

Chapter 3. Large Number Laws for Sequences of Random Variables

Let X_1, X_2, \dots be a sequence of random variables, and, for every n , define the empirical average of X_1, \dots, X_n as

$$\bar{X}_n := \frac{1}{n} \sum_1^n X_i.$$

In this chapter we shall study large number laws for the sequence of empirical averages. We say that a *weak law* of large numbers holds for the empirical averages if there is a deterministic sequence $\{b_n\}$ such that

$$(1) \quad \bar{X}_n - b_n \rightarrow 0 \quad \text{in probability as } n \rightarrow \infty.$$

We say that a *strong law* of large numbers holds if

$$(2) \quad \bar{X}_n - b_n \rightarrow 0 \quad \text{almost surely as } n \rightarrow \infty.$$

In the classical statements of large number laws, the random variables have an identical and finite mean $\mu = E[X_i]$, and $b_n = \mu$ for every n . This is the case, in particular, when the random variables have identical laws.

In this chapter, we state and prove some basic strong and weak laws. Our interest lies as much with the techniques of proof as with the results. We shall see that there are only two simple tools, Tchebyshev's inequality and the Borel-Cantelli lemma, but by skillfully combining them, we can derive some deep results. Then, in section D, we study the different but related question of convergence of a random series $\sum_1^\infty Z_i$ for a sequence of independent random variables $\{Z_n\}$. We state Kolmogorov's three series theorem giving necessary and sufficient conditions for convergence and prove sufficiency but not necessity. A second proof of the strong law of large numbers is then presented

Throughout the chapter we use the standard abbreviation *i.i.d.* to denote *independent and identically distributed*.

A. The Weak Law of Large Numbers

We have already proved the weak law for empirical averages of Bernoulli random variables in Chapter 2, section E, by using Chernoff's inequality. In this section we consider a more general case.

Theorem A.1 (Weak law in the uncorrelated case) Let $\{X_n\}$ be a sequence of uncorrelated random variables with common mean μ and common variance $\sigma^2 < \infty$; (For every i , $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$.) Then

$$(3) \quad (P) \lim_{n \rightarrow \infty} \bar{X} = \mu.$$

Proof: To simplify notation we shall assume that $\mu = 0$. This entails no loss of generality, since (3) is equivalent to $(P) \lim(1/n) \sum_1^n (X_i - \mu) = 0$ and $X_i - \mu$ has mean zero.

From Tchebyshev's inequality and equation (5) of Proposition C.4 in Chapter 2, we obtain,

$$(4) \quad \mathbb{P}(|\bar{X}| > \epsilon) \leq \frac{\text{Var}(\bar{X})}{\epsilon^2} \leq \frac{\sum_1^n \text{Var}(X_i)}{n^2 \epsilon^2} = \frac{\sigma^2}{n \epsilon^2}.$$

Thus, for every $\epsilon > 0$, $\mathbb{P}(|\bar{X}| > \epsilon) \rightarrow 0$ as $n \rightarrow \infty$, proving convergence in probability. \diamond

The argument we used in the proof extends easily to more general situations. For example, let X_1, X_2, \dots be uncorrelated with means respective means μ_1, μ_2, \dots , and assume that $\sup_n \text{Var}(X_n) < \infty$. Then

$$(P) \lim_{n \rightarrow \infty} \bar{X}_n - \frac{1}{n} \sum_1^n \mu_i = 0.$$

We emphasize the simple bound (4) based on Tchebyshev's inequality for the probability of deviation of the empirical mean from the true mean. For the case of uncorrelated random variables with arbitrary means and common variance σ^2 , we restate it as

$$(5) \quad \mathbb{P}\left(|\bar{X} - \frac{1}{n} \sum_1^n \mu_i| > \epsilon\right) \leq \frac{\sigma^2}{n \epsilon^2}.$$

This bound is ultimately the basis for all our proofs of large number laws.

B. The Borel-Cantelli Lemma

To pass from weak laws to strong laws of large numbers, we need a tool for for passing from convergence in probability to almost sure convergence. The Borel-Cantelli lemma serves precisely this function. Actually, the Borel-Cantelli lemma has two parts, the second of which, concerning independent events, is not directly relevant to our treatment of large number laws. Nevertheless, we include its statement and proof.

Let $\{A_n\}$ be a countable sequence of events, and define the event

$$\begin{aligned} & \{ \{A_n\} \text{ occurs infinitely often} \} \\ & := \{ \omega \mid \exists \text{ a subsequence } \{n_k\} \text{ (depending on } \omega) \text{ such that } \omega \in A_{n_k}, \forall k \}. \end{aligned}$$

We shall use the abbreviation $\{ \{A_n\} \text{ i.o.} \}$ for $\{ \{A_n\} \text{ occurs infinitely often} \}$. An equivalent definition is

$$\{ \{A_n\} \text{ i.o.} \} = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n.$$

In set theory $\{\{A_n\} \text{ i.o.}\}$ is denoted $\limsup A_n$. The complementary set construction to the limsup is

$$\liminf A_n := \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n.$$

Notice that

$$\{\{A_n \text{ occurs finitely often}\}\} := \{\{A_n \text{ i.o.}\}\}^c = \liminf A_n^c.$$

Example. This example shows the immediate relevance of the $\limsup A_n$ construction to questions of convergence. Let X_1, X_2, \dots and Y be random variables. By the statement, “ $\lim X_n(\omega) \neq Y(\omega)$ ”, we mean either $\lim X_n(\omega)$ does not exist, or the limit exists and is not equal to $Y(\omega)$. Now suppose that $\lim X_n(\omega) \neq Y(\omega)$. Then given any $\epsilon > 0$ there is a subsequence $n_1 < n_2 < \dots$ along which $|X_{n_k}(\omega) - Y(\omega)| > \epsilon$, for every k . Conversely, if there exists such an ϵ and such a subsequence, then $\lim X_n(\omega) \neq Y(\omega)$. Thus

$$(1) \quad \{\omega \mid \lim X_n(\omega) \neq Y(\omega)\} = \bigcup_{j=1}^{\infty} \{|X_n - Y| > \frac{1}{j} \text{ i.o.}\}.$$

Similarly, let $\epsilon_n \downarrow 0$ be a sequence of positive integers decreasing down to 0. Then

$$(2) \quad \{\lim X_n \neq Y\} \subset \{|X_n - Y| > \epsilon_n \text{ i.o.}\}. \quad \diamond$$

The Borel-Cantelli Lemma gives simple conditions for concluding when the probability of $\{A_n \text{ i.o.}\}$ is 0 or 1.

