7

Laws of Large Numbers

7.1 Definitions, the Borel-Cantelli Lemmas, and the Kolmogorov Inequality

We again turn our discussion to sequences of independent random variables. Let $\xi_1, \xi_2, \ldots$ be a sequence of random variables with finite expectations $m_n = \mathrm{E}\xi_n$, $n = 1, 2, \ldots$. Let $\zeta_n = (\xi_1 + \ldots + \xi_n)/n$ and $\overline{\zeta}_n = (m_1 + \ldots + m_n)/n$.

Definition 7.1. *The sequence of random variables ξ_n satisfies the Law of Large Numbers if $\zeta_n - \overline{\zeta}_n$ converges to zero in probability, that is $\mathrm{P}(|\zeta_n - \overline{\zeta}_n| > \varepsilon) \to 0$ as $n \to \infty$ for any $\varepsilon > 0$.*

It satisfies the Strong Law of Large Numbers if $\zeta_n - \overline{\zeta}_n$ converges to zero almost surely, that is $\lim_{n\to\infty}(\zeta_n - \overline{\zeta}_n) = 0$ for almost all ω.

If the random variables ξ_n are independent, and if $\mathrm{Var}(\xi_i) \leq V < \infty$, then by the Chebyshev Inequality, the Law of Large Numbers holds:

$$\mathrm{P}(|\zeta_n - \overline{\zeta}_n| > \varepsilon) = \mathrm{P}(|\xi_1 + \ldots + \xi_n - (m_1 + \ldots + m_n)| \geq \varepsilon n)$$
$$\leq \frac{\mathrm{Var}(\xi_1 + \ldots + \xi_n)}{\varepsilon^2 n^2} \leq \frac{V}{\varepsilon^2 n},$$

which tends to zero as $n \to \infty$. There is a stronger statement due to Khinchin:

Theorem 7.2 (Khinchin). *A sequence ξ_n of independent identically distributed random variables with finite mathematical expectation satisfies the Law of Large Numbers.*

Historically, the Khinchin Theorem was one of the first theorems related to the Law of Large Numbers. We shall not prove it now, but obtain it later as a consequence of the Birkhoff Ergodic Theorem, which will be discussed in Chap. 16.

We shall need the following three general statements.

L. Koralov and Y.G. Sinai, *Theory of Probability and Random Processes*,
Universitext, DOI 10.1007/978-3-540-68829-7_7,

Lemma 7.3 (First Borel-Cantelli Lemma). *Let* $(\Omega, \mathcal{F}, \mathrm{P})$ *be a probability space and* $\{A_n\}$ *an infinite sequence of events,* $A_n \subseteq \Omega$*, such that* $\sum_{n=1}^{\infty} \mathrm{P}(A_n) < \infty$*. Define*

$$A = \{\omega : \textit{there is an infinite sequence } n_i(\omega) \textit{ such that } \omega \in A_{n_i}, i = 1, 2, \ldots\}.$$

Then $\mathrm{P}(A) = 0$*.*

Proof. Clearly,

$$A = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n.$$

Then $\mathrm{P}(A) \leq \mathrm{P}(\bigcup_{n=k}^{\infty} A_n) \leq \sum_{n=k}^{\infty} \mathrm{P}(A_n) \rightarrow 0$ as $k \rightarrow \infty$. □

Lemma 7.4 (Second Borel-Cantelli Lemma). *Let* A_n *be an infinite sequence of independent events with* $\sum_{n=1}^{\infty} \mathrm{P}(A_n) = \infty$*, and let*

$$A = \{\omega : \textit{there is an infinite sequence } n_i(\omega) \textit{ such that } \omega \in A_{n_i}, i = 1, 2, \ldots\}.$$

