Chapter 2

Independence and Strong Convergence

This chapter is devoted to the fundamental concept of independence and to several results based on it, including the Kolmogorov strong laws and his three series theorem. Some applications to empiric distributions, densities, queueing sequences and random walk are also given. A number of important results, included in the problems section, indicate the profound impact of the concept of independence on the subject. All these facts provide deep motivation for further study and development of probability theory.

2.1 Independence

If A and B are two events of a probability space (Ω, Σ, P), it is natural to say that A is independent of B whenever the occurrence or nonoccurrence of A has no influence on the occurrence or nonoccurrence of B. Consequently the uncertainty about joint occurrence of both A and B must be higher than either of the individual events. This means that the probability of a joint occurrence of A and B should be "much smaller" than either of the individual probabilities. This intuitive feeling can be formalized mathematically by the equation

$$P(A \cap B) = P(A)P(B)$$

for a pair of events A, B. How should intuition translate for three events A, B, C if every pair among them is independent? The following ancient example, due to S. Bernstein, shows that, for a satisfactory mathematical abstraction, more care is necessary. Thus if $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\Sigma = \mathcal{P}(\Omega)$, the power set, let each point carry the same weight, so that

$$P(\{\omega_i\}) = \frac{1}{4}, i = 1, ..., 4.$$

Let $A = \{\omega_1, \omega_2\}$, $B = \{\omega_1, \omega_3\}$, and $C = \{\omega_4, \omega_1\}$. Then clearly $P(A \cap B) = P(A)P(B) = \frac{1}{4}$, $P(B \cap C) = P(B)P(C) = \frac{1}{4}$, and $P(C \cap A) = P(C)P(A) = \frac{1}{4}$.

But $P(A \cap B \cap C) = \frac{1}{4}$, and $P(A)P(B)P(C) = \frac{1}{8}$. Thus A, B, C are not independent. Also $A, (B \cap C)$ are not independent, and similarly $B, (C \cap A)$ and $C, (A \cap B)$ are not independent.

These considerations lead us to introduce the precise concept of mutual independence of a collection of events by not pairwise but by systems of equations so that the above anomaly cannot occur.

Definition 1 Let (Ω, Σ, P) be a probability space and $\{A_i, i \in I\} \subset \mathcal{P}(\Omega)$ be a family of events. They are said to be *pairwise independent* if for each distinct i, j in I we have $P(A_i \cap A_j) = P(A_i)P(A_j)$. If $A_{i_1}, \ldots, A_{i_n}$ are n (distinct) events, $n \geq 2$, then they are *mutually independent* if

$$P\left(\bigcap_{k=1}^{m} A_{i_k}\right) = \prod_{k=1}^{m} P(A_{i_k}) \tag{1}$$

holds simultaneously for each $m = 2, 3, ..., n$. The whole class $\{A_i, i \in I\}$ is said to be mutually independent if each finite subcollection is mutually independent in the above sense, i.e., equations (1) hold for each $n \geq 2$. Similarly if $\{\mathcal{A}_i, i \in I\}$ is a collection of families of events from Σ then they are mutually independent if for each n, $A_{i_k} \in \mathcal{A}_{i_k}$ we have the set of equations (1) holding for A_{i_k}, $k = 1, ..., m, 1 < m \leq n$. Thus if $A_i \in \mathcal{A}_i$ then $\{A_i, i \in I\}$ is a mutually independent family. [Following custom, we usually omit the word "mutually".]

It is clear that the (mutual) independence concept is given by a system of equations (1) which can be arbitrarily large depending on the richness of Σ. Indeed for each n events, (1) is a set of $2^n - n - 1$ equations, whereas the pairwise case needs only $\binom{n}{2}$ equations. Similarly "m-wise" independence has $\binom{n}{m}$ equations, and it does not imply other independences if $2 \leq m < n$ is a fixed number m. It is the strength of the (mutual) concept that allows all $n \geq 2$. This is the mathematical abstraction of the intuitive feeling of independence that experience has shown to be the best possible one. It seems to give a satisfactory approximation to the heuristic idea of independence in the physical world. In addition, this mathematical formulation has been found successful in applications to such areas as number theory, and Fourier analysis. **The notion of independence is fundamental to probability theory and distinguishes it from measure theory.** The concept translates itself to random variables in the following form.

Definition 2 Let (Ω, Σ, P) be a probability space and $\{X_i, i \in I\}$ be abstract random variables on Ω into a measurable space $(S, \mathcal{A})$. Then they are said to be *mutually independent* if the class $\{\mathcal{B}, i \in I\}$ of σ-algebras in Σ is mutually independent in the sense of Definition 1, where $\mathcal{B}_i = X_i^{-1}(\mathcal{A})$, the σ-algebra generated by $X_i, i \in I$. Pairwise independence is defined similarly.

Taking $S = \mathbb{R}(or\mathbb{R}^n)$ and $\mathcal{A}$ as its Borel σ-algebra, one gets the corresponding concept for real (or vector) random families.

It is perhaps appropriate at this place to observe that many such (independent) families of events or random variables on an (Ω, Σ, P) need not exist if (Ω, Σ) is not rich enough. Since $\emptyset$ and Ω are clearly independent of each event $A \in \Omega$, the set of equations (1) is non vacuous. Consider the trivial example $\Omega = \{0, 1\}$, $\Sigma = \mathcal{P}(\Omega) = \{\emptyset, \{0\}, \{1\}, \Omega\}$, $P(\{0\}) = p = 1 - P(\{1\})$, $0 < p < 1$. Then, omitting the $\emptyset$, Ω, there are no other independent events, and if $X_i : \Omega \to \mathbb{R}$, $i = 1, 2$, defined as $X_1(0) = 1 = X_2(1)$ and $X_1(1) = 2 = X_2(0)$, then X_1, X_2 are distinct random variables, but they are not independent. Any other random variables defined on Ω can be obtained as functions of these two, and it is easily seen that there are no nonconstant independent random variables on this Ω. Thus (Ω, Σ, P) is not rich enough to support nontrivial (i.e., nonconstant) independent random variables. We show later that a probability space can be enlarged to have more sets, so that one can always assume the existence of enough independent families of events or random variables. We now consider some of the profound consequences of this mathematical formalization of the natural concept of mutual independence. It may be noted that the latter is also termed *statistical* (*stochastic* or *probabilistic*) *independence* to contrast it with other concepts such as *linear independence* and *functional independence*. [The functions X_1, X_2 in the above illustration are linearly independent but *not* mutually (or statistically) independent! See also Problem 1.]

To understand the implications of equations (1), we consider different forms (or consequences) of Definitions 1 and 2. First note that if $\{A_i, i \in I\} \subset \Sigma$ is a class of mutually independent events, then it is evident that $\{\sigma(A_i), i \in I\}$ is an independent class. However, the same cannot be said if the singleton A_i is replaced by a bigger family $\mathcal{G}_i = \{A^i_j, j \in J_i\} \subset \Sigma$, where each J_i has at least two elements, $i \in I$, as simple examples show. Thus $\{\sigma(\mathcal{G}_i), i \in I\}$ need not be independent. On the other hand, we can make the following statements.

Theorem 3 (a) *Let $\{\mathcal{A}, \mathcal{B}_i, i \in I\}$ be classes of events from (Ω, Σ, P) such that they are all mutually independent in the sense of Definition* 1. *If each $\mathcal{B}_i, i \in I$, is a π-class, then for any subset J of I, the generated σ-algebra $\sigma(\mathcal{B}_i, i \in J)$ and $\mathcal{A}$ are independent of each other.*

(b) *Definition* 2 *with $S = \mathbb{R}$ reduces to the statement that for each finite subset $i_1, \ldots, i_n$ of I and random variables $X_{i_1}, \ldots, X_{i_n}$, the collection of events $\{[X_{i_1} < x_1, \ldots, X_{i_n} < x_n], x_j \in \mathbb{R}, j = 1, ..., n, n \geq 1\}$ forms an independent class.*

Proof (a) Let $\mathcal{B} = \sigma(\mathcal{B}_i, i \in J), J \subset I$. If $A \in \mathcal{A}, B_j \in \mathcal{B}_j$,

$$j \in \{j_1, ..., j_n\} \subset J,$$

then

$$\{A, B_{j_1}, ..., B_{j_n}\}$$

are independent by hypothesis, i.e., (1) holds. We need to show that

$$P(A \cap B) = P(A)P(B), \quad A \in \mathcal{A}, B \in \mathcal{B}. \tag{2}$$

If B is of the form $B_1 \cap \ldots \cap B_m$, where $B_i \in \mathcal{B}_i$, $i \in J$, then (2) holds by (1). Let $\mathcal{D}$ be the collection of all sets B which are finite intersections of sets each belonging to a $\mathcal{B}_j, j \in J$. Since each $\mathcal{B}_j$ is a π-class, it follows that $\mathcal{D}$ is also a π-class, and by the preceding observation, (2) holds for $\mathcal{A}$ and $\mathcal{D}$, so that they are independent. Also it is clear that $\mathcal{B}_j \subset \mathcal{D}, j \in J$. Thus $\sigma(\mathcal{B}_j, j \in J) \subset \sigma(\mathcal{D})$. We establish (2) for $\mathcal{A}$ and $\sigma(\mathcal{D})$ to complete the proof of this part, and it involves another idea often used in the subject in similar arguments.

Define a class $\mathcal{G}$ as follows:

$$\mathcal{G} = \{B \in \sigma(\mathcal{D}) : P(A \cap B) = P(A)P(B), A \in \mathcal{A}\}. \tag{3}$$

Evidently $\mathcal{D} \subset \mathcal{G}$. Also $\Omega \in \mathcal{G}$, and if $B_1, B_2 \in \mathcal{G}$ with $B_1 \cap B_2 = \emptyset$, then

$$\begin{aligned} P((B_1 \cup B_2) \cap A) &= P(B_1 \cap A) + P(B_2 \cap A) \quad \text{(since the } B_i \cap A \text{ are disjoint)} \\ &= P(B_1)P(A) + P(B_2)P(A) \quad [\text{ by definition of (3)}] \\ &= P(B_1 \cup B_2)P(A). \end{aligned}$$

Hence $B_1 \cup B_2 \in \mathcal{G}$. Similarly if $B_1 \supset B_2$, $B_i \in \mathcal{G}$, then

$$\begin{aligned} P((B_1 - B_2) \cap A) &= P(B_1 \cap A) - P(B_2 \cap A) \text{ (since } B_1 \cap A \supset B_2 \cap A) \\ &= (P(B_1) - P(B_2))P(A) \\ &= P(B_1 - B_2)P(A). \end{aligned}$$

Thus $B_1 - B_2 \in \mathcal{G}$. Finally, if $B_n \in \mathcal{G}, B_n \subset B_{n+1}$, we can show, from the fact that P is σ-additive, that $\lim_n B_n = \cup_{n \geq 1} B_n \in \mathcal{G}$. Hence $\mathcal{G}$ is a λ-class. Since $\mathcal{G} \supset \mathcal{D}$, by Proposition 1.2.8b, $\mathcal{G} \supset \sigma(\mathcal{D})$. But (3) implies $\mathcal{G}$ and $\mathcal{A}$ are independent. Thus $\mathcal{A}$ and $\sigma(\mathcal{D})$ are independent also, as asserted. Note that since $J \subset I$ is an arbitrary subset, we need the full hypothesis that $\{\mathcal{A}, \mathcal{B}_i, i \in I\}$ is a mutually independent collection, and *not* a mere two-by-two independence.

(b) It is clear that Definition 2 implies the statement here. Conversely, let $\mathcal{B}_1$ be the collection of sets $\{[X_{i_1} < x], x \in \mathbb{R}\}$, and

$$\mathcal{B}_2 = \left\{ \bigcup_{j=2}^{n} [X_{i_j} < x_j], x_j \in \mathbb{R} \right\}$$

It is evident that $\mathcal{B}_1$ and $\mathcal{B}_2$ are π-classes. Indeed,

$$[X_{i_1} < x] \cap [X_{i_1} < y] = [X_{i_1} < \min(x, y)] \in \mathcal{B}_1,$$

and similarly for $\mathcal{B}_2$. Hence by (a), $\mathcal{B}_1$ and $\sigma(\mathcal{B}_2)$ are independent. Since $\mathcal{B}_1$ is a π-class, we also get, by (a) again, that $\sigma(\mathcal{B}_1)$ and $\sigma(\mathcal{B}_2)$ are independent. But $\sigma(\mathcal{B}_1) = X_{i_1}^{-1}(\mathcal{R})[= \sigma(X_{i_1})]$, and $\sigma(\mathcal{B}_2) = \sigma(\cup_{j=2}^{n} X_{i_j}^{-1}(\mathcal{R}))[= \sigma(X_{i_2}, \ldots, X_{i_n})]$, where $\mathcal{R}$ is the Borel σ-algebra of $\mathbb{R}$.

Hence if $A_1 \in \sigma(X_{i_1}), A_j \in X_{i_j}^{-1}(\mathcal{R})(= \sigma(X_{i_j})) \subset \sigma(\mathcal{B}_2)$, then A_1 and $\{A_2, ..., A_n\}$ are independent. Thus

$$P(A_1 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2 \cap \ldots \cap A_n). \tag{4}$$

Next consider X_{i_2} and $(X_{i_3}, \ldots, X_{i_n})$. The above argument can be applied to get

$$P(A_2 \cap \ldots \cap A_n) = P(A_2) \cdot P(A_3 \cap \ldots \cap A_n).$$

Continuing this finitely many times and substituting in (4), we get (1). Hence Definition 2 holds. This completes the proof.

The above result says that we can obtain (1) for random variables if we assume the apparently weaker condition in part (b) of the above theorem. This is particularly useful in computations. Let us record some consequences.

Corollary 4 *Let $\{\mathcal{B}_i, i \in I\}$ be an arbitrary collection of mutually independent π-classes in (Ω, Σ, P), and $J_i \subset I, J_1 \cap J_2 = \emptyset$. If*

$$\mathcal{G}_i = \sigma(\mathcal{B}_j, j \in J_i), i = 1, 2,$$

then $\mathcal{G}_1$ and $\mathcal{G}_2$ are independent. The same is true if $\mathcal{G}_i' = \pi(\mathcal{B}_j, j \in J_i), i = 1, 2$, are the generated π-classes.

If X, Y are independent random variables, f, g are any pair of real Borel functions on $\mathbb{R}$, then $f \circ X$, $g \circ Y$ are also independent random variables. This is because $(f \circ X)^{-1}(\mathcal{R}) = X^{-1}(f^{-1}(\mathcal{R})) \subset X^{-1}(\mathcal{R})$, and similarly $(g \circ Y)^{-1}(\mathcal{R}) \subset Y^{-1}(\mathcal{R})$; and $X^{-1}(\mathcal{R}), Y^{-1}(\mathcal{R})$ are independent σ-subalgebras of Σ. The same argument leads to the following:

Corollary 5 *If $X_1, \ldots, X_n$ are mutually independent random variables on (Ω, Σ, P) and $f : \mathbb{R}^k \to \mathbb{R}$, $g : \mathbb{R}^{n-k} \to \mathbb{R}$ are any Borel functions, then the random variables $f(X_1, \ldots, X_k)$, $g(X_{k+1}, \ldots, X_n)$ are independent; and $\sigma(X_1, \ldots, X_k)$, $\sigma(X_{k+1}, \ldots, X_n)$ are independent σ-algebras, for any $k \geq 1$.*

Another consequence relates to distribution functions and expectations when the latter exist.

Corollary 6 *If $X_1 \ldots X_n$ are independent random variables on (Ω, Σ, P), then their joint distribution is the product of their individual distributions:*

$$\begin{aligned} F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) &= P[X_1 < x_1,\ldots,X_n < x_n] \\ &= \prod_{i=1}^{n} P[X_i < x_i] \\ &= \prod_{i=1}^{n} F_{X_i}(x_i), x_i \in \mathbb{R}. \end{aligned} \tag{5}$$

If, moreover, each of the random variables is integrable, then their product is integrable and we have

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i). \tag{6}$$

Proof By Theorem 3b, (1) and (5) is each equivalent to independence, and so the image functions $F_{X_1,\ldots,X_n}$ and $\prod_{i=1}^{n} F_{X_i}$ are identical. In Definition 2.2 the distribution function of a single random variable is given. The same holds for a (finite) random vector, and $F_{X_1,\ldots,X_n}$ is termed a *joint distribution function* of $X_1,\ldots,X_n$. The result on image measures (Theorem 1.4.1) connects the integrals on the Ω-space with those on $\mathbb{R}^n$, the range space of $(X_1,\ldots,X_n)$.

We now prove (6). Taking $f(x) = |x|$, $f : \mathbb{R} \to \mathbb{R}^+$ being a Borel function, by Corollary 5, $|X_1|,\ldots,|X_n|$ are also mutually independent. Then by (5) and Tonelli's theorem,

$$\begin{aligned} E\left(\prod_{i=1}^{n} |X_i|\right) &= \int_{\Omega} |X_1|\ldots|X_n| dP \\ &= \int_{\mathbb{R}^n_+} x_1 \ldots x_n dG_{|X_1|,\ldots,|X_n|}(x_1,\ldots,x_n), \\ &\qquad \text{[by Theorem 1.4.1i with } G \text{ as the image law]} \\ &= \int_{\mathbb{R}^n_+} x_1 \ldots x_n dG_{|X_1|}(x_1)\ldots, dG_{|X_n|}(x_n), \ \text{[by (5)]}, \\ &= \prod_{i=1}^{n} \int_{\mathbb{R}^+} x_i dG_{|X_i|(x_i)}, \ \text{(by Tonelli's theorem)} \\ &= \prod_{i=1}^{n} E(|X_i|), \ \text{[by Theorem 1.4.1i]}. \end{aligned} \tag{7}$$

Since the right side is finite by hypothesis, so is the left side. Now that $\prod_{i=1}^n = |X_i|$ is integrable we can use the same computation above for X_i and $F_{X_1,\ldots,X_n}(= \prod_{i=1}^n F_{X_i})$, and this time use Fubini's theorem in place of Tonelli's. Then we get (6) in place of (7). This proves the result.

Note. It must be remembered that a direct application of Fubini's theorem is *not* possible in the above argument since the integrability of $|\prod_{i=1}^n X_i|$ has to be established *first* for this result (cf. Theorem 1.3.11). In this task we need Tonelli's theorem for nonnegative random variables, and thus the proof cannot be shortened. Alternatively, one can prove (6) first for simple random variables with Theorem 3b, and then use the Lebesgue monotone (or dominated) convergence theorem, essentially repeating part of the proof for Tonelli's theorem.

We shall now establish one of the most surprising consequences of the independence concept, the zero-one law. If $X_1, X_2, \ldots$ is a sequence of random variables, then $\bigcap_{n=1}^\infty \sigma(X_i, i \geq n)$ is called the *tail σ-algebra* of $\{X_n, n \geq 1\}$.

Theorem 7 (Kolmogorov's Zero-One Law) *Any event belonging to the tail σ-algebra of a sequence of independent random variables on (Ω, Σ, P) has probability either zero or one.*

Proof Denote by $\mathcal{T} = \bigcap_{n=1}^\infty \sigma(X_k, k \geq n)$, the tail σ-algebra of the sequence. Then by Theorem 3a, $\sigma(X_n)$ and $\sigma(X_k, k \geq n+1)$ are independent σ-algebras for each $n \geq 1$. But $\mathcal{T} \subset \sigma(X_k, k \geq n+1)$, so that $\sigma(X_n)$ and $\mathcal{T}$ are independent for each n. By Theorem 3a again $\mathcal{T}$ is independent of $\sigma(\sigma(X_n), n \geq 1) = \sigma(X_n, n \geq 1)$. However, $\mathcal{T} \subset \sigma(X_n, n \geq 1)$ also, so that $\mathcal{T}$ is independent of itself! Hence $A \in \mathcal{T}$ implies

$$P(A) = P(A \cap A) = P(A)P(A) = P(A)^2;$$

thus we must have $P(A) = 0$ or 1, completing the proof.

An immediate consequence is that any function measurable relative to $\mathcal{T}$ of the theorem must be a constant with probability one. Thus $\limsup_n X_n$, $\liminf_n X_n$ (and $\lim_n X_n$ itself, if this exists) of independent random variables are constants with probability one. Similarly if

$$A_n = \{\omega : |\sum_{k \geq n} X_k(\omega)| < \infty\},$$

then $\sum_{n=1}^\infty X_n(\omega)$ converges iff $\omega \in A_n$ for each n, i.e., iff $\omega \in A = \bigcap_{n=1}^\infty A_n$. Since clearly $A_n \in \sigma(X_k, k \geq n), A \in \mathcal{T}$, so that $P(A) = 0$ or 1. Thus for independent X_n the series $\sum_{n=1}^\infty X_n$ converges with probability 0 or 1. The following form of the above theorem is given in Tucker (1967).

Corollary 8 *Let I be an arbitrary infinite index set, and $\{X_i, i \in I\}$ be a family of independent random variables on (Ω, Σ, P). If $\mathcal{F}$ is the directed (by inclusion) set of all finite subsets of I, the (generalized) tail σ-algebra is defined as*

$$\mathcal{T}_0 = \bigcap \{\sigma(X_i, i \notin J) : J \in \mathcal{F}\}. \tag{8}$$

Then P takes only 0 and 1 values on $\mathcal{T}_0$.

Proof The argument is similar to that of the theorem. Note that $\mathcal{T}_0$ and $\mathcal{B}_J = \sigma(X_i, i \in J)$ are independent for each $J \in \mathcal{F}$, as in the above proof. So by Theorem 3a, $\mathcal{T}_0$ and $\mathcal{B} = \sigma(\mathcal{B}_J, J \in \mathcal{F})$ are independent. But clearly $\mathcal{B} = \sigma(X_i, i \in I)$, so that $\mathcal{T}_0 \subset \mathcal{B}$. Hence the result follows as before.

Let us now show that independent random variables can be assumed to exist on a probability space by a process of *enlargement of the space by adjunction.* The procedure is as follows: Let (Ω, Σ, P) be a probability space. If this is not rich enough, let $(\Omega_i, \Sigma_i, P_i), i = 1, ..., n$, be n copies of the given space. Let $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{P}) = (\times_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \Sigma_i, \bigotimes_{i=1}^n P_i)$ be their Cartesian product. If $X_1, ..., X_n$ are random variables on (Ω, Σ, P), define a "new" set of functions $\tilde{X}_1, ..., \tilde{X}_n$ on (Ω, Σ, P) by the equations

$$\tilde{X}_i(\omega) = X_i(\omega_i), \quad \omega = (\omega_1, ..., \omega_n) \in \tilde{\Omega}, \quad i = 1, ..., n.$$

Then for each $a \in \mathbb{R}$,

$$\begin{aligned}\{\omega : \tilde{X}_i(\omega) < a\} &= \{\omega : X_i(\omega_i) < a\} \\ &= \Omega_1 \times \ldots \times \Omega_{i-1} \times [X_i < a] \times \Omega_{i+1} \times \ldots \times \Omega_n,\end{aligned}$$

which is a measurable rectangle and hence is in $\tilde{\Sigma}$. Thus $\tilde{X}_i$ is a random variable. Also, since $P_i = P$, we deduce that

$$\tilde{P}[\tilde{X}_1 < a_1, \tilde{X}_2 < a_2, ..., \tilde{X}_n < a_n] = \prod_{i=1}^n P[X_i < a_i], \tag{9}$$

by Fubini's theorem and the fact that $P_i(\Omega_i) = 1$. Consequently the $\tilde{X}_i$ are independent (cf. Theorem 3b) and each $\tilde{X}_i$ has the same distribution as X_i. Thus by enlargement of (Ω, Σ, P) to $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{P})$, we have n independent random variables. This procedure can be employed for the existence of any finite collection of independent random variables without altering the probability structure (see also Problem 5 (a)). The results of Section 3.4 establishing the Kolmogorov- Bochner theorem will show that this enlargement can be used for *any* collection of random variables (countable or not). Consequently, we can and do develop the theory without any question of the richness of the underlying σ-algebra or of the existence of families of independent random variables.

The following elementary but powerful results, known as the *Borel-Cantelli lemmas*, are true even for the weaker pairwise independent events. Recall that $\limsup_n A_n = \{\omega : \omega \in A_n \text{ for infinitely many } n\}$. This set is abbreviated as $\{A_n, \text{ i.o.}\}$ $[= \{A_n \text{ occurs infinitely often}\}]$.

Theorem 9 (i) (First Borel-Cantelli Lemma). *Let $\{A_n, n \geq 1\}$ be a sequence of events in (Ω, Σ, P) such that $\sum_{n=1}^{\infty} P(A_n) < \infty$. Then*

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

(ii) (Second Borel-Cantelli Lemma). *Let $\{A_n, n \geq 1\}$ be a sequence of pairwise independent events in (Ω, Σ, P) such that $\sum_{n=1}^{\infty} P(A_n) = \infty$. Then $P(A_n, \text{ i.o.}) = 1$.*

(iii) *In particular, if $\{A_n, n \geq 1\}$ is a sequence of (pairwise or mutually) independent events, then $P(A_n, \text{ i.o.}) = 0$ or 1 according to whether $\sum_{n=1}^{\infty} P(A_n)$ is $< \infty$ or $= \infty$.*

Proof (i) This simple result is used more often than the other more involved parts, since the events need not be (even pairwise) independent. By definition, $A = \limsup_n A_n = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k \subseteq \bigcup_{k \geq n} A_k$ for all $n \geq 1$. Hence by the σ-subadditivity of P, we have

$$P(A) \leq P(\bigcup_{k \geq n} A_k) \leq \sum_{k=n}^{\infty} P(A_k), n \geq 1.$$

Letting $n \to \infty$, and using the convergence of the series $\sum_{k=1}^{\infty} P(A_k)$, the result follows.

(ii) (After Chung, 1974) Let $\{A_n, n \geq 1\}$ be *pairwise* independent. By Problem 1 of Chapter 1, we have

$$A = [A_n, \text{ i.o.}] \quad \text{iff} \quad \chi_A = \limsup_n \chi_{A_n}.$$

Hence

$$P(A) = 1 \quad \text{iff} \quad P[\limsup_n \chi_{A_n} = 1] = 1.$$

Thus

$$P([\chi_{A_n} = 1, \text{ i.o. }]) = 1 \quad \text{iff} \quad P\left(\left[\sum_{i=1}^{\infty} \chi_{A_n} = \infty\right]\right) = 1. \tag{10}$$

Now we use the hypothesis that the series diverges:

$$\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} E(\chi_{A_n}) = \lim_{n \to \infty} E(S_n) = +\infty, \tag{11}$$

where $S_n = \sum_{k=1}^{n} \chi_{A_k}$ and the monotonicity of S_n is used above. With (11) and the pairwise independence of A_n, we shall show that

$$P([\lim_{n\to\infty} S_n = \infty]) = 1,$$

which in view of (10) proves the assertion.

