STUDENT MATHEMATICAL LIBRARY
Volume 28

Heads or Tails

An Introduction to Limit Theorems in Probability

Emmanuel Lesigne

AMS
American Mathematical Society

http://dx.doi.org/10.1090/stml/028

Heads or Tails

An Introduction to Limit Theorems in Probability

This page intentionally left blank

STUDENT MATHEMATICAL LIBRARY
Volume 28

Heads or Tails

An Introduction to Limit Theorems in Probability

Emmanuel Lesigne

Translated by
Anna Pierrehumbert

AMS
American Mathematical Society

2000 *Mathematics Subject Classification.* Primary 60–01, 60Fxx.

The work was originally published in French by ELLIPSES under the title *PILE OU FACE—Une introduction aux théorèmes limites du Calcul des Probabilités*, Copyright 2001 Edition Marketing S.A. The present translation was created under license for the American Mathematical Society and is published by permission.

The photograph of the coin toss on the cover was created by Professor Andrew Davidhazy of the Rochester Institute of Technology; it is published by permission.

For additional information and updates on this book, visit
www.ams.org/bookpages/stml-28

Library of Congress Cataloging-in-Publication Data

Lesigne, Emmanuel, 1956–
[Pile ou face. English]
Heads or tails : an introduction to limit theorems in probability / Emmanuel Lesigne.
p. cm. — (Student mathematical library, ISSN 1520-9121 ; v. 28)
Includes bibliographical references and index.
ISBN 0-8218-3714-1 (alk. paper)
1. Limit theorems (Probability theory) I. Title. II. Series.

QA273.67.L4713 2005
519.2—dc22   2005041058

10 9 8 7 6 5 4 3 2 1   10 09 08 07 06 05

Contents

Preface

If you toss a fair coin many times, you would expect the coin to land *heads* as often as *tails*. The goal of this book is to make this intuition precise. As the number of tosses increases, the proportion of heads approaches $1/2$, but in what way, how quickly, and what deviations should we expect? *Heads or Tails* is an introduction to probability theory; in particular, it is an introduction to the study of convergence properties of sequences of observations. In this book, I will present an area of mathematics that has both utility and beauty.

Probability theory is the branch of mathematics concerned with the study of random phenomena. A *random phenomenon* is an experiment with an outcome that depends on chance, either because the exact conditions for its outcome are not known or because the randomness of the experiment actually exists. However, we will not discuss the sources of randomness in random phenomena; instead, we will start with a mathematical model of probability. *Heads or Tails* presents an introduction to the mathematical models of these phenomena and to the rigorous deduction of the laws we expect the outcomes of sequences of independent experiments to follow.

While writing this book, I kept the following three points in mind.

1. A freshman- or sophomore-level analysis course is all that is needed to understand the material in this book. In particular, a knowledge of measure theory is not necessary. This book is aimed

toward undergraduate students in math, science, and engineering programs, as well as teachers and all people with a basic knowledge of upper-level mathematics.

2. The level of rigor is that of most mathematics textbooks. The definitions and statements are precise and the proofs are complete.

3. Our discussion will essentially be limited to studying the game of Heads or Tails with a possibly unfair coin: we will study the laws that describe the result of sequences of identical, independent experiments with two possible outcomes. Although this choice may appear too restrictive, the simple game of Heads or Tails actually harbors much of the complexity of the general study of probability. This opinion is evident in Borel's statement that "The game of Heads or Tails, which seems so simple, is characterized by great generality and leads, when studied in detail, to the most sophisticated mathematics."[1]

This book is an invitation to probability theory. Some of the concepts and theorems that it contains are difficult because "elementary" is not a synonym for "easy". The reader should not expect to find strategies for winning the lottery or for maximizing returns from slot machines. On the contrary, the mathematics that we will study shows that the best strategy for such games of chance is abstinence.

Following the excellent suggestion of Pierre Damphousse, the founder and editor of the series in which the French edition of this book appears, I included precise historical background and biographical sketches. A brief bibliography is also included.

To conclude this Preface, I would like to thank the people who helped me pursue mathematical knowledge; the list of colleagues and students who should be thanked is too long to include here. In particular, however, I would like to acknowledge Jean Blanchard and Jean-Pierre Conze, who sparked my interest in mathematics, as well as my friends and colleagues Pierre Damphousse, Marc Peigné, and Elisabeth Rouy, who helped me while writing this short work.

Emmanuel Lesigne

Tours, February 2001

[1]Émile Borel, *Principes et formules classiques du Calcul des Probabilités*, Chapitre V: *Jeu de pile ou face*; 1924.

http://dx.doi.org/10.1090/stml/028/01

Prerequisites and Overview

Throughout this book, $\mathbb{R}$ is the set of real numbers, $\mathbb{Z}$ is the set of all integers, $\mathbb{N}$ is the set of nonnegative integers, and $\mathbb{N}^*$ is the set of positive integers.

The prerequisite knowledge used in this book is generally covered in the first two years of college:

- Elementary set theory: sets, product sets, functions;
- Combinatorics: countability, combinations;
- Real numbers: sequences, limits, comparison of sequences (the meaning of the symbols $\sim$, o, and O is reviewed in Chapter 6);
- Real functions of a real variable: limits and continuity, classical functions, integration of a continuous function over a real interval, Riemann sums.

In probability theory, a *limit theorem* is a theorem about convergence that relates to the outcome of a sequence of trials of a probabilistic experiment. Chapters 5 through 13 are each centered around a type of limit theorem.

Heads or Tails is composed of three parts. In the first part, consisting of Chapters 1 through 4, we provide the mathematical model used to describe a *finite probabilistic experiment* (that is, a probabilistic experiment with a finite number of possible outcomes); in the first and third sections of Chapter 11, we extend this discussion to infinite sequences of probabilistic experiments. In the second part, consisting of Chapters 5 through 10, we discuss theorems concerned with the probabilities associated to finite experiments. Two main results contained in these chapters are the weak law of large numbers and the central limit theorem. In addition, we discuss the large and moderate deviations estimates, which add precision to the weak law of large numbers and the central limit theorem, as well as the arcsine law and the local limit theorem. In the third part, consisting of Chapters 11 through 14, we model infinite probabilistic experiments. Here we provide various forms of the strong law of large numbers, a proof of the law of the iterated logarithm, and some results about the recurrence of random walks.

Starting with Chapter 5, each chapter opens with an introduction to the material of that chapter. A summary of this book can be obtained by assembling these introductions.

When combined with a presentation of countability, continuous and discrete probability distributions, and conditional probability, Chapters 1 through 7 would be appropriate for a first course in probability.

http://dx.doi.org/10.1090/stml/028/02

Chapter 1

Modeling a Probabilistic Experiment

1.1. Elementary Experiments

We will start by presenting the mathematical model that describes a probabilistic experiment having a finite number d of possible outcomes. Each outcome is represented by a variable ω^i, and the *sample space* is the set $\Omega := \{\omega^1, \omega^2, \ldots, \omega^d\}$ of all possible outcomes. To each outcome ω^i we associate a *probability* p_i. Each probability p_i is a nonnegative real number and $\sum_{i=1}^d p_i = 1$.

It is important to note that we assume that the probability of each outcome is given a priori. The work consisting of determining these probabilities from observations belongs to the study of statistics, a branch of mathematics that is related to but distinct from probability theory. The study of statistics uses tools and results that are presented in this book, but we will not deal with statistics directly.

Let us return to our model in order to introduce some vocabulary. A subset of Ω is called an *event* and the *probability* of an event is the sum of the probabilities of the outcomes belonging to that event. In symbols, if $A \subset \Omega$ is an event, then its probability $P(A)$ is defined by

$$P(A) := \sum_{\omega^i \in A} p_i.$$

In particular, we have that $P(\{\omega^i\}) = p_i$, which we will write simply as $P(\omega^i) = p_i$.

We let χ_A be the *characteristic function* of A; that is, χ_A is the function mapping Ω to $\{0, 1\}$ that takes the value 1 on A and the value 0 on its complement A^c. Thus

$$P(A) = \sum_{i=1}^{d} p_i \chi_A(\omega^i).$$

In summary, our mathematical model is defined by a pair (Ω, P) where Ω is a finite set and P is a function from the set of subsets of Ω to the interval [0,1] satisfying the following two conditions:

(1) $P(\Omega) = 1$.

(2) If A and B are disjoint subsets of Ω, then $P(A \cup B) = P(A) + P(B)$.

A pair (Ω, P) satisfying these conditions is called a *finite probability space* and the function P is called a *probability*. It is easy to check the following properties of P:

(1) $P(\emptyset) = 0$.

(2) If $A \subset \Omega$, then $P(A^c) = 1 - P(A)$.

(3) If $A, B \subset \Omega$, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

In the special case where all the outcomes are equally likely, we say that the space Ω has a *uniform probability*. In this case, it is easy to calculate the probability of an event: this probability is simply the number of elements in the event divided by d, the number of elements of Ω. This situation is described by the well-known rule that "the probability of an event is the ratio of the number of favorable outcomes to the total number of possible outcomes".

Here are a few examples.

Example. The flip of a fair coin is described by a set Ω of two elements and a probability giving the same value to each of the two outcomes. If we let 1 represent the outcome *heads* and 0 represent the outcome *tails*, then $\Omega = \{0, 1\}$ and $P(0) = P(1) = \frac{1}{2}$. We say that the space $\{0, 1\}$ is equipped with the uniform probability $(\frac{1}{2}, \frac{1}{2})$.

Example. More generally, the model describing a probabilistic experiment with two possible outcomes, which we call *success* and *failure*, is determined by a real parameter p between 0 and 1 that represents the probability of success. Writing 1 for the outcome success and 0 for the outcome failure, we have that $\Omega = \{0, 1\}$, $P(0) = 1 - p$ and $P(1) = p$. We say that the space $\{0, 1\}$ is equipped with the probability $(1 - p, p)$.

Example. The drawing of a number in a lottery (where the numbers range from, say, 1 to 49) is modeled by the pair (Ω, P), where $\Omega = \{1, 2, 3, \ldots, 49\}$ and P is the uniform probability on Ω (in this case, $P(\omega) = \frac{1}{49}$ for each $\omega \in \Omega$). If we only care about the parity of the number drawn, the relevant model would be a space Ω of two elements equipped with the probability $(\frac{24}{49}, \frac{25}{49})$.

Example. Even simple experiments can yield enormous probability spaces. For example, the space Ω needed to describe the drawing of a bridge hand, that is a choice of 13 cards out of 52, has $635{,}013{,}559{,}600$ elements (this is the binomial coefficient $\binom{52}{13}$; see Chapter 4). If the deck is randomly shuffled before the cards are distributed, this space will have a uniform probability.

As in every branch of mathematics, there are a few notations that are specific to probability theory. If X is a function from Ω to a set E and if F is a subset of E we write the inverse image of F by X as $(X \in F)$. In symbols, we have

$$(X \in F) := \{\omega \in \Omega : X(\omega) \in F\}.$$

Here we treat X as an element of the set E and $X(\omega)$ as the value of this element at the outcome ω. The probability of the event $(X \in F)$ is written as $P(X \in F)$.

1.2. Sequences of Elementary Experiments

In this book, we will mostly deal with sequences of identical and independent experiments. We will only consider finite sequences of experiments in the first part of the book, and we will start studying infinite sequences in Chapter 11.

We thus consider a *composite* experiment that consists of repeating an elementary experiment n times. We will suppose that the elementary experiment has two possible outcomes: success, denoted by the digit 1, and failure, denoted by the digit 0. Our model will incorporate the fact that these n elementary experiments are identical and independent. Let p be the probability of success and $q = 1 - p$ be the probability of failure. An outcome of the composite experiment is represented by a sequence of n zeros and ones. The space Ω, which we will write as Ω_n, is the set of ordered n-tuples of zeros and ones; that is, $\Omega_n = \{0, 1\}^n$. We denote the elements of Ω_n by $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$, where each ω_i equals 0 or 1. Finally, the probability on the space Ω_n is denoted by P_n. The fact that all the elementary experiments in the sequence are identical is conveyed by the rule

for each i between 1 and n ,

$$P_n(\omega_i = 0) = q \text{ and } P_n(\omega_i = 1) = p.$$

The fact that the outcome of the $(i+1)$-st trial is independent of the results of the i previous trials is conveyed by the rule

$$\begin{aligned} &\text{for each } (e_1, e_2, \ldots, e_i) \in \{0, 1\}^i, \\ &P_n(\omega_{i+1} = 1 \text{ and } (\omega_1, \omega_2, \ldots, \omega_i) = (e_1, e_2, \ldots, e_i)) \\ &\qquad = P_n(\omega_{i+1} = 1) \times P_n((\omega_1, \omega_2, \ldots, \omega_i) = (e_1, e_2, \ldots, e_i)). \end{aligned}$$

Inducting on n implies that these two rules uniquely define the probability P_n. In fact, if we let $S_n(\omega)$ be the number of successes for each outcome ω of the composite experiment, then the probability P_n is given by

$$P_n(\omega) = p^{S_n(\omega)} q^{n - S_n(\omega)}.$$

We say that the space $\Omega_n = \{0, 1\}^n$ is equipped with the *product probability* $P_n = (q, p)^{\otimes n}$.

If the probability of success equals the probability of failure, then the space Ω_n is equipped with a uniform probability. This agrees with what intuition suggests: in the experiment that consists of tossing a fair coin n times and recording the successive results, all the outcomes have equal probability. In this case, the probability of an event is simply equal to its cardinality divided by 2^n.

http://dx.doi.org/10.1090/stml/028/03

Chapter 2

Random Variables

Let (Ω, P) be a finite probability space. A function defined from Ω to $\mathbb{R}$ is called a *random variable*. Random variables are traditionally denoted by capital letters; for example,

$$\begin{aligned} X : \Omega &\to \mathbb{R} \\ \omega &\mapsto X(\omega). \end{aligned}$$

The *probability distribution* of the random variable X is given by the probabilities of the events corresponding to the values of X. If the random variable X takes the values $x_1, x_2, \dots x_k$, then the events $(X = x_i)$ for i from 1 to k form a partition of Ω and the distribution of X is given by the pairs $(x_i, P(X = x_i))$ for i ranging from 1 to k.

The *expected value* $E[X]$ of a random variable X is given by the formula

$$E[X] = \sum_{i=1}^{k} x_i P(X = x_i).$$

The concept of expected value as well as its name (*expectatio*, in Latin) were introduced by Christiaan Huygens in an analysis of bets in games of chance.[1]

[1]C. Huygens, *De ratiociniis in aleae ludo*, 1657.

This number $E[X]$ represents the average, under the probability P, of the values taken by the random variable X. The following properties follow easily from the definition of expected value.

- If X is a constant function, then $E[X] = X$; in particular, $E[E[X]] = E[X]$ for any random variable X.
- If $X \geq 0$, then $E[X] \geq 0$.
- $|E[X]| \leq E[|X|]$.
- The function E acts linearly on the real vector space of random variables on Ω (that is, if X and X' are two random variables and λ is a real number, then $E[X + X'] = E[X] + E[X']$ and $E[\lambda X] = \lambda E[X]$).

To verify the last statement, note that a random variable X can be represented in several ways as a linear combination of characteristic functions of events. If $X = \sum_i y_i \chi_{A_i}$ is such a representation, then $E[X] = \sum_i y_i P(A_i)$.

This remark also provides a proof of the formula for the expected value of a function of a random variable. If X is a random variable and if f is a real function defined on the image A of X, then $f(X) := f \circ X$ is a random variable and

$$E[f(X)] = \sum_{x \in A} f(x) P(X = x). \tag{2.1}$$

The following two inequalities are simple to prove and very useful.

