta$ for every θ, meaning T_n converges to θ in probability P_θ for any $\theta \in \Theta$. The estimator T_n is consistent and asymptotically normal, if for any $\theta \in \Theta$,

$$\lim_{n\to\infty} P_\theta[\sigma_n(T_n - \theta) \leqslant x] = N(0,1,x).$$

From CLT, we get

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow N(0,1).$$

Equivalently,

$$\sqrt{n}\left(\frac{\overline{X} - \mu}{\sigma}\right) \Rightarrow N(0,1).$$

So $\overline{X}$ is consistent and an asymptotically normal estimator of μ. The delta method asserts that if $g(x)$ has a non-zero derivative $g'(\mu)$ at μ, then

$$\sqrt{n}\left(\frac{g(\overline{X})-g(\mu)}{\sigma g'(u)}\right) \Rightarrow N(0,1),$$

and so $g\left(\overline{X}\right)$ is also consistent and asymptotically normal for $g(\mu)$.

Remark. The proof does not depende on the limiting r.v. being $N(0,1)$ and would work equally well if $N(0,1)$ were replaced by any random variable Y.

Proof. By the Baby Skorohod Theorem, there exist random variable $\widetilde{Z}_n$ for any $n \in \mathbb{N}$ and $\widetilde{N}$ on the probability space $([0,1], B([0,1]), m)$ such that

$$\widetilde{Z}_n \stackrel{d}{=} \sqrt{n}\left(\frac{\overline{X}-\mu}{\sigma}\right) \text{ and } \widetilde{N} \stackrel{d}{=} N,$$

and $\widetilde{Z}_n \xrightarrow{\text{a.s.}} \widetilde{N}$. Then $\overline{X} \stackrel{d}{=} \mu + \sigma \widetilde{Z}_n/\sqrt{n}$. Then

$$\begin{aligned}
\sqrt{n}\left(\frac{g\left(\overline{X}\right)-g(\mu)}{\sigma g'(u)}\right) &\stackrel{d}{=} \sqrt{n}\left(\frac{g\left(\mu+\sigma \widetilde{Z}_n/\sqrt{n}\right)-g(\mu)}{\sigma g'(u)}\right)\\
&= \frac{g\left(\mu+\sigma \widetilde{Z}_n/\sqrt{n}\right)-g(\mu)}{\sigma \widetilde{Z}_n/\sqrt{n}} \frac{\widetilde{Z}_n}{g'(\mu)}\\
&\xrightarrow{\text{a.s }(m)} g'(\mu)\frac{\widetilde{N}}{g'(\mu)} = \widetilde{N} \stackrel{d}{=} N,
\end{aligned}$$

since $\widetilde{Z}_n/\sqrt{n} \to 0$ almost surely. This completes the proof. □

Theorem 8.30 (Portmanteau Theorem)**.** *Let $\{F_n\}_{n\geqslant 0}$ be a family of proper distributions. The following are equivalent.*

(i) $F_n \Longrightarrow F_0$.

(ii) For all $f: \mathbb{R} \to \mathbb{R}$ which are bounded and continuous, $\int f dF_n \to \int f dF_0$. Equivalently, if X_n is a r.v. with d.f. F_n for any $n \in \mathbb{N}$, then for f bounded and continuous, $\mathbb{E}\left[f(X_n)\right] \to \mathbb{E}\left[f(X_0)\right]$.

(iii) If $A \in \mathcal{B}(\mathbb{R})$ satisfies $F_0(\delta A) = 0$, then $F_n(A) \to F_0(A)$, where $\delta A = \overline{A} \setminus$ interior(A), and $\overline{A}$ is the intersection of all closed sets containing A and interior(A) is the union of all open sets contained in A.

Proof. (i)$\Longrightarrow$(ii) Suppose $f: \mathbb{R} \to \mathbb{R}$ is bounded and ocntinuous on $\mathbb{R}$. Since $X_n \Rightarrow X_0$, the continuous mapping theorem implies $f(X_n) \to f(X_0)$. Since f is bounded on $\mathbb{R}$, by DCT,

$$\lim_{n\to\infty} E[f(X_n)] = E\left[f(X_0)\right].$$

(ii)$\Longrightarrow$(i) Let $a, b \in \mathcal{C}(F_0)$, it suffices to show $F_n(a,b] \to F_0(a,b]$. Defined the bounded continuous function g_k whose graph is the trapezoid of height 1 obtained by takng a rectangle of height 1 with base $[a,b]$ and extending the base sysmmetrically to $[a-k^{-1}, b+k^{-1}]$. Then $g_k \downarrow \mathbb{1}_{[a,b]}$. For all $k \in \mathbb{N}$,

$$F_n(a,b] = \int_{\mathbb{R}} \mathbb{1}_{(a,b]} dF_n \leqslant \int_{\mathbb{R}} g_k dF_n \to \int_{\mathbb{R}} g_k dF_0.$$

Since $|g_k| \leqslant 1$ and $g_k \downarrow \mathbb{1}_{[a,b]}$, by DCT, $\int_{\mathbb{R}} g_k dF_0 \downarrow F_0([a,b]) = F_0((a,b])$, where the last equality follows since $a \in \mathcal{C}(F_0)$. We conclude that $\lim_{n\to\infty} \sup F_n(a,b] \leqslant F_0(a,b]$. Next, define new functions h_k whose graphs are trapezoids of height 1 obtained by taking a rectangle of height 1 with base $[a+k^{-1}, b-k^{-1}]$ and stretching the base symmetically to obtain $[a,b]$. Then $h_k \uparrow \mathbb{1}_{[a,b]}$ and for any $k \in \mathbb{N}$,

$$F_n(a,b] \geqslant \int_{\mathbb{R}} h_k dF_n \to \int_{\mathbb{R}} h_k dF_0,$$

By MCT, since $a \in \mathcal{C}(F_0)$, $\int_{\mathbb{R}} h_k dF_0 \uparrow F_0([a,b]) = F_0((a,b])$ so that $\lim_{n\to\infty} \inf F_n(a,b] \leqslant F_0(a,b]$. □

Remark. (ii) allows for the easy generalziation of the notion of weak convergence of random elements $\{\xi_n, n \geqslant 0\}$ whose range $\mathbb{S}$ is a subset of the metric space $\mathbb{R}^2$. The definition is $\xi_n \Rightarrow \xi_0$ if and only if $E(f(\xi_n)) \to E(f(\xi_0))$, for all test functions $f : \mathbb{S} \to \mathbb{R}$ which are bounded and continuous. (The notion of continuity is natural since $\mathbb{S}$ is a metric space.)

Example 8.31. Suppose F_n has atoms at i/n, $1 \leqslant i \leqslant n$ of size $1/n$. Let F_0 be the uniform distribution on $[0,1]$. Then $F_n \Rightarrow F_0$. It suffices to show integrals of arbitrary bounded continuous test functions converge. Let f be real valued, bounded and continuous with domain $[0,1]$. Observe that

$$\int f dF_n = \sum_{i=1}^{n} f(i/n)\frac{1}{n} = \text{ Riemman approximating sum} \to \int_0^1 f(x)dx \ (n\to\infty) = \int f dF_0,$$

where F_0 is the uniform distribution on $[0,1]$.

It is possible to restrict the test function in the portmanteau theorem to be uniformly continuous and not just continuous.

Corollary 8.32. TFAE:

(i) $F_n \Longrightarrow F_0$.

(ii) If X_n is a random variable with distribution F_n for any $n \in \mathbb{N}$, then for f bounded and uniformly continuous $Ef(X_n) \to Ef(X_0)$.

Proof. The the proof of (ii)$\Longrightarrow$(i) in the portmanteau theorem, the trapezoid functions are each bounded, continuous, vanish off a compact set, and are hence uniformly continuous. This observation suffices. □

8.7 More Relations Among Modes of Convergence

Proposition 8.33. Let $\{X, X_n,\ n \geqslant 1\}$ be random variables on the probability space $(\Omega, \mathcal{B}, P)$

(a) If $X_n \xrightarrow{\text{a.s.}} X$, then $X_n \xrightarrow{p} X$.

(b) If $X_n \xrightarrow{p} X$, then $X_n \Rightarrow X$.

Proof. We have shown (a). To verify (b), suppose $X_n \xrightarrow{p} X$ and f is bounded and continuous function. Then $f(X_n) \xrightarrow{p} f(X)$. By DCT, $E[f(X_n)] \to E(f(X))$. So $X_n \Rightarrow X$, by the portmanteau theorem. □

Proposition 8.34. If $X_n \Rightarrow a \in \mathbb{R}$, then $X_n \xrightarrow{p} a$.

Proof. Let F_n be the d.f.'s of X_n for any $n \in \mathbb{N}$. Fix $\epsilon > 0$,

$$\begin{aligned}
\lim_{n\to\infty} P\left(|X_n - a| > \epsilon\right) &= 1 - \lim_{n\to\infty} P(-\epsilon \leqslant X_n - a \leqslant \epsilon) \\
&= 1 - \lim_{n\to\infty} P(a-\epsilon \leqslant X_n \leqslant a+\epsilon) \\
&= 1 + \lim_{n\to\infty} P(X_n < a-\epsilon) - \lim_{n\to\infty} P(X_n \leqslant a+\epsilon) \\
&\leqslant 1 + \lim_{n\to\infty} P(X_n \leqslant a-\epsilon) - \lim_{n\to\infty} P(X_n \leqslant a+\epsilon) \\
&= 1 + \lim_{n\to\infty} F_n(a-\epsilon) - \lim_{n\to\infty} F_n(a+\epsilon) \\
&= 1 + F(a-\epsilon) - F(a+\epsilon) \\
&= 1 + 0 - 1 \\
&= 0,
\end{aligned}$$

since the constant function a is continuous on $\mathbb{R}$ and

$$F(x) = \begin{cases} 1, & x \geqslant a \\ 0, & x < a. \end{cases}$$

Alternative: Fix $\epsilon > 0$,

$$\begin{aligned}
\lim_{n\to\infty} P\left(|X_n - a| > \epsilon\right) &= \lim_{n\to\infty} P(X_n > a+\epsilon) + \lim_{n\to\infty} P(X_n < a-\epsilon) \\
&= \lim_{n\to\infty} P(X_n > a+\epsilon) + \lim_{n\to\infty} P(X_n < a-\epsilon) \\
&= 1 - F(a+\epsilon) + F(a-\epsilon) \\
&= 1 - 1 + 0 \\
&= 0.
\end{aligned}$$

□

Theorem 8.35 (Slutsky's Theorem)**.** *Suppose X, $\{X_n\}$ and $\{Y_n\}$ are all real-valued r.v's s.t. $X_n \Rightarrow X$ as $n \to \infty$, and $Y_n \xrightarrow{p} 0$ as $n \to 0$. Then $X_n + Y_n \Rightarrow X$.*

Proof. Fix a bounded and uniformly continuous $f : \mathbb{R} \to \mathbb{R}$. Fix $\epsilon > 0$, define

$$w_\epsilon(f) = \sup_{|x-y|\leqslant\epsilon} |f(x) - f(y)|.$$

Then since $X_n \Rightarrow X$,

$$\begin{aligned}
&\lim_{n\to\infty} |\mathbb{E}[f(X_n+Y_n)] - \mathbb{E}[f(X_n)]| \\
&\leqslant \lim_{n\to\infty} |\mathbb{E}[f(X_n+Y_n)] - \mathbb{E}[f(X_n)]| + \lim_{n\to\infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \\
&= \lim_{n\to\infty} \mathbb{E}\left(|f(X_n+Y_n) - f(X_n)|\mathbb{1}(|Y_n| \leqslant \epsilon)\right) + \lim_{n\to\infty} \mathbb{E}\left(|f(X_n+Y_n) - f(X_n)|\mathbb{1}(|Y_n| > \epsilon)\right) \\
&\leqslant \lim_{n\to\infty} E(w_\epsilon(f)) + 2M \lim_{n\to\infty} P\left(|Y_n| > \epsilon\right) \\
&= w_\epsilon(f) \to 0 \text{ as } \epsilon \downarrow 0.
\end{aligned}$$

Thus, by portmanteau theorem, $X_n + Y_n \Rightarrow X$. ☐

Remark. Since the slutsky's theorem follows from the portmanteau theorem, it allows for the easy generalziation of the notion of weak convergence of random vector $\{(X_n, Y_n), n \geqslant 0\}$ whose range $\mathbb{S}$ is a metric space.

Lemma 8.36. If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$, then $(X_n, Y_n) \xrightarrow{p}\!\!\to (X, Y)$.

Proof. Since $x, y, z, w \in (X, d)$,

$$d\left((x,y),(z,w)\right) = \left(d^2(x,z) + d^2(y,w)\right)^{\frac{1}{2}} = \left(d^2(x,z) + d^2(y,w)\right)^{\frac{1}{2}} \leqslant d(x,z) + d(y,w).$$

So

$$\begin{aligned}
P\left(d_2\left((X_n,Y_n),(X,Y)\right) \geqslant \epsilon\right) &\leqslant P\left(|X_n - X| + |Y_n - Y| \geqslant \epsilon\right) \\
&\leqslant P\left(|X_n - X| \geqslant \epsilon/2\right) + P\left(|Y_n - Y| \geqslant \epsilon/2\right) \\
&\to 0.
\end{aligned}$$

☐

Lemma 8.37. If $X_n \Rightarrow X$ and $Y_n \Rightarrow c$, where $c \in \mathbb{R}$ is a constant, then $(X_n, Y_n) \Rightarrow (X, c)$.

Proof. First we will show that $(X_n, c) \Rightarrow (X, c)$. By ther portmanteau theoerm, it is equivalent to show for any bounded and continuous function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, we have $E\left[f(X_n, c)\right] \to E\left[f(X, c)\right]$. So let f be such arbitrary function. Now consider the function of a single variable $g(x) = f(x, c)$, which is also bounded and continuous. By the portmanteau theorem, since $X_n \Rightarrow X$, $E\left[g(X_n)\right] \to E[g(X)]$. However the above expression is equivalent to $E\left[f(X_n, c)\right] \to E[f(X, c)]$. Hence $(X_n, c) \Rightarrow (X, c)$. Since $Y_n \Rightarrow c$, by Proposition 8.34, $Y_n \xrightarrow{p} c$. Similar to Lemma 8.36,

$$\begin{aligned}
P\left(d_2\left((X_n,Y_n),(X_n,c)\right) \geqslant \epsilon\right) &\leqslant P\left(|X_n - X_n| + |Y_n - c| \geqslant \epsilon\right) \\
&= P\left(|Y_n - c| \geqslant \epsilon\right) \\
&\to 0.
\end{aligned}$$

Thus, $(X_n, Y_n) \xrightarrow{p} (X_n, c)$. By the slutsky's theorem, $(X_n, Y_n) \Rightarrow (X, c)$. ☐

Corollary 8.38. If $X_n \Rightarrow X$, $Y_n \Rightarrow c$, where $c \in \mathbb{R}$ is a constant, then $X_n + Y_n \Rightarrow X + c$; $X_n Y_n \Rightarrow cX$; $X_n/Y_n \Rightarrow X/c$, provided that c is invertible.

Proof. By Lemma 8.37, we have $(X_n, Y_n) \Rightarrow (X_n, c)$. Since $+, \cdot, / : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ are continuous functions on the metric space $\mathbb{R}^2$, by the portmanteau theorem, we have the above conclusion. ☐

8.8 Convergence of types theorem

Definition 8.39. Two random variables X and Y are said to be of the *same type* if there exists $a > 0$ and $b \in \mathbb{R}$ such that

$$\frac{Y-b}{a} \stackrel{d}{=} X.$$

In terms of dfs, this is written as

$$F_X(x) = F_Y(ax+b).$$

Example 8.40. Suppose $X \sim N(\mu_1, \sigma^2)$ and $Y \sim N(\mu_2, \sigma^2)$. Since $\frac{X-\mu_1}{\sigma_1} \stackrel{d}{=} \frac{Y-\mu_2}{\sigma_2}$,

$$Y \stackrel{d}{=} \frac{\sigma_2}{\sigma_1} X - \left(\frac{\sigma_2}{\sigma_1}\mu_1 - \mu_2\right) = \frac{X - (\mu_1 - \sigma_1/\sigma_2 \mu_2)}{\sigma_1/\sigma_2}.$$

Then X and Y are of the same type.

Theorem 8.41. *Let U and V be df.'s, neither of which are degenerate.*

(a) Let $\{F_n\}_{n\geqslant 1}$ be dfs and $a_n, \alpha_n > 0$, $b_n, \beta_n \in \mathbb{R}$ and $F_n(a_n x + b_n) \to U(x)$, $F_n(\alpha_n x + \beta_n) \to V(x)$. Then

$$\frac{\alpha_n}{a_n} \to A > 0 \text{ and } \frac{\beta_n - b_n}{a_n} \to B \in \mathbb{R},$$

and $V(x) = U(Ax+B)$, i.e., U and V are the same type.

(a′) In terms of random variables $X_n, n = 1, \cdots, X$ and Y, $\frac{X_n - b_n}{a_n} \Rightarrow X$ and $\frac{X_n-\beta_n}{\alpha_n} \Rightarrow Y$, then

$$\frac{\alpha_n}{a_n} \to A > 0, \text{ and } \frac{\beta_n - b_n}{a_n} \to B \in \mathbb{R},$$

and $Y \stackrel{d}{=} \frac{X-B}{A}$.

(b) If $\frac{\alpha_n}{a_n} \to A > 0$ and $\frac{\beta_n - b_n}{a_n} \to B \in \mathbb{R}$ and either $F_n(a_n X + b_n) \Rightarrow U(x)$ or $F_n(\alpha_n x + \beta_n) \Rightarrow V(x)$, then does the other and so does $V(x) = U(Ax+B)$.

(b′) In terms of random variables $X_n, n = 1, \cdots, X$ and Y, if $\frac{\alpha_n}{a_n} \to A > 0$, and $\frac{\beta_n - b_n}{a_n} \to B \in \mathbb{R}$, and either

$$\frac{X_n - b_n}{a_n} \Rightarrow X \text{ or } \frac{X_n - \beta_n}{\alpha_n} \Rightarrow Y,$$

then does the other and so does $Y \stackrel{d}{=} \frac{X-B}{A}$.

Proof. Proof of (b′) by Baby Skorohod's Theorem. Suppose $Y_n = \frac{X_n - b_n}{a_n} \Rightarrow X$ and $\frac{\alpha_n}{a_n} \to A > 0$, and $\frac{\beta_n - b_n}{a_n} \to B \in \mathbb{R}$, then we need to show that $\frac{X_n - \beta_n}{\alpha_n} \Rightarrow \frac{X-B}{A}$. Use Baby Skorohod's Theorem, $\widetilde{X} \stackrel{d}{=} X$, $\widetilde{Y}_n \stackrel{d}{=} Y_n$ and $\widetilde{Y}_n \to \widetilde{X}$. Let $\widetilde{X}_n = a_n \widetilde{Y}_n + b_n$ for any $n \in \mathbb{N}$. Then $\widetilde{X}_n \stackrel{d}{=} a_n Y_n + b_n = X_n$. So

$$\begin{aligned} \frac{X_n - \beta_n}{\alpha_n} &\stackrel{d}{=} \frac{\widetilde{X}_n - \beta_n}{\alpha_n} = \frac{a_n}{\alpha_n}\left(\frac{\widetilde{X}_n - b_n + b_n - \beta_n}{a_n}\right) = \frac{a_n}{\alpha_n}\left(\widetilde{Y}_n + \frac{b_n - \beta_n}{a_n}\right) \\ &\to \frac{1}{A}\left(\widetilde{X} - B\right) \stackrel{d}{=} \frac{X-B}{A}. \end{aligned}$$

Thus, $\frac{X_n-\beta_n}{\alpha_n} \Rightarrow \frac{X-B}{A}$. Alternative: Suppose $\frac{X_n-b_n}{a_n} \Rightarrow X$, and $\frac{\alpha_n}{a_n} \to A$ and $\frac{\beta_n-b_n}{a_n} = B$. Then

$$\frac{X_n-\beta_n}{\alpha_n} = \frac{a_n}{\alpha_n}\frac{X_n-\beta_n}{a_n} = \frac{a_n}{\alpha_n}\left(\frac{X_n-b_n}{a_n} + \frac{b_n-\beta_n}{a_n}\right) \Rightarrow \frac{1}{A}(W-B),$$

by the corollary of slutsky's theorem.

Proof of (a) by inverse functions. Assume $F_n(a_nx+b_n) \to U(x)$, and $F_n(\alpha_nx+\beta_n) \to V(x)$. Claim.

$$\frac{\alpha_n}{a_n} \to A > 0, \text{ and } \frac{\beta_n-b_n}{a_n} \to B \in \mathbb{R},$$

and

$$V(x) = U(Ax+B).$$

Let G_n be the d.f. of $\frac{X_n-b_n}{a_n}$ for any $n \in \mathbb{N}$. Then

$$G_n(x) = P\left(\frac{X_n-b_n}{a_n} \leqslant x\right) = P(X_n \leqslant a_nX + b_n) = F_n(a_nx+b_n).$$

Since $G_n \to U$, for any $y \in (0,1) \cap C(U^{\leftarrow})$, $G_n^{\leftarrow}(y) \to U^{\leftarrow}(y)$. Then for any $y \in (0,1) \cap C(U^{\leftarrow})$,

$$\begin{aligned} F_n^{\leftarrow}(y) &= \inf\{x : F_n(x) \geqslant y\} \\ &= \inf\{x : F_n(a_nx+b_n) \geqslant y\} \\ &= \frac{\inf\{a_nx+b_n : F_n(a_nx+b_n) \geqslant y\} - b_n}{a_n} \ (b/c: \ a_n > 0) \\ &= \frac{F_n^{\leftarrow}(y) - b_n}{a_n}. \end{aligned}$$

Then for any $y \in (0,1) \cap C(U^{\leftarrow})$,

$$\frac{F_n^{\leftarrow}(y) - b_n}{a_n} \to U^{\leftarrow}(y) \tag{8.1}$$

Likewise, for any $y \in (0,1) \cap C(V^{\leftarrow})$,

$$\frac{F_n^{\leftarrow}(y) - b_n}{a_n} \to V^{\leftarrow}(y),$$

Choose $y_1 < y_2$, so they are in $C(U^{\leftarrow}) \cap C(V^{\leftarrow})$ and

$$-\infty < U^{\leftarrow}(y_1) \leqslant U^{\leftarrow}(y_2) < \infty \text{ and } -\infty < V^{\leftarrow}(y_1) < V^{\leftarrow}(y_2) < \infty.$$

(8.1) holds for both y_1, y_2, so

$$\frac{F_n^{\leftarrow}(y_2) - F_n^{\leftarrow}(y_1)}{a_n} \to U^{\leftarrow}(y_2) - U^{\leftarrow}(y_1) \geqslant 0.$$

Likewise,

$$\frac{F_n^{\leftarrow}(y_2) - F_n^{\leftarrow}(y_1)}{\alpha_n} \to V^{\leftarrow}(y_2) - V^{\leftarrow}(y_1) > 0.$$

Hence

$$\frac{\alpha_n}{a_n} \to \frac{U^{\leftarrow}(y_2) - U^{\leftarrow}(y_1)}{V^{\leftarrow}(y_2) - V^{\leftarrow}(y_1)} = A.$$

Moreover,

$$\begin{aligned}\frac{\beta_n - b_n}{a_n} &= \frac{F_n^{\leftarrow}(y_1) - b_n}{a_n} - \frac{F_n^{\leftarrow}(y_1) - \beta_n}{a_n} \\ &= \frac{F_n^{\leftarrow}(y_1) - b_n}{a_n} - \frac{\alpha_n}{a_n}\frac{F_n^{\leftarrow}(y_1) - \beta_n}{\alpha_n} \\ &\to U^{\leftarrow}(y_1) - AV^{\leftarrow}(y_1) \\ &= B.\end{aligned}$$

At last, use (b) to show they are of the same type. □

Given $X_1, X_2, \dots, X_n$ iid with d.f. F. Set

$$M_n = \max\{X_1, X_2, \dots, X_n\}.$$

Definition 8.42 (Max-**stable** Distribution)**.** A non-degenerate d.f. F on $\mathbb{R}$ is called max-stable, if for i.i.d. random variables $(X_i)_{i\in\mathbb{N}}$ with d.f. F and for any $n \in \mathbb{Z}^{>0}$, $\exists a_n > 0,\ b_n \in \mathbb{R}$ such that $\frac{M_n - b_n}{a_n}$ also has distribution F. Equivalently, M is max-stable if its d.f. F satisfies: for each $n \in \mathbb{Z}^{>0}$, there exists $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$F^n(a_n x + b_n) = F(x) \forall x \in \mathbb{R}.$$

Theorem 8.43 (Max-stable Distribution are Weak Limits of Maximal)**.** *A non-degenerate d.f. F on $\mathbb{R}$ is max-stable if and only if $\exists$ i.i.d. random variables $(X_i)_{i\in\mathbb{Z}^{>0}},\ a_n > 0,\ b_n \in \mathbb{R}$, such that the d.f. of $\frac{M_n - b_n}{a_n}$ converges to F.*

Proof. "$\Leftarrow$". Assume

$$Z_n := \frac{M_n - b_n}{a_n} \to Z,$$

for some $a_n > 0$ and $b_n \in \mathbb{R}$ and $\mathbb{Z}$ has d.f. F. Let $\left(Z^{(j)}\right)_{j\in\mathbb{Z}^{>0}}$ be i.i.d. copies of Z. For $j \in \mathbb{Z}^{>0}$, let $X^{(j)} := \left(X_i^{(j)}\right)_{i\in\mathbb{Z}^{>0}}$ be i.i.d. copies of the sequence $(X_i)_{i\in\mathbb{Z}^{>0}}$. Let

$$M_n^{(j)} := \max\left\{X_1^{(j)}, \dots, X_n^{(j)}\right\},$$

and denote

$$Z_n^j = \frac{M_n^{(j)} - b_n}{a_n}.$$

Then for each $m \in \mathbb{Z}^{>0}$,

$$\left(Z_n^{(1)}, \dots, Z_n^{(m)}\right) \Rightarrow \left(Z^{(1)}, \dots, Z^{(m)}\right).$$

Therefore, since "max" is continuous, by the Continuous Mapping Theorem for the weak convergence,

$$\max_{1\leqslant j\leqslant m} Z_n^{(j)} \Rightarrow \max_{1\leqslant j\leqslant m} Z^{(j)}.$$

On the other hand,

$$\begin{aligned}\max_{1\leqslant j\leqslant m} Z_n^{(j)} &= \max_{1\leqslant j\leqslant m} \frac{M_n^{(j)}-b_n}{a_n}\\ &\overset{d}{=} \frac{M_{mn}-b_n}{a_n}\\ &= c_{m,n}\left(\frac{M_{mn}-b_{mn}}{a_{mn}}\right)+d_{m,n}\\ &= c_{m,n}Z_{mn}+d_{m,n},\end{aligned}$$

where

$$M_{mn}=\max_{1\leqslant j\leqslant m} M_n^{(j)} \text{ and } mn = \arg max_{1\leqslant j\leqslant m} M_n^{(j)}$$

and

$$c_{m,n}=\frac{a_{mn}}{a_n} \text{ and } d_{mn}=\frac{b_{mn}-b_n}{a_n}.$$

Since

$$Z_{mn}\Rightarrow Z \text{ and } c_{m,n}Z_{mn}+d_{m,n}\Rightarrow \max_{1\leqslant j\leqslant m} Z^{(j)},$$

where both limits are non-degenerate, we can apply the Convergence of Types Theorem, to conclude that

$$c_{m,n}\to c_m>0 \text{ and } d_{m,n}\to d_m$$

and

$$Z\overset{d}{=}\frac{\max_{1\leqslant j\leqslant m} Z^{(j)}-d_m}{c_m},$$

and hence the distribution of Z is max-stable.

$\Longrightarrow$ It is obvious by def since for any $n\in\mathbb{Z}^{>0}$, $\exists a_n>0$ and $b_n\in\mathbb{R}$ such that

$$\frac{M_n-b_n}{a_n}\overset{d}{=}Z.$$

□

Theorem 8.44. *Suppose there exists $a_n>0$ and $b_n\in\mathbb{R}$ such that*

$$P\left(\frac{M_n-b_n}{a_n}\leqslant x\right)=F^n(a_nx+b_n)\overset{d}{\longrightarrow}G(x),$$

where $G(x)$ is proper and non-degenerate. Then G is of one of the following types.

(a)

$$G(x)=L(x)=e^{-e^{-x}},\ x\in\mathbb{R}.$$

(b) There is $\alpha>0$ such that

$$G(x)=\Phi_\alpha(x)=\begin{cases}0, & x<0\\ e^{-x^{-\alpha}}, & x>0\end{cases}.$$

(c) There is $\alpha > 0$ such that

$$G(x) = \Psi_\alpha(x) = \begin{cases} e^{-(-x)^\alpha}, & x < 0 \\ 1, & x \geqslant 0 \end{cases}.$$

G above is a max-stable distribution.

Remark.

Proof.

- Claim I: For all $t > 0$,

$$G^t(x) = G\left(\alpha(t)x + \beta(t)\right).$$

(Then G^t is of the same type as G.) Suppose $F^n(a_n x + b_n) \to G(x)$. Then

(a)
$$F^{\lfloor nt \rfloor}\left(a_{\lfloor nt \rfloor}x + b_{\lfloor nt \rfloor}\right) \to G(x),$$
where $\lfloor nt \rfloor \in \mathbb{Z}^{\geqslant 0}$.

(b)
$$F^{\lfloor nt \rfloor}\left(a_n x + b_n\right) = \left[F^n(a_n x + b_n)\right]^{\frac{\lfloor nt \rfloor}{n}} \to G^t(x).$$

Proof. $\forall\ \epsilon > 0$, $\exists N \in \mathbb{N}$, as $n > N$,

$$G(x) - \epsilon \leqslant F^n(a_n x + b_n) \leqslant G(x) + \epsilon.$$

Then
$$\left[G(x) - \epsilon\right]^{\frac{\lfloor nt \rfloor}{n}} \leqslant F^{\lfloor nt \rfloor}(a_n x + b_n) \leqslant \left[G(x) + \epsilon\right]^{\frac{\lfloor nt \rfloor}{n}}.$$

Since $\frac{\lfloor nt \rfloor}{n} \to t$, letting $n \to \infty$,

$$\left[G(x) - \epsilon\right]^t \leqslant \lim_{n\to\infty} F^{\lfloor nt \rfloor}(a_n x + b_n) \leqslant \left[G(x) + \epsilon\right]^t.$$

Let $\epsilon \downarrow 0$, then
$$\lim_{n\to\infty} F^{\lfloor nt \rfloor}(a_n x + b_n) = G^t(x).$$

Thus, G and G^t are of the same type since $G^{\frac{1}{\lfloor nt \rfloor}}$ and $G^{\frac{t}{\lfloor nt \rfloor}}$ are the same type and we can raise both sides to a power $\lfloor nt \rfloor$. Then since $G(t)$ and $G^t(x)$ are not degenerate, by the convergence of types theorem, there exist two functions $\alpha(t) > 0$ and $\beta(t)$, $t \geqslant 0$ such that for any $t > 0$,

$$\frac{a_n}{a_{\lfloor nt \rfloor}} \to \alpha(t) \text{ and } \frac{b_n - b_{\lfloor nt \rfloor}}{a_{\lfloor nt \rfloor}} \to \beta(t),$$

and also
$$G^t(x) = G\left(\alpha(t)x + \beta(t)\right),\ t > 0.$$

(Note that since a_n, b_n are constant and $a_{\lfloor nt \rfloor}$, $b_{\lfloor nt \rfloor}$ are step functions, $\alpha(t)$ and $\beta(t)$ are Lebesgue measurable.)

- Claim II. α and β are Lebesgue measurable function on $(0, \infty)$.

Proof. Define

$$\alpha_n(t) = \frac{a_n}{a_{\lfloor nt \rfloor}}, \forall t > 0.$$

Then

$$\alpha_n(t) = \sum_{k=0}^{\infty} \frac{a_n}{a_k} \mathbb{1}_{\{k \leqslant nt < k+1\}} = \sum_{k=0}^{\infty} \frac{a_n}{a_k} \mathbb{1}_{\{\frac{k}{n} \leqslant t < \frac{k+1}{n}\}}.$$

Since $\left\{\frac{k}{n} \leqslant t < \frac{k+1}{n}\right\}$ is a measurable set for any $k \in \mathbb{Z}$, we have $\frac{a_n}{a_k} \mathbb{1}_{\{k \leqslant nt < k+1\}}$ is a simple function for any $k \in \mathbb{Z}$. Then

$$\alpha_{n,m}(t) := \sum_{k=0}^{m} \frac{a_n}{a_k} \mathbb{1}_{\{k \leqslant nt < k+1\}} \to \alpha_n(t), \forall t > 0, \text{ as } m \to \infty.$$

Since the sum is convergent, α_n is measurable. Thus, α is measurable. Likewise, β is measurable. ☐

- Claim III. for any $s, t > 0$,

$$\alpha(st) = \alpha(s)\alpha(t),$$

and

$$\beta(st) = \alpha(t)\beta(s) + \beta(t) = \alpha(s)\beta(t) + \beta(s).$$

Since

$$G^{st}(x) = G(\alpha(st)x + \beta(st)),$$

and

$$\begin{aligned} G^{st}(x) &= \left(G^t(x)\right)^s \\ &= G^s(\alpha(t)x + \beta(t)) \\ &= G(\alpha(s)(\alpha(t)x + \beta(t)) + \beta(s)) \\ &= G\left(\alpha(s)\alpha(t)x + \alpha(s)\beta(t) + \beta(s)\right), \end{aligned}$$

also G is non-degenerate and $\alpha(t) > 0$ for any $t > 0$ we have

$$\alpha(st) = \alpha(s)\alpha(t),$$

and

$$\beta(st) = \alpha(s)\beta(t) + \beta(s),$$

where we used the fact that if F is a non-degenerate d.f. and

$$F(ax + b) = F(cx + d), \forall x \in \mathbb{R},$$

for some $a, c \in \mathbb{R}^+,\ b, d \in \mathbb{R}$, then $a = c$ and $b = d$.
Proof. Define

$$H_{a,b}(x) := H(ax + b), \forall x \in \mathbb{R},$$

$$H_{c,d}(x) := H(cx + d), \forall x \in \mathbb{R}.$$

$\forall\ y \in (0,1)$,

$$\begin{aligned} H^{\leftarrow}_{a,b}(y) &= \inf\{x \in \mathbb{R} | H_{a,b}(x) \geqslant y\} \\ &= \inf\{x \in \mathbb{R}, H(ax+b) \geqslant y\} \\ &= \frac{\inf\{ax+b \in \mathbb{R} | H(ax+b) \geqslant y\} - b}{a} \\ &= \frac{H^{\leftarrow}(y) - b}{a}. \end{aligned}$$

Likewise,

$$H^{\leftarrow}_{c,d}(y) = \frac{H^{\leftarrow}(y) - d}{c}.$$

Since

$$F(ax+b) = F(cx+d), \forall x \in \mathbb{R},$$

we have

$$H^{\leftarrow}_{a,b}(y) = H^{\leftarrow}_{c,d}(y).$$

So

$$H^{\leftarrow}(y)\left(\frac{1}{a} - \frac{1}{c}\right) = \frac{b}{a} - \frac{d}{c}, \forall t \in (0,1).$$

Thus,

$$a = c \text{ and } b = d.$$

The only measurable solution to the Hamel's equation

$$\alpha(st) = \alpha(s)\alpha(t)$$

is

$$\alpha(t) = t^{\theta}$$

for some $\theta \in \mathbb{R}$ since $\alpha : (0,\infty) \to \mathbb{R}$ is Lebesgue measurable.

(a) Assume $\theta = 0$. Then $\alpha(t) = 1$ for any $t > 0$, and then $\beta(st) = \alpha(t)\beta(t) + \beta(s) = \beta(s) + \beta(t)$. Define $\Omega(t) = e^{\beta(t)},\ t > 0$. Then $\Omega(st) = e^{\beta(st)} = e^{\beta(s)+\beta(t)} = \Omega(s)\Omega(t)$. Then $e^{\beta(t)} = \Omega(t) = t^c$, for some $c \in \mathbb{R}$. Thus, $\beta(t) = c\log t,\ t > 0$. Then $G^t(x) = G(\alpha(t)x + \beta(t)) = G(x + c\log(t)),\ t > 0$. Assume $c = 0$, then for $t > 0$, $G^t(x) = G(x)$, which implies for any $x \in \mathbb{R}$, $G(x) \in \{0,1\}$. Then clearly G is degenerate or not proper, a contradiction. Thus, $c \neq 0$. Suppose $\exists x_0 \in \mathbb{R}$ such that $G(x_0) = 0$. Then for any $t > 0$, $0 = G^t(x_0) = G(x_0 + c\log t)$. Setting $u = x_0 + c\log t$, since $\log t \in (-\infty,\infty)$ and $c \neq 0$, $G(u) = 0$ for any $u \in \mathbb{R}$, then G is not proper, which is a contradiction. Thus, $G(x) \neq 0$ for any $x \in \mathbb{R}$. Similarly, $G(x) \neq 1$ for any $x \in \mathbb{R}$. Therefore, $0 < G(x) < 1$ for any $x \in \mathbb{R}$. Then $G^t(x)$ is decreasing in t. Since $G^t(x) = G(x + c\log t)$ for any $t > 0$, $G(x + c\log t)$ is decreasing. Thus, $c < 0$. Let $x = 0$, then $G^t(0) = G(c\log t) \in (0,1)$. We can set $e^{-k} = G(0)$ for some $k \in \mathbb{R}^+$. Set $y = c\log t$, then $e^{\frac{y}{c}} = t$. Then

$$\begin{aligned} G(y) &= e^{-kt} \\ &= e^{-ke^{\frac{y}{c}}} \\ &= e^{-e^{-\left(\frac{1}{|c|}y - \log k\right)}} \\ &= L\left(\frac{1}{|c|}y - \log k\right). \end{aligned}$$

(b) Assume $\theta \neq 0$. Recall $\alpha(t) = t^\theta$. We claim

$$\beta(t) = c\left(1 - t^\theta\right),\ c \in \mathbb{R}.$$

Then

$$\beta(ts) = \alpha(t)\beta(s) + \beta(t) = \alpha(s)\beta(t) + \beta(s),$$

or

$$\beta(t)(1-\alpha(s)) = \beta(s)(1-\alpha(t)),$$

or

$$\frac{\beta(t)}{1-\alpha(t)} = \frac{\beta(s)}{1-\alpha(s)}, \forall s \neq 1, t \neq 1.$$

Thus, $\frac{\beta(t)}{1-\alpha(t)}$ is a constant function of t. Set $\frac{\beta(t)}{1-\alpha(t)} = c$ and then

$$\beta(t) = \frac{\beta(s)}{1-\alpha(s)}(1-\alpha(t)) = c\left(1-t^\theta\right).$$

Then

$$G^t(x) = G(\alpha(t)x + \beta(t) = G\left(t^\theta x + c\left(1 - t^{-\theta}\right)\right) = G\left(t^\theta(x-c) + c\right).$$

Let $y + c = x$, then $G^t(y+c) = G\left(t^{-\theta}y + c\right)$. Set $H(x) = G(x+c)$. (Note H and G are of the same type.) Then

$$H^t(x) = G^t(x+c) = G\left(t^\theta x + c\right) = H\left(t^\theta x\right), \forall t > 0.$$

- Assume $\theta < 0$. Claim. $H(0) = 0$. Set $x = 0$, then $H^t(0) = H(0)$ for any $t > 0$. So $H(0) \in \{0, 1\}$. Suppose $H(0) = 1$. Consider $x < 0$, since H is not degenerate, $\exists x_0 < 0$ such that $0 < H(x_0) < 1$, and $H^t(x_0) = H\left(t^\theta x_0\right)$. (If $H(x) = 0$ for any $x < 0$, then H is degenerate; if $H(x) = 1$ for any $x < 0$, then H is not proper.) For all $t > 0$, the LHS is strictly decreasing in t and the RHS is nondecreasing in t since $x_0 < 0$, a contradiction. Thus, $H(0) \neq 1$, then $H(0) = 0$. Therefore, $H(x) = 0,\ x \leqslant 0$. Next, we can consider $x > 0$. Suppose $H(x_1) = 0$ for some $x_1 > 0$, then $0 = H^t(x_1) = H\left(t^\theta x_1\right)$ for any $t > 0$. Thus, $H \equiv 0$, contradiction. Suppose $H(x_2) = 1$ for some $x_2 > 0$, then $1 = H^t(x_2) = H\left(t^\theta x_2\right)$ for any $t > 0$. Then

  $$H(t) = \begin{cases} 0, & \text{if } t \leqslant 0 \\ 1, & \text{if } t > 0 \end{cases},$$

  so H is degenerate, a contradiction. Hence $H(t) \in (0,1)$ for any $t > 0$. Let $x = 1$, then $H^t(1) = H(t^\theta) \in (0,1)$. We can set $e^{-k} = H(1)$ for some $k \in \mathbb{R}^+$. Set $y = t^\theta$, then $y^{\frac{1}{\theta}} = t$. Then

  $$H(y) = e^{-kt} = e^{-ky^{\frac{1}{\theta}}} = e^{-ky^{-\frac{1}{|\theta|}}} = e^{-(k^{-|\theta|}y)^{-\frac{1}{|\theta|}}}.$$

  Besides, H and G are of the same type.

- Assume $\theta > 0$. Claim $H(0) = 1$. Set $x = 0$, then $H^t(0) = H(0)$ for any $t > 0$. So $H(0) \in \{0, 1\}$. Suppose $H(0) = 0$. Consider $x > 0$, since H is not degenerate, $\exists x_0 < 0$ such that $0 < H(x_0) < 1$, and $H^t(x_0) = H\left(t^\theta x_0\right)$. (If $H(x) = 0$ for any $x > 0$, then H is not proper; if $H(x) = 1$ for any $x > 0$, then H is degenerate.) For all $t > 0$, the lhs is strictly decreasing in t and the rhs is nondecreasing in t since $x_0 > 0$, a contradiction. Thus, $H(0) \neq 0$, then $H(0) = 1$. Therefore, $H(x) = 1,\ x \geqslant 0$. Next, we can consider $x < 0$. Suppose $H(x_1) = 0$ for some $x_1 < 0$, then $0 = H^t(x_1) = H\left(t^\theta x_1\right)$ for all $t > 0$. Then

$$H(t) = \begin{cases} 0, & \text{if } t \leqslant 0 \\ 1, & \text{if } t > 0 \end{cases},$$

so H is degenerate, a contradiction. $0 = H^t(x_1) = H\left(t^\theta x_1\right)$ for any $t > 0$. Suppose $H(x_2) = 1$ for some $x_2 < 0$, then $1 = H^t(x_2) = H\left(t^\theta x_2\right)$ for any $t > 0$. Thus, $H \equiv 1$, contradiction. Hence $H(t) \in (0,1)$ for any $t > 0$. Let $x = -1$, then $H^t(-1) = H(-t^\theta) \in (0,1)$. We can set $e^{-k} = H(-1)$ for some $k \in \mathbb{R}^+$. Set $y = -t^\theta$, then $(-y)^{\frac{1}{\theta}} = t$. Then

$$H(y) = e^{-kt} = e^{-k(-y)^{\frac{1}{\theta}}} = e^{-(-k^\theta y)^{\frac{1}{\theta}}}.$$

Besides, H and G are of the same type. $\square$

Lemma 8.45. If $a_n \sim b_n$, then

$$\frac{a_n}{b_n} \to 1.$$

If $b_n \to b$ and $a_n \sim b_n$. Then

$$a \leftarrow a_n = \frac{a_n}{b_n} \cdot b_n.$$

Theorem 8.46. *Let $X_1, X_2, \cdots$ be iid and $0 \leqslant \tau \leqslant \infty$. Suppose $\{u_n\}$ is a sequence of real numbers (Think of $u_n = a_n x + b_n$) such that*

$$n(1 - F(u_n)) \to \tau. \tag{8.2}$$

Then

$$P(M_n \leqslant u_n) \to e^{-\tau}. \tag{8.3}$$

Conversely, (8.3) *holds, so does* (8.2).

Proof. Assume $0 \leqslant \tau \leqslant \infty$ and (8.2) holds. Then $1 - F(u_n) = \frac{\tau}{n} + 0(1)$. So

$$\begin{aligned} P(M_n \leqslant u_n) &= \prod_{k=1}^n P(X_k \leqslant u_n) \\ &= F^n(u_n) \\ &= [1 - (1 - F(u_n))]^n \\ &= \left[1 - \left(\frac{\tau}{n} + 0(1)\right)\right]^n \\ &\to e^{-\tau} \text{ as } n \to \infty. \end{aligned}$$

Suppose $0 \leqslant \tau < \infty$ and (8.3) holds. Claim. $1 - F(u_n) \to 0$. Suppose there exists a subsequence $\{u_{n_k}\}_{k \geqslant 1}$ so that $1 - F(u_k)$ is bounded away from 0. Then as $k \to \infty$,

$$P(M_{n_k} \leqslant u_{n_k}) = [1 - (1 - F(u_{n_k}))]^{n_k} \to 0 = e^{-\infty},$$

since $0 < 1 - F(u_{n_k}) < 1$, which means $\tau = \infty$, a contradiction. We know as $x \to 0$, $-\log(1-x) \sim x$. Since 8.3 holds and by Lemma 8.45,

$$\tau \leftarrow -\log P(M_n \leqslant u_n) = -n \log(1 - (1 - F(u_n))) \sim n(1 - F(u_n)).$$

Let $\tau = \infty$ and suppose (8.3) holds, but (8.2) does not hold. If $n(1-F(u_n))$ does not converges to infinite, then there exists a subsequence $\{u_{n_k}\}$ such that

$$n_k(1-F(u_{n_k})) \to C,$$

where $0 \leqslant C < \infty$. By the proof of the finite case of τ, similarly, we have

$$P(M_{n_k} \leqslant u_k) \to e^{-C} > 0 = e^{-\infty},$$

which is a contradiction. □

Example 8.47. Suppose $X_1, X_2, \cdots$ iid and $X_1 \sim N(0,1)$. Let

$$M_n = \max\{X_1, \ldots, X_n\}.$$

Then

$$P\left(\frac{M_n - b_n}{a_n}\right) \to L(x) = e^{-e^{-x}},$$

where $a_n = (2\log n)^{-\frac{1}{2}}$ and

$$b_n = (2\log n)^{\frac{1}{2}} - \frac{1}{2}(2\log n)^{-\frac{1}{2}}(\log\log n + \log(4\pi)).$$

The proof is to show

$$n(1-\Phi(u_n)) \to e^{-x},$$

where $u_n = a_n x + b_n$. Let $\varphi(x) = \frac{1}{2\pi}e^{-\frac{x^2}{2}}$. We claim

$$1-\Phi(u_n) \sim \frac{\varphi(u_n)}{u_n}.$$

On one hand,

$$\begin{aligned}
1-\Phi(x) &= \int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\,dt \\
&< \frac{1}{x}\frac{1}{\sqrt{2\pi}}\int_x^\infty te^{-\frac{t^2}{2}}\,dt \\
&= \frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \\
&= \frac{\varphi(x)}{x}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\,dt &= \int_x^\infty \frac{1}{\sqrt{2\pi}}\frac{1}{t}d\left(-e^{-\frac{t^2}{2}}\right) \\
&= \frac{\varphi(x)}{x} - \int_x^\infty \frac{1}{t^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}.
\end{aligned}$$

Then

$$\begin{aligned}\frac{\varphi(x)}{x} &= \int_x^\infty \frac{1}{\sqrt{2\pi}}\left(1+\frac{1}{t^2}\right)e^{-\frac{t^2}{2}}\\ &< \int_x^\infty \frac{1}{\sqrt{2\pi}}\left(1+\frac{1}{x^2}\right)e^{-\frac{t^2}{2}}\\ &= \frac{x^2+1}{x^2}(1-\Phi(x)).\end{aligned}$$

So

$$\frac{x\varphi(x)}{x^2+1} < 1-\Phi(x) < \frac{\varphi(x)}{x}.$$

Since

$$\lim_{x\to\infty}\frac{\frac{\varphi(x)}{x}}{\frac{x\varphi(x)}{x^2+1}} = \lim_{x\to\infty}\frac{x^2+1}{x^2} = 1,$$

$$1-\Phi(x) \sim \frac{\varphi(x)}{x}.$$

Want to choose u_n so that $n(1-\Phi(u_n)) = e^{-x}$, but we replace $1-\Phi(u_n)$ with $\frac{\varphi(u_n)}{u_n}$. Then

$$\frac{1}{n}e^{-x} = 1-\Phi(u_n) \sim \frac{\varphi(u_n)}{u_n} = \frac{1}{\sqrt{2\pi}}\frac{e^{-\frac{u_n^2}{2}}}{u_n}.$$

So

$$\frac{\sqrt{2\pi}u_n}{n}\frac{e^{-x}}{e^{-\frac{u_n^2}{2}}} \to 1 \text{ as } n\to\infty.$$

Then

$$\frac{1}{2}\log(2\pi) + \log u_n - x - \log n + \frac{u_n^2}{2} \to 0 \text{ as } n\to\infty.$$

Since $\frac{u_n^2}{2}$ dominates $\log u_n$,

$$\frac{u_n^2}{2} \to -\frac{1}{2}\log(2\pi) + x + \log n.$$

So

$$\frac{u_n^2}{2\log n} \to 1.$$

Then

$$\log u_n = \frac{1}{2}(\log\log n + \log 2) + o(1).$$

Thus,

$$\begin{aligned}u_n^2 &= -\log(2\pi) - (\log\log n + \log 2) + 2x + 2\log n + o(1).\\ &= -\log(4\pi) - \log\log n + 2x + 2\log n + o(1).\\ &= 2\log n\left[1 + \frac{x - \frac{1}{2}(\log\log n + \log(4\pi))}{\log n} + o\left(\frac{1}{2\log n}\right)\right].\end{aligned}$$

Since $\sqrt{1+y} = 1 + \frac{y}{2} + o(y)$,

$$\begin{aligned} u_n &= (2\log n)^{\frac{1}{2}} \left[1 + \frac{x - \frac{1}{2}(\log\log n + \log(4\pi))}{2\log n} + o\left(\frac{1}{2\log n}\right)\right]. \\ &= (2\log n)^{-\frac{1}{2}}x + (2\log n)^{\frac{1}{2}} - \frac{1}{2}(2\log n)^{-\frac{1}{2}}(\log\log n + \log(4\pi)) + o\left((2\log n)^{-\frac{1}{2}}\right). \end{aligned}$$

Take $a_n = (2\log n)^{-\frac{1}{2}}$ and

$$b_n = (2\log n)^{\frac{1}{2}} - \frac{1}{2}(2\log n)^{-\frac{1}{2}}(\log\log n + \log(4\pi)),$$

we have

$$u_n = a_n x + b_n + o(a_n).$$

Take $\alpha_n = a_n$ and $\beta_n = b_n + o(a_n)$. Then

$$P\left(\frac{M_n - \beta_n}{a_n} \leqslant x\right) = P(M_n \leqslant a_n x + \beta_n) \to e^{e^{-x}}.$$

Since $\frac{\alpha_n}{a_n} = 1 > 0$ and $\frac{\beta_n - a_n}{a_n} = 0 \in \mathbb{R}$, by convergence of types theorem,

$$P\left(\frac{M_n - b_n}{a_n} \leqslant x\right) = P(M_n \leqslant a_n x + b_n) \to e^{e^{-x}}.$$

Given a sequence of $X_1, X_2, \cdots$ of r.v.'s that are $N(0,1)$. Assume the sequence is stationary.

Definition 8.48. A sequence $X_1, X_2, \cdots$ is said to be (strictly) stationary if for all positive integers n and k,

$$X_1, \ldots, X_k \stackrel{d}{=} X_{n+1}, \ldots, X_{n+k}.$$

Aside: Suppose the sequence forms a Markov chain with transition matrix P. Suppose π is a prob. vector that satisfies $\pi = \pi P$. Let π be the initial dist.
Then

$$P(X_0 = i_0, \ldots, X_k = i_k) = \pi_{i_0} P_{i_0 i_1} P_{i_1 i_2} \cdots P_{i_{k-1} i_k}.$$

Also,

$$P(X_n = i_0, \ldots, X_{n+k} = i_k) = P(X_n = i_0) P_{i_0 i_1} \cdots P_{i_{k-1} i_k}.$$

But

$$\begin{aligned} P(X_n = i_0) &= \sum_{j \in E} P(X_0 = j, X_n = i_0) \\ &= \sum_{j \in E} P(X_0 = j) P(X_n = i_0 | X_0 = j) \\ &= \sum_{j \in E} \pi_j P^n_{j i_0} \\ &= \pi_{i_0}, \end{aligned}$$

since $\pi = \pi P$ implies

$$\pi P^n = \pi, \forall n \in \mathbb{Z}^+.$$

Thus, if $\exists \pi$ with $\pi_j > 0$ for any $j \in E$, satisfying $\pi = \pi P$, let $X_0 \stackrel{d}{=} \pi$, then the sequence $X_1, X_2, \cdots$ in a Markov chain is stationary.

Definition 8.49. A sequence $(X_1, \ldots, X_n)$ is multivariate normal if for **every** $\alpha = (\alpha_1, \ldots, \alpha_n)^T \in \mathbb{R}^n$, the r.v. $\sum_{i=1}^n \alpha_i X_i$ is normally distributed.
Then (alternative)

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^n \alpha_i X_i\right) &= E\left[\sum_{j=1}^n \alpha_j X_j \sum_{k=1}^n \alpha_k X_k\right] - E\left[\sum_{j=1}^n \alpha_j X_j\right] E\left[\sum_{k=1}^n \alpha_k X_k\right] \\
&= \sum_j \sum_k \alpha_j E[X_j X_k]\alpha_k - \sum_j \sum_k \alpha_j E[X_j]E[X_k]\alpha_k \\
&= \sum_j \sum_k \alpha_j (E[X_j X_k] - E[X_j]E[X_k])\alpha_k \\
&= \sum_j \sum_k \alpha_j \mathrm{Cov}(X_j, X_k)\alpha_k \\
&= \alpha^T \Sigma \alpha \geqslant 0, \forall \alpha \in \mathbb{R}^n,
\end{aligned}$$

where $\Sigma_{ij} = \mathrm{Cov}(X_i, X_j)$. Σ is symmetric and nonnegative definite. Provided the covariance matrix Σ is nonsingular, the random vector $\vec{X} = (X_1, \ldots, X_n)$ has a **joint** (Gaussian) pdf given by

$$f_{\vec{X}}(x) = (2\pi)^{-\frac{n}{2}} (\det\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})'\Sigma^{-1}(\vec{x} - \vec{\mu})\right).$$

Remark. It is obvious that $T_{(a_1,\ldots,a_n)}$ is in the dual space of $\mathbb{R}^n$.

Definition 8.50. A sequence $X_1, X_2, \cdots$ is said to be a *Gaussian random sequence* (Gaussian process) if for $n \in \mathbb{Z}^+$ and $(k_1, \ldots, k_n)^T \in (\mathbb{Z}^n)^+$, $(X_{k_1}, \ldots, X_{k_n})$ is multivariate normal.
Hence a Gaussian process depend only on

(a)
$$m(i) = E[X_i],$$

(b)
$$\gamma(i, j) = \mathrm{Cov}(X_i, X_j).$$

Definition 8.51. A Gaussian process $\{X_i\}_{i \in \mathbb{Z}^+}$ is stationary if

(a)
$$m(i) = E[X_i] = \mu, \tag{8.4}$$

(b)
$$\gamma(i, j) = \mathrm{Cov}(X_i, X_j). \tag{8.5}$$

only depends on the difference $j - i$, so denote it as

$$\gamma(i, j) = \gamma_{|j-i|},$$

where $\gamma.$ is autocovariance function. Or another explanation is

$$(X_i, X_j) \stackrel{d}{=} (X_0, X_{j-i}), \forall i \leqslant j,\ i, j \in \mathbb{Z}^+.$$

As a result, the (joint) distribution of a stationary Gaussian process is only determined by u and γ_j for any $j \in \mathbb{Z}$, since the (joint) pdf is just related to μ and Σ, where $\Sigma_{ij} = \gamma_{|i-j|}$.

Remark. A general stochastic process $\{X_i\}_{i\in\mathbb{Z}^+}$ satisfying conditions (8.4) and (8.5) is said to be weakly or second-order stationary. The first-order and second-order moments of weakly stationary processes are invariant with respect to time translations. For Gaussian time series, the concepts of weak and strict stationarity coalesce.

Let $M_n = \max\{X_1, \dots, X_n\}$, where $\{X_i\}_{i\in\mathbb{Z}^+}$ is some sequence. Are there constants a_n and b_n such that

$$\frac{M_n - b_n}{a_n} \to \gamma \text{ (Gumble)?}$$

$$\vdots$$

Given is a regenerative process on the nonnegative integers, i.e $\{X_n, n = 0, 1, 2\cdots\}$. There is a sequence of r.v's

$$0 \leqslant \tau_0 \leqslant \tau_1 < \tau_2 \cdots$$

that forms a delayed renewal process. Then we have

(a) $X_{\tau_n}, X_{\tau_n+1}\cdots$ is independent of τ_n, and $X_0, \dots, X_{\tau_n-1}$.

(b) For $n = 1, 2, \cdots$, $X_{\tau_n}, X_{\tau_n+1}, \cdots$ has the same distribution as $X_{\tau_0}, X_{\tau_0+1}, \cdots$.

Example 8.52. Let $\{X_n, n = 0, 1, \cdots\}$ be a Markov chain with state space $\{0, 1, \cdots\}$. Let $\tau_0 = \inf\{n \geqslant 0 | X_n = 12\}$. For $k = 1, 2, \cdots$,

$$\tau_k = \inf\{n > \tau_{k-1} : X_n = 12\}.$$

τ_k is the time of the k^{th} return of the MC to state 12. Note that $\{\tau_k, k = 0, 1, 2, \cdots\}$ forms a *delayed renewal process*, and w.r.t. that renewal process, $\{X_n, n = 0, 1, \cdots\}$ is a *regenerative process*. Thus,

$$(X_0, \dots, X_{\tau_1-1}),\ (X_{\tau_1}, \dots, X_{\tau_2-1}),\ \cdots$$

are iid random elements. Intuitively, it means a regenerative process can be split into i.i.d. cycles. Take for $k \in \mathbb{Z}^+$,

$$Y_k = \max\left(X_{\tau_{k-1}}, X_{\tau_{k-1}+1}, \dots, X_{\tau_k-1}\right).$$

Suppose the renewal sequence $\{\tau_n, n = 0, 1, 2\cdots\}$ is recurrent, and aperiodic. Assume $\tau_0 = 0$ with probability 1. Find a renewal equation for $P(X_t \in B)$.

$$P(X_t \in B) = P(X_t \in B, \tau_1 \leqslant t) + P(X_t \in B, \tau_1 > t).$$

Note

$$\begin{aligned}P(X_t \in B, \tau_1 \leqslant t) &= \sum_{k=1}^{t} P(X_t \in B, \tau_1 = k)\\ &= \sum_{k=1}^{t} P(X_{t-k+\tau_1} \in B, \tau_1 = k)\\ &= \sum_{k=1}^{t} P(X_{t-k+\tau_1} \in B) P(\tau_1 = k)\\ &= \sum_{k=1}^{t} P(X_{t-k} \in B) P(\tau_1 = k),\end{aligned}$$

since $-k+\tau_1 = 0$ and $X_{t-k+\tau_1}$ is independent of $(\tau_1 - \tau_0) = \tau_1$ and has the same distribution as X_{t-k}. Thus,

$$P(X_t \in B) = P(X_t \in B, \tau_1 > t) + \sum_{k=1}^{t} P(X_{t-k} \in B) P(\tau_1 = k).$$

If one set $h(t) = P(X_k \in B)$ and $g(t) = P(X_t \in B, \tau_1 > t)$, get the renewal equation

$$h(t) = g(t) + \sum_{k=1}^{t} P(\tau_1 = k) h(t-k).$$

Thus,

$$\lim_{t\to\infty} h(t) = \frac{\sum_{t=0}^{\infty} g(t)}{E[\tau_1]}.$$

Note

$$\begin{aligned}\sum_{t=0}^{\infty} g(t) &= \sum_{t=0}^{\infty} P(X_t \in B, \tau_1 > t)\\ &= \sum_{t=0}^{\infty} E\left[\mathbb{1}_B(X_t)\mathbb{1}(\tau_1 > t)\right]\\ &= E\left[\sum_{t=0}^{\infty} \mathbb{1}_B(X_t)\mathbb{1}(\tau_1 > t)\right]\\ &= E\left[\sum_{t=0}^{\tau_1-1} \mathbb{1}_B(X_t)\right]\end{aligned}$$

Note $\sum_{t=0}^{\tau_1-1} \mathbb{1}_B(X_t)$ is the number of times the regenerative process visits the set B during the first cycle. Thus,

$$\lim_{t\to\infty} P(X_t \in B) = \frac{E\left[\sum_{t=0}^{\tau_1-1} \mathbb{1}_B(X_t)\right]}{E[\tau_1]}.$$

It equals to expected number of visits the regenerative process makes to set B in the first cycle divided by the expected cycle time. Set X to be a r.v. with d.f. F_X give by

$$F_X(x) = \frac{E\left[\sum_{t=0}^{\tau_1-1} \mathbb{1}_{(-\infty,x]}(X_k)\right]}{E[\tau_1]},$$

where F_X is the limiting distribution.

Proposition 8.53. Continuing the last example, where the sequence of iid copies of $\{Y_k\}_{k\geqslant 1}$ are the maximums of those iid cycles, suppose there exists a_n and b_n such that

$$[F_X(a_n x + b_n)]^n \to G(x),$$

where $G(x)$ is the extreme value distribution.
Set

$$A_{n,k} = \{X_k > a_n x + b_n\}.$$

Consider the condition

$$\lim_{n\to\infty} n \sum_{0\leqslant i<k<\tau_1} P\left(A_{n,i} \cap A_{n,k}\right) = 0. \tag{8.6}$$

Then

$$\lim_{n\to\infty} P\left(\max_{1\leqslant k\leqslant n} Y_k \leqslant a_n x + b_n\right) = [G(x)]^u,$$

where $u = E[\tau_1]$.