Now given $N > 0$, we have by Čebyšev's inequality, with $\varepsilon = N\sqrt{VarS_n}$,

$$P[|S_n - E(S_n)| > \varepsilon] \leq \frac{VarS_n}{\varepsilon^2} = \frac{1}{N^2}.$$

Equivalently,

$$P[E(S_n) - N\sqrt{VarS_n} \leq S_n \leq E(S_n) + N\sqrt{VarS_n}] \geq 1 - \frac{1}{N^2}. \tag{12}$$

To simplify this we need to evaluate $Var\ S_n$. Let $p_n = P(A_n)$. Then

$$E(S_n) = \sum_{k=1}^{n} E(\chi_{A_k}) = \sum_{k=1}^{n} p_k. \tag{13}$$

If $I_n = \chi_{A_n} - p_n$, then the I_n are orthogonal random variables. In fact, using the inner product notation,

$$\begin{aligned}(I_n, I_m) &= \int_{\Omega} (\chi_{A_n} - p_n)(\chi_{A_m} - p_m)dP \\ &= P(A_n \cap A_m) - p_n p_m = 0, \quad \text{if } n \neq m \text{ (by pairwise independence)} \\ &= p_n(1 - p_n), \quad \text{if } n = m. \end{aligned} \tag{14}$$

Thus

$$\begin{aligned} VarS_n = E(S_n - E(S_n))^2 &= E\left(\sum_{k=1}^{n} I_k\right)^2 \\ &= \sum_{k=1}^{n} E(I_k^2) = \sum_{k=1}^{n} p_k(1 - p_k), \quad [\text{by (14) }], \\ &\leq \sum_{k=1}^{n} p_k = E(S_n), \quad [\text{by (13) }]. \end{aligned} \tag{15}$$

Since by (11) $E(S_n) \nearrow \infty$, (15) yields $\sqrt{VarS_n/E(S_n)} \leq (E(S_n))^{-1/2} \to 0$. Thus given $N > 1$, and $0 < \alpha_1 = \alpha/N$ for $0 < \alpha < 1$, there exists $n_0 = n_0(\alpha, N)$ such that $n \geq n_0 \Rightarrow \sqrt{VarS_n/E(S_n)} \leq \alpha_1 < 1$. Since $\alpha_1 = \alpha/N$, we get

$$N\sqrt{VarS_n} \leq \alpha E(S_n), \quad n \geq n_0. \tag{16}$$

Consequently (12) implies, with $1 > \beta = 1 - \alpha > 0$ and the monotonicity of S_n, (i.e., $S_n \uparrow$)

$$P[\beta E(S_n) < \lim_n S_n] \geq P[\beta E(S_n) < S_n] \geq 1 - \frac{1}{N^2}, \quad n \geq n_0. \tag{17}$$

Let $n \to \infty$, and then $N \to \infty$ (so that $\beta \to 1$); (17) gives $P[\lim_{n\to\infty} S_n = \infty] = 1$. This establishes the result because of (10).

(iii) This is an immediate consequence of (ii), and again gives a zero-one phenomenon! However, in the case of mutual independence, the proof is simpler than that of (ii), and we give the easy argument here, for variety. Let $A_n^c = B_n$. Then $B_n, n \geq 1$, are independent, since $\{\sigma(A_n), n \geq 1\}$ forms an independent class. Let $P(A_n) = \alpha_n$. To show that

$$P[A_n, \text{ i.o. }] = P\left[\bigcap_{k\geq 1}\bigcup_{n\geq k} A_n\right] = 1,$$

it suffices to verify that, for each $n \geq 1$, $P(\bigcup_{k>n} A_k) = 1$, or equivalently

$$P\left[\bigcap_{k>n} B_k\right] = 0, \quad n \geq 1.$$

Now for any $n \geq 1$,

$$\begin{aligned}
P\left[\bigcap_{k>n} B_k\right] &= \lim_{m\to\infty} P\left[\bigcap_{k=n+1}^{m} B_k\right] \\
&= \lim_{m\to\infty} \prod_{k=n+1}^{m} (1-\alpha_k) \quad \text{(by independence of } B_k\text{)} \\
&\leq \prod_{k=n+1}^{\infty} e^{-\alpha_k} \quad \left(\text{since } x \geq \int_0^x e^{-t}dt = 1 - e^{-x}\ \right) \\
&= \exp\left(-\sum_{k=n+1}^{\infty} \alpha_k\right) = 0 \quad \left(\text{ since } \sum_{k=1}^{\infty} \alpha_k = \infty \text{ by hypothesis}\right).
\end{aligned}$$

This completes the proof of the theorem.

Note 10 The estimates in the proof of (ii) yield a stronger statement than we have asserted. *One can actually show that*

$$P[\lim_{n\to\infty} \frac{S_n}{E(S_n)} = 1] = 1.$$

In fact, (12) implies for each n and N,

$$P\left[\frac{S_n}{E(S_n)} \leq 1 + \frac{N\sqrt{VarS_n}}{E(S_n)}\right] \geq 1 - \frac{1}{N^2}.$$

Since $VarS_n/(E(S_n))^2 \to 0$ as $n \to \infty$, for each fixed N, this gives

$$p[\limsup_n \frac{S_n}{E(S_n)} \leq 1] \geq 1 - \frac{1}{N^2}, \tag{18}$$

and letting $N \to \infty$, we get

$$\limsup_n \frac{S_n}{E(S_n)} \leq 1 \quad \text{a.e.}$$

On the other hand by (17), $P[\beta \leq S_n/E(S_n)] \geq 1 - 1/N^2$, $n \geq n_0$. Hence for each fixed N, this yields

$$P\left[\beta \leq \liminf_n \frac{S_n}{E(S_n)}\right] \geq 1 - \frac{1}{N^2}. \tag{19}$$

Now let $N \to \infty$ and note that $\beta \to 1$; then by the monotonicity of events in brackets of (19) we get $1 \leq \liminf_n [S_n/E(S_n)]$ a.e. These two statements imply the assertion.

Before leaving this section, we present, under a stronger hypothesis than that of Theorem 7, a zero-one law due to Hewitt and Savage (1955), which is useful in applications. We include a short proof as in Feller (1966).

Definition 11 If $X_1, \ldots, X_n$ are random variables on (Ω, Σ, P), then they are *symmetric* (or *symmetrically dependent*) if for each permutation $i_1, \ldots, i_n$ of $(1, 2, \ldots, n)$, the vectors $(X_{i_1}, \ldots, X_{i_n})$ and $(X_1, \ldots, X_n)$ have the same joint distribution. A sequence $\{X_n, n \geq 1\}$ is *symmetric* if $\{X_k, 1 \leq k \leq n\}$ is symmetric for each $n \geq 1$.

We want to consider some functions of $X = \{X_n, n \geq 1\}$. Now $X : \Omega \to \mathbb{R}^\infty = \times_{i=1}^\infty \mathbb{R}_i$, where $\mathbb{R}_i \equiv \mathbb{R}$ is an infinite vector. If $\mathcal{B}^\infty = \otimes_{i=1}^\infty \mathcal{B}_i$ is the (usual) product σ-algebra, then

$$\Sigma_0 = \sigma(X_n, n \geq 1) = X^{-1}(\mathcal{B}^\infty) = \sigma\left(\bigcup_{n=1}^\infty \sigma(X_1, \ldots, X_n)\right).$$

Let $g : \Omega \to \mathbb{R}$ be Σ_0-measurable. Then by Proposition 1.2.3 there is a Borel function $h : \mathbb{R}^\infty \to \mathbb{R}$ (i.e., h is $\mathcal{B}^\infty$-measurable) such that $g = h \circ X = h(X_1, X_2, \ldots)$. Thus if $\{X_n, n \geq 1\}$ is a symmetric sequence, then each Σ_0-measurable g is symmetric [1], so that

[1] In detail, this means if $g : \mathbb{R}^\infty \to \mathbb{R}$, then $g(X_1, \ldots, X_n, X_{n+1}, \ldots) = g(X_{i_1}, \ldots, X_{i_n}, X_{n+1}, \ldots)$ for each permutation $(i_1, \ldots, i_n)$ of $(1, 2, \ldots, n)$, each $n > 1$.

$$g = h(X_1, X_2, \ldots) = h(X_{i_1} \ldots, X_{i_n}, X_{i_{n+1}}, \ldots)$$

for each finite permutation. Let $A \in \Sigma_0$. Then A is a *symmetric event* if χ_A is a symmetric function in the above sense. The following result is true:

Theorem 12 (Hewitt-Savage Zero-One Law). *If $X_1, X_2, \ldots$ are independent with a common distribution, then every symmetric set in*

$$\Sigma_0 = \sigma(X_n, n \geq 1)$$

has probability zero or one.

Proof Recall that if $\rho : \Sigma_0 \times \Sigma_0 \to \mathbb{R}^+$ defined by $\rho(A, B) = P(A \triangle B)$ with $\triangle$ as symmetric difference, then (Σ, ρ) is a (semi) metric space on which the operations $\cup, \cap$, and $\triangle$ are continuous. Also, $\bigcup_{n=1}^{\infty} \sigma(X_1, \ldots, X_n) \subset \Sigma_0$ is a dense subspace, in this metric.

Hence if $A \in \Sigma_0$, there exists $A_n \in \sigma(X_1, \ldots, X_n)$ such that $\rho(A, A_n) \to 0$, and by the definition of $\sigma(X_1, \ldots, X_n)$ there is a Borel set $B_n \subset \mathbb{R}^n$ such that $A_n = [(X_1, \ldots, X_n) \in B_n]$. Since

$$\overline{X} = (X_{i_1}, \ldots, X_{i_n}, X_{n+1} \ldots) \text{ and } X = (X_1, \ldots, X_n, X_{n+1} \ldots)$$

have the same (finite dimensional) distributions, because the X_n are identically distributed, and we have for any $B \in \mathcal{B}^\infty$, $P(\overline{X} \in B) = P(X \in B)$. In particular, if the permutation is such that $\tilde{A}_n = [(X_{2n}, X_{2n-1}, \ldots, X_{n+1}) \in B_n]$, then A_n and $\tilde{A}_n$ are independent and $\rho(A, \tilde{A}_n) \to 0$ as $n \to \infty$ again.

Indeed, let τ be the 1-1 measurable permutation mapping $\tau A_n = \tilde{A}_n$ and $\tau A = A$ since A is symmetric. So

$$\rho(A, \tilde{A}_n) = \rho(\tau A, \tau A_n) = \rho(A, A_n) \to 0.$$

Hence also $A_n \cap \tilde{A}_n \to A \cap A = A$, by the continuity of $\cap$ in the ρ-metric. But

$$P(A_n \cap \tilde{A}_n) = P(A_n)P(\tilde{A}_n),$$

by independence.

Letting $n \to \infty$, and noting that the metric function is also continuous in the resulting topology, it follows that $A_n \to A$ in $\rho \Rightarrow P(A_n) \to P(A)$. Hence

$$\lim_{n\to\infty} P(A_n \cap \tilde{A}_n) = P(A \cap A) = \lim_{n\to\infty} P(A_n) \cdot P(\tilde{A}_n) = P(A)^2.$$

Thus $P(A) = P(A)^2$ so that $P(A) = 0$ or 1, as asserted.

Remarks (1) It is not difficult to verify that if $S_n = \sum_{k=1}^n X_k$, X_k as in the theorem, then for any Borel set B, the event $[S_n \in B \text{ i.o.}]$ is not necessarily a tail event but is a symmetric one. Thus this is covered by the above theorem, but not by the Kolmogorov zero-one law.

(2) Note 10, as well as part (ii) of Theorem 9, indicate how several weakenings of the independence condition can be formulated. A number of different extensions of Borel-Cantelli lemmas have appeared in the literature, and they are useful for special problems. The point here is that the concept of independence, as given in Definitions 1 and 2, leads to some very striking results, which then motivate the introduction of different types of dependences for a sustained study. In this chapter we present only the basic results founded on the independence hypothesis; later on we discuss how some natural extensions suggest themselves.

2.2 Convergence Concepts, Series, and Inequalities

There are four convergence concepts often used in probability theory. They are pointwise a.e., in mean, in probability, and in distribution. Some of these have already appeared in Chapter 1. We state them again and give some interrelations here. It turns out that for sums of independent (integrable) random variables, these are all equivalent, but this is a relatively deep result. A partial solution is given in Problem 16. Several inequalities are needed for the proof of the general case. We start with the basic Kolmogorov inequality and a few of its variants. As consequences, some important "strong limit laws" will be established. Applications are given in Section 2.4.

Definition 1 Let $\{X, X_n, n \geq 1\}$ be a family of random variables on a probability space (Ω, Σ, P).

(a) $X_n \to X$ *pointwise a.e.* if there is a set $N \in \Sigma$, $P(N) = 0$ and $X_n(\omega) \to X(\omega)$, as $n \to \infty$, for each $\omega \in \Omega - N$.

(b) The sequence is said to converge to X *in probability* if for each $\varepsilon > 0$, we have $\lim_{n\to\infty} P[|X_n - X| \geq \varepsilon] = 0$, symbolically written as $X_n \xrightarrow{P} X$ (or as $p\lim_n X_n = X$).

(c) The sequence is said to converge *in distribution* to X, often written $X_n \xrightarrow{D} X$ if $F_{X_n}(x) \to F_x(x)$ at all points $x \in \mathbb{R}$ for which x is a continuity point of F_X, where F_{X_n}, F_X are distribution functions of X_n and X (cf. Definition 2.2).

(d) Finally, if $\{X, X_n, n \geq 1\}$ have p-moments, $0 < p < \infty$, then the sequence is said to tend to X in *pth order mean*, written $X_n \xrightarrow{\mathcal{L}^p} X$, if $E(|X_n - X|^p) \to 0$. If $p = 1$, we simply say that $X_n \to X$ in mean.

The first two as well as the last convergences have already appeared, and these are defined and profitably employed in general analysis on arbitrary measure spaces. However, on finite measure spaces there are some additional relations which are of particular interest in our study. The third concept, on the other hand, is somewhat special to probability theory since distribution functions are image probability measures on $\mathbb{R}$. This plays a pivotal role in probability theory, and so we study the concept in greater detail.

Some amplification of the conditions for "in distribution" is in order. If $X = a$ a.e., then $F_X(x) = 0$ for $x < a$, $= 1$ for $x \geq a$. Thus we are asking that for $X_n \xrightarrow{D} a$, $F_{X_n}(x) \to F_x(x)$ for $x < a$ and for $x > a$ but not at $x = a$, the discontinuity point of F_X. Why? The restriction on the set is that it should be only a "continuity set" for the limit function F_X. This condition is arrived at after noting the "natural-looking" conditions proved themselves useless. For instance, if $X_n = a_n$ a.e., and $a_n \to a$ as numbers, then $F_{X_n}(x) \to F_X(x)$ for all $x \in \mathbb{R} - \{a\}$, but $F_{X_n}(a) \not\to F_X(a)$, since $\{F_{X_n}(a), n \geq 1\}$ is an oscillating sequence if there are infinitely many n on both sides of a. Similarly, if $\{X_n = a_n, n \geq 1\}$ diverges, it is possible that $\{F_{X_n}(x), n \geq 1\}$ may converge for each $x \in \mathbb{R}$ to a function taking values in the open interval $(0, 1)$. Other unwanted exclusions may appear. Thus the stipulated condition is weak enough to ignore such uninteresting behavior. But it is not too weak, since we do want the convergence on a suitable dense set of $\mathbb{R}$. (Note that the set of discontinuity points of a monotone function is at most countable, so that the continuity set of F_X is $\mathbb{R} - \{$ that countable set $\}$.) Actually, the condition comes from the so-called simple convergence on $C_{00}(\mathbb{R})$, the space of continuous functions with compact supports, which translates to the condition we gave for the distribution functions on $\mathbb{R}$ according to a theorem in abstract analysis. For this reason N. Bourbaki actually calls it the *vague convergence*, and others call it the *weak-star convergence*. We shall use the terminology introduced in the definition and the later work shows how these last two terms can also be justifiably used.

The first three convergences are related as follows:

Proposition 2 *Let X_n and X be random variables on (Ω, Σ, P). Then $X_n \to X$ a.e. $\Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{D} X$. If, moreover, $X = a$ a.e., where $a \in \mathbb{R}$, then $X_n \xrightarrow{D} X \Rightarrow X_n \xrightarrow{P} X$ also. In general these implications are not reversible.* (Here, as usual, the limits are taken as $n \to \infty$.)

Proof The first implication is a standard result for any finite measure. In fact, if $X_n \to X$ a.e., then there is a set $N \in \Sigma$, $P(N) = 0$, and on $\Omega - N$, $X_n(\omega) \to X(\omega)$. Thus $\limsup_n X_n(\omega) = X(\omega)$, $\omega \in \Omega - N$, and for each $\varepsilon > 0$,

$$\left\{\omega : \limsup_n |X_n - X|(\omega) > \varepsilon\right\} \subset \bigcap_{k\geq 1} \bigcup_{n\geq k} \{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\} \subset N.$$

Hence the set has measure zero. Since P is a finite measure, this implies

$$P\left(\lim_{k\to\infty}\bigcup_{n\geq k}\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}\right) = \lim_{k\to\infty} P\left(\bigcup_{n\geq k}[|X_n - X| > \varepsilon]\right)$$
$$\leq P(N) = 0. \tag{1}$$

Consequently,

$$P([|X_n - X| > \varepsilon]) \leq P\left(\bigcup_{j\geq n}[|X_j - X| > \varepsilon]\right) \to 0 \text{ as } n \to \infty. \tag{2}$$

Thus $X_n \xrightarrow{P} X$, and the first assertion is proved.

For the next implication, let F_X, F_{X_n} be the distribution functions of X and X_n, and let a, b be continuity points of F_X with $a < b$. Then

$$[X < a] = [X < a, X_n < b] \cup [X < a, X_n \geq b]$$
$$\subset [X_n < b] \cup [X < a, X_n \geq b],$$

so that computing probabilities of these sets gives

$$F_X(a) \leq F_{X_n}(b) + P[X < a, X_n \geq b]. \tag{3}$$

Also, since $X_n \xrightarrow{P} X$, with $\varepsilon = b - a > 0$, one has from the inclusion

$$[X < a, X_n \geq b] \subset [|X_n - X| \geq b - a],$$
$$\lim_n P[X < a, X_n \geq b] = 0. \tag{4}$$

Thus (3) becomes

$$F_X(a) \leq \liminf_n F_{X_n}(b). \tag{5}$$

Next, by an identical computation, but with c, d $(c < d)$ in place of a, b and X_n, X in place of X, X_n in (3), one gets

$$F_{X_n}(c) \leq F_X(d) + P[X_n < c, X \geq d]. \tag{6}$$

The last term tends to zero as $n \to \infty$, as in (4). Consequently (6) becomes

$$\limsup_n F_{X_n}(c) \leq F_X(d). \tag{7}$$

From (5) and (7) we get for $a < b \leq c < d$,

$$F_X(a) \leq \liminf_n F_{X_n}(b) \leq \limsup_n F_{X_n}(b)$$
$$\leq \limsup_n F_{X_n}(c) \leq F_X(d). \tag{8}$$

Letting $a \uparrow b = c$ and $d \downarrow c$, where $b = c$ is a continuity point of F_X, (8) gives $\lim_n F_{X_n}(b) = F_X(b)$, so that $X_n \xrightarrow{D} X$, since such points of continuity of F_X are everywhere dense in $\mathbb{R}$.

If now $X = \alpha$ a.e., then for each $\varepsilon > 0$,

$$\begin{aligned} P([|X_n - \alpha| \geq \varepsilon]) &= P([X_n \geq \alpha + \varepsilon]) + P([X_n \leq \alpha - \varepsilon]) \\ &= 1 - F_{X_n}(\alpha + \varepsilon) + F_{X_n}(\alpha - \varepsilon) \to 0 \text{ as } n \to \infty, \end{aligned}$$

since

$$F_{X_n}(x) \to F_X(x) = \begin{cases} 0, & x < \alpha \\ 1, & x \geq \alpha, \end{cases}$$

and $\alpha \pm \varepsilon$ are points of continuity of F_X for each $\varepsilon > 0$. Thus $X \xrightarrow{P} \alpha$. This completes the proof except for the last comment, which is illustrated by the following simple pair of standard counter-examples.

Let X_n, X be defined on (Ω, Σ, P) as two-valued random variables such that $P([X_n = a]) = \frac{1}{2} = P([X_n = b]), a < b$, for all n. Next let $P([X = b]) = \frac{1}{2} = P([X = a])$. Then for each n, $\omega \in \Omega$, for which $X_n(\omega) = a$ (or b), we set $X(\omega) = b$ (or a), respectively. Thus $\{\omega : |X_n - X|(\omega) \geq \varepsilon\} = \Omega$ if $0 < \varepsilon < b - a$, and $X_n \not\to X$ in probability. But $F_{X_n} = F_X$, so that $X_n \xrightarrow{D} X$ trivially. This shows that the last implication cannot be reversed in general. Next, consider the first one. Let $\Omega = [0, 1]$, $\Sigma =$ Borel σ-algebra of Ω, and $P =$ Lebesgue measure. For each $n > 1$, express n in a binary expansion, $n = 2^r + k$, $0 \leq k \leq 2^r$, $r \geq 0$. Define $f_n = \chi_{A_n}$, where $A_n = [k/2^r, (k+1)/2^r]$. It is clear that f_n is measurable, and for $0 < \varepsilon < 1$,

$$P[|f_n - 0| > \varepsilon] \leq \frac{1}{2^r} < \frac{2}{n} \to 0.$$

But $f_n(\omega) \not\to 0$ for any $\omega \in \Omega$. This establishes all assertions. (If we are allowed to change probability spaces, keeping the same image measures of the random variables, these problems become less significant. Cf. Problem 5 (b).)

In spite of the last part, we shall be able to prove the equivalence to a subclass of random variables, namely, if the X_n form a sequence of partial sums of *independent* random variables. For this result we need to develop probability theory much further, and thus it is postponed until Chapter 4. (For a partial result, see Problem 16.) Here we proceed with the implications that do not refer to "convergence in distribution."

The following result is of interest in many calculations.

Proposition 3 (**F. Riesz**). *Let $\{X, X_n, n \geq 1\}$ be random variables on (Ω, Σ, P) such that $X_n \xrightarrow{P} X$. Then there exists a subsequence $\{X_{n_k}, k \geq 1\}$ with $X_{n_k} \to X$ a.e. as $k \to \infty$.*

Proof Since for each $\varepsilon > 0$, $P[|X_n - X| \geq \varepsilon] \to 0$, let n_1 be chosen such that $n \geq n_1 \Rightarrow P[|X_n - X| \geq 1] < \frac{1}{2}$,and if $n_1 < n_2 < \ldots < n_k$ are selected, let $n_{k+1} > n_k$ be chosen such that

$$P[|X_n - X| \geq 1/2^k] < 1/2^{k+1}, \quad n \geq n_{k+1}. \tag{9}$$

If $A_k = [|X_{n_k} - X| \geq 1/2^{k-1}]$, $B_k = \bigcup_{n \geq k} A_n$, then for $\omega \in B_k^c$, $|X_{n_r} - X|(\omega) < 1/2^{r-1}$ for all $r \geq k$. Hence if $B = \lim_n B_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k$, then for $\omega \in B^c$, $X_{n_r}(\omega) \to X(\omega)$ as $r \to \infty$. But we also have $B \subset B_n$ for all n, so that

$$P(B) \leq P(B_n) \leq \sum_{k \geq n} P(A_k) < \sum_{k \geq n} 2^{-k} = 2^{-n+1} \to 0 \text{ as } n \to \infty.$$

Thus $\{X_{n_r}, r \geq 1\}$ is the desired subsequence, completing the proof.

Remark We have not used the finiteness of P in the above proof, and the result holds on nonfinite measure spaces as well. (Also there can be infinitely many such a.e. convergent subsequences.) But the next result is strictly for (finite or) probability measures only.

Recall that a sequence $\{X_n, n \geq 1\}$ on (Ω, Σ, P) *converges P-uniformly* to X if for each $\varepsilon > 0$, there is a set $A_\varepsilon \in \Sigma$ such that $P(A_\varepsilon) < \varepsilon$ and on $\Omega - A_\varepsilon$, $X_n \to X$ uniformly. We then have

Theorem 4 **(Egorov)**. *Let $\{X, X_n, n \geq 1\}$ be a sequence of random variables on (Ω, Σ, P). Then $X_n \to X$ a.e. iff the sequence converges to X P-uniformly.*

Proof One direction is simple. In fact, if $X_n \to X$ P-uniformly, then for $\varepsilon = 1/n_0$ there is an $A_{n_0} \in \Sigma$ with $P(A_{n_0}) < 1/n_0$ and $X_n(\omega) \to X(\omega)$ uniformly on $\Omega - A_{n_0}$. If $A = \bigcap_{n \geq 1} A_n$, then $P(A) = 0$, and if $\omega \in \Omega - A$, then $X_n(\omega) \to X(\omega)$, i.e., the sequence converges a.e. The other direction is non-trivial.

Thus let $X_n \to X$ a.e. Then there is an $N \in \Sigma$, $P(N) = 0$, and $X_n(\omega) \to X(\omega)$ for each $\omega \in \Omega - N$. If $k \geq 1$, $m \geq 1$ are integers and we define

$$A_{k,m} = \{\omega \in \Omega - N : |X_n(\omega) - X(\omega)| < \frac{1}{m} \text{ for all } n \geq k\},$$

then the facts that $X_n \to X$ on $\Omega - N$ and $A_{k,m} \subset A_{k+1,m}$ imply that $\Omega - N = \bigcup_{k=1}^{\infty} A_{k,m}$ for all $m \geq 1$. Consequently for each $\varepsilon > 0$, and each $m \geq 1$, we can find a large enough $k_0 = k_0(\varepsilon, m)$ such that $A_{k_0,m}$ has large measure, i.e., $P(\Omega - A_{k_0,m}) < \varepsilon/2^m$. If $A_\varepsilon = \bigcup_{m=1}^{\infty} A^c_{k_0(\varepsilon,m),m}$ then

$$P(A_\varepsilon) \leq \sum_{m=1}^{\infty} P(A^c_{k_0(\varepsilon,m),m}) < \sum_{m=1}^{\infty} \frac{\varepsilon}{2^m} = \varepsilon.$$

On the other hand, $n \geq k_0(\varepsilon, m) \Rightarrow |X_n(\omega) - X(\omega)| < 1/m$ for $\omega \in A_{k_0,m}$. Thus

$$n \geq k_0(\varepsilon, m) \Rightarrow \sup_{\omega \in A_\varepsilon^c} |X_n(\omega) - X(\omega)| \leq \sup_{\omega \in A_{k_0,m}} |X_n(\omega) - X(\omega)| \leq \frac{1}{m}$$

for every $m \geq 1$, so that $X_n \to X$ uniformly on A_ε^c. This completes the proof.

Also, the following is a simple consequence of Markov's inequality.

Remark *Let* $\{X, X_n, n \geq 1\} \subset \mathcal{L}^p(\Omega, \Sigma, P)$ *such that* $X_n \xrightarrow{\mathcal{L}^p} X$, $p > 0$. *Then* $X_n \xrightarrow{P} X$.

Proof Given $\varepsilon > 0$, we have

$$\begin{aligned} P[|X_n - X| > \varepsilon] &= P[|X_n - X|^p > \varepsilon^p] \\ &\leq \frac{1}{\varepsilon^p} E(|X_n - X|^p) \to 0 \text{ as } n \to \infty, \end{aligned}$$

by the pth mean convergence hypothesis. Note that there is generally no relation between mean convergence and pointwise a.e., since for the latter the random variables need not be in any $\mathcal{L}^p$, $p > 0$.