Proposition 2.1 (Markov's inequality). *Let X be a random variable taking only nonnegative values. Then, for each $a > 0$,*

$$P(X \geq a) \leq \frac{E[X]}{a}.$$

Proof. Let $x_1, x_2, \ldots, x_k$ be the values taken by X. These are nonnegative real numbers, so

$$P(X \geq a) = \sum_{i : x_i \geq a} P(X = x_i) \leq \sum_{i : x_i \geq a} \frac{x_i}{a} P(X = x_i) \leq \frac{1}{a} E[X].$$

Corollary 2.2 (Bienaymé[2]–Chebyshev[3] inequality). *Let X be a random variable. Then, for each $a > 0$,*

$$P(|X - E[X]| \geq a) \leq \frac{1}{a^2} E\left[(X - E[X])^2\right].$$

This follows from Proposition 2.1 by applying Markov's inequality to the random variable $(X - E[X])^2$.

The value $E[(X - E[X])^2]$ is called the *variance* of the random variable X and is denoted by $\mathrm{var}(X)$. By expanding the square and using the linearity of the expected value function, we see that

$$\mathrm{var}(X) = E\left[(X - E[X])^2\right] = E\left[X^2\right] - \left(E[X]\right)^2.$$

The square root of the variance of X is called the *standard deviation* of X. The standard deviation measures the average deviation of the values of the random variable from the expected value.

We conclude with a remark about notation: the expected value associated with a probability P_n is naturally denoted by E_n.

[2]M. Bienaymé, *Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moidres carrés*, Journal de Mathématiques pures et appliquées, vol. 12, pp. 158–176, 1867.

[3]P. L. Chebyshev, *Des valeurs moyennes*, Journal de Mathématiques pures et appliquées, vol. 12, pp. 177–184, 1867.

This page intentionally left blank

http://dx.doi.org/10.1090/stml/028/04

Chapter 3

Independence

We will introduce the concept of independence first for a family of events and then for a family of random variables. As above, we consider a finite probability space (Ω, P).

In intuitive terms, we say that two events are independent if the result of one does not affect the result of the other. We can make this rigorous in the following way. Two sets A and B of Ω are *independent events* if $P(A \cap B) = P(A) \times P(B)$. Except in the trivial case where $P(B) = 0$, we can write this as $P(A \cap B)/P(B) = P(A)/P(\Omega)$.

Example. The model that describes the outcome of two independent and identical trials of an elementary experiment with probability of success p and probability of failure $q = 1 - p$ is

$$\Omega = \{(0,0),(0,1),(1,0),(1,1)\} = \{(\omega_1, \omega_2) : \omega_i = 0 \text{ or } 1\},$$

$$P(0,0) = q^2, \quad P(0,1) = P(1,0) = pq, \quad P(1,1) = p^2.$$

The model for the outcome of two experiments (each having two possible outcomes) that are independent but not necessarily identical and for which the probability of success are respectively denoted p_1 and p_2, is given by the same space Ω as above but equipped with the probability

$$P(0,0) = (1 - p_1)(1 - p_2), \qquad P(0,1) = (1 - p_1)p_2,$$

$$P(1,0) = p_1(1 - p_2), \qquad P(1,1) = p_1 p_2.$$

The reader should verify the elementary fact that the probability P is uniquely determined by the values of $P(\omega_i = 1)$ for $i = 1, 2$ and the condition that the events $(w_1 = 1)$ and $(w_2 = 1)$ are independent.

A family of events $\{A_1, A_2, \ldots, A_k\}$ is called a *family of independent events* if

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_\ell}) = P(A_{i_1}) \times P(A_{i_2}) \times \cdots \times P(A_{i_\ell})$$

for all integers $i_1, i_2, \ldots, i_l$ such that $1 \leq i_1 < i_2 < \cdots < i_l \leq k$. There are a few important observations to make about families of independent events.

- Events $A_1, A_2, \ldots, A_k$ satisfying $P\left(\bigcap_{i=1}^k A_i\right) = \prod_{i=1}^k P(A_i)$ are not necessarily independent. (In particular, this is clear if one of the events is empty.)
- The events in a family of pairwise independent events are not necessarily mutually independent. Here is the simplest illustration of this fact, due to Bernstein. Consider two tosses of fair coin, with success corresponding to heads and failure corresponding to tails. We examine the following three events: $A_1 := (\omega_1 = 1)$, that is, in the first toss the coin lands heads; $A_2 := (\omega_2 = 1)$, that is, in the second toss the coin lands heads; and $A_3 := (\omega_1 = \omega_2)$, that is, the results of the two tosses are the same. Then each pair of these events is a pair of independent events. However, the events A_1, A_2, and A_3 are not mutually independent, since $P(A_1 \cap A_2 \cap A_3) = \frac{1}{4}$ and $P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}$.

The concept of independence for random variables is a natural extension of the concept of independent events. Let $\{X_1, X_2, \ldots, X_k\}$ be a family of random variables defined on (Ω, P). These variables are *independent* if the events $(X_1 = x_1), (X_2 = x_2), \ldots, (X_k = x_k)$ are independent for all real $x_1, x_2, \ldots, x_k$. The reader should verify the following statements.

- The random variables $X_1, X_2, \ldots, X_k$ are independent if and only if the events $(X_1 \in B_1), (X_2 \in B_2), \ldots, (X_k \in B_k)$ are independent for all subsets $B_1, B_2, \ldots, B_k$ of $\mathbb{R}$.

- The property of independence does not depend on the order of the random variables, and a subset of a family of independent random variables is also a family of independent events.
- The random variables $X_1, X_2, \ldots, X_k$ are independent if and only if the events ($X_1 = x_1$ and $\cdots$ and $X_{j-1} = x_{j-1}$) and $(X_j = x_j)$ are independent for all j between 2 and k and for all $x_1, x_2, \ldots, x_j$. (If this is the case, we say that the random variable X_j is independent of the random vector $(X_1, X_2, \ldots, X_{j-1})$.)
- A given family of subsets of Ω forms a family of independent events if and only if their characteristic functions form a family of independent random variables.

Remark. In the penultimate assertion above, we wrote the event $(X_1 = x_1) \cap (X_2 = x_2)$ as ($X_1 = x_1$ and $X_2 = x_2$). This use of the conjuction "and" to correspond to the intersection of sets is natural, and we will use this throughout the rest of the book.

The following two propositions will be very useful.

Proposition 3.1. *Suppose that $(X_i)_{i\in I}$ is a finite family of independent random variables and that J and K are subsets of I. If $J \cap K = \emptyset$, Y is a real function defined on $(X_i)_{i\in J}$, and Z is a real function defined on $(X_i)_{i\in K}$, then Y and Z are independent random variables.*

Proof. Let f and g be real functions defined respectively on $\mathbb{R}^J$ and $\mathbb{R}^K$. Suppose that $Y = f\big((X_j)_{j\in J}\big)$ and $Z = g\big((X_k)_{k\in K}\big)$, and let $a, b \in \mathbb{R}$. Let A be the (finite) set of $(x_j)_{j\in J}$ such that $f\big((x_j)_{j\in J}\big) = a$ and $\big(X_j(\omega)\big)_{j\in J} = (x_j)_{j\in J}$ for some $\omega \in \Omega$. Let B be the set of $(x_k)_{k\in K}$ such that $g\big((x_k)_{k\in K}\big) = b$ and $\big(X_k(\omega)\big)_{k\in K} = (x_k)_{k\in K}$ for some $\omega \in \Omega$. Then

$$(Y = a) = \bigcup_{(x_j)\in A} \bigcap_{j\in J} (X_j = x_j)\,, \quad (Z = b) = \bigcup_{(x_k)\in B} \bigcap_{k\in K} (X_k = x_k)$$

and

$$(Y = a \text{ and } Z = b) = \bigcup_{(x_j)\in A,\,(x_k)\in B} \; \bigcap_{i\in J\cup K} (X_i = x_i).$$

Since this is the union of pairwise disjoint events,

$$P(Y = a \text{ and } Z = b) = \sum_{(x_j)\in A,\,(x_k)\in B} P\Bigg(\bigcap_{i\in J\cup K} (X_i = x_i)\Bigg).$$

By the independence of the random variables X_i, we obtain

$$\begin{aligned} P(Y = a \text{ and } Z = b) &= \sum_{(x_j)\in A,\,(x_k)\in B} \prod_{i\in J\cup K} P(X_i = x_i) \\ &= \Big(\sum_{(x_j)\in A} \prod_{j\in J} P(X_j = x_j)\Big)\Big(\sum_{(x_k)\in B} \prod_{k\in K} P(X_k = x_k)\Big) \\ &= P(Y = a)\, P(Z = b). \end{aligned}$$

This proves that the variables Y and Z are independent. □

Proposition 3.2. *If X and Y are independent random variables, then*

$$E[XY] = E[X] \times E[Y]$$

and

$$\operatorname{var}(X + Y) = \operatorname{var}(X) + \operatorname{var}(Y).$$

Proof. Let A be the (finite) set of values taken by X, and let B be the (finite) set of values taken by Y. Then

$$E[XY] = \sum_{(a,b)\in A\times B} abP(X = a \text{ and } Y = b),$$

and the condition of independence implies that

$$\begin{aligned} E[XY] &= \sum_{a\in A, b\in B} abP(X = a)P(Y = b) \\ &= \sum_{a\in A} aP(X = a) \sum_{b\in B} bP(Y = b) = E[X] \times E[Y]. \end{aligned}$$

Using this formula, we see that

$$\begin{aligned} &\operatorname{var}(X + Y) = E\left[(X + Y)^2\right] - \big(E[X + Y]\big)^2 \\ &= \left(E\left[X^2\right] + E\left[Y^2\right] + 2E[XY]\right) - \left(E[X]^2 + E[Y]^2 + 2E[X]E[Y]\right) \\ &= E\left[X^2\right] - E[X]^2 + E\left[Y^2\right] - E[Y]^2 = \operatorname{var}(X) + \operatorname{var}(Y). \end{aligned}$$

□

http://dx.doi.org/10.1090/stml/028/05

Chapter 4

The Binomial Distribution

The rest of this book will be focused on the following problem. Let S_n be the number of successful outcomes in a sequence of n identical and independent random trials, each having two possible outcomes. What can we say about the behavior of the sequence (S_n) as n approaches infinity?

Let p be the probability of success for each of the elementary experiments, and consider the probability space (Ω_n, P_n) described in Section 1.2. Let the function S_n defined on $\Omega_n = \{0,1\}^n$ be the random variable defined by $S_n(\omega) = \sum_{i=1}^{n} \omega_i$.

Before proceeding, we must review the concept of the *binomial coefficients* $\binom{n}{k}$. If k and n are two integers such that $0 \le k \le n$, then $\binom{n}{k}$ is the number of k-element subsets of an n-element set. These numbers satisfy the recurrence relation of Pascal's triangle

$$\binom{n}{0} = \binom{n}{n} = 1; \qquad \binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1} \text{ for } 0 \le k < n.$$

They are given by the formula

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

and they appear in Newton's binomial theorem

$$(4.1) \qquad (a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Proposition 4.1. *The random variable S_n only takes integer values between 0 and n and*

$$P_n(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for every k between 0 and p.

We say that the random variable S_n follows a *binomial distribution* with parameters n and p.

Proof. Let $q = 1 - p$. For each $\omega \in \Omega_n$,

$$P_n(\omega) = p^{S_n(\omega)} q^{n-S_n(\omega)}.$$

The event $(S_n = k)$ includes exactly the outcomes ω with probability $p^k q^{n-k}$. Therefore, the probability of this event equals its cardinality multiplied by $p^k q^{n-k}$. Since the event's cardinality is the number of ways to obtain k successes among the outcomes of n trials, the cardinality is $\binom{n}{k}$. □

A random variable X follows the *Bernoulli distribution* with parameter p if it only takes the values 0 and 1 and if $P(X = 1) = p$. This is the same as the binomial distribution with parameters 1 and p. The following proposition, which follows easily from the results of the previous chapter, shows how the binomial distribution with parameters n and p relates to the Bernoulli distribution with parameter p.

Proposition 4.2. *If $X_1, X_2, \ldots, X_n$ are independent random variables following a Bernoulli distribution with parameter p, then their sum $X_1 + X_2 + \cdots + X_n$ is a random variable following a binomial distribution with parameters n and p.*

Proposition 4.3. $E[S_n] = np$ *and* $\operatorname{var}(S_n) = np(1-p)$.

Proof of Proposition 4.3. Proposition 4.1 implies that

$$E[S_n] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$$

and, by using formula (2.1), that

$$\operatorname{var}(S_n) = \sum_{k=0}^{n} (k - E[S_n])^2 \binom{n}{k} p^k (1-p)^{n-k}.$$

From these, the formulas given in Proposition 4.3 follow easily from Newton's binomial theorem. (Starting with the binomial formula (4.1), use the formulas obtained by differentiating twice with respect to the variable a.)

Note that Proposition 4.3 also follows immediately from Propositions 4.2 and 3.2. Indeed, if X is a random variable following a Bernoulli distribution with parameter p, then $E[X] = p$ and $\operatorname{var}(X) = p(1-p)$. □

http://dx.doi.org/10.1090/stml/028/06

Chapter 5

The Weak Law of Large Numbers

The probability of success for a random experiment is usually obtained experimentally from the frequency of success in a sequence of (identical and independent) repetitions of that experiment. However, in our mathematical model the probability of success is given a priori. The law of large numbers reconciles these two notions of probability. In fact, it shows that for a large number of trials, it is very probable that the frequency of success is close to the theoretical probability. This is the first justification for our mathematical model of probability.

Consider the situation studied in the preceding section in which $\Omega_n = \{0,1\}^n$ is the space of outcomes for n trials of the experiment. Supposing that these trials are identical and independent, the space Ω_n is equipped with the product probability $P_n = (1-p,p)^{\otimes n}$, where p is a parameter between 0 and 1 that represents the theoretic probability of success (see Section 1.2). As before, we let S_n be the random variable that counts the number of successes

$$S_n(\omega) = \omega_1 + \omega_2 + \cdots + \omega_n, \qquad \text{where } \omega = (\omega_1, \omega_2, \ldots, \omega_n) \in \Omega_n.$$

The random variable S_n follows a binomial distribution with parameters n and p.

The *empirical* (or *experimental*) *probability* of success is $\frac{S_n}{n}$. For large n, we expect this frequency to be close to p. The following theorem says that this is very likely to be true. The result appeared for the first time in a posthumously published work by Jacob Bernoulli.[1]

Theorem 5.1 (weak law of large numbers). *For each* $\epsilon > 0$,

$$P_n\left(\left|\frac{S_n}{n} - p\right| > \epsilon\right) \longrightarrow 0$$

as n *approaches infinity, and this convergence is uniform in* p.

This was named the *weak law of large numbers* by Siméon Denis Poisson,[2] who generalized Bernoulli's result to cases where the probability of success varies from trial to trial. (We will discuss this result later; see Theorem 11.12.)

Proof. The variance of the random variable S_n is $\text{var}(S_n)=np(1-p)$. By the Bienaymé–Chebyshev inequality (Corollary 2.2),

$$P_n\left(|S_n - np| > n\epsilon\right) \leq \frac{1}{(n\epsilon)^2}\text{var}(S_n) = \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2},$$

which proves the theorem. □

In the case of tossing a fair coin ($p = 1/2$), the weak law of large numbers says that the proportion of sequences of 0's and 1's of length n in which the frequency of 1's differs from $1/2$ by less than ϵ tends to 1 as n approaches infinity.