Proposition B.1 Borel-Cantelli Lemma. Let $\{A_n\}$ be a sequence of events.

(a). If $\sum_1^{\infty} \mathbb{P}(A_n) < \infty$ then

$$\mathbb{P}(\{A_n\} \text{ i.o.}) = 0.$$

(b) If the events A_1, A_2, \dots are independent and $\sum_1^{\infty} \mathbb{P}(A_n) = \infty$, then

$$\mathbb{P}(\{A_n\} \text{ i.o.}) = 1.$$

Proof: The proof of (a) uses the subadditivity of \mathbb{P} . For every m ,

$$\begin{aligned} \mathbb{P}(\{A_n\} \text{ i.o.}) &= \mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n\right) \\ &\leq \mathbb{P}\left(\bigcup_{n=m}^{\infty} A_n\right) \\ &\leq \sum_{n=m}^{\infty} \mathbb{P}(A_n). \end{aligned}$$

The condition $\sum_1^{\infty} \mathbb{P}(A_n) < \infty$ thus implies that

$$\mathbb{P}(\{A_n\} \text{ i.o.}) \leq \lim_{m \rightarrow \infty} \sum_{n=m}^{\infty} \mathbb{P}(A_n) = 0.$$

The proof of (b) is more involved. We shall do it by showing that $\mathbb{P}(\{\{A_n\} \text{ i.o.}\}^c) = 0$. First notice that because of the continuity of the probability measure under limits of increasing and decreasing sets

$$(3) \quad \mathbb{P}(\{\{A_n\} \text{ i.o.}\}^c) = \mathbb{P}\left(\bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} A_n^c\right) = \lim_{m \rightarrow \infty} \lim_{N \rightarrow \infty} \mathbb{P}\left(\bigcap_{n=m}^N A_n^c\right).$$

Now we use the independence of A_1, A_2, \dots and the inequality $\log(1-x) < -x$, $0 < x < 1$, to obtain

$$\begin{aligned} \log \mathbb{P}\left(\bigcap_{n=m}^N A_n^c\right) &= \log \prod_{n=m}^N \mathbb{P}(A_n^c) = \sum_{n=m}^N \log \mathbb{P}(A_n^c) \\ &= \sum_{n=m}^N \log(1 - \mathbb{P}(A_n)) \leq - \sum_{n=m}^N \mathbb{P}(A_n). \end{aligned}$$

The assumption that $\sum_1^{\infty} \mathbb{P}(A_n) = \infty$ therefore implies that

$$\lim_{N \rightarrow \infty} \log \mathbb{P}\left(\bigcap_{n=m}^N A_n^c\right) \leq - \lim_{N \rightarrow \infty} \sum_{n=m}^N \mathbb{P}(A_n) = -\infty,$$

and thus $\lim_{N \rightarrow \infty} \mathbb{P}\left(\bigcap_{n=m}^N A_n^c\right) = 0$ for every $m \geq 1$. Statement (b) follows from this conclusion and equation (3). \diamond

We give a preliminary application of the Borel-Cantelli lemma, part (a), to a result on the relationship between almost sure convergence and convergence in probability.

Proposition B.2

- (a) If X_n converges to Y in probability then there exists a subsequence $\{X_{n_k}\}$ such that (a.s.) $\lim X_{n_k} = Y$.
- (b) Suppose that $\{X_n\}$ is a sequence of random variables such that every subsequence $\{X_{n_k}\}$ contains a sub-subsequence converging almost surely to the random variable Y . Then $(P) \lim X_n = Y$.

Proof: It suffices to consider the case in which $Y \equiv 0$. Thus, suppose that $(P) \lim X_n = 0$. Choose a subsequence $\{n_k\}$ of the integers such that for every k ,

$$\mathbb{P}(|X_{n_k}| > \frac{1}{k^2}) < \frac{1}{k^2}.$$

Since $\sum_k k^{-2} < \infty$, the Borel-Cantelli lemma implies that

$$\mathbb{P}(|X_{n_k}| > \frac{1}{k^2} \text{ i.o.}) = 0.$$

The inclusion (2) then implies that

$$\mathbb{P}(\lim_{k \rightarrow \infty} X_{n_k} \neq 0) = 0,$$

thus proving (a).

A proof by contradiction, using the fact that almost sure convergence implies convergence in probability, proves part (b). \diamond

C. Strong Large Number Laws

The technique used here to obtain strong laws rests on the following observation.

Lemma C.1 Let $\{Y_k\}$ be sequence of random variables, and suppose that for every $\epsilon > 0$, $\mathbb{P}(|Y_k| > \epsilon \text{ i.o.}) = 0$. Then (a.s.) $\lim Y_k = 0$.

Proof: By using identity (1) of section B and the hypothesis of the lemma,

$$\begin{aligned} \mathbb{P}(\lim_{k \rightarrow \infty} Y_k \neq 0) &= \mathbb{P}\left(\bigcup_1^{\infty} \{|Y_k| > \frac{1}{j} \text{ i.o.}\}\right) \\ &\leq \sum_{j=1}^{\infty} \mathbb{P}(|Y_k| > \frac{1}{j} \text{ i.o.}) = 0. \quad \diamond \end{aligned}$$

The use of the Lemma C.1 is simple. To show (a.s.) $\lim Y_k = 0$, we first employ Tchebyshev's inequality to bound $\mathbb{P}(|Y_k| > \epsilon)$. If for every $\epsilon > 0$, $\sum_1^{\infty} \mathbb{P}(|Y_k| > \epsilon) < \infty$, then the Borel-Cantelli lemma implies that for every $\epsilon > 0$

$$\mathbb{P}(|Y_k| > \epsilon \text{ i.o.}) = 0.$$

By Lemma C.1, the almost sure convergence of $\{Y_n\}$ to 0 as $n \rightarrow \infty$ follows.