We now specialize the convergence theory if the sequences are partial sums of *independent* random variables, and present important consequences. Some further, less sharp, assertions in the general case are possible. Some of these are included as problems at the end of the chapter.

At the root of the pointwise convergence theory, there is usually a "maximal inequality," for a set of random variables. Here is a generalized version of Čebyšev's inequality. The latter was proved for only one r.v. We thus start with the fundamental result:

Theorem 5 (Kolmogorov's Inequality). *Let* $X_1, X_2, \ldots$ *be a sequence of independent random variables on* (Ω, Σ, P) *with means* $\mu_k = E(X_k)$ *and variances* $\sigma_k^2 = Var X_k$. *If* $S_n = \sum_{k=1}^n X_k$ *and* $\varepsilon > 0$, *then*

$$P[\max_{1 \leq k \leq n} |S_k - E(S_k)| \geq \varepsilon] \leq \frac{1}{\varepsilon^2} \sum_{k=1}^n \sigma_k^2. \tag{10}$$

Proof If $n = 1$, then (10) is Čebyšev's inequality, but the present result is deeper than the former. The proof shows how the result may be generalized to certain nonindependent cases, particularly to martingale sequences, to be studied in the next chapter.

Let $A = \{\omega : \max_{1 \leq k \leq n} |S_k(\omega) - E(S_k)| \geq \varepsilon\}$. We express A as a disjoint union of n events; such a decomposition appears in our subject on several occasions. It became one of the standard tools. [It is often called a process of *disjunctification* of a compound event such as A.] Thus let

$$A_1 = \{\omega : |S_1(\omega) - E(S_1)| \geq \varepsilon\}$$

and for $1 < k \leq n$,

$$A_k = \{\omega : |S_i(\omega) - E(S_i)| < \varepsilon, 1 \leq i \leq k-1, |S_k(\omega) - E(S_k) \geq \varepsilon\}.$$

In words, A_k is the set of ω such that $|S_k(\omega) - E(S_k)|$ exceeds ε for the first time. It is clear that the A_k are disjoint, $A_k \in \Sigma$, and $A = \bigcup_{k=1}^n A_k$. Let $Y_i = X_i - \mu_i$ and $\tilde{S}_n = \sum_{k=1}^n Y_k$, so that $E(\tilde{S}_n) = 0$, $Var\tilde{S}_n = VarS_n$. Now consider

$$\begin{aligned}
\int_{A_k} \tilde{S}_n^2 \, dP &= \int_{A_k} [\tilde{S}_k^2 + (\tilde{S}_n^2 - \tilde{S}_k^2)] \, dP \\
&= \int_{A_k} \tilde{S}_k^2 \, dP + 2 \int_{A_k} \tilde{S}_k (Y_{k+1} + \ldots + Y_n) \, dP \\
&\quad + \int_{A_k} (Y_{k+1} + \ldots + Y_n)^2 \, dP, \;\; \text{since } S_n = S_k + \sum_{i=k+1}^n Y_i, \\
&\geq \varepsilon^2 \int_{A_k} dP + 2 \int_\Omega (\chi_{A_k} \tilde{S}_k)(Y_{k+1} + \ldots + Y_n) \, dP \qquad (11) \\
&= \varepsilon^2 P(A_k) + 2E\left(\chi_{A_k} \tilde{S}_k\right) E\left(\sum_{i=k+1}^n Y_i\right) \\
&\qquad \text{(since } \chi_{A_k}\tilde{S}_k \text{ and } Y_i, i \geq k+1, \text{ are independent)} \\
&= \varepsilon^2 P(A_k) \quad [\text{since} E(Y_i) = 0\,].
\end{aligned}$$

Adding on $1 \leq k \leq n$, we get

$$Var(S_n) = Var(\tilde{S}_n) = \int_\Omega \tilde{S}_n^2 \, dP \geq \varepsilon^2 \sum_{k=1}^n P(A_k) = \varepsilon^2 P(A).$$

Since $VarS_n = \sum_{i=1}^n VarX_i$, by independence of the X_i, this gives (10), and completes the proof.

Remark The only place in the above proof where we use the independence hypothesis is to go from (11) to the next line to conclude that

$$E(\chi_{A_k} \tilde{S}_k (Y_{k+1} + \ldots + Y_n)) = 0.$$

Any other hypothesis that guarantees the nonnegativity of this term gives the corresponding maximal inequality. There are several classes of nonindependent random variables including (positive sub-) $\mathcal{L}^2$-martingale sequences giving such a result. This will be seen in the next chapter.

All the strong convergence theorems that follow in this section are due to Kolmogorov.

Theorem 6 *Let $X_1, X_2, \ldots$ be a sequence of independent random variables on (Ω, Σ, P) with means $\mu_1, \mu_2, \ldots,$ and variances $\sigma_1^2, \sigma_2^2, \ldots$. Let*

$$S_n = \sum_{k=1}^{n} (X_k - \mu_k)$$

and $\sigma^2 = \sum_{n=1}^{\infty} \sigma_n^2$. Suppose that $\sigma^2 < \infty$ and $\sum_{k=1}^{\infty} \mu_k$ converges. Then $\sum_{k=1}^{\infty} X_k$ converges a.e. and in the mean of order 2 to an r.v. X. Moreover, $E(X) = \sum_{k=1}^{\infty} \mu_k$, $Var X = \sigma^2$, and for any $\varepsilon > 0$,

$$P\left[\sup_{n \geq 1} |S_n| \geq \varepsilon\right] \leq \frac{\sigma^2}{\varepsilon^2} \tag{12}$$

Proof It should be shown that $\lim_n S_n$ exists a.e. If this is proved, since $\sum_{k=1}^{\infty} \mu_k$ converges, we get

$$\lim_n \sum_{k=1}^{n} X_k = \lim_{n \to \infty} S_n + \lim_{n \to \infty} \sum_{k=1}^{n} \mu_k = X \quad \text{exists a.e.}$$

But the sequence $\{S_n(\omega), n \geq 1\}$ of scalars converges iff it satisfies the Cauchy criterion, i.e., iff $\inf_m \sup_k |S_{m+k}(\omega) - S_m(\omega)| = 0$ a.e. Thus let $\varepsilon > 0$ be given, and by Theorem 5,

$$P\left[\max_{m \leq n \leq m+k} |S_n - S_m| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^2} \sum_{i=m}^{m+k} \sigma_i^2 \leq \frac{1}{\varepsilon^2} \sum_{i \geq m} \sigma_i^2. \tag{13}$$

Hence letting $k \to \infty$ in (13) and noting that the events

$$\left[\max_{m \leq n \leq m+k} |S_n - S_m| \geq \varepsilon\right]$$

form an increasing sequence, we get

$$P\left[\sup_{n \geq 1} |S_{m+n} - S_m| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^2} \sum_{k \geq m} \sigma_k^2. \tag{14}$$

It follows that

$$P\left[\inf_m \sup_{n \geq 1} |S_{m+n} - S_m| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^2} \sum_{k \geq m} \sigma_k^2. \tag{15}$$

Letting $\varepsilon \nearrow \infty$, since $\sum_{k=1}^{\infty} \sigma_k^2 < \infty$, the right side of (15) goes to zero, so that $\limsup_{n,m} |S_n - S_m| < \infty$ a.e. But $|S_n| \leq |S_n - S_m| + |S_m|$, so

$$\begin{aligned}\limsup_{n\geq m}|S_n| &\leq \limsup_{n\geq m}[|S_n - S_m| + |S_m|] \\ &\leq |S_m| + \limsup_{n\geq m}|S_n - S_m| \\ &\leq |S_m| + \sup_{n\geq m}|S_n - S_m| < \infty \quad \text{a.e.}\end{aligned}$$

Thus $\limsup_n S_n$, $\liminf_n S_n$ must be finite a.e. Also

$$\left[\limsup_n S_n - \liminf_n S_n \geq 2\varepsilon\right] \subset \left[\sup_{n\geq m}|S_n - S_m| \geq \varepsilon\right], m \geq 1.$$

Hence by (14)

$$P\left[\limsup_n S_n - \liminf_n S_n \geq 2\varepsilon\right] \leq \frac{1}{\varepsilon^2}\sum_{k=m}^{\infty}\sigma_k^2 \to 0 \tag{16}$$

as $m \to \infty$ for each $\varepsilon > 0$. It follows that $\limsup_n S_n = \liminf_n S_n$ a.e., and the limit exists as asserted.

If we let $m = 0$ in (14) and $X_0 = 0$, then (14) implies (12). It remains to establish mean convergence. In fact, consider for $m < n$, with $\tilde{X}_n = X_n - \mu_n$,

$$E((S_n - S_m)^2) = E((\tilde{X}_{m+1} + \ldots + \tilde{X}_n)^2) = \sum_{k=m+1}^{n}\sigma_k^2 \to 0 \text{ as } m, n \to \infty. \tag{17}$$

Thus $S_n \to S$ in $\mathcal{L}^2(P)$, and hence also in $\mathcal{L}^1(P)$, since $||f||_1 \leq ||f||_2$ for any $f \in \mathcal{L}^2$. It follows that $E(S^2) = \lim_n E(S_n^2) = \lim_n \sum_{k=1}^n \sigma_k^2 = \sigma^2$, and $E(S) = \lim_n E(S_n) = 0$. But $X = S + \sum_{n=1}^{\infty}\mu_n$, so that $E(X) = \sum_{n=1}^{\infty}\mu_n$. This completes the proof.

Remarks (1) If we are given that $\lim_n \sum_{k=1}^n X_k$ exists in $\mathcal{L}^2$ and $\sum_{n=1}^{\infty}\mu_n$ converges, then $S_n = \sum_{k=1}^n (X_k - \mu_k) \to S$ in $\mathcal{L}^2$ also, so that $\sum_{k=1}^n \sigma_k^2 = E(S_n^2) \to E(S^2) = \sigma^2$. Thus $\sum_{k=1}\sigma_k^2 < \infty$. Hence by the theorem $\sum_{k=1}^{\infty} X_k$ also exists a.e.

(2) If the hypothesis of independence is simply dropped in the above theorem, the result is certainly false. In fact let $X_n = X/n$, where $E(X) = 0$, $0 < VarX = \sigma^2 < \infty$, so that $\sum_{k=1}^{\infty}\mu_k = 0$ and

$$\sum_{k\geq 1}\sigma_k^2 = \sigma^2 \sum_{n\geq 1}\frac{1}{n^2} < \infty.$$

But $\sum_{n=1}^{\infty} X_n = X\sum_{n=1}^{\infty} 1/n$, diverges a.e., on the set where $|X| > 0$, a.e.

A partial converse of the above theorem is as follows.

Theorem 7 *Let $\{X_n, n \geq 1\}$ be a uniformly bounded sequence of independent random variables on (Ω, Σ, P) with means zero and variances*

$\{\sigma_n^2, n \geq 1\}$. *If* $\sum_{n=1}^{\infty} X_n$ *converges on a set of positive measure, then* $\sum_{n=1}^{\infty} \sigma_n^2 < \infty$, *and hence the series actually converges a.e. on the whole space* Ω.

Proof Let $X_0 = 0$ and $S_n = \sum_{i=1}^{n} X_i$. If A is the set of positive measure on which $S_n \to S$ a.e., then by Theorem 4 (of Egorov), there is a measurable subset $\tilde{A} \subset A$ of arbitrarily small measure such that if $B_0 = A - \tilde{A} \subset A$, we have $P(B_0) > 0$ and $S_n \to S$ on B_0 uniformly. Since S is an r.v., we can find a set $B \subset B_0$ of positive measure (arbitrarily close to that of B_0), and a positive number d such that $|S_n| \leq d < \infty$ on B. Thus if $\overline{A} = \bigcap_{n=0}^{\infty}[|S_n| \leq d]$, then $\overline{A} \in \Sigma$, $\overline{A} \supset B$, and $P(\overline{A}) \geq P(B) > 0$.

Let $A_n = \bigcap_{k=0}^{n}[|S_k| \leq d]$, so that $A_n \downarrow A$. If $C_n = A_n - A_{n+1}$, and $C_0 = \bigcup_{n=1}^{\infty} C_n$, which is a disjoint union, let $a_n = \int_{A_n} S_n^2 dP$. Clearly $a_n \leq d^2 P(A_n) \leq d^2$, so that $\{a_n, n \geq l\}$ is a bounded sequence. Consider

$$\begin{aligned} a_n - a_{n-1} &= \int_{A_n} S_n^2 dP - \int_{A_{n-1}} S_{n-1}^2 dP \\ &= \int_{A_{n-1}} (S_{n-1} + X_n)^2 dP \\ &- \int_{C_{n-1}} S_n^2 dP - \int_{A_{n-1}} S_{n-1}^2 dP \quad (\text{since} A_n = A_{n-1} - C_{n-1}\ ) \\ &= \int_{A_{n-1}} X_n^2 dP + 2\int_{A_{n-1}} S_{n-1}X_n dP - \int_{C_{n-1}} S_n^2 dP. \end{aligned} \tag{18}$$

However,

$$\int_{A_{n-1}} X_n^2 dP = E(\chi_{A_{n-1}} X_n^2) = \sigma_n^2 P(A_{n-1}) \quad \text{by independence of } \chi_{A_{n-1}} \text{ and } X_n,$$

and

$$\int_{A_{n-1}} X_n S_{n-1} dP = E(X_n) E(\chi_{A_{n-1}} S_{n-1}) = 0,$$

since $E(X_n) = 0$. Thus by noting that $P(A_{n-1}) \geq P(A_n)$, (18) becomes, with these simplifications and the hypothesis that $|X_n| \leq c < \infty$ a.e.,

$$\begin{aligned} a_n - a_{n-1} &\geq \sigma_n^2 P(A_n) - \int_{C_{n-1}} (|S_{n-1}| + |X_n|)^2 dP \\ &\geq \sigma_n^2 P(A_n) - (c+d)^2 P(C_{n-1}) \\ &\qquad (\text{since } |S_n| \leq |S_{n-1}| + |X_n| \leq d + c\ ), \\ &\geq \sigma_n^2 P(\overline{A}) - (c+d)^2 P(C_{n-1}). \end{aligned}$$

Summing over $n = 1, 2, \ldots, m$, we get ($a_0 = 0$)

$$a_m \geq P(\overline{A}) \sum_{n=1}^{m} \sigma_n^2 - (c+d)^2 P(\bigcup_{n=1}^{m} C_{n-1}).$$

Hence recalling that $a_m \leq d^2$, one has

$$d^2 \geq P(\overline{A}) \sum_{n=1}^{m} \sigma_n^2 - (c+d)^2, \;\; m \geq 1. \tag{19}$$

Since $P(\overline{A}) > 0$, (19) implies that $\sum_{n=1}^{\infty} \sigma_n^2 < \infty$. This yields the last statement and, in view of Theorem 6, completes the proof.

As an immediate consequence, we have

Corollary 8 *If* $\{X_n, n \geq 1\}$ *is a uniformly bounded sequence of independent random variables on* (Ω, Σ, P) *with* $E(X_n) = 0, n \geq 1$, *then* $\sum_{n=1}^{\infty} X_n$ *converges with probability* 0 *or* 1.

We are now in a position to establish a very general result on this topic.

Theorem 9 (Three Series Theorem). *Let* $\{X_n, n \geq 1\}$ *be a sequence of independent random variables on* (Ω, Σ, P). *Then* $\sum_{n=1}^{\infty} X_n$ *converges a.e. iff the following three series converge. For some (and then every)* $0 < c < \infty$,

(i) $\sum_{n=1}^{\infty} P([|X_n| > c])$,

(ii) $\sum_{n=1}^{\infty} E(X_n^c)$,

(iii) $\sum_{n=1}^{\infty} \sigma^2(X_n^c)$,

where X_n^c *is the truncation of* X_n *at* c, *so that* $X_n^c = X_n$ *if* $|X_n| \leq c$, *and* $= 0$ *otherwise.*

Proof Sufficiency is immediate. In fact, suppose the three series converge. By (i), and the first Borel-Cantelli lemma, $P[\limsup_n |X_n| > c] = 0$, so that for large enough n, $X_n = X_n^c$ a.e. Next, the convergence of (ii) and (iii) imply, by Theorem 6, $\sum_{n=1}^{\infty} X_n^c$ converges a.e. Since $X_n = X_n^c$ for large n, $\sum_{m=1}^{\infty} X_n$ itself converges a.e. Note that $c > 0$ is arbitrarily fixed.

Conversely, suppose $\sum_{i=1}^{\infty} X_i$ converges a.e. Then $\lim_n X_n = 0$ a.e. Hence if $A_{n,c} = [X_n \neq X_n^c] = [|X_n| > c]$ for any fixed $c > 0$, then the $A_{n,c}$ are independent and $P[\limsup_n A_{n,c}] = 0$. Thus by the second Borel-Cantelli lemma (cf. Theorem 1.9iii), $\sum_{n=1}^{\infty} P(A_{n,c}) < \infty$, which proves (i). Also, $\sum_{n=1}^{\infty} X_n^c$ converges a.e., since for large enough n, X_n^c and X_n are equal a.e. But now the X_n^c are uniformly bounded. We would like to reduce the result to Theorem 7. However, $E(X_n^c)$ is not necessarily zero. Thus we need a new idea for this reduction. One considers a sequence of independent random variables $\tilde{X}_n^c$ which are also independent of, but with the same distributions as, the

X_n^c-sequence. Now, the given probability space may not support two such sequences. In that case, we enlarge it by adjunction as explained after Corollary 8 in the last section. The details are as follows.

Let $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{P}) = (\Omega, \Sigma, P) \otimes (\Omega, \Sigma, P)$, and let X_n^1, X_n^2; be defined on $\tilde{\Omega}$ by the equations

$$X_n^1(\omega) = X_n^c(\omega_1), X_n^2(\omega) = X_n^c(\omega_2), \quad \text{where } \omega = (\omega_1, \omega_2) \in \tilde{\Omega}. \tag{20}$$

It is trivial to verify that $\{X_n^1, n \geq 1\}$, $\{X_n^2, n \geq 1\}$ are two mutually independent sequences of random variables on $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{P})$, $|X_n^i| \leq c$, $i = 1, 2$, and have the same distributions. Thus if $Z_n = X_n^1 - X_n^2$, $n \geq 1$, then $E(Z_n) = 0$, $Var Z_n = Var X_n^1 + Var X_n^2 = 2\sigma_n^2(X_n^c)$, and $\{Z_n, n \geq 1\}$ is a uniformly bounded (by 2c) independent sequence to which Theorem 7 applies. Hence, by that result, $\sum_{n=1}^{\infty} Var Z_n < \infty$, so that $\sum_{n=1}^{\infty} \sigma_n^2(X_n^c) < \infty$, which is (iii).

Next, if $Y_n = X_n^c - E(X_n^c)$, then $E(Y_n) = 0$, $Var Y_n = Var X_n^c$, so that $\sum_{n=1}^{\infty} \sigma^2(Y_n) < \infty$. Hence by Theorem 6, $\sum_{n=1}^{\infty} Y_n$ converges a.e. Thus we have $\sum_{n=1}^{\infty} E(X_n^c) = \sum_{n=1}^{\infty} X_n^c - \sum_{n=1}^{\infty} Y_n$, and both the series on the right converge a.e. Thus the left side, which is a series of constants, simply converges and (ii) holds. Observe that if the result is true for one $0 < c < \infty$, then by this part the three series must converge for every $0 < c < \infty$. This completes the proof.

Remarks (1) If any one of the three series of the above theorem diverges, then $\sum_{n \geq 1} X_n$ diverges a.e. This means the set $[\sum_{n=1}^{\infty} X_n$ converges$]$ has probability zero, so that the zero-one criterion obtains. The proof of this statement is a simple consequence of the preceding results (since the convergence is determined by $\sum_{k \geq n} X_k$ for large n), but not of Theorem 1.12.

(2) Observe that the convergence statements on series in all these theorems relate to *unconditional convergence*. It is *not absolute* convergence, as simple examples show. For instance, if $a_n > 0, \sum_{n=1}^{\infty} a_n = \infty$, but

$$\sum_{n=1}^{\infty} a_n^2 < \infty,$$

then the independent random variables $X_n = \pm a_n$ with equal probability on (Ω, Σ, P) satisfy the hypothesis of Corollary 8 and so $\sum_{n=1}^{\infty} X_n$ converges a.e. But it is clear that $\sum_{n=1}^{\infty} |X_n| = \sum_{n=1}^{\infty} |a_n| = \infty$ a.e. The point is that $X_n \in L^2(\Omega, \Sigma, P)$ and the series $\sum_{n=1}^{\infty} X_n$ converges unconditionally in $L^2(P)$, but not absolutely there if the space is infinite dimensional. In fact, it is a general result of the Banach space theory that the above two convergences are unequal in general.

(3) One can present easy sufficient conditions for absolute convergence of a series of random variables on (Ω, Σ, P). Indeed, $\sum_{n=1}^{\infty} X_n$ converges absolutely a.e. if $\sum_{n=1}^{\infty} E(|X_n|) < \infty$. This is true since $E(\sum_{n=1}^{\infty} |X_n|) = \sum_{n=1}^{\infty} E(|X_n|) < \infty$ by the Lebesgue dominated convergence theorem, and since $Y = \sum_{n=1}^{\infty} |X_n|$ is a (positive) r.v. with finite expectation, $P[Y > \lambda] \leq$

$E(Y)/\lambda \to 0$ as $\lambda \to \infty$, so that $0 \le Y < \infty$ a.e. Here X_n need not be independent. But the integrability condition is very stringent. Such results are "nonprobabilistic" in nature, and are not of interest in our subject.

A natural question now is to know the properties of the limit r.v. $X = \sum_{n=1}^{\infty} X_n$ in Theorem 9 when the series converges. For example: if each X_n has a countable range, which is a simple case, what can one say about the distribution of X? What can one say about $Y = \sum_{n=1}^{\infty} a_n X_n$, where $\sum_{n=1}^{\infty} a_n^2 < \infty$, $E(X_n) = 0$, $E(X_n^2) = 1$, and X_n are independent?

Not much is known about these queries. Some special cases are studied, and a sample result is discussed in the problems section. For a deeper analysis of special types of random series, one may refer to Kahane (1985). We now turn to the next important aspect of averages of independent random variables, which has opened up interesting avenues for probability theory.

2.3 Laws of Large Numbers

Very early in Section 1.1 we indicated that probability is a "long-term average." This means that the averages of "successes" in a sequence of independent trials "converge" to a number. As the preceding section shows, there are three frequently used types of convergences, namely, the pointwise a.e., the stochastic (or "in probability") convergence, and the distributional convergence, each one being strictly weaker than the preceding one. The example following the proof of Proposition 2.2 shows that $X_n \xrightarrow{D} X$ does not imply that the $X_n(\omega)$ need to approach $X(\omega)$ for any $\omega \in \Omega$. So in general it is better to consider the a.e. and "in probability" types for statements relating to outcomes of ω. Results asserting a.e. convergence always imply the "in probability" statements, so that the former are called *strong laws* and the latter, *weak laws*. If the random variables take only two values $\{0, 1\}$, say, then the desired convergence in probability of the averages was first rigorously established by James Bernoulli in about the year 1713, and the a.e. convergence result for the same sequence was obtained by E. Borel only in 1909.

Attempts to prove the same statements for general random variables, with range space $\mathbb{R}$, and the success thus achieved constitute a general story of the subject at hand. In fact, P. L. Čebyšev seems to have devised his inequality for extending the Bernoulli theorem, and established the following result in 1882.

Proposition 1 (**Čebyšev**). *Let $X_1, X_2, \ldots$ be a sequence of independent random variables on (Ω, Σ, P) with means $\mu_1, \mu_2, \ldots$ and variances $\sigma_1^2, \sigma_2^2, \ldots$, such that if $S_n = \sum_{i=1}^{n} X_i$, one has $\sigma^2(S_n)/n^2 \to 0$ as $n \to \infty$. Then the sequence obeys the weak law of large numbers (WLLN), which means, given $\varepsilon > 0$, we have*

$$\lim_{n\to\infty} P\left[\left|\frac{S_n - E(S_n)}{n}\right| \geq \varepsilon\right] = 0. \tag{1}$$

Proof By Čebyšev's inequality (1) follows at once.

Note that if all the X_n have the same distribution, then they have equal moments, i.e., $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma^2$, so that $\sigma^2(S_n) = \sum_{i=1}^n \sigma_i^2 = n\sigma^2$, and $\sigma^2(S_n)/n^2 = \sigma^2/n \to 0$ is automatically satisfied. The result has been improved in 1928 by A. Khintchine, by assuming just one moment. For the proof, he used a truncation argument, originally introduced in 1913 by A. A. Markov. Here we present this proof as it became a powerful tool. Later we see that the result can be proved, using the characteristic function technique, in a very elementary manner, and even with a slightly *weaker* hypothesis than the existence of the first moment [i.e., only with the existence of a derivative at the origin for its Fourier transform; that does *not* imply $E(X)$ exists].