Bernoulli's weak law of large numbers has a nice application to the problem of uniformly approximating a continuous function on an interval of $\mathbb{R}$ by a polynomial. The Weierstrass approximation theorem states that, for any real function f defined and continuous on a closed and bounded interval $[a, b]$ of $\mathbb{R}$ and for every $\epsilon > 0$, there exists a real polynomial function g such that

$$\sup_{a \leq x \leq b} |f(x) - g(x)| < \epsilon.$$

---

[1] J. Bernoulli, *Ars conjectandi*, 1713.

[2] S. D. Poisson, *Recherches sur la probabilité des jugements en matière criminelle et en matiére civile*, Paris, 1837.

Serge Bernstein[3] gave a proof of this result using the weak law of large numbers. His method has the advantage that it gives an explicit formula for the approximating polynomials.

Proposition 5.2 (Bernstein polynomials). *Let f be a real function that is defined and continuous on the interval $[0, 1]$. Then*

$$\sup_{0\leq x\leq 1}\left|f(x)-\sum_{k=0}^{n}\binom{n}{k}f\left(\frac{k}{n}\right)x^k(1-x)^{n-k}\right|\longrightarrow 0$$

as n approaches infinity.

This theorem implies that the polynomials approximating f on an arbitrary interval $[a, b]$ are given by

$$x\mapsto\sum_{k=0}^{n}\binom{n}{k}f\left(a+(b-a)\frac{k}{n}\right)\left(\frac{x-a}{b-a}\right)^k\left(\frac{b-x}{b-a}\right)^{n-k}.$$

Proof. Fix an $\epsilon > 0$. Since f is uniformly continuous, there exists an $\eta > 0$ such that

$$0\leq x,y\leq 1 \text{ and } |x-y|<\eta \Rightarrow |f(x)-f(y)|<\epsilon.$$

Now, consider the probability space (Ω_n, P_n) as previously defined and the random variable $f(\frac{S_n}{n})$. Bernstein's polynomials appear as the expected value of this random variable; indeed,

$$\begin{aligned}E_n\left[f\left(\frac{S_n}{n}\right)\right]&=\sum_{k=0}^{n}f\left(\frac{k}{n}\right)P_n(S_n=k)\\&=\sum_{k=0}^{n}\binom{n}{k}f\left(\frac{k}{n}\right)p^k(1-p)^{n-k}.\end{aligned}$$

By the law of large numbers, there exists an integer n_0, independent of the parameter p, such that

$$P_n\left(\left|\frac{S_n}{n}-p\right|>\eta\right)<\epsilon$$

for every $n \geq n_0$. We have

$$\left|E_n\left[f\left(\frac{S_n}{n}\right)\right]-f(p)\right|=\left|\sum_{k=0}^{n}\left(f\left(\frac{k}{n}\right)-f(p)\right)P_n(S_n=k)\right|,$$

[3] S. Bernstein, *Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités*, Soobsch. Charkovskovo Mat. Obsch., vol. 13, pp. 1–2, 1912.

and the triangle inequality implies that an upper bound for this expression is

$$\sum_{|\frac{k}{n}-p|\le\eta} \left|f\left(\frac{k}{n}\right)-f(p)\right| P_n(S_n=k)$$
$$+\sum_{|\frac{k}{n}-p|>\eta} \left(\left|f\left(\frac{k}{n}\right)\right|+|f(p)|\right) P_n(S_n=k)$$
$$\le \sum_{|\frac{k}{n}-p|\le\eta} \epsilon P_n(S_n=k)+\sum_{|\frac{k}{n}-p|>\eta} 2\sup_{0\le x\le 1}|f(x)|P_n(S_n=k)$$
$$=\epsilon P_n\left(\left|\frac{S_n}{n}-p\right|\le\eta\right)+2\sup_{0\le x\le 1}|f(x)|P_n\left(\left|\frac{S_n}{n}-p\right|>\eta\right).$$

Thus for every $n\ge n_0$,

$$\left|E_n\left[f\left(\frac{S_n}{n}\right)\right]-f(p)\right|\le\epsilon+2\epsilon\sup_{0\le x\le 1}|f(x)|.$$

This proves that $\left|E_n\left[f\left(\frac{S_n}{n}\right)\right]-f(p)\right|$ can be made arbitrarily small, uniformly with respect to p, by picking a large enough n. □

http://dx.doi.org/10.1090/stml/028/07

Chapter 6

The Large Deviations Estimate

Many of the results we will discuss are refinements of the weak law of large numbers. We will now present the first of these refinements, which establishes that the rate of convergence in this law is exponential.

By using the Bienaymé–Chebyshev inequality in the proof of the weak law of large numbers, we obtained the estimate

$$P_n\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq \frac{p(1-p)}{n\epsilon^2}.$$

As Serge Bernstein[1] remarked, this upper bound can be greatly improved for large values of n.

For every $\epsilon \in (0, 1-p)$, we define

$$h_+(\epsilon) := (p+\epsilon)\ln\frac{p+\epsilon}{p} + (1-p-\epsilon)\ln\frac{1-p-\epsilon}{1-p}.$$

This function appears in the statement of the following theorem.

[1] S. Bernstein, *Sur une modification de l'inégalité de Tchebychev*, Ann. Sc. Instit. Sav. Ukraine, Sect. Math I, 1924.

Theorem 6.1 (large deviations estimate). *For every* $\epsilon \in (0, 1-p)$ *and* $n \geq 1$, $h_+(\epsilon) > 0$ *and*

$$P_n\left(\frac{S_n}{n} \geq p + \epsilon\right) \leq e^{-nh_+(\epsilon)}.$$

This theorem immediately yields the following two results.

(1) If $0 < \epsilon < p$, then $h_-(\epsilon) := h_+(-\epsilon)$ is well defined and positive, and

$$P_n\left(\frac{S_n}{n} \leq p - \epsilon\right) \leq e^{-nh_-(\epsilon)}.$$

(This follows by applying the theorem after interchanging success and failure. In this case, S_n is replaced by $n - S_n$ and p is replaced by $1-p$.)

(2) If $0 < \epsilon < \min(p, 1-p)$, then

$$P_n\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq e^{-nh_+(\epsilon)} + e^{-nh_-(\epsilon)},$$

which approaches 0 exponentially as n approaches infinity.

Proof. Fix a $t > 0$. We have

$$P_n\left(\frac{S_n}{n} \geq p + \epsilon\right) = P_n\left(e^{t(S_n - np - n\epsilon)} \geq 1\right),$$

and Markov's inequality implies that

$$\begin{aligned} P_n\left(\frac{S_n}{n} \geq p + \epsilon\right) &\leq E_n\left[e^{t(S_n - np - n\epsilon)}\right] \\ &= e^{-nt(p+\epsilon)} E_n\left[e^{tS_n}\right] \\ &= e^{-nt(p+\epsilon)} \sum_{k=0}^{n} e^{tk}\binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

By Newton's binomial formula, then

$$P_n\left(\frac{S_n}{n} \geq p + \epsilon\right) \leq e^{-nt(p+\epsilon)}(1-p+pe^t)^n = e^{-n(t(p+\epsilon) - \ln(1-p+pe^t))}.$$

Since this is true for every $t > 0$, we obtain

$$P_n\left(\frac{S_n}{n} \geq p + \epsilon\right) \leq e^{-nh},$$

where $h = \sup_{t>0}(t(p+\epsilon) - \ln(1-p+pe^t))$. To complete the proof, we only need to prove that

$$h = (p+\epsilon)\ln\frac{p+\epsilon}{p} + (1-p-\epsilon)\ln\frac{1-p-\epsilon}{1-p} > 0.$$

This follows from an analysis of the behavior of the function $g : t \mapsto t(p+\epsilon) - \ln(1-p+pe^t)$ on $[0, +\infty)$. We leave it to the reader to check that $g(0) = 0$ and $g'(0) = \epsilon > 0$. Together, these imply that the supremum of g is strictly positive. The derivative of g vanishes at the point $t = \ln\frac{(p+\epsilon)(1-p)}{p(1-p-\epsilon)}$; g is maximized at this point, and its value is in fact $h_+(\epsilon)$. □

It is easy to illustrate the large deviations estimate with numerical applications. Here are a few.

(1) If a fair coin is tossed 100 times, the probability that the number of heads is at least 60 is less than .14. Here $(p, n, \epsilon) = (.5, 100, .1)$.

(2) If a fair coin is tossed 1000 times, the probability that the number of heads is at least 600 is less than $2 \cdot 10^{-9}$. Here $(p, n, \epsilon) = (.5, 1000, .1)$.

(3) If a fair coin is tossed 1000 times, the probability that the number of heads is at least 540 is less than .05. Here $(p, n, \epsilon) = (.5, 1000, .04)$.

It is useful to understand the behavior of the exponent $h_+(\epsilon)$ as ϵ approaches zero. A second order expansion of the logarithm function immediately yields the following result.

Proposition 6.2. *As ϵ approaches 0,*

$$h_+(\epsilon) = \frac{\epsilon^2}{2p(1-p)} + O(\epsilon^3).$$

We conclude this chapter with the following proposition, which says that the large deviations estimate given by Theorem 6.1 is optimal in the sense that the exponent cannot be improved.

Proposition 6.3. *For all $\epsilon \in (0, 1-p)$,*

$$\lim_{n\to\infty}\frac{1}{n}\ln\left(P_n\left(\frac{S_n}{n} \geq p+\epsilon\right)\right) = -h_+(\epsilon). \tag{6.1}$$

To prove this proposition, we will use Stirling's result that the sequence $\left(n^{-(n+1/2)}e^n n!\right)_{n\geq 1}$ converges to a positive number c. We will prove this result in a more precise form in the following chapter (see Proposition 7.2).

Before proving Proposition 6.3, let us take a moment to review asymptotic notation used for comparing sequences. Let (u_n) and (v_n) be two real sequences. We write $u_n = O(v_n)$ if there exists a constant $k > 0$ such that $|u_n| \leq k v_n$ for every n; if this is the case, we say that u_n is of order at most v_n. We write $u_n = o(v_n)$ if for every $\epsilon > 0$ there exists an $n_0 \geq 1$ such that $|u_n| \leq \epsilon v_n$ for each $n \geq n_0$; if this is the case, we say that u_n is negligible compared to v_n. Finally, we write $u_n \sim v_n$ if $u_n - v_n = o(|v_n|)$ (for sequences that do not vanish, this is equivalent to saying the quotient u_n/v_n tends to 1 as n tends to infinity); if this is the case, we say that the two sequences are aymptotically equal. Using this notation, we can rewrite Stirling's formula as $n! \sim c n^{n+1/2} e^{-n}$.

Proof of Proposition 6.3. For each $n \geq 1$, the large deviations estimate can be written as

$$\frac{1}{n} \ln\left(P_n\left(S_n \geq n(p+\epsilon)\right)\right) \leq -h_+(\epsilon). \tag{6.2}$$

We need to find a lower bound. Let $K_n := 1 + [n(p+\epsilon)]$ be the least integer greater than $n(p+\epsilon)$. It is clear that

$$P_n\left(S_n \geq n(p+\epsilon)\right) \geq P_n\left(S_n = k_n\right),$$

and we will show that

$$\lim_{n\to\infty} \frac{1}{n} \ln\left(P_n\left(S_n = k_n\right)\right) = -h_+(\epsilon). \tag{6.3}$$

This will imply that

$$\liminf_{n\to\infty} \frac{1}{n} \ln\left(P_n\left(S_n \geq n(p+\epsilon)\right)\right) \geq -h_+(\epsilon),$$

which, with (6.2), yields the desired result.

The sequences (k_n) and $(n - k_n)$ approach infinity with n and

$$P_n\left(S_n = k_n\right) = \frac{n!}{k_n!(n-k_n)!} p^{k_n}(1-p)^{n-k_n}.$$

By applying Stirling's formula to the three factorial sequences, we obtain

$$(6.4)\quad P_n\left(S_n = k_n\right) \sim \frac{1}{c}\sqrt{\frac{1}{n}k_n(n-k_n)}\left(\frac{np}{k_n}\right)^{k_n}\left(\frac{n(1-p)}{n-k_n}\right)^{n-k_n}.$$

Since these two sequences that approach zero as n approaches infinity are asymptotically equal, so are their logarithms. Let us first analyze the behavior of the logarithm of the right hand side. Since $k_n \sim n(p+\epsilon)$ and $n-k_n \sim n(1-p-\epsilon)$,

$$\sqrt{\frac{n}{k_n(n-k_n)}} \sim \frac{1}{\sqrt{n}};$$

hence,

$$\lim \frac{1}{n}\ln\left(\frac{1}{c}\sqrt{\frac{n}{k_n(n-k_n)}}\right) = 0.$$

In addition,

$$k_n \ln\left(\frac{np}{k_n}\right) = n(p+\epsilon)\ln\left(\frac{p}{p+\epsilon}\right) + n(p+\epsilon)\ln\left(\frac{n(p+\epsilon)}{k_n}\right)$$
$$+ (k_n - n(p+\epsilon))\ln\left(\frac{np}{k_n}\right),$$

and the fact that $k_n - n(p+\epsilon)$ is bounded implies that

$$\lim \frac{1}{n}\ln\left(\left(\frac{np}{k_n}\right)^{k_n}\right) = (p+\epsilon)\ln\left(\frac{p}{p+\epsilon}\right).$$

By the same method we can show that

$$\lim \frac{1}{n}\ln\left(\left(\frac{n(1-p)}{n-k_n}\right)^{n-k_n}\right) = (1-p-\epsilon)\ln\left(\frac{1-p}{1-p-\epsilon}\right).$$

We deduce (6.3) from (6.4) and the last three limit calculations. This proves the theorem. □

This page intentionally left blank

http://dx.doi.org/10.1090/stml/028/08

Chapter 7

The Central Limit Theorem

7.1. Statement of the Theorem

Obtained through the successive work of Abraham de Moivre,[1] Pierre Simon Laplace,[2] and Carl Friedrich Gauss,[3] the *central limit theorem* plays a central role in probability theory. The theorem is fascinating because of the extremely wide range of applications, and it establishes the universal role of the normal (or Gaussian) distribution, the famous *bell curve.* In view of the goals of this book, we will limit our study to sequences of identically distributed independent random variables, each taking only two values.

Let us now return to the previously introduced setting. The space $\Omega_n = \{0,1\}$ is equipped with the product probability $P_n = (p, 1-p)^{\otimes n}$ and $S_n(\omega) = \omega_1 + \omega_2 + \cdots + \omega_n$, where $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$. In addition, we suppose that $0 < p < 1$, which eliminates the two trivial cases from our analysis.

The weak law of large numbers gave us a restriction on the fluctuations of S_n around its average value np: if n is large enough,

[1]A. de Moivre, *The Doctrine of Chances*, London, 1718.

[2]P. S. Laplace, *Théorie analytique des probabilités*, Paris, 1812.

[3]C. F. Gauss, *Theoria combinationis observationum erroribus minimus obnoxiae*, 1821.

fluctuations of order n are very improbable. In addition, we know that the expected value of $(S_n - np)^2$ equals $np(1-p)$. This suggests that likely deviations of S_n from its average have order $\sqrt{n}$. The central limit theorem confirms this intuition and gives quantitative (but asymptotic) information about these deviations. Here is the precise statement of the central limit theorem.

