## 2.3. LAWS OF LARGE NUMBERS

If we carry out independent copies of the same experiment or, if in the course of the same experiment, we observe repeatedly independent copies $X_1, X_2, \cdots$ of a certain numerical characteristic $X$, then we expect that the arithmetic mean or "sample average"

$$\overline{X}_n := \frac{1}{n}\sum_{j=1}^{n} X_j \;\text{ should converge, in some sense, to the ensemble average }\; \mathbf{E}(X)\,, \tag{3.1}$$

as the number $n$ of observations becomes large ($n \to \infty$). Here by "ensemble average" we mean the expectation

$$\mathbf{E}(X) = \int_\Omega X(\omega)\, d\mathbf{P}(\omega)\,.$$

In the special context of the coin-tossing Example 2.1, we saw a manifestation of this principle in (2.5) and in Exercise 2.3.

More generally, we obtained in Example 2.6 the a.e. convergence of the random sequence $\{\overline{X}_n\}_{n\in\mathbf{N}}$ of *sample averages*, to the number $\mathbf{E}(X)$, the *ensemble average*, under the condition $\mathbf{E}(|X|) < \infty$, based on the Birkhoff pointwise ergodic theorem.

A celebrated result of Kolmogorov (1931) shows that (3.1) holds $\mathbf{P}-$almost everywhere under the condition $\mathbf{E}(|X|) < \infty$, for random variables $X_1, X_2, \cdots$ that have the same distribution as $X$ but are only *pairwise* independent. This result is known as the **Kolmogorov Strong Law of Large Numbers.**

**3.1 THEOREM: KOLMOGOROV's STRONG LAW OF LARGE NUMBERS.** *Let $X_1, X_2, \cdots$ be a sequence of pairwise-independent random variables, with the same distribution and $\mathbf{E}(|X_1|) < \infty$. Setting $S_n := \sum_{k=1}^n X_k$ and $\overline{X}_n := S_n / n$, we have:*

$$\lim_{n\to\infty} \overline{X}_n(\omega) \;=\; \mathbf{E}(X_1) \quad \text{for a.e. } \omega \in \Omega\,. \tag{3.3}$$

We shall defer the (rather technical) proof of this result until the next section. An elegant, very direct, stronger and "natural" argument can be given with the help of the theory of Martingales in the case of independent random variables; see Theorem 4.2.4.

An earlier result in the development of this subject, due to Ȟinčin (1928), shows that the convergence in (3.1) holds "in probability", in the sense of Definition 2.1. This result is known as the *Weak Law of Large Numbers*, and is of course subsumed by the strong law. However, in order to illustrate some useful techniques in the theory, we shall give the proof of this weak law (Proposition 3.1) straightaway.

**3.1 Proposition: Ȟinčin's Weak Law of Large Numbers.** *Let the random variables* $X_1, X_2, \cdots$ *be pairwise-independent, with the same distribution and* $\mathbf{E}(|X_1|) < \infty$*. Setting* $S_n := \sum_{k=1}^n X_k$ *and* $\overline{X}_n := S_n \,/\, n$*, we have for any* $\varepsilon > 0$:

$$\mathbf{P}\left(|\overline{X}_n - \mathbf{E}(X_1)| > \varepsilon\right) \longrightarrow 0\,, \quad \text{as } n \to \infty\,. \tag{3.3}$$

*Proof*: We have already established this result in the case $X_1 \in \mathbf{L}^2$ (Exercise 2.3), so we need now to argue its validity under the weaker condition $X_1 \in \mathbf{L}^1$. For this we shall use the *method of truncation:* we consider the "truncated variables"

$$Z_n := X_n\,, \quad \text{if } |X_n| \le n\,; \qquad Z_n := 0\,, \quad \text{if } |X_n| > n\,,$$

for all $n \in \mathbf{N}$. We have from Exercise 1.2:

$$\sum_{n\in\mathbf{N}} \mathbf{P}(X_n \neq Z_n) = \sum_{n\in\mathbf{N}} \mathbf{P}(|X_n| > n) = \sum_{n\in\mathbf{N}} \mathbf{P}(|X_1| > n) \le 1 + \mathbf{E}(|X_1|) < \infty\,.$$

Thus, thanks to the Borel-Cantelli Lemma, the original sequence $\{X_n\}_{n\in\mathbf{N}}$ and its "truncated version" $\{Z_n\}_{n\in\mathbf{N}}$ *differ on at most finitely many indices:* $\mathbf{P}(X_n \neq Z_n\,,\ \text{i.o.}) = 0$.

