8
Laws of Large Numbers

When measuring a physical quantity such as the mass of an object, it is commonly believed that the average of several measurements is more reliable than a single one. Similarly, in applications of statistical inference when estimating a population mean μ, a random sample $\{X_1, X_2, \ldots, X_n\}$ of size n is drawn from the population, and the sample average $\bar{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$ is used as an estimator for the parameter μ. This is based on the idea that as n gets large, $\bar{X}_n$ will be close to μ in some suitable sense. In many time-evolving physical systems $\{f(t) : 0 \leq t < \infty\}$, where $f(t)$ is an element in the phase space $\mathbb{S}$, "time averages" of the form $\frac{1}{T}\int_0^T h(f(t))dt$ (where h is a bounded function on $\mathbb{S}$) converge, as T gets large, to the "space average" of the form $\int_{\mathbb{S}} h(x)\pi(dx)$ for some appropriate measure π on $\mathbb{S}$. The above three are examples of a general phenomenon known as the *law of large numbers*. This chapter is devoted to a systematic development of this topic for sequences of independent random variables and also to some important refinements of the law of large numbers.

8.1 Weak laws of large numbers

Let $\{Z_n\}_{n\geq 1}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$. Recall that the sequence $\{Z_n\}_{n\geq 1}$ is said to *converge in probability* to a random variable Z if for each $\epsilon > 0$,

$$\lim_{n\to\infty} P(|Z_n - Z| \geq \epsilon) = 0. \qquad (1.1)$$

This is written as $Z_n \longrightarrow^p Z$. The sequence $\{Z_n\}_{n\geq 1}$ is said to *converge with probability one* or *almost surely* (a.s.) to Z if there exists a set A in $\mathcal{F}$ such that

$$P(A) = 1 \quad \text{and for all} \quad \omega \quad \text{in} \quad A, \quad \lim_{n\to\infty} Z_n(\omega) = Z(\omega). \tag{1.2}$$

This is written as $Z_n \to Z$ w.p. 1 or $Z_n \to Z$ a.s.

Definition 8.1.1: A sequence $\{X_n\}_{n\geq 1}$ of random variables on a probability space $(\Omega, \mathcal{F}, P)$ is said to obey the *weak law of large numbers* (WLLN) with normalizing sequences of real numbers $\{a_n\}_{n\geq 1}$ and $\{b_n\}_{n\geq 1}$ if

$$\frac{S_n - a_n}{b_n} \longrightarrow^p 0 \quad \text{as} \quad n \to \infty \tag{1.3}$$

where $S_n = \sum_{i=1}^n X_i$ for $n \geq 1$.

The following theorem says that if $\{X_n\}_{n\geq 1}$ is a sequence of iid random variables with $EX_1^2 < \infty$, then it obeys the weak law of large numbers with $a_n = nEX_1$ and $b_n = n$.

Theorem 8.1.1: *Let $\{X_n\}_{n\geq 1}$ be a sequence of iid random variables such that $EX_1^2 < \infty$. Then*

$$\bar{X}_n \equiv \frac{X_1 + \ldots + X_n}{n} \longrightarrow^p EX_1. \tag{1.4}$$

Proof: By Chebychev's inequality, for any $\epsilon > 0$,

$$P(|\bar{X}_n - EX_1| > \epsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = \frac{1}{\epsilon^2} \cdot \frac{\sigma^2}{n}, \tag{1.5}$$

where $\sigma^2 = \text{Var}(X_1)$. Since $\frac{\sigma^2}{n\epsilon^2} \to 0$ as $n \to \infty$, (1.4) follows. $\square$

Corollary 8.1.2: *Let $\{X_n\}_{n\geq 1}$ be a sequence of iid Bernoulli (p) random variables, i.e., $P(X_1 = 1) = p = 1 - P(X_1 = 0)$. Let*

$$\hat{p}_n = \frac{\#\{i : 1 \leq i \leq n, X_i = 1\}}{n}, \quad n \geq 1, \tag{1.6}$$

where for a finite set A, $\#A$ denotes the number of elements in A. Then $\hat{p}_n \longrightarrow^p p$.

Proof: Check that $EX_1 = p$ and $\hat{p}_n = \bar{X}_n$. $\square$

This says that one can estimate the probability p of getting a "head" of a coin by tossing it n times and calculating the proportion of "heads." This is also the basis of public opinion polls. Since the proof of Theorem 8.1.1 depended only on Chebychev's inequality, the following generalization is immediate (Problem 8.1).

Theorem 8.1.3: *Let $\{X_n\}_{n\ge1}$ be a sequence of random variables on a probability space such that*

(i) $EX_n^2 < \infty$ *for all* $n \ge 1$,

(ii) $EX_iX_j = (EX_i)(EX_j)$ *for all* $i \ne j$
(i.e., $\{X_n\}_{n\ge1}$ are uncorrelated),

(iii) $\frac{1}{n^2}\sum_{i=1}^n \sigma_i^2 \to 0$ *as* $n \to \infty$, *where* $\sigma_i^2 = Var(X_i)$, $i \ge 1$.

Then

$$\bar{X}_n - \bar{\mu}_n \longrightarrow^p 0 \tag{1.7}$$

where $\bar{\mu}_n \equiv \frac{1}{n}\sum_{i=1}^n EX_i$.

Corollary 8.1.4: *Let $\{X_n\}_{n\ge1}$ satisfy (i) and (ii) of the above theorem and let the sequence $\{\sigma_n^2\}_{n\ge1}$ be bounded. Let $\bar{\mu}_n \equiv \frac{1}{n}\sum_{i=1}^n EX_i \to \mu$ as $n \to \infty$. Then $\bar{X}_n \longrightarrow^p \mu$.*

An Application to Real Analysis

Let $f : [0,1] \to \mathbb{R}$ be a continuous function. K. Weierstrass showed that f can be approximated uniformly over $[0,1]$ by polynomials. S.N. Bernstein constructed a special class of such polynomials. A proof of Bernstein's result using the WLLN (Theorem 8.1.1) is given below.

Theorem 8.1.5: *Let $f : [0,1] \to \mathbb{R}$ be a continuous function. Let*

$$B_{n,f}(x) \equiv \sum_{r=0}^{n} f\Big(\frac{r}{n}\Big)\binom{n}{r} x^r(1-x)^{n-r}, \quad 0 \le x \le 1 \tag{1.8}$$

be the Bernstein polynomial of order n for the function f. Then

$$\lim_{n\to\infty} \sup\Big\{|f(x) - B_{n,f}(x)| : 0 \le x \le 1\Big\} = 0.$$

Proof: Since f is continuous on the closed and bounded interval $[0,1]$, it is uniformly continuous and hence for any $\epsilon > 0$, there exists a $\delta_\epsilon > 0$ such that

$$|x - y| < \delta_\epsilon \Rightarrow |f(x) - f(y)| < \epsilon. \tag{1.9}$$

Fix x in $[0,1]$. Let $\{X_n\}_{n\ge1}$ be a sequence of iid Bernoulli (x) random variables. Let $\hat{p}_n$ be as in (1.6). Then $B_{n,f}(x) = Ef(\hat{p}_n)$. Hence,

$$\begin{aligned} |f(x) - B_{n,f}(x)| &\le E|f(\hat{p}_n) - f(x)| \\ &= E\Big\{|f(\hat{p}_n) - f(x)| I(|\hat{p}_n - x| < \delta_\epsilon)\Big\} \\ &\quad + E\Big\{|f(\hat{p}_n) - f(x)| I(|\hat{p}_n - x| \ge \delta_\epsilon)\Big\} \\ &\le \epsilon + 2\|f\| P(|\hat{p}_n - x| \ge \delta_\epsilon) \end{aligned}$$

where $\|f\| = \sup\{|f(x)| : 0 \le x \le 1\}$. But by Chebychev's inequality,

$$\begin{aligned} P(|\hat{p}_n - x| \ge \delta_\epsilon) &\le \frac{1}{\delta_\epsilon^2}\mathrm{Var}(\hat{p}_n) \\ &= \frac{x(1-x)}{n\delta_\epsilon^2} \le \frac{1}{4n\delta_\epsilon^2} \quad \text{for all} \quad 0 \le x \le 1. \end{aligned}$$

Thus, $\sup\Big\{|f(x) - B_{n,f}(x)| : 0 \le x \le 1\Big\} \le \epsilon + 2\|f\|\frac{1}{4n\delta_\epsilon^2}$. Letting $n \to \infty$ first and then $\epsilon \downarrow 0$ completes the proof. □

8.2 Strong laws of large numbers

.

Definition 8.2.1: A sequence $\{X_n\}_{n\ge1}$ of random variables on a probability space $(\Omega, \mathcal{F}, P)$ is said to obey the *strong law of large numbers* (SLLN) with normalizing sequences of real numbers $\{a_n\}_{n\ge1}$ and $\{b_n\}_{n\ge1}$ if

$$\frac{S_n - a_n}{b_n} \to 0 \quad \text{as} \quad n \to \infty \quad \text{w.p. 1}, \tag{2.1}$$

where $S_n = \sum_{i=1}^n X_i$ for $n \ge 1$.

The following theorem says that if $\{X_n\}_{n\ge1}$ is a sequence of iid random variables with $EX_1^4 < \infty$, then the strong law of large numbers holds with $a_n = nEX_1$ and $b_n = n$. This result is referred to as Borel's SLLN.

Theorem 8.2.1: (*Borel's SLLN*). *Let $\{X_n\}_{n\ge1}$ be a sequence of iid random variables such that $EX_1^4 < \infty$. Then*

$$\bar{X}_n \equiv \frac{X_1 + X_2 + \ldots + X_n}{n} \to EX_1 \quad w.p.\ 1. \tag{2.2}$$

Proof: Fix $\epsilon > 0$ and let $A_n \equiv \{|\bar{X}_n - EX_1| \ge \epsilon\}$, $n \ge 1$. To establish (2.2), by Proposition 7.2.3 (a), it suffices to show that

$$\sum_{n=1}^{\infty} P(A_n) < \infty. \tag{2.3}$$

By Markov's inequality

$$P(A_n) \le \frac{E|\bar{X}_n - EX_1|^4}{\epsilon^4}. \tag{2.4}$$

Let $Y_i = X_i - EX_1$ for $i \ge 1$. Since the X_i's are independent, it is easy to check that

$$E|\bar{X}_n - EX_1|^4 = \frac{1}{n^4}E\left(\left(\sum_{i=1}^n Y_i\right)^4\right)$$

$$\begin{aligned} &= \frac{1}{n^4}\Big(nEY_1^4 + 3n(n-1)(EY_1^2)^2\Big) \\ &= O(n^{-2}). \end{aligned}$$

By (2.4) this implies (2.3). □

The following two results are easy consequences of the above theorem.

Corollary 8.2.2: *Let $\{X_n\}_{n\ge1}$ be a sequence of iid random variables that are bounded, i.e., there exists a $C < \infty$ such that $P(|X_1| \le C) = 1$. Then*

$$\bar{X}_n \to EX_1 \quad \textit{w.p. 1.}$$

Corollary 8.2.3: *Let $\{X_n\}_{n\ge1}$ be a sequence of iid Bernoulli(p) random variables. Then*

$$\hat{p}_n \equiv \frac{\#\{i : 1 \le i \le n, X_i = 1\}}{n} \to p \quad \textit{w.p. 1.} \tag{2.5}$$

An application of the above result yields the following theorem on the uniform convergence of the empirical cdf to the true cdf.

Theorem 8.2.4: (*Glivenko-Cantelli*). *Let $\{X_n\}_{n\ge1}$ be a sequence of iid random variables with a common cdf $F(\cdot)$. Let $F_n(\cdot)$, the empirical cdf based on $\{X_1, X_2, \ldots, X_n\}$, be defined by*

$$F_n(x) \equiv \frac{1}{n}\sum_{j=1}^{n} I(X_j \le x), \quad x \in \mathbb{R}. \tag{2.6}$$

Then,

$$\tilde{\Delta}_n \equiv \sup_x |F_n(x) - F(x)| \to 0 \quad \textit{w.p. } 1. \tag{2.7}$$

Remark 8.2.1: Note that by applying Corollary 8.2.3 to the sequence of Bernoulli random variables $\{Y_n \equiv I(X_n \le x)\}_{n\ge1}$, one may conclude that $F_n(x) \to F(x)$ w.p. 1 for each *fixed* x. So the main thrust of this theorem is the *uniform* convergence on $\mathbb{R}$ of F_n to F w.p. 1. It can be shown that (2.7) holds for sequences $\{X_n\}_{n\ge1}$ that are identically distributed and only pairwise independent. The proof is based on Etemadi's SLLN (Theorem 8.2.7) below.

The proof of Theorem 8.2.4 makes use of the following two lemmas.

Lemma 8.2.5: (*Scheffe's theorem: A generalized version*). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $\{f_n\}_{n\ge1}$ and f be nonnegative μ-integrable functions such that as $n \to \infty$, (i) $f_n \to f$ a.e. (μ) and (ii) $\int f_n d\mu \to \int f d\mu$. Then $\int |f - f_n| d\mu \to 0$ as $n \to \infty$.*

Proof: See Theorem 2.5.4. □

For any bounded monotone function $H: \mathbb{R} \to \mathbb{R}$, define

$$H(\infty) \equiv \lim_{x \uparrow \infty} H(x), \quad H(-\infty) \equiv \lim_{x \downarrow -\infty} H(x).$$

Lemma 8.2.6: *(Polyā's theorem). Let $\{G_n\}_{n \geq 1}$ and G be a collection of bounded nondecreasing functions on $\mathbb{R} \to \mathbb{R}$ such that $G(\cdot)$ is continuous on $\mathbb{R}$ and*

$$G_n(x) \to G(x) \text{ for all } x \text{ in } D \cup \{-\infty, +\infty\},$$

where D is dense in $\mathbb{R}$. Then $\Delta_n \equiv \sup\{|G_n(x) - G(x)| : x \in \mathbb{R}\} \to 0$. That is, $G_n \to G$ uniformly on $\mathbb{R}$.

Proof: Fix $\epsilon > 0$. By the definitions of $G(\infty)$ and $G(-\infty)$, there exist C_1 and C_2 in D such that

$$G(C_1) - G(-\infty) < \epsilon, \quad \text{and} \quad G(\infty) - G(C_2) < \epsilon. \tag{2.8}$$

Since $G(\cdot)$ is continuous, it is uniformly continuous on $[C_1, C_2]$ and so there exists a $\delta > 0$ such that

$$x, y \in [C_1, C_2], |x - y| < \delta \Rightarrow |G(x) - G(y)| < \epsilon. \tag{2.9}$$

Also, there exist points $a_1 = C_1 < a_2 < \ldots < a_k = C_2$, $1 < k < \infty$, in D such that

$$\max\{(a_{i+1} - a_i) : 1 \leq i \leq k-1\} < \delta.$$

Let $a_0 = -\infty$, $a_{k+1} = \infty$. By the convergence of $G_n(\cdot)$ to $G(\cdot)$, on $D \cup \{-\infty, \infty\}$,

$$\Delta_{n1} \equiv \max\{|G_n(a_i) - G(a_i)| : 0 \leq i \leq k+1\} \to 0 \tag{2.10}$$

as $n \to \infty$. Now note that for any x in $[a_i, a_{i+1}]$, $1 \leq i \leq k-1$, by the monotonicity of $G_n(\cdot)$ and $G(\cdot)$, and by (2.9) and (2.10),

$$\begin{aligned} G_n(x) - G(x) &\leq G_n(a_{i+1}) - G(a_i) \\ &\leq G_n(a_{i+1}) - G(a_{i+1}) + G(a_{i+1}) - G(a_i) \\ &\leq \Delta_{n1} + \epsilon, \end{aligned}$$

and similarly,

$$G_n(x) - G(x) \geq -\Delta_{n1} - \epsilon.$$

Thus

$$\sup\{|G_n(x) - G(x)| : a_1 \leq x \leq a_k\} \leq \Delta_{n1} + \epsilon. \tag{2.11}$$

For $x < a_1$, by (2.8) and (2.10),

$$\begin{aligned} |G_n(x) - G(x)| &\leq |G_n(x) - G_n(-\infty)| + |G_n(-\infty) - G(-\infty)| \\ &\quad + |G(-\infty) - G(x)| \\ &\leq (G_n(a_1) - G_n(-\infty)) + |G_n(-\infty) - G(-\infty)| + \epsilon \\ &\leq |G_n(a_1) - G(a_1)| + |G(a_1) - G(-\infty)| \\ &\quad + 2|G_n(-\infty) - G(-\infty)| + \epsilon \\ &\leq 3\Delta_{n1} + 2\epsilon. \end{aligned}$$

Similarly, for $x > a_k$,

$$|G_n(x) - G(x)| \leq 3\Delta_{n1} + 2\epsilon.$$

Combining the above with (2.11) yields

$$\Delta_n \leq 3\Delta_{n1} + 2\epsilon.$$

By (2.10),

$$\limsup_{n\to\infty} \Delta_n \leq 2\epsilon,$$

and $\epsilon > 0$ being arbitrary, the proof is complete. □

Proof of Theorem 8.2.4: First note that $\tilde{\Delta}_n = \sup_{x\in\mathbb{Q}} |F_n(x) - F(x)|$ and hence, it is a random variable. Let $B \equiv \{b_j : j \in J\}$ be the set of jump discontinuity points of F with the corresponding jump sizes $\{p_j : j \in J\}$, where J is a subset of $\mathbb{N}$. Let $p = \sum_{j\in J} p_j$.

Note that

$$\begin{aligned} F_n(x) &= \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x) \\ &= \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x, X_i \in B) + \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x, X_i \notin B) \\ &= F_{nd}(x) + F_{nc}(x), \text{ say.} \end{aligned} \tag{2.12}$$

Then, $F_{nd}(x) = \sum_{j\in J} \hat{p}_{nj} I(b_j \leq x)$, where

$$\hat{p}_{nj} = \frac{\#\{i : 1 \leq i \leq n, X_i = b_j\}}{n}.$$

Let $\hat{p}_n = \sum_{j\in J} \hat{p}_{nj} = \frac{1}{n} \cdot \#\{i : 1 \leq i \leq n, X_i \in B\}$. By Corollary 8.2.3, for each $j \in J$,

$$\hat{p}_{nj} \to p_j \quad \text{w.p. 1 and} \quad \hat{p}_n \to p \quad \text{w.p. 1.}$$

Since B is countable, there exists a set A_0 in $\mathcal{F}$ such that $P(A_0) = 1$ and for all ω in A_0, $\hat{p}_{nj} \to p_j$ for all $j \in J$ and $\sum_{j\in J} \hat{p}_{nj} = \hat{p}_n \to p = \sum_{j\in J} p_j$.