Theorem 7.1 (central limit theorem). *Let a and b be two elements of $\mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$ such that $a < b$. Then*

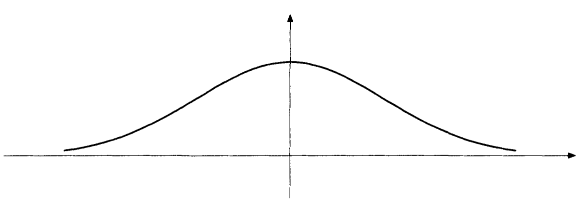
$$(7.1)\quad P_n\left(a \le \frac{S_n - np}{\sqrt{p(1-p)}\sqrt{n}} \le b\right) \longrightarrow \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-x^2/2\right)\, dx$$

as n approaches infinity.

Before proving the central limit theorem, we will make a few remarks about the theorem and present some of its applications.

7.2. Remarks

The graph of the function $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ is the Gaussian curve.



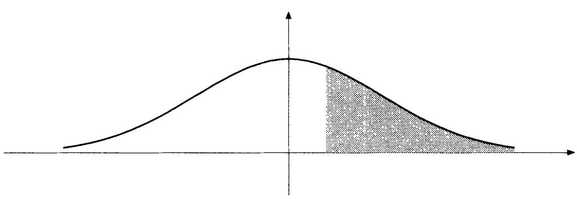
In the proof of the theorem, we will need the fact that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2}\, dx = 1.$$

(This shows that the function $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ is a *probability density* on $\mathbb{R}$; however, we will not elaborate on this viewpoint.)

The integral of the function $x \mapsto e^{-x^2/2}$ cannot be expressed in terms of classical functions, so integrals of the form $\frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx$ cannot written in a simpler way. Nevertheless, the numerical values of such functions can be calculated with good precision. Tables of these values can be found in many books because they play such an important role in the numerous applications of the central limit theorem in statistics and probability. In addition, mathematical software, such as Mathematica® and Maple®, can be used to compute these values.

The function $\Phi : y \mapsto \frac{1}{\sqrt{2\pi}} \int_y^{+\infty} e^{-x^2/2}\, dx$, which measures the area of the *tail* of the Gaussian curve, decreases very quickly as y increases in $(0, +\infty)$.



For example,

$$\Phi(1) \approx 0.1587, \quad \Phi(2) \approx 0.0228, \quad \text{and } \Phi(4) \approx 3.2 \cdot 10^{-5}.$$

For each $y > 0$,

$$\frac{1}{\sqrt{2\pi}(y+1)} \left(e^{-y^2/2} - e^{-(y+1)^2/2}\right) \le \Phi(y) \le \frac{1}{\sqrt{2\pi} y} e^{-y^2/2},$$

so as y approaches $+\infty$,

$$\Phi(y) \sim \frac{1}{\sqrt{2\pi} y} e^{-y^2/2}.$$

The verification of these statements is left to the reader as a little exercise in analysis.

The central limit theorem can be strengthened by a result about uniform convergence: the convergence in (7.1) is uniform in a and b. A classical result of elementary analysis justifies our assertion: every sequence of real monotonic functions on $\mathbb{R}$ taking values in $[0, 1]$ that converges to a continuous function with image containing $(0, 1)$ converges uniformly.

By using the expression for the binomial distribution, we can rewrite the central limit theorem in the following way: if $y \in \mathbb{R}$ and if $k(n)$ is the integer part of $np + y\sqrt{np(1-p)}$, then

$$\sum_{j=0}^{k(n)} \binom{n}{j} p^j (1-p)^{n-j} \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-x^2/2}\, dx$$

as n approaches infinity. This follows from that fact that

$$\sum_{j=0}^{k(n)} \binom{n}{j} p^j (1-p)^{n-j} = P_n\left(S_n \leq np + y\sqrt{np(1-p)}\right).$$

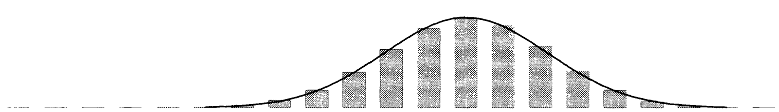
We could also write

$$\sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j} \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{k-np}{\sqrt{np(1-p)}}} e^{-x^2/2}\, dx,$$

which, by making the change of variable $t = np + x\sqrt{np(1-p)}$, is equivalent to

$$\sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j} \approx \frac{1}{\sqrt{2\pi np(1-p)}} \int_{-\infty}^{k} \exp\left(-\frac{(t-np)^2}{2np(1-p)}\right) dt.$$

We can illustrate this result by showing the similarity between the bell curve and histograms representing the binomial distribution: it suffices to plot the function taking the value $\frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{(t-np)^2}{2np(1-p)}\right)$ for each real t and the function taking the value $\binom{n}{j} p^j (1-p)^{n-j}$ for each integer j between 1 and n on the same axes. The following figure shows these two functions where the parameters take values $n = 20$ and $p = .6$.



This graph suggests that

$$\binom{n}{j} p^j (1-p)^{n-j} \approx \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{(j-np)^2}{2np(1-p)}\right).$$

This will be confirmed by the de Moivre–Laplace theorem, which will be stated and proved later in this chapter as a step towards proving the central limit theorem.

7.3. Applications

Here are a few examples of the applications of the central limit theorem. We will give a theoretical application to the law of large numbers as well as more practical examples of its applications in estimating probabilities, calculating confidence intervals, and performing statistical tests.

The weak law of large numbers is an immediate consequence of the central limit theorem. Fix $\epsilon > 0$ and $\delta > 0$. There exists an $a > 0$ such that $\Phi(a) < \delta$ and $\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}} \geq a$ for large enough n. Then, for such an integer n,

$$P_n\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) = P_n\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq -\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) + P_n\left(\frac{S_n - np}{\sqrt{np(1-p)}} \geq \frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right),$$

so

$$P_n\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq P_n\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq -a\right) + P_n\left(\frac{S_n - np}{\sqrt{np(1-p)}} \geq a\right).$$

By the central limit theorem,

$$\left|P_n\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq -a\right) - \Phi(a)\right| < \delta$$

and

$$\left|P_n\left(\frac{S_n - np}{\sqrt{np(1-p)}} \geq a\right) - \Phi(a)\right| < \delta$$

for large enough n. From the choice of a and the above inequalities, we conclude that

$$P_n\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq 4\delta$$

for large n. We have thus shown that

$$P_n\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \longrightarrow 0$$

as n approaches infinity.

After examining this argument, it is evident that we can actually prove a stronger result: for every real sequence (u_n) such that $\lim_{n\to\infty} \frac{u_n}{\sqrt{n}} = 0$,

$$P_n\left(u_n\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \longrightarrow 0.$$

We will present other consequences of the central limit theorem over the course of this book. In particular, in Chapter 10 we will explain the meaning of the assertion that "the sequence (S_n) is almost surely not bounded", and we will prove this assertion.

Now we will explain some more practical applications of the central limit theorem. These applications are based on estimating probabilities of the form

$$P_n\left(np + \sqrt{n}a \leq S_n \leq np + \sqrt{n}b\right)$$

by the value

$$\frac{1}{\sqrt{2\pi}}\int_{\frac{a}{\sqrt{p(1-p)}}}^{\frac{b}{\sqrt{p(1-p)}}} e^{-x^2/2}\,dx.$$

However, the statement of the central limit theorem does not say how good of an approximation this is. There are theoretical results, such as the Berry–Esseen inequality, that give explicit upper bounds for the error in the approximation. These theorems are addressed in more advanced texts about probability theory, and we will not discuss them further here. In general, the approximation given by the central limit theorem is considered to be acceptable whenever $np(1-p) > 18$. This statement deserves to be studied more closely, which is certainly possible with modern methods of calculus. A discussion of this statement appears in William Feller's book.[4]

In the previous chapter, we determined an upper bound for the probability that the number of heads would be at least 60 after tossing a fair coin 100 times. The central limit theorem provides an estimate for this probability: here $p = .5$, so

$$P_{100}(S_{100} \geq 60) = P_{100}\left(\frac{S_{100} - 50}{5} \geq 2\right) \approx \Phi(2) \approx .02.$$

[4]W. Feller, *Introduction to Probability Theory and its Applications*, volume 1, chap. VII.3.

In the same way, we can estimate the probability that the number of heads is at least 540 after 1000 coin tosses:

$$P_{1000}(S_{1000} \geq 540) = P_{1000}\left(\frac{S_{1000} - 500}{5\sqrt{10}} \geq \frac{8}{\sqrt{10}}\right)$$
$$\approx \Phi\left(\frac{8}{\sqrt{10}}\right) \approx .006.$$

Now, here is an example which involves two concepts from statistics: confidence intervals and hypothesis tests. (However, we will not give precise definitions for these concepts.)

Between 1871 and 1900, 1,359,670 boys and 1,285,086 girls were born in Switzerland. Are these numbers compatible with the hypothesis that the genders of newborns are independent random characteristics and that the two genders appear with equal probability?

Set $n = 1,359,670 + 1,285,086 = 2,644,756$ and suppose that n trials were performed. Without the intention of appearing sexist, we will call the birth of a boy *success* and the birth of a girl *failure*. Supposing for the moment that the n trials are independent and the probability of success is 1/2, we will determine a number C such that the probability that the number of successes is greater than C is less than 10^{-5}.

Since $\Phi(4.5) < 10^{-5}$ and

$$P_n(S_n \geq C) \approx \Phi\left(\frac{2C - n}{\sqrt{n}}\right)$$

by the central limit theorem, we can choose $C \geq \frac{1}{2}(n + 4.5\sqrt{n})$. This yields $C = 1,326,037$.

The number of successes (or male births) is actually much higher than this number. If our hypothesis about the distribution of births were true, the figures observed in Switzerland would be highly improbable. Therefore, it is reasonable to reject the hypothesis. We conclude that our demographic figures contradict the assumption that the gender of newborns is an independent random characteristic with uniform probability distribution.

Remark. One might be bothered by the use of the central limit theorem, a purely asymptotic result, in this example; however, this is how the theorem is used in practice. Nonetheless, we note that the large deviations estimate provides a similar result for this last example: in this case, we write the estimation as

$$P_n(S_n \geq C) \leq \exp\left(-nh_+\left(\frac{C}{n} - \frac{1}{2}\right)\right),$$

which yields

$$P_n(S_n \geq C) \leq 5 \cdot 10^{-5}$$

for $n = 2,644,756$ and $C = 1,326,037$.

7.4. Proof of the Theorem

The proof is composed of several steps. The first step, which is also the most important ingredient of the proof, is Stirling's formula. This formula provides a sequence that is asymptotically equal to the sequence $(n!)$ and thus allows us to estimate the binomial coefficients $\binom{n}{k}$ for large n, k, and $n - k$. The proof will continue with the estimation of the probability that S_n is within a certain interval $[np + a\sqrt{p(1-p)}\sqrt{n}, np + b\sqrt{p(1-p)}\sqrt{n}]$. To do this, we will estimate $P_n(S_n = k)$ for large n and integer k within the given interval with help from the de Moivre–Laplace theorem (in Chapter 9, we will discuss the *local limit theorem*, which is a generalized form of this theorem). At the last step, we will attain the central limit theorem by using Riemann sums.

The proof that we provide uses only elementary tools of real analysis. Another proof uses the Fourier transform of measures, which is called the *characteristic function* in probability theory (this is distinct from the characteristic function introduced in Chapter 1). Although this Fourier analysis method can be used to prove the central limit theorem in much more general settings than the one we are studying, we will not present it here because it is not elementary.

Proposition 7.2 (Stirling's formula). *For each integer $n > 0$, set*

$$n! = \sqrt{2\pi}n^{n+1/2}e^{-n}(1 + \epsilon_n).$$

There exists a real constant A such that $|\epsilon_n| < \frac{A}{n}$.

Proof. First, we will show that there exists a $c_1 \in \mathbb{R}$ such that

$$\ln(n!) = c_1 + \left(n + \frac{1}{2}\right)\ln n - n + O\left(\frac{1}{n}\right).$$

This estimate is based on a comparison of the series with general term $\ln n$ to the logarithmic integral. We write

$$\ln(n!) = \int_{1/2}^{n+1/2} \ln t\, dt + \sum_{k=1}^{n}\left(\ln k - \int_{k-1/2}^{k+1/2} \ln t\, dt\right). \tag{7.2}$$

On one hand,

$$\int_{1/2}^{n+1/2} \ln t\, dt = [t\ln t - t]_{1/2}^{n+1/2} = \left(n + \frac{1}{2}\right)\ln\left(n + \frac{1}{2}\right) - n + c_2$$

for some constant c_2. The Taylor series for the logarithm function about 1 yields

$$\ln\left(n + \frac{1}{2}\right) = \ln(n) + \ln\left(1 + \frac{1}{2n}\right) = \ln(n) + \frac{1}{2n} + O\left(\frac{1}{n^2}\right),$$

which yields

$$\int_{1/2}^{n+1/2} \ln t\, dt = \left(n + \frac{1}{2}\right)\ln(n) - n + c_3 + O\left(\frac{1}{n}\right) \tag{7.3}$$

for a constant c_3.

On the other hand,

$$\begin{aligned}\ln k - \int_{k-1/2}^{k+1/2} \ln t\, dt &= \ln k - [t\ln t - t]_{k-1/2}^{k+1/2} \\ &= \ln k - \left(k + \frac{1}{2}\right) ln\left(k + \frac{1}{2}\right) + \left(k - \frac{1}{2}\right) ln\left(k - \frac{1}{2}\right) + 1 \\ &= -\frac{1}{2}\ln\left(1 - \frac{1}{4k^2}\right) - k\left(\ln\left(1 + \frac{1}{2k}\right) - \ln\left(1 - \frac{1}{2k}\right)\right) + 1.\end{aligned}$$

The Taylor series expansion of the logarithm function allows us to write this as

$$\begin{aligned}\ln k - \int_{k-1/2}^{k+1/2} \ln t\, dt = O\left(\frac{1}{k^2}\right) - k\Bigg[&\left(\frac{1}{2k} - \frac{1}{8k^2} + O\left(\frac{1}{k^3}\right)\right) \\ &- \left(-\frac{1}{2k} - \frac{1}{8k^2} + O\left(\frac{1}{k^3}\right)\right)\Bigg] + 1,\end{aligned}$$

which shows that there exists a constant $c_4 \in \mathbb{R}$ such that

$$\left| \ln k - \int_{k-1/2}^{k+1/2} \ln t \, dt \right| \leq \frac{c_4}{k^2}.$$

Thus $\left(\ln k - \int_{k-1/2}^{k+1/2} \ln t \, dt\right)$ is the general term of an absolutely convergent series. Let

$$c_5 := \sum_{k=1}^{+\infty} \left(\ln k - \int_{k-1/2}^{k+1/2} \ln t \, dt \right).$$

Now, note that

$$\begin{aligned} \sum_{k=n+1}^{+\infty} \left| \ln k - \int_{k-1/2}^{k+1/2} \ln t \, dt \right| &\leq c_4 \sum_{k=n+1}^{+\infty} \frac{1}{k^2} \\ &< c_4 \sum_{k=n+1}^{+\infty} \frac{1}{k(k-1)} = \frac{c_4}{n}, \end{aligned}$$

which proves that

$$\sum_{k=1}^{n} \left(\ln k - \int_{k-1/2}^{k+1/2} \ln t \, dt \right) = c_5 + O\left(\frac{1}{n}\right). \tag{7.4}$$

From (7.2), (7.3), and (7.4), we obtain

$$\ln(n!) = \left(n + \frac{1}{2}\right) \ln(n) - n + c_3 + c_5 + O\left(\frac{1}{n}\right),$$

which completes our first step. By writing $d := e^{c_3 + c_5}$, we have obtained

$$n! = d n^{n+1/2} e^{-n} (1 + \epsilon_n),$$

where $\epsilon_n = O\left(\frac{1}{n}\right)$.