Here are two simple applications of the strategy. The first is a proof of the strong law of large numbers for Bernoulli random variables, stated previously in Chapters 1 and 2.

Corollary C.1 Let X_1, X_2, \dots be i.i.d. Bernoulli random variables with distribution $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 0) = 1 - p$. Then (a.s.) $\lim \bar{X}_n = p$.

PROOF: From Chernoff's inequality (equation (5) in section E of chapter 2),

$$\sum_1^{\infty} \mathbb{P}(|\bar{X}_n - p| > \epsilon) \leq \sum_1^{\infty} 2e^{-2n\epsilon^2} < \infty,$$

for every $\epsilon > 0$. The strong law follows immediately from Lemma C.1 and the Borel-Cantelli lemma. \diamond

Corollary C.2 Let X_1, X_2, \dots be uncorrelated random variables with identical mean μ and identical variance $\sigma^2 < \infty$. Then (a.s.) $\lim_{k \rightarrow \infty} \bar{X}_{k^2} = \mu$

PROOF: By the inequality (5) of section A,

$$\mathbb{P}(|\bar{X}_{k^2} - p| > \epsilon) \leq \frac{\sigma^2}{k^2 \epsilon^2}.$$

Since $\{k^{-2}\}_{k \geq 1}$ is a summable sequence, the corollary follows from Lemma C.1 and Messrs. Borel and Cantelli. \diamond

We can go much beyond these naive applications of Borel-Cantelli. The first result we prove is the strong law for uncorrelated random variables. The idea is to first extract a subsequence of $\{\bar{X}\}$ that converges almost-surely. This we have already done in Corollary 2. Then we show \bar{X}_n at values of n between successive points of the subsequence, again using the Borel-Cantelli machine.

Theorem C.1 Let $\{X_n\}$ be a sequence of uncorrelated random variables with common mean μ and common variance $\sigma^2 < \infty$. Then (a.s.) $\lim \bar{X} = \mu$.

PROOF: Without loss of generality we may assume that $\mu = 0$; otherwise we just replace X_i by $X_i - \mu$. We know already from Corollary C.2 that

$$(1) \quad (\text{a.s.}) \lim \bar{X}_{k^2} = 0.$$

Now, consider n and k such that $k^2 < n < (k+1)^2$. We wish to compare \bar{X}_n to \bar{X}_{k^2} . For this purpose, define

$$W_k = \frac{1}{k^2} \max\left\{ \left| \sum_{k^2+1}^n X_i \right| ; k^2 < n < (k+1)^2 \right\}.$$

For $k^2 < n < (k+1)^2$, we have

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_1^{k^2} X_i + \frac{1}{n} \sum_{k^2+1}^n X_i \\ &= \frac{k^2}{n} \bar{X}_{k^2} + \frac{k^2}{n} \left(\frac{1}{k^2} \sum_{k^2+1}^n X_i \right), \end{aligned}$$

from which it follows that

$$(2) \quad |\bar{X}| \leq |\bar{X}_{k^2}| + W_k.$$

If we can show that $(\text{a.s.}) \lim W_k = 0$, the strong law of large numbers will clearly follow from (1) and (2).

To show $(\text{a.s.}) \lim W_k = 0$, it suffices to prove $\mathbb{P}(W_k > \epsilon \text{ i.o.}) = 0$ for any $\epsilon > 0$, because of Lemma C.1. By definition of W_k ,

$$(3) \quad \{W_k > \epsilon\} = \bigcup_{n=k^2+1}^{(k+1)^2-1} \left\{ \left| \sum_{k^2+1}^n X_i \right| > k^2 \epsilon \right\}.$$

Notice that for $k^2 < n < (k+1)^2$,

$$(4) \quad \text{Var}\left(\sum_{k^2+1}^n X_i\right) = (n - k^2)\sigma^2 \leq 2k\sigma^2.$$

From (3), (4), and Tchebyshev's inequality

$$\mathbb{P}(W_k > \epsilon) \leq \sum_{k^2+1}^{(k+1)^2-1} \frac{2k\sigma^2}{k^4\epsilon^2} \leq \frac{(2k)^2\sigma^2}{k^4\epsilon^2} = \frac{4\sigma^2}{k^2\epsilon^2},$$

where the second inequality follows from the fact that the sum contains $2k$ terms. Since $\sum k^{-2} < \infty$, the Borel-Cantelli lemma implies that $\mathbb{P}(W_k > \epsilon \text{ i.o.}) = 0$, as desired. \diamond

The next step in pushing forward the strong law is to relax the assumption that second moments be bounded. This was first achieved by Kolmogorov, who dropped the second moment condition at the expense of assuming that the random variables be independent and identically distributed.

Theorem C.2 (The Strong Law of Large Numbers) Let $\{X_n\}$ be a sequence of independent, identically distributed, integrable random variables with mean μ . Then

$$(\text{a.s.}) \lim \bar{X} = \mu.$$

We will not present Kolmogorov's proof here. Instead, we will prove the following stronger theorem due to Etemadi. We follow the treatment in *Probability and Measure*, P. Billingsley.

Theorem C.3 Let $\{X_n\}$ be a sequence of identically distributed, *pairwise independent* random variables with finite mean μ . Then $(\text{a.s.}) \lim \bar{X} = \mu$.

The rest of this section will be devoted to the proof of this theorem. Again we will make use only of Tchebyshev's inequality and the Borel-Cantelli lemma. However, we must introduce another, important technique – truncation – and the concept of equivalent sequences.

Definition. Two sequences of random variables $\{U_n\}$ and $\{V_n\}$ are equivalent, written $\{U_n\} \sim \{V_n\}$, if $\sum_1^\infty \mathbb{P}(U_n \neq V_n) < \infty$.