Proof. It suffices to show

$$\lim_{n\to\infty} nP\left(Y_1 > a_n x + b_n\right) \to -u \log G(x),$$

by the theorem 8.46. It is obvious that

$$\{Y_1 > a_n x + b_n\} = \bigcup_{k=0}^{\tau_1-1} \{X_k > a_n x + b_n\} = \bigcup_{k=0}^{\tau_1-1} A_{n,k}.$$

By the inclusion-exclusion formula, we have

$$E\left[\sum_{k=0}^{\tau_1-1} \mathbb{1}_{A_{n,k}}\right] - E\left[\sum_{0\leqslant i<k<\tau_1} \mathbb{1}_{A_{n,i}\cap A_{n,k}}\right] \leqslant P\left(\bigcup_{k=0}^{\tau_1-1} A_{n,k}\right) \leqslant E\left[\sum_{k=0}^{\tau_1-1} \mathbb{1}_{A_{n,k}}\right].$$

Using condition (8.6),

$$\lim_{n\to\infty} nE\left[\sum_{k=0}^{\tau_1-1} \mathbb{1}_{A_{n,k}}\right] \leqslant \lim_{n\to\infty} nP\left(\bigcup_{k=0}^{\tau_1-1} A_{n,k}\right) \leqslant \lim_{n\to\infty} nE\left[\sum_{k=0}^{\tau_1-1} \mathbb{1}_{A_{n,k}}\right].$$

Since $\bigcup_{k=0}^{\tau_1-1} A_{n,k} = \{Y_1 > a_n x + b_n\}$,

$$\lim_{n\to\infty} nP\left(Y_1 > a_n x + b_n\right) = \lim_{n\to\infty} nE\left[\sum_{k=0}^{\tau_1-1} \mathbb{1}_{A_{n,k}}\right].$$

Divide by $E[\tau_1]$, we obtain

$$\frac{1}{E[\tau_1]} \lim_{n\to\infty} P\left(Y_1 > a_n x + b_n\right) = \lim_{n\to\infty} n(1 - F_X(a_n x + b_n)),$$

since

$$\frac{E\left[\sum_{k=0}^{\tau_1 - 1} \mathbb{1}_{A_{n,k}}\right]}{E[\tau_1]} = 1 - F_X(a_n x + b_n).$$

Since

$$[1 - (1 - F_X(a_n x + b_n))]^n = [F_X(a_n x + b_n)]^n \to G(x),$$

we have

$$\lim_{n\to\infty} n(1 - F_X(a_n x + b_n)) = -\log G(x).$$

Thus,

$$\lim_{n\to\infty} nP\left(Y_1 > a_n x + b_n\right) \to -E[\tau_1] \log G(x). \qquad \square$$

8.8.1 Maximum process

Let $X_1, X_2, \cdots$ be iid r.v.'s with d.f. F such that there are continuous a_n and b_n for which

$$F^n(a_n x + b_n) \to G(x).$$

Set $M_n = \max(X_1, \ldots, X_n)$, what types of process is $\{M_1, M_2, \cdots\}$?

Calculate for $t_1 < t_2$ and $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_{t_1}$,

$$P\left(M_{t_2} \leqslant x_{t_2} | M_1 = x_1, \ldots, M_{t_1} = x_{t_1}\right).$$

If $x_{t_2} < x_{t_1}$, then $M_{t_2} \leqslant x_{t_2} < x_{t_1} = M_{t_1}$, which has prob. 0.

If $x_{t_2} \geqslant x_{t_1}$, then given $M_{t_1} = x_{t_1} \leqslant x_{t_2}$,

$$M_{t_2} \leqslant x_{t_2} \text{ if and only if } \left(X_{t_1+1} \leqslant X_{t_2}, \ldots, X_{t_2} \leqslant X_{t_2}\right).$$

Since $X_{t_1+1}, \ldots, X_{t_2}$ are independent of $M_1, \ldots, M_{t_1}$,

$$P\left(M_{t_2} \leqslant x_{t_2} | M_1 = x_1, \ldots, M_{t_1} = x_{t_1}\right) = P\left(X_{t_1+1} \leqslant x_{t_2}, \ldots, X_{t_2} \leqslant x_{t_2}\right) = F^{t_2 - t_1}(x_{t_2}).$$

Thus,

$$P\left(M_{t_2} \leqslant x_{t_2} | M_1 = x_1, \ldots, M_{t_1} = x_{t_1}\right) = \begin{cases} 0, & \text{if } x_{t_2} < x_{t_1}, \\ F^{t_2 - t_1}(x_{t_2}), & \text{otherwise.} \end{cases}$$

Hence

$$P\left(M_{t_2} \leqslant x \middle| M_1, \ldots, M_{t_1}\right) = \mathbb{1}_{[0,x]}(M_{t_1}) F^{t_2 - t_1}(x).$$

Therefore,

$$\begin{aligned} P(M_{t_1} \leqslant x_1, M_{t_2} \leqslant x_2) &= E\left[\mathbb{1}_{\{M_{t_1} \leqslant x_1\} \cap \{M_{t_2} \leqslant x_2\}}\right] \\ &= E\left[\mathbb{1}_{[0,x_1]}(M_{t_1})\mathbb{1}_{[0,x_2]}(M_{t_2})\right] \\ &= E\left[E\left[\mathbb{1}_{[0,x_1]}(M_{t_1})\mathbb{1}_{[0,x_2]}(M_{t_2})\middle| M_{t_1}\right]\right] \\ &= E\left[\mathbb{1}_{[0,x_1]}(M_{t_1})E\left[\mathbb{1}_{[0,x_2]}(M_{t_2})\middle| M_{t_1}\right]\right] \\ &= E\left[\mathbb{1}_{[0,x_1]}(M_{t_1})\mathbb{1}_{[0,x_2]}(M_{t_1})F^{t_2-t_1}(x_2)\right] \\ &= E\left[\mathbb{1}_{[0,x_1\wedge x_2]}(M_{t_1})F^{t_2-t_1}(x_2)\right] \\ &= P\left(M_{t_1} \leqslant x_1 \wedge x_2\right)F^{t_2-t_1}(x_2) \\ &= F^{t_1}(x_1 \wedge x_2)F^{t_2-t_1}(x_2). \end{aligned}$$

For $t_1 < t_2 < \cdots < t_n$,

$$P(M_{t_1} \leqslant x_1, M_{t_2} \leqslant x_2, \ldots, M_{t_n} \leqslant x_n) = F^{t_1}(x_1 \wedge \cdots \wedge x_n)F^{t_2-t_1}(x_2 \wedge \cdots \wedge x_n)\cdots F^{t_n-t_{n-1}}(x_n).$$

8.8.2 Extremal process

An extremal process $\{U_t; t \geqslant 0\}$ is a **continuous** time Markov chain such that

$$P(U_{t_1} \leqslant x_1, Ut_2 \leqslant x_2, \cdots, U_{t_n} \leqslant x_n) = G^{t_1}(x_1 \wedge \cdots \wedge x_n)G^{t_2-t_1}(x_2 \wedge \cdots \wedge x_n)\cdots G^{t_n-t_{n-1}}(x_n),$$

where G is any extreme value d.f. Set $Q(x) = -\log G(x)$. The holding time in a state will be exponential with rate $Q(x)$. For $y \geqslant x$, the d.f. that the new state will be less than y is

$$1 - \frac{Q(y)}{Q(x)}.$$

Suppose there exists a_n and b_n such that

$$P\left(\frac{M_n - b_n}{a_n} \leqslant x\right) \to G(x).$$

Define for $n = 1, 2, \cdots$, the process $\{M_t^n; t \geqslant 0\}$ such that

$$M_t^n = \frac{M_{\lfloor nt \rfloor} - b_n}{a_n},$$

where the time and space are scaled. Then one can show

$$\{M_t^n\} \Rightarrow \{U_t\},$$

where $\{U_t\}$is a extremal process. (Process convergence in metric space of sequence space.)

Let $X_1, X_2, \cdots$ be a "nice" regenerative process so that

$$\lim_{n\to\infty} P(X_n \leqslant x) = \frac{E\left[\sum_{k=0}^{\tau-1} \mathbb{1}_{(-\infty,x](X_k)}\right]}{E[\tau_1]} = F(x).$$

Suppose F is in the domain of attraction of an extreme value distribution G, that is, there exists a_n and b_n such that

$$F^n(a_n x + b_n) \to G(x).$$

For $n = 1, 2, \cdots$, set

$$Y_n = \max\left\{X_{\tau_{n-1},\ldots,X_{\tau_n}-1}\right\},$$

which is the maximum of the nth cycle. We've shown that if $\widetilde{M}_n = \max\{Y_1, \ldots, Y_n\}$, then

$$P(\widetilde{M}_n \leqslant a_n x + b_n) \to G^u(x),$$

where $u = E[\tau_1]$. Set

$$\widetilde{M_t^n} = \frac{\widetilde{M}_{\lfloor nt \rfloor} - b_n}{a_n}.$$

Then

$$\left\{\widetilde{M_t^n}\right\} \to \left\{\widetilde{U}_t\right\},$$

where $\left\{\widetilde{U}_t\right\}$ is an extremal process such that

$$P(\widetilde{U}_t \leqslant x) = G^t(x).$$

Proof. Let

$$M_n = \max\left\{X_1, \ldots, X_n\right\}.$$

Let

$$W_n := \max\{k : \tau_k \leqslant n\}.$$

Then $\{W_n\}$ is a renewal counting proces. For large n,

$$M_n = M_{W_n} = \widetilde{M}_{W_n},\ (a.s.)$$

since essentially, the maximum of $X_1, \ldots, X_n$ can not occur in $X_{W_n+1}, \cdots X_n$.
Set

$$M_t^n = M_{\lfloor nt \rfloor} = \widetilde{M}_{W_{\lfloor nt \rfloor}} = \widetilde{M}_{\lfloor n \frac{W_{\lfloor nt \rfloor}}{n} \rfloor} = \widetilde{M}^n{}_{\frac{W_{\lfloor nt \rfloor}}{n}}.$$

Set

$$\widetilde{\phi_t^n} = \frac{W_{\lfloor nt \rfloor}}{n},$$

so that

$$M_t^n = \widetilde{M}^n \circ \widetilde{\phi_t^n}.$$

Know

$$\frac{W_n}{n} \to \frac{1}{E[\tau_1]} = \frac{1}{\mu}.$$

Also,

$$\frac{\lfloor nt \rfloor}{n} \to t.$$

Thus,

$$\frac{W_{\lfloor nt \rfloor}}{n} = \frac{W_{\lfloor nt \rfloor}}{\lfloor nt \rfloor} \frac{\lfloor nt \rfloor}{n} \to \frac{t}{\mu}, \text{ w.p. 1.}$$

As a result,

$$\lfloor \widetilde{\phi}_t^n \rfloor \to \left\lfloor \frac{t}{\mu} \right\rfloor, \text{ w.p. 1.}$$

Let

$$\widetilde{\phi}_t := \frac{t}{\mu}.$$

If

$$\widetilde{M}^n \circ \widetilde{\phi}_t^n \Rightarrow \widetilde{U} \circ \widetilde{\phi}_t, \text{ w.p. 1.}$$

But

$$P\left(\widetilde{U} \circ \widetilde{\phi}_t \leqslant x\right) = \{[G(x)]^\mu\}^{\widetilde{\phi}_t} = \{[G(x)]^\mu\}^{\frac{t}{\mu}} = G^t(x).$$

If we take care of the edge effects, then

$$\{\widetilde{M^n}_t\} \Rightarrow \{\widetilde{U}_t\},$$

wher $\widetilde{U}_t = \widetilde{U} \circ \widetilde{\phi}_t$.

□

Chapter 9

Charateristic Functions and the Central Limit Theorem

9.1 Characteristic Functions

Suppose X is a real-valued r.v on the probability space $(\Omega, \mathcal{B}, P)$.

$$X : \Omega \to \mathbb{R}$$

is $\mathcal{B}/\mathcal{B}(\mathbb{R})$ measurable.

Definition 9.1. The moment generating function $\psi : \mathbb{R} \to \overline{\mathbb{R}}$ is defined as

$$\psi(u) = \mathbb{E}[e^{uX}].$$

Proposition 9.2. Suppose X is a random variable satisfying

$$\mathbb{E}\left[e^{tX}\right] < \infty,$$

Suppose that there exists a set $(-u_1, u_2), u_1, u_2 > 0$ such that for each $u \in (-u_1, u_2)$, $\psi(u) < \infty$. Notice that for $|u| < \min(u_1, u_2)$,

$$\mathbb{E}\left[e^{u|X|}\right] = \mathbb{E}[e^{uX}\mathbb{1}(X \geqslant 0)] + \mathbb{E}[e^{uX}\mathbb{1}(X < 0)] \leqslant \mathbb{E}[e^{uX}] + \mathbb{E}[e^{-uX}] = \psi(u) + \psi(-u) < \infty.$$

Furthermore, since for n satisfying $0 < n < \min(u_1, u_2)$, we have

$$e^{u|X|} = \sum_{n=0}^{\infty} \frac{u^n |X|^n}{n!} \geqslant \frac{u^n |X|^n}{n!}.$$

Thus,

$$\frac{u^n \mathbb{E}[|X|^n]}{n!} \leqslant \mathbb{E}\left[e^{u|X|}\right] < \infty,$$

and so for each integer $n \geqslant 1$,

$$\mathbb{E}\left(|X|^n\right) < \infty.$$

Definition 9.3. The characteristic function (chf) $\phi : \mathbb{R} \to \mathbb{C}$ of X with distribution F is defined as

$$\phi(t) = \mathbb{E}[e^{itx}] = \mathbb{E}[cos(tx)] + i\mathbb{E}[sin(tx)] = \int_{-\infty}^{\infty} cos(tx)Fdx + i\int_{-\infty}^{\infty} sin(tx)Fdx.$$

Proposition 9.4. Suppose $X : \Omega \to \mathbb{C}$ is of the form $X = U + iV$, where $U, V : \Omega \to \mathbb{R}$ are $\mathcal{B}/\mathcal{B}(\mathbb{R})$ measurable random variables having finite first moment. Then

$$|\mathbb{E}[X]| \leqslant \mathbb{E}\left[|X|\right].$$

Proof. Let $c = \mathbb{E}[X] \neq 0$.
Then

$$|c|^2 = \bar{c}c = \mathbb{E}[\bar{c}X] = \mathbb{E}[Re(\bar{c}X)] + i\mathbb{E}[Im(\bar{c}X)] = \mathbb{E}[Re(c\bar{X})],$$

since $|c|^2$ must be real.
Furthermore,

$$\left|\mathbb{E}[Re(c\bar{X})]\right| \leqslant \mathbb{E}\left[\left|Re(c\bar{X})\right|\right] \leqslant \mathbb{E}\left[|\bar{c}X|\right] = |\bar{c}|\mathbb{E}\left[|X|\right].$$

Thus,

$$|c|^2 \leqslant |\bar{c}|\mathbb{E}\left[|X|\right],$$

or

$$|\mathbb{E}[X]| = |c| \leqslant \mathbb{E}\left[|X|\right].$$

□

Remark. Properties of ϕ.

(a) $\phi(0) = 1$.

(b) $|\phi(t)| \leqslant 1$ for each $t \in \mathbb{R}$.

Proof.

$$\begin{aligned}
|\phi(t)|^2 &= \phi(t)\overline{\phi(t)} \\
&= (\mathbb{E}[cos(tx)] + i\mathbb{E}[sin(tx)])(\mathbb{E}[cos(tx)] - i\mathbb{E}[sin(tx)]) \\
&= \mathbb{E}^2[cos(tx)] + \mathbb{E}^2[sin(tx)] \\
&\leqslant \mathbb{E}[cos^2(tx)] + \mathbb{E}[sin^2(tx)] = 1. \qquad (9.1)
\end{aligned}$$

□

(c) ϕ is uniformly continuous on $\mathbb{R}$.

Proof. For $t, h \in \mathbb{R}$,

$$\begin{aligned}
|\phi(t+h) - \phi(t)| &= \left|\mathbb{E}[e^{i(t+h)x}] - \mathbb{E}[e^{itx}]\right| \\
&= \left|\mathbb{E}[e^{itx}(e^{ihx} - 1)]\right| \\
&\leqslant \mathbb{E}\left[\left|e^{itx}\right|\left|e^{ihx} - 1\right|\right] \\
&= \mathbb{E}\left[\left|e^{ihx} - 1\right|\right] \\
&\to 0 \text{ as } h \to 0. \qquad (9.2)
\end{aligned}$$

□

(d) For $a, b \in \mathbb{R}$, $t \in \mathbb{R}$,

$$\phi_{aX+b}(t) = \mathbb{E}[e^{it(aX+b)}] = \mathbb{E}[e^{itb}e^{i(at)X}] = e^{itb}\phi(at).$$

(e)

$$\phi_{-X}(t) = \phi_X(-t) = \mathbb{E}[cos(-tX)] + i\mathbb{E}[sin(-tX)] = \mathbb{E}[cos(tX)] - i\mathbb{E}[sin(tX)] = \overline{\phi_X(t)}$$

(f) X_1 and X_2 are independent, real-valued,

$$\phi_{X_1+X_2} = \phi_{X_2}(t)\phi_{X_2}(t)$$

(g) The chf ϕ is real if and only if

$$X \stackrel{d}{=} -X$$

if and only if the cdf F is a symmetric function. This follows since ϕ is real if and only if $\phi = \overline{\phi}$ if and only if X and $-X$ have the same chf by (5).

Lemma 9.5. For each integer $n \geqslant 0$, we have the following identity

$$\int_0^x (x-s)^n e^{is} ds = \frac{x^{n+1}}{n+1} + \frac{i}{n+1}\int_0^x (x-s)^{n+1} e^{is} ds,$$

for each $x \in \mathbb{R}$.

Proposition 9.6.

$$e^{ix} = \sum_{k=0}^{n} \frac{(ik)^x}{k!} + \frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is} ds,$$

for each $x \in \mathbb{R}$.

Proof.

$$\frac{1}{i}(e^{ix} - 1) = \int_0^x e^{is} ds = x + i\int_0^x (x-s)e^{is} ds$$

by setting $n = 0$ in Lemma 9.5. So

$$\begin{aligned} e^{ix} &= 1 + ix + i^2 \int_0^x (x-s)e^{is} ds \\ &= 1 + ix + i^2 \left(\frac{x^2}{2} + \frac{i}{2}\int_0^x (x-s)^2 e^{is} ds \right) \\ &= 1 + ix + \frac{(ix)^2}{2!} + \frac{i^3}{2!}\int_0^x (x-s)^2 e^{is} ds. \end{aligned}$$

by setting $n = 1$ in Lemma 9.5.

$$\vdots$$

□

Corollary 9.7. $\forall x \in \mathbb{R}$,

$$\begin{aligned}\left|e^{ix} - \sum_{k=1}^{n} \frac{(ik)^x}{k!}\right| &= \left|\frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is} ds\right| \\ &= \frac{1}{n!}\left|\int_0^x (x-s)^n e^{is} ds\right| \\ &\leqslant \frac{1}{n!}\int_0^x |x-s|^n ds \\ &= \frac{|x|^{n+1}}{(n+1)!}\end{aligned}$$

Corollary 9.8. For any $x \in \mathbb{R}$,

$$\left|e^{ix} - \sum_{k=1}^{n} \frac{(ik)^x}{k!}\right| \leqslant \frac{2|x|^n}{n!}$$

Proof. For $n \geqslant 1$,

$$\int_0^x e^{is}(x-s)^{n-1} ds = \frac{x^n}{n} + \frac{i}{n}\int_0^x (x-s)^n e^{is} ds, \tag{9.3}$$

by Lemma 9.5. Multiply both sides of equation 9.3 by $\frac{i^n}{(n-1)!}$,

$$\frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is} ds = \frac{i^n}{(n-1)!}\int_0^x (x-s)^{n-1} e^{is} ds - \frac{(ix)^n}{n!}. \tag{9.4}$$

Furthermore,

$$e^{ix} = \sum_{k=0}^{n} \frac{(ik)^x}{k!} + \frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is} ds. \tag{9.5}$$

Plugging 9.4 into 9.5 yields

$$e^{ix} - \sum_{k=0}^{n} \frac{(ik)^x}{k!} = -\frac{(ix)^n}{n!} + \frac{i^n}{(n-1)!}\int_0^x (x-s)^{n-1} e^{is} ds.$$

Then

$$\left|e^{ix} - \sum_{k=0}^{n} \frac{(ik)^x}{k!}\right| \leqslant \left|\frac{x^n}{n!}\right| + \frac{|x|^n}{n!} = \frac{2|x|^n}{n!}.$$

□

Proposition 9.9.

$$\left|e^{ix} - \sum_{k=0}^{n} \frac{(ik)^x}{k!}\right| \leqslant \min\left\{\frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!}\right\},$$

for each $x \in \mathbb{R}$.

Proposition 9.10. Suppose X is a random variable that satisfies

$$\psi(t) = \mathbb{E}\left[e^{tX}\right] < \infty, \forall t \in (-a, a),$$

for some $a > 0$. Then

$$\phi(t) = \mathbb{E}\left[e^{itX}\right] = \sum_{k=0}^{\infty} \frac{i^k \mathbb{E}\left[X^k\right]}{k!} t^k.$$

Proof. For any $t \in (-a, a)$,

$$\begin{aligned}
\mathbb{E}\left[e^{|tX|}\right] &= \mathbb{E}[e^{|t|X} \mathbb{1}(X \geqslant 0)] + \mathbb{E}[e^{-|t|X} \mathbb{1}(X \leqslant 0)] \\
&\leqslant \mathbb{E}[e^{|t|X}] + \mathbb{E}[e^{-|t|X}] \\
&= \psi(|t|) + \psi(-|t|) \\
&< \infty.
\end{aligned}$$

Since

$$\sum_{k=0}^{\infty} \frac{|t|^k |X|^k}{k!} = e^{|tX|},$$

by MCT,

$$\sum_{k=0}^{\infty} \frac{|t|^k \mathbb{E}\left[|X|^k\right]}{k!} = \mathbb{E}\left[e^{|tX|}\right] < \infty.$$

Then

$$\mathbb{E}\left[|X|^n\right] < \infty, \forall n \in \mathbb{N},$$

and

$$\lim_{n\to\infty} \frac{|t|^n |X|^n}{n!} = 0.$$

Then

$$\begin{aligned}
&\left|\mathbb{E}\left[e^{itX}\right] - \sum_{k=0}^{n} \frac{(it)^k \mathbb{E}\left[X^k\right]}{k!}\right| \\
&= \left|\mathbb{E}\left[e^{itX} - \sum_{k=0}^{n} \frac{(it)^k X^k}{k!}\right]\right| \\
&\leqslant \mathbb{E}\left|e^{itX} - \sum_{k=0}^{n} \frac{(it)^k X^k}{k!}\right| \\
&\leqslant \mathbb{E}\left[\frac{2|t|^n |X|^n}{n!} \bigwedge \frac{|t|^{n+1} |X|^{n+1}}{(n+1)!}\right] \\
&\leqslant 2\mathbb{E}\left[\frac{|t|^n |X|^n}{n!}\right] \\
&\to 0 \text{ as } n \to \infty,
\end{aligned}$$

where we use the DCT since for n large enough,

$$\frac{|t|^n |X|^n}{n!} \leqslant |X|^n,$$

and

$$E[|X|^n] < \infty, \forall n \in \mathbb{N}.$$

Thus, $\forall t \in (-a, a)$,

$$\phi(t) = \mathbb{E}\left[e^{itX}\right] = \sum_{k=0}^{\infty} \frac{i^k \mathbb{E}\left[x^k\right]}{k!} t^k. \qquad \square$$

Remark. Suppose X is a r.v that satisfies

$$\mathbb{E}\left[e^{tX}\right] < \infty, \forall t \in (-a, a),$$

for some $a > 0$. Then

$$\mathbb{E}\left[e^{tX}\right] = \sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k]}{k!} t^k, \forall t \in (-a, a).$$

Example 9.11. Suppose $Z \sim N(0, 1)$, compute

$$\phi_Z(t) = \mathbb{E}[e^{itZ}], \forall t \in \mathbb{R}.$$

For $t \in \mathbb{R}$,

$$\psi_Z(t) = \mathbb{E}[e^{tZ}] = e^{\frac{t^2}{2}} < \infty.$$

By Proposition 9.10,

$$\phi_Z(t) = \sum_{k=0}^{\infty} \frac{i^k \mathbb{E}\left[Z^k\right]}{k!} t^k, \forall t \in \mathbb{R}.$$

But

$$\psi_Z(t) = e^{\frac{t^2}{2}} = \sum_{k=0}^{\infty} \frac{t^{2k}}{2^k k!} = \sum_{k=0}^{\infty} \frac{(2k)!}{2^k k!} \frac{t^{2k}}{(2k)!},$$

and we know

$$\psi_Z(t) = \sum_{k=0}^{\infty} \mathbb{E}\left[Z^k\right] \frac{t^k}{k!}.$$

Thus,

$$E\left[Z^{2k}\right] = \frac{(2k)!}{2^k k!}, \forall k \geqslant 0,$$

and

$$E\left[Z^{2k+1}\right] = 0, \forall k \geqslant 0.$$

Then

$$\phi_Z(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \mathbb{E}\left[Z^k\right] = \sum_{k=0}^{\infty} \frac{(it)^{2k}}{(2k)!} \frac{(2k)!}{2^k k!} = \sum_{k=0}^{\infty} \frac{\left(-\frac{t^2}{2}\right)^k}{k!} = e^{-\frac{t^2}{2}} \in \mathbb{R}, \forall t \in \mathbb{R},$$

because $f_Z(t) = f_Z(-t)$.

Proposition 9.12. Suppose X is a r.v. satisfying $\mathbb{E}[|X|] < \infty$. Then $\phi'(0) = i\mathbb{E}[X]$. Furthermore, if $\mathbb{E}[|X|^n] < \infty$, then

$$\phi^{(n)}(0) = i^n \mathbb{E}[X^n].$$

Proof. We will focus on the case when $n = 1$. Fix $t, h \in \mathbb{R}$, then

$$\begin{aligned}
&\left| \frac{\phi(t+h) - \phi(t)}{h} - \mathbb{E}\left[(iX)e^{itX}\right] \right| \\
&= \left| \frac{1}{h}\mathbb{E}\left[e^{i(t+h)X} - e^{itX}\right] - \mathbb{E}\left[(iX)e^{itX}\right] \right| \\
&= \frac{1}{|h|}\left|\mathbb{E}\left[e^{itX}\left(e^{ihX} - 1 - (ihX)\right)\right]\right| \\
&\leqslant \frac{1}{|h|}\mathbb{E}\left[\left|e^{ihX} - 1 - (ihX)\right|\right] \\
&= \frac{1}{|h|}\mathbb{E}\left[\left|e^{ihX} - \sum_{k=0}^{1}\frac{(ihX)^k}{k!}\right|\right] \\
&\leqslant \frac{1}{|h|}\mathbb{E}\left[2\frac{|hX|^1}{1!} \bigwedge \frac{|hX|^2}{2!}\right] \\
&= \mathbb{E}\left[2|X| \bigwedge \frac{|h||X|^2}{2}\right] \\
&\to 0 \text{ as } h \to 0,
\end{aligned}$$

where we use the DCT since

$$2|X| \bigwedge \frac{|h||X|^2}{2} \leqslant 2|X|,$$

and

$$\mathbb{E}[X] < \infty.$$

Thus,

$$\lim_{h\to 0}\sup\left|\frac{\phi(t+h) - \phi(t)}{h} - \mathbb{E}[(iX)e^{itX}]\right| = 0.$$

So

$$\phi'(t) = \lim_{h\to 0}\frac{\phi(t+h) - \phi(t)}{h} = \mathbb{E}[(ix)e^{itx}].$$

Hence,

$$\phi(0) = i\mathbb{E}[X]. \qquad \square$$

9.2 Uniqueness and Continuity Theorem

Theorem 9.13 (Uniqueness)**.** *The characteristic function of a probability distribution on $\mathbb{R}$ uniquely determines the probability distribution function.*

Proof. Let X be a real-valued r.v. with cdf F and characteristic function (chf) ϕ. For any cdf G with chf γ and $\forall\theta \in \mathbb{R}$, we have the possible relation

$$\int_{\mathbb{R}} e^{-i\theta y}\phi(y)G(dy) = \int_{\mathbb{R}} \gamma(x-\theta)F(dx), \tag{9.6}$$

since

$$\begin{aligned}
&\int_{\mathbb{R}} e^{-i\theta y}\phi(y)G(dy) \\
&= \int_{\mathbb{R}} e^{-i\theta y}\int_{\mathbb{R}} e^{iyx}F(dx)G(dy) \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} e^{iy(x-\theta)}F(dx)G(dy) \\
&= \int_{\mathbb{R}}\left[\int_{\mathbb{R}} e^{iy(x-\theta)}G(dy)\right]F(dx) \\
&= \int_{\mathbb{R}} \gamma(x-\theta)Fdx.
\end{aligned}$$

Next, let G satisfying

$$G(dy) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}dy,$$

where G is a cdf of $N(0,\sigma^2)$. Let $Z = \sigma N$, where $N \sim N(0,1)$. Then chf of G is

$$\gamma = \phi_Z(x) = \phi_{\sigma N}(x) = \phi_N(\sigma x) = e^{-\frac{\sigma^2 x^2}{2}}, x \in \mathbb{R}.$$

Put γ and G into 9.6, we have

$$\int_{\mathbb{R}} e^{-i\theta y}\phi(y)\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}dy = \int_{\mathbb{R}} e^{-\frac{\sigma^2(z-\theta)^2}{2}}F(dz).$$

Then integrate both sides of the above equation over θ,

$$\int_{-\infty}^{x}\int_{\mathbb{R}} e^{-i\theta y}\phi(y)\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}dyd\theta = \int_{-\infty}^{x}\int_{\mathbb{R}} e^{-\frac{\sigma^2(z-\theta)^2}{2}}F(dz)d\theta.$$

By Toneli Theorem,

$$
\begin{aligned}
&\int_{-\infty}^{x}\int_{\mathbb{R}} e^{-\frac{\sigma^2(z-\theta)^2}{2}} F(dz)d\theta \\
&= \int_{\mathbb{R}}\int_{-\infty}^{x} e^{-\frac{\sigma^2(\theta-z)^2}{2}} F(dz)d\theta \\
&= \int_{\mathbb{R}}\left[\int_{-\infty}^{x-z} e^{-\frac{\sigma^2\theta^2}{2}} d\theta\right] F(dz) \\
&= \frac{\sqrt{2\pi}}{\sigma}\int_{\mathbb{R}}\left[\int_{-\infty}^{x-z} \frac{1}{\sqrt{2\pi\frac{1}{\sigma^2}}} e^{-\frac{\theta^2}{2\frac{1}{\sigma^2}}} d\theta\right] F(dz) \\
&= \frac{\sqrt{2\pi}}{\sigma}\int_{\mathbb{R}} P\left(\frac{N}{\sigma}\leqslant x-z\right) F(dz) \\
&= \frac{\sqrt{2\pi}}{\sigma} P\left(X+\frac{N}{\sigma}\leqslant x\right),
\end{aligned}
$$

since the last but one equation is a convolution and z is a instance of X and $N\sim N(0,1)$. Thus,

$$
\begin{aligned}
&P\left(X+\frac{N}{\sigma}\leqslant x\right) \\
&= \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{x}\int_{\mathbb{R}} e^{-i\theta y}\phi(y)\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}dyd\theta \\
&= \frac{1}{2\pi}\int_{-\infty}^{x}\int_{\mathbb{R}} e^{-i\theta y}\phi(y)e^{-\frac{y^2}{2\sigma^2}}dyd\theta.
\end{aligned}
$$

Let $\sigma_n = n, \forall n\in\mathbb{N}^{>0}$. Then

$$X+\frac{N}{\sigma_n}\to X \text{ as } n\to\infty.$$

So

$$X+\frac{N}{\sigma_n}\Rightarrow X.$$

Then

$$
\begin{aligned}
F(x) &= P(X\leqslant x) \\
&= \lim_{n\to\infty} P\left(X+\frac{N}{n}\leqslant x\right) \\
&= \frac{1}{2\pi}\lim_{n\to\infty}\int_{-\infty}^{x}\int_{\mathbb{R}} e^{-i\theta y}\phi(y)e^{-\frac{y^2}{2n^2}}dyd\theta,
\end{aligned}
$$

which implies the chf $\phi(t)$ uniquely determines the cdf $F(x)$ since we know the limit must exist. Or use Slutsky Theorem, since $X\Rightarrow X$ and $\frac{N}{\sigma_n}\xrightarrow{p}0$, $X+\frac{N}{\sigma_n}\Rightarrow X$.

□

Lemma 9.14 (Diagnolization)**.** Given a sequence $\{a_j\}_{j\geqslant 1}$ of distinct real numbers and a family $\{u_n\}_{n\geqslant 1}$ of real valued functions defined on $\mathbb{R}$, there exists a subsequence $\{u_{n_k}\}_{k\geqslant 1}$ of functions such that

$$\lim_{k\to\infty} u_{n_k}(a_j) \text{ exists (could be } \pm\infty), \forall j \in \mathbb{N}^{>0}.$$

Proof. The sequence $\{u_n(a_1)\}_{n\geqslant 1}$ contains a subsequence $\{u_{1,k}(a_1)\}_{k\geqslant 1}$ converges, i.e.

$$\lim_{k\to\infty} u_{1,k}(a_1) \text{ exists.}$$

Similarly, the sequence $\{u_{n_{1,k}}(a_1)\}$ contains a subsequence $\{u_{n_{2,k}}(a_2)\}$ such that

$$\lim_{k\to\infty} u_{2,k}(a_1) = \lim_{k\to\infty} u_{1,k}(a_1) \text{ exists.}$$

Continuing in this way,$\forall j \in \mathbb{N}^{>0}$, we have for $1 \leqslant l \leqslant j$,

$$\{u_{n_{j,k}}(a_l)\}_{k\geqslant 1} \subseteq \{u_{n_{l,k}}(a_l)\}_{k\geqslant 1}$$

and

$$\lim_{k\to\infty} u_{n_{j,k}}(a_l) = \lim_{k\to\infty} u_{n_{l,k}}(a_l) \text{ exists.}$$

Construct a new seuqnce of integers $\{m_j\}_{j\geqslant 1}$ as $m_j = n_{j,j},\ j \geqslant 1$. Then for each fixed $l \geqslant 1$,

$$\{m_j\}_{j\geqslant l} = \{n_{j,j}\}_{j\geqslant l} \subseteq \{n_{j,k}\}_{k\geqslant l}.$$

Then for any $l \in \mathbb{N}^{>0}$,

$$\lim_{k\to\infty} u_{m_k}(a_l) = \lim_{k\to\infty} u_{n_{l,k}}(a_l).$$ □

Lemma 9.15. If $D = \{a_i\}_{i\geqslant 1}$ is a countable dense subset of $\mathbb{R}$ and if $\{F_n\}_{n\geqslant 1}$ are distribution functions such that

$$\lim_{n\to\infty} F_n(a_i) \text{ exists}, \forall i \in \mathbb{N}^{>0},$$

then define for any $i \in \mathbb{N}^{>0}$,

$$F_\infty(a_i) = \lim_{n\to\infty} F_n(a_i).$$

This determines a distribution function F_∞ on $\mathbb{R}$ and

$$F_n \Rightarrow F_\infty \text{ as } n \to \infty.$$

Proof. Then F_∞ is non-decreasing, and $0 \leqslant F_\infty(x) \leqslant 1$ for any $x \in D$, and

$$\lim_{x\in D, x\to+\infty} F_\infty(x) = 1, \quad \lim_{x\in D, x\to-\infty} F_\infty(x) = 0.$$

For each $x \in \mathbb{R}$, define

$$F_\infty(x) = \inf_{a_i\in D, a_i\geqslant x} F_\infty(a_i).$$

By Lemma 8.2, F_∞ is a right continuous probability distribution function. Next, let $x \in C(F_\infty)$. Since D is dense, there exists two subsequences $\{a_i\}_{i\geqslant 1} \subseteq D$ and $\{a_i'\}_{i\geqslant 1} \subseteq D$ such that

$$a_i \downarrow x \text{ and } a_i' \uparrow x.$$

Then for any $k, i \in \mathbb{N}$,

$$F_k(a_i') \leqslant F_k(x) \leqslant F_k(a_i).$$

Then for any $i \in \mathbb{N}$, taking limit on k, we have

$$F_\infty(a_i') = \lim_{k\to\infty} \inf F_k(a_i') \leqslant \lim_{k\to\infty} \inf F_k(x) \leqslant \lim_{k\to\infty} \sup F_k(x) \leqslant \lim_{k\to\infty} \sup F_k(a_i) = F_\infty(a_i).$$

Since $x \in C(F_\infty)$,

$$\lim_{i\to\infty} F_\infty(a_i')(= F_\infty(x^-)) = F_\infty(x)(= F_\infty(x^+)) = \lim_{i\to\infty} F_\infty(a_i).$$

Thus,

$$F_\infty(x) = \lim_{k\to\infty} \inf F_k(x) = \lim_{k\to\infty} \sup F_k(x).$$

Hence

$$\lim_{k\to\infty} F_k(x) = F_\infty(x), \forall x \in C(F_\infty).$$ □

Theorem 9.16 (Section). *Any sequence of dfs $\{F_n\}_{n\geqslant 1}$ contains a weakly convergent subsequence (but the limit may not be proper).*

Proof. Define $D = \{a_i\}_{a_i \in \mathbb{Q}}$. By the Diagnolization Lemma, there exists a subsequence $\{F_{n_k}\}_{k\geqslant 1}$ such that $\lim_{k\to\infty} F_{n_k}(a_i)$ exists, $\forall i \in \mathbb{N}^{>0}$. By Lemma 9.15 there is a df F_∞ such that $F_{n_k} \Rightarrow F_\infty$. □

Example 9.17. Consider $\{F_n\}_{n\geqslant 1}$, where

$$F_n(x) = \begin{cases} 0, & x < n \\ 1, & x \geqslant n. \end{cases}$$

For each $q \in \mathbb{Q}$, $\lim_{n\to\infty} F_n(a) = 0$. By the Lemma 9.15, $F_\infty(x) = 0, \forall x \in \mathbb{R}$. So the limit is not proper.

Alternative: by the def of convergence in dist.

Definition 9.18. A collEction of distribution functions Π is relatively compact if every sequence $\{F_n\}_{n\geqslant 1}$ contains a subsequence that converges to a proper d.f..

Definition 9.19. A collection of distribution function Π is tight if$\forall \epsilon > 0$, there exists a compact set $K_\epsilon \subseteq \mathbb{R}$ such that$\forall F \in \Pi$,

$$F(K_\epsilon) > 1 - \epsilon.$$

Remark. Tightness of Π implies that each of d.f. $F \in \Pi$ is proper.

Theorem 9.20 (Prohorov's Theorem). *A family Π of distribution functions is relatively compact if and only if it is tight.*

Proof. $\Longleftarrow$ Suppose first Π is tight, and choose an arbitrary sequence $\{F_n\}_{n\geqslant 1} \subseteq \Pi$. Next, fix $\epsilon > 0$. Tightness of $\{F_n\}_{n\geqslant 1}$ implies $\exists M_\epsilon \in \mathbb{R}^+$ such that

$$F_n([-M_\epsilon, M_\epsilon]) > 1 - \epsilon, \forall n \in \mathbb{Z}^{>0}.$$

Furthermore, the Selection Theorem says there exists a subsequence $\{F_{n_k}\}_{k\geqslant 1}$ satisfying

$$F_{n_k} \Rightarrow F_\infty,$$

where F_∞ is a d.f.. We claim F_∞ is proper. Since $C(F_\infty)^c$ is at most countable, given M_ϵ, we can always find $M'_\epsilon > M_\epsilon$ such that $M'_\epsilon, -M'_\epsilon \in C(F_\infty)$. Clearly,

$$F_n([-M'_\epsilon, M'_\epsilon]) > 1-\epsilon, \forall n \in \mathbb{Z}^{>0}.$$

Then

$$\begin{aligned} F_\infty\left([-M'_\epsilon, M'_\epsilon]^c\right) &= 1 - F_\infty(M'_\epsilon) + F_\infty(-M'_\epsilon) \\ &= \lim_{k\to\infty}\left[1 - F_{n_k}(M'_\epsilon) + F_{n_k}(-M'_\epsilon)\right] \\ &= \lim_{k\to\infty} F_{n_k}\left([-M'_\epsilon, M'_\epsilon]^c\right) \\ &< \epsilon. \end{aligned}$$

So

$$F_\infty\left([-M'_\epsilon, M'_\epsilon]^c\right) \leqslant \epsilon.$$

Hence

$$F_\infty\left([-M'_\epsilon, M'_\epsilon]\right) > 1-\epsilon.$$

This holds for each ϵ, so $F_\infty(\mathbb{R}) = 1$.

$\Longrightarrow$ Assume Π is not tight. Then $\exists \epsilon > 0$ such that$\forall n \in \mathbb{Z}^{>0}$, we can pick $F_n \in \Pi$ satisfying

$$F_n\left([-n, n]\right) \leqslant 1-\epsilon.$$

This defines a sequence $\{F_n\}_{n\geqslant 1} \subseteq \Pi$. The Selection Theorem says $\exists \{F_{n_k}\}_{k\geqslant 1} \subseteq \{F_n\}_{n\geqslant 1}$ such that

$$F_{n_k} \Rightarrow G,$$

where G is a d.f.. The goal now is to show G is not proper. Choose $a, b \in C(G)$ such that $a < b$, then

$$G((a, b]) = \lim_{k\to\infty} F_{n_k}((a, b]).$$

If n_k is large enough,

$$F_k((a, b]) \leqslant F_{n_k}([-n_k, n_k]) \leqslant 1-\epsilon.$$

So

$$\lim_{k\to\infty} F_{n_k}((a, b]) \leqslant 1-\epsilon,$$

meaning

$$G((a, b]) \leqslant 1-\epsilon.$$

This holds for any $a, b \in C(G)$. Hence $G(\mathbb{R}) \leqslant 1-\epsilon$. Thus, G is not proper, proving Π is not relatively compact. $\square$

Remark. In the above theorem, we uses $\mathbb{R}$ is a polish space, which is a separable completely metrizable topological space when we use compactness of $\mathbb{R}$ to show there is a limit point for every bound and infinite set in Diagnolization Theorem and use separability in Selection Theorem.

Lemma 9.21. Suppose F is d.f. on $\mathbb{R}$ with chf ϕ. Then there exists $\alpha > 0$ such that for each $x > 0$,

$$F\left([-x,x]^c\right) < \alpha x \int_0^{\frac{1}{x}} \left[1 - \mathrm{Re}(\phi(t))\right] dt,$$

where α does not depend on F.

Proof. Recall first that

$$\mathrm{Re}(\phi(t)) = \int_{-\infty}^{\infty} cos(ty) F(dy).$$

This means for fixed $x > 0$,

$$\begin{aligned} x \int_0^{\frac{1}{x}} \left[1 - \mathrm{Re}(\phi(t))\right] dt &= x \int_0^{\frac{1}{x}} \int_{-\infty}^{\infty} (1 - cos(ty)) F(dy) dt \\ &= x \int_{-\infty}^{\infty} \left(\frac{1}{x} - \frac{1}{y} \sin \frac{y}{x}\right) F(dy) \\ &= \int_{-\infty}^{\infty} \left(1 - \frac{\sin \frac{y}{x}}{\frac{y}{x}}\right) F(dy). \end{aligned}$$

Since $1 - \frac{\sin u}{u} \geqslant 0, \forall u \in \mathbb{R}$,

$$\begin{aligned} x \int_0^{\frac{1}{x}} \left[1 - \mathrm{Re}(\phi(t))\right] dt &\geqslant \int_{|y| > x} \left(1 - \frac{\sin \frac{y}{x}}{\frac{y}{x}}\right) F(dy) \\ &\geqslant \alpha^{-1} F\left([-x,x]^c\right), \end{aligned}$$

where

$$\alpha^{-1} = \inf_{|y| \geqslant 1} \left(1 - \frac{\sin y}{y}\right).$$

□

Theorem 9.22 (Continuity Theorem)**.** *Let $\{X_n\}_{n\geqslant 1}$ be a sequence of real-valued r.v.'s with CDF F_n and chf ϕ_n.*

(a) If $X_n \Rightarrow X_0$ for some real-valued r.v. X_0, then

$$\lim_{n\to\infty} \phi_n(t) = \phi_0(t), \forall t \in \mathbb{R},$$

where ϕ_0 is the chf of X_0.

(b) Deeper part: Suppose

(1) $\lim_{n\to\infty} \phi_n(t) (= \phi_\infty(t))$ exists$\forall t \in \mathbb{R}$.

(2) $\phi_\infty(t)$ is continuous at $t = 0$.

Then for some d.f. F_∞, $F_n \Rightarrow F_\infty$, and ϕ_∞ is the chf of F_∞. If $\phi_\infty(0) = 1$, then F_∞ is proper.

Proof. Suppose $X_n \Rightarrow X_0$. Then by Portmanteau Theorem,$\forall t \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{E}[\cos(tX_n)] = \mathbb{E}[\cos(tX_0)],$$

and

$$\lim_{n\to\infty} \mathbb{E}[\sin(tX_n)] = \mathbb{E}[\sin(tX_0)].$$

Hence

$$\lim_{n\to\infty} \mathbb{E}\left[e^{itX_n}\right] = \mathbb{E}\left[e^{itX_0}\right].$$

Suppose$\forall t \in \mathbb{R}$,

$$\lim_{n\to\infty} \phi_n(t) = \phi_\infty(t),$$

where ϕ_∞ is continuous at $t = 0$. We claim $\{F_n\}_{n\geqslant 1}$ is tight. For $M \in \mathbb{R}^+$, by Lemma 9.21,

$$F_n\left([-M, M]^c\right) \leqslant \alpha M \int_0^{\frac{1}{M}} \left[1 - \mathrm{Re}(\phi_n(t))\right] dt, \forall n \in \mathbb{Z}^{>0}.$$

Then

$$\lim_{n\to\infty} \sup F_n\left([-M, M]^c\right) \leqslant \alpha x \lim_{n\to\infty} \int_0^{\frac{1}{M}} \left[1 - \mathrm{Re}(\phi_n(t))\right] dt.$$

Observe that

$$|1 - \mathrm{Re}(\phi_n(t))| = |1 - \mathbb{E}[cos(tX_n)]| \leqslant \mathbb{E}\left[|1 - cost(tX_n)|\right] \leqslant 2.$$

Then by the DCT/BCT,

$$\lim_{n\to\infty} \int_0^{\frac{1}{M}} \left[1 - \mathrm{Re}(\phi_n(t))\right] dt = \int_0^{\frac{1}{M}} \left[1 - \mathrm{Re}(\phi_\infty(t))\right] dt$$

Hence,

$$\lim_{n\to\infty} \sup F_n\left([-M, M]^c\right) \leqslant \alpha M \int_0^{\frac{1}{M}} \left[1 - \mathrm{Re}(\phi_\infty(t))\right] dt.$$

Since ϕ_∞ is continuous at $t = 0$,

$$\lim_{t\to 0} \phi_\infty(t) = \phi_\infty(0) = \lim_{n\to\infty} \phi_n(0) = 1,$$

since $\{X_n\}_{n\geqslant 1}$ are real-valued r.v. and then $\{PX_n^{-1}\}_{n\geqslant 1}$ are probability measures (d.f.'s) and by the definition of the measure, $PX_n^{-1}(\mathbb{R}) = 1 = F_n(\mathbb{R})$ for any $n \in \mathbb{Z}^{>0}$, so we have $\{F_n\}_{n\geqslant 1}$ are proper.

Next, fix $\epsilon > 0$ and choose large $M_\epsilon > 0$ such that

$$\sup_{0\leqslant t\leqslant \frac{1}{M_\epsilon}} \left[1 - \mathrm{Re}(\phi_\infty(t))\right] \leqslant \epsilon.$$

Then

$$\lim_{n\to\infty} \sup F_n\left([-M_\epsilon, M_\epsilon]^c\right) \leqslant \alpha M_\epsilon \int_0^{\frac{1}{M}} \epsilon dt = \alpha M_\epsilon.$$

This means that there exists an $n_0(\epsilon) \in \mathbb{Z}^{>0}$ such that as $n \geqslant n_0(\epsilon)$,

$$F_n\left([-M_\epsilon, M_\epsilon]^c\right) \leqslant \epsilon.$$

Not only that, $\exists M'_\epsilon > M_\epsilon$ such that

$$F_n\left([-M'_\epsilon, M'_\epsilon]^c\right) \leqslant \epsilon, \forall n \in \mathbb{Z}^{>0}.$$

Thus, $\{F_n\}_{n\geqslant 1}$ is tight. By Prohorov' Theorem, $\{F_n\}$ is also relative compact. Suppose $\{F_{n_{1,k}}\}_{k\geqslant 1}$ and $\{F_{n_{2,k}}\}_{k\geqslant 1}$ are both two subsequences of $\{F_n\}_{n\geqslant 1}$, where

$$F_{n_{1,k}} \Rightarrow G_1 \text{ and } F_{n_{2,k}} \Rightarrow G_2.$$

Then

$$\phi_\infty(t) = \lim_{k\to\infty} \phi_{n_{1,k}}(t) = \phi_{G_1}(t),$$

and

$$\phi_\infty(t) = \lim_{k\to\infty} \phi_{n_{2,k}}(t) = \phi_{G_2}(t).$$

Thus,

$$\phi_{G_1}(t) = \phi_{G_2}(t), \forall t \in \mathbb{R}.$$

□

9.3 The Classical CLT for iid Random Variables

Lemma 9.23. Suppose $\{a_n\}_{n\geqslant 1} \subseteq \mathbb{C}$ and $\lim_{k\to\infty} a_k = a \in \mathbb{C}$. Then

$$\lim_{n\to\infty}\left(1+\frac{a_n}{n}\right)^n = e^a.$$

Proof.

$$\begin{aligned}\left(1+\frac{a_n}{n}\right)^n &= \sum_{k=0}^{n}\binom{n}{k}\left(\frac{a_n}{n}\right)^k \\ &= \sum_{k=0}^{n}\frac{n(n-1)\cdots(n-k+1)}{n^k}\frac{a_n^k}{k!} \\ &= \sum_{k=0}^{\infty}\mathbb{1}(k\leqslant n)\frac{n(n-1)\cdots(n-k+1)}{n^k}\frac{a_n^k}{k!}.\end{aligned}$$

Theb by DCT,

$$\begin{aligned}\lim_{n\to\infty}\left(1+\frac{a_n}{n}\right)^n &= \sum_{k=0}^{\infty}\lim_{n\to\infty}\mathbb{1}(k\leqslant n)\frac{n(n-1)\cdots(n-k+1)}{n^k}\frac{a_n^k}{k!} \\ &= \sum_{k=0}^{\infty}\lim_{n\to\infty}\frac{a_n^k}{k!} \\ &= \sum_{k=0}^{\infty}\frac{a^k}{k!} \\ &= e^a.\end{aligned}$$

□

Theorem 9.24 (Weak Law of Large Numbers)**.** *Suppose $\{X_k\}_{k\geqslant 1}$ are iid sequence of r.v's, having cdf F and satisfying $\mathbb{E}\left[|X_1|\right] < \infty$. Define $\mu = \mathbb{E}[X_1]$ and $S_n = \sum_{k=1}^{n} X_k$, $n \geqslant 1$. Then $\frac{S_n}{n} \xrightarrow{p} \mu$.*

Proof. For $t \in \mathbb{R}$, since $\{X_k\}_{k\geqslant 1}$ are iid,

$$\begin{aligned}\mathbb{E}\left[e^{it\frac{S_n}{n}}\right] &= \mathbb{E}\left[\prod_{k=1}^{n} e^{\frac{it}{n}X_k}\right]\\ &= \mathbb{E}^n\left[e^{\frac{it}{n}X_1}\right].\end{aligned}$$

Next,

$$\begin{aligned}\mathbb{E}\left[e^{\frac{it}{n}X_1}\right] &= \mathbb{E}\left[\sum_{k=0}^{1}\frac{1}{k!}\left(\frac{it}{n}\right)^k X_1^k\right] + h(t,n)\\ &= 1 + \frac{it}{n}\mu + h(t,n).\end{aligned}$$

where

$$\begin{aligned}|h(t,n)| &= \left|\mathbb{E}\left[e^{\frac{it}{n}X_1} - \sum_{k=0}^{1}\frac{1}{k!}\left(\frac{it}{n}\right)^k X_1^k\right]\right|\\ &\leqslant \mathbb{E}\left|e^{\frac{it}{n}X_1} - \sum_{k=0}^{1}\frac{1}{k!}\left(\frac{it}{n}\right)^k X_1^k\right|\\ &\leqslant \mathbb{E}\left[\min\left\{\frac{2\left|\frac{tX_1}{n}\right|}{1!}, \frac{\left(\frac{tX_1}{n}\right)^2}{2!}\right\}\right]\\ &\leqslant \mathbb{E}\left[\min\left\{\frac{2|t||X_1|}{n}, \frac{t^2|X_1|^2}{2n^2}\right\}\right].\end{aligned}$$

Since

$$\begin{aligned}\lim_{n\to\infty}|nh(t,n)| &\leqslant \lim_{n\to\infty}\mathbb{E}\left[\min\left\{2|t||X_1|, \frac{t^2|X_1|^2}{2n}\right\}\right]\\ &\overset{DCT}{=\!=} \mathbb{E}\left[\lim_{n\to\infty}\min\left\{2|t||X_1|, \frac{t^2|X_1|^2}{2n}\right\}\right]\\ &\leqslant \mathbb{E}\left[\lim_{n\to\infty}\frac{t^2|X_1|^2}{2n}\right]\\ &= 0,\end{aligned}$$

since $\min\left\{2|t||X_1|, \frac{t^2|X_1|^2}{2n}\right\} \leqslant 2|t||X_1|$ and $\mathbb{E}\left[2|t||X_1|\right] = 2|t|\mathbb{E}\left[X_1\right] < \infty, \forall t \in \mathbb{R}$.
So

$$\lim_{n\to\infty} nh(t,n) = 0.$$

Hence,

$$\lim_{n\to\infty} itu + nh(t,n) = itu.$$

By Lemma 9.23,

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}^n\left[e^{\frac{it}{n}X_1}\right] &= \left(1+\frac{itu+nh(t,n)}{n}\right)^n \\ &= e^{it\mu} \\ &= \phi(t).\end{aligned}$$

Also $\phi(0)=1$, by the continuity theorem,

$$X_n \Rightarrow \mu.$$

By Proposition 8.34, we have

$$X_n \xrightarrow{p} \mu.$$

□

Theorem 9.25 (Central Limit Theorem)**.** *Suppose* $\{X_k\}_{k\geqslant 1}$ *is an iid sequence, where* $\mathbb{E}[X_1] = \mu < \infty$*,* $Var(X_1) = \sigma^2 < \infty$*. Define* $S_n = \sum_{k=1}^n X_k,\ n \geqslant 1$*. Then*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \to N(0,1).$$

Proof. Assume, wlog, that $\mu = 0,\ \sigma = 1$ by transforming theorem. Then $E\left[X_1\right] = 0,\ \mathbb{E}\left[X_1^2\right] = 1$ and we need to show $\frac{S_n}{\sqrt{n}} \to N(0,1)$. For $t \in \mathbb{R}$, since $\{X_k\}_{k\geqslant 1}$ are iid,

$$\begin{aligned}\mathbb{E}\left[e^{it\frac{S_n}{\sqrt{n}}}\right] &= \mathbb{E}\left[\prod_{k=1}^n e^{\frac{it}{n}X_k}\right] \\ &= \mathbb{E}^n\left[e^{\frac{it}{\sqrt{n}}X_1}\right].\end{aligned}$$

Next,

$$\begin{aligned}\mathbb{E}\left[e^{\frac{it}{\sqrt{n}}X_1}\right] &= \mathbb{E}\left[\sum_{k=0}^2 \frac{1}{k!}\left(\frac{it}{\sqrt{n}}\right)^k X_1^k\right] + h(t,n) \\ &= 1 + \frac{it}{\sqrt{n}}E\left[X_1\right] + \frac{1}{2}\left(\frac{it}{\sqrt{n}}\right)^2 E\left[X_1^2\right] + h(t,n). \\ &= 1 - \frac{t^2}{2n} + h(t,n),\end{aligned}$$

where

$$\begin{aligned}h(t,n) &\leqslant \mathbb{E}\left[\min\left\{2\frac{\left|\frac{tX_1}{\sqrt{n}}\right|^2}{2!}, \frac{\left|\frac{tX_1}{\sqrt{n}}\right|^3}{3!}\right\}\right] \\ &= \mathbb{E}\left[\min\left\{\frac{t^2X_1^2}{n}, \frac{|t|^3|X_1|^3}{6n^{\frac{3}{2}}}\right\}\right].\end{aligned}$$

Since

$$\begin{aligned}\lim_{n\to\infty} |nh(t,n)| &\leqslant \lim_{n\to\infty} \mathbb{E}\left[\min\left\{t^2X_1^2, \frac{|t|^3|X_1|^3}{6n^{\frac{1}{2}}}\right\}\right] \\ &\overset{DCT}{=\!=\!=} \mathbb{E}\left[\lim_{n\to\infty}\min\left\{t^2X_1^2, \frac{|t|^3|X_1|^3}{6n^{\frac{1}{2}}}\right\}\right] \\ &= 0,\end{aligned}$$

since $\min\left\{t^2X_1^2, \frac{|t|^3|X_1|^3}{6n^{\frac{1}{2}}}\right\} \leqslant t^2X_1^2$ and $\mathbb{E}\left[t^2X_1^2\right] = t^2\mathbb{E}\left[X_1^2\right] < \infty, \forall t \in \mathbb{R}$. So

$$\lim_{n\to\infty} nh(t,n) = 0.$$

Hence ,

$$\lim_{n\to\infty} -\frac{t^2}{2} + nh(t,n) = -\frac{t^2}{2}.$$

By Lemma 9.23,

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}^n\left[e^{\frac{it}{n}X_1}\right] &= \left(1 + \frac{-\frac{t^2}{2} + nh(t,n)}{n}\right)^n \\ &= e^{-\frac{t^2}{2}} \\ &= \phi(t).\end{aligned}$$

Also $\phi(0) = 1$, by the continuity theorem,

$$\frac{S_n}{\sqrt{n}} \Rightarrow N(0,1).$$

□

9.4 Lindeberg CLT

Let $\{X_n\}$ be independent, but not necessarily identically distributed, and suppose X_k has cdf F_k and chf $\phi(k)$ with $E[X_k] = 0$, $Var(X_k) = \sigma_k^2$. Define

$$s_n^2 = \sum_{k=1}^n \sigma_k^2 = \mathrm{Var}\left(\sum_{k=1}^n X_k\right).$$

Definition 9.26. $\{X_k\}_{k\geqslant 1}$ satisfies the lindeberg condition if for each $t > 0$,

$$\frac{1}{s_n^2}\sum_{k=1}^n E\left[X_k^2 \mathbb{1}\left(\left|\frac{X_k}{s_n}\right| > t\right)\right] \to 0.$$

Corollary 9.27. Consequences of the Lindeberg are

(a)

$$\lim_{n\to\infty} \max_{1\leqslant k\leqslant n} \frac{\sigma_k^2}{S_n^2} = 0.$$

Proof. For $\epsilon > 0$ fixed,$\forall k \in [n]$,

$$\begin{aligned}
\frac{\sigma_k^2}{s_n^2} &= \frac{1}{s_n^2}\mathbb{E}[X_k^2] \\
&= \mathbb{E}\left[\left|\frac{X_k}{s_n}\right|^2\right] \\
&= \mathbb{E}\left[\left|\frac{X_k}{s_n}\right|^2 \mathbb{1}\left(\left|\frac{X_k}{s_n}\right| \leqslant \epsilon\right)\right] + \mathbb{E}\left[\left|\frac{X_k}{s_n}\right|^2 \mathbb{1}\left(\left|\frac{X_k}{s_n}\right| > \epsilon\right)\right] \\
&\leqslant \epsilon^2 + \frac{1}{|s_n|^2}\mathbb{E}\left[|X_k|^2 \mathbb{1}\left(\left|\frac{X_k}{s_n}\right| > \epsilon\right)\right] \\
&\leqslant \epsilon^2 + \frac{1}{|s_n|^2}\sum_{l=1}^{n}\mathbb{E}\left[|X_k|^2 \mathbb{1}\left(\left|\frac{X_k}{s_n}\right| > \epsilon\right)\right].
\end{aligned}$$

Then

$$\max_{1\leqslant k\leqslant n}\frac{\sigma_k^2}{s_n^2} \leqslant \epsilon^2 + \frac{1}{|s_n|^2}\sum_{l=1}^{n}\mathbb{E}\left[|X_k|^2 \mathbb{1}\left(\left|\frac{X_k}{s_n}\right| > \epsilon\right)\right].$$

Then

$$\lim_{n\to\infty}\sup\max\frac{\sigma_k^2}{s_n^2} \leqslant \epsilon^2.$$

So

$$\lim_{n\to\infty}\max_{1\leqslant k\leqslant n}\frac{\sigma_k^2}{s_n^2} = 0.$$

□

(b) If

$$\lim_{n\to\infty}\max_{1\leqslant k\leqslant n}\frac{\sigma^k}{s_n^2} = 0,$$

then$\forall \epsilon > 0$,

$$\max_{1\leqslant\leqslant k\leqslant n} P\left(\left|\frac{X_k}{s_n}\right| > \epsilon\right) \to 0,$$

which is called uniform asymptotic negligibility (UAN). It is typical in CLT that the UAN condition holds so that no one summand dominates but each sumand contributes a small amount to the total.

Proof.

$$\begin{aligned}
\max_{1\leqslant k\leqslant n} P\left(\frac{X_k}{s_n} > \epsilon\right) &= \max_{1\leqslant k\leqslant n} P\left(|X_k| > s_n\epsilon\right) \\
&\leqslant \max_{1\leqslant k\leqslant n}\frac{\mathbb{E}[X_k^2]}{\epsilon^2 s_n^2} \\
&\leqslant \frac{1}{\epsilon^2}\max_{1\leqslant k\leqslant n}\frac{\mathbb{E}[X_k^2]}{s_n^2} \\
&\to 0.
\end{aligned}$$

□

Lemma 9.28. Suppose $\{a_k\}_{k=1}^n \subseteq \mathbb{C}$ and $\{b_k\}_{k=1}^n \subseteq \mathbb{C}$, where $|a_k| \leqslant 1$ and $|b_k| \leqslant 1, \forall k \in [n]$. Then

$$\left|\prod_{k=1}^n a_k - \prod_{k=1}^n b_k\right| \leqslant \sum_{k=1}^n |a_k - b_k|.$$

Proof. When $n = 1$, it clearly holds. Assume it holds for n.
Let $\{a_k\}_{k=1}^{n+1} \subseteq \mathbb{C}$ and $\{b_k\}_{k=1}^{n+1} \subseteq \mathbb{C}$, $|a_k| \leqslant 1$ and $\{b_k\} \leqslant 1, \forall k \in [n+1]$. Then

$$\left|\prod_{k=1}^n a_k - \prod_{k=1}^n b_k\right| = |c_n a_{n+1} - d_n b_{n+1}|,$$

where $c_n = \prod_{k=1}^n a_k$, $d_n = \sum_{k=1}^n b_k$. We have

$$|c_n| = \left|\prod_{k=1}^n a_k\right| = \prod_{k=1}^n |a_k| \leqslant 1.$$

Similarly,

$$|d_n| \leqslant 1.$$

Furthermore,

$$\begin{aligned} |c_n a_{n+1} - d_n b_{n+1}| &= |c_n a_{n+1} - d_n a_{n+1} + d_n a_{n+1} - d_n b_{n+1}| \\ &\leqslant |c_n - d_n||a_{n+1}| + |a_{n+1} - b_{n+1}||d_n| \\ &\leqslant |c_n - d_n| + |a_{n+1} - b_{n+1}| \\ &\leqslant \sum_{k=1}^{n+1} |a_k - b_k|. \end{aligned}$$

So it also holds for $n + 1$. □

Lemma 9.29. Suppose $\{Y_n\}$ is an iid sequence of r.v.'s with common cdf F and chf ϕ.
Let N be independent of $\{Y_n\}_{n\geqslant 1}$.
Suppose $N \sim Poisson(c)$, then the chf of $\sum_{k=1}^N Y_k$ is

$$\phi(t) = \mathbb{E}\left[e^{it\sum_{k=1}^N Y_k}\right] = e^{c(\phi(t)-1)}, \forall t \in \mathbb{R}.$$

Proof.

$$\begin{aligned}
\mathbb{E}\left[e^{it\sum_{k=1}^{N} Y_k}\right] &= \sum_{n=0}^{\infty} \mathbb{E}\left[e^{it\sum_{k=1}^{N} Y_k} \mathbb{1}(N=n)\right] \\
&= \sum_{n=0}^{\infty} \mathbb{E}\left[e^{it\sum_{k=1}^{n} Y_k} \mathbb{1}(N=n)\right] \\
&= \sum_{n=0}^{\infty} \mathbb{E}\left[e^{it\sum_{k=1}^{n} Y_k}\right] P(N=n) \\
&= \sum_{n=0}^{\infty} \phi(t)^n \frac{c^n e^{-c}}{c!} \\
&= e^{-c} \sum_{n=0}^{\infty} \frac{(\phi(t)c)^n}{n!} \\
&= e^{c(\phi(t)-1)}.
\end{aligned}$$

□

Theorem 9.30 (Lindeberg-FellEr CLT)**.** *The Lindeberg condition implies*

$$\frac{S_n}{s_n} \Rightarrow N(0,1),$$

where $S_n = \sum_{k=1}^{n} X_k$.