To complete our proof, we must show that $d = \sqrt{2\pi}$. To do this, we will use Wallis integrals, which are defined for each $n \in \mathbb{N}$ by

$$I_n := \int_0^{\frac{\pi}{2}} \sin^n t \, dt.$$

Integration by parts yields

$$I_n = \int_0^{\frac{\pi}{2}} \sin^{n-1} t \cdot \sin t \, dt = \int_0^{\frac{\pi}{2}} (n-1) \sin^{n-2} t \, \cos t \cdot \cos t \, dt$$

for each $n \geq 2$. Thus

$$I_n = (n-1) \int_0^{\frac{\pi}{2}} \sin^{n-2} t \ (1 - \sin^2 t) \ dt,$$

which yields the recurrence

$$nI_n = (n-1)I_{n-2}.$$

Since $I_0 = \frac{\pi}{2}$ and $I_1 = 1$, for each $n \geq 1$, we can write

$$I_{2n} = \frac{(2n-1)(2n-3)\cdots 1}{(2n)(2n-2)\cdots 2} I_0 = \frac{(2n)!}{2^{2n}(n!)^2}\frac{\pi}{2}$$

and

$$I_{2n+1} = \frac{(2n)(2n-2)\cdots 2}{(2n+1)(2n-1)\cdots 3} I_1 = \frac{2^{2n}(n!)^2}{(2n+1)!}.$$

Now, note that the recurrence relation yields

$$\lim_{n\to\infty} \frac{I_n}{I_{n-2}} = 1.$$

Since $I_{n-2} \geq I_{n-1} \geq I_n$, this implies that

$$\lim_{n\to\infty} \frac{I_n}{I_{n-1}} = 1,$$

which can be written as

$$\lim_{n\to\infty} \frac{((2n)!)^2(2n+1)}{2^{4n}(n!)^4}\frac{\pi}{2} = 1.$$

From the first step of our proof, we know that $n! \sim dn^{n+1/2}e^{-n}$ as n approaches infinity. By replacing the factorials in the above limit with the expressions given by Stirling's formula, we conclude that $d^2 = 2\pi$. This completes the proof. □

Proposition 7.3 (de Moivre–Laplace theorem). *For each* $0 \leq k \leq n$, *set*

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{1}{\sqrt{2\pi p(1-p)n}} \exp\left(-\frac{(k-np)^2}{2np(1-p)}\right)(1+\delta_n(k)).$$

Then, for each $a > 0$,

$$\lim_{n\to\infty} \max_{k:|k-np|<a\sqrt{n}} |\delta_n(k)| = 0.$$

Proof. Fix a real number $a > 0$ and let I_n be the set of integers between $np - a\sqrt{n}$ and $np + a\sqrt{n}$. Let $(s_{n,k})_{n>0,k\in I_n}$ and $(t_n)_n > 0$ be two families of real numbers. We will write $s_{n,k} = O_u(t_n)$ if there exists a real constant C such that $|s_{n,k}| < Ct_n$ for every $n > 0$ and $k \in I_n$ (The subscript u of O indicates that the estimate is uniform in k.)

Stirling's formula implies that

(7.5)
$$\binom{n}{k} p^k(1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k(1-p)^{n-k}$$
$$= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \left(\frac{n}{k}p\right)^k \left(\frac{n}{n-k}(1-p)\right)^{n-k} \frac{1+\epsilon_n}{(1+\epsilon_k)(1+\epsilon_{n-k})}.$$

If $k \in I_n$, then

(7.6)
$$\frac{n}{(np+a\sqrt{n})(n(1-p)+a\sqrt{n})} \leq \frac{n}{k(n-k)}$$
$$\leq \frac{n}{(np-a\sqrt{n})(n(1-p)-a\sqrt{n})}.$$

From this we conclude that

$$\frac{n}{k(n-k)} = \frac{1}{np(1-p)}\left(1 + O_u\left(n^{-1/2}\right)\right), \tag{7.7}$$

and thus

$$\sqrt{\frac{n}{k(n-k)}} = \frac{1}{\sqrt{np(1-p)}}\left(1 + O_u\left(n^{-1/2}\right)\right). \tag{7.8}$$

Now, by using the Taylor series $\ln(1+t) = t - \frac{t^2}{2} + O(t^3)$, equation (7.7), and the facts that $\frac{k-np}{k} = O_u(n^{-1/2})$ and $\frac{k-np}{n-k} = O_u(n^{-1/2})$, we can write

$$
\begin{aligned}
&\ln\left(\left(\frac{n}{k}p\right)^k\left(\frac{n}{n-k}(1-p)\right)^{n-k}\right)\\
&= k\ln\left(1-\frac{k-np}{k}\right)+(n-k)\ln\left(1+\frac{k-np}{n-k}\right)\\
&= -\frac{1}{2}(k-np)^2\left(\frac{1}{k}+\frac{1}{n-k}\right)+kO_u\left(n^{-3/2}\right)+(n-k)O_u\left(n^{-3/2}\right)\\
&= -\frac{1}{2}(k-np)^2\frac{1}{np(1-p)}+O_u(n^{-1/2}).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
(7.9)\qquad &\left(\frac{n}{k}p\right)^k\left(\frac{n}{n-k}(1-p)\right)^{n-k}\\
&= \exp\left(-\frac{(k-np)^2}{2np(1-p)}\right)\left(1+O_u\left(n^{-1/2}\right)\right).
\end{aligned}
$$

Finally, since

$$
\epsilon_n < \frac{A}{n},\qquad \epsilon_k < \frac{A}{k},\qquad \epsilon_{n-k} < \frac{A}{n-k},
$$

$$
\frac{1}{k} = O_u\left(\frac{1}{n}\right),\qquad \frac{1}{n-k} = O_u\left(\frac{1}{n}\right),
$$

we have that

$$
(7.10)\qquad \frac{1+\epsilon_n}{(1+\epsilon_k)(1+\epsilon_{n-k})} = \left(1+O_u\left(\frac{1}{n}\right)\right).
$$

By combining the estimates (7.8), (7.9), and (7.10) in the identity (7.5), we obtain

$$
\begin{aligned}
&\binom{n}{k}p^k(1-p)^{n-k}\\
&= \frac{1}{\sqrt{2\pi p(1-p)n}}\exp\left(-\frac{(k-np)^2}{2np(1-p)}\right)\left(1+O_u\left(n^{-1/2}\right)\right).
\end{aligned}
$$

□

We will use the following lemmas to deduce the central limit theorem from the de Moivre–Laplace theorem.

Lemma 7.4. *Let $[a,b]$ be an interval of $\mathbb{R}$, and let f be a function defined on $\mathbb{R}$ that is zero outside of $[a,b]$ and continuous on $[a,b]$. Then*

$$\lim_{h\to 0, h>0} h \sum_{k=-\infty}^{+\infty} f(t+kh) = \int_a^b f(x)\ dx$$

uniformly in $t \in \mathbb{R}$.

Proof. Note that the sum appearing in the statement of this lemma is nearly a Riemann sum for $f(x)$. The result follows easily from the uniform continuity of the function f on the interval $[a,b]$. Fix $\epsilon > 0$ and choose h small enough so that

$$|f(x) - f(y)| < \epsilon \quad \text{whenever } x, y \in [a,b] \text{ and } |x-y| < h.$$

Letting

$$\{k \in \mathbb{Z} \mid a \le t + kh \le b\} = \{i, i+1, i+2, \ldots, j\}$$

and

$$M := \sup_{a \le x \le b} |f(x)|,$$

we have that

$$\begin{aligned}
&\left| h \sum_{k:t+kh\in[a,b]} f(t+kh) - \int_a^b f(x)\ dx \right| \\
&\quad \le hM + \sum_{k=i}^{j-1} \left| hf(t+kh) - \int_{t+kh}^{t+(k+1)h} f(x)\ dx \right| + 2hM \\
&\quad \le 3hM + (j-i)h\epsilon \le 3hM + (b-a)\epsilon.
\end{aligned}$$

□

The following lemma is well known; we leave its proof as an exercise for the reader.

Lemma 7.5.

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2}\ dx = 1.$$

Proof of the Central Limit Theorem. We start with the case in which a and b are two real numbers. Let K_n be the interval

$$[a\sqrt{np(1-p)}, b\sqrt{np(1-p)}].$$

We have

$$\begin{aligned} P_n(S_n - np \in K_n) &= \sum_{k=0}^{n} \chi_{K_n}(k-np) \cdot P_n(S_n = k) \\ &= \frac{1}{\sqrt{2\pi p(1-p)n}} \sum_{k=0}^{n} \Bigg[\chi_{K_n}(k-np) \\ &\qquad \cdot \exp\left(-\frac{(k-np)^2}{2np(1-p)}\right) \cdot (1+\delta_n(k)) \Bigg], \end{aligned}$$

where $\lim_{n\to\infty} \max_k |\delta_n(k)| = 0$ by Proposition 7.3, since $k - np$ is always in K_n. It follows that

$$\begin{aligned} (7.11)\quad & P_n(S_n - np \in K_n) \\ &= \frac{1+\delta_n}{\sqrt{2\pi p(1-p)n}} \sum_{k=0}^{n} \chi_{K_n}(k-np) \cdot \exp\left(-\frac{(k-np)^2}{2np(1-p)}\right), \end{aligned}$$

where $\lim_{n\to\infty} \delta_n = 0$. When n is large enough, the expression (7.11) equals the expression obtained by replacing $\sum_{k=0}^{n}$ by $\sum_{k\in\mathbb{Z}}$. This expression is equivalent to

(7.12)

$$\begin{aligned} &\frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{p(1-p)n}} \\ &\quad \times \sum_{k\in\mathbb{Z}} \Bigg[\chi_{[a,b]}\left(\frac{k}{\sqrt{np(1-p)}} - \sqrt{\frac{np}{1-p}}\right) \\ &\qquad\qquad \times \exp\left(-\frac{1}{2}\left(\frac{k}{\sqrt{np(1-p)}} - \sqrt{\frac{np}{1-p}}\right)^2\right)\Bigg]. \end{aligned}$$

By setting $h = (np(1-p))^{-1/2}$ and $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, we see that this expression has the form of the one in Lemma 7.4. Therefore, the expression (7.12) approaches $\frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx$ as n approaches infinity.

This proves the theorem for real a and b. Lemma 7.5 allows us to treat the case where $a = -\infty$ by proving that

$$(7.13) \qquad P_n\left(S_n - np \le b\sqrt{np(1-p)}\right) \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b} e^{-x^2/2}\, dx$$

as n approaches infinity for all real b. The case where $b = +\infty$ is treated similarly.

Let $b \in \mathbb{R}$ and $\epsilon > 0$. Fix $c > \max(0, b)$ such that

$$\frac{1}{\sqrt{2\pi}} \int_{c}^{+\infty} e^{-x^2/2}\, dx < \epsilon.$$

Then $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-c} e^{-x^2/2}\, dx < \epsilon$, and Lemma 7.5 implies that

$$\frac{1}{\sqrt{2\pi}} \int_{-c}^{c} e^{-x^2/2}\, dx > 1 - 2\epsilon.$$

We can write

$$\left| P_n\left(S_n - np \le b\sqrt{np(1-p)}\right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b} e^{-x^2/2}\, dx \right| \le A_n + B_n + C,$$

where

$$A_n := P_n\left(S_n - np < -c\sqrt{np(1-p)}\right),$$

$$B_n := \left| P_n\left(-c \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right) - \frac{1}{\sqrt{2\pi}} \int_{-c}^{b} e^{-x^2/2} dx \right|,$$

$$C := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-c} e^{-x^2/2}\, dx.$$

We have that

$$0 \le A_n \le 1 - P_n\left(-c\sqrt{np(1-p)} \le S_n - np \le c\sqrt{np(1-p)}\right),$$

and, as in the first part of the proof,

$$\lim_{n\to\infty} P_n\left(-c\sqrt{np(1-p)} \le S_n - np \le c\sqrt{np(1-p)}\right)$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-c}^{c} e^{-x^2/2}\, dx > 1 - 2\epsilon.$$

This shows that $A_n < 2\epsilon$ for large enough n. Similarly, $\lim_{n\to\infty} B_n = 0$, and $C < \epsilon$ by our choice of c. This proves (7.13). □

http://dx.doi.org/10.1090/stml/028/09

Chapter 8

The Moderate Deviations Estimate

We will continue using the notation introduced in the previous chapters. Thus S_n denotes the number of successes observed in a sequence of n independent trials of a probabilistic experiment with success p. The empirical probability of success is S_n/n. The estimation of large deviations tells us the size of the probability that the empirical probability deviates from p as n tends to infinity. In addition, the central limit theorem tells us that if (a_n) is a real sequence that approaches infinity with n, then the probability that the empirical probability deviates from p by about $a_n/\sqrt{n}$ approaches 0 as n approaches infinity. Under certain hypotheses about the sequence (a_n), the moderate deviations estimate will provide more detail about this convergence. To obtain our estimate, we will use a generalization of the central limit theorem to a form with *variable bounds*. In Chapter 12 we will use the moderate deviations estimate to establish the famous law of the iterated logarithm.

What kind of estimate can we hope for? The central limit theorem tells us that as n approaches infinity,

$$P_n\left(\frac{S_n}{n} - p \geq \sqrt{p(1-p)}\frac{a}{\sqrt{n}}\right) \sim \Phi(a).$$

The *moderate deviations estimate* asserts that this remains true when a is allowed to approach infinity with n at a slow enough rate. Recall that if $\lim_{n\to\infty} a_n = +\infty$, then

$$\Phi(a_n) = \frac{1}{\sqrt{2\pi}} \int_{a_n}^{+\infty} e^{-x^2/2} dx \sim \frac{1}{a_n\sqrt{2\pi}} \exp\left(-\frac{a_n^2}{2}\right).$$

The following theorem first appeared in an article by Harald Cramér.[1]

Theorem 8.1 (moderate deviations estimate). *Suppose that* (a_n) *is a sequence of real numbers such that* $\lim_{n\to\infty} a_n = +\infty$ *and* $\lim_{n\to\infty} a_n n^{-1/6} = 0$. *Then*

$$P_n\left(\frac{S_n}{n} - p \geq \sqrt{p(1-p)}\frac{a_n}{\sqrt{n}}\right) \sim \frac{1}{a_n\sqrt{2\pi}} \exp\left(-\frac{a_n^2}{2}\right).$$

Note that the theorem's conclusion does not hold if the sequence (a_n) approaches infinity too quickly. For example (taking $a_n = \sqrt{n}$), the sequences $(\Phi(\sqrt{n})$ and $\left(P_n\left(\frac{S_n}{n} - p \geq \sqrt{p(1-p)}\right)\right)$ are not equivalent. In fact,

$$\lim \frac{1}{n} \ln \Phi(\sqrt{n}) = -\frac{1}{2},$$

but if $p < \frac{1}{2}$, then $\sqrt{p(1-p)} < 1 - p$ and

$$\lim \frac{1}{n} \ln P_n\left(\frac{S_n}{n} - p \geq \sqrt{p(1-p)}\right) = -h_+\left(\sqrt{p(1-p)}\right)$$

by Proposition 6.3.