Theorem 2 (**Khintchine**) *Let $X_1, X_2, \ldots$ be independent random variables on (Ω, Σ, P) with a common distribution [i.e., $P[X_n < x] = F(x), x \in \mathbb{R}$, for $n \geq 1$] and with one moment finite. Then the sequence obeys the WLLN.*

Proof We use the preceding result in the proof for the truncated functions and then complete the argument with a detailed analysis. Let $\varepsilon > 0, \delta > 0$ be given. Define

$$U_k^n = X_k \chi_{[|X_k| \leq n\delta]}, \quad V_k^n = X_k \chi_{[|X_k| > n\delta]}, \tag{2}$$

so that $X_k = U_k + V_k$. Let F be the common distribution function of the X_k. Since $E(|X_k|) < \infty$, we have $M = E(|X_k|) = \int_{\mathbb{R}} |x| dF(x) < \infty$, by the fundamental (image) law of probability. If $\mu = E(X_k) = \int_{\mathbb{R}} x dF(x)$ and $\mu_n' = E(U_k^n)$, then

$$\mu_n' = E(X_k \chi_{[|X_k| \leq n\delta]}) = \int_{[|x| \leq n\delta]} x dF(x)$$

and by the dominated convergence theorem, we have

$$\int_{[|x| \leq n\delta]} x dF(x) = \int_{\mathbb{R}} \chi_{[|X_k| \leq n\delta]} x dF(x) \to \int_{\mathbb{R}} x dF(x) = \mu \tag{3}$$

Thus there is N_1 such that $n \geq N_1 \Rightarrow |\mu_n' - \mu| < \varepsilon/2$. Note that μ_n' depends only on n, and hence not on k, because of the common distribution of the X_k. Similarly

$$Var U_k^n = E(U_k^n)^2 - (\mu_n')^2 \leq \int_{|x| \leq n\delta} x^2 dF(x) \leq n\delta \int_{\mathbb{R}} |x| dF(x) = n\delta M.$$

By hypothesis $U_1^n, U_2^n, \ldots$ are independent (bounded) random variables with means μ_n' and variances bounded by $n\delta M$. Let $T_m^n = U_1^n + \ldots + U_m^n$ and

$W_m^n = V_1^n + \ldots + V_m^n$. Then by the preceding proposition, or rather the Čebyšev's inequality,

$$P\left[\left|\frac{T_n^n - n\mu_n'}{n}\right| \geq \frac{\varepsilon}{2}\right] \leq \frac{4n^2\delta M}{n^2\varepsilon^2} = 4\delta M\varepsilon^{-2}. \tag{4}$$

On the other hand, adding and subtracting $n\mu$ and using the triangle inequality gives

$$\left|\frac{T_n^n - n\mu_n'}{n}\right| \geq \left|\frac{T_n^n - n\mu}{n}\right| - |\mu - \mu_n'|.$$

Thus if $n \geq N_1$, we have, with the choice of N_1 after (2), on the set

$$\left|\frac{T_n^n - n\mu}{n}\right| \geq \varepsilon,$$

the following:

$$\left|\frac{T_n^n - n\mu_n'}{n}\right| \geq \varepsilon - \frac{\varepsilon}{2} = \frac{\varepsilon}{2}.$$

Hence for $n \geq N_1$ this yields

$$P\left[\left|\frac{T_n^n - n\mu}{n}\right| \geq \varepsilon\right] \leq P\left[\left|\frac{T_n^n - n\mu_n'}{n}\right| \geq \frac{\varepsilon}{2}\right] \leq \frac{4\delta M}{\varepsilon^2}, \text{ by (4)}. \tag{5}$$

But by definition $S_n = T_n^n + W_n^n, n \geq 1$, so that

$$\begin{aligned}\omega \in \left[\left|\frac{S_n - n\mu}{n}\right| \geq \varepsilon\right] &= \left[\left|\frac{T_n^n - n\mu + W_n^n}{n}\right| \geq \varepsilon\right] \\ &\Rightarrow \omega \in \left[\left|\frac{T_n^n - n\mu}{n}\right| \geq \frac{\varepsilon}{2}\right] \cup [W_n^n \neq 0].\end{aligned}$$

Thus

$$P\left[\left|\frac{S_n - n\mu}{n}\right| \geq \varepsilon\right] \leq P\left[\left|\frac{T_n^n - n\mu}{n}\right| \geq \frac{\varepsilon}{2}\right] + P[W_n^n \neq 0]. \tag{6}$$

However,

$$\begin{aligned}P[V_k^n \neq 0] &= P[|V_k^n| > 0] = P[|X_k| > n\delta] \\ &= \frac{1}{n\delta}\int_{[|X_k|>n\delta]} n\delta dP \leq \frac{1}{n\delta}\int_{[|X_k|>n\delta]} |X_k| dP \\ &= \frac{1}{n\delta}\int_{[|x|>n\delta]} |x| dP. \end{aligned} \tag{7}$$

Choose N_2 such that $n \geq N_2 \Rightarrow \int_{[|x|>n\delta]} |x| dF(x) < \delta^2$, which is possible since $M = E(|X_k|) < \infty$. Thus for $n \geq N_2$, $P[V_k^n \neq 0] \leq \delta^2/(n\delta) = \delta/n$ by (7). Consequently,

$$P[W_n^n \neq 0] \leq P\left(\bigcup_{k=1}^{n}[V_k^n \neq 0]\right) \leq \sum_{k=1}^{n} P[V_k^n \neq 0] < \delta. \tag{8}$$

If $N = \max(N_1, N_2)$ and $n \geq N$, then (5) and (8) give for (6)

$$P\left[\left|\frac{S_n - n\mu}{n}\right| \geq \varepsilon\right] \leq \frac{4\delta M}{\varepsilon^2} + \delta = \left[\frac{4M}{\varepsilon^2} + 1\right]\delta. \tag{9}$$

Letting $n \to \infty$ and then $\delta \to 0$ in (9), we get the desired conclusion.

It is important to notice that the independence hypothesis is used only in (4) in the above proof in deducing that the variance of T_n^n = the sum of variances of U_k^n. But this will follow if the U_k^n are uncorrelated for each n. In other words, we used only that

$$\begin{aligned} E(U_k^n U_j^m) &= \int_{[|x| \geq n\delta]} \int_{[|y| \leq m\delta]} xy dF_{X_k, X_j}(x, y) \\ &= \int_{[|x| \geq n\delta]} x dF_{X_k}(x) \int_{[|y| \leq m\delta]} y dF_{X_j}(y) = \mu_n' \mu_m'. \end{aligned}$$

Now this holds if X_i, X_j are independent when $i \neq j$. Thus the above proof actually yields the following stronger result, stated for reference.

Corollary 3 *Let $X_1, X_2, \ldots$ be a pairwise independent sequence of random variables on (Ω, Σ, P) with a common distribution having one moment finite. Then the sequence obeys the WLLN.*

In our development of the subject, the next result serves as a link between the preceding considerations and the "strong laws." It was obtained by A. Rajchman in the early 1930's. The hypothesis is weaker than pairwise independence, but demands the existence of a uniform bound on variances, and then yields a stronger conclusion. The proof uses a different technique, of interest in the subject.

Theorem 4 *Let $\{X_n, n \geq 1\}$ be a sequence of uncorrelated random variables on (Ω, Σ, P) such that $\sigma^2(X_n) \leq M < \infty$, $n \geq 1$. Then $[S_n - E(S_n)]/n \to 0$ in L^2-mean, as well as a.e.* [*The pointwise convergence statement is the definition of the strong law of large numbers (SLLN) of a sequence.*]

Proof The first statement is immediate, since

$$E\left(\left[\frac{S_n - E(S_n)}{n}\right]^2\right) = \frac{1}{n^2} \sum_{k=1}^{n} \sigma^2(X_k) \leq \frac{nM}{n^2} \to 0,$$

by the uncorrelatedness hypothesis of the X_n and the uniform boundedness of $\sigma^2(X_k)$. This, of course, implies by Proposition 1 that the *WLLN* holds for the sequence. The point is that the a.e. convergence also holds.

Consider now, by Čebyšev's inequality, for any $\varepsilon > 0$,

$$P[|S_{n^2} - E(S_{n^2})| > n^2\varepsilon] \leq \frac{VarS_{n^2}}{n^4\varepsilon^2} \leq \frac{M}{n^2\varepsilon^2},$$

so that

$$\sum_{n\geq 1} P[|S_{n^2} - E(S_{n^2})| > n^2\varepsilon] \leq \frac{M}{\varepsilon^2}\sum_{n\geq 1}\frac{1}{n^2} < \infty. \tag{10}$$

Hence by the first Borel-Cantelli lemma, letting $Y_k = X_k - E(X_k)$ and $\tilde{S}_n = \sum_{k=1}^n Y_k$ (so that the Y_k are orthogonal), one has $P([|\tilde{S}_{n^2}| > n^2\varepsilon], i.o.) = 0$, which means $\tilde{S}_{n^2}/n^2 \to 0$ a.e. This is just an illustration of Proposition 2.3.

With the boundedness hypothesis we show that the result holds for the full sequence $\tilde{S}_n/n$ and not merely for a subsequence, noted above.

For each $n \geq 1$, consider $n^2 \leq k < (n+1)^2$ and $\tilde{S}_k/k$. Then

$$\frac{|\tilde{S}_k|}{k} \leq \frac{|\tilde{S}_k - \tilde{S}_{n^2}| + |\tilde{S}_{n^2}|}{k} \leq \frac{|\tilde{S}_{n^2}|}{n^2} + \max_{n^2\leq k<(n+1)^2}\frac{|\tilde{S}_k - \tilde{S}_{n^2}|}{n^2} \tag{11}$$

and let $T_n = \max_{n^2\leq k<(n+1)^2}|\tilde{S}_k - \tilde{S}_{n^2}|$. Since as $k \to \infty$ the first term on the right $\to 0$ a.e., (shown above) it suffices to establish $T_n/n^2 \to 0$ a.e. To use the orthogonality property of the Y_k, consider T_n^2. We have

$$T_n^2 = \max_{n^2\leq k\leq(n+1)^2}|\tilde{S}_k - \tilde{S}_{n^2}|^2 \leq \sum_{k=n^2}^{(n+1)^2}|\tilde{S}_k - \tilde{S}_{n^2}|^2 = \sum_{k=n^2}^{(n+1)^2}\left(\sum_{i=n^2+1}^{k} Y_i\right)^2$$

and so for $n \geq 2$, since $\sigma^2(Y_i) = \sigma^2(X_i) \leq M$,

$$\begin{aligned} E(T_n^2) &\leq \sum_{k=n^2}^{(n+1)^2}\sum_{i=n^2+1}^{k}\sigma^2(Y_i) \\ &\leq \sum_{k=n^2}^{(n+1)^2}\sum_{i=n^2+1}^{k} M = \sum_{k=n^2}^{(n+1)^2}(k-(n^2+1))M \\ &= M(2n^2+n) \leq 3n^2M < (2n)^2M \end{aligned} \tag{12}$$

This crude estimate is sufficient to show, as before, that

$$P[|T_n| > n^2\varepsilon] \leq P[T_n^2 \geq n^4\varepsilon^2] \leq E(T_n^2)/n^4\varepsilon^2 \leq 4M/n^2\varepsilon^2, \tag{13}$$

by Markov's inequality and (12). Thus $\sum_{n>2} P[|T_n| > n^2\varepsilon] < \infty$ and the Borel-Cantelli lemma again yields $P[|T_n/n^2| > \varepsilon, i.o.] = 0$. Hence $T_n/n^2 \to 0$ a.e. and by (11) $\tilde{S}_k/k \to 0$ a.e., proving the result.

We now strengthen the probabilistic hypothesis from uncorrelatedness to mutual independence and weaken the moment condition. The resulting statement is significantly harder to establish. It will be obtained in two stages, and both are of independent interest. They have been proved in 1928 by A. Kolmogorov, and are sharp. We begin with an elementary but powerful result from classical summability theory.

Proposition 5 (Kronecker's Lemma). *Let $a_1, a_2, \ldots$ be a sequence of numbers such that $\sum_{n\geq 1}(a_n/n)$ converges. Then*

$$\frac{1}{n}\sum_{k=1}^{n} a_k \to 0 \text{ as } n \to \infty.$$

Proof Let $s_0 = 0$, $s_n = \sum_{k=1}^{n}(a_k/k)$, and $R_n = \sum_{k=1}^{n} a_k$. Then $s_n \to s$ by hypothesis. Also, $a_k = k(s_k - s_{k-1})$, so that

$$R_{n+1} = \sum_{k=1}^{n+1} ks_k - \sum_{k=1}^{n+1} ks_{k-1} = -\sum_{k=1}^{n} s_k + (n+1)s_{n+1}.$$

Hence

$$\frac{1}{n+1}R_{n+1} = s_{n+1} - \frac{n}{n+1}\frac{1}{n}\sum_{k=1}^{n} s_k \to s - 1\cdot s = 0 \text{ as } n \to \infty$$

because $s_n \to s \Rightarrow$ for any $\varepsilon > 0$, there is $n_0 [= n_o(\varepsilon)]$ such that $n > n_0 \Rightarrow |s_n - s| < \varepsilon$, and hence

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=1}^{n} s_k - s\right| &\leq \frac{|s_1 + \ldots + s_{n_0}|}{n} + \frac{|(s_{n_0+1} - s) + \ldots (s_n - s)|}{n} + \frac{n_0|s|}{n} \\ &\leq \frac{|s_1 + \ldots + s_{n_0}|}{n} + \frac{(n - n_0)}{n}\varepsilon + \frac{n_0|s|}{n} \to \varepsilon \qquad (14)\end{aligned}$$

as $n \to \infty$. [This is called $(c, 1)$-convergence or *Cesàro summability* of s_n.] Since $\varepsilon > 0$ is arbitrary, the result follows.

Theorem 6 (First form of SLLN). *If $X_1, X_2, \ldots$ is a sequence of independent random variables on (Ω, Σ, P) with means zero and variances $\sigma_1^2, \sigma_2^2, \ldots$, satisfying $\sum_{n=1}^{\infty}(\sigma_n^2/n^2) < \infty$, then the sequence obeys the SLLN, i.e.,*

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \to 0$$

a.e. as $n \to \infty$.

Proof Let $Y_n = X_n/n$. Then the $Y_n, n \geq 1$, are independent with means zero, and $\sum_{n=1}^{\infty} \sigma^2(Y_n) = \sum_{n=1}^{\infty}(\sigma_n^2/n^2) < \infty$. Thus by Theorem 2.6, $\sum_{n=1}^{\infty} Y_n$

converges a.e. Hence $\sum_{n=1}^{\infty}(X_n/n)$ converges a.e. By Kronecker's lemma, $(1/n)\sum_{k=1}^{n} X_k \to 0$ a.e., proving the theorem.

This result is very general in that there are sequences of independent random variables $\{X_n, n \geq 1\}$ with means zero and finite variances $\sigma_1^2, \sigma_2^2, \ldots$ satisfying $\sum_{n=1}^{\infty}(\sigma_n^2/n^2) = \infty$ for which the SLLN does not hold. Here is a simple example. Let $X_1, X_2, \ldots$ be independent two-valued random variables, defined as

$$P([X_n = n]) = P([X_n = -n]) = \frac{1}{2}.$$

Hence $E(X_n) = 0, \sigma^2(X_n) = n^2$, so that $\sum_{n=1}^{\infty}[\sigma^2(X_n)/n^2] = +\infty$. If the sequence obeys the SLLN, then $(\sum_{k=1}^{n} X_k)/n \to 0$ a.e. This implies

$$\frac{X_n}{n} = \frac{1}{n}\sum_{k=1}^{n} X_k - \frac{n-1}{n} \cdot \frac{1}{n-1}\sum_{k=1}^{n-1} X_k \to 0 \text{ a.e.},$$

hence $P[|X_n| \geq n, i.o.] = 0$. By independence, this and the second Borel-Cantelli lemma yield $\sum_{n=1}^{\infty} P[|X_n| \geq n] < \infty$. However, by definition $P[|X_n| \geq n] = 1$, and this contradicts the preceding statement. Thus $(1/n)\sum_{k=1}^{n} X_k \not\to 0$ a.e., and SLLN is not obeyed.

On the other hand, the above theorem is still true if we make minor relaxations on the means. For instance, if $\{X_n, n \geq 1\}$ is independent with means $\{\mu_n, n \geq 1\}$ and variances $\{\sigma_n^2;, n \geq 1\}$ such that (i) $\sum_{n=1}^{\infty} \sigma_n^2 < \infty$ and (ii) if either $\mu_n \to 0$ or just $(1/n)\sum_{k=1}^{n} \mu_k \to 0$ as $n \to \infty$, then $(1/n)\sum_{k=1}^{n} X_k \to 0$ a.e. Indeed, if $Y_n = X_n - \mu_n$, then $\{Y_n, n \geq 1\}$ satisfies the conditions of the above result. Thus $(1/n)\sum_{k=1}^{n} Y_k = (1/n)\sum_{k=1}^{n} X_k - (1/n)\sum_{k=1}^{n} \mu_k \to 0$ a.e. If $\mu_k \to \mu$, then $(1/n)\sum_{k=1}^{n} \mu_k \to \mu$ by (14). Here $\mu = 0$. The same holds if we only demanded $(1/n)\sum_{k=1}^{n} \mu_k \to 0$. In either case, then, $(1/n)\sum_{k=1}^{n} X_k \to 0$ a.e. However, it should be remarked that there exist independent symmetric two-valued X_n, $n \geq 1$, with $\sum_{n\geq 1}[\sigma^2(X_n)/n^2] = \infty$ obeying the SLLN. Examples can be given to this effect, if we have more information on the growth of the partial sums $\{S_n, n \geq 1\}$, through, for instance, the laws of the iterated logarithm. An important result on the latter subject will be established in Chapter 5.

The following is the celebrated SLLN of Kolmogorov.

Theorem 7 (Main SLLN). *Let $\{X_n, n \geq 1\}$ be independent random variables on (Ω, Σ, P) with a common distribution and $S_n = \sum_{k=1}^{n} X_k$. Then $S_n/n \to \alpha_0$, a constant, a.e. iff $E(|X_1|) < +\infty$, in which case $\alpha_0 = E(X_1)$. On the other hand, if $E(|X_1|) = +\infty$, then $\limsup_n(|S_n|/n) = +\infty$ a.e.*

Proof To prove the sufficiency of the first part, suppose $E(|X_1|) < \infty$. We use the truncation method of Theorem 2. For simplicity, let $E(X_1) = 0$, since otherwise we consider the sequence $Y_k = X_k - E(X_1)$. For each n, define

$$U_n = X_n\chi_{[|X_n|\le n]}, \quad V_n = X_n\chi_{[|X_n|>n]}.$$

Thus $X_n = U_n + V_n$ and $\{U_n, n \ge 1\}$, $\{V_n, n \ge 1\}$ are independent sequences. First we claim that $\limsup_n |V_n| = 0$ a.e., implying $(1/n)\sum_{k=1}^n V_k \to 0$ a.e. That is to say, $P([V_n \ne 0], i.o.) = 0$. By independence, and the Borel-Cantelli lemma, this is equivalent to showing $\sum_{n=1}^\infty P[|V_n| > 0] < \infty$.

Let us verify the convergence of this series:

$$\begin{aligned}
\sum_{n=1}^\infty P[V_n \ne 0] &= \sum_{n=1}^\infty P[|X_n| > n] \\
&= \sum_{n=1}^\infty P[|X_1| > n] \text{ (since the } X_i \text{ have the same distribution)} \\
&= \sum_{n=1}^\infty \sum_{k=n}^\infty P[k < |X_1| \le k+1] \\
&= \sum_{n=1}^\infty n a_{n,n+1} \text{ (where } a_{k,k+1} = P[k < |X_1| \le k+1]) \\
&= \sum_{n=1}^\infty \int_{[n<|X_1|\le n+1]} n\,dP \le \int_\Omega |X_1|dP = E(|X_1|) < \infty \qquad (15)
\end{aligned}$$

Next consider the bounded sequence $\{U_n, n \ge 1\}$ of independent random variables. If $\mu_n = E(U_n)$, then

$$\mu_n = \int_{[|X_1|\le n]} X_1 dP \to \int_\Omega X_1 dP = E(X_1) = 0$$

by the dominated convergence theorem. Hence $(1/n)\sum_{k=1}^n \mu_k \to 0$. Thus by the remark preceding the statement of the theorem, if $\sum_{n=1}^\infty [\sigma^2(U_n)/n^2] < \infty$, then $(1/n)\sum_{n=1}^\infty U_n \to 0$ a.e., and the result follows.

We verify the desired convergence by a computation similar to that used in (15). Thus

$$\begin{aligned}
\sigma^2(U_n) &= E(U_n^2) - \mu_n^2 \le E(U_n^2) \\
&= \int_{[|X_1|\le n]} X_1^2 dP \text{ (by the common distribution of the } X_n) \\
&\le \sum_{k=0}^{n-1} (k+1)^2 P[k < |X_1| \le k+1].
\end{aligned}$$

Hence

$$\sum_{n=1}^{\infty} \frac{\sigma^2(U^2)}{n^2} \le \sum_{n=1}^{\infty}\sum_{k=0}^{n-1} \frac{(k+1)^2}{n^2} a_{k,k+1} \text{ [using the notation of (15)]}$$

$$= \sum_{k=1}^{\infty} k^2 a_{k-1,k} \sum_{n=k}^{\infty} \frac{1}{n^2}$$

$$\le \sum_{k=1}^{\infty} k^2 \cdot \frac{2}{k} a_{k-1,k} \quad \left[\text{since } \sum_{n=k}^{\infty} \frac{1}{n^2} \le \frac{1}{k} + \int_{k+1}^{\infty} \frac{dx}{(x-1)^2} \le \frac{2}{k} \right]$$

$$= 2 \sum_{k=1}^{\infty} k a_{k-1,k} \le 2E(|X_1|) + 2 < \infty \quad [\text{ by}(15)].$$

Thus

$$\frac{1}{n}\sum_{k=1}^{n} X_k = \frac{1}{n}\sum_{k=1}^{n} U_k + \frac{1}{n}\sum_{k=1}^{n} V_k \to 0 \text{ a.e.}$$

as $n \to \infty$.

Conversely, suppose that $S_n/n \to \alpha_0$, a constant, a.e. We observe that

$$\frac{X_n}{n} = \frac{S_n}{n} - \frac{n-1}{n} \cdot \frac{1}{n-1} S_{n-1} \to 0 \text{ a.e.,}$$

so that $\limsup_n (|X_n|/n) = 0$ a.e. Again by the Borel-Cantelli lemma, this is equivalent to saying that $\sum_{n=1}^{\infty} P[|X_n| > n] < \infty$. But

$$\sum_{n=1}^{\infty} P[|X_n| > n] = \sum_{n=1}^{\infty} P[|X_1| > n]$$

$$= \sum_{n=1}^{\infty} n a_{n,n+1} \text{ [as shown for (15)]}$$

$$= \sum_{n=1}^{\infty} (n+1) P[n < |X_1| \le n+1]$$

$$- \sum_{n=1}^{\infty} P[n < |X_1| \le n+1]$$

$$\ge E(|X_1|) - 2. \tag{16}$$

Hence $E(|X_1|) < \infty$. Then by the sufficiency $(1/n)S_n \to E(X_1)$ a.e., so that $\alpha_0 = E(X_1)$, as asserted.

For the last part, suppose that $E(|X_1|) = +\infty$, so that $E(|X_1|/\alpha) = +\infty$ for any $\alpha > 0$. Then the computation for (16) implies

$$\sum_{n=1}^{\infty} P[|X_n| > \alpha n] = +\infty,$$

since the X_n have the same distribution. Consequently, by the second Borel-Cantelli lemma, we have

$$P([|X_n| > \alpha n], i.o.) = 1. \tag{17}$$

But $|S_n - S_{n-1}| = |X_n| > \alpha n$ implies either $|S_n| > \alpha n/2$ or $|S_{n-1}| > \alpha n/2$. Thus (17) and this give

$$P\left(\left[|S_n| > \frac{\alpha n}{2}\right], i.o.\right) = 1. \tag{18}$$

Hence for each $\alpha > 0$ we can find an $A_\alpha \in \Sigma$, $P(A_\alpha) = 0$, such that

$$\limsup_n \frac{|S_n|}{n} > \frac{\alpha}{2} \quad \text{on } \Omega - A_\alpha.$$

Letting α run through the rationals and setting $A = \bigcup_{\alpha \in \text{ rationals}} A_\alpha$, we get $P(A) = 0$, and on $\Omega - A$, $\limsup_n (|S_n|/n) > k$ for every $k > 0$. Hence $\limsup_n (|S_n|/n) = +\infty$ a.e. This completes the proof of the theorem.

The above result contains slightly more information. In fact, we have the following:

Corollary 8 *Let $\{X_n, n \geq 1\}$ be as in the theorem with $E(|X_1|) < \infty$. Then $|S_n|/n \to |E(X_1)|$ in $L^1(P)$-mean (in addition to the a.e. convergence).*

Proof Since the X_n are i.i.d., so are the $|X_n|, n \geq 1$, and they are clearly independent. Moreover, by i.i.d. $P[-x < X_n < x] = P[-x < X_1 < x]$. Indeed

$$\begin{aligned} P[|X_n| < x] &= \int_\Omega \chi_{[|X_n|<x]} dP \\ &= \int_{\mathbb{R}} \chi_{[|\lambda|<x]} dF(\lambda) (\text{ since } X_n \text{ has } F \text{ as its d.f. for all } n \geq 1) \\ &= P[|X_1| < x] \text{ (by the image law).} \end{aligned}$$

By the SLLN, $S_n/n \to E(X_1)$ a.e., so that $|S_n|/n \to |E(X_1)|$ a.e. Given $\varepsilon > 0$, choose $x_0 > 0$ such that

$$E(|X_n|\chi_{[|X_n|>x_0]}) = E(|X_1|\chi_{[|X_1|>x_0]}) < \varepsilon.$$

If $S_n' = \sum_{k=1}^n X_k \chi_{[|X_k| \leq x_0]}$ and $S_n'' = S_n - S_n'$, then $\{S_n'/n, n \geq 1\}$ is uniformly bounded, so that it is uniformly integrable. But

$$\frac{1}{n} E(|S_n''|) \leq E(|X_1|\chi_{[|X_1|>x_0]}) < \varepsilon$$

uniformly in n. Thus $\{(1/n)S_n'', n \geq 1\}$ is also uniformly integrable. Consequently $\{(1/n)|S_n|, n \geq 1\}$ is a uniformly integrable set. Hence the result follows by Vitali's theorem and the limits must agree, as asserted.

Remark See also Problem 10 for similar (but restricted to finite measure or probability spaces) convergence statements of real analysis, without mention of independence.

These results and their methods of proofs have been extended in various directions. The idea of investigating the averages (both the WLLN and SLLN) has served an important role in creating the modern ergodic theory. Here the random variables X_n are derived from one fixed function $X_1 : \Omega \to \mathbb{R}$ in terms of a measurable mapping $T : \Omega \to \Omega$ $[T^{-1}(\Sigma) \subset \Sigma]$ which preserves measure, meaning $P = P \circ T^{-1}$, or $P(A) = P(T^{-1}(A)), A \in \Sigma$. Then

$$X_{n+1}(\omega) = (X_1 \circ T^n)(\omega), \omega \in \Omega, n \geq 1,$$

where $T^2 = T \circ T$ and $T^n = T \circ T^{n-1}, n \geq 1$. Since $X_1 : \Omega \to \mathbb{R}$ and $T : \Omega \to \Omega$ are both measurable, so that $(X_1 \circ T)^{-1}(\mathcal{B}) = T^{-1}(X^{-1}(\mathcal{B})) \subset T^{-1}(\Sigma) \subset \Sigma$, where $\mathcal{B}$ is the Borel σ-algebra of $\mathbb{R}$, X_2 is an r.v., and similarly X_n is an r.v. For such a sequence, which is no longer independent, the prototypes of the laws of large numbers have been proved. These are called *ergodic* theorems. The correspondents of weak laws are called mean ergodic theorems and those of the strong laws are termed individual ergodic theorems. This theory has branched out into a separate discipline, leaning more toward measure theoretic functional analysis than probability, but still retaining important connections with the latter. For a brief account, see Section 3 of Chapter 7.

Another result suggested by the above theorem is to investigate the growth of sums of independent random variables. How fast does S_n cross some prescribed bound? The laws of the iterated logarithm are of this type, for which more tools are needed. We consider some of them in Chapters 5 and later. We now turn to some applications.

2.4 Applications to Empiric Distributions, Densities, Queueing, and Random Walk

(A) Empiric Distributions

One of the important and popular applications of the SLLN is to show that the empiric distribution converges a.e. and uniformly to the distribution of the random variable. To make this statement precise, consider a sequence of random variables $X_1, X_2, \ldots$ on (Ω, Σ, P) such that $P[X_n < x] = F(x)$,

$x \in \mathbb{R}, n \geq 1$; i.e., they are identically distributed. If we observe "the segment" $X_1, \ldots, X_n$, then the *empiric distribution* is defined as the "natural" proportion for each outcome $\omega \in \Omega$:

$$F_n(x, \omega) = \frac{1}{n} \{ \text{ number of } X_i(\omega) < x\} . \tag{1}$$

Equivalently, let us define

$$Y_j(\omega) = \chi_{[X_j < x]}(\omega), j = 1, ..., n.$$

Then

$$F_n(x, \cdot) = \frac{1}{n}(Y_1 + \ldots + Y_n). \tag{2}$$

We have the following important result, obtained in about 1933.