By Lemma 8.2.5 (applied with μ being the counting measure on the set J), it follows that for ω in A_0,

$$\sum_{j\in J} |\hat{p}_{nj} - p_j| \to 0. \tag{2.13}$$

Let $F_d(x) \equiv \sum_{j\in J} p_j I(b_j \le x),\ x \in \mathbb{R}$. Then,

$$\sup_{x\in\mathbb{R}} |F_{nd}(x) - F_d(x)| \le \sum_{j\in J} |\hat{p}_{nj} - p_j|, \tag{2.14}$$

which $\to 0$ as $n \to \infty$ for all ω in A_0, by (2.13).

Next let,

$$F_c(x) \equiv F(x) - F_d(x), \quad x \in \mathbb{R}.$$

Then, it is easy to check that, $F_c(\cdot)$ is continuous and nondecreasing on $\mathbb{R}$, $F_c(-\infty) = 0$ and $F_c(\infty) = 1 - p$.

Again, by Corollary 8.2.3, there exists a set A_1 in $\mathcal{F}$ such that $P(A_1) = 1$ and for all ω in A_1,

$$F_{nc}(x) \to F_c(x)$$

for all rational x in $\mathbb{R}$ and

$$F_{nc}(\infty) \equiv 1 - \hat{p}_n \to 1 - p = F_c(\infty).$$

Also, $F_{nc}(-\infty) = 0 = F_c(-\infty)$. So by Lemma 8.2.6, with $D = \mathbb{Q}$, for ω in A_1,

$$\sup_{x\in\mathbb{R}} |F_{nc}(x) - F_c(x)| \to 0 \quad \text{as} \quad n \to \infty. \tag{2.15}$$

Since $P(A_0 \cap A_1) = 1$, the theorem follows from (2.12)–(2.15). $\square$

Borel's SLLN for iid random variables requires that $E|X_1|^4 < \infty$. Kolmogorov (1956) improved on this significantly by using his "3-series" theorem and reduced the moment condition to $E|X_1| < \infty$. More recently, Etemadi (1981) N. improved this further by assuming only that the $\{X_n\}_{n\ge1}$ are *pairwise independent* and identically distributed with $E|X_1| < \infty$. More precisely, he proved the following.

Theorem 8.2.7: (*Etemadi's SLLN*). *Let $\{X_n\}_{n\ge1}$ be a sequence of pairwise independent and identically distributed random variables with $E|X_1| < \infty$. Then*

$$\bar{X}_n \to EX_1 \quad \textit{w.p. 1.} \tag{2.16}$$

Proof: The main steps in the proof are

(I) reduction to the nonnegative case,

(II) proof of convergence of $\bar{Y}_n$ along a geometrically growing subsequence using the Borel-Cantelli lemma and Chebychev's inequality, where $\bar{Y}_n$ is the average of certain truncated versions of $X_1, \ldots, X_n$, and extending the convergence from the geometric subsequence to the full sequence.

STEP I: Since the $\{X_n\}_{n \geq 1}$ are pairwise independent and identically distributed with $E|X_1| < \infty$, it follows that $\{X_n^+\}_{n\geq 1}$ and $\{X_n^-\}_{n\geq 1}$ are both sequences of pairwise independent and identically distributed nonnegative random variables with $EX_1^+ < \infty$ and $EX_1^- < \infty$. Also, since

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^+\right) - \left(\frac{1}{n}\sum_{i=1}^{n} X_i^-\right),$$

it is enough to prove the theorem under the assumption that the X_i's are nonnegative.

STEP II: Now let X_i's be nonnegative and let

$$Y_i = X_i I(X_i \leq i), \quad i \geq 1.$$

Then,

$$\begin{aligned}
\sum_{i=1}^{\infty} P(X_i \neq Y_i) &= \sum_{i=1}^{\infty} P(X_i > i) \\
&= \sum_{i=1}^{\infty} P(X_1 > i) \leq \sum_{i=1}^{\infty} \int_{i-1}^{i} P(X_1 > t)dt \\
&= \int_0^{\infty} P(X_1 > t)dt \\
&= EX_1 < \infty.
\end{aligned}$$

Hence, by the Borel-Cantelli lemma,

$$P(X_i \neq Y_i, \text{ infinitely often}) = 0.$$

This implies that w.p. 1, $X_i = Y_i$ for all but finitely many i's and hence, it suffices to show that

$$\bar{Y}_n \equiv \frac{1}{n}\sum_{i=1}^{n} Y_i \to EX_1 \text{ w.p. 1.} \tag{2.17}$$

Next, $EY_i = E(X_i I(X_i \leq i) = E(X_1 I(X_1 \leq i)) \to EX_1$ (by the MCT) and hence

$$E\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} EY_i \to EX_1 \quad \text{as} \quad n \to \infty. \tag{2.18}$$

Suppose for the moment that for each fixed $1 < \rho < \infty$, it is shown that

$$\bar{Y}_{n_k} \to EX_1 \text{ as } k \to \infty \quad \text{w.p. } 1 \tag{2.19}$$

where $n_k = \lfloor \rho^k \rfloor$ = the greatest integer less than or equal to ρ^k, $k \in \mathbb{N}$. Then, since the Y_i's are nonnegative, for any n and k satisfying $\rho^k \le n < \rho^{k+1}$, one gets

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n_k} Y_i \le \bar{Y}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j \le \frac{1}{n}\sum_{i=1}^{n_{k+1}} Y_i \\ \Longrightarrow \quad & \frac{n_k}{n}\bar{Y}_{n_k} \le \bar{Y}_n \le \frac{n_{k+1}}{n}\bar{Y}_{n_{k+1}} \\ \Longrightarrow \quad & \frac{1}{\rho}\bar{Y}_{n_k} \le \bar{Y}_n \le \rho\bar{Y}_{n_{k+1}}. \end{aligned}$$

From (2.19), it follows that

$$\frac{1}{\rho} EX_1 \le \liminf_{n\to\infty} \bar{Y}_n \le \limsup_{n\to\infty} \bar{Y}_n \le \rho EX \quad \text{w.p. } 1.$$

Since this is true for each $1 < \rho < \infty$, by taking $\rho = 1 + \frac{1}{r}$ for $r = 1, 2, \ldots$, it follows that

$$EX_1 \le \liminf_{n\to\infty} \bar{Y}_n \le \limsup_{n\to\infty} \bar{Y}_n \le EX_1 \quad \text{w.p. } 1,$$

establishing (2.17).

It now remains to prove (2.19). By (2.18), it is enough to show that

$$\bar{Y}_{n_k} - E\bar{Y}_{n_k} \to 0 \quad \text{as} \quad k \to \infty, \quad \text{w.p. } 1. \tag{2.20}$$

By Chebychev's inequality and the pairwise independence of the variables $\{Y_n\}_{n\ge 1}$, for any $\epsilon > 0$,

$$\begin{aligned} P(|\bar{Y}_{n_k} - E\bar{Y}_{n_k}| > \epsilon) \quad & \le \quad \frac{1}{\epsilon^2}\operatorname{Var}(\bar{Y}_{n_k}) = \frac{1}{\epsilon^2}\frac{1}{n_k^2}\sum_{i=1}^{n_k}\operatorname{Var}(Y_i) \\ & \le \quad \frac{1}{\epsilon^2}\frac{1}{n_k^2}\sum_{i=1}^{n_k} EY_i^2. \end{aligned}$$

Thus,

$$\begin{aligned} \sum_{k=1}^{\infty} P(|\bar{Y}_{n_k} - E\bar{Y}_{n_k}| > \epsilon) \quad & \le \quad \frac{1}{\epsilon^2}\sum_{k=1}^{\infty}\frac{1}{n_k^2}\sum_{i=1}^{n_k} EY_i^2 \\ & = \quad \frac{1}{\epsilon^2}\sum_{i=1}^{\infty} EY_i^2 \Big(\sum_{k:n_k\ge i}\frac{1}{n_k^2}\Big). \end{aligned} \tag{2.21}$$

Since $n_k = \lfloor \rho^k \rfloor > \rho^{k-1}$ for $1 < \rho < \infty$,

$$\sum_{k:n_k \geq i} \frac{1}{n_k^2} \leq \sum_{k:\rho^{k-1} \geq i} \frac{1}{\rho^{(k-1)2}} \leq \frac{C_1}{i^2} \tag{2.22}$$

for some constant C_1, $0 < C_1 < \infty$.

Next, since the X_i's are identically distributed,

$$\begin{aligned}
\sum_{i=1}^{\infty} \frac{EY_i^2}{i^2} &= \sum_{i=1}^{\infty} \frac{EX_1^2 I(0 \leq X_1 \leq i)}{i^2} \\
&= \sum_{i=1}^{\infty} \sum_{j=1}^{i} \frac{EX_1^2 I(j-1 < X_1 \leq j)}{i^2} \\
&= \sum_{j=1}^{\infty} \big(EX_1^2 I(j-1 < X_1 \leq j)\big) \sum_{i=j}^{\infty} i^{-2} \\
&\leq \sum_{j=1}^{\infty} \big(jEX_1 I(j-1 < X_1 \leq j)\big) \cdot C_2 j^{-1} \\
&= C_2 EX_1 < \infty, \qquad (2.23)
\end{aligned}$$

for some constant C_2, $0 < C_2 < \infty$.

Now (2.21)–(2.23) imply that

$$\sum_{k=1}^{\infty} P(|\bar{Y}_{n_k} - E\bar{Y}_{n_k}| > \epsilon) < \infty$$

for each $\epsilon > 0$. By the Borel-Cantelli lemma and Proposition 7.2.3 (a), (2.20) follows and the proof is complete. □

The following corollary is immediate from the above theorem.

Corollary 8.2.8: (*Extension to the vector case*). *Let* $\{X_n = (X_{n1}, \ldots, X_{nk})\}_{n \geq 1}$ *be a sequence of* k*-dimensional random vectors defined on a probability space* $(\Omega, \mathcal{F}, P)$ *such that for each* i, $1 \leq i \leq k$, *the sequence* $\{X_{ni}\}_{n \geq 1}$ *are pairwise independent and identically distributed with* $E|X_{1i}| < \infty$. *Let* $\mu = (EX_{11}, EX_{12}, \ldots, EX_{1k})$ *and* $f : \mathbb{R}^k \to \mathbb{R}$ *be continuous at* μ. *Then*

(i) $\bar{X}_n \equiv (\bar{X}_{n1}, \bar{X}_{n2}, \ldots, \bar{X}_{nk}) \to \mu$ *w.p. 1, where* $\bar{X}_{ni} = \frac{1}{n} \sum_{j=1}^{n} X_{ji}$ *for* $1 \leq i \leq k$.

(ii) $f(\bar{X}_n) \to f(\mu)$ *w.p. 1.*

Example 8.2.1: Let (X_n, Y_n), $n = 1, 2, \ldots$ be a sequence of bivariate iid random vectors with $EX_1^2 < \infty$, $EY_1^2 < \infty$. Then the *sample correlation* *coefficient* $\hat{\rho}_n$, defined by,

$$\hat{\rho}_n \equiv \frac{\left(\frac{1}{n}\sum_{i=1}^n X_iY_i - \bar{X}_n\bar{Y}_n\right)}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2\right)\left(\frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y}_n)^2\right)}}$$

is a *strongly consistent* estimator of the *population correlation coefficient* ρ, defined by,

$$\rho = \frac{\mathrm{Cov}(X_1, Y_1)}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(Y_1)}},$$

i.e., $\hat{\rho}_n \to \rho$ w.p. 1. This follows from the above corollary by taking $f : \mathbb{R}^5 \to \mathbb{R}$ to be

$$f(t_1, t_2, t_3, t_4, t_5) = \begin{cases} \dfrac{t_5 - t_1t_2}{\sqrt{(t_3 - t_1^2)(t_4 - t_2^2)}}, & \text{for } t_3 > t_1^2,\ t_4 > t_2^2 \\ 0, & \text{otherwise,} \end{cases}$$

and the vector $(X_{n1}, X_{n2}, \ldots, X_{n5})$ to be

$$X_{n1} = X_n,\ X_{n2} = Y_n,\ X_{n3} = X_n^2,\ X_{n4} = Y_n^2,\ X_{n5} = X_nY_n.$$

Corollary 8.2.9: (*Extension to the pairwise m-dependent case*). *Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$ such that for an integer m, $1 \leq m < \infty$, and for each i, $1 \leq i \leq m$, the random variables $\{X_i, X_{i+m}, X_{i+2m}, \ldots\}$ are identically distributed and pairwise independent with $E|X_i| < \infty$. Then*

$$\bar{X}_n \to \frac{1}{m}\sum_{i=1}^m EX_i \quad \textit{w.p. } 1.$$

The proof is left as an exercise (Problem 8.2). For an application of the above result to a discussion on *normal numbers*, see Problem 8.15.

Example 8.2.2: (*IID Monte Carlo*). Let $(\mathbb{S}, \mathcal{S}, \pi)$ be a probability space, $f \in L^1(\mathbb{S}, \mathcal{S}, \pi)$ and $\lambda = \int_{\mathbb{S}} f d\pi$. Let $\{X_n\}_{n\geq 1}$ be a sequence of iid $\mathbb{S}$-valued random variables with distribution π. Then, the *IID Monte Carlo approximation* to λ is defined as

$$\hat{\lambda}_n \equiv \frac{1}{n}\sum_{i=1}^n f(X_i).$$

Note that by the SLLN, $\hat{\lambda}_n \to \lambda$ w.p. 1.

An extension of this to the case where $\{X_i\}_{i\geq 1}$ is a Markov chain, known as Markov chain Monte Carlo (MCMC), is discussed in Chapter 14.

8.3 Series of independent random variables

Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables on a probability space $(\Omega, \mathcal{F}, P)$. The goal of this section is to investigate the convergence of the infinite series $\sum_{n=1}^{\infty} X_n$, i.e., that of the partial sum sequence, $S_n = \sum_{i=1}^{n} X_i$, $n \geq 1$.

The main result of this section is Kolmogorov's 3-series theorem (Theorem 8.3.5). The following two inequalities play a fundamental role in the proof of this theorem and also have other important applications.

Theorem 8.3.1: *Let $\{X_j : 1 \leq j \leq n\}$ be a collection of independent random variables. Let $S_i = \sum_{j=1}^{i} X_j$ for $1 \leq i \leq n$.*

(i) (Kolmogorov's first inequality). Suppose that $EX_j = 0v$ and $EX_j^2 < \infty$, $1 \leq j \leq n$. Then, for $0 < \lambda < \infty$,

$$P\Big(\max_{1\leq i\leq n} |S_i| \geq \lambda\Big) \leq \frac{\mathrm{Var}(S_n)}{\lambda^2}. \tag{3.1}$$

(ii) (Kolmogorov's second inequality). Suppose that there exists a constant $C \in (0,\infty)$ such that $P(|X_j - EX_j| \leq C) = 1$ for $1 \leq j \leq n$. Then, for any $0 < \lambda < \infty$,

$$P\Big(\max_{1\leq i\leq n} |S_i| \leq \lambda\Big) \leq \frac{(2C + 4\lambda)^2}{\mathrm{Var}(S_n)}.$$

Proof: Let $A = \{\max_{1\leq i\leq n} |S_i| \geq \lambda\}$ and let

$$\begin{aligned} A_1 &= \{|S_1| \geq \lambda\}, \\ A_j &= \{|S_1| < \lambda, |S_2| < \lambda, \ldots, |S_{j-1}| < \lambda, |S_j| \geq \lambda\} \end{aligned}$$

for $j = 2, \ldots, n$. Then $A_1, \ldots, A_n$ are disjoint, $\bigcup_{j=1}^{n} A_j = A$ and $P(A) = \sum_{j=1}^{n} P(A_j)$. Since $EX_j = 0$ for all j,

$$\begin{aligned} \mathrm{Var}(S_n) = ES_n^2 &\geq E(S_n^2 I_A) = \sum_{j=1}^{n} E(S_n^2 I_{A_j}) \\ &= \sum_{j=1}^{n} E\Big[\Big((S_n - S_j)^2 + S_j^2 + 2(S_n - S_j)S_j\Big) I_{A_j}\Big] \\ &\geq \sum_{j=1}^{n} E(S_j^2 I_{A_j}) + 2\sum_{j=1}^{n-1} E\Big((S_n - S_j)S_j I_{A_j}\Big). \end{aligned} \tag{3.2}$$

Note that since $\{X_1, \ldots, X_n\}$ are independent, $(S_n - S_j) \equiv \sum_{i=j+1}^{n} X_i$ and $S_j I_{A_j}$ are independent for $1 \leq j \leq n-1$. Hence,

$$E\Big((S_n - S_j)S_j I_{A_j}\Big) = E(S_n - S_j)E(S_j I_{A_j}) = 0.$$

Also on A_j, $S_j^2 \ge \lambda^2$. Therefore, by (3.2),

$$\mathrm{Var}(S_n) \ge \sum_{j=1}^{n} \lambda^2 P(A_j) = \lambda^2 P(A).$$

This establishes (i). For a proof of (ii), see Chung (1974), p. 117. □

Remark 8.3.1: Recall that Chebychev's inequality asserts that for $\lambda > 0$, $P(|S_n| \ge \lambda) \le \frac{\mathrm{Var}(S_n)}{\lambda^2}$ and thus Kolmogorov's first inequality is significantly stronger. Kolmogorov's first inequality has an extension known as Doob's maximal inequality to a class of dependent random variables, called *martingales* (see Chapter 13). The next inequality is due to P. Levy.

Definition 8.3.1: For any random variable X, a real number c is called a *median of* X if

$$P(X < c) \le \frac{1}{2} \le P(X \le c). \tag{3.3}$$

Such a c always exists. It can be verified that $c_0 \equiv \inf\{x : P(X \le x) \ge \frac{1}{2}\}$ is a median. Note that if c is a median of X and α is a real number, then αc is a median of αX and $\alpha + c$ is a median of $\alpha + X$. Further, if $P(|X| \ge \alpha) < \frac{1}{2}$ for some $\alpha > 0$, then any median c of X satisfies $|c| \le \alpha$ (Problem 8.4).