We will prove Theorem 8.1 by using a refinement of the techniques we used to prove the central limit theorem. The first step is to optimize the proof of the de Moivre–Laplace theorem (Proposition 7.3). This theorem gives a sequence that is asymptotically equal to $P_n(S_n = k_n)$ as n tends to infinity, where $k_n = np + O(\sqrt{n})$. The careful reader will realize that the theorem's proof can be extended to the case where $k_n = np + o(n^{2/3})$. This extension of the de Moivre–Laplace theorem is the goal of the following proposition.

---

[1] H. Cramér, *Sur un nouveau théorème-limite de la théorie des probabilités*, Acualités Scientifiques et Industrielles, vol. 736, pp. 5–23, Hermann, Paris, 1938.

Proposition 8.2. *For $0 \le k \le n$, set*

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{1}{\sqrt{2\pi p(1-p)n}} \exp\left(-\frac{(k-np)^2}{2np(1-p)}\right) \cdot (1+\delta_n(k)).$$

Then for every positive real sequence (c_n) approaching zero,

$$\lim_{n\to\infty} \max_{k:|k-np|<c_n n^{2/3}} |\delta_n(k)| = 0.$$

Proof. Let $J_n{}'$ be the set of integers k such that $|k-np| \le c_n n^{2/3}$. For each $k \in J_n{}'$, formula (7.6) becomes

$$\begin{aligned} \frac{n}{(np+c_n n^{2/3})(n(1-p)+c_n n^{2/3})} &\le \frac{n}{k(n-k)} \\ &\le \frac{n}{(np-c_n n^{2/3})(n(1-p)-c_n n^{2/3})}. \end{aligned}$$

We deduce that

$$\frac{n}{k(n-k)} = \frac{1}{np(1-p)}\left(1+O_u\left(c_n n^{-1/3}\right)\right),$$

and thus

$$\sqrt{\frac{n}{k(n-k)}} = \frac{1}{\sqrt{np(1-p)}}\left(1+O_u\left(c_n n^{-1/3}\right)\right). \tag{8.1}$$

Now, since, $\frac{k-np}{k} = O_u(c_n n^{-1/3})$ and $\frac{k-np}{n-k} = O_u(c_n n^{-1/3})$, the Taylor series for the logarithm function yields

$$\begin{aligned} &\ln\left(\left(\frac{n}{k}p\right)^k \left(\frac{n}{n-k}(1-p)\right)^{n-k}\right) \\ &= -\frac{1}{2}(k-np)^2\left(\frac{1}{k}+\frac{1}{n-k}\right) + kO_u\left(c_n^3 n^{-1}\right) + (n-k)O_u\left(c_n^3 n^{-1}\right) \\ &= -\frac{1}{2}(k-np)^2 \frac{1}{np(1-p)} + O_u(c_n^3). \end{aligned}$$

Hence,

$$\begin{aligned} &\left(\frac{n}{k}p\right)^k \left(\frac{n}{n-k}(1-p)\right)^{n-k} \\ &\qquad = \exp\left(-\frac{(k-np)^2}{2np(1-p)}\right)\left(1+O_u\left(c_n^3\right)\right). \end{aligned} \tag{8.2}$$

Finally, as in the proof of Proposition 7.3,

$$\frac{1+\epsilon_n}{(1+\epsilon_k)(1+\epsilon_{n-k})} = \left(1+O_u\left(\frac{1}{n}\right)\right). \tag{8.3}$$

By combining the estimates (8.1), (8.2), and (8.3) in the identity (7.5), we obtain

$$\binom{n}{k}p^k(1-p)^{n-k}$$
$$= \frac{1}{\sqrt{2\pi p(1-p)n}}\exp\left(-\frac{(k-np)^2}{2np(1-p)}\right)\cdot\left(1+O_u\left(c_n'\right)\right),$$

where $c_n' := \max\left(c_n n^{-1/3}, c_n^3, n^{-1}\right)$. □

By following the steps we used to obtain the central limit theorem from the de Moivre–Laplace theorem, we will obtain a version of the central limit theorem with variable bounds from Proposition 8.2.

Proposition 8.3. *Let (k_n) and (ℓ_n) be two integer sequences such that $k_n < \ell_n$ and $k_n, \ell_n = np + o\left(n^{2/3}\right)$ as n approaches infinity. Set $a_n := \frac{k_n - np}{\sqrt{np(1-p)}}$ and $b_n := \frac{\ell_n - np}{\sqrt{np(1-p)}}$. Then*

$$P_n\left(k_n \le S_n < \ell_n\right) \sim \frac{1}{\sqrt{2\pi}}\int_{a_n}^{b_n} e^{-x^2/2}\,dx$$

as n approaches infinity.

If the sequences (a_n) and (b_n) converge respectively to a and b such that $a < b$, this theorem becomes the central limit theorem.

Proof. Set $h(n) := \frac{1}{\sqrt{np(1-p)}}$. We will only consider integers n that are large enough so that $0 \le k_n < \ell_n \le n$. Write

$$P_n(S_n = j) = \frac{h(n)}{\sqrt{2\pi}}\exp\left(-\frac{(j-np)^2}{2np(1-p)}\right)\cdot\left(1+\delta_n(j)\right)$$

and

$$P_n\left(k_n \le S_n < \ell_n\right) = \frac{h(n)}{\sqrt{2\pi}}\sum_{j=k_n}^{\ell_n - 1}\exp\left(-\frac{(j-np)^2}{2np(1-p)}\right)\cdot\left(1+\delta_n(j)\right).$$

The hypotheses on the sequences (k_n) and (ℓ_n) along with Proposition 8.2 imply that the sequence $(\delta_n(j))_{n\geq 1}$ converges uniformly to zero when $k_n \leq j \leq \ell_n$. Therefore, we just need to show that

$$(8.4) \qquad h(n) \sum_{j=k_n}^{\ell_n - 1} \exp\left(-\frac{(j-np)^2}{2np(1-p)}\right) \sim \int_{a_n}^{b_n} e^{-x^2/2}\, dx.$$

This follows easily by considering the Riemann sum of $e^{-x^2/2}$. Set

$$x(j) := \frac{j - np}{\sqrt{np(1-p)}}.$$

Then $a_n = x(k_n)$ and $b_n = x(\ell_n)$. We can rewrite formula (8.4) as

$$(8.5) \qquad h(n) \sum_{j=k_n}^{\ell_n - 1} \exp\left(-\frac{x(j)^2}{2}\right) - \int_{a_n}^{b_n} e^{-x^2/2}\, dx = o\left(\int_{a_n}^{b_n} e^{-x^2/2}\, dx\right).$$

Suppose for the moment that $a_n \geq 0$. By noting that

$$h(n) \exp\left(-\frac{x(j+1)^2}{2}\right) < \int_{x(j)}^{x(j+1)} e^{-x^2/2}\, dx < h(n) \exp\left(-\frac{x(j)^2}{2}\right)$$

for $k_n \leq j < \ell_n$, we obtain

$$0 \leq h(n) \sum_{j=k_n}^{\ell_n - 1} e^{-x(j)^2/2} - \int_{a_n}^{b_n} e^{-x^2/2}\, dx \leq h(n)\left(e^{-a_n^2/2} - e^{-b_n^2/2}\right).$$

We also know that

$$\int_{a_n}^{b_n} e^{-x^2/2}\, dx \geq \frac{1}{b_n} \int_{a_n}^{b_n} x e^{-x^2/2}\, dx = \frac{1}{b_n}\left(e^{-a_n^2/2} - e^{-b_n^2/2}\right).$$

Now, $h(n) = o(b_n^{-1})$ since $b_n = o(n^{1/6})$. Thus, the last two inequalities yield (8.5).

We have proved the theorem in the case where $a_n \geq 0$. By symmetry, the case where $b_n \leq 0$ is proven identically. In the case where $a_n < 0 < b_n$, we can use the same argument by separating the comparison of the integral with the discrete sum into three parts: one part for each of the intervals $\mathbb{R}^+$ and $\mathbb{R}^-$ and one part for the potential term at zero. We leave the details to the reader. □

Proof of the Moderate Deviations Estimate. Since the limit of $a_n/n^{1/6}$ as n approaches infinity is 0, (a_n) approaches infinity less quickly than $(n^{1/6})$ does. Consider a real sequence (d_n) such that $d_n = o(a_n)$ and $d_n \geq \sqrt{2\ln a_n}$ (for example, we could choose $d_n = \ln a_n$ or $d_n = \sqrt{a_n}$). For each n, let k_n be the least integer that is greater than or equal to $np + \sqrt{np(1-p)}a_n$, and let ℓ_n be the least integer that is greater than or equal to $np + \sqrt{np(1-p)}(a_n + d_n)$. We have

$$P_n\left(\frac{S_n}{n} - p \geq \sqrt{p(1-p)}\frac{a_n}{\sqrt{n}}\right) = P_n\left(S_n \geq k_n\right),$$

so

$$P_n\left(\frac{S_n}{n} - p \geq \sqrt{p(1-p)}\frac{a_n}{\sqrt{n}}\right) = P_n\left(k_n \leq S_n < \ell_n\right) + P_n\left(S_n \geq \ell_n\right).$$

From the fact that $a_n = o(n^{1/6})$, we deduce that $k_n, \ell_n = np + o\left(n^{2/3}\right)$, so the hypotheses of Proposition 8.3 are satisfied. Setting $a'_n := \frac{k_n - np}{\sqrt{np(1-p)}}$ and $b_n := \frac{\ell_n - np}{\sqrt{np(1-p)}}$, yields

$$P_n\left(k_n \leq S_n < \ell_n\right) \sim \frac{1}{\sqrt{2\pi}}\int_{a'_n}^{b_n} e^{-x^2/2}\,dx,$$

so

$$P_n\left(k_n \leq S_n < \ell_n\right) \sim \frac{1}{\sqrt{2\pi}}\int_{a_n}^{b_n} e^{-x^2/2}\,dx - \frac{1}{\sqrt{2\pi}}\int_{a_n}^{a'_n} e^{-x^2/2}\,dx.$$

To finish the proof, we just need to show that

$$\int_{a_n}^{b_n} e^{-x^2/2}\,dx \sim \frac{1}{a_n}\exp\left(-\frac{a_n^2}{2}\right), \tag{8.6}$$

$$\int_{a_n}^{a'_n} e^{-x^2/2}\,dx = o\left(\frac{1}{a_n}\exp\left(-\frac{a_n^2}{2}\right)\right), \tag{8.7}$$

and

$$P_n\left(S_n \geq \ell_n\right) = o\left(\frac{1}{a_n}\exp\left(-\frac{a_n^2}{2}\right)\right). \tag{8.8}$$

We have

$$\int_{a_n}^{b_n} e^{-x^2/2}\,dx \leq \frac{1}{a_n}\int_{a_n}^{+\infty} xe^{-x^2/2}\,dx = \frac{1}{a_n}\exp\left(-\frac{a_n^2}{2}\right),$$

and, since $b_n \geq a_n + d_n$,

$$\begin{aligned}\int_{a_n}^{b_n} e^{-x^2/2}\,dx &\geq \int_{a_n}^{a_n+d_n} e^{-x^2/2}\,dx \\ &\geq \frac{1}{a_n+d_n}\int_{a_n}^{a_n+d_n} xe^{-x^2/2}\,dx \\ &= \frac{1}{a_n+d_n}\left(\exp\left(-\frac{a_n^2}{2}\right) - \exp\left(-\frac{(a_n+d_n)^2}{2}\right)\right) \\ &\sim \frac{1}{a_n}\exp\left(-\frac{a_n^2}{2}\right).\end{aligned}$$

This proves (8.6).

The fact that $0 \leq a_n' - a_n \leq (np(1-p))^{-1/2}$ directly implies that

$$\int_{a_n}^{a_n'} e^{-x^2/2}\,dx \leq \frac{1}{\sqrt{np(1-p)}}\exp\left(-\frac{a_n^2}{2}\right),$$

which in turn implies (8.7) because $a_n = o(\sqrt{n})$.

The last estimate (8.8) follows from the large deviations estimate (Theorem 6.1). We have

$$P_n\left(S_n \geq \ell_n\right) \leq \exp\left(-nh_+\left(\sqrt{p(1-p)}\frac{b_n}{\sqrt{n}}\right)\right).$$

Now, $h_+(\epsilon) = \frac{\epsilon^2}{2p(1-p)} + O\left(\epsilon^3\right)$ when ϵ is close to 0, so

$$P_n\left(S_n \geq \ell_n\right) \leq \exp\left(-\frac{b_n^2}{2} + O\left(\frac{b_n^3}{\sqrt{n}}\right)\right) \sim \exp\left(-\frac{b_n^2}{2}\right)$$

since $b_n = o(n^{1/6})$. Finally,

$$\begin{aligned}\exp\left(-\frac{b_n^2}{2}\right) &\leq \exp\left(-\frac{(a_n+d_n)^2}{2}\right) \\ &= o\left(\exp\left(-\frac{d_n^2}{2}\right)\exp\left(-\frac{a_n^2}{2}\right)\right).\end{aligned}$$

Now, $\exp\left(-\frac{d_n^2}{2}\right) \leq \frac{1}{a_n}$ by the choice of d_n, and (8.8) follows. □

This page intentionally left blank

http://dx.doi.org/10.1090/stml/028/10

Chapter 9

The Local Limit Theorem

We will use the notation introduced in the previous chapters; in particular, S_n is the number of successes observed in a sequence of n independent trials of an experiment with probability of success p. For each k between 0 and n, we know that

$$P_n(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

The weak law of large numbers and the central limit theorem describe the behavior of the sequence (S_n) and the size of the fluctuations around the average value np. We will return to these questions with the strong law of large numbers and the law of the iterated logarithm.

In this chapter, we will study the size of $P_n(S_n = k)$ as n approaches infinity. We will describe two estimates, each uniform in k, that are two forms of the *local limit theorem.* Consider the random variable $M_n := 2S_n - n$ based on the sequence of random coin tosses. This sequence (M_n) describes, for example, the fortune held by a player who receives 1 at each outcome of success and loses 1 at each outcome of failure. (The sequence (M_n) will be studied further in the chapters about the arcsine law and the recurrence of random walks.)

The first form of the local limit theorem is useful for estimating the probability that M_n takes a given value close to its average value

$(2p-1)n$. The second form is useful for estimating the probability that M_n takes a fixed value.

Theorem 9.1 (local limit theorem, first form).

$$P_n(S_n = k) = \frac{1}{\sqrt{2\pi p(1-p)n}}\left(\exp\left(-\frac{(k-np)^2}{2np(1-p)}\right) + o(1)\right)$$

uniformly in $k \in \mathbb{Z}$.

Theorem 9.2 (local limit theorem, second form).

$$P_n(M_n = k)$$
$$= \sqrt{\frac{2}{\pi}}\left(\frac{p}{1-p}\right)^{k/2}\frac{\left(2\sqrt{p(1-p)}\right)^n}{\sqrt{n}}\left(\exp\left(-\frac{k^2}{2n}\right) + o(1)\right)$$

uniformly in $k \in \mathbb{Z}$ *such that* $n+k$ *is even.*

Note that these two statements are equivalent for the *centered* case (that is, when $p = \frac{1}{2}$). On the other hand, if $p \neq \frac{1}{2}$, then $2\sqrt{p(1-p)} < 1$ and the second theorem says that the probability that M_n is in a fixed finite subset of $\mathbb{Z}$ decreases exponentially as n approaches infinity. We can write this as the following corollary.