This means that it suffices to show $\mathbf{P}\left(|\overline{Z}_n - \mathbf{E}(X_1)| > \varepsilon\right) \longrightarrow 0$, as $n \to \infty$, for $\overline{Z}_n = T_n \,/\, n$, $T_n := \sum_{k=1}^n Z_k$, or in fact

$$\mathbf{P}\left(|T_n - \mathbf{E}(T_n)| > \varepsilon\, n\right) \longrightarrow 0\,, \quad \text{as } n \to \infty\,. \tag{3.4}$$

This is because $\mathbf{E}(T_n) = \sum_{k=1}^n \mathbf{E}(Z_k)$ and $\mathbf{E}(Z_n) = \mathbf{E}\left(X_1\, \chi_{\{|X_1|\le n\}}\right) \to \mathbf{E}(X_1)$ by the Dominated Convergence Theorem, thus $\mathbf{E}(T_n)\,/\,n \to \mathbf{E}(X_1)$ as $n \to \infty$.

To prove (3.4), let us start by noticing that the variables $\{Z_n\}_{n\in\mathbf{N}}$ are also pairwise independent, so that by the Čebyšev inequality:

$$(\varepsilon\, n)^2 \cdot \mathbf{P}(|T_n - \mathbf{E}(T_n)| > \varepsilon n) \le \mathrm{Var}(T_n) = \sum_{k=1}^n \mathrm{Var}(Z_k) \le \sum_{k=1}^n \mathbf{E}(Z_k^2)\,,$$

so it suffices to show $\sum_{k=1}^n \mathbf{E}(Z_k^2) = o(n^2)$. A rather crude estimate gives

$$\sum_{k=1}^n \mathbf{E}(Z_k^2) \le \sum_{k=1}^n k\ \mathbf{E}\left(|X_1|\, \chi_{\{|X_1|\le k\}}\right) \le \mathbf{E}\left(|X_1|\right) \cdot n(n+1)\,/\,2 = O(n^2)\,,$$

so we need to do something more elaborate.

For this purpose, consider an increasing sequence $\{a_n\}_{n\in\mathbf{N}}$ with $0 < a_n < n$ and such that $a_n \to \infty$, $a_n = o(n)$ as $n \to \infty$. If we write $\sum_{k=1}^n \mathbf{E}(Z_k^2)$ in the form $\left(\sum_{k\le a_n} + \sum_{a_n<k\le n}\right)\ \mathbf{E}\left(X_1^2\, \chi_{\{|X_1|\le k\}}\right)$, then this last expression is dominated by

$$\sum_{k\le a_n} a_n\, \mathbf{E}\left(|X_1|\, \chi_{\{|X_1|\le a_n\}}\right) + \sum_{a_n<k\le n} \Big(a_n\, \mathbf{E}\left(|X_1|\, \chi_{\{|X_1|\le a_n\}}\right) + n\, \mathbf{E}\left(|X_1|\, \chi_{\{a_n<|X_1|\le n\}}\right)\Big)$$

which, in turn, is dominated by $a_n \cdot n\,\mathbf{E}(|X_1|) + n^2\,\mathbf{E}\left(|X_1|\,\chi_{\{a_n<|X_1|\le n\}}\right)$. Therefore,

$$\frac{1}{n^2}\sum_{k=1}^{n}\mathbf{E}(Z_k^2) \;\le\; \frac{a_n}{n}\,\mathbf{E}(|X_1|) + \mathbf{E}\left(|X_1|\,\chi_{\{a_n<|X_1|\}}\right) \longrightarrow 0 \quad \text{as} \;\; n\to\infty\,,$$

since we assume $a_n \to \infty$ and $(a_n/n)\to 0$, as $n\to\infty$. This finishes the argument. ◇

The method of truncation that served us so well in the above proof suggests also the following result due to Kolmogorov and Feller.

**3.2 Proposition:** *Let $X_1, X_2, \cdots$ be independent variables, and suppose that we have*

$$\lim_{n\to\infty}\left(\sum_{j=1}^{n}\mathbf{P}(|X_j|>b_n)\right) \;=\; 0\,, \qquad \lim_{n\to\infty}\left(\frac{1}{b_n^2}\sum_{j=1}^{n}\mathbf{E}\left(X_j^2\,\chi_{\{|X_j|\le b_n\}}\right)\right) \;=\; 0$$

*for some sequence $\{b_n\}_{n\in\mathbf{N}}$ of positive numbers with $\lim_{n\to\infty} b_n = \infty$. Then with $a_n := \sum_{j=1}^{n}\mathbf{E}\left(X_j\,\chi_{\{|X_j|\le b_n\}}\right)$ we have*

$$\lim_{n\to\infty}\frac{1}{b_n}\left(\sum_{j=1}^{n}X_j - a_n\right) \;=\; 0\,, \qquad \textit{in probability.}$$