Theorem 1 (**Glivenko-Cantelli**). *Let $X_1, X_2, \ldots$ be independent and identically distributed (i.i.d.) random variables on (Ω, Σ, P). Let F be their common distribution function, and if the first n-random variables are "observed" (termed a random sample of size n), let F_n be the empiric distribution determined by (1) [or (2)] for this segment. Then*

$$P\left[\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x, \cdot) - F(x)| = 0\right] = 1. \tag{3}$$

Proof Since the X_i are identically distributed with a common distribution, the same is clearly true of the Y_i given by (2). Indeed,

$$P[Y_i = 1] = P[X_i < x] = F(x) = 1 - P[Y_i = 0] = 1 - P[X_i \geq x]$$

for all $i \geq 1$. Hence by the (special case of) SLLN, we get

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} Y_i \to E(Y_1) = F(x) \text{ a.e.} \tag{4}$$

We need to prove the stronger assertion on a.e. uniform convergence in x for (4), which is (3). This is more involved and is presented in three steps.

1. Let $0 \leq k \leq r$ be integers and $x_{k,r}$ be a real number such that

$$F(x_{k,r}) \leq k/r \leq F(x_{k,r} + 0), \quad k = 1, 2, \ldots, r, \tag{5}$$

and set

$$x_{0,r} = -\infty, x_{r,r} = +\infty$$

for definiteness [and use $F(-\infty) = \lim_{x\to-\infty} F(x) = 0, F(+\infty) = \lim_{x\to+\infty} F(x) = 1$]. Also define

$$E_{k,r} = \left\{\omega : \lim_{n\to\infty} F_n(x_{k,r}, \omega) = F(x_{k,r})\right\},$$

and

$$H_{k,r} = \left\{\omega : \lim_{n\to\infty} F_n(x_{k,r}+0,\omega) = F(x_{k,r}+0)\right\}.$$

Then by (4), $P(E_{k,r}) = 1 = P(H_{k,r}), 1 \le k \le r$. Let

$$E_r = \bigcap_{k=1}^{r} E_{k,r} \cap \bigcap_{k=1}^{r} H_{k,r}.$$

2. We have $P(E_r) = 1$ and if $\tilde{E} = \bigcap_{r=1}^{\infty} E_r$, then $P(\tilde{E}) = 1$. In fact, if $A, B \in \Sigma$, $P(A) = 1 = P(B)$, then clearly

$$1 = P(A \cup B) = P(A) + P(B) - P(A \cap B) = 2 - P(A \cap B).$$

Hence $P(A \cap B) = 1$. By induction, with $A = E_{k,r}, B = H_{k,r}, k = 1, \ldots, r$, it follows that $P(E_r) = 1, r \ge 1$. Since $\tilde{E} = \bigcap_{n=1}^{\infty} B_n = \lim_{n\to\infty} B_n$, where $B_n = \bigcap_{r=1}^{n} E_r$, it also follows that $P(E) = \lim_n P(B_n) = 1$.

Let us express $\tilde{E}$ in a different form. First note that E_r is given by

$$E_r = \{\omega : \lim_{n\to\infty} \max_{\substack{1\le k\le r\\ 1\le j\le r}} [\ |F_n(x_{j,r},\omega) - F(x_{j,r})|,$$
$$|F_n(x_{k,r}+0,\omega) - F(x_{k,r}+0)|] = 0\}.$$

If we let

$$S = \left\{\omega : \lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x,\omega) - F(x)| = 0\right\},$$

then $S \in \Sigma$, because if

$$\tilde{S} = \left\{\omega : \lim_{n\to\infty} \sup_{\substack{-\infty<r_i<\infty\\ r_i-rational}} |F_n(r_i,\omega) - F(r_i) = 0\right\},$$

clearly $S \subset \tilde{S}$, and by the density of rationals in $\mathbb{R}$, $\tilde{S} \subset S$ also follows. Since $\tilde{S} \in \Sigma$, so is $S \in \Sigma$. We need to establish the following result.

3. $\tilde{E} \subset S$, so that $1 = P(\tilde{E}) \le P(S) \le 1$. For, if we let $x \in (x_{k,r}, x_{k+1,r})$, then by the monotonicity of F_n and F, we get

$$F_n(x_{k,r}+0)(\omega) \le F_n(x)(\omega) \le F_n(x_{k+1,r})(\omega) \ \ a.a.(\omega),$$
$$F(x_{k,r}+0) \le F(x) \le F(x_{k+1,r}). \tag{6}$$

Let k, r be chosen in (5) such that $k \le r$ and $x_{k,r}$ satisfies

$$0 \le F(x_{k+1,r}) - F(x_{k,r}+0) \le \frac{k+1}{r} - \frac{k}{r} = \frac{1}{r}. \tag{7}$$

This is clearly possible since $F(x_{k+1,r}) \leq (k+1)/r$ and $F(x_{k,r}+0) \geq k/r$. Hence (6) may be written

$$\begin{aligned} F_n(x) - F(x) &\leq F_n(x_{k+1,r}) - F(x_{k,r}+0) \\ &\leq F_n(x_{k+1,r}) - F(x_{k+1,r}) + \frac{1}{r} \ a.e., \end{aligned}$$

and in a similar way

$$\begin{aligned} F_n(x) - F(x) &\geq F_n(x_{k,r}+0) - F(x_{k+1,r}) \\ &\geq F_n(x_{k,r}+0) - F(x_{k,r}+0) - \frac{1}{r} \ a.e. \end{aligned}$$

Combining these two sets of inequalities we get for a.a. (ω)

$$\begin{aligned} &\sup_{\substack{-\infty<x<\infty \\ x_{k,r}\leq x\leq x_{k+1,r} \\ 1\leq k\leq r}} \{|F_n(x) - F(x)|\}(\omega) \\ &\quad \leq \max_{\substack{0\leq j\leq r-1 \\ 0\leq k\leq r-1}} \{|F_n(x_{k+1,r}) - F(x_{k+1,r})|(\omega), \\ &\qquad\qquad |F_n(x_{j+1,r}+0) - F(x_{j+1,r}+0)|(\omega)\} + \frac{1}{r}. \end{aligned}$$

Since $r \geq 1$ is arbitrary, the left-side inequality holds if the right-side inequality does, for almost all ω. Hence $\omega \in \tilde{E} \Rightarrow \omega \in \tilde{S} = S$. Thus $\tilde{E} \subset S$, and the theorem is proved.

Remark: The empiric distribution has found substantial use in the statistical method known as the "Bootstrap". In the theory of statistics, bootstrapping is a method for estimating the sampling distribution of an estimator by "resampling" with replacement from the original sample.

In the proof of the theorem, one notes that the detailed analysis was needed above in extending the a.e. convergence of (4) for each x to uniform convergence in x over $\mathbb{R}$. This extension does not involve any real probabilistic ideas. It is essentially classical analysis. If we denote by $\mathcal{C}$ the class of all intervals $(-\infty, x)$, and denote by

$$\mu_n(A)(\omega) = \int_A F_n(dx, \omega)$$

and similarly

$$\mu(A) = \int_A F(dx),$$

then $\mu_n(A)$ is a sample "probability" of A [i.e., $\mu_n(\cdot)(\omega)$ is a probability for each $\omega \in \Omega$, and $\mu_n(A)(\cdot)$ is a measurable function for each Borel set A]; and μ is an ordinary probability (that is, determined by the common image measure). Then (3) says the following:

$$P[\lim_{n\to\infty} \sup_{A\in\mathcal{C}} |\mu_n(A) - \mu(A)| = 0] = 1. \tag{8}$$

This form admits an extension if $X_1, X_2, \ldots$ are random vectors. But here the correspondent for $\mathcal{C}$ must be chosen carefully, as the result will not be true for all collections because of the special sets demanded in Definition 2.1 (see the counterexample following it). For instance, the result will be true if $\mathcal{C}$ is the (corresponding) family of all half-spaces of $\mathbb{R}^n$. But the following is much more general and is due to R. Ranga Rao, *Ann. Math. Statist.* **33** (1962), 659-680.

Theorem 2 *Let $X_1, X_2, \ldots$ be a sequence of independent random vectors on (Ω, Σ, P) with values in $\mathbb{R}^m$, and for each Borel set $A \subset \mathbb{R}^m$, we have $\mu(A) = P[X_n \in A]$, $n \geq 1$, so that they have the common image measure μ (or distribution). Let $\mu_n(A)$ be the empiric distribution based on the sample (or initial segment) of size n (i.e. on, $X_1, \ldots, X_n$) so that*

$$\mu_n(A) = \frac{1}{n}\{\text{ number of } X_k \in A, 1 \leq k \leq n\}.$$

If $\mathcal{C}$ is the class of measurable convex sets from $\mathbb{R}^m$ whose boundaries have zero measure relative to the nonatomic part of μ, then

$$P\left[\lim_{n\to\infty} \sup_{A\in\mathcal{C}} |\mu_n(A) - \mu(A)| = 0\right] = 1. \tag{9}$$

We shall not present a proof of this result, since it needs several other auxiliary facts related to convergence in distribution, which have not been established thus far. However, this result, just as the preceding one, also starts its analysis from the basic SLLN for its probabilistic part.

(B) **Density Estimation**

Another application of this idea is to estimate the probability density by a method that is essentially due to Parzen (1962).

Suppose that $P[X < x] = F_X(x)$ is absolutely continuous relative to the Lebesgue measure on $\mathbb{R}$, with density $f(u) = (dF_X/dx)(u)$, and one wants to find an "empiric density" of $f(\cdot)$ in the manner of the Glivenko-Cantelli theorem. One might then consider the "empirical density"

$$f_n(x,h) = \frac{F_n(x+h) - F_n(x)}{h}$$

and find conditions for $f_n(x,h) \to f(x)$ a.e. as $n \to \infty$ and $h \to 0$. In contrast to the last problem, we have two limiting processes here which need additional work. Thus we replace h by h_n so that as $n \to \infty$, $h_n \to 0$. Since $F_n(x)$ itself is an ω-function, we still need extra conditions. Writing $\tilde{f}_n(x)$ for $f_n(x, h_n)$, this quotient is of the form

$$\tilde{f}_n(x)(\cdot) = \int_{\mathbb{R}} K\left(\frac{x-t}{h_n}\right) F_n(dt, \cdot) = \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{x - X_j(\cdot)}{h_n}\right), \tag{10}$$

for a suitable nonnegative function $K(\cdot)$, called a *kernel.* The approximations employed in Fourier integrals [cf. Bochner (1955)], Chapter I) give us some clues. Examples of kernels $K(t)$ are (i) e^{-t^2}, (ii) $e^{-t}\chi_{[t \geq 0]}$, (iii) $\chi_{[0,1]}$, and (iv) $1/(1+t^2)$. In this way we arrive at the following result of Parzen. [Actually he assumed a little more on K, namely, that $\hat{K}$, the Fourier transform of K, is also absolutely integrable, so that the examples (ii) and (iii) are not admitted. These are included in the following result. However, the ideas of proof are essentially his.]

Theorem 3 *Let $X_1, X_2, \ldots$ be independent identically distributed random variables on (Ω, Σ, P) whose common distribution admits a uniformly continuous density f relative to the Lebesgue measure on the line. Suppose that $K : \mathbb{R} \to \mathbb{R}^+$ is a bounded continuous function, except for a finite set of discontinuities, satisfying the conditions: (i) $\int_{\mathbb{R}} K(t)dt = 1$ and (ii) $|tK(t)| \to 0$ as $|t| \to \infty$. Define the "empiric density" $f_n : \mathbb{R} \times \Omega \to \mathbb{R}^+$ by*

$$f_n(x) = \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right), \quad x \in \mathbb{R} \tag{11}$$

where h_n is a sequence of numbers such that $nh_n^2 \to \infty$, but $h_n \to 0$ as $n \to \infty$. Then

$$P\left[\lim_{n\to\infty} \sup_{-\infty<x<\infty} |f_n(x) - f(x)| = 0\right] = 1. \tag{12}$$

Proof The argument here is somewhat different from the previous one, and it will be presented again in steps for convenience. As usual, let E be the expectation operator.

1. Consider

$$g_n(x) = E(f_n(x)) = \frac{1}{nh_n} \sum_{j=1}^{n} E\left(K\left(\frac{x - X_j}{h_n}\right)\right).$$

We assert that $g_n(x) \to f(x)$ uniformly in $x \in \mathbb{R}$ as $n \to \infty$. For using the i.i.d. hypothesis,

$$g_n(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_n}\int_{\mathbb{R}} K\left(\frac{x-y}{h_n}\right) f(y)dy. \tag{13}$$

If

$$v_n(x) = \frac{1}{h_n}\int_{\mathbb{R}} K\left(\frac{x-y}{h_n}\right) f(y)dy,$$

it suffices to show that $v_n(x) \to f(x)$ uniformly. Then since g_n is a $(C,1)$-average like v_n, it follows that $g_n(x) \to f(x)$ in the same sense. Since f is assumed to be uniformly continuous and integrable (and a probability density), it is easily seen that f is also bounded. Thus consider

$$v_n(x) - f(x) = \int_{\mathbb{R}} \frac{1}{h_n} K\left(\frac{t}{h_n}\right)[f(x-t) - f(x)]dt. \tag{14}$$

But, given $\varepsilon > 0$, there is a $\delta_\varepsilon > 0$ such that $|f(x-t) - f(x)| \le \varepsilon$ for $|t| \le \delta$ by the uniform continuity of f. Thus

$$\begin{aligned}
|v_n(x) - f(x)| \le & \int_{[|t|\le\delta]} \frac{1}{h_n} K\left(\frac{t}{h_n}\right) \max_{|t|\le\delta} |f(x-t) - f(x)| dt \\
& + \int_{[|t|\ge\delta]} K\left(\frac{t}{h_n}\right) \frac{f(x-t)}{|t|}\frac{|t|}{h_n} dt \\
& + f(x)\int_{[|t|\ge\delta]} K\left(\frac{t}{h_n}\right)\frac{1}{h_n}dt \\
& \le \max_{|t|\le\delta} |f(x-t) - f(x)| \int_{\mathbb{R}} K(u)du \\
& + \sup_{|u|\ge\delta/h_n} |uK(u)|\frac{1}{\delta}\int_{\mathbb{R}} f(t)dt \\
& + f(x)\int_{[|u|\ge\delta/h_n]} K(u)du \\
& \le \varepsilon + \frac{1}{\delta}\sup_{|u|\ge\delta/h_n} |uK(u)| + M\int_{[|u|\ge\delta/h_n]} K(u)du, \qquad (15)
\end{aligned}$$

since f is bounded. Letting $n \to \infty$, so that $h_n \to 0$, by (i) and (ii) both the second and third terms go to zero. Since the right side is independent of x, it follows that $v_n(x) \to f(x)$ uniformly in x, as $n \to \infty$.

2. We use now a result from Fourier transform theory. It is the following. Let $\hat{K}(u) = \int_{\mathbb{R}} e^{iux}K(x)dx$; then one has the inversion, in the sense that for almost every $x \in \mathbb{R}$ (i.e., except for a set of Lebesgue measure zero)

$$\lim_{a\to\infty}\frac{1}{2\pi}\int_{-a}^{a}\left(1-\frac{|u|}{a}\right)e^{-iux}\hat{K}(u)du = K(x). \tag{16}$$

Results of this type for distribution functions, called "inversion formulas," will be established in Chapter 4. If K is assumed integrable, then the above integral can be replaced by $(1/2\pi)\int_{\mathbb{R}} e^{-iux}\hat{K}(u)du = K(x)$ a.e. so (16) is the $(C,1)$-summability result for integrals, an exact analog for series that we noted in the preceding step.

Let $\phi_n(u) = (1/n)\sum_{j=1}^{n} e^{iuX_j}$. Then $e^{iuX_j} = \cos uX_j + i\sin uX_j$ is a bounded complex random variable and, for different j, these are identically distributed. Thus applying the SLLN to the real and imaginary parts, we get

$$\lim_{n\to\infty}\phi_n(u) = E(\phi_1(u)) = E(e^{iuX_1}) \ a.e.[P]. \tag{17}$$

If $\omega \in \Omega$ is arbitrarily fixed, then $\phi_n(u)$ can be regarded as

$$\phi_n(u)(\omega) = \int_{\mathbb{R}} e^{iux}F_n(dx,\omega), \tag{18}$$

where F_n is the empiric distribution of the X_n. Now using the "inversion formula" (16) for $\hat{K}$, we can express f_n as follows:

$$\begin{aligned}
&\lim_{a\to\infty}\tfrac{1}{2\pi}\int_{-a}^{a}\left(1-\frac{|u|}{a}\right)\hat{K}(h_n u)e^{-iux}\phi_n(u)(\omega)du\\
&= \frac{1}{n}\sum_{j=1}^{n}\lim_{a\to\infty}\frac{1}{2\pi}\int_{-a}^{a}\left(1-\frac{|u|}{a}\right)e^{-iu(x-X_j(\omega))}\hat{K}(h_n u)du\\
&= \frac{1}{n}\sum_{j=1}^{n}\lim_{\tilde{a}\to\infty}\frac{1}{2\pi}\int_{-\tilde{a}}^{\tilde{a}}\left(1-\frac{|r|}{\tilde{a}}\right)\exp\left(-i\frac{r}{h_n}t_j\right)\hat{K}(r)\frac{dr}{h_n}\\
&\qquad \text{with } t_j = x - X_j(\omega) \text{ and } \ \tilde{a} = h_n a,\\
&= \frac{1}{nh_n}\sum_{j=1}^{n} K\left(\frac{t_j}{h_n}\right)\\
&\qquad [\text{ by the inversion formula, a.e. (Lebesgue measure)}]\\
&= f_n(x)(\omega) \ [\text{ by}(12)].
\end{aligned} \tag{19}$$

We need this formula to get uniform convergence of $f_n(x)$ to $f(x)$.

3. The preceding work can be used in our proof in the following manner. By Markov's inequality

$$P\left[\lim_{n\to\infty}\sup_{x\in\mathbb{R}}|f_n(x)-f(x)| \ge \varepsilon\right] \le \lim_{n\to\infty}\frac{E\left[\sup_{x\in\mathbb{R}}|f_n(x)-f(x)|\right]}{\varepsilon}, \tag{20}$$

where the limit can be brought outside of the P-measure by Fatou's lemma. (Note that the sup inside the square brackets is bounded by hypothesis and is a measurable function, by the same argument as in step 2 of the proof of Theorem 1. The existence of limit in (20) will be proved.) We now show that the right side of (20) is zero, so that (12) results. But if $||\cdot||_u$ is the uniform (or supremum) norm over $\mathbb{R}$, then

$$||f_n(\cdot) - f(\cdot)||_u \leq ||f_n(\cdot) - g_n(\cdot)||_u + ||g_n(\cdot) - f(\cdot)||_u, \tag{21}$$

and $x \mapsto g_n(x) = E(f_n(x))$ is a constant function (independent of ω). By step 1, the last term goes to zero as $n \to \infty$, and hence its expectation will go to zero by the dominated convergence since the terms are bounded. Thus it suffices to show that the expectation of the first term also tends to zero uniformly in x. Consider

$$\begin{aligned}||f_n(\cdot) - g_n(\cdot)||_u = \sup_{x\in\mathbb{R}} \Bigg| \frac{1}{2\pi}\int_{-a}^{a}\left(1 - \frac{|u|}{a}\right)\hat{K}(h_n u) \\ \times\, e^{-iux}[\phi_n(u) - E(\phi_n(u))]du\Bigg|,\end{aligned}$$

where we used (19) and the fact that $g_n(x) = E(f_n(x))$, which is again obtained from (19) with $E(\phi_n(u))$. With the same computation, first using the Fubini theorem and then the dominated convergence theorem to interchange integrals on $[-a, a]\times\Omega$, we can pass to the limit as $a \to \infty$ through a sequence under the expectation. Thus

$$||f_n(\cdot) - g_n(\cdot)||_u \leq \lim_{a\to\infty} \frac{1}{2\pi}\int_{-a}^{a}\left(1 - \frac{|u|}{a}\right)\hat{K}(h_n u)|\phi_n(u) - E(\phi_n(u))|du. \tag{22}$$

But by (17), $|\phi_n(u) - E(\phi_n(u))| = |\phi_n(u) - E(e^{iuX_1})| \to 0$ a.e., and since these quantities are bounded, this is also true boundedly. Thus by letting $n \to \infty$ in both sides of (22) and noting that the limits on a and n are on independent sets, it follows that the right side of (22) is zero a.e. By the uniform boundedness of the left-side norms in (22), we can take expectations, and the result is zero.

Thus $E(||f_n(\cdot) - f(\cdot)||_u) \to 0$ as $n \to \infty$, and the right side of (20) is zero. This completes the proof.

Remark Evidently, instead of (17), even WLLN is sufficient for (22). Also, using the CBS-inequality in (22) and taking expectations, one finds that $Var(\phi_n) \leq M_1/n$ and this yields the same conclusion without even using WLLN. (However, this last step is simply the proof of the WLLN, as given by Čebyšev.) It is clear that considerable analysis is needed in these results, after employing the probabilistic theorems in key places. Many of the applications use such procedures.

(C) Queueing

We next present a typical application to queueing theory. Such a result was originally considered by A. Kolmogorov in 1936 and is equivalent to a one-server queueing model. It admits extensions and raises many other problems. The formulation using the current terminology appears to be due to D. V. Lindley.

A general queueing system consists of three elements: (i) customers, (ii) service, and (iii) a queue. These are generic terms; they can refer to people at a service counter, or planes or ships arriving at a port facility, etc. The arrival of customers is assumed to be random, and the same is true of the service times as well as waiting times in a queue. Let a_k be the interarrival time between the kth and the $(k+1)$th customer, b_k the service time, and W_k the waiting time of the kth customer. When customer one arrives, we assume that there is no waiting, since there is nobody ahead of this person. Thus it is reasonable to assume $a_0 = W_0 = 0$. Now $b_k + W_k$ is the length of time that the $(k+1)$th customer has to wait in the queue before the turn comes at the service counter. We assume that the interarrival times a_k are independent nonnegative random variables with a common distribution, and similarly, the b_k are nonnegative i.i.d. and independent of the a_k. As noted before, we can assume that the basic probability space is rich enough to support such independent sequences, as otherwise we can enlarge it by adjunction to accomplish this. The waiting times are also positive random variables. If $a_{k+1} > b_k + W_k$,then the $(k+1)$th customer obviously does not need to wait on arrival, but if $a_{k+1} \leq b_k + W_k$ then the person has to wait $b_k + W_k - a_{k+1}$ units of time. Thus

$$W_{k+1} = \max(W_k + b_k - a_{k+1}, 0), \quad k \geq 0. \tag{23}$$

If we let $X_k = b_{k-1} - a_k$, then the X_k are i.i.d. random variables, and (23) becomes $W_0 = 0$ and $W_{k+1} = \max(W_k + X_{k+1}, 0), k \geq 0$. Note that whenever $W_k = 0$ for some k, the server is free and the situation is like the one at the beginning, so that we have a recurrent pattern. This recurrence is a key ingredient of the solution of the problem of finding the limiting behavior of the W_k-sequence. It is called the *single server queueing* problem.

Consider $S_0 = 0$, $S_n = \sum_{k=1}^{n} X_k$. Then the sequence $\{S_n, n \geq 0\}$ is also said to perform a *random walk* on $\mathbb{R}$, and if $S_k \in A$ for some $k \geq 0$ and Borel set A, one says that the walk S_n visits A at step k. In the queueing situation, we have the following statement about the process $\{W_n, n \geq 0\}$.

Theorem 4 *Let $X_k = b_{k-1} - a_k$, $k \geq 1$, and $\{S_n, n \geq 0\}$ be as above. Then for each $n \geq 0$, the quantities W_n and $M_n = \max\{S_j, 0 \leq j \leq n\}$ are identically distributed random variables. Moreover, if $F_n(x) = P[W_n < x]$, then*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

exists for each x, but $F(x) = 0$ is possible. If $E(X_1)$ exists, then $F(x) \equiv 0$, $x \in \mathbb{R}$, whenever $E(X_1) \geq 0$, and $F(\cdot)$ defines an honest distribution function when $E(X_1) < 0$, i.e., $F(+\infty) = 1$.

The last statement says that if $E(b_k) \geq E(a_k), k \geq 1$, so that the expected service time is not smaller than that of the interarrival time, then the line of customers is certain to grow longer without bound (i.e., with probability 1).

Proof For the first part of the proof we follow Feller (1966), even though it can also be proved by using the method of convolutions and the fact that W_k and X_{k+1} are independent. The argument to be given is probabilistic and has independent interest.

Since $W_0 = 0 = S_0$, we may express W_n in an alternative form as $W_n = \max\{(S_n - S_k) : 0 \leq k \leq n\}$. In fact, this is trivial for $n = 0$; suppose it is verified for $n = m$. Then consider the case $n = m + 1$. Writing $S_{n+1} - S_k = S_n - S_k + X_{n+1}$, we have with $\vee$ for "max"

$$\begin{aligned}\max_{0\leq k\leq m+1}(S_{m+1} - S_k) &= \left(\max_{0\leq k\leq m+1}(S_m - S_k) + X_{m+1}\right) \vee 0 \\ &= (W_m + X_{m+1}) \vee 0 = W_{m+1}. \qquad (24)\end{aligned}$$

Hence the statement is true for all $m \geq 0$. On the other hand, $X_1, \ldots, X_n$ are i.i.d. random variables. Thus the joint distribution of $X_1, X_2, \ldots, X_n$ is the same as that of $X_1', X_2', \ldots, X_n'$,where $X_1' = X_n$, $X_2' = X_{n-1}, \ldots, X_n' = X_1$. But the joint distribution of $S_0', S_1', \ldots, S_n'$, where $S_n' = \sum_{k=1}^n X_k' (S_0' = 0)$, and that of $S_0, S_1, \ldots, S_n$ must also be the same. This in turn means, on substituting the unprimed variables, that $S_0, S_1, \ldots, S_n$ and S_0', $S_1' = S_n - S_{n-1}$, $S_2' = S_n - S_{n-2}, \ldots, S_n' = S_n - S_0$ are identically distributed. Putting these two facts together, we get $\max_{0\leq k\leq n} S_k'$ and $\max_{0\leq k\leq n}(S_n - S_k) = W_n$ are identically distributed. But the $S_0', S_1', \ldots, S_n'$ and $S_0, S_1, \ldots, S_n$ were noted to have the same distribution, so that $\max_{0\leq k\leq n} S_k' \stackrel{D}{=} \max_{0\leq k\leq n} S_k$ or M_n and W_n have the same distribution. This is the first assertion in which we only used the i.i.d. property of the X_n but not the fact that $X_n = b_{n-1} - a_n$.