Proof. To show $\frac{S_n}{s_n} \Rightarrow N(0,1)$, it is equivalent to show

$$\phi_{S_n/s_n}(t) = \prod_{k=1}^{n} \phi_{X_k}(t/s_n) \to e^{-\frac{t^2}{2}},$$

by the Continuity Theorem. Then it suffices to show

$$\begin{aligned}
&\left|\prod_{k=1}^{n} \phi_{X_k}(t/s_n) - e^{-\frac{t^2}{2}}\right| \\
&\leqslant \left|\prod_{k=1}^{n} e^{\phi_{X_k}(t/s_n)-1} - \prod_{k=1}^{n} \phi_{X_k}(t/s_n)\right| \\
&\quad + \left|\prod_{k=1}^{n} e^{\phi_{X_k}(t/s_n)-1} - e^{-\frac{t^2}{2}}\right| \\
&= \left|\prod_{k=1}^{n} e^{\phi_{X_k}(t/s_n)-1} - \prod_{k=1}^{n} \phi_{X_k}(t/s_n)\right| \\
&\quad + \left|e^{\sum_{k=1}^{n}(\phi_{X_k}(t/s_n)-1)} - e^{-\frac{t^2}{2}}\right| \to 0.
\end{aligned}$$

Then it suffices to show

$$\left|\prod_{k=1}^{n} e^{\phi_{X_k}(t/s_n)-1} - \prod_{k=1}^{n} \phi_{X_k}(t/s_n)\right| \to 0,$$

and (since $|\cdot|$ and $e^{\cdot}$ is continuous,)

$$\sum_{k=1}^{n}(\phi_{X_k}(t/s_n)-1)-\left(-t^2/2\right)\to 0.$$

Since

$$\left|e^{\phi_{X_k}(t/s_n)-1}\right|\leqslant\frac{1}{e}e^{|\phi_{X_k}(t/s_n)|}\leqslant\frac{1}{e}e^{1}=1,$$

and

$$|\phi_{X_k}(t/s_n)|\leqslant 1,$$

by Lemma 9.28,

$$\begin{aligned}&\left|\prod_{k=1}^{n}e^{\phi_{X_k}(t/s_n)-1}-\prod_{k=1}^{n}\phi_{X_k}(t/s_n)\right|\\ \leqslant&\sum_{k=1}^{n}\left|e^{\phi_{X_k}(t/s_n)-1}-\phi_{X_k}(t/s_n)\right|\\ =&\sum_{k=1}^{n}\left|e^{\phi_{X_k}(t/s_n)-1}-1-(\phi_{X_k}(t/s_n)-1)\right|.\end{aligned}$$

Note that for $z\in\mathbb{C}$,

$$\begin{aligned}|e^z-1-z|&=\left|\sum_{k=2}^{\infty}\frac{z^k}{k!}\right|\\ &\leqslant\sum_{k=2}^{\infty}|z|^k\\ &=\frac{|z|^2}{1-|z|}\ \text{ if } |z|<1,\\ &\leqslant 2|z|^2,\ \text{ if } |z|\leqslant\frac{1}{2},\\ &\leqslant\delta|z|,\ \text{ if } |z|\leqslant\frac{\delta}{2}<\frac{1}{2},\end{aligned}$$

for any $0<\delta<1$.

Now for fixed $t \in \mathbb{R}$,

$$\begin{aligned}
|\phi_{X_k}(t/s_n) - 1| &= \left|\mathbb{E}\left[e^{itX_k/s_n}\right] - 1\right| \\
&= \left|\mathbb{E}\left[e^{itX_k/s_n} - itX_k/s_n - 1\right]\right| \\
&\leqslant \mathbb{E}\left[\left|e^{itX_k/s_n} - itX_k/s_n - 1\right|\right] \\
&\leqslant \mathbb{E}\left[\frac{(X_k t/s_n)^2}{2!}\right] \\
&= \frac{t^2}{2s_n^2}\sigma_k^2 \\
&\leqslant \frac{t^2}{2}\max_{1\leqslant k\leqslant n}\frac{\sigma_k^2}{s_n^2} < \frac{\delta}{2}, \forall 0 < \delta < 1,
\end{aligned}$$

when n is sufficiently large by CorollAry 9.27 from the consequence of Lindeberg condition. Then since $|\phi_{X_k}(t/s_n) - 1| < \frac{\delta}{2}, \forall k \in [n]$,

$$\begin{aligned}
&\left|\prod_{k=1}^n e^{\phi_{X_k}(t/s_n)-1} - \prod_{k=1}^n \phi_{X_k}(t/s_n)\right| \\
\leqslant &\sum_{k=1}^n \left|e^{\phi_{X_k}(t/s_n)-1} - 1 - (\phi_{X_k}(t/s_n) - 1)\right| \\
\leqslant &\sum_{k=1}^n \delta|\phi_{X_k}(t/s_n) - 1)| \\
\leqslant &\delta\frac{t^2}{2s_n^2}\sum_{k=1}^n \sigma_k^2 = \delta\frac{t^2}{2}, \forall 0 < \delta < 1,
\end{aligned}$$

for n is sufficiently large.
Thus,

$$\left|\prod_{k=1}^n e^{\phi_{X_k}(t/s_n)-1} - \prod_{k=1}^n \phi_{X_k}(t/s_n)\right| \to 0.$$

Next, since for any $k \in [n]$,

$$E[X_k] = 0 \text{ and } E[X_k^2] = \sigma_k^2,$$

we can write

$$\begin{aligned}
&\sum_{k=1}^{n}(\phi_{X_k}(t/s_n - 1)) - \left(-t^2/2\right) \\
&= \sum_{k=1}^{n} E\left(e^{itX_k/s_n} - 1\right) - \sum_{k=1}^{n} i\frac{t}{s_n}E[X_k] - \sum_{k=1}^{n}\frac{(it)^2E[X_k^2]}{2s_n^2} \\
&= \sum_{k=1}^{n} E\left(e^{itX_k/s_n} - 1 - i\frac{t}{s_n}X_k - \frac{1}{2}\left(\frac{it}{s_n}\right)^2 X_k^2\right) \\
&= \sum_{k=1}^{n} E(*) \\
&= \sum_{k=1}^{n}\left(E(*)\mathbb{1}\left(\left|\frac{X_k}{s_n}\right| \leqslant \epsilon\right) + E(*)\mathbb{1}\left(\left|\frac{X_k}{s_n}\right| > \epsilon\right)\right) \\
&= \sum_{k=1}^{n} E(*)\mathbb{1}\left(\left|\frac{X_k}{s_n}\right| \leqslant \epsilon\right) + \sum_{k=1}^{n} E(*)\mathbb{1}\left(\left|\frac{X_k}{s_n}\right| > \epsilon\right) \\
&= I + II.
\end{aligned}$$

Since

$$\begin{aligned}
|I| &\leqslant \sum_{k=1}^{n} E(|*|)\mathbb{1}\left(\left|\frac{X_k}{s_n}\right| \leqslant \epsilon\right) \\
&\leqslant \sum_{k=1}^{n} E\left(\frac{1}{3!}\left|\frac{t}{s_n}X_k\right|^3 \mathbb{1}\left(\left|\frac{X_k}{s_n}\right| \leqslant \epsilon\right)\right) \\
&\leqslant \frac{|t|^3}{6}\sum_{k=1}^{n} E\left(\left|\frac{X_k}{s_n}\right|^3 \mathbb{1}\left(\left|\frac{X_k}{s_n}\right| \leqslant \epsilon\right)\right) \\
&\leqslant \frac{|t|^3}{6}\epsilon\sum_{k=1}^{n} E\left(\left|\frac{X_k}{s_n}\right|^2 \mathbb{1}\left(\left|\frac{X_k}{s_n}\right| \leqslant \epsilon\right)\right) \\
&\leqslant \frac{|t|^3}{6}\epsilon\sum_{k=1}^{n}\frac{\sigma_k^2}{s_n^2} = \frac{|t|^3}{6}\epsilon,
\end{aligned}$$

and

$$\begin{aligned}
|II| &\leqslant \sum_{k=1}^{n} E\left(|\mathbb{1}|\left(\left|\frac{X_k}{s_n}\right| > \epsilon\right)\right) \\
&\leqslant 2\sum_{k=1}^{n} E\left(\frac{1}{2!}\left|\frac{tX_k}{s_n}\right|^2 \mathbb{1}\left(\left|\frac{X_k}{s_n}\right| > \epsilon\right)\right) \\
&= \frac{t^2}{s_n^2}\sum_{k=1}^{n} E\left(X_k^2\mathbb{1}\left(\left|\frac{X_k}{s_n}\right| > \epsilon\right)\right) \\
&\to 0,
\end{aligned}$$

by the Lindeberg condition. □

Theorem 9.31 (Second Converging Together Theorem)**.** *Let us suppose that* $\{X_{u,n}\}_{n\geqslant 1, u\geqslant 1}$. $\{X_u\}_{u\geqslant 1}$, $\{Y_n\}_{n\geqslant 1}$ *and* X *are random variables such that for each* $n \geqslant 1$, Y_n, $\{X_{u,n}\}$ *are all defined on the same prob. space. Suppose for each* $u \geqslant 1$ *that* $X_{u,n} \Rightarrow X_u$ *as* $n \to \infty$. *Furthmore,* $X_u \Rightarrow X$ *as* $u \to \infty$. *If* $\forall\ \epsilon > 0$, *we have*

$$\lim_{\mu\to\infty} \lim_{n\to\infty} \sup P\left(|Y_n - X_{u,n}| > \epsilon\right) = 0,$$

then

$$Y_n \Rightarrow X \text{ as } n \to \infty.$$

Proof. Suppose $f : \mathbb{R} \to \mathbb{R}$ is bounded and uniformly continuous on $\mathbb{R}$. Our goal is to show

$$\lim_{n\to\infty} \sup|E\left[f(Y_n)\right] - E\left[f(X)\right]| = 0.$$

Fix $\epsilon > 0$ and define

$$w_f(\epsilon) := \sup_{|x-y|<\epsilon} |f(x) - f(y)|.$$

$$\begin{aligned}|E\left[f(Y_n)\right] - E\left[f(X)\right]| = |E\left[f(Y_n)\right] - E\left[f(X_{u,n})\right]| + |E\left[f(X_{u,n})\right] - E\left[f(X_u)\right]| \\ + |E\left[f(X_u)\right] - E\left[f(X)\right]|.\end{aligned}$$

Letting $n \to \infty$,

$$\begin{aligned}\lim_{n\to\infty} \sup|E\left[f(Y_n)\right] - E\left[f(X)\right]| \leqslant \lim_{n\to\infty} \sup|E\left[f(Y_n)\right] - E\left[f(X_{u,n})\right]| + 0 \\ + |E\left[f(X_u)\right] - E\left[f(X)\right]|.\end{aligned}$$

Letting $u \to \infty$,

$$\lim_{n\to\infty} \sup|E\left[f(Y_n)\right] - E\left[f(X)\right]| \leqslant \lim_{u\to\infty} \lim_{n\to\infty} \sup|E\left[f(Y_n)\right] - E\left[f(X_{u,n})\right]|.$$

Since

$$\begin{aligned}|E\left[f(Y_n)\right] - E\left[f(X_{u,n})\right]| &\leqslant E\left(|f(Y_n) - f(X_{u,n})|\mathbb{1}_{|f(Y_n)-f(X_{u,n})|\leqslant\epsilon}\right) \\ &\quad + E\left(|f(Y_n) - f(X_{u,n})|\mathbb{1}_{|f(Y_n)-f(X_{u,n})|>\epsilon}\right) \\ &\leqslant E(w_f(\epsilon)) + 2MP\left(|Y_n - X_{u,n}| > \epsilon\right).\end{aligned}$$

So

$$\begin{aligned}\lim_{u\to\infty} \lim_{n\to\infty} \sup|E\left[f(Y_n)\right] - E\left[f(X_{u,n})\right]| &\leqslant w_f(\epsilon) + 2MP \lim_{u\to\infty} \lim_{n\to\infty} \sup\left(|Y_n - X_{u,n}| > \epsilon\right) \\ &= w_f(\epsilon).\end{aligned}$$

Let $\epsilon \to 0$, then $w_f(\epsilon) \to 0$, so

$$\lim_{n\to\infty} \sup|E\left[f(Y_n)\right] - E\left[f(X)\right]| = 0.$$ □

9.5 CLT for m-dependent random variables

Definition 9.32. A sequence of r.v.'s $\{X_n\}_{n\geqslant 1}$ is m-dependent ($m \in \mathbb{Z}^{\geqslant 0}$ fixed.) if$\forall t \in \mathbb{Z}^+$, $\sigma(X_j, j \leqslant t)$ and $\sigma(X_j, j \geqslant t+m+1)$ are independent.

The most common example of a stationary m-dependent sequence is the time series model called the moving average of order m.

Definition 9.33. Let $\{Z_n\}$ be iid and define for given constants $c_1, \ldots, c_m \in \mathbb{R}$ the process

$$X_t = \sum_{j=1}^{m} c_j Z_{t-j},\ t = 0, 1. \cdots .$$

Example 9.34. $X_t = \sum_{j=1}^m c_j Z_{t-j}$ and $X_{t+m} = \sum_{j=1}^m c_j Z_{t+m-j}$ are independent. Then X_k and X_l are independent$\forall |k-l| \geqslant m$. Thus, the sequence $\{X_n\}_{n\geqslant 1}$ are $(m-1)$-dependent.

Remark. m-dependent implies k-independet$\forall\, k > m$.

Theorem 9.35 (Hoeffding and Robbins)**.** *Suppose $\{X_n, n \geqslant 1\}$ is a strictly stationary and m-dependent sequence with $E(X_1) = 0$ and*

$$Cov(X_t, X_{t+h}) = E(X_t X_{t+h}) := \gamma(h).$$

Suppose

$$\nu_m := \gamma(0) + 2\sum_{j=1}^{m} \gamma(j) \neq 0.$$

Then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \Rightarrow N(0, \nu_m).$$

Proof. (a) Part 1: Variance calculation.

$$\begin{aligned}
n\mathrm{Var}(\overline{X}_n) &= n\mathrm{Cov}\left(\frac{1}{n}\sum_{i=1}^{n} X_i, \frac{1}{n}\sum_{i=j}^{n} X_j\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j) \\
&= \gamma(0) + \frac{2}{n}\sum_{i<j} \gamma(j-i) \\
&= \gamma(0) + \frac{2}{n}\sum_{i=1}^{n-1} i\gamma(n-i) \\
&= \gamma(0) + \frac{2}{n}\sum_{k=1}^{n-1} (n-k)\gamma(k) \\
&= \gamma(0) + 2\sum_{k=1}^{n-1} (1-\frac{k}{n})\gamma(k).
\end{aligned}$$

Since $\gamma(l) = 0$ if $|l| > m$, letting $n \to \infty$,

$$n\text{Var}(\overline{X}_n) \to \gamma(0) + 2\sum_{k=1}^{m} \gamma(k).$$

(b) Part 2: The big block-little block method. Pick $n > u > 2m$. Consider the following diagram.

$$(1) \xrightarrow{fb} (u-m) \xrightarrow{fl} (u) \xrightarrow{sb} (2u-m) \xrightarrow{sl} (2u) \cdots\cdots (r-1)u \xrightarrow{rb} (ru-m) \to (rn) \xrightarrow{remainder} (n).$$

Let

$$r = \left\lfloor \frac{n}{u} \right\rfloor$$

so that $\frac{r}{n} \to \frac{1}{u}$ and define

$$\begin{aligned}
\xi_1 &= X_1 + \cdots + X_{u-m}, \\
\xi_2 &= X_{u+1} + \cdots + X_{2u-m}, \\
&\vdots \\
\xi_r &= X_{(r-1)u+1} + \cdots + X_{ru-m},
\end{aligned}$$

which are the "big block" sums. Note by stationarity and m-dependence that $\xi_i, \ldots, \xi_r$ are iid because the little blocks have been removed. Define

$$X_{un} := \frac{\xi_1 + \cdots + \xi_r}{\sqrt{n}} = \frac{\xi + \cdots + \xi_r}{\sqrt{r}} \sqrt{\frac{r}{n}}.$$

Note that

$$\sqrt{\frac{r}{n}} \to \sqrt{\frac{1}{u}}.$$

From the CLT for iid summands, as $n \to \infty$,

$$X_{un} \Rightarrow N\left(0, \frac{\text{Var}(\xi_1)}{u}\right) =: X_u.$$

Now observe that as $u \to \infty$,

$$\begin{aligned}
\frac{\text{Var}(\xi_1)}{u} &= \frac{\text{Var}\left(\sum_{i=1}^{u-m} X_i\right)}{u} \\
&= \frac{(u-m)^2}{u}\text{Var}\left(\frac{\sum_{i=1}^{u-m} X_i}{u-m}\right) \\
&= (u-m)\text{Var}\left(\overline{X}_{u-m}\right)\frac{u-m}{u} \\
&\to \nu_m \cdot 1 = \nu_m,
\end{aligned}$$

by the part (1). Thus, as $u \to \infty$,

$$X_u = N\left(0, \frac{\mathrm{Var}(\xi_1)}{u}\right) \Rightarrow N(0, \nu_m) =: X,$$

since a sequence of normal distributions converges weakly if their means (in this case, all zero) and variance converge (Continuity theorem). By the second convergence together theorem, it remains to show that

$$\lim_{u\to\infty} \lim_{n\to\infty} \sup P\left(\left|X_{un} - \frac{\sum_{i=1}^n X_i}{\sqrt{n}}\right| > \epsilon\right) = 0.$$

For $i = 1, \ldots, r$, let

$$B_i = \{iu - m + 1, \ldots, iu\}$$

be the m integers in the $i$the little block, and let

$$B_r = \{ru - m + 1, \ldots, n\}$$

be the integers in the last little block **coupled** with the remainder due to u not dividing n exactly. Then we have

$$\left|\frac{\sum_{i=1}^n X_i}{\sqrt{n}} - X_{un}\right| = \frac{1}{\sqrt{n}}\left|\sum_{i\in B_1} X_1 + \cdots + \sum_{i\in B_{r-1}} X_i + \sum_{i\in B_r} X_i\right|,$$

and all sums on the right side are independent by m-dependence. Also, by the stationarity ($(r-1)$iid),

$$\mathrm{Var}\left(\frac{\sum_{i=1}^n X_i}{\sqrt{n}} - X_{un}\right) = \frac{1}{n}\left((r-1)\mathrm{Var}\left(\sum_{i=1}^m X_i\right) + \mathrm{Var}\left(\sum_{i=1}^{n-ru+m+1} X_i\right)\right).$$

Note that

$$\begin{aligned} h(n) :&= n - ru + m \\ &= n - \left[\frac{n}{u}\right] u + m \\ &\leqslant n - \left(\frac{n}{u} - 1\right) u + m \\ &= n - n + u + m \\ &= u + m. \end{aligned}$$

Thus for fixed u, as $n \to \infty$,

$$\frac{1}{n}\mathrm{Var}\left(\sum_{i=1}^{h(n)} X_i\right) \leqslant \frac{\sup_{j\in[u+m]} \mathrm{Var}\left(\sum_{i=1}^{j} X_i\right)}{n} \to 0.$$

Also, since $r/n \to 1/u$ as $n \to \infty$,

$$\frac{1}{n}\left((r-1)\mathrm{Var}\left(\sum_{i=1}^m X_i\right)\right) \sim \frac{1}{u}\mathrm{Var}\left(\sum_{i=1}^m X_i\right) \to 0.$$

By Chebychev's inequality

$$\lim_{u\to\infty}\lim_{n\to\infty}\sup P\left(\left|\frac{\sum_{i=1}^{n}X_i}{\sqrt{n}}-X_{un}\right|>\epsilon\right)\leqslant\lim_{u\to\infty}\lim_{n\to\infty}\sup\frac{1}{\epsilon^2}\mathrm{Var}\left(\frac{\sum_{i=1}^{n}X_i}{\sqrt{n}}-X_{un}\right)$$
$$=0.$$

□

Chapter 10

Point process

10.1 Borel Measure

Definition 10.1. Let E be a non-empty set, A family of subsets $\mathcal{E}$ of E is called a *σ-algebra* if

(a) $E \subseteq \mathcal{E}$,

(b) if $A \in \mathcal{E}$, so is A^c,

(c) if $A_1, A_2, \cdots \in \mathcal{E}$, so is $\bigcup_{n=1}^{\infty} A_n$.

Remark. (a) We can measure the whole set E.

(b) $A \in \mathcal{E}$ measurable, then A^c should be measurable. If $\mu(A) < \infty$, $\mu(A^c) = \mu(E) - \mu(A)$.

(c) If A_1 and A_2 are disjoint, $\mu(A_1 \sqcup A_2) = \mu(A_1) + \mu(A_2)$.

(d)

$$A_1 \cap A_2^c,\ A_1^c \cap A_2,\ A_1 \cap A_2 \in \mathcal{E}.$$

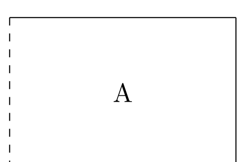
$$\begin{aligned}
\mu(A_1 \cup A_2) &= \mu\left((A_1 \cap A_2^c) \sqcup (A_1^c \cap A_2) \sqcup (\cap A_1 \cap A_2)\right) \\
&= \mu(A_1 \cap A_2^c) + \mu(A_1^c \cap A_2) + \mu(A_1 \cap A_2) \\
&= (\mu(A_1 \cap A_2^c) + \mu(A_1 \cap A_2)) + (\mu(A_1^c \cap A_2) + \mu(A_1 \cap A_2)) - \mu(A_1 \cap A_2) \\
&= \mu(A_1) + \mu(A_2) - \mu(A_1 \cap A_2).
\end{aligned}$$

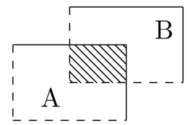
Definition 10.2. Suppose $E = \mathbb{R}^2$,

$$\mathcal{B} := \{\, (a_1, b_1] \times (a_2, b_2] |\, a_1 < b_1, a_2 < b_2\} \bigcup \emptyset,$$

which are all the *rectangles*.

Remark. A^c is not a rectangle.





Since for any $A, B \in \mathcal{C}$, $A \cap B \in \mathcal{B}$, so $\mathcal{B}$ is a π-system. Since $A \in \mathcal{B}$, $A^c \notin \mathcal{B}$, and $A \sqcup B$ may not in $\mathcal{B}$, $\mathcal{B}$ is not a λ-system. Thus, $\mathcal{B}$ is not a σ-algebra. The borel subsets of $\mathbb{R}^2$, denoted as $\mathcal{B}(\mathbb{R}^2)$, is the smallest σ-algebra that contains $\mathcal{B}$ (proved by the axiom of choice).

Definition 10.3. A *Hausdorff space* X is a topological space in which distinct points have disjoint neighbourhoods, which means for any $x, y \in X$, there exist $r_x, r_y \in \mathbb{R}^+$, and $B^d_{r_x}(x)$ and $B^d_{r_y}(y)$ such that

$$B^d_{r_x}(x) \cap B^d_{r_y}(y) = \emptyset,$$

where d is the corresponding *Hausdorff distance* or *Hausdorff metric*.

Definition 10.4. A *borel measure* μ on the Hausdorff space $\mathbb{R}^2$ is a mapping from $\mathcal{B}(\mathbb{R}^2)$ to $[0, \infty]$ that satisfies if $A_1, A_2, \cdots$ are disjoint subsets of $\mathcal{B}(\mathbb{R}^2)$, then

$$\mu\left(\bigsqcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Remark. The choice of Borel measure which assigns

$$\mu(a, b] \times (c, d]) = (c - a)(d - b), \forall (a, b], (c, d] \in B$$

is sometimes called "the" Borel measure on $\mathbb{R}^2$. The measure is actually the Lebesgue measure restricted on the Borel algebra $\mathcal{B}(\mathbb{R}^2)$, which is a complete measure and is defined on the Lebesgue σ-algebra (all Lebesgue measurable sets form a σ-algebra).

Definition 10.5. Let (X, T, d) be a topological metric space. The *Diameter* of a $B \in (X, T)$,

$$D(B) = \sup\{d(x, y) \mid x, y \in B\}.$$

If B is bounded, then $D(B) < \infty$.

10.2 Measure

Theorem 10.6. *Suppose μ_1 and μ_2 are two measures on $\left(\mathbb{R}^2, \mathcal{B}\left(\mathbb{R}^2\right)\right)$ such that for any $B \in \mathcal{B}$, $\mu_1(B) = \mu_2(B)$. Then*

$$\mu_1(A) = \mu_2(A), \forall A \in \mathcal{B}(\mathbb{R}^2).$$

Definition 10.7 (Measurable sets)**.** A classical solution to the measure problem consists in attempting to approximate the measure of a complicated set using simple sets. More precisely, suppose we have a class of simple sets S which we know how to measure (these would contain events, rectangles and finite unions of rectangle for example). Then, given some arbitrary set A, we can define an inner measure $\mu_I(A)$ and an outer measure $\mu_O(A)$ of A by letting

$$\mu_I(A) = \sup\{\mu(E) : E \subseteq A, E \in S\} \text{ and } \mu_O(A) = \inf\{\mu(E) : E \supseteq A, E \in S\}.$$

Note that the inner and outer measures of sets in S are clearly the same as the measure we have already assigned to them. In this framework, one calls a set A *measurable* if $\mu_O(A) = \mu_I(A)$, in which case we assign $\mu(A) = \mu_O(A) = \mu_I(A)$.

Definition 10.8. Let (X, T) be a topological space, and let Σ be a σ-algebra on X that contains the topology T. Let M be a collection of (possibly signed or comlex) measure defined on Σ. The collection M is called *tight* (or uniformly tight) if for any $\epsilon > 0$, there is a compact subset K_ϵ of X such that, for all measures $\mu \in M$,

$$|\mu|(K \smallsetminus K_\epsilon) < \epsilon,$$

where $|\mu|$ is the total variation measure of μ. Very often, the measures in question are probability measure, so the last part can be written as

$$\mu(K_\epsilon) > 1 - \epsilon.$$

If a tight collection M consists of a single measure μ, then μ may either be said to be a tight measure or to be an inner regular measure. If Y is an X-valued random variable whose probability distribution on X is a tight measure, then Y is said to be a seperable random variable or a Radon random variable.

Example 10.9. Let $Y \in (\Omega, \mathcal{B}, P)$. Then Y is tight.

Proof. Let $\epsilon > 0$. Since P is nondecreasing and

$$\lim_{x \to \infty} P([-x, x]) = 1,$$

there exists $K_\epsilon = [-n, n] \subsetneq \mathbb{R}^+$ such that $P(K_\epsilon) > 1 - \epsilon$. □

Example 10.10. Consider $\mathbb{R}$ with its usual Borel topology. Let δ_x denote the Dirac measure, a unit mass at the point x in $\mathbb{R}$. The collection $M_1 := \{\delta_n \mid n \in \mathbb{N}\}$ is not tight.

Proof. Assume there is such a comact K_ϵ. Then there exists $N \in \mathbb{Z}^+$ such that $\sup K_\epsilon \leqslant N < \infty$. However, $\delta_{N+1}(K_\epsilon) = 0 \leqslant 1 - \epsilon$ when $0 < \epsilon \leqslant 1$. On the other hand, the collection $M_2 := \{\delta_{1/n} \mid n \in \mathbb{N}\}$ is tight: the compact interval $[0, 1]$ will work as K_ϵ for any $\epsilon > 0$. In general, a collection of Dirac delta measures on $\mathbb{R}^n$ is tight if and only if the collection of their supports is bounded. □

Definition 10.11. Let μ be a measure on the σ-algebra of Borel sets of a Hausdorff space X.

- μ is called *inner regular* or *tight*, if for any Borel set B,

$$\mu(B) = \sup\{\mu(K) \mid K \subseteq B \text{ and } K \text{ is compact}\}.$$

- μ is called outer *outer regular* if, for any Borel set B,

$$\mu(B) = \inf\{\mu(U) \mid B \subseteq U \text{ and } U \text{ is open}\}.$$

- μ is called *locally finite* if every point X has a neighborhood U for which $\mu(U)$ is finite. (If μ is locally finite, then it follows that μ is finite on compact sets.)

Definition 10.12. A measure μ is called *Radon measure* if it is inner regular and locally finite.

Remark. The Lebesgue measure is a radon measure, instead of writing $\mu(B)$ for $B \in \mathcal{B}\left(\mathbb{R}^2\right)$, we will write

$$|B|.$$

10.3 Random measure and Point process

Definition 10.13. Let M be the collection of radon measures on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$. Set $M_p \subseteq M$ be these $\mu \in M$ such that

$$\mu(A) \in \{0, 1 \cdots, \infty\}, \forall A \in \mathcal{B}(\mathbb{R}^2).$$

Example 10.14. Let $x \in \mathbb{R}^2$. Set $\delta_x \in M_p$ by

$$\delta_x(A) = \begin{cases} 1, & x \in A \\ 0, & x \notin A. \end{cases}$$

Suppose $x_1, x_2, \cdots$ are in $\mathbb{R}^2$. Set

$$\mu(A) = \sum_{k=1}^{\infty} \delta_{x_k}(A).$$

Then μ is a counting measure and $\mu \in M_p$. So for any bounded (finite?) $B \in \mathcal{B}(\mathbb{R})$, $\mu(B) < \infty$. In our compact set, we do not want any accumulation point, so we just consider finite sets.

Let $(\Omega, \mathcal{A}, P)$ be a probability space.

Definition 10.15 (Random measure)**.** Let $\left(\mathbb{R}^2, \mathcal{B}\left(\mathbb{R}^2\right)\right)$ be a measurable space. A *random measure* on $\left(\mathbb{R}^2, \mathcal{B}\left(\mathbb{R}^2\right)\right)$ is a transition kernel from $(\Omega, \mathcal{A})$ into $\left(\mathbb{R}^2, \mathcal{B}\left(\mathbb{R}^2\right)\right)$. More explicitly, a mapping

$$N : \Omega \times \mathcal{B}\left(\mathbb{R}^2\right) \to \overline{R}_+$$

is called a *random measure* if

(a) for any $B \in \mathcal{B}$,

$$\begin{aligned} N(\cdot, B) : \Omega &\to \overline{\mathbb{R}}_+ \\ \omega &\mapsto N(\omega, B) \end{aligned}$$

is a random variable,

(b) for any $\omega \in \Omega$,

$$\begin{aligned} N(\omega, \cdot) : \mathcal{B}\left(\mathbb{R}^2\right) &\to \overline{\mathbb{R}}_+ \\ B &\mapsto N(\omega, B) \end{aligned}$$

is a *radon* measure on $\left(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)\right)$.

Call a function

$$K : \Omega_1 \times \mathcal{B}_2 \to [0, 1]$$

a transition function if

(a) for all $B_2 \in \mathcal{B}_2$,

$$\begin{aligned} K(\cdot, B_2) : \Omega &\to [0, 1] \\ \omega &\mapsto K(\omega, B_2) \end{aligned}$$

is $\mathcal{B}_1/\mathcal{B}([0, 1])$-measurable, (random variable taking value on $[0, 1]$, e.g. $P(Y = 1|X) = f(X)$.) and

(b) for any $\omega \in \Omega$,

$$\begin{aligned} K(\omega, \cdot) : \mathcal{B}_2 &\to [0,1] \\ B_2 &\mapsto K(\omega, B_2) \end{aligned}$$

is a *probability* measure on $(\Omega_2, \mathcal{B}_2)$. (e.g. $P(\cdot|X=1)$ on some sub-σ-field $\mathcal{B}_2$ of $\mathcal{B}_1$ with $\Omega_1 = \Omega_2$.) Transition functions are used to define discrete time Markov processes where $K(\omega_1, B_2)$ represents the conditional probability that, starting from ω_1, the next movement of the system results in a state in $B_2 \in \mathcal{B}_2 \subseteq \mathcal{B}_1$.

Remark. We have the following.

- We may denote $N(\cdot, B)$ as $N(B)$.
- Then we may regard N as the collections of random variables $\left\{N(B), B \in \mathcal{B}\left(\mathbb{R}^2\right)\right\}$.
- We shall denote the radon measure $N(\omega, \cdot)$ as N_ω.
- We may regard N is a random measure that assigns a measure N_ω to every outcome $\omega \in \Omega$. (A random element N of a collection of Radon measures is called a random measure.)
- $N(B)$ is a r.v., meaning for any $B \in \mathcal{B}$ and any $C \in \mathcal{B}(\mathbb{R})$,

$$\{\omega : N(\omega, B) \in C\} \in \mathcal{A}.$$

Let μ be Lebesgue measure on $\mathbb{R}^2$ and X be a nonnegative r.v.,

$$N(\omega, B) = X(\omega)\mu(B).$$

Definition 10.16. A random measure N is a *point process* if for any $\omega \in \Omega$, $N(\omega, \cdot) \in M_p$.

Example 10.17. Let X be a multivariate (or d-variate) r.v. on $\mathbb{R}^2$. Define a point process $N := \delta_X$, where for $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\delta_X(A) = \begin{cases} 1, & \text{if } X \in A; \\ 0, & \text{if } X \notin A. \end{cases}$$

Then for any $B \in \mathcal{B}(\mathbb{R}^2)$,

$$P(N(B) = 1) = P(\delta_X(B) = 1) = P(X \in B) =: \mu(B).$$

Hence $\delta_X(B) = N(B)$ is a r.v. and $N(B) \sim \text{Bernoulli}\,(\mu(B))$. Hence

$$E\left[e^{-s\delta_X(B)}\right] = 1 + \mu(A)\left(e^{-s} - 1\right).$$

Theorem 10.18. *Suppose $X_1, X_2, \ldots$ are r.v.'s in $\mathbb{R}^d$ and τ is a nonnegative integer value r.v., independent of $X_1, X_2, \ldots$. Define*

$$\begin{aligned} N : \mathcal{B}(\mathbb{R}^d) &\longrightarrow \mathbb{Z}^{\geqslant 0} \cup \{\infty\} \\ A &\longmapsto \sum_{i=1}^{\tau} \delta_{X_i}(A). \end{aligned}$$

Then $N(A)$ is a r.v., which counts the number of points $X_1, \ldots, X_\tau$ that belongs to A.

Remark. Regard X_i as the position of the i-th particle.

Example 10.19. Let $X_1, X_2, \ldots, X_n$ be an iid such that $P(X_1 \in A) = \mu(A)$ for any $A \in \mathcal{B}(\mathbb{R}^d)$. Define

$$N(A) := \sum_{i=1}^{n} \delta_{X_i}(A), \forall A \in \mathcal{B}(\mathbb{R}^d).$$

Since $\delta_{X_1}(A), \ldots, \delta_{X_n}(A) \overset{iid}{\sim} \text{Bernoulli}\,(\mu(A))$, we have

$$N(A) = \sum_{i=1}^{n} \delta_{X_i}(A) \sim \text{Binomial}(n, \mu(A)).$$

Moreover, $E[N(A)] = n\mu(A)$, $\text{Var}(N(A)) = n\mu(A)(1-\mu(A))$, and

$$E\left[e^{-sN(A)}\right] = \left[1 + \mu(A)\left(e^{-s} - 1\right)\right]^n.$$

Suppose sets $A_1, \ldots, A_k \in \mathcal{B}(\mathbb{R}^d)$ forms a partition of $\mathbb{R}^d$. Note

$$n = \sum_{i=1}^{n} \delta_{X_i}(\mathbb{R}^d) = N(\mathbb{R}^d) = N\left(\bigsqcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} N(A_i).$$

The random vector $(N(A_1), \ldots, N(A_k))$ has a multinormal distribution with parameters $n, u(A_1), \cdots \mu(A_k)$. (Consider there are k candidates and n voters.) Then for $n_i > 0$ and $n_1 + \cdots + n_k = n$,

$$P\left(N(A_1) = n_1, N(A_k) = n_k\right) = \frac{n!}{n_1! \cdots n_k!} \left[\mu(A_1)\right]^{n_1} \cdots \left[\mu(A_k)\right]^{n_k}.$$

Proof.

$$P(N(A_1) = n_1) = \frac{n!}{n_1!(n-n_1)!} \left[\mu(A_1)\right]^{n_1} \left[1 - \mu(A_1)\right]^{n-n_1}.$$

Given $N(A_1) = n_1$, there are $n - n_1$ candidates remaining, then

$$N(A_2) \sim B\left(n - n_1,\ \frac{\mu(A_2)}{1 - \mu(A_1)}\right),$$

where $\frac{\mu(A_2)}{1-\mu(A_1)} = P(\text{votes for 2} \mid \text{not votes for 1})$. Hence

$$\begin{aligned}
&P\left(N(A_1) = n_1, N(A_2) = n_2\right) \\
&= P\left(N(A_1) = n_1\right) P\left(N(A_2) = n_2 | N(A_1) = n_1\right) \\
&= \frac{n!}{n_1!(n-n_1)!} \left[\mu(A_1)\right]^{n_1} \left[1 - \mu(A_1)\right]^{n-n_1} \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \left[\frac{\mu(A_2)}{1-\mu(A_1)}\right]^{n_2} \left[1 - \frac{\mu(A_2)}{1-\mu(A_1)}\right]^{n-n_1-n_2} \\
&= \frac{n!}{n_1! n_2! (n-n_1-n_2)!} \left[\mu(A_1)\right]^{n_1} \left[\mu(A_2)\right]^{n_2} \left[1 - \mu(A_1) - \mu(A_2)\right]^{n-n_1-n_2}.
\end{aligned}$$

Then use induction to finish the proof. ∎

Remark. $\{N(A_i)\}_{i=1,\ldots,n}$ are not independent.

Example 10.20. Let $X_1, X_2, \ldots$ be iid, and τ a nonnegative integer value r.v., independent of $X_1, X_2, \ldots$. Define

$$\begin{aligned} N : \mathcal{B}(\mathbb{R}^d) &\to \mathbb{Z}^{\geqslant 0} \cup \{\infty\} \\ A &\mapsto \sum_{i=1}^{\tau} \delta_{X_i}(A). \end{aligned}$$

Then $N(A)$ is a random sum of iid Bernoulli r.v.'s. Moreover, $E[N(A)] = E[\tau]\mu(A)$, $\operatorname{Var}(N(A)) = E[\tau]\mu(A)(1-\mu(A)) + \mu^2(A)\operatorname{Var}(\tau)$, and

$$\begin{aligned} E\left[e^{-sN(A)}\right] &= E\left[E\left[e^{-sN(A)}\middle|\,\tau\right]\right] \\ &= E\left(\left[E\left(e^{-s\delta_{X_1}(A)}\right)\right]^{\tau}\right) \\ &= E\left(\left[1+\left(e^{-s}-1\right)\mu(A)\right]^{\tau}\right) \\ &= G_\tau\left(1+\left(e^{-s}-1\right)\mu(A)\right), \end{aligned}$$

since the generating function of τ in terms of z is $G_\tau(z) = E[z^\tau]$. If $A_1, \ldots, A_k$ forms a partition of $\mathbb{R}^d$,

$$\begin{aligned} E\left[\exp\left(-\sum_{i=1}^k s_i N(A_i)\right)\right] &= E\left[E\left[\exp\left(-\sum_{i=1}^k s_i N(A_i)\right)\middle|\,\tau\right]\right] \\ &= E\left[\left(\sum_{i=1}^k e^{-s_i}\mu(A_i)\right)^{\tau}\right] = G_\tau\left(\sum_{i=1}^k e^{-s_i}\mu(A_i)\right), \end{aligned}$$

since the joint laplace transform of $\{N(A_i)\}_{i=1}^k$ given $\tau = n$ is

$$\begin{aligned} & E\left[\exp\left(-\sum_{i=1}^k s_i N(A_i)\right)\middle|\,\tau = n\right] \\ &= \sum \exp\left(-\sum_{i=1}^k s_i n_i\right)\frac{n!}{n_1!\cdots n_k!}\left[\mu(A_1)\right]^{n_1}\cdots\left[\mu(A_k)\right]^{n_k} \\ &= \sum \frac{n!}{n_1!\cdots n_k!}\left[e^{-s_1}\mu(A_1)\right]^{n_1}\cdots\left[e^{-s_k}\mu(A_k)\right]^{n_k} \\ &= \left(\sum_{i=1}^k e^{-s_i}\mu(A_i)\right)^n. \end{aligned}$$

Assume $\tau \sim$Poisson(λ). Then $E[N(A)] = \lambda\mu(A)$ and

$$\operatorname{Var}(N(A)) = \lambda\mu(A)(1-\mu(A)) + \lambda\mu^2(A) = \lambda\mu(A).$$

Since

$$G_\tau(z) = E[z^\tau] = \sum_{\tau=1}^{\infty}\frac{\lambda^\tau}{\tau!}e^{-\lambda}z^\tau = \sum_{\tau=0}^{\infty}\frac{(\lambda z)^\tau}{\tau!}e^{-\lambda} = e^{\lambda z}e^{-\lambda} = e^{\lambda(z-1)},$$

we have

$$E\left[e^{-sN(A)}\right] = G_\tau\left(1+\left(e^{-s}-1\right)\mu(A)\right) = e^{\lambda\left(e^{-s}-1\right)\mu(A)} = e^{\lambda\mu(A)\left(e^{-s}-1\right)}.$$

Thus, $N(A) \sim \text{Poisson}\,(\lambda\mu(A))$. Suppose $A_1, \ldots, A_k$ forms a partition $\mathbb{R}^d$. Then

$$\begin{aligned}
E\left[\exp\left(-\sum_{i=1}^k s_i N(A_i)\right)\right] &= G_\tau\left(\sum_{i=1}^k e^{-s_i}\mu(A_i)\right) \\
&= \exp\left(\lambda\left(\sum_{i=1}^k e^{-s_i}\mu(A_i) - 1\right)\right) \\
&= \exp\left(\lambda\left(\sum_{i=1}^k e^{-s_i}\mu(A_i) - \sum_{i=1}^k \mu(A_i)\right)\right) \\
&= \prod_{i=1}^k e^{\lambda\mu(A_i)\left(e^{-s_i}-1\right)} \\
&= \prod_{i=1}^k E\left[e^{-s_i N(A_i)}\right].
\end{aligned}$$

Thus, $N(A_1), \ldots, N(A_k)$ are independent Poisson r.v., and $N(A_i) \sim \text{Poisson}\,(\lambda\mu(A_i))$.

10.4 Poisson random measure

Let N be a random measure on $(\Omega, \mathcal{F}, P)$ such that for each bounded $A \in \mathcal{B}(\mathbb{R}^d)$, $N(A)(\cdot) \in \mathbb{Z}^{\geqslant 0} \cup \{\infty\}$ on Ω, i.e., $N(A) \in M_p$. The point process

$$N : \Omega \to M_p$$

is called a *Poisson random measure* or *Poisson point process* if

(a) for disjoint **bounded** Borel sets $A_1, \ldots, A_n$, $N(A_1), \ldots, N(A_n)$ are independent random variables,

(b) there exists a measure $\mu : \mathcal{B}(\mathbb{R}^d) \to [0, \infty]$ such that for all bounded $B \in \mathcal{B}(\mathbb{R}^d)$, $\mu(B) < \infty$ and $N(B) \sim$Poisson$(\mu(B))$. The measure μ is called the intensity of N.

Example 10.21. Let $\{X_t, t \geqslant 0\}$ be a homogeneous Poisson process having rate $\lambda > 0$. For $I = (a, b]$, define $N(I) := X_b - X_a$, denoting the number of points in interval I. Define for $I = (a, b]$, $\mu(I) := \lambda(b - a)$.

(a) Let $0 \leqslant a_1 < b_1 \leqslant \cdots \leqslant a_n < b_n < \infty$. For $k = 1, \ldots, n$, set $I_k = (a_k, b_k]$. Since the intevals $I_1, \ldots, I_n$ are pointwise disjoint and a Poisson proces has independent increment, the r.v.'s $X_{b_1} - X_{a_1}, \ldots, X_{b_n} - X_{a_n}$ are independent. So $N(I_1), \ldots, N(I_n)$ are independent random variables.

(b) For $I = (a, b]$, $N(I) \sim \text{Poisson}\,(\lambda(b - a))$, where $\lambda(b - a) = \mu(I)$.

Thus, N is a Poisson random measure. Using the above measure N, μ can be extended to $\mathcal{B}([0,\infty))$ in a way so that N is a Poisson random measure.

Example 10.22. Let $X_1, X_2, \cdots$ be iid d-variate random variables having distribution ν, that is, for any $B \in \mathcal{B}(\mathbb{R}^d)$, $P(X \in B) = \nu(B)$. Let $\tau \sim \text{Poisson}(\lambda)$, independent of $X_1, X_2, \cdots$. Define

$$N : \mathcal{B}(\mathbb{R}^d) \to \mathbb{Z}^{\geqslant 0}$$
$$B \mapsto \sum_{k=1}^{\tau} \delta_{X_k}(B).$$

From Example 10.20, we have N is a Poisson random measure with intensity $\mu = \lambda\nu$, where $\lambda\nu$ denotes a scalar multiple of the measure ν.

Example 10.23. Define

$$I_{n,m} = (n-1, n] \times (m-1, m], \forall n, m \in \mathbb{Z}.$$

Then the sets $\{I_{n,m}\}_{n,m\in\mathbb{Z}}$ form a partition of $\mathbb{R}^2$. Let $\nu_{n,m}$ be a probability distribution on $I_{n,m}$. For example, $\nu_{n,m}$ can be the uniform distribution on $I_{n,m}$. Let $\lambda_{n,m} > 0$. Using the example 10.22,$\forall n, m \in \mathbb{Z}$, there exists a Poisson random measure $N_{n,m}$ on $I_{n,m}$ having indensity

$$\mu_{n,m} = \lambda_{n,m}\nu_{n,m}.$$

Assume that the Poisson random measures $\{N_{n,m}\}_{n,m\in\mathbb{Z}}$ are independent. Now define a process N and a measure μ on $\mathbb{R}^2$ by setting for each $A \in \mathcal{B}(\mathbb{R}^2)$,

$$N(A) = \sum_{n,m} N_{n,m}(A \cap I_{n,m}),$$
$$\mu(A) = \sum_{n,m} \mu_{n,m}(A \cap I_{n,m}).$$

(a) Let $B_1, \ldots, B_k \subseteq \mathcal{B}(\mathbb{R}^2)$ be bounded disjoint. Then for any $n, m \in \mathbb{Z}$, $\{B_1 \cap I_{n,m}, \ldots, B_k \cap I_{n,m}\}$ are disjoint subsets of $I_{n,m}$. So for any $n, m \in \mathbb{Z}$, the r.v.'s $N_{n,m}(B_1 \cap I_{n,m}), \ldots, N_{n,m}(B_k \cap I_{n,m})$ are independent. Since the Poisson random measures $N_{n,m}$ are independent. Since for $i = 1, \ldots, k$,

$$N(B_i) = \sum_{n,m} N_{n,m}(B_i) \cap I_{n,m},$$

and by assumption, $\{N_{n,m}\}_{n,m\in\mathbb{Z}}$ are independent, $N(B_1), \ldots, N(B_k)$ are independent.

(b) Let $B \in \mathcal{B}(\mathbb{R}^2)$ be bounded, then $\{\{n,m\} : B \cap I_{n,m} \neq \emptyset\}$ is finite. Hence

$$N(B) = \sum_{n,m} N_{n,m}(B \cap I_{n,m})$$

is a finite sum of independent Poisson r.v.'s. Since r.v.'s $\{N_{n,m}(B \cap I_{n,m})\}$ are independent and

$$N_{n,m}(B \cap I_{n,m}) \sim \text{Poisson}\left(\mu_{n,m}(B \cap I_{n,m})\right),$$

and the sum of independent Poisson distributed random variables is Poisson whose parameter is sum of the parameters, i.e.,

$$N(B) \sim \text{Poisson}(\mu(B)),$$

where

$$\mu(B) = \sum_{n,m} \mu_{n,m}(B \cap I_{n,m}).$$

Example 10.24 (Cox Random Measure)**.** Let N be a point process on $\mathbb{R}^d$ and let η be a random measure on $\mathbb{R}^d$ such that $\eta(\{x\}) = 0$. For example, for bounded $B \in \mathcal{B}(\mathbb{R}^d)$, define

$$\nu(B) := \int_B \left(\frac{1}{2\pi}\right)^{\frac{d}{2}} e^{-\frac{\|x\|_d^2}{2}} dx,$$

to be the standard normal distribution in $\mathbb{R}^d$. Let Λ be a positive random variable, say $\Lambda \sim \exp(1)$. Define $\eta := \Lambda\nu$, or $\eta(B) = \Lambda\nu(B)$ for any $B \in \mathcal{B}(\mathbb{R}^d)$. Then η is a Cox random measure if conditioned on $\eta = \mu$, N is a Poisson random measure with tensity μ. In this case,

$$\begin{aligned} P(N(B) = k) &= E\left(P(N(B) = k|\eta)\right) \\ &= E\left(\frac{\eta^k(B)}{k!} e^{-\eta(B)}\right). \end{aligned}$$

Then N are said to be conditionally Poisson, or doubly stochastic Poisson, or Cox processes.
In the case when $\eta = \Lambda\nu$ with Λ and ν defined above, we have

$$\begin{aligned} P(N(B) = k) &= E\left(\frac{\Lambda^k \nu^k(B)}{k!} e^{-\Lambda\nu(B)}\right) \\ &= \int_0^\infty \frac{\lambda^k \nu^k(B)}{k!} e^{-\nu(B)\lambda} e^{-\lambda} d\lambda \\ &= \frac{\nu^k(B)}{(\nu(B)+1)^{k+1}} \int_0^\infty \frac{(\nu(B)+1)^{k+1} \lambda^{k+1-1}}{\Gamma(k+1)} e^{-(\nu(B)+1)\lambda} d\lambda \\ &= \left(1 - \frac{\nu(B)}{\nu(B)+1}\right)\left(\frac{\nu(B)}{\nu(B)+1}\right)^k. \end{aligned}$$

Thus, $N(B)$ is geometrically distributed. Since $N(B)$ is not Poisson distributed, N is not a Poisson random measure.

10.5 Integration w.r.t Measure

Let S be a set and $\mathcal{S}$ an sigma algebra on S and μ is a measure on $(S, \mathcal{S})$. (Think of S to be $\mathbb{R}^d$, $\mathcal{S} = \mathcal{B}(\mathbb{R}^d)$ and μ point measure.) The goal is to give meaning to integral of f with repect to the measure μ denoted by

$$\int_S f(x)\mu(dx) = \int f d\mu =: \mu f.$$

Definition 10.25. Let N be a random measure on $(S, \mathcal{S})$.

(a)

$$Nf := \int_S f(x)N(dx),$$

or

$$Nf(\omega) := \int_S f(x)N(\omega, dx), \forall \omega \in \Omega,$$

defines a positive random variable Nf for any $f \in \mathcal{S}/\mathcal{B}(\overline{\mathbb{R}}_+)$.

(b)

$$\mu(B) = E[N(B)] = \int_\Omega N(\omega, B)P(d\omega), \forall B \in \mathcal{S},$$

defines a measure μ on $(S, \mathcal{S})$. Hence

$$\mu(dx) = E[N(dx)] = \int_\Omega N(\omega, dx)P(d\omega).$$

(c) $E[Nf] = \mu f$.

Proof. By Fubini theorem,

$$\mu f = \int_S f(x)\mu(dx) = \int_S f(x) \int_\Omega N(\omega, dx)P(d\omega) = \int_\Omega \int_S f(x)N(\omega, dx)P(d\omega) = E[Nf]. \qquad \square$$

Definition 10.26. Call a function $f : S \to \mathbb{R}$ *measurable* if for any $t \in \mathbb{R}$, $\{x \in S : f(x) \leqslant t\} \in \mathcal{S}$. For example, if $(\Omega, \mathcal{F}, P)$ is a probability space, a measurable function X is just a r.v. and $\{\omega \in \Omega : X(\omega) \leqslant t\} \in \mathcal{F}$. Then taking the probability measure, $P(\{\omega \in \Omega : X(\omega) \leqslant t\}) = F_X(t)$, where $F_X(t)$ is the cdf of X, since $(P \circ X^{-1})((-\infty, t]) = F_X(t)$.

How does one integrate f w.r.t. μ ? How does one construct

$$\int_S f(x)\mu(dx) = \int f d\mu =: \mu f.$$

Recall in calculus, how did we construct $\int_a^b f(x)dx$ for a continuous function $f \geqslant 0$? Thought of the integral as the area under the curve and make approximation using rectangle (if you will) and check that limit exists as one make finite partitions. Let $A \in \mathcal{S}$ and recall

$$\mathbb{1}_A(s) = \begin{cases} 1, & s \in A, \\ 0, & s \in A^c. \end{cases}$$

Then define

$$\int_S \mathbb{1}_A(s)\mu(ds) = E[\mathbb{1}_A] = \mu(A).$$

Let $A_1, \ldots, A_n \in \mathcal{S}$ be pairwise disjoint and let $c_1, \ldots, c_n \in \mathbb{R}$. Define

$$f(x) = \sum_{i=1}^n c_i \mathbb{1}_{A_i}(x),$$

which is a simple function. Then

$$\int_S f(x)\mu(dx) = \int_S \sum_{i=1}^{n} c_i \mathbb{1}_{A_i}(x)\mu(dx) = \sum_{i=1}^{n} c_i \int \mathbb{1}_{A_i}(x)\mu(dx) = \sum_{i=1}^{n} c_i \mu(A_i).$$

Recall $f : \mathcal{S} \to [0, \infty)$ is measurable, then for $a < b$,

$$\{s \in S : a < f(s) \leqslant b\} = \{s \in S : f(s) \leqslant b\} \setminus \{s \in S : f(s) \leqslant a\} \in \mathcal{S}.$$

For $n = 1, 2, \cdots$, construct simple functions $\{f_n\}_{n \in \mathbb{Z}^+}$ such that $f_1 \leqslant f_2 \leqslant \cdots$ and $f_n \uparrow f$. Then we define

$$\int_S f(s)\mu(ds) = \lim_{n\to\infty} \int_S f_n(s)\mu(ds).$$

For $k = 1, \ldots, n2^n$,

$$A_k = \left\{ s \in S : \frac{k-1}{2^n} < f(s) \leqslant \frac{k}{2^n} \right\}.$$

$$A_{n2^n+1} = \{s \in S : f(s) > n\}.$$

Set for $k = 1, \ldots, n2^n$, $c_k = \frac{k-1}{2^n}$, $c_{n2^n+1} = n$. The function

$$f_n(s) = \sum_{k=1}^{n2^n+1} c_k \mathbb{1}_{A_k}(s)$$

is a simple function. When n going from n to $n+1$, each interval $(\frac{k-1}{2^n}, \frac{k}{2^n}]$ gets split into two disjoint intervals

$$\left(\frac{2k-2}{2^{n+1}}, \frac{2k-1}{2^{n+1}} \right] \bigsqcup \left(\frac{2k-1}{2^{n+1}}, \frac{2k}{2^{n+1}} \right].$$

A_k gets split into two disjoint sets

$$f^{-1}\left(\left(\frac{2k-2}{2^{n+1}}, \frac{2k-1}{2^{n+1}} \right] \right) \bigsqcup f^{-1}\left(\left(\frac{2k-1}{2^{n+1}}, \frac{2k}{2^{n+1}} \right] \right).$$

If s is in the former, then $f_{n+1}(s) = f_n(s) = \frac{k-1}{2^n}$. If s is in the latter,

$$f_{n+1}(s) = \frac{2k-1}{2^{n+1}} > f_n(s) = \frac{k-1}{2^n}.$$

Thus for any $s \in S$, $f_n(s) \leqslant f_{n+1}(s)$ and $f_n(s) \uparrow f(s)$. Note $\lim_{n\to\infty} f_n(s)\mu(ds)$ exists but may be ∞. Set

$$\int_S f(s)\mu d(s) = \lim_{n\to\infty} \int f_n(s)\mu(ds),$$

where

$$\int_S f_n(s)\mu(ds) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n}\mu(A_k) + n\mu(A_{n2^n+1}).$$

Let f be a measurable real function on S. Then both f^+ and f^- are nonnegative measurable functions, implying their integrals exist. If at least one of the integrals μf^+ and μf^- is finite, then

$$\int_S f(s)\mu(ds) = \int_S f^+\mu(ds) - \int_S f^-(s)\mu(ds).$$

Example 10.27. Let $S = \mathbb{R}^2$ and $\mathcal{S} = \mathcal{B}(\mathbb{R})$ and $\mu(dx) = g(x)dx$, where g is a nonnegative and piecewise continuous function whose Riemman integral is 1. So g is the density for a r.v. and then μ is a induced probabolity measure. Then for any $A \in \mathcal{B}(\mathbb{R}^2)$,

$$\mu(A) = \int_{\mathbb{R}^2} \mathbb{1}_A(x)\mu(dx) = \int_{\mathbb{R}^2} \mathbb{1}_A(x)g(x)dx = P(X \in A).$$

Now if f is an integrable function w.r.t μ,

$$\int_{\mathbb{R}} f(x)\mu(dx) = \int_R f(x)g(x)dx = E[f(x)].$$

(Think of the density $g(x)$ of the Lebegue measure is 1, which implies the random variable is unformly distributed.) Then

$$\lambda(dx) = 1 \cdot dx = dx \text{ and } \int_{\mathbb{R}} 1dx = \mu(\mathbb{R}) = \infty,$$

in space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ and

$$E[f] = \int_{\mathbb{R}} f d\lambda = \int_{\mathbb{R}} f\lambda(dx) = \int_{\mathbb{R}} f \cdot 1 \cdot dx = \int f dx.$$

Example 10.28. Let $(S, \mathcal{S})$ be a measurable space. Let $x \in S$, for measurable $f \geqslant 0$ and f is defined at x, find $\int_S f(t)\delta_x(dt)$. Recall the measure $\delta_x \in M_p$ and

$$\delta_x(A) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

Let A in $\mathcal{S}$. If $f(x) = \mathbb{1}_A(x)$,

$$\int_S f(s)\delta_x(ds) = \int_S \mathbb{1}_A(s)\delta_x(ds) = \delta_x(A) = \mathbb{1}_A(x) = f(x).$$

Example 10.29. Let $A_1, \ldots, A_n \in \mathcal{S}$ be pairwise disjoint, and let $c_1, \ldots, c_n \in [0, \infty)$, set

$$f_n(x) = \sum_{i=1}^n c_i \mathbb{1}_{A_i}(x).$$

If $x \notin \sqcup_{i=1}^n A_i$,

$$\int_S f_n(t)\delta_x(dt) = \sum_{i=1}^n c_i\delta_x(A_i) = 0 = f_n(x).$$

If $x \in A_j$ for some j,

$$\int_S f_n(t)\delta_x(dt) = \sum_{i=1}^n c_i\delta_x(A_j) = c_j = f_n(x).$$

Like before, let

$$f_n(x) = \sum_{k=1}^{n2^n+1} \frac{k-1}{2^n}\mathbb{1}_{A_k}(x) + n\mathbb{1}_{A_{n2^n+1}}(x) \uparrow f(x).$$

Assume f is defined at x. If $\frac{k-1}{2^n} < f(x) \leqslant \frac{k}{2^n}$ for some $k = 1, \ldots, n2^n$, then

$$\int_S f_n(t)\delta_x(dt) = \frac{k-1}{2^n}\delta_x\left(f^{-1}\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right]\right) = \frac{k-1}{2^n} = f_n(x).$$

If $f(x) > n$,

$$\int_S f_n(x)\delta_x(dt) = n\delta_x\left(f^{-1}(n, \infty)\right) = n = f_n(x).$$

Thus,

$$\int_S f_n(t)\delta_x(dt) = f_n(x).$$

Then by MCT,

$$\int_S f(t)\delta_x(dt) = \lim_{n\to\infty}\int_S f_n(t)\delta_x(dt) = \lim_{n\to\infty} f_n(x) = f(x),$$

when a f is nonnegative and measurable function.

Example 10.30. If f is measurable and f is defined at x,

$$\int_S f(t)\delta_x(dt) = \delta_x(f^+) - \delta_x(f^-) = f^+(x) - f^-(x) = (f^+ - f^-)(x) = f(x),$$

since $f^+(x)$ or $f^-(x)$ must be 0. Theorefore, the integration of a function f w.r.t. to the point measure δ_x is the evaluation of the function f at that point x, i.e., $f(x)$.

Example 10.31. Let $x_1, \ldots, x_n$ be points in S and define a measure μ by

$$\mu(A) = \sum_{k=1}^{n} \delta_{x_k}(A), \forall A \in \mathcal{S},$$

which counts the number of $x_1, \ldots, x_n$ that are in the set A. So μ is a counting measure. Then for measurable f and f is defined at $x_1, \ldots, x_k$,

$$\mu f := \int_S f(t)\mu(dt) = \sum_{k=1}^{n}\int_S f(t)\delta_{x_k}(dt) = \sum_{k=1}^{n} f(x_k).$$

Example 10.32. Let N be a point process on $(S, \mathcal{S})$. For measurable $f \geqslant 0$, the *Laplace functional* of N is given by

$$L_N(f) := E\left[e^{-Nf}\right] = E\left[\exp\left(-\int f(t)N(dt)\right)\right].$$

If μ is a measure,

$$\int_S f(x)\mu(dx) \in \mathbb{R}.$$

$$\int_S (af(x) + bg(x)\mu(dx) = a\int f(x)\mu(dx) + b\int g(x)\mu(dx),$$

where we think of the integral w.r.t μ as a linear functional on the msble functions on $\mathcal{S}$.

Example 10.33. Let X be an S-valued random variable having distribution ν, that is $\nu(B) = P(X \in B)$. Define a point process $N := \delta_X$. Let f be measurable. Since $\int_S f(x)\delta_X(dx) = f(X)$, we have the Laplace functional of f is

$$L_{\delta_X}(f) = E\left[\exp\left(-\int f(x)\delta_X(dx)\right)\right] = E\left[\exp\left(-f(X)\right)\right] = \int_S e^{-f(x)}\nu(dx).$$

If $S = [0,\infty)$ and $f(x) = s \cdot x$ for some $s > 0$, then

$$L_{\delta_X}(f) = \int_0^\infty e^{-st}\nu(dt) = E[e^{-sX}],$$

which is just the Laplace transform of X.

