Elementary Probability Theory

Richard B. Sowers

Department of Mathematics, University of Illinois at Urbana–Champaign, Urbana, IL 61801

E-mail address: r-sowers@illinois.edu

DO NOT CIRCULATE

January 18, 2012

Contents

CHAPTER 1

Connections with Measure Theory

The grammar of probability theory is *measure theory*. We always start with

i) A measurable space $(\Omega, \mathscr{F})$. Thus Ω is a set, often called the event space, and $\mathscr{F}$ is a sigma-algebra of subsets of Ω. The elements of Ω are often denoted by ω, and dependence on ω is often suppressed.
ii) A measure $\mathbb{P}$ on $(\Omega, \mathscr{F})$ for which $\mathbb{P}(\Omega) = 1$; this is called a *probability measure*.

We then call $(\Omega, \mathscr{F}, \mathbb{P})$ a *probability triple*. Often the existence of $(\Omega, \mathscr{F}, \mathbb{P})$ is assumed or implicit. Often, we will assume that Ω has a topology $\mathscr{T}$, and then we will let $\mathscr{F} = \sigma(\mathscr{T})$, the smallest sigma-algebra containing the open subsets of Ω[1]; this is called the **Borel** sigma algebra of subsets of Ω.

Now let's define random variables. Let $(S, \mathscr{S})$ be a second measurable space[2].

DEFINITION 0.1 (Random variables). A *random variable* is a measurable mapping from $(\Omega, \mathscr{F})$ to $(S, \mathscr{S})$; i.e., $X^{-1}S \in \mathscr{F}$ for all $S \in \mathscr{S}$.

We also can take expectations.

DEFINITION 0.2 (Expectation). If X is an $\mathbb{R}$-valued random variable, we define

$$\mathbb{E}[X] \stackrel{\text{def}}{=} \int_{\omega \in \Omega} X(\omega) \mathbb{P}(d\omega)$$

when the quantity on the right is defined. We say that this quantity is the expectation of X.

Note that if X is any set which contains another set A, we can define the indicator function $\chi_A : X \to \{0, 1\}$ as

$$\chi_A(x) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \in X \setminus A \end{cases}$$

Then for any $A \in \mathscr{F}$,

$$\mathbb{E}[\chi_A] = \mathbb{P}(A).$$

Some other common expectations are as follows.

DEFINITION 0.3 (Moments). The p-th moment of a $\mathbb{R}$-valued random variable is defined as $\mathbb{E}[X^p]$, when this expectation exists.

DEFINITION 0.4 (Mean and variance). If X is a $\mathbb{R}$-valued random variable, we define its mean to be $\mathbb{E}[X]$ (if it exists) and its variance to be

$$\mathbb{E}[(X - \mathbb{E}[X])^2].$$

DEFINITION 0.5 (Characteristic function). If X is an $\mathbb{R}^d$-valued random variable, we define its characteristic function φ as

$$\varphi(\theta) \stackrel{\text{def}}{=} \mathbb{E}\left[\exp[i\langle X, \theta \rangle_{\mathbb{R}^d}]\right]. \qquad \theta \in \mathbb{R}^d$$

where $\langle \cdot, \cdot \rangle_{\mathbb{R}^d}$ is the standard inner product in $\mathbb{R}^d$.

---

[1] See Problem 1.

[2] If $S = \mathbb{R}^d$, we usually endow $\mathbb{R}^d$ with the standard topology and then take $\mathscr{S} = \mathscr{B}(\mathbb{R}^d)$.

1. Convergence

Now let's consider how random variables can converge. We assume that $(S, \mathscr{S})$ is a measurable space and that $\{X_n;\, n = 1, 2 \dots\}$ and X be $(S, \mathscr{S})$-valued random variables. First, we assume that S has a metric d (which generates a topology) and that $\mathscr{S} \supset \mathscr{B}(S)$.

Definition 1.1 (Almost sure convergence). We say that X_n converges to X almost surely (written $X_n \to X$ a.s.) if

$$\mathbb{P}\left\{\lim_{n\to\infty} d(X_n, X) = 0\right\} = 1$$

(in the language of measure theory, X_n converges to X almost everywhere).

Definition 1.2 (Convergence in probability). We say that X_n converges to X in probability if

$$\lim_{n\to\infty} \mathbb{P}\{d(X_n, X) \geq \varepsilon\} = 0$$

for every $\varepsilon > 0$ (in the language of measure theory, X_n converges to X in measure).

Definition 1.3 (Weak convergence[3]). We say that X_n converges to X weakly, in law, or in distribution, if

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

for all $\phi \in C_b(X)$, the vector space of bounded real-valued continuous functions on X. More generally, we say that a collection $\{\mu_n;\, n = 1, 2 \dots\}$ on $(S, \mathscr{S})$ converges weakly to another probability measure μ on $(S, \mathscr{S})$ if

$$\lim_{n\to\infty} \int_X \phi(x)\mu_n(dx) = \int_X \phi(x)\mu(dx)$$

for all $\phi \in C_b(X)$.

Now we assume that the metric d comes from a norm $\|\cdot\|$.

Definition 1.4 (Convergence in L^p). Fix $1 \leq p < \infty$[4]. We say an $\mathbb{R}$-valued random variable is in L^p if

$$\mathbb{E}[\|X\|^p] < \infty;$$

and we define

$$\|X\|_{L^p} \stackrel{\text{def}}{=} \mathbb{E}[\|X\|^p]^{1/p}$$

for all $X \in L^p$. If the X_n's and X are in L^p, we say that X_n converges to X in L^p if $\lim_{n\to\infty} \|X - X_n\|_{L^p} = 0$.

Let's understand how these types of convergence are related. The proofs are given as exercises.

Proposition 1.5. *Assume that S has metric d and that $\mathscr{S} \supset \mathscr{B}(S)$.*

- *If X_n tends to X a.s., then X_n tends to X in probability.*
- *If X_n tends to X in probability, then X_n tends to X in law.*
- *If the metric d comes from a norm $\|\cdot\|$, then if X_n tends to X in L^p (for $1 < p < \infty$), then X_n tends to X in probability.*

We also have a partial converse, whose proof is also one of the exercises.

Proposition 1.6. *Assume that S has metric d and that $\mathscr{S} \supset \mathscr{B}(S)$. If X_n tends to X in probability, then X_{n_k} tends to X a.s., where $\{X_{n_k}\}$ is some subsequence of $\{X_n\}$.*

We will later on need to know more about the relationship between almost-sure convergence and convergence in L^1. We will start with

Definition 1.7 (Uniform Integrability). Let A be an index set. A collection $\{X_\alpha;\, \alpha \in A\}$ of real-valued random variables is said to be uniformly integrable if

$$\overline{\lim_{K\to\infty}} \sup_{\alpha\in A} \mathbb{E}\left[|X_\alpha| \chi_{\{|X_\alpha| \geq K\}}\right] = 0.$$

[3]We will take this up in more detail in Chapter 2.

[4]There is typically little use for L^∞ in probability theory.

A slightly easier condition which is sufficient for uniform integrability is given by the following.

PROPOSITION 1.8. *Let $\{X_\alpha;\, \alpha \in A\}$ be a collection of random variables. Suppose that there is a function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ such that*

$$\lim_{t\to\infty} \frac{\phi(t)}{t} = \infty$$

and such that

$$\sup_{\alpha\in A} \mathbb{E}[\phi(|X_\alpha|)] < \infty$$

Then $\{X_\alpha;\, \alpha \in A\}$ is uniformly integrable.

PROOF. Fix $M > 0$ and $K' > 0$ such that $\phi(t)/t \geq M$ if $t \geq K'$. Then for any $K > K'$,

$$t\chi_{\{t\geq K\}} \leq M^{-1}\phi(t)\chi_{\{t\geq K\}} \leq M^{-1}\phi(t).$$

Thus

$$\overline{\lim_{K\to\infty}}\, \mathbb{E}[|X_\alpha|\chi_{\{|X_\alpha|\geq K\}}] \leq M^{-1} \sup_{\alpha\in A} \mathbb{E}[\phi(|X_\alpha|)].$$

Now let M tend to infinity. □

It turns out that uniform integrability is exactly the condition needed to strengthen almost-sure convergence to L^1 convergence. The following lemma will be useful in proving this.

LEMMA 1.9. *If $\{X_n;\, n \in \mathbb{N}\}$ is a uniformly integrable collection of real-valued random variables, then*

$$\mathbb{E}\left[\underline{\lim_{n\to\infty}}\, X_n\right] \leq \underline{\lim_{n\to\infty}}\, \mathbb{E}[X_n] \leq \overline{\lim_{n\to\infty}}\, \mathbb{E}[X_n] \leq \mathbb{E}\left[\overline{\lim_{n\to\infty}}\, X_n\right].$$

PROOF. By putting negative signs in the obvious places, we see that it is sufficient to prove the first inequality.

Fix $\varepsilon > 0$ and $K > 0$ such that

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|\chi_{\{|X_n|\geq K\}}] < \varepsilon.$$

Set

$$Y_n \stackrel{\text{def}}{=} X_n\chi_{\{X_n\geq -K\}} = X_n - X_n\chi_{\{X_n<-K\}}.$$

Then it is easy to see that $Y_n \geq X_n$ and $Y_n \geq -K$. By these observations and Fatou's lemma, we have that

$$\mathbb{E}[\underline{\lim_{n\to\infty}}\, X_n] \leq \mathbb{E}\left[\underline{\lim_{n\to\infty}}\, Y_n\right] \leq \underline{\lim_{n\to\infty}}\, \mathbb{E}[Y_n] \leq \underline{\lim_{n\to\infty}}\, \mathbb{E}[X_n] + \sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|\chi_{\{|X_n|\geq K\}}] \leq \underline{\lim_{n\to\infty}}\, \mathbb{E}[X_n] + \varepsilon.$$

Now let ε tend to zero. □

We can now prove that in the presence of uniform integrability, a.s. convergence implies L^1 convergence.

PROPOSITION 1.10. *Suppose that $\{X_n;\, n \in \mathbb{N} \cup \{\infty\}\} \subset L^1$ and $X_n \to X$ $\mathbb{P}$-a.s.. Then $X_n \to X_\infty$ in L^1 if and only if $\{X_n;\, n \in \mathbb{N}\}$ is uniformly integrable.*

PROOF. First, assume that $\{X_n;\, n \in \mathbb{N}\}$ is uniformly integrable. Then (as it is easy to see) $\{|X_n - X_\infty|;\, n \in \mathbb{N}\}$ is uniformly integrable. Since $|X_n - X_\infty| \to 0$ $\mathbb{P}$-a.s., the previous lemma implies that

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X_\infty|] = \mathbb{E}\left[\lim_{n\to\infty} |X_n - X_\infty|\right] = 0.$$

Now assume that $X_n \to X_\infty$ in L^1. Fix $\varepsilon \in (0, 1)$. Then there is an integer $N_\varepsilon \geq 1$ such that $\mathbb{E}[|X_n - X_\infty|] \leq \varepsilon$ for all $n \geq N_\varepsilon$. Since $\{X_n;\, 1 \leq n \leq N_\varepsilon\}$ is a finite subset of L^1,

$$\lim_{K\nearrow\infty} \sup_{1\leq n\leq N_\varepsilon} \mathbb{E}\left[|X_n|\chi_{\{|X_n|\geq K\}}\right] = 0. \tag{1}$$

Next note that for any $K > 0$,

$$\begin{aligned}\sup_{n\geq N_\varepsilon} \mathbb{E}\left[|X_n|\chi_{\{|X_n|\geq K\}}\right] &\leq \sup_{n\geq N_\varepsilon} \mathbb{E}\left[|X_\infty|\chi_{\{|X_n|\geq K\}}\right] + \sup_{n\geq N_\varepsilon} \mathbb{E}\left[|X_n - X_\infty|\right] \\ &\leq \sup_{n\geq N_\varepsilon} \mathbb{E}\left[|X_\infty|\chi_{\{|X_n|\geq K\}}\right] + \sup_{n\geq N_\varepsilon} \mathbb{E}\left[|X_n - X_\infty|\right] \leq \sup_{n\geq N_\varepsilon} \mathbb{E}\left[|X_\infty|\chi_{\{|X_n|\geq K\}}\right] + \varepsilon.\end{aligned}$$

We also have that

$$\begin{aligned}\overline{\lim_{K\nearrow\infty}} \sup_{n\geq N_\varepsilon} \mathbb{P}\{|X_\infty| \geq K\} &\leq \overline{\lim_{K\nearrow\infty}} \sup_{n\geq N_\varepsilon} \frac{\mathbb{E}[|X_n|]}{K} \\ &\leq \overline{\lim_{K\nearrow\infty}} \frac{\mathbb{E}|X_\infty|]}{K} + \overline{\lim_{K\nearrow\infty}} \sup_{n\geq N_\varepsilon} \frac{\mathbb{E}[|X_n - X_\infty|]}{K} \\ &\leq \overline{\lim_{K\nearrow\infty}} \frac{\mathbb{E}|X_\infty|]}{K} + \overline{\lim_{K\nearrow\infty}} \frac{\varepsilon}{K} \\ &= 0.\end{aligned}$$

Since X_∞ is integrable,

$$\overline{\lim_{K\nearrow\infty}} \sup_{n\geq N_\varepsilon} \mathbb{E}\left[|X_n|\chi_{\{|X_n|\geq K\}}\right] \leq \overline{\lim_{K\nearrow\infty}} \sup_{n\geq N_\varepsilon} \mathbb{E}\left[|X_\infty|\chi_{\{|X_n|\geq K\}}\right] + \varepsilon = \varepsilon.$$

Combine this with (1) to get the claim. □

Exercises

We assume a probability triple $(\Omega, \mathscr{F}, \mathbb{P})$ on which are defined all random variables. We also assume that $(S, \mathscr{S})$ is a second measurable space.

(1) Let $\mathscr{S}$ be a collection of subsets of Ω. Show that

$$\sigma(\mathscr{S}) \stackrel{\text{def}}{=} \bigcap_{\substack{\mathscr{S}' \text{ a sigma-algebra} \\ \mathscr{S}' \supset \mathscr{S}}} \mathscr{S}'$$

is the smallest sigma-algebra containing $\mathscr{S}$.

(2) Let $\{A_n;\, n = 1, 2 \dots\}$ be a collection of measurable subsets of Ω. Show that if

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty,$$

then

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = 0.$$

Hint: Note that

$$\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k \subset \cup_{k=N}^{\infty} A_k$$

for any N and use the monotonicity and subadditivity of $\mathbb{P}$. This is the first half of the **Borel-Cantelli** law. The second part is in Chapter 3.

(3) Fix a mapping $X : \Omega \to S$, where $(S, \mathscr{S})$ is some measurable space. Show that if $\mathscr{A} \subset \mathscr{S}$, $\sigma(\mathscr{A}) = \mathscr{S}$, and $X^{-1}A \in \mathscr{F}$ for all $A \in \mathscr{A}$, then X is a random variable.

(4) Let X be an S-valued random variable. Define

$$\mu(A) \stackrel{\text{def}}{=} \mathbb{P}\{X \in A\}. \qquad A \in \mathscr{S}$$

Show that μ is a probability measure on $(S, \mathscr{S})$ and that for any bounded and measurable function $\varphi : S \to \mathbb{R}$,

$$\mathbb{E}[\varphi(X)] = \int_S \varphi(z)\mu(dz).$$

The measure μ is called the **law** of X and is often denoted by $\mathbb{P}X^{-1}$.

(5) Let $F : \mathbb{R} \to [0, 1]$ be right-continuous and nondecreasing and have the following limits:

$$\lim_{x\to-\infty} F(x) = 0 \quad \text{and} \quad \lim_{x\to\infty} F(x) = 1.$$

We want to find a random variable X (on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$) such that $\mathbb{P}\{X \leq t\} = F(t)$ for all $t \in \mathbb{R}$; then F is called the **cumulative distribution function** of X. Consider the probability triple $([0, 1], \mathscr{B}([0, 1], \mathfrak{L}^1\big|_{\mathscr{B}([0,1])})$, where $\mathfrak{L}^1\big|_{\mathscr{B}([0,1])})$ is one-dimensional Lebesgue measure restricted to

$\mathscr{B}([0,1])$. Set $X(\omega) = \omega$ for all $\omega \in [0,1]$. The random variable is said to be **uniformly distributed** on $[0,1]$. Set

$$G(t) \stackrel{\text{def}}{=} \inf\{s : F(s) \geq t\}. \qquad t \in [0,1]$$

Prove that $Y = G(X)$ has distribution F.

(6) Assume that X is a nonnegative real-valued random variable. Assume that $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ is a nondecreasing function. Show that for any $L > 0$, such that $\phi(L) > 0$,

$$\mathbb{P}\{X \geq L\} \leq \mathbb{E}[\phi(X)\chi_{\{X \geq L\}}]/\phi(L) \leq \mathbb{E}[\phi(X)]/\phi(L).$$

This is a generalized form of the **Chebychev inequality**.

(7) Let X be a nonnegative real-valued random variable and fix $1 \leq p < \infty$. Show that

$$\mathbb{E}[X^p] = p \int_0^\infty t^{p-1} \mathbb{P}\{X \geq t\} dt.$$

Hint: Note that $x^p = p \int_0^x t^{p-1} dt$ for any $x \geq 0$.

(8) Let $(S_1, \mathscr{S}_2)$ and $(S_2, \mathscr{S}_2)$ be two measurable spaces. Assume that X is an S_1-measurable random variable and that Y is an S_2-measurable random variable. Show that $\tilde{X}(\omega) \stackrel{\text{def}}{=} (X(\omega), Y(\omega))$ is an $(S_1 \times S_1, \mathscr{S}_1 \times \mathscr{S}_2)$-random variable[5]. Hint: Consider the collection $\mathscr{R}$ of rectangle sets and the set

$$\mathscr{A} \stackrel{\text{def}}{=} \{S \subset S_1 \times S_2 : \tilde{X}^{-1} S \in \mathscr{F}\}.$$

(9) Assume that S has metric d and that $\mathscr{S} \supset \mathscr{B}(S)$. Show that if X_n tends to X a.s., then X_n tends to X in probability.

(10) Assume that S has metric d and that $\mathscr{S} \supset \mathscr{B}(S)$. Show that if X_n tends to X in probability, then X_n tends to X in law.

(11) Assume that S has metric d and that $\mathscr{S} \supset \mathscr{B}(S)$. Show that if X_n tends to X in probability, then X_{n_k} tends to X a.s., where $\{X_{n_k}\}$ is some subsequence of $\{X_n\}$. Hint: Consider the sets $A_k \stackrel{\text{def}}{=} \{d(X_{n_k}, X) \geq 1/k\}$, where n_k is large enough that $\mathbb{P}(A_{n_k}) \leq 2^{-k}$. Then use Borel-Cantelli

Next, assume that the metric d comes from a norm $\|\cdot\|$.

(12) Show that if X_n tends to X in L^p (for $1 < p < \infty$), then X_n tends to X in probability. Hint: use Chebychev's inequality.

(13) Let $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathscr{B}([0,1], \mathfrak{L}^1\big|_{\mathscr{B}([0,1])})$, and set $X_n = n\chi_{[0,1/n)}$ for all $n \in \mathbb{N}$ and $X = 0$. Show that X_n tends to X a.s. but not in L^1. Thus, neither almost-sure convergence nor convergence in probability imply L^1 convergence.

(14) Let $(\Omega, \mathscr{F}, \mathbb{P}) = ([0,1], \mathscr{B}([0,1], \mathfrak{L}^1\big|_{\mathscr{B}([0,1])})$, and set $A_{k,n} = [k/n, (k+1)/n)$ for all $n \in \mathbb{N}$ and $1 \leq k \leq n-1$. Let $\{B_n;\, n \in \mathbb{N}\}$ be some enumeration of the $A_{k,n}$'s and set $X_n = \chi_{B_n}$ and $X = 0$. Show that X_n converges to X in probability but not almost-surely.

(15) Show that if $1 \leq p < p' < \infty$, then $\|\cdot\|_{L^{p'}}$ is stronger than $\|\cdot\|_{L^p}$; i.e., convergence in $L^{p'}$ implies convergence in L^p.

(16) Show that if $1 \leq p_1 < p_2 < p_3 < \infty$, $\lim_n \|X_n - X\|_{p_1} = 0$, and $\sup_n \|X_n - X\|_{p_3} < \infty$, then $\lim_n \|X_n - X\|_{p_2} = 0$. This ends up using a simple *interpolation inequality* (Hint: use Hölder's inequality).

(17) Assume that $\{X_n\}$ are identically distributed (i.e., they have the same law) square-integrable random variables with common expectation μ and which are *uncorrelated*, i.e.,

$$\mathbb{E}\left[(X_j - \mu)(X_k - \mu)\right] = 0$$

if $j \neq k$. Then show that $n^{-1} \sum_{j=1}^n X_j$ tends to μ in L^2. This implies the **weak law of large numbers** (see Theorem 1.1).

(18) Show that the variance of a random variable X is also equal to $\mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

(19) Show that if φ is the characteristic function of some random variable, then
   (a) φ is continuous
   (b) $\varphi(0) = 1$

[5]Thus $d(X_n, X)$ is measurable in the definitions of almost-sure convergence and convergence in probability.

(c) For any $N \geq 0$, any $\theta_1, \theta_2 \ldots \theta_N$ in $\mathbb{R}$ and any $\alpha_1, \alpha_2 \ldots \alpha_N$ in $\mathbb{C}$,

$$\sum_{1 \leq i,j \leq N} \alpha_i \alpha_j^* \varphi(\theta_i - \theta_j) \geq 0.$$

(20) Show that if X is an $\mathbb{R}$-valued random variable in L^p for some positive integer p, then $\mathbb{E}[X^p] = (-i)^p \varphi^{(p)}(0)$.

(21) Let X be an $\mathbb{R}^d$-valued random variable with characteristic function φ. Show if f is bounded, integrable, and continuous, then

$$\mathbb{E}[f(X)] = \lim_{\varepsilon \to 0} \int_{\mathbb{R}^d} f(x) \left\{ \frac{1}{(2\pi)^d} \int_{\theta \in \mathbb{R}^d} \exp\left[-\frac{\varepsilon}{2} \|\theta\|^2 - \sqrt{-1} \langle \theta, x \rangle_{\mathbb{R}^d} \right] \varphi(\theta) d\theta \right\} dx.$$

Thus, characteristic functions are unique.