The above analysis implies

$$F_n(x) = P[W_n < x] = P[M_n < x] = P\left[\max_{0\leq k\leq n} S_k < x\right].$$

But

$$\left[\max_{0\leq k\leq n} S_k < x\right] \downarrow \left[\lim_{n\to\infty}\max_{0\leq k\leq n} S_k < x\right],$$

so that

$$F(x) = \lim_{n\to\infty} F_n(x) = P\left[\lim_{n\to\infty}\max_{0\leq k\leq n} S_k \leq x\right] = P\left[\sup_{k\geq 0} S_k \leq x\right], \qquad (25)$$

exists, and $0 \leq F(x) \leq 1, x \in \mathbb{R}$. Clearly $F(x) = 0$ for $x < 0$. On the other hand, if $E(|X_1|) = \infty$, then by Theorem 2.7, $\limsup_n |S_n| = +\infty$ a.e. which implies (since $S_0 = 0$) that either $\limsup_n S_n = +\infty$ a.e., so that $\sup_n S_n = +\infty$ a.e., or this can happen with probability zero. Since $F(x) = 0$ for $x < 0$, we only need to consider $x \geq 0$. Thus if $\sup_n S_n = +\infty$ a.e., then $1 - F(x) = P[\sup_{k \geq 0} S_k > x] = 1$, so that $F(x) = 0, x \in \mathbb{R}$. If $\sup_n S_n < \infty$ a.e., then $\lim_{x \to \infty} F(x) = 1$ and F is a distribution function. Note that, since $\sup_{n \geq 0} S_n$ is a symmetric function of the random variables X_n, which are i.i.d., we can deduce that $\sup_n S_n = \infty$ has probability zero or one by Theorem 1.12 so that (25) can be obtained in this way also.

Suppose that $E(|X_1|) < \infty$. Then we consider the cases (i) $E(X_1) > 0$, (ii) $E(X_1) < 0$, and (iii) $E(X_1) = 0$ separately for calculating the probability of A_x, where

$$A_x = \left[\sup_{n \geq 0} S_n \leq x\right].$$

Case (i): $\mu = E(X_1) > 0$: By the SLLN, $S_n/n \to E(X_1)$ a.e., so that for sufficiently large n, $S_n > E(X_1) \cdot \frac{n}{2}$ a.e. Thus

$$A_x = \bigcap_{n \geq 0} [S_n \leq x] \subset \left\{\omega : S_n(\omega) > \frac{n}{2}\mu, n \geq n_\omega\right\}^c$$

for any $x \in \mathbb{R}^+$, and hence $P(A_x) = 0$, or $F(x) = 0, x \in \mathbb{R}^+$, in this case.

Case (ii): $\mu = E(X_1) < 0$: Again by the SLLN, $S_n/n \to E(X_1)$ a.e., and given $\varepsilon > 0$, and $\delta > 0$, one can choose $N_{\varepsilon\delta}$ such that $n \geq N_{\varepsilon\delta}$ implies

$$P[|(S_n/n) - E(X_1)| \leq \varepsilon, n \geq N_{\varepsilon\delta}] \geq 1 - \delta \tag{26}$$

This may be expressed in the following manner. Let $\varepsilon > 0$ be small enough so that $E(X_1) + \varepsilon < 0$. Then for $0 < \delta < \frac{1}{2}$, choose $N_{\varepsilon\delta}$ such that with (26),

$$\begin{aligned} P[S_n < 0, n \geq N_{\varepsilon\delta}] &\geq P[S_n \leq n(\mu + \varepsilon), n \geq N_{\varepsilon\delta}] \\ &\geq P[n(\mu - \varepsilon) \leq S_n \leq (\mu + \varepsilon)n, n \geq N_{\varepsilon\delta}] \\ &\geq 1 - \delta. \end{aligned} \tag{27}$$

For this $N_{\varepsilon\delta}$, consider the finite set $S_1, S_2, \ldots, S_{N_{\varepsilon\delta}-1}$. Since these are real random variables, we can find an $x_\delta \in \mathbb{R}^+$ such that $x \geq x_\delta$ implies

$$P[S_1 < x, \ldots, S_{N_{\varepsilon\delta}-1} < x] > 1 - \delta. \tag{28}$$

If now

$$\tilde{A}_x = \bigcap_{n \geq N_{\varepsilon\delta}} [S_n < x], \quad B_x = \bigcap_{k=1}^{N_{\varepsilon\delta-1}} [S_k < x], \quad A_x = \bigcap_{n=1}^{\infty} [S_n < x],$$

then $A_x = \tilde{A}_x \cap B_x$, for $x \geq 0$. Hence we have

$$\begin{aligned} F(x) &= P(A_x) = P(\tilde{A}_x \cap B_x) \\ &= P(\tilde{A}_x) + P(B_x) - P(\tilde{A}_x \cup B_x) \\ &\geq 2(1-\delta) - 1 = 1 - 2\delta \quad [\text{ by } (27\ ) \text{ and } (28)\ ]. \end{aligned}$$

Since $0 < \delta < \frac{1}{2}$ is arbitrary, we conclude that $\lim_{x\to\infty} F(x) = 1$, and hence F gives an honest distribution in this case.

Case (iii): $E(X_1) = 0$: Now $S_n = \sum_{i=1}^{n} X_i$, $n \geq 1$, is a symmetrically dependent sequence of random variables and $S_0 = 0$. Thus $\sup_{n\geq 0} S_n \geq 0$ a.e., and since we can assume that $X_1 \not\equiv 0$ a.e., all the S_n do not vanish identically a.e. Consider the r.v. $Y = \limsup_n S_n$. Then $Y[= Y(S_n, n \geq 1)]$ is symmetrically dependent on the S_n and is measurable for the tail σ-algebra. Hence, by Theorem 1.12 it is a constant $= k_0$ a.e. It will be seen later (cf. Theorem 8 below) that, since $S_n/n \to 0$, a.e. by the SLLN, S_n takes both positive and negative values infinitely often. Thus $k_0 \geq 0$. But then

$$\begin{aligned} 0 \leq Y &= \limsup_{n\geq 1} S_n \\ &= \limsup_{n\geq 1}(X_1 + \ldots + X_n) = X_1 + \limsup_{n\geq 2}(X_2 + \ldots + X_n) \\ &\overset{D}{=} X_1 + Y. \end{aligned} \tag{29}$$

Since $Y = k_0$ a.e. and X_1 is a real nonzero r.v., (29) can hold only if $k_0 = +\infty$.

Now $[\limsup_{n\geq 1} S_n = +\infty] \subset [\sup_{n\geq 0} S_n = \infty]$, and so we are back in the situation treated in case (i), i.e., $F(x) = 0, x \in \mathbb{R}$. This completes the proof of the theorem.

The preceding result raises several related questions, some of which are the following. When $E(X_1) < 0$, we saw that the waiting times $W_n \to W$ in distribution where W is a (proper) r. v. Thus, in this case, if Q_n is the number of customers in the queue when the service of the nth customer is completed, then Q_n is an r.v. But then what is the distribution of Q_n, and does $Q_n \xrightarrow{D} Q$? Since Q_n is no more than k iff the completion of the nth customer service time is no more than the interarrival times of the last k customers, we get

$$P[Q_n < k] = P[W_n + b_n \leq a_{n+1} + \ldots + a_{n+k}]. \tag{30}$$

The random variables on the right are all independent, and thus this may be calculated explicitly in principle. Moreover, it can be shown, since $W_n \xrightarrow{D} W$ and the b_n and a_n are identically distributed, that $Q_n \xrightarrow{D} Q$ from this expression.

Other questions, such as the distribution of the times that $W_n = 0$, suggest themselves. Many of these results use some properties of convolutions of the image measures (i.e., distribution functions) on $\mathbb{R}$, and we shall omit consideration of these specializations here.

All of the above discussions concerned a single-server queueing problem. But what about the analogous problem with many servers? This is more involved. The study of these problems has branched out into a separate discipline because of its great usefulness in real applications. Here we consider only one other aspect of the above result.

(D) **Fluctuation Phenomena**

In Theorem 4 we saw that the behavior of the waiting time sequence is governed by $S_n = \sum_{k=1}^n X_k$, the sequence of partial sums of i.i.d. random variables. In Section 2 we considered the convergence of sums of general independent random variables, but the surprising behavior of i.i.d. sums was not analyzed more thoroughly. Such a sequence is called a *random walk*. Here we include an introduction to the subject that will elaborate on the proof of Theorem 4 and complete it. The results are due to Chung and Fuchs. We refer to Chung (1974). For a detailed analysis of the subject, and its relation to the group structure of the range space, see Spitzer (1964).

Thus if $X_n, n \geq 1$, are i.i.d., and $\{S_n = \sum_{k=1}^n X_k, n \geq 1\}$ is a random walk sequence, let $Y = \limsup_n X_n$. We showed in the proof of Theorem 4 [Case (iii)] that $Y = X_1 + Y$ and Y is a "permutation invariant" r.v. Then this equation implies $Y = k_0$ a.e. ($= \pm\infty$ possibly), by the Hewitt-Savage zero-one law. If $X_1 = 0$ a.e., then by the i.i.d. condition, all $X_n = 0$ a.e., so that $S_n = 0$ for all n (and $Y = 0$ a.e.). If $X_1 \not\equiv 0$ a.e., then $k_0 = -\infty$ or $+\infty$ only. If $k_0 = -\infty$, then clearly $-\infty \leq \liminf_n S_n \leq \limsup_n S_n = -\infty$, so that $\lim_{n\to\infty} S_n = -\infty$ a.e.; or if $k_0 = +\infty$, then $\liminf_n S_n$ can be $+\infty$, in which case $S_n \to +\infty$ a.e., or $\liminf_n S_n = -\infty < \limsup_n S_n = +\infty$. Since $\limsup_n(S_n) = -\liminf_n(-S_n)$, no other possibilities can occur. In the case $-\infty = \liminf_n S_n < \limsup_n S_n = +\infty$ a.e. (the interesting case), we can look into the behavior of $\{S_n, n \geq 1\}$ and analyze its fluctuations.

A state $x \in \mathbb{R}$ is called a *recurrent point* of the range of the sequence if for each $\varepsilon > 0$, $P[|S_n - x| < \varepsilon, i.o.] = 1$, i.e., the random walk visits x infinitely often with probability one. Let R be the set of all recurrent points of $\mathbb{R}$. A point $y \in \mathbb{R}$ is termed a *possible value* of the sequence if for each $\varepsilon > 0$, there is a k such that $P[|S_k - y| < \varepsilon] > 0$. We remark that by Cases (i) and (ii) of the proof of Theorem 4, if $E(X_1) > 0$ or < 0, then $\lim_{n\to\infty} S_n = +\infty$ or $= -\infty$ respectively. Thus fluctuations show up only in the case $E(X_1) = 0$ when the expectation exists. However, $E(|X_1|) < \infty$ will not be assumed for the present discussion.

Theorem 5 *For the random walk $\{S_n, n \geq 1\}$, the set R of recurrent values (or points) has the following description: Either $R = \emptyset$ or $R \subset \mathbb{R}$ is a* *closed subgroup. In the case $R \neq \emptyset$, $R = \{0\}$ iff $X_1 = 0$ a.e., and if $X_1 \neq 0$ a.e., we have either $R = \mathbb{R}$ or else $R = \{nd : n = 0, \pm 1, \pm 2, \ldots\}$, the infinite cyclic group generated by a number $d > 0$.*

Proof Suppose $R \neq \emptyset$. If $x_n \in R$ and $x_n \to x \in \mathbb{R}$, then given $\varepsilon > 0$, there is n_ε such that $n \geq n_\varepsilon \Rightarrow |x_n - x_1| < \varepsilon$. Thus letting $S_n(\omega) = x_n$, we get $|S_n(\omega) - x| < \varepsilon$, $n \geq n_\varepsilon(\omega)$, for almost all ω, and hence if $I = (x - \varepsilon, x + \varepsilon)$, then $P[S_n \in I, i.o.] = 1$. Since $\varepsilon > 0$ is arbitrary, $x \in R$, and so R is closed.

To prove the group property, let $x \in R$ and $y \in \mathbb{R}$ be a possible value of the random walk. We claim that $x - y \in R$. Indeed for each $\varepsilon > 0$, choose m such that $P[|S_m - y| < \varepsilon] > 0$. Since x is recurrent, $P[|S_n - x| < \varepsilon, i.o.] = 1$. Or equivalently $P[|S_n - x| < \varepsilon$, finitely many n only $] = 0$. Let us consider, since $[|S_n - x| < \varepsilon$ for finitely many $n] = [|S_n - x| \geq \varepsilon$ for all but finitely many $n]$,

$$\begin{aligned} &P[|S_n - x| < \varepsilon, \text{ finitely often }] \\ &\geq P[|S_m - y| < \varepsilon, |S_{m+n} - S_m - (x-y)| \geq 2\varepsilon, \text{ all } n \geq k_0],\ k_0 \geq 1, \\ &= P[|S_m - y| < \varepsilon] P\left[\left|\sum_{k=m+1}^{m+n} X_k - (x-y)\right| < 2\varepsilon, \text{ finitely often }\right], \\ &\quad \text{(by the independence of } S_m \text{ and } S_{m+n} - S_m). \end{aligned} \tag{31}$$

By hypothesis $P[|S_m - y| < \varepsilon] > 0$, and this shows that the second factor of (31) is zero. But by the i.i.d. hypothesis, S_n and $S_{m+n} - S_m$ have the same distribution. Hence $P[|S_n - (x-y)| < 2\varepsilon$, finitely many $n] = 0$, and $x - y \in R$. Since $y = x$ is a possible value, $0 \in R$ always, and $x - (x - y) = y \in R$. Similarly $0 - y \in R$ and so R is a group. As is well known, the only closed subgroups of $\mathbb{R}$ are those of the form stated in the theorem,[2] and $R = \{0\}$ if $X_1 \equiv 0$ a.e. In the case that $X_1 \neq 0$ a.e., there is a possible value $y \in \mathbb{R}$ of the random walk, and $y \in R$ by the above analysis. Thus $R = \{0\}$ iff $X_1 \equiv 0$ a.e. It is of interest also to note that unless the values of the r.v. X_1 are of the form $nd, n = 0, \pm 1, \pm 2, \ldots, R = \mathbb{R}$ itself. This completes the proof.

It is clear from the above result that 0 plays a key role in the recurrence phenomenon of the random walk. A characterization of this is available:

Theorem 6 *Let $\{X_n, n \geq 1\}$ be i.i.d. random variables on (Ω, Σ, P) and $\{S_n, n \geq 0\}$ be the corresponding random walk sequence. If for an $\varepsilon > 0$ we*

[2] Indeed,if $R \neq \emptyset$, because it is a closed subgroup of $\mathbb{R}$, let $d = \inf\{x \in R, x > 0\}$. Then $d \geq 0$ and there exist $d_n \in R, d_n \downarrow d$. If $d = 0$, we can verify that $\{kd_n, k = 0, \pm 1, \pm 2, \ldots; n \geq 1\}$ is dense in $\mathbb{R}$ and $\subset R \Rightarrow R = \mathbb{R}$. If $d > 0$,then $\{nd, n = 0, \pm 1, \ldots\} \subset R$ and is all of R. There are no other kinds of groups. Note that if $R \neq \emptyset$ every possible value is also a recurrent value of the random walk.

have

$$\sum_{n=1}^{\infty} P[|S_n| < \varepsilon] < \infty, \tag{32}$$

then 0 *is not a recurrent value of* $\{S_n, n \geq 0\}$. *If, on the other hand,for every* $\varepsilon > 0$ *it is true that the series in (32) diverges, then* 0 *is recurrent.* [*It follows from (36) below that if the series (32) diverges for one* $\varepsilon > 0$, *then the same is true for all* $\varepsilon > 0$.]

Proof If the series in (32) converges, then the first Borel-Cantelli lemma implies $P[|S_n| < \varepsilon$, finitely often $] = 1$ so that $0 \notin R$. The second part is harder, since the events $\{[|S_n| < \varepsilon], n \geq 1\}$ are *not* independent. Here one needs to show that $P[|S_n| < \varepsilon, i.o.] = 1$. We consider the complementary event and verify that it has probability zero, after using the structure of the S_n sequence.

Consider for any fixed $k \geq 1$ the event A_m^k defined as

$$A_m^k = [|S_m| < \varepsilon, |S_n| \geq \varepsilon, n \geq m + k]. \tag{33}$$

Then A_m^k is the event that the S_n will not visit $(-\varepsilon, \varepsilon)$ after the $(m+k-1)$th trial, but visits at the mth trial [from the $(m+1)$th to $(m+k-1)$th trials, it may or may not visit]. Hence $A_m^k, A_{m+k}^k, A_{m+2k}^k, \ldots$ are disjoint events for $m \geq 1$ and fixed $k \geq 1$. Thus

$$\sum_{m=1}^{\infty} P(A_m^k) = k + \sum_{m=k+1}^{\infty} P(A_m^k) \leq k + \sum_{\ell=1}^{k} \sum_{j=1}^{\infty} P(A_{\ell+jk}^k)$$

$$= \sum_{\ell=1}^{k} P\left(\bigcup_{j=1}^{\infty} A_{\ell+jk}^k \right) \leq 2k. \tag{34}$$

But for each $k \geq 1$, $[|S_n| < \varepsilon]$ and $[|S_n - S_m| \leq 2\varepsilon, n \geq m+k]$ are independent, and $A_m^k \supset [|S_m| < \varepsilon] \cap [|S_n - S_m| \geq 2\varepsilon, n \geq m + k]$, $k \geq 1$, since $|S_n| \geq (|S_n - S_m| - |S_m|) \geq 2\varepsilon - \varepsilon = \varepsilon$, on the displayed set. Hence, with independence, (34) becomes

$$\sum_{m=1}^{\infty} P[|S_m| < \varepsilon] P[|S_n - S_m| \geq 2\varepsilon, n \geq m + k]$$

$$= \sum_{m=1}^{\infty} P[|S_m| < \varepsilon, |S_n - S_m| \geq 2\varepsilon, n \geq m + k] \leq \sum_{m=1}^{\infty} P(A_m^k) \leq 2k.$$

But

$$P[|S_n - S_m| \geq 2\varepsilon, n \geq m + k] = P[|X_{m+1} + \ldots + X_n| \geq 2\varepsilon, n \geq m + k]$$

$$= P[|S_n| \geq 2\varepsilon, n \geq k] \quad \text{(by the i.i.d. condition)}.$$

Hence

$$\sum_{m=1}^{\infty} P[|S_m| < \varepsilon] \cdot P[|S_n| \geq 2\varepsilon, n \geq k] \leq 2k.$$

Since we may take the second factor on the left out of the summation, and since the sum is divergent by hypothesis, we must have $P[|S_n| \geq 2\varepsilon, n \geq k] = 0$ for each k. Hence taking the limit as $k \to \infty$, we get

$$P[|S_n| > 2\varepsilon, \quad \text{finitely often }] = P\left[\bigcup_{k\geq 1}\bigcap_{n\geq k}[|S_n| \geq 2\varepsilon]\right] = 0,$$

or $P[|S_n| < \varepsilon, i.o.] = 1$ for any $\varepsilon > 0$. This means $0 \in R$ and completes the proof of the theorem.

Suppose, in the above, the $X_n : \Omega \to \mathbb{R}^k$ are i.i.d. random vectors and $S_n = \sum_{i=1}^n X_i$. If $|X_i|$ is interpreted as the maximum absolute value of the k components of X_i, and S_n visits $(-\varepsilon, \varepsilon)$ means it visits the cube $(-\varepsilon, \varepsilon)^k \subset \mathbb{R}^k$ (i.e., $|S_n| < \varepsilon$), then the preceding proof holds verbatim for the k-dimensional random variables, and establishes the corresponding result for the k-dimensional random walk. We state the result for reference as follows:

Theorem 7 *Let $\{X_n, n \geq 1\}$ be i.i.d. k-vector random variables on (Ω, Σ, P) and $S_n = \sum_{i=1}^n X_i$, $S_0 = 0$, where $k \geq 1$. Then 0 is a recurrent value of the k-random walk $\{S_n, n \geq 0\}$ iff for any $\varepsilon > 0$,*

$$\sum_{n=1}^{\infty} P[|S_n| < \varepsilon] = +\infty. \tag{35}$$

Moreover, the set of all recurrent values R forms a closed subgroup of the additive group $\mathbb{R}^k$.

The proof of the last statement is the same as that for Theorem 5, which has a more precise description of R in case $k = 1$.

If $R = \emptyset$, then the random walk is called *transient*, and is termed *recurrent*, (or *persistent*) if $R \neq \emptyset$.

We can now present a sufficient condition for the recurrence of a random walk, and this completes the proof of case (iii) of Theorem 4.

Theorem 8 *Let $S_n = X_1 + \ldots + X_n$, $\{X_n, n \geq 1$, i.i.d.$\}$ be a (real) random walk sequence on (Ω, Σ, P) such that $S_n/n \xrightarrow{P} 0$. Then the walk is recurrent.*

Remark As noted prior to Theorem 3.2, this condition holds for certain symmetric random variables without the existence of the first moment. On the other hand, if $E(|X_1|) < \infty$, then it is always true by the WLLN (or SLLN).

We shall establish the result with the weaker hypothesis as stated. The proof uses the linear order structure of the range of S_n. Actually the result itself is not valid in higher dimensions (≥ 3). It is true in 2-dimensions, but needs a different method with characteristic functions (cf. Problem 21.)

Proof We first establish an auxiliary inequality, namely, for each $\varepsilon > 0$,

$$\sum_{m=0}^{r} P[|S_m| < k\varepsilon] \leq 2k \sum_{m=0}^{r} P[|S_m| < \varepsilon], \quad r, k \geq 1, \text{ integers.} \tag{36}$$

If this is granted, the result can be verified (using an argument essentially due to Chung and Ornstein (1962)) as follows: We want to show that (32) fails. Thus for any integer $b > 0$, let $r = kb$ in (36). Then

$$\sum_{m=0}^{kb} P[|S_m| < \varepsilon] \geq \frac{1}{2k} \sum_{m=0}^{kb} P[|S_m| < k\varepsilon] \geq \frac{1}{2k} \sum_{m=0}^{kb} P\left[|S_m| < \frac{m\varepsilon}{b}\right], \tag{37}$$

because $(m/b) \leq k$. By hypothesis $S_m/m \xrightarrow{P} 0$, so that $P[|S_m|/m < \varepsilon/b] \to 1$ as $m \to \infty$. By the $(C, 1)$-summability,

$$\lim_{k\to\infty} \frac{1}{kb} \sum_{m=0}^{kb} P\left[\frac{|S_m|}{m} < \frac{\varepsilon}{b}\right] = 1, \quad \text{for each } b > 0.$$

Hence (37) becomes on letting $k \to \infty$

$$\sum_{m=0}^{\infty} P[|S_m| < \varepsilon] \geq \frac{b}{2} \lim_{k\to\infty} \frac{1}{kb} \sum_{m=0}^{kb} P\left[\frac{|S_m|}{m} < \frac{\varepsilon}{b}\right] = \frac{b}{2}.$$

Since $b > 0$ is arbitrary, (32) fails for each $\varepsilon > 0$, and so $\{S_n, n \geq 1\}$ is recurrent.

It remains to establish (36). Consider, for each integer m, $[m\varepsilon \leq S_n < (m+1)\varepsilon]$ and write it as a disjoint union:

$$[m\varepsilon \leq S_n < (m+1)\varepsilon] = \bigcup_{k=0}^{n} [m\varepsilon \leq S_n < (m+1)\varepsilon] \cap A_k, \tag{38}$$

where $A_0 = [m\varepsilon \leq S_0 < (m+1)\varepsilon]$ and for $k \geq 1$, $A_k = [S_k \in [m\varepsilon, (m+1)\varepsilon), S_j \notin [m\varepsilon, (m+1)\varepsilon), 0 \leq j \leq k-1]$. Thus A_k is the ω-set for which S_k enters the interval $[m\varepsilon, (m+1)\varepsilon]$ for the first time. Then

$$\begin{aligned}
&\sum_{n=0}^{r} P[S_n \in [m\varepsilon, (m+1)\varepsilon)] \\
&\qquad = \sum_{n=0}^{r} \sum_{k=0}^{n} P[(S_n \in [m\varepsilon, (m+1)\varepsilon) \cap A_k] \quad [\text{ by } (38)]
\end{aligned}$$

$$\leq \sum_{n=0}^{r} \sum_{k=0}^{n} P[A_k \cap [(S_n - S_k) \in (-\varepsilon, \varepsilon)]]$$

$$[\text{since on} A_k, m\varepsilon \leq S_n, S_k < (m+1)\varepsilon \Rightarrow |S_n - S_k| < \varepsilon]$$

$$= \sum_{n=0}^{r} \sum_{k=0}^{n} P(A_k) P[|S_n - S_k| < \varepsilon]$$

$$(\text{since } A_k \text{ is determined by } X_1, \ldots, X_k$$

$$\text{and hence is independent of } S_n - S_k \text{ for } n \geq k)$$

$$= \sum_{n=0}^{r} \sum_{k=0}^{n} P(A_k) P[|S_{n-k}| < \varepsilon] \quad (\text{by the i.i.d. property})$$

$$= \sum_{n=0}^{r} P(A_n) \sum_{k=n}^{r} P[|S_{r-k}| < \varepsilon]$$

$$\leq \sum_{j=0}^{r} P[|S_j| < \varepsilon] \quad (\text{since the } A_k \text{ are disjoint}). \tag{39}$$

Summing for $m = -k$ to $k - 1$, we get

$$\sum_{n=0}^{r} \sum_{m=-k}^{k-1} P[S_n \in [m\varepsilon, (m+1)\varepsilon)] = \sum_{n=0}^{r} P[S_n \in [-k\varepsilon, k\varepsilon)] \leq 2k \sum_{j=0}^{r} P[|S_j| < \varepsilon].$$

This proves the inequality (36), and hence also the theorem.

It is now natural to investigate several other properties of recurrent random walks, such as the distribution of the first entrance time T_A of the process into a given Borel set $A \subset \mathbb{R}$, finding conditions on X in order that $E(T_A) < \infty$ or $= \infty$, and $P[T_A < \infty] = 1$. Conversely, the recurrence and transience of a random walk determines the structure of the range space $\mathbb{R}$ or $\mathbb{R}^n$ on a general locally compact group G. However, these questions need for their consideration certain analytic tools that we have not yet developed. In particular, a detailed study of characteristic functions and distribution functions is an essential first step, and this is undertaken in Chapter 4. It is then necessary to study further properties of sums of independent but not necessarily identically distributed random variables, continuing the work of Section 2. Here the most striking result, which we have not yet touched upon, is the law of the iterated logarithm. This is a strong limit theorem, based on the existence of two moments, but for its proof we also need the work on the central limit problem. Thus the results of this chapter are those obtainable only by means of the basic techniques. We need to continue expanding the subject. First a weakening of the concept of independence is needed. Then one proceeds to a study of the central limit problem and the (distributional or) weak limit laws.

Exercises

1. (a) Let (Ω, Σ, P) be a probability space with Ω having at least three points. If $X : \Omega \to \mathbb{R}$ is a random variable taking three or more distinct values, verify that $1, X, X^2$ are linearly independent (in the sense of linear algebra) but will be stochastically independent only if X is two valued and X^2 is a constant with probability 1, in which case $1, X, X^2$ are not linearly independent. Give an example satisfying the latter conditions. On the other hand, if X, Y are stochastically independent and not both are constant, then they are linearly independent, whenever $X \neq 0$ and $Y \neq 0$.