Then $\mathrm{P}(A) = 1$*.*

Proof. We have $\Omega \backslash A = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} (\Omega \backslash A_n)$. Then

$$\mathrm{P}(\Omega \backslash A) \leq \sum_{k=1}^{\infty} \mathrm{P}(\bigcap_{n=k}^{\infty} (\Omega \backslash A_n))$$

for any n. By the independence of A_n we have the independence of $\Omega \backslash A_n$, and therefore

$$\mathrm{P}(\bigcap_{n=k}^{\infty} (\Omega \backslash A_n)) = \prod_{n=k}^{\infty} (1 - \mathrm{P}(A_n)).$$

The fact that $\sum_{n=k}^{\infty} \mathrm{P}(A_n) = \infty$ for any k implies that $\prod_{n=k}^{\infty} (1 - \mathrm{P}(A_n)) = 0$ (see Problem 1). □

Theorem 7.5 (Kolmogorov Inequality). *Let* $\xi_1, \xi_2, \ldots$ *be a sequence of independent random variables which have finite mathematical expectations and variances,* $m_i = \mathrm{E}\xi_i$*,* $V_i = \mathrm{Var}(\xi_i)$*. Then*

$$\mathrm{P}(\max_{1 \leq k \leq n} |(\xi_1 + \ldots + \xi_k) - (m_1 + \ldots + m_k)| \geq t) \leq \frac{1}{t^2} \sum_{i=1}^{n} V_i.$$

Proof. We consider the events $C_k = \{\omega : |(\xi_1 + \ldots + \xi_i) - (m_1 + \ldots + m_i)| < t$ for $1 \leq i < k, |(\xi_1 + \ldots + \xi_k) - (m_1 + \ldots + m_k)| \geq t\}$, $C = \bigcup_{k=1}^{n} C_k$. It is clear that C is the event whose probability is estimated in the Kolmogorov Inequality, and that C_k are pair-wise disjoint. Thus

$$\sum_{i=1}^{n} V_i = \mathrm{Var}(\xi_1 + \ldots + \xi_n) = \int_{\Omega} ((\xi_1 + \ldots + \xi_n) - (m_1 + \ldots + m_n))^2 d\mathrm{P} \geq$$

$$\sum_{k=1}^{n} \int_{C_k} ((\xi_1 + \ldots + \xi_n) - (m_1 + \ldots + m_n))^2 d\mathrm{P} =$$

$$\sum_{k=1}^{n} [\int_{C_k} ((\xi_1 + \ldots + \xi_k) - (m_1 + \ldots + m_k))^2 d\mathrm{P} +$$

$$2 \int_{C_k} ((\xi_1 + \ldots + \xi_k) - (m_1 + \ldots + m_k))((\xi_{k+1} + \ldots + \xi_n) - (m_{k+1} + \ldots + m_n)) d\mathrm{P} +$$

$$\int_{C_k} ((\xi_{k+1} + \ldots + \xi_n) - (m_{k+1} + \ldots + m_n))^2 d\mathrm{P}].$$

The last integral on the right-hand side is non-negative. Most importantly, the middle integral is equal to zero. Indeed, by Lemma 4.15, the random variables

$$\eta_1 = ((\xi_1 + \ldots + \xi_k) - (m_1 + \ldots + m_k))\chi_{C_k}$$

and

$$\eta_2 = (\xi_{k+1} + \ldots + \xi_n) - (m_{k+1} + \ldots + m_n)$$

are independent. By Theorem 4.8, the expectation of their product is equal to the product of the expectations. Thus, the middle integral is equal to

$$\mathrm{E}(\eta_1 \eta_2) = \mathrm{E}\eta_1 \mathrm{E}\eta_2 = 0.$$

Therefore,

$$\sum_{i=1}^{n} V_i \geq \sum_{k=1}^{n} \int_{C_k} ((\xi_1 + \ldots + \xi_k) - (m_1 + \ldots + m_k))^2 d\mathrm{P} \geq$$

$$t^2 \sum_{k=1}^{n} \mathrm{P}(C_k) = t^2 \mathrm{P}(C).$$

That is $\mathrm{P}(C) \leq \frac{1}{t^2} \sum_{i=1}^{n} V_i$. □

7.2 Kolmogorov Theorems on the Strong Law of Large Numbers

Theorem 7.6 (First Kolmogorov Theorem). *A sequence of independent random variables ξ_i, such that $\sum_{i=1}^{\infty} \mathrm{Var}(\xi_i)/i^2 < \infty$, satisfies the Strong Law of Large Numbers.*