Theorem 8.3.2: (*Levy's inequality*). *Let X_j, $j = 1, \ldots, n$ be independent random variables. Let $S_j = \sum_{j=1}^{n} X_i$, and $c_{j,n}$ be a median of $(S_n - S_j)$ for $1 \le j \le n$, where $c_{n,n}$ is set equal to 0. Then, for any $0 < \lambda < \infty$,*

(i) $P\Big(\max_{1 \le j \le n} (S_j - c_{j,n}) \ge \lambda\Big) \le 2P(S_n \ge \lambda)$;

(ii) $P\Big(\max_{1 \le j \le n} |S_j - c_{j,n}| \ge \lambda\Big) \le 2P(|S_n| \ge \lambda)$.

Proof: Let

$$\begin{aligned} A_j &= \{S_j - S_n \le c_{j,n}\} \quad \text{for} \quad 1 \le j \le n, \\ B &= \Big\{\max_{1 \le j \le n} (S_j - c_{j,n}) \ge \lambda\Big\}, \\ B_1 &= \{S_1 - c_{1,n} \ge \lambda\} \\ B_j &= \{S_i - c_{i,n} < \lambda \quad \text{for} \quad 1 \le i \le j-1,\ S_j - c_{j,n} \ge \lambda\}, \end{aligned}$$

for $j = 2, \ldots, n$. Then $B_1, \ldots, B_n$ are disjoint and $\bigcup_{j=1}^{n} B_j = B$. Since $X_1, \ldots, X_n$ are independent, A_j and B_j are independent for each $j = 1, 2, \ldots, n$. Also for each j, $A_j = \{S_j - c_{j,n} \le S_n\}$, and hence on $A_j \cap B_j$, $S_n \ge \lambda$ holds. Thus,

$$P(S_n \ge \lambda) \quad \ge \quad \sum_{j=1}^{n} P(A_j \cap B_j)$$

$$
\begin{aligned}
&= \sum_{j=1}^{n} P(A_j)P(B_j) \\
&\geq \frac{1}{2} P\Big(\bigcup_{j=1}^{n} B_j\Big) \\
&= \frac{1}{2} P(B),
\end{aligned}
$$

proving part (i). Now part (ii) follows by applying part (i) to both $\{X_i\}_{i=1}^n$ and $\{-X_i\}_{i=1}^n$. $\square$

Recall that if $\{Y_n\}_{n\geq 1}$ is a sequence of random variables, then $\{Y_n\}_{n\geq 1}$ converges w.p. 1 implies that $\{Y_n\}_{n\geq 1}$ converges in probability as well. A remarkable result of P. Levy is that if $\{S_n\}_{n\geq 1}$ is the sequence of partial sums of *independent* random variables and $\{S_n\}_{n\geq 1}$ converges in probability, then $\{S_n\}_{n\geq 1}$ must converge w.p. 1 as well. The proof of this uses Levy's inequality proved above.

Theorem 8.3.3: *Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables. Let $S_n = \sum_{j=1}^n X_j$ for $1 \leq n < \infty$ and let $\{S_n\}_{n\geq 1}$ converge in probability to a random variable S. Then $S_n \to S$ w.p. 1.*

Proof: Recall that a sequence $\{x_n\}_{n\geq 1}$ of real numbers converges iff it is Cauchy iff $\delta_n \equiv \sup\{|x_k - x_\ell| : k, \ell \geq n\} \to 0$ as $n \to \infty$. Let

$$
\begin{aligned}
\tilde{\Delta}_n &\equiv \sup\{|S_k - S_\ell| : k, \ell \geq n\} \text{ and} \\
\Delta_n &\equiv \sup\{|S_k - S_n| : k \geq n\}.
\end{aligned}
$$

Then, $\tilde{\Delta}_n \leq 2\Delta_n$ and $\tilde{\Delta}_n$ is decreasing in n. Suppose it is shown that

$$
\Delta_n \longrightarrow^p 0. \tag{3.4}
$$

Then, $\tilde{\Delta}_n \longrightarrow^p 0$ and hence there is a subsequence $\{n_k\}_{k\geq 1}$ such that $\tilde{\Delta}_{n_k} \to 0$ as $k \to \infty$ w.p. 1. Since $\tilde{\Delta}_n$ is decreasing in n, this implies that $\tilde{\Delta}_n \to 0$ w.p. 1. Thus it suffices to establish (3.4). Fix $0 < \epsilon < 1$. Let

$$
\begin{aligned}
S_{n,\ell} &= S_{n+\ell} - S_n \quad \text{for} \quad \ell \geq 1, \\
\Delta_{n,k} &= \max\{|S_{n,\ell}| : 1 \leq \ell \leq k\}, \ k \geq 1.
\end{aligned}
$$

Note that for each $n \geq 1$, $\{\Delta_{n,k}\}_{k\geq 1}$ is a nondecreasing sequence, $\lim\limits_{k\to\infty} \Delta_{n,k} = \Delta_n$ and hence, for any $n \geq 1$,

$$
P(\Delta_n > \epsilon) = \lim_{k\to\infty} P(\Delta_{n,k} > \epsilon). \tag{3.5}
$$

Levy's inequality (Theorem 8.3.2) will now be used to bound $P(\Delta_{n,k} > \epsilon)$ uniformly in k. Since $S_n \longrightarrow^p S$, for any $\eta > 0$, there exists an $n_0 \geq 1$ such that for all $n \geq n_0$,

$$
P(|S_n - S| > \eta) < \eta.
$$

This implies that for all $k \geq \ell \geq n_0$,

$$P(|S_k - S_\ell| > 2\eta) < 2\eta. \tag{3.6}$$

If $0 < \eta < \frac{1}{4}$, then the medians of $S_k - S_\ell$ for $k \geq \ell \geq n_0$ are bounded by 2η. Hence, for $n \geq n_0$ and $k \geq 1$, applying Levy's inequality (i.e., the above theorem) to $\{X_i : n+1 \leq i \leq n+k\}$,

$$\begin{aligned} P(\Delta_{n,k} > \epsilon) &= P\Big(\max_{1\leq j\leq k} |S_{n,j}| > \epsilon\Big) \\ &\leq P\Big(\max_{1\leq j\leq k} |S_{n,j} - c_{n+j,n+k}| \geq \epsilon - 2\eta\Big) \\ &\leq 2P(|S_{n,k}| \geq \epsilon - 2\eta). \end{aligned}$$

Now, choosing $0 < \eta < \frac{\epsilon}{4}$, (3.6) yields $P(\Delta_{n,k} > \epsilon) < 4\eta < \epsilon$ for all $n \geq n_0$, $k \geq 1$. Then, by (3.5), $P(\Delta_n > \epsilon) \leq \epsilon$ for all $n \geq n_0$. Hence, (3.4) holds. □

The following result on convergence of infinite series of independent random variables is an immediate consequence of the above theorem.

Theorem 8.3.4: (*Khinchine-Kolmogorov's 1-series theorem*). *Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables on a probability space $(\Omega, \mathcal{F}, P)$ such that $EX_n = 0$ for all $n \geq 1$ and $\sum_{n=1}^\infty EX_n^2 < \infty$. Then $S_n \equiv \sum_{j=1}^n X_j$ converges in mean square and almost surely, as $n \to \infty$.*

Proof: For any n, $k \in \mathbb{N}$,

$$E(S_n - S_{n+k})^2 = \text{Var}(S_n - S_{n+k}) = \sum_{j=n+1}^{n+k} \text{Var}(X_j) = \sum_{j=k+1}^{n+k} EX_j^2,$$

by independence. Since $\sum_{n=1}^\infty EX_n^2 < \infty$, $\{S_n\}_{n\geq 1}$ is a Cauchy sequence in $L^2(\Omega, \mathcal{F}, P)$ and hence converges in mean square to some S in $L^2(\Omega, \mathcal{F}, P)$. This implies that $S_n \longrightarrow^p S$, and by the above theorem $S_n \to S$ w.p. 1. □

Remark 8.3.2: It is possible to give another proof of the above theorem using Kolmogorov's inequality. See Problem 8.5.

Theorem 8.3.5: (*Kolmogorov's 3-series theorem*). *Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables on a probability space $(\Omega, \mathcal{F}, P)$ and let $S_n = \sum_{i=1}^n X_i$, $n \geq 1$. Then the sequence $\{S_n\}_{n\geq 1}$ converges w.p. 1 iff the following 3-series converge for some $0 < c < \infty$:*

(i) $\sum_{i=1}^\infty P(|X_i| > c) < \infty$,

(ii) $\sum_{i=1}^\infty E(Y_i)$ *converges,*

(iii) $\sum_{i=1}^\infty \text{Var}(Y_i) < \infty$,

where $Y_i = X_i I(|X_i| \le c)$, $i \ge 1$.

Proof: (*Sufficiency*). By (i) and the Borel-Cantelli lemma, $P(X_i \neq Y_i \text{ i.o.}) = P(|X_i| > c \text{ i.o.}) = 0$. Hence $\{S_n\}_{n\ge1}$ converges w.p. 1 iff $\{T_n\}_{n\ge1}$ converges w.p. 1, where $T_n = \sum_{i=1}^n Y_i$, $n \ge 1$. By (iii) and the 1-series theorem, the sequence $\{\sum_{i=1}^n (Y_i - EY_i)\}_{n\ge1}$ converges w.p. 1. Hence, by (ii), $\{T_n\}_{n\ge1}$ converges w.p. 1 and hence $\{S_n\}_{n\ge1}$ converges w.p. 1.

(*Necessity*). Suppose $\{S_n\}_{n\ge1}$ converges w.p. 1. Fix $0 < c < \infty$ and let $Y_i = X_i I(|X_i| \le c)$, $i \ge 1$. Since $\{S_n\}_{n\ge1}$ converges w.p. 1, $X_n \to 0$ w.p. 1. Hence, w.p. 1, $|X_i| \le c$ for all but a finite number of i's. If $A_i \equiv \{X_i \neq Y_i\} = \{|X_i| > c\}$, then by the second Borel-Cantelli lemma,

$$\sum_{i=1}^{\infty} P(A_i) < \infty, \text{ establishing (i).}$$

To establish (ii) and (iii), the following construction and the second inequality of Kolmogorov will be used. Without loss of generality, assume that there is another sequence $\{\tilde{X}_n\}_{n\ge1}$ of random variables on the same probability space $(\Omega, \mathcal{F}, P)$ such that (a) $\{\tilde{X}_n\}_{n\ge1}$ are independent, (b) $\{\tilde{X}_n\}_{n\ge1}$ is independent of $\{X_n\}_{n\ge1}$, and (c) for each $n \ge 1$, $X_n =^d \tilde{X}_n$, i.e., X_n and $\tilde{X}_n$ have the same distribution. Let

$$\begin{aligned}
\tilde{Y}_i &= \tilde{X}_i I(|\tilde{X}_i| \le c), \\
Z_i &= Y_i - \tilde{Y}_i, \ i \ge 1, \\
T_n &\equiv \sum_{i=1}^n Y_i, \\
\tilde{T}_n &\equiv \sum_{i=1}^n \tilde{Y}_i,
\end{aligned}$$

and

$$R_n \equiv \sum_{i=1}^n Z_i, \ n \ge 1.$$

Since $\{S_n \equiv \sum_{i=1}^n X_i\}_{n\ge1}$ converges w.p. 1, and $X_i = Y_i$ for all but a finite number of i, $\{T_n\}_{n\ge1}$ converges w.p. 1. Since $\{Y_i\}_{n\ge1}$ and $\{\tilde{Y}_i\}_{n\ge1}$ have the same distribution on $\mathbb{R}^\infty$, $\{\tilde{T}_n\}_{n\ge1}$ converges w.p. 1. Thus the difference sequence $\{R_n\}_{n\ge1}$ converges w.p. 1.

Next, note that $\{Z_n\}_{n\ge1}$ are independent random variables with mean 0 and $\{Z_n\}_{n\ge1}$ are uniformly bounded by $2c$. Applying Kolmogorov's second inequality (Theorem 8.3.1 (b)) to $\{Z_j : m < j \le m+n\}$ yields

$$P\left(\max_{m<j\le m+n} |R_j - R_m| \le \epsilon\right) \le \frac{(2c+4\epsilon)^2}{\sum_{i=m+1}^{m+n} \operatorname{Var}(Z_i)} \tag{3.7}$$

for all $m \ge 1$, $n \ge 1$, $0 < \epsilon < \infty$.

Let $\Delta_m \equiv \max_{m<j} |R_j - R_m|$. Let $n \to \infty$ in (3.7) to conclude that

$$P(\Delta_m \leq \epsilon) \leq \frac{(2c+4\epsilon)^2}{\sum_{i=m+1}^{\infty} \mathrm{Var}(Z_i)}.$$

Now suppose (iii) does not hold. Then, since Y_i and $\tilde{Y}_i$ are iid, $\mathrm{Var}(Z_i) = 2\mathrm{Var}(Y_i)$ for all $i \geq 1$, and thus $\sum_{i=m+1}^{\infty} \mathrm{Var}(Z_i) = \infty$ and hence $P(\Delta_m \leq \epsilon) = 0$ for each $m \geq 1$, $0 < \epsilon < \infty$. This implies that $P(\Delta_m > \epsilon) = 1$ for each $\epsilon > 0$ and hence that $\Delta_m = \infty$ w.p. 1 for all $m \geq 1$. This contradicts the convergence w.p. 1 of the sequence $\{R_n\}_{n\geq 1}$. Hence (iii) holds.

By the 1-series theorem, $\{\sum_{i=1}^{n}(Y_i - EY_i)\}_{n\geq 1}$ converges w.p. 1. Since $\{\sum_{i=1}^{n} Y_i\}_{n\geq 1}$ converges w.p. 1, $\sum_{i=1}^{\infty} EY_i$ converges, establishing (ii). This completes the proof of necessity part and of the theorem. □

Remark 8.3.3: To go from the convergence w.p. 1 of $\{R_n\}_{n\geq 1}$ to (iii), it suffices to show that if (iii) fails, then for each $0 < A < \infty$, $P(|R_n| \leq A) \to 0$ as $n \to \infty$. This can be established without the use of (3.7) but using the central limit theorem (to be proved later in Chapter 11), which shows that if $\mathrm{Var}(R_n) \to \infty$, then

$$P\left(\frac{R_n}{\sqrt{\mathrm{Var}(R_n)}} \leq x\right) \to \Phi(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt,$$

for all x in $\mathbb{R}$. (Also see Billingsley (1995), p. 290.)

8.4 Kolmogorov and Marcinkiewz-Zygmund SLLNs

For a sequence of independent and identically distributed random variables $\{X_n\}_{n\geq 1}$, Kolmogorov showed that $\{X_n\}_{n\geq 1}$ obeys the SLLN with $b_n = n$ iff $E|X_1| < \infty$. Marcinkiewz and Zygmund generalized this result and proved a class of SLLNs for $\{X_n\}_{n\geq 1}$ when $E|X|^p < \infty$ for some $p \in (0, 2)$. The proof uses Kolmogorov's 3-series theorem and some results from real analysis. This approach is to be contrasted with Etemadi's proof of the SLLN, which uses a decomposition of the random variables $\{X_n\}_{n\geq 1}$ into positive and negative parts and uses monotonicity of the sum to establish almost sure convergence along a subsequence by an application of the Borel-Cantelli lemma. The alternative approach presented in this section is also useful for proving SLLNs for sums of independent random variables that are not necessarily identically distributed.

The next three are preparatory results for Theorem 8.4.4.

Lemma 8.4.1: (*Abel's summation formula*). *Let $\{a_n\}_{n\geq 1}$ and $\{b_n\}_{n\geq 1}$ be two sequences of real numbers. Then, for all $n \geq 2$,*

$$\sum_{j=1}^{n} a_j b_j = A_n b_n - \sum_{j=1}^{n-1} A_j (b_{j+1} - b_j) \tag{4.1}$$

where $A_k = \sum_{j=1}^k a_j$, $k \geq 1$.

Proof: Let $A_0 = 0$. Then, $a_j = A_j - A_{j-1}$, $j \geq 1$. Hence,

$$\begin{aligned}\sum_{j=1}^n a_j b_j &= \sum_{j=1}^n (A_j - A_{j-1}) b_j = \sum_{j=1}^n A_j b_j - \sum_{j=1}^n A_{j-1} b_j \\ &= \sum_{j=1}^n A_j b_j - \sum_{j=1}^{n-1} A_j b_{j+1},\end{aligned}$$

yielding (4.1). □

Lemma 8.4.2: (*Kronecker's lemma*). *Let* $\{a_n\}_{n\geq 1}$ *and* $\{b_n\}_{n\geq 1}$ *be sequences of real numbers such that* $0 < b_n \uparrow \infty$ *as* $n \to \infty$ *and* $\sum_{j=1}^\infty a_j$ *converges. Then,*

$$\frac{1}{b_n} \sum_{j=1}^n a_j b_j \longrightarrow 0 \quad as \quad n \to \infty. \tag{4.2}$$

Proof: Let $A_k = \sum_{j=1}^k a_j$, $A \equiv \sum_{j=1}^\infty a_j = \lim_{k\to\infty} A_k$ and $R_k = A - A_k$, $k \geq 1$. Then, by Lemma 8.4.1 for $n \geq 2$,

$$\begin{aligned}\sum_{j=1}^n a_j b_j &= A_n b_n - \sum_{j=1}^{n-1} A_j (b_{j+1} - b_j) \\ &= A_n b_n - \sum_{j=1}^{n-1} (A - R_j)(b_{j+1} - b_j) \\ &= A_n b_n - A \sum_{j=1}^{n-1} (b_{j+1} - b_j) + \sum_{j=1}^{n-1} R_j (b_{j+1} - b_j) \\ &= A_n b_n - A b_n + A b_1 + \sum_{j=1}^{n-1} R_j (b_{j+1} - b_j) \\ &= -R_n b_n + A b_1 + \sum_{j=1}^{n-1} R_j (b_{j+1} - b_j). \end{aligned} \tag{4.3}$$

Since $\sum_{n=1}^\infty a_n$ converges, $R_n \to 0$ as $n \to \infty$. Hence, given any $\epsilon > 0$, there exists $N = N_\epsilon > 1$ such that $|R_n| \leq \epsilon$ for all $n \geq N$. Since $0 < b_n \uparrow \infty$, for all $n > N$,

$$\left| b_n^{-1} \sum_{j=1}^{n-1} R_j (b_{j+1} - b_j) \right|$$

$$\begin{aligned} &\leq b_n^{-1} \sum_{j=1}^{N-1} |R_j|\,|b_{j+1} - b_j| + \epsilon\, b_n^{-1} \sum_{j=N}^{n-1} (b_{j+1} - b_j) \\ &= b_n^{-1} \sum_{j=1}^{N-1} |R_j|\,|b_{j+1} - b_j| + \epsilon. \end{aligned}$$

Now letting $n \to \infty$ and then letting $\epsilon \downarrow 0$, yields

$$\limsup_{n\to\infty} \left| b_n^{-1} \sum_{j=1}^{n-1} R_j (b_{j+1} - b_j) \right| = 0.$$

Hence, (4.2) follows from (4.3). □

Lemma 8.4.3: *For any random variable X,*

$$\sum_{n=1}^{\infty} P(|X| > n) \leq E|X| \leq \sum_{n=0}^{\infty} P(|X| > n). \tag{4.4}$$

Proof: For $n \geq 1$, let $A_n = \{n - 1 < |X| \leq n\}$. Define the random variables

$$Y = \sum_{n=1}^{\infty} (n-1)\; I_{A_n} \quad \text{and} \quad Z = \sum_{n=1}^{\infty} n\; I_{A_n}.$$

Then, it is clear that $Y \leq |X| \leq Z$, so that

$$EY \leq E|X| \leq EZ. \tag{4.5}$$

Note that

$$\begin{aligned} EY &= \sum_{n=1}^{\infty} (n-1) P(A_n) \\ &= \sum_{n=2}^{\infty} \sum_{j=1}^{n-1} P(A_n) \\ &= \sum_{j=1}^{\infty} \sum_{n=j+1}^{\infty} P(n - 1 < |X| \leq n) \\ &= \sum_{j=1}^{\infty} P(|X| > j). \end{aligned}$$

Similarly, one can show that $EZ = \sum_{j=0}^{\infty} P(|X| > j)$. Hence, (4.4) follows. □

Theorem 8.4.4: (*Marcinkiewz-Zygmund SLLNs*). *Let $\{X_n\}_{n\geq 1}$ be a sequence of identically distributed random variables and let $p \in (0, 2)$. Write $S_n = \sum_{i=1}^{n} X_i$, $n \geq 1$.*

(i) If $\{X_n\}_{n\geq 1}$ *are pairwise independent and*

$$\frac{S_n - nc}{n^{1/p}} \quad \text{converges w.p. 1} \tag{4.6}$$

for some $c \in \mathbb{R}$*, then* $E|X_1|^p < \infty$.

