Replacing t by λt and integrating with respect to F one gets (6.7). It follows, in particular, that

$$\varphi(\lambda) = \sum_{k=0}^{\infty} \frac{(-1)^k \mu_k \lambda^k}{k!} \tag{6.9}$$

in any interval $0 \leq \lambda < \lambda_0$ *in which the series on the right converges.* It is known from analytic function theory that in this case the series in (6.9) uniquely determines $\varphi(\lambda)$ for *all* $\lambda > 0$, and hence *the moments* $\mu_1, \mu_2, \ldots$ *determine the distribution* F *uniquely whenever the series in* (6.9) *converges in some interval* $|\lambda| < \lambda_0$. This useful criterion holds also for distributions not concentrated on $\overline{0, \infty}$, but the proof depends on the use of characteristic functions (see XV,4).

*7. LAWS OF LARGE NUMBERS FOR IDENTICALLY DISTRIBUTED VARIABLES

Throughout this section we use the notation $\mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n$. The oldest version of the law of large numbers states that if the $\mathbf{X}_k$ are independent and have a common distribution with expectation μ and finite variance then[10] for fixed $\epsilon > 0$ as $n \to \infty$

$$\mathbf{P}\{|n^{-1}\mathbf{S}_n - \mu| > \epsilon\} \to 0. \tag{7.1}$$

This chapter started from the remark that (7.1) is contained in Chebyshev's inequality. To obtain sharper results we derive a variant of Chebyshev's inequality applicable even when no expectation exists. Define new random variables $\mathbf{X}'_k$ by truncation of $\mathbf{X}_k$ at an arbitrary, but fixed, level $\pm s_n$. Thus

$$\mathbf{X}'_k = \begin{matrix} \mathbf{X}_k & \textit{when} \quad |\mathbf{X}_k| \leq s_n \\ 0 & \textit{when} \quad |\mathbf{X}_k| > s_n. \end{matrix} \tag{7.2}$$

Put

$$\mathbf{S}'_n = \mathbf{X}'_1 + \cdots + \mathbf{X}'_n, \qquad m'_n = \mathbf{E}(\mathbf{S}'_n) = n\mathbf{E}(\mathbf{X}'_1). \tag{7.3}$$

Then obviously

$$\mathbf{P}\{|\mathbf{S}_n - m'_n| > t\} \leq \mathbf{P}\{|\mathbf{S}'_n - m'_n| > t\} + \mathbf{P}\{\mathbf{S}_n \neq \mathbf{S}'_n\} \tag{7.4}$$

because the event on the left cannot occur unless one of the events on the right occurs.

* The topics of this section are related to the oldest probabilistic theory but are of no particular significance in the remainder of this book. They are treated for their historical and methodological interest and because many papers are devoted to partial converses of the law of large numbers.

[10] (7.1) is equivalent to $n^{-1}\mathbf{S}_n - \mu \xrightarrow{\mathrm{P}} 0$, where $\xrightarrow{\mathrm{P}}$ signifies "tends in probability to." (See VIII,2.)

This inequality is valid also for dependent variables with varying distributions, but here we are interested only in identically distributed independent variables. Putting $t = nx$ and applying Chebyshev's inequality to the first term on the right, we get from (7.4) the following

Lemma. *Let the* $\mathbf{X}_k$ *be independent with a common distribution* F. *Then for* $x > 0$

$$\mathbf{P}\left\{\left|\frac{1}{n}\mathbf{S}_n - \mathbf{E}(\mathbf{X}_1')\right| > x\right\} \leq \frac{1}{n^2x^2}\mathbf{E}(\mathbf{X}_1'^2) + nP\{|\mathbf{X}_1| > s_n\}. \tag{7.5}$$

As an application we could derive *Khintchine's law of large numbers* which states that (7.1) holds for all $\epsilon > 0$ whenever the $\mathbf{X}_k$ have finite expectation μ. The proof would be essentially a repetition of the proof for the discrete case given in **1**; X,2. We pass therefore directly to a stronger version which includes a necessary and sufficient condition. For its formulation we put for $t > 0$

$$\tau(t) = [1-F(t)+F(-t)]t \tag{7.6}$$

and

$$\sigma(t) = \frac{1}{t}\int_{-t}^{t} x^2\, F(dx) = -\tau(t) + \frac{2}{t}\int_0^t x\tau(x)\, dx. \tag{7.7}$$

(The identity of these two expressions follows by a simple integration by parts.)