Proof. Without loss of generality we may assume that $m_i = \mathrm{E}\xi_i = 0$ for all i. Otherwise we could define a new sequence of random variables $\xi_i' = \xi_i - m_i$. We need to show that $\zeta_n = (\xi_1 + \ldots + \xi_n)/n \to 0$ almost surely. Let $\varepsilon > 0$, and consider the event

$$B(\varepsilon) = \{\omega : \text{there is } N = N(\omega) \text{ such that for all } n \geq N(\omega) \text{ we have } |\zeta_n| < \varepsilon\}.$$

Clearly

$$B(\varepsilon) = \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{\omega : |\zeta_n| < \varepsilon\}.$$

Let

$$B_k(\varepsilon) = \{\omega : \max_{2^{k-1} \leq n < 2^k} |\zeta_n| \geq \varepsilon\}.$$

By the Kolmogorov Inequality,

$$\mathrm{P}(B_k(\varepsilon)) = \mathrm{P}(\max_{2^{k-1} \leq n < 2^k} \frac{1}{n} |\sum_{i=1}^{n} \xi_i| \geq \varepsilon|) \leq$$

$$\mathrm{P}(\max_{2^{k-1} \leq n < 2^k} |\sum_{i=1}^{n} \xi_i| \geq \varepsilon 2^{k-1}) \leq$$

$$\mathrm{P}(\max_{1 \leq n < 2^k} |\sum_{i=1}^{n} \xi_i| \geq \varepsilon 2^{k-1}) \leq \frac{1}{\varepsilon^2 2^{2k-2}} \sum_{i=1}^{2^k} \mathrm{Var}(\xi_i).$$

Therefore,

$$\sum_{k=1}^{\infty} \mathrm{P}(B_k(\varepsilon)) \leq \sum_{k=1}^{\infty} \frac{1}{\varepsilon^2 2^{2k-2}} \sum_{i=1}^{2^k} \mathrm{Var}(\xi_i) =$$

$$\frac{1}{\varepsilon^2} \sum_{i=1}^{\infty} \mathrm{Var}(\xi_i) \sum_{k \geq [\log_2 i]} \frac{1}{2^{2k-2}} \leq \frac{c}{\varepsilon^2} \sum_{i=1}^{\infty} \frac{\mathrm{Var}(\xi_i)}{i^2} < \infty,$$

where c is some constant. By the First Borel-Cantelli Lemma, for almost every ω there exists an integer $k_0 = k_0(\omega)$ such that $\max_{2^{k-1} \leq n \leq 2^k} |\zeta_n| < \varepsilon$ for all $k \geq k_0$. Therefore $\mathrm{P}(B(\varepsilon)) = 1$ for any $\varepsilon > 0$. In particular $\mathrm{P}(B(\frac{1}{m})) = 1$ and $\mathrm{P}(\bigcap_m B(\frac{1}{m})) = 1$. But if $\omega \in \bigcap_m B(\frac{1}{m})$, then for any m there exists $N = N(\omega, m)$ such that for all $n \geq N(\omega, m)$ we have $|\zeta_n| < \frac{1}{m}$. In other words, $\lim_{n \to \infty} \zeta_n = 0$ for such ω. □

Theorem 7.7 (Second Kolmogorov Theorem). *A sequence ξ_i of independent identically distributed random variables with finite mathematical expectation $m = \mathrm{E}\xi_i$ satisfies the Strong Law of Large Numbers.*

This theorem follows from the Birkhoff Ergodic Theorem, which is discussed in Chap. 16. For this reason we do not provide its proof now.