Corollary 9.3. *Let* K *be a fixed finite subset of* $\mathbb{Z}$ *that contains at least one even number and one odd number. Then*

$$\lim_{n\to\infty}\frac{1}{n}\ln\left(P_n(M_n \in K)\right) = \ln\left(2\sqrt{p(1-p)}\right).$$

Proof of Theorem 9.1. Fix a real t such that $\frac{1}{2} < t < \frac{2}{3}$. Proposition 8.2 implies that

$$\frac{1}{\sqrt{2\pi p(1-p)n}}\exp\left(-\frac{(k-np)^2}{2np(1-p)}\right)\cdot\left(1+\delta_n(k)\right),$$

where

$$\lim_{n\to\infty}\max_{k:|k-np|<n^t}|\delta_n(k)| = 0.$$

Since the exponential term is at most 1, we have

$$(9.1)\qquad P_n(S_n = k)$$
$$= \frac{1}{\sqrt{2\pi p(1-p)n}}\exp\left(-\frac{(k-np)^2}{2np(1-p)}\right) + o\left(n^{-1/2}\right),$$

where the estimate $o(n^{-1/2})$ is uniform in k when $k - np < n^t$.

On the other hand, we can apply the moderate deviations estimate (Theorem 8.1) to the sequence $a_n = n^{t-\frac{1}{2}}$ to obtain

$$P_n\left(S_n > np + n^t\right) = P_n\left(\frac{S_n}{n} - p > \frac{n^{t-\frac{1}{2}}}{\sqrt{n}}\right)$$
$$= o\left(\exp\left(-\frac{1}{2}n^{2t-1}\right)\right).$$

This implies the considerably weaker estimate

$$P_n\left(S_n > np + n^t\right) = o\left(n^{-1/2}\right).$$

Thus

$$P_n\left(S_n = k\right) = o\left(n^{-1/2}\right),$$

where the estimate $o(n^{-1/2})$ is uniform in k when $k - np > n^t$. Since $t > \frac{1}{2}$, then

$$\exp\left(-\frac{(k-np)^2}{2np(1-p)}\right) = o(1)$$

unformly in k when $k - np > n^t$. This implies that

$$\frac{1}{\sqrt{2\pi p(1-p)n}}\exp\left(-\frac{(k-np)^2}{2np(1-p)}\right) = o\left(n^{-1/2}\right),$$

where the estimate $o(n^{-1/2})$ is uniform in k when $k - np > n^t$. Thus

$$P_n\left(S_n = k\right) = \frac{1}{\sqrt{2\pi p(1-p)n}}\exp\left(-\frac{(k-np)^2}{2np(1-p)}\right) + o\left(n^{-1/2}\right),$$

where the estimate $o(n^{-1/2})$ is uniform in k when $k - np > n^t$.

We obtain the same estimate when k satisfies $k - np < -n^t$.

By combining these two estimates with (9.1), we obtain the result stated in Theorem 9.1. □

Proof of Theorem 9.2. When $p = 1/2$, Theorem 9.1 implies that

$$\binom{n}{k}\frac{1}{2^n} = \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{n}}\left(\exp\left(-\frac{2}{n}\left(k-\frac{n}{2}\right)^2\right) + o(1)\right)$$

uniformly in $k \in \mathbb{Z}$.

For an arbitrary parameter p in the interval $(0,1)$, we can write

$$P_n(S_n = k) = \sqrt{\frac{2}{\pi}} p^k (1-p)^{n-k} 2^n \frac{1}{\sqrt{n}} \left(\exp\left(-\frac{2}{n} \left(k - \frac{n}{2} \right)^2 \right) + o(1) \right)$$

uniformly in $k \in \mathbb{Z}$.

Replacing k by $\frac{n+k}{2}$ yields the desired result. □

Remark. The transition from one form of the local limit theorem to the other, in the special case of variables with a binomial distribution, is deceptive. Like all the other limit theorems presented in this book, these local limit theorems extend to the much more general setting of sums of arbitrary, independent, identically distributed random variables. The two versions of the local limit theorem coincide in the case of centered variables. The first form of the local limit theorem, like the central limit theorem, can be proved using Fourier analysis. The first form implies the second form, but via a much less trivial argument than the one we use here: a method developed by Cramér connects the study of noncentered random variables to the study of centered random variables, and it establishes the transition from the first form of the local limit theorem to the second form.

Proof of Corollary 9.3. If k is a fixed even integer, Theorem 9.2 implies that

$$P_{2n}\left(M_{2n} = k\right) = \sqrt{\frac{2}{\pi}} \left(\frac{p}{1-p} \right)^{k/2} \frac{\left(2\sqrt{p(1-p)} \right)^{2n}}{\sqrt{2n}} \left(1 + o\left(1\right)\right),$$

which implies that

$$\frac{1}{2n} \ln\left(P_{2n}\left(M_{2n} = k\right)\right) \longrightarrow \ln\left(2\sqrt{p(1-p)}\right)$$

as n approaches infinity. Similarly,

$$\frac{1}{2n+1} \ln\left(P_{2n+1}\left(M_{2n+1} = k+1\right)\right) \longrightarrow \ln\left(2\sqrt{p(1-p)}\right)$$

as n approaches infinity. If the finite subset K of $\mathbb{Z}$ contains an even integer k_0, then

$$P_{2n}\left(M_{2n} \in K\right) \geq P_{2n}\left(M_{2n} = k_0\right)$$

and

$$P_{2n}\left(M_{2n}\in K\right)$$
$$\le \sqrt{\frac{2}{\pi}}\left(\sum_{k\in K}\left(\frac{p}{1-p}\right)^{k/2}\right)\frac{\left(2\sqrt{p(1-p)}\right)^{2n}}{\sqrt{2n}}\left(1+o\left(1\right)\right).$$

When combined, these two inequalities imply that

$$\frac{1}{2n}\ln\left(P_{2n}\left(M_{2n}\in K\right)\right)\longrightarrow\ln\left(2\sqrt{p(1-p)}\right)$$

as n approaches infinity. The behavior along the sequence of odd integers is identical. □

This page intentionally left blank

http://dx.doi.org/10.1090/stml/028/11

Chapter 10

The Arcsine Law

10.1. Introduction

Wally and Andre are playing the game of Heads or Tails. This game consists of successive tosses of a fair coin; Wally gains one point when the coin lands heads and Andre gains one point when the coin lands tails. Since the coin is fair, we would expect Wally to be ahead about half the time and Andre to be ahead about half the time in a long enough game. However, this expectation is false: ties become more rare as the game lasts longer, and in fact one of the two players will probably be ahead most of the time!

For example, we will show that one of the two players will be ahead more than 75% of the time with a probability greater than 2/3, and one of the players will be ahead more than 85% of the time with a probability greater than 1/2. It is not even unlikely that one of the players will be ahead more than 97% of the time, which occurs with a probability greater than 1/5.

These results follow from Theorem 10.1, called the *arcsine law*, which provides a precise asymptotic estimate of the probability that Wally will be ahead a fixed proportion of the time. This law was discovered by Paul Lévy.[1]

[1] P. Lévy, *Sur certains processus stochastiques homogènes*, Compositio Mathematica, vol. 27, pp. 283–339, 1939.

Here is a second surprising property of the game of Heads or Tails. Over the course of n tosses, Wally and Andre will be tied Z_n times. Naively, one might expect that the random variable Z_n would increase proportionally with n. However, we will see that Z_n grows much more slowly than that, at a rate of order $\sqrt{n}$. This is the result of Theorem 10.2, which we call the *law of returns to the origin.*

10.2. Statement of the Theorems

Our mathematical model is the same as in the previous chapters. We set $p = 1/2$ to reflect the fact that the coin is fair; thus, the space $\Omega_n = \{0,1\}^n$ is equipped with the uniform probability P_n. We set $S_n(\omega) = \omega_1 + \omega_2 + \cdots + \omega_n$, where $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$, and

$$M_n := 2S_n - n$$

(by convention, we set $M_0 = S_0 = 0$).

If S_n is the number of heads appearing in a sequence of n tosses, then M_n is the number of points by which Wally is ahead of Andre at the end of the n tosses (if Wally is winning after the nth toss, M_n is positive; if Wally is losing after the nth toss, M_n is negative). The sequence (M_n) is called the *simple random walk.* For large n, the arcsine law describes the behavior of the random variable

$$T_n := \#\{k \ : \ 0 \le k \le n \text{ and } M_k > 0\},$$

which is the number of tosses after which Wally is winning, and the law of returns to zero describes the behavior of the random variable

$$U_n := \#\{k \ : \ 0 < k \le n \text{ and } M_k = 0\},$$

which is the number of instances where the two players are tied.

Theorem 10.1 (arcsine law). *For each real α between* 0 *and* 1*,*

$$P_n\left(T_n < n\alpha\right) \longrightarrow \frac{1}{\pi}\int_0^\alpha \frac{1}{\sqrt{x(1-x)}}\,dx = \frac{2}{\pi}\arcsin\sqrt{\alpha}$$

as n approaches infinity.

Here is a numerical illustration of this law. Let W be the proportion of game time during which Wally is ahead, and let p_i be the probability that W lies between $\frac{i}{10}$ and $\frac{i+1}{10}$. If the game lasts long enough, then $p_0 = p_9 \approx .20 > p_1 = p_8 \approx .090 > p_2 = p_7 \approx .074 > p_3 = p_6 \approx .067 > p_4 = p_5 \approx .064$.

Theorem 10.2 (law of returns to the origin). *For each $\alpha > 0$,*

$$P_n\left(U_n < \alpha\sqrt{n}\right) \longrightarrow \sqrt{\frac{2}{\pi}} \int_0^{\alpha} e^{-x^2/2}\, dx$$

as n approaches infinity.

This theorem tells us, for example, that if Wally and Andre toss a coin 10,000 times, then they will be tied less than 68 times with probability greater than 1/2. If their game consists of one hundred times as many tosses, then the same is true for 680 ties instead of 68.

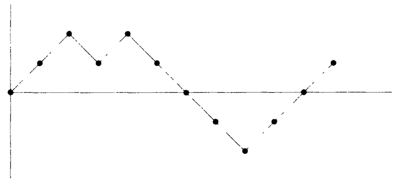
An immediate corollary of the law of returns to zero is that $\lim_{n\to\infty} P_n(U_n > n\epsilon) = 0$ for every $\epsilon > 0$. This implies that the random variable $T_n + U_n = \#\{k : 0 < k \leq n \text{ and } M_k \geq 0\}$ satisfies the same arcsine law as the random variable T_n.

Like the proof of central limit theorem, the proofs of these two theorems consist of three steps: an explicit calculation of probabilities, an approximation using Stirling's formula, and an application of Riemann sums.

We also note that numerous results similar to these two theorems can be found in the literature on probability theory.

10.3. The Reflection Principle

We will use a graphical representation of the sample space $\Omega_n = \{0,1\}^n$. Any continuous curve in $\mathbb{R}^2$ consisting of a finite union of segments of the form $[(i,j),(i+1,j+1)]$ or $[(i,j),(i+1,j-1)]$, where i and j integers, is called a *path.* Such a path has an origin (a,b) and an endpoint (c,d), which are points on C with integer coordinates satisfying $a \leq i \leq c$ for all $(i,j) \in C$. The length of the path C is $c-a$ (the Euclidean length of C is $(c-a)\sqrt{2}$). To each element $\omega \in \Omega_n$, we associate a path $\bigcup_{i=0}^{n-1}[(i, M_i(\omega)), (i+1, M_{i+1}(\omega)]$ with origin $(0,0)$ and endpoint $(n, M_n(\omega))$. It is clear that this representation of Ω_n, which describes the lead of Wally over Andre during the course of the game, is injective. Also, note that each path with origin $(0,0)$ and an endpoint with x-coordinate n represents an element of Ω_n. For example, the following graph represents the element $\{1\,1\,0\,1\,0\,0\,0\,0\,1\,1\,1\}$ of Ω_{11}.



If c and d are two integers such that $0 \le |d| \le c$, then the number of paths with origin $(0,0)$ and endpoint (c,d) is zero if $c+d$ is odd and $2^c P_c\,(M_c = d) = 2^c P_c\,(S_c = (c+d)/2) = \binom{c}{(c+d)/2}$ if $c+d$ is even. More generally, if $a, b, c,$ and d are integers such that $0 \le |d-b| \le c-a$ and $c-a+d-b$ is even, then the number of paths with origin (a,b) and endpoint (c,d) is $\binom{c-a}{(c-a+d-b)/2}$.

The following proposition, which is equivalent to a result proved by Désiré André,[2] will play an essential role in our analysis.

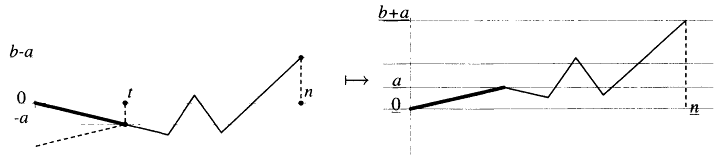
Proposition 10.3 (reflection principle). *Let $a, b \ge 0$ and $n > 0$ be integers. Then*

$$P_n\,(M_n = b-a \text{ and } M_k = -a \text{ for a } k \in [0,n]) = P_n\,(M_n = b+a)\,.$$

Proof. The case $a = 0$ is trivial, so suppose that $a > 0$. By using our graphical representation of Ω_n, we can prove the theorem by proving that the number of paths with origin $(0,0)$ and endpoint $(n, b-a)$ that cross the horizontal line $y = -a$ is equal to the number of paths with origin $(0,0)$ and endpoint $(n, a+b)$. By translating each path of the first type up by a units, we see that this is equivalent to showing that the number of paths with origin $(0,a)$ and endpoint (n,b) crossing the x-axis (which we will call *paths of the first type*) equals the number of paths with origin $(0,-a)$ and endpoint (n,b) (which we will call *paths of the second type*). If C is a path of the first type, we let $t(C)$ be the smallest $i > 0$ such that $(i,0) \in C$. The path C is a union of a path

[2]D. André, *Solution directe du problème résolu par M. Bertrand*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 105, pp. 436–437, 1887. The paper is about the *ballot problem*: for an election between two candidates, what is the probability that the winning candidate is always ahead as the votes are counted?

C_1 with origin $(0,0)$ and endpoint $(t(C),0)$ and a path C_2 with origin $(t(C),0)$ and endpoint (n,b). Then to each C, we associate the path C' that is the union of C_2 with the reflection of C_1 across the x-axis. The path C' is a path of the second type and the correspondence $C \leftrightarrow C'$ is a bijection between the sets of paths of each of the two types.



□

We know the probability that the two players will be tied at the end of $2n$ coin tosses:

$$P_{2n}(M_{2n}=0) = P_{2n}(S_{2n}=n) = 2^{-2n}\binom{2n}{n}.$$

In the next two corollaries, we will see that this equals the probability of the event *the same player remains ahead over the course of* $2n$ *consecutive tosses* and the probability of the event *Andre is winning at no point during the game.*

Corollary 10.4.

$$P_{2n}(M_1 \neq 0, M_2 \neq 0, \dots, \text{ and } M_{2n} \neq 0)$$
$$= 2P_{2n}(M_1 > 0, M_2 > 0, \dots, \text{ and } M_{2n} > 0)$$

and

$$P_{2n}(M_1 > 0, M_2 > 0, \dots, \text{ and } M_{2n} > 0) = 2^{-(2n+1)}\binom{2n}{n}.$$

Proof. The first identity is obvious by symmetry.