Example 10.34. Let $X_1, \dots, X_n$ be iid S-valued r.v.'s with distribution ν. Define a point process N by

$$N = \sum_{k=1}^n \delta_{X_k}.$$

For $f \geqslant 0$ and measurable,

$$L_N(f) = e^{-\int_S f(x)N(dx)} = E\left[\exp\left(-\sum_{k=1}^n \int_S f(x)\delta_{X_k}(dx)\right)\right] = E\left[\exp\left(-\sum_{k=1}^n f(X_k)\right)\right]$$
$$= \prod_{k=1}^n E\left[e^{-f(X_k)}\right] = \left(\int_S e^{-f(x)}\nu(dx)\right)^n = \left(\nu e^{-f}\right)^n.$$

Example 10.35. Let τ be a nonnegative S-valued r.v., independent of $X_1, \dots, X_n$. Set

$$N = \sum_{k=1}^\tau \delta_{X_k}.$$

Let $f \geqslant 0$ be measurable and consider

$$E[e^{-Nf}|\tau = n] = \left(\nu e^{-f}\right)^\tau.$$

Then

$$L_N(f) = E\left[e^{-Nf}\right] = E\left[E\left[e^{-Nf}\middle|\,\tau\right]\right] = E\left[\left(\nu e^{-f}\right)^\tau\right].$$

Let G_τ be the probability generating function for τ, i.e., $G_\tau(z) = E[z^\tau]$. Then

$$L_N(f) = G_\tau\left(\nu e^{-f}\right).$$

If, for example, $\tau \sim$ Poisson (1), $G_\tau(z) = e^{-\lambda(1-z)}$. Then

$$L_N(f) = G_\tau\left(\nu e^{-f}\right) = e^{-\lambda\left(1-\nu e^{-f}\right)}.$$

Since

$$\nu 1 = \int_S 1 \cdot \nu(dx) = \int_S \nu(dx) = \nu(S) = 1,$$

we have $L_N(f) = e^{-\lambda\nu\left(1-e^{-f}\right)}$. Let $m = \lambda\nu$, then $L_N(f) = e^{-m\left(1-e^{-f}\right)}$. Suppose $X \sim$ Poisson(λ), then

$$E[e^{-sX}] = \sum_{k=0}^{\infty} e^{-sk}\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda\left(1-e^{-s}\right)}.$$

Example 10.36. A poisson random measure on $(S, \mathcal{S})$ is a point process N so that $A_1, \ldots, A_n$ are pairwise disjoint in $\mathcal{S}$. Then $N(A_1), \ldots, N(A_n)$ are independent poisson distributed r.v.'s with parameter $m(A_1), \ldots, m(A_n)$, respectively, where the intensity m is a measure on $(S, \mathcal{S})$. Suppose $f = \mathbb{1}_A$ for $A \in \mathcal{S}$ and $N(A) \sim$Poisson($m(A)$), then

$$Nf = N\mathbb{1}_A = \int_S \mathbb{1}_A(x)N(dx) = N(A).$$

In this case,

$$L_N(\mathbb{1}_A) = E\left[e^{-Nf}\right] = E\left[e^{-N(A)}\right] = e^{-m(A)\left(1-e^{-1}\right)}. \text{ (Let } s = 1).$$

Note

$$m\left(1-e^{-\mathbb{1}_A}\right) = \int_S \left(1-e^{-\mathbb{1}_A(x)}\right)m(dx).$$

If $x \notin A$, $1-e^{-\mathbb{1}_A(x)} = 1-e^{-0} = 0$. If $x \in A$, $1-e^{-\mathbb{1}_A(x)} = 1-e^{-1}$. Then

$$m\left(1-e^{-\mathbb{1}_A}\right) = \int_A \left(1-e^{-1}\right)m(dx) = \left(1-e^{-1}\right)m(A).$$

Thus, $L_N(\mathbb{1}_A) = e^{-m\left(1-e^{-\mathbb{1}_A}\right)}$. In this case, $L_N(f) = e^{-m\left(1-e^{-f}\right)}$.

Example 10.37. Let $A_1, \ldots, A_n$ be disjoint sets in $\mathcal{S}$ and let $c_1, \ldots, c_n$ be positive number, and

$$f(x) = \sum_{k=1}^{n} c_k \mathbb{1}_{A_k}(x).$$

Then $Nf = \int_S f(x)N(dx) = \sum_{k=1}^n c_k N(A_k)$. Then

$$\begin{aligned} L_N(f) = E\left[e^{-Nf}\right] &= \prod_{k=1}^{n} E\left[e^{-c_k N(A_k)}\right] = \exp\left(-\sum_{k=1}^{n} m(A_k)\left(1-e^{-c_k}\right)\right) \quad (s = c_k) \\ &= \exp\left(-\sum_{k=1}^{n} m\left(1-e^{-c_k\mathbb{1}_{A_k}}\right)\right) = e^{m\left(1-e^{-f}\right)}??? \end{aligned}$$

10.6 Kernel

Definition 10.38. Let $(S, \mathcal{S})$ and $(T, \mathcal{T})$ be two meaurable Polish spaces, e.g., $S = \mathbb{R}^d$.
Kernel is a mapping

$$\nu : S \times \mathcal{T} \to \mathbb{R}_+$$

satisfying

(a) for any $s \in S$, $\nu(s, \cdot)$ is a measure on $\mathcal{T}$.

(b) for any $B \in \mathcal{T}$, $\nu(\cdot, B)$ is a mesurable function on S.

Example 10.39. Let X be S-valued and Y be T-valued random variables, respectively. Let μ be the distribution of X, i.e.,

$$\mu(A) = P(X \in A), \forall A \in \mathcal{S}.$$

Let $A \in S$ and $B \in T$. Then

$$\begin{aligned} P(X \in A, Y \in B) &= E[\mathbb{1}_A(X)\mathbb{1}_B(Y)] \\ &= E[E[\mathbb{1}_A(X)\mathbb{1}_B(Y)|X]] \\ &= E[\mathbb{1}_A(X)E[\mathbb{1}_B(Y)|X]] \\ &= E[\mathbb{1}_A(X)P(Y \in B|X)] \end{aligned}$$

If S is nice enough, there exists a kernel ν such that

$$\nu(s, B) = P(Y \in B|X = s).$$

Then $\nu(X, B)$ is a random variable and

$$P(X \in A, Y \in B) = E[\mathbb{1}_A(X)\nu(X, B)] = \int_S \mathbb{1}_A(s)\nu(s, B)\mu(ds) = \int_A \nu(s, B)\mu(ds).$$

Recall the discrete case,

$$\begin{aligned} P(X \in A, Y \in B) &= \sum_{i \in A} \sum_{j \in B} P(X = i, Y = j) \\ &= \sum_{i \in A} \sum_{j \in B} P(Y = j|X = i)P(X = i) \\ &= \sum_{i \in A} P(Y \in B|X = i)P(X = i). \end{aligned}$$

So we can regard the sum as the integral, $P(Y \in B|X = i)$ as $\nu(s, B)$ and $P(X = i)$ as $\mu(ds)$.

10.7 Randomization of a Point Process

Let N be a point process on $(S, \mathcal{S})$. Write

$$N(A) = \sum_k \delta_{\xi_k}(A), \forall A \in \mathcal{S}.$$

Here $\xi_1, \xi_2, \cdots$ are S-valued random variables and ξ_k can be thought of the location of the k-th points. Now let $\tau_1, \tau_2, \cdots$ be a sequence of random variables on $(T, \mathcal{T})$, which are conditionally independent given N and such that

$$P(\tau_k \in B|N) = P(\tau_k \in B|\xi_k) = \nu(\xi_k, B), \forall B \in \mathcal{T}.$$

Here when $\xi_k = s$, $P(\tau_k|\xi_k = s)$ is distribution for τ_k, i.e.,

$$P(\tau_k \in B|\xi_k = s) \in \mathbb{R}_+, \forall B \in \mathcal{T}.$$

Example 10.40. Let N be a Poisson random measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with intensity m, which is Lebesgue measure. Then

$$E[N(a,b)] = m(b-a) = b-a.$$

Think of N as being the arrival process to a queue system.

$$N(A) = \sum_{k \in \mathbb{Z}} \delta_{\xi_k}(A), \forall A \in \mathcal{S}.$$

Let $\{\tau_k\}_{k\in\mathbb{Z}}$ be iid random variables independent of N that are nonnegative. Think of τ_k as the amount of work brought by the arrival at ξ_k. Construct a new point process

$$N_\nu(A \times B) = \sum_k \delta_{(\xi_k,\tau_k)}(A \times B)$$

for $A \times B \in \mathcal{S} \times T$, i.e., N_ν is a point process on the product space $(S \times T, \mathcal{S} \times \mathcal{T})$.

Let μ_k be the distribution of ξ_k. Since given N, τ_k just depends on ξ_k,

$$\begin{aligned} P\left(\delta_{(\xi_k,\tau_k)}(A,B) = 1\right) &= P(\xi_k \in A, \tau_k \in B) \\ &= E[E[P(\xi_k \in A, \tau_k \in B)|N]] \\ &= E[E[\mathbb{1}_A(\xi_k)\mathbb{1}_B(\tau_k)|N]] \\ &= E[\mathbb{1}_A(\xi_k)E[\mathbb{1}_B(\tau_k)|N]] \\ &= E[\mathbb{1}_A(\xi_k)P(\tau_k \in B|N)] \\ &= E[\mathbb{1}_A(\xi_k)\nu(\xi_k, B)] \\ &= \int_S \nu(s,B)\mu_k(ds)\mathbb{1}_A(s) \\ &= \int_A \nu(s,B)\mu_k(ds). \end{aligned}$$

Define a (product) measure γ on $(S \times T, \mathcal{S} \times \mathcal{T})$ by

$$\begin{aligned} \gamma(A \times B) &= \int_{A\times B} \gamma(ds,dt) = \int_A \nu(s,B)\mu(ds) = \int_A \int_B \nu(s,dt)\mu(ds) \\ &= \int_S \int_T \nu(s,dt)\mu(ds)\mathbb{1}_A(s)\mathbb{1}_B(t), \forall A \times B \in \mathcal{S} \times \mathcal{T}. \end{aligned}$$

Suppose $f(X,Y) = g(X)h(Y)$, where g and h are nonnegative. Then

$$\begin{aligned} E[f(X,Y)] &= E[g(X)h(Y)] = E[E[g(X)h(Y)|X]] = E[g(X)E[h(Y)|X]] \\ &= E\left[g(X)\int_T h(t)\nu(X,dt)\right] = \int_T E\left[g(X)h(t)\nu(X,dt)\right] \\ &= \int_T \int_S g(s)h(t)\nu(s,dt)\mu(ds) = \int_S \int_T f(s,t)\nu(s,dt)\mu(ds). \\ &= \int_{S\times T} f(s,t)\gamma(ds,dt). \end{aligned}$$

Let μ be the intensity of N and μ_k be the distribution of ξ_k. Recall that μ is a measure on $(S, \mathcal{S})$ defined by $\mu(A) = E[N(A)]$. Thus, for nonnegative measurable f on $(S, \mathcal{S})$,

$$\begin{aligned}
\mu f &= \int_S f(s)\mu(ds) = \int_S f(s) E[N(ds)] \\
&= E\left[\int_S f(s) N(ds)\right] = E[Nf] \\
&= E\left[\sum_k \delta_{\xi_k} f\right] = E\left[\sum_k f(\xi_k)\right] = \sum_k E[f(\xi_k)] \\
&= \sum_k \int_S f(s)\mu_k(ds).
\end{aligned}$$

So $\mu f = \sum_k \mu_k f$. Note $\sum_k E[\mathbb{1}_{A_k}] = \sum_k P(A_k)$, $E[\delta_k(A)] = P(\xi_k \in A)$, and

$$\mu(A) = E[N(A)] = \sum_k E[\delta_{\xi_k}(A)] = \sum_k P(\xi_k \in A) = \sum_k \mu_k(A).$$

Since $\mu f = \sum_k \mu_k f$, $\mu\nu(\cdot, B) = \sum_k \mu_k \nu(\cdot, B)$.

$$\begin{aligned}
E[N_\nu(A \times B)] &= E\left[\sum_k \delta_{(\xi_k, \tau_k)}(A \times B)\right] = \sum_k E\left[\delta_{\xi_k}(A)\delta_{\tau_k}(B)\right] = \sum_k E\left[\mathbb{1}_A(\xi_k)\mathbb{1}_B(\tau_k)\right] \\
&= \sum_k P(\xi_k \in A, \tau_k \in B) = \sum_k \int_S \nu(s, B)\mu_k(ds)\mathbb{1}_A(s) = \sum_k \int_A \nu(s, B)\mu_k(ds) \\
&= \int_A \nu(s, B)\mu(ds) = \gamma(A \times B).
\end{aligned}$$

Suppose $f : S \times T \to \mathbb{R}_+$ is measurable. Then $E[f(\xi_k, \tau_k)|\xi_k] = \int_T f(\xi_k, t)\nu(\xi_k, dt)$. The above equation is a function of the random variable ξ_k and for notational convenience, we write it as $E[f(\xi_k, \tau_k)|\xi_k] = \hat{\nu} f(\xi_k)$. The ν-randomization of the point process N is the point process N_ν on the product space $(S \times T, \mathcal{S} \times \mathcal{T})$ given by

$$N_\nu(A \times B) = \sum_k \delta_{(\xi_k, \tau_k)}(A \times B).$$

Let $f : S \times T \to \mathbb{R}_+$ is measurable. Note

$$\begin{aligned}
N_\nu f &= \int_{S \times T} f(s, t) N_\nu(ds, dt) = \int_{S \times T} f(s, t) \sum_k \delta_{(\xi_k, \tau_k)}(ds, dt) \\
&= \sum_k \int_{S \times T} f(s, t)\delta_{(\xi_k, \tau_k)}(ds, dt) = \sum_k f(\xi_k, \tau_k).
\end{aligned}$$

Since τ_k's are conditionally independent given N and ξ_k's are known given N,

$$\begin{aligned} E\left[e^{-N_\nu f}\middle| N\right] &= E\left[\exp\left(-\sum_k f(\xi_k,\tau_k)\right)\middle| N\right] \\ &= \prod_k E\left[\exp\left(-f(\xi_k,\tau_k)\right)\middle| N\right] \\ &= \prod_k E\left[\exp\left(-f(\xi_k,\tau_k)\right)\middle| \xi_k\right] \\ &= \prod_k \hat{\nu}\left(e^{-f}\right)(\xi_k) \\ &= \exp\left(\sum_k \log\left(\hat{\nu}\left(e^{-f}\right)(\xi_k)\right)\right). \end{aligned}$$

Hence

$$L_{N_\nu}(f) = E\left[E\left[e^{-N_\nu f}\middle| N\right]\right] = E\left[\exp\left(\sum_k \log\left(\hat{\nu}\left(e^{-f}\right)(\xi_k)\right)\right)\right] = L_N\left(-\log\left(\hat{\nu}\left(e^{-f}\right)\right)\right),$$

since $L_N(f) = E\left[\exp\left(-\sum_k f(\xi_k)\right)\right]$. Thus, the Laplace functional of N_ν can be written in terms of Laplace functional of N and the kernel ν.

Example 10.41. Suppose N is a Poisson random measure with intensity μ. Oberve that

$$\mu\hat{\nu}f = \int_S \left[\hat{\nu}f(s)\right]\mu(ds) = \int_S\int_T f(s,t)\nu(s,dt)\mu(ds).$$

Since $L_N(f) = e^{-\mu(1-e^{-f})}$, and $\hat{\nu}\cdot 1 = E[1|\xi_k] = 1$, we have

$$\begin{aligned} L_{N_\nu}(f) &= L_N\left(-\log\left(\hat{\nu}\left(e^{-f}\right)\right)\right) = \exp\left(-\mu\left(1-\hat{\nu}\left(e^{-f}\right)\right)\right) = \exp\left(-\mu\hat{\nu}\left(1-\left(e^{-f}\right)\right)\right) \\ &= \exp\left[-\int_S\int_T\left(1-e^{-f(s,t)}\right)\nu(s,dt)\mu(ds)\right] = \exp\left[-\gamma\left(1-e^{-f(s,t)}\right)\right]. \end{aligned}$$

It follows that N_ν is a Poisson random measure on the product space $(S\times T, \mathcal{S}\times\mathcal{T})$ having intensity $\gamma(ds,dt) = \nu(s,dt)\mu(ds)$.

Example 10.42. Let $A\in\mathcal{S}$ and $B\in\mathcal{T}$.
Note

$$\begin{aligned} E[N_\nu(A\times B)|N] &= E\left[\sum_k \delta_{(\xi_k,\tau_k)}(A\times B)\middle| N\right] = E\left[\sum_k \delta_{\xi_k}(A)\delta_{\tau_k}(B)\middle| N\right] \\ &= \sum_k E\left[\delta_{\xi_k}(A)\delta_{\tau_k}(B)\middle| N\right] = \sum_k \delta_{\xi_k}(A)P(\tau_k\in B|\xi_k) \\ &= \sum_k \delta_{\xi_k}(A)\nu(\xi_k,B). \end{aligned}$$

Also, if $f(s) = \mathbb{1}_A(s)\nu(s,B)$,

$$\begin{aligned}(Nf = \int_S f(s)N(ds) =) \int \mathbb{1}_A(s)\nu(s,B)N(ds) \\ = \sum_k \int_S \mathbb{1}_A(s)\nu(s,B)\delta_{\xi_k}(ds) = \sum_k \int_A \nu(s,B)\delta_{\xi_k}(ds) \\ = \sum_k \int_A \nu(\xi_k,B)\delta_{\xi_k}(ds) = \sum_k \delta_{\xi_k}(A)\nu(\xi_k,B).\end{aligned}$$

Then

$$\begin{aligned}E[N_\nu(A\times B)] = E\left[\int_A \nu(s,B)N(ds)\right] = \int_A \nu(s,B)\mu(ds)\ (\star) \\ = \int_A\int_B \nu(s,dt)\mu(ds) = \int_S\int_T \mathbb{1}_{A\times B}\nu(s,dt)\mu(ds),\end{aligned}$$

where $\gamma(ds,dt) = \nu(s,dt)\mu(ds)$ is the joint distribution of ξ_k and τ_k.

Example 10.43. Let $N = \sum_k \delta_{\xi_k}$ be a Poisson random measure and partition the random measure N into j groups $N_1,\ldots,N_J$. An arrival ξ_k is put into group j w/prob $P_j(\xi_k)$. Let τ_k be random variables taking values in the sets $T = \{e_1,\ldots,e_j\}$, where e_k is the kth unit vector such that

$$P(\tau_k = e_j|N) = \gamma(\xi_k,\{e_j\}) = P_j(\xi_k).$$

Let N_ν be the randomization

$$N_\nu = \sum_k \delta(\xi_k,\tau_k)$$

Let $A\in\mathcal{S}$ and consider the set $A\times\{e_j\}$,

$$N_\nu(A\times\{e_j\}) = \sum_k \delta_{(\xi_k,\tau_k)}(A\times\{e_j\}) := N_j(A),$$

the number of points in group j. Let $f: S\times T\to[0,\infty)$ be measurable. Set $f_j(s) = f(s,e_j)$. Then

$$\begin{aligned}N_\nu f = \sum_k \delta(\xi_k,\tau_k)f = \sum_k f(\xi_k,\tau_k) = \sum_k\sum_{j=1}^J f(\xi_k,e_j)\mathbb{1}_{\{\tau_k=e_j\}} = \sum_k\sum_{j=1}^J f_j(\xi_k)\mathbb{1}_{\{\tau_k=e_j\}} \\ = \sum_{j=1}^J\sum_k f_j(\xi_k)\mathbb{1}_{\{\tau_k=e_j\}} = \sum_{j=1}^J\sum_{k=1} (\delta_{\xi_k}f_j)\,\mathbb{1}_{\{\tau_k=e_j\}} = \sum_{j=1}^J N_j f_j.\end{aligned}$$

Hence $E[N_\nu f] = \sum_{j=1}^J E[N_j f_j]$. Note

$$\begin{aligned}
E[N_j f_j] &= E\left[\sum_k f_j(\xi_k)\mathbb{1}_{\{\tau_k=e_j\}}\right] = E\left[E\left[\sum_k f_j(\xi_k)\mathbb{1}_{\{\tau_k=e_j\}}\middle| N\right]\right] \\
&= E\left[\sum_k f_j(\xi_k)E\left[\mathbb{1}_{\{\tau_k=e_j\}}\middle| N\right]\right] = E\left[\sum_k f_j(\xi_k)P_j(\xi_k)\right] \\
&= E\left[\sum_k \delta_{\xi_k}(f_j\cdot P_j)\right] = E[Nf] \\
&= E\left[\int_S (f_j\cdot P_j)(s)N(ds)\right] = \int_S f_j(s)P_j(s)\mu(ds) \\
&= [P_j(\cdot)\mu]\, f_j
\end{aligned}$$

So the intensity of the jth point process is $P_j(\cdot)\mu$, where μ is the intensity of N. Note

$$E[N_\nu f] = \sum_{j=1}^J P_j(\cdot)\mu] f_j.$$

So the intensity of N_ν is $\sum_{j=1}^J P_j(\cdot)\mu = \mu$. Next, since N_ν has intensity $\gamma(ds,dt) = \mu(ds)\nu(s,dt)$,

$$\begin{aligned}
E\left[\prod_{j=1}^J e^{-N_j f_j}\right] &= E\left[e^{-\sum_{j=1}^J N_j f_j}\right] = E\left[e^{-N_\nu f}\right] \\
&= \exp\left[-\int_S\int_T \nu(s,dt)\mu(ds)\left(1-e^{-f(s,t)}\right)\right] \\
&= \exp\left[-\int_S\sum_{j=1}^J \nu(s,e_j)\mu(ds)\left(1-e^{-f(s,e_j)}\right)\right] \\
&= \exp\left[-\sum_{j=1}^J\int_S P_j(s)\mu(ds)\left(1-e^{-f_j(s)}\right)\right] \\
&= \prod_{j=1}^J \exp\left[-\int_S P_j(s)\mu(ds)\left(1-e^{-f_j(s)}\right)\right].
\end{aligned}$$

Since

$$\exp\left[-\int_S P_j(s)\mu(ds)\left(1-e^{-f_j(s)}\right)\right]$$

is the Laplace functional of a Possion random measure with intensity $P_j(s)\mu(ds)$, we get $N_1,\ldots,N_j$ are independent Poisson random measure with intensity $P_j(\cdot)\mu$.

10.7.1 $M/G/\infty$

Arrival forms a Poisson process $\{N_t, t\geqslant 0\}$ having rate $\lambda>0$. Service time $s_1, s_2, \cdots$ are iid having df F. Fix a time t, an arrival at time ξ_k is still in the system if $\xi_k + s_k > t$ and has departed if

$\xi_k + s_k \leqslant t$. Let

$$N = \{N_s; 0 \leqslant s \leqslant t\}.$$

Let F be the distribution of S_1 and set

$$P_1(\xi_k) = 1 - F(t - \xi_k) = P(s_1 + \xi_k > t),$$

$$P_2(\xi_k) = F(t - \xi_k) = P(s_1 + \xi_k \leqslant t).$$

Then $P_1(\xi_k)$ is the prob. a customer which arrives at time ξ_k is still in the system at time t. $P_2(\xi_k)$ is the prob. a customer departs by time given arrival at ξ_k. Then defining N_1 and N_2 as the number that belongs to group i, we get N_1 and N_2 are independent Poisson random measures with intensities $\lambda(1 - F(s))ds$ and $\lambda F(s)ds$. The expected number of customer still in system at time t has expectation

$$E\left[N_1(0,t]\right] = \int_0^t \lambda(1 - F(t-s))ds = \lambda \int_0^t [1 - F(s)]ds,$$

and similarly, $E\left[N_2(0,t]\right] = \lambda \int_0^t F(s)ds$.

10.8 Transformation of random measures

Definition 10.44. Suppose $(S, \mathcal{S})$ and $(T, \mathcal{T})$ are measurable space. A mapping $f : S \to T$ is said to be measurable if for any $B \in \mathcal{T}$,

$$f^{-1}(B) = \{s \in S | f(s) \in B\} \in \mathcal{S}.$$

Remark. If $\{B_n\}_{n\in\mathbb{Z}^+} \subseteq \mathcal{T}$ are disjoint, then $\left\{f^{-1}(B_n)\right\}_{n\in\mathbb{Z}^+}$ are also disjoint.

Proof. Assume $x = f^{-1}(\mathcal{B}_1) = f^{-1}(\mathcal{B}_2)$. Then $f(x) \in \mathcal{B}_1$ and $f(x) \in \mathcal{B}_2$, which is contradicted by $\mathcal{B}_1$ and $\mathcal{B}_2$ are disjoint. □

Theorem 10.45. *Let μ be a measure on $(S, \mathcal{S})$. Given a measurable mappping $f : S \to T$ and $\mathcal{T} = \sigma(T)$. Define*

$$\begin{aligned} \mu \circ f^{-1} : \mathcal{T} &\to \mathbb{R} \\ B &\to \mu\left(f^{-1}(B)\right) \end{aligned}$$

Then $\mu \circ f^{-1}$ is a measure on $(T, \mathcal{T})$.

Proof. (a)

$$\mu \circ f^{-1}(\emptyset) = \mu\left(f^{-1}(\emptyset)\right) = \mu(\emptyset) = 0.$$

(b) Let $\{B_n\}_{n\in\mathbb{Z}^+} \subseteq \mathcal{T}$ be disjoint. Since μ is σ-addtive,

$$(\mu \circ f^{-1})\left(\bigsqcup_{n=1}^{\infty} B_n\right) = \mu\left(f^{-1}\left(\bigsqcup_{n=1}^{\infty} B_n\right)\right) = \mu\left(\bigsqcup_{n=1}^{\infty} f^{-1}(B_n)\right) = \sum_{n=1}^{\infty} (\mu \circ f^{-1})(B_n).$$ □

Example 10.46. Let $X : \Omega \to \mathbb{R}$ be a random variable. Then $P \circ X^{-1} : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$ is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Since

$$P \circ X^{-1}(\mathbb{R}) = P\left(X^{-1}(\mathbb{R})\right) = P(\Omega) = 1,$$

$P \circ X^{-1}$ is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, which is called a distribution of X.

Theorem 10.47. *Let g be a nonnegative and measurable function on T. That is*

$$g : T \to \mathbb{R}_+$$

so that $g^{-1}(B) \in \mathcal{T}$ for all $B \in \mathcal{B}(\mathbb{R}_+)$. Then

$$\int_T g(t) \left[\mu \circ f^{-1}\right](dt) = \int_S g \circ f(s)\mu(ds) = \int_S g\left(f(s)\right)\mu(ds).$$

$$\mathbb{1}_{f^{-1}(A)}(s) = \begin{cases} 1, & s \in f^{-1}(A), \\ 0, & s \notin f^{-1}(A) \end{cases} = \begin{cases} 1, & f(s) \in A, \\ 0, & f(s) \notin A. \end{cases} = \mathbb{1}_A f(s).$$

Proof. Let $g = \mathbb{1}_A$. Then

$$\begin{aligned}\int_T \mathbb{1}_A(t) \ \left[\mu \circ f^{-1}\right](dt) &= \mu \circ f^{-1}(A) = \mu\left(f^{-1}(A)\right) = \int_S \mathbb{1}_{f^{-1}(A)}\mu(ds) \\ &= \int_S \mathbb{1}_A(f(s))\mu(ds) = \int_S [\mathbb{1}_A \circ f](s)\mu(ds).\end{aligned}$$

Set $g = \sum_{i=1}^n c_i \mathbb{1}_{A_i}$, where $c_i \in \mathbb{R}$ and $A_i \in \mathcal{S}$ for $i = 1, \dots, n$. Then

$$\begin{aligned}\int_T g(t)\left[\mu \circ f^{-1}\right](dt) &= \sum_{i=1}^n c_i \int_T \mathbb{1}_{A_i}\left[\mu \circ f^{-1}\right](dt) \\ &= \sum_{i=1}^n c_i \int_S \left[\mathbb{1}_{A_i} \circ f\right](s)\mu(ds) \\ &= \int_S \left(\sum_{i=1}^n c_i \mathbb{1}_{A_i}\right) f(s)\mu(ds) \\ &= \int_S g \circ f(s)\mu(ds).\end{aligned}$$

Let $g \geqslant 0$ be measurable, then $\exists \{g_n\}_{n \in \mathbb{Z}^+}$ simple such that $g_n \uparrow g$. Then $g_n \circ f^{-1} \uparrow g \circ f^{-1}$. By MCT,

$$\begin{aligned}\int_S g \circ f(s)\mu(ds) &= \int_S \lim_{n\to\infty} g_n \circ f(s)\mu(ds) \\ &= \lim_{n\to\infty} \int_S g_n \circ f(s)\mu(ds) \\ &= \lim_{n\to\infty} \int_T g_n(t)\left[\mu \circ f^{-1}\right](dt) \\ &= \int_T g(t)\left[\mu \circ f^{-1}\right](dt).\end{aligned}$$

□

Example 10.48. By laws of the unconscious statistician,

$$E[g(x)] = \int_\Omega g(X(\omega))P(d\omega) = \int_\mathbb{R} g(x)\left[P\circ X^{-1}\right](dx) = \int_\mathbb{R} g(x)F(dx).$$

Example 10.49. Suppose N is a point process on $(S,\mathcal{S})$. Write $N=\sum_k \delta_{\xi_k}$. Then $N\circ f^{-1}$ is a point process on $(T,\mathcal{T})$. Let g be a nonnegative measurable function on T. Then

$$\begin{aligned}(N\circ f^{-1})g &= \int_T g(t)N\circ f^{-1}(dt) = \int_S g\circ f(s)N(ds) = \int_S f\circ f(s)\sum_k \delta_{\xi_k}(ds)\\ &= \sum_k g\circ f(\xi_k) = \sum_k g\left(f(\xi_k)\right).\end{aligned}$$

So setting $\tau_k = f(\xi_k)$, we have $N\circ f^{-1} = \sum_k \delta_{\tau_k} = \sum_k \delta_{f(\xi_k)}$. Let m be the intensity of N, then $m\circ f^{-1}$ is the intensity of $N\circ f^{-1}$ since

$$E[N\circ f^{-1}(A)] = E\left[N(f^{-1}(A))\right] = m(f^{-1}(A)) = m\circ f^{-1}(A).$$

Next the Laplace functional of $N\circ f^{-1}$ is

$$L_{N\circ f^{-1}}(g) = E\left[\exp\left(-N\circ f^{-1}g\right)\right] = E\left[-\int_S g\circ f(s)N(ds)\right] = L_N(g\circ f).$$

Finally, let N be a Poisson random measure on $(S,\mathcal{S})$ with intensity m. Then

$$\begin{aligned}L_{N\circ f^{-1}}(g) &= L_N(g\circ f) = \exp\left[-m\left(1-e^{-f\circ g}\right)\right]\\ &= \exp\left[-m\left(1-e^{-g}\right)\circ f\right] = \exp\left[-m\circ f^{-1}\left(1-e^{-g}\right)\right],\end{aligned}$$

since

$$1-e^{-g\circ f(s)} = \left(1-e^{-g}\right)f(s)(=1[f(s)]-\left(e^{-g}\right)(f(s)) = 1-e^{-g\circ f(s)}, \forall s\in\mathcal{S}.)$$

So $N\circ f^{-1}$ is a Poisson random measure with intensity mf^{-1}.

Example 10.50. Suppose N is a Poisson random measure on $\mathbb{R}\times\mathbb{R}_+$ with intensity

$$m(dt,dx) = \lambda(dt)\mu(dx),$$

where μ is the distribution of a nonnegative random variable. Let $N=\sum_k \delta_{(\xi_k,\tau_k)}$, where $\xi_k\in R$ and $\tau_k\in\mathbb{R}^+$. Consider an infinite server queue, and let ξ_k be the arrival time and τ_k be the service time of the arrival ξ_k. Define

$$\begin{aligned}f:\mathbb{R}\times\mathbb{R}_+ &\to \mathbb{R}\\ (s,x) &\to s+x.\end{aligned}$$

Then $Nf^{-1} = \sum_k \delta_{f(\xi_k,\tau_k)} = \sum_k \delta_{\xi_k+\tau_k}$, is the point process giving the departure time from the system. It is a Poisson random measure with intensity $m\circ f^{-1}$. Let $g:\mathbb{R}\to\mathbb{R}_+$. Then

$$\left(m\circ f^{-1}\right)g = \int_\mathbb{R} g(y)m\circ f^{-1}(dy) = \int_{\mathbb{R}\times\mathbb{R}_+} g\circ f(s,t)m(ds,dt) = \int_{\mathbb{R}^+}\int_\mathbb{R} g(s+t)\lambda(ds)\mu(dt).$$

Let $x=s+t$, then $dx=ds$. Then

$$\left(m\circ f^{-1}\right)g = \int_{\mathbb{R}^+}\int_\mathbb{R} g(x)\lambda dx\mu(dt) = \int_\mathbb{R} g(x)\left[\int_{\mathbb{R}^+}\mu(dt)\right]\lambda dx = \int_\mathbb{R} g(x)\lambda dx.$$

$m\circ f^{-1}(dy) = \lambda dy.$???

10.9 The Distribution of a Point Process

Definition 10.51. Let $\mathcal{N}$ be the set of all counting measures on S given $(\Omega, \mathcal{A}, P)$, consisting of all events of the form

$$E_{A,k} = \{m \in \mathcal{N} \mid m(A) = k\}, \forall \text{compact } A \in \mathcal{S}, \forall k \in \mathbb{N},$$

i.e, it is the event that there are exactly k points in the region A. The measurable space $(\mathcal{N}, \mathcal{B}(\mathcal{N}))$ is called the canonical space or outcome space for a point process in S.

Remark. The σ-field $\mathcal{B}(\mathcal{N})$ includes events such that

$$E_{A_1,k_1} \cap \cdots E_{A_n,k_n} = \{m \in \mathcal{N} : m(A_1) = k_1, \ldots, m(A_n) = k_n\},$$

i.e., the event that there are exactly k_i points in the region B_i for $i = 1, \ldots, m$. It also includes, for example, the event that the point process has no points at all,

$$\{N = 0\} = \{m \in \mathcal{N} : m(A) = 0, \forall A \in \mathcal{S}\},$$

since this event can be represented as the intersection of the countable sequence of events $\{E_{B(0,n),0}\}_{n \in \mathbb{Z}^+}$. Here $B(0, r)$ denotes the ball of radius r and center 0 in S.

A point process N may now be defined formly as a measurable mapping from a probability space to an outcome space

$$N : (\Omega, \mathcal{A}, P) \to (\mathcal{N}, \mathcal{B}(\mathcal{N})).$$

Thus, each elementary outcome $\omega \in \Omega$ determines an outcome $N_\omega \in M_p$ for the entire point process. Measurability is the requirement that, $\forall E \in \mathcal{B}(\mathcal{N})$, the event

$$\{N \in E\} = \{\omega \in \Omega : N_\omega \in E\} \in \mathcal{A}.$$

This implies that any event has a well-defined probability $P(N \in E)$.

Definition 10.52. The distribution of a point process N is the probability measure P_N on the outcome space $(\mathcal{N}, \mathcal{B}(\mathcal{N}))$, defined by

$$P_N(A) = P(N \in A) = P(\omega \in \Omega : N_\omega \in A), \forall A \in \mathcal{B}(\mathcal{N}).$$

10.10 Stationary Random measure

Let $\{N(t); t \geqslant 0\}$ be a time homogeneous Poisson process having rate λ so that it has the following properties:

(a) $N(0) = 0$.

(b) It has stationary increments.

(c) It has independent increments.

(d) $N(t) \sim \text{Poi}(\lambda t)$.

Remark. Property (2) stats that if $A = (a, b]$ and $t > 0$, then with $t + A = \{x + t : x \in A\}$, we have $N(t + A) = N(A)$, i.e., $N(t + b - a) - N(t) = N(b - a)$.

Definition 10.53. For any $t \in \mathbb{R}$, define

$$\begin{aligned}\theta_t : \mathcal{N} &\to \mathcal{N}\\ m &\mapsto \theta_t m : \mathcal{B}(\mathbb{R}) \to \overline{\mathbb{R}}_+\\ &\qquad\qquad A \mapsto m(t + A).\end{aligned}$$

Therefore, on $\mathbb{R}^d$, $\theta_t x = x - t$? (Shift all points to the left or shift the origin to the right.)

Theorem 10.54. *θ_t is $\mathcal{B}(\mathcal{N})/\mathcal{B}(\mathcal{N})$-measurable.*

Proof. NTS: for any $B \in \mathcal{B}(\mathcal{N})$,

$$\theta_t^{-1}(B) = \{m \in \mathcal{N} : \theta_t m \in B\} \in \mathcal{B}(\mathcal{N}).$$

Define

$$\zeta := \{B \in \mathcal{B}(\mathcal{N}) : \theta_t^{-1}(B) \in \mathcal{B}(\mathcal{N})\} \subseteq \mathcal{B}(\mathcal{N}).$$

Since preimages are nice, it is very easy to show ζ is an σ-algebra. Next, we show $\mathcal{B}(\mathcal{N}) \subseteq \zeta$, we need to show ζ contains all sets of the form $\{m \in \mathcal{N} : m(A) = k\}$. Suppose this is done. Thus, $\zeta = \mathcal{B}(\mathcal{N})$. Let $E_{A,k} = \{m \in \mathcal{N} : m(A) = k\}$, where $A \in \mathcal{B}(\mathbb{R}^d)$ and $k \in \mathbb{N}$. Since $\theta_t : \mathcal{N} \to \mathcal{N}$,

$$\theta_t^{-1}(E_{A,k}) = \{m \in \mathcal{N} : (\theta_t m)(A) = k\} = \{m \in \mathcal{N} : m(t + A) = k\}.$$

Since $t + A \in \mathcal{B}(\mathbb{R}^d)$, it makes sense for $m(t + A) = k$ and we are done. □

Definition 10.55. A point process

$$N : (\Omega, \mathcal{A}, P) \to (\mathcal{N}, \mathcal{B}(\mathcal{N}))$$

is said to be *stationary* if for any $t \in \mathbb{R}^d$, $\theta_t N$ has the same distribution as N, where

$$P_N(A) = P(\omega \in \Omega \mid N_\omega \in A), \forall\, A \in \mathcal{B}(\mathcal{N}),$$

$$P_{\theta_t N}(A) = P(\omega \in \Omega \mid (\theta_t N)_\omega \in A), \forall\, A \in \mathcal{B}(\mathcal{N}).$$

This is equivalent to the condition that for any $n \in \mathbb{Z}^+$, if $\{A_i\}_{i=1}^n \subseteq \mathcal{B}(\mathbb{R}^d)$ are disjoint,

$$\begin{aligned}((\theta_t N)(A_1), \dots, (\theta_t N)(A_n)) &= (N(t + A_1), \dots, N(t + A_n))\\ &\overset{d}{=} (N(A_1), \dots, N(A_n)).\end{aligned}$$

For a stationary point process N, let $u \in M_p$ be the intensity of N. Then$\forall\, A \in \mathcal{B}(\mathbb{R}^d)$,

$$\mu(A) = E[N(A)] = E[(\theta_t N)(A)] = E[N(t + A)] = \mu(t + A),$$

which is translation invariant. Moreover, if $A \in \mathcal{B}(\mathbb{R})$ is bounded, then $\mu(A) < \infty$. Since for all Radon measures, only the multiple of Lebesgue measure on $\mathbb{R}^d$ satisfying the translation invariance property,

$$\exists \lambda > 0, \ s.t.\ \mu(A) = \lambda |A|,$$

where $|\cdot|$ denotes the Lebesgue measure. Note $\lambda|\cdot|$ is still a invariant measure. Let $\{N_t; t \geqslant 0\}$ be a time homogeneous Poisson process having rate λ. By above, we can think of $\{N_t, t \geqslant 0\}$ as a stationary point process having the intensity $\lambda \times |\cdot|$. Recall that

$$\lim_{t\to\infty} \frac{N(t)}{t} = \lambda \text{ w.p.1.}$$

One method of proving the above result is to apply the strong law of large numbers to the inter-arrival and invert. A slightly weaker result is to show that $\frac{N(t)}{t} \xrightarrow{p} \lambda$. Let $\epsilon > 0$ and $t > 0$, then

$$\begin{aligned} P(|N(t)/t - \lambda| > \epsilon) = P(|N(t) - \lambda t| > \lambda t) &\leqslant \frac{\operatorname{Var}(N_t - \lambda t)}{(\lambda t)^2} \\ &= \frac{\operatorname{Var}(N_t)}{(\lambda t)^2} = \frac{1}{\lambda t} \to 0, \text{ as } t \to \infty. \end{aligned}$$

Our goal is to provide a similar result for stationary point process. Before doing so consider the following result from Markov chain theory. Let $\{(X_n, Y_n); n = 0, 1, \cdots\}$ be a time homogeneous Markov chain with state space $\{(0,0), (1,0), (0,1), (1,1)\}$ and one-step transition matrix P given by

$$\begin{bmatrix} 0.3 & 0.7 & 0 & 0 \\ 0.7 & 0.3 & 0 & 0 \\ 0 & 0 & 0.3 & 0.7 \\ 0 & 0 & 0.2 & 0.8 \end{bmatrix}$$

Then

$$\begin{aligned} \pi^{(0)} &= (1/2, 1/2, 0, 0), \text{ (sub chain is doubly MC)} \\ \pi^{(1)} &= (0, 0, 2/9, 7/9), \end{aligned}$$

are stationary distribution, as are their convex combinations. Let $S_n = \sum_{m=1}^{n} \mathbb{1}_{\{X_m = 0\}}$. Then

$$\frac{S_n}{n} \to 1/2\mathbb{1}_{\{Y_0=0\}} + 2/9\mathbb{1}_{\{Y_0=1\}},$$

which is a random variable.

10.11 Invariant Sets and the ergodic theorem for random measures

Definition 10.56. For $\theta_t : \mathcal{N} \to \mathcal{N}$, a set $I \in \mathcal{B}(\mathcal{N})$ is called *shift-invariant* if

$$\theta_t^{-1} I = \{m \in \mathcal{N} \mid \theta_t m \in I\} = I.$$

Let

$$B_n = [0, n),$$

and

$$I_c = \left\{ \frac{N(B_n)}{n} \to c \right\} \text{ as } n \to \infty.$$

Then

$$\theta_t I_c = \left\{ \frac{N(t+B_n)}{n} \to c \right\} \text{ as } n \to \infty.$$

Note $t + B_n = [t, n+t)$. Let $\omega \in I_c$, then

$$\frac{N_\omega(B_n)}{n} \to c.$$

Since as $n \to \infty$,

$$c \leftarrow \frac{N_\omega([0, \lceil n+t \rceil))}{\lceil n+t \rceil} = \frac{N_\omega([0,t))}{\lceil n+t \rceil} + \frac{N_\omega([t, n+t))}{\lceil n+t \rceil} + \frac{N_\omega([n+t, \lceil n+t \rceil))}{\lceil n+t \rceil},$$

we have as $n \to \infty$,

$$\frac{N_\omega([t, n+t))}{\lceil n+t \rceil} \to c.$$

Then as $n \to \infty$,

$$\frac{N_\omega(t+B_n)}{n} = \frac{N_\omega(t+\mathcal{B}_n)}{\lceil n+t \rceil} \frac{\lceil n+t \rceil}{n} = \frac{N_\omega([t, n+t))}{\lceil n+t \rceil} \frac{\lceil n+t \rceil}{n} \to c.$$

So $I_c \subseteq \theta_t I_c$. Similarly, $T_t I_c \subseteq I_c$. Hence $T_t I_c = I_c$. Thus, I_c is invariant?

Definition 10.57. Let $I \in \mathcal{B}(\mathcal{N})$ be invariant. Then

$$\mathcal{I} := \mathcal{B}(\{I \in \mathcal{B}(\mathcal{N}) \mid I \text{ is invariant}\}) \subseteq \mathcal{B}(\mathcal{N})$$

is a σ-algebra. Then

$$\mathcal{I}_N := N^{-1}(\mathcal{I}) = \{N^{-1}(I) \mid I \in \mathcal{I}\} \subseteq \mathcal{A}$$

is a σ-algebra.

Definition 10.58. A stationary point process N is called *ergodic* if

$$P\left(N^{-1}(I)\right) \in \{0,1\}, \forall I \in \mathcal{I}.$$

Theorem 10.59. *In general, it can be shown that there is an $\mathcal{I}_N$ measurable random variable $\bar{\xi}$ such that*

$$E[N(B)|\mathcal{I}_N] = \bar{\xi} \cdot |B|, \forall B \in \mathcal{B}(\mathbb{R}^d).$$

If N is ergodic, $\bar{\xi}$ can be taken to be a constant.

Proof. Since $\{\bar{\xi} \leqslant t\} \subseteq \mathcal{I}_N, \forall t \in \mathbb{R}$, then $P(\bar{\xi} \leqslant t) \in \{0,1\}, \forall t \in \mathbb{R}$. Let $c := \inf\{t : P(\bar{\xi} \leqslant t) = 1\}$. If $t > c$, then $P(\bar{\xi} \leqslant t) = 1$; If $t < c$, then $P(\bar{\xi} \leqslant t) = 0$. Thus, $\bar{\xi} = c$, *w.p.*1. □

Theorem 10.60. *The ergodic theorem states that if $\{B_n\}_{n \in \mathbb{Z}^+}$ are rectangles such that $B_1 \subseteq B_2 \subseteq \cdots$ and $|B_n| \uparrow \infty$, then*

$$\frac{N(B_n)}{|B_n|} \to \bar{\xi}.$$

Example 10.61. Recall a stationary Cox process N is a point process for which there exists a random variable Λ such that conditional on $\Lambda = \lambda$, N is a homogeneous Poisson process with rate λ. Taking $B_n = [0, n)$. Since

$$\left. \frac{N([0,n))}{n} \right| \Lambda = \lambda \longrightarrow \lambda, \text{ w.p.1}.$$

Then

$$\left.\frac{N([0,n))}{n}\right| \Lambda \longrightarrow \Lambda \text{ w.p.1.}$$

By ergodic theorem, $\bar{\xi} = \Lambda$.

10.12 Stochastic process

Definition 10.62. A stochastic process is defined as a collection of random variables defined on a common probability space $(\Omega, \mathcal{B}, P)$ and the random variables, indexed by some set T, all take values in the same space $(S, \mathcal{S})$. In other words, for a given probability space $(\Omega, \mathcal{B}, P)$ and a measurable space $(S, \mathcal{S})$, a stochastic process is a collection of S-valued random variables, which can be written as

$$\{X(t) : t \in T\}.$$

Definition 10.63. The space S is called the state space of the stochastic process.

Definition 10.64. If $\{X(t); t \in T\}$ is a stochastic process, then for any $\omega \in \Omega$, the mapping

$$X(\cdot, \omega) : T \to S,$$

is called a sample function, a realization, or, particularly when T is interpreted as time, as sample path. This means that for a fixed $\omega \in \Omega$, there exits a sample function that maps the index set T to the state space S.

Remark. (a) t is typically time, but can also be a spatial dimension.

(b) t can be discrete or continuous.

(c) The range of t can be finite, but more often is finite, which means the process contains an infinite number of r.v's.

Example 10.65.

- The wireless signal received by a cell phone over time
- The daily stock price
- The number of packets arriving at a router in 1-second intervals.
- the image intensity over $1cm^2$ regions.

Definition 10.66. A stochastic process

$$X : (\Omega, \mathcal{A}, P) \to \left(S^{\mathbb{R}}, \mathcal{S}^{\mathbb{R}}\right)$$

is said to be *stationary* if for any $t \in \mathbb{R}^d$, $\theta_t X$ has the same distribution as X, where

$$P_X(A) = P(\omega \in \Omega \mid X_\omega \in A), \forall A \in \mathcal{S}^{\mathbb{R}},$$

$$P_{\theta_t X}(A) = P(\omega \in \Omega \mid (\theta_t X)_\omega \in A), \forall A \in \mathcal{S}^{\mathbb{R}}.$$

This is equivalent to the condition that for any $n \in \mathbb{Z}^+$, if $\{A_i\}_{i=1}^n \subseteq \mathcal{S}^{\mathbb{R}}$ are disjoint,

$$\begin{aligned}((\theta_t X)(A_1), \ldots, (\theta_t X)(A_n)) &= (X(t + A_1), \ldots, X(t + A_n)) \\ &\stackrel{d}{=} (X(A_1), \ldots, X(A_n)).\end{aligned}$$

10.13 Jointly Stationary Random Measures and Stochastic Processes

10.13.1 Event and Time Averges

Theorem 10.67 (Ergodic theorem)**.** *Let $\{X(t); t \geqslant 0\}$ be a ergodic stochastic process with $E[X(t)] < \infty$. Then*

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t f(X_\omega(s))ds = E\left[f\left(X(0)\right)\right], \forall \omega \in \Omega,$$

for all bounded functions f, which is called time average. Suppose we observe X_t at times $\{T_i\}_{i\in\mathbb{Z}^+}$ that correspond to a ergoic point process $\{N(t); t \geqslant 0\}$. Define the following if it exists:

$$E_N[f\left(X(0)\right)] \stackrel{d}{=} \lim_{t\to\infty} \frac{1}{N_\omega(t)} \sum_{k=1}^{N_\omega(t)} f\left(X(T_k(\omega))\right) = \lim_{t\to\infty} \frac{1}{N_\omega(t)} \int_0^t f\left(X_\omega(s)\right) N_\omega(ds),$$

where we note that by the definiton of the Stieltjes integral,

$$\int_0^t f(X_\omega(s)) N_\omega(ds) = \int_0^t f(X_\omega(s)) \left(\sum_{k=1}^{N_\omega(t)} \delta_{T_k}\right)(ds) = \sum_{k=1}^{N_\omega(t)} f\left(X(T_k(\omega))\right).$$

Let N be a point process. Let the index $T = \mathbb{R}$ and X be a S-valued stochastic process $X = \{X_t; t \in \mathbb{R}\}$. Let $S^{\mathbb{R}}$ be the space consisting of all functions $f : \mathbb{R} \to S$. Think X of

$$X : (\Omega, \mathcal{A}, P) \to (S^{\mathbb{R}}, \mathcal{S}^{\mathbb{R}}),$$

which usually is not onto.

Definition 10.68. For any $t \in \mathbb{R}$, define the evaluation

$$\begin{aligned} \pi_t : S^{\mathbb{R}} &\to S \\ x &\mapsto x(t). \end{aligned}$$

Definition 10.69. Take $\mathcal{S}^{\mathbb{R}}$ to be $\sigma(S^{\mathbb{R}})$ that makes π_t's measurable, whose element is of the form

$$\{x \in S^{\mathbb{R}} \mid \pi_t x \in A\}, \forall t \in \mathbb{R}, \forall A \in \mathcal{S},$$

Then the distribution of X is determined by the finite dimensional distribution

$$P(X(t_1) \in B_1, \ldots, X(t_n) \in B_n),$$

where $t_1, \ldots, t_n \in \mathbb{R}$ and $B_1, \ldots, B_n \in \mathcal{S}$.

10.13.2 Joint processes

Consider the joint process (X, N) on $\left(S^{\mathbb{R}} \times \mathcal{N}, \mathcal{S}^{\mathbb{R}} \otimes \mathcal{B}(\mathcal{N})\right)$.

Definition 10.70. (a) For any $t \in \mathbb{R}$,

$$\begin{aligned} \theta_t : (\mathcal{N}, \mathcal{B}(\mathcal{N})) &\to (\mathcal{N}, \mathcal{B}(\mathcal{N})) \\ m &\mapsto m_t, \end{aligned}$$

where $m_t(A) = m(t + A)$ for any $A \in \mathcal{B}(\mathbb{R})$. For example, $m = N_\omega$.

(b) For any $t \in \mathbb{R}$, define

$$\begin{aligned} \theta_t : (S^{\mathbb{R}}, \mathcal{S}^{\mathbb{R}}) &\to (S^{\mathbb{R}}, \mathcal{S}^{\mathbb{R}}) \\ x &\mapsto \theta_t x, \end{aligned}$$

where $(\theta_t x)(s) = x(t + s)$ for any $s \in S$. For example, $x = X_\omega$.

(c) For any $t \in \mathbb{R}$, define

$$\begin{aligned} \theta_t : (S^{\mathbb{R}} \times \mathcal{N}, \mathcal{S}^{\mathbb{R}} \times \mathcal{B}(\mathcal{N})) &\to (\mathcal{S}^{\mathbb{R}} \times \mathcal{N}, \mathcal{S}^{\mathbb{R}} \times \mathcal{B}(\mathcal{N})) \\ (x, m) &\mapsto (\theta_t x, \theta_t m). \end{aligned}$$

Definition 10.71. The joint process (X, N) is called *stationary* if

$$\theta_t(X, N) \stackrel{d}{=} (X, N), \forall t \in \mathbb{R},$$

where

$$P_{(X,N)}(A) = P((X, N) \in A), \forall A \in \mathcal{S}^{\mathbb{R}} \times \mathcal{B}(\mathcal{N}).$$

Example 10.72. Let $X = \{X(t); t \in \mathbb{R}_+\}$ be a CTMC having a countable state space S. Assume X has a **unique** stationary distribution which is also a limiting distribution. If Q is the generator of X, then $\pi Q = 0$, is the "stationary dist" in that if X_0 has distribution π, then

$$\theta_t \left(X(t_1), \ldots, X(t_n)\right) = \left(X(t + t_1), \ldots, X(t + t_n)\right) \stackrel{d}{=} \left(X(t_1), \ldots, X(t_n)\right).$$

Thus,

$$\theta_t X \stackrel{d}{=} X$$

and X is stationary on $\mathbb{R}_+$. One can extend X to a stationary process on $\mathbb{R}$ having the same dist as on $\mathbb{R}_+$. Label the jump times of the MC so that

$$\cdots < \tau_{-1}(N) < \tau_0(N) \leqslant 0 < \tau_1(N) < \cdots.$$

Let N be the point process whose jump times are given by the $\{\tau_k\}_{k\in\mathbb{Z}}$. The N is a measurable function of X and it turns out (X, N) is stationary. Since N is stationary, there exists $\lambda > 0$ such that

$$E[N(A)] = \lambda |A|.$$

Note

$$E[N(0, t)|X_0 = i] = 1 \cdot P(N(0, t) = 1|X_0 = i) + \sum_{n=2}^{\infty} nP(N(0, t) = n|X_0 = i),$$

and $P(N(0,t)=1|X_0=i)=\lambda_i t+o(t)$, where $\lambda_i=-q_{ii}=\sum_{j\neq i}q_{ij}$. Next, (and one takes a little work)

$$\sum_{n=2}^{\infty} nP(N(0,t)=n)=o(t).$$

Hence

$$E[N(0,t)|X_0=i]=\lambda_i t+o(t).$$

Thus,

$$\begin{aligned} E[N(0,t)] = E[E[N(0,t)|X_0]] &= E\left[\sum_{i\in S} E[N(0,t)|X_0=i]\mathbb{1}_{\{X_0=i\}}\right] \\ &= \sum_{i\in S} E[N(0,t)|X_0=i]P(X_0=i) = \sum_{i\in S}\pi_i\left(\lambda_i t+o(t)\right)=\lambda t. \end{aligned}$$

where we divide by t and letting $t\to 0$ to obtain $\lambda=\sum_{i\in S}\pi_i\lambda_i$.

10.14 Palm distribution

The palm distribution of (X,N). The stationary pair (X,N) involves a dist. $P_{(X,N)}$ on $(S^{\mathbb{R}}\times\mathcal{N},\mathcal{S}^{\mathbb{R}}\times\mathcal{B}(\mathcal{N}))$ via

$$P_{(X,N)}(A)=P((X,N)\in A),\forall A\in\mathcal{S}^{\mathbb{R}}\times\mathcal{B}(\mathcal{N}).$$

For nonnegative measurable f on $S^{\mathbb{R}}\times\mathcal{N}$,

$$P_{(X,N)}f=\int_{S^{\mathbb{R}}\times\mathcal{N}} f(x,m)P_{(X,N)}(dx,dm)=E[f(X,N)].$$

Define a palm distribution $Q_{(X,N)}$ on $(S^{\mathbb{R}}\times\mathcal{N},\mathcal{S}^{\mathbb{R}}\times\mathcal{B}(\mathcal{N}))$ by setting for all nonnegative measurable f,

$$Q_{(X,N)}f=\int_{S^{\mathbb{R}}\times N} f(x,m)Q(dx,dm):=\frac{E\left[\int_B f\left(\theta_s(X,N)\right)N(ds)\right]}{E[N(B)]},\forall B\in\mathcal{B}(\mathbb{R})\text{ and }|B|\neq 0.$$

Note $Q_{(X,N)}f$ does not depend on choice of B since (X,N) is a stationary pair. To gain some insight into the palm distribution, label the points of N as

$$\cdots<\tau_{-1}(N)<\tau_0(N)\leqslant 0<\tau_1(N)<\cdots.$$

The notation reminds us that the points of N are functions (indeed measurable) of N. Since $N=\sum_{k=-\infty}^{N(t)}\delta_{\tau_k(N)}$ is a point process,

$$\int_B f\left(\theta_s(X,N)\right)N(ds)=\sum_{\tau_k(N)\in B} f\left(\theta_{\tau_k(N)}(X,N)\right),\forall B\in\mathcal{B}(\mathbb{R}).$$

For $\omega\in\Omega$, $\theta_{\tau_k(N(\omega))}(X(\omega),N(\omega))$ is the outcome formed from $(X(\omega,N(\omega)))$ by shifting the origin to $\tau_k(N(\omega))$. If one thinks of the $\tau_k(N)$'s as arrival times, $\theta_{\tau_k(N)}(X,N)$ is the process viewed by from the perspective of the arrival at $\tau_k(N)$. If we now think of f as being some measurement made on the process (X, N), then $f\left(\theta_{\tau_k(N)}(X, N)\right)$ is the measurement one makes if the arrival at time $\tau_k(N)$ occurs at times 0. If one sums over the arrivals that occur in B, then takes the expectation and finally divides by the expected value of B, one gets a type of average of the measurement f as seen by an arriving customer.

Theorem 10.73. *Suppose P is a probability measure and $f \geqslant 0$ is measurable, define*

$$\mu(A) = \int_A f(x)P(dx), \forall A \in \mathcal{B}.$$

Then by definition, μ is a measure.

Example 10.74. Let X be a CTMC and let N be the point process consisting of the jumps $\{\tau_k\}_{k\in\mathbb{Z}}$. Assume X is stationary. Since N is a function of X, (X, N) is jointly stationary. Note given a distribution, there is always a random variable which has the distribution. Let (Y, M) be a pair that has joint distribution $Q_{(X,N)}$.

$$E[f(Y, M)] = Q_{(X,N)}f.$$

Let $f(x, m) = \mathbb{1}_{\{j\}}(x_0)$. Take $B = [0, t]$. Since $\theta s X_0 = X_s$, we have

$$E\left[\mathbb{1}_{\{Y_0=j\}}\right] = P(Y_0 = j) = \frac{E\left[\int_0^t \mathbb{1}_{\{j\}}(X_s)N(ds)\right]}{E[N(0,t)]} = \frac{E\left[\sum_{k=1}^{N(0,t)} \mathbb{1}_{\{j\}}\left(X_{\tau_k(N)}\right)\right]}{\left(\sum_{i\in S}\pi_i\lambda_i\right)t}.$$

Now

$$\begin{aligned}
E\left[\sum_{k=0}^{N(0,t)} \mathbb{1}_{\{j\}}\left(X_{\tau_k}(N)\right)\right] &= E\left[E\left[\sum_{k=0}^{N(0,t)} \mathbb{1}_{\{j\}}\left(X_{\tau_k}(N)\right)\middle| X_0\right]\right] \\
&= E\left[\sum_{k=1}^{\infty} P\left(X_{\tau_k(N)} = j, N(0,t) \geqslant k \middle| X_0\right)\right] \\
&= E\left[P\left(X_{\tau_k(N)} = j, N(0,t) \geqslant 1 \middle| X_0\right)\right] \\
&\quad + E\left[\sum_{k=2}^{\infty} P\left(X_{\tau_k(N)} = j, N(0,t) \geqslant k \middle| X_0\right)\right] \\
&= E[q_{X_0 j}t + o(t)] + E[o(t)] \\
&= \sum_{i\neq j} \pi_i q_{ij} t + o(t) \\
&= \pi_j \lambda_j t + o(t).
\end{aligned}$$

since the CTMC cannot make a jump to itself and then when $X_0 = j$, the conditional expectation is just $o(t)$ and finally, by the balance equation $\pi Q = 0$, we get the result. Letting $t \to 0$, we have

$$P(Y_0 = j) = \frac{\pi_j \lambda_j}{\sum_{i\in S}\pi_i\lambda_i}.$$

Given a joint stationary pair (X, N), the palm measure is

$$E[f(Y, M)] = Q_{(X,N)}f = \frac{E\left[\int_B f\left(\theta_s(X, N)\right) N(ds)\right]}{E[N(B)]} = \frac{E\left[\int_B f\left(\theta_s(X, N)\right) N(ds)\right]}{E[\overline{N}]|B|},$$

where $B \in \mathcal{B}(\mathbb{R})$, $\overline{N} = E[N(0,1)|\mathcal{I}_N]$, $E[N(A)] = \lambda|A|$. stationary: invariant, multiple, Lebesgue.

Lemma 10.75. Let f be a nonnegative measurable function of $S^{\mathbb{R}} \times \mathcal{N}$ and define a random measure N_f on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by

$$N_f(B) = \int_B f\left(\theta_s(X, N)\right) N(ds), \forall B \text{ bounded Borel set.}$$

Then N_f is also stationary.

Corollary 10.76. Given the lemma, since N_f is stationary,

$$E\left[\int_B f\left(\theta_s(X, N)\right) N(ds)\right] = E[N_f(B)] = c|B|,$$

where $c = E[N_f(0,1)]$. Then

$$Q_{(X,N)}f = \frac{c|B|}{E[\overline{N}]|B|} = \frac{c}{E[\overline{N}]} = \frac{c}{E[N(0,1)]} = \frac{c}{\lambda},$$

which does not depend on B.