The Law of Large Numbers, as well as the Strong Law of Large Numbers, is related to theorems known as Ergodic Theorems. These theorems give general conditions under which the averages of random variables have a limit.

Both Laws of Large Numbers state that for a sequence of random variables ξ_n, the average $\frac{1}{n}\sum_{i=1}^n \xi_i$ is close to its mathematical expectation, and therefore does not depend asymptotically on ω, i.e., it is not random. In other words, deterministic regularity appears with high probability in long series of random variables.

Let c be a constant and define

$$\xi^c(\omega) = \begin{cases} \xi(\omega) & \text{if } |\xi(\omega)| \leq c, \\ 0 & \text{if } |\xi(\omega)| > c. \end{cases}$$

Theorem 7.8 (Three Series Theorem). *Let ξ_i be a sequence of independent random variables. If for some $c > 0$ each of the three series*

$$\sum_{i=1}^{\infty} \mathrm{E}\xi_i^c, \qquad \sum_{i=1}^{\infty} \mathrm{Var}(\xi_i^c), \qquad \sum_{i=1}^{\infty} \mathrm{P}(|\xi_i| \geq c)$$

converges, then the series $\sum_{i=1}^\infty \xi_i$ converges almost surely.

Conversely, if the series $\sum_{i=1}^\infty \xi_i$ converges almost surely, then each of the three series above also converges for each $c > 0$.

Proof. We'll only prove the direct statement, leaving the converse as an exercise for the reader.

We first establish the almost sure convergence of the series $\sum_{i=1}^\infty (\xi_i^c - \mathrm{E}\xi_i^c)$. Let $S_n = \sum_{i=1}^n (\xi_i^c - \mathrm{E}\xi_i^c)$. Then, by the Kolmogorov Inequality, for any $\varepsilon > 0$

$$\mathrm{P}(\sup_{i \geq 1} |S_{n+i} - S_n| \geq \varepsilon) = \lim_{N\to\infty} \mathrm{P}(\max_{1 \leq i \leq N} |S_{n+i} - S_n| \geq \varepsilon) \leq$$

$$\lim_{N\to\infty} \frac{\sum_{i=n+1}^{n+N} \mathrm{E}(\xi_i^c)^2}{\varepsilon^2} = \frac{\sum_{i=n+1}^{\infty} \mathrm{E}(\xi_i^c)^2}{\varepsilon^2}.$$

The right-hand side can be made arbitrarily small by choosing n large enough. Therefore

$$\lim_{n\to\infty} \mathrm{P}(\sup_{i \geq 1} |S_{n+i} - S_n| \geq \varepsilon) = 0.$$

Hence the sequence S_n is fundamental almost surely. Otherwise a set of positive measure would exist where $\sup_{i \geq 1} |S_{n+i} - S_n| \geq \varepsilon$ for some $\varepsilon > 0$. We have therefore proved that the series $\sum_{i=1}^\infty (\xi_i^c - \mathrm{E}\xi_i^c)$ converges almost surely. By the hypothesis, the series $\sum_{i=1}^\infty \mathrm{E}\xi_i^c$ converges almost surely. Therefore $\sum_{i=1}^\infty \xi_i^c$ converges almost surely.