It is immediate from the Borel-Cantelli lemma that if $\{U_n\} \sim \{V_n\}$,

$$\mathbb{P}(U_n \neq V_n \text{ i.o.}) = 0, \quad \text{or, equivalently,} \quad \mathbb{P}(\{U_n \neq V_n \text{ i.o.}\}^c) = 1.$$

In other words, because

$$(5) \quad \{U_n \neq V_n \text{ i.o.}\}^c = \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \{U_n = V_n\},$$

equivalent sequences $\{U_n\}$ and $\{V_n\}$ eventually coincide with probability one. Thus, in discussing tail events we may replace $\{U_n\}$ by $\{V_n\}$ and vice-versa. For example, we have

LEMMA C.2 Suppose $\{U_n\} \sim \{V_n\}$. Then $\{\lim \bar{U}_n \text{ exists}\} = \{\lim \bar{V}_n \text{ exists}\}$ and $\lim \bar{U}_n = \lim \bar{V}_n$, \mathbb{P} -almost surely on the set $\{\lim \bar{U}_n \text{ exists}\}$. In particular, if Z is a random variable, then (a.s.) $\lim \bar{U}_n = Z$ if and only if (a.s.) $\lim \bar{V}_n = Z$. Also, $(P) \lim \bar{U}_n = Z$ if and only if $(P) \lim \bar{V}_n = Z$.

The proof is left as an exercise. The claim about convergence in probability follows from the claim that almost sure limits of equivalent sequences are almost surely equal and the characterization of convergence in probability in part (b) of Proposition B.2. \diamond

Now let $\{X_n\}$ be a given sequence of pairwise independent, identically distributed random variables with finite mean μ . The first step in the proof of the strong law is to replace $\{X_n\}$ by an equivalent sequence $\{Y_n\}$ such that each Y_n is bounded. Then we can apply Tchebyshev's inequality to these bounded random variables and obtain a weak law of large numbers for the equivalent sequence. Finally, we pass from the weak law to the strong law for the equivalent sequence by using the Borel-Cantelli lemma along appropriate subsequences. The strong law for the original sequence $\{X_n\}$ then follows from Lemma C.2.

To construct an equivalent sequence, we truncate each X_n at level n ; that is, we define $Y_n = X_n \mathbf{1}_{\{|X_n| \leq n\}}$ for each positive integer n . This truncation produces an equivalent sequence precisely because we are assuming that the X_i 's have finite mean. Indeed, from the identity, $E[|X|] = \int_0^\infty \mathbb{P}(|X| > x) dx$, which is obtained by integration by parts, we get that

$$(6) \quad E[|X|] < \infty \quad \text{if and only if} \quad \sum_1^\infty \mathbb{P}(|X| > n) < \infty.$$

Thus, since the X_n , $n \geq 1$, are identically distributed,

$$\sum_1^\infty \mathbb{P}(X_n \neq Y_n) = \sum_1^\infty \mathbb{P}(|X_n| > n) = \sum_1^\infty \mathbb{P}(|X_1| > n) < \infty,$$

and it follows that $\{Y_n\} \sim \{X_n\}$. Note that the independence assumption is not used in this argument.

Now let $\mu_n = E[Y_n]$ and define $\bar{\mu}_n = n^{-1} \sum_1^n \mu_i$. Notice that

$$(7) \quad \lim_{n \rightarrow \infty} \mu_n = \mu \quad \text{and hence} \quad \lim_{n \rightarrow \infty} \bar{\mu}_n = \mu.$$

We shall prove a weak law for $\{Y_n\}$ using the assumption of pairwise independence of X_1, X_2, \dots . In what follows \bar{Y}_n is the empirical mean of Y_1, \dots, Y_n .

LEMMA C.3 If $\{X_n\}$ is a sequence of pairwise independent identically distributed random variables with finite mean μ , then

$$(8) \quad \mathbb{P}(|\bar{Y}_n - \bar{\mu}_n| > \epsilon) \leq \frac{1}{\epsilon^2 n} E[X_1^2 \mathbf{1}_{\{|X_1| \leq n\}}],$$

and hence $(P) \lim \bar{Y}_n = \mu$.

PROOF: Since X_1, X_2, \dots are pairwise independent, Y_1, Y_2, \dots are uncorrelated. Thus, using Proposition C.4 of Chapter 2, and the inequality $\text{Var}(Z) \leq E[Z^2]$,

$$\begin{aligned}
 \text{Var}(\bar{Y}_n) &= \frac{1}{n^2} \sum_1^n \text{Var}(Y_k) \\
 (9) \quad &\leq \frac{1}{n^2} \sum_1^n E[X_k^2 1_{\{|X_k| \leq k\}}] \\
 &\leq \frac{1}{n^2} (n E[X_1^2 1_{\{|X_1| \leq n\}}]) = \frac{1}{n} E[X_1^2 1_{\{|X_1| \leq n\}}].
 \end{aligned}$$

Inequality (8) then follows from (9) and Tchebyshev's inequality. An application of dominated convergence shows that the right hand side of (8) converges to 0 as $n \rightarrow \infty$, and hence that $(P) \lim \bar{Y}_n - \bar{\mu}_n = 0$. From (7), it follows that $(P) \lim \bar{Y}_n = \mu$. \diamond

We turn now to the completion of the proof of Etemadi's strong law. First note that it suffices to consider only the case in which the random variables are almost surely non-negative, i.e. $\mathbb{P}(X_n \geq 0) = 1$. To see why, write $X_n = X_n^+ - X_n^-$, where $X_n^+ = X_n 1_{\{X_n \geq 0\}}$ and $X_n^- = -X_n 1_{\{X_n < 0\}}$. The sequences $\{X_n^+\}$ and $\{X_n^-\}$ both satisfy the hypotheses of Theorem C.3 and consist of non-negative random variables. Since $\bar{X}_n = n^{-1} \sum_1^n X_i^+ - n^{-1} \sum_1^n X_i^-$, it suffices to establish Theorem C.3 for $\{X_n^+\}$ and $\{X_n^-\}$ separately in order to prove it for $\{X_n\}$. The assumption of positivity will be technically convenient, and we shall henceforth assume it is satisfied.