(22) This is Widder's inversion formula; see [**?**]. Fix $\mu \in \mathscr{P}[0, \infty)$ and $f \in C_b(\mathbb{R})$. For each $\lambda > 0$, define

$$\Phi(\lambda) \stackrel{\text{def}}{=} \int_{t \in [0,\infty)} e^{-\lambda t} \mu(dt);$$

this is the Laplace transform of μ (and note that μ puts full measure on $\mathbb{R}_+$).

(a) For each $\alpha > 0$, let X_α be Poisson with parameter α. Fix $f \in C_b(\mathbb{R})$. Show that for each $t \geq 0$,

$$\lim_{\lambda \to \infty} \mathbb{E}\left[f\left(\frac{X_{t\lambda}}{\lambda} \right) \right] = f(t).$$

(b) Use the result of the first part to compute

$$\lim_{\lambda \to \infty} \sum_{n=0}^{\infty} f\left(\frac{n}{\lambda} \right) \frac{(-\lambda)^n}{n!} \Phi^{(n)}(\lambda)$$

where $\Phi^{(n)}$ is the n-derivative of Φ.

CHAPTER 2

Independence, and Conditioning

1. Independence

Next, let's consider *independence*. We now assume a probability measure $\mathbb{P}$ on $(\Omega, \mathscr{F})$ and an index set Λ.

DEFINITION 1.1 (Independence). Let $\{\mathscr{G}_i;\, i \in \Lambda\}$ be a collection of sub-sigma-fields of $\mathscr{F}$. We say that these sigma-algebras are independent if

$$\mathbb{P}\left(\bigcap_{i\in\lambda} A_i\right) = \prod_{i\in\lambda} \mathbb{P}(A_i)$$

for all $A_i \in \mathscr{G}_i$ and for all finite subsets λ of Λ. If $\{X_i;\, i \in \Lambda\}$ is some collection of random variables, we say that they are independent if $\{\sigma\{X_i\};\, i \in \Lambda\}$ are independent

This reduces to the requirement that two sigma-fields $\mathscr{G}_1$ and $\mathscr{G}_2$ are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all $A \in \mathscr{G}_1$ and all $B \in \mathscr{G}_2$.

The following is an interesting consequence of having an infinite number of independent sigma-algebras:

THEOREM 1.2 (Kolmogorov's $0-1$ Law). *Let $\{\mathscr{G}_n;\, n = 1, 2 \dots\}$ be a collection of independent sigma-algebras. Define*

$$\mathscr{I} = \bigcap_{j=1}^{\infty} \bigvee_{k\geq j} \mathscr{G}_k;$$

*this is called the **tail** sigma-algebra. Then either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$ for any $A \in \mathscr{I}$.*

PROOF. We will show that $\mathscr{I}$ is independent of itself; then $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A)$, which implies the result. Fix $1 \leq j < j'$. Then $\bigvee_{j\leq k\leq j'} \mathscr{G}_k$ is independent of $\bigvee_{k\geq j'+1} \mathscr{G}_k \supset \mathscr{I}$. Now let j' tend to infinity. Thus $\bigvee_{k\geq j} \mathscr{G}_k \subset \mathscr{I}$ is independent of $\mathscr{I}$. □

We also have the second half of the Borel-Cantelli law[1]

THEOREM 1.3 (Borel-Cantelli, second half). *Assume that $\{A_1, A_2 \dots\}$ are independent events. Then $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ implies that*

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = 1. \tag{2}$$

PROOF. It of course suffices to show that

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^C\right) = 0.$$

To this end, recall that $1 - x \leq e^{-x}$ for any $x \geq 0$, and calculate that for any $0 \leq n \leq m$,

$$\mathbb{P}\left(\bigcap_{k=n}^{m} A_k^C\right) = \prod_{k=n}^{m} \mathbb{P}(A_k^C) = \prod_{k=n}^{m} (1 - \mathbb{P}(A_k)) \leq \exp\left[-\sum_{k=n}^{m} \mathbb{P}(A_k)\right].$$

[1]The first half was in one of the problems in Chapter 1.

Let m tend to infinity to see that

$$\mathbb{P}\left(\bigcap_{k=n}^{m} A_k^C\right) = 0.$$

This then implies (2). □

2. Conditional probability

If a random variable X is independent of a sub sigma-algebra $\mathscr{G}$, then we expect that $\mathscr{G}$ should have no effect on X. What happens if X and $\mathscr{G}$ are not independent?

Definition 2.1 (Conditional expectation). Fix $\mathscr{G}$ a sub sigma-algebra of $\mathscr{F}$ and X an integrable real-valued random variable. We say that a second integrable real-valued random variable ξ is a *version* of $\mathbb{E}[X|\mathscr{G}]$, the conditional expectation of X given $\mathscr{G}$, if

(a) ξ is $\mathscr{G}$-measurable.
(b) For every $A \in \mathscr{G}$, $\mathbb{E}[\chi_A \xi] = \mathbb{E}[\chi_A X]$.

The following result answers the obvious question of existence.

Theorem 2.2 (Existence of versions of conditional expectations). *For any real-valued integrable random variable X and any sub sigma-algebra $\mathscr{G}$ of $\mathscr{F}$, a version of $\mathbb{E}[X|\mathscr{G}]$ exists.*

Proof. Define the two measures

$$\mu_\pm(A) \stackrel{\text{def}}{=} \mathbb{E}[\chi_A X^\pm] \qquad A \in \mathscr{G}$$

where $X^+ \stackrel{\text{def}}{=} \max\{X, 0\}$ and $X^- \stackrel{\text{def}}{=} \max\{-X, 0\}$. Then $\mu_\pm$ is a measure on $(\Omega, \mathscr{G})$ which is absolutely continuous with respect to $\mathbb{P}\big|_{\mathscr{G}}$ (if $A \in \mathscr{G}$ and $\mathbb{P}(A) = 0$, then $\mu_\pm(A) = 0$). Thus applying the Radon-Nikodym theorem to measures on $(\Omega, \mathscr{G})$, we get the existence of two $\mathscr{G}$-measurable integrable random variables ξ_+ and ξ_- such that

$$\mathbb{E}[\chi_A \xi_\pm] = \mu_\pm(A) = \mathbb{E}[\chi_A X^\pm]$$

for all $A \in \mathscr{G}$. Since $X = X^+ - X_-$ and since expectations are linear, we get that $\xi \stackrel{\text{def}}{=} \xi^+ - \xi^-$ is a version of $\mathbb{E}[X|\mathscr{G}]$. □

One of the problems tells us that all versions of $\mathbb{E}[X|\mathscr{G}]$ differ only on a set of measure zero, so we can safely refer to $\mathbb{E}[X|\mathscr{G}]$ as an equivalence class of integrable functions.

Definition 2.3 (Conditional probability). For any $A \in \mathscr{F}$ and any sub sigma-algebra $\mathscr{G}$ of $\mathscr{F}$, we define

$$\mathbb{P}(A|\mathscr{G}) \stackrel{\text{def}}{=} \mathbb{E}[\chi_A|\mathscr{G}].$$

Exercises

As usual, we consider an underlying probability triple $(\Omega, \mathscr{F}, \mathbb{P})$. We also assume an index set $I \subset \mathbb{R}$ and a filtration $\{\mathscr{F}_t;\, t \in I\}$.

(1) Suppose if X is a random variable taking values in a measurable space $(S, \mathscr{S})$. We define $\sigma\{X\}$ as the smallest sub-sigma-algebra of $\mathscr{F}$ with respect to which X is measurable. Prove that $\sigma\{X\} = X^{-1}(\mathscr{S}) \stackrel{\text{def}}{=} \{X^{-1}(A) :\, A \in \mathscr{S}\}$. The point of this is that the definition of $\sigma\{X\}$ relied upon a generated sigma-field, whereas we can also define it directly as $X^{-1}\mathscr{S}$.
(2) Prove **Halmos' monotone class theorem**. A collection $\mathscr{F}_0$ of subsets of Ω is called a *field* if it contains Ω, is closed under complementation and (finite) unions. A collection $\mathscr{M}$ of subsets of Ω is called a *monotone class* if it is closed under monotone limits (i.e., if $\{A_n\} \subset \mathscr{M}$ and if $A_n \nearrow A$ or $A_n \searrow A$, then $A \in \mathscr{M}$). Show that if $\mathscr{F}_0 \subset \mathscr{M}$, then $\sigma(\mathscr{F}_0) \subset \mathscr{M}$. Hints:
   (a) Define

$$m(\mathscr{F}_0) \stackrel{\text{def}}{=} \bigcap_{\substack{\mathscr{M}' \text{ a monotone class} \\ \mathscr{M}' \supset \mathscr{F}_0}} \mathscr{M}'.$$

Show that $m(\mathscr{F}_0)$ is a monotone class and that if it is a field, then it is a sigma-algebra (and hence $\sigma(\mathscr{F}_0) \subset m(\mathscr{F}_0)$).

(b) Show that $\Omega \in m(\mathscr{F}_0)$.

(c) Show that $m(\mathscr{F}_0)$ is closed under complementation by showing that

$$m(\mathscr{F}_0) \subset \{A \subset \Omega : A^c \in m(\mathscr{F}_0)\}.$$

(d) Show that $m(\mathscr{F}_0)$ is closed under (finite) unions by showing that

$$m(\mathscr{F}_0) \subset \{A \subset \Omega : A \cup B \in m(\mathscr{F}_0) \text{ for all } B \in \mathscr{F}_0\}$$
$$m(\mathscr{F}_0) \subset \{A \subset \Omega : A \cup B \in m(\mathscr{F}_0) \text{ for all } B \in m(\mathscr{F}_0)\}.$$

(3) Prove **Dynkin's $\pi - \lambda$ theorem**. A collection $\mathscr{P}$ of subsets of Ω is called a *π-system* if it is closed under (finite) intersections. A collection $\mathscr{L}$ of subsets of Ω is called a *λ-system* if it contains Ω, is closed under complementation and countable disjoint unions (i.e., if $\{A_n\} \subset \mathscr{L}$ are disjoint, then $\cup_n A_n \in \mathscr{L}$). Show that if $\mathscr{P} \subset \mathscr{L}$, then $\sigma(\mathscr{P}) \subset \mathscr{L}$. Hints:

(a) Define

$$l(\mathscr{P}) \stackrel{\text{def}}{=} \bigcap_{\substack{\mathscr{L}' \text{ a } \lambda\text{-system} \\ \mathscr{L}' \supset \mathscr{P}}} \mathscr{L}'.$$

Show that $l(\mathscr{P})$ is a λ-system and that if it is a π-system, then it is a sigma-algebra (and hence $\sigma(\mathscr{P}) \subset l(\mathscr{P})$).

(b) Show that $l(\mathscr{P})$ is a π-system by showing that

$$l(\mathscr{P}) \subset \{A \subset \Omega : A \cap B \in l(\mathscr{P}) \text{ for all } B \in \mathscr{P}\}$$
$$l(\mathscr{P}) \subset \{A \subset \Omega : A \cap B \in l(\mathscr{P}) \text{ for all } B \in l(\mathscr{P})\}.$$

To show that the collections on the right are closed under complementation, note that $(A^c \cap B)^c = A \cup B^c = (A \cap B) \cup B^c$.

(4) Let X be a metric space and let μ be a probability measure on $(X, \mathscr{B}(X))$. Show that μ is *regular*; i.e., that for all $A \in \mathscr{B}(X)$,

$$\mu(A) = \mu^*(A) = \mu_*(A),$$

where

$$\mu^*(A) \stackrel{\text{def}}{=} \inf\{\mu(O) : O \supset A \text{ is open}\}$$
$$\mu_*(A) \stackrel{\text{def}}{=} \sup\{\mu(F) : F \subset A \text{ is closed}\}$$

for all $A \in \mathscr{B}(X)$. Hint: consider the collection

$$\mathscr{G} \stackrel{\text{def}}{=} \{A \in \mathscr{B}(X) : \mu^*(A) = \mu(A) = \mu_*(A)\}$$

of subsets of X.

(5) Let E be a metric space and $T > 0$ and consider the measurable space $(\Omega, \mathscr{F})$, where $\Omega \stackrel{\text{def}}{=} C([0,T]; E)$, this being endowed with the topology generated by the supremum norm, and $\mathscr{F} \stackrel{\text{def}}{=} \mathscr{B}(C([0,T]; E))$. For each $0 \le t \le T$, let $X_t(\omega) = \omega(t)$ for all $\omega \in \Omega$. Show that the X_t's are E-valued random variables and that $\mathscr{B}(C([0,T]; E)) = \sigma\{X_t;\, t \in [0,T]\}$.

(6) Fix A and B in $\mathscr{F}$, and set $\mathscr{G}_1 \stackrel{\text{def}}{=} \{\emptyset, \Omega, A, A^c\}$ and $\mathscr{G}_2 \stackrel{\text{def}}{=} \{\emptyset, \Omega, B, B^c\}$. Show that A and B are independent if and only if $\mathscr{G}_1$ and $\mathscr{G}_2$ are independent. This reduces the general definition of independence to the most elementary one.

(7) Let Λ be an index set and let $\{\mathscr{G}_i;\, i \in \Lambda\}$ be a collection of sub sigma-algebras of $\mathscr{F}$. We define

$$\bigvee_{t \in \Lambda} \mathscr{G}_t \stackrel{\text{def}}{=} \sigma\left\{\bigcup_{t \in \Lambda} \mathscr{G}_t\right\}.$$

Show that

$$\bigvee_{t \in \Lambda} \mathscr{G}_t = \sigma\left\{\cap_{t \in \lambda} A_t : \lambda \subset \Lambda \text{ is countable and } A_t \in \mathscr{G}_t \text{ for all } t \in \lambda\right\}.$$

(8) Assume that $\mathscr{A}_1$ and $\mathscr{A}_2$ are either π-systems or fields, and that all of the sets in $\mathscr{A}_1$ are independent of all of the sets in $\mathscr{A}_2$. Show that $\sigma(\mathscr{A}_1)$ is independent of $\sigma(\mathscr{A}_2)$.

(9) Let $(\Omega_i, \mathscr{G}_i, \mathbb{P}_i)$ be a probability triple for $i = 1, 2 \dots n$. Define $\Omega \stackrel{\text{def}}{=} \times_{i=1}^n \Omega_i$, $\mathscr{G} \stackrel{\text{def}}{=} \times_{i=1}^n \mathscr{G}_i$, and $\mathbb{P} \stackrel{\text{def}}{=} \times_{i=1}^n \mathbb{P}_i$. For each $1 \leq i \leq n$, define

$$\tilde{\mathscr{G}_i} \stackrel{\text{def}}{=} \{(x_1, x_2 \dots x_n) \in \Omega;\, x_i \in A_i :\, A_i \in \mathscr{G}_i\}.$$

Show that the $\tilde{\mathscr{G}_i}$'s are independent. This proves that independent random variables give rise to product measures.

(10) Assume that X and Y are independent random variables taking values, respectively, in two measurable spaces $(S_1, \mathscr{S}_1)$ and $(S_2, \mathscr{S}_2)$ with laws, respectively, of μ_1 and μ_2. Define the $(S_1 \times S_2, \mathscr{S}_1 \times \mathscr{S}_2)$-valued random variable $Z \stackrel{\text{def}}{=} (X, Y)$. Let μ be the law of Z. Show that $\mu = \mu_1 \times \mu_2$. Hint: start with rectangle sets. This proves that product measures give rise to independence.

(11) Assume that $\{X_1, X_2 \dots\}$ are independent real-valued random variables. Show that

$$\mathbb{P}\left\{\lim_{n\to\infty} n^{-1} \sum_{j=1}^n X_j \text{ exists}\right\} \in \{0, 1\}.$$

Hint: show that the existence of the limit does not depend on any finite number of terms.

Now let X be a real-valued integrable random variable and let $\mathscr{G}$ be a sub sigma-algebra of $\mathscr{F}$.

(12) Any two versions of $\mathbb{E}[X|\mathscr{G}]$ differ only on a set of measure zero.

(13) The mapping $X \mapsto \mathbb{E}[X|\mathscr{G}]$ is a linear mapping from L^1 into itself of norm 1.

(14) If $\varphi : \mathbb{R} \to \mathbb{R}$ is any convex function such that $\varphi(X)$ is integrable, then $\varphi(\mathbb{E}[X|\mathscr{G}]) \leq \mathbb{E}[\varphi(X)|\mathscr{G}]$. Hint: write φ as the supremum of all linear minorants.

(15) If $X \geq 0$, then $\mathbb{E}[X|\mathscr{G}] \geq 0$.

(16) If X is $\mathscr{G}$-measurable, then $\mathbb{E}[X|\mathscr{G}] = X$.

(17) If $\mathscr{G}'$ is a second sub sigma-algebra of $\mathscr{F}$ such that $\mathscr{G} \subset \mathscr{G}'$, then $\mathbb{E}[\mathbb{E}[X|\mathscr{G}']|\mathscr{G}] = \mathbb{E}[X|\mathscr{G}]$. This is *iterated conditioning*.

(18) Let X and Y be random variables which take values in measurable spaces $(S_1, \mathscr{S}_1)$ and $(S_2, \mathscr{S}_2)$ respectively. Suppose that Y is measurable with respect to some sigma-algebra $\mathscr{G}$ but that X is independent of $\mathscr{G}$. Let $\varphi : S_1 \times S_2 \to \mathbb{R}$ be a bounded function. Then $\mathbb{E}[\varphi(X, Y)|\mathscr{G}] = \Phi(Y)$, where $\Phi(y) \stackrel{\text{def}}{=} \mathbb{E}[\phi(X, y)]$ for all $y \in S_2$. Hint: first consider functions which are indicators of rectangle sets.

(19) Suppose that X is independent of $\mathscr{G}$. Use the previous question to compute $\mathbb{E}[X|\mathscr{G}]$.

(20) Let $\mathscr{G} \stackrel{\text{def}}{=} \sigma\{A_1, A_2 \dots A_n\}$, where $\{A_i\} \subset \mathscr{F}$ are disjoint and $\cup_{i=1}^n A_i = \Omega$. Let X be an integrable random variable. Find $\mathbb{E}[X|\mathscr{G}]$.

(21) Let X be a bounded or nonnegative random variable, and let $\mathscr{G}$ be a sub sigma-algebra of $\mathscr{F}$. Let $\mathbb{P}'$ be a second probability measure on $(\Omega, \mathscr{F})$ which is absolutely continuous with respect to $\mathbb{P}$, and let $\mathbb{E}'$ be the expecation operator associated with $\mathbb{P}'$. Show that

$$\mathbb{E}'[X|\mathscr{G}] = \frac{\mathbb{E}\left[X \frac{d\mathbb{P}'}{d\mathbb{P}} \middle| \mathscr{G}\right]}{\mathbb{E}\left[\frac{d\mathbb{P}'}{d\mathbb{P}} \middle| \mathscr{G}\right]}.$$

Make sure to prove that $\mathbb{P}'$-a.s., $\mathbb{E}\left[\frac{d\mathbb{P}'}{d\mathbb{P}} \middle| \mathscr{G}\right] > 0$. This is a form of **Bayes' rule**.

(22) Let X be an $\mathbb{R}^d$-valued random variable and let $\mathscr{G}$ be a sub-sigma-algebra of $\mathscr{F}$. Suppose that

$$\mathbb{E}\left[\exp\left[\sqrt{-1}\langle\theta, X\rangle_{\mathbb{R}^d}\right] \middle| \mathscr{G}\right] = \mathbb{E}\left[\exp\left[\sqrt{-1}\langle\theta, X\rangle_{\mathbb{R}^d}\right]\right]$$

for all $\theta \in \mathbb{R}$. Show that X is independent of $\mathscr{G}$.

CHAPTER 3

Asymptotics: Limit theorems

We now take up some *asymptotic* questions. Throughout this section, we will let $\{\xi_1, \xi_2 \dots\}$ be an *independent and identically distributed* (i.i.d.) collection of $\mathbb{R}$-valued random variables with common law μ. Also define

$$S_n \stackrel{\text{def}}{=} \sum_{j=1}^{n} \xi_j$$

for all n.

1. The weak law of large numbers

First, let's assume that

$$\int_{\mathbb{R}} |x| \mu(dx) < \infty \qquad \text{and} \qquad \int_{\mathbb{R}} x \mu(dx) = 0 \tag{3}$$

(if the second condition is not true, we translate).

THEOREM 1.1 (Weak Law of Large Numbers for L^2 random variables). *Assume that*

$$\int_{x \in \mathbb{R}} x^2 \mu(dx) < \infty.$$

Then we have that

$$\lim_{n \to \infty} \frac{S_n}{n} = 0$$

in probability.

PROOF. We use problem 15 in chapter 1. Since the ξ_i's are independent, they are uncorrelated. Problem 15 in chapter 1 implies that $n^{-1} S_n$ tends to zero in L^2. Since L^2 convergence implies convergence in probability, we have the desired result. □

We can fairly easily remove the requirement of square-integrability.