Since $\sum_{i=1}^\infty \mathrm{P}(|\xi_i| \geq c) < \infty$ almost surely, the First Borel-Cantelli Lemma implies that $\mathrm{P}(\{\omega : |\xi_i| \geq c \text{ for infinitely many } i\}) = 0$. Therefore, $\xi_i^c = \xi_i$ for all but finitely many i with probability one. Thus the series $\sum_{i=1}^\infty \xi_i$ also converges almost surely. □

7.3 Problems

1. Let $y_1, y_2, \ldots$ be a sequence such that $0 \le y_n \le 1$ for all n, and $\sum_{n=1}^{\infty} y_n = \infty$. Prove that $\prod_{n=1}^{\infty}(1 - y_n) = 0$.
2. Let $\xi_1, \xi_2, \ldots$ be independent identically distributed random variables. Prove that $\sup_n \xi_n = \infty$ almost surely if and only if $\mathrm{P}(\xi_1 > A) > 0$ for every A.
3. Let $\xi_1, \xi_2, \ldots$ be a sequence of random variables defined on the same probability space. Prove that there exists a numeric sequence $c_1, c_2, \ldots$ such that $\xi_n / c_n \to 0$ almost surely as $n \to \infty$.
4. For each $\gamma > 2$, define the set $D_\gamma \subset [0,1]$ as follows: $x \in D_\gamma$ if there is $K_\gamma(x) > 0$ such that for each $q \in \mathbb{N}$

$$\min_{p \in \mathbb{N}} |x - \frac{p}{q}| \ge \frac{K_\gamma(x)}{q^\gamma}.$$

(The numbers x which satisfy this inequality for some $\gamma > 2$, $K_\gamma(x) > 0$, and all $q \in \mathbb{N}$ are called Diophantine.) Prove that $\lambda(D_\gamma) = 1$, where λ is the Lebesgue measure on $([0,1], \mathcal{B}([0,1]))$.
5. Let $\xi_1, \ldots, \xi_n$ be a sequence of n independent random variables, each ξ_i having a symmetric distribution. That is, $\mathrm{P}(\xi_i \in A) = \mathrm{P}(\xi_i \in -A)$ for any Borel set $A \subseteq \mathbb{R}$. Assume that $\mathrm{E}\xi_i^{2m} < \infty$, $i = 1, 2, \ldots, n$. Prove the stronger version of the Kolmogorov Inequality:

$$\mathrm{P}(\max_{1 \le k \le n} |\xi_1 + \ldots + \xi_k| \ge t) \le \frac{\mathrm{E}(\xi_1 + \ldots + \xi_n)^{2m}}{t^{2m}}.$$

6. Let $\xi_1, \xi_2, \ldots$ be independent random variables with non-negative values. Prove that the series $\sum_{i=1}^{\infty} \xi_i$ converges almost surely if and only if

$$\sum_{i=1}^{\infty} \mathrm{E}\frac{\xi_i}{1 + \xi_i} < \infty.$$

7. Let $\xi_1, \xi_2, \ldots$ be a sequence of independent identically distributed random variables with uniform distribution on $[0,1]$. Prove that the limit

$$\lim_{n \to \infty} \sqrt[n]{\xi_1 \cdot \ldots \cdot \xi_n}$$

exists with probability one. Find its value.
8. Let $\xi_1, \xi_2, \ldots$ be a sequence of independent random variables, $\mathrm{P}(\xi_i = 2^i) = 1/2^i$, $\mathrm{P}(\xi_i = 0) = 1 - 1/2^i$, $i \ge 1$. Find the almost sure value of the limit $\lim_{n \to \infty} (\xi_1 + \ldots + \xi_n)/n$.
9. Let $\xi_1, \xi_2, \ldots$ be a sequence of independent identically distributed random variables for which $\mathrm{E}\xi_i = 0$ and $\mathrm{E}\xi_i^2 = V < \infty$. Prove that for any $\gamma > 1/2$, the series $\sum_{i \ge 1} \xi_i / i^\gamma$ converges almost surely.

10. Let $\xi_1, \xi_2, \ldots$ be independent random variables uniformly distributed on the interval $[-1, 1]$. Let $a_1, a_2, \ldots$ be a sequence of real numbers such that $\sum_{n=1}^{\infty} a_n^2$ converges. Prove that the series $\sum_{n=1}^{\infty} a_n \xi_n$ converges almost surely.