(b) Consider $\Omega = \{1,2,3,4,5\}$ with $P(\{i\}) = 1/5$ for $i = 1,2,3,4,5$. Is it possible to find events A, B of Ω so that A and B are independent? The answer to this simple and interesting problem is no. A probability space (Ω, Σ, P) is called a *"dependent probability space"* if there are no nontrivial independent events in Ω, (Ω, Σ, P) is called an *independent space* otherwise. R. Shiflett and H. Schultz (1979) introduced this concept where they studied both finite and countably infinite settings for Ω. Show that if Ω is finite so that $\Omega = \{1, 2, \ldots, n\}$ with $P(\{i\}) = 1/n$ for $i = 1, 2, \ldots, n$ then (Ω, Σ, P) is a dependent probability space if and only if n is prime. Additional results on finite dependent spaces with uniform probabilities can be found in the article by Shiflett and Schultz and in the work of Eisenberg and Ghosh (1987). Recently, W.F. Edwards (2004) investigated the case of the space (Ω, Σ, P) with $\Omega = \{1, 2, 3, \ldots\}$ and the measure P not uniform as follows. Show that if $\Omega = \{1, 2, 3, \ldots\}$ with $P(\{i\}) = p_i \geq p_{i+1} = P(\{i+1\})$ for all i and if

$$p_i \leq \sum_{k=1}^{\infty} p_{i+k}$$

then (Ω, Σ, P) is an independent space. The hypothesis in this last statement is sufficient but not necessary which can be seen by showing if $\Omega = \{1, 2, 3, \ldots\}$ with $p_i = (1-r)r^{i-1}$ for $0 < r < 1, i = 1, 2, \ldots$, then (Ω, Σ, P) is an independent space. These results give an idea of the interest that is associated with the question, "Are there necessary and sufficient conditions for a probability space (Ω, Σ, P) to be dependent?"

(c) One result without a restriction on the cardinality of Ω can be obtained by showing that (Ω, Σ, P) is an independent probability space if and only if there exists a partition of Ω into four nontrivial events A, B, C and D

for which $P(A)P(B) = P(C)P(D)$. [A related idea was considered by Chen, Rubin and Vitale (1997) who show that if the collection of pairwise independent events are identical for two measures, then the measures coincide. These are just some of the ideas associated with independent probability spaces. This type of inquiry can be continued with a serious investigation.]

2. Let $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ be an increasing continuous convex or concave function such that $\phi(0) = 0$, with $\phi(-x) = \phi(x)$, and in the convex case $\phi(2x) \leq c\phi(x)$, $x \geq 0$, $0 < c < \infty$. If $X_i : \omega \to \mathbb{R}$, $i = 1, 2$, are two random variables on (Ω, Σ, P) such that $E(\phi(X_i)) < \infty, i = 1, 2$, then verify that $E(\phi(X_1 + X_2)) < \infty$ and that the converse holds if X_1, X_2 are (stochastically) independent. [*Hint:* For the converse, it suffices to consider $|X_2| > n_0 > 1$. Thus

$$\begin{aligned} E(\phi(|X_1 + X_2|)) &\geq E(\phi(|X_1| - |X_2|) \geq E(\phi(|X_1| - n)\chi_{A_n})) \\ &= E(\phi(|X_1| - n))P(A_n) \end{aligned}$$

for an $A_n = [|X_2| \leq n], n_0 \geq n \geq 0$. Note that the converse becomes trivial if $X_i \geq 0$ instead of the independence condition.]

3. The preceding problem can be strengthened if the hypothesis there is strengthened. Thus let X_1, X_2 be independent and $E(X_1) = 0$. If now $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ there is restricted to a continuous convex function and $E(\phi(X_1 + X_2)) < \infty$, then $E(\phi(X_2)) \leq E(\phi(X_1 + X_2))$. If $E(X_2) = 0$ is also assumed, then $E(\phi(X_i)) \leq E(\phi(X_1 + X_2)), i = 1, 2$. [*Hint*: Use Jensen's inequality, and the fundamental law of probability, (Theorem 1.4.1) in $\phi(x) = \phi(E(X_2 + x)) \leq E(\phi(x + X_2))$ and integrate relative to $dF_{X_1}(x)$, then use Fubini's theorem.]

4. (a) Let $I = [0, 1]$, $\mathcal{B}$ = Borel σ-algebra of I, and P = Lebesgue measure on I. Let $X_1, \ldots, X_n$ be i.i.d. random variables on (Ω, Σ, P) with their common distribution $F(x) = P[X_1 < x] = x, 0 \leq x \leq 1$ (and $= 0$ for $x < 0$, $= 1$ for $x \geq 1$). Define $Y_1 = \min(X_1, ..., X_n)$, and if Y_i is defined, let $Y_{i+1} = \min\{X_k > Y_i : 1 \leq k \leq n\}$. Then (verify that) $Y_1 < Y_2 < \ldots < Y_n$ are random variables, called *order statistics* from the d.f. F, and are not independent. If $F_{Y_1,\ldots,Y_n}$ is their joint distribution, show that

$$F_{Y_1,\ldots,Y_n}(x_1, \ldots, x_n) = \begin{cases} n! \int \cdots \int_{0<x_1<\ldots<x_n<1} dt_1 \ldots dt_n \\ 0, & \text{otherwise} \end{cases}.$$

From this deduce that, for $0 \leq a < b \leq 1, i = 1, \ldots, n$,

$$P[a \leq Y_i < b] = \frac{n!}{(i-1)!(n-i)!} \int_a^b y^{i-1}(1-y)^{n-i} dy,$$

and that for $0 \leq a < b < c \leq 1, 1 \leq i < j \leq n$,

$$P[a \le Y_i < b < Y_j < c]$$
$$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!}$$
$$\times \int\int_{0\le a\le y_1\le b\le y_2\le c} y_1^{i-1}(y_2-y_1)^{j-i-1}(1-y_2)^{n-j}dy_2dy_1.$$

[Note that for $0 \le y_1 < y_2 < \ldots < y_n \le 1$, for small enough $\varepsilon > 0$ such that $[y_i, y_i+\varepsilon]$ are disjoint for $1 \le i \le n$, we have

$$P[y_i \le Y_i \le y_i + \varepsilon_i, 1 \le i \le n]$$
$$= \sum_{\substack{\text{all permutations} \\ (i_1,\ldots,i_n) \text{ of } (1,2,\ldots,n)}} P[y_j \le X_{i_j} \le y_j + \varepsilon_j, 1 \le j \le n],$$

where the X_i are i.i.d. for each permutation, and that there are $n!$ permutations.]

(b) Let $Z_1, \ldots, Z_n$ be i.i.d, random variables on (Ω, Σ, P) with their common distribution F on $\mathbb{R}$ continuous and strictly increasing. If $X_i = F(Z_i)$, $1 \le i \le n$, show that $X_1, \ldots, X_n$ are random variables satisfying the hypothesis of **(a)**. Deduce from the above that if Z_i is the ith-order statistic of $(Z_1, \ldots, Z_n)$, then

$$P[\tilde{Z}_i < x] = \frac{n!}{(i-1)!(n-i)!}\int_0^{F(x)} y^{i-1}(1-y)^{n-j}dy.$$

Similarly, obtain the corresponding formulas of **(a)** for the $\tilde{Z}_i$-sequence.

5. (a) Following Corollary 1.8 we have discussed the adjunction procedure. Let $X_1, X_2, \ldots,$ be any sequence of random variables on (Ω, Σ, P). Let $F_i(x) = P[X_i < x], i = 1, 2, \ldots$. Then using the same procedure, show that there is another probability space $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{P})$ and a mutually independent sequence of random variables $Y_1, Y_2, \ldots$ on it such that $P[Y_n < x] = F_n(x), x \in \mathbb{R}, n \ge 1$. [*Hint:* Since F_n is a d.f., let $\mu_n(A) = \int_A dF_n(x), A \subset \mathbb{R}$ Borel, $X_n =$ identity on $\mathbb{R}$. Then $(\mathbb{R}, \mathcal{B}, \mu_n)$ is a probability space and $\tilde{X}_n$ is an r.v. with F_n as its d.f. Consider, with the Fubini-Jessen theorem, the product probability space $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{P}) = \bigotimes_{n\ge1}(\mathbb{R}_n, \mathcal{B}_n, \mu_n)$, where $\mathbb{R}_n = \mathbb{R}$, $\mathcal{B}_n = \mathcal{B}$. If $\tilde{\omega} = (x_1, x_2, \ldots) \in \tilde{\Omega} = \mathbb{R}^{\mathbb{N}}$, let $Y_n(\tilde{\omega}) = n$th coordinate of $\tilde{\omega}$ $[= x_n = \tilde{X}_n(\tilde{\omega})]$. Note that the Y_n are independent random variables on $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{P})$ and $\tilde{P}[\tilde{Y}_n < x] = \mu_n[\tilde{X}_n < x] = F_n(x), x \in \mathbb{R}, n \ge 1$.]

(b) (Skorokhod) With a somewhat different specialization, we can make, the following assertion: Let $X_1, X_2 \ldots$ be a sequence of random variables on (Ω, Σ, P) which converge in distribution to an r.v. X. Then there is another probability space (Ω', Σ', P') and random variables $Y_1, Y_2, \ldots$ on it such that $Y_n \to Y$ a.e. and $P[X_n < x] = P'[Y_n < x], x \in \mathbb{R}$,for $n \ge 1$. Thus X_n, Y_n

have the same distributions and the (stronger) pointwise convergence is true for the Y_n-sequence. (Compare this with Proposition 2.2.) [*Sketch of proof:* Let $F_n(x) = P[X_n < x], F(x) = P[X < x], x \in \mathbb{R}, n \geq 1$. If Y_n, Y are inverses to F_n, F, then $Y_n(x) = \inf\{y \in \mathbb{R} : F_n(y) > x\}$; and similarly for Y. Clearly Y_n, Y are Borel functions on $(0,1) \to \overline{\mathbb{R}}$. Since $Y_n(x) < y$ iff $F_n(y) > x$, we have, on letting $\Omega' = (0,1)$, $\Sigma' =$ Borel σ-algebra of Ω', with P' as the Lebesgue measure, $P'[Y_n < y] = P'[x : x < F_n(y)] = F_n(y)$; and similarly for $P'[Y < y] = F(y)$. Since $F_n(x) \to F(x)$ at all continuity points of F, let x be a continuity point of F. If the F_n are strictly increasing, then $Y_n = F_n^{-1}$ and the result is immediate. In the general case, follow the argument of Proposition 2.2, by showing that for $a < b \leq c < d$,

$$Y(a) \leq \liminf_n Y_n(b) \leq \limsup_n Y_n(b) \leq \limsup_n Y_n(c) \leq Y(d),$$

and then setting $b = c$, a continuity point of F; let $a \uparrow b$ and $d \downarrow c$, so that $Y_n(c) \to Y(c)$. Since the discontinuities of F are countable and form a set of P' measure zero, the assertion follows. *Warning:* In this setup the Y_n will not be independent if Y is nonconstant (or X is nonconstant).]

(c) The following well-known construction shows that the preceding part is an illustration of an important aspect of our subject. Let (Ω, Σ) be a measurable space and $B_i \in \Sigma$ be a family of sets indexed by $D \subset \mathbb{R}$ such that for $i, j \in D, i < j \Rightarrow B_i \subset B_j$. Then there exists a unique random variable $X : \Omega \to \mathbb{R}$ such that $\{\omega : X(\omega) \leq i\} \subset B_i$ and $\{\omega : X(\omega) > i\} \subset B_i^c$. [Verify this by defining $X(\omega) = \inf\{i \in D : \omega \in B_i\}$ and that X is measurable for Σ.] If $P : \Sigma \to \mathbb{R}^+$ is a probability and D is countable, $\{B_i, i \in D\}$ is increasing P a.e. (i.e., for $i < j$, $P(B_i - B_j) = 0$), then the variable X above satisfies $\{\omega : X(\omega) \leq i\} = B_i$, a.e. and $\{\omega : X(\omega) \geq i\} = B_i^c$, $i \in D$. (See e.g., Royden (1968, 1988), 11.2.10.) Suppose that there is a collection of such families $\{B_i^n, i \in D = \mathbb{R}, n \geq 1\} \subset \Sigma$. Let X_n be the corresponding random variable constructed for each n, and let $F_n(x) = P(B_x^n)$ where $-\infty < x < \infty$. Show that $F_n = P \circ X_n^{-1}$, determined by the collection, and that for $n_1, \ldots, n_m, x_i \in \mathbb{R}, m \geq 1$ one has

$$P(B_{x_1}^{n_1} \cap \cdots \cap B_{x_m}^{n_m}) = F_{n_1,\ldots,n_m}(x_1, \ldots, x_m)$$

defines an m-dimensional (joint) distribution of $(X_{n_1}, \ldots, X_{n_m})$ so constructed. [This construction of distributions will play a key role in establishing a general family of random variables, or processes, later (cf., Theorem 3.4.10).

(d) Here is a concrete generation of independent families of random variables already employed by N. Wiener (cf. Paley and Wiener (1934), p. 143), and emphasized by P. Lévy ((1953), Sec. 2.3). It also shows where the probabilistic concept enters the construction. Let $Y_1, \ldots, Y_n$ be functions on (0,1) each represented by its decimal expansion

$$Y_n = \sum_{\nu=1}^{\infty} \frac{a_{n,\nu}}{10^\nu},$$

$a_{n,\nu}$ taking values $0, 1, \ldots, 9$ each with probability $\frac{1}{10}$, independent of one another. (This is where probability enters!) Then each Y_n is uniformly distributed and they are mutually independent. (Clearly binary or ternary etc. expansions can be used in lieu of decimal expansion. Unfortunately, no recipe exists for choosing $a_{n,\nu}$ here. A similar frustration was (reportedly) expressed by A. Einstein regarding his inability to find a recipe for a particular Brownian particle to be in a prescribed region, but only a probability of the event can be given. [cf., Science, **30** (2005), pp. 865-890, special issue on Einstein's legacy].) If $\{F_n, n \geq 1\}$ is a sequence of distribution functions on $\mathbb{R}$, let F_n^{-1} be the generalized inverse of F, as defined (in part (b)) above. Let $X_n = F_n^{-1}(Y_n), n \geq 1$. Then $\{X_n, n \geq 1\}$ is a sequence (of mutually independent) random variables with distributions F_n. [It is even possible to take a single uniformly distributed random variable Y by reordering $a_{n,\nu}$ into a single sequence $\{b_k, k \geq 1\}$ so that $Y = \sum_{k=1}^{\infty} \frac{b_k}{10^k}$, by excluding the terminating decimal expansions which are countable and hence constitute a set of (Lebesgue) measure zero, and then $X_n = F_n^{-1}(Y), n \geq 1$.] It should be observed that in the representation of X_n as a mapping of $(Y_1, \ldots, Y_n)$ [or of Y] by I_n which is one-to-one, there are infinitely many representations, while a unique distribution obtains if it is nondecreasing, such as F_n^{-1}. This fact is of interest in applications such as those implied in part (b) above.

The following example is considered by Wiener (in the book cited above, p. 146). Let Y_1, Y_2 be independent uniformly distributed random variables on (0,1) and define $R = (-\log Y_1)^{\frac{1}{2}}$, and $\theta = 2\pi Y_2$ and let $X_1 = R\cos\theta, X_2 = R\sin\theta$. Then the Jacobian is easily computed, and one has $dy_1 dy_2 = \frac{1}{\pi}e^{-(x_1^2+x_2^2)}$ $dx_1 dx_2$ so that X_1, X_2 are independent normal random variables generated by Y_1, Y_2. Extending this procedure establish the following n-dimensional version. Let $Y_1, \ldots, Y_n$ be independent uniformly distributed random variables on (0,1), $\theta_k = 2\pi Y_{k+1}$ and $X_1 = R\sin\theta_{n-1}\ldots\sin\theta_2\sin\theta_1; X_2 = R\sin\theta_{n-1}\ldots\sin\theta_2\cos\theta_1, \ldots, X_{n-1} = R\sin\theta_{n-1}\cos\theta_{n-2}$, and $X_n = R\cos\theta_{n-1}$ where $R = (-2\log\, Y_1)^{\frac{1}{2}}$. The Jacobian is much more difficult, [use induction], but is nonvanishing, giving a one-to-one mapping. (With $R = 1$, the transformation has Jacobian to be $(-1)^n(\sin\theta_1)^n(\sin\theta_2)^n \ldots \sin\theta_{n-1}\cos\theta_n$ so that it is 1-1 between the open unit n-ball and the open rectangle $0 < \theta_i < \pi, i = 1, \ldots n$.) This shows that the Φ_n sequence (different from the F_n) can be somewhat involved, but the procedure is quite general as noted by N. Wiener whose use in a construction of Brownian motion is now legendary, and was emphasized by P. Lévy later. [In the last chapter we again consider the Brownian motion construction with a more recent and (hopefully) simpler method.]

6. (a) (Jessen-Wintner) If $\{X_n, n \geq 1\}$ is a sequence of independent countably valued random values on (Ω, Σ, P) such that $S_n = \sum_{k=1}^{n} X_k \to S$ a.e., then the distribution of S on $\mathbb{R}$ is either (i) absolutely continuous or singular relative to the Lebesgue measure or (ii) $P[S = j] > 0$ for a countable set of points $j \in \mathbb{R}$, and no mixed types can occur. [*Hints*: Let $G \subset \mathbb{R}$ be the group generated by the ranges of the X_n, so that G is countable. Note that for any Borel set B, the vector sum $G + B = \{x + y : x \in G, y \in B\}$ is again Borel. If $\Omega_0 = \{\omega : S_n(\omega) \to S(\omega)\}$, then let $A = \{\omega : S(\omega) \in (G + B) \cap \Omega_0\}$, and verify that A is a tail event, so that $P(A) = 0$ or 1 by Theorem 1.7. Indeed, if $g_1 - g_2 \in G$, then $g_1 \in G + B$ for some Borel set B iff $g_2 \in G + B$. Now if $S_n = S - (S - S_n) \in G$, then $S - S_n \in G + B$, and conversely. But $S - S_n \in \Omega_0$. Hence $A = [S - S_n \in G + B] \cap \Omega_0$, so that A is a tail event, and $P(A) = 0$ or 1. This implies either S is countably valued or else, since $P(\Omega_0) = 1$, $P[S \in G+B] = 0$ for each countable B. In this case $P[S \in B] = 0$ for each countable B, so that S has a continuous distribution, with range non-countable. Consequently, either the distribution of S is singular relative to the Lebesgue measure, or it satisfies $P[S \in G + B] = 0$ for all Borel B of zero Lebesgue measure. Since G is countable, this again implies $P[S \in G+B] = 0$, so that $P[S \in B] = 0$ for all Lebesgue null sets. This means the distribution of S is absolutely continuous. To see what type is the distribution of S, we have to exclude the other two cases, and no recipe is provided in this result. In fact this is the last result of Jessen-Wintner's long paper (1935).]

(b) To decide on the types above, we need to resort to other tricks, and some will be noted here. Let $\{X_n, n \geq 1\}$ be i.i.d. random variables with

$$P[X_1 = +1] = \frac{1}{2} = P[X_1 = -1].$$

Let $S_n = \sum_{k=1}^{n} X_k/2^k$. Then $S_n \to S$ a.e. (by Theorem 2.6). Also $|S| \leq 1$ a.e. Prove that the S distribution is related to that of $U - V$, where U and V are independent random variables on the Lebesgue unit interval $[0, 1]$, with the uniform distribution F, i.e., $F(x) = 0$ if $x \leq 0, = x$ if $0 < x \leq 1$, and $F(x) = 1$ for $x > 1$, and hence has an absolutely continuous distribution. [*Hints*: Note that if F_U, F_V are the distributions of U, V, then F_{U+V} can be obtained by the image law (cf. Theorem 1.4.1) as a convolution:

$$\begin{aligned} F_{U+V}(x) = P[U + V < x] &= \int_{\Omega} \chi_{[U+V<x]} dP \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \chi_{[\lambda_1+\lambda_2<x]} F_U(d\lambda_1) F_V(d\lambda_2) \\ &\quad \text{(since } F_{U,V} = F_U \cdot F_V \text{ by independence)} \\ &= \int_{\mathbb{R}} F_U(d\lambda_1) F_V(x - \lambda_1). \end{aligned}$$

Thus F_{U+V} is continuous if at least one of F_U, F_V is continuous. Next verify that if $x = \sum_{k=1}^{\infty} \varepsilon_k/2^k$, where $\varepsilon_k = 0, 1$ is the dyadic expansion of $0 < x < 1$, then (as in the construction of Problem 5 (d) above)

$$\mu\{x : \varepsilon_k(x) = 0\} = \frac{1}{2} = \mu\{x : \varepsilon_k(x) = 1\}$$

with μ as the Lebesgue measure. Deduce that U has the same distribution as the identity mapping $I : (0,1) \to (0,1)$ with Lebesgue measure.](Explicit calculation with ch.f. is easier and will be noted in Exercise 4.11.)

(c) By similar indirect arguments verify the following: (i) If $\{X_n, n \geq 1\}$ is as above, then $S_n = \sum_{k=1}^{n} X_k/3^k \to S$ a.e. and S has a singular distribution. (ii) (P. Lévy) If $Y_n, n = 1, 2, \ldots$, are independent with values in a countable set $C \subset \mathbb{R}$, and if there is a convergent set of numbers $c_n \in C$ such that

$$\sum_{n=1}^{\infty} P[Y_n \in C - c_n] < \infty,$$

then $S = \sum_{k=1}^{\infty} Y_k$ exists a.e., and S takes only countably many values with positive probability.

(d) The proofs of Theorems 2.6 and 2.7 used the Kronecker lemma and the $(c,1)$-summability. Thus the Kolomogorov SLLN (Theorem 2.7) can be considered as a probabilistic analog of the classical $(c,1)$-summability in the sense that a sequence $\{X_n, n \geq 1\}$ of i.i.d. r.v.s on (Ω, Σ, P) obeys the $(c,1)$-pointwise a.e. iff $E(X_1) = \mu \in \mathbb{R}$ exists. Since classical analysis shows that $(c,1)$-summability implies (c,p)-summability for $p \geq 1$, one can expect a similar result for i.i.d sequences. In fact the following precise version holds. Let $p, \mu \in \mathbb{R}$, $p \geq 1$. Verify the following equivalences for i.i.d. r.v.s:

(i) $\{X_n, n \geq 1\}$ obeys the SLLN,

(ii) $E(X_1) = \mu$,

(iii) $\{X_n, n \geq 1\}$ obeys $(c,1)$-summability a.e. with limit μ,

(iv) $\{X_n, n \geq 1\}$ obeys (c,p)-summability a.e. with limit μ,

$$\text{i.e.,}\ \lim_{n\to\infty} \frac{1}{\binom{n+p}{n}} \sum_{k=0}^{n-1} \binom{k+p-1}{k} X_{n-k} = \mu \text{ a.e.,}$$

(v) $\{X_n, n \geq 1\}$ obeys Abel mean a.e. with value μ,

$$\text{i.e.,}\ \lim_{0 \leq \lambda \uparrow 1} (1-\lambda) \sum_{i=1}^{\infty} \lambda^i X_i = \mu \text{ a.e..}$$

[*Hints*: The classical theories on summability imply that (i) $\Rightarrow$ (iii) $\Rightarrow$ (iv) $\Rightarrow$ (v) and Theorem 2.7 gives (i) $\Leftrightarrow$ (ii). So it suffices to show (v) $\Rightarrow$ (ii). For ordinary sequences of reals, Abel convergence does not imply even $(c,1)$-convergence. (Here the converse holds if the sequence is bounded in addition, as shown by J.E. Littlewood.) But the i.i.d. hypothesis implies the converse a.e. as follows. Using the method of Theorem 2.9, called symmetrization, let $X_n^s = X_n - X_n'$ where X_n and X_n' are i.i.d. (one may use enlargement of the basic probability space as in the proof of 2.9, where X_n^s is denoted as Z_n there), and (v) can be expressed if $1 - \lambda = 1/m, m \geq 1$ as

$$\lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{\infty} (1 - \frac{1}{m})^i X_i^s = \mu - \mu = 0,$$

or alternately

$$\lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{\infty} e^{i \log(1-1/m)} X_i^s = \lim_{m\to\infty} \frac{1}{m} \sum_{j=1}^{\infty} e^{-j/m} X_j^s = 0 \text{ a.e.}$$

Let

$$Y_m = \frac{1}{m} \sum_{j=1}^{m} e^{-j/m} X_j^s, \quad Z_m = \frac{1}{m} \sum_{j=m+1}^{\infty} e^{-j/m} X_j^s.$$

Then $Y_m + Z_m \to 0$ a.e. as $m \to \infty$, and Y_m, Z_m are independent. Verify that for each $\varepsilon > 0$, $P[|Z_m| \geq \varepsilon] \to 0$ as $m \to \infty$. Then using Slutzky's Theorem and stochastic calculus (Problems 9(b) and 11(c) below) suitably conclude that $Y_m \to 0$. Next $\overline{Y}_m = Y_m - \frac{1}{em} X_m^2 \to 0$ and finally that $X_m^s/m \to 0$ also as $m \to \infty$. [This needs some more work!] Then by the Borel-Cantelli lemma, deduce that $E(|X_1|) < \infty$, as in the proof of Theorem 3.7. Hence SLLN holds. Thus the equivalence follows. The above sketch is a paraphrase of T. L. Lai (1974). Can we replace mutual independence here by pairwise independence, as in Corollary 3.3 if we only ask for WLLN?]