(ii) Conversely, if $E|X_1|^p < \infty$ *and* $\{X_n\}_{n\geq 1}$ *are independent, then (4.6) holds with* $c = EX_1$ *for* $p \in [1, 2)$ *and with any* $c \in \mathbb{R}$ *for* $p \in (0, 1)$.

Corollary 8.4.5: (*Kolmogorov's SLLN*). *Let* $\{X_n\}_{n\geq 1}$ *be a sequence of iid random variables. Then,*

$$\frac{S_n - nc}{n} \to 0 \quad w.p.\ 1$$

for some $c \in \mathbb{R}$ *iff* $E|X_1| < \infty$*, in which case,* $c = EX_1$.

Thus, Kolmogorov's SLLN corresponds with the special case $p = 1$ of Theorem 8.4.4. Note that compared with the WLLN and Borel's SLLN of Sections 8.1 and 8.2, Kolmogorov's SLLN presents a significant improvement in the moment condition, i.e., it assumes the finiteness of only the first absolute moment. Further, both the Kolmogorov's SLLN and the Marcinkiewz-Zygmund SLLN are proved under minimal moment conditions, since the corresponding moment conditions are shown to be necessary.

Proof of Theorem 8.4.4: (i) Suppose that (4.6) holds for some $c \in \mathbb{R}$. Then,

$$\begin{aligned} \frac{X_n}{n^{1/p}} &= \frac{S_n - S_{n-1}}{n^{1/p}} \\ &= \frac{S_n - nc}{n^{1/p}} - \frac{S_{n-1} - (n-1)c}{n^{1/p}} + \frac{c}{n^{1/p}} \\ &\to 0 \quad \text{as} \quad n \to \infty, \quad \text{a.s.} \end{aligned}$$

Hence, $P(|X_n/n^{1/p}| > 1 \ \text{ i.o.}) = 0$. By the second Borel-Cantelli lemma and by the pairwise independence of $\{X_n\}_{n\geq 1}$, this implies

$$\sum_{n=1}^{\infty} P\bigg(\frac{|X_n|}{n^{1/p}} > 1\bigg) < \infty,$$

i.e.,

$$\sum_{n=1}^{\infty} P\big(|X_1|^p > n\big) < \infty.$$

Hence, by Lemma 8.4.3, $E|X_1|^p < \infty$.

To prove (ii), suppose that $E|X_1|^p < \infty$ for some $p \in (0,2)$. For $1 \le p < 2$, w.l.o.g. assume that $EX_1 = 0$. Next, define the variables $Z_n = X_n I(|X_n|^p \le n)$, $n \ge 1$. Then, by Lemma 8.4.3,

$$\begin{aligned}&\sum_{n=1}^{\infty} P(X_n \neq Z_n)\\ = \quad & \sum_{n=1}^{\infty} P(|X_n|^p > n) = \sum_{n=1}^{\infty} P(|X_1|^p > n) \le E|X_1|^p < \infty.\end{aligned}$$

Hence, by the Borel-Cantelli lemma,

$$P(X_n \neq Z_n \quad \text{i.o.}) = 0. \tag{4.7}$$

Note that, in view of (4.7), (4.6) holds with $c = 0$ if and only if

$$n^{1/p} \sum_{i=1}^{n} Z_i \to 0 \quad \text{as} \quad n \to \infty, \quad \text{w.p. 1.} \tag{4.8}$$

Note that for any $j \in \mathbb{N}$, $\theta > 1$ and $\beta \in (-\infty, 0)\backslash\{-1\}$,

$$\begin{aligned}\sum_{n=j}^{\infty} n^{-\theta} \quad &\le \quad j^{-\theta} + \sum_{n=j+1}^{\infty} \int_{n-1}^{n} x^{-\theta} dx\\ &= \quad j^{-\theta} + \frac{1}{\theta - 1} \cdot j^{-(\theta-1)}\\ &\le \quad \frac{\theta}{\theta-1} \cdot j^{-(\theta-1)} \end{aligned} \tag{4.9}$$

and similarly,

$$\begin{aligned}\sum_{n=1}^{j} n^{\beta} \quad &\le \quad \big[\beta + j^{(\beta+1)}\big]/(\beta+1)\\ &\le \quad \frac{\beta}{\beta+1} I(\beta < -1) + \frac{j^{\beta+1}}{\beta+1} I(-1 < \beta < 0). \end{aligned} \tag{4.10}$$

Now,

$$\begin{aligned}&\sum_{n=1}^{\infty} \text{Var}(Z_n/n^{1/p})\\ \le \quad & \sum_{n=1}^{\infty} EX_1^2 I(|X_1|^p \le n) \cdot n^{-2/p}\\ = \quad & \sum_{n=1}^{\infty} \sum_{j=1}^{n} EX_1^2 I(j-1 < |X_1|^p \le j) \cdot n^{-2/p}\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{\infty}\left(\sum_{n=j}^{\infty} n^{-2/p}\right) \cdot EX_1^2 I(j-1 < |X_1|^p \le j) \\
&\le \frac{2}{2-p}\sum_{j=1}^{\infty} j^{-(\frac{2}{p}-1)} \cdot EX_1^2 I\big((j-1) < |X_1|^p \le j\big) \quad \text{(by (4.9))} \\
&\le \frac{2}{2-p}\sum_{j=1}^{\infty} j^{-(\frac{2}{p}-1)} \cdot E|X_1|^p I(j-1 < |X_1|^p \le j) \cdot (j^{1/p})^{2-p} \\
&= \frac{2}{2-p} E|X_1|^p < \infty.
\end{aligned}
$$

Hence, by Theorem 8.3.4, $\sum_{n=1}^{\infty}(Z_n - EZ_n)/n^{1/p}$ converges w.p. 1. By Kronecker's lemma (viz. Lemma 8.4.2),

$$
n^{-1/p}\sum_{j=1}^{n}(Z_j - EZ_j) \to 0 \quad \text{as} \quad n \to \infty, \quad \text{w.p. 1.} \tag{4.11}
$$

Now consider the case $p = 1$. In this case, $E|X_1| < \infty$ and by the DCT, $EZ_n = EX_1 I(|X_1| \le n) \to EX_1 = 0$ as $n \to \infty$. Hence, $n^{-1}\sum_{i=1}^{n} EZ_i \to 0$. Part (ii) of the theorem now follows from (4.8) and (4.11) for $p = 1$.

Next consider the case $p \in (0, 2)$, $p \neq 1$. Using (4.9) and (4.10), one can show (cf. Problem 8.12) that

$$
n^{-1/p}\sum_{j=1}^{n} EZ_j \to 0 \quad \text{as} \quad n \to \infty. \tag{4.12}
$$

Hence, by (4.8), (4.11), and (4.12), one gets (4.6) with $c = 0$ for $p \in (0, 2)\backslash\{1\}$. Finally, note that for $p \in (0, 1)$, and for any $c \in \mathbb{R}$,

$$
\begin{aligned}
\frac{S_n - nc}{n^{1/p}} &= \frac{S_n}{n^{1/p}} - \frac{nc}{n^{1/p}} \\
&\to 0 \quad \text{as} \quad n \to \infty, \quad \text{a.s.},
\end{aligned}
$$

whenever $S_n/n^{1/p} \to 0$ as $n \to \infty$, w.p. 1. Hence, (4.6) holds with an arbitrary $c \in \mathbb{R}$ for $p \in (0, 1)$. This completes the proof of part (ii) for $p \in (0, 2)\backslash\{1\}$ and hence of the theorem. □

The next result gives a SLLN for independent random variables that are not necessarily identically distributed.

Theorem 8.4.6: *Let $\{X_n\}_{n\ge1}$ be a sequence of independent random variables. If $\sum_{n=1}^{\infty} E|X_n|^{\alpha_n}/n^{\alpha_n} < \infty$ for some $\alpha_n \in [1, 2]$, $n \ge 1$, then*

$$
n^{-1}\sum_{j=1}^{n}(X_j - EX_j) \to 0 \quad \textit{as} \quad n \to \infty, \quad \textit{w.p. 1.} \tag{4.13}
$$

Proof: W.l.o.g. suppose that $EX_n = 0$ for all $n \geq 1$. Let $Y_n = X_n I(|X_n| \leq n)/n$. Note that $|EY_n| = |n^{-1}(EX_n - EX_n I(|X_n| > n))| = n^{-1}|EX_n I(|X_n| > n)|$, $n \geq 1$. Since $1 \leq \alpha_n \leq 2$,

$$\begin{aligned}
&\sum_{n=1}^{\infty}\{P(|X_n| > n) + |EY_n|\} \\
&\quad \leq \; 2\sum_{n=1}^{\infty} n^{-1}E|X_n|I(|X_n| > n) \\
&\quad \leq \; 2\sum_{n=1}^{\infty} E|X_n|^{\alpha_n}/n^{\alpha_n} < \infty
\end{aligned}$$

and

$$\begin{aligned}
\sum_{n=1}^{\infty}\mathrm{Var}(Y_n) &\leq \sum_{n=1}^{\infty} n^{-2}EX_n^2 I(|X_n| \leq n) \\
&\leq \sum_{n=1}^{\infty} n^{-\alpha_n}EX_n^{\alpha_n} < \infty.
\end{aligned}$$

Hence, by Kolmogorov's 3-series theorem, $\sum_{n=1}^{\infty}(X_n/n)$ converges w.p. 1. Now the theorem follows from Lemma 8.4.2. □

Corollary 8.4.7: *Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables such that for some $\alpha \in [1, 2]$, $\sum_{n=1}^{\infty}(n^{-\alpha}E|X_n|^{\alpha}) < \infty$. Then (4.13) holds.*

8.5 Renewal theory

8.5.1 Definitions and basic properties

Let $\{X_n\}_{n\geq 0}$ be a sequence of nonnegative random variables that are independent and, for $i \geq 1$, identically distributed with cdf F. Let $S_n = \sum_{i=0}^{n} X_i$ for $n \geq 0$. Imagine a system where a component in operation at time $t = 0$ lasts X_0 units of time and then is replaced by a new one that lasts X_1 units of time, which, at failure, is replaced by yet another new one that lasts X_2 units of time and so on. The sequence $\{S_n\}_{n\geq 0}$ represents the sequence of epochs when 'renewal' takes place and is called a *renewal sequence*. Assume that $P(X_1 = 0) < 1$. Then, since $P(X_1 < \infty) = 1$, it follows that for each n, $P(S_n < \infty) = 1$ and $\lim_{n\to\infty} S_n = \infty$ w.p. 1 (Problem 8.16). Now define the counting process $\{N(t) : t \geq 0\}$ by the relation

$$N(t) = k \quad \text{if} \quad S_{k-1} \leq t < S_k \quad \text{for} \quad k = 0, 1, 2, \ldots \tag{5.1}$$

where $S_{-1} = 0$. Thus $N(t)$ counts the number of renewals up to time t.

Definition 8.5.1: The stochastic process $\{N(t) : t \geq 0\}$ is called a *renewal process* with lifetime distribution F. The renewal sequence $\{S_n\}_{n\geq 0}$ and the renewal process $\{N(t) : t \geq 0\}$ are called *nondelayed* or *standard* if X_0 has the same distribution as X_1 and are called *delayed* otherwise.

Since $P(X_1 \geq 0) = 1$, $\{S_n\}_{n\geq 0}$ is nondecreasing in n and for each $t \geq 0$, the event $\{N(t) = k\} = \{S_{k-1} \leq t < S_k\}$ belongs to the σ-algebra $\sigma\langle\{X_j : 0 \leq j \leq k\}\rangle$ and hence $N(t)$ is a random variable. Using the nontriviality hypothesis that $P(X_1 = 0) < 1$, it is shown below that for each $t > 0$, the random variable $N(t)$ has finite moments of all order.

Proposition 8.5.1: *Let $P(X_1 = 0) < 1$. Then there exists $0 < \lambda < 1$ (not depending on t) and a constant $C(t) \in (0, \infty)$ such that*

$$P(N(t) > k) \leq C(t)\lambda^k \quad \textit{for all} \quad k > 0. \tag{5.2}$$

Proof: For $t > 0$, $k \in \mathbb{N}$,

$$\begin{aligned} P(N(t) > k) &= P(S_k \leq t) \\ &= P\big(e^{-\theta S_k} \geq e^{-\theta t}\big) \quad \text{for} \quad \theta > 0 \\ &\leq e^{\theta t} E\big(e^{-\theta S_k}\big) \quad \text{(by Markov's inequality)} \\ &= e^{\theta t} E\big(e^{-\theta X_0}\big)\Big(E\big(e^{-\theta X_1}\big)\Big)^k. \end{aligned}$$

By BCT, $\lim_{\theta\uparrow\infty} E(e^{-\theta X_1}) = P(X_1 = 0) < 1$. Hence, there exists a θ large such that $\lambda \equiv E(e^{-\theta X_1})$ is less than one, thus, completing the proof. □

Corollary 8.5.2: *There exists an $s_0 > 0$ such that the moment generating function (m.g.f.) $E(e^{sN(t)}) < \infty$ for all $s < s_0$ and $t \geq 0$.*

Proof: From (5.2), for any $t > 0$, it follows that $P\big(N(t) = k\big) = O(\lambda^k)$ as $k \to \infty$ for some $0 < \lambda < 1$ and hence $E\big(e^{sN(t)}\big) = \sum_{k=0}^{\infty}(e^s)^k P\big(N(t) = k\big) < \infty$ for any s such that $e^s\lambda < 1$, i.e., for all $s < s_0 \equiv -\log\lambda$. □

From (5.1), it follows that for $t > 0$,

$$\begin{aligned} & S_{N(t)-1} \leq t < S_{N(t)} \\ \Rightarrow \quad & \Big(\frac{N(t)-1}{N(t)}\Big)\frac{S_{N(t)-1}}{(N(t)-1)} \leq \frac{t}{N(t)} \leq \Big(\frac{S_{N(t)}}{N(t)}\Big). \end{aligned} \tag{5.3}$$

Let A be the event that $\frac{S_n}{n} \to EX_1$ as $n \to \infty$ and let B be the event that $N(t) \to \infty$ as $t \to \infty$. Since $S_n \to \infty$ w.p. 1, it follows that $P(B) = 1$. Also, by the SLLN, $P(A) = 1$. On the event $C = A \cap B$, it holds that

$$\frac{S_{N(t)}}{N(t)} \to EX_1 \quad \text{as} \quad t \to \infty.$$

This together with (5.3) yields the following result.

Proposition 8.5.3: *Suppose that $P(X_1 = 0) < 1$. Then,*

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{EX_1} \quad \textit{w.p. 1.} \tag{5.4}$$

Definition 8.5.2: The function $U(t) \equiv EN(t)$ for the nondelayed process is called the *renewal function*.

An explicit expression for $U(\cdot)$ is given by (5.13) below.

Next consider the convergence of $EN(t)/t$. By (5.4) and Fatou's lemma, one gets

$$\liminf_{t\to\infty} \frac{EN(t)}{t} \geq \frac{1}{EX_1}. \tag{5.5}$$

It turns out that the $\liminf_{t\to\infty}$ in (5.5) can be replaced by $\lim_{t\to\infty}$ and $\geq$ by equality. To do this it suffices to show that the family $\{\frac{N(t)}{t} : t \geq k\}$ is uniformly integrable for some $k < \infty$. This can be done by showing $E(\frac{N(t)}{t})^2$ is bounded in t (see Chung (1974), Chapter 5). An alternate approach is to bound the $\limsup$. For this one can use an identity known as Wald's equation (see also Chapter 13).

8.5.2 Wald's equation

Let $\{X_j\}_{j\geq 1}$ be independent random variables with $EX_j = 0$ for all $j \geq 1$. Also, let $S_0 = 0$, $S_n = \sum_{j=1}^n X_j$, $n \geq 1$.

Definition 8.5.3: A positive integer valued random variable N is called a *stopping time* with respect to $\{X_j\}_{j\geq 1}$ if for every $j \geq 1$, the event $\{N = j\} \in \sigma\langle\{X_1, \ldots, X_j\}\rangle$. A stopping time N is called *bounded* if there exists a $K < \infty$ such that $P(N \leq K) = 1$.

Example 8.5.1: $N \equiv \min\{n : \sum_{j=1}^n X_j \geq 25\}$ is a stopping time w.r.t. $\{X_j\}_{j\geq 1}$, but $M \equiv \max\{n : \sum_{j=1}^n X_j \geq 25\}$ is not.