Since $\{Y_n\} \sim \{X_n\}$, we know from Lemma C.2 that to complete the proof of the strong law, it suffices to prove $(\text{a.s.}) \lim \bar{Y}_n = \mu$. The proof proceeds in two steps. Step 1 is the next lemma establishing almost sure convergence of a geometrically increasing subsequence. For positive numbers a , let $\lfloor a \rfloor$ denote the greatest integer less than or equal to a .

Lemma C.4 For every $a > 1$, we have that $(\text{a.s.}) \lim \bar{Y}_{\lfloor a^n \rfloor} = \mu$.

The second step is to deduce the strong law from Lemma C.4. Here is where we use the assumed non-negativity of the X_n .

Lemma C.5 If the random variables X_n are non-negative, the statement that $(\text{a.s.}) \lim \bar{Y}_{\lfloor a^n \rfloor} = \mu$ for every $a > 1$ implies $(\text{a.s.}) \lim \bar{Y}_n = \mu$.

We prove Lemma C.5 first, as it is easier. Thus, assume $(\text{a.s.}) \lim \bar{Y}_{\lfloor a^n \rfloor} = \mu$. Let $\lfloor a^n \rfloor < k < \lfloor a^{n+1} \rfloor$. Since the Y_k random variables are non-negative,

$$\frac{\lfloor a^n \rfloor}{\lfloor a^{n+1} \rfloor} \bar{Y}_{\lfloor a^n \rfloor} = \frac{1}{\lfloor a^{n+1} \rfloor} \sum_1^{\lfloor a^n \rfloor} Y_i < \frac{1}{k} \sum_1^k Y_i = \bar{Y}_k \leq \frac{1}{\lfloor a^n \rfloor} \sum_1^{\lfloor a^{n+1} \rfloor} Y_i = \frac{\lfloor a^{n+1} \rfloor}{\lfloor a^n \rfloor} \bar{Y}_{\lfloor a^{n+1} \rfloor}.$$

By letting $n \rightarrow \infty$ in this string of equalities and inequalities and using $\lim [a^{n+1}]/[a^n] = a$, it follows that for every $a > 1$,

$$\mathbb{P}\left(\frac{\mu}{a} \leq \liminf \bar{Y}_k \leq \limsup \bar{Y}_k \leq a\mu\right) = 1.$$

Letting $a \downarrow 1$ gives $\mathbb{P}(\lim \bar{Y}_k = \mu) = 1$, as desired. \diamond

Proof of Lemma C.4: we shall need the following simple, analytic fact. For each $a > 1$, there is a constant C_a such that

$$(10) \quad \sum_{n=1}^{\infty} \frac{1}{[a^n]} 1_{\{x \leq [a^n]\}} \leq \frac{C_a}{x} \quad \text{for all } x > 0.$$

Indeed, let $N_x = \min\{n \mid [a^n] \geq x\}$. Then

$$\sum_{n=1}^{\infty} \frac{1}{[a^n]} 1_{\{x \leq [a^n]\}} = \frac{1}{[a^{N_x}]} \sum_{n=N_x}^{\infty} \frac{[a^{N_x}]}{[a^n]} \leq \frac{1}{x} \sum_{n=N_x}^{\infty} \frac{a^{N_x}}{a^n - 1},$$

since $[a^{N_x}] \geq x$ and $a^n \geq [a^n] \geq a^n - 1$. A simple calculation shows that

$$\sum_{n=N_x}^{\infty} \frac{a^{N_x}}{a^n - 1} \leq \sum_{n=0}^{\infty} \frac{1}{a^n - a^{-1}} < \infty$$

for any N_x , which proves (10).

Now, using (8), applied at $[a^n]$ instead of n , and Tchebyshev's inequality

$$\mathbb{P}(|\bar{Y}_{[a^n]} - \bar{\mu}_{[a^n]}| > \epsilon) \leq \frac{1}{\epsilon^2} E[X_1^2] \frac{1}{[a^n]} 1_{\{|X_1| \leq [a^n]\}}.$$

Thus, from (10),

$$\begin{aligned} \sum_1^{\infty} \mathbb{P}(|\bar{Y}_{[a^n]} - \bar{\mu}_{[a^n]}| > \epsilon) &\leq \frac{1}{\epsilon^2} E[X_1^2] \sum_1^{\infty} \frac{1}{[a^n]} 1_{\{|X_1| \leq [a^n]\}} \\ &\leq \frac{1}{\epsilon^2} E[X_1^2 C_a |X_1|^{-1}] \\ &= \frac{C_a}{\epsilon^2} E[|X_1|] < \infty. \end{aligned}$$

The Borel-Cantelli lemma then implies that (a.s.) $\lim \bar{Y}_{[a^n]} - \bar{\mu}_{[a^n]} = 0$, thus completing the proof of Lemma C.4. \diamond

D. Convergence of infinite series of independent random variables and applications

Let $\{X_n\}$ be an infinite sequence of independent random variables. Throughout S_n will denote the partial sum $\sum_1^n X_i$. We say that $\sum_1^\infty X_i$ converges a.s. if $\lim_{N \rightarrow \infty} S_N$ exists and is finite almost-surely. We pose the question: under what conditions does $\sum_1^\infty X_i$ converge a.s.? An example of this problem is the random signs problem: if $\{c_n\}_1^\infty$ is a sequence of real numbers and if $\{\xi_n\}$ is a sequence of i.i.d. symmetric Bernoulli random variables ($\mathbb{P}(\xi_n = 1) = \mathbb{P}(\xi_n = -1) = 1/2$), what can we say about the convergence of $\sum_1^\infty c_n \xi_n$? In fact there is a necessary and sufficient condition for the a.s. convergence of $\sum_1^\infty X_i$.

Theorem D.1 (Kolmogorov's Three Series Theorem) Let $\{X_n\}$ be a sequence of independent random variables. Then $\sum_1^\infty X_i$ converges a.s. if and only if for some $a > 0$ all of the following three series converges:

- (i) $\sum_1^\infty \mathbb{P}(|X_i| \geq a)$;
- (ii) $\sum_1^\infty E[X_i \mathbf{1}_{\{|X_i| \leq a\}}]$;
- (iii) $\sum_1^\infty \text{Var}(X_i \mathbf{1}_{\{|X_i| \leq a\}})$.