Theorem 10.77. *Let (X, N) be a stationary pair as above. Let Y be an S valued stochastic process and M a point process on $\mathbb{R}$ having joint distribution $Q_{(X,N)}$. Then for any $f \geqslant 0$,*

$$E[f(X, N)] = E[\overline{N}]E\left[\int_0^{\tau_1(M)} f(\theta_s(Y, M))ds\right].$$

Proof. By

$$E[\overline{N}]|B|E[f(Y, M)] = E\left[\int_B f\left(\theta_s(X, N)\right) N(ds)\right],$$

we have

$$E[\overline{N}]\int_{\Omega\times\mathbb{R}} \mathbb{1}_B(s) f\left(Y(\omega), M(\omega)\right) P(d\omega)ds = \int_{\mathbb{R}\times\Omega} \mathbb{1}_B(s) f\left(\theta_s(X(\omega), N(\omega))\right) N(\omega, ds)P(d\omega),$$

where $dsP(d\omega)$ is the product measure on $\mathbb{R}\times\Omega$ and $N(\omega, ds)P(d\omega)$ is the measure on $\mathbb{R}\times\Omega$. From here, one can extend the relationship to

$$E[\overline{N}]E\left[\int_{\mathbb{R}} h(Y, M, s)ds\right] = E\left[\int_{\mathbb{R}} h\left(\theta_s(X, N), s\right) N(ds)\right],$$

where $h \geqslant 0$ is a measurable function on $S^{\mathbb{R}} \times \mathcal{N} \times \mathbb{R}$. Setting $h(x, m, s) = f\left(\theta_s(x, m), s\right)$ for nonnegative measurable f. Making the change of variable $t = -s$ on the left hand side yields

$$\begin{aligned} E[\overline{N}]E\left[\int_{\mathbb{R}} f\left(\theta_s(Y, M), s\right) ds\right] &= E[\overline{N}]E\left[\int_{\mathbb{R}} f\left(\theta_{-t}(Y, M), -t\right) dt\right] \\ &= E\left[\int_{\mathbb{R}} f(X, N, s)N(ds)\right]. \end{aligned}$$

Set $f(x,m,s) = h(x,m)\mathbb{1}_{\{\tau_0(m)=s\}}$. Since $M(\{0\}) = 1$,

$$\tau_0(\theta_s(M)) = -s \text{ if and only if } 0 < s < \tau_1(M).$$

The the left hand side of the above equation becomes

$$\begin{aligned} E[\overline{N}]E\left[\int_{\mathbb{R}} h\left(\theta_s(Y,M)\right)\mathbb{1}_{\{\tau_0(\theta_s(M))=-s\}}ds\right] &= E[\overline{N}]E\left[\int_0^{\tau_1(M)} h\left(\theta_s(Y,M)\right)ds\right] \\ &= E\left[\int_{\mathbb{R}} h(X,N)\mathbb{1}_{\{\tau_0(N)=s\}}N(ds)\right] \\ &= E\left[h(X,N)\int_{\mathbb{R}} \mathbb{1}_{\{\tau_0(N)=s\}}N(ds)\right] \\ &= E[h(X,N)], \end{aligned}$$

which is equal to the right hand side of the above equation. □

Example 10.78. Take

$$f(x,m) = \mathbb{1}_B(x_0), \text{ for some } B \in \mathcal{S}.$$

Since $E[\overline{N}] = \frac{1}{E[\tau_1(M)]}$, by the theorem above,

$$P(X_0 \in B) = \frac{E\left[\int_0^{\tau_1(M)}(\theta_s\mathbb{1}_B)(Y_0)ds\right]}{E[\tau_1(M)]} = \frac{E\left[\int_0^{\tau_1(M)}\mathbb{1}_B(Y_s)ds\right]}{E[\tau_1(M)]}.$$

This result should look familar. For suppose X is a stationary regenerative process and N is associated renewal sequence, then (Y,M) is the pair when M is the counting process for an ordinary renewal process. Note $\tau_1(M)$ is time the first renewal occurs. If the interrenewal distribution is F, then $P(\tau_1(M) > t) = 1 - F(t)$. Suppose the interrenewal distribution is nonarithemtic so that a limiting distribution exists and is the stationary distribution. In this case, the above equation becomes

$$\lim_{t\to\infty} P(X_t \in B) = \frac{E\left[\int_0^{\tau_1(M)}\mathbb{1}_B(Y_s)ds\right]}{E[\tau_1(M)]},$$

which is precisely the limiting distribution of a regenerative process.

Example 10.79. Let (X,N) be the stationary pair, where X is a CTMC and N is the point process which counts the jumps of X. Let (Y,M) be the pair when using the measure $Q_{(X,N)}$. Let $f(x,m) = \mathbb{1}_j(x_0)$. Then

$$\pi_j = P(X_0 = j) = \frac{E\left[\int_0^{\tau_1(M)}\mathbb{1}_{\{j\}}(Y_s)ds\right]}{E[\tau_1(M)]}.$$

Note the process Y is constant in the interval $[0,\tau_1(M))$ when given Y_0, so the numerator equals

$$\begin{aligned} E\left[\int_0^{\tau_1(M)}\mathbb{1}_{\{j\}}(Y_s)\tau_1(M)\right] &= E\left[E\left[\int_0^{\tau_1(M)}\mathbb{1}_{\{j\}}(Y_s)\tau_1(M)\middle| Y_0\right]\right] \\ &= E\left[E\left[\mathbb{1}_{\{j\}}(Y_0)\tau_1(M)\middle| Y_0\right]\right] = E\left[E\left[\mathbb{1}_{\{j\}}(Y_0)\tau_1(M)\middle| Y_0 = j\right]\mathbb{1}_{\{Y_0=j\}}\right] \\ &= E\left[\mathbb{1}_{\{j\}}(Y_0)\tau_1(M)\middle| Y_0 = j\right]E\left[\mathbb{1}_{\{Y_0=j\}}\right] = E\left[\tau_1(M)|Y_0 = j\right]P(Y_0 = j) \\ &= \nu_j/\lambda_j. \end{aligned}$$

A similar argument shows the denominator equals $\sum_i \nu_i/\lambda_i$. We obtain

$$\pi_j = \frac{\nu_j/\lambda_j}{\sum_i \nu_i/\lambda_i}.$$

Chapter 11

Stochastic intensity

Let $X \geqslant 0$ be a random variable with distribution function F and a density f. Then the hazard function (failure rate)

$$h(x) = \lim_{y\to 0} \frac{P(x < X \leqslant x+y|X > x)}{y} = \lim_{y\to 0} \frac{P(x < X \leqslant x+y)}{yP(X > x)}$$
$$= \lim_{y\to 0} \frac{F(x+y)-F(x)}{y}\frac{1}{1-F(x)} = \frac{f(x)}{1-F(x)},\ x > 0.$$

Since $F(0) = 0$, integrating both sides from 0 to x,

$$-\int_0^x h(y)dy = \log(1-F(x)) - \log(1-F(0)),\ x > 0.$$

So

$$1 - F(x) = \exp\left(-\int_0^x h(y)dy\right),\ x > 0.$$

Theorem 11.1.

$$E\left[\int_0^X h(x)dx\right] = \int_\Omega \int_0^{X(\omega)} h(x)dxP(d\omega),$$

or

$$E\left[\int_0^X h(x)dx\right] = E\left[\int_0^\infty h(x)\mathbb{1}_{\{X>x\}}dx\right] = \int_0^\infty h(x)E\left[\mathbb{1}_{\{X>x\}}\right]dx$$
$$= \int_0^\infty h(x)\left[1-F(x)\right]dx = \int_0^\infty f(x) = 1.$$

Let $\{N(t), t \geqslant 0\}$ be a point process on $\mathbb{R}^{\geqslant 0}$. Assume $N(0) = 0$ and the jump times occurs at $0 < T_1 < T_2 < \cdots$, and with probability 1, $\lim_{n\to\infty} T_n = \infty$ and $N(t) \to \infty$ as $t \to \infty$. Assume there exists a stochastic process $\{\lambda(t), t \geqslant 0\}$ such that

$$P\big(N(t+s) - N(t) = 1|\mathcal{F}_t^0\big) = \lambda(t)s + o(s),$$

$$P\left(N(t+s)-N(t)>1|\mathcal{F}_t^0\right)=o(s),$$

where $\mathcal{F}_t^0$ is the smallest σ-algebra that makes each $N(h)$ measurable for $0 \leqslant h \leqslant t$. Also

$$\mathcal{F}_t^0=\bigcap_{h>t}\mathcal{F}_h^0.$$

The family $\{\mathcal{F}_t^0, t \geqslant 0\}$ is called the natural filtration (history) of the process $\{N(t), t \geqslant 0\}$. The filtration $\{\mathcal{F}_t, t \geqslant 0\}$ is called the right continuous version of $\{\mathcal{F}_t^0, t \geqslant 0\}$. A random variable τ taking values in $[0,\infty]$ is called a $\{\mathcal{F}_t^0\}$ *stopping time* if $\{\tau \leqslant t\} \in \mathcal{F}_t, \forall t > 0$. If τ is an $\{\mathcal{F}_t\}$-stopping time, then for any $t > 0$,

$$\{\tau<t\}=\bigcap_{n=1}^{\infty}\left\{\tau \leqslant t+\frac{1}{n}\right\}\ ?$$

But

$$\left\{\tau \leqslant t+\frac{1}{n}\right\} \in \mathcal{F}_{t+\frac{1}{n}},$$

hence

$$\{\tau<t\} \in \bigcap_{n=1}^{\infty}\mathcal{F}_{t+\frac{1}{n}}=\mathcal{F}_t.$$

Thus, S is an $\{\mathcal{F}_t\}$ stopping if $\{\tau < t\} \in \mathcal{F}_t, \forall t \geqslant 0$. Each jump time T_n is a stopping time since $\{T_n \leqslant t\} = \{N(t) \geqslant n\} \in \mathcal{F}_t$.

Example 11.2. Suppose $\{N(t), t \geqslant 0\}$ is a homogeneous Poisson process having rate λ.

(a) $N(0) = 0$.

(b) Sample path are right-continuous, step process with jumps of size 1.

(c) stationary and independent increments.

(d) $P(N(t) = 1) = \lambda t + o(t)$ and $P(N(t) > 1) = o(t)$.

Note by property (3),

$$P(N(t+s)-N(t)=1|\mathcal{F}_t)=P(N(t+s)-N(t)=1)=P(N(s)=1)=\lambda s+o(s),$$

$$P(N(t+s)-N(t)>1|\mathcal{F}_t)=P(N(t+s)-N(t)>1)=P(N(s)>1)=o(s).$$

Choose n_0 so that $\frac{1}{n_0} < s$. For $n \geqslant n_0$, since $N(t+s)-N(t+\frac{1}{n})$ is independent of $\mathcal{F}_{t+\frac{1}{n}}$ and hence independent of $\mathcal{F}_t$, then

$$P(N(t+s)-N(t+1/n)=1|\mathcal{F}_t)=P(N(s-1/n)=1)=\lambda(s-1/n)e^{-\lambda(s-1/n)},$$

and right-continuity of a sample path N_ω in terms of t, where $\omega \in A \subseteq \Omega$ and $P(A) = 1$, and since $\lim_{n\to\infty} T_n = \infty$,

$$N(t+s)-N(t+1/n) \leqslant N(t+s) < \infty, \forall n \in \mathbb{N},$$

by BCT,

$$
\begin{aligned}
P(N(t+s)-N(t)=1|\mathcal{F}_t) &= P\left(\lim_{n\to\infty}(N(t+s)-N(t+1/n))=1\right)\\
&= \lim_{n\to\infty} P(N(t+s)-N(t+1/n)=1|\mathcal{F}_t)\\
&= \lambda s e^{-\lambda s} \xlongequal{\text{small}} \lambda s + o(s),
\end{aligned}
$$

Hence the stochastic intensity of a homogeneous Poisson process is

$$
\lambda(t) = \lim_{s\to 0}\frac{P(t<T_{n+1}\leqslant t+s|\mathcal{F}_t)}{s} = \lim_{s\to 0}\frac{P(N(t+s)-N(t)=1|\mathcal{F}_t)}{s} = \lambda, \forall n\in\mathbb{N}.
$$

General form of the stochastic intensity. Set $\mathcal{B}_0=\{\emptyset,\Omega\}$, and $\mathcal{B}_n=\sigma(T_1,\ldots,T_n), \forall n\in\mathbb{Z}^+$. Set

$$
F_n(t)=P(T_n\leqslant t\mid \mathcal{B}_{n-1}), \forall n\in\mathbb{Z}^+,
$$

i.e., F_n is the conditional distribution of T_n given $\mathcal{B}_{n-1}$ and f_n is the corresponding density function. Assume F_n has a density f_n. Think of the information is available at time $T_1,\ldots,T_n$ and $t-T_n$. Fix $n\in\mathbb{N}$, assume $T_n<t\leqslant T_{n+1}$, the failure rate at t is

$$
\begin{aligned}
\lambda(t) &= \lim_{y\to 0}\frac{P(t<T_{n+1}\leqslant t+y|T_{n+1}>t)}{y} = \lim_{y\to 0}\frac{P(t<T_{n+1}\leqslant t+y|\mathcal{B}_n)}{yP(T_{n+1}>t|\mathcal{B}_n)}\\
&= \lim_{y\to 0}\frac{F_{n+1}(t+y)-F_{n+1}(t)}{y}\frac{1}{1-F_{n+1}(t)} = \frac{f_{n+1}(t)}{1-F_{n+1}(t)},\ T_n<t\leqslant T_{n+1},
\end{aligned}
$$

So

$$
\lambda(t)=\sum_{n=0}^{\infty}\frac{f_{n+1}(t)}{1-F_{n+1}(t)}\mathbb{1}_{\{T_n<t\leqslant T_{n+1}\}}.
$$

Example 11.3. For a renewal process with **interrenewal** distribution F, since

$$
\begin{aligned}
P(T_{n+1}-T_n>t|T_n=s) &= P(T_{n+1}-T_n>t-T_n|T_n=s)\\
&= P(T_{n+1}-T_n>t-s|T_n=s)\\
&= P(T_{n+1}-T_n>t-s)\\
&= 1-F(t-s),
\end{aligned}
$$

we have

$$
\begin{aligned}
P(T_{n+1}>t|\mathcal{B}_n) &= P(T_{n+1}-T_n>t-T_n|\mathcal{B}_n) = P(T_{n+1}-T_n>t-T_n|T_n)\\
&= 1-F(t-T_n).
\end{aligned}
$$

Similarly,

$$
F_{n+1}(t)=P(T_{n+1}<t|\mathcal{B}_n)=F(t-T_n),
$$

$$
F_{n+1}(t+y)=P(T_{n+1}<t+y|\mathcal{B}_n)=F(t+y-T_n).
$$

Hence

$$
\lim_{y\to 0}\frac{F(t+y-T_n)-F(t-T_n)}{y}=F'(t-T_n)=f(t-T_n).
$$

Hence

$$
\begin{aligned}
\lambda(t) &= \sum_{n=0}^{\infty} \frac{f(t-T_n)}{1-F(t-T_n)} \mathbb{1}_{\{T_n < t \leqslant T_{n+1}\}} \\
&= \sum_{n=1}^{\infty} h(t-T_n) \mathbb{1}_{\{T_n < t \leqslant T_{n+1}\}} (\text{conditioning on } \sigma\text{-algebra}) \\
&= h(t - T_{N(t)}),
\end{aligned}
$$

where h is the hazard function of F. Plot the grap of the hazard function, we can see between any interrenewal, the tend are the same.

11.1 Hawkes Process

The Hawkes process is also called the self-exciting Poisson process. Given a Poisson process $\{N^p(t), t \geqslant 0\}$ with parameter λ and for $k \in \mathbb{Z}^+$, we have independent Poisson process $\{N^a_k(t), t \geqslant 0\}$ that are independent of $\{N^p(t), t \geqslant 0\}$ having intensity function $\{\varphi(t), t \geqslant 0\}$, which is decreasing over time. Define

$$
N(t) = \sum_{k=1}^{N^p(t)} N^a_k(t - T^p_k) + N^p(t), \ t \geqslant 0,
$$

where T^p_k is the time if the kth primary event occurs and we can think of $\{N^a_k, t \geqslant 0\}$ is the after Poisson process initiated by the kth primary events. The we will get the stochastic intensity of $\{N(t), t \geqslant 0\}$.

(a) Assume the filtration $\{\mathcal{F}_t, t \geqslant 0\}$ can distinguish after events from primary events. In this case,

$$
P(N(t+s) - N(t) = 1 | \mathcal{F}_t) = \lambda s + \sum_{k=1}^{N^p(t)} \varphi(t - T^p_k)s + o(s),
$$

$$
P(N(t+s) - N(t) > 1 | \mathcal{F}_t) = o(s).
$$

The intensity of $\{N^a_k(t), t \geqslant 0\}$ at time t given it started at time T^p_k is the intensity of $\{N^a_k(t), t \geqslant 0\}$ at time $t - T^p_k$ given it started at time 0, i,e., $\gamma(t - T^p_k)$, where $\gamma : [0, \infty) \to [0, \infty)$ s.t. $\gamma(t) \to 0$ as $t \to \infty$. Hence,

$$
\lambda(t) = \lambda + \sum_{k=1}^{N^p(t)} \gamma(t - T^p_k) = \lambda + \int_0^t \gamma(t-s) dN(s).
$$

Set

$$
Y(t) = \int_0^t \gamma(t-s) dN(s).
$$

Then $\{Y(t), t \geqslant 0\}$ is a shot noise process.

(b) Allow all events to generate the after events. Then with the same notation

$$N_1(t) = \sum_{k=1}^{N(t)} N(t - T_k) + N(t).$$

The intensity function is

$$\lambda(t) = \lambda + \int_0^t \gamma(t-s)dN(s).$$

Suppose $T_1, T_2, \cdots$ are iid random variables having density function $f(\cdots, \theta^d)$, where $\theta \in \mathbb{R}^d$ is a parameter. Suppose we take a sample of size n and observe $t_1, \ldots, t_n$. The likelihood function is

$$L(x_1, \ldots, x_n; \theta) = f(x_1, \theta) \cdots f(x_n, \theta).$$

Maximizing over Θ gives that parameter which is most likely to give the data $t_1, \ldots, t_n$.

Example 11.4. Suppose $\{N(t); t \geqslant 0\}$ is a homogeous Poisson process with rate λ. Observe the process over an interval $[0, t]$. The arrivals occurs at times

$$0 \leqslant t_1 < \cdots < t_n \leqslant t.$$

Construct the likelihood function. Suppose one just saw one arrival at $t_1 \in [0, t)$. Then

$$\begin{aligned} &P(N(t_1) = 0, N(t_1 + \Delta t) - N(t_1) = 1, N(t) - N(t_1 + \Delta t) = 0) \\ &= e^{-\lambda t_1} \cdot (\lambda \Delta t + o(\Delta t)) \cdot e^{-\lambda(t - t_1 - \Delta t)} \\ &= (\lambda \Delta t + o(\Delta t)) \cdot e^{-\lambda(t - \Delta t)}. \end{aligned}$$

Then

$$L(t_1, \lambda) = \lim_{\Delta t \to 0} \frac{(\lambda \Delta t + o(\Delta t)) \cdot e^{-\lambda(t-\Delta t)})}{\Delta t} = \lambda e^{-\lambda t}.$$

Extend the same argument,

$$L(t_1, \ldots, t_n, \lambda) = \lambda^n e^{-\lambda t}.$$

Since given t, $N(t)$ is a random variable.!!! By the conditional uniformity, the arrivals the rhs does not depend on $t_1, \ldots, t_n$. The log-likelihood function is

$$\log L(t_1, \ldots, t_n, \lambda) = n \log \lambda - \lambda t,$$

Then the MLE

$$\hat{\lambda}_M = \frac{n}{t}.$$

Can the idea be extended to a P.P with stochastic intensity function $\lambda(t)$? Suppose observe one arrival and it occurs at time t_1. Partition the interval $[0, t)$ into n intervals, each of the length is $\frac{t}{n}$, i.e.,

$$0 < \frac{t}{n} < \frac{2t}{n} < \cdots < \frac{(n-1)t}{n} < t.$$

Know t_1 is in one of the intervals $\left[\frac{(i-1)t}{n}, \frac{it}{n}\right)$. Suppose it is in $\left[\frac{k-1}{n}, \frac{kt}{n}\right)$, then probability of observing t_1 as the only point on $[0, t)$ is

$$\begin{aligned}
&L(t_1, \lambda) \\
&= P\left(N(t/n) = 0, \dots, N(kt/n) - N((k-1)t/n) = 1, \dots, N(t) - N((n-1)t/n) = 0\right) \\
&= P(N(t/n) = 0) P(N(2t/n) - N(t/n) = 0 | N(t/n) = 0) \\
&\cdot P(N(kt/n) - N((k-1)t/n) = 1 | N((k-1)t/n) - N((k-2)t/n) = 0, \dots, N(t/n) = 0) \\
&\cdot P(N(t) - N((n-1)t/n) = 0 | N((n-1)t/n) - N((n-2)t/n) = 0, \dots, N(t/n) = 0).
\end{aligned}$$

Since

$$P(N(t/n) = 0) = 1 - \lambda(t/n)1/n + o(1/n),$$
$$P(N(2t/n) - N(t/n) = 0 | N(t/n) = 0) = 1 - \lambda(2t/n)1/n + o(1/n),$$
$$\vdots$$

$$\begin{aligned}
&P(N(kt/n) - N((k-1)t/n) = 1 | N((k-1)t/n) - N((k-2)t/n) = 0, \dots, N(t/n) = 0) \\
&= \lambda(kt/n)1/n + o(1/n).
\end{aligned}$$

$$\begin{aligned}
&P(N(t) - N((n-1)t/n) = 0 | N((n-1)t/n) - N((n-2)t/n) = 0, \dots, N(t/n) = 0) \\
&= 1 - \lambda(t)1/n + o(1/n).
\end{aligned}$$

Hence

$$L(t_1, \lambda) = \prod_{j=1, j\neq k}^{n} \left(1 - \lambda(jt/n)1/n + o(1/n)\right) \left(\lambda(kt/n)1/n + o(1/n)\right)$$

Taking log,

$$\begin{aligned}
&\sum_{j=1, j\neq k}^{n} \log\left(1 - \lambda(jt/n)1/n + o(1/n)\right) + \log\left(\lambda(kt/n)1/n + o(1/n)\right) \\
&\sim -\sum_{j=1, j\neq k}^{n} \lambda(jt/n)1/n + \log \frac{\lambda(kt/n)1/n}{1 - \lambda(kt/n)1/n} \\
&\to -\int_0^t \lambda(s)ds + \log(\lambda(t_1))?
\end{aligned}$$

Observe n points at $0 < t_1 < \cdots t_n \leqslant t$, the log-likelihood function

$$\log L(t_1, \dots, t_n, \lambda) = \sum_{k=1}^{n} \log \lambda(t_i) - \int_0^t \lambda(s)ds.$$

Hawkes process is a point process. $\{N(t)\}_{t\geqslant 0}$ is defined as follows

(a) There is a homogeous Poisson process which $\{N_p(t)\}_{t\geqslant 0}$ which are primary jumps.

(b) There are independent Poisson process $\{N_a(t)\}_{t\geqslant 0}$ which have rate function $\varphi(t)dt$ and is independent of $\{N_p(t)\}_{t\geqslant 0}$. Then

$$N(t) = N_p(t) + \int_0^t N_a(t-s) N_p(ds).$$

11.1.1 Probability generating functional

Recall the probability generating function for a nonnegative integer valued random variable N is

$$G_N(z) = E[z^N] = \sum_{n=0}^{\infty} P(N=n) z^n.$$

Let h be a measurable function defined on $[0,\infty)$ such that

(a) $0 < h(t) \leqslant 1$.

(b) $1-h$ has compact support, which means that there exists $[a,b]$ such that $1-h(t)=0$ for any $t \notin [a,b]$.

The probability generating functional of a point process $\{N(t)\}_{t\geqslant 0}$ is

$$G_N(h) = E\left[\exp\left\{\int_0^\infty \log(h(t)) N(dt)\right\}\right].$$

Let $A=[a,b]$ and define $h = 1-(1-z)\mathbb{1}_A, \forall 0<z<1$. Then $1-h(t)=0$ for any $t \in A^c$, and

$$G_N(h) = E\left[\exp\left\{\int_0^\infty \log\left(1-(1-z)\mathbb{1}_{[a,b]}(t)\right) N(dt)\right\}\right].$$

Note

$$\begin{aligned}
\int_0^\infty \log\left(1-(1-z)\mathbb{1}_{[a,b]}(t)\right) N(dt) &= \sum_i \log\left(1-(1-z)\mathbb{1}_{[a,b]}(T_i)\right) \\
&= \sum_i (\log z)\mathbb{1}_{[a,b]}(T_i) \\
&= (\log z) N(a,b) \\
&= \log z^{N(a,b)}.
\end{aligned}$$

But then $G_N(h) = E\left[z^{N(a,b)}\right]$. Suppose $A_1,\ldots,A_k$ are disjoint intervals and $z_i \in (0,1)$ for any $i \in [k]$. Define

$$h(t) = 1 - \sum_{j=1}^k (1-z_j)\mathbb{1}_{A_j}(t).$$

Then the previous argument (modified) gives

$$G_N(h) = E\left[\prod_{j=1}^k z_j^{N(A_j)}\right].$$

Example 11.5. Let $\{N(t)\}_{t\geqslant 0}$ be a Poisson process with intensity measure μ, that is, $\mu(A) = E[N(A)]$. Then $E[z^N] = e^{-\lambda(1-z)}$. Let

$$h(t) = 1 - \sum_{j=1}^k (1-z_j)\mathbb{1}_{A_j}(t).$$

Then

$$\begin{aligned}
G_N(h) &= E\left[\prod_{j=1}^{k} z_j^{N(A_j)}\right] = \prod_{j=1}^{k} E\left[Z_j^{N(A_j)}\right] \\
&= \prod_{j=1}^{k} e^{-\mu(A_j)(1-z_j)} = \exp\left\{-\sum_{j=1}^{k}\mu(A_j)(1-z_j)\right\} \\
&= \exp\left\{-\int_0^\infty \sum_{j=1}^{k}(1-z_j)\mathbb{1}_{A_j}(x)\mu(dx)\right\} \\
&= \exp\left\{-\int_0^\infty (1-h(t))\mu(dt)\right\}.
\end{aligned}$$

Define

$$\begin{aligned}
\tau_x : \mathbb{R}^{\geqslant 0} &\to \mathbb{R}^{\geqslant 0} \\
y &\mapsto x+y.
\end{aligned}$$

Suppose that μ is a measure on a measurable space $(E, \mathcal{E})$ and let $(G, \mathcal{G})$ be another measurable space. Suppose that $f : E \to G$ is a measurable function. For $B \in \mathcal{G}$ define

$$\mu \circ f^{-1}(B) = \mu(f^{-1}(B)) = \mu\{x \in E : f(x) \in B\}.$$

Then, $\mu \circ f^{-1}$ define a measure on $(E, \mathcal{E})$. If g is nonnegative measurable function of G, then

$$\int_G g(y)\mu \circ f^{-1}(dy) = \int_E g \circ f(x)\mu(dx),$$

which is true by the definition of $\mu \circ f^{-1}$ when $g = \mathbb{1}_B$ for $B \in \mathcal{G}$ and extends to nonnegative measurable functions by the usual arguments. Also,

$$\int_G g(y)\mu \circ \tau_x(dy) = \int_G g \circ \tau_x(y)\mu(dy) = \int_G g(x+y)\mu(dy).$$

If μ is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}^{\geqslant 0}))$, define a new measure $\mu \circ \tau_x^{-1}$ on $(\mathbb{R}^{\geqslant 0}, \mathcal{B}(\mathbb{R}^{\geqslant 0}))$ by

$$\mu \circ \tau_x^{-1}(B) = \mu(\tau_x^{-1}(B)) = \mu(B - x),$$

where $B \in \mathcal{B}(\mathbb{R}^{\geqslant 0})$ and

$$\tau_x^{-1}(B) = \{y \in \mathbb{R}^{\geqslant 0} : \tau_x(y) \in B\} = \{y \in \mathbb{R}^{\geqslant 0} : x + y \in B\} = B - x.$$

Let δ_x be the dirac delta function. Then

$$\delta_s \circ \tau_x^{-1}(A) = \delta_s(A - x) = \begin{cases} 1 & s \in A - x \text{ or } s + x \in A, \\ 0 & \text{otherwise.} \end{cases} = \delta_{s+x}(A).$$

If $\mu = \sum_i \delta_{s_i}$ (a realization of a point process), then

$$\mu \circ \tau_x^{-1}(A) = \sum_i \delta_{s_i + x}(A) = \mu(A - x).$$

Consider Hawkes process. Let N_c be a stationary Poisson random measure having intensity $\lambda \cdot l$, where l is the Lebesgue measure. Let $\{N_k\}_{k\in\mathbb{Z}^+}$ be a iid Poisson random measures, independent of N_c with common intensity $\varphi(x)dx$, where $\varphi : \mathbb{R}^{\geqslant 0} \to \mathbb{R}^{\geqslant 0}$, and $\varphi(x) \to 0$ as $x \to \infty$. The Hawkes process is the random measure on $(\mathbb{R}^{\geqslant 0}, \mathcal{B}(\mathbb{R}^{\geqslant 0}))$ given by

$$\begin{aligned} N(A) &= N_c(A) + \int_{\mathbb{R}^{\geqslant 0}} N_{N_c((0,s])} \circ \tau_s^{-1}(A) N_c(ds) \\ &= N_c(A) + \int_{\mathbb{R}^+} N_{N_c((0,s])} \circ \tau_s(A-s) N_c(ds) \\ &= N_c(A) + \int_{\mathbb{R}^+} N_{N_c((0,s])}(\tau_s(A-s)) N_c(ds), \end{aligned}$$

where $N_c(\{s\}) \in \{0,1\}$, and then we can see as if N_k starts at 0. Setting $N(t) = N(0,t]$, one gets

$$\begin{aligned} N(t) &= N_c(t) + \int_{\mathbb{R}^{\geqslant 0}} N_{N_c((0,s])}((0,t-s]) N_c(ds) \\ &= N_c(t) + \int_{\mathbb{R}^{\geqslant 0}} N_{N_c(s)}(t-s) N_c(ds). \end{aligned}$$

So if $0 < T_1 < T_2 < \cdots$ are the event time of N_c, i.e., $N_c = \sum_{i=1}^{\infty} \delta_{T_i}$. We have

$$\begin{aligned} N(A) &= \sum_{i=1}^{\infty} \delta_{T_i}(A) + \int_{\mathbb{R}^{\geqslant 0}} N_{N_c((0,s])} \circ \tau_s^{-1}(A) \sum_{i=1}^{\infty} \delta_{T_i}(ds) \\ &= \sum_{i=1}^{\infty} \delta_{T_i}(A) + \sum_{i=1}^{\infty} \int_{\mathbb{R}^{\geqslant 0}} N_{N_c((0,s])} \circ \tau_s^{-1}(A) \delta_{T_i}(ds) \\ &= \sum_{i=1}^{\infty} \delta_{T_i}(A) + \sum_{i=1}^{\infty} N_i \circ \tau_{T_i}^{-1}(A) \\ &= \sum_{i=1}^{\infty} \left(\delta_{T_i} + \sum_{i=1}^{\infty} N_i \circ \tau_{T_i}^{-1} \right)(A). \end{aligned}$$

Consider the probability generating functional

$$G_N(h) = E\left[\exp\left\{ \int_0^{\infty} \log(h(t)) N(dt) \right\} \right].$$

By the definition of N,

$$\begin{aligned}
&E\left[\exp\left\{\int_{\mathbb{R}_{\geqslant 0}}\log(h(t))N(dt)\right\}\middle| N_c=\sum_{i=1}^{\infty}\delta_{t_i}\right]\\
&=E\left[\exp\left\{\sum_{i=1}^{\infty}\left[\log(h(t_i))+\int_{\mathbb{R}_{\geqslant 0}}\log(h(t))N_i\circ\tau_{t_i}^{-1}(dt)\right]\right\}\middle| N_c=\sum_{i=1}^{\infty}\delta_{t_i}\right]\\
&=E\left[\exp\left\{\sum_{i=1}^{\infty}\left[\log(h(t_i))+\int_{\mathbb{R}_{\geqslant 0}}\log(h(t))N_i\circ\tau_{t_i}^{-1}(dt)\right]\right\}\right]\\
&=E\left[\prod_{i=1}^{\infty}\left[h(t_i)\exp\left\{\int_{\mathbb{R}_{\geqslant 0}}\log(h(t))N_i\circ\tau_{t_i}^{-1}(dt)\right\}\right]\right]\\
&=\prod_{i=1}^{\infty}h(t_i)E\left[\exp\left\{\int_{\mathbb{R}_{\geqslant 0}}\log(h(t))N_i\circ\tau_{t_i}^{-1}(dt)\right\}\right]\\
&=\prod_{i=1}^{\infty}h(t_i)G_{N_1\circ\tau_{t_i}^{-1}}(h).
\end{aligned}$$

Thus,

$$E\left[\exp\left\{\int_{\mathbb{R}_{\geqslant 0}}\log(h(t))N(dt)\right\}\middle| N_c=\sum_{i=1}^{\infty}\delta_{T_i}\right]=\prod_{i=1}^{\infty}h(t_i)G_{N_1\circ\tau_{T_i}^{-1}}(h).$$

Taking expectations, yield

$$\begin{aligned}
G_N(h)&=E\left[\exp\left\{\sum_{i=1}^{\infty}\log\left(h(T_i)\circ G_{N_1\circ\tau_{T_i}^{-1}}(h)\right)\right\}\right]\\
&=E\left[\exp\left\{\int_0^{\infty}\log\left(h(s)G_{N_1\circ\tau_s^{-1}}(h)N_c(ds)\right)\right\}\right]\\
&=G_{N_c}\left(h(\cdot)G_{N_1}\circ\tau_{(\cdot)}^{-1}\right).
\end{aligned}$$

Note that

$$\begin{aligned}
G_{N_1\circ\tau_s^{-1}}(h)&=E\left[\exp\left\{\int_{\mathbb{R}_{\geqslant 0}}\log(h(t))N_1\circ\tau_s^{-1}(dt)\right\}\right]\\
&=E\left[\exp\left\{\int_{\mathbb{R}_{\geqslant 0}}\log(h\circ\tau_s(x))N_1(dx)\right\}.\right]
\end{aligned}$$

Since N_1 is a Poisson random measure with intensity $\varphi(x)dx$, by previous result,

$$G_{N_1\circ\tau_s^{-1}}(h)=\exp\left\{-\int_{\mathbb{R}_{\geqslant 0}}(1-h(s+x))\,\varphi(x)\right\}.$$

Since N_c is a stationary Poisson random measure with intensity λl, by previosu result,

$$\begin{aligned} G_N(h) &= G_{N_c}\left(h(\cdot)G_{N_1}\circ\tau_{(\cdot)}^{-1}\right) \\ &= \exp\left\{-\int_{\mathbb{R}^{\geqslant 0}}\left(1-h(s)G_{N_1\circ\tau_s^{-1}}(h)\right)\lambda ds\right\} \\ &= \exp\left\{-\int_{\mathbb{R}^{\geqslant 0}}\left(1-h(s)\exp\left\{-\int_{\mathbb{R}^{\geqslant 0}}\left(1-h(s+x)\right)\varphi(x)\right\}\right)\lambda ds\right\}. \end{aligned}$$

For $a, c \in \mathbb{R}^{\geqslant 0}$,

$$\begin{aligned} E[N((a,a+c])) &= E[N_c((a,a+c])] + E\left[\int_{\mathbb{R}^{\geqslant 0}]} N_{N_c((0,s])}\circ\tau_s^{-1}((a,a+c]N_c(ds)\right] \\ &= \lambda c + E\left[\int_{\mathbb{R}^{\geqslant 0}]} N_{N_c((0,s])}\circ\tau_s^{-1}((a,a+c]N_c(ds)\right]. \end{aligned}$$

Let $U_1,\ldots,U_n$ be the unordered event times of N_c in $(0,a+c]$. Then

$$\begin{aligned} &E\left[\int_{\mathbb{R}^{\geqslant 0}]} N_{N_c((0,s])}\circ\tau_s^{-1}((a,a+c]N_c(ds)\,\middle|\, N_c(0,a+c]=n\right] \\ &= E\left[\sum_{i=1}^n N_i\circ\tau_{u_i}^{-1}((a,a+c])\,\middle|\, N_c((0,a+c])=n\right], \end{aligned}$$

Given $N((0,a+c])=n$, $U_1,\ldots,U_n$ are iid Uniform$(0,a+c)$ distributed random variables and since $N_1,\ldots,N_n$ are iid Poisson random measures having intensity $\varphi\cdot l$,

$$\begin{aligned} &E\left[\sum_{i=1}^n N_i\circ\tau_{U_i}^{-1}((a,a+c])\,\middle|\, N_c((0,a+c])=n\right] \\ &= E\left[\sum_{i=1}^n N_i\circ\tau_{U_i}^{-1}((a,a+c])\right] = \sum_{i=1}^n E\left[N_i\circ\tau_{U_i}^{-1}((a,a+c])\right] \\ &= \sum_{i=1}^n\left[\int_0^a E[N_i((a-s,a+c-s])]\frac{1}{a+c}ds + \int_a^{a+c} E[N_i((0,a+c-s])]\frac{1}{a+c}ds\right] \\ &= \frac{n}{a+c}\left[\int_0^a\int_{a-s}^{a+c-s}\varphi(x)dxds + \int_a^{a+c}\int_0^{a+c-s}\varphi(x)dxds\right]. \end{aligned}$$

Thus,

$$\begin{aligned} &E\left[\sum_{i=1}^n N_i\circ\tau_{\mu_i}^{-1}((a,a+c])\,\middle|\, N_c((0,a+c])\right] \\ &= \frac{N_c((0,a+c])}{a+c}\left[\int_0^a\int_a^{a+c-s} + \int_a^{a+c}\int_0^{a+c-s}\right]\varphi(x)dxds. \end{aligned}$$

Therefore,

$$\begin{aligned}E[N((a,a+c])] &= \lambda c + E\left[\frac{N_c((0,a+c])}{a+c}\left[\int_0^a\int_a^{a+c-s}+\int_a^{a+c}\int_0^{a+c-s}\right]\varphi(x)dxds\right]\\ &= \lambda c + \lambda\left[\int_0^a\int_a^{a+c-s}+\int_a^{a+c}\int_0^{a+c-s}\right]\varphi(x)dxds.\end{aligned}$$

To discuss limiting properties of the Hawkes process, we assume that the function φ is uniformly bounded and has compact support, i.e., there exists $T>0$ such that $\varphi(t)=0$ for any $t>T$. Goal: Find

$$\lim_{t\to\infty} P(N(t,t+c)=k).$$

Let $0=T_0<T_1<T_2<\cdots$ be the event times of N_c. Define $K_0=0$ and

$$K_n=\min\{n>K_{n-1}:T_n-T_{K_{n-1}}>T\}.$$

For $t\geqslant 0$, set

$$\mathcal{F}_t=\sigma\{N(A):A\in\mathcal{B}([0,t])\}.$$

Let

$$\mathcal{F}_\infty=\sigma(\{\mathcal{F}_t\}_{t\in\mathbb{R}^{\geqslant 0}}).$$

Since $\{K_n\}_{n\in\mathbb{N}}$ are stopping times,

$$\mathcal{F}_{K_n}=\{A\in\mathcal{F}_\infty : A\cap\{T_{K_n}\leqslant t\}\in\mathcal{F}_t,\forall t\in\mathbb{R}^{\geqslant 0}\}.$$

Theorem 11.6. *Let $A\in\mathcal{B}(\mathbb{R}^{\geqslant 0})$. Then the random measures $\{N(T_{K_n}+A)\}_{n\in\mathbb{Z}^+}$ are independent of $\mathcal{F}_{K_n}$. Moreover, $\{N(T_{K_n}+A)\}_{n\in\mathbb{Z}^+}$ and $N(A)$ are identically distributed.*

Proof. Let $n\in\mathbb{Z}^+$. By definition,

$$\begin{aligned}N(T_{K_n}+A) &= N_c(T_{K_n}+A)+\int_{\mathbb{R}^{\geqslant 0}} N_{N_c((0,s])}\circ\tau_s^{-1}(T_{K_n}+A)N_c(ds)\\ &= N_c(T_{K_n}+A)+\sum_{j=1}^{\infty}N_j\circ\tau_{T_j}^{-1}(T_{K_n}+A)\\ &= N_c(T_{K_n}+A)+\sum_{j=1}^{\infty}N_j(T_{K_n}-T_j+A)\\ &= N_c(T_{K_n}+A)+\sum_{j=K_n}^{\infty}N_j(T_{K_n}-T_j+A),\end{aligned}$$

since $T_{K_n}-T_j>T$ for any $1\leqslant j\leqslant K_n-1$, and $\varphi(t)=0$ for any $t>T$, then we have

$$N_j(T_{K_n}-T_j+A)=0,\forall 1\leqslant j\leqslant K_n-1.$$

(!!!) Since N_c is stationary, $\{T_j-T_{K_n}\}_{j\geqslant K_n}$ has the same distribution as $\{T_n\}_{n\in\mathbb{N}}$? Also the processes $\{N_n\}_{n\in\mathbb{N}}$ are independent of N_c, we obtain

$$N(T_{K_n}+A)\stackrel{d}{=}N_c(A)+\sum_{n=0}^{\infty}N_n(T_n+A)=N(A).$$

□

Chapter 12

Martingales

12.1 Signed Measure (Royden)

Definition 12.1 (Mutually singular)**.** Two measures ν_1 and ν_2 on (X, M) are said to be *mutually singular* if there are disjoint measurable sets A and B with $X = A \sqcup B$ for which

$$\nu_1(B) = \nu_2(A) = 0.$$

Two measures ν^+ and ν^- defined below are mutually singular.

Definition 12.2 (The Jordan Decomposition Theorem)**.** Let ν be a signed measure on the measurable space (X, M). Then there are two mutually singular measure ν^+ and ν^- on (X, M) for which $\nu = \nu^+ - \nu^-$. Moreover, there is only one such pair of mutually singular measures.

Example 12.3. Let $f : \mathbb{R} \to \mathbb{R}$ be a function that is Lebesgue integrable over $\mathbb{R}$. For a Lebesgue measurable set E, define $\nu(E) = \int_E f dm$. We infer form the countable additivity of integration that ν is a signed measure on the measurable space $(\mathbb{R}, \mathcal{L})$, where $\mathcal{L}$ is the collEction of Lebesgue measurable subsets of $\mathbb{R}$. Define $A = \{x \in \mathbb{R} | f(x) \geqslant 0\}$ and $B = \{x \in \mathbb{R} | f(x) < 0\}$ and define, for each Lebesgue measurable set E,

$$\nu^+(E) = \int_{A \cap E} f dm$$

and

$$\nu^-(E) = -\int_{B \cap E} f dm.$$

Then $\{A, B\}$ is a Hahn decomposition of $\mathbb{R}$ with respect to the signed measure ν. Moreover,

$\nu := \nu^+ - \nu^-$ is a Jordan decomposition of ν. Then

$$\begin{aligned}
\nu^+(E) - \nu^-(E) &= \int_{A\cap E} f dm - \int_{B\cap E} f dm \\
&= \int_{A\cap E} f^+ dm + \int_{B\cap E} \left(-f^-\right) dm \\
&= \int_{A\cap E} \left(f^+ - f^-\right) dm + \int_{B\cap E} \left(f^+ - f^-\right) dm \\
&= \int_{A\cap E} f dm + \int_{B\cap E} f dm \\
&= \int_{(A\cap E)\cup(B\cap E)} f dm \\
&= \int_E f dm \\
&= \nu(E),
\end{aligned}$$

where $\nu := \nu^+ - \nu^-$.

Definition 12.4 (The absolute value of measure)**.** The measure $|\nu|$ is defined on M by

$$|\nu|(E) = \nu^+(E) + \nu^-(E), \forall E \in M.$$

Let (X, M) be a measurable space. For μ a measure on (X, M) and f a nonnegative function on X that is measurable with respect to M, defind the set function ν on M by

$$\nu(E) = \int_E f d\mu, \forall E \in M.$$

Then from the linearity of integration and the monotone convergence theorem that ν is a measure on the measurable space (X, M), and it has the property

$$\text{if } E \in M \text{ and } \mu(E) = 0, \text{ then } \nu(E) = 0,$$

which implies $\nu \ll \mu$.

Proposition 12.5. *$\nu \ll \mu$ if and only if for any $\epsilon > 0$, there is a $\delta > 0$ such that for any set $E \in M$, if $\mu(E) < \delta$, then $\nu(E) < \epsilon$.*

Proof. $\Longleftarrow$ is obvious.

"$\Rightarrow$". Assume there is an $\epsilon > 0$ and a $\{E_n\}_{n=1}^{\infty}$ with $\mu(E_n) < \frac{1}{2^n}$ while $\nu(E_n) \geqslant \epsilon_0, \forall n \in \mathbb{N}$. Define

$$A = \bigcap_{k=1}^{\infty} \bigcup_{k=n}^{\infty} E_k.$$

Then

$$\mu(A) = \lim_{n\to\infty} \mu\left(\bigcup_{k=n}^{\infty} E_k\right) \leqslant \lim_{n\to\infty} \sum_{k=n}^{\infty} \mu(E_k) \leqslant \lim_{n\to\infty} \sum_{k=n}^{\infty} \frac{1}{2^n} = 0.$$

But

$$\nu(A) = \lim_{n\to\infty} \nu\left(\bigcup_{k=n}^{\infty} E_k\right) \geqslant \epsilon_0,$$

since $\nu(E_n) \geqslant \epsilon_0$ for any $n \in \mathbb{N}$, contradiction. □

Theorem 12.6 (Radon-Nikodym Theorem)**.** *Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space and ν a σ-finite measure defined on the measurable space $(X, \mathcal{M})$ and $\nu << \mu$. Then there is a nonnegative function f on X that is measurable w.r.t $\mathcal{M}$ for which*

$$\nu(E) = \int_E f d\mu \ \forall\ E \in \mathcal{M}.$$

The function f is unique in the sense that if g also has this property, then $g = f$ μ a.e..

12.2 Radon Nikodym Derivative

Let $(\Omega, \mathcal{B})$ be a measurable space, and let μ, λ be positive and bounded measures on $(\Omega, \mathcal{B})$.

Definition 12.7. We say λ is *absolutely continuous* w.r.t μ (written as $\lambda << u$) if

$$\mu(A) = 0 \Longrightarrow \lambda(A) = 0, \forall A \in \mathcal{B}.$$

Definition 12.8. We say that λ *concentrate* on a set $A \in \mathcal{B}$ if

$$\lambda(A^c) = 0.$$

Definition 12.9. We say that λ and μ are *mutually singular* if there exist sets $A, B \in \mathcal{B}$ such that $A \cap B = \emptyset$, $\lambda(A^c) = 0$ and $\lambda(B^c) = 0$.

Let's suppose H is a real Hilbert space.

Proposition 12.10 (Riesz Representation Theorem)**.** Let $L : H \to \mathbb{R}$ be a continuous linear functional, then there exists a unique $y \in H$ such that

$$L(x) = \langle x, y\rangle, \forall x \in H.$$

Lemma 12.11 (Integral Comparison Lemma)**.** Suppose $X, Y \in L(\Omega, \mathcal{B}, P)$. Then

$$\int_A X dP \leqslant \int_A Y dP, \forall A \in \mathcal{B} \Longrightarrow X \leqslant Y, \text{a.e..}$$

Complete version:

$$\int_A X dP \leqslant (\geqslant / =) \int_A Y dP, \forall A \in \mathcal{B} \Longleftrightarrow X \leqslant (\geqslant / =) Y, \text{a.e..}$$

Proof. Claim. if $X \geqslant 0$, then $P(X > 0) > 0$ implies $E(X) > 0$. Since $P(X > 0) > 0$, $\exists n \in \mathbb{N}^+$ such that

$$P\left(X > \frac{1}{n}\right) > 0,$$

otherwise,

$$P(X>0) \leqslant \lim_{n\to\infty} P\left(X>\frac{1}{n}\right) = P\left(\bigcup_{n=1}^{\infty}\left\{X>\frac{1}{n}\right\}\right) \leqslant \sum_{n=1}^{\infty} P\left(X>\frac{1}{n}\right) = 0.$$

Then

$$E(X) \geqslant \frac{1}{n} P\left(X>\frac{1}{n}\right) > 0,$$

which is a contradiction. Assume $P(X>Y)>0$. Let

$$A := \{X>Y\} \text{ and consider } \{(X-Y)\mathbb{1}_A \geqslant 0\}.$$

Then

$$P\left((X-Y)\mathbb{1}_A > 0)\right) \geqslant P(A) > 0.$$

So

$$E\left((X-Y)\mathbb{1}_A > 0\right) > 0,$$

which is a contradiction. Thus, $X \leqslant Y$ a.e.. ☐

Theorem 12.12 (Radon Nikodym)**.** *Let $(\Omega, \mathcal{B}, P)$ be probability space. Suppose ν is a **positive** bounded measure and $\nu \ll P$. Then there exists integrable random variable X satisfying*

$$\nu(A) = \int_A X dP, \forall A \in \mathcal{B}.$$

X is Radon-Nikodym derivative and is written $X = \frac{d\nu}{dP}$ or $d\nu = X dP$.

Proof. Define

$$\begin{aligned} Q : \mathcal{B} &\longrightarrow [0,1] \\ A &\longmapsto \frac{\nu(A)}{\nu(\Omega)}. \end{aligned}$$

Then Q is a probability measure on $(\Omega, \mathcal{B})$ and $Q \ll P$ since $Q \ll \nu \ll P$. Define

$$\begin{aligned} P^* : \mathcal{B} &\to [0,1] \\ A &\mapsto \frac{P(A)+Q(A)}{2}. \end{aligned}$$

which is also a probability measure on $(\Omega, \mathcal{B})$ since $P^*(\Omega) = 1$. Then

$$H := L_2(\Omega, \mathcal{B}, P^*),$$

where we say a random variable

$$(\mathbb{R}, \mathcal{B}(\mathbb{R})) \leftarrow (\Omega, \mathcal{B}) : X \in L^2(\Omega, \mathcal{B}, P^*)$$

if $X \in \mathcal{B}$ and $\int_\Omega |X|^2 dP^* < \infty$. Then H is a Hilbert space (up to equivalent classes since $\langle X, X\rangle = \int |X|^2 dP^* = 0 \not\Rightarrow X = 0$) with inner product

$$\langle X, Y\rangle = \int_\Omega XY dP^*.$$

Then

$$\|Y\|_2 = \sqrt{\langle Y, Y\rangle} = \left(\int_\Omega |Y|^2 dP^*\right)^{\frac{1}{2}}, \forall Y \in H.$$

Note that all elements of $H = L_2(\Omega, \mathcal{B}, P^*)$ are $\mathcal{B}/\mathcal{B}(\mathbb{R})$-measurable (random variable). Then we can define the functional

$$\begin{aligned} L : H &\longrightarrow \mathbb{R} \\ Y &\longmapsto \int_\Omega Y dQ \end{aligned}$$

so that $L : L_2(P^*) \to \mathbb{R}$ is linear and bounded (and hence continuous). Since

$$|L(Y)| \leqslant \int |Y| dQ \leqslant \int |Y| dQ + \int |Y| dP \leqslant 2 \int |Y| dP^* \leqslant 2 \left(\int Y^2 dP^*\right)^{\frac{1}{2}} = 2\|Y\|_2.$$

So L is bounded. Then L is (Lipschitz) continuous since H is a normed linear space. Thus, by Proposition 12.10, there exists $Z \in H$ such that for any $Y \in H$,

$$\int Y dQ = L(Y) = \langle Y, Z\rangle = \int Y Z dP^*.$$

Let $Y = \mathbb{1}_A$, where $A \in \mathcal{B}$, then

$$Q(A) = \int_A dQ = \int_A Z dP^*.$$

Then assuming P^* is positive,

$$0 \leqslant \frac{Q(A)}{P^*(A)} = \frac{\int_A Z dP^*}{P^*(A)} \leqslant \frac{\int_A Z dP^*}{Q(A)/2} = 2,$$

since $2P^* = P + Q \geqslant Q$. Then

$$0 \leqslant \int_A Z dP^* = Q(A) \leqslant 2P^*(A), \forall A \in \mathcal{B},$$

that is,

$$0 \leqslant \int_A Z dP^* \leqslant \int_A 2 dP^*, \forall A \in \mathcal{B}.$$

From the Integral Comparison Lemma 12.11,

$$0 \leqslant Z \leqslant 2, \ P^*\text{-a.s.}.$$

Since

$$\int Y dQ = \int Y Z dP^* = \int \frac{1}{2} Y Z dP + \int \frac{1}{2} Y Z dQ,$$

we have

$$\int Y \left(1 - \frac{Z}{2}\right) dQ = \int \frac{YZ}{2} dP, \forall Y \in H.$$

Set $Y = \mathbb{1}_{\{Z=2\}}$ to get

$$\int_{\{Z=2\}} \left(1 - \frac{Z}{2}\right) dQ = \int_{\{Z=2\}} \frac{Z}{2} dP,$$

that is, $0 = P(Z = 2)$. Since $Q << P$, we have $0 = Q(Z = 2)$. Hence $P^*(Z = 2) = 0$. Thus,

$$0 \leqslant Z < 2, \; P^*\text{-a.s.}.$$

Let

$$Y = \left(\frac{Z}{2}\right)^n \mathbb{1}_A, A \in \mathcal{B}.$$

Then

$$Y \in H = L_2(P^*) \text{ and } 0 \leqslant Y < 1 \; P/Q/P^*\text{-a.s.}$$

Since

$$\int Y \left(1 - \frac{Z}{2}\right) dQ = \int \frac{YZ}{2} dP, \forall Y \in H,$$

$$\int_A \left(\frac{Z}{2}\right)^n \left(1 - \frac{Z}{2}\right) dQ = \int_A \left(\frac{Z}{2}\right)^{n+1} dP.$$

Sum both sides over $n = 0$ to $n = N$ to get

$$\int_A \left(1 - \left(\frac{Z}{2}\right)^{N+1}\right) dQ = \int_A \frac{\frac{Z}{2}\left(1 - \left(\frac{Z}{2}\right)^{N+1}\right)}{1 - \frac{Z}{2}} dP.$$

Since $1 - \left(\frac{Z}{2}\right)^{N+1} \uparrow 1$ Q-a.s., by MCT,

$$\frac{\nu(A)}{\nu(\Omega)} = Q(A) = \int_A \frac{Z}{2 - Z} dP := \int_A \frac{X}{\nu(\Omega)} dP, \forall A \in \mathcal{B}.$$

Thus,

$$\nu(A) = \int_A X dP, \forall A \in \mathcal{B}.$$

□

Corollary 12.13. Suppose Q and P are probability measures on $(\Omega, \mathcal{B})$ such that $Q \ll P$. Let $\mathcal{G} \subseteq \mathcal{B}$ be a subσ-algebra. Let $Q|_{\mathcal{G}}$, $P|_{\mathcal{G}}$ be the restrictions of Q and P to $\mathcal{G}$. Then in $(\Omega, \mathcal{G})$,

$$Q|_{\mathcal{G}} \ll P|_{\mathcal{G}}$$

and

$$\frac{dQ|_{\mathcal{G}}}{dP|_{\mathcal{G}}}$$

is $\mathcal{G}$-measurable.

Proof. Check the proof of the Radon-Nikodym theorem. Since $Z \in H$,

$$X = \frac{Z}{2 - Z} \nu(\Omega) \in \mathcal{B}.$$

□

12.3 Conditional Expectation

Definition 12.14. Suppose $X \in L_1(\Omega, \mathcal{B}, P)$ and let $\mathcal{G} \subseteq \mathcal{B}$ be a sub-σ-field. Then there exists a random variable $E(X|\mathcal{G})$, called the conditional expectation of X w.r.t to $\mathcal{G}$ such that

(a) $E(X|\mathcal{G})$ is $\mathcal{G}/\mathcal{B}(\mathbb{R})$-measurable and integrable.

(b) $\forall\, A \in \mathcal{G}$, we have
$$\int_A X dP = \int_A E(X|\mathcal{G}) dP.$$

Or
$$\int_A \frac{d\nu}{dP} = \int_A \frac{d\nu|_{\mathcal{G}}}{dP|_{\mathcal{G}}} dP,$$
where
$$\nu(A) = \int_A X dP, \forall A \in \mathcal{B}.$$

Why does this defintion of conditional expectation make mathematical sense? We will show the existence.

Proof. Suppose $X \in L_1(\Omega, \mathcal{B}, P)$. Suppose initially that $X \geqslant 0$. Define
$$\nu(A) = \int_A X dP, \forall A \in \mathcal{B}.$$
Then since $\nu(\Omega) = E[X] < \infty$, ν is a bounded measure. Moreover, $\nu \ll P$ on $\mathcal{B}$, so
$$\nu|_{\mathcal{G}} \ll P|_{\mathcal{G}}.$$
From the Radon-Nikodym theorem, the derivative exists and we set
$$E[X|\mathcal{G}] := \frac{d\nu|_{\mathcal{G}}}{dP|_{\mathcal{G}}},$$
or
$$d\nu = E[X|\mathcal{G}] dP \text{ on } \mathcal{G},$$
or for any $A \in \mathcal{G}$,
$$\int_A X dP = \nu(A) = \nu|_{\mathcal{G}}(A) = \int_A d\nu|_{\mathcal{G}} = \int_A \frac{d\nu|_{\mathcal{G}}}{dP|_{\mathcal{G}}} dP|_{\mathcal{G}} = \int_A \frac{d\nu|_{\mathcal{G}}}{dP|_{\mathcal{G}}} dP := \int_A E[X|\mathcal{G}] dP,$$
which is (ii) and part of (i) of the definition of conditional expectation. Besides, by 12.13, $\frac{d\nu|_{\mathcal{G}}}{dP|_{\mathcal{G}}}$ is $\mathcal{G}$-measurable. Thus, $\frac{d\nu|_{\mathcal{G}}}{dP|_{\mathcal{G}}}$ is the conditional expectation of X. Without the condition $X \geqslant 0$, define
$$\nu(A) = \int_A X dP,\ A \in \mathcal{B}.$$
Since $X \in L_1(\Omega, \mathcal{B}, P)$, $X^{\pm} \in L_1(\Omega, \mathcal{B}, P)$. Define
$$\nu^{\pm}(A) = \int_A X^{\pm} dP, \forall A \in \mathcal{G}.$$

Then $\nu^{\pm}$ are bounded measure and $\nu^{\pm}|_{\mathcal{G}} \ll P|_{\mathcal{G}}$. Smilarly, by the Rodan-Nikodym theorem, $E[X^+|\mathcal{G}]$ and $E[X^{-1}|\mathcal{G}]$ satisfy the two conditions of the conditional expectation. Then after defining

$$E[X|\mathcal{G}] :=]E[X^+|\mathcal{G}] - E[X^-|\mathcal{G}],$$

we have it satisfies (i) of the definition of the conditional expectation. Also, $\nu = \nu^+ - \nu^-$ is a Jordan decomposition of ν. So $\nu^+(A) - \nu^-(A) = \nu(A)$. Then for any $A \in \mathcal{G}$,

$$\begin{aligned}
\int_A X dP = \nu(A) &\left(= \int_A X dP = \int_A (X^+ - X^-) dP = \int_A X^+ dP - \int_A X^- dP \right) \\
&= \nu^+(A) - \nu^-(A) = \int_A E[X^+|\mathcal{G}] dP - \int_A E[X^-|\mathcal{G}] dP \\
&= \int_A \left(E[X^+|\mathcal{G}] - E[X^-|\mathcal{G}] \right) dP \\
&=: \int_A E[X|\mathcal{G}] dP, (!!!)
\end{aligned}$$

which implies $E[X^+|\mathcal{G}] - E[X^-|\mathcal{G}]$ satisfies (ii) of the definition of the conditional expectation. Thus, $E[X^+|\mathcal{G}] - E[X^-\mathcal{G}]$ is the conditional expectation of X. □

Remark. We can write

$$E[\cdot|\mathcal{G}] : L_1(\mathcal{F}) \to L_1(\mathcal{G}).$$

Definition 12.15 (Conditional expectation w.r.t. an event)**.**

$$E[X|A] = \frac{E[\mathbb{1}_A X]}{P(A)} = \int_{x \in \mathcal{X}} x P(dx|A).$$

So

$$P[B|A] = \frac{P(A \cap B)}{P(A)}.$$

Definition 12.16 (Conditional expectation w.r.t. a r.v.)**.** If Y is a discrete random variable with range $\mathcal{Y}$, then we can define

$$\begin{aligned}
g : \mathcal{Y} &\to \mathbb{R} \\
y &\mapsto E[X|Y = y]
\end{aligned}$$

Then $g \circ Y$ is called the conditional expectation of X w.r.t. Y so that we have

$$\begin{aligned}
E[X|Y] : \Omega &\to \mathbb{R} \\
\omega &\mapsto E[X|Y = Y(\omega)],
\end{aligned}$$

which is a random variable. If Y is a continuous random variable, as explained in the Borel-Kolmogorov paradox, we have to specify what limiting procedure produces the set $Y = y$. This can be naturally done by defining the set

$$H_y^{\epsilon} = \{\omega \mid \|Y(\omega) - y\| < \epsilon\},$$

so that if $P(H_y^\epsilon) > 0$ for all $\epsilon > 0$, then

$$\begin{aligned} g : \mathcal{Y} &\to \mathbb{R} \\ y &\mapsto \lim_{\epsilon\to 0} E\left[X|H_y^\epsilon\right]. \end{aligned}$$

The modern definition is anlogous to the above except that the above limiting process is replaced by the Radon-Nikodym derivative just instroduced, so the result holds more generally.

Definition 12.17 (Conditional expectation w.r.t. a σ-algebra)**.** Let $X \in L_1(\Omega, \mathcal{B}, P)$. The function $X : \Omega \to \mathbb{R}$ is ususally not $\mathcal{G}$-measurable, thus the existence of the integrals of the form $\int_A X dP_{\mathcal{G}}$, where $A \in \mathcal{G}$ and $P_{\mathcal{G}}$ is the restriction of P to $\mathcal{G}$ cannot be stated in general . However, the local averages $\int_A X dP$ can be recovered in $(\Omega, \mathcal{G}, P|_{\mathcal{G}})$ with the help of the conditional expectation. A conditional expectation of X given $\mathcal{G}$, denoted as $E[X|\mathcal{G}]$, is any $\mathcal{G}$-measurable function which satisfies:

$$\int_A E[X|\mathcal{G}]dP = \int_A X dP, \forall A \in \mathcal{G}.$$

The existence of $E[X|\mathcal{G}]$ can be established by noting that

$$\begin{aligned} \mu^X : \mathcal{B} &\to \mathbb{R} \\ A &\mapsto \int_A X dP, \end{aligned}$$

which is a finite measure. Then $\mu^X \ll P$. Furthermore, if π is the natural injection from $\mathcal{G}$ to $\mathcal{B}$, then $\mu^X \circ \pi = \mu^X_{|\mathcal{G}}$ is the restriction of μ^X to $\mathcal{G}$ and $P \circ \pi = P_{|\mathcal{G}}$ is the restriction of P to $\mathcal{G}$ and $\mu^X \circ \pi \ll P \circ \pi$ since for any $A \in \mathcal{G}$,

$$P \circ \pi(A) = 0 \iff P(\pi(A)) = 0 \implies \mu^X(\pi(A)) = 0 \iff \mu^X \circ \pi(A) = 0.$$

Thus, we have

$$E[X|\mathcal{G}] = \frac{d\mu^X_{|\mathcal{G}}}{dP_{\mathcal{G}}} = \frac{d(\mu^X \circ \pi)}{d(P \circ \pi)},$$

where the derivative are Radon-Nikodym derivatives of measures.

Definition 12.18 (Conditional probability)**.** Given $(\Omega, \mathcal{B}, P)$, with $\mathcal{G}$ a sub-σ-field of $\mathcal{B}$, define

$$P(A|\mathcal{G}) = E[\mathbb{1}_A|\mathcal{G}], \forall A \in \mathcal{B}.$$

Definition 12.19 (Conditioning on random variables)**.** Suppose $\{X_t, t \in T\}$ is a family of random variables defined on $(\Omega, \mathcal{B})$ and indexed by some index set T. Define

$$\mathcal{G} := \sigma(X_t, t \in T)$$

to be the σ-field generated by the process $\{X_t, t \in T\}$. Then define

$$E(X|X_t, t \in T) = E(X|\mathcal{G}).$$

(This definition saves us from having to make seperate defintions for $E[X|X_1], E(X|X_1, X_2)$, etc).