THEOREM 1.2 (Weak Law of Large Numbers). *Under the assumptions of* (3) *we have the weak law of large numbers.*

PROOF. Fix $\delta > 0$; we want to bound

$$\mathbb{P}\left\{ \left| \frac{1}{n} S_n \right| \geq \delta \right\}.$$

Fix next $L > 0$ (to be determined in a moment) and *truncate* the ξ_i'; define

$$\begin{aligned} \xi_j^L &\stackrel{\text{def}}{=} \xi_j \chi_{[-L,L]}(\xi_j) \\ \hat{\xi}_j^L &\stackrel{\text{def}}{=} \xi_j^L - \mathbb{E}\left[\xi_j^L\right] \\ \tilde{\xi}_j^L &\stackrel{\text{def}}{=} \xi_j - \xi_j^L \end{aligned}$$

for all $j \in \mathbb{N}$. Then we have that

$$S_n \stackrel{\text{def}}{=} S_n^{L,1} + S_n^{L,2} + S_n^{L,3}$$

where

$$S_n^{L,1} \stackrel{\text{def}}{=} \sum_{j=1}^n \hat{\xi}_j^L \qquad S_n^{L,2} \stackrel{\text{def}}{=} \sum_{j=1}^n \mathbb{E}[\xi_j^L] \qquad \text{and} \qquad S_n^{L,3} \stackrel{\text{def}}{=} \sum_{j=1}^n \tilde{\xi}_j^L$$

We first note that $S_n^{L,2}$ is in fact not random;

$$\frac{1}{n} S_n^{L,2} = 0\mu\{[-L,L]^c\} + \int_{|z|\le L} z\mu(dz);$$

by dominated convergence,

$$\lim_{L\nearrow\infty} \int_{|z|\le L} z\mu(dz) = 0,$$

so there is an $\bar{L}$ such that

$$\left| \int_{|z|\le L} z\mu(dz) \right| < \frac{\delta}{3}$$

if $L \ge \bar{L}$. We next note that

$$\mathbb{P}\left\{ \left| \frac{1}{n} S_n^{L,3} \right| \ge \frac{\delta}{3} \right\} \le \frac{3}{\delta} \mathbb{E}[|\frac{1}{n} S_n^{L,3}|] \le \frac{3}{\delta n} \sum_{j=1}^n \mathbb{E}[|\tilde{\xi}_j^L|]$$

$$\le \frac{3}{\delta} \left\{ 0\mu\{[-L,L]\} + \int_{|z|>L} z\mu(dz) \right\} = \frac{3}{\delta} \int_{|z|>L} z\mu(dz).$$

By the weak law of large numbers for L^2 random variables, we have that

$$\lim_{n\to\infty} \mathbb{P}\left\{ \left| \frac{1}{n} S_n^{L,1} \right| \ge \frac{\delta}{3} \right\} = 0,$$

so for $L > \bar{L}$,

$$\mathbb{P}\left\{ \left| \frac{1}{n} S_n \right| \ge \delta \right\} \le \frac{3}{\delta} \int_{|z|>L} z\mu(dz).$$

By dominated convergence,

$$\lim_{L\to\infty} \int_{|z|>L} z\mu(dz) = 0,$$

so we have the claimed result. □

We will later use martingale theory to prove the *strong* law of large numbers, which gives almost-sure convergence if μ is integrable.

2. The Central Limit Theorem

We now consider the central limit theorem for the ξ_j's. Let us begin by defining

Definition 2.1 (Gaussian Random Variables). Define

$$\mathfrak{G}(A) \stackrel{\text{def}}{=} \int_A \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2} \right] dx. \qquad A \in \mathscr{B}(\mathbb{R})$$

For any $m \in \mathbb{R}$ and $\sigma \ge 0$, we say that an $\mathbb{R}$-valued random variable η (defined on some probability triple $(\Omega, \mathscr{F}, \mathbb{P})$) is $\mathfrak{N}(m, \sigma^2)$ if

$$\mathbb{P}\{\eta \in A\} = \mathfrak{G}(\{x \in \mathbb{R} : \sigma x + m \in A\}). \qquad A \in \mathscr{B}(\mathbb{R})$$

We now assume that

$$\int_{\mathbb{R}} x\mu(dx) = 0 \qquad \text{and} \qquad \int_{\mathbb{R}} x^2\mu(dx) = 1; \tag{4}$$

i.e., the ξ's have mean zero and standard deviation 1[1]. Then we have

[1]This is not really a big restriction; as long as the standard deviation σ of the ξ_j's is positive, we can get (4) by replacing ξ_j by $(\xi_j - \bar{\xi})/\sigma$.

THEOREM 2.2 (Central Limit Theorem). *Let ν_n be the law of $S_n/\sqrt{n}$; i.e.,*

$$\nu_n(A) \stackrel{def}{=} \mathbb{P}\left\{n^{-1/2}S_n \in A\right\}; \qquad A \in \mathscr{B}(\mathbb{R})$$

then $\lim_{n\to\infty}\mu_n = \mathfrak{G}$ in the Prohorov topology. Equivalently,

$$\lim_{n\to\infty}\mathbb{E}\left[\varphi\left(n^{-1/2}S_n\right)\right] = \int_{\mathbb{R}}\varphi(x)\mathfrak{G}(dx)$$

for all $\varphi \in C_b(\mathbb{R})$.

PROOF. The key observation is that the result is identically true for all n if the ξ_j's are Gaussian. To make this precise, enlarge $(\Omega, \mathscr{F}, \mathbb{P})$ as necessary to include a collection $\{\eta_j;\, j = 1, 2\dots\}$ of independent $\mathfrak{N}(0,1)$-random variable which are also independent of the ξ_j's. Then $n^{-1/2}\sum_{j=1}^n \eta_j$ is $\mathfrak{N}(0,1)$ for all n. We want to sequentially replace the ξ_j's by the η_j's. For each n, define

$$\bar{S}_n^k \stackrel{\text{def}}{=} \sum_{\substack{j\in\{1,2\dots n\}\\ j\le k}}\xi_j + \sum_{\substack{j\in\{1,2\dots n\}\\ j>k}}\eta_j.$$

Then $\bar{S}_n^n = S_n$ and that $\bar{S}_n^0/\sqrt{n}$ is $\mathfrak{N}(0,1)$.

Fix now $\varphi \in C_b^3(\mathbb{R})$ (i.e., φ and its first three derivatives are bounded. Thus

$$\begin{aligned}\mathbb{E}\left[\varphi(n^{-1/2}S_n)\right] - \int_{\mathbb{R}}\varphi(x)\mathfrak{G}(dx) &= \mathbb{E}\left[\varphi(n^{-1/2}\bar{S}_n^n)\right] - \mathbb{E}\left[\varphi(n^{-1/2}\bar{S}_n^0)\right]\\ &= \sum_{k=0}^{n-1}\left\{\mathbb{E}\left[\varphi(n^{-1/2}\bar{S}_k^n)\right] - \mathbb{E}\left[\varphi(n^{-1/2}\bar{S}_n^{k+1})\right]\right\}.\end{aligned}$$

To take full advantage of independence, define

$$U_n^k \stackrel{\text{def}}{=} \sum_{\substack{j\in\{1,2\dots n\}\\ j<k}}\xi_j + \sum_{\substack{j\in\{1,2\dots n\}\\ j>k}}\eta_j.$$

then for each $k = 1, 2\dots n$,

$$S_n^k = U_n^k + \xi_k \quad \text{and} \quad S_n^{k-1} = U_n^k + \eta_k$$

and ξ_k and η_k are independent of U_n^k. Define now

$$R(x;\delta) \stackrel{\text{def}}{=} \varphi(x+\delta) - \varphi(x) - \dot{\varphi}(x)\delta - \frac{1}{2}\ddot{\varphi}(x)\delta^2$$

for all x and y in $\mathbb{R}$. Then for any $k = 1, 2\dots n$,

$$\begin{aligned}\mathbb{E}\left[\varphi(n^{-1/2}\bar{S}_n^k)\right] &= \mathbb{E}\left[\varphi(n^{-1/2}U_n^k)\right] + \frac{1}{\sqrt{n}}\mathbb{E}\left[\dot{\varphi}(n^{-1/2}U_n^k)\xi_k\right]\\ &\quad + \frac{1}{2n}\mathbb{E}\left[\ddot{\varphi}(n^{-1/2}U_n^k)(\xi_k)^2\right] + \mathbb{E}\left[R(n^{-1/2}U_n^k;\xi_k/\sqrt{n})\right]\\ &= \mathbb{E}\left[\varphi(n^{-1/2}U_n^k)\right] + \frac{1}{2n}\mathbb{E}\left[\ddot{\varphi}(n^{-1/2}U_n^k)\right] + \mathbb{E}\left[R(U_n^k/\sqrt{n};\xi_k/\sqrt{n})\right]\\ \mathbb{E}\left[\varphi(n^{-1/2}\bar{S}_n^{k+1})\right] &= \mathbb{E}\left[\varphi(n^{-1/2}U_n^k)\right] + \frac{1}{\sqrt{n}}\mathbb{E}\left[\dot{\varphi}(n^{-1/2}U_n^k)\eta_k\right]\\ &\quad + \frac{1}{2n}\mathbb{E}\left[\ddot{\varphi}(n^{-1/2}U_n^k)(\eta_k)^2\right] + \mathbb{E}\left[R(n^{-1/2}U_n^k;\eta_k/\sqrt{n})\right]\\ &= \mathbb{E}\left[\varphi(n^{-1/2}U_n^k)\right] + \frac{1}{2n}\mathbb{E}\left[\ddot{\varphi}(n^{-1/2}U_n^k)\right] + \mathbb{E}\left[R(n^{-1/2}U_n^k;\eta_k/\sqrt{n})\right]\end{aligned}$$

Thus

$$\left|\mathbb{E}\left[\varphi(n^{-1/2}S_n)\right] - \int_{\mathbb{R}}\varphi(x)\mathfrak{G}(dx)\right| \le \sum_{k=1}^n\left\{\mathbb{E}\left[|R(n^{-1/2}U_n^k;\xi_k/\sqrt{n})\right] + \mathbb{E}\left[|R(n^{-1/2}U_n^k;\eta_k/\sqrt{n})\right]\right\}.$$

We will now bound $R(x;y)$ in two ways, for large y and then for small y. We have that

$$R(x;\delta) = \frac{\delta^3}{2}\int_{s=0}^1 (1-s)^2\varphi^{(3)}(x+s\delta)ds = \delta^2\int_{s=0}^1 (1-s)\{\ddot{\varphi}(x+s\delta)-\ddot{\varphi}(x)\}ds$$

for all x and y. Thus

$$|R(x;\delta)| \le \min\left\{\frac{|\delta|^3}{3}\|\varphi\|_{C^3(\mathbb{R})}, \frac{\delta^2}{2}\|\varphi\|_{C^2(\mathbb{R})}\right\} \le \kappa\min\{|\delta|^3,|\delta|^2\}$$

where

$$\kappa \stackrel{\text{def}}{=} \frac{1}{2}\|\varphi\|_{C^3(\mathbb{R})}.$$

Thus,

$$\begin{aligned}
&\left|\mathbb{E}\left[\varphi(S_n/\sqrt{n})\right] - \int_{\mathbb{R}}\varphi(x)\mathfrak{G}(dx)\right| \\
&\qquad\le \kappa\sum_{k=1}^n\left\{\mathbb{E}\left[\min\left\{\frac{|\xi_j|^2}{n},\frac{|\xi_j|^3}{n^{3/2}}\right\}\right] + \mathbb{E}\left[\min\left\{\frac{|\eta_j|^2}{n},\frac{|\eta_j|^3}{n^{3/2}}\right\}\right]\right\} \\
&\qquad= \frac{\kappa}{n}\sum_{k=1}^n\left\{\mathbb{E}\left[\min\left\{|\xi_j|,n^{-1/2}|\xi_j|^3\right\}\right] + \mathbb{E}\left[\min\left\{|\eta_j|,n^{-1/2}|\eta_j|^3\right\}\right]\right\} \\
&\qquad= \kappa\left\{\int_{x\in\mathbb{R}}\min\left\{|x|,n^{-1/2}|x|^3\right\}\mu(dx) + \int_{x\in\mathbb{R}}\min\left\{|x|,n^{-1/2}|x|^3\right\}\mathfrak{G}(dx)\right\}.
\end{aligned}$$

By dominated convergence,

$$\lim_{n\to\infty}\left\{\int_{\mathbb{R}}|z|^2(1\wedge|z/\sqrt{n}|\mu(dz) + \int_{\mathbb{R}}|z|^2(1\wedge|z/\sqrt{n}|\mathfrak{G}(dz)\right\} = 0,$$

which completes the proof if $\varphi\in C_b^3(\mathbb{R})$.

Assume now that $\varphi\in C_b(\mathbb{R})$. For each $\delta>0$, define

$$\varphi_\delta(x) \stackrel{\text{def}}{=} \int_{y\in\mathbb{R}}\varphi(y)(2\pi\delta)^{-1/2}\exp\left[-\frac{(y-x)}{2\delta}\right]dy.$$

Then $\varphi_\delta\in C_b^3(\mathbb{R})$. We also have that

$$\lim_{\delta\to 0}\sup_{|x|\le L}|\varphi_\delta(x)-\varphi(x)| = 0.$$

for each $L>0$. We write that

$$\begin{aligned}
\left|\mathbb{E}\left[\varphi(S_n/\sqrt{n})\right] - \int_{\mathbb{R}}\varphi(x)\mathfrak{G}(dx)\right| &\le \mathbb{E}\left[\left|\varphi(S_n/\sqrt{n})-\varphi_\delta(S_n/\sqrt{n})\right|\right] \\
&\quad + \left|\mathbb{E}\left[\varphi_\delta(S_n/\sqrt{n})\right] - \int_{\mathbb{R}}\varphi_\delta(x)\mathfrak{G}(dx)\right| + \int_{\mathbb{R}}|\varphi_\delta(x)-\varphi(x)|\,\mathfrak{G}(dx).
\end{aligned}$$

From the above, we have that

$$\lim_{n\to\infty}\left|\mathbb{E}\left[\varphi_\delta(S_n/\sqrt{n})\right] - \int_{\mathbb{R}}\varphi_\delta(x)\mathfrak{G}(dx)\right| = 0$$

for each $\delta>0$. We then have that

$$\begin{aligned}
\int_{\mathbb{R}}|\varphi_\delta(x)-\varphi(x)|\,\mathfrak{G}(dx) &\le 2\|\varphi\|_{C(\mathbb{R})}\mathfrak{G}(\mathbb{R}\setminus[-L,L]) + \sup_{|x|\le L}|\varphi_\delta(x)-\varphi(x)| \\
\mathbb{E}\left[\left|\varphi(S_n/\sqrt{n})-\varphi_\delta(S_n/\sqrt{n})\right|\right] &\le 2\|\varphi\|_{C(\mathbb{R})}\mathbb{P}\left\{|S_n/\sqrt{n}|\ge L\right\} + \sup_{|x|\le L}|\varphi_\delta(x)-\varphi(x)|.
\end{aligned}$$

We thus have that

$$\lim_{n\to\infty}\left|\mathbb{E}\left[\varphi(S_n/\sqrt{n})\right] - \int_{\mathbb{R}}\varphi(x)\mathfrak{G}(dx)\right| \;\le\; 2\sup_{|x|\le L}|\varphi_\delta(x)-\varphi(x)| + 2\|\varphi\|_{C(\mathbb{R})}\left\{\mathfrak{G}(\mathbb{R}\setminus[-L,L]) + \frac{1}{L^2}\right\}.$$

Let $\delta \searrow 0$ and $L \to \infty$. □

3. Large Deviations

Let's now go back to the weak law of large numbers, and study *how quickly does convergence occurs*. This is the subject of *large deviations*.

Let's first carry out a Gaussian calculation, just because we can. Let $\{\eta_i;\, i \in \mathbb{N}\}$ be an independent collection of identically-distributed $\mathfrak{N}(0,1)$ random variables. Define

$$\tilde{S}_n \stackrel{\text{def}}{=} \sum_{j=1}^{n} \eta_j$$

Then $n^{-1}\tilde{S}_n$ is $\mathfrak{N}(0, 1/n)$. We have

LEMMA 3.1. *For any $L > 0$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{|n^{-1}\tilde{S}_n| \geq L\} = -\frac{L^2}{2}. \tag{5}$$

PROOF. First, we note that

$$\mathbb{P}\{|\tilde{S}_n/n| \geq L\} = \int_{|z|\geq L} \frac{1}{\sqrt{2\pi/n}} \exp\left[-nz^2/2\right] dz.$$

For any $\delta > 0$, we thus have that

$$\mathbb{P}\{|\tilde{S}_n/n| \geq L\} \geq \int_{L\leq|z|\leq L+\delta} \frac{1}{\sqrt{2\pi/n}} \exp\left[-nz^2/2\right] dz \geq \exp\left[-\frac{1}{2}(L+\delta)^2\right] \frac{2\delta}{\sqrt{2\pi/n}}.$$

Thus

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{|n^{-1}\tilde{S}_n| \geq L\} \geq -\frac{(L+\delta)^2}{2}.$$

so in fact we have the *lower bound*

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{|n^{-1}\tilde{S}_n| \geq L\} \geq -\frac{L^2}{2}.$$

We also have that for each $\delta > 0$

$$\begin{aligned}
\mathbb{P}\{|\tilde{S}_n/n| \geq L\} &\leq \int_{|z|\geq L} \frac{1}{\sqrt{2\pi/n}} \exp\left[-n(1-\delta)z^2/2\right] \exp\left[-n\delta z^2/2\right] dz \\
&\leq \exp\left[-n(1-\delta)L^2/2\right] \int_{|z|\geq L} \frac{1}{\sqrt{2\pi/n}} \exp\left[-n\delta z^2/2\right] dz \\
&\leq \exp\left[-n(1-\delta)L^2/2\right] \int_{z\in\mathbb{R}} \frac{1}{\sqrt{2\pi/n}} \exp\left[-n\delta z^2/2\right] dz = \sqrt{\frac{n}{\delta}} \exp\left[-n(1-\delta)L^2/2\right].
\end{aligned}$$

Thus

$$\overline{\lim}_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{|n^{-1}\tilde{S}_n| \geq L\} \geq -\frac{L^2(1-\delta)^2}{2}$$

so in fact we have the *upper bound*

$$\overline{\lim}_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{|n^{-1}\tilde{S}_n| \geq L\} \geq -\frac{L^2}{2}.$$

□

If we define $I_G(z) \stackrel{\text{def}}{=} \frac{z^2}{2}$ for all $z \in \mathbb{R}$, we thus can rewrite (5) as

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{|n^{-1}\tilde{S}_n| \geq L\} = -\inf_{|z|\geq L} I_G(z).$$

DEFINITION 3.2 (Large deviations). A collection $\{X_n\}$ of random variables taking values in some Polish space X has a large deviations principle with rate function $I : X \to [0,\infty]$ if

(a) For every $s \geq 0$, the set

$$\Phi(s) \stackrel{\text{def}}{=} \{x \in X : I(x) \leq s\}$$

is compact.

(b) For every $F \subset \mathbb{R}$ closed,

$$\overline{\lim_{n\to\infty}} n^{-1} \log \mathbb{P}\{S_n/n \in F\} \leq -\inf_{x\in F} I(x).$$

(c) For every $G \subset \mathbb{R}$ open,

$$\underline{\lim_{n\to\infty}} n^{-1} \log \mathbb{P}\{S_n/n \in G\} \geq -\inf_{x\in G} I(x).$$

Heuristically, we can write that

$$\mathbb{P}\{X_n \in A\} \asymp \exp\left[-n \inf_{x\in A} I(x)\right]$$

To return to our setting, assume now that

$$M(\theta) \stackrel{\text{def}}{=} \int_{\mathbb{R}} e^{\theta x} \mu(dx) < \infty$$

for all $\theta \in \mathbb{R}$. We want to show that under this assumption, S_n/n has a large deviations principle. To guess what the action functional is, let's first observe a certain way to use the exponential Chebychev inequality.

LEMMA 3.3. *For any $L > 0$ and any $\theta \in \mathbb{R}$,*

$$\mathbb{P}\left\{\theta S_n/n > L\right\} \leq \exp\left[-n(L - \log M(\theta))\right]$$

for all n.

PROOF. We have that

$$\mathbb{P}\left\{\theta S_n > nL\right\} \leq e^{-nL} \mathbb{E}\left[\exp\left[n\theta X_n\right]\right].$$

Note that

$$\mathbb{E}\left[\exp\left[n\theta X_n\right]\right] = \mathbb{E}\left[\prod_{i=1}^{n} e^{\theta \xi_i}\right] = M(\theta)^n.$$

□

From this we can immediately see that for any $L > 0$ and any $\theta > 0$,

$$\mathbb{P}\{S_n/n \geq L\} = \mathbb{P}\{\theta S_n/n \geq \theta L\} \leq \exp\left[-n(L\theta - \log M(\theta))\right]$$

Thus

$$\overline{\lim_{n\to\infty}} n^{-1} \log \mathbb{P}\{S_n/n \geq L\} \leq \inf_{\theta>0}\{-\theta x + \log M(\theta)\} = -\sup_{\theta>0}\{\theta x \log M(\theta)\}.$$

Define now

$$I(x) \stackrel{\text{def}}{=} \sup_{\theta\in\mathbb{R}} \{\theta x - \log M(\theta)\} \qquad \theta \in \mathbb{R}$$

(I is the *Legendre-Fenchel* transform of $\log M$). We will show that $X_n \stackrel{\text{def}}{=} S_n/n$ has a large deviations principle (in $\mathbb{R}$) with rate function I.

LEMMA 3.4. *For every $s \geq 0$, $\Phi(s)$ is compact.*

PROOF. For convenience, define

$$f_\theta(x) \stackrel{\text{def}}{=} \theta x - \log M(\theta) \qquad x \in \mathbb{R}$$

for each $\theta \in \mathbb{R}$. Then

$$\left\{x \in \mathbb{R} : \sup_{\theta\in\mathbb{R}} f_\theta(x) \leq s\right\} = \bigcap_{\theta\in\mathbb{R}} \left\{x \in \mathbb{R} : f_\theta(x) \leq s\right\}.$$

Since f_θ is continuous for each θ, we have written $\Phi(s)$ as an intersection of closed sets. Thus $\Phi(s)$ is clearly closed. We also note that

$$\Phi(s) \subset \left\{x \in \mathbb{R} : f_1(x) \leq s\right\} \cap \left\{x \in \mathbb{R} : f_{-1}(x) \leq s\right\}.$$

If $f_1(x) \le s$ and $f_{-1}(x) \le s$, then $x \le s + \log M(1)$ and $-x \le s + \log M(-1)$, so

$$\Phi(s) \subset [-s - \log M(-1), s + \log M(1)]$$

so $\Phi(s)$ is also bounded. Thus $\Phi(s)$ is compact. □

We now want to prove the large deviations upper bound. We begin with *compact* sets.