**3.1 Exercise:** Prove Proposition 3.2.

**3.2 Exercise : Wald's Identity.** Let $X_1, X_2, \cdots$ be independent and integrable random variables, and define $S_0 = 0$, $S_n = \sum_{j=1}^{n} X_j \;\; (n\in\mathbf{N})$. Suppose that $T:\Omega\to\mathbf{N}$ is measurable and satisfies

$$\{T\le n\} \,\in\, \sigma(X_1,\cdots,X_n)\,, \qquad \forall\;\; n\in\mathbf{N}\,.$$

(Such an integer-valued random variable will be called *stopping time* in Chapter 4.)

(i) If $\mathbf{E}(X_j)=0$ for all $j\in\mathbf{N}$, and if $T$ is bounded (that is, $\mathbf{P}(T\le\kappa)=1$ for some $\kappa\in\mathbf{N}$), then $S_T$ is integrable and $\mathbf{E}(S_T)=0$.

(ii) If $X_1, X_2, \cdots$ are identically distributed and non-negative, then

$$\mathbf{E}(S_T) \;=\; \mathbf{E}(X_1)\cdot\mathbf{E}(T)\,. \qquad (3.5)$$

(iii) If $X_1, X_2, \cdots$ are identically distributed, and if $\mathbf{E}(T)<\infty$, then $S_T$ is integrable and (3.5) holds.

**3.3 Exercise : Renewal Theory.** Let $X_1, X_2, \cdots$ be *I.I.D.* random variables with $\mathbf{P}(X_1 > 0) = 1$, $0 < m := \mathbf{E}(X_1) < \infty$, and define $S_0 = 0$, $S_n = \sum_{j=1}^n X_j$, $n \in \mathbf{N}$, as well as the family of random variables $\mathcal{N}$ given as

$$N_t := \max\{n \in \mathbf{N}_0 \,|\, S_n \le t\} \;=\; \sum_{n=1}^{\infty} \chi_{\{S_n \le t\}}\,, \qquad 0 \le t < \infty\,.$$

If one thinks of the $X$'s as "inter-arrival times" of customers in a certain facility, then $S_n$ is the arrival-time of the $n^{th}$ customer, and $N_t$ the number of customers who have arrived by the fixed time $t$. Show that

(i) $\{N_t = n\} = \{S_n \le t < S_{n+1}\}\,,\;\; \{N_t < n\} = \{S_n > t\}\,,$ for all $n \in \mathbf{N}_0\,,\; t \in [0, \infty)\,;$

(ii) $\lim_{t\to\infty} N_t \;=\; \infty\,,$ $\mathbf{P}-$a.e. ;

(iii) $\lim_{t\to\infty} (N_t \,/\, t) \;=\; (1 \,/\, m)\,,$ $\mathbf{P}-$a.e. ;

(iii) $\lim_{t\to\infty} (\,\mathbf{E}(N_t) \,/\, t) \;=\; (1 \,/\, m)$ .

For the last claim, you may find the Wald identity of Exercise 3.2 useful.
(The family $\mathcal{N}$ is the prototype for a *Stochastic Process:* a family of random variables indexed by "time" $t \in [0, \infty)$. This process $\mathcal{N}$ is called **Renewal Process**. In the special case where the $X$'s have the exponential distribution $\lambda\, e^{-\lambda x}\, dx$ on $(0, \infty)$, then $\mathcal{N} = \{N_t\,,\, 0 \le t < \infty\}$ is the so-called **Poisson Process** with rate $\lambda$. We shall encounter Stochastic Processes again in section 2.9, where we take up the study of Brownian Motion.)

**3.4 Exercise:** For any bounded, continuous function $f : [0, \infty) \to \mathbf{R}$, compute the limits

$$\lim_{n\to\infty} \int_0^1 \cdots \int_0^1 f\left(\frac{x_1 + \cdots + x_n}{n}\right) dx_1 \cdots dx_n\,,$$

$$\lim_{n\to\infty}\; 2^n \int_0^\infty \cdots \int_0^\infty f\left(\frac{x_1 + \cdots + x_n}{n}\right) e^{-2(x_1 + \cdots + x_n)}\, dx_1 \cdots dx_n\,.$$

**3.5 Exercise:** Consider a sequence $X_1, X_2, \cdots$ of *I.I.D.* random variables with

$$\mathbf{P}(X_1 = \pm\, n) \;=\; \frac{c}{n^2\, \log\, n}\,, \qquad n = 3, 4, \cdots$$

where $c$ is a normalizing constant, and set $S_n = \sum_{j=1}^n X_j\,,\; n \in \mathbf{N}\,.$

(i) Observe that $\mathbf{P}(|X_1| > n) \sim \text{constant} \,/\, (n \log n)\,,$ $\mathbf{E}(|X_1|) = \infty\,.$

(ii) $\lim_{n\to\infty} (S_n \,/ n) \;=\; 0\,,$ in probability.