It is actually true that convergence of $\sum_1^\infty X_i$ implies the convergence of the three series in (i)-(iii) for every $a > 0$, but for the converse, convergence of (i)-(iii) for any one $a > 0$ suffices. We shall prove the sufficiency part of this theorem only. It is an easy consequence of the following special case, which is important enough to state separately as a theorem.

Theorem D.2 If $\{X_n\}$ are independent random variables which all have 0 mean, then $\sum_1^\infty \text{Var}(X_i) < \infty$ implies that $\sum_1^\infty X_i$ converges a.s.

Notice that $E[(S_N - S_M)^2] = \sum_{M+1}^N \text{Var}(X_i)$ because $S_N - S_M = \sum_{M+1}^N X_i$ and the X_i are independent and have zero mean. The condition $\sum_1^\infty \text{Var}(X_i) < \infty$ then implies that $\lim_{M, N \rightarrow \infty} E[(S_N - S_M)^2] = 0$, or, in words, that the sequence $\{S_n\}_1^\infty$ is Cauchy in $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Hence, there is a random variable $Z \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ such that $E[(S_n - Z)^2] \rightarrow 0$ as $n \rightarrow \infty$. Theorem D.2 asserts that S_n converges to Z almost surely as well.

Theorem D.2 resolves the random signs problem posed above. If ξ_1, ξ_2, \dots are independent, symmetric Bernoulli random variables, and if c_1, c_2, \dots is a sequence of real numbers, then

$$(1) \quad \sum_1^\infty \text{Var}(c_i X_i) = \sum_1^\infty c_i^2$$

and hence convergence of this sum is sufficient for almost sure convergence of $\sum_1^\infty c_i X_i$. The "if" part of Theorem D.1 implies that convergence of (1) is also necessary for almost sure convergence of the random sum. For a particular example, consider the sequence

$c_n = 1/n$. Since $\sum_1^\infty n^{-2} < \infty$, we learn that $\sum_1^\infty \xi_n/n$ converges almost surely, even though it is almost surely *not* absolutely convergent.

We shall prove Theorem D.2 using an important inequality of Kolmogorov which is a remarkable strengthening of Tchebyshev's inequality for sums of independent random variables. Kolmogorov's inequality depends on the following simple observation concerning sums of random variables.

Lemma D.1 Let $\{X_n\}$ be a sequence of independent, zero mean random variables with finite variances. If $m < n$ and if $A \in \sigma\{X_1, \dots, X_m\}$ ($=\sigma$ -algebra generated by X_1, \dots, X_m), then

$$(2) \quad E[S_m^2 \mathbf{1}_A] \leq E[S_n^2 \mathbf{1}_A].$$

Proof. Observe that $S_n - S_m = \sum_{m+1}^n X_i$ is independent of X_1, \dots, X_m , and hence $E[(S_n - S_m)S_m \mathbf{1}_A] = E[S_n - S_m]E[S_m \mathbf{1}_A] = 0$. Thus

$$\begin{aligned} E[S_n^2 \mathbf{1}_A] &= E[(S_n - S_m)^2 \mathbf{1}_A] + 2E[(S_n - S_m)S_m \mathbf{1}_A] + E[S_m^2 \mathbf{1}_A] \\ &= E[(S_n - S_m)^2 \mathbf{1}_A] + E[S_m^2 \mathbf{1}_A], \end{aligned}$$

and inequality (2) follows. \diamond

Proposition D.1 (Kolmogorov's inequality) Let $\{X_n\}$ be a sequence of independent random variables with zero mean. Then

$$\mathbb{P}(\max_{1 \leq n \leq N} |S_n| \geq \lambda) \leq \frac{E[S_N^2]}{\lambda^2}.$$

Proof. Let A_k be the event $\{|S_n| < \lambda \text{ for } 1 \leq n < k, |S_k| \geq \lambda\}$. In words, A_k is the event that the first time n such that $|S_n|$ rises above λ is $n = k$. The sets A_1, \dots, A_N are disjoint and

$$\{\max_{1 \leq n \leq N} |S_n| \geq \lambda\} = \cup_1^N A_k.$$

Now for each k , $1 \leq k \leq N$, Lemma D.1 and the Markov inequality imply

$$\mathbb{P}(A_k) \leq \lambda^{-2} E[S_k^2 \mathbf{1}_{A_k}] \leq \lambda^{-2} E[S_N^2 \mathbf{1}_{A_k}].$$

Thus

$$\mathbb{P}(\max_{1 \leq n \leq N} |S_n| \geq \lambda) \leq \lambda^{-2} \sum_1^N E[S_N^2 \mathbf{1}_{A_k}] \leq \lambda^{-2} E[S_N^2].$$

In the last step we used the disjointness of the A_k several times. \diamond

Proof of Theorem D.2: Kolmogorov's inequality implies that

$$\mathbb{P}(\max_{M < k \leq N} |S_k - S_M| > \lambda) \leq \lambda^{-2} E[(S_N - S_M)^2] = \lambda^{-2} \sum_{M+1}^N \text{Var}(X_i).$$

Taking limits as $N \rightarrow \infty$,

$$\mathbb{P}\left(\sup_{M < k \leq \infty} |S_k - S_M| > \lambda\right) \leq \lambda^{-2} \sum_{M+1}^{\infty} \text{Var}(X_i)$$

Because $\sum_1^{\infty} \text{Var}(X_i) < \infty$, we conclude that $\sup_{M < k \leq \infty} |S_k - S_M|$ converges to 0 in probability. It follows that $\sup_{M < k, M < \ell} |S_k - S_\ell|$ converges to 0 in probability. But any decreasing sequence of random variables that converges to 0 in probability must converge to 0 almost surely as well. Hence $\sup_{M < k, M < \ell} |S_k - S_\ell|$ converges to 0 almost surely. This implies that the sequence $\{S_n\}_1^{\infty}$ is almost surely a Cauchy sequence. Hence $\lim_{n \rightarrow \infty} S_n = \sum_1^{\infty} X_i$ converges almost surely. \diamond