LEMMA 3.5. *For every $K \subset\subset \mathbb{R}$,*

$$\overline{\lim_{n\to\infty}}\, n^{-1} \log \mathbb{P}\{S_n/n \in K\} \le -\inf_{x\in K} I(x). \tag{6}$$

PROOF. First fix $s < \inf_{x\in K} I(x)$, and note that

$$K \subset \{x \in \mathbb{R} : I(x) > s\} = \bigcup_{\theta\in\mathbb{R}} \{x \in \mathbb{R} : \theta x - \log M(\theta) > s\}$$

Since we thus cover K by a collection of open sets, we can extract a finite subcover; there is a finite subset Θ of $\mathbb{R}$ such that

$$K \subset \bigcup_{\theta\in\Theta} \{x \in \mathbb{R} : \theta x > s + \log M(\theta)\}$$

Now note that for any $\theta \in \Theta$, Lemma 3.3 implies that

$$\mathbb{P}\{\theta S_n/n > s + \log M(\theta)\} \le e^{-sn}.$$

Thus, $\mathbb{P}\{S_n/n \in K\} \le |\Theta| e^{-sn}$ for all n, and hence

$$\mathbb{P}\{S_n/n \in K\} \le |\Theta| e^{-sn},$$

so

$$\overline{\lim_{n\to\infty}}\, n^{-1} \log \mathbb{P}\{S_n/n \in K\} \le -s$$

and thus (6) holds. □

Now we prove *exponential tightness*;

LEMMA 3.6. *We have that*

$$\overline{\lim_{L\to\infty}}\, \overline{\lim_{n\to\infty}}\, n^{-1} \log \mathbb{P}\{|S_n/n| \ge L\} = -\infty. \tag{7}$$

PROOF. This is an easy consequence of Lemma 3.3. Take $\theta = \pm 1$ and any $L > 0$. Then

$$\overline{\lim_{n\to\infty}}\, n^{-1} \log \mathbb{P}\{S_n/n \ge L\} \le -(L - \log M(1))$$
$$\overline{\lim_{n\to\infty}}\, \log \mathbb{P}\{S_n/n \le -L\} \le -(L - \log M(-1))$$

and this proves (7). □

We now have the full upper bound

PROPOSITION 3.7. *For every closed subset F of $\mathbb{R}$,*

$$\overline{\lim_{n\to\infty}}\, n^{-1} \log \mathbb{P}\{S_n/n \in F\} \le -\inf_{x\in F} I(x). \tag{8}$$

PROOF. For any $L > 0$, we thus have that $F \subset (F \cap [-L, L]) \cup (-L, L)^c$, so

$$\overline{\lim_{n\to\infty}}\, n^{-1} \log \mathbb{P}\{S_n/n \in F\} \le \overline{\lim_{n\to\infty}}\, n^{-1} \log \{\mathbb{P}\{S_n/n \in F \cap [-L, L]\} + \mathbb{P}\{|S_n/n| \ge L\}\}$$
$$\le \max\left\{ -\inf_{z\in F\cap[-L,L]} I(z), \omega(L) \right\}$$

where

$$\omega(L) \stackrel{\text{def}}{=} \overline{\lim_{n\to\infty}}\, n^{-1} \log \mathbb{P}\{|S_n/n| \ge L\};$$

we have from (7) that $\lim_{L\to\infty} \omega(L) = -\infty$. We also have that

$$\lim_{L\to\infty} \inf_{z\in F\cap[-L,L]} I(z) = \inf_{z\in F} I(z).$$

Thus we have the desired upper bound (8). □

To prove the lower bound, let's first study I a bit more closely under a regularity assumption.

LEMMA 3.8. *Assume that* $\mu((\alpha,\beta)) > 0$ *for all* $\alpha < \beta$ (*i.e.,* $\operatorname{supp}\mu = \mathbb{R}$). *Then*

$$I(x) = \max_{\theta\in\mathbb{R}} \{\theta x - \log M(\theta)\}.$$

PROOF. For each $x \in \mathbb{R}$, define

$$\tilde{f}_x(\theta) \stackrel{\text{def}}{=} \theta x - \log M(\theta) = -\log \int_{z\in\mathbb{R}} e^{\theta(z-x)}\mu(dz) \qquad \theta \in \mathbb{R}$$

Each $\tilde{f}_x$ is continuous. We also note that $x \mapsto -\log x$ is decreasing on $(0,\infty)$. If $\theta > 0$, then

$$\tilde{f}_x(\theta) \le -\log \int_{z>x+1} e^{\theta(z-x)}\mu(dz) \le -\log \int_{z>x+1} e^{\theta}\mu(dz) = -\theta - \log\mu((x+1,\infty))$$

so $\lim_{\theta\to\infty} \tilde{f}_x(\theta) = -\infty$, and if $\theta < 0$, then

$$\tilde{f}_x(\theta) \le -\log \int_{z<x-1} e^{\theta(z-x)}\mu(dz) \le -\log \int_{z<x+1} e^{-\theta}\mu(dz) = \theta - \log\mu((-\infty,x+1))$$

so $\lim_{\theta\to-\infty} \tilde{f}_x(\theta) = -\infty$. Thus the claim follows. □

This allows us to prove the lower bound under the above regularity assumption.

LEMMA 3.9. *Assume that* $\operatorname{supp}\mu = \mathbb{R}$. *Then for every* $G \subset \mathbb{R}$ *open,*

$$\varliminf_{n\to\infty} n^{-1}\log \mathbb{P}\{S_n/n \in G\} \le -\inf_{x\in G} I(x).$$

PROOF. It is sufficient to prove that for any $x \in G$,

$$\varliminf_{\delta\to 0} \varliminf_{n\to\infty} n^{-1}\log \mathbb{P}\{S_n/n \in G\} \ge -I(x). \tag{9}$$

This is trivial if $I(x) = \infty$, so we further assume that $I(x) < \infty$.

Fix $\theta \in \mathbb{R}$ such that

$$I(x) = \theta x - \log M(\theta).$$

Note that the first-order condition of optimality for θ is that

$$\frac{\dot{M}(\theta)}{M(\theta)} = x; \tag{10}$$

we will use this later. Now fix any $\delta > 0$ such that $(x-\delta, x+\delta) \subset G$ (possible since G is open). We now write that

$$\begin{aligned}\mathbb{P}\{S_n/n \in G\} &\ge \mathbb{P}\{|S_n/n - x| < \delta\} \\ &= \mathbb{E}\left[\chi_{\{|S_n/n-x|<\delta\}} \exp\left[\theta S_n - n\log M(\theta)\right] \exp\left[-n\left(\theta S_n/n - \log M(\theta)\right)\right]\right]\end{aligned}$$

Note that

$$\mathbb{P}'_{n,\theta}(A) \stackrel{\text{def}}{=} \mathbb{E}\left[\chi_A \exp\left[\theta S_n - n\log M(\theta)\right]\right] = \frac{\mathbb{E}\left[\chi_A \prod_{i=1}^n e^{\theta\xi_i}\right]}{\mathbb{E}\left[\prod_{i=1}^n e^{\theta\xi_i}\right]} \qquad A \in \mathscr{F}$$

is a probability measure on $(\Omega, \mathscr{F})$. Also note that if $|S_n/n - x| < \delta$, then by our choice of θ,

$$\theta S_n/n - \log M(\theta) \le I(x) + \|\theta\|\delta.$$

Thus

$$\mathbb{P}\{|S_n/n - x| < \delta\} \ge \mathbb{P}'_{n,\theta}\{|S_n/n - x| < \delta\} \exp\left[-n\left(I(x) + \|\theta\|\delta\right)\right].$$

Let's next look at the statistics of the ξ_i's under $\mathbb{P}'_{n,\theta}$. For any $\{A_i;\, i = 1, 2 \ldots n\}$ in $\mathscr{B}(\mathbb{R})$,

$$\mathbb{P}'_{n,\theta}\left(\bigcap_{i=1}^n \{\xi_i \in A_i\}\right) = \frac{\mathbb{E}\left[\prod_{i=1}^n \chi_{A_i}(\xi_i) \prod_{i=1}^n e^{\theta\xi_i}\right]}{\mathbb{E}\left[\prod_{i=1}^n e^{\theta\xi_i}\right]} = \prod_{i=1}^n \frac{\mathbb{E}\left[\chi_{A_i}(\xi_i) e^{\theta\xi_i}\right]}{\mathbb{E}\left[e^{\theta\xi_i}\right]} = \prod_{i=1}^n \mu_\theta(A_i)$$

where

$$\mu_\theta(A) \stackrel{\text{def}}{=} \frac{\int_{z\in A} e^{\theta z}\mu(dz)}{M(\theta)}. \qquad A \in \mathscr{B}(\mathbb{R})$$

Thus, under $\mathbb{P}'_{\theta,n}$, the $\xi_1, \xi_2 \dots \xi_n$ are i.i.d. with law μ_θ. Note that the expected value of the ξ_i's (for $1 \le i \le n$) under $\mathbb{P}'_{n,\theta}$ is exactly

$$\int_{z\in\mathbb{R}} z\mu_\theta(dz) = \frac{\int_{z\in\mathbb{R}} ze^{\theta z}\mu(dz)}{M(\theta)} = \frac{\dot{M}(\theta)}{M(\theta)} = x$$

by (10). Note also that

$$\int_{z\in\mathbb{R}} z^2\mu_\theta(dz) = \frac{\int_{z\in\mathbb{R}} z^2e^{\theta z}\mu(dz)}{M(\theta)} = \frac{\ddot{M}(\theta)}{M(\theta)},$$

so the variance of the ξ_i's (for $1 \le i \le n$) is

$$\frac{\ddot{M}(\theta)}{M(\theta)} - \frac{\dot{M}(\theta)}{M(\theta)}$$

which is finite. By Chebychev's inequality,

$$\mathbb{P}'_{n,\theta}\left\{|S_n/n - x| < \delta\right\} \ge 1 - \mathbb{P}'_{n,\theta}\left\{|S_n/n - x| \ge \delta\right\} \ge 1 - n^{-1}\left(\frac{\ddot{M}(\theta)}{M(\theta)} - \frac{\dot{M}(\theta)}{M(\theta)}\right).$$

Thus

$$\underline{\lim}_{n\to\infty}\, n^{-1}\log\mathbb{P}\left\{|S_n/n - x| < \delta\right\} \ge -n(I(x) + \|\theta\|\delta).$$

Now let δ tend to zero; we get (9). □

Let's now prove the full lower bound.

LEMMA 3.10. *For every $G \subset \mathbb{R}$ open,*

$$\overline{\lim}_{n\to\infty}\, n^{-1}\log\mathbb{P}\{S_n/n \in G\} \le -\inf_{x\in G} I(x).$$

This holds even if $\operatorname{supp}\mu \ne \mathbb{R}$.

PROOF. Enlarge $(\Omega, \mathscr{F}, \mathbb{P})$ as necessary to support a independent collection $\{\eta_1, \eta_2 \dots\}$ of $\mathfrak{N}(0,1)$ random variables which are independent of the ξ_j's. For each $\varepsilon > 0$, define

$$\xi_j^\varepsilon \stackrel{\text{def}}{=} \xi_j + \varepsilon\eta_j \qquad j \in \mathbb{N}$$
$$S_n^\varepsilon \stackrel{\text{def}}{=} \sum_{j=1}^n \xi_j^\varepsilon = S_n + \varepsilon\sum_{j=1}^n \eta_j. \qquad n \in \mathbb{N}$$

Then the ξ_j^ε's are independent and identically distributed with common law

$$\mu^\varepsilon(A) \stackrel{\text{def}}{=} \int_{z\in A} f^\varepsilon(z)dz \qquad A \in \mathscr{B}(\mathbb{R})$$

where

$$f^\varepsilon(x) \stackrel{\text{def}}{=} \int_{z\in\mathbb{R}} (2\pi\varepsilon^2)^{-1/2}\exp\left[-\frac{x^2}{2\varepsilon^2}\right]\mu(dz) \qquad x \in \mathbb{R}$$

Thus $\operatorname{supp}\mu^\varepsilon = \mathbb{R}$. We furthermore have that

$$M_\varepsilon(\theta) \stackrel{\text{def}}{=} \int_{z\in\mathbb{R}} e^{\theta z}\mu_\varepsilon(dz) = M(\theta)e^{-\varepsilon^2\theta^2/2}$$

for all $\theta \in \mathbb{R}$ and we define

$$I_\varepsilon(x) \stackrel{\text{def}}{=} \sup_{\theta\in\mathbb{R}}\{\theta x - \log M_\varepsilon(\theta)\} = \sup_{\theta\in\mathbb{R}}\left\{\theta x - \log M(\theta) - \frac{\theta^2\varepsilon^2}{2}\right\} \le I(x).$$

Fix now $\delta > 0$ and $x^* \in G$ such that $I(x^*) < \inf_{x\in G} I(x) + \delta$. Fix next $\delta' > 0$ such that $(x^* - \delta', x^* + \delta') \subset G$. Then by Lemma 3.9

$$\underline{\lim}_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left\{\left|\frac{S_n^\varepsilon}{n} - x^*\right| < \delta'/2\right\} \ge -\inf_{|z-z^*|<\delta'/2} I_\varepsilon(z) \ge -I_\varepsilon(x^*) \ge -I(x^*) + \delta \ge -\inf_{z\in G} I(z) + \delta. \tag{11}$$

We now compute that

$$\mathbb{P}\left\{\left|\frac{S_n}{n}-x^*\right|<\delta'/2\right\}\le\mathbb{P}\left\{\left|\frac{S_n}{n}-x^*\right|<\delta'\right\}+\mathbb{P}\left\{\left|\frac{S_n-S_n^\varepsilon}{n}\right|\ge\delta'/2\right\}.$$

and thus

$$\begin{aligned}\underline{\lim}_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left\{S_n/n\in G\right\}&\ge\underline{\lim}_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left\{\left|\frac{S_n}{n}-x^*\right|<\delta'\right\}\\&\ge\underline{\lim}_{n\to\infty}\frac{1}{n}\log\left\{\mathbb{P}\left\{\left|\frac{S_n^\varepsilon}{n}-x^*\right|<\delta'/2\right\}-\mathbb{P}\left\{\left|\frac{S_n-S_n^\varepsilon}{n}\right|\ge\delta'/2\right\}\right\}.\end{aligned}\tag{12}$$

From Lemma L:GaussLDP we have that

$$\overline{\lim_{\varepsilon\searrow 0}}\,\overline{\lim_{n\to\infty}}\frac{1}{n}\log\mathbb{P}\left\{\left|\frac{S_n-S_n^\varepsilon}{n}\right|\ge\delta'/2\right\}=-\infty$$

and combining this and (11) in (12), we get the desired result. □

Exercises

(1) Show that if M is finite on a region (a,b), it is infinitely differentiable on (a,b).

(2) Fix $f\in C(\mathbb{R})$ is such that $\lim_{|x|\to\infty}f(x)=-\infty$ and $\int_{\mathbb{R}}e^{f(x)}dx<\infty$. Show (directly) that for any measurable subset A of $\mathbb{R}$,

$$\underline{\lim}_{n\to\infty}\,n^{-1}\log\int_{x\in A}e^{nf(x)}dx\ge\sup_{x\in A^\circ}f(x)$$
$$\overline{\lim}_{n\to\infty}\,n^{-1}\log\int_{x\in A}e^{nf(x)}dx\le\sup_{x\in\bar{A}}f(x)$$

(3) Let $\mu=\lambda\delta_a+(1-\lambda)\delta_b$, for fixed $\lambda\in[0,1]$ and a and b in $\mathbb{R}$ such that $a<b$. Compute $M(\theta)$ and $I(x)$.

(4) Show that $I(\bar{\xi})=0$, where $\bar{\xi}\stackrel{\text{def}}{=}\int_{\mathbb{R}}x\mu(dx)$.

(5) Show that η is $\mathfrak{N}(m,\sigma^2)$ if and only if $\mathbb{E}[e^{i\theta\eta}]=e^{-\sigma^2\theta^2/2+m\theta}$ for all $\theta\in\mathbb{R}$.

(6) Show that if η is $\mathfrak{N}(m,\sigma^2)$, then its mean is m and its variance is σ^2. Hint: use the characteristic function.

(7) Show that if $\sigma>0$, then η is $\mathfrak{N}(m,\sigma^2)$ if and only if

$$\mathbb{P}\{\eta\in A\}=\int_A\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]dx.\qquad A\in\mathscr{B}(\mathbb{R})$$

Show that the law of a $\mathfrak{N}(m,0)$ random variable is δ_m.

CHAPTER 4

Martingales

As usual, we assume the existence of an underlying measurable space $(\Omega, \mathscr{F})$; note that for the moment, we are *not* requiring a probability measure on $(\Omega, \mathscr{F})$. Throughout, we fix an index set $I \subset \mathbb{R}$.

DEFINITION 0.11 (Filtration). A collection $\{\mathscr{F}_t;\, t \in I\}$ of sub-sigma-algebras of $\mathscr{F}$ is called a filtration of $\mathscr{F}$ if $\mathscr{F}_s \subset \mathscr{F}_t$ for all s and t in I such that $s \leq t$.

DEFINITION 0.12 (Stochastic process). Fix a measurable space $(S, \mathscr{S})$. Then a collection $\{X_t;\, t \in I\}$ of S-valued random variables is called a stochastic process.

In practice, filtrations are often related to stochastic processes. We can ask that a stochastic process "follow" a filtration

DEFINITION 0.13 (Adapted process). If we have a filtration $\{\mathscr{F}_t;\, t \in I\}$ and a stochastic process $\{X_t;\, t \in I\}$, then the stochastic process is said to be adapted to the filtration if X_t is $\mathscr{F}_t$-measurable for each $t \in I$; i.e., if $\sigma\{X_t\} \subset \mathscr{F}_t$ for all $t \in I$.

We can also generate a filtration by a stochastic process. To so so, let's first make a definition.

DEFINITION 0.14 (Filtration defined by a process). Let $\{X_t;\, t \in I\}$ be a stochastic process. We then define

$$\mathscr{F}_t^X \stackrel{\text{def}}{=} \sigma\{X_s;\, s \leq t\}.$$

Often it is useful to *stop* based upon current knowledge.

DEFINITION 0.15 (Stopping time). Let $\{\mathscr{F}_t;\, t \in I\}$ be a filtration. An I-valued random variable τ is a called stopping time if $\{\tau \leq t\} \in \mathscr{F}_t$ for all $t \in I$.

We can then randomly "stop" the filtration

DEFINITION 0.16 (Stopped filtration). Let $\{\mathscr{F}_t;\, t \in I\}$ be a filtration and let τ be a stopping time. We define

$$\mathscr{F}_\tau \stackrel{\text{def}}{=} \{A \in \mathscr{F} :\, A \cap \{\tau \leq t\} \in \mathscr{F}_t \text{ for all } t \in I\}.$$

We will have a lot more to say about stopping times when we consider martingales.

Let's now consider the following setup. Let $\{\mathscr{F}_n;\, n \in \mathbb{N}\}$ be a filtration of $(\Omega, \mathscr{F})$. We then define

DEFINITION 0.17 (Martingale). An adapted collection $X = \{X_n;\, n \in \mathbb{N}\}$ of integrable random variables is a martingale if

$$\mathbb{E}[X_{n+1}|\mathscr{F}_n] = X_n$$

for all $n \in \mathbb{N}$.

DEFINITION 0.18 (Supermartingale). An adapted collection $X = \{X_n;\, n \in \mathbb{N}\}$ of integrable random variables is a supermartingale if

$$\mathbb{E}[X_{n+1}|\mathscr{F}_n] \leq X_n$$

for all $n \in \mathbb{N}$.

DEFINITION 0.19 (Submartingale). An adapted collection $X = \{X_n;\, n \in \mathbb{N}\}$ of integrable random variables is a submartingale if

$$\mathbb{E}[X_{n+1}|\mathscr{F}_n] \geq X_n$$

for all $n \in \mathbb{N}$.

Note, of course that the negative of a submartingale is a supermartingale and vice versa, and that a process is a martingale if and only if it is a supermartingale and submartingale.

We will study martingale *inequalities* and *convergence*. It turns out that this topic has connections to a lot of probability theory. One may look at the above setup as a general framework for considering the evolution of information. This is important in itself, and it is also important for the characterization of Markov processes (which we shall not touch upon in this course).

First of all, let's extend the above properties from fixed times to stopping times.

PROPOSITION 0.20 (Optional Sampling Theorem (Doob)). *Suppose that X is a supermartingale and ρ and τ are bounded stopping times with $\rho \leq \tau$. Then*

$$\mathbb{E}[X_\tau|\mathscr{F}_\rho] \leq X_\rho. \tag{13}$$

PROOF. We will prove the result via three steps

Step 1. First, assume that $\rho \equiv k$ and $k \leq \tau \leq k+1$ (i.e., ρ is constant and τ can vary at most by 1. Then

$$\begin{aligned}\mathbb{E}[X_\tau|\mathscr{F}_\rho] = \mathbb{E}[X_\tau|\mathscr{F}_k] &= \mathbb{E}[X_\tau\chi_{\{\tau>k\}}|\mathscr{F}_k] + \mathbb{E}[X_\tau\chi_{\{\tau\leq k\}}|\mathscr{F}_k] \\ &= \mathbb{E}[X_{k+1}|\mathscr{F}_k]\chi_{\{\tau>k\}} + X_k\chi_{\{\tau\leq k\}} \leq X_k\chi_{\{\tau>k\}} + X_k\chi_{\{\tau\leq k\}} = X_k = X_\rho\end{aligned}$$

The third equality comes from the fact that if $\tau > k$, then $\tau = k+1$ and if $\tau \leq k$, then $\tau = k$, and that $\{\tau > k\}$, $\{\tau \leq k\}$, and X_k are $\mathscr{F}_k$-measurable.