Proposition 8.5.4: *Let $\{X_j\}_{j\geq 1}$ be independent random variables with $EX_j = 0$. Let N be a bounded stopping time w.r.t. $\{X_j\}_{j\geq 1}$. Then*

$$E(|S_N|) < \infty \quad \textit{and} \quad ES_N = 0.$$

Proof: Let $K \in \mathbb{N}$ be such that $P(N \leq K) = 1$. Then $|S_N| \leq \sum_{j=1}^K |X_i|$ and hence $E|S_N| < \infty$. Next, $S_N = \sum_{j=1}^K X_j I(N \geq j)$ and hence

$$ES_N = \sum_{j=1}^K E\big(X_j I(N \geq j)\big).$$

But the event $\{N \geq j\} = \{N \leq j-1\}^c \in \sigma\langle\{X_1, X_2, \ldots, X_{j-1}\}\rangle$. Since X_j is independent of $\sigma\langle X_1, X_2, \ldots, X_{j-1}\rangle$,

$$E\big(X_j I(N \geq j)\big) = 0 \quad \text{for} \quad 1 \leq j \leq K.$$

Thus $ES_N = 0$. □

Corollary 8.5.5: *Let $\{X_j\}_{j\geq 1}$ be iid random variables with $E|X_1| < \infty$. Let N be a bounded stopping time w.r.t. $\{X_j\}_{j\geq 1}$. Then*

$$ES_N = (EN)EX_1.$$

Corollary 8.5.6: *Let $\{X_j\}_{j\geq 1}$ be iid nonnegative random variable with $E|X_1| < \infty$. Let N be a stopping time w.r.t. $\{X_j\}_{j\geq 1}$. Then*

$$ES_N = (EN)EX_1.$$

Proof: Let $N_k = N \wedge k$, $k = 1, 2, \ldots$. Then N_k is a bounded stopping time. By Corollary 8.5.5,

$$E(S_{N_k}) = (EN_k)EX_1.$$

Let $k \uparrow \infty$. Then $0 \leq S_{N_k} \uparrow S_N$ and $N_k \uparrow N$. By the MCT, $ES_{N_k} \uparrow ES_N$ and $EN_k \uparrow EN$. □

Theorem 8.5.7: (*Wald's equation*). *Let $\{X_j\}_{j\geq 1}$ be iid random variables with $E|X_1| < \infty$. Let N be a stopping time w.r.t. $\{X_j\}_{j\geq 1}$ such that $EN < \infty$. Then*

$$ES_N = (EN)EX_1.$$

Proof: Let $T_n = \sum_{j=1}^{n} |X_j|$, $n \geq 1$. Let $N_k = N \wedge k$, $k = 1, 2, \ldots$. Then by Corollary 8.5.5,

$$E(S_{N_k}) = (EN_k)EX_1.$$

Also, $|S_{N_k}| \leq T_{N_k}$ and

$$ET_{N_k} = (EN_k)E|X_1|.$$

Further, as $k \to \infty$, $N_k \to N$, $S_{N_k} \to S_N$, $T_{N_k} \to T_N$, and

$$ET_{N_k} \to ET_N = (EN)E|X_1| < \infty.$$

So, by the extended *DCT* (Theorem 2.3.11)

$$\begin{aligned} &ES_{N_k} \to ES_N \\ \text{i.e.,} \quad &(EN_k)EX_1 \to ES_N \\ \text{i.e.,} \quad &ES_N = (EN)EX_1. \end{aligned}$$

□

8.5.3 The renewal theorems

In this section, two versions of the renewal theorem will be proved. For this, the notation and concepts introduced in Sections 8.5.1 and 8.5.2 will be used without further explanation. Note that for each $t > 0$ and $j = 0, 1, 2, \ldots$, the event $\{N(t) = j\} = \{S_{j-1} \leq t < S_j\}$ belongs to $\sigma\langle\{X_0, \ldots, X_j\}\rangle$. Thus, by Wald's equation (Theorem 8.5.7 above)

$$E(S_{N(t)}) = \big(EN(t)\big)EX_1 + EX_0.$$

Let $m \in (0, \infty)$ and $\tilde{X}_i = \min\{X_i, m\}$, $i \geq 0$. Let $\{\tilde{S}_n\}_{n\geq 0}$ and $\{\tilde{N}(t)\}_{t\geq 0}$ be the associated renewal sequence and renewal process, respectively. Again, by Wald's equation,

$$E\big(\tilde{S}_{\tilde{N}(t)}\big) = \big(E\tilde{N}(t)\big)E\tilde{X}_1 + E\tilde{X}_0.$$

But since $\tilde{S}_{\tilde{N}(t)-1} \leq t < \tilde{S}_{\tilde{N}(t)}$, it follows that $\tilde{S}_{\tilde{N}(t)} \leq t + m$ and hence

$$(E\tilde{N}(t))E\tilde{X}_1 + E\tilde{X}_0 \leq t + m.$$

This yields

$$\limsup_{t\to\infty} \frac{E\tilde{N}(t)}{t} \leq \frac{1}{E\tilde{X}_1}.$$

Clearly, for all $t > 0$, $\tilde{N}(t) \geq N(t)$ and hence

$$\limsup_{t\to\infty} E\frac{N(t)}{t} \leq \frac{1}{E\tilde{X}_1}. \tag{5.6}$$

Since this is true for each $m \in (0, \infty)$ and by the MCT, $E\tilde{X}_1 \to EX_1$ as $m \to \infty$, it follows that

$$\limsup_{t\to\infty} \frac{EN(t)}{t} \leq \frac{1}{EX_1}.$$

Combining this with (5.5) leads to the following result.

Theorem 8.5.8: (*The weak renewal theorem*). *Let $\{N(t) : t \geq 0\}$ be a renewal process with distribution F. Let $\mu = \int_{[0,\infty)} x dF(x) \in (0, \infty)$. Then,*

$$\lim_{t\to\infty} \frac{EN(t)}{t} = \frac{1}{\mu}. \tag{5.7}$$

The above result is also valid when $\mu = \infty$ when $\frac{1}{\mu}$ is interpreted as zero.

Definition 8.5.4: A random variable X (and its probability distribution) is called *arithmetic* (or *lattice*) if there exists $a \in \mathbb{R}$ and $d > 0$ such that $\frac{X-a}{d}$

is integer valued. The largest such d is called the *span* of (the distribution of) X.

Definition 8.5.5: A random variable X (and its distribution distribution) is called *nonarithmetic* (or *nonlattice*) if it is not arithmetic.

The weak renewal theorem (Theorem 8.5.8) implies that $EN(t) = t/\mu + o(t)$ as $t \to \infty$. This suggests that $E\big(N(t+h) - N(t)\big) = (t+h)/\mu - t/\mu + o(t) = h/\mu + o(t)$. A strengthening of the above result is as follows.

Theorem 8.5.9: (*The strong renewal theorem*). *Let* $\{N(t) : t \geq 0\}$ *be a renewal process with a nonarithmetic distribution* F *with a finite positive mean* μ. *Then, for each* $h > 0$,

$$\lim_{t\to\infty} E\big(N(t+h) - N(t)\big) = \frac{h}{\mu}. \tag{5.8}$$

Remark 8.5.1: Since

$$N(t) = \sum_{j=0}^{k-1} \big(N(t-j) - N(t-j-1)\big) + N(t-k)$$

where $k \leq t < k+1$, the weak renewal theorem follows from the strong renewal theorem.

The following are the "arithmetic versions" of Theorems 8.5.8 and 8.5.9. Let $\{X_i\}_{i\geq 0}$ be independent positive integer valued random variables such that $\{X_i\}_{i\geq 1}$ are iid with distribution $\{p_j\}_{j\geq 1}$. Let $S_n = \sum_{j=0}^{n} X_j$, $n \geq 0$, $S_{-1} = 0$. Let $N_n = k$ if $S_{k-1} \leq n < S_k$, $k = 0, 1, 2, \ldots$. Let

$$\begin{aligned} u_n &= P(\text{there is a renewal at time } n) \\ &= P(S_k = n \text{ for some } k \geq 0). \end{aligned}$$

Theorem 8.5.10: *Let* $\mu = \sum_{j=1}^{\infty} jp_j \in (0, \infty)$. *Then*

$$\frac{1}{n}\sum_{j=0}^{n} u_j \to \frac{1}{\mu} \quad \textit{as} \quad n \to \infty. \tag{5.9}$$

Theorem 8.5.11: *Let* $\mu = \sum_{j=1}^{\infty} jp_j \in (0, \infty)$ *and g.c.d.* $\{k : p_k > 0\} = 1$. *Then*

$$u_n \to \frac{1}{\mu} \quad \textit{as} \quad n \to \infty. \tag{5.10}$$

For proofs of Theorems 8.5.9 and 8.5.11, see Feller (1966) for an analytic proof or Lindvall (1992) for a proof using the coupling method. The proof of Theorem 8.5.10 is similar to that of Theorem 8.5.8.

8.5.4 Renewal equations

The above strong renewal theorems have many applications. These are via what are known as *renewal equations*.

Let $F(\cdot)$ be a cdf such that $F(0) = 0$. Let $\boldsymbol{B}_0 \equiv \{f \mid f : [0, \infty) \to \mathbb{R},\ f$ is Borel measurable and bounded on bounded intervals$\}$.

Definition 8.5.6: A function $a(\cdot)$ is said to satisfy the *renewal equation* with *distribution* $F(\cdot)$ and *forcing function* $b(\cdot) \in \boldsymbol{B}_0$ if $a \in \boldsymbol{B}_0$ and

$$a(t) = b(t) + \int_{(0,t]} a(t-u)dF(u) \quad \text{for} \quad t \geq 0. \tag{5.11}$$

Theorem 8.5.12: *Let F be a cdf such that $F(0) = 0$ and let $b(\cdot) \in \boldsymbol{B}_0$. Then there is a unique solution $a_0(\cdot) \in \boldsymbol{B}_0$ to (5.11) given by*

$$a_0(t) = \int_{[0,t]} b(t-u)U(du) \tag{5.12}$$

where $U(\cdot)$ is the Lebesgue-Stieltjes measure induced by the nondecreasing function

$$U(t) \equiv \sum_{n=0}^{\infty} F^{(n)}(t), \tag{5.13}$$

with $F^{(n)}(\cdot)$, $n \geq 0$ being defined by the relations

$$\begin{aligned} F^{(n)}(t) &= \int_{(0,t]} F^{(n-1)}(t-u)dF(u),\ t \in \mathbb{R},\ n \geq 1, \\ F^{(0)}(t) &= \begin{cases} 1 & \text{if} \quad t \geq 0 \\ 0 & \quad\quad t < 0. \end{cases} \end{aligned}$$

It will be shown below that the function $U(\cdot)$ defined in (5.13) is the *renewal function* $EN(t)$ as in Definition 8.5.2.

Proof: For any function $b \in \boldsymbol{B}_0$ and any nondecreasing right continuous function $G : [0, \infty) \to \mathbb{R}$, let

$$(b * G)(t) \equiv \int_{[0,t]} b(t-u)dG(u).$$

Then since $F(0) = 0$, the equation (5.11) can be rewritten as

$$a = b + a * F. \tag{5.14}$$

Let $\{X_i\}_{i\geq 1}$ be iid random variables with cdf F. Then it is easy to verify that $F^{(n)}(t) = P(S_n \leq t)$, where $S_0 = 0$, and $S_n = \sum_{i=1}^{n} X_i$ for $n \geq 1$. Let

$\{N(t) : t \geq 0\}$ be as defined by (5.1). Then, for $t \in (0, \infty)$,

$$EN(t) = \sum_{j=1}^{\infty} P\big(N(t) \geq j\big) = \sum_{j=1}^{\infty} P(S_{j-1} \leq t) = \sum_{n=0}^{\infty} F^{(n)}(t) = U(t).$$

By Proposition 8.5.1, $U(t) < \infty$ for all $t > 0$ and is nondecreasing. Since $b \in \boldsymbol{B}_0$ for each $0 < t < \infty$, a_0 defined by (5.12) is well-defined. By definition $a_0 = b * U$ and by (5.13), a_0 satisfies (5.14) and hence (5.11). If a_1 and a_2 from $\boldsymbol{B}_0$ are two solutions to (5.14) then $\tilde{a} \equiv a_1 - a_2$ satisfies

$$\tilde{a} = \tilde{a} * F$$

and hence

$$\tilde{a} = \tilde{a} * F^{(n)} \quad \text{for all} \quad n \geq 1.$$

This implies

$$M(t) \equiv \sup\{|\tilde{a}(u)| : 0 \leq u \leq t\} \leq M(t) F^{(n)}(t).$$

But $F^{(n)}(t) \to 0$ as $n \to \infty$. Hence $|\tilde{a}| = 0$ on $(0, t]$ for each t. Thus $a_0 = b * U$ is the unique solution to (5.11). □

The discrete or arithmetic analog of the renewal equation (5.11) is as follows. Let $\{X_i\}_{i \geq 1}$ be iid positive integer valued random variables with distribution $\{p_j\}_{j \geq 1}$. Let $S_0 = 0$, and $S_n = \sum_{i=1}^n X_i$ for $n \geq 1$. Let $u_n = P(S_j = n \text{ for some } j \geq 0)$. Then, $u_0 = 1$ and u_n satisfies $u_n = \sum_{j=1}^n p_j u_{n-j}$ for $n \geq 1$. For any sequence $\{b_j\}_{j \geq 0}$, the equation

$$a_n = b_n + \sum_{j=1}^{n} a_{n-j} p_j, \; n = 0, 1, 2, \ldots \tag{5.15}$$

is called the *discrete renewal equation*. As in the general case, it can be shown (Problem 8.17 (a)) that the unique solution to (5.15) is given by

$$a_n = \sum_{j=0}^{n} b_{n-j} u_j. \tag{5.16}$$

The following convergence results are easy to establish from Theorem 8.5.11 (Problem 8.17 (b)).

Theorem 8.5.13: (*The key renewal theorem, discrete case*). *Let $\{p_j\}_{j \geq 1}$ be aperiodic, i.e., g.c.d. $\{k : p_k > 0\} = 1$ and $\mu \equiv \sum_{j=1}^{\infty} j p_j \in (0, \infty)$. Let $\{u_n\}_{n \geq 0}$ be the renewal sequence associated with $\{p_j\}_{j \geq 1}$. That is, $u_0 = 1$ and $u_n = \sum_{j=1}^n p_j u_{n-j}$ for $n \geq 1$. Let $\{b_j\}_{j \geq 0}$ be such that $\sum_{j=1}^{\infty} |b_j| < \infty$. Let $\{a_n\}_{n \geq 0}$ satisfy $a_0 = b_0$ and*

$$a_n = b_n + \sum_{j=1}^{\infty} a_{n-j} p_j \quad n \geq 1. \tag{5.17}$$

Then $a_n = \sum_{j=0}^{\infty} b_j u_{n-j}$, $n \geq 0$ *and* $\lim_{n\to\infty} a_n = \frac{1}{\mu}\sum_{j=0}^{\infty} b_j$.

The nonarithmetic analog of the above is as follows.

Definition 8.5.7: A function $b(\cdot) \in \boldsymbol{B}_0$ is *directly Riemann integrable* (*dri*) on $[0,\infty)$ iff (i) for all $h > 0$, $\sum_{n=0}^{\infty} \sup\{|b(u)| : nh \leq u \leq (n+1)h\} < \infty$, and (ii) $\lim_{h\to 0} \sum_{n=0}^{\infty} h\big(\overline{m}_n(h) - \underline{m}_n(h)\big) = 0$ where

$$\begin{aligned} \overline{m}_n(h) &= \sup\{b(u) : nh \leq u \leq (n+1)h\} \\ \underline{m}_n(h) &= \inf\{b(u) : nh \leq u \leq (n+1)h\}. \end{aligned}$$

Theorem 8.5.14: (*The key renewal theorem, nonarithmetic case*). *Let* $F(\cdot)$ *be a nonarithmetic distribution with* $F(0) = 0$ *and* $\mu = \int_{[0,\infty)} u dF(u) < \infty$. *Let* $U(\cdot) = \sum_{n=0}^{\infty} F^{(n)}(\cdot)$ *be the renewal function associated with* F. *Let* $b(\cdot) \in \boldsymbol{B}_0$ *be directly Riemann integrable.*

Then the unique solution to the renewal equation

$$a = b + a * F \tag{5.18}$$

is given by $a = b * U$ *and*

$$\lim_{t\to\infty} a(t) = \frac{c(b)}{\mu} \tag{5.19}$$

where $c(b) \equiv \lim_{h\to 0} \sum_{n=0}^{\infty} h\overline{m}_n(h)$.

Remark 8.5.2: A sufficient condition for $b(\cdot)$ to be *dri* is that it is Riemann integrable on bounded intervals and that there exists a nonincreasing integrable function $h(\cdot)$ on $[0,\infty)$ and a constant C such that $|b(\cdot)| \leq Ch(\cdot)$ (Problem 8.18 (b)).

8.5.5 Applications

Here are two important applications of the above two theorems to a class of stochastic processes known as *regenerative processes*.

Definition 8.5.8:

(a) A sequence of random variables $\{Y_n\}_{n\geq 0}$ is called *regenerative* if there exists a renewal sequence $\{T_j\}_{j\geq 0}$ such that the random cycles and cycle length variables $\eta_j = \big(\{Y_i : T_j \leq i < T_{j+1}\}, T_{j+1} - T_j\big)$ for $j = 0, 1, 2, \ldots$ are iid.

(b) A stochastic process $\{Y(t) : t \geq 0\}$ is called *regenerative* if there exists a renewal sequence $\{T_j\}_{j\geq 0}$ such that the random cycles and cycle length variables $\eta_j \equiv \{Y(t) : T_j \leq t < T_{j+1}, T_{j+1} - T_j\}$ for $j = 0, 1, 2, \ldots$ are iid.

(c) In both (a) and (b), the sequence $\{T_j\}_{j\geq 0}$ are called the *regeneration times*.

Example 8.5.2: Let $\{Y_n\}_{n\geq 0}$ be a countable state space Markov chain (see Chapter 14) that is irreducible and recurrent. Fix a state Δ. Let

$$\begin{aligned} T_0 &= \min\{n : n > 0,\ Y_n = \Delta\} \\ T_{j+1} &= \min\{n : n > T_j,\ Y_n = \Delta\},\ n \geq 0. \end{aligned}$$

Then $\{Y_n\}_{n\geq 0}$ is regenerative (Problem 8.19).