Proof of Theorem D.1, "only if" part: In the set up of Theorem D.1, let $Y_i = X_i \mathbf{1}_{|X_i| > a}$ and let $\mu_i = E[Y_i]$. Convergence of series (iii) implies by Theorem D.1 that $\sum_1^{\infty} (Y_i - \mu_i)$ converges almost surely. Convergence of series (ii) says precisely that $\sum_1^{\infty} \mu_i$ converges, and it follows that $\sum_1^{\infty} Y_i$ converges a.s. Now by convergence of series (i) and the Borel-Cantelli lemma $\mathbb{P}(X_i \neq Y_i \text{ i.o.}) = 0$ and hence convergence of $\sum_1^{\infty} Y_i$ implies that of $\sum_1^{\infty} X_i$. Since we have established a.s. convergence of $\sum_1^{\infty} Y_i$, a.s. convergence of $\sum_1^{\infty} X_i$ follows. \diamond

We shall give several applications of Theorem D.2 to large number law theorems. First we give a refinement of the second part of the Borel-Cantelli lemma. Then we give another proof of the strong law of large numbers, but under the slightly more restrictive hypothesis of mutual, rather than pairwise, independence of random variables. The applications depend on a simple lemma in the theory of infinite series.

Lemma D.2 (Kronecker's Lemma) Let b_n be an increasing sequence of positive numbers such that $b_n \uparrow \infty$. Then if $\sum_1^{\infty} z_n/b_n$ converges to a finite limit,

$$\lim_{n \rightarrow \infty} \frac{1}{b_n} \sum_1^n z_i = 0.$$

Proof (Sketch) A summation by parts shows that, with $f_k := \sum_1^k z_i/b_i$, $f_0 = 0$, $b_0 = 0$,

$$(3) \quad \frac{1}{b_n} \sum_1^n z_i = f_n - \frac{1}{b_n} \sum_1^n f_{i-1}(b_i - b_{i-1}).$$

Since $\lim_{n \rightarrow \infty} \frac{1}{b_n} \sum_1^n f_{i-1}(b_i - b_{i-1}) = \lim_{n \rightarrow \infty} f_n$, the right hand side of (3) tends to 0 as $n \rightarrow \infty$. \diamond

We shall use Kronecker's lemma in conjunction with Theorem D.2 to prove a stronger version of the second part of the Borel-Cantelli Lemma.

Theorem D.3. Let A_1, A_2, \dots be independent events and let $p_i = \mathbb{P}(A_i)$, $1 \leq i < \infty$. If $\sum_1^{\infty} p_i = \infty$,

$$(4) \quad \frac{\sum_1^n \mathbf{1}_{A_i}}{\sum_1^n p_i} \rightarrow 1 \quad \text{almost surely as } n \rightarrow \infty.$$

Proof: Let

$$X_k = \frac{\mathbf{1}_{A_k} - p_k}{\sum_1^k p_i}.$$

Then the random variables X_1, X_2, \dots are independent, have zero mean, and

$$(5) \quad \sum_1^\infty \text{Var}(X_k) = \sum_1^\infty \frac{p_k - p_k^2}{(\sum_1^k p_i)^2} \leq \sum_1^\infty \frac{p_k}{(\sum_1^k p_i)^2}.$$

Note that for $k \geq 2$

$$\frac{p_k}{(\sum_1^k p_i)^2} \leq \frac{p_k}{\sum_1^k p_i \sum_1^{k-1} p_i} \leq \frac{1}{\sum_1^k p_i} - \frac{1}{\sum_1^{k-1} p_i}.$$

Hence the last sum in (5) is bounded by the collapsing sum

$$\sum_1^\infty \frac{1}{\sum_1^k p_i} - \frac{1}{\sum_1^{k-1} p_i} = 1 - p_1^{-1},$$

and this is finite. Therefore Theorem D.2 applies to $\{X_n\}$, implying that $\sum_1^\infty X_i$ converges almost-surely. But Kronecker's lemma, applied with $\mathbf{1}_{A_k} - p_k$ in the role of x_k and $\sum_1^n p_i$ in the role of b_n , then yields

$$\frac{\sum_1^n \mathbf{1}_{A_i} - p_i}{\sum_1^n p_i} \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty,$$

and (4) follows immediately. \diamond

It is a simple corollary of the last result that when $\sum_1^\infty \mathbb{P}(A_i) = \infty$, we have $\sum_1^\infty \mathbf{1}_{A_i} = \infty$ almost surely, and hence $\mathbb{P}(A_n \text{ i.o.}) = 1$. Thus Theorem D.3 includes the second half of the Borel-Cantelli lemma.

Next we shall apply Kronecker's Lemma and Theorem D.2 to prove the classical Strong Law of Large Numbers for an i.i.d. sequence of random variables with finite mean.

Theorem D.4 Let X_1, X_2, \dots be i.i.d. random variables with $E[|X_i|] < \infty$, $\mu := E[X_i]$. Then $(\text{a.s.}) \lim \bar{X}_n = \mu_X$.

Of course we have proved in Theorem C.3 a stronger result requiring only pairwise independence. We offer the following proof of Theorem D.4 as an illustration of technique, and for cultural reasons—it is the "classical proof."

Proof: We shall use again the truncation $Y_n := X_n \mathbf{1}_{\{|X_n| \leq n\}}$ and its centering $Z_n = Y_n - E[Y_n]$. We know from the work in section 3.2 that $\{Y_n\}$ is equivalent to $\{X_n\}$, and hence that to prove Theorem D.4 it suffices to prove that $\bar{Z}_n \rightarrow 0$ almost surely. Now Kronecker's Lemma says that $\bar{Z}_n \rightarrow 0$ almost surely if

$$(6) \quad \sum_1^\infty Z_n/n \quad \text{converges almost surely.}$$

On the other hand, Theorem D.2 says that (6) is true if

$$(7) \quad \sum_1^{\infty} n^{-2} \text{Var}(Z_n) < \infty.$$

We will show that this is true and hence complete the proof of Theorem D.4. To verify (7), first observe that $\text{Var}(Z_n) \leq E[X_1^2 \mathbf{1}_{\{|X_1| \leq n\}}]$ because the X_i 's are identically distributed. Then notice that there is a constant C such that for any $x > 0$,

$$\sum_1^{\infty} \frac{1}{n^2} \mathbf{1}_{\{x \leq n\}} \leq \frac{C}{x}.$$

Indeed, we see easily that

$$\sup_{x>0} x \sum_1^{\infty} \frac{1}{n^2} \mathbf{1}_{\{x \leq n\}} < \infty$$

because for $x > 2$

$$\sum_1^{\infty} \frac{1}{n^2} \mathbf{1}_{\{x \leq n\}} \leq \int_{x-1}^{\infty} y^{-2} dy = (x-1)^{-1}.$$