Step 2. Now assume that ρ is simply a bounded stopping time (assume M is an upper bound for ρ and that $\rho \leq \tau \leq \rho + 1$). To show (13), we will show that for any $A \in \mathscr{F}_\rho$,

$$\mathbb{E}[X_\tau\chi_A] \leq \mathbb{E}[X_\rho\chi_X].$$

Define the stopping time $\tilde{\tau}_k \stackrel{\text{def}}{=} \min\{\max\{\tau, k\}, k+1\}$ for each $k \geq 0$; then if $\rho = k$, $\tau = \tilde{\tau}_k$. We calculate that

$$\begin{aligned}\mathbb{E}[X_\tau\chi_A] = \sum_{k=0}^{M}\mathbb{E}[X_\tau\chi_A\chi_{\{\rho=k\}}] &= \sum_{k=0}^{M}\mathbb{E}[\mathbb{E}[X_\tau\chi_{A\cap\{\rho=k\}}|\mathscr{F}_k]] \\ &= \sum_{k=0}^{M}\mathbb{E}[\mathbb{E}[X_{\tilde{\tau}_k}|\mathscr{F}_k]\chi_{A\cap\{\rho=k\}}] \leq \sum_{k=0}^{M}\mathbb{E}[X_k\chi_{A\cap\{\rho=k\}}] = \mathbb{E}[X_\rho\chi_A].\end{aligned}$$

The third equality uses several facts. First, note that $A\cap\{\rho=k\} = (A\cap\{\rho\leq k\})\setminus(A\cap\{\rho\leq k-1\})$ and that $A\cap\{\rho\leq k\} \in \mathscr{F}_k$ and $A\cap\{\rho\leq k-1\} \in \mathscr{F}_{k-1} \subset \mathscr{F}_k$; thus $A\cap\{\rho=k\}$ is $\mathscr{F}_k$ measurable. Also, note that $\tau = \tilde{\tau}_k$ on $\rho = k$. The first inequality stems from Step 1 (since $k \leq \tilde{\tau}_k \leq k+1$).

Step 3. Now assume that ρ and τ are simply bounded stopping times and $\rho \leq \tau$. For each $j \geq 0$, define the stopping time $\hat{\tau}_j \stackrel{\text{def}}{=} \min\{\tau, \rho + j\}$. Then the $\hat{\tau}_j$'s are stopping times, $\hat{\tau}_0 = \rho$ and $\hat{\tau}_M = \tau$, and $\tau_j \leq \tau_{j+1} \leq \tau_j + 1$. Thus for every $0 \leq j \leq M-1$,

$$\mathbb{E}[X_{\hat{\tau}_{j+1}}|\mathscr{F}_\rho] = \mathbb{E}[\mathbb{E}[X_{\hat{\tau}_{j+1}}|\mathscr{F}_{\hat{\tau}_j}]|\mathscr{F}_\rho] \leq \mathbb{E}[X_{\hat{\tau}_j}|\mathscr{F}_\rho].$$

By induction, we then get (13) in full generality. □

Note that if X is a submartingale, then $\mathbb{E}[X_{n+1}] \geq \mathbb{E}[X_n]$, so X is in some sense "increasing". Let's investigate this. Note that an increasing function has two obvious properties: that it is bounded from above on any interval by its value at the end of that interval, and it is not increasing.

PROPOSITION 0.21 (Doob's Maximal Inequality). *Suppose that X is either a martingale or a nonnegative submartingale. Then for any n and any L,*

$$\mathbb{P}\left\{\max_{0\leq k\leq n}|X_k| \geq L\right\} \leq \frac{\mathbb{E}\left[|X_n|\chi_{\{\max_{0\leq k\leq n}|X_k|\geq L\}}\right]}{L}.$$

PROOF. Set

$$\tau \stackrel{\text{def}}{=} \min\{k \geq 0 : |X_k| \geq L\}.$$

Then

$$\left\{\max_{0\le k\le n}|X_k|\ge L\right\}=\{|X_{\tau\wedge n}|\ge L\}=\{\tau\le n\}\in\mathscr{F}_\tau\cap\mathscr{F}_n=\mathscr{F}_{\tau\wedge n}.$$

By Chebychev's inequality,

$$\mathbb{P}\left\{\max_{0\le k\le n}|X_k|\ge L\right\}=\mathbb{P}\{|X_{\tau\wedge n}|\ge L\}\le\frac{\mathbb{E}\left[|X_{\tau\wedge n}|\chi_{\{|X_{\tau\wedge n}|\ge L\}}\right]}{L}. \tag{14}$$

By Optional Sampling (and Jensen's inequality if X is a martingale),

$$\mathbb{E}[|X_n||\mathscr{F}_{\tau\wedge n}]\ge|X_{\tau\wedge n}|.$$

Thus

$$\mathbb{E}\left[|X_n|\chi_{\{|X_{\tau\wedge n}|\ge L\}}\right]\le\mathbb{E}\left[\mathbb{E}[|X_n||\mathscr{F}_{\tau\wedge n}]\chi_{\{\tau\le n\}}\right]=\mathbb{E}\left[\mathbb{E}[|X_n|\chi_{\{\tau\le n\}}|\mathscr{F}_{\tau\wedge n}]\right]=\mathbb{E}[|X_n|\chi_{\{\tau\le n\}}].$$

Use this in (14) to complete the proof. □

Let's use this to get

COROLLARY 0.22. *Under the same assumptions as in Doob's maximal inequality,*

$$\mathbb{E}\left[\max_{0\le k\le n}|X_k|^p\right]^{1/p}\le\left(\frac{p}{p-1}\right)\mathbb{E}[|X_n|^p]^{1/p}$$

for any $1<p<\infty$.

PROOF. For convenience, define the maximal function

$$X_n^*\stackrel{\text{def}}{=}\max_{0\le k\le n}|X_k|.$$

If $X_n^*=0$, the result is trivial, so we assume that $X_n^*>0$. Let $q=p/(p-1)$ (i.e., $p^{-1}+q^{-1}=1$). We now calculate that

$$\begin{aligned}\mathbb{E}[(X_n^*)^p]&=p\int_0^\infty t^{p-1}\mathbb{P}\{X_n^*\ge t\}dt\le p\int_0^\infty t^{p-2}\mathbb{E}[|X_n|\chi_{\{X_n^*\ge t\}}]dt=p\mathbb{E}\left[|X_n|\int_0^{X_n^*}t^{p-2}dt\right]\\&=\frac{p}{p-1}\mathbb{E}\left[|X_n|(X_n^*)^{p-1}\right]\le\frac{p}{p-1}\mathbb{E}[|X_n|^p]^{1/p}\mathbb{E}\left[(X_n^*)^{p-1}q\right]^{1/q}.\end{aligned}$$

We use Doob's maximal inequality to get the first inequality and we use Hölder's inequality to get the last inequality. Noting that $(p-1)q=p$ and rearranging, we get the desired result. □

Now note that if a function is "increasing", it crosses any interval at most once. We can also generalize this to submartingales.

PROPOSITION 0.23 (Doob's Upcrossing Inequality). *Let X be a submartingale and fix $a<b$. Define*

$$\begin{aligned}\sigma_1&\stackrel{def}{=}\min\{n\ge0:\,X_n\le a\}\\\tau_1&\stackrel{def}{=}\min\{n\ge\sigma_1:\,X_n\ge b\}\\\sigma_k&\stackrel{def}{=}\min\{n\ge\tau_{k-1}:\,X_n\le a\}\qquad k\ge2\\\tau_k&\stackrel{def}{=}\min\{n\ge\sigma_{k-1}:\,X_n\ge b\}\qquad k\ge2\end{aligned}$$

For each n, define now

$$U_n^{a,b}\stackrel{def}{=}|\{k\ge1:\,\tau_k\le n\}|.$$

($U_n^{a,b}$ is the number of upcrossings of (a,b) by X by the time n). Then $U_n^{a,b}$ is measurable and

$$\mathbb{E}[U_n^{a,b}]\le\frac{\mathbb{E}[(X_n-a)^+]}{(b-a)}. \tag{15}$$

PROOF. First, note that for any L,

$$\{U_n^{a,b} \geq L\} = \{\tau_{\lceil L \rceil} \leq n\},$$

so $U_n^{a,b}$ is indeed measurable.

Define now

$$Y_n \stackrel{\text{def}}{=} (X_n - a)^+; \qquad n \in \mathbb{N}$$

since $x \mapsto (x-a)^+$ is nondecreasing, convex, and nonnegative, Y is a nonnegative submartingale (this reduces our original problem to a simpler one). We first claim that for any $k \leq n$,

$$Y_{\tau_k \wedge n} - Y_{\sigma_k \wedge n} \geq (b-a)\chi_{\{\tau_k \leq n\}}.$$

If $\sigma_k \geq n$, this is clearly true since then both sides are zero. If $\tau_k \leq n$, it is also true since then $Y_{\tau_k \wedge n} = Y_{\tau_k} = (b-a)$ and $Y_{\sigma_k \wedge n} = Y_{\sigma_k} = 0$. Finally, if $\sigma_k < n < \tau_k$, then $Y_{\tau_k \wedge n} = Y_n \geq 0$ and $Y_{\sigma_k \wedge n} = Y_{\sigma_k} = 0$. Thus

$$\sum_{k=1}^{n} (Y_{\tau_k \wedge n} - Y_{\sigma_k \wedge n}) \geq (b-a) U_n^{a,b}.$$

Thus

$$(b-a)\mathbb{E}[U_n^{a,b}] \leq \sum_{k=1}^{n} \{\mathbb{E}[Y_{\tau_k \wedge n}] - \mathbb{E}[Y_{\sigma_k \wedge n}]\} = \mathbb{E}[Y_{\tau_n \wedge n}] - \mathbb{E}[Y_{\sigma_n \wedge n}] - \sum_{k=1}^{n-1} \left\{\mathbb{E}[Y_{\sigma_{k+1} \wedge n}] - \mathbb{E}[Y_{\tau_k \wedge n}]\right\} \leq \mathbb{E}[Y_{\tau_n \wedge n}].$$

To get the first equality, we simply rearranged terms. To get the second inequality, we used the fact that Y is nonnegative (hence $\mathbb{E}[Y_{\sigma_n \wedge n}] \geq 0$) and Doob's optional sampling theorem (to show that the sum is nonnegative). Finally, note that $\tau_n \geq 2n$, so $\tau_n \wedge n = n$, and this gives us (15). □

Now note that limits exist if and only if there is no oscillation. We now have

PROPOSITION 0.24 (Submartingale Convergence Theorem). *Suppose that X is a submartingale and $\sup_{n \geq 0} \mathbb{E}[X_n^+] < \infty$. Then $X \stackrel{def}{=} \lim_{n \to \infty} X_n$ exists $\mathbb{P}$-a.s. and $\mathbb{E}[|X|] < \infty$.*

PROOF. For any $a < b$,

$$\mathbb{E}[\lim_{n \to \infty} U_n^{a,b}] = \lim_{n \to \infty} \mathbb{E}[U_n^{a,b}] \leq (b-a)^{-1} \sup_{n \geq 0} \mathbb{E}[(X_n - a)^+] \leq (b-a)^+ \left\{\sup_{n \geq 0} \mathbb{E}[X_n^+] + |a|\right\} < \infty.$$

The first equality comes from monotone convergence and the last inequality comes from the fact that $(x-a)^+ \leq |x| + a$ for all $x \in \mathbb{R}$. Thus

$$\mathbb{P}\left\{\underline{\lim}_{n \to \infty} X_n \leq a \text{ and } \overline{\lim}_{n \to \infty} X_n \geq b\right\} = 0.$$

But

$$\mathbb{P}\left\{\lim_{n \to \infty} X_n \text{ does not exist}\right\} \leq \bigcup_{\substack{a<b \\ a,\, b \text{ rational}}} \mathbb{P}\left\{\underline{\lim}_{n \to \infty} X_n \leq a \text{ and } \overline{\lim}_{n \to \infty} X_n \geq b\right\} = 0.$$

Now note that by Fatou's lemma,

$$\mathbb{E}[X^+] \leq \overline{\lim}_{n \to \infty} \mathbb{E}[X_n^+] \leq \sup_{n \geq 0} \mathbb{E}[X_n^+] < \infty$$

and also

$$\mathbb{E}[X^-] \leq \underline{\lim}_{n \to \infty} \mathbb{E}[X_n^-] \leq \sup_{n \geq 0} \mathbb{E}[X_n^+] - \overline{\lim}_{n \to \infty} \mathbb{E}[X_n] \leq \sup_{n \geq 0} \mathbb{E}[X_n^+] - \mathbb{E}[X_0].$$

The second inequality comes from the fact that $x = x^+ - x^-$ and the third comes from the fact that X is a submartingale. □

From the submartingale convergence theorem, we see that martingales want to converge. Let's follow this thought for a while.

PROPOSITION 0.25. *Let X be a martingale such that X_n converges to X in L^1. Then*

$$X_n = \mathbb{E}[X|\mathscr{F}_n] \qquad n \in \mathbb{N}$$

PROOF. By the definition of conditional expectation, it suffices to show that for any n and any $A \in \mathscr{F}_n$,

$$\mathbb{E}[X\chi_A] = \mathbb{E}[X_n\chi_A].$$

Fix any $m \geq n$. Then

$$\begin{aligned}|\mathbb{E}[X\chi_A] - \mathbb{E}[X_n\chi_A]| &= |\mathbb{E}[X\chi_A] - \mathbb{E}[\mathbb{E}[X_m|\mathscr{F}_n]\chi_A]| \\ &= |\mathbb{E}[X\chi_A] - \mathbb{E}[X_m\chi_A]| = |\mathbb{E}[(X - X_m)\chi_A]| \leq \|X - X_m\|_{L_1}.\end{aligned}$$

Let m tend to infinity. □

We have already understood how to strengthen almost-sure convergence to convergence in L^1—the criterion is uniform integrability; see Proposition 1.10. Thus we have

COROLLARY 0.26. *If X is a uniformly integrable martingale, then $X = \lim_{n\to\infty} X_n$ exists $\mathbb{P}$-a.s. and in L^1, and $X_n = \mathbb{E}[X|\mathscr{F}_n]$ for all n.*

We also have

THEOREM 0.27 (Martingale continuity theorem). *Fix $X \in L^1$ and $\{\mathscr{F}_n;\, n \in \mathbb{N}\}$ a filtration. Set $\mathscr{F}_\infty \stackrel{def}{=} \bigvee_{n\in\mathbb{N}} \mathscr{F}_n$. Then $\lim_{n\to\infty} \mathbb{E}[X|\mathscr{F}_n] = \mathbb{E}[X|\mathscr{F}_\infty]$ $\mathbb{P}$-a.s. and in L^1.*

PROOF. First, consider the martingale

$$Y_n \stackrel{\text{def}}{=} \mathbb{E}[X|\mathscr{F}_n]. \qquad n \in \mathbb{N}$$

We claim that Y is uniformly integrable. Note that for any $n \in \mathbb{N}$ and $K > 0$,

$$\mathbb{E}[|Y_n|\chi_{\{|Y_n|\geq K\}}] = \mathbb{E}[|\mathbb{E}[X|\mathscr{F}_n]|\chi_{\{|Y_n|\geq K\}}] \leq \mathbb{E}[\mathbb{E}[|X||\mathscr{F}_n]\chi_{\{|Y_n|\geq K\}}] = \mathbb{E}|X|\chi_{\{|Y_n|\geq K\}}]. \tag{16}$$

Note that

$$\mathbb{P}\{|Y_n| \geq K\} \leq K^{-1}\mathbb{E}[|Y_n|] \leq K^{-1}\mathbb{E}[|X|].$$

Since X is integrable, this means that

$$\overline{\lim_{K\to\infty}} \sup_{n\in\mathbb{N}} \mathbb{P}\{|Y_n| \geq K\} = 0,$$

which, since X is integrable, allows us to see that Y is uniformly integrable from (16).

Thus $Y = \lim_{n\to\infty} Y_n$ exists $\mathbb{P}$-a.s. and in L^1 and $Y_n = \mathbb{E}[Y|\mathscr{F}_n]$. It remains only to show that $Y = \mathbb{E}[X|\mathscr{F}_\infty]$. For any $n \in \mathbb{N}$ and any $A \in \mathscr{F}_n$,

$$\mathbb{E}[X\chi_A] = \mathbb{E}[\mathbb{E}[X|\mathscr{F}_n]\chi_A] = \mathbb{E}[Y\chi_A].$$

Thus $\mathbb{E}[X\chi_A] = \mathbb{E}[Y\chi_A]$ for any $A \in \cup_{n\in\mathbb{N}}\mathscr{F}_n$ and thus for any $A \in \mathscr{F}_\infty$ (use the monotone class theorem). Thus

$$\mathbb{E}[X|\mathscr{F}_\infty] = \mathbb{E}[Y|\mathscr{F}_\infty] = Y,$$

the last equality holding since Y, which is the $\mathbb{P}$-a.s. limit of a sequence of $\mathscr{F}_\infty$-measurable functions, is itself $\mathscr{F}_\infty$-measurable. □

Finally, let's prove an alternate characterization for submartingales.

THEOREM 0.28 (Doob-Meyer decomposition). *An adapted stochastic process X is a submartingale if and only if*

$$X_n = M_n + A_n, \qquad n \in \mathbb{N} \tag{17}$$

where M is a martingale and A is a nondecreasing integrable process such that A_n is $\mathscr{F}_{n-1}$-measurable for each $n \geq 1$. If A_0 (or alternately M_0) is specified, then this decomposition is unique.

PROOF. First, assume that we have (17). Then for any $n \geq 1$,

$$\mathbb{E}[X_{n+1}|\mathscr{F}_n] = \mathbb{E}[M_{n+1}|\mathscr{F}_n] + \mathbb{E}[A_{n+1}|\mathscr{F}_n] \geq M_n + \mathbb{E}[A_n|\mathscr{F}_n] = M_n + A_n = X_n.$$

Assume now that X is a submartingale. Set

$$\begin{aligned} A_n &\stackrel{\text{def}}{=} \begin{cases} \sum_{j=1}^{n} \left(\mathbb{E}[X_j|\mathscr{F}_{j-1}] - X_{j-1}\right) & \text{if } n \geq 1 \\ 0 & \text{if } n = 0 \end{cases} \qquad n \in \mathbb{N} \\ M_n &\stackrel{\text{def}}{=} X_n - A_n. \end{aligned} \tag{18}$$

Then A is clearly nondecreasing (by the submartingale property of X) and integrable and A_n is $\mathscr{F}_{n-1}$-measurable for each $n \geq 1$. Also X is integrable and adapted and for any $n \geq 0$,

$$\mathbb{E}[M_{n+1}|\mathscr{F}_n] = \mathbb{E}[X_{n+1}|\mathscr{F}_n] - \mathbb{E}[A_{n+1}|\mathscr{F}_n] = \mathbb{E}[X_{n+1}|\mathscr{F}_n] - (\mathbb{E}[X_{n+1}|\mathscr{F}_n] - X_n) - A_n = M_n$$

Finally, consider any decomposition $X = M' + A'$ as in (17). Then for any $n \geq 0$,

$$\mathbb{E}[X_{n+1}|\mathscr{F}_n] = \mathbb{E}[M'_{n+1}|\mathscr{F}_n] + \mathbb{E}[A'_{n+1}|\mathscr{F}_n] = M'_n + A'_{n+1} = X_n + (A'_{n+1} - A'_n).$$

Thus

$$A'_{n+1} = A'_n + \mathbb{E}[X_{n+1}|\mathscr{F}_n] - X_n, \qquad n \in \mathbb{N}$$

and so

$$A'_n = A_n + A'_0 \qquad n \in \mathbb{N}$$

where A is as in (18). □

We simply note for future reference the following definition

DEFINITION 0.29 (Bracket). Let M be a square-integrable martingale. Then $\langle M \rangle$ is the increasing part of the Doob-Meyer decomposition of the submartingale M^2.; i.e., $\langle M \rangle$ is the unique process such that

- $\langle M \rangle_n$ is $\mathscr{F}_{n-1}$-measurable,
- $\langle M \rangle_0 = 0$
- $M^2 - \langle M \rangle$ is a martingale.

Finally, let's discuss some issues of *backward martingales*. The essential difference is that before we were interested in convergence as n tended to ∞; now we are interested in convergence as n tends to $-\infty$. We will

PROPOSITION 0.30 (Backward submartingale convergence theorem[1]). *Let $\{\mathscr{G}_n; n \in -\mathbb{N}\}$ be a filtration and let X be a submartingale with respect to $\{\mathscr{G}_n; n \in -\mathbb{N}\}$. Then $X_{-\infty} \stackrel{def}{=} \lim_{n\to-\infty} X_n$ exists $\mathbb{P}$-a.s. If $\sup_n \mathbb{E}[|X_n|] < \infty$, then X is uniformly integrable, $X_{-\infty} = \lim_{n\to-\infty} X_n$, the limit now being both $\mathbb{P}$-a.s. and in L^1, and $X_{-\infty} \leq \mathbb{E}[X_n|\mathscr{G}_{-\infty}]$ for all $n \in -\mathbb{N}$, where $\mathscr{G}_{-\infty} \stackrel{def}{=} \cap_{n\in-\mathbb{N}}\mathscr{G}_n$.*

PROOF. Since X_0 is assumed to be integrable, Doob's Upcrossing Inequality implies that for every $(a, b) \subset \mathbb{R}$,

$$\mathbb{E}[|\{\text{upcrossings of } (a,b) \text{ by } X \text{ in between times } n \text{ and } 0\}|] \leq \frac{\mathbb{E}[(X_0 - a)^+]}{(b-a)} < \infty$$

for all $n \in -\mathbb{N}$. This implies that $X_{-\infty}$ exists $\mathbb{P}$-a.s., just as for the regular submartingale convergence theorem.