Example 12.20. Fix $B \in \mathcal{B}$, and let

$$\mathcal{G} = \{\emptyset, B, B^c, \Omega\}.$$

What is the probability of A given $\mathcal{G}$? Assume $P(B) > 0$. Then

$$\begin{aligned}
\int_B E[\mathbb{1}_A|\mathcal{G}] &= \int_B \mathbb{1}_A dP \\
&= P(A \cap B) \\
&= P(A|B)P(B) \\
&= \int_B P(A|B)\mathbb{1}_B dP \\
&= \int_B \left[P(A|B)\mathbb{1}_B + P(A|B^c)\mathbb{1}_{B^c}\right] dP.
\end{aligned}$$

Smilarly,

$$\int_{B^c} E[\mathbb{1}_A|\mathcal{G}] = \int_{B^c} \mathbb{1}_A dP = \int_{B^c} \left[P(A|B)\mathbb{1}_B + P(A|B^c)\mathbb{1}_{B^c}\right] dP.$$

So

$$\begin{aligned}
\int_\Omega E[\mathbb{1}_A|\mathcal{G}] &= \int_\Omega \mathbb{1}_A dP \\
&= \int_B \mathbb{1}_A dP + \int_{B^c} \mathbb{1}_A dP \\
&= \int_B \left[P(A|B)\mathbb{1}_B + P(A|B^c)\mathbb{1}_{B^c}\right] dP + \int_{B^c} \left[P(A|B)\mathbb{1}_B + P(A|B^c)\mathbb{1}_{B^c}\right] dP \\
&= \int_\Omega \left[P(A|B)\mathbb{1}_B + P(A|B^c)\mathbb{1}_{B^c}\right] dP.
\end{aligned}$$

Finally,

$$\int_\emptyset E[\mathbb{1}_A|\mathcal{G}] = \int_\emptyset \mathbb{1}_A dP = \int_\emptyset \left[P(A|B)\mathbb{1}_B + P(A|B^c)\mathbb{1}_{B^c}\right] dP.$$

Thus, by integral comparison lemma,

$$P(A|\mathcal{G}) = E[\mathbb{1}_A|\mathcal{G}] = P(A|B)\mathbb{1}_B + P(A|B^c)\mathbb{1}_{B^c}.$$

Example 12.21 (countable partition)**.** Suppose $X \in L_1(\Omega, \mathcal{B}, P)$. Let $\{A_n\}$ be a partition of Ω.

(We can define the discrete r.v. $Y = \sum_{n=1}^{\infty} c_i \mathbb{1}_{A_n}$, then $\{\{Y = c_i\}\}$ forms a partition of Ω.)

Define

$$\mathcal{G} = \sigma(A_n, n \geqslant 1)$$

so that

$$\mathcal{G} = \left\{ \bigsqcup_{i \in J} A_i, J \subseteq \{1, 2, \cdots\} \right\}.$$

For $X \in L_1(\Omega, \mathcal{B}, P)$, define

$$E(X|A_n) = \begin{cases} \frac{\int_{A_n} X dP}{P(A_n)}, & P(A_n) > 0 \\ 17, & P(A_n) = 0. \end{cases} = \begin{cases} \frac{E[X \mathbb{1}_{A_n}]}{P(A_n)}, & P(A_n) > 0 \\ 17, & P(A_n) = 0. \end{cases}$$

We claim

$$E[X|\mathcal{G}] \overset{\text{a.s.}}{=} \sum_{n=1}^{\infty} E[X|A_n] \mathbb{1}_{A_n} \in \mathcal{G}.$$

If this holds, let $X = \mathbb{1}_A, \forall A \in \mathcal{B}$, then

$$P(A|\mathcal{G}) \overset{\text{a.s.}}{=} \sum_{n=1}^{\infty} P(A|A_n) \mathbb{1}_{A_n}.$$

Proof. Let $A = \bigsqcup_{i \in J} A_i$ for some $J \subseteq \{1, 2, \cdots\}$.

$$\begin{aligned} \int_A \sum_{n=1}^{\infty} E[X|A_n] \mathbb{1}_{A_n} dP &= \sum_{n=1}^{\infty} \sum_{i \in J} \int_{A_i} E[X|A_n] \mathbb{1}_{A_n} dP \\ &= \sum_{n=1}^{\infty} \sum_{i \in J} E[X|\mathbb{1}_{A_n}] P(A_i A_n) \\ &= \sum_{i \in J} E[X|\mathbb{1}_{A_i}] P(A_i) \\ &= \sum_{i \in J} \left(\frac{E[X \mathbb{1}_{A_i}]}{P(A_i)} P(A_i) + 17 \cdot 0 \right) \\ &= \sum_{i \in J} E[X \mathbb{1}_{A_i}] \\ &= E\left[X \mathbb{1}_{\bigsqcup_{i \in J} A_i} \right] \\ &= E[X \mathbb{1}_A] \\ &= \int_A X dP \\ &= \int_A E[X|\mathcal{G}] dP. \end{aligned}$$

Thus,

$$E[X|\mathcal{G}] \overset{\text{a.s.}}{=} \sum_{n=1}^{\infty} E[X|A_n] \mathbb{1}_{A_n}. \qquad \square$$

Interpretation: Consider an experiment with sample space Ω. Condition on the information that "some event in $\mathcal{G}$ occurs". Imaine that at a future time you will be told which set A_n the outcome ω fallS in (but you will not be told ω). At time 0

$$\sum_{n=1}^{\infty} P(A|A_n) \mathbb{1}_{A_n}$$

is the best you can do to evaluate conditional probabilities.

Example 12.22 (Discrete case)**.** Let X be a discrete random variable with possible values $x_1, x_2, \cdots$. Then for $A \in \mathcal{B}$,

$$P(A|X) = P(A|\sigma(X)) = \sum_{i=1}^{\infty} P(A|X = x_i)\mathbb{1}_{\{X=x_i\}}.$$

Example 12.23 (Absolutely continuous case)**.** Let $\Omega = \mathbb{R}^2$ and suppose X and Y are r.v.'s whose joint distribution is absolutely continuous with density $f(x, y)$ so that for $A \in \mathcal{B}(\mathbb{R}^2)$,

$$P((X, Y) \in B) = \int\int_B f_{X,Y}(s, t)dsdt,$$

or

$$P(X \in B_x, Y \in B_y) = \int_{B_x}\int_{B_y} f_{X,Y}(s, t)dsdt,$$

where $f_{X,Y}$ is the joint pdf of X and Y. Find for $C \in \mathcal{B}(\mathbb{R})$,

$$P(Y \in C|\sigma(X)).$$

Proof. We use $\mathcal{G} = \sigma(X)$. Let

$$f_X(x) := \int_{\mathbb{R}} f_{X,Y}(x, t)dt$$

be the marginal density of X. Define

$$\phi_C(X) = \begin{cases} \frac{\int_C f_{X,Y}(x,t)dt}{f_X(x)}, & f_X(x) > 0, \\ 17, & f_X(x) = 0. \end{cases}$$

We claim that

$$P(Y \in C|\sigma(X)) = P(Y \in C|\mathcal{G}) = \phi_C(X).$$

First of all, note

$$\int_C f_{X,Y}(x, t)dt$$

is $\sigma(X)$-measurable and hence $\phi(X)$ is $\sigma(X)$-meaurable. Since

$$P(Y \in C|\mathcal{G}) = E[\mathbb{1}_{\{Y\in C\}}|\mathcal{G}],$$

and by the definition of conditional expection, for any $A \in \mathcal{B}(\mathbb{R})$,

$$\int_{X\in A} E[\mathbb{1}_{\{Y\in C\}}|\mathcal{G}]dP = \int_{\{X\in A\}} \mathbb{1}_{\{Y\in C\}}dP,$$

it suffices to show that for any $A \in \mathcal{B}(\mathbb{R})$,

$$\int_{\{X\in A\}} \phi_C(X)dP = \int_{\{X\in A\}} \mathbb{1}_{\{Y\in C\}}dP.$$

Since

$$E(g(X)) = \int_\Omega g(X(\omega))P(d\omega) = \int_{x\in\mathbb{R}} g(x)F(dx),$$

we have

$$\begin{aligned}
\int_{\{X\in A\}} \phi_C(X)dP &= \int_\Omega \mathbb{1}_A(X)\phi_C(X)dP \\
&= \int_{\mathbb{R}} \mathbb{1}_A(x)\phi_C(x)F(dx) \\
&\overset{TR}{=} \int_{\mathbb{R}} \mathbb{1}_A(x)\phi_C(x)f_X(x)dx \\
&= \int_{\mathbb{R}} \mathbb{1}_{\{A\cap f_X(x)>0\}}(x)\phi_C(x)f_X(x)dx + \int_{\mathbb{R}} \mathbb{1}_{\{A\cap f_X(x)=0\}}(x)\phi_C(x)f_X(x)dx \\
&= \int_{\mathbb{R}} \mathbb{1}_A(x)\frac{\int_C f_{X,Y}(x,t)dt}{f_X(x)} f_X(x)dx + 0 \\
&= \int_{(x\in)A}\int_C f_{X,Y}(x,t)dtdx \\
&\overset{def}{=} P((X,Y)\in A\times C) \\
&= P(X\in A, Y\in C) = \int_{\{X\in A\}} \mathbb{1}_{\{Y\in C\}}dP.
\end{aligned}$$

Since $\{X\in A\}\in\sigma(X)$ is arbitray, □

12.4 Properties of conditional expectation

Our probability space $(\Omega,\mathcal{B},P)$ and $\mathcal{G}$ is a sub-field of $\mathcal{B}$. $L_1(\Omega,\mathcal{B},P)$ is the set of $r.v.$' that are $\mathcal{B}$-measurable and satisfies $E[|X|]<\infty$.

(a) Linearity: If $X, Y\in L_1$ and $\alpha_1,\alpha_2\in\mathbb{R}$, then

$$E[\alpha_1 X+\alpha_2 Y|\mathcal{G}] = \alpha_1 E[X|\mathcal{G}]+\alpha_2 E[Y|\mathcal{G}].$$

Proof. For any $A\in\mathcal{G}$,

$$\begin{aligned}
\int_A E[\alpha_1 X+\alpha_2 Y|\mathcal{G}]dP &= \int_A (\alpha_1 X+\alpha_2 Y)dP \\
&= \alpha_1\int_A XdP + \alpha_2\int_A YdP \\
&= \alpha_1\int_A E[X|\mathcal{G}]dP + \alpha_2\int_A E[Y|\mathcal{G}]dP \\
&= \int_A \left(\alpha_1 E[X|\mathcal{G}]+\alpha_2 E[Y|G]\right)dP.
\end{aligned}$$

So by integral comparison lemma, $E[\alpha_1 X+\alpha_2 Y|\mathcal{G}] = \alpha_1 E[X|\mathcal{G}]+\alpha_2 E[Y|\mathcal{G}]$. □

(b) If $X\in L_1$ and X is $\mathcal{G}$-measurable, then

$$E[X|\mathcal{G}] = X,\ P\text{-a.s.}.$$

Proof. Note for any $A \in \mathcal{G}$,

$$\int_A E[X|\mathcal{G}]dP = \int_A XdP.$$

Since X is $\mathcal{G}$-measurable (for any $A \in \mathcal{B}(\mathbb{R})$, $X^{-1}(A) \in \mathcal{G}$), by the integral comparision lemma,

$$X = E[X|G] \ P\text{-a.s.}.$$

Since constant functions are always measurable on any σ-algebra (the only one on the trival σ-algebra), let $c \in \mathbb{R}$, $E[c|\mathcal{G}] = c$ P-a.s.. □

(c) Suppose $X \in L_1$ and $\mathcal{G} = \{\emptyset, \Omega\}$. Then

$$E[X|\mathcal{G}] = E[X]. \ P\text{-a.s.}$$

Proof. Here

$$\int_\Omega E[X|\mathcal{G}]dP = \int_\Omega XdP = E[X] = \int_\Omega E[X]dP,$$

$$\int_\emptyset E[X|\mathcal{G}]dP = \int_\emptyset XdP = 0 = \int_\emptyset E[X]dP.$$ □

(d) Monotonicity: Suppose $X, Y \in L_1$ and $X \leqslant Y$, then

$$E[X|\mathcal{G}] \leqslant E[Y|\mathcal{G}] \ P\text{-a.s.}$$

Proof. For any $A \in \mathcal{G}$,

$$\int_A E[X|\mathcal{G}]dP = \int_A XdP \leqslant \int_A YdP = \int_A E[Y|\mathcal{G}]dP.$$

Then by the integral comparison lemma,

$$E[X|\mathcal{G}] \leqslant E[Y|\mathcal{G}], P\text{-a.s.}.$$ □

(e) Modulus inequality: Suppose $X \in L_1$, then

$$|E[X|\mathcal{G}]| \leqslant E\left[|X||\mathcal{G}\right] P\text{-a.s.}.$$

Proof. Since $X \in L_1$, $X^- \in L_1$. So $E[X^-|\mathcal{G}] < \infty$, and then

$$E[X|\mathcal{G}] = E[X^+|\mathcal{G}] - E[X^-|\mathcal{G}].$$

Since $E[X^+|\mathcal{G}]$ and $E[X^-|\mathcal{G}]$ are nonnegative,

$$\begin{aligned} |E[X|\mathcal{G}]| &= \left|E[X^+|\mathcal{G}] - E[X^-|\mathcal{G}]\right| \\ &\leqslant E[X^+|\mathcal{G}] + E[X^-|\mathcal{G}] \\ &= E[X^+ + X^-|\mathcal{G}] \\ &= E\left[|X||\mathcal{G}\right] \ P\text{-a.s.} \end{aligned}$$ □

(f) Monotone convergence theorem: For nonnegative and monotone increasing sequence $\{X_n\}$, if $X := \lim_{n\to\infty} X_n \in L_1$, then

$$\lim_{n\to\infty} E[X_n|\mathcal{G}] = E\left[\lim_{n\to\infty} X_n|\mathcal{G}\right].$$

Proof. Note that by the property of monotonicity,

$$E[X_n|\mathcal{G}] \leqslant E[X_{n+1}|\mathcal{G}] \ P\text{-a.s.}.$$

Then

$$\lim_{n\to\infty} E[X_n|\mathcal{G}] \text{ exists } P\text{-a.s.}.$$

Note that if $P(A_n) = 1$ for any $n \in \mathbb{Z}^+$, then

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = 1.$$

Then for any $A \in \mathcal{G}$,

$$\begin{aligned}
\int_A \lim_{n\to\infty} E[X_n|\mathcal{G}]dP &= \lim_{n\to\infty} \int_A E[X_n|\mathcal{G}]dP \\
&= \lim_{n\to\infty} \int_A X_n dP \\
&= \int_A \lim_{n\to\infty} X_n dP \\
&= \int_A E\left[\lim_{n\to\infty} X_n|\mathcal{G}\right] dP, \ P\text{-a.s.}.
\end{aligned}$$

□

(g) For nonnegative sequence $\{X_n\}$ satisfies $X_n \in L_1, \forall n \in \mathbb{Z}$,

$$E\left[\lim_{n\to\infty} \inf X_n|\mathcal{G}\right] \leqslant \lim_{n\to\infty} \inf E[X_n|\mathcal{G}].$$

Proof. For any $A \in \mathcal{G}$,

$$\begin{aligned}
E\left[\lim_{n\to\infty} \inf X_n|\mathcal{G}\right] &= E\left[\lim_{n\to\infty} \inf_{k\geqslant n} X_k|\mathcal{G}\right] \\
&\overset{MCT}{=} \lim_{n\to\infty} E\left[\inf_{k\geqslant n} X_k|\mathcal{G}\right] \\
&\leqslant \lim_{n\to\infty} \inf_{k\geqslant n} E[X_k|\mathcal{G}] \ (\text{ bounded inside the expectation, then take inf}) \\
&= \lim_{n\to\infty} \inf E[X_n|\mathcal{G}].
\end{aligned}$$

□

(h) If $\{X_n\} \subseteq L_1$, $|X_n| \leqslant Z$ for $n \geqslant 1$ and $\mathbb{Z} \in L^1$, and $X_n \to X_\infty$ as $n \to \infty$. Then

$$\lim_{n\to\infty} E[X_n|\mathcal{G}] = E[X_\infty|\mathcal{G}].$$

(i) Product rule. Let X and Y be random varibales satisfying $X_1 \in L_1$, $YX \in L_1$,. If $Y \in \mathcal{G}$ ($\mathcal{G}$-measurable), then

$$E[XY|\mathcal{G}] \overset{a.s.}{=} YE(X|\mathcal{G}).$$

Proof. Case 1: Assume $Y = \mathbb{1}_B$ for some $B \in \mathcal{G}$. Then$\forall\, A \in \mathcal{G}$, since $A \cap B \in \mathcal{G}$,

$$\int_A E[X\mathbb{1}_B|\mathcal{G}]dP = \int_A X\mathbb{1}_B dP = \int_{A\cap B} XdP = \int_{A\cap B} E[X|\mathcal{G}]dP = \int_A \mathbb{1}_B E[X|\mathcal{G}]dP.$$

So

$$E[X\mathbb{1}_B|\mathcal{G}] \overset{a.s.}{=} \mathbb{1}_B E(X|\mathcal{G}).$$

Case 2: Assume $Y = \sum_{k=1}^n c_k \mathbb{1}_{B_k}$, where $c_1, \dots, c_k \in \mathbb{R}$, $B_1 \cdots B_k \in \mathcal{G}$. Then $Y \in \mathcal{G}$ and

$$E\left[X\sum_{k=1}^n c_k\mathbb{1}_{B_k}\middle|\,\mathcal{G}\right] = \sum_{k=1}^n E[c_k\mathbb{1}_{B_k}X|\mathcal{G}] = \sum_{k=1}^n c_k\mathbb{1}_{B_k}E[X|\mathcal{G}],$$

by the case 1.
Case 3: Assume $Y \geqslant 0$. Then there exists simple random variables $\{Y_n\}$ s.t. $\{Y_n\} \subseteq \mathcal{G}$ and $Y_n \uparrow Y$ as $n \to \infty$. However $XY \in L_1$, $(XY)^- = X^-Y \in L_1$. However $X \in L_1$, $X^- = X^- \in L_1$. So $E[(XY)^-|\mathcal{G}] < \infty$. Then by the case 2,

$$\begin{aligned}
E[XY|\mathcal{G}] &= E[(XY)^+|\mathcal{G}] - E[(XY)^-|\mathcal{G}] \\
&= E[X^+Y|\mathcal{G}] - E[X^-Y|\mathcal{G}] \\
&= \lim_{n\to\infty} E[X^+Y_n|\mathcal{G}] - \lim_{n\to\infty} E[X^-Y_n|\mathcal{G}] \\
&= \lim_{n\to\infty} Y_nE[X^+|\mathcal{G}] - \lim_{n\to\infty} Y_nE[X^-|\mathcal{G}] \\
&= YE[X^+|\mathcal{G}] - YE[X^-|\mathcal{G}] \\
&= YE[X|\mathcal{G}].
\end{aligned}$$

Last case: extend to $\mathcal{G}$-measurable. □

(j) The tower (smooth) property: Suppose $X \in L_1$, and $\mathcal{G}_1$ and $\mathcal{G}_2$ are both sub-σ-fields of $\mathcal{B}$ satisfying

$$\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{B}\ P\text{-a.s.}.$$

Then

$$E[E[X|\mathcal{G}_2]|\mathcal{G}_1] = E[X|\mathcal{G}_1],\ P\text{-a.s.},$$

$$E[E[X|\mathcal{G}_1]|\mathcal{G}_2] = E[X|\mathcal{G}_1],\ P\text{-a.s.}.$$

Proof. Since $E[X|\mathcal{G}_1] \in \mathcal{G}_2$, by the product rule,

$$E[E[X|\mathcal{G}_1]|\mathcal{G}_2] = E[X|\mathcal{G}_1]E[1|\mathcal{G}_2] = E[X|\mathcal{G}_1].$$

Note that for any $A \in \mathcal{G}_1 \subseteq \mathcal{G}_2$,

$$\begin{aligned}\int_A E[E[X|G_2]|\mathcal{G}_1]dP &\overset{def}{=} \int_A E[X|\mathcal{G}_2]dP \\ &= \int_A XdP \ (A \in \mathcal{G}_1 \subseteq \mathcal{G}_2) \\ &= \int_A E[X|\mathcal{G}_1]dP.\end{aligned}$$

Since $E[X|\mathcal{G}_1]$ is $\mathcal{G}_1$-measurable, by the integral comparison lemma,

$$E[E[X|\mathcal{G}_2]|\mathcal{G}_1] = E[X|\mathcal{G}_1], \ P\text{-a.s.}.$$ □

A special case: $\mathcal{G}_1 = \{\emptyset, \Omega\}$. Then by property (3), $E[X|\mathcal{G}_1] = E[X]$, and then

$$\begin{aligned}E[E[X|\mathcal{G}_2]] &= E[E[X|\mathcal{G}_2]|\mathcal{G}_1] \\ &= E[X|\mathcal{G}_1] \\ &= E[X].\end{aligned}$$

Then let $\mathcal{G}_2 = \sigma(Y)$, then

$$E[X] = E[E[X|Y]].$$

A common use of the tower property i the calculation for $\mathcal{G}$-measurable $Y \in L_1$,

$$E[XY] = E\left[E[XY|\mathcal{G}]\right] = E\left[YE[X|\mathcal{G}]\right].$$

(k) Projections: Suppose $X \in L_2(\mathcal{B})$. Then $E[X|\mathcal{G}]$ is the projection of X onto $L_2(\mathcal{G})$, a subspace of $L_2(\mathcal{B})$. The projection of X onto $L_2(\mathcal{G})$ is the unique element of $L_2(\mathcal{G})$ achieving

$$\inf_{Z \in L_2(\mathcal{G})} \|X - Z\|_2.$$

It is computed by solving the prediction equations for $Z \in L_2(\mathcal{G})$:

$$(Y, X - Z) = 0, \ \forall Y \in L_2(\mathcal{G}).$$

By trying a solution of $Z = E[X|\mathcal{G}]$, we get

$$\begin{aligned}\int Y(X - Z)dP &= E[Y(X - E[X|\mathcal{G}])] \\ &= E[YX] - E[YE[X|\mathcal{G}]] \\ &= E[YX] - E[E[YX|\mathcal{G}]] \\ &= E[YX] - E[YX] = 0.\end{aligned}$$

In time series analysis, $E[X|\mathcal{G}]$ is the best predictor of X in $L_2(\mathcal{G})$. It is not often used when $\mathcal{G} = \sigma(X_1, \ldots, X_n)$ and $X = X_{n+1}$ because of its lack of linearity and hence its computational difficulty.

(l) Conditioning and independence: If $X \in L_1$ and $X \perp\!\!\!\perp \mathcal{G}$, then

$$E[X|\mathcal{G}] = E[X].$$

Proof. Recall X is independent of $\mathcal{G}$ if $\sigma(X)$ is independent of $\mathcal{G}$ w.r.t the underlying measure P. Let $A \in \mathcal{G}$, since $E[X]$ is $\mathcal{G}$-measurable,

$$\int_A E[X|\mathcal{G}]dP = \int_A XdP = E[X\mathbb{1}_A] = E[X]P(A) = \int_A E[X]dP.$$

So $E[X|\mathcal{G}] = E[X]$. □

(m) Conditional Jensen's inequality: Let ϕ be a convex function, $X \in L_1$ and $\phi(X) \in L_1$. Then almost surely

$$\phi[E(X|\mathcal{G})] \leqslant E[\phi(X)|\mathcal{G}].$$

(n) Conditional expectation is continuous on L_p: Assume $X_n \xrightarrow{L_p} X$ as $n \to \infty$ for $p \geqslant 1$. Then

$$E[X_n|\mathcal{G}] \xrightarrow{L_p} E[X|\mathcal{G}].$$

Proof. When $p \geqslant 1$, $\phi(x) = |x|^p$ is a convex function. Then

$$\begin{aligned} E\left[|E[X_n|\mathcal{G}] - E[X|\mathcal{G}]|^p\right] &= E\left[|E[X_n - X|\mathcal{G}]|^p\right] \\ &\leqslant E\left[E|X_n - X|^p|\mathcal{G}\right] \\ &= E\left[|X_n - X|^p\right] \\ &\to 0. \end{aligned}$$

□

(o) Conditional expectation is L_p norm reducing, i.e.,

$$\|E[X|\mathcal{B}]\|_p \leqslant \|X\|_p.$$

Proof. Similarly, by Jensen's inequality. □

12.5 Martingale

Let $\{X_n\}_{n \in \mathbb{Z}^{\geqslant 0}} \subseteq L_1(\Omega, \mathcal{B}, P)$ and let $\{\mathcal{B}_n\}_{n \in \mathbb{Z}^{\geqslant 0}}$ be sub-σ-field of $\mathcal{B}$.
Consider the following statements.

(a)

$$\mathcal{B}_n \subseteq \mathcal{B}_{n+1}, \forall n \in \mathbb{Z}^{\geqslant 0}.$$

(b)

$$X_n \in \mathcal{B}_n, \forall n \in \mathbb{Z}^{\geqslant 0}.$$

(c) (1) For any $n \in \mathbb{Z}^{\geqslant 0}$ and any $m \in \mathbb{Z}^+$,

$$E[X_{n+m}|\mathcal{B}_n] = X_n, \ P\text{-a.s.}.$$

(2) For any $n \in \mathbb{Z}^{\geqslant 0}$ and any $m \in \mathbb{Z}^+$,

$$E[X_{n+m}|\mathcal{B}_n] \geqslant X_n,\ P\text{-a.s.}.$$

(3) For any $n \in \mathbb{Z}^{\geqslant 0}$ and any $m \in \mathbb{Z}^+$,

$$E[X_{n+m}|\mathcal{B}_n] \leqslant X_n,\ P\text{-a.s.}.$$

Then

(a) $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is called a martingale if it satisfies (i), (ii) and (iii)(a).

(b) $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is called a submartingale if it satisfies (i), (ii) and (iii)(b).

(c) $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is called a supmartingale if it satisfies (i), (ii) and (iii)(c).

Remark. (a) $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is called a martingale if and only if it is both submartingle and supmartingale.

(b) Postulate (iii) could be replaced by

$$E[X_{n+1}|\mathcal{B}_n] = X_n\ P\text{-a.s.}\forall n \in \mathbb{Z}^{\geqslant 0}.$$

Proof. For any $m \geqslant 2$,

$$\begin{aligned} E[X_{n+m}|\mathcal{B}_n] &= E[E[X_{n+m}|\mathcal{B}_{n+m-1}]|\mathcal{B}_n] \\ &= E[X_{n+m-1}|\mathcal{B}_n] \\ &= \vdots \\ &= E[X_{n+1}|\mathcal{B}_n] \\ &= X_n. \end{aligned}$$ □

(c) For any $n \geqslant 0$, let $\mathcal{C}_n = \sigma(X_0, \dots, X_n)$. If $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is a martingale, then $\{(X_n, \mathcal{C}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is a martingale.

Proof. (1)

$$\mathcal{C}_n = \sigma(X_0, \dots, X_n) \subseteq \sigma(X_0, \dots, X_{n+1}) = \mathcal{C}_{n+1}.$$

(2) Since $X_k \in \mathcal{B}_k \subseteq \mathcal{B}_n$ for any $0 \leqslant k \leqslant n$, we have $\mathcal{C}_n \in \mathcal{B}_n$.

(3) For any $n \in \mathbb{Z}^{\geqslant 0}, m \in \mathbb{Z}^+$, since $X_n \in \mathcal{C}_n$,

$$E[X_{n+m}|\mathcal{C}_n] = E[E[X_{n+m}|\mathcal{B}_n]|\mathcal{C}_n] = E[X_n|\mathcal{C}_n] = X_n,\ P\text{-a.s.}.$$ □

Definition 12.24. A sequence $\{\mathcal{B}_n\}_{n\in\mathbb{Z}^{\geqslant 0}}$ of σ-fields satisfying $\mathcal{B}_n \subseteq \mathcal{B}_{n+1}$ for any $n \in \mathbb{Z}^{\geqslant 0}$ is called a *filtration*.

Definition 12.25 (Martingale difference)**.** For $\{(d_j, \mathcal{B}_j)\}_{j\in\mathbb{Z}^{\geqslant 0}}$, where $\{d_j\}_{j\in\mathbb{Z}^{\geqslant 0}} \subseteq L_1(\Omega, \mathcal{B}, P)$ and $\{\mathcal{B}_j\}_{j\in\mathbb{Z}^{\geqslant 0}}$ are sub-σ-field of $\mathcal{B}$. Consider the following statements,

(a) $\mathcal{B}_j \subseteq \mathcal{B}_{j+1}$ for any $j \in \mathbb{Z}^{\geqslant 0}$.

(b) $d_j \in \mathcal{B}_j$ for any $j \in \mathbb{Z}^{\geqslant 0}$.

(c) (1)

$$E[d_{j+1}|\mathcal{B}_j] = 0,\ P\text{-a.s.}\forall j \in \mathbb{Z}^{\geqslant 0}.$$

(2)

$$E[d_{j+1}|\mathcal{B}_j] \geqslant 0,\ P\text{-a.s.}\forall j \in \mathbb{Z}^{\geqslant 0}.$$

(3)

$$E[d_{j+1}|\mathcal{B}_j] \leqslant 0,\ P\text{-a.s.}\forall j \in \mathbb{Z}^{\geqslant 0}.$$

Then

(a) $\{(d_j, \mathcal{B}_j)\}_{j\in\mathbb{Z}^{\geqslant 0}}$ is called a *martingale difference sequence* or a *fair sequence* if it satisfies (i), (ii) and (iii)(a).

(b) $\{(d_j, \mathcal{B}_j)\}_{j\in\mathbb{Z}^{\geqslant 0}}$ is called a sub-fair sequence if it satisfies (i), (ii) and (iii)(b).

(c) $\{(d_j, \mathcal{B}_j)\}_{j\in\mathbb{Z}^{\geqslant 0}}$ is called a sup-fair sequence if it satisfies (i), (ii) and (iii)(c).

Proposition 12.26. Facts about martingale differences:

(a) If $\{(d_j, \mathcal{B}_j)\}_{j\in\mathbb{Z}^{\geqslant 0}}$ is (sub,super) fair sequence, then

$$\left\{\left(X_n := \sum_{j=0}^{n} d_j,\ \mathcal{B}_n\right)\right\}_{n\in\mathbb{Z}^{\geqslant 0}}$$

is a (sub, sup) fair martingale.

Proof. Proof of the fair sequence cases. Suppose $\{(d_j, \mathcal{B}_j)\}_{j\in\mathbb{Z}^{\geqslant 0}}$ is fair. Then $d_j \in L_1(\Omega, \mathcal{B}, P)$ for any $j \in \mathbb{Z}^{\geqslant 0}$ and since L_1 is closed under finite sum,

$$X_n \in L_1(\Omega, \mathcal{B}, P), \forall n \in \mathbb{Z}^{\geqslant 0}.$$

Furthermore,

(1) Clearly, $\mathcal{B}_n \subseteq \mathcal{B}_{n+1}$ for any $n \in \mathbb{Z}^{\geqslant 0}$.

(2) Since $d_j \in \mathcal{B}_j \subseteq \mathcal{B}_n$ for any $0 \leqslant j \leqslant n$ and $j \in \mathbb{Z}$, $X_n = \sum_{j=0}^{n} d_j \in \mathcal{B}_n$.

(3) For any $n \in \mathbb{Z}^{\geqslant 0}$,

$$\begin{aligned} E[X_{n+1}|\mathcal{B}_n] &= E[X_n + d_{n+1}|\mathcal{B}_n] \\ &= E[X_n|\mathcal{B}_n] + E[d_{n+1}|\mathcal{B}_n] \\ &= X_n + 0 \\ &= X_n,\ P\text{-a.s.}. \end{aligned}$$

□

(b) Suppose $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is a (sub, sup) martingale. Define

$$d_0 := X_0 - E[X_0],$$

$$d_j := X_j - X_{j-1}, \forall j \in \mathbb{Z}^+.$$

Then $\{(d_j, \mathcal{B}_j)\}_{j\in\mathbb{Z}^{\geqslant 0}}$ is a (sub, sup) fair sequence.

Proof. Proof of the martingale case. Suppose $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is a martingale. Then $X_n \in L_1(\Omega, \mathcal{B}, P)$ for any $n \in \mathbb{Z}^{\geqslant 0}$ and since L_1 is closed linear combinations,

$$d_j \in L_1(\Omega, \mathcal{B}, P), \forall j \in \mathbb{Z}^{\geqslant 0}.$$

(1) Clearly, $\mathcal{B}_j \subseteq \mathcal{B}_{j+1}, \forall n \in \mathbb{Z}^{\geqslant 0}$.

(2) Clearly, $d_0 = X_0 - E[X_0] \in \mathcal{B}_0$. Since $X_j \in \mathcal{B}_j$ and $X_{j-1} \in \mathcal{B}_{j-1} \subseteq \mathcal{B}_j$ for any $j \in \mathbb{Z}^+$,

$$d_j = X_j - X_{j-1} \in \mathcal{B}_j, \forall j \in \mathbb{Z}^+.$$

(3) For any $j \in \mathbb{Z}^{\geqslant 0}$,

$$E[d_{j+1}|\mathcal{B}_j] = E[X_{j+1} - X_j|\mathcal{B}_j] = E[X_{j+1}|\mathcal{B}_j] - E[X_j|\mathcal{B}_j] = X_j - X_j = 0, \ P\text{-a.s.}. \qquad \square$$

(c) (Orthogonaility of martingale difference.) Suppose $\{(d_j, \mathcal{B}_j)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is a fair sequence and $\{d_j\}_{n\in\mathbb{Z}^{\geqslant 0}} \subseteq L_2(\Omega, \mathcal{B}, P)$. Then

$$E[d_i d_j] = 0, \forall i \neq j.$$

Proof. By holder's inequality,

$$E\left[|d_i d_j|\right] \leqslant \sqrt{E[d_i^2]}\sqrt{E[d_j^2]} < \infty.$$

So

$$\{d_i d_j\}_{i\neq j, i, j\in\mathbb{Z}^{\geqslant 0}} \subseteq L_1(\Omega, \mathcal{B}, P).$$

When $i \neq j$, without loss of generality, assume $j > i$, since

$$E[d_j|\mathcal{B}_i] = E[E[d_j|\mathcal{B}_{j-1}]|\mathcal{B}_i] = E[0|\mathcal{B}_i] = 0,$$

we have

$$E[d_i d_j] = E[E[d_i d_j|\mathcal{B}_i]] = E[d_i E[d_j|\mathcal{B}_i]] = E[d_i \cdot 0] = 0, \ P\text{-a.s.}. \qquad \square$$

12.5.1 Examples of Martingales

Example 12.27. Suppose $X \in L_1(\Omega, \mathcal{B}, P)$ and $\{\mathcal{B}_n\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is an increasing family of sub-σ-field of $\mathcal{B}$. Define for any $n \in \mathbb{Z}^{\geqslant 0}$.

$$X_n := E[X|\mathcal{B}_n].$$

Then $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is a martingle.

Proof. Since $X \in L_1$ and $X_n = E[X|\mathcal{B}_n]$, $X_n \in L_1(\Omega, \mathcal{B}, P)$ for any $n \in \mathbb{Z}^{\geqslant 0}$.

(a) Clearly, $\mathcal{B}_n \subseteq \mathcal{B}_{n+1}, \forall n \geqslant 0$.

(b) Since $X \in L_1$ and $X_n = E[X|\mathcal{B}_n]$, $X_n \in \mathcal{B}_n$ for any $n \in \mathbb{Z}^{\geqslant 0}$.

(c) For any $n \in \mathbb{Z}^{\geqslant 0}$,

$$E[X_{n+1}|\mathcal{B}_n] = E[E[X|\mathcal{B}_{n+1}]|\mathcal{B}_n] = E[X|\mathcal{B}_n] = X_n, \ P\text{-a.s.}. \qquad \square$$

Example 12.28 (Martingales's and sums of independent random variables.)**.** Suppose $\{Z_n\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is an independent sequence satisfying (random walk)

$$Z_n \in L_1(\Omega, \mathcal{B}, P), \forall n \in \mathbb{Z}^{\geqslant 0}.$$

$$E[Z_n] = 0, \forall n \in \mathbb{Z}^{\geqslant 0}.$$

Set

$$X_n = \sum_{i=0}^{n} Z_i, \forall n \in \mathbb{Z}^{\geqslant 0},$$

$$\mathcal{B}_n = \sigma(Z_0, \ldots, Z_n).$$

Then $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is a martingale.

Proof. It suffices to show $\{Z_n\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is a fair sequence.

(a) $\mathcal{B}_n = \sigma(Z_0, \ldots, Z_n) \subseteq \sigma(Z_0, \ldots, Z_n) = \mathcal{B}_{n+1}$ for any $n \in \mathbb{Z}^{\geqslant 0}$.

(b) $Z_n \in L_1$ for any $n \in \mathbb{Z}^{\geqslant 0}$.

(c) Since $\{Z_n\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is an independent sequence for any $n \in \mathbb{Z}^{\geqslant 0}$,

$$E[Z_{n+1}|\mathcal{B}_n] = E[Z_{n+1}] = 0, \ P\text{-a.s.}. \qquad \square$$

Example 12.29. Suppose $\{X_n\}_{n\in\mathbb{Z}^{\geqslant 0}}$ have a countable space $E = \{0, 1, \cdots\}$ and transition matrix

$$P = [P_{ij}]_{i,j\in E}.$$

For any $n \in \mathbb{Z}^{\geqslant 0}$, define

$$\mathcal{B}_n = \sigma(X_0, \ldots, X_n).$$

By Markov property, For any $n \in \mathbb{Z}^{\geqslant 0}$ and any $j \in E$,

$$P(X_{n+1} = j|\mathcal{B}_n) = P\left(X_{n+1} = j|\sigma(X_n)\right) = P(X_{n+1} = j|X_n), \ P\text{-a.s.}.$$

By example 12.22,

$$P\left(X_{n+1} = j|X_n\right) = \sum_{i\in E} P(X_{n+1} = j|X_n = i)\mathbb{1}_{\{X_n=i\}}.$$

On the set $\{X_n = i, \ldots, X_0 = x_0\}$,

$$P(X_{n+1} = j|\mathcal{B}_n) = P(X_{n+1} = j|X_n = i) = P_{ij}.$$

Note that

$$X_n : (\Omega, \mathcal{B}_n) \to (E, \mathcal{P}(E)), \forall n \in \mathbb{Z}^+.$$

Suppose there exists an eigenvalue λ and a corresponding eigenvector

$$f : (E, \mathcal{P}(E)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})).$$

satisfying $Pf = \lambda f$, and $f = (f(0), f(1), \cdots)$. In component form, this is

$$\sum_{j \in E} P_{ij} f(j) = \lambda f(i).$$

In terms of expectation,

$$E[f(X_{n+1}|X_n = i)] = \lambda f(i).$$

or

$$E[f(X_{n+1}|X_n)] = E[f(X_{n+1}|\sigma(X_n)] = E[f(X_{n+1}|\mathcal{B}_n)] = \lambda f(X_n),$$

by (inverse) Markov property. Assume $\lambda \neq 0$, define

$$Z_n = \frac{f(X_n)}{\lambda^n}, \forall n \in \mathbb{Z}^{\geqslant 0}.$$

Claim $\{(Z_n, \mathcal{B}_n)\}_{n \in \mathbb{Z}^{\geqslant 0}}$ is a martingale. Clearly, $f \in \mathcal{P}(E)$. Then $f(X_n) \in \mathcal{B}_n$ for any $n \in \mathbb{Z}^{\geqslant 0}$. So $Z_n \in \mathcal{B}_n \subseteq \mathcal{B}, \forall n \in \mathbb{Z}^{\geqslant 0}$. Also, assume

$$E\left[|f(X_0)|\right] = \sum_{i \in E} |f(i)| P(X_0 = i) < \infty,$$

meaning $f(X_0) \in L_1$, where X_0 is the intial distribution. Then

$$\begin{aligned} E\left[|f(X_n)|\right] &= \sum_{i \in E} |f(i)| P(X_n = i) \\ &= \sum_{i \in E} |f(i)| \sum_{j \in E} P(X_0 = j) P_{ji}^{(n)} \\ &= \sum_{i \in E} |f(i)| P(X_0 = i) \ (\textit{stationary?}) \\ &< \infty. \end{aligned}$$

So $Z_n = \frac{f(X_n)}{\lambda^n} \in L_1$ for any $n \in \mathbb{Z}^{\geqslant 0}$.

(a) $\mathcal{B}_n \subseteq \mathcal{B}_{n+1}$ for any $n \in \mathbb{Z}^{\geqslant 0}$.

(b) We already showed $Z_n \in \mathcal{B}_n$ for any $n \in \mathbb{Z}^{\geqslant 0}$.

(c) For any $n \in \mathbb{Z}^{\geqslant 0}$,

$$E[Z_{n+1}|\mathcal{B}_n] = \frac{E[f(X_{n+1})|\mathcal{B}_n]}{\lambda^{n+1}} = \frac{\lambda f(X_n)}{\lambda^{n+1}} = \frac{f(X_n)}{\lambda^n} = Z_n, \ P\text{-a.s.}.$$

So $\{(Z_n, \mathcal{B}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is a martingale.

Example 12.30 (Simple branching process)**.** Let X_n be the number of organisms of the n-th generation. Each organisms reproduces asexually and let

$$P_i = P(\text{an organism has } i \text{ offspring}), \forall i \in \mathbb{Z}^{\geqslant 0}.$$

Let $Z_{n,k}$ be the number of offspring produced by the k-th organism in the n-th generation. Assume $\{Z_{n,k}\}_{n\in\mathbb{Z}^{\geqslant 0}, k\in\mathbb{Z}^+}$ are iid. Then

$$X_{n+1} = \begin{cases} \sum_{j=1}^{X_n} Z_{n,j}, & X_n > 0, \\ 0, & X_n = 0. \end{cases}$$

So $\{X_n\}_{n\geqslant 0}$ is a DTMC and

$$P_{ij} = \begin{cases} \delta_{0j}, & i = 0 \\ P_j^{*i}, & i \geqslant 1, \end{cases}$$

where P_j^{*i} is the j-th component of the i-fold convolution of the sequence $\{p_n\}$, i.e.,

$$p * p[i] = \sum_{m=0}^{\infty} p_m p_{i-m} = \sum_{m=0}^{\infty} p_{i-m} p_m.$$

Let

$$m = \sum_{k=1}^{\infty} k p_k$$

be the mean number of offspring per organism. Note that for any $i \in \mathbb{Z}^+$,

$$mi = \sum_{j=0}^{\infty} P_{ij} j = \sum_{j=0}^{\infty} P_j^{*i} j,$$

while for $i = 0$,

$$\sum_{j=0}^{\infty} p_{ij} j = P_{00} \cdot 0 = 0 = mi.$$

With $f(j) = j$, we have $Pf = mf$.

Or note when $X_n > 0$,

$$\begin{aligned}
E[X_{n+1}] &= E\left[\sum_{k=1}^{X_n} Z_{n,k}\right] \\
&= E\left[\sum_{k=0}^{\infty} Z_{n,k}\mathbb{1}_{\{X_n \geqslant k\}}\right] \\
&= \sum_{k=0}^{\infty} E[Z_{n,k}\mathbb{1}_{\{X_n \geqslant k\}}.] \\
&= \sum_{k=0}^{\infty} E[Z_{n,k}]P(X_n \geqslant k) \text{ (naturally independent)} \\
&= \sum_{k=0}^{\infty} mP(X_n \geqslant k) \\
&= mE[X_n],
\end{aligned}$$

while for $X_n = 0$, $E[X_{n+1}] = 0 = mE[X_n]$. Thus, the process $\{(X_n/m^n, \sigma(X_0, \dots, X_n), n \geqslant 0\}$ is a martingale.

Example 12.31. (Likelihood ratios). Suppose $\{Y_n, n \geqslant 0\}$ are iid random variables and suppose the true density of Y_1 is f_0. (The word "density" can be understood with respect to some fixed reference meaure μ.) Let f_1 be some other probability density. For simplicity suppose $f_0(y) > 0$, for all y. Then for $n \geqslant 0$,

$$X_n = \frac{\prod_{i=0}^n f_1(Y_i)}{\prod_{i=0}^n f_0(Y_i)}$$

is a martingale.

Proof. We check the condition (iii).

$$\begin{aligned}
E(X_{n+1}|Y_0, \dots, Y_n) &= E\left(\frac{\prod_{i=0}^n f_1(Y_i)}{\prod_{i=0}^n f_0(Y_i)}\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}\middle| Y_0, \dots, Y_n\right) \\
&= X_n E\left(\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}\middle| Y_0, \dots, Y_n\right) \ P\text{-a.s.}
\end{aligned}$$

By independence this becomes

$$\begin{aligned}
E(X_{n+1}|Y_0, \dots, Y_n) &= X_n E\left(\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}\right) \ P\text{-a.s.} \\
&= X_n \int \frac{f_1(y)}{f_0(y)} f_0(y)\mu(dy) \\
&= X_n \int f_1 d\mu \\
&= X_n \cdot 1 \\
&= X_n \ P\text{-a.s.}
\end{aligned}$$

since f_1 is a density. ∎

Definition 12.32. A sequence of random variables $\{U_j\}_{j\geqslant 0}$ is said to be predictable w.r.t a filtration $\{\mathcal{B}_n\}$ if

(a) $U_0 \in \mathcal{B}_0$,

(b) $U_j \in \mathcal{B}_{j-1}$ for any $j \in \mathbb{Z}^+$.

Definition 12.33. A sequence of random variables $\{U_j\}_{j\geqslant 0}$ is *adapt* to the filtration $\{\mathcal{B}_j\}_{j\geqslant 0}$ if

$$U_j \in \mathcal{B}_j, \forall j \in \mathbb{Z}^+.$$

Remark. If $\{U_j\}_{j\in\mathbb{Z}^{\geqslant 0}}$ is predictable w.r.t. $\{\mathcal{B}_j\}_{j\in\mathbb{Z}^{\geqslant 0}}$, it is adapt to $\{\mathcal{B}_j\}_{j\geqslant 0}$.

Example 12.34 (Discrete stochastic integration)**.** Suppose $\{(d_j, \mathcal{B}_j)\}_{j\in\mathbb{Z}^{\geqslant 0}}$ is a fair sequence. Let $\{(U_j)\}_{j\in\mathbb{Z}^{\geqslant 0}}$ be predictable w.r.t. $\{\mathcal{B}_j\}_{j\in\mathbb{Z}^{\geqslant 0}}$ and assume that

$$\{U_j\}_{j\in\mathbb{Z}^{\geqslant 0}} \subseteq L_\infty(\Omega, \mathcal{B}, P).$$

Show that $\{(U_j d_j, \mathcal{B}_j)\}_{j\in\mathbb{Z}^{\geqslant 0}}$ is still a fair sequence.

Proof. Since $d_j \in \mathcal{B}_j \subseteq \mathcal{B}$ and $U_j \in \mathcal{B}_{j-1} \subseteq \mathcal{B}$, $U_j d_j \in \mathcal{B}$. It is easy to find (or use Holder's inequality) $U_j d_j \in L_1(\Omega, \mathcal{B}, P)$.

(a) Clearly, $\mathcal{B}_n \subseteq \mathcal{B}_{n+1}$.

(b) Since $d_j \in \mathcal{B}_j$ and $U_j \in \mathcal{B}_{j-1} \subseteq \mathcal{B}_j$, $U_j d_j \in \mathcal{B}_j$.

(c) For any $j \in \mathbb{Z}^{\geqslant 0}$, since $U_{j+1} \subseteq \mathcal{B}_j$,

$$E[U_{j+1}d_{j+1}|\mathcal{B}_j] = U_{j+1}E[d_{j+1}|\mathcal{B}_j] = 0 \ P\text{-a.s.}.$$

So $\{(U_j d_j, \mathcal{B}_j)\}_{j\in\mathbb{Z}^{\geqslant 0}}$ is a fair sequence. Define

$$X_n := \sum_{m=0}^{n} U_m d_m, \forall n \in \mathbb{Z}^{\geqslant 0}.$$

Then $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is martingale. □

In gambling models, d_j might be ± 1 and U_j is how much you gamble so that U_j is a strategy based on previous gambles. In investment model, d_j might be the change in price of a risky asset and U_j is the number of shares of the asset held by the investor. In stochastic integration, ths d_j's are increment of Brownian motion. Refer to

(a) Stochastic Differential Equations by Okendal

(b) Stochastic Integration and Differential Equations by Protter

(c) Introduction to Stochastic Integration by K.L. Chung.

Lemma 12.35. Suppose $\{\mathcal{B}_n\}_{n\geqslant 0}$ is a filtration and let

$$\{M_n\}_{n\geqslant 0} \subseteq L_1(\Omega, \mathcal{B}, P)$$

be adapt w.r.t $\{\mathcal{B}_n\}_{n\in\mathbb{Z}^{\geqslant 0}}$. Define

$$\begin{aligned} d_0 &= M_0, \\ d_n &= M_n - M_{n-1}, \forall n \in \mathbb{Z}^+. \end{aligned}$$

Then $\{(M_n, \mathcal{B}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is a martingale if and only if for every bounded predictable sequence $\{U_n\}_{n\in\mathbb{Z}^{\geqslant 0}}$, we have

$$E\left(\sum_{n=1}^{N} U_n d_n\right) = 0, \forall N \in \mathbb{Z}^+.$$

Proof. Suppose $\{(M_n, \mathcal{B}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is a martingale and $\{U_j\}_{j\in\mathbb{Z}^{\geqslant 0}}$ is bounded predictable sequence, where "bounded" means

$$\sup_{n\in\mathbb{Z}^{\geqslant 0}} \sup_{\omega\in\Omega} |U_n(\omega)| \leqslant A < \infty.$$

Then for any $N \in \mathbb{Z}^+$,

$$E\left[\sum_{n=1}^{N} U_n d_n\right] = \sum_{n=1}^{N} E[U_n d_n] = \sum_{n=1}^{N} E\left[E[U_n d_n|\mathcal{B}_{n-1}]\right] = \sum_{n=1}^{N} E\left[U_n E[d_n|\mathcal{B}_{n-1}]\right] = 0.$$

Assume the other direction.

(a) $\{\mathcal{B}_n\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is a filtration.

(b) $\{M_n\}_{n\in\mathbb{Z}^{\geqslant 0}} \subseteq L_1(\Omega, \mathcal{B}, P)$ is adapt w.r.t. $\{\mathcal{B}_n\}_{n\in\mathbb{Z}^{\geqslant 0}}$.

(c) Fix $j \in \mathbb{Z}^{\geqslant 0}$ and let $A \in \mathcal{B}_j$.
Define for any $n \in \mathbb{Z}^{\geqslant 0}$,

$$U_n = \begin{cases} 0, & n \neq j+1 \\ \mathbb{1}_A, & n = j+1. \end{cases}$$

Then $\{U_n\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is bounded and predictable w.r.t. $\{\mathcal{B}_n\}_{n\in\mathbb{Z}^{\geqslant 0}}$ since

(1) If $n \neq j+1$, $U_n = 0 \in \mathcal{B}_{n-1}$.
(2) If $n = j+1$, $U_{j+1} = \mathbb{1}_A \in \mathcal{B}_j$ since $A \in \mathcal{B}_j$.

Then for any $N \geqslant j+1$ and $N \in \mathbb{Z}^+$, by assummption,

$$0 = E\left(\sum_{n=1}^{N} U_n d_n\right) = E[U_{j+1}(M_{j+1} - M_j)] = E[\mathbb{1}_A(M_{j+1} - M_j)] = E[\mathbb{1}_A M_{j+1}] - E[\mathbb{1}_A M_j].$$

So $E[\mathbb{1}_A M_{j+1}] = E[\mathbb{1}_A M_j]$. Since $A \in \mathcal{B}_j$ is arbitrary,

$$E[\mathbb{1}_A M_j] = E[\mathbb{1}_A M_{j+1}] = E\left[\mathbb{1}_{A_j} E[M_{j+1}|\mathcal{B}_j]\right] \ P\text{-a.s.}.$$

By the integral comparision lemma,

$$E[M_{j+1}|\mathcal{B}_j] = M_j \ P\text{-a.s.}.$$

Or use when for any $N \geqslant j+1$ and $N \in \mathbb{Z}^+$, by assumption,

$$0 = E\left(\sum_{n=1}^{N} U_n d_n\right) = E[U_{j+1}d_{j+1}] = E[\mathbb{1}_{A_j} d_{j+1}]$$

so that

$$0 = E[\mathbb{1}_{A_j} d_{j+1}] = E\left[\mathbb{1}_{A_j} E[d_{j+1}|\mathcal{B}_j]\right].$$

By the integral comparision lemma, $E[d_{j+1}|\mathcal{B}_j] = 0$ P-a.s..

So $\{(M_n, \mathcal{B}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is a martingale. ☐

Definition 12.36. A collection of random variables that is predictable w.r.t. the filtration $\{\mathcal{B}_n\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is said to be *increasing* if

$$0 = A_0 \leqslant A_1 \leqslant A_2 \leqslant \cdots \ P\text{-a.s.}.$$

Theorem 12.37 (Doob decomposition)**.** *Any submartingale $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ can be written in a unique way as the sum of a martingale $\{(M_n, \mathcal{B}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ and an increasing process $\{A_n\}_{n\in\mathbb{Z}^{\geqslant 0}}$. That is $X_n = M_n + A_n,\ n \in \mathbb{Z}^{\geqslant 0}$.*

Proof. Define

$$M_n = X_0 + \sum_{k=1}^{n} \left[X_k - E[X_k|\mathcal{B}_{k-1}]\right].$$

(a) $\mathcal{B}_n \subseteq \mathcal{B}_{n+1}, \forall n \in \mathbb{Z}^{\geqslant 0}$.

(b) $M_0 = X_0 \in L_1$ and $M_0 \in \mathcal{B}_0$. Let $n \in \mathbb{Z}^+$. Since conditional expectation

$$E[X_k|\mathcal{B}_{k-1}] \in \mathcal{B}_{k-1} \subseteq \mathcal{B}_n \subseteq \mathcal{B}, \forall k \in [n],$$

and conditional expectations are always integrable,

$$E[X_k|\mathcal{B}_{k-1}] \in L_1(\Omega, \mathcal{B}, P), \forall k \in [n].$$

So $M_n \in L_1$. Similarly, $M_n \in \mathcal{B}_n$.

(c) For any $n \in \mathbb{Z}^{\geqslant 0}$,

$$\begin{aligned} E[M_{n+1}|\mathcal{B}_n] &= E\left[M_n + X_{n+1} - E[X_{n+1}|\mathcal{B}_n] | \mathcal{B}_n\right] \\ &= E[M_n|\mathcal{B}_n] + E[X_{n+1}|\mathcal{B}_n] - E[X_{n+1}|\mathcal{B}_n] \\ &= M_n. \end{aligned}$$

So $\{(M_n, \mathcal{B}_n)\}$ is a martingale. Define

$$A_n = X_n - M_n, \forall n \in \mathbb{Z}^{\geqslant 0}.$$

Then

$$A_0 = X_0 - M_0 = 0 \in \text{any } \sigma \text{ field.}$$

$$\begin{aligned}
A_n &= X_n - X_0 - \sum_{k=1}^{n} (X_k - E[X_k|\mathcal{B}_{k-1}]) \\
&= E[X_n|\mathcal{B}_{n-1}] - X_0 - \sum_{k=1}^{n-1} (X_k - E[X_k|\mathcal{B}_{k-1}]) \\
&= E[X_n|\mathcal{B}_{n-1}] - M_{n-1} \\
&= E[X_n|\mathcal{B}_{n-1}] - X_{n-1} + A_{n-1} \\
&\geqslant A_{n-1} \ P\text{-a.s.},
\end{aligned}$$

since $\{(X_n, \mathcal{B}_n)\}_{n\geqslant 0}$ is a submartingale. Besides,

$$A_n = E[X_n|\mathcal{B}_{n-1}] - X_{n-1} + A_{n-1} \in \mathcal{B}_{n-1}.$$

So $\{A_n\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is increasing. Next, we prove the uniqueness. Suppose there is another decomposition

$$X_n = M_n' + A_n',$$

where $\{(M_n', \mathcal{B}_n)\}_{n\in\mathbb{Z}^{\geqslant 0}}$ is a martingale and $\{A_n'\}$ is an increasing process. Then$\forall n \geqslant 0$,

$$\begin{aligned}
A_{n+1}' - A_n' &= X_{n+1} - M_{n+1}' - (X_n - M_n') \\
&= X_{n+1} - X_n - (M_{n+1}' - M_n').
\end{aligned}$$

Since $\{A_n'\}$ is predictable and $\{M_n'\}$ is a martingale,

$$\begin{aligned}
A_{n+1}' - A_n' &= E[A_{n+1}' - A_n'|\mathcal{B}_n] \\
&= E[X_{n+1} - X_n - (M_{n+1}' - M_n')|\mathcal{B}_n] \\
&= E[X_{n+1}|\mathcal{B}_n] - X_n \ P\text{-a.s..}
\end{aligned}$$

Similarly,

$$A_{n+1} - A_n = E[X_{n+1}|\mathcal{B}_n] - X_n \ P\text{-a.s..}$$

Thus, since $A_0 = A_0' = 0$,

$$\begin{aligned}
A_n' &= A_0' + \sum_{k=1}^{n}(A_k' - A_{k-1}') \\
&= A_0 + \sum_{k=1}^{n}(A_k - A_{k-1}) \\
&= A_n \ P\text{-a.s..}
\end{aligned}$$

Finally,

$$M_n = X_n - A_n = X_n - A_n' = M_n' \ P\text{-a.s.}, \forall n \in \mathbb{Z}^{\geqslant 0}. \qquad \square$$

Example 12.38. Suppose $\{B_k\}_{k\geqslant 1}$ is an iid sequence of Poisson random variables having rate λ. Define

$$N_n = \sum_{k=1}^{n} B_k, \forall n \in \mathbb{Z}^{\geqslant 0}.$$

Define

$$\mathcal{B}_n = \sigma(N_0, \ldots, N_n), n \geqslant 0.$$

Then $\{(N_n, \mathcal{B}_n)\}_{n\geqslant 0}$ is a submartingale.

Proof. Note

$$\sigma(N_0, \ldots, N_n) = \sigma(B_1, \ldots, B_n).$$

(a) Clearly, $\mathcal{B}_n \subseteq \mathcal{B}_{n+1}$ for any $n \in \mathbb{Z}^{\geqslant 0}$.

(b) Since $B_k \in \mathcal{B}_k$ for any $k \in \mathbb{Z}^{\geqslant 0}$ and $E[N_n] = n\lambda < \infty$, $N_n \in \mathcal{B}_n$ and $N_n \in L_1$ for any $n \in \mathbb{Z}^{\geqslant 0}$.

(c)

$$E[N_{n+1}|\mathcal{B}_n] = E\left[\sum_{k=1}^{n+1} B_n|\mathcal{B}_n\right] = E[N_n|\mathcal{B}_n] + E[B_{n+1}|\mathcal{B}_n] = N_n + E[B_{n+1}] = N_n + \lambda \geqslant N_n.$$

What's the Doob's decomposition?

$$\begin{aligned} M_n &= N_0 + \sum_{k=1}^{n} (N_k - E[N_k|\mathcal{B}_{k-1}]) \\ &= \sum_{k=1}^{n} (N_k - (N_{k-1} + E[\mathcal{B}_k])) \\ &= \sum_{k=1}^{n} (N_k - N_{k-1}) + \sum_{k=1}^{n} E[B_k] \\ &= N_n - n\lambda. \end{aligned}$$

Furthermore, $A_n = N_n - M_n = n\lambda$. □

Proposition 12.39. (a) Let $\{X_n, \mathcal{B}_n\}_{n\geqslant 0}$ be a martingale, and let $\phi : \mathbb{R} \to \mathbb{R}$ be convex on $\mathbb{R}$, where

$$\phi(X_n) \in L_1(\Omega, \mathcal{B}, P), \forall n \in \mathbb{Z}^{\geqslant 0}.$$

Then $\{\phi(X_n), \mathcal{B}_n\}_{n\geqslant 0}$ is a submartingale.

(b) Let $\{(X_n, \mathcal{B}_n)\}$ be a submartingale and $\phi : \mathbb{R} \to \mathbb{R}$ be a convex, nondecreasing function that satisfies

$$\phi(X_n) \in L_1(\Omega, \mathcal{B}, P), \forall n \in \mathbb{Z}^{\geqslant 0}.$$

Then $\{(\phi(X_n), \mathcal{B}_n)\}_{\geqslant 0}$ is a submartingale.

Proof. By Jensen's inequality,

$$E[\phi(X_{n+1})|\mathcal{B}_n] \geqslant \phi\left(E[X_{n+1}|\mathcal{B}_n]\right) \geqslant \phi(X_n).$$

□

12.6 Stopping time

Define

$$\mathbb{N} := \{0, 1, 2, \cdots\},$$
$$\overline{\mathbb{N}} = \{0, 1, 2, \ldots, \infty\},$$

and suppose $\{\mathcal{B}_n\}_{n\geqslant 0}$ is a filtration.

Definition 12.40. A random variable $\nu : \Omega \to \overline{\mathbb{N}}$ is a *stopping time* (w.r.t $\{\mathcal{B}_n\}$) if

$$\{\nu = n\} \in \mathcal{B}_n, \forall n \in \mathbb{N}.$$

Remark. To fix ideas, image a sequence of gambles. Then ν is the rule for when to stop and $\mathcal{B}_n$ is the information accumulated up to time n. You decide whether or not to stop after the nth gamble based on information availble up to and including the nth gamble.

Definition 12.41. Define

$$\mathcal{B}_\infty = \bigvee_{n\in\mathbb{N}} \mathcal{B}_n = \sigma(\mathcal{B}_n, n \in \mathbb{N}).$$

Then

$$\{\nu = \infty\} = [\nu < \infty]^c = \left(\bigcup_{n\in\mathbb{N}} [\nu = n]\right)^c = \bigcap_{n\in\mathbb{N}} [\nu = n]^c \in \mathcal{B}_\infty.$$

Requiring

$$\{\nu = n\} \in \mathcal{B}_n,\ n \in \mathbb{N}$$

implies

$$\{\nu = n\} \in \mathcal{B}_n,\ n \in \overline{\mathbb{N}}.$$

Example 12.42. Suppose $\{X_n\}_{n\in\mathbb{N}}$ is adapt to a filtration $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$. For any $A \in \mathcal{B}(\mathbb{R})$, define

$$\nu = \inf\{n \in \mathbb{N} \mid X_n \in A\}.$$

with the convention that $\inf \emptyset = \infty$. Then ν is a stopping time w.r.t $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$ since

$$\{\nu = n\} = \{X_0 \notin A, \ldots, X_{n-1} \notin A, X_n \in A\} \in \mathcal{B}_n.$$

Example 12.43. Suppose ν is a stopping time w.r.t. $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$. Define

$$\mathcal{B}_\nu = \{A \in \mathcal{B}_\infty \mid A \cap \{\nu = n\} \in \mathcal{B}_n, \forall n \in \mathbb{N}\}.$$

$\mathcal{B}_\nu$ consists of all events that have the property that adding the information of when ν occurred, places the intersection in the appropprite σ-field. Claim. $\mathcal{B}_\nu$ is a σ-field.

Proof. (a) For any $n \in \mathbb{N}$, $\Omega \cap \{\nu = n\} = \{\nu = n\} \in \mathcal{B}_n$.

(b) Suppose $A \in \mathcal{B}_\nu$. For any $n \in \mathbb{N}$,

$$A^c \cap \{\nu = n\} = \{\nu = n\} \setminus (A \cap \{\nu = n\}) \in \mathcal{B}_n.$$

So $A^c \in \mathcal{B}_n$.

(c) Suppose $\{A_k\}_{k=1}^{\infty} \subseteq \mathcal{B}_\nu$. Then$\forall n \in \mathbb{N}$,

$$\left(\bigcup_{k=1}^{\infty} A_k\right) \cap \{\nu = n\} = \bigcup_{k=1}^{\infty} A_k \cap \{\nu = n\} \in \mathcal{B}_n.$$

So

$$\bigcup_{k=1}^{\infty} A_k \in \mathcal{B}_n.$$ □

Basic Facts:

(a) If $\nu \equiv k$, for some $k \in \mathbb{N}$, then ν is a stopping time and $\mathcal{B}_\nu = \mathcal{B}_k$.