Theorem 1. (*Generalized weak law of large numbers.*) *Let the* $\mathbf{X}_k$ *be independent with a common distribution* F. *In order that there exist constants* μ_n *such that for each* $\epsilon > 0$

$$\mathbf{P}\{|n^{-1}\mathbf{S}_n - \mu_n| > \epsilon\} \to 0 \tag{7.8}$$

it is necessary and sufficient that[11] $\tau(t) \to 0$ *as* $t \to \infty$. *In this case* (7.8) *holds with*

$$\mu_n = \int_{-n}^{n} x\, F\{dx\}. \tag{7.9}$$

Proof. (*a*) *Sufficiency.* Define μ_n by (7.9). We use the truncation (7.2) with $s = n$. Then $\mu_n = \mathbf{E}(\mathbf{X}_1')$ and the preceding lemma the left side of (7.8) is $< \epsilon^{-2}\sigma(n) + \tau(n)$, which tends to 0 whenever $\tau(t) \to 0$. Thus this condition is sufficient.

(*b*) *Necessity.* Assume (7.8). As in V,5 we introduce the variables ${}^0\mathbf{X}_k$ obtained directly by symmetrization of $\mathbf{X}_k$. Their sum ${}^0\mathbf{S}_n$ can be obtained

[11] It follows from (7.7) that $\tau(t) \to 0$ implies $\sigma(t) \to 0$. The converse is also true; see problem 11. For a different proof of theorem 1 see XVII,2*a*.

by symmetrization of $S_n - n\mu$. Let a be a median of the variables X_k. Using the inequalities V,(5.6), V,(5.10), and V,(5.7) in that order we get

$$2P\{|S_n - n\mu| > n\epsilon\} \geq P\{|{}^0S_n| > 2n\epsilon\} \geq \tfrac{1}{2}[1 - \exp(-nP\{|{}^0X_1| > 2n\epsilon\})]$$
$$\geq \tfrac{1}{2}[1 - \exp(-\tfrac{1}{2}nP\{|X_1| > 2n\epsilon + |a|\})].$$

In view of (7.8) the left side tends to 0. It follows that the exponent on the right tends to 0, and this is manifestly impossible unless $\tau(t) \to 0$. ▶

The condition $\tau(t) \to 0$ is satisfied whenever F has an expectation μ. The truncated moment μ_n then tends to μ and so in this case (7.8) is equivalent with the classical law of large numbers (7.1). However, *the classical law of large numbers in the form* (7.1) *holds also for certain variables without expectation.* For example, if F is a symmetric distribution such that $t[1 - F(t)] \to 0$ then $P\{|n^{-1}S_n| > \epsilon\} \to 0$. But an expectation exists only if $1 - F(t)$ is integrable between 0 and ∞, which is a stronger condition.

(It is interesting to note, that the *strong* law of large numbers holds *only* for variables with expectations. See theorem 4 of section 8).

The empirical meaning of the law of large numbers was discussed in **1**; X with special attention to the classical theory of "fair games." We saw in particular that even when expectations exist a participant in a "fair game" may be strongly on the losing side. On the other hand, the analysis of the St. Petersburg game showed that the classical theory applies also to certain games with infinite expectations except that the "fair entrance fee" will depend on the contemplated number of trials. The following theorem renders this more precise.

We consider independent *positive* variables X_k with a common distribution F. [Thus $F(0) = 0$.]. The X_k may be interpreted as possible gains, and a_n as the total entrance fee for n trials. We put

$$\mu(s) = \int_0^s x\, F\{dx\}, \qquad \frac{\mu(s)}{s[1 - F(s)]} = \rho(s). \tag{7.10}$$

Theorem 2. *In order that there exist constants a_n such that*

$$P\{|a_n^{-1}S_n - 1| > \epsilon\} \to 0 \tag{7.11}$$

it is necessary and sufficient that[12] $\rho(s) \to \infty$ *as* $s \to \infty$. *In this case there exist numbers* s_n *such that*

$$n\mu(s_n) = s_n \tag{7.12}$$

and (7.11) *holds with* $a_n = n\mu(s_n)$.