Now assume that $\sup_{n\in-\mathbb{N}} \mathbb{E}[|X_n|] < \infty$; then $\sup_{n\in-\mathbb{N}} |\mathbb{E}[X_n]| < \infty$. Since $n \mapsto \mathbb{E}[X_n]$ is increasing, $\lim_{n\to-\infty} \mathbb{E}[X_n]$ exists and is finite. Fix $\varepsilon > 0$; then there is an $N_0 \in -\mathbb{N}$ such that $\mathbb{E}[X_n] > \mathbb{E}[X_N] - \varepsilon$ if $n \leq N$. For any $K > 0$, we calculate that for any $n \leq N$,

$$\begin{aligned} \mathbb{E}\left[|X_n|\chi_{\{|X_n|\geq K\}}\right] &= \mathbb{E}\left[X_n\chi_{\{X_n\geq K\}}\right] - \mathbb{E}\left[X_n\chi_{\{X_n\leq -K\}}\right] = \mathbb{E}\left[X_n\left(\chi_{\{X_n\geq K\}} + \chi_{\{X_n > -K\}}\right)\right] - \mathbb{E}[X_n] \\ &\leq \mathbb{E}\left[X_N\left(\chi_{\{X_n\geq K\}} + \chi_{\{X_n > -K\}}\right)\right] - \mathbb{E}[X_N] + \varepsilon/2 = \mathbb{E}\left[|X_N|\chi_{\{|X_n|\geq K\}}\right] + \varepsilon/2 \end{aligned}$$

[1]Taken from Revuz and Yor

In the last inequality, the first term comes from the submartingale inequality, and the last term comes from the choice of N. By Markov's inequality,

$$\sup_{n\in-\mathbb{N}} \mathbb{P}\{|X_n| \geq K\} \leq K^{-1} \sup_{n\in-\mathbb{N}} \mathbb{E}[|X_n|].$$

Since X_N is integrable, we have that

$$\overline{\lim_{K\to\infty}} \sup_{n\leq N} \mathbb{E}\left[|X_n|\chi_{\{|X_n|\geq K\}}\right] \leq \varepsilon;$$

since all of the X_n's are integrable, we also have that

$$\overline{\lim_{K\to\infty}} \sup_{N<n\leq 0} \mathbb{E}\left[|X_n|\chi_{\{|X_n|\geq K\}}\right] \leq \varepsilon;$$

putting these two together, we have that the X_n's are uniformly integrable. Hence, the $\mathbb{P}$-a.s. convergence of X_n to $X_{-\infty}$ also holds in L^1. For any $A \in \mathscr{G}_{-\infty}$ and any n, we thus have

$$\mathbb{E}[X_{-\infty}\chi_A] = \lim_{m\to-\infty} \mathbb{E}[X_m\chi_A] \leq \mathbb{E}[X_n\chi_A].$$

The last inequality follows from the submartingale inequality. Thus we get the final claim. □

We can use this result to easily prove the Strong Law of Large Numbers

THEOREM 0.31 (Strong Law of Large Numbers). *Let $\{\xi_1, \xi_2 \dots\}$ be a collection of independent and identically distributed integrable random variables with common law μ. Then*

$$\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} \xi_k = \int_{\mathbb{R}} x\mu(dx),$$

this limit being both almost-sure and in L^1.

PROOF. Set

$$S_n \stackrel{\text{def}}{=} \sum_{k=1}^{n} \xi_k \qquad n \in \mathbb{N}$$

For each $n \in \mathbb{N}$, define

$$\mathscr{G}_{-n} \stackrel{\text{def}}{=} \sigma\{S_k;\, k \geq n\}.$$

Set

$$X_n \stackrel{\text{def}}{=} \mathbb{E}[\xi_1|\mathscr{G}_n]. \qquad n \in -\mathbb{N}$$

Then X is a martingale. Clearly

$$\sup_{n\in-\mathbb{N}} \mathbb{E}[|X_n|] \leq \mathbb{E}[|\xi_1|],$$

so $X_{-\infty} \stackrel{\text{def}}{=} \lim_{n\to-\infty} X_n$ exists both in L^1 and $\mathbb{P}$-a.s. Note that for any $n \geq 1$ and any $1 \leq k \leq n$,

$$\mathbb{E}[\xi_k|\mathscr{G}_{-n}] = \mathbb{E}[\xi_1|\mathscr{G}_{-n}];$$

thus

$$\sum_{k=1}^{n} X_{-n} = \sum_{k=1}^{n} \mathbb{E}[\xi_1|\mathscr{G}_{-n}] = \sum_{k=1}^{n} \mathbb{E}[\xi_k|\mathscr{G}_{-n}] = \mathbb{E}[S_n|\mathscr{G}_{-n}] = S_n$$

so in fact

$$X_{-n} = \frac{1}{n} S_n$$

for all $n \geq 1$. Thus $S_n = X_{-n}$ for all $n \in \mathbb{N}$. Hence

$$X_{-\infty} = \lim_{n\to\infty} \frac{S_n}{n}$$

this limit being both $\mathbb{P}$-a.s. and in L^1. We thus only need show that $X_{-\infty} = \int_{\mathbb{R}} x\mu(dx)$. Note that for every k,

$$X_{-\infty} = \overline{\lim_{n\to\infty}} \frac{S_n}{n} = \overline{\lim_{n\to\infty}} \frac{\sum_{j=1}^{n} \xi_{k+j}}{n}.$$

so by Kolmogorov's zero-one law, $X_{-\infty}$ is almost-surely constant. Hence

$$X_{-\infty} = \mathbb{E}[X_{-\infty}] = \lim_{n\to\infty} n^{-1}\mathbb{E}[S_n] = \int_{\mathbb{R}} x\mu(dx)$$

This completes the proof. □

Exercises

(1) Show that for any stopping time τ, $\mathscr{F}_\tau$ is indeed a sigma-algebra.
(2) Show that for any stopping time τ, τ is itself $\mathscr{F}_\tau$-measurable.
(3) Show that for any fixed $t \in I$, the mapping $\tau : \Omega \mapsto t$ is a stopping time. Show that $\mathscr{F}_\tau = \mathscr{F}_t$.
(4) Let $\{\tau_1, \tau_2 \dots\}$ be a countable collection of stopping times. Show that $\sup_{n\geq 1} \tau_n$ is also a stopping time. Show that if $\{\tau_1, \tau_2 \dots, \tau_n\}$ is a finite collection of stopping times, then $\min_{1\leq k\leq n} \tau_k$ is also a stopping time. stopping times.
(5) Let τ be a stopping time and s a nonnegative number. Show that $\tau + s$ is a stopping time.
(6) Suppose that I is discrete, X is an adapted process taking values in a measurable space $(S, \mathscr{S})$, $A \in \mathscr{S}$, and τ is a stopping time. Define

$$\tau' \stackrel{\text{def}}{=} \min\{t \in I : t \geq \tau \text{ and } X_t \in A\}$$

and show that τ' is a stopping time.
(7) Suppose that $I = \mathbb{R}_+$, X is a continuous process taking values in a topological space S, and $F \subset S$ is closed. Define

$$\tau \stackrel{\text{def}}{=} \inf\{t \geq 0 : X_t \in F\}$$

and show that τ is a stopping time.
(8) Suppose that τ_1 and τ_2 are stopping times. Show that
   (a) If $\tau_1 \leq \tau_2$, then $\mathscr{F}_{\tau_1} \subset \mathscr{F}_{\tau_2}$.
   (b) If $A \in \mathscr{F}_{\tau_1}$, then $A \cap \{\tau_1 \leq \tau_2\} \in \mathscr{F}_{\tau_2}$. Hint:

$$\{\tau_1 \leq t\} \cap \{\tau_2 \leq t\} \cap \{\tau_1 \leq \tau_2\} = \{\tau_1 \leq t\} \cap \{\tau_2 \leq t\} \cap \{\tau_1 \wedge t \leq \tau_2 \wedge t\}.$$

   (c) $\mathscr{F}_{\tau_1\wedge\tau_2} = \mathscr{F}_{\tau_1} \cap \mathscr{F}_{\tau_2}$.
   (d) $\{\tau_1 \leq \tau_2\} \in \mathscr{F}_{\tau_1} \cap \mathscr{F}_{\tau_2}$ (and thus $\{\tau_1 i \tau_2\} \in \mathscr{F}_{\tau_1} \cap \mathscr{F}_{\tau_2}$ where i is any inequality or equality.
(9) Let τ be a stopping time and that the $I = [0, \infty)$. For each n, define $\tau_n \stackrel{\text{def}}{=} \lceil \tau n \rceil / n$. Show that each τ_n is also a stopping time.
(10) If $I \subset \mathbb{Z}$, X is an adapted process, and τ is a stopping time, then X_τ is $\mathscr{F}_\tau$-measurable. Hint: partition the space according to the values of τ.
(11) Let $\{\xi_n;\, n \in \mathbb{N}\}$ be a collection of independent and identically distributed integrable random variables. Define

$$\mathscr{F}_n \stackrel{\text{def}}{=} \sigma\{\xi_k;\, k \leq n\}$$

and show that $X_n \stackrel{\text{def}}{=} \sum_{k=1}^n \xi_k$ is a martingale with respect to $\{\mathscr{F}_n;\, n \in \mathbb{N}\}$. Show that $\{X_n;\, n \in \mathbb{N}\}$ is a martingale, submartingale, or supermartingale, respectively, if the mean of the ξ's is zero, positive, or negative.
(12) Let X be a martingale with respect to a filtration, and suppose that $\{Y_n;\, n \in \mathbb{N}\}$ is a collection of integrable random variables such that Y_n is $\mathscr{F}_{n-1}$-measurable for all n. Show that

$$Z_n \stackrel{\text{def}}{=} \begin{cases} \sum_{k=1}^n Y_k(X_k - X_{k-1}) & \text{if } n \geq 1 \\ 0 & \text{if } n = 0 \end{cases}$$

is also a martingale (with respect to the same filtration). This new process is called a **martingale transform** of X and is a simple case of a stochastic integral.
(13) Let X be a martingale and $\varphi : \mathbb{R} \to \mathbb{R}$ be convex and such that $\mathbb{E}[\varphi(X_n)] < \infty$ for all n. Show that $Z_n \stackrel{\text{def}}{=} \varphi(X_n)$ is a submartingale.

(14) Let X be a martingale, submartingale, or supermartingale (respectively), and let τ be a stopping time. Show that $Z_n \stackrel{\text{def}}{=} X_{\tau \wedge n}$ is also a martingale, submartingale, or supermartingale, respectively (with respect to the original filtration $\{\mathscr{F}_n\}$). Also show that (assuming that $X_0 = 0$) Z can be written as a martingale transform of X.

CHAPTER 5

Weak convergence

1. The topology of weak convergence

It turns out that the notion of *weak convergence*, defined in Chapter 1, is an important one. In this chapter, we will assume that X is *Polish*, i.e., that it has a metric d under which it is complete and separable. We let $\mathscr{T}$ denote the collection of open subsets of X defined by d. Our goal is to study $\mathscr{P}(X)$, the collection of probability measures on $(X, \mathscr{B}(X))$. As a first simple observation, note that $\mathscr{P}(X)$ is convex. Note that we can also topologize it as a dual of $C_b(X)$. In other words, this topology, the *weak topology* on $\mathscr{P}(X)$, is the smallest topology on $\mathscr{P}(X)$ with respect to which all of the mappings $I_\varphi : \mathscr{P}(X) \to \mathbb{R}$ defined by

$$I_\varphi(\mu) \stackrel{\text{def}}{=} \int_X \varphi(x)\mu(dx) \qquad \mu \in \mathscr{P}(X) \tag{19}$$

are continuous, as φ varies over $C_b(X)$. It turns out that $\mathscr{P}(X)$, endowed with this topology, is itself Polish, and that we can characterize its compact sets.

2. The Prohorov metric

To begin, let's write down the *Prohorov* metric. Let $\mathscr{C}$ be the collection of closed subsets of X, and for any subset A of X and any $\varepsilon > 0$, define

$$A^\varepsilon \stackrel{\text{def}}{=} \{x \in X : \operatorname{dist}(x, A) < \varepsilon\},$$

where $\operatorname{dist}(x, A) \stackrel{\text{def}}{=} \inf_{y \in A} d(x, y)$ for all $x \in X$. We define

$$\rho(\mu, \nu) \stackrel{\text{def}}{=} \inf\{\varepsilon > 0 : \mu(F) \le \nu(F^\varepsilon) + \varepsilon \text{ for all } F \in \mathscr{C}\}$$

for all μ and ν in $\mathscr{P}(X)$.

We have the following theorem which connects the Prohorov metric and different definitions of weak convergence. For future reference, we now define the open balls

$$B(x, \varepsilon) \stackrel{\text{def}}{=} \{x' \in X : d(x', x) < \varepsilon\}$$

for each $x \in X$ and $\varepsilon > 0$.

PROPOSITION 2.1. *Fix $\{\mu_1, \mu_2 \dots\}$ and μ in $\mathscr{P}(X)$. The following are equivalent.*

(a) $\lim_{n\to\infty} \rho(\mu_n, \mu) = 0$.
(b) $\overline{\lim}_{n\to\infty} \mu_n(F) \le \mu(F)$ *for all* $F \subset X$ *closed.*
(c) $\underline{\lim}_{n\to\infty} \mu_n(G) \ge \mu(G)$ *for all* $G \subset X$ *open.*
(d) $\lim_{n\to\infty} \int_X \varphi(x)\mu_n(dx) = \int_X \varphi(x)\mu(dx)$ *for all* $\varphi \in C_b(X)$.

PROOF. Most of the work is done in the problems; namely that (a) implies (b) and that (b) and (c) are equivalent and are in turn equivalent to (d). The only remaining part is that (b) implies (a).

Since X is separable and metric, we can find $\{x_1, x_2 \dots\}$ a countable dense subset of X. Fix $\varepsilon > 0$ and define $E_0 \stackrel{\text{def}}{=} B(x_1, \varepsilon/4)$ and $E_n \stackrel{\text{def}}{=} B(x_n, \varepsilon/4) \setminus \cup_{j=1}^n E_j$ for all $n \ge 2$. Then the E_n's are disjoint and the diameter of each of them is $\varepsilon/2$ or less. Let L be an integer large enough that

$$\mu\left(\overline{\left(\cup_{j=1}^L E_j\right)^c}\right) < \varepsilon/2. \tag{20}$$

Thus by assumption, for n sufficiently large

$$\mu_n\left(\overline{\left(\left(\cup_{j=1}^{L} E_j\right)^c\right)}\right) < \varepsilon/2. \tag{21}$$

Also, by assumption, for n sufficiently large

$$\mu_n\left(\overline{\cup_{i\in I} E_i}\right) \leq \mu\left(\overline{\cup_{i\in I} E_i}\right) + \varepsilon/2$$

for all subsets I of the finite set $\{1, 2 \dots L\}$. Fix now n large enough that (20) and (21) hold. Fix any closed subset F of X. It is easy to see that

$$F \subset \overline{\cup_{\substack{1\leq i\leq L \\ E_i\cap F\neq\emptyset}} E_i} \cup \overline{\left(\cup_{i=1}^{L} E_i\right)^c}$$
$$\overline{\cup_{\substack{1\leq i\leq L \\ E_i\cap F\neq\emptyset}} E_i} \subset F^{\varepsilon}.$$

Then

$$\mu_n(F) \leq \mu\left(\overline{\cup_{\substack{1\leq i\leq L \\ E_i\cap F\neq\emptyset}} E_i}\right) + \varepsilon \leq \mu(F^{\varepsilon}) + \varepsilon.$$

Vary F to get that $\rho(\mu_n, \mu) \leq \varepsilon$ for all n large. Let n tend to infinity, and then ε tend to zero to see that $\lim_{n\to\infty} \rho(\mu_n, \mu) = 0$. □

3. Tightness and compactness in the Prohorov topology

A natural next question is: what do the compact subsets of $\mathscr{P}(X)$ look like? It turns out that the following has a lot to do with compactness.

Definition 3.1 (Tightness). A subset $\mathscr{M}$ of $\mathscr{P}(X)$ is tight if for each $\varepsilon > 0$ there is a compact subset K of X (denoted by $K \subset\subset X$) such that $\mu(K^c) < \varepsilon$ for all $\mu \in \mathscr{M}$.

Below we will conclude that $\mathscr{M} \subset \mathscr{P}(X)$ *is tight if and only if* $\overline{\mathscr{M}}$ *is compact.* First we will show that if $\mathscr{M} \subset \mathscr{P}(X)$ is compact, it must be tight. Then we will show that if $\mathscr{M} \subset \mathscr{P}(X)$ is tight, then $\overline{\mathscr{M}}$ is compact.

Let's start by showing that the simplest possible compact subsets of $\mathscr{P}(X)$ are tight.

Lemma 3.2. *Any $\mu \in \mathscr{P}(X)$ is tight.*

Proof. Fix $\varepsilon > 0$. Since X is separable, it contains a countable dense subset $\{x_1, x_2 \dots\}$. For each n, let L_n be such that

$$\mu\left(\bigcup_{i=1}^{L_n} B(x_i, 1/n)\right) \geq 1 - \varepsilon/2^{n+1}.$$

Set

$$K \stackrel{\text{def}}{=} \overline{\bigcap_{n=1}^{\infty} \bigcup_{i=1}^{L_n} B(x_i, 1/n)}.$$

Then K is closed and totally bounded and is thus compact. Note that

$$\begin{aligned}\mu(K) &\geq \mu\left(\bigcap_{n=1}^{\infty} \bigcup_{i=1}^{L_n} B(x_i, 1/n)\right) = 1 - \mu\left(\bigcup_{n=1}^{\infty} \bigcap_{i=1}^{L_n} B(x_i, 1/n)^c\right) \\ &\geq 1 - \sum_{n=1}^{\infty} \mu\left(\bigcap_{i=1}^{L_n} B(x_i, 1/n)^c\right) \geq 1 - \sum_{n=1}^{\infty}\left(1 - \mu\left(\bigcup_{i=1}^{L_n} B(x_i, 1/n)\right)\right) \geq 1 - \sum_{n=1}^{\infty} \varepsilon/2^{n+1} = 1 - \varepsilon.\end{aligned}$$

Thus μ is tight. □

From here, we can show that compactness implies tightness.

Proposition 3.3. *If $\mathscr{M} \subset \mathscr{P}(X)$ is compact, then it is tight.*

PROOF. Fix $\varepsilon > 0$. Since $\mathscr{M}$ is compact, it is totally bounded, so we can for each n, find a finite subset $\mathscr{N}_n$ or $\mathscr{M}$ such that

$$\mathscr{M} \subset \bigcup_{\tilde{\mu} \in \mathscr{N}_n} \left\{ \mu \in \mathscr{P}(X) : \rho(\mu, \tilde{\mu}) < \varepsilon/2^{n+1} \right\}.$$

By Lemma 2, each of the $\tilde{\mu}$'s in each of the $\mathscr{N}_n$'s are tight, so for each n, we can find a $K_n \subset\subset X$ such that $\tilde{\mu}(K^C) < \varepsilon/2^{n+1}$ for all $\tilde{\mu} \in \mathscr{N}_n$. Thus, for all n and all $\mu \in \mathscr{P}(X)$ and $\tilde{\mu} \in \mathscr{N}_n$ such that $\rho(\mu, \tilde{\mu}) < \varepsilon/2^{n+1}$,

$$1 - \varepsilon/2^{n+1} \leq \tilde{\mu}(K_n) \leq \mu\left(K_n^{\varepsilon/2^{n+1}}\right) + \varepsilon/2^{n+1}$$

so $\mu\left(K_n^{\varepsilon/2^{n+1}}\right) \geq 1 - \varepsilon/2^n$ for all $\mu \in \mathscr{M}$. Set now

$$K \stackrel{\text{def}}{=} \overline{\bigcap_{n=1}^{\infty} K_n^{\varepsilon/2^{n+1}}}.$$

Then K is compact and $\mu(K) \geq 1 - \varepsilon$, by a calculation similar to that of Lemma 2. □

The other direction is a bit more complicated. We will show that if $\mathscr{M} \subset \mathscr{P}(X)$ is tight, then it is totally bounded. Secondly, we will show that if $\mathscr{M}$ is tight, its closure is complete.

PROPOSITION 3.4. *If $\mathscr{M} \subset \mathscr{P}(X)$ is tight, then it is totally bounded.*

PROOF. Since X is separable, we can find a countable dense subset $\{x_1, x_2 \dots\}$ of X. For each N and m, define the finite subset

$$\mathscr{N}_{N,m} \stackrel{\text{def}}{=} \left\{ \sum_{k=1}^{N+1} a_k \delta_{x_k} \in \mathscr{P}(X) : a_k \in \mathbb{Z}_+/m \text{ for all } 1 \leq k \leq N+1 \right\}$$

of $\mathscr{P}(X)$. For any ε, we will show that for sufficiently large N and m, any element of $\mathscr{M}$ is within ε (in the ρ metric) of some element of $\mathscr{N}_{N,m}$. Indeed, fix $\varepsilon > 0$ and then $K \subset\subset X$ such that $\mu(K^c) < \varepsilon/2$ for all $\mu \in \mathscr{M}$. Since $K \subset \cup_{k=1}^{\infty} B(x_k, \varepsilon)$ and $K \subset\subset X$, we can let N be any number such that $K \subset \cup_{k=1}^{N} B(x_k, \varepsilon)$, and we then fix $m \geq 2N/\varepsilon$. Consider any $\mu \in \mathscr{M}$. Set

$$E_1 \stackrel{\text{def}}{=} B(x_1, \varepsilon)$$

$$E_n \stackrel{\text{def}}{=} B(x_n, 2\varepsilon) \setminus \cup_{k=1}^{n-1} B(x_k, \varepsilon) \quad n = 1, 2 \dots$$

$$a_i \stackrel{\text{def}}{=} \lfloor \mu(E_i) m \rfloor / m \quad 1 \leq i \leq N$$

$$a_{N+1} \stackrel{\text{def}}{=} 1 - \sum_{i=1}^{N} a_i.$$

Define $\tilde{\mu} \in \mathscr{N}_{N,m}$ by

$$\tilde{\mu} \stackrel{\text{def}}{=} \sum_{k=1}^{N+1} a_k \delta_{x_k}.$$

Note that

$$1 - a_{N+1} = \sum_{i=1}^{N} a_i \geq \sum_{i=1}^{N} \frac{\mu(E_i)m - 1}{m} = \mu\left(\cup_{i=1}^{N} E_i\right) - N/m \geq \mu(K) - N/m \geq 1 - \varepsilon/2 - N/m.$$

From here we get that $a_{N+1} \leq \varepsilon/2 + N/m < \varepsilon$. Fix now any $F \subset X$ closed. Then

$$\tilde{\mu}(F) \leq \sum_{\substack{1 \leq i \leq N \\ x_i \in F}} a_i + a_{N+1} \leq \sum_{\substack{1 \leq i \leq N \\ x_i \in F}} \mu(E_i) + \varepsilon \leq \mu\left(\bigcup_{\substack{1 \leq i \leq N \\ x_i \in F}} E_i \right) + \varepsilon$$

Note that

$$\bigcup_{\substack{1\le i\le N\\ x_i\in F}} E_i \subset F^\varepsilon;$$

Thus $\tilde{\mu}(F) \le \mu(F^\varepsilon) + \varepsilon$ for all closed $F \subset X$ and thus $\rho(\mu, \tilde{\mu}) \le \varepsilon$, which completes the proof. □

Next, we start to prove that if $\mathscr{M}$ is tight, it is complete. We start with

PROPOSITION 3.5. *If $\mathscr{M} \subset \mathscr{P}(X)$ is tight and $\{\mu_n\} \subset \mathscr{M}$ is Cauchy, then*

$$\Lambda(\varphi) \stackrel{def}{=} \lim_{n\to\infty} \int_X \varphi(x)\mu_n(dx)$$

exists for all $\varphi \in C_b(X)$. Furthermore, if $\{\varphi_n\} \subset C_b(X)$ is such that $\varphi_n \searrow 0$, then $\lim_{n\to\infty} \Lambda(\varphi_n) = 0$.