7. This problem illustrates the strengths and limitations of our a.e. convergence statements. Let (Ω, Σ, P) be the Lebesgue unit interval, so that $\Omega = (0,1)$ and $P =$ Lebesgue measure on the completed Borel σ-algebra Σ. If $\omega \in \Omega$, expand this in decimals: $\omega = 0.x_1x_2\ldots$ so that if $X_n(\omega) = x_n$, then $X_n : \Omega \to \{0, 1, \ldots, 9\}$ is a r.v. Verify that $\{X_n, n \geq 1\}$ is an i.i.d. sequence with the common distribution F, given by $F(y) = (k+1)/10$, for $k \leq y < k+1$, $k = 0, 1, \ldots, 9$; $= 0$ if $y < 0$; $= 1$ for $y > 9$. Let $\delta_k(\cdot)$ be the Dirac delta function, and consider $\delta_k(X_n)$. Then $P[\delta_k(X_n) = 1] = 1/10, P[\delta_k(X_n) = 0] = 9/10$, and $\delta_k(X_n), n \geq 1$, are i.i.d., for each $k = 0, 1, \ldots, 9$. If $k_1, k_2, \ldots, k_r$ are a fixed r-tuple of integers such that $0 \leq k_i \leq 9$, define (cf. Problem 5 (d) also)

$$\varepsilon_{n,r} = \delta_{k_1}(X_{rn})\delta_{k_2}(X_{rn+1}) \ldots \delta_{k_r}(X_{rn+r-1}).$$

Show that the $\varepsilon_{n,r}, n \geq 1$, are bounded uncorrelated random variables for which we have $(1/m)\sum_{n=1}^{m} \varepsilon_{n,r} \to 1/10^r$ a.e. as $m \to \infty$ (apply Theorem 3.4), $r = 1, 2, \ldots$. This means for a.a. $\omega \in \Omega$, the ordered set of numbers $(k_1, \ldots, k_r)$ appears in the decimal expansion of ω with the asymptotic relative frequency of $1/10^r$. Every number $\omega \in \Omega$ for which this holds is called a *normal number*. It follows that $\sum_{n=1}^{m} \varepsilon_{n,r} \to \infty$ as $m \to \infty$ for a.a.(ω) (as in the proof of Theorem 4.4); thus $\varepsilon_{n,r} = 1$ infinitely often, which means that the given set $(k_1, \ldots, k_r)$ in the same order occurs infinitely often in the expansion of each normal number, and that almost all $\omega \in \Omega$ are normal. [This fact was established by E. Borel in 1909.] However, there is no known recipe to find which numbers in Ω are normal. Since the transcendental $(\pi - e) \in (0,1)$, it is not known whether $\pi - e$ is normal; otherwise it would have settled the old question of H. Weyl: Is it true or false that in the decimal expansion of the irrational number π, the integers $0, 1, \ldots, 9$ occur somewhere in their natural order? This question was raised in the 1920's to counter the assertion of the logicians of Hilbert's school asserting that every statement is either "true" or "false," i.e., has only two truth values. As of now we do not know the definitive answer to Weyl's question, even though π has been expanded to over 10^5 decimal places and the above sequence still did not appear! [See D. Shanks and J. W. Wrench, Jr. (1962). *Math. Computation* 16, 76-89, for such an expansion of π. On the other hand, it is known that 0.1234567891011121314151617..., using all the natural numbers, is normal. Recently two Japanese computer scientists seem to have shown that the answer is 'yes' after expanding π for several billions of decimal places. See, e.g. J.M. Borwain (1998), *Math. Intelligencer*, **20**, 14-15.]

8. The WLLN of Theorem 3.2 does not hold if (even) the symmetric moment does not exist. To see this, we present the classical *St. Petersburg game*, called a "paradox," since people applied the WLLN without satisfying its hypothesis. Let X be an r.v. such that

$$P[X = 2^n] = \frac{1}{2^n}, \quad \text{for } n \geq 1,$$

on (Ω, Σ, P). Let $\{X_n, n \geq 1\}$ be i.i.d. random variables with the distribution of X. If $S_n = \sum_{k=1}^{n} X_k$, show that $S_n/n \not\to \alpha$, as $n \to \infty$, for any $\alpha \in \mathbb{R}$, either in probability or a.e. for any subsequence. (Use the last part of Theorem 3.7.) The game interpretation is that a player tosses a fair coin until the head shows up. If this happens on the nth toss, the player gets 2^n dollars. If any fixed entrance fee per game is charged, the player ultimately wins and the house is ruined. Thus the "fair" fee will have to be "infinite," and this is the paradox! Show however, by the truncation argument, that $S_n/(n \log_2 n) \xrightarrow{P} 2$ as $n \to \infty$, where $\log_2 n$ is the logarithm of n to base 2. If the denominator is replaced by $h(n)$ so that $(n \log_2 n)/h(n) \to 0$, then $S_n/h(n) \xrightarrow{P} 0$ and a.e. In fact show that for any sequence of random variables $\{Y_n, n \geq 1\}$ there exists an increasing sequence k_n such that $P[|Y_n| > k_n, i.o.] = 0$, so that $Y_n/k_n \to 0$ a.e. Thus $n \log_2 n$ is the correct "normalization" for the St. Petersburg game. (An interesting and elementary variation of the St. Petersburg game can be found in D.K. Neal, & R.J. Swift, (1999) *Missouri J. Math. Sciences*, **11**, No. 2, 93-102.)

9. (Mann-Wald). A calculus of "in probability" will be presented here. (Except for the sums, most of the other assertions do not hold on infinite measure spaces!) Let $\{X_n, n \geq 1\}$ and $\{Y_n, n \geq 1\}$ be two sequences of random variables on (Ω, Σ, P). Then we have the following, in which ***no*** assumption of independence appears:

(a) $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y \Rightarrow X_n \pm Y_n \xrightarrow{P} X \pm Y$, and $X_nY_n \xrightarrow{P} XY$.

(b) If $f : \mathbb{R}^2 \to \mathbb{R}$ is a Borel function such that the set of discontinuities of f is measurable and is of measure zero relative to the Stieltjes measure determined by the d.f. $F_{X,Y}$ of the limit vector (X, Y) of **(a)**, then $f(X_n, Y_n) \xrightarrow{D} (X, Y)$ under either of the conditions: (i) $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$ or (ii) $\alpha X_n + \beta Y_n \xrightarrow{P} \alpha X + \beta Y$ for all real α, β. If f is continuous, then strengthen this to the assertion that $f(X_n, Y_n) \xrightarrow{P} f(X, Y)$ if condition (i) holds. [Hint: For (ii), use Problem 5(b) and the fact that $(X_n, Y_n) \xrightarrow{D} (X, Y)$ iff $\alpha X_n + \beta Y_n \xrightarrow{D} \alpha X + \beta Y$ for all real α, β.]

10. Suppose that for a sequence $\{X_n, n \geq 1, X\}$ in $L^1(P)$ we have $X_n \xrightarrow{D} X$. Show it is true that $E(|X|) \leq \liminf_n E(|X_n|)$, and if, *further*, the set is uniformly integrable, then $E(X) = \lim_n E(X_n)$. [*Hint*: Use Problem 5 (b) and the image probability Theorem 1.4.1. This strengthening of Vitali's convergence theorem (and Fatou's lemma) is a nontrivial contribution of Probability Theory to Real Analysis!]

11. **(a)** If X is an r.v. on (Ω, Σ, P), then $\mu(X)$, called a *median* of the distribution of X, is any number which satisfies the inequalities

$$P[X \leq \mu(X)] \geq \frac{1}{2}, \quad P[X \geq \mu(X)] \geq \frac{1}{2}.$$

Note that a median of X always exists [let $\mu(X) = \inf\{\alpha \in \mathbb{R} : P[X \leq \alpha] \geq \frac{1}{2}\}$ and verify that $\mu(X)$ is a median and $\mu(aX + b) = a\mu(X) + b$, for $a, b \in \mathbb{R}$]. If $X_n \xrightarrow{D} a_0$, $a_0 \in \mathbb{R}$, show that $\mu(X_n) \to a_0$

(b) A sequence $\{X_n, n \geq 1\}$ of random variables is *bounded in probability* if for each $\varepsilon > 0$ there is an $n_0[= n_0(\varepsilon)]$ and a constant $M_0[= M_0(\varepsilon)] > 0$ such that $P[|X_n| \geq M_\varepsilon] \leq \varepsilon$ for all $n \geq n_0$. Show that if $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{P} 0$ are two sequences of random variables, then $X_nY_n \xrightarrow{P} 0$, $X_n + Y_n \xrightarrow{D} X$, as $n \to \infty$ and $\{X_n, n \geq 1\}$ is bounded in probability. If $\{X_n, n \geq 1\}$ has the latter property and $Y_n \xrightarrow{P} 0$, then $X_nY_n \xrightarrow{P} 0$ as $n \to \infty$.

(c) (Cramér-Slutsky) Let $X_n \xrightarrow{D} X$, $Y_n \xrightarrow{D} a$, where $a \in \mathbb{R}$ and $n \to \infty$. Then $X_nY_n \xrightarrow{D} aX$, and if $a \neq 0$, $X_n/Y_n \xrightarrow{D} X/a$, so that the distributions of aX and X/a are $F(x/a)$ and $F(ax)$ for $a > 0$, $1 - F(x/a)$ and $1 - F(ax)$ for $a < 0$. Here again the sequences $\{X_n\}$ and $\{Y_n\}$ need not be independent.

(d) Let $\{X_n, n \geq 1\}$ and $\{Y_n, n \geq 1\}$ be two sequences of random variables on (Ω, Σ, P) and $\alpha_n \downarrow 0$, $\beta_n \downarrow 0$ be numbers, such that $(X_n - a)/\alpha_n \xrightarrow{D} X$ and $(Y_n - b)/\beta_n \xrightarrow{D} Y$, where $a, b \in \mathbb{R}$, $b \neq 0$. Show that $(X_n - a)/\alpha_nY_n \xrightarrow{D} X/b$. All limits are taken as $n \to \infty$.

12. (Kolmogorov). Using the method of proof of Theorem 2.7, show that if $\{X_n, n \geq 1\}$ is an independent sequence of bounded random variables on

(Ω, Σ, P), common bound M and means zero, then for any $d > 0$ we have, with $S_n = \sum_{k=1}^n X_k$,

$$P\left[\max_{1\le k\le n} |S_k| \le d\right] \le \frac{(2M+d)^2}{Var(S_n)}.$$

Deduce that if $Var(S_n) \to \infty$, then for each $d > 0$, $P[|S_n| \le d] \to 0$ as $n \to \infty$.

13. (Ottaviani). Let $\{X_n, n \ge 1\}$ be independent random variables on (Ω, Σ, P) and $\varepsilon > 0$ be given. If $S_n = \sum_{k=1}^n X_k$, $P[|X_k + \ldots + X_n| \le \varepsilon] \ge \eta > 0$, $1 \le k \le n$, show that

$$P\left[\max_{1\le k\le n} |S_k| \ge \varepsilon\right] \le \frac{1}{\eta} P[|S_n| \ge \frac{\varepsilon}{2}].$$

[Note that if $A_1 = [|S_1| \ge \varepsilon]$, and for $k > 1$, $A_k = [|S_k| \ge \varepsilon, |S_j| < \varepsilon, 1 \le j \le k-1]$, then $[|S_n| \ge \varepsilon/2] \supset \bigcup_k (A_k \cap [|X_{k+1} + \ldots + X_n| \le \varepsilon/2])$. The decomposition of $[\max_{k\le n} |S_k| \ge \varepsilon]$ is analogous to that used for the proof of Theorem 2.5.]

14. We present two extensions of Kolmogorov's inequality for applications. (a) Let $X_1, \ldots, X_n$ be independent random variables on (Ω, Σ, P) with means zero and variances $\sigma_1^2, \ldots, \sigma_n^2$. Then the following improved one-sided inequality [similar to that of Čebyšev's; this improvement in 1960 is due to A. W. Marshall] holds: for $\varepsilon > 0$, and $S_k = \sum_{i=1}^k X_i$, one has

$$P\left[\max_{1\le k\le n} S_k \ge \varepsilon\right] \le \frac{\sum_{k=1}^n \sigma_k^2}{\varepsilon^2 + \sum_{k=1}^n \sigma_k^2} \le \frac{1}{\varepsilon^2}\sum_{k=1}^n \sigma_k^2.$$

[*Hint*: Consider $f : \mathbb{R}^n \to \mathbb{R}$ defined by $f(x_1, \ldots x_n) = [\sum_{i=1}^n (\varepsilon x_i + \sigma_i^2)/(\varepsilon^2 + \sum_{i=1}^n \sigma_i^2)]^2$, and evaluate $E(f(X_1, \ldots, X_n))$ with the same decomposition as in Theorem 2.5. If $n = 1$, this reduces to Problem 6 (a) of Chapter 1.]

(b) Let $\{X_n, n \ge 1\}$ be independent random variables on (Ω, Σ, P) as above, with zero means and $\{\sigma_n^2, n \ge 1\}$ as respective variances. If $\varepsilon > 0$, $S_n = \sum_{k=1}^n X_k$, and $a_1 \ge a_2 \ge \ldots \to 0$, show that with simple modifications of the proof of Theorem 2.5,

$$P\left[\sup_{n\ge 1} a_n^{\frac{1}{2}} |S_n| > \varepsilon\right] \le \frac{1}{\varepsilon^2} \sum_{n\ge 1} (a_n - a_{n+1}) \sum_{k=1}^n \sigma_k^2.$$

[This inequality was noted by J. Hájek and A. Rényi.]

(c) If in (b) we take $a_k = (n_0 + k - 1)^{-2}$ for any fixed but arbitrary $n_0 \ge 1$, deduce that

$$P\left[\sup_{n\ge n_0} \frac{|S_n|}{n} > \varepsilon\right] \le \frac{1}{\varepsilon^2}\left(\frac{1}{n_0^2}\sum_{i=1}^{n_0} \sigma_i^2 + \sum_{n\ge n_0+1} \frac{\sigma_n^2}{n^2}\right).$$

Hence, if $\sum_{n\geq 1}(\sigma_n^2/n^2) < \infty$, conclude that the sequence $\{X_n, n \geq 1\}$ obeys the SLLN. (Thus we need not use Kronecker's lemma.)

15. In some problems of classical analysis, the demonstration is facilitated by a suitable application of certain probabilistic ideas and results. This was long known in proving the Weierstrass approximation of a continuous function by Bernstein polynomials. Several other results were noted by K. L. Chung for analogous probabilistic proofs. The following is one such: an inversion formula for Laplace transforms. Let $X_1(\lambda), \ldots, X_n(\lambda)$ be i.i.d. random variables on (Ω, Σ, P), depending on a parameter $\lambda > 0$, whose common d.f. F is given by $F(x) = 0$ if $x < 0$; and $= \lambda \int_0^x e^{-\lambda t} dt$ if $x \geq 0$. If $S_n(\lambda) = \sum_{k=1}^n X_k(\lambda)$, using the hints given for Problem 6(b) show that the d.f. of $S_n(\lambda)$ is F_n, where $F_n(x) = 0$ for $x < 0$, and $= [\lambda^n/(n-1)!] \int_0^x t^{n-1} e^{-\lambda t} dt$ for $x \geq 0$. Deduce that $E(S_n(\lambda)) = n/\lambda$, $Var S_n(\lambda) = n/\lambda^2$, so that $S_n(n/x) \xrightarrow{P} x$ as $n \to \infty$. Using the fundamental law of probability, verify that for any bounded continuous mapping $f : \mathbb{R}^+ \to \mathbb{R}^+$, or f Borel satisfying $E(f(S_n))^2 < k_0 < \infty$ (cf., also Proposition 4.1.3 later) then $E(f(S_n)) \to E(f(x)) = f(x)$, by uniform integrability, (use Scheffè's lemma, Proposition 1.4.6), where

$$E(f(S_n(\lambda))) = \frac{(-\lambda)^{n-1}}{(n-1)!} \lambda \int_0^\infty \frac{d^{n-1}}{d\lambda^{n-1}} (e^{-\lambda t} f(t)) dt.$$

Hence prove, using Problem 6(b), that for any continuous $f \in L^2(\mathbb{R}^+)$ if $\hat{f}$ is the Laplace transform of f $[\hat{f}(u) = \int_0^\infty e^{-ut} f(t) dt, u > 0]$ one has the inversion

$$f(t) = \lim_{n\to\infty} \left(\frac{-n}{t}\right)^{n-1} \cdot \frac{1}{(n-1)!} \frac{n}{t} \cdot \left(\frac{d^{n-1}\hat{f}}{d\lambda^{n-1}}\right) \left(\frac{n}{\lambda}\right) |_{\lambda=t} \ ,$$

the limit existing uniformly on compact intervals of $\mathbb{R}^+$. [Actually f can be in any $L^p(\mathbb{R}^+)$, $1 < p < \infty$, not just $p = 2$. The distribution of X_1 above is called the *exponential*, and that of $S_n(\lambda)$, the *gamma* with parameters (n, λ). More d.f.s are discussed in Section 4.2 later.] The result above is the *classical Post-Widder formula.*

16. Let $\{X_n, n \geq 1\}$ be independent random variables on (Ω, Σ, P) and $S_n = \sum_{k=1}^n X_k$. Then $S_n \to S$ a.e. iff $S_n \xrightarrow{P} S$. This result is due to P. Lévy. (We shall prove later that $S_n \xrightarrow{P} S$ can be replaced here by $S_n \xrightarrow{D} S$, but more tools are needed for it.) [*Hints*: In view of Proposition 2.2, it suffices to prove the converse. Now $S_n - S \xrightarrow{P} 0 \Rightarrow \{S_n, n \geq 1\}$ is Cauchy in probability, so for $1 > \varepsilon > 0$, there is an $n_0[= n_0(\varepsilon)]$ such that $m, n > n_0 \Rightarrow P[|S_n - S_m| > \varepsilon] < \varepsilon$. Thus $P[|S_k - S_m| \geq \varepsilon] > 1 - \varepsilon$ for all $m < k \leq n$. Hence by Problem 13 applied to the set $\{X_j, j \geq m \geq n_0\}$, we get

$$P[\max_{m<k\leq n} |S_k - S_m| \geq 2\varepsilon] \leq \frac{1}{1-\varepsilon} P[|S_n - S_m| \geq \varepsilon] < \frac{\varepsilon}{1-\varepsilon}.$$

This implies upon first letting $n \to \infty$, and then letting $m \to \infty$, since the $0 < \varepsilon < 1$ is arbitrary, that $\{S_k, k \geq 1\}$ is pointwise Cauchy and hence converges a.e.]

17.(P. Lévy Inequalities). Let $X_1, \ldots, X_n$ be independent random variables on (Ω, Σ, P) and $S_j = \sum_{k=1}^{j} X_k$. If $\sigma(X)$ denotes a median (cf. Problem 11) of X, show that for each $\varepsilon > 0$ the following inequalities obtain:

(a) $P[\max_{1\leq j\leq n}(S_j - \mu(S_j - S_n)) \geq \varepsilon] \leq 2P[S_n \geq \varepsilon]$;

(b) $P[\max_{1\leq j\leq n} |S_j - \mu(S_j - S_n)| \geq \varepsilon] \leq 2P[|S_n| \geq \varepsilon]$.

[*Hints*: Use the same decomposition for max as we did before. Thus, let $A_j = [S_j - S_n \leq \mu(S_j - S_n)]$, so that $P(A_j) \geq \frac{1}{2}, 1 \leq j \leq n$, and

$$B_j = [S_j - \mu(S_j - S_n) \geq \varepsilon, \text{ for the first time at } j].$$

Then $B_j \in \sigma(X_1, \ldots, X_j), A_j \in \sigma(X_{j+1}, \ldots, X_n)$, and they are independent; $\bigcup_{j=1}^{n} B_j = B = [\max(S_j - \mu(S_j - S_n)) \geq \varepsilon]$, a disjoint union. Thus $P[S_n \geq \varepsilon] \geq \sum_{j=1}^{n} P(B_j \cap A_j) \geq \frac{1}{2}P(B)$, giving (a). Since $\mu(-X) = -\mu(X)$, write $-X_j$ for $X_j, 1 \leq j \leq n$, in (a) and add it to (a) to obtain (b). Hence if the X_j are also symmetric, so that $\mu(X) = 0$, (a) and (b) take the following simpler form:

(a') $P[\max_{1\leq j\leq n} S_j \geq \varepsilon] \leq 2P[S_n \geq \varepsilon]$;

(b') $P[\max_{1\leq j\leq n} |S_j| \geq \varepsilon] \leq 2P[|S_n| \geq \varepsilon]$.]

18. Let $\{X_n, n \geq 1\}$ be independent random variables on (Ω, Σ, P) with zero means and variances $\{\sigma_n^2, n \geq 1\}$ such that $\sum_{n\geq 1} \sigma_n^2/b_n^2 < \infty$ for some $0 < b_n \leq b_{n+1} \nearrow \infty$. Then $(1/b_n)\sum_{k=1}^{n} X_k \to 0$ a.e. [*Hint*: Follow the proof of Theorem 3.6 except that in using Kronecker's lemma (Proposition 3.5) replace the sequence $\{n\}_{n\geq 1}$ there by the $\{b_n\}_{n\geq 1}$-sequence here. The same argument holds again.]

19. Let $\{X_n, n \geq 1\}$ be i.i.d. and be symmetric, based on (Ω, Σ, P). If $S_n = \sum_{k=1}^{n} X_k$, show that for each $\varepsilon > 0$,

$$\sum_{n\geq 1} P[|S_n| \geq n\varepsilon] < \infty \Rightarrow \{X_n, n \geq 1\} \subset L^2(P) \Leftrightarrow \sum_{n\geq 1}\sum_{j=1}^{n} P[|X_j| > n] < \infty.$$

[*Hints*: By Problem 17b' and the i.i.d. hypothesis, we have, with $S_0 = 0 = X_0$,

$$\begin{aligned} 2P[|S_n| \geq \varepsilon] &\geq P\left[\max_{0\leq j\leq n} |S_j| \geq \varepsilon\right] \geq P\left[\max_{0\leq j\leq n} |X_j| \geq 2\varepsilon\right] \\ &= 1 - \prod_{j=1}^{n} P[X_j < 2\varepsilon], \end{aligned} \tag{40}$$

since $[\max_{j\le n}|S_j|\ge\varepsilon]\supset[\max_{j\le n}|X_j|\ge 2\varepsilon]$. Summing and using the hypothesis with n for ε, and $\alpha_n=P[X_1<2n]$ in (40), we get

$$\begin{aligned}\infty>\sum_{n=1}^{\infty}(1-\alpha_n^n)&=\sum_{n=1}^{\infty}(1-\alpha_n)(1+\alpha_n+\ldots+\alpha_n^{n-1})\\&=\sum_{n=1}^{\infty}P[|X_1|\ge 2n]\sum_{j=0}^{n-1}\prod_{i=0}^{j}P[|X_i|<2n]\\&=\sum_{n=1}^{\infty}P[|X_1|\ge 2n]\sum_{j=0}^{n-1}P\left[\max_{i\le j}|X_i|<2n\right]\\&\ge\sum_{n=1}^{\infty}nP[|X_1|\ge 2n]\\&\quad\times\frac{1}{n}\sum_{j=0}^{n-1}(1-2P[|S_j|\ge j])\ \ [\text{ by (40)}].\end{aligned}\tag{41}$$

The convergence of the given series implies $P[|S_n|\ge n\varepsilon]\to 0$ as $n\to\infty$, and then by the $(C,1)$-summability the second term in (41) $\to 1$. Hence $\sum_{n=1}^{\infty}nP[|X_1|\ge 2n]<\infty$. But this is the same as the last series (by i.i.d.). Rewriting $P[|X_1|\ge 2n]$ as $\sum_{k\ge 2n}P[k\le|X_1|<k+1]$ and changing the order of summation one gets $X_1\in L^2(P)$, and by the i.i.d. hypothesis

$$\{X_n,n\ge 1\}\subset L^2(P).$$

The converse here is similar, so that the last equivalence follows. It should be remarked that actually all the implications are equivalences. The difficult part (the first one) needs additional computations, and we have not yet developed the necessary tools for its proof. This (harder) implication is due to Hsu and Robbins (1947), and we establish it later, in Chapter 4.] Show, however, what has been given is valid if the symmetry assumption is dropped in the hypothesis.

20. In the context of the preceding problem, we say [after Hsu and Robbins (1947)] that a sequence $\{Y_n,n\ge 1\}$ of random variables on (Ω,Σ,P) *converges completely* if for each $\varepsilon>0$, $(*)\sum_{n=1}^{\infty}P[|Y_n|>\varepsilon]<\infty$. Show that complete convergence implies convergence a.e. Also, verify that (*) implies that the a.e. limit of Y_n is necessarily zero. Establish by simple examples that the converse fails. [For example, consider the Lebesgue unit interval and $Y_n=n\chi_{[0,1/n]}$.] Show, however, that the converse implication does hold if there is a probability space (Ω',Σ',P'), a sequence $\{Z_n,n\ge 1\}$ of independent random variables on it such that $P[Y_n<x]=P'[Z_n<x],x\in\mathbb{R},n\ge 1$,

and $Z_n \to 0$ a.e. Compare this strengthening with Problem 5. [*Hint*: Note that $\limsup_n Z_n = 0$ a.e., and apply the second Borel-Cantelli lemma.]

21. The following surprising behavior of the symmetric random walk sequence was discovered by G. Pólya in 1921. Consider a symmetric random walk of a particle in the space $\mathbb{R}^k$. If $k = 1$, the particle moves in unit steps to the left or right, from the origin, with equal probability. If $k = 2$, it moves in unit steps in one of the four directions parallel to the natural coordinate axes with equal probability, which is $1/4$. In general, it moves in unit steps in the $2k$ directions parallel to the natural coordinate axes each step with probability $1/2k$. Show that the particle visits the origin infinitely often if $k = 1$ or 2, and only finitely often for $k = 3$. (The last is also true if $k > 3$.) [*Hints*: If $e_1, \ldots, e_k$ are the unit vectors in $\mathbb{R}^k$, so that $e_i = (0, \ldots, 1, 0, \ldots, 0)$ with 1 in the ith place, and $X_n : \Omega \to \mathbb{R}^k$ are i.i.d., then

$$P[X_n = e_i] = P[X_n = -e_i] = 1/2k, i = 1, \ldots, k.$$

Let $S_n = \sum_{j=1}^n X_j$. Then if $k = 1$, the result follows from Theorem 4.7, and if $k = 2$ or 3, we need to use Theorem 4.8 and verify the convergence or divergence of (35) there. If $p_n = P[|S_n| = 0]$, so that the particle visits 0 at step n with probability p_n, then the particle can visit 0 only if the positive and negative steps are equal. Thus $p_n = 0$ for odd n and $p_{2n} > 0$. However, by a counting argument ("multinomial distribution "), we see that

$$p_{2n} = \sum_{j=0}^{n} \frac{(2n)!}{[j!(n-j)!]^2} \left(\frac{1}{2 \cdot 2}\right)^{2n} = 4^{-2n} \binom{2n}{n} \sum_{j=0}^{n} \binom{n}{j}^2 = 4^{-2n} \binom{2n}{n}^2.$$

Using Stirling's approximation, $n! \sim \sqrt{2\pi n} n^n e^{-n}$, one sees that $p_n \sim 1/n$, and so $\sum_{n \geq 1} p_{2n} = \infty$, as desired. If $k = 3$, one gets by a similar computation

$$\begin{aligned} p_{2n} &= \sum_{\substack{0 \leq i,j \leq n \\ i+j \leq n}} \frac{(2n)!}{[i!j!(n-i-j)!]^2} \left(\frac{1}{2 \cdot 3}\right)^{2n} \\ &= \sum_{0 \leq i+j \leq n} \left(\frac{n!}{i!j!(n-i-j)!} \cdot \frac{1}{3^n}\right) \binom{2n}{n} \frac{1}{2^{2n}}. \end{aligned}$$

Again simplification by Stirling's formula shows that $p_{2n} \sim 1/n^{3/2}$, so that $\sum_{n \geq 1} p_{2n} < \infty$ (in fact, the series is approximately $= 0.53$), and S_n is not recurrent. By more sophisticated computations, Chung and Fuchs in their work on random walk showed that the same is true if the X_n are just i.i.d., with $E(X_1) = 0, 0 < E(|X_1|^2) < \infty$, and no component of X_1 is degenerate. This problem also shows an intimate relation between the structure of random walks and the group theoretical properties of its range (or state space), and deeper connections with convolution operators on these spaces or the group. For a recent contribution on the subject, and several references to the related literature on the problems, the reader is referred to Rao (2004a).]