Example 8.5.3: Let $\{Y(t) : t \geq 0\}$ be a continuous time Markov chain (see Chapter 14) with a countable state space that is irreducible and recurrent. Fix a state Δ. Let

$$\begin{aligned} T_0 &= \inf\{t : t > 0,\ Y(t) = \Delta\} \\ T_{j+1} &= \inf\{t : t > T_j,\ Y(t) = \Delta\}. \end{aligned}$$

Then $\{Y(t) : t \geq 0\}$ is regenerative (Problem 8.19).

Theorem 8.5.15: *Let $\{Y_n\}_{n\geq 0}$ be a regenerative sequence of random variables with some state space $(\mathbb{S}, \mathcal{S})$ where $\mathcal{S}$ is a σ-algebra on $\mathbb{S}$ with regeneration times $\{T_j\}_{j\geq 0}$. Let $f : \mathbb{S} \to \mathbb{R}$ be bounded and $\langle \mathcal{S}, \mathcal{B}(\mathbb{R})\rangle$-measurable. Let*

$$\begin{aligned} a_n &\equiv Ef(Y_{n+T_0}), \\ b_n &\equiv Ef(Y_{T_0+n})I(T_1 > T_0 + n). \end{aligned} \tag{5.20}$$

Let $\mu = E(T_1 - T_0) \in (0, \infty)$ and g.c.d. $\{j : p_j \equiv P(T_1 - T_0 = j) > 0\} = 1$. Then

(i)
$$a_n \to \int_{\mathbb{S}} f(y)\pi(dy)$$

where $\pi(A) \equiv \frac{1}{\mu} E\Big(\sum_{j=T_0}^{T_1-1} I_A(Y_j)\Big)$, $A \in \mathcal{S}$.

(ii) In particular,

$$\|P(Y_n \in \cdot) - \pi(\cdot)\| \to 0 \quad \text{as} \quad n \to \infty, \tag{5.21}$$

where $\|\cdot\|$ denotes the total variation norm.

Proof: By the regenerative property, $\{a_n\}_{n\geq 1}$ satisfies the renewal equation

$$a_n = b_n + \sum_{j=0}^{n} a_{n-j} p_j$$

and hence, part (i) of the theorem follows from Theorem 8.5.13 and the fact $\sum_{n=0}^{\infty} b_n = \mu\pi(A)$.

To prove (ii) note that $\tilde{a}_n \equiv Ef(Y_n) = E(f(Y_n)I(T_0 > n)) + \sum_{j=0}^{n} a_{n-j}P(T_0 = j)$ and by DCT $\lim_{n\to\infty} \tilde{a}_n = \lim_{n\to\infty} a_n$.

It is not difficult to show that for any two probability measures μ and ν on $(\mathbb{S}, \mathcal{S})$, the total variation norm

$$\|\mu - \nu\| = \sup\Big\{\Big|\int f d\mu - \int f d\nu\Big| : f \in \boldsymbol{B}(\mathbb{S}, \mathbb{R})\Big\}$$

where $\boldsymbol{B}(\mathbb{S}, \mathbb{R}) = \{f : f : \mathbb{S} \to \mathbb{R},\ \mathcal{F} \text{ measurable}, \sup\{|f(s)| : s \in \mathbb{S}\} \leq 1\}$ (Problem 4.10 (b)). Thus,

$$\begin{aligned} &\|P(Y_{n+T_0} \in \cdot) - \pi(\cdot)\| \\ &\quad\leq \sup\Big\{\Big|Ef(Y_{n_0+T}) - \int f d\pi\Big| : f \in \boldsymbol{B}(\mathbb{S}, \mathbb{R})\Big\}. \end{aligned} \tag{5.22}$$

Now, for any $f \in \boldsymbol{B}(\mathbb{S}, \mathbb{R})$ and any integer $K \geq 1$, from Theorem 8.5.13,

$$\begin{aligned} &\Big|Ef(Y_{n_0+T}) - \int f d\pi\Big| \\ &\quad\leq \sum_{j=0}^{K} b_j\Big|u_{n-j} - \frac{1}{\mu}\Big| + 2\sum_{j=(K+1)}^{\infty} P(T_1 - T_0 > j) \equiv \delta_n, \text{ say} \end{aligned} \tag{5.23}$$

where $\{b_j\}$ is defined in (5.20). Since $E(T_1 - T_0) < \infty$, given $\epsilon > 0$, there exists a K such that

$$\sum_{j=(K+1)}^{\infty} P(T_1 - T_0 > j) < \epsilon/2.$$

By Theorem (8.5.11), $u_n \to \frac{1}{\mu}$. Thus, in (5.23), $\overline{\lim}\, \delta_n \leq \epsilon$ and so from (5.22), (ii) follows. □

Theorem 8.5.16: *Let $\{Y(t) : t \geq 0\}$ be a regenerative stochastic process with state space $(\mathbb{S}, \mathcal{S})$ where $\mathcal{S}$ is a σ-algebra on $\mathbb{S}$. Let $f : \mathbb{S} \to \mathbb{R}$ be bounded and $\langle \mathcal{S}, \mathcal{B}(\mathbb{R})\rangle$-measurable. Let*

$$\begin{aligned} a(t) &= Ef(Y_{T_0+t}),\ t \geq 0, \\ b(t) &\equiv Ef(Y_{T_0+t})I(T_1 > T_0 + t),\ t \geq 0. \end{aligned}$$

Let $\mu = E(T_1 - T_0) \in (0, \infty)$ and the distribution of $T_1 - T_0$ be nonarithmetic. Then

(i) $$a(t) \to \int_{\mathbb{S}} f(y)\pi(dy)$$

where $\pi(A) = \frac{1}{\mu} E\Big(\int_{T_0}^{T} I_A(Y(u))du\Big)$, $A \in \mathcal{S}$.

(ii) In particular,

$$\|P(Y_t \in \cdot) - \pi(\cdot)\| \to 0 \quad as \quad t \to \infty \tag{5.24}$$

where $\|\cdot\|$ is the total variation norm.

The proof of this is similar to that of the previous theorem but uses Theorem 8.5.14. □

8.6 Ergodic theorems

8.6.1 Basic definitions and examples

The law of large numbers proved in Section 8.2 states that if $\{X_i\}_{i\geq 1}$ are pairwise independent and identically distributed and if $h(\cdot)$ is a Borel measurable function, then

$$\begin{aligned} &\text{the time average, i.e., } \frac{1}{n}\sum_{i=1}^{n} h(X_i) \\ &\to Eh(X_1), \text{ i.e., space average w.p. 1} \end{aligned} \tag{6.1}$$

as $n \to \infty$, provided $E|h(X_1)| < \infty$.

The goal of this section is to investigate how far the independence assumption can be relaxed.

Definition 8.6.1: (*Stationary sequences*). A sequence of random variables $\{X_i\}_{i\geq 1}$ on a probability space $(\Omega, \mathcal{F}, P)$ is called *strictly stationary* if for each $k \geq 1$ the joint distribution of $(X_{i+j} : j = 1, 2, \ldots, k)$ is the same for all $i \geq 0$.

Example 8.6.1: $\{X_i\}_{i\geq 1}$ iid.

Example 8.6.2: Let $\{X_i\}_{i\geq 1}$ be iid. Fix $1 \leq \ell < \infty$. Let $h : \mathbb{R}^\ell \to \mathbb{R}$ be a Borel function and $Y_i = h(X_i, X_{i+1}, \ldots, X_{i+\ell-1})$, $i \geq 1$. Then $\{Y_i\}_{i\geq 1}$ is strictly stationary.

Example 8.6.3: Let $\{X_i\}_{i\geq 1}$ be a Markov chain with a stationary distribution π. If $X_1 \sim \pi$ then $\{X_i\}_{i\geq 1}$ is strictly stationary (see Chapter 14).

It will be shown that if $\{X_i\}_{i\geq 1}$ is a strictly stationary sequence that is not a mixture of two other strictly stationary sequences, then (6.1) holds. This is known as *the ergodic theorem* (Theorem 8.6.1 below).

Definition 8.6.2: (*Measure preserving transformations*). Let $(\Omega, \mathcal{F}, P)$ be a probability space and $T : \Omega \to \Omega$ be $\langle \mathcal{F}, \mathcal{F} \rangle$ measurable. Then, T is called P*-preserving* (or simply *measure preserving* on $(\Omega, \mathcal{F}, P)$) if for all $A \in \mathcal{F}$, $P(T^{-1}(A)) = P(A)$. That is, the random point $T(\omega)$ has the same distribution as ω.

Let X be a real valued random variable on $(\Omega, \mathcal{F}, P)$. Let $X_i \equiv X(T^{(i-1)}(\omega))$ where $T^{(0)}(\omega) = \omega$, $T^{(i)}(\omega) = T(T^{(i-1)}(\omega))$, $i \geq 1$. Then $\{X_i\}_{i\geq 1}$ is a strictly stationary sequence.

It turns out that every strictly stationary sequence arises this way. Let $\{X_i\}_{i\geq 1}$ be a strictly stationary sequence defined on some probability space $(\Omega, \mathcal{F}, P)$. Let $\tilde{P}$ be the probability measure induced by $\tilde{X} \equiv \{X_i(\omega)\}_{i\geq 1}$ on $\big(\tilde{\Omega} \equiv \mathbb{R}^\infty, \tilde{\mathcal{F}} \equiv \mathcal{B}(\mathbb{R}^\infty)\big)$ where $\mathbb{R}^\infty$ is the space of all sequences of real numbers and $\mathcal{B}(\mathbb{R}^\infty)$ is the σ-algebra generated by finite dimensional cylinder sets of the form $\{x : (x_j : j = 1, 2, \ldots, k) \in A_k\}$, $1 \leq k < \infty$, $A_k \in \mathcal{B}(\mathbb{R}^k)$. Let $T : \mathbb{R}^\infty \to \mathbb{R}^\infty$ be the *unilateral (one sided) shift to the right*, i.e., $T\big((x_i)_{i\geq 1}\big) = (x_i)_{i\geq 2}$. Then T is measure preserving on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$. Let $Y_1(\tilde{\omega}) = x_1$, and $Y_i(\tilde{\omega}) = Y_1(T^{i-1}\tilde{\omega}) = x_i$ for $i \geq 2$ if $\tilde{\omega} = (x_1, x_2, x_3, \ldots)$. Then $\{Y_i\}_{i\geq 1}$ is a strictly stationary sequence on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ and has the same distribution as $\{X_i\}_{i\geq 1}$.

Example 8.6.4: Let $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}([0, 1])$, $P =$ Lebesgue measure. Let $T\omega \equiv 2\omega$ mod 1, i.e.,

$$T\omega = \begin{cases} 2\omega & \text{if} \quad 0 \leq \omega < \frac{1}{2} \\ 2\omega - 1 & \text{if} \quad \frac{1}{2} \leq \omega < 1 \\ 0 & \quad\;\; \omega = 1. \end{cases}$$

Then T is measure preserving since $P(\{\omega : a < T\omega < b\}) = (b - a)$ for all $0 < a < b < 1$ (Problem 8.20).

This example is an equivalent version of the iid sequence $\{\delta_i\}_{i\geq 1}$ of Bernoulli (1/2) random variables. To see this, let $\omega = \sum_{i=1}^\infty \frac{\delta_i(\omega)}{2^i}$ be the binary expansion of ω. Then $\{\delta_i\}_{i\geq 1}$ is iid Bernoulli (1/2) and $T\omega = 2\omega$ mod $1 = \sum_{i=2}^\infty \frac{\delta_i(\omega)}{2^{i-1}}$ (cf. Problem 7.4). Thus T corresponds with the unilateral shift to right on the iid sequence $\{\delta_i\}_{i\geq 1}$. For this reason, T is called the *Bernoulli shift*.

Example 8.6.5: (*Rotation*). Let $\Omega = \{(x, y) : x^2 + y^2 = 1\}$ be the unit circle. Fix θ_0 in $[0, 2\pi)$. If $\omega = (\cos\theta, \sin\theta)$, θ in $[0, 2\pi)$ set $T\omega = \big(\cos(\theta + \theta_0), \sin(\theta + \theta_0)\big)$. That is, T rotates any point ω on Ω counterclockwise through an angle θ_0. Then T is measure preserving w.r.t. the Uniform distribution on $[0, 2\pi]$.

Definition 8.6.3: Let $(\Omega, \mathcal{F}, P)$ be a probability space and $T : \Omega \to \Omega$ be a $\langle \mathcal{F}, \mathcal{F} \rangle$ measurable map. A set $A \in \mathcal{F}$ is T*-invariant* if $A = T^{-1}A$. A set $A \in \mathcal{F}$ is *almost* T*-invariant* w.r.t. P if $P(A \triangle T^{-1}A) = 0$ where $A_1 \triangle A_2 = (A_1 \cap A_2^c) \cup (A_1^c \cap A_2)$ is the symmetric difference of A_1 and A_2.

It can be shown that A is almost T-invariant w.r.t. P iff there exists a set A' that is T-invariant and $P(A \triangle A') = 0$ (Problem 8.21).

Examples of T-invariant sets are $A_1 = \{\omega : T^j\omega \in A_0$ for infinitely many $i \geq 1\}$ where $A_0 \in \mathcal{F}$; $A_2 = \big\{\omega : \frac{1}{n}\sum_{j=1}^n h(T^j\omega)$ converges as $n \to \infty\big\}$ where $h : \Omega \to \mathbb{R}$ is a $\mathcal{F}$ measurable function. On the other hand, the event $\{x : x_1 \leq 0\}$ is not shift invariant in $\big(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty)\big)$ nor is it almost shift invariant if $\tilde{P}$ corresponds to the iid case with a nondegenerate distribution.

The collection $\mathcal{I}$ of T-invariant sets is a σ-algebra and is called the *invariant σ-algebra*. A function $h : \Omega \to \mathbb{R}$ is $\mathcal{I}$-measurable iff $h(\omega) = h(T\omega)$ for all ω (Problem 8.22).

Definition 8.6.4: A measure preserving transformation T on a probability space $(\Omega, \mathcal{F}, P)$ is *ergodic* or *irreducible* (w.r.t. P) if A is T-invariant implies $P(A) = 0$ or 1.

Definition 8.6.5: A stationary sequence of random variables $\{X_i\}_{i\geq 1}$ is *ergodic* if the unilateral shift T is *ergodic* on the sequence space $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), \tilde{P})$ where $\tilde{P}$ is the measure on $\mathbb{R}^\infty$ induced by $\{X_i\}_{i\geq 1}$.

Example 8.6.6: Consider the above sequence space. Then $A \in \tilde{\mathcal{F}}$ is invariant with respect to the unilateral shift implies that A is in the tail σ-algebra $\mathcal{T} \equiv \bigcap_{n=1}^\infty \sigma(\tilde{X}_j(\omega), j \geq n)$ (Problem 8.23). If $\{X_i\}_{i\geq 1}$ are independent then by the Kolmogorov's zero-one law, $A \in \mathcal{T}$ implies $P(A) = 0$ or 1. Thus, if $\{X_i\}_{i\geq 1}$ are iid then it is *ergodic*.

On the other hand, mixtures of iid sequences are not ergodic as seen below.

Example 8.6.7: Let $\{X_i\}_{i\geq 1}$ and $\{Y_i\}_{i\geq 1}$ be two iid sequences with different distributions. Let δ be Bernoulli (p), $0 < p < 1$ and independent of both $\{X_i\}_{i\geq 1}$ and $\{Y_i\}_{i\geq 1}$. Let $Z_i \equiv \delta X_i + (1-\delta)Y_i$, $i \geq 1$. Then $\{Z_i\}_{i\geq 1}$ is a stationary sequence and is not *ergodic* (Problem 8.24).

The above example can be extended to mixtures of irreducible positive recurrent discrete state space Markov chains (Problem 8.25 (a)). Another example is Example 8.6.5, i.e., rotation of the circle when θ is rational (Problem 8.25 (b)).

Remark 8.6.1: There is a simple example of a measure preserving transformation T that is ergodic but T^2 is not. Let $\Omega = \{\omega_1, \omega_2\}$, $\omega_1 \neq \omega_2$. Let $T\omega_1 = \omega_2$, $T\omega_2 = \omega_1$, P be the distribution $P(\{\omega_1\}) = P(\{\omega_2\}) = \frac{1}{2}$. Then T is ergodic but T^2 is not (Problem 8.26).

8.6.2 Birkhoff's ergodic theorem

Theorem 8.6.1: *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $T : \Omega \to \Omega$ be a measure preserving ergodic map on $(\Omega, \mathcal{F}, P)$ and $X \in L^1(\Omega, \mathcal{F}, P)$. Then*

$$\frac{1}{n}\sum_{j=0}^{n-1} X(T^j\omega) \to EX \equiv \int_\Omega X dP \tag{6.2}$$

w.p. 1 and in L^1 as $n \to \infty$.

Remark 8.6.2: A more general version is without the assumption of T being ergodic. In this case, the right side of (6.2) is a random variable $Y(\omega)$ that is T-invariant, i.e., $Y(\omega) = Y(T(\omega))$ w.p. 1 and satisfies $\int_A X dP = \int_A Y dP$ for all T-invariant sets A. This Y is called the conditional expectation of X given $\mathcal{I}$, the σ-algebra of invariant sets (Chapter 13).

For a proof of this version, see Durrett (2004).

The proof of Theorem 8.6.1 depends on the following inequality.

Lemma 8.6.2: (*Maximal ergodic inequality*). *Let T be measure preserving on a probability space $(\Omega, \mathcal{F}, P)$ and $X \in L^1(\Omega, \mathcal{F}, P)$. Let $S_0(\omega) = 0$, $S_n(\omega) = \sum_{j=0}^{n-1} X(T^j\omega)$, $n \geq 1$, $M_n(\omega) = \max\{S_j(\omega) : 0 \leq j \leq n\}$. Then*

$$E\big(X(\omega) I\big(M_n(\omega) > 0\big)\big) \geq 0.$$

Proof: By definition of $M_n(\omega)$, $S_j(\omega) \leq M_n(\omega)$ for $1 \leq j \leq n$. Thus

$$X(\omega) + M_n(T\omega) \geq X(\omega) + S_j(T\omega) = S_{j+1}(\omega).$$

Also, since $M_n(T\omega) \geq 0$,

$$X(\omega) \geq X(\omega) - M_n(T\omega) = S_1(\omega) - M_n(T\omega).$$

Thus $X(\omega) \geq \max\big\{S_j(\omega) : 1 \leq j \leq n\big\} - M_n(T\omega)$. For ω such that $M_n(\omega) > 0$, $M_n(\omega) = \max\big\{S_j(\omega) : 1 \leq j \leq n\big\}$ and hence $X(\omega) \geq M_n(\omega) - M_n(T\omega)$. Also, since $X \in L^1(\Omega, \mathcal{F}, P)$ it follows that $M_n \in L^1(\Omega, \mathcal{F}, P)$ for all $n \geq 1$. Taking expectations yields

$$\begin{aligned}
& E\big(X(\omega) I\big(M_n(\omega) > 0\big)\big) \\
\geq \;& E\big(M_n(\omega) - M_n(T\omega) I\big(M_n(\omega) > 0\big)\big) \\
\geq \;& E\big(M_n(\omega) - M_n(T\omega) I\big(M_n(\omega) \geq 0\big)\big) \text{ (since } M_n(T\omega) \geq 0) \\
= \;& E\big(M_n(\omega) - M_n(T\omega)\big) = 0,
\end{aligned}$$

since T is measure preserving. $\square$

Remark 8.6.3: Note that the measure preserving property of T is used only at the last step.