PROOF. For $\mu \in \mathscr{P}(X)$ and $\varphi \in C_b(X)$, define

$$\|\varphi\| \stackrel{\text{def}}{=} \sup_{x\in X} |\varphi(x)|$$

$$\Lambda_\mu(\varphi) \stackrel{\text{def}}{=} \int_X \varphi(x)\mu(dx).$$

Noting that $\varphi + \|\varphi\|$ is nonnegative, we can arguments similar to those of Problem 1.7 to see that

$$\Lambda_\mu(\varphi) = \int_{-\|\varphi\|}^{\|\varphi\|} \mu\{x \in X : \varphi(x) \ge t\}dt - \|\varphi\| \tag{22}$$

for all $\varphi \in C_b(X)$ and all $\mu \in \mathscr{P}(X)$. Fix now $K \subset\subset X$ and set

$$\omega_K(\delta) \stackrel{\text{def}}{=} \sup_{\substack{x,y\in K\\ d(x,y)\le\delta}} |\varphi(x) - \varphi(y)|;$$

i.e., ω_K is the modulus of continuity of φ restricted to K. For any $t \in \mathbb{R}$ and $\delta > 0$,

$$\mu\{x \in X : \varphi(x) \ge t\} \le \mu(K^c) + \mu\{x \in K : \varphi(x) \ge t\}$$
$$\{x \in K : \varphi(x) \ge t\}^\delta \subset K^c \cup \{x \in X : \varphi(x) \ge t - \omega_K(\delta)\}.$$

Thus, for any $\nu \in \mathscr{P}(X)$ and any $t \in \mathbb{R}$,

$$\mu\{x \in X : \varphi(x) \ge t\} \le \mu(K^c) + \nu(K^c) + \nu\{x \in X : \varphi(x) \ge t - \omega_K(\rho(\mu,\nu))\} + \rho(\mu,\nu)$$

so

$$\begin{aligned}
\Lambda_\mu(\varphi) &\le \int_{-\|\varphi\|}^{\|\varphi\|} \nu\{x \in X : \varphi(x) \ge t - \omega_K(\rho(\mu,\nu))\}dt - \|\varphi\| + 2\|\varphi\|\left(\mu(K^c) + \nu(K^c) + \rho(\mu,\nu)\right)\\
&\le \int_{-\|\varphi\|}^{\|\varphi\|} \nu\{x \in X : \varphi(x) \ge t\}dt - \|\varphi\| + 2\|\varphi\|\left(\mu(K^c) + \nu(K^c) + \rho(\mu,\nu)\right) + \omega_K(\rho(\mu,\nu))\\
&= \Lambda_\nu(\varphi) + 2\|\varphi\|\left(\mu(K^c) + \nu(K^c) + \rho(\mu,\nu)\right) + \omega_K(\rho(\mu,\nu)).
\end{aligned}$$

Thus

$$|\Lambda_\nu(\varphi) - \Lambda_\mu(\varphi)| \le 2\|\varphi\|\left(\mu(K^c) + \nu(K^c) + \rho(\mu,\nu)\right) + \omega_K(\rho(\mu,\nu))$$

for any μ and ν in $\mathscr{P}(X)$ and any $K \subset\subset X$. In particular, for $\{\mu_n\} \subset \mathscr{M}$ which is tight and Cauchy, we can fix $\varepsilon > 0$ and then find $K \subset\subset X$ such that $\mu_n(K^c) \le \varepsilon$ for all n. Then

$$\overline{\lim_{m,n\to\infty}} |\Lambda_{\mu_n}(\varphi) - \Lambda_{\mu_m}(\varphi)| \le 4\|\varphi\|\varepsilon.$$

Now let ε tend to zero to see that $\{\Lambda_{\mu_n}(\varphi)\}$ is Cauchy and thus convergent.

Now assume that $\{\varphi_n\} \subset C_b(X)$ is such that $\varphi_n \searrow 0$. Clearly $\underline{\lim}_{n\to\infty} \Lambda(\varphi_n) \ge 0$. Fix next any $\varepsilon > 0$ and then $K \subset\subset X$ such that $\mu(K^c) \le \varepsilon$ for all $\mu \in \mathscr{M}$. Then

$$\Lambda(\varphi_n) \le \lim_{n\to\infty} \sup_{x\in K} \varphi_n(x) + \|\varphi_n\| \sup_{\mu\in\mathscr{M}} \mu(K^c) \le \|\varphi_1\|\varepsilon.$$

Here we have used Dini's theorem to see that $\overline{\lim}_{n\to\infty} \sup_{x\in K} \varphi_n(x) = 0$. Now let ε tend to zero to see that $\overline{\lim}_{n\to\infty} \Lambda(\varphi_n) \le 0$. □

From here, we could directly appeal to Daniell's theory of linear functionals. In the interest of completeness, however, we will write out a proof. Let's define

$$\mu^*(O) \stackrel{\text{def}}{=} \sup\{\Lambda(\varphi) : \varphi \in C_c(X;[0,1]) \text{ and } \operatorname{supp}\varphi \subset O\} \qquad O \text{ open}$$

$$\bar{\mu}(A) \stackrel{\text{def}}{=} \inf\{\mu^*(O) : O \supset A \text{ is open}\}. \qquad A \in \mathscr{B}(X)$$

Here $C_c(X:[0,1])$ is the collection of continuous mappings from X into $[0,1]$ of compact support. The idea here is to approximate indicators of open sets from below by continuous functions to get μ^* and to use the notion of regularity to define the measure of any measurable set.

LEMMA 3.6. *The set function μ^* is an inner content which agrees with $\bar{\mu}$ on open sets and such that $\mu^*(\emptyset) = 0$ and $\mu^*(X) = 1$. By an inner content, we mean*

(a) *If $O_1 \subset O_2$ are open subsets of X, then $\mu^*(O_1) \le \mu^*(O_2)$.*
(b) *If $O \subset \cup_{i=1}^{\infty} O_i$, where O and the O_i's are open subsets of X, then $\mu^*(O) \le \sum_{i=1}^{\infty} \mu^*(O_i)$.*
(c) *If O_1 and O_2 are two disjoint open subsets of X, then $\mu(O_1 \cup O_2) \ge \mu(O_1) + \mu(O_2)$.*
(d) *For any open subset O of X,*

$$\mu^*(O) \le \sup\left\{\mu^*(G) : G \text{ is open, } \bar{G} \text{ is compact, and } \bar{G} \subset O\right\}.$$

PROOF. The facts that μ^* agrees with μ on open sets and that $\mu^*(\emptyset) = 0$ and $\mu^*(X) = 1$ are obvious. For convenience, define

$$\varrho(x) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } x \le 1/3 \\ 3x - 1 & \text{if } 1/3 < x \le 2/3 \\ 1 & \text{if } x > 2/3. \end{cases}$$

Proof of (a) Obvious.

Proof of (b) Fix $\varphi \in C_c(X:[0,1])$ such that $\operatorname{supp}\varphi \subset O$. Then since $\operatorname{supp}\varphi$ is compact, $\operatorname{supp}\varphi \subset \cup_{i=1}^{N} O_i$ for some N, and then

$$\eta \stackrel{\text{def}}{=} \operatorname{dist}\left(\operatorname{supp}\varphi, \left(\cup_{i=1}^{N} O_i\right)^c\right) > 0.$$

Define

$$\varphi_i(x) \stackrel{\text{def}}{=} \frac{\varrho(\eta^{-1}\operatorname{dist}(x, O_i^c))}{\sum_{j=1}^{N} \varrho(\eta^{-1}\operatorname{dist}(x, O_j^c))}\varphi(x). \qquad x \in X,\ 1 \le i \le N$$

Then for each i, φ_i is a well-defined element of $C_c(X;[0,1])$ and $\operatorname{supp}\varphi_i \subset O_i$, and $\sum_{i=1}^{N} \varphi_i = \varphi$. Thus

$$\Lambda(\varphi) = \sum_{i=1}^{N} \Lambda(\varphi_i) \le \sum_{i=1}^{N} \mu^*(O_i) \le \sum_{i=1}^{\infty} \mu^*(O_i).$$

Now vary φ to get (a).

Proof of (c) Fix φ_1 and φ_2 in $C_c(X:[0,1])$ such that $\operatorname{supp}\varphi_1 \subset O_1$ and $\operatorname{supp}\varphi_2 \subset O_2$. Then $\varphi \stackrel{\text{def}}{=} \varphi_1 + \varphi_2$ is in $C_c(X:[0,1])$ and $\operatorname{supp}\varphi \subset O_1 \subset O_2$. Thus

$$\Lambda(\varphi_1) + \Lambda(\varphi_2) = \Lambda(\varphi) \le \mu^*(O_1 \cup O_2).$$

Vary φ_1 and φ_2 to get (b).

Proof of (d) Here is where we use the continuity of Λ. Fix $\varphi \in C_c(X:[0,1])$ such that $\operatorname{supp}\varphi \subset O$. Set

$$G \stackrel{\text{def}}{=} \{x \in X : \varphi(x) > 0\}$$

$$\varphi_n(x) \stackrel{\text{def}}{=} \varrho(\operatorname{dist}(x, G^c)/n)\varphi(x). \qquad x \in X,\ n = 1, 2\ldots$$

Then $\bar{G} = \operatorname{supp}\varphi \subset\subset X$ and $G \subset O$ is open. Furthermore, $\varphi_n \in C_c(X:[0,1])$ for all n, $\operatorname{supp}\varphi_n \subset G$, and $\varphi_n \nearrow \varphi$. Thus $\varphi - \varphi_n \searrow 0$, so

$$\Lambda(\varphi) = \lim_{n\to\infty} \Lambda(\varphi_n) \le \mu^*(G) \le \sup\left\{\mu^*(G) : G \text{ is open, } \bar{G} \text{ is compact, and } \bar{G} \subset O\right\}$$

Vary φ to get (c). $\square$

Next, let's consider $\bar{\mu}$.

LEMMA 3.7. *For $A \subset B$ in $\mathscr{B}(X)$, $\bar{\mu}(A) \leq \bar{\mu}(B)$. For any $\{A_n\} \subset \mathscr{B}(X)$ which are disjoint,*

$$\bar{\mu}\left(\cup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \bar{\mu}(A_n).$$

Secondly,

(a) *If $O_1 \subset X$ and $O_2 \subset$ are open subsets of X, then $\bar{\mu}(O_1) \geq \bar{\mu}(O_1 \cap O_2) + \bar{\mu}(O_1 \setminus O_2)$.*
(b) *If O is an open subset of X and $A \in \mathscr{B}(X)$, $\bar{\mu}(A) \geq \bar{\mu}(A \cap O) + \bar{\mu}(A \setminus O)$.*
(c) *For A and B in $\mathscr{B}(X)$, $\bar{\mu}(A) \geq \bar{\mu}(A \cap B) + \bar{\mu}(A \setminus B)$.*

Then for any $\{A_i\} \subset \mathscr{B}(X)$ which are disjoint,

$$\bar{\mu}\left(\cup_{i=1}^{\infty} A_i\right) \geq \sum_{i=1}^{\infty} \bar{\mu}(A_i). \tag{23}$$

Thus $\bar{\mu} \in \mathscr{P}(X)$.

PROOF. The monotonicity of $\bar{\mu}$ is obviously inherited from the monotonicity of μ^*.

To prove subadditivity, fix $\varepsilon > 0$. For each n, let $O_n \supset A_n$ be open and such that $\mu^*(O_n) \leq \bar{\mu}(O_n) + \varepsilon/2^n$. Then

$$\bar{\mu}\left(\cup_{n=1} A_n\right) \leq \mu^*\left(\cup_{n=1} O_n\right) \leq \sum_{n=1}^{\infty} \mu^*(O_n) \leq \sum_{n=1}^{\infty} \left(\bar{\mu}(A_n) + \varepsilon/2^n\right) \leq \sum_{n=1}^{\infty} \bar{\mu}(A_n) + \varepsilon$$

Let ε tend to zero.

Proof of (a) Fix G open such that $\bar{G} \subset\subset X$ and $\bar{G} \subset O_1 \cap O_2$. Note that G and $O_1 \setminus \bar{G}$ are disjoint, that $G \cup (O_1 \setminus \bar{G}) \subset O_1$, and that $O_1 \setminus \bar{G} \supset O_1 \setminus O_2$. Thus

$$\bar{\mu}(O_1) = \mu^*(O_1) \geq \mu^*(G) + \mu^*(O_1 \setminus \bar{G}) \geq \mu^*(G) + \bar{\mu}(O_1 \setminus O_2).$$

Vary now G.

Proof of (b) Fix $G \supset A$ open. Then

$$\mu^*(G) = \bar{\mu}(G) \geq \bar{\mu}(G \cap O) + \bar{\mu}(G \setminus O) \geq \bar{\mu}(A \cap O) + \bar{\mu}(A \setminus O).$$

Now vary G.

Proof of (c) Define

$$\mathscr{G} \stackrel{\text{def}}{=} \left\{A \in \mathscr{B}(X) : \bar{\mu}(S) \geq \bar{\mu}(S \cap A) + \bar{\mu}(S \setminus A) \text{ for all } S \in \mathscr{B}(X)\right\}.$$

We claim that $\mathscr{G}$ is a sub sigma-algebra of $\mathscr{B}(X)$, much like the collection of Lebesgue-measurable subsets of $\mathbb{R}$ is a sigma-algebra. It is clear that $\mathscr{G}$ contains X and is closed under complementation. Next, fix any A and B in $\mathscr{G}$. Then for any $S \in \mathscr{B}(X)$,

$$\begin{aligned} \bar{\mu}(S) &\geq \bar{\mu}(S \cap A) + \bar{\mu}(S \setminus A) \geq \bar{\mu}(S \cap A \cap B) + \bar{\mu}(S \cap A \setminus B) + \bar{\mu}(S \setminus A) \\ &\geq \bar{\mu}(S \cap (A \cap B)) + \bar{\mu}(S \setminus (A \cap B)) \end{aligned} \tag{24}$$

The first inequality comes from the fact that $A \in \mathscr{G}$ and the second from the fact that $B \in \mathscr{G}$. The last line comes from the fact that

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

this implies that $S \setminus (A \cap B) = (S \setminus A) \cup (S \cap A \setminus B)$, and the subadditivity of $\bar{\mu}$ then gives us the last line of (24). Thus $\mathscr{G}$ is closed under finite intersections and complements, and thus $\mathscr{G}$ is a field. Finally, fix $\{A_n\} \subset \mathscr{G}$, and set $B_1 \stackrel{\text{def}}{=} A_1$ and $B_n \stackrel{\text{def}}{=} A_n \setminus \cup_{j=1}^{n-1} A_j$ for all $n \geq 1$. Then the B_n's are in $\mathscr{G}$, are disjoint, and $\cup_{n=1}^{\infty} B_n = \cup_{n=1}^{\infty} A_n$. For any n and any $S \in \mathscr{B}(X)$,

$$\bar{\mu}(S) \geq \bar{\mu}\left(S \cap \left(\cup_{j=1}^{n} B_j\right)\right) + \bar{\mu}(S \setminus \cup_{j=1}^{n} B_j) \geq \sum_{j=1}^{n} \bar{\mu}(S \cap B_j) + \bar{\mu}(S \setminus \cup_{j=1}^{\infty} A_j).$$

The first inequality comes from the fact that $\cup_{j=1}^n B_j \in \mathscr{G}$. The second inequality comes from the fact that the B_j's are in $\mathscr{G}$ and that

$$\left(\cup_{j=1}^n B_j\right) c = \left(\cup_{j=1}^n A_j\right)^c \supset \left(\cup_{j=1}^n A_j\right)^c.$$

Now let n tend to infinity and use the subadditivity of $\bar{\mu}$ to see that

$$\sum_{j=1}^{\infty} \bar{\mu}(S \cap B_j) \geq \bar{\mu}\left(\cup_{j=1}^{\infty} S \cap B_j\right) = \bar{\mu}\left(S \cap (\cup_{j=1}^{\infty} A_j)\right).$$

Thus $\cup_{j=1}^{\infty} A_j \in \mathscr{G}$, so $\mathscr{G}$ is a sigma-algebra. Since the open subsets are in $\mathscr{G}$ by part (b), we have that $\mathscr{B}(X) \subset \mathscr{G}$. This implies (c).

Now we can prove (23). From a repeated application of (c), we get that

$$\bar{\mu}\left(\cup_{j=1}^{\infty} A_j\right) \geq \bar{\mu}\left(\cup_{j=1}^{n} A_j\right) = \sum_{j=1}^{n} \bar{\mu}(A_j).$$

Let n tend to infinity.

Finally, we claim that $\bar{\mu} \in \mathscr{P}(X)$. From Lemma 6, we see that $\bar{\mu}(\emptyset) = 0$ and $\bar{\mu}(X) = 1$. From the subadditivity and (23), we see that $\bar{\mu}$ is indeed additive, implying that $\bar{\mu} \in \mathscr{P}(X)$. □

LEMMA 3.8. *We have that* $\lim_{n\to\infty} \mu_n = \bar{\mu}$.

PROOF. it suffices to show that

$$\underline{\lim}_{n\to\infty} \mu_n(O) \geq \mu^*(O) \tag{25}$$

for any open subset O of X. Fix $\varphi \in C_c(X : [0,1])$ with $\operatorname{supp} \varphi \subset O$. Then

$$\underline{\lim}_{n\to\infty} \mu_n(O) \geq \mu^*(O) \geq \underline{\lim}_{n\to\infty} \int_X \varphi(x)\mu_n(dx) = \Lambda(\varphi).$$

Vary φ to get (25). □

Finally, we can get

PROPOSITION 3.9. *If $\mathscr{M} \subset \mathscr{P}(X)$ is tight, its closure is complete.*

PROOF. If $\mathscr{M}$ is tight, so is its closure $\overline{\mathscr{M}}$ (one of the problems). From the above, we know that if $\{\mu_n\} \subset \overline{\mathscr{M}}$ is Cauchy, it has a limit point $\bar{\mu} \in \overline{\mathscr{M}}$. □

This give us

THEOREM 3.10 (Prohorov). *$\mathscr{M} \subset \mathscr{P}(X)$ is tight if and only if $\overline{\mathscr{M}}$ is compact.*

PROOF. Compactness of $\overline{\mathscr{M}}$ implies tightness by Propositidon 3. Tightness of $\overline{\mathscr{M}}$ implies compactness by Propositions 4 and 9. □

Finally, we can complete the proof that $\mathscr{P}(X)$ is itself Polish

THEOREM 3.11. *The space $\mathscr{P}(X)$, endowed with the weak topology, is Polish with metric ρ.*

PROOF. Theorem 1 tells us that $\mathscr{P}(X)$ with the topology of weak convergence is metric. Separability is in one of the questions. To prove completeness, fix a Cauchy sequence $\{\mu_n\}$ in $\mathscr{P}(X)$. We claim that $\{\mu_n;\, n = 1, 2 \dots\}$ is tight. Fix $\varepsilon > 0$, and for every l, let a number m_l and a $K_l \subset\subset X$ such that

$$\sup_{n \geq m_l} \rho(\mu_n, \mu_{n_l}) < \varepsilon/2^{l+1}$$

$$\min_{1 \leq n \leq m_l} \mu_n(K_l) \geq 1 - \varepsilon/2^{l+1}.$$

Then for $n > m_l$,

$$1 - \varepsilon/2^{l+1} \leq \mu_{m_l}(K_l) \leq \mu_n\left(K_l^{\varepsilon/2^{l+1}}\right) + \varepsilon/2^{l+1},$$

which implies that $\nu_n\left(K_l^{\varepsilon/2^{l+1}}\right) \geq 1 - \varepsilon/2^l$. Thus

$$\inf_n \mu_n\left(K_l^{\varepsilon/2^{l+1}}\right) \geq 1 - \varepsilon/2^l.$$

Set

$$K \stackrel{\text{def}}{=} \overline{\cap_{l=1}^{\infty} K_l^{\varepsilon/2^{l+1}}}.$$

Then $\mu_n(K^c) \leq \varepsilon$ for all n, so $\{\mu_n;\, n = 1, 2 \dots\}$ is indeed tight, so its closure is compact, so $\lim_n \mu_n$ must exist. □

Exercises

(1) Show that the Prohorov metric is a metric.

(2) Show that $\mathscr{P}(X)$ is separable. Hint: consider probability measures of the form $\sum_{x\in I} a_x \delta_x$, where I is a finite subset of a dense subset of X and where the $a(x)$'s are nonnegative rationals.

(3) Show that if $\mathscr{M}$ is tight, then so is $\overline{\mathscr{M}}$.

(4) Let $\{x_n\}$ be a collection of points in X and fix some $x \in X$. Consider the probability measures δ_{x_n}. Show that δ_{x_n} converges to δ_x in the Prohorov metric if and only if x_n converges to x.