Putting these observations together

$$\begin{aligned} \sum_1^{\infty} n^{-2} \text{Var}(Z_n) &\leq \sum_1^{\infty} n^{-2} E[X_1^2 \mathbf{1}_{\{|X_1| \leq n\}}] \\ &= E[X_1^2 \sum_1^{\infty} n^{-2} \mathbf{1}_{\{|X_1| \leq n\}}] \\ &\leq CE[|X_1|] < \infty. \end{aligned}$$

This completes the proof. \diamond .

E. Generalized weak laws

If X_1, X_2, \dots are independent and identically distributed but $E[|X_i|] = \infty$, it can be shown that

$$\frac{X_1 + \dots + X_n}{n} \quad \text{does not have a finite limit, almost-surely.}$$

One may then ask if there exist increasing sequences of positive integers, $\{a_n\}$ and $\{b_n\}$, such that

$$\frac{(X_1 + \dots + X_n) - a_n}{b_n} \quad \text{admits a finite limit.}$$

We shall give some results on this question when convergence is understood in the sense of probability. We do not aim for ultimate generality, but give only examples of the type of theorem possible. The techniques are entirely elementary.

We shall assume throughout that X_1, X_2, \dots are independent, but not necessarily identically distributed. For convenience, set $S_n = X_1 + \dots + X_n$. The basic idea is again truncation. For each n , we choose a truncation level k_n , where $\{k_n\}$ is an increasing sequence of positive numbers. This time, instead of truncating each X_j at k_j , we shall, for each n , truncate X_1, \dots, X_n at k_n . Thus we define

$$X_{j,n} := X_j \mathbf{1}_{\{|X_j| \leq k_n\}}$$

and we form

$$S'_n := X_{1,n} + \dots + X_{n,n}.$$

The mean of S'_n is

$$a_n = \sum_1^n E [X_j \mathbf{1}_{\{|X_j| \leq k_n\}}].$$

Notice that

$$(1) \quad \mathbb{P}(S_n \neq S'_n) \leq \mathbb{P}(\cup_1^n \{|X_j| > k_n\}) \leq \sum_1^n \mathbb{P}(|X_j| > k_n).$$

By independence, Tchebyshev, and the fact that $\text{Var}(X) \leq E[X^2]$, we derive

$$(2) \quad \mathbb{P}\left(\left|\frac{S'_n - a_n}{b_n}\right| \geq \epsilon\right) \leq \frac{1}{b_n^2} \sum_1^n E [X_j^2 \mathbf{1}_{\{|X_j| \leq k_n\}}].$$

Observe that $\left\{\left|\frac{S_n - a_n}{b_n}\right| \geq \epsilon\right\} \subset \left\{\left|\frac{S'_n - a_n}{b_n}\right| \geq \epsilon\right\} \cup \{S_n \neq S'_n\}$, and hence

$$(3) \quad \mathbb{P}\left(\left|\frac{S_n - a_n}{b_n}\right| \geq \epsilon\right) \leq \mathbb{P}\left(\left|\frac{S'_n - a_n}{b_n}\right| \geq \epsilon\right) + \mathbb{P}(S_n \neq S'_n).$$

The next lemma is an immediate consequence of inequalities (1), (2) and (3).

Lemma E.1 Assume that

$$(4) \quad \sum_1^n \mathbb{P}(|X_j| \leq k_n) \rightarrow 0 \quad \text{and} \quad \frac{1}{b_n^2} \sum_1^n E[X_j^2 \mathbf{1}_{\{|X_j| \leq k_n\}}] \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

as $n \rightarrow \infty$. Then $(P) \lim \frac{S_n - a_n}{b_N} = 0$.

By setting $k_n = b_n = n$ in this lemma, one obtains the sufficiency part of the following theorem.

Theorem E.1 If X_1, X_2, \dots are independent *and* identically distributed then

$$(P) \lim \left(\frac{S_n}{n} - E[X_1 \mathbf{1}_{\{|X_1| \leq n\}}] \right) = 0$$

if and only if $\lim_{x \rightarrow \infty} x \mathbb{P}(|X| \geq x) = 0$.

Proof: Because the random variables are identically distributed,

$$\sum_1^n \mathbb{P}(|X_j| \leq n) = n \mathbb{P}(|X_1| \leq n) \quad \text{and} \quad a_n = n E[X_1 \mathbf{1}_{\{|X_1| \leq n\}}].$$

The first expression converges to 0 as $n \rightarrow \infty$ by assumption, verifying the first condition of (4). To check the second condition of (4), let \mathbb{F} denote the distribution law of $|X_i|$ and observe

$$\begin{aligned} \frac{1}{n^2} \sum_1^n E[X_j^2 \mathbf{1}_{\{|X_j| \leq n\}}] &= \frac{1}{n} [X_1^2 \mathbf{1}_{\{|X_1| \leq n\}}] \\ &= \frac{1}{n} \int_0^n x^2 d\mathbb{F}(x) = \frac{1}{n} \int_0^n 2y \mathbb{F}([y, n]) \\ &\leq \frac{1}{n} \int_0^n 2y \mathbb{P}(|X_1| \geq y) dy \end{aligned}$$

The third equality is derived by integration by parts and the last inequality by definition of \mathbb{F} . The assumption $\lim_{x \rightarrow \infty} x \mathbb{P}(|X| \geq x) = 0$ implies (L'Hopital's rule!)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \int_0^n 2y \mathbb{P}(|X_1| \geq y) dy = 0,$$

and so the proof of sufficiency is complete. The proof of necessity is more sophisticated and is omitted. \diamond