Proof of Theorem 8.6.1: W.l.o.g. assume that $EX = 0$. Let $Z(\omega) \equiv \limsup_{n\to\infty} \frac{S_n(\omega)}{n}$. Fix $\epsilon > 0$ and set $A_\epsilon \equiv \{\omega : Z(\omega) > \epsilon\}$. It will be shown that $P(A_\varepsilon) = 0$. Clearly, A_ϵ is T invariant. Since T is ergodic, $P(A_\epsilon) = 0$ or 1. Suppose $P(A_\epsilon) = 1$. Let $Y(\omega) = X(\omega) - \epsilon$. Let $M_{n,Y}(\omega) \equiv \max\{S_{j,Y}(\omega) : 0 \le j \le n\}$ where $S_{0,Y}(\omega) \equiv 0$, $S_{j,Y}(\omega) \equiv \sum_{k=0}^{j-1} Y(T^k\omega)$, $j \ge 1$. Then by Lemma 8.6.2 applied to $Y(\omega)$

$$E\big(Y(\omega)I\big(M_{n,Y}(\omega) > 0\big)\big) \ge 0.$$

But $B_n \equiv \{\omega : M_{n,Y}(\omega) > 0\} = \{\omega : \sup_{1\le j\le n} \frac{1}{j} S_{j,Y}(\omega) > 0\}$. Clearly, $B_n \uparrow B \equiv \{\omega : \sup_{1\le j<\infty} \frac{1}{j} S_{j,Y}(\omega) > 0\}$. Since $\frac{1}{j} S_{j,Y}(\omega) = \frac{1}{j} S_j(\omega) - \epsilon$ for $j \ge 1$, $B \supset A_\epsilon$ and since $P(A_\epsilon) = 1$, it follows that $P(B) = 1$. Also $|Y| \le |X| + \epsilon \in L^1(\Omega, \mathcal{F}, P)$. So by the bounded convergence theorem, $0 \le E(YI_{B_n}) \to E(YI_B) = EY = 0 - \epsilon < 0$, which is a contradiction. Thus $P(A_\epsilon) = 0$. This being true for every $\epsilon > 0$ it follows that $P(\overline{\lim}_{n\to\infty} \frac{S_n(\omega)}{n} \le 0) = 1$. Applying this to $-X(\omega)$ yields

$$P\Big(\varliminf_{n\to\infty} \frac{S_n(\omega)}{n} \ge 0\Big) = 1$$

and hence $P\big(\lim_{n\to\infty} \frac{S_n(\omega)}{n} = 0\big) = 1$.

To prove L^1-convergence, note that applying the above to X^+ and X^- yields

$$f_n(\omega) \equiv \frac{1}{n}\sum_{i=1}^n X^+(T^i\omega) \to EX^+(\omega) \quad \text{w.p. 1.}$$

Since T is measure preserving $\int f_n(\omega)dP = EX^+(\omega)$ for all n. So by Scheffe's theorem (Lemma 8.2.5), $\int |f_n(\omega) - EX^+(\omega)|dP \to 0$, i.e., $E\left|\frac{1}{n}\sum_{i=1}^n X^+(T^i\omega) - EX^+\right| \to 0$. Similarly, $E\left|\frac{1}{n}\sum_{i=1}^n X^-(T^i\omega) - EX^-\right| \to 0$. This yields L^1 convergence. □

Corollary 8.6.3: *Let $\{X_i\}_{i\ge1}$ be a stationary ergodic sequence of $\mathbb{R}^k$ valued random variables on some probability space $(\Omega, \mathcal{F}, P)$. Let $h : \mathbb{R}^k \to \mathbb{R}$ be Borel measurable and let $E|h(X_1, X_2, \ldots, X_k)| < \infty$. Then*

$$\frac{1}{n}\sum_{i=1}^n h(X_i, X_{i+1}, \ldots, X_{i+k-1}) \to Eh(X_1, X_2, \ldots, X_k) \quad \textit{w.p. 1.}$$

Proof: Consider the probability space $\tilde{\Omega} = (\mathbb{R}^k)^\infty$, $\tilde{\mathcal{F}} \equiv \mathcal{B}\big((\mathbb{R}^k)^\infty\big)$ and $\tilde{P}$ the probability measure induced by the map $\omega \to (X_i(\omega))_{i\ge1}$ and the unilateral shift map $\tilde{T}$ on $\tilde{\Omega}$ defined by $\tilde{T}(x_i)_{i\ge1} = (x_i)_{i\ge2}$. Then $\tilde{T}$ is measure preserving and ergodic. So the corollary follows from Theorem 8.6.1. □

Remark 8.6.4: This corollary is useful in statistical time series analysis. If $\{X_i\}_{i\geq 1}$ is a real valued stationary ergodic sequence, then the mean $m \equiv EX_1$, variance $\mathrm{Var}(X_1)$, and covariance $\mathrm{Cov}(X_1, X_2)$ can all be estimated consistently by the corresponding sample functions

$$\frac{1}{n}\sum_{i=1}^{n} X_i, \quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big)^2, \quad \text{and}$$
$$\frac{1}{n}\sum_{i=1}^{n} X_i X_{i+1} - \Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big)^2.$$

Further, the joint distribution of $(X_1, X_2, \ldots, X_k)$ for any $k \geq 1$, can be estimated consistently by the corresponding empirical measure, i.e., $L_n(A_1, A_2, \ldots, A_k) \equiv \frac{1}{n}\sum_{i=1}^{n} I(X_{i+k} \in A_k,\ j = 1, 2, \ldots, k)$, which converges to

$$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_k \in A_k) \quad \text{w.p. 1}$$

where $A_i \in \mathcal{B}(\mathbb{R})$, $i = 1, 2, \ldots, k$.

The next three results (Theorems 8.6.4–8.6.6) are consequences and extensions of the ergodic theorem, Theorem 8.6.1. For proofs, see Durrett (2004).

The first one is the following result on the behavior of the log-likelihood function of a stationary ergodic sequence of random variables with a finite range.

Theorem 8.6.4: (*Shannon-McMillan-Breiman theorem*). *Let $\{X_i\}_{i\geq 1}$ be a stationary ergodic sequence of random variables with values in a finite set $S \equiv \{a_1, a_2, \ldots, a_k\}$. For each n, $x_1, x_2, \ldots, x_n$ in S, let*

$$\begin{aligned} p(x_n \mid x_{n-1}, x_{n-2}, \ldots, x_1) &= P(X_n = x_n \mid X_j = x_j, 1 \leq j \leq n-1) \\ &\equiv \frac{P(X_j = x_j : 1 \leq j \leq n)}{P(X_j = x_j : 1 \leq j \leq n-1)} \end{aligned}$$

whenever the denominator is positive and let $p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$. Then

$$\lim_{n\to\infty} \frac{1}{n} \log p(X_1, X_2, \ldots, X_n) = -H \quad \textit{exists w.p. 1}$$

where $H \equiv \lim_{n\to\infty} E\big(-\log p(X_n \mid X_{n-1}, X_{n-2}, \ldots, X_1)\big)$ is called the entropy rate of $\{X_i\}_{i\geq 1}$.

Remark 8.6.5: In the iid case this is a consequence of the strong law of large numbers, and H can be identified as $\sum_{j=1}^{k}(-\log p_j)p_j$ where $p_j =$

$P(X_1 = a_j)$, $1 \le j \le k$. This is called the Kolmogorov-Shannon entropy of the distribution $\{p_j : 1 \le j \le k\}$.

If $\{X_i\}_{i \ge 1}$ is a stationary ergodic Markov chain, then again it is a consequence of the strong law of large numbers, and H can be identified with

$$E\big(-\log p(X_2 \mid X_1)\big) = \sum_{i=1}^{k} \pi_i \sum_{j=1}^{k} (-\log p_{ij}) p_{ij}$$

where $\pi \equiv \{\pi_i : 1 \le i \le k\}$ is the stationary distribution and $P \equiv \big((p_{ij})\big)$ is the transition probability matrix of the Markov chain $\{X_i\}_{i \ge 1}$. See Problem 8.27.

A more general version of the ergodic Theorem 8.6.1 is the following.

Theorem 8.6.5: (*Kingman's subadditive ergodic theorem*). *Let* $\{X_{m,n} : 0 \le m < n\}_{n \ge 1}$ *be a collection of random variables such that*

(i) $X_{0,m} + X_{m,n} \ge X_{0,n}$ *for all* $0 \le m < n$, $n \ge 1$.

(ii) For all $k \ge 1$, $\{X_{nk,(n+1)k}\}_{n \ge 1}$ *is a stationary sequence.*

(iii) The sequence $\{X_{m,m+k}, k \ge 1\}$ *has a distribution that does not depend on* $m \ge 0$.

(iv) $EX_{0,1}^{+} < \infty$ *and for all* n, $\frac{EX_{0,n}}{n} \ge \gamma_0$, *where* $\gamma_0 > -\infty$.

Then

(i) $\lim_{n \to \infty} \frac{EX_{0,n}}{n} = \inf_{n \ge 1} \frac{EX_{0,n}}{n} \equiv \gamma$.

(ii) $\lim_{n \to \infty} \frac{X_{0,n}}{n} \equiv X$ *exists w.p. 1 and in* L^1, *and* $EX = \gamma$.

(iii) If $\{X_{nk,(n+1)k}\}_{n \ge 1}$ *is ergodic for each* $k \ge 1$, *then* $X \equiv \gamma$ *w.p. 1.*

A nice application of this is a result on products of random matrices.

Theorem 8.6.6: *Let* $\{A_i\}_{i \ge 1}$ *be a stationary sequence of* $k \times k$ *random matrices with nonnegative entries. Let* $\alpha_{m,n}(i,j)$ *be the* (i,j)*th entry in* $A_{m+1}, \cdots, A_n$. *Suppose* $E|\log \alpha_{1,2}(i,j)| < \infty$ *for all* i, j. *Then*

(i) $\lim_{n \to \infty} \frac{1}{n} \log \alpha_{0,n}(i,j) = \eta$ *exists w.p. 1.*

(ii) For any m, $\lim_{n \to \infty} \frac{1}{n} \log \|A_{m+1} \cdots, A_n\| = \eta$ *w.p. 1, where for any* $k \times k$ *matrix* $B \equiv ((b_{ij}))$, $\|B\| = \max\big\{ \sum_{j=1}^{k} |b_{ij}| : 1 \le i \le k \big\}$.

Remark 8.6.6: A concept related to ergodicity is that of mixing. A measure preserving transformation T on a probability space $(\Omega, \mathcal{F}, P)$ is *mixing* if for all A, $B \in \mathcal{B}$

$$\lim_{n\to\infty} \left|P(A \cap T^{-n}B) - P(A)P(T^{-n}B)\right| = 0.$$

A stationary sequence of random variables $\{X_i\}_{i\geq 1}$ is *mixing* if the unilateral shift on the sequence space $\mathbb{R}^\infty$ induced by $\{X_i\}_{i\geq 1}$ is mixing. If T is mixing and A is T-invariant, then taking $B = A$ in the above yields

$$P(A) = P^2(A)$$

i.e., $P(A) = 0$ or 1. Thus, if T is mixing, then T is ergodic. Conversely, if T is ergodic, then by Theorem 8.6.1, for any B in $\mathcal{B}$

$$\frac{1}{n}\sum_{j=1}^{n} I_B(T^j\omega) \to P(B) \text{ w.p. 1.}$$

Integrating both sides over A w.r.t. P yields $\frac{1}{n}\sum_{j=1}^{n} P(A \cap T^{-j}B) \to P(A)P(B)$, i.e., T is mixing in an average sense, i.e., the Cesaro sense. A sufficient condition for a stationary sequence to be mixing is that the tail σ-algebra be trivial. If $\{X_i\}_{i\geq 1}$ is a stationary irreducible Markov chain with a countable state space, then it is *mixing* iff it is aperiodic.

For proofs of the above results, see Durrett (2004).

8.7 Law of the iterated logarithm

Let $\{X_n\}_{n\geq 1}$ be a sequence of iid random variables with $EX_1 = 0$, $EX_1^2 = 1$. The SLLN asserts that the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to 0$ w.p. 1. The central limit theorem (to be proved later) asserts that for all $-\infty < a < b < \infty$, $P(a \leq \sqrt{n}\bar{X}_n \leq b) \to \Phi(b) - \Phi(a)$ where $\Phi(\cdot)$ is the standard Normal cdf. This suggests that $S_n = \sum_{i=1}^{n} X_i$ is of the order magnitude $\sqrt{n}$ for large n. This raises the question of how large does $\frac{S_n}{\sqrt{n}}$ get as a function of n. It turns out that it is of the order $\sqrt{2n \log\log n}$. More precisely, the following holds:

Theorem 8.7.1: (*Law of the iterated logarithm*). *Let $\{X_i(\omega)\}_{i\geq 1}$ be iid random variables on a probability space $(\Omega, \mathcal{F}, P)$ with mean zero and variance one. Let $S_0(\omega) = 0$, $S_n(\omega) = \sum_{i=1}^{n} X_i(\omega)$, $n \geq 1$. For each ω, let $A(\omega)$ be the set of limit points of $\left\{\frac{S_n(\omega)}{\sqrt{2n\log\log n}}\right\}_{n\geq 1}$. Then $P\{\omega : A(\omega) = [-1, +1]\} = 1$.*

For a proof, see Durrett (2004).

A deep generalization of the above was obtained by Strassen (1964).

Theorem 8.7.2: *Under the setup of Theorem 8.7.1, the following holds: Let* $Y_n(\frac{j}{n};\omega) = \frac{S_j(\omega)}{\sqrt{2n\log\log n}}$, $j = 0, 1, 2, \ldots, n$ *and* $Y_n(t,\omega)$ *be the function obtained by linearly interpolating the above values on* $[0,1]$. *For each* ω, *let* $B(\omega)$ *be the set of limit points of* $\{Y_n(\cdot,\omega)\}_{n\geq 1}$ *in the function space* $C[0,1]$ *of all continuous functions on* $[0,1]$ *with the supnorm. Then*

$$P\{\omega : B(\omega) = K\} = 1$$

where $K \equiv \big\{f : f : [0,1] \to \mathbb{R},\ f$ *is continuously differentiable,* $f(0) = 0$ *and* $\frac{1}{2}\int_0^1 (f'(t))^2 dt \leq 1\big\}$.

8.8 Problems

8.1 Prove Theorem 8.1.3 and Corollary 8.1.4.

(**Hint:** Use Chebychev's inequality.)

8.2 Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$ such that for some $m \in \mathbb{N}$ and for each $i = 1, \ldots, m$, $\{X_i, X_{i+m}, X_{i+2m}, \ldots\}$ are identically distributed and pairwise independent. Furthermore, suppose that $E(|X_1| + \cdots + |X_m|) < \infty$. Show that

$$\overline{X}_n \longrightarrow \frac{1}{m}\sum_{i=1}^{m} EX_i, \quad \text{w.p. } 1.$$

(**Hint:** Reduce the problem to nonnegative X_n's and apply Theorem 8.2.7 for each $i = 1, \ldots, m$.)

8.3 Let f be a bounded measurable function on [0,1] that is continuous at $\frac{1}{2}$. Evaluate $\lim\limits_{n\to\infty} \int_0^1 \int_0^1 \cdots \int_0^1 f\Big(\frac{x_1+x_2+\cdots+x_n}{n}\Big) dx_1 dx_2 \ldots dx_n$.

8.4 Show that if $P(|X| > \alpha) < \frac{1}{2}$ for some real number α, then any median of X must lie in the interval $[-\alpha, \alpha]$.

8.5 Prove Theorem 8.3.4 using Kolmogorov's first inequality (Theorem 8.3.1 (a)).

(**Hint:** Apply Theorem 8.3.1 to $\Delta_{n,k}$ defined in the proof of Theorem 8.3.3 to establish (3.4).)

8.6 Let $\{X_n\}_{n\geq 1}$ be a sequence of iid random variables with $E|X_1|^\alpha < \infty$ for some $\alpha > 0$. Derive a necessary and sufficient condition on α for almost sure convergence of the series $\sum_{n=1}^{\infty} X_n \sin 2\pi nt$ for all $t \in (0,1)$.

8.7 Show that for any given sequence of random variables $\{X_n\}_{n\geq 1}$, there exists a sequence of real numbers $\{a_n\}_{n\geq 1} \subset (0,\infty)$ such that $\frac{X_n}{a_n} \to 0$ w.p. 1.

8.8 Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables with

$$P(X_n = 2) = P(X_n = n^\beta) = a_n, \ P(X_n = a_n) = 1 - 2a_n$$

for some $a_n \in (0, \frac{1}{3})$ and $\beta \in \mathbb{R}$. Show that $\sum_{n=1}^\infty X_n$ converges if and only if $\sum_{n=1}^\infty a_n < \infty$.

8.9 Let $\{X_n\}_{n\geq 1}$ be a sequence of iid random variables with $E|X_1|^p = \infty$ for some $p \in (0,2)$. Then $P(\limsup_{n\to\infty} |n^{-1/p}\sum_{i=1}^n X_i| = \infty) = 1$.

8.10 For any random variable X and any $r \in (0,\infty)$, $E|X|^r < \infty$ iff $\sum_{n=1}^\infty n^{r-1}(\log n)^r P(|X| > n\log n) < \infty$.

(**Hint:** Check that $\sum_{n=1}^m n^{r-1}(\log n)^r \sim r^{-1}m^r(\log m)^r$ as $m \to \infty$.)