(5) Consider $\{\mu_n\} \subset \mathscr{P}(X)$. Show that

(a) if $\rho(\mu_n, \mu) \to 0$, then $\overline{\lim}_n \mu_n(F) \leq \mu(F)$ for all closed $F \subset X$.

(b) $\overline{\lim}_n \mu_n(F) \leq \mu(F)$ for all closed $F \subset X$ if and only if $\underline{\lim}_n \mu_n(G) \geq \mu(G)$ for all open $G \subset X$.

(c) if $\overline{\lim}_n \mu_n(F) \leq \mu(F)$ for all closed $F \subset X$ (and consequently $\overline{\lim}_n \mu_n(G) \geq \mu(G)$ for all open $G \subset X$), then

$$\lim_n \int_X \varphi(x)\mu_n(dx) = \int_X \varphi(x)\mu(dx)$$

for $\varphi \in C_b(X)$. Hint: use (22).

(d) if $\lim_n \int_X \varphi(x)\mu_n(dx) = \int_X \varphi(x)\mu(dx)$ for all bounded and continuous $\varphi : X \to \mathbb{R}$, then $\overline{\lim}_n \mu_n(F) \leq \mu(F)$ for all closed $F \subset X$. Hint: consider functions of the form $\varphi_\varepsilon(x) \stackrel{\text{def}}{=} 1 - (\varepsilon^{-1} \operatorname{dist}(x, F) \wedge 1)$.

(6) Fix $k \geq 0$. Show that $\{\mu_n\} \subset \mathscr{P}(\mathbb{R}^d)$ converges to $\mu \in \mathscr{P}(\mathbb{R}^d)$ if and only if $\lim_{n\to\infty} I_\varphi(\mu_n) = I_\varphi(\mu)$ (using the notation of (19)) for all $\varphi \in C_b$ which are k-differentiable and for which all k derivatives are bounded.

(7) Fix $T > 0$ and two probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ in $\mathscr{P}(C([0,T];\mathbb{R}^d))$, where $C([0,T];\mathbb{R}^d))$ is endowed with the standard supremum-norm topology (and is thus Polish). Show that if

$$\mathbb{P}_1\left(\bigcap_{t\in I}\{\omega \in C([0,T];\mathbb{R}^d) :\, \omega(t) \in A_t\}\right) = \mathbb{P}_2\left(\bigcap_{t\in I}\{\omega \in C([0,T];\mathbb{R}^d) :\, \omega(t) \in A_t\}\right)$$

for all finite subsets I of $[0,T]$ and $\{A_t;\, t \in I\} \subset \mathscr{B}(\mathbb{R}^d)$, then $\mathbb{P}_1 = \mathbb{P}_2$.

(8) Fix $T > 0$ and consider the set $C_0([0,T];\mathbb{R}^d)$, which is collection of elements ω of $C([0,T];\mathbb{R}^d)$ for which $\omega(0) = 0$. Then $C_0([0,T];\mathbb{R}^d)$ inherits a Polish structure from $C([0,T];\mathbb{R}^d)$. Show that $\mathscr{M} \subset \mathscr{P}(C_0([0,T];\mathbb{R}^d))$ is tight if

$$\overline{\lim_{\delta\to 0}} \sup_{\mu\in\mathscr{M}} \mu\left\{ \sup_{\substack{|t-s|\leq\delta \\ s,t\in[0,T]}} |\omega(t) - \omega(s)| \geq \varepsilon \right\} = 0.$$

(9) Let X be a Polish space with metric ρ. Assume that for each $n \in \mathbb{N}$ (where $\mathbb{N}$ is the set of positive integers), we have a $\mu_n \in \mathscr{P}(X^n)$ (X^n is the n-fold product of X, which is Polish with the product topology). Assume furthermore that these μ_n's are *consistent*; that for any $n \in \mathbb{N}$ and any $A \in \mathscr{B}(X^n)$,

$$\mu_{n+1}\{(x_1, x_2 \dots x_{n+1}) \in X^{n+1} :\, (x_1, x_2 \dots x_n) \in A\} = \mu_n(A).$$

We will show that the μ_n's have a limit in the proper sense. If the μ_n's are themselves product measures (i.e., the law of independent random variables), this means that we can find a probability triple on which is defined a countably infinite collection of independent random variables). We can also use this to show

that, given a discrete-time Markov transition function (which we will not discuss in this course), we can find a probability space on which is defined a Markov process for all (discrete) time.

(a) Set $X^\infty \stackrel{\text{def}}{=} \times_{n\in\mathbb{N}} X$ and endow it with its natural product topology. For each $n \in \mathbb{N}$, define $\pi_n : X^\infty \to X$ as the natural projection operator (i.e., π_n gives the n-th coordinate). Show that X^∞ is metrizable with metric

$$\rho_\infty(x,y) \stackrel{\text{def}}{=} \sum_{n\in\mathbb{N}} 2^{-n} \frac{\rho(\pi_n(x),\pi_n(y))}{1+\rho(\pi_n(x),\pi_n(y))}. \qquad x,y \in X^\infty$$

(b) Fix $x^* \in X$, and for each $n \in \mathbb{N}$, define $\Phi_n : X^n \to X^\infty$ as

$$\Phi_n(x_1,x_2\dots x_n) \stackrel{\text{def}}{=} (x_1,x_2\dots x_n,x^*,x^*\dots).$$

In other words,

$$\pi_m\Phi_n(x_1,x_2\dots x_n) = \begin{cases} x_m & \text{if } m \le n \\ x^* & \text{else} \end{cases}$$

For each $n \in \mathbb{N}$, define

$$\tilde{\mu}_n(A) \stackrel{\text{def}}{=} \mu_n\{(x_1,x_2\dots x_n) \in X^n : \Phi_n(x_1,x_2\dots x_n) \in A\}. \qquad A \in \mathscr{B}(X^\infty)$$

Show that $\mu_n \in \mathscr{P}(X^\infty)$ by showing that each of the Φ_n's is measurable.

(c) Show that $\{\tilde{\mu}_n\}$ is tight. Note that by Tychonoff's theorem

$$K \stackrel{\text{def}}{=} \bigcap_{n\in\mathbb{N}} \pi_n^{-1}(K_n)$$

is compact for any collection $\{K_n\}$ of compact subsets of X.

(d) Let μ_∞ be a limit point of the $\tilde{\mu}_n$'s (in $\mathscr{P}(X^\infty)$). Show that for any $n \in \mathbb{N}$ and any $A \in \mathscr{B}(X^n)$,

$$\mu_\infty\{(x(1),x(2)\dots) \in X^\infty : (x(1),x(2)\dots x(n)) \in A\} = \mu_n(A).$$

Hint: first, note that if $\varphi \in C_b(X^n)$, then the mapping

$$\tilde{\varphi}_n(x(1),x(2)\dots) \stackrel{\text{def}}{=} \varphi(x(1),x(2)\dots x(n))$$

is an element of $C_b(X^\infty)$.

(e) Show that μ_∞ is unique. Hint: use problem 1.

CHAPTER 6

Construction of Wiener Measure

Let's now use a number of the tools we have developed to construct *Wiener measure*.

We will start out with a collection $\{\xi_k;\, k = 1, 2 \dots\}$ of independent and indentically distributed random variables with common law μ. We assume that

$$K_4 \stackrel{\text{def}}{=} \int_{\mathbb{R}} x^4 \mu(dx) < \infty \qquad \int_{\mathbb{R}} x\mu(dx) = 0 \qquad \int_{\mathbb{R}} x^2 \mu(dx) = 1.$$

Set

$$S_n \stackrel{\text{def}}{=} \begin{cases} \sum_{j=1}^n \xi_j & \text{if } n \geq 1 \\ 0 & \text{if } n = 0 \end{cases} \qquad n \in \mathbb{N}$$

and for each n, define a $C([0,1])$-valued random variable X^n by

$$X_t^n \stackrel{\text{def}}{=} \frac{1}{\sqrt{n}} \left(S_{\lfloor nt \rfloor} + (tn - \lfloor tn \rfloor) \xi_{\lfloor nt \rfloor + 1} \right) \qquad t \geq 0 \tag{26}$$

(note that X^n can be represented as a continuous mapping of the S_n's, so X^n is indeed measurable).

We are interested in the behavior of the law of X^n as n tends to infinity. To be more specific, define for each $n \in N$ an element $\mu_n \in \mathscr{P}(C([0,1]))$ (where $C([0,1])$) is endowed with the standard supremum-norm topology) by

$$\mu_n(A) \stackrel{\text{def}}{=} \mathbb{P}\left\{X^n \in A\right\}. \qquad A \in \mathscr{B}(C([0,1]))$$

We are interested in the limit of the μ_n's in the sense of weak convergence.

First of all, let's use a simple martingale inequality;

LEMMA 0.12. *For every $m \in \mathbb{N}$ and $L > 0$,*

$$\mathbb{P}\left\{ \max_{0 \leq j \leq m} |S_j| \geq L \right\} \leq (4/3)^4 (K_4 + 1) \frac{m^2}{L^4}. \tag{27}$$

PROOF. By Doob's maximal inequality,

$$\mathbb{P}\left\{ \max_{0 \leq j \leq m} |S_j| \geq L \right\} \leq (4/3)^4 \frac{\mathbb{E}[|S_m|^4]}{L^4}.$$

Now note that for any $n \geq 0$,

$$\mathbb{E}[S_m^4] = \sum_{i=1}^m \mathbb{E}[\xi_i^4] + \sum_{\substack{1 \leq i,j \leq m \\ i \neq j}} \mathbb{E}[\xi_i^2 \xi_j^2] = mK_4 + \binom{m}{2} \leq m^2 (K_4 + 1).$$

This gives us the proof. □

We can translate this into

LEMMA 0.13. *For every $n \in \mathbb{N}$, $m \in \mathbb{N}$, and $L > 0$,*

$$\mathbb{P}\left\{ \max_{\substack{0 \leq j,k \leq n \\ |j-k| \leq m}} |S_j - S_k| \geq L \right\} \leq (4/3)^4 (K_4 + 1) \frac{nm}{L^4}.$$

PROOF. We break n up into intervals of length m. Note that for any $0 \le j,k \le n$ with $|j-k| \le m$,

$$|S_k - S_j| \le |S_k - S_{\lfloor k/m \rfloor m}| + |S_j - S_{\lfloor j/m \rfloor m}| + |S_{\lfloor k/m \rfloor m} - S_{\lfloor j/m \rfloor m}| \le 3 \max_{0 \le l \le \lfloor n/m \rfloor} \max_{0 \le j \le m} |S_{lm+j} - S_{lm}|.$$

Thus

$$\mathbb{P}\left\{ \max_{\substack{0 \le j,k \le n \\ |j-k| \le m}} |S_j - S_k| \ge L \right\} \le \sum_{l=0}^{\lfloor n/m \rfloor} \mathbb{P}\left\{ \max_{0 \le j \le m} |S_{lm+j} - S_{lm}| \ge L/3 \right\}$$

$$= \lfloor n/m \rfloor \mathbb{P}\left\{ \max_{0 \le j \le m} |S_j| \ge L/3 \right\} \le (4/3)^4 (K_4 + 1) \lfloor n/m \rfloor \frac{m^2}{(L/3)^4}.$$

This yields the desired bound. □

From here we note

PROPOSITION 0.14. *For any $\delta > 0$, any $n \in \mathbb{N}$, and any $L > 0$,*

$$\mathbb{P}\left\{ \sup_{\substack{0 \le s,t \le 1 \\ |t-s| \le \delta}} |X_t^n - X_s^n| \ge L \right\} \le 4^4 (3 + K_4) \frac{\delta + n^{-1}}{L^4}$$

PROOF. First, for any s and t in $[0,1]$ with $|t-s| \le \delta$,

$$|X_t^n - X_s^n| \le |X^n_{\lfloor tn \rfloor / n} - X_t^n| + |X^n_{\lfloor sn \rfloor / n} - X_s^n| + |X^n_{\lfloor tn \rfloor / n} - X^n_{\lfloor sn \rfloor / n}| \le 3n^{-1/2} \max_{\substack{0 \le j,k \le n \\ |j-k| \le \delta n + 1}} |S_j - S_k|$$

Now use Lemma 2 with $m = \lfloor n\delta + 1 \rfloor \le n\delta + 1$. □

From here we get

PROPOSITION 0.15. *The collection $\{\mu_n;\, n \in \mathbb{N}\}$ is tight.*

PROOF. First, we obviously get that

$$\lim_{\delta \to 0} \overline{\lim_{n \to \infty}} \mathbb{P}\left\{ \sup_{\substack{0 \le s,t \le 1 \\ |t-s| \le \delta}} |X_t^n - X_s^n| \ge \varepsilon \right\} = 0$$

for each $\varepsilon > 0$. Fix now $\eta > 0$. For each $k \in \mathbb{N}$, fix $\delta_{1,k}$ and then n_k such that

$$\sup_{n \ge n_k} \mathbb{P}\left\{ \sup_{\substack{0 \le s,t \le 1 \\ |t-s| \le \delta_{1,k}}} |X_t^n - X_s^n| \ge 1/k \right\} \le \eta / 2^k.$$

Since

$$\lim_{\delta \to 0} \sup_{1 \le n \le n_k} \mathbb{P}\left\{ \sup_{\substack{0 \le s,t \le 1 \\ |t-s| \le \delta_{1,k}}} |X_t^n - X_s^n| \ge 1/k \right\} = 0,$$

we can find a $\delta_{2,k}$ such that

$$\sup_{1 \le n \le n_k} \mathbb{P}\left\{ \sup_{\substack{0 \le s,t \le 1 \\ |t-s| \le \delta_{2,k}}} |X_t^n - X_s^n| \ge 1/k \right\} \le \eta / 2^k.$$

Thus, upon setting $\delta_k \stackrel{\text{def}}{=} \min\{\delta_{1,k}, \delta_{2,k}\}$, we have that

$$\sup_{n \in \mathbb{N}} \mathbb{P}\left\{ \sup_{\substack{0 \le s,t \le 1 \\ |t-s| \le \delta_k}} |X_t^n - X_s^n| \ge 1/k \right\} \le \eta / 2^k.$$

From this, we set

$$C_k \stackrel{\text{def}}{=} \left\{ \varphi \in C([0,1]) : \varphi(0) = 0, \sup_{\substack{0 \le s,t \le 1 \\ |t-s| \le \delta_k}} |\varphi(t) - \varphi(s)| \le 1/k \right\} \qquad k \in \mathbb{N}$$

$$K \stackrel{\text{def}}{=} \overline{\bigcap_{k=1}^{\infty} C_k}$$

Thus

$$\sup_{n \in \mathbb{N}} \mu_n(K^c) \le \eta,$$

so the claimed result is true. □

Let's now define what will turn out to be the limit point. As usual, we define the coordinate random variables

$$X_t(\omega) \stackrel{\text{def}}{=} \omega(t). \qquad t \in [0,1],\, \omega \in C([0,1])$$

DEFINITION 0.16 (Wiener measure). A measure $\mu \in \mathscr{P}(C([0,1])$ is said to be Wiener measure on $C([0,1])$ if

- $X_0 = 0$ μ-a.s.
- For any $0 = t_0 < t_1 < t_2 \dots t_K \le 1$ $\{X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1} \dots X_{t_K} - X_{t_{K-1}}\}$ are jointly Gaussian and independent and $X_{t_j} - X_{t_{j-1}}$ is $\mathfrak{N}(0, t_j - t_{j-1})$.

Let's first show that Wiener measure, if it exists, is unique.

PROPOSITION 0.17. *Wiener measure, if it exists, is unique.*

PROOF. Let μ_1 and μ_2 in $\mathscr{P}(C([0,1])$ be Wiener measures. Fix $0 = t_0 < t_1 \cdots < t_K \le 1$ and $\{A_0, A_1 \dots A_K\} \in \mathscr{B}(\mathbb{R})$. Then for $i \in \{1,2\}$,

$$\mu_i\left(\cap_{i=1}^K \{X_{t_i} - X_{t_{i-1}} \in A_i\} \cap \{X_{t_0} \in A_0\}\right) = \delta_0(A_0) \prod_{i=1}^K \int_{z_i \in A_i} \frac{\exp\left[-\frac{z_i^2}{2(t_i - t_{i-1})}\right]}{\sqrt{2\pi(t_i - t_{i-1})}} dz$$

By the Dynkin $\pi - \lambda$ theorem, we thus have that

$$\mu_1\left\{\left(X_{t_0}, X_{t_1} - X_{t_0} \dots X_{t_K} - X_{t_{K-1}}\right) \in A\right\} = \mu_2\left\{\left(X_{t_0}, X_{t_1} - X_{t_0} \dots X_{t_K} - X_{t_{K-1}}\right) \in A\right\}$$

for all $A \in \mathscr{B}(\mathbb{R}^{K+1})$. Define now $T : \mathbb{R}^{K+1} \to \mathbb{R}^{K+1}$ as

$$T(x_0, x_1 \dots x_K) \stackrel{\text{def}}{=} (x_0, x_1 - x_0 \dots x_K - x_{K-1}). \qquad (x_0, x_1 \dots x_K) \in \mathbb{R}^{K+1}$$

Note that T is invertible with

$$T^{-1}(z_0, z_1 \dots z_K) \stackrel{\text{def}}{=} (z_0, z_1 + z_0 \cdots \sum_{j=1}^K z_j) \qquad (z_0, z_1 \dots z_K) \in \mathbb{R}^{K+1}$$

and that both T and T^{-1} are measurable from $\mathscr{B}(\mathbb{R}^{K+1})$ to itself. Thus

$$\mu_1\left\{(X_{t_0}, X_{t_1} \dots X_{t_K}) \in A\right\} = \mu_2\left\{(X_{t_0}, X_{t_1} \dots X_{t_K}) \in A\right\}$$

for all $A \in \mathscr{B}(\mathbb{R}^{K+1})$. Thus by one of the problems in Chapter 2, we know that $\mu_1 = \mu_2$. □

Finally, we claim that any limit points of the μ_n's of (26) is Wiener measure.

PROPOSITION 0.18. $\mu \stackrel{def}{=} \lim_{n \to \infty} \mu_n$ *exists and is Wiener measure.*

PROOF. First of all, Proposition 4 ensures that all subsequences of $\{\mu_n;\, n \in \mathbb{N}\}$ have convergent subsequences. It suffices to show that all convergent subsequences converge to Wiener measure. Let $\mu = \lim_{m\to\infty} \mu_{n_m}$. First, since $\mu \in \mathscr{P}(C([0,1]))$, we clearly have that $t \mapsto X_t$ is μ-a.s. continuous. Secondly, for any $\varepsilon > 0$, the set

$$F_\varepsilon \stackrel{\text{def}}{=} \{\varphi \in C([0,1]) : |\varphi(0)| \geq \varepsilon\}$$

is a closed subset of $C([0,1])$. Thus by weak convergence

$$\mu\{|X_0| \geq \varepsilon\} = \mu(F_\varepsilon) \leq \overline{\lim_{m\to\infty}}\, \mu_{m_n}(F_\varepsilon) \leq \overline{\lim_{m\to\infty}}\, \mathbb{P}\{|X_0^{n_m}| \geq \varepsilon\} = 0.$$

Thus, by taking ε to zero, we get that $X_0 = 0$ μ-a.s. Thirdly, fix $0 = t_0 < t_1 \cdots < t_K \leq 1$ and $\{\varphi_1, \varphi_2 \ldots \varphi_K\} \subset C_b(\mathbb{R})$ and define

$$\Phi(\omega) \stackrel{\text{def}}{=} \prod_{i=1}^K \varphi_i(\omega(t_i) - \omega(t_{i-1})); \qquad \omega \in C([0,1])$$

then $\Phi \in C_b(C([0,1])$. Thus

$$\begin{aligned}
\int_{C([0,1])} \Phi(\omega)\mu(d\omega) &= \lim_{m\to\infty} \mathbb{E}\left[\prod_{i=1}^K \varphi_i(X_{t_i}^{n_m} - X_{t_{i-1}}^{n_m})\right] = \lim_{m\to\infty} \mathbb{E}\left[\prod_{i=1}^K \varphi_i(X_{\lfloor t_i n_m\rfloor/n_m}^{n_m} - X_{\lfloor t_{i-1} n_m\rfloor/n_m}^{n_m})\right] \\
&= \prod_{i=1}^K \lim_{m\to\infty} \mathbb{E}\left[\varphi_i(X_{\lfloor t_i n_m\rfloor/n_m}^{n_m} - X_{\lfloor t_{i-1} n_m\rfloor/n_m}^{n_m})\right] \\
&= \prod_{i=1}^K \lim_{m\to\infty} \int_{\mathbb{R}} \varphi_i(z)(2\pi(t_i - t_{i-1}))^{-1/2} \exp\left[-\frac{z^2}{2(t_k - t_{k-1})}\right] dz.
\end{aligned}$$

The second equality holds because the φ_i's are in $C_b(C([0,1])$ and since for any $L > 0$,

$$\begin{aligned}
\overline{\lim_{n\to\infty}}\, \mathbb{P}\left\{\sup_{0\leq s,t\leq 1} \left|(X_t^n - X_s^n) - (X_{\lfloor tn\rfloor/n}^n - X_{\lfloor sn\rfloor/n}^n)\right| \geq L\right\} &\leq \overline{\lim_{n\to\infty}}\, \mathbb{P}\left\{\sup_{0\leq t\leq 1} |X_t^n - X_{\lfloor tn\rfloor/n}^n| \geq L/2\right\} \\
&\leq \overline{\lim_{n\to\infty}}\, \mathbb{P}\left\{\sup_{\substack{0\leq s,t\leq 1 \\ |t-s|\leq 1/n}} |X_t^n - X_{\lfloor tn\rfloor/n}^n| \geq L/2\right\} = 0
\end{aligned}$$

by using Proposition 3. The third equality comes from independence. The last equality comes from the central limit theorem. □

CHAPTER 7

References

Most of the material and questions in these notes come from Stroock's [**?**] and Billingsley's [**?**] books on probability. Some of the real analysis used comes from Royden [**?**].