Proof. (*a*) *Sufficiency.* Assume $\rho(s) \to \infty$. For large n the function $n\mu(s)/s$ assumes values > 1, but it tends to 0 as $s \to \infty$. The function is right continuous, and the limit from the left cannot exceed the limit from the right. If s_n is the lower bound of all s such that $n\mu(s)s^{-1} \leq 1$ it follows that (7.12) holds.

[12] It will be seen in VIII,9 (theorem 2) that $\rho(s) \to \infty$ iff $\mu(s)$ *varies slowly* at infinity. The relation (7.11) is equivalent to $a_n^{-1}S_n \xrightarrow{P} 1$ (see VIII,2).

Put $\mu_n = \mu(s_n) = \mathbf{E}(\mathbf{X}_1')$. We use the inequality (7.5) of the lemma with $x = \epsilon\mu_n$ to obtain

$$\mathbf{P}\left\{\left|\frac{\mathbf{S}_n}{n\mu_n} - 1\right| > \epsilon\right\} \leq \frac{1}{\epsilon^2 n\mu_n^2}\mathbf{E}(\mathbf{X}_1'^2) + n[1-F(s_n)]. \tag{7.13}$$

An integration by parts reduces $\mathbf{E}(\mathbf{X}_1'^2)$ to an integral with integrand $x[1 - F(x)]$, and by assumption this function is $o(\mu(x))$. Thus $\mathbf{E}(\mathbf{X}_1'^2) = o(s_n\mu_n)$, and in view of (7.12) this means that the first term on the right in (7.13) tends to 0. Similarly (7.12) and the definition (7.10) of $\rho(s)$ show that $n[1 - F(s_n)] \to 0$. Thus (7.13) reduces to (7.11) with $a_n = n\mu_n$.

(*b*) *Necessity.* We now assume (7.11) and use the truncation (7.2) with $s_n = 2a_n$. Since $\mathbf{E}(\mathbf{X}_1'^2) \leq s_n\mu_n$ we get from the basic inequality (7.5) with $x = \epsilon a_n/n$

$$\mathbf{P}\{\mathbf{S}_n > n\mu_n + \epsilon a_n\} \leq \frac{2}{\epsilon^2}\cdot\frac{n\mu_n}{a_n} + n[1-F(2a_n)]. \tag{7.14}$$

Since we are dealing with positive variables

$$\mathbf{P}\{\mathbf{S}_n < 2a_n\} \leq \mathbf{P}\{\max_{k\leq n} \mathbf{X}_k \leq 2a_n\} = F^n(2a_n). \tag{7.15}$$

By assumption the left side tends to 1, and this implies $n[1 - F(2a_n)] \to 0$ (because $x \leq e^{-(1-x)}$ for $x \leq 1$). If $n\mu_n/a_n$ tended to zero the same would be true of the right side in (7.14) and this inequality would manifestly contradict the assumption (7.11). This argument applies also to subsequences and shows that $n\mu_n/a_n$ remains bounded away from zero; this in turn implies that $\rho(2a_n) \to \infty$.

To show that $\rho(x) \to \infty$ for any approach $x \to \infty$ choose a_n such that $2a_n < x \leq 2a_{n+1}$. Then $\rho(x) \geq (2a_n)a_n/a_{n+1}$, and it is obvious that (7.11) necessitates the boundedness of the ratios a_{n+1}/a_n. ▶

*8. STRONG LAWS

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be mutually independent random variables with a common distribution F and $\mathbf{E}(\mathbf{X}_k) = 0$. As usual we put $\mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n$. The weak law of large numbers states that for every $\epsilon > 0$

$$\mathbf{P}\{n^{-1}|\mathbf{S}_n| > \epsilon\} \to 0. \tag{8.1}$$

This fact does not eliminate the possibility that $n^{-1}\mathbf{S}_n$ may become arbitrarily large for infinitely many n. For example, in a symmetric random walk the probability that the particle passes through the origin at the nth step tends to 0, and yet it is certain that infinitely many such passages will occur. In practice one is rarely interested in the probability in (8.1) for any particular large value of n. A more interesting question is whether $n^{-1}|\mathbf{S}_n|$ will ultimately become and remain small, that is, whether $n^{-1}|\mathbf{S}_n| < \epsilon$ simultaneously for all $n \geq N$. Accordingly we ask for the probability of the event[13] that $n^{-1}\mathbf{S}_n \to 0$.

* This section may be omitted at the first reading.

[13] It follows from the zero-or-one law of IV,6 that this probability equals 0 or 1, but we shall not use this fact.