Proof. For any $n \in \mathbb{N}$,

$$\{\nu = n\} = \begin{cases} \emptyset, & n \neq k \\ \Omega, & n = k \end{cases} \in \mathcal{B}_n \ .$$

Note

$$\mathcal{B}_\nu = \{A \in \mathcal{B}_\infty : A \cap \{\nu = n\} \in \mathcal{B}_n, \forall n \in \mathbb{N}\}.$$

If $A \in \mathcal{B}_\nu$, by the definition of $\mathcal{B}_\nu$,

$$A = A \cap \Omega = A \cap \{\nu = k\} \in \mathcal{B}_k.$$

So $\mathcal{B}_\nu \subseteq \mathcal{B}_k$. Suppose $A \in \mathcal{B}_k$. If $0 \leqslant n < k$ or $n > k$,

$$A \cap \{\nu = k\} = A \cap \emptyset = \emptyset \in \mathcal{B}_n.$$

If $n = k$, $A \cap \{\nu = k\} = A \in \mathcal{B}_k = \mathcal{B}_n$. So $A \subseteq \mathcal{B}_\nu$. Hence $\mathcal{B}_k \subseteq \mathcal{B}_\nu$. Thus, $\mathcal{B}_k = \mathcal{B}_\nu$. □

(b) If ν is a stopping time and $B \in \mathcal{B}_\nu$, then $B \cap \{\nu = \infty\} \in B_\infty$, and hence

$$B \cap \{\nu = n\} \in \mathcal{B}_n, \forall n \in \overline{\mathbb{N}}.$$

Proof.

$$\begin{aligned} B \cap \{\nu = \infty\} &= B \cap \{\nu < \infty\}^c \\ &= B \cap \left(\bigcup_{n\in\mathbb{N}} \{\nu = n\}\right)^c \\ &= B \cap \bigcap_{n\in\mathbb{N}} \{\nu \neq n\} \\ &= \bigcap_{n\in\mathbb{N}} B \cap \{\nu \neq n\} \\ &\in \mathcal{B}_\infty, \end{aligned}$$

since $B \in \mathcal{B}_\nu \subseteq \mathcal{B}_\infty$ and $\{\nu \neq n\} = \{\nu = n\}^c \in \mathcal{B}_n \subseteq \mathcal{B}_\infty$. □

(c) If ν is a stopping time, then $\nu \in \mathcal{B}_\nu \subseteq \mathcal{B}_\infty$.

Proof. $\{\nu = n\} \in \mathcal{B}_n \subseteq \mathcal{B}_\infty$ for any $n \in \overline{\mathbb{N}}$. Since for any $n \in \overline{\mathbb{N}}$,

$$\{\nu = n\} \cap \{\nu = k\} = \left\{ \begin{array}{ll} \{\nu = k\}, & k \leqslant n \\ \{\nu = n\}, & k > n \end{array} \right. \in \mathcal{B}_k, \forall k \in \overline{\mathbb{N}}\ ,$$

we have $\{\nu = n\} \in \mathcal{B}_\nu, \forall n \in \overline{\mathbb{N}}$. Since the range of ν is $\overline{\mathbb{N}}$, $\nu \in \mathcal{B}_\nu$. □

(d) ν is a stopping time if and only if

$$\{\nu \leqslant n\} \in \mathcal{B}_n, \forall n \in \mathbb{N},$$

if and only if

$$\{\nu > n\} \in \mathcal{B}_n, \forall n \in \mathbb{N}.$$

Proof. Since

$$\{\nu \leqslant n\} = \bigcup_{0 \leqslant j \leqslant n} \{\nu = j\},$$

we have $\{\nu = n\} = \{\nu \leqslant n\} - \{\nu \leqslant n-1\}$ and $\{\nu > n\} = \{\nu \leqslant n\}^c$. □

(e) If $B \in \mathcal{B}_\infty$, then

$$B \in \mathcal{B}_\nu \Longleftrightarrow B \cap \{\nu \leqslant n\} \in \mathcal{B}_n, \forall n \in \mathbb{N}.$$

(f) If $\{\nu_k\}_{k \in \mathbb{Z}^+}$, then so is

$$\min_{k \in \mathbb{Z}^+} \{\nu_k\} \text{ and } \max_{k \in \mathbb{Z}^+} \{\nu_k\}.$$

Proof.

$$\left\{ \bigwedge_{k \in \mathbb{Z}^+} \nu_k > n \right\} = \bigcap_{k \in \mathbb{Z}^+} \{\nu_k > n\} \in \mathcal{B}_n, \forall n \in \mathbb{N},$$

$$\left\{ \bigvee_{k \in \mathbb{Z}^+} \nu_k \leqslant n \right\} = \bigcap_{k \in \mathbb{Z}^+} \{\nu_k \leqslant n\} \in \mathcal{B}_n, \forall n \in \mathbb{N}.$$

□

(g) If $\{\nu_k\}$ is a monotone family of stopping times, $\lim_{k \to \infty} \nu_k$ is a stopping time since the limit is

$$\bigvee_{k \in \mathbb{Z}^+} \nu_k \quad \text{or} \quad \bigwedge_{k \in \mathbb{Z}^+} \nu_k.$$

(h) If $\nu_i, i = 1, 2$ are stopping times, so is $\nu_1 + \nu_2$.

Example 12.44. If ν is a stopping times, then

$$\nu_n = \nu \wedge n, \forall n \in \mathbb{Z}^+$$

is a stopping time (which is bounded) since both ν and n are stopping times.

Facts concerning the comparison of two stopping times ν and ν'. We assume ν and ν' are both stopping time w.r.t. the same filtration $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$.

(a)

$$\{\nu < \nu'\},\ \{\nu = \nu'\},\ \{\nu \leqslant \nu'\} \in \mathcal{B}_\nu \cap \mathcal{B}_{\nu'}.$$

Proof. Note

$$\{\nu = \nu'\} \cap \{\nu = n\} = \{n = \nu'\} \cap \{\nu = n\} \in \mathcal{B}_n, \forall n \in \mathcal{B}_n,$$

so $\{\nu = \nu'\} \in \mathcal{B}_\nu$. Note that

$$\{\nu < \nu'\} \cap \{\nu = n\} = \{n < \nu'\} \cap \{\nu = n\} \in \mathcal{B}_n, \forall n \in \mathbb{N},$$

so $\{\nu < \nu'\} \in \mathcal{B}_\nu$. Thus,

$$\{\nu \leqslant \nu'\} = \{\nu < \nu'\} \cup \{\nu = \nu'\} \in \mathcal{B}_\nu.$$

or use $(\star\star\star)$

$$\{\nu \leqslant \nu'\} \cap \{\nu = n\} = \{n \leqslant \nu'\} \cap \{\nu = n\} = \{n - 1 < \nu'\} \cap \{\nu = n\} \in \mathcal{B}_n, \forall n \in \mathbb{N},$$

Similarly, note

$$\{\nu \leqslant \nu'\} \cap \{\nu' = n\} = \{\nu \leqslant n\} \cap \{\nu' = n\} \in \mathcal{B}_n,$$

so $\{\nu \leqslant \nu'\} \in \mathcal{B}_{\nu'}$. □

(b) If $B \in \mathcal{B}_\nu$, then

$$B \cap \{\nu \leqslant \nu'\} \in \mathcal{B}_{\nu'},$$
$$B \cap \{\nu < \nu'\} \in \mathcal{B}_{\nu'}.$$

Proof.

$$B \cap \{\nu \leqslant \nu'\} \cap \{\nu' = n\} = (B \cap \{\nu \leqslant n\}) \cap \{\nu' = n\} \in \mathcal{B}_n,$$

since

$$B \cap \{\nu \leqslant n\} = B \cap \bigcup_{k=1}^{n} \{\nu = k\} = \bigcup_{k=1}^{n} B \cap \{\nu = k\} \in \mathcal{B}_n.$$ □

(c) If $\nu \leqslant \nu'$ on Ω, then $\mathcal{B}_\nu \subseteq \mathcal{B}_{\nu'}$.

Proof. This follows from (2) since $\{\nu \leqslant \nu'\} = \Omega$. □

12.7 Positive Supermartingale

Suppose $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a positive supermartingale so that

$$X_n \geqslant 0\ P\text{-a.s.},$$
$$X_n \in \mathcal{B}_n,$$

and

$$E[X_{n+1}|\mathcal{B}_n] \leqslant X_n,\ n \geqslant 0,\ P\text{-a.s.}.$$

Consider the following questions.

(a) When does $\lim_{n\to\infty} X_n$ exist? In what sense does convergence take place if some form of convergence holds? Since supermartingale tend to decrease, at least on the average, one expects that under reasonbale conditions, supermartingale bounded below by 0 should converge.

(b) Is fairness preserved under random stopping? If $\{X_n\}$ is a martingale, we know that we have constant mean; that is

$$E(X_n) = E(X_0), \forall n \in \mathbb{N}.$$

Is

$$E(X_\nu) = E(X_0)$$

for some reasonable class of stopping times ν?

When it holds, preservation of the mean under random stopping is quite useful. However, we can quickly see that preservation of the mean under random stopping does not always hold.

Example 12.45. Let $\{X_0 = 0, X_n = \sum_{i=1}^{n} Y_i, n \geqslant 1\}$ be the BernoullI random walk so that $\{Y_i\}_{i\in\mathbb{Z}^+}$ are iid and

$$P(Y_i = \pm 1) = \frac{1}{2},\ i \in \mathbb{Z}^+.$$

Then $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a martingale, where $\mathcal{B}_n = \sigma(X_0, \ldots, X_n)$. Let

$$\nu = \inf\{n \in \mathbb{Z}^+ | : X_n = 1\}$$

be the first time the random walks hits 1. Standard Markov chain analysis asserts that

$$P(\nu < \infty) = 1.$$

But $X_\nu = 1$ so that

$$E[X_\nu] = 1 \neq E(X_0) = 0.$$

$$\big(X_\nu(\omega) := X_{\nu(\omega)}(\omega), \forall \omega \in \Omega.\big)$$

Thus, for random stopping to preserve the process mean, we need restrictions either on $\{X_n\}$ or on ν or both.

12.7.1 Operations on Supermartingale

We consider two transformations of supermartingles which yield supermartingales.

Proposition 12.46 (Pasting of supermartingales)**.** For $i = 1, 2$, let

$$\left\{\left(X_n^{(i)}, \mathcal{B}_n\right), n \geqslant 0\right\}$$

be positive supermartingales. Let ν be a stopping time such that

$$X_\nu^{(1)}(\omega) \geqslant X_\nu^{(2)}(\omega), \forall \omega \in \{\nu < \infty\}.$$

For any $n \in \mathbb{N}$, define

$$X_n(\omega) = \begin{cases} X_n^{(1)}(\omega), & n < \nu(\omega) \\ X_n^{(2)}(\omega), & n \geqslant \nu(\omega). \end{cases}$$

Then $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a positive supermartingale, called the pasted supermartingale.

Proof. Write

$$X_n = X_n^{(1)} \mathbb{1}_{\{\nu > n\}} + X_n^{(2)} \mathbb{1}_{\{\nu \leqslant n\}}.$$

(a) Clearly, $\mathcal{B}_n \subseteq \mathcal{B}_{n+1}, \forall n \in \mathbb{N}$.

(b) For any $n \geqslant 0$, $X_n \in \mathcal{B}_n$, $X_n \geqslant 0$ and $X_n \in L_1$.

(c) Since on the set $\{\nu < \infty\}$, $X_n^{(1)} \geqslant X_n^{(2)}$, then for any $k \in \mathbb{Z}^+$, on the set $\{\nu = k\}$: $X_n^{(1)} \geqslant X_n^{(2)}$. Then for any $n \in \mathbb{N}$,

$$\begin{aligned}
X_n &= X_n^{(1)} \mathbb{1}_{\{\nu > n\}} + X_n^{(2)} \mathbb{1}_{\{\nu \leqslant n\}} \\
&\geqslant E\left[X_{n+1}^{(1)} \middle| \mathcal{B}_n \right] \mathbb{1}_{\{\nu > n\}} + E\left[X_{n+1}^{(2)} \middle| \mathcal{B}_n \right] \mathbb{1}_{\{\nu \leqslant n\}} \\
&= E\left[X_{n+1}^{(1)} \mathbb{1}_{\{\nu > n\}} + X_{n+1}^{(2)} \mathbb{1}_{\{\nu \leqslant n\}} \middle| \mathcal{B}_n \right] \\
&= E\left[X_{n+1}^{(1)} \mathbb{1}_{\{\nu > n+1\}} + X_{n+1}^{(1)} \mathbb{1}_{\{\nu = n+1\}} + X_{n+1}^{(2)} \mathbb{1}_{\{\nu \leqslant n\}} \middle| \mathcal{B}_n \right] \\
&\geqslant E\left[X_{n+1}^{(1)} \mathbb{1}_{\{\nu > n+1\}} + X_{n+1}^{(2)} \mathbb{1}_{\{\nu = n+1\}} + X_{n+1}^{(2)} \mathbb{1}_{\{\nu \leqslant n\}} \middle| \mathcal{B}_n \right] \\
&= E\left[X_{n+1}^{(1)} \mathbb{1}_{\{\nu > n+1\}} + X_{n+1}^{(2)} \mathbb{1}_{\{\nu \leqslant n+1\}} \middle| \mathcal{B}_n \right] \\
&= E[X_{n+1} | \mathcal{B}_n], \ P\text{-a.s.}.
\end{aligned}$$

□

Our second operation is to freeze the supermartingale after n steps. We show that if $\{X_n\}$ is a supermartingale (martingale), $\{X_{\nu \wedge n}\}$ is still a supermartingale (martingale). Note that

$$(X_{\nu \wedge n, n \in \mathbb{N}}) = (X_0, X_1, \ldots, X_{n-1}, X_\nu, X_\nu, X_\nu, \cdots).$$

Proposition 12.47. If $\{(X_n, \mathcal{B}_n)\}_{n \in \mathbb{N}}$ is a supermartingale (martingale), then $\{(X_{\nu \wedge n}, \mathcal{B}_n)\}_{n \in \mathbb{N}}$ is also a supermartingale (martingale).

Proof. Assume $\{(X_n, \mathcal{B}_n)\}_{n \in \mathbb{N}}$ is a supermartingale.

(a) Clearly, $\mathcal{B}_n \subseteq \mathcal{B}_{n+1}, \forall n \in \mathbb{N}$.

(b) For any $n \in \mathbb{N}$,

$$X_{\nu \wedge n} = X_\nu \mathbb{1}_{\{n > \nu\}} + X_n \mathbb{1}_{\{\nu \geqslant n\}} = \sum_{k=0}^{n-1} X_k \mathbb{1}_{\{\nu = j\}} + X_n \mathbb{1}_{\{\nu \geqslant n\}} \in \mathcal{B}_n.$$

Also, $X_{\nu \wedge n} \in L_1$.

(c) For any $n \in \mathbb{N}$,

$$\begin{aligned}
E\left[X_{\nu\wedge\{n+1\}}\middle|\mathcal{B}_n\right] &= E\left[\sum_{k=0}^{n} X_k \mathbb{1}_{\{\nu=n\}} + X_{n+1}\mathbb{1}_{\{\nu\geqslant n+1\}}\middle|\mathcal{B}_n\right] \\
&= \sum_{k=0}^{n} E[X_k|\mathcal{B}_n]\mathbb{1}_{\{\nu=k\}} + E[X_{n+1}|\mathcal{B}_n]\mathbb{1}_{\{\nu\geqslant n+1\}} \\
&= \sum_{k=0}^{n} X_k|\mathbb{1}_{\{\nu=k\}} + X_{n+1}\mathbb{1}_{\{\nu\geqslant n+1\}} \\
&\leqslant \sum_{k=0}^{n} X_k \mathbb{1}_{\{\nu=k\}} + X_n \mathbb{1}_{\{\nu\geqslant n+1\}} \\
&= X_\nu \mathbb{1}_{\{\nu<n+1\}} + X_n \mathbb{1}_{\{\nu\geqslant n+1\}} \\
&= X_\nu \mathbb{1}_{\{\nu\leqslant n\}} + X_n \mathbb{1}_{\{\nu\geqslant n+1\}} \\
&= X_{\nu\wedge n}.
\end{aligned}$$

□

12.7.2 Upcrossings

Let $\{x_n\}_{n\in\mathbb{N}} \subseteq \overline{\mathbb{R}}$, and let $-\infty < a < b < \infty$. Define the crossing times of $[a,b]$ by the sequence $\{x_n\}_{n\in\mathbb{N}}$ as

$$\begin{aligned}
\nu_1 &= \inf\{n \geqslant 0 : x_n \leqslant a\} \\
\nu_2 &= \inf\{n \geqslant \nu_1 : x_n \geqslant b\} \\
\nu_3 &= \inf\{n \geqslant \nu_2 : x_n \leqslant a\} \\
\nu_4 &= \inf\{n \geqslant \nu_3 : x_n \geqslant b\} \\
&\vdots \\
\nu_{2k-1} &= \inf\{n \geqslant \nu_{2k-2} : x_n \leqslant a\} \\
\nu_{2k} &= \inf\{n \geqslant \nu_{2k-1} : x_n \geqslant b\} \\
&\vdots
\end{aligned}$$

Define

$$\beta_{a,b} = \sup\{k \in \mathbb{Z}^+ : \nu_{2k} < \infty\},$$

the number of upcrossings of $[a,b]$ by $\{x_n\}$ (from $[-\infty, a]$ to $[b, \infty]$).

Lemma 12.48 (Upcrossing and Convergence)**.** The sequence $\{x_n\}_{n\in\mathbb{N}} \subseteq \overline{\mathbb{R}}$ converges in $\overline{\mathbb{R}}$ if and only if

$$\beta_{a,b} < \infty, \forall a, b \in \mathbb{Q} \text{ and } a < b.$$

Proof. $\Longleftarrow$ Assume

$$\liminf_{n\to\infty} x_n < \limsup_{n\to\infty} x_n.$$

Then there exist $a, b \in \mathbb{Q}$ and $a < b$ such that

$$\liminf_{n\to\infty} x_n < a < b < \limsup_{n\to\infty} x_n.$$

So $x_n < a$ i.o. and $x_n > b$ i.o.. Thus, $\beta_{a,b} = \infty$, a contradiction.

$\Longrightarrow$ Suppose there exists $a, b \in \mathbb{Q}$ and $a < b$ such that $\beta_{a,b} = \infty$. Then $x_n \leqslant a$ i.o. and $x_n \leqslant b$ i.o. so that

$$\liminf_{n\to\infty} x_n \leqslant a < b \leqslant \limsup_{n\to\infty} x_n,$$

a contradition. □

12.7.3 Boundedness Properties

This section considers how to prove the following intuitive fact: A positive supermartingale tends to decrease but must stay non-negative, so the process should be bounded.

Proposition 12.49. Let $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ be a positive supermartingale ($X_n \geqslant 0$ a.s.). We have that

$$\sup_{n\in\mathbb{N}} X_n < \infty \text{ } a.s. \text{ on } \{X_0 < \infty\},$$

and

$$P\left(\sup_{n\in\mathbb{N}} X_n > a|\mathcal{B}_0\right) \leqslant \min\left\{\frac{X_0}{a}, 1\right\}, \forall \text{constants } a > 0.$$

Proof. Consider two supermartingales

$$\{(X_n^{(i)}, \mathcal{B}_n)\}_{n\in\mathbb{N}},\ i = 1, 2,$$

defined by $X_n^{(1)} = X_n$ and $X_n^{(2)} = a$ for any $n \in \mathbb{N}$,. Define a stopping time

$$\nu_a = \inf\{n \in \mathbb{N} : X_n \geqslant a\}.$$

Note on the set $\{\nu_a < \infty\}$, $X_{\nu_a}^{(1)} = X_{\nu_a} \geqslant a = X_{\nu_a}^{(2)}$. Define

$$Y_n := \begin{cases} X_n, & n < \nu_a, \\ a, & n \geqslant \nu_a \end{cases}.$$

By the pasting property, $\{(Y_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a positive supermartingale. This means

$$E[Y_n|\mathcal{B}_0] \leqslant Y_0, \forall n \in \mathbb{N}.$$

Furthermore,

$$Y_n = X_n \mathbb{1}_{\{\nu_a > n\}} + a\mathbb{1}_{\{\nu_a \leqslant n\}} \geqslant a\mathbb{1}_{\{\nu_a \leqslant n\}},$$

and

$$Y_0 = X_0 \mathbb{1}_{\{\nu_a > 0\}} + a\mathbb{1}_{\{\nu_a = 0\}} = X_0 \mathbb{1}_{\{X_0 < a\}} + a\mathbb{1}_{\{X_0 \geqslant a\}} = \min\{X_0, a\}.$$

Then

$$\min\{X_0, a\} = Y_0 \geqslant E[Y_n|\mathcal{B}_0] \geqslant E[a\mathbb{1}_{\{\nu_a \leqslant n\}}|\mathcal{B}_0] = aP(\nu_a \leqslant n|\mathcal{B}_0) = aP\left(\sup_{0\leqslant k\leqslant n} X_k \geqslant a|\mathcal{B}_0\right).$$

So

$$P\left(\sup_{0\leqslant k\leqslant n} X_k \geqslant a|\mathcal{B}_0\right) \leqslant \min\left\{\frac{X_0}{a}, 1\right\}.$$

By the conditional MCT,

$$P\left(\sup_{n\in\mathbb{N}} X_n > a|\mathcal{B}_0\right) \leqslant \min\left\{\frac{X_0}{a}, 1\right\}.$$

Next, notice that

$$\begin{aligned} P\left(\sup_{n\in\mathbb{N}} X_n \geqslant a, X_0 < \infty\right) &= E\left[E\left[\mathbb{1}_{\{\sup_{n\in\mathbb{N}} X_n \geqslant a\}}\mathbb{1}_{\{X_0<\infty\}}\middle|\mathcal{B}_0\right]\right] \\ &= E\left[\mathbb{1}_{\{X_0<\infty\}} P\left(\sup_{n\in\mathbb{N}} X_n > a|\mathcal{B}_0\right)\right] \\ &\leqslant E\left[\mathbb{1}_{\{X_0<\infty\}} \min\left\{\frac{X_0}{a}, 1\right\}\right] \\ &\to 0 \text{ as } a \to \infty. \end{aligned}$$

by the DCT. So

$$P\left(\sup_{n\in\mathbb{N}} X_n = \infty, X_0 < \infty\right) \leqslant 0.$$

Thus,

$$P\left(\sup_{n\in\mathbb{N}} X_n = \infty, X_0 < \infty\right) = 0.$$

As a result, on $\{X_0 < \infty\}$, $\sup_{n\in\mathbb{N}} X_n < \infty$ a.s.. □

12.7.4 Convergence of Positive Supermartingale

Let $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ be a positive supermartingale. For $a, b \in \mathbb{R}$ and $a < b$, define

$$\begin{aligned} \nu_1(\omega) &= \inf\{n \geqslant 0 : X_n(\omega) \leqslant a\} \\ \nu_2(\omega) &= \inf\{n \geqslant \nu_1(\omega) : X_n(\omega) \geqslant b\} \\ \nu_3(\omega) &= \inf\{n \geqslant \nu_2(\omega) : X_n(\omega) \leqslant a\} \\ \nu_4(\omega) &= \inf\{n \geqslant \nu_3(\omega) : X_n(\omega) \geqslant b\} \\ &\vdots \\ \nu_{2k-1}(\omega) &= \inf\{n \geqslant \nu_{2k-2}(\omega) : X_n(\omega) \leqslant a\} \\ \nu_{2k}(\omega) &= \inf\{n \geqslant \nu_{2k-1}(\omega) : X_n(\omega) \geqslant b\} \\ &\vdots \end{aligned}$$

and define

$$\beta_{a,b}(\omega) = \sup\{k \in \mathbb{Z}^+ : \nu_{2k}(\omega) < \infty\},$$

the number of upcrossings of $[a, b]$ by $\{X_n(\omega)\}$ (from $[-\infty, a]$ to $[b, \infty]$). Note we have the fact

$$\{\omega : \lim_{n\to\infty} X_n(\omega) \text{ exists}\} = \bigcap_{a,b\in\mathbb{Q},a<b} \{\omega : \beta_{a,b}(\omega) < \infty\}.$$

So

$$\lim_{n\to\infty} X_n \text{ exists a.s.} \Longleftrightarrow \beta_{a,b} < \infty, \forall\, a, b \in \mathbb{Q} \text{ and } a < b.$$

Thus, to show $P\left(\lim_{n\to\infty} X_n \text{ exists}\right) = 1$, it suffices to show $P\left(\beta_{a,b} < \infty\right) = 1$ for any $a, b \in \mathbb{Q}$ and $a < b$.

Proposition 12.50 (Dubin's inequality)**.** Let $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ be a positive supermartingale. Suppose $0 < a < b$. Then

(a)

$$P\left(\beta_{a,b} \geqslant k|\mathcal{B}_0\right) \leqslant \left(\frac{a}{b}\right)^k \min\left(\frac{X_0}{a}, 1\right), \forall k \in \mathbb{Z}^+.$$

(b) $\beta_{a,b} < \infty$ a.s..

Proof. (2) follows from (1) since

$$E[\beta_{a,b}] = \sum_{k=1}^{\infty} P(\beta_{a,b} \geqslant k) \leqslant \sum_{k=1}^{\infty} \left(\frac{a}{b}\right)^k < \infty.$$

Start by considering the supermartingales

$$X_n^{(1)} = 1, \forall n \in \mathbb{N},$$

$$X_n^{(2)} = \frac{X_n}{a}, \forall n \in \mathbb{N},$$

and paste at ν_1. Note on $\{\nu_1 < \infty\}$,

$$X_{\nu_1}^{(1)} = 1 \geqslant \frac{X_{\nu_1}}{a} = X_{\nu_1}^{(2)}.$$

So

$$Y_n^{(1)} = \begin{cases} 1, & n < \nu_1, \\ \frac{X_n}{a}, & n \geqslant \nu_1, \end{cases}$$

is a positive supermartingale w.r.t $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$. Now compare and paste

$$X_n^{(3)} = Y_n^{(1)},$$

$$X_n^{(4)} = b/a$$

at the stopping time ν_2. On $\{\nu_2 < \infty\}$,

$$X_{\nu_2}^{(3)} = Y_{\nu_2}^{(1)} = \frac{X_{\nu_2}}{a} \geqslant \frac{b}{a} = X_{\nu_2}^{(4)}.$$

So

$$\begin{aligned} Y_n^{(2)} &= \begin{cases} Y_n^{(1)}, & n < \nu_2, \\ \frac{b}{a}, & n \geqslant \nu_2, \end{cases} \\ &= \begin{cases} 1, & n < \nu_1, \\ \frac{X_n}{a}, & \nu_1 < n < \nu_2, \\ \frac{b}{a}, & n \geqslant \nu_2, \end{cases} \end{aligned}$$

is a positive supermartingale w.r.t $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$. Now compare $Y_n^{(2)}$ and $\frac{b}{a}\frac{X_n}{a}$. On $\{\nu_3<\infty\}$,

$$Y_{\nu_3}^{(2)}=\frac{b}{a}\geqslant\frac{b}{a}\frac{X_{\nu_3}}{a},$$

and so

$$Y_n^{(3)}=\begin{cases}Y_n^{(2)}, & n<\nu_3,\\ \frac{b}{a}\frac{X_n}{a}, & n\geqslant\nu_3,\end{cases}$$

is a positive supermartingale w.r.t $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$. Now compare $Y_n^{(3)}$ and $\left(\frac{b}{a}\right)^2$. On $\{\nu_4<\infty\}$,

$$Y_{\nu_4}^{(3)}=\frac{b}{a}\frac{X_{\nu_4}}{a}\geqslant\left(\frac{b}{a}\right)^2,$$

and so

$$Y_n^{(4)}=\begin{cases}Y_n^{(3)}, & n<\nu_4,\\ \left(\frac{b}{a}\right)^2, & n\geqslant\nu_4,\end{cases}$$

$$=\begin{cases}1, & n<\nu_1,\\ \frac{X_n}{a}, & \nu_1\leqslant n<\nu_2,\\ \frac{b}{a}, & \nu_2\leqslant n<\nu_3,\\ \frac{b}{a}\frac{X_n}{a}, & \nu_3\leqslant n<\nu_4,\\ \left(\frac{b}{a}\right)^2, & n\geqslant\nu_4,\end{cases}$$

is a positive supermartingale w.r.t $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$. Continuing on in this manner, we see for any $k\in\mathbb{Z}^+$,

$$Y_n^{(2k)}=\begin{cases}1, & n<\nu_1,\\ \frac{X_n}{a}, & \nu_1\leqslant n<\nu_2,\\ \frac{b}{a}, & \nu_2\leqslant n<\nu_3,\\ \frac{b}{a}\frac{X_n}{a}, & \nu_3\leqslant n<\nu_4,\\ \vdots & \\ \left(\frac{b}{a}\right)^{k-1}\frac{X_n}{a}, & \nu_{2k-1}\leqslant n<\nu_{2k},\\ \left(\frac{b}{a}\right)^k, & n\geqslant\nu_{2k},\end{cases}$$

is a positive supermartingale w.r.t $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$. Note that

$$Y_0^{(2k)}=Y_0^{(2k)}\mathbb{1}_{\{\nu_1>0\}}+Y_0^{(2k)}\mathbb{1}_{\{\nu_1=0\}}=\mathbb{1}_{\{\nu_1>0\}}+\frac{X_0}{a}\mathbb{1}_{\{\nu_1=0\}}=\mathbb{1}_{\{X_0\geqslant a\}}+\frac{X_0}{a}\mathbb{1}_{\{X_0\leqslant a\}}=\min\left\{\frac{X_0}{a},1\right\}.$$

Also for any $n\in\mathbb{N}$,

$$Y_n^{(2k)}\geqslant Y_n^{(2k)}\mathbb{1}_{\{\nu_{2k}\leqslant n\}}=\left(\frac{b}{a}\right)^k\mathbb{1}_{\{\nu_{2k}\leqslant n\}}\ P\text{-a.s.}.$$

Since $\{(Y_n^{2k},\mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a positive supermartingale,

$$\left(\frac{b}{a}\right)^k P\left(\nu_{2k}\leqslant n|\mathcal{B}_0\right)\leqslant E[Y_n^{(2k)}|\mathcal{B}_0]\leqslant Y_0^{(2k)}=\min\left\{\frac{X_0}{a},1\right\}.$$

Thus,

$$P\left(\nu_{2k}\leqslant n|\mathcal{B}_0\right)\leqslant\left(\frac{a}{b}\right)^k\min\left\{\frac{X_0}{a},1\right\},\forall n\in\mathbb{N}.\qquad\square$$

By the conditional MCT,

$$P(\nu_{2k} < \infty|\mathcal{B}_0) \leqslant \left(\frac{a}{b}\right)^k \min\left\{\frac{X_0}{a}, 1\right\}.$$

So

$$P(\beta_{a,b} \geqslant k|\mathcal{B}_0) \leqslant \left(\frac{a}{b}\right)^k \min\left\{\frac{X_0}{a}, 1\right\}.$$

Let $k \to \infty$ and we see $P(\beta_{a,b} = \infty|\mathcal{B}_0) = 0$.

Remark. Since any martingale is also a supermartingale and -1 times a supermartingale is a submartingale, we have the same conclusion for martingales and submartingale.

Theorem 12.51 (Martingale Convergence Theorem)**.** *If $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a positive supermartingale, then $\lim_{n\to\infty} X_n =: X_\infty$ exists a.s.. Furthermore, $E[X_\infty|\mathcal{B}_n] \leqslant X_n$ for any $n \in \mathbb{N}$, so $\{(X_n, \mathcal{B}_n)\}_{n\in\overline{\mathbb{N}}}$ is a positive supermartingale. Moverover, $E[X_\infty] \leqslant E[X_n] < \infty$ for any $n \in \mathbb{N}$. So $X_\infty \in L_1$.*

Proof. It suffices to show X_∞ can be added to $\{X_n\}_{n\in\mathbb{N}}$ while preserving the supermartingale property. Fix $a, b \in \mathbb{R}$ and $a < b$. By Dubin's Inequality, $P(\beta_{a,b} < \infty) = 1$. So $X_\infty := \lim_{n\to\infty} X_n$ exists a.s.. Let $p \in \mathbb{N}$. For $n \geqslant p$,

$$E\left[\inf_{m\geqslant n} X_m \middle| \mathcal{B}_p\right] \leqslant E\left[X_n|\mathcal{B}_p\right] \leqslant X_p \ P\text{-a.s.}.$$

Since $X_\infty = \lim_{n\to\infty} \inf_{m\geqslant n} X_m$ P-a.s., by the conditional MCT, $E[X_\infty|\mathcal{B}_p] \leqslant X_p$ P-a.s.. □

Remark. The last statement says we can add a last variable preserves the supermartingale property. This is the *closure* property to be discussed in the next subsection and is an essential concept for the stopping theorems.

12.7.5 Closure

If $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is positive martingale, then we know it is almost surely convergent. But when is it the case that

(a) $X_n \stackrel{L_1}{\to} X_\infty$,

(b) $E[X_\infty|\mathcal{B}_n] = X_n$ so that $\{(X_n, \mathcal{B}_n)\}_{n\in\overline{\mathbb{N}}}$ is a positive martingale?

Even though it is true that $X_n \stackrel{a.s.}{\to} X_\infty$ and $E(X_m|\mathcal{B}_n) = X_n$ for any $m > n$, it is not necessarily the case that

$$E(X_\infty|\mathcal{B}_n) = X_n.$$

Extra conditions are needed. Consider, for instance, the example of the simple branching process. If $\{Z_n\}_{n\in\mathbb{N}}$ is the process with $Z_0 = 1$ and Z_n representing the number of particles in the nth generation and $m = E(Z_1)$ is the mean offspring number per individual, then $\{Z_n/m^n\}$ is a non-negative martingale so the almost sure limit exists:

$$W_n := Z_n/m^n \xrightarrow{\text{a.s.}} W.$$

However, if $m \leqslant 1$, then extinction is sure so $W \equiv 0$ and we do not have

$$E[W|\mathcal{B}_n] = Z_n/m^n.$$

Definition 12.52 (Closed Martingale)**.** A martingale $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is closed on the right if there exists an integrable random variable $X_\infty \in \mathcal{B}_\infty$ such that

$$X_n = E[X_\infty|\mathcal{B}_n], \forall n \in \mathbb{N}.$$

In this case, $\{(X_n, \mathcal{B}_n)\}_{n\in\overline{\mathbb{N}}}$ is a martingale.

In what follows, we write L_p^+ for the random variables $\xi \in L_p$ and $X \geqslant 0$ P-a.s..

Proposition 12.53. Let $p \geqslant 1$, $X \in L_p^+$ and define $X_n := E[X|\mathcal{B}_n], \forall n \in \mathbb{N}$, and $X_\infty := E[X|\mathcal{B}_\infty]$. Then

$$X_n \xrightarrow{\text{a.s.}} X_\infty,\ X_n \xrightarrow{L_p} X_\infty,$$

and $\{(X_n, \mathcal{B}_n)\}_{n\in\overline{\mathbb{N}}}$ is a closed martingale.

Proof. Clearly, $\{(E[X|\mathcal{B}_n], \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a positive martingale and thus by the Maringale convergence theorem, there exists $X_\infty^\#$ such that $X_n \xrightarrow{\text{a.s.}}\rightarrow X_\infty^\#$. Since $X_n \in \mathcal{B}_n \subseteq \mathcal{B}_\infty$, $X_\infty^\# \in \mathcal{B}_\infty$. NTS:

$$X_\infty^\# = E[X|\mathcal{B}_\infty],\ P\text{-a.s.}.$$

- Case 1: Suppose there exists $\lambda < \infty$ satisfying $P(X \leqslant \lambda) = 1$. Then since $X \leqslant \lambda$ a.s.,

$$E[X|\mathcal{B}_n] \leqslant \lambda\ P\text{-a.s}, \forall n \in \overline{\mathbb{N}}.$$

  Assume the probability space is $(\Omega, \mathcal{B}, P)$. By DCT,

$$\lim_{n\to\infty}\int_A E[X|\mathcal{B}_n]dP = \int_A \lim_{n\to\infty} E[X|\mathcal{B}_n]dP = \int_A E[X|\mathcal{B}_\infty]dP = \int_A X_\infty dP, \forall A \in \mathcal{B}.$$

  Fix $m \in \mathbb{Z}^+$ and let $A \in \mathcal{B}_m$, then$\forall n > m$, we have $A \in \mathcal{B}_m \subseteq \mathcal{B}_n$ and by the definition of the conditional expectation,

$$\int_A E[X|\mathcal{B}_n]dP = \int_A X dP.$$

  So

$$\int_A X_\infty dP = \lim_{n\to\infty}\int_A E[X|\mathcal{B}_n]dP = \int_A X dP.$$

  Thus,

$$\int_A X_\infty dP = \int_A X dP, \forall A \in \bigcup_{m=1}^\infty \mathcal{B}_m.$$

  To show

$$X_\infty^\# = E[X|\mathcal{B}_\infty],\ P\text{-a.s.},$$

  we need to show

$$\int_A X dP = \int_A E[X|\mathcal{B}_\infty] = \int_A X_\infty^\# dP, \forall A \in \mathcal{B}_\infty.$$

  It suffices to show

$$\int_A X dP = \int_A X_\infty^\# dP, \forall A \in \mathcal{B}_\infty.$$

Define

$$\mathcal{C} = \bigcup_{n=1}^{\infty} \mathcal{B}_n,$$

which is a π system. Define

$$\mathcal{D} = \left\{ A \in \mathcal{B} : \int_A X dP = \int_A X_\infty^\# dP \right\}.$$

Note that $\mathcal{C} \subseteq \mathcal{D}$. If we can show $\mathcal{D}$ is a λ-system, by Dykin's theorem, $\mathcal{B}_\infty = \sigma(\mathcal{C}) \subseteq \mathcal{D}$.

(a) Clearly, $\Omega \in D$.

(b) Suppose $A, B \in \mathcal{D}$ and $A \subseteq \mathcal{B}$. Then

$$\int_{A \smallsetminus B} X dP = \int_B X dP - \int_A X dP = \int_B X_\infty^\# - \int_A X_\infty^\# = \int_{B \smallsetminus A} X_\infty^\# dP.$$

So $B \smallsetminus A \in D$.

(c) Suppose $\{A_n\}_{n \in \mathbb{Z}^+} \subseteq \mathcal{D}$ and $A_n \subseteq A_{n+1}, \forall n \in \mathbb{Z}^+$. Then by MCT,

$$\begin{aligned} \int_{\bigcup_{n=1}^\infty An} X dP &= \int_\Omega \mathbb{1}_{\bigcup_{n=1}^\infty A_n} X dP = \lim_{m \to \infty} \int_\Omega \mathbb{1}_{\bigcup_{n=1}^m A_n} X dP \\ &= \lim_{m \to \infty} \int_{A_m} X dP = \lim_{m \to \infty} \int_{A_m} X_\infty^\# dP \\ &= \lim_{m \to \infty} \int_\Omega \mathbb{1}_{\bigcup_{n=1}^m A_n} X_\infty^\# dP = \int_\Omega \mathbb{1}_{\bigcup_{n=1}^\infty A_n} X_\infty^\# dP \\ &= \int_{\bigcup_{n=1}^\infty A_n} X_\infty^\# dP. \end{aligned}$$

Thus, $\mathcal{D}$ is a λ-system. Hence $X_\infty^\# = E[X|\mathcal{B}_\infty]$, P-a.s.. Next, since

$$E[X|\mathcal{B}_n] \leqslant \lambda, \ P\text{-a.s.}, \forall n \in \overline{\mathbb{N}},$$

we have

$$\begin{aligned} \lim_{n \to \infty} \| E[X|\mathcal{B}_n] - E[X|\mathcal{B}_\infty] \|_p &= \lim_{n \to \infty} \left(E \left(|E[X|\mathcal{B}_n] - E[X|\mathcal{B}_\infty]|^p \right) \right)^{\frac{1}{p}} \\ &= \left(\lim_{n \to \infty} E \left(|E[X|\mathcal{B}_n] - E[X|\mathcal{B}_\infty]|^p \right) \right)^{\frac{1}{p}} \\ &\overset{\text{DCT}}{=\!=} \left(E \left(\lim_{n \to \infty} |E[X|\mathcal{B}_n] - E[X|\mathcal{B}_\infty]|^p \right) \right)^{\frac{1}{p}} = 0. \end{aligned}$$

So $X_n \xrightarrow{L_p} X_\infty$. Note that

$$E[X_\infty|\mathcal{B}_n] = E\left[E[X|\mathcal{B}_\infty]|\mathcal{B}_n\right] = E[X|\mathcal{B}_n] = X_n, \forall n \in \mathcal{N},$$

proving $\{(X_n, \mathcal{B}_n)\}_{n \in \overline{\mathbb{N}}}$ is a closed martingale.

- Fix $\lambda > 0$, and write $X = X \wedge \lambda + (X-\lambda)^+$. Since $E(\cdot|\mathcal{B}_n)$ is L_p-norm reducing, we have

$$\begin{aligned}
&\|E[X|\mathcal{B}_n] - E[X|\mathcal{B}_\infty]\|_p \\
&\leqslant \|E[X\wedge\lambda|\mathcal{B}_n] - E[X\wedge\lambda|\mathcal{B}_\infty]\|_p + \left\|E[(X-\lambda)^+|\mathcal{B}_n]\right\|_p + \left\|E[(X-\lambda)^+|\mathcal{B}_\infty]\right\|_p \\
&\leqslant \|E[X\wedge\lambda|\mathcal{B}_n] - E[X\wedge\lambda|\mathcal{B}_\infty]\|_p + 2\left\|(X-\lambda)^+\right\|_p \\
&= I + II.
\end{aligned}$$

Since $0 \leqslant X\wedge\lambda \leqslant \lambda$, by case 1, $I \to 0$. For II, note as $\lambda \to \infty$, $(X-\lambda)^+ \to 0$ and $(X-\lambda)^+ \leqslant X \in L_p$. By DCT, $\|(X-\lambda)^+\|_p \to 0$ as $\lambda \to \infty$. We may conclude that

$$\limsup_{n\to\infty} \|E[X|\mathcal{B}_n] - E[X|\mathcal{B}_\infty]\|_p \leqslant 2\left\|(X-\lambda)^+\right\|_p.$$

The left side is independent of λ, so let $\lambda\to\infty$ to get

$$\limsup_{n\to\infty} \|E[X|\mathcal{B}_n] - E[X|\mathcal{B}_\infty]\|_p = 0.$$

Thus, $E[X|\mathcal{B}_n] \xrightarrow{L_p} E[X|\mathcal{B}_\infty]$. Then $E[X|\mathcal{B}_n] \xrightarrow{p} E[X|\mathcal{B}_\infty]$. Also, we already have

$$E[X|\mathcal{B}_n] \xrightarrow{\text{a.s.}} X_\infty^\#.$$

So $E[X|\mathcal{B}_n] \xrightarrow{p} X_\infty^\#$. Thus, $X_\infty^\# = E[X|\mathcal{B}_\infty]$, P-a.s.. □

Remark. (a) Since $E[X_\infty|\mathcal{B}_n] = E[E[X|\mathcal{B}_\infty]|\mathcal{B}_n] = E[X|\mathcal{B}_n]$, we have $X_n = E[X_\infty|\mathcal{B}_n]$, P-a.s..

(b) We can extend it to the cases where the closing random variable is not necessarily non-negative by writting $X = X^+ - X^-$.

Corollary 12.54. For $p \in \mathbb{Z}^+$, the class of L_p convergence positive martingales is the class of the form

$$\{(E[X|\mathcal{B}_n], \mathcal{B}_n)\}_{n\in\mathbb{N}}$$

with $X \in L_p^+$.

Proof. If $X \in L_p^+$, then $\{(E[X|\mathcal{B}_n], B_n)\}_{n\in\mathbb{N}}$ is a positive martingale and

$$E[X|\mathcal{B}_n] \xrightarrow{L_p} E[X|\mathcal{B}_\infty].$$

Conversely, suppose $\{X_n\}_{n\in\mathbb{N}}$ is a positive martingale and L_p convergent. For $n < r$, then

$$E[X_r|\mathcal{B}_n] = X_n.$$

Now by assumption, $X_r \xrightarrow{L_p} X_\infty$ and $E(\cdot|\mathcal{B}_n)$ is continuous in the L_p-metric by the property of conditional expectation. Thus, as $r\to\infty$,

$$X_n = E[X_r|\mathcal{B}_n] \xrightarrow{L_p} E[X_\infty|\mathcal{B}_n]$$

by continuity. Therefore, since $X_n = E[X_r|\mathcal{B}_n]$ for any $r > n$, $X_n = E[X_\infty|\mathcal{B}_n]$. Thus, it is of the form

$$\{(E[X_\infty|\mathcal{B}_n], \mathcal{B}_n)\}_{n\in\mathbb{N}}.$$

□

12.8 Stopping Supermartingales

What happens to the supermartingale property if deterministic indices are replaced by stopping times?

Lemma 12.55. If ν is a stopping time and $\xi \in L_1$, then

$$E[\xi|\mathcal{B}_\nu] = \sum_{n\in\overline{N}} E[\xi|\mathcal{B}_n]\mathbb{1}_{\{\nu=n\}}.$$

Proof. Note the right side is $\mathcal{B}_\nu$-measurable and$\forall A \in \mathcal{B}_\nu$, since $A \cap \{\nu = n\} \in \mathcal{B}_n$,

$$\begin{aligned}\int_A E[\xi|\mathcal{B}_\nu]dP = \int_A \xi dP &= \sum_{n\in\overline{\mathbb{N}}} \int_{A\cap\{\nu=n\}} \xi dP \\ &= \sum_{n\in\overline{\mathbb{N}}} \int_{A\cap\{\nu=n\}} E[\xi|\mathcal{B}_n]dP \\ &= \int_A \sum_{n\in\overline{\mathbb{N}}} E[\xi|\mathcal{B}_n]\mathbb{1}_{\{\nu=n\}}\xi dP.\end{aligned}$$

So

$$E[\xi|\mathcal{B}_\nu] = \sum_{n\in\overline{\mathbb{N}}} E[\xi|\mathcal{B}_n]\mathbb{1}_{\{\nu=n\}}.$$

□

Theorem 12.56 (Random Stopping). *Suppose $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a positive supermartingale and also suppose $X_n \overset{a.s.}{\to} X_\infty$. Let ν_1, ν_2 be two stopping times. Then*

$$E[X_{\nu_2}|\mathcal{B}_{\nu_1}] \leqslant X_{\nu_1},\ P\text{-a.s. on } [\nu_1 \leqslant \nu_2].$$

Proof. By previous lemma,

$$E[X_{\nu_2}|\mathcal{B}_{\nu_1}] = \sum_{n\in\overline{\mathbb{N}}} E[X_{\nu_2}|\mathcal{B}_n]\mathbb{1}_{\{\nu_1=n\}}.$$

It suffices to show on the set $[\nu_1 \leqslant \nu_2] \cap \{\nu_1 = n\}$,

$$E[X_{\nu_2}|\mathcal{B}_n] \leqslant X_n,\ P\text{-a.s.}, \forall n \in \overline{\mathbb{N}}.$$

Define $Y_n = X_{\nu_2\wedge n}$ for any $n \in \mathbb{N}$. Then $\{(Y_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a positive supermartingale and claim

$$Y_n \xrightarrow{\text{a.s.}} Y_\infty = X_{\nu_2}.$$

Note that if $\nu_2(\omega) < \infty$, then for n large, we have $n \wedge \nu_2(\omega) = \nu_2(\omega)$. On the other hand, if $\nu_2(\omega) = \infty$, then

$$Y_n(\omega) = X_n(\omega) \to X_\infty(\omega) = X_{\nu_2}(\omega).$$

By convergence theorem, $\{(Y_n, \mathcal{B}_n)\}_{n\in\overline{\mathbb{N}}}$ is a positive supermartingale. So

$$E[Y_\infty|\mathcal{B}_n] \leqslant Y_n, \forall n \in \overline{\mathbb{N}}\ ?$$

That is

$$E[X_{\nu_2}|\mathcal{B}_n] \leqslant X_{\nu_2 \wedge n}, \forall n \in \overline{\mathbb{N}}.$$

Hence on the set $[\nu_1 \leqslant \nu_2] \cap \{\nu_1 = n\}$,

$$E[X_{\nu_2}|\mathcal{B}_n] \leqslant X_n,\ P\text{-a.s.}, \forall n \in \overline{\mathbb{N}}. \qquad \square$$

Some special cases:

(a) If $\nu_1 = 0$, then $\nu_2 \geqslant 0 = \nu_1$ and $E[X_{\nu_2}|\mathcal{B}_0] \leqslant X_0$ and then $E[X_{\nu_2}] \leqslant E[X_0]$.

(b) If $\nu_1 \leqslant \nu_2$ pointwise everywhere, then $E[X_{\nu_2}|\mathcal{B}_{\nu_1}] \leqslant X_{\nu_1},\ P$-a.s., and then

$$E[X_{\nu_2}] \leqslant E[X_{\nu_1}],\ P\text{-a.s.}.$$

For martingale, we will see that it is useful to know when equality holds. Unfortunately, this does not always hold and conditions must be present to guarantee preservation of the martingale property under random stopping.

12.8.1 Gambler's Ruin

Suppose $\{Z_n\}_{n\in\mathbb{Z}^+}$ are idd BernoullI random variables satisfying

$$P(Z_1 = \pm 1) = \frac{1}{2},$$

and assuming a fixed $j_0 \in \{0, 1, \ldots, N\}$, let

$$X_0 = j_0,\ X_n = \sum_{i=1}^{n} Z_i + j_0,\ n \in \mathbb{Z}^+$$

be the simple random walk starting from j_0. We ask: starting from j_0, will be the random walk hit 0 or N first? Define

$$\mathcal{B}_n = \sigma(Z_1, \ldots, Z_n),\ \mathcal{B}_0 = \{\emptyset, \Omega\}.$$

Then $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a martingale. Define

$$\nu = \inf\{n > 0, X_n \in \{0, N\}\}.$$

Then $\{(X_{\nu\wedge n}, \mathcal{B}_n)\}$ is a positive martingale. If random stopping preserves the martingale property (to be verified later), then

$$j_0 = E[X_0] = E[X_\nu] = 0P(X_\nu = 0) + NP(X_\nu = N).$$

So

$$P(X_\nu = N) = \frac{j_0}{N},$$

$$P(X_\nu = 0) = 1 - \frac{j_0}{N}.$$

12.9 Martingale and Submartingale Convergence

This section begins by discussing another relation between martingales and submartingales called the Krickeberg decomposition. This decomposition is used to extend convergence properties of positive supermartingale to more general martingal structures.

12.9.1 Krickeberg Decomposition

Theorem 12.57 (Krickeberg Decomposition)**.** *If* $\{(X_n, \mathcal{B}_n)\}_{n\geqslant 0}$ *is a submartingale such that*

$$\sup_{n\in\mathbb{N}} E[X_n^+] < \infty, \ (u.i)$$

then there exists a positive martingale $\{(M_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ *and a positive supermartingale* $\{(Y_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ *and*

$$X_n = M_n - Y_n . \forall n \in \mathbb{N}.$$

Proof. Recall if $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is submartingale, so is $\{(X_n^+, \mathcal{B}_n)\}_{n\in\mathbb{N}}$. Also, $\{E\left[X_p^+|\mathcal{B}_n\right]\}_{p\geqslant n}$ is monotone non-decreasing in p since

$$E\left[X_{p+1}^+|\mathcal{B}_n\right] = E\left[E\left[X_{p+1}^+|\mathcal{B}_p\right]|\mathcal{B}_n\right] \geqslant E\left[X_p^+|\mathcal{B}_n\right], \ P\text{-a.s..}$$

Define

$$M_n := \lim_{p\to\infty} E\left[X_p^+|\mathcal{B}_n\right], \forall n \in \mathbb{N},$$

which is well-defined by MCT. Claim. $\{(M_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a positive martingale.

(a) $M_n \in \mathcal{B}_n, \ M_n \geqslant 0, \ P$-a.s. for any $n \in \mathbb{N}$.

(b) By MCT and since the expetations of submartingale increas,

$$\begin{aligned} E[M_n] &= E\left[\lim_{p\to\infty} E\left[X_p^+|\mathcal{B}_n\right]\right] \\ &= \lim_{p\to\infty} E\left[E\left[X_p^+|\mathcal{B}_n\right]\right] \\ &= \lim_{p\to\infty} E\left[X_p^+\right] \\ &= \sup_{n\geqslant 0} E\left[X_n^+\right] < \infty, \ P\text{-a.s.}, \forall n \in \mathbb{N}. \end{aligned}$$

(c)

$$\begin{aligned} E[M_{n+1}|\mathcal{B}_n] &= E\left[\lim_{p\to\infty} E\left[X_p^+|\mathcal{B}_{n+1}\right]\middle|\mathcal{B}_n\right] \\ &= \lim_{p\to\infty} E\left[E\left[X_p^+|\mathcal{B}_{n+1}\right]\middle|\mathcal{B}_n\right] \\ &= \lim_{p\to\infty} E\left[X_p^+|\mathcal{B}_n\right] = M_n, \ P\text{-a.s.}, \forall n \in \mathbb{N}. \end{aligned}$$

Define $Y_n := M_n - X_n$ for any $n \in \mathbb{N}$. Claim. $\{(Y_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a positive supermartingale.

(a) $Y_n \in \mathcal{B}_n, \forall n \in \mathbb{N}$.

$$M_n = \lim_{p\to\infty} E\left[X_p^+|\mathcal{B}_n\right] \geqslant E[X_n^+|\mathcal{B}_n] = X_n^+ \geqslant X_n^+ - X_n^- = X_n,\ P\text{-a.s.}, \forall n \in \mathbb{N}.$$

So $Y_n \geqslant 0,\ P$-a.s., $\forall n \in \mathbb{N}$. Clearly, $Y_n \in L_1,\ P$-a.s. for any $n \in \mathbb{N}$.

(b)

$$\begin{aligned} E[Y_{n+1}|\mathcal{B}_n] &= E[M_{n+1}\mathcal{B}_n] - E[X_{+1}|\mathcal{B}_n] \\ &= M_n - E[X_{n+1}|\mathcal{B}_n] \\ &\leqslant M_n - X_n = Y_n,\ P\text{-a.s.}, \forall n \in \mathbb{N}. \end{aligned}$$

□

12.9.2 Doob's (Sub)martingale Convergence Theorem

Krickeberg's decomposition leads to the Doob submartingale convergence theorem.

Theorem 12.58 (Submartingale Convergence). *If $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a (sub)-martingale satisfying L_1-bounded, i.e.,*

$$\sup_{n\in\mathbb{N}} E[X_n^+] < \infty,$$

then there exists $X_\infty \in L_1$ such that

$$X_n \overset{a.s.}{X}_\infty.$$

Proof. From the Krickberg decomposition, there exists a positive martingale $\{(M_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ and a positive supermartingale $\{(Y_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ such that $X_n = M_n - Y_n$. Since a martingale is also a supermartingale, by martingale convergence theorem,

$$M_n \xrightarrow{\text{a.s.}} M_\infty \in L_1,\ Y_n \xrightarrow{\text{a.s.}} Y_\infty \in L_1.$$

So M_∞ and Y_∞ are finite a.s., $X_\infty := M_\infty - Y_\infty$ exists a.s., $X_n \xrightarrow{\text{a.s.}} X_\infty$. □

Remark. If $\{(X_n, \mathcal{B}_n)\}$ is a submartingale, then

$$\sup_{n\in\mathbb{N}} E\left[X_n^+\right] < \infty \text{ if and only if } \sup_{n\in\mathbb{N}} E\left[|X_n|\right] < \infty,$$

in which case the submartingale is called L_1-bounded. To see this equivalence, observe that if $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a submartingale, then

$$E\left[|X_n|\right] = E\left[X_n^+\right] + E\left[X_n^-\right] = 2\left[X_n^+\right] - E\left[X_n\right] \leqslant 2\left[X_n^+\right] - E[X_0].$$

So

$$\sup_{n\in\mathbb{N}} E\left[|X_n|\right] \leqslant 2 \sup_{n\in\mathbb{N}} E\left[X_n^+\right] - E[X_0].$$

On the other hand,

$$\sup_{n\in\mathbb{N}} E\left[X_n^+\right] \leqslant \sup_{n\in\mathbb{N}} E\left[|X_n|\right].$$

12.10 Regularity and Closure

Question: Given a martingale $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$, under what conditions, is it true that there exists a random variable X satisfying

$$X_n = E[X|\mathcal{B}_n] \ P\text{-a.s..}$$

This means the martingale $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is closed.

Definition 12.59. A family of random variable's $\{X_t\}_{t\in T}$ is unformly integrable if

$$\lim_{b\to\infty}\sup_{t\in I} E\left[|X_t|\mathbb{1}_{\{|X_t|>b\}}\right] = 0.$$

Clearly, this means $X_t \in L_1, \forall t \in T$.

Recall the fact: If $X_n \xrightarrow{\text{a.s.}} X_\infty$, and $\{X_n\}_{n\in\mathbb{N}}$ is u.i., then $X_n \xrightarrow{L_1} X_\infty$.

Proposition 12.60. Let $X \in L_1$, and let $\mathcal{K}$ be collections of subfields of $\mathcal{B}$. Then $\{E[X|\mathcal{G}]\}_{\mathcal{G}\in\mathcal{K}}$ is u.i..

Proof. Fix $b > 0$ and $\mathcal{G} \in \mathcal{K}$ and let $c > 0$, by Markov's inequality,

$$\begin{aligned}
\int_{\{|E[X|\mathcal{G}]|>b\}} |E[X|\mathcal{G}]|dP &\leqslant \int_{\{E[|X||\mathcal{G}]>b\}} E\left[|X||\mathcal{G}\right]dP \\
&\overset{def}{=} \int_{\{E[|X||\mathcal{G}]>b\}} |X|dP \\
&= \int_{\{E[|X||\mathcal{G}]>b\}\cap\{|X|\leqslant c\}} |X|dP + \int_{\{E[|X||\mathcal{G}]>b\}\cap\{|X|>c\}} |X|dP \\
&\leqslant cP\left(E\left[|X||\mathcal{G}\right] > b\right) + \int_{\{|X|>c\}} |X|dP \\
&\leqslant \frac{K}{b} E\left[E\left[|X||\mathcal{G}\right]\right] + \int_{\{|X|>c\}} |X|dP \\
&= \frac{k}{b} E\left[|X|\right] + \int_{\{|X|>c\}} |X|dP,
\end{aligned}$$

that is,

$$\lim_{b\to\infty}\sup\sup_{\mathcal{G}} \int_{\{|E[X|\mathcal{G}]|>b\}} |E[X|\mathcal{G}]|dP \leqslant \int_{\{|X|>c\}} |X|dP.$$

Since $c > 0$ is arbitrary chosen, letting $c \to \infty$ by DCT, proving the claim. ☐

Proposition 12.61 (Uniformly integrability Martingales)**.** Suppose that $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a martingale. The following are equivalent:

(a) $\{X_n\}$ is L_1-convergent.

(b) $\sup_{n\in\mathbb{N}} E\left[|X_n|\right] < \infty$, and there exists a random variable X_∞ such that $X_n \xrightarrow{\text{a.s.}} X_\infty$ which satisfyies $X_n = E[X_\infty|\mathcal{B}_n]$ for any $n \in \mathbb{N}$.

(c) $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is closed, that is, $\exists X \in L_1$ such that $X_n = E[X|\mathcal{B}_n]$ for any $n \in \mathbb{N}$.

(d) The sequence $\{X_n\}_{n\in\mathbb{N}}$ is u.i..

If one of the statements $(a) - (d)$ is true, the martingale is called regular or closable.

Proof. "(a) $\Longrightarrow$ (b)". Since X_n is L_1-convergent, $\lim_{n\to\infty} E(|X_n|)$ exists. So $E[|X_n|] < \infty$ for any $n \in \mathbb{N}$. Since the sequence is bounded and its limit exists, it is uniformly bounded, i.e.,

$$\sup_{n\in\mathbb{N}} E[|X_n|] < \infty.$$

By the martingale convergence theorem, $X_n \xrightarrow{\text{a.s.}} X_\infty$. Finllay, for any $n \in \mathbb{N}$, we have for any $p > n$,

$$\begin{aligned} E[|X_n - E[X_\infty|\mathcal{B}_n]|] &= E[|E[X_p|\mathcal{B}_n] - E[X_\infty|\mathcal{B}_n]|] \\ &= E[|E[X_p - X_\infty|\mathcal{B}_n]|] \\ &\leqslant E[E[|X_p - X_\infty||\mathcal{B}_n]] \\ &= E[|X_p - X_\infty|]. \end{aligned}$$

Since $p > n$ is arbitraryliy chosen, letting $p \to \infty$, we have

$$E\big[|X_n - E[X_\infty|\mathcal{B}_n]|\big] = 0, \forall n \in \mathbb{N}.$$

$$X_n = E[X_\infty|\mathcal{B}_n], \ P\text{-a.s.}, \forall n \in \mathbb{N}.$$

"(b) $\Longrightarrow$ (c)". The random variable variable $X = X_\infty$ serves the purpose and note (by trangale inequality)

$$E[|X_\infty|] = E\left[\lim_{n\to\infty}\inf|X_n|\right] \leqslant \lim_{n\to\infty}\inf E[|X_n|] \leqslant \sup_{n\in\mathbb{N}} E[|X_n|] < \infty.$$

"(c) $\Longrightarrow$ (d)". The family $\{E[X|\mathcal{B}_n]\}_{n\in\mathbb{N}}$ is u.i..

"(d) $\Longrightarrow$ (a)". If $\{X_n\}_{n\in\mathbb{N}}$ is u.i., $\sup_{n\in\mathbb{N}} E[|X_n|] < \infty$ by the characterization of uniform integrability. By the martingale convergence theorem, $X_n \xrightarrow{\text{a.s.}} X_\infty$. Therefore, $X_n \xrightarrow{L_1} X_\infty$. ☐

12.11 Regularity and Stopping

We now discuss when a stopped martingale retains the martingale characteristics.

Theorem 12.62. *Let $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ be a regular martingale.*

(a) If ν is a stopping time, then $X_\nu \in L_1$.

(b) If ν_1 and ν_2 are stopping times and $\nu_1 \leqslant \nu_2$, then

$$E[X_{\nu_2}|\mathcal{B}_{\nu_1}] = X_{\nu_1}, \ P\text{-a.s.}.$$

$$E[X_{\nu_2}] = E[X_{\nu_1}] = E[X_0].$$

Proof. (a) Let $\nu : \Omega \to N \cup \{\infty\}$ be a stopping time. The martingale is assumed regular so that there exists a random variable X such that $X_n = E[X|\mathcal{B}_n]$ for any $\in \mathbb{N}$. Moreover, $X = X_\infty$, and

$$X_n \xrightarrow{\text{a.s.}} X_\infty,$$

$$X_n \xrightarrow{L_1} X_\infty.$$

Then

$$E[X_\infty|\mathcal{B}_\nu] = \sum_{n \in \overline{\mathbb{N}}} E[X_\infty|\mathcal{B}_n]\mathbb{1}_{\{\nu=n\}} = \sum_{n \in \overline{\mathbb{N}}} X_n \mathbb{1}_{\{\nu=n\}} = X_\nu.$$

Since $X_\infty \in L_1$,

$$E\left[|X_\nu|\right] = E\left[|E[X_\infty|\mathcal{B}_\nu]|\right] \leqslant E\left[E\left[|X_\infty||\mathcal{B}_\nu\right]\right] = E\left[|X_\infty|\right] < \infty.$$

(b) If $\nu_1 \leqslant \nu_2$, then $\mathcal{B}_{\nu_1} \subseteq \mathcal{B}_{\nu_2}$. From (a), we have for any stopping time ν, $X_\nu = E[X_\infty|\mathcal{B}_\nu]$. Then

$$E[X_{\nu_2}|\mathcal{B}_{\nu_1}] = E\left[E\left[X_\infty|\mathcal{B}_{\nu_2}\right]|\,\mathcal{B}_{\nu_1}\right] = E[X_\infty|\mathcal{B}_{\nu_1}] = X_{\nu_1}.$$ □

Remark. For regular martingales, random stopping preserves fairness and for a stoppping fime ν, we have $E[X_\nu] = E[X_0]$ since we may take $\nu = \nu_2$ and $\nu_1 = 0$.