8.11 Let $\{X_n\}_{n\geq 1}$ be a sequence of independent random variables with $EX_n = 0$, $EX_n^2 = \sigma_n^2$, $s_n^2 = \sum_{j=1}^n \sigma_j^2 \to \infty$. Then, show that for any $a > \frac{1}{2}$,

$$s_n^{-2}(\log s_n^2)^{-a}\sum_{i=1}^n X_i \to 0 \quad \text{w.p. 1.}$$

8.12 Show that for $p \in (0,2)$, $p \neq 1$, (4.12) holds.

(**Hint:** For $p \in (1,2)$, $\sum_{n=1}^\infty |EZ_n/n^{1/p}| \leq \sum_{n=1}^\infty E|X_1|I(|X_1| > n)n^{-1/p} = \sum_{j=1}^\infty \sum_{n=1}^j n^{-1/p} \cdot E|X_1|I(j < |X_1|^p \leq j+1) \leq \frac{p}{p-1}E|X_1|^p < \infty$, by (4.10). For $p \in (0,1)$, $\sum_{n=1}^\infty |EZ_n/n^{1/p}| \leq \sum_{j=1}^\infty(\sum_{n=j}^\infty n^{-1/p})E|X_1|I(j-1 < |X_1|^p \leq j) \leq \frac{1}{1-p}E|X_1|^p$, by (4.9).)

8.13 Let $Y_i = x_i\beta + \epsilon_i$, $i \geq 1$ where $\{\epsilon_n\}_{n\geq 1}$ is a sequence of iid random vectors, $\{x_n\}_{n\geq 1}$ is a sequence of constants, and $\beta \in \mathbb{R}$ is a constant (the regression parameter). Let $\hat\beta_n = \sum_{i=1}^n x_iY_i/\sum_{i=1}^n x_i^2$ denote (the least squares) estimator of β. Let $n^{-1}\sum_{i=1}^n x_i^2 \to c \in (0,\infty)$ and $E\epsilon_1 = 0$.

(a) If $E|\epsilon_1|^{1+\delta} < \infty$ for some $\delta \in (0,\infty)$, then show that

$$\hat\beta_n \longrightarrow \beta \quad \text{as} \quad n \to \infty, \quad \text{w.p. 1.} \tag{8.1}$$

(b) Suppose $\sup\{|x_i| : i \geq 1\} < \infty$ and $E|\epsilon_1| < \infty$. Show that (8.1) holds.

8.14 (*Strongly consistent estimation.*) Let $\{X_i\}_{i\geq 1}$ be random variables on some probability space $(\Omega, \mathcal{F}, P)$ such that (i) for some integer $m \geq 1$ the collections $\{X_i : i \leq n\}$ and $\{X_i : i \geq n+m\}$ are independent for each $n \geq 1$, and (ii) the distribution of $\{X_{i+j}; 0 \leq j \leq k\}$ is independent of i, for all $k \geq 0$.

(a) Show that for every $\ell \geq 1$ and $h : \mathbb{R}^\ell \to \mathbb{R}$ with $E|h(X_1, X_2, \ldots, X_\ell)| < \infty$, there are functions $\{f_n : \mathbb{R}^n \to \mathbb{R}\}_{n\geq 1}$ such that $f_n(X_1, X_2, \ldots, X_n) \to \lambda \equiv Eh(X_1, X_2, \ldots, X_\ell)$ w.p. 1. In this case, one says λ is estimable from $\{X_i\}_{i\geq 1}$ in a strongly consistent manner.

(b) Now suppose the distribution $\mu(\cdot)$ of X_1 is a mixture of the form $\mu = \sum_{i=1}^k \alpha_i \mu_i$. Suppose there exist disjoint Borel sets $\{A_i\}_{1\leq i\leq k}$ in $\mathbb{R}$ such that $\mu_i(A_i) = 1$ for each i. Show that all the α_i's as well as $\lambda_i \equiv \int h_i(x) d\mu_i$ where $h_i \in L_1(\mu_i)$ are estimable from $\{X_i\}_{i\geq 1}$ in a strongly consistent manner.

8.15 (*Normal numbers*). Recall that in Section 4.5 it was shown that for any positive integer $p > 1$ and for any $0 \leq \omega \leq 1$, it is possible to write ω as

$$\omega = \sum_{i=1}^{\infty} \frac{X_i(\omega)}{p^i} \tag{8.2}$$

where for each i, $X_i(\omega) \in \{0, 1, 2, \ldots, p-1\}$. Recall also that such an expansion is unique except for ω of the form q/p^n, $q = 1, 2, \ldots, p^n - 1$, $n \geq 1$ in which case there are exactly two expansions, one of which is recurring. In what follows, for such ω's the recurrent expansion will be the one used in (8.2). A number ω in [0,1] is called *normal* w.r.t. the integer p if for every finite pattern $a_1 a_2 \ldots a_k$ where $k \geq 1$ is a positive integer and $a_i \in \{0, 1, 2, \ldots, p-1\}$ for $1 \leq i \leq k$ the relative frequency $\frac{1}{n}\sum_{i=1}^n \delta_i(\omega)$ where

$$\delta_i(\omega) = \begin{cases} 1 & \text{if} \quad X_{i+j}(\omega) = a_{j+1}, \ j = 0, 1, 2, \ldots, k-1 \\ 0 & \text{otherwise} \end{cases}$$

converges to p^{-k} as $n \to \infty$. A number ω in [0,1] is called *absolutely normal* if it is normal w.r.t. p for every integer $p > 1$. Show that the set A of all numbers ω in [0,1] that are absolutely normal has Lebesgue measure one.

(**Hint:** Note that in (8.2), the function $\{X_i(\omega)\}_{i\geq 1}$ are iid random variables. Now use Problem 8.14 repeatedly.)

8.16 Show that for the renewal sequence $\{S_n\}_{n=0}^{\infty}$, if $P(X_1 > 0) > 0$, then $\lim_{n\to\infty} S_n = \infty$ w.p. 1.

8.17 (a) Show that $\{a_n\}_{n\geq 0}$ of (5.16) is the unique solution to (5.15) by using generating functions (cf. Section 5.5).

(b) Deduce Theorems 8.5.13 and 8.5.14 from Theorems 8.5.11 and 8.5.12, respectively.

(**Hint:** For Theorems 8.5.13 use the *DCT*, and for Theorem 8.5.14, show first that

$$\begin{aligned} & \sum_{n=0}^{k} \underline{m}_n(h)\big(U((n+1)h) - U(nh)\big) \\ \leq\ & a(kh) \\ \leq\ & \sum_{n=0}^{k} \overline{m}_n(h)\big(U((n+1)h) - U(nh)\big).) \end{aligned}$$

8.18 (a) Let $b(\cdot) : [0, \infty) \to \mathbb{R}$ be *dri*. Show that $b(\cdot)$ is Riemann integrable on every bounded interval. Conclude that if $b(\cdot)$ is *dri* it must be continuous almost everywhere w.r.t. Lebesgue measure.

(b) Let $b(\cdot) : [0, \infty) \to \mathbb{R}$ be Riemann integrable on $[0, K]$ for each $K < \infty$. Let $h(\cdot) : [0, \infty) \to \mathbb{R}^+$ be nonincreasing and integrable w.r.t. Lebesgue measure and $|b(\cdot)| \leq h(\cdot)$ on $[0, \infty)$. Show that $b(\cdot)$ is *dri*.

8.19 Verify that the sequence $\{Y_n\}_{n\geq 0}$ in Example 8.5.2 and the process $\{Y(t) : t \geq 0\}$ in Example 8.5.3 are both regenerative.

8.20 Show that the map T in Example 8.6.4 in Section 8.6 is measure preserving.

(**Hint:** Show that for $0 < a < b < 1$, $P\big(\omega : T\omega \in (a, b)\big) = (b - a)$.)

8.21 Let T be a measure preserving map on a probability space $(\Omega, \mathcal{F}, P)$. Show that A is almost T-invariant w.r.t. P iff there exists a set A_1 such that $A_1 = T^{-1}A_1$ and $P(A \triangle A_1) = 0$.

(**Hint:** Consider $A_1 = \bigcup_{n=0}^{\infty} T^{-n}A$.)

8.22 Show that a function $h : \Omega \to \mathbb{R}$ is $\mathcal{I}$-measurable iff $h(\omega) = h(T\omega)$ for all ω where $\mathcal{I}$ is the σ-algebra of T-invariant sets.

8.23 Consider the sequence space $\big(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty)\big)$. Show that $A \in \mathcal{B}(\mathbb{R}^\infty)$ is invariant w.r.t. the unilateral shift T implies that A is in the tail σ-algebra.

8.24 In Example 8.6.7 of Section 8.6, show that $\{Z_i\}_{i\geq 1}$ is a stationary sequence that is not ergodic.

(**Hint:** Assuming it is ergodic, derive a contradiction using the ergodic Theorem 8.6.1.)

8.25 (a) Extend Example 8.6.7 to the Markov chain case with two disjoint irreducible positive recurrent subsets.

(b) Show that in Example 8.6.5, if θ_0 is rational, then T is not ergodic.

8.26 (a) Verify that in Remark 8.6.1, T is ergodic but T^2 is not.

(b) Construct a Markov chain with four states for which T is ergodic but T^2 is not.

8.27 In Remark 8.6.5, prove the Shannon-McMillan-Breiman theorem directly for the Markov chain case.

(**Hint:** Express $p(X_1, X_2, \ldots, X_n)$ as $\left(\prod_{i=1}^{n-1} p_{X_i X_{i+1}}\right) p(X_1)$.)

8.28 Let $\{X_i\}_{i\geq 1}$ be iid Bernoulli (1/2) random variables. Let

$$\begin{aligned} W_1 &= \sum_{i=1}^{\infty} \frac{2X_{2i}}{4^i} \\ W_2 &= \sum_{i=1}^{\infty} \frac{X_{2i-1}}{4^i}. \end{aligned}$$

(a) Show that W_1 and W_2 are independent.

(b) Let $A_1 = \{\omega : \omega \in (0,1)$ such that in the expansion of ω in base 4 only the digits 0 and 2 appear$\}$ and $A_2 = \{\omega : \omega \in (0,1)$ such that in the expansion of ω in base 4 only the digits 0 and 1 appear$\}$. Show that $m(A_1) = m(A_2) = 0$ where $m(\cdot)$ is Lebesgue measure and hence that the distribution of W_1 and W_2 are singular w.r.t. $m(\cdot)$.

(c) Let $W \equiv W_1 + W_2$. Then show that W has uniform (0,1) distribution.

(**Hint:** For (b) use the SLLN.)

Remark: This example shows that the convolution of two singular probability measures can be absolutely continuous w.r.t. Lebesgue measure.

8.29 Let $\{X_n\}_{n\geq 1}$ be a sequence of pairwise independent and identically distributed random variables with $P(X_1 \leq x) = F(x)$, $x \in \mathbb{R}$. Fix $0 < p < 1$. Suppose that $F(\zeta_p + \epsilon) > p$ for all $\epsilon > 0$ where

$$\zeta_p = F^{-1}(p) \equiv \inf\{x : F(x) \geq p\}.$$

Show that $\hat{\zeta}_n \equiv F_n^{-1}(p) \equiv \inf\{x : F_n(x) \geq p\}$ converges to ζ_p w.p. 1 where $F_n(x) \equiv n^{-1} \sum_{i=1}^{n} I(X_i \leq x)$, $x \in \mathbb{R}$ is the empirical distribution function of $X_1, \ldots, X_n$.

8.30 Let $\{X_i\}_{i\geq 1}$ be random variables such that $EX_i^2 < \infty$ for all $i \geq 1$. Suppose $\frac{1}{n}\sum_{i=1}^{n} EX_i \to 0$ and $a_n \equiv \frac{1}{n^2}\sum_{j=0}^{n}(n-j)v(j) \to 0$ as $n \to \infty$ where $v(j) = \sup_i \left|\text{Cov}(X_i, X_{i+j})\right|$.

(a) Show that $\bar{X}_n \longrightarrow^p 0$.

(b) Suppose further that $\sum_{n=1}^{\infty} a_n < \infty$. Show that $\bar{X}_n \to 0$ w.p. 1.

(c) Show that as $n \to \infty$, $v(n) \to 0$ implies $a_n \to 0$ but the converse need not hold.

8.31 Let $\{X_i\}_{i\geq 1}$ be iid random variables with cdf $F(\cdot)$. Let $F_n(x) \equiv \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x)$ be the empirical cdf. Suppose $x_n \to x_0$ and $F(\cdot)$ is continuous at x_0. Show that $F_n(x_n) \to F(x_0)$ w.p. 1.

8.32 Let p be a positive integer > 1. Let $\{\delta_i\}_{i\geq 1}$ be iid random variable with distribution $P(\delta_1 = j) = p_j$, $0 \leq j \leq p-1$, $p_j \geq 0$, $\sum_0^{p-1} p_j = 1$. Let $X = \sum_{i=1}^{\infty} \frac{\delta_i}{p^i}$. Show that

(a) $P(X \in (0,1)) = 1$.

(b) $F_X(x) \equiv P(X \leq x)$ is continuous and strictly increasing in (0,1) if $0 < p_j < 1$ for any $0 \leq j \leq p-1$.

(c) $F_X(\cdot)$ is absolutely continuous iff $p_j = \frac{1}{j}$ for all $0 \leq j \leq p-1$ in which case $F_X(x) \equiv x$, $0 \leq x \leq 1$.

8.33 (*Random AR-series*). Let $\{X_n\}_{n\geq 0}$ be a sequence of random variables such that

$$X_{n+1} = \rho_{n+1}X_n + \epsilon_{n+1}, \quad n \geq 0$$

where the sequence $\{(\rho_n, \epsilon_n)\}_{n\geq 1}$ are iid and independent of X_0.

(a) Show that if $E(\log|\rho_1|) < 0$ and $E(\log|\epsilon_1|)^+ < \infty$ then

$$\hat{X}_n \equiv \sum_{j=0}^{n} \rho_1\rho_2\ldots\rho_j, \epsilon_{j+1} \quad \text{converges w.p. 1.}$$

(b) Show that under the hypothesis of (a), for any bounded continuous function $h : \mathbb{R} \to \mathbb{R}$ and for any distribution of X_0

$$Eh(X_n) \to Eh(\hat{X}_\infty).$$

(**Hint:** Show by SLLN that there is a $0 < \lambda < 1$ such that $\rho_1, \rho_2, \ldots, \rho_j = 0(\lambda^j)$ w.p. 1 as $j \to \infty$ and by Borel-Cantelli $|\epsilon_j| = 0(\lambda'^j)$ for some $\lambda' > 0 \ni \lambda'\lambda < 1$.)

8.34 (*Iterated random functions*). Let $(\mathbb{S}, \rho)$ be a complete separable metric space. Let $(G, \mathcal{G})$ be a measurable space. Let $f : G \times \mathbb{S} \to \mathbb{S}$ be

$\langle \mathcal{G} \times \mathcal{B}(\mathbb{S}), \mathcal{B}(\mathbb{S}) \rangle$ measurable function. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{\theta_i\}_{i \geq 1}$ be iid G-valued random variables on $(\Omega, \mathcal{F}, P)$. Let X_0 be an $\mathbb{S}$-valued random variable on $(\Omega, \mathcal{F}, P)$ independent of $\{\theta_i\}_{i \geq 1}$. Define $\{X_n\}_{n \geq 0}$ by the random iteration scheme,

$$X_0(x, \omega) \equiv x$$

$$X_{n+1}(x, \omega) = f\big(\theta_{n+1}(\omega), X_n(x, \omega)\big) \ n \geq 0.$$

(a) Show that for each $n \geq 0$, the map $X_n = \mathbb{S} \times \Omega \to \mathbb{S}$ is $\langle \mathcal{B}(\mathbb{S}) \times \mathcal{F}, \mathcal{B}(\mathbb{S}) \rangle$ measurable.

(b) Let $f_n(x) \equiv f_n(x, \omega) \equiv f(\theta_n(\omega), x)$. Let $\hat{X}_n(x, \omega) = f_1(f_2, \ldots, f_n(x))$. Show that for each x and n, $\hat{X}_n(x, \omega)$ and $X_n(x, \omega)$ have the same distribution.

(c) Now assume that for all ω, $f(\theta_1(\omega), x)$ is Lipschitz from $\mathbb{S}$ to $\mathbb{S}$, i.e.,

$$\ell_i(\omega) \equiv \sup_{x \neq y} \frac{d(f(\theta_i(\omega), x), f(\theta_i(\omega), y))}{d(x, y)} < \infty.$$

Show that $\ell_i(\omega)$ is a random variable on $(\Omega, \mathcal{F}, P)$, i.e. that $\ell_i(\cdot) : \Omega \to \mathbb{R}^+$ is $\langle \mathcal{F}, \mathcal{B}(\mathbb{R}) \rangle$ measurable.

(d) Suppose that $E|\log \ell_1(\omega)| < \infty$ and $E \log \ell_1(\omega) < 0$, $E|\log d(f(\theta_1, x), x)| < \infty$ for all x. Show that $\lim_n \hat{X}_n(x, \omega) = \hat{X}_\infty(\omega)$ exists w.p. 1 and is independent of x w.p. 1.

(**Hint:** Use Borel-Cantelli to show that for each x, $\{\hat{X}_n(x, \omega)\}_{n \geq 1}$ is Cauchy in $(\mathbb{S}, \rho)$.)

(e) Under the hypothesis in (d) show that for any bounded continuous $h : \mathbb{S} \to \mathbb{R}$ and for any $x \in \mathbb{S}$, $\lim_{n \to \infty} Eh(X_n(x, \omega)) = Eh(\hat{X}_\infty(\omega))$.

(f) Deduce the results in Problems 7.15 and 8.33 as special cases.

8.35 (*Extension of Gilvenko-Cantelli (Theorem 8.2.4) to the multivariate case*). Let $\{X_n\}_{n \geq 1}$ be a sequence of pairwise independent and identically distributed random vectors taking values in $\mathbb{R}^k$ with cdf $F(x) \equiv P\big(X_{11} \leq x_1, X_{12} \leq x_2, \ldots, X_{1k} \leq x_k\big)$ where $X_1 = (X_{11}, X_{12}, \ldots, X_{1k})$ and $x = (x_1, x_2, \ldots, x_k) \in \mathbb{R}$. Let $F_n(x) \equiv \frac{1}{n} \sum_{i=1}^n I(X_i \leq x)$ be the *empirical cdf* based on $\{X_i\}_{1 \leq i \leq n}$. Show that $\sup\{|F_n(x) - F(x)| : x \in \mathbb{R}\} \to 0$ w.p. 1.

(**Hint:** First prove an extension of Polyā's theorem (Lemma 8.2.6) to the multivariate case.)