In other words, we have a weak-law-of-large-numbers-type result, despite the fact that the expectation of the random variables is infinite.

(iii) Show that $\mathbf{P}\big(|X_n| > n\,,\ \text{i.o.}\big) = 1$, and from this argue that

$$\mathbf{P}\Big(\,|S_n| > (n/2)\,,\ \text{i.o.}\Big) = 1\,, \qquad \mathbf{P}\bigg(\lim_{n\to\infty}(S_n\,/\,n)\,=0\bigg) < 1\,.$$

(iv) Show that, in fact: $\limsup_{n\to\infty}(S_n\,/\,n)\;=\;+\infty\,,\quad \liminf_{n\to\infty}(S_n\,/\,n)\;=\;-\infty\,,$ a.e.

**3.6 Exercise:** Let $X_1, X_2, \cdots$ are *I.I.D.* random variables, and denote $S_n := \sum_{j=1}^{n} X_j$, $\overline{X}_n := S_n\,/\,n$ for $n \geq 1$.

(i) We have $\mathbf{E}(|X_1|)\;<\;\infty$, if and only if $\mathbf{P}(\,|X_n| > c\,n\,,\ \text{i.o.}\,)\;=\;0$ for every constant $c \in (0,\infty)$.

(ii) If $\mathbf{E}(|X_1|)\;=\;\infty$, then

- $\mathbf{P}(\,|X_n| > c\,n\,,\ \text{i.o.}\,)\;=\;1\,,\quad \forall\; c \in (0,\infty).$
- $\big\{\lim_{n\to\infty}\overline{X}_n\ \text{ exists in }\mathbf{R}\big\} \subseteq \big\{\lim_{n\to\infty}\overline{X}_n = 0\big\}\,,\quad$ mod. $\mathbf{P}$.
- $\mathbf{P}\left(\lim_{n\to\infty}\overline{X}_n\ \text{ exists in }\mathbf{R}\right)\;=\;0\,,\quad \mathbf{P}\left(\limsup_{n\to\infty}|\overline{X}_n|\;=\;\infty\right)\;=\;1\,.$

**3.7 Exercise: Glivenko-Cantelli Theorem:** Let $X_1, X_2, \cdots$ be independent random variables with common distribution function $F(x) = \mathbf{P}[\,X_n \leq x\,]\,,\ x \in \mathbf{R}$. For any *given* real number $x$ consider the relative frequency

$$F_n(x,\omega)\;:=\;\frac{1}{n}\sum_{j=1}^{n}\chi_{(-\infty,x]}\big(X_j(\omega)\big)$$

of values from among $\big(X_1(\omega),\cdots,X_n(\omega)\big)$ that do not exceed the number $x$.

(i) Show that $\lim_{n\to\infty}F_n(x,\omega)\;=\;F(x)$ holds for a.e. $\omega \in \Omega$.

(ii) The random function $x \mapsto F_n(x,\omega)$ is called the **empirical distribution function** of $\big(X_1(\omega),\cdots,X_n(\omega)\big)$; show that it converges uniformly to $F(\cdot)$, that is, for a.e. $\omega \in \Omega$ we have

$$\lim_{n\to\infty}\left(\sup_{x\in\mathbf{R}}\big|\,F_n(x,\omega) - F(x)\,\big|\right)\;=\;0\,.$$

**3.8 Exercise: The Saint Petersburg Game.** Let $X_1, X_2, \cdots$ be independent random variables with common distribution $\mathbf{P}(X_n = 2^j) = 2^{-j}\,,\ j \in \mathbf{N}$ (you win $2^j$ dollars if it takes $j$ tosses of a fair coin to obtain a success). Clearly $\mathbf{E}(X_n) = \infty$. Show that

$$\limsup_{n\to\infty}\left(\frac{\sum_{k=1}^{n}X_k}{n\log_2 n}\right)\;=\;1\,,\qquad \text{in probability}$$

but not a.e. In fact,

$$\limsup_{n\to\infty}\left(\frac{X_n}{n\log_2 n}\right)\;=\;\limsup_{n\to\infty}\left(\frac{\sum_{k=1}^{n}X_k}{n\log_2 n}\right)\;=\;\infty\,,\qquad\qquad \text{a.e.}$$