Remark. By Crystal Ball condtition, if $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a martingale and L_p bounded, i.e.,

$$\sup_n E\left[|X_n|^p\right] < \infty,\ p > 1,$$

then $\{X_n\}$ u.i. and hence regular.

12.12 Stopping Theorems

We now examine more flxible condtions for a stopped martingale to retain martingale characteristics. In order for this to be the case, either one must impose conditions on the sequence (such as the ui condition) or the on the stopping time or both.

Definition 12.63. A stopping time ν is regular for a martingale $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ if $\{(X_{\nu\wedge n}, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a regular martingale since we've shown $\{(X_{\nu\wedge n}, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a martingale.

Proposition 12.64 (Regularity)**.** Let $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ be a martingale and suppose ν is a stopping time, then ν is regular for $\{(X_n, \mathcal{B}_n)\}$ if and only if the following 3 conditions hold.

(a) $X_\infty := \lim_{n\to\infty} X_n$ exists a.s. on $\{\nu = \infty\}$. This means $\lim_{n\to\infty} X_{\nu\wedge n}$ exists a.s. on Ω.

(b) $X_\nu \in L_1$. (Note from (i), we know X_ν is defined a.s. on Ω.)

(c) $X_{\nu\wedge n} = E[X_\nu|\mathcal{B}_n]$ for any $n \in \mathbb{N}$.

Proof. Suppose ν is regular for $\{(X_n, \mathcal{B}_n)\}$. Then $\{(Y_n = X_{\nu\wedge n}, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a regular martingale.

(a) There exists a random variable Y_∞ such that

$$Y_n \xrightarrow{\text{a.s.}} Y_\infty.$$

$$Y_n \xrightarrow{L_1} Y_\infty.$$

On the set $\{\nu = \infty\}$, $Y_n = X_{\nu\wedge n} = X_n$, and so $\lim_{n\to\infty} X_{\nu\wedge n}$ exists a.s. on Ω.

(b) Note

$$X_\nu = \lim_{n\to\infty} X_{\nu\wedge n} = \lim_{n\to\infty} Y_n = Y_\infty \in L_1,\ P\text{-a.s.}.$$

(c) By (ii), $X_\nu = Y_\infty$, so

$$E[X_\nu|\mathcal{B}_n] = E[Y_\infty|\mathcal{B}_n] = Y_n = X_{\nu\wedge n},\ P\text{-a.s.}.$$

Next, suppose (i),(ii),(iii) hold. By (i), X_ν is defined a.s. on Ω. By (ii), $X_\nu \in L_1$. By(iii), $X_{\nu\wedge n} = E[X_\nu|\mathcal{B}_n]$. So we get X_ν is a closing random variable for the martingale $\{(X_{\nu\wedge n}, \mathcal{B}_n)\}_{n\in\mathbb{N}}$. Thus, $\{(X_{\nu\wedge n}, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a regular martingale. □

Here are two circumstances which guarantee that ν is regular.

(a) If $\nu \leqslant M$ a.s., then ν is regular since

$$|X_{\nu\wedge n}| \leqslant \sup_{m\in\mathbb{N}} |X_{\nu\wedge m}| = \sup_{m\leqslant M} |X_m| \in L_1.$$

Recall that domination by an integrable random variable is sufficient for uniform integrability. Or it suffices to show $X_{\nu\wedge n}$ is u.i.. By BCT in terms of b,

$$\lim_{b\to\infty} \sup_{n\in\mathbb{N}} E\left[|X_{\nu\wedge n}|\mathbb{1}_{\{|X_{\nu\wedge n}|>b\}}\right] \leqslant \lim_{b\to\infty} E\left[\max_{0\leqslant n\leqslant M} |X_n|\mathbb{1}_{\{\max_{0\leqslant n\leqslant M}|X_n|>b\}}\right] = 0.$$

(b) If $\{X_n\}$, then any stopping time ν is regular. (See the Corollary below.)

Theorem 12.65. *If ν is regular and $\nu_1 \leqslant \nu_2 \leqslant \nu$ for stopping time ν_1 and ν_2, then X_{ν_1} and X_{ν_2} exists, $X_{\nu_1} \in L_1$, $X_{\nu_2} \in L_1$ and*

$$E[X_{\nu_2}|\mathcal{B}_{\nu_1}] = X_{\nu_1},\ P\text{-a.s.}.$$

Proof. Define $Y_n = X_{\nu\wedge n}$. So $\{(Y_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a regular martingale. Then by previous theorem, $Y_{\nu_1}, Y_{\nu_2} \in L_1$, and $E[Y_{\nu_2}|\mathcal{B}_{\nu_1}] = Y_{\nu_1},\ P$-a.s.. Also,

$$L_1 \ni Y_{\nu_1} = X_{\nu_1\wedge\nu} = X_{\nu_1},$$

$$L_1 \ni Y_{\nu_2} = X_{\nu_2\wedge\nu} = X_{\nu_2}.$$

So $E[X_{\nu_2}|\mathcal{B}_{\nu_1}] = X_{\nu_1},\ P$-a.s.. □

Remark. Suppose ν is regular, $\nu_1 = 0$ and $\nu_2 = \nu$. Then $E[X_\nu|\mathcal{B}_0] = X_0$ and $E[X_\nu] = E[X_0]$.

Corollary 12.66. (a) Suppose ν_1 and ν_2 are stopping times and $\nu_1 \leqslant \nu_2$. If ν_2 is regular for the martingale $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$, so is ν_1.

(b) If $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a regular martingale, every stopping time ν is regular.

Proof. (b) follows from (a). Set $\nu_1 = \nu$ and $\nu_2 = \infty$. Then $X_{\nu_2\wedge n} = X_n$. So $\{(X_{\nu_2\wedge n}, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a regular martingale. Then ν_2 is regular. By (a), ν_1 is regular. Next, we prove (a). Assume ν_1 and ν_2 are stopping time, where ν_2 is regular. It suffices to show $\{(X_{\nu_1\wedge n}, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a regular martingale. It is enough to show $\{X_{\nu_1\wedge n}\}_{n\in\mathbb{N}}$ is u.i.. Fix $b > 0$ and $n \in \mathbb{N}$, note

$$\begin{aligned}
\int_{\{|X_{\nu_1\wedge n}|>b\}} |X_{\nu_1\wedge n}|dP &= \int_{\{|X_{\nu_1\wedge n}|>b,\nu_1\leqslant n\}} |X_{\nu_1\wedge n}|dP + \int_{\{|X_{\nu_1\wedge n}|>b,\nu_1>n\}} |X_{\nu_1\wedge n}|dP \\
&= \int_{\{|X_{\nu_1}|>b,\nu_1\leqslant n\}} |X_{\nu_1}|dP + \int_{\{|X_n|>b,\nu_1>n\}} |X_n|dP \\
&\leqslant \int_{\{|X_{\nu_1}|>b\}} |X_{\nu_1}|dP + \int_{\{|X_n|>b,\nu_2>n\}} |X_n|dP \\
&\leqslant \int_{\{|X_{\nu_1}|>b\}} |X_{\nu_1}|dP + \int_{\{|X_n\wedge n|>b,\nu_2>n\}} |X_{\nu_2\wedge n}|dP \\
&\leqslant \int_{\{|X_{\nu_1}|>b\}} |X_{\nu_1}|dP + \int_{\{|X_n\wedge n|>b\}} |X_{\nu_2\wedge n}|dP.
\end{aligned}$$

Since $X_{\nu_1} \in L_1$ by previous theorem, and $\{X_{\nu_2\wedge n}\}_{n\in\mathbb{N}}$ is u.i.,

$$\begin{aligned}
\sup_{n\in\mathbb{N}} \int_{\{|X_{\nu_1\wedge n}|>b\}} |X_{\nu_1\wedge n}|dP &\leqslant \int_{\{|X_{\nu_1}|>b\}} |X_{\nu_1}|dP + \sup_{n\in\mathbb{Z}^+} \int_{\{|X_{\nu_2\wedge n}|>b\}} |X_{\nu_2\wedge n}|dP \\
&\to 0 \text{ as } b \to \infty.
\end{aligned}$$

□

Theorem 12.67. *In order for the stopping time ν to be regular for the martingale $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$, it is necessary and sufficient that*

(a) $\int_{\{\nu<\infty\}} |X_\nu|dP < \infty$, *and*

(b) $\{X_n \mathbb{1}_{\{\nu>n\}}\}_{n\in\mathbb{N}}$ *is u.i..*

Proof. Sufficiency: It suffices to show $\{X_{\nu\wedge n}\}_{n\in\mathbb{N}}$ is u.i.. Fix $b > 0$ and $n \in \mathbb{N}$, note

$$\begin{aligned}
\int_{\{|X_{\nu\wedge n}|>b\}} |X_{\nu\wedge n}|dP &= \int_{\{|X_{\nu\wedge n}|>b,\nu\leqslant n\}} |X_{\nu\wedge n}|dP + \int_{\{|X_{\nu\wedge n}|>b,\nu>n\}} |X_{\nu\wedge n}|dP \\
&= \int_{\{|X_\nu|>b,\nu\leqslant n\}} |X_\nu|dP + \int_{\{|X_n|>b,\nu>n\}} |X_n|\mathbb{1}_{\{\nu>n\}}dP \\
&\leqslant \int_{\{|X_\nu|>b,\nu<\infty\}} |X_\nu|\mathbb{1}_{\{\nu<\infty\}}dP + \int_{\{|X_n\mathbb{1}_{\{\nu>n\}}|>b\}} \left|X_n\mathbb{1}_{\{\nu>n\}}\right|dP \\
&= \int_{\{|X_\nu\mathbb{1}_{\{\nu<\infty\}}|>b\}} \left|X_\nu\mathbb{1}_{\{\nu<\infty\}}\right|dP + \int_{\{|X_n\mathbb{1}_{\{\nu>n\}}|>b\}} \left|X_n\mathbb{1}_{\{\nu>n\}}\right|dP.
\end{aligned}$$

Since $X_\nu\mathbb{1}_{\{\nu<\infty\}} \in L_1$ by (a), and $\{X_n\mathbb{1}_{\{\nu>n\}}\}_{n\in\mathbb{N}}$ is u.i. by (b),

$$\begin{aligned}
\sup_{n\in\mathbb{N}} \int_{\{|X_{\nu\wedge n}|>b\}} |X_{\nu\wedge n}|dP &\leqslant \int_{\{|X_\nu\mathbb{1}_{\{\nu<\infty\}}|>b\}} \left|X_\nu\mathbb{1}_{\{\nu<\infty\}}\right|dP + \sup_{n\in\mathbb{N}} \int_{\{|X_n\mathbb{1}_{\{\nu>n\}}|>b\}} \left|X_n\mathbb{1}_{\{\nu>n\}}\right|dP \\
&\to 0, \text{ as } b \to \infty.
\end{aligned}$$

Necessity: Suppose ν is regular.

(a) Since $\{X_{\nu\wedge n}\}_{n\in\mathbb{N}}$ is u.i.,

$$\int_{\{\nu<\infty\}}|X_\nu|dP=\lim_{n\to\infty}\int_{\{\nu\leqslant n\}}|X_\nu|dP=\lim_{n\to\infty}\int_{\{\nu\leqslant n\}}|X_{\nu\wedge n}|dP\leqslant\sup_{n\in\mathbb{N}}E\left[X_{\nu\wedge n}\right]<\infty.$$

(b) Since $\{X_{\{}\nu\wedge n\}\}_{n\in\mathbb{N}}$ is u.i.,

$$\begin{aligned}\sup_{n\in\mathbb{N}}\int_{\{|X_n\mathbb{1}_{\{\nu>n\}}|>b\}}\left|X_n\mathbb{1}_{\{\nu>n\}}\right|dP&=\sup_{n\in\mathbb{N}}\int_{\{|X_{\nu\wedge n}\mathbb{1}_{\{\nu>n\}}|>b\}}\left|X_{\nu\wedge n}\mathbb{1}_{\{\nu>n\}}\right|dP\\&\leqslant\sup_{n\in\mathbb{N}}\int_{\{|X_{\nu\wedge n}|>b\}}|X_{\nu\wedge n}|dP\\&\to 0,\ \text{as } b\to\infty.\end{aligned}$$

□

Remark. Question: is there a simple sufficient condition for (a) to hold? Fact: If $\{X_n\}_{n\geqslant 0}$ is L_1-bounded, we have that

$$\sup_{n\in\mathbb{N}}E\left[|X_n|\right]<\infty.$$

Proof. By the martingale convergence theorem, $X_n\xrightarrow{\text{a.s.}}X_\infty$ as $n\to\infty$. Thus, X_ν is defined a.e.. We claim $X_\nu\in L_1$ and then $X_\nu\mathbb{1}_{\{\nu<\infty\}}\in L_1$. To verify the claim, observe that $X_{\nu\wedge n}\xrightarrow{\text{a.s.}}X_\nu$, and so by Fatou's lemma

$$E\left[|X_\nu|\right]=E\left[\lim_{n\to\infty}|X_{\nu\wedge n}|\right]\leqslant\lim_{n\to\infty}\inf E\left[|X_{\nu\wedge n}|\right].$$

Then for any $n\in\mathbb{N}$,

$$E[X_n|\mathcal{B}_{\nu\wedge n}]=\sum_{j=0}^{n}E[X_n|\mathcal{B}_j]\mathbb{1}_{\{v\wedge n=j\}}\overset{a.s.}{=}\sum_{j=0}^{n}X_j\mathbb{1}_{\{v\wedge n=j\}}=X_{\nu\wedge n}.$$

Thus,

$$E\left(|X_{\nu\wedge n}|\right)\leqslant E\left[|E\left[X_n|\mathcal{B}_{\nu\wedge n}\right]|\right]\leqslant E\left[E\left[|X_n||\mathcal{B}_{\nu\wedge n}\right]\right]=E\left[|X_n|\right].$$

Thus,

$$E\left[|X_v|\right]\leqslant\lim_{n\to\infty}\inf E\left[|X_{\nu\wedge n}|\right]\leqslant\lim_{n\to\infty}E\left[|X_n|\right]\leqslant\sup_{n\in\mathbb{N}}E\left[|X_n|\right]<\infty.$$

□

Remark. If the martingale $\{X_n\}_{n\in\mathbb{N}}$ is non-negative, then it is automatically L_1-bounded since

$$\sup_{n\in\mathbb{N}}E\left[|X_n|\right]=\sup_{n\in\mathbb{N}}E[X_n]=E[X_0].$$

Corollary 12.68. Let $\{(X_n,\mathcal{B}_n),n\geqslant 0\}$ be an L_1-bounded martingale.

(a) For any level $a>0$, the escape time

$$\nu_a=\inf\{n\geqslant 0:|X_n|>a\}$$

is regular. In particular, this holds if $\{X_n\}$ is a positive martingale.

(b) For $a, b \in \mathbb{R}$ and $b < 0 < a$,

$$\nu_{a,b} = \inf\{n \geqslant 0 : X_n > a \text{ or } X_n < b\}$$

is regular.

Proof. Since $\{X_n\}$ is L_1 bounded and $\nu_a, \nu_{a,b}$ are stopping time, we have

$$\int_{\{\nu_a < \infty\}} |X_{\nu_a}| dP < \infty,$$

$$\int_{\{\nu_{a,b} < \infty\}} |X_{\nu_{a,b}}| dP < \infty.$$

Then it suffices to show that $\{X_n \mathbb{1}_{\{\nu_a > n\}}\}$ and $\{X_n \mathbb{1}_{\{\nu_{a,b} > n\}}\}$ are both u.i.. Observe that

$$\left|X_n \mathbb{1}_{\{\nu_a > n\}}\right| = |X_n| \mathbb{1}_{\{\nu_a > n\}} \leqslant a \mathbb{1}_{\{\nu_a > n\}} \leqslant a,$$

and

$$\left|X_n \mathbb{1}_{\nu_{a,b} > n}\right| = |X_n| \mathbb{1}_{\{\nu_{a,b} > n\}} \leqslant \max\{|a|, |b|\} \mathbb{1}_{\{\nu_{a,b} > n\}} \leqslant \max\{|a|, |b|\}. \qquad \square$$

Proposition 12.69. Suppose $\{(X_n, \mathcal{B}_n)\}_{n \geqslant 0}$ is a martingale. Then

(a) ν is regular for $\{(X_n, \mathcal{B}_n)\}_{n \geqslant 0}$ and

(b) $X_n \xrightarrow{\text{a.s.}} 0$ on $\{\nu = \infty\}$

is equivalent to

(a)' $\int_{\{\nu < \infty\}} |X_\nu| dP < \infty$ and

(b)' $\lim_{n \to \infty} \int_{\{\nu > n\}} |X_n| dP = 0$.

Proof. Assume (a) and (b) hold. Then $\{(X_{\nu \wedge, \mathcal{B}_n})\}_{n \in \mathbb{N}}$ is a regular martingale, and hence

$$X_{\nu \wedge n} \xrightarrow{\text{a.s.}} X_\nu, \ X_{\nu \wedge n} \xrightarrow{L_1} X_\nu.$$

From (b), $X_\nu = 0$ on $\{\nu = \infty\}$. Then

$$\int_{\{\nu < \infty\}} |X_\nu| dP = \int_\Omega |X_\nu| < \infty.$$

Claim. If $X_n \xrightarrow{L_1} X$ and $Y_n \xrightarrow{\text{a.s.}} Y$, then

$$\begin{aligned}
|E[X_n Y_n] - E[XY]| &\leqslant |E[X_n Y_n] - E[X Y_n]| + |E[X Y_n] - E[XY]| \\
&= E\left[|Y_n||X_n - X|\right] + E\left[|X||Y_n - Y|\right] \\
&\to 0.
\end{aligned}$$

Then

$$\int_{\{\nu > n\}} |X_n| dP = \int_\Omega |X_{\nu \wedge n}| \mathbb{1}_{\{\nu > n\}} dP \to \int_{\{\nu = \infty\}} |X_\nu| dP = 0.$$

Assume (a)' and (b)' hold. To prove (a), it suffices to show $\{X_n \mathbb{1}_{\{\nu>n\}}\}_{n\geqslant 0}$ is u.i.. Fix $\epsilon > 0$, by (b)', there exists n_0 such that when $n \geqslant n_0$,

$$\int_{\{\nu>n\}} |X_n| dP \leqslant \epsilon.$$

Then for $b > 0$,

$$\begin{aligned}
&\sup_{n\geqslant 0} \int_{\{|X_n|\mathbb{1}_{\{\nu>n\}}\}} |X_n| dP \\
&= \max\left\{ \max_{0\leqslant n\leqslant n_0-1} \int_{\{|X_n|>b,\nu>n\}} |X_n| dP, \sup_{n\geqslant n_0} \int_{\{|X_n|>b,\nu>n\}} |X_n| dP \right\} \\
&\leqslant \max\left\{ \max_{0\leqslant n\leqslant n_0-1} \int_{\{|X_n|>b,\nu>n\}} |X_n| dP, \epsilon \right\}.
\end{aligned}$$

Therefore,

$$\limsup_{b\to\infty} \sup_{n\in\mathbb{N}} \int_{\{|X_n \mathbb{1}_{\{\nu>n\}}|>b\}} \left|X_n \mathbb{1}_{\{\nu>n\}}\right| dP \leqslant \epsilon.$$

So $\{X_n \mathbb{1}_{\{\nu>n\}}\}_{n\geqslant 0}$ is u.i.. Thus, ν is regular. Since ν is regular, X_ν is defined on Ω and

$$X_{\nu\wedge n} \xrightarrow{\text{a.s.}} X_\nu, \text{ and } X_{\nu\wedge n} \xrightarrow{L_1} X_\nu.$$

Then

$$0 = \lim_{n\to\infty} \int_{\nu>n} |X_n| dP = \lim_{n\to\infty} \int_{\nu>n} |X_{\nu\wedge n}| dP = \int_{\nu=\infty} |X_\nu| dP.$$

So $X_\nu \mathbb{1}_{\{\nu=\infty\}} = 0$ a.s.. That is, $X_n \to 0$ on $\{\nu = \infty\}$. □

12.13 Wald's identity and random walks

This section discuss a martingale approach to some facts about the random walk. Consider a sequence of iid random variables $\{Y_n\}_{n\in\mathbb{Z}^+}$ which are not a.s. constant and define the random walk $\{X_n\}_{n\in\mathbb{N}}$ by

$$X_0 = 0, \ X_n = \sum_{i=1}^n Y_i, \ n \in \mathbb{Z}^+,$$

with associated σ-fields

$$\mathcal{B}_0 = \{\emptyset, \Omega\}, \ \mathcal{B}_n = \sigma(Y_1, \ldots, Y_n) = \sigma(X_0, \ldots, X_n), \ n \in \mathbb{Z}^+.$$

Definition 12.70. The cumulant generating function of Y_1 is a funtion

$$\begin{aligned}
\phi : \mathbb{R} &\to \overline{\mathbb{R}} \\
u &\mapsto \log\left(E\left[e^{uY_1}\right]\right).
\end{aligned}$$

Facts about ϕ:

(a) ϕ is convex on $\mathbb{R}$.

Proof. Let $u_1, u_2 \in \mathbb{R}$, and fix $\alpha \in (0,1)$. By Holder's inequality,

$$\begin{aligned}\phi(\alpha u_1 + (1-\alpha)u_2) &= \log E\left[e^{\alpha u_1 Y_1} e^{(1-\alpha)u_2 Y_2}\right] \\ &\leqslant \log\left[E\left[e^{u_1 Y_1}\right]\right]^{\alpha}\left[E\left[e^{u_2 Y_1}\right]\right]^{1-\alpha} \\ &= \alpha\phi(u_1) + (1-\alpha)\phi(u_2).\end{aligned}$$

In fact, we can show that ϕ is strictly convex on $\{\phi < \infty\}$. Hence, on $\{\phi < \infty\}$, ϕ' is strictly increasing. □

(b) $\{\phi < \infty\}$ is an interval containing 0. (This interval might be $[0,0] = [0]$, as would be the case if Y_1 were Cauchy distributed.)

Proof. Clearly, $\phi(0) = 0 < \infty$. So $0 \in \{\phi < \infty\}$. Fix $u_1, u_2 \in \{\phi < \infty\}$, then for $\alpha \in [0,1]$,

$$\phi(\alpha u_1 + (1-\alpha)u_2) \leqslant \alpha\phi(u_1) + (1-\alpha)\phi(u_2) < \infty.$$

Then $\alpha u_1 + (1-\alpha)u_2 \in \{\phi < \infty\}$. Thus, $[u_1, u_2] \in \{\phi < \infty\}$. □

(c) If the interior of $\{\phi < \infty\}$ is non-empty, ϕ is analytic there, hence infinitely differentiable, and

$$\phi'(u) = E\left[Y_1 e^{uY_1 - \phi(u)}\right].$$

Hence, $\phi'(0) = E[Y_1]$. One may also check that $\phi''(0) = \mathrm{Var}(Y_1)$.

12.13.1 The basic Martingales

Here is a basic connection between martingales and the random walk.

Proposition 12.71. For any $u \in \{\phi < \infty\}$, define

$$M_n(u) = e^{uX_n - n\phi(u)} = e^{uX_n} e^{-n\log E\left[e^{uY_1}\right]} = \frac{e^{uX_n}}{\left[E\left[e^{uY_1}\right]\right]^n}.$$

Then $\{(M_n(u), \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a positive martingale with $E[M_n(u)] = 1$. Since $M_n(u) > 0$, $M_n(u) \in L_1$. Also,

$$M_n(u) \xrightarrow{\text{a.s.}} 0 \text{ as } n \to \infty.$$

Hence $\{M_n(u)\}$ is a non-regular martingale.

Proof. Clearly, $M_n(u) \in \mathcal{B}_n$.

$$E\left[M_{n+1}(u)|\mathcal{B}_n\right] = E\left[M_n(u)\frac{e^{uY_{n+1}}}{E\left[e^{uY_1}\right]}\middle|\mathcal{B}_n\right] = M_n(\mu).$$

Since $0, u \in \{\phi < \infty\}$, we have $\frac{u}{2} \in \{\phi < \infty\}$. Then

$$\phi(u/2) = \phi(u/2 + 0/2) < 1/2\phi(u) + 1/2\phi(0) = 1/2\phi(u).$$

Also, $\{M_n(u/2)\}_{n\in\mathbb{N}}$ is a positive martingale and then L_1-bounded, so there exists a random variable Z such that

$$M_n(u/2) = e^{u/2X_n - n\phi(u/2)} \overset{a.s.}{\to} Z < \infty,$$

and by continuity,

$$M_n^2(u/2) = e^{uX_n - 2n\phi(u/2)} \overset{a.s.}{\to} Z^2 < \infty.$$

Therefore,

$$\begin{aligned} M_n(u) = e^{uX_n - n\phi(u)} &= e^{uX_n - 2n\phi(u/2) + n[2\phi(u/2) - \phi(u)]} \\ &= e^{2[u/2X_n - n\phi(u/2)]} e^{-2n[1/2\phi(u) - \phi(u/2)]} \\ &= M_n^2(u/2) e^{-2n[1/2\phi(u) - \phi(u/2)]} \\ &\to Z^2 \cdot 0 = 0. \end{aligned}$$

□

12.13.2 Regular stopping times

We will call the martingale $\{(M_n(u), \mathcal{B}_n)\}_{n\in\mathbb{N}}$, where

$$M_n(u) = e^{uX_n - u\phi(u)} = \frac{e^{uX_n}}{[E[e^{uY_1}]]^n}$$

the exponential martingale. Recall that if $u \neq 0$ and $u \in \{\phi < \infty\}$, then

$$M_n(u) \xrightarrow{\text{a.s.}} 0.$$

Here is Wald's identity for the exponential martingale.

Proposition 12.72 (Wald's Identity)**.** Let $u \in \{\phi < \infty\}$ and suppose $\phi'(u) \geqslant 0$. Then for $a > 0$,

$$\nu_a^+ := \inf\{n > 0 : X_n \geqslant a\}$$

is regular for the martingale $\{(M_n(u), \mathcal{B}_n)\}_{n\in\mathbb{N}}$. Consequently, any stopping time $\nu \leqslant \nu_a^+$ is regular and hence Wald's identity holds

$$1 = E[M_0(u)] = E[M_\nu(u)] = \int_\Omega e^{uX_\nu - \nu\phi(u)} dP = \int_{\{\nu<\infty\}} e^{uX_\nu - \nu\phi(u)} dP.$$

Proof. To show that ν_a^+ is regular, since $M_n(u) \xrightarrow{\text{a.s.}} 0$, by the equivalence proposition, it suffices to show that

(a)

$$\int_{\{\nu_a^+<\infty\}} M_{\nu_a^+}(u) dP < \infty,$$

(b)

$$\lim_{n\to\infty} \int_{\{\nu_a^+>n\}} M_n(u) dP = 0.$$

(a) is true since $\{(M_n(u), \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is L_1-bounded for $\sup_{n\in\mathbb{N}} E[|M_n(u)|] = 1$. It remains to check

$$\lim_{n\to\infty}\int_{\{\nu_a^+>n\}} M_n(u)dP = \lim_{n\to\infty}\int_{\{\nu_a^+>n\}} e^{uX_n-n\phi(u)}dP = 0.$$

We need the following random walk fact. Let $\{Y_i\}_{i\in\mathbb{Z}^+}$ be iid with d.f. F and $E[Y_i] \geqslant 0$. After defining

$$X_n = \sum_{i=1}^n Y_i,$$

we have $\limsup_{n\to\infty} X_n = +\infty$. If $E[Y_1] > 0$, then by the SLLN,

$$\frac{X_n}{n} \xrightarrow{\text{a.s.}} E[Y_1] > 0 \text{ as } n\to\infty.$$

So $X_n \xrightarrow{\text{a.s.}} \infty$. It is still true that if $E[Y_1] = 0$ but one must use standard random walk theory as discussed in, for example, Chung (1974), Feller (1971), Resnick (1994). Also

$$\lim\inf_{n\to\infty} X_n = -\infty,\ P\text{-a.s.}.$$

Since

$$\{\lim\sup_{n\to\infty} X_n = \infty\} \subseteq \{\nu_a^+ < \infty\}, \forall a > 0,$$

we have $P(\nu_a^+ < \infty) = 1$. Thus, $\lim_{n\to\infty} P(\nu_a^+ > n) = 0$. Exponentia tilting: Construct a probability space $(\Omega^\#, \mathcal{B}^\#, P^\#)$, and on the space, construct iid random variables $\{Y_i^\#\}_{i\in\mathbb{Z}^+}$ with distribution $F^\#$ defined by

$$F^\#(dy) = e^{uy-\phi(u)}F(dy).$$

Note $F^\#$ is a probability distribution since

$$F^\#(\mathbb{R}) = \int_{\mathbb{R}} e^{uy-\phi(u)}F(dy) = \int_\Omega e^{uY_1-\phi(u)}dP = E\left[e^{uY_1}/e^{\phi(u)}\right] = 1.$$

$F^\#$ is sometimes called the Esscher transform of F. Also,

$$E^\#[Y_1^\#] = \int_{\mathbb{R}} yF^\#(dy) = \int_{\mathbb{R}} ye^{uy-\phi(u)}F(dy) = \frac{E\left[Y_1e^{uY_1}\right]}{E\left[e^{uY_1}\right]} = \frac{d}{du}\log E\left[e^{uY_1}\right] = \phi'(u) \geqslant 0.$$

Finally,

$$\begin{aligned}\int_{\{\nu_a^+>n\}} e^{uX_n-n\phi(u)}dP &= \int_\Omega e^{uX_n-n\phi(u)}\mathbb{1}_{\{\nu_a^+>n\}}dP\\ &= \int\cdots\int_{(y_1,\ldots,y_n):\sum_{i=1}^k y_i<a, 1\leqslant k\leqslant n} \prod_{l=1}^n e^{uy_l-\phi(u)}F(dy_n)\cdots F(dy_1)\\ &= \int\cdots\int_{(y_1,\ldots,y_n):\sum_{i=1}^k y_i<a, 1\leqslant k\leqslant n} F^\#(dy_n)\cdots F^\#(dy_1)\\ &= P\left(\sum_{i=1}^k y_i^\# < a, j = 1,\ldots,n\right)\\ &= P^\#(\nu_a^{+\#} > n) \to 0 \text{ a.s. } n\to\infty,\end{aligned}$$

since $E^\#[Y_1^\#] \geqslant 0$. ☐

Corollary 12.73. Let $b < 0 < a$, $u \in \{\phi < \infty\}$ and define

$$\nu_{a,b} = \inf\{n \geqslant 0 : X_n \leqslant b \text{ or } X_n \geqslant a\}.$$

Then $\nu_{a,b}$ is regular for $\{(M_n(u), \mathcal{B}_n)\}_{n\in\mathbb{N}}$ and thus satisfies Wald's identity.

Proof. If $\phi'(u) \geqslant 0$, then ν_a^+ is regular for $\{(M_n(u), \mathcal{B}_n)\}_{n\in\mathbb{N}}$, and hence $\nu_{a,b} < \nu_a^+$ is regular. If $\phi'(u) \leqslant 0$, check the previous Proposition to convince yourself that

$$\nu_b^- := \inf\{n : X_n \leqslant b\}$$

is regular and hence $\nu_{a,b} \leqslant \nu_b^-$ is also regular. □

Example 12.74 (Skip free random walk)**.** Suppose the range of Y_1 is $\{1, 0, -1, -2, \cdots\}$ and

$$P(Y_1 = 1) \in (0, 1].$$

Then the random walk $\{X_n\}_{n\in\mathbb{N}}$ with steps $\{Y_j\}_{j\in\mathbb{Z}^+}$ is skip free positive random walk since it cannnot jump over states in the upward direction. Let $a \in \mathbb{Z}^+$, since $\{X_n\}_{n\in\mathbb{N}}$ is skip free positive,

$$X_{\nu_a^+} = a \text{ on } \{\nu_a^+ < \infty\}.$$

Next,

$$\phi(u) = \log E\left[e^{uY_1}\right] = \log\left(e^u P(Y_1 = 1) + \sum_{j=0}^{\infty} e^{-uj} P(Y_1 = -j)\right) < \infty, \forall u \in (0, \infty).$$

Also, since $P(Y_1 = 1) > 0$, $\lim_{u\to\infty} \phi(u) = \infty$. $\phi : [0, \infty) \to \mathbb{R}$ is strictly convex on $[0, \infty)$ and ϕ' strictly increasing on $[0, \infty)$. Then there exists $u^* \in [0, \infty)$ such that

$$\phi(u^*) = \inf_{u\geqslant 0} \phi(u),$$

u^* is in the parabola or straight line. Thus, $\phi'(u) \geqslant 0, \forall u \geqslant u^*$.

Goal: find the Laplace transform of ν_a^+. For $u \geqslant u^*$, since on $\{\nu_a^+ < \infty\}$, $X_{\nu_a^+} = a$, and Wald's identity says

$$\begin{aligned} 1 &= E\left[e^{uX_{\nu_a^+} - \nu_a^+\phi(u)} \mathbb{1}_{\{\nu_a^+ < \infty\}}\right] \\ &= \int_{\{\nu_a^+ < \infty\}} e^{uX_{\nu_a^+} - \nu_a^+\phi(u)} dP \\ &= e^{ua} \int_{\{\nu_a^+ < \infty\}} e^{-\nu_a^+\phi(u)} dP \\ &= e^{ua} E\left[e^{-\phi(u)\nu_a^+} \mathbb{1}_{\{\nu_a^+ < \infty\}}\right]. \end{aligned}$$

So

$$E\left[e^{-\phi(u)\nu_a^+} \mathbb{1}_{\{\nu_a^+ < \infty\}}\right] = e^{-ua}, \forall u \in \{\phi < \infty\}.$$

Thus,

$$E\left[e^{-\phi(u)\nu_a^+} \mathbb{1}_{\{\nu_a^+ < \infty\}}\right] = e^{-ua}, \forall u \in [u^*, \infty). \tag{12.1}$$

Consider the following cases.

(a) Suppose $\phi'(0) = E[Y_1] \geqslant 0$. Then (straight line) $u^* = 0$. Since $E[Y_1] \geqslant 0$, we have $\nu_a^+ < \infty$ a.s.. Then by 12.1,

$$E\left[e^{-\phi(u)\nu_a^+}\right] = e^{-ua}, \forall u \geqslant 0.$$

Setting $\lambda = \phi(u)$ gives

$$E\left[e^{-\lambda\nu_a^+}\right] = e^{-\phi^{\leftarrow}(\lambda)a}.$$

In this case, Wald's identity gives a formula for the Laplace transform of ν_a^+.

(b) Suppose $E[Y_1] = \phi'(0) < 0$. Since $\phi(0) = 0$, convexity requires $\phi(u^*) < 0$. So there exists ! $u_0 > u^* > 0$ such that $\phi(u_0) = 0$ (parabola). Thus if we substitude u_0 in 12.1, we get

$$e^{-ua} = E\left[e^{-\phi(u_0)\nu_a^+}\mathbb{1}_{\{\nu^+<\infty\}}\right] = E\left[\mathbb{1}_{\{\nu_a^+<\infty\}}\right] = P(\nu_a^+ < \infty) < 1.$$

In this case, Wald's identity gives a formula for $P(\nu_a^+ < \infty)$.

We now examine the following martingales:

$$\{X_n - nE[Y_1]\}_{n\in\mathbb{N}} \text{ and } \{(X_n - nE[Y_1])^2 - n\text{Var}(Y_1)\}_{n\in\mathbb{N}}.$$

Suppose $\{Y_k\}_{k\geqslant}$ is an iid random varibles. Define

$$X_0 = 0,\ X_n = \sum_{k=1}^{n} Y_k, \forall n \in \mathbb{Z}^+.$$

Define

$$\mathcal{B}_n = \sigma(X_1, \ldots, X_n), \forall n \in \mathbb{Z}^+.$$

Clearly, $\mathcal{B}_0 = \{\emptyset, \Omega\}$. If $E[|Y_1|] < \infty$, then $\{(X_n - nE[Y_1], \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a martingale. If $E[Y_1^2] < \infty$, then

$$\left\{\left((X_n - nE[Y_1])^2 - n\,\text{Var}(Y_1), \mathcal{B}_n\right)\right\}_{n\in\mathbb{N}}$$

is also a martingale. Neither is regular but we can find regular stopping times.

Proposition 12.75. Let ν be a stopping time which satisfies $E[\nu] < \infty$. Then

(a) ν is regular for $\{(X_n, nE[Y_1], \mathcal{B}_n)\}_{n\in\mathbb{N}}$ assuming $E[|Y_1|] < \infty$.

(b) ν is regular for $\{(X_n - nE[Y_1])^2 - n\,\text{Var}(Y_1), \mathcal{B}_n)\}_{n\in\mathbb{N}}$ assuming $E[Y_1^2] < \infty$.

Proof. (a) Since $\{X_n - nE[Y_1]\}$ has mean 0, wlog, we can assume that $E[Y_1] = 0$. If $E[\nu] < \infty$, $P(\nu < \infty) = 1$ and so $X_{\nu\wedge n} \xrightarrow{\text{a.s.}} X_\nu$. NTS:

$$X_{\nu\wedge n} \xrightarrow{L_1} X_\nu \text{ as } n \to \infty.$$

Note

$$\begin{aligned}|X_\nu - X_{\nu\wedge n}| &= \left|\sum_{k=1}^{\nu} Y_k - \sum_{k=1}^{\nu\wedge n} Y_k\right| = \left|\sum_{k=1}^{\infty} \mathbb{1}_{\{\nu\geqslant k\}} - \sum_{k=1}^{n} Y_k\mathbb{1}_{\{\nu\geqslant k\}}\right| \\ &= \left|\sum_{k=n+1}^{\infty} Y_k\mathbb{1}_{\{\nu\geqslant k\}}\right| \leqslant \sum_{k=n+1}^{\infty} |Y_k|\mathbb{1}_{\{\nu\geqslant k\}}.\end{aligned}$$

Define

$$\xi_n = \sum_{k=n}^{\infty} |Y_k| \mathbb{1}_{\{\nu \geqslant k\}}.$$

Then $\xi_n \leqslant \xi_1, \forall n \geqslant 1$, and since $\nu < \infty$ a.e., $\xi_n \to 0$ as $n \to \infty$. Since $\{\nu \geqslant j\} \in \mathcal{B}_{j-1}$,

$$Y_j \perp\!\!\!\perp \mathbb{1}_{\{\nu \geqslant j\}}.$$

Then

$$E[\xi_1] = E\left[\sum_{j=1}^{\infty} |Y_1| \mathbb{1}_{\{\nu \geqslant j\}}\right] = \sum_{j=1}^{\infty} E\left[|Y_1| \mathbb{1}_{\{\nu \geqslant j\}}\right] = \sum_{j=1}^{\infty} E\left[|Y_1|\right] P(\nu \geqslant j) = E\left[|Y_1|\right] E[\nu] < \infty.$$

By DCT, $\lim_{n\to\infty} E[\xi_n] = E[0] = 0$. So

$$E\left[|X_{\nu\wedge n} - X_\nu|\right] \leqslant E[\xi_{n+1}] \to 0 \text{ as } n \to \infty.$$

Thus, $X_{\nu\wedge n} \xrightarrow{L_1} X_\nu, \text{ as } n \to \infty.$ □

12.13.3 Examples of integrable stopping times

Previous Proposition has a hypothesis that the stopping time be integrable. In this subsection, we give sufficient conditions for first passage times and first escape times from strips to be integrable.

Proposition 12.76. Consider the random walk with steps $\{Y_j\}_{j\in\mathbb{Z}^+}$.

(a) If $E[Y_1] > 0$, then for $a > 0$,

$$\nu_a^+ := \inf\{n : X_n \geqslant a\} \in L_1.$$

(b) If $E[Y_1] < 0$, then for $b > 0$,

$$\nu_b^- = \inf\{n \geqslant 1 : X_n \leqslant -b\} \in L_1.$$

(c) If $E[Y_1] \neq 0$ amd $Y_1 \in L_1$, then

$$\nu_{a,b} := \inf\{n \geqslant 1 : X_n \leqslant -b \text{ or } X_n \geqslant a\} \subseteq L_1.$$

Proof. Suppose (a) holds. Define

$$Y_i' = -Y_i, \forall i \in \mathbb{N},$$

$$X_n' = \sum_{i=1}^{n} Y_k', n \in \mathbb{N}.$$

Then $E[Y_k'] > 0$, and

$$\nu_b'^{+} = \inf\{n \geqslant 1 : X_n' \geqslant b\} \in L_1.$$

Also,

$$\nu_b'^{+} = \inf\{n \geqslant 1 : X_n \leqslant -b\} = \nu_b^-.$$

So (b) follows from (a).

(c) follows from (a) since $\nu_{a,b} \leqslant \max\{\nu_a^+, \nu_b^-\}$. It remains to establish (a). Since $\{(X_n - nE[Y_1])\}_{n\in\mathbb{N}}$ is a martingale, then so is

$$\{(X_{\nu_a^+\wedge n} - (\nu_a^+ \wedge n)E[Y_1], \mathcal{B}_n)\}_{n\in\mathbb{N}}.$$

Hence, since

$$X_{\nu_a^+\wedge n},\ (\nu_a^+ \wedge n)E[Y_1] \in L_1,$$

we have

$$E\left[X_{\nu_a^+\wedge n} - (\nu_a^+ \wedge n)E[Y_1]\right] = 0.$$

Then

$$E\left[X_{\nu_a^+\wedge n}\right] = E[Y_1]E[\nu_a^+ \wedge n].$$

Since $\nu_a^+ \wedge n \uparrow \nu_a^+$, by MCT, $E\left[\nu_a^+ \wedge n\right] \uparrow E[\nu_a^+]$. Then we need a bound on $E\left[X_{\nu_a^+\wedge n}\right]$.

(a) Case 1: Suppose there exists $c \in \mathbb{R}$ such that $P(Y_1 \leqslant c) = 1$. On $\{\nu_a^+ < \infty\}$, we have $X_{\nu_a^+-1} < a$ and $Y_{\nu_a^+} \leqslant c$ so that

$$X_{\nu_a^+} = X_{\nu_a^+-1} + Y_{\nu_a^+} \leqslant a + c.$$

If $n \geqslant \nu_a^+$, $X_{\nu_a^+\wedge n} = X_{\nu_a^+} \leqslant a + c$; if $n < \nu_a^+$, $X_{\nu_a^+\wedge n} \leqslant a$. Thus, $X_{\nu_a^+\wedge n} \leqslant a + c$. Then

$$E\left[X_{\nu_a^+\wedge n}\right] \leqslant a + c.$$

Since $Y_1 > 0$,

$$\frac{a+c}{E[Y_1]} \geqslant E\left[\nu_a^+ \wedge n\right] \uparrow E[\nu_a^+].$$

So $\nu_a^+ \in L_1$.

(b) If Y_1 is not bounded above by c, given $\{Y_k\}_{k\geqslant 1}$, define

$$Y_i^{(c)} = Y_i \wedge c, \forall i \in \mathbb{N},$$

$$X_n^{(c)} = \sum_{i=1}^{n} Y_i^{(c)}, \forall n \in \mathbb{Z}^+,$$

$$\nu_a^{+(c)} = \inf\{n \in \mathbb{Z}^+ : X_n^{(c)} \geqslant a\} \geqslant \nu_a^+.$$

From Case 1, we have $\nu_a^{+(c)} \in L_1$. Thus, $\nu_a^+ \in L_1$. □

12.13.4 The simple random walk

Suppose $\{Y_n\}_{n\in\mathbb{Z}^+}$ are iid random variables and $P[Y_1 = \pm 1] = \frac{1}{2}$. Define

$$X_0 = 0,\ X_n = \sum_{i=1}^{n} Y_i,\ n \in \mathbb{Z}^+,$$

and think of X_n as your fortune after the nth gamble. Let $a \in \mathbb{Z}^+$, define

$$\nu_a^+ = \inf\{n : X_n = a\}.$$

Claim.

$$P(\nu_a^+ < \infty) = 1.$$

Proof. Clearly,

$$\nu_a^+ < \infty \text{ a.s. } \iff \nu_1^+ < \infty \text{ a.s.},$$

since if the random walk can reach state 1 in finite time, then it can start afresh and advance to state 2 with the same probability that governed its transition from 0 to 1. Define

$$p := P(\nu_a^+ = \infty).$$

Then

$$\begin{aligned} 1 - p &= P(\nu_1^+ < \infty) \\ &= P(\nu_1^+ < \infty, X_1 = -1) + P(\nu_1^+ < \infty, X_1 = 1) \\ &= \frac{1}{2}(1-p)(1-p) + \frac{1}{2} \end{aligned}$$

since $(1-p)(1-p)$ is the probability the random walk starts from -1, ultimately hits 0, and then starting from 0 ultimately hits 1. Therefore, $p = 0$. Notice that even though $P(\nu_a^+ < \infty) = 1$, $E[\nu_a^+] = \infty$ since otherwise, by Walds's equation

$$a = E\left[X_{\nu_a^+}\right] = E[Y_1]E[\nu_a^+] = 0,$$

a contradiction. □

Example 12.77 (Gambler's ruin)**.** Fix $a, b \in \mathbb{Z}^+$, define

$$\nu_{a,b} = \inf\{n \in \mathbb{Z}^+ : X_n \leqslant -b \text{ or } X_n \geqslant a\},$$

where we can use equal sign since the increments are -1 or $+1$. $\nu_{a,b}$ is regular w.r.t. $\{X_n\}_{n\geqslant 1}$ since

$$\int_{\{\nu_{a,b}\}} \left|X_{\nu_{a,b}}\right| dP \leqslant \max\{|a|, |b|\} < \infty,$$

and

$$|X_n| \mathbb{1}_{\{\nu_{a,b} > n\}} \leqslant \max\{|a|, |b|\} < \infty$$

so that $\{X_n \mathbb{1}_{\{\nu_{a,b} > n\}}\}$ is u.i..

Now regularity of the stopping time allows optimal stopping

$$\begin{aligned} 0 = E[X_0] = E\left[X_{\nu_{a,b}}\right] &= aP(\nu_a^+ < \nu_b^-) - bP(\nu_b^- < \nu_a^+) \\ &= aP(\nu_a^+ < \nu_b^-) - b(1 - P(\nu_a^+ < \nu_b^-)). \end{aligned}$$

So

$$P(\nu_a^+ < \nu_b^-) = \frac{b}{a+b}.$$

$$P(\nu_b^- < \nu_a^+) = \frac{a}{a+b}.$$

We now compute the expected duration of the game $E[\nu_{a,b}]$. Recall $\{X_n^2 - n\}_{n\in\mathbb{N}}$ is a martingale and $E[X_n^2 - n] = 0$. Also $\{(X_{\nu_{\nu_{a,b}\wedge n}-(\nu_{a,b}\wedge n)}, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a zero mean martingale so that

$$0 = E\left[X^2_{\nu_{a,b}\wedge n} - (\nu_{a,b}\wedge n)\right];$$

and since $X^2_{\nu_{a,b}\wedge n},\ \nu_{a,b}\wedge n \in L_1$, we have

$$E\left[X^2_{\nu_{a,b}\wedge n}\right] = E[\nu_{a,b}\wedge n].$$

Note that $\nu_{a,b}\wedge n \uparrow \nu_{a,b}$ as $n\to\infty$. By MCT, $E[\nu_{a,b}\wedge n]\uparrow E[\nu_{a,b}]$. Also,

$$\left|X^2_{\nu_{a,b}\wedge n}\right| \leqslant \max\left\{|a|^2, |b|^2\right\},$$

and $X_{\nu_{a,b}\wedge n}\to X_{\nu_{a,b}}$. By DCT, $E\left[X^2_{\nu_{a,b}\wedge n}\right]\to E\left[X^2_{\nu_{a,b}}\right]$. Thus, $E[\nu_{a,b}] = E\left[X^2_{\nu_{a,b}}\right]$. So $\nu_{a,b}\in L_1$, and then $\nu_{a,b}$ is regular. Hence

$$E[\nu_{a,b}] = E\left[X^2_{\nu_{a,b}}\right] = a^2P(X_{\nu_{a,b}} = a) + b^2P(X_{\nu_{a,b}=b}) = a^2\frac{b}{a+b} + b^2\frac{a}{a+b} = ab.$$

12.14 Reversed martingales

Lemma 12.78. Let X be an integrable random variable and let $\mathcal{F}$ and $\mathcal{G}$ be sub σ-fields of $\mathcal{B}$, where $X, G \perp\!\!\!\perp \mathcal{F}$. Then $E[X|\mathcal{G}\vee\mathcal{F}] = E[X|\mathcal{G}]$.

Proof.

$$\mathcal{G}\vee\mathcal{F} = \sigma\{A\cap B : A\in\mathcal{G}, B\in\mathcal{F}\}.$$

Then for $A\in\mathcal{G}$ and $B\in\mathcal{F}$,

$$\begin{aligned}\int_{A\cap B} E[X|\mathcal{G}]dP &= \int_\Omega E[X|\mathcal{G}]\mathbb{1}_{A\cap B}dP = E[E[X|\mathcal{G}]\mathbb{1}_A\mathbb{1}_B]\\ &= E[E[X|\mathcal{G}]\mathbb{1}_A]P(B) = E[E[X\mathbb{1}_A|\mathcal{G}]]P(B)\\ &= E[X\mathbb{1}_A]P(B) = E[X\mathbb{1}_A\mathbb{1}_B]\\ &= \int_{A\cap B} XdP.\end{aligned}$$

Since by the definition of conditional expectation,

$$\int_{A\cap B} E[X|\mathcal{G}\wedge\mathcal{F}]dP = \int_{A\cap B} XdP.$$

Thus,

$$\int_{A\cap B} E[X|\mathcal{G}]dP = \int_{A\cap B} E[X|\mathcal{G}\wedge\mathcal{F}]dP.$$

Since $\{A\cap B : A\in\mathcal{G}, B\in\mathcal{F}\}$ is a π-system and $\{C\in\mathcal{B} : \int_C E[X|\mathcal{G}]dP = \int_C E[X|\mathcal{G}\wedge\mathcal{F}]dP\}$ is a λ-system, by the Dynkin's π-λ theorem,

$$\int_C E[X|\mathcal{G}]dP = \int_C E[X|\mathcal{G}\wedge\mathcal{F}]dP, \forall C\in\mathcal{G}\vee\mathcal{F}.$$

By the integral comparision lemma, $E[X|\mathcal{G}] = E[X|\mathcal{G}\vee\mathcal{F}],\ P$-a.s.. □

Suppose that $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$ is a decreasing family of σ-fields, i.e., $\mathcal{B}_n \supseteq \mathcal{B}_{n+1}$ for any $n\in\mathbb{N}$.

Definition 12.79. Call $\{(X_n,\mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a *reversed martingale* if

$$X_n\in\mathcal{B}_n,\ X_n\in L_1, \forall n\in\mathbb{N},$$

$$E[X_n|\mathcal{B}_{n+1}] = X_{n+1}, \forall n\in\mathbb{N}.$$

From a reversed martingale, we can construct a martingale. Define

$$X'_n := X_{-n}, \forall n\in\mathbb{Z}^{\leqslant 0},$$

$$\mathcal{B}'_n := \mathcal{B}_{-n}, \forall n\in\mathbb{Z}^{\leqslant 0}.$$

Then $\mathcal{B}'_n\subseteq\mathcal{B}'_m$ for any $n<m<0$, and $\{(X'_n,\mathcal{B}'_n)\}_{n\in\mathbb{Z}^{\leqslant 0}}$ is a martingale on the index set

$$T = \{\cdots,-2,-1,0\}$$

with time flowing as usual from left to right. Note that

$$E[X'_{n+1}|\mathcal{B}'_n] = E[X_{-n-1}|\mathcal{B}_{-n}] = X_{-n} = X'_n,\ P\text{-a.s.}, \forall n\in\mathbb{Z}^{\leqslant 0}.$$

Also, this martingale is closed on the right by X'_0 and $E[X'_0|\mathcal{B}'_n] = X'_n$ for any $n\in\mathbb{Z}^{\leqslant 0}$. Therefore, the martingale $\{(X'_n,\mathcal{B}'_n)\}_{n\in\mathbb{Z}^{\leqslant 0}}$ is u.i. and as we will see, this implies the original sequence is convergent a.s. and in L_1.

Example 12.80. Let $\{\xi_k, k\in\mathbb{Z}^+\}$ be iid, L_1 random variables. Define

$$S_n = \sum_{i=1}^n \xi_i, \forall n\in\mathbb{Z}^+,$$

$$\mathcal{B}_n = \sigma(S_n, S_{n+1},\cdots).$$

Hence $\{\mathcal{B}_n\}_{n\in\mathbb{Z}^+}$ is a decreasing family of σ-fields and $S_n\in\mathcal{B}_n,\ S_n\in L_1$. Also,

$$\mathcal{B}_n = \sigma(S_n, \xi_{n+1}, \xi_{n+1},\cdots).$$

Furthermore, by symmetry, for any $n\in\mathbb{Z}^+$,

$$E[\xi_k|\mathcal{B}_n] = E[\xi_1|\mathcal{B}_n],\ P\text{-a.s.}, \forall 1\leqslant k\leqslant n.$$

By previous lemma,

$$E[\xi_k|\mathcal{B}_n] = E[\xi_k|\sigma(S_n)],\ P\text{-a.s.}, \forall 1 \leqslant k \leqslant n.$$

Next, need to show

$$E[\xi_k|\sigma(S_n) = E[\xi_1|\sigma(S_n)]\ P\text{-a.s.}, \forall 1 \leqslant k \leqslant n.$$

Fix $t \in \mathbb{R}$, then (or use $S_n \leqslant t$, π system: $\{S_n \leqslant t_1\} \cap \{S_n \leqslant t_2\} = \{S_n \leqslant t_1 \wedge t_2\}$.)

$$\int_{\{S_n>t\}} \xi_k dP = E[\xi_k \mathbb{1}_{\{S_n>t\}}] = E[\xi_k \mathbb{1}_{\{\xi_1+\cdots+\xi_n>t\}}] = E[\xi_1 \mathbb{1}_{\{S_n>t\}}] = \int_{\{S_n>t\}} \xi_1 dP.$$

By π-λ theorem,

$$E[\xi_k|\sigma(S_n)] = E[\xi_1|\sigma(S_n)],\ P\text{-a.s.}, \forall 1 \leqslant k \leqslant n.$$

Finally, for any $n \in \mathbb{Z}^+$,

$$S_n = E[S_n|\mathcal{B}_n] = \sum_{i=1}^n E[\xi_i|\mathcal{B}_n] = \sum_{i=1}^n E[\xi_1|\mathcal{B}_n] = nE[\xi_1|\mathcal{B}_n].$$

So $\frac{S_n}{n} = E[\xi_1|\mathcal{B}_n]$. Note

$$E\left[\frac{S_n}{n}\middle|\mathcal{B}_{n+1}\right] = E[E[\xi_1|\mathcal{B}_n]|\mathcal{B}_{n+1}] = E[\xi_1|\mathcal{B}_{n+1}] = \frac{S_{n+1}}{n+1},\ \text{etc.}$$

This implies that $\{(S_n/n, \mathcal{B}_n)\}_{n\in\mathbb{Z}^+}$ is a reversed martingale and thus uniformly integrable. From the theorem below, this sequence is almost surely convergent. The Kolmogorov 0-1 law gives

$$\lim_{n\to\infty} \frac{S_n}{n} = c(\text{constant}).$$

But this means ?$c = \frac{1}{n}E[S_n] = E[\xi_1]$. Thus, the Reversed Matingale Convergence Theorem below provides a very short proof of the SLLN.

Theorem 12.81 (Reversed Martingale Convergence Theorem)**.** *Suppose that $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$ is a decreasing family of σ-fields and suppose $\{(X_n, \mathcal{B}_n)\}_{n\in\mathbb{N}}$ is a positive reversed martingale. Set*

$$\mathcal{B}_\infty = \bigcap_{n\geqslant 0} \mathcal{B}_n.$$

Then

(a) there exists $X_\infty \in \mathcal{B}_\infty$ such that $X_n \xrightarrow{a.s.} X_\infty$,

(b) $E[X_n|\mathcal{B}_\infty] = X_\infty$, P-a.s. for any $n \in \mathbb{N}$,

(c) $\{X_n\}_{n\in\mathbb{N}}$ is u.i.. and $X_n \xrightarrow{L_1} X_\infty$.

Proof. (a) Define

$$X'_n := X_{-n},\ \mathcal{B}'_n := \mathcal{B}_{-n}, \forall n \in \mathbb{Z}^{\leqslant 0}.$$

Then $\{(X'_n, \mathcal{B}'_n)\}_{n\in\mathbb{Z}^{\leqslant 0}}$ is a martingale. Define

$$\begin{aligned}\delta_{a,b}^{(n)} :&= \#\text{ of downcrossing of } [a,b] \text{ by } X_0,\dots,X_n \\ &= \#\text{ of upcrossing of } [a,b] \text{ by } X_n,\dots,X_0 \\ &= \#\text{ of upcrossing of } [a,b] \text{ by } X'_{-n},\dots,X'_0 \\ &=: \gamma_{a,b}^{(n)}.\end{aligned}$$

Now apply Dubin's inequality to the positive martingale $X'_{-n},\dots,X'_0$ to get for any $k\in\mathbb{Z}^+$,

$$P(\gamma_{a,b}^{(n)} \geqslant k|\mathcal{B}'_{-n}) \leqslant \left(\frac{a}{b}\right)^k \left(\frac{X'_{-n}}{a}\wedge 1\right) = \left(\frac{a}{b}\right)^k \left(\frac{X_n}{a}\wedge 1\right).$$

Taking $E[\cdot|\mathcal{B}_\infty]$ on both sides yields

$$P(\delta_{a,b}^{(n)} \geqslant k|\mathcal{B}_\infty) \leqslant \left(\frac{a}{b}\right)^k E\left[\left(\frac{X_n}{a}\wedge 1\right)\middle|\mathcal{B}_\infty\right].$$

Note for any $k\in\mathbb{Z}^+$,

$$\begin{aligned}P(\delta_{a,b}^{(n)} \geqslant k|\mathcal{B}_\infty) &= E\left[E\left[\mathbb{1}_{\{\delta_{a,b}^{(n)}\geqslant k\}}\middle|\mathcal{B}_n\right]\middle|\mathcal{B}_\infty\right] \\ &\leqslant E\left[\left(\frac{a}{b}\right)^k\left(\frac{X_n}{a}\wedge 1\right)\middle|\mathcal{B}_\infty\right] \\ &\leqslant \left(\frac{a}{b}\right)^k \sup_{n\in\mathbb{N}} E\left[\frac{X_n}{a}\wedge 1\middle|\mathcal{B}_\infty\right] \\ &\leqslant \left(\frac{a}{b}\right)^k.\end{aligned}$$

Then $P(\delta_{a,b}^{(n)} \geqslant k) \leqslant \left(\frac{a}{b}\right)^k$. Also, as $n\uparrow\infty$.

$$\delta_{a,b}^{(n)} \uparrow \delta_{a,b} = \#\text{ downcrossings of } [a,b] \text{ by } \{X_0, X_1,\cdots\}.$$

So $P(\delta_{a,b}\geqslant k|) \leqslant \left(\frac{a}{b}\right)^k$. Thus, $P(\delta_{a,b}<\infty) = 1$. Therefore, $\{X_n\}$ converges a.s.. Set

$$X_\infty = \limsup_{n\to\infty} X_n.$$

Note for a fixed $p\in\mathbb{N}$, since $X_n\in\mathcal{B}_n$ for any $n\in\mathbb{N}$, $\sup_{n\geqslant p} X_n \in \mathcal{B}_p$. So

$$\limsup_{n\to\infty} X_n \in \mathcal{B}_p, \forall p\in\mathbb{N}.$$

Thus, $X_\infty = \limsup_{n\to\infty} X_n \in \mathcal{B}_\infty$.

(c) Note $X_n = E[X_0|\mathcal{B}_n]$. So $\{X_n\}$ is u.i.. U.i. and a.s. convergence imply L_1 convergence.

(b) Notice that

$$E[X_n|\mathcal{B}_\infty] = E[E[X_0|\mathcal{B}_n]|\mathcal{B}_\infty] = E[X_0|\mathcal{B}_\infty],\ P\text{-a.s.}.$$

Then

$$\lim_{n\to\infty} E[X_n|\mathcal{B}_\infty] = E[X_0|\mathcal{B}_\infty],\ P\text{-a.s.}.$$

Define

$$A_n = \{\omega : E[X_n|\mathcal{B}_\infty] = E[X_0|\mathcal{B}_\infty]\}, \forall n \in \mathbb{N}.$$

Then $P(A_n) = 1$ for any $n \in \mathbb{N}$. Define

$$A = \bigcap_{n\in\mathbb{N}} A_n.$$

Then $P(A) = 1$, and on A, since $X_\infty \in \mathcal{B}_\infty$, $X_n \xrightarrow{\text{a.s.}} X_\infty$, and $X_n \xrightarrow{L_1} X_\infty$, we have

$$X_\infty = E[X_\infty|\mathcal{B}_\infty] = \lim_{n\to\infty} E[X_n|\mathcal{B}_\infty] = E[X_0|\mathcal{B}_\infty].$$ □

Remark. These results are easily extended when we drop the assumption of positivity which was only assumed in order to be able to apply Dubin's inequality.

Corollary 12.82. Suppose $\{\mathcal{B}_n\}_{n\in\mathbb{N}}$ is a decreasing family of σ-fields and $X \in L_1$. Then

$$E[X|\mathcal{B}_n] \stackrel{a.s.}{\to} E[X|\mathcal{B}_\infty].$$

$$E[X|\mathcal{B}_n] \stackrel{L_1}{\to} E[X|\mathcal{B}_\infty].$$

(The result also holds if $\{\mathcal{B}_n\}$ is an increasing family of σ-fields.)

Proof. Observe that if we define $X_n := E[X|\mathcal{B}_n]$ for any $n \in \mathbb{N}$, then this sequence is a reversed martingale from smoothing. From the previous theorem, we know

$$X_n \to X_\infty \in \mathcal{B}_\infty \text{ a.s. and in } L_1.$$

We must identify X_∞. From L_1-convergence we have that for any $A \in \mathcal{B}$,

$$\int_A E[X|\mathcal{B}_n]dP \to \int_A X_\infty dP.$$

Thus for any $A \in \mathcal{B}_\infty \subseteq \mathcal{B}_n$,

$$\int_A E[X|\mathcal{B}_n]dP \stackrel{\text{def}}{=\!=} \int_A X dP \stackrel{\text{def}}{=\!=} \int_A E[X|\mathcal{B}_\infty]dP \to \int_A X_\infty dP.$$

Thus, by the integral Comparison Lemma, $X_\infty = E[X|\mathcal{B}_\infty]$. □