To prove the second identity, we will again use our graphical representation of Ω_n. We need to count the paths with origin $(0,0)$ and length $2n$ that are contained in the upper half-plane (and that do not return to the x-axis). We will thus need to sum, for $k=1$ to $k=n$, the number of paths with origin $(1,1)$ and endpoint $(2n,2k)$ that do not touch the x-axis. There is only one path that connects the point $(1,1)$ to the point $(2n, 2n)$, and this path does not return to the x-axis. If $1 \leq k < n$, the number of paths connecting the point $(1,1)$ to the point $(2n, 2k)$ that do not return to the x-axis equals the number of paths connecting the point $(1,1)$ to the point $(2n, 2k)$ minus the number of paths connecting the point $(1,1)$ to the point $(2n, 2k)$ that do return to the x-axis. The number of paths connecting the point $(1,1)$ to the point $(2n, 2k)$ equals $\binom{2n-1}{n+k-1}$. By the reflection principle, the number of paths connecting the point $(1,1)$ to the point $(2n, 2k)$ that return to the x-axis equals the number of paths connecting the point $(1,-1)$ to the point $(2n, 2k)$, which is $\binom{2n-1}{n+k}$.

In conclusion, the number of paths with origin $(0,0)$ and length $2n$ that are contained in the upper half-plane is

$$1 + \sum_{k=1}^{n-1} \left(\binom{2n-1}{n+k-1} - \binom{2n-1}{n+k} \right).$$

This number simplifies to $\binom{2n-1}{n}$, and it is easy to check that $\binom{2n-1}{n} = \frac{1}{2}\binom{2n}{n}$. This proves the corollary. □

Now we will compute the probability that Andre is ahead at no point in the game.

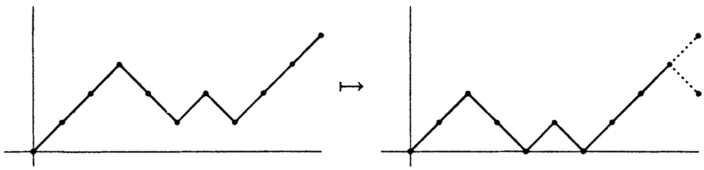
Corollary 10.5.

$$P_{2n}(M_1 \geq 0, M_2 \geq 0, \ldots, \text{ and } M_{2n} \geq 0) = 2^{-2n}\binom{2n}{n}.$$

Proof. We want to show that

$$\begin{aligned} &P_{2n}(M_1 \geq 0, M_2 \geq 0, \ldots, \text{ and } M_{2n} \geq 0) \\ &\qquad = 2P_{2n}(M_1 > 0, M_2 > 0, \ldots, \text{ and } M_{2n} > 0). \end{aligned}$$

The number of paths with origin $(0,0)$ and length $2n$ that are contained in the upper half-plane and that never return to the x-axis equals the number of paths with origin $(0,0)$ and length $2n-1$ that are contained in the upper half-plane (including the x-axis). This is true because there is a bijection between the sets of these two types of paths: we associate a path of the second type to each path of the first type by removing the initial segment and translating the path 1 unit left and 1 unit down.



Note that M_{2n-1} is never zero. Then to each path with origin $(0,0)$ and length $2n-1$ that is contained in the upper half-plane (including the x-axis), we can associate exactly two paths with origin $(0,0)$ and length $2n$ that are contained in the upper half-plane (including the x-axis). To do this, we add a segment of length 1 and slope 1 or -1 to the end of the path of length $2n-1$. Therefore, the cardinality of the event $(M_1 \geq 0, M_2 \geq 0, \ldots,$ and $M_{2n} \geq 0)$ is twice the cardinality of the event $(M_1 > 0, M_2 > 0, \ldots,$ and $M_{2n} > 0)$. □

The reflection principle also allows us to calculate the probability that Wally is ahead until the last toss in the game, which ties Wally and Andre.

Corollary 10.6.

$$P_{2n}(M_1 > 0, \ldots, M_{2n-1} > 0 \text{ and } M_{2n} = 0) = \frac{1}{n2^{2n}}\binom{2n-2}{n-1}.$$

Proof. We need to count the paths with origin $(0,0)$ and endpoint $(2n,0)$ that are contained in the open upper half-plane (which does not include the x-axis). In other words, we need to count the paths with origin $(1,1)$ and endpoint $(2n-1,1)$ that are contained in the open upper half-plane. This number equals the number of paths with origin $(1,1)$ and endpoint $(2n-1,1)$ minus the number of paths with origin $(1,1)$ and endpoint $(2n-1,1)$ that touch the x-axis at some point. The number of paths with origin $(1,1)$ and endpoint $(2n-1,1)$ equals $\binom{2n-2}{n-1}$. By the reflection principle, the number of paths with origin $(1,1)$ and endpoint $(2n-1,1)$ that touch the x-axis equals the number of paths with origin $(1,-1)$ and endpoint $(2n-1,1)$; this number is $\binom{2n-2}{n-2}$. Finally, it is easy to check that $\binom{2n-2}{n-1} - \binom{2n-2}{n-2} = \frac{1}{n}\binom{2n-2}{n-1}$. □

Corollary 10.6 implies the following useful combinatorial identity.

Lemma 10.7. *If n and k are integers such that $0 \leq k < n$, then*

$$\sum_{j=1}^{n-k} \frac{1}{j}\binom{2j-2}{j-1}\binom{2(n-j)-k}{n-j} = \binom{2n-k-1}{n}.$$

In particular,

$$2\sum_{j=1}^{n} \frac{1}{j}\binom{2j-2}{j-1}\binom{2(n-j)}{n-j} = \binom{2n}{n}.$$

Proof. We will count the number of paths with origin $(0,0)$ and endpoint $(2n-k,k)$ in two different ways. On one hand, the number of such paths is $\binom{2n-k}{n}$. On the other hand, the number of such paths equals the number of paths with origin $(1,1)$ and endpoint $(2n-k,k)$ plus the number of paths with origin $(1,-1)$ and endpoint $(2n-k,k)$. First, the number of paths with origin $(1,1)$ and endpoint $(2n-k,k)$ is $\binom{2n-k-1}{n-1}$. Next, for every path with origin $(1,-1)$ and endpoint $(2n-k,k)$, there exists a minimum integer $1 \leq j \leq n-k$ such that the path passes through $(2j,0)$. The number of paths with origin $(1,-1)$ and endpoint $(2n-k,k)$, therefore, is equal to the sum for j from 1 to $n-k$ of the product of the number of paths with origin $(0,0)$ and endpoint $(2j,0)$ that are contained in the lower half-plane and the number of paths with origin $(2j,0)$ and endpoint $(2n-k,k)$. By Corollary 10.6, this implies that the number of paths with origin $(1,-1)$ and endpoint $(2n-k,k)$ is

$$\sum_{j=1}^{n-k} \left(\frac{1}{j}\binom{2j-2}{j-1}\right)\left(\binom{2(n-j)-k}{n-j}\right).$$

Considering all of the information above, we see that this equals $\binom{2n-k}{n} - \binom{2n-k-1}{n-1}$, which simplifies to $\binom{2n-k-1}{n}$.

The second formula of Lemma 10.7 is obtained by considering the special case where $k=0$ and using the fact that $\binom{2n}{n} = 2\binom{2n-1}{n}$. □

10.4. Proof of the Arcsine Law

To help us prove the arcsine law, we will define a new sequence of random variables (T'_n) that are closely related to the variables T_n. Let T'_n be the number of integers k between 1 and n such that there were more heads than tails in the first k or $k-1$ tosses of the coin.

Using the graphical representation of Ω_n introduced in the beginning of Section 10.3, we see that T'_n associates to each path of length n and origin $(0,0)$ the number of elementary segments of the path that lie in the upper half-plane. More formally,

$$T'_n := \#\{k : 0 < k \leq n \text{ and } (M_k > 0 \text{ or } M_{k-1} > 0)\}.$$

Every elementary segment has the form

$$[(2k-1, M_{2k-1}), (2k, M_{2k-1} \pm 1)]$$

or the form

$$[(2k-2, M_{2k-1} \pm 1), (2k-1, M_{2k-1})],$$

where M_{2k-1} is a nonzero integer. This segment is in the upper half-plane if and only if $M_{2k-1} > 0$. Therefore,

$$T'_{2n} = 2\#\{k : 0 < k \leq n \text{ and } M_{2k-1} > 0\}.$$

Proposition 10.8. *For each $n > 0$ and k between 0 and n,*

$$P_{2n}(T'_{2n} = 2k) = 2^{-2n}\binom{2k}{k}\binom{2(n-k)}{n-k}. \tag{10.1}$$

Proof. We have

$$(T'_{2n} = 2n) = (M_k \geq 0 \text{ for every } k \text{ between } 1 \text{ and } 2n),$$

and the probability of this event is given by Corollary 10.5. We will prove the general formula (10.1) by induction on n. The statement is clearly true for $n = 1$. Fix an $N > 1$ and suppose that (10.1) is true for every n between 1 and $N-1$ and every k between 0 and n.

The probability of the event

$$(T'_{2N} = 0) = (M_k \leq 0 \text{ for every } k \text{ between } 1 \text{ and } 2N)$$

is given by Corollary 10.5. If $0 < T'_{2N} < 2N$, then there exists a j between 1 and N such that $M_{2j} = 0$. For each $\omega \in \Omega_{2n}$ such that $0 < T'_{2n}(\omega) < 2N$, we set

$$t(\omega) := \min\{j > 0 : M_{2j}(\omega) = 0\}.$$

Now fix a k between 0 and N. Then

$$P_{2N}\left(T'_{2N}=2k\right)=\sum_{j=1}^{N} P_{2N}\left(T'_{2N}=2k \text{ and } t=2j \text{ and } M_1>0\right)$$
$$+\sum_{j=1}^{N} P_{2N}\left(T'_{2N}=2k \text{ and } t=2j \text{ and } M_1<0\right).$$

If $j>k$, then the event $(T'_{2N}=2k$ and $t=2j$ and $M_1>0)$ is empty. If $j\le k$, the cardinality of this event equals the number of paths with origin $(0,0)$ and endpoint $(2j,0)$ that are contained in the open upper half-plane between the origin and the endpoint multiplied by the number of paths with origin $(2j,0)$ and length $2(N-j)$ in which $2(k-j)$ elementary segments are contained in the upper half-plane. The number of paths with origin $(0,0)$ and endpoint $(2j,0)$ that are contained in the open upper half-plane between the origin and the endpoint is given by Corollary 10.6; the number of such paths is $\frac{1}{j}\binom{2j-2}{j-1}$. The number of paths with origin $(2j,0)$ and length $2(N-j)$ in which $2(k-j)$ elementary segments are contained in the upper half-plane is given by the induction hypothesis; the number of such paths is $\binom{2(k-j)}{k-j}\binom{2(N-k)}{N-k}$. Therefore,

$$P_{2N}\left(T'_{2N}=2k \text{ and } t=2j \text{ and } M_1>0\right)$$
$$=\frac{1}{j2^{2N}}\binom{2j-2}{j-1}\binom{2(k-j)}{k-j}\binom{2(N-k)}{N-k}.$$

If $j>N-k$, then the event $(T'_{2N}=2k$ and $t=2j$ and $M_1<0)$ is empty. If $j\le N-k$, the cardinality of this event equals the number of paths with origin $(0,0)$ and endpoint $(2j,0)$ that are contained in the open lower half-plane between the origin and the endpoint multiplied by the number of paths with origin $(2j,0)$ and length $2(N-j)$ in which $2k$ elementary segments lie in the upper half-plane. The number of paths with origin $(0,0)$ and endpoint $(2j,0)$ that are contained in the open lower half-plane between the origin and the endpoint is given by Corollary 10.6; the number of such paths is $\frac{1}{j}\binom{2j-2}{j-1}$. The number of paths with origin $(2j,0)$ and length $2(N-j)$ in which $2k$ elementary segments lie in the upper half-plane is given by the induction hypothesis; the number of such paths is $\binom{2k}{k}\binom{2(N-j-k)}{N-j-k}$.

Thus

$$P_{2N}\left(T'_{2N} = 2k \text{ and } t = 2j \text{ and } M_1 < 0\right)$$
$$= \frac{1}{j2^{2N}}\binom{2j-2}{j-1}\binom{2k}{k}\binom{2(N-j-k)}{N-j-k}.$$

We complete the proof by using the second formula of Lemma 10.7:

$$P_{2N}(T'_{2N} = 2k)$$
$$= \sum_{j=1}^{k} \frac{1}{j2^{2N}}\binom{2j-2}{j-1}\binom{2(k-j)}{k-j}\binom{2(N-k)}{N-k}$$
$$+ \sum_{j=1}^{N-k} \frac{1}{j2^{2N}}\binom{2j-2}{j-1}\binom{2k}{k}\binom{2(N-j-k)}{N-j-k}$$
$$= \left(\frac{1}{2^{2N}}\binom{2(N-k)}{N-k}\right)\left(\frac{1}{2}\binom{2k}{k}\right)$$
$$+ \left(\frac{1}{2^{2N}}\binom{2k}{k}\right)\left(\frac{1}{2}\binom{2(N-k)}{N-k}\right)$$
$$= \frac{1}{2^{2N}}\binom{2(N-k)}{N-k}\binom{2k}{k}.$$

The result of the proposition follows by induction. □

Now we will show that the random variables T'_{2n} satisfy the arcsine law.

Proposition 10.9. *For all a and b such that $0 \leq a \leq b \leq 1$,*

$$\lim_{n\to\infty} P_{2n}\left(2na \leq T'_{2n} \leq 2nb\right) = \frac{1}{\pi}\int_a^b \frac{1}{\sqrt{x(1-x)}}\, dx.$$

Proof. First we analyze the case where $0 < a < b < 1$. From Proposition 10.8, a direct application of Stirling's formula (Proposition 7.2) yields that

$$P_{2n}\left(T'_{2n} = 2k\right) = \frac{1}{\pi}\frac{1}{\sqrt{k(n-k)}}(1+\epsilon(k))(1+\epsilon(n-k)),$$

where $\lim_{k\to\infty} \epsilon(k) = 0$, for $0 < k < n$. Thus

$$P_{2n}\left(T'_{2n} = 2k\right) = \frac{1}{\pi}\frac{1}{\sqrt{k(n-k)}}(1+\epsilon(n,k)),$$

where $\lim_{n\to\infty} \epsilon(n,k) = 0$ uniformly when k is an integer in $[na, nb]$. This implies that

$$\begin{aligned} P_{2n}\left(2na \le T'_{2n} \le 2nb\right) &= \sum_{na \le k \le nb} P_{2n}\left(T'_{2n} = 2k\right) \\ &\sim \sum_{na \le k \le nb} \frac{1}{\pi} \frac{1}{\sqrt{k(n-k)}}. \end{aligned}$$

The last summation can be written as

$$\frac{1}{n\pi} \sum_{k=0}^{n} \chi_{[a,b]}\left(\frac{k}{n}\right) \frac{1}{\sqrt{\frac{k}{n}\left(1 - \frac{k}{n}\right)}},$$

which is a Riemann sum that approaches $\frac{1}{\pi}\int_a^b \frac{1}{\sqrt{x(1-x)}}\, dx$ as n approaches infinity. This completes the first step of the proof.

Now, fix $\epsilon > 0$. Since the integral $\int_0^1 \frac{1}{\sqrt{x(1-x)}}\, dx$ converges for all small enough $a > 0$,

$$\frac{1}{\pi}\int_0^a \frac{1}{\sqrt{x(1-x)}}\, dx < \epsilon$$

and

$$\frac{1}{\pi}\int_{1-a}^1 \frac{1}{\sqrt{x(1-x)}}\, dx < \epsilon.$$

Fix such a number a. From the first step of the proof, we have that

$$\left| \frac{1}{\pi}\int_a^{1-a} \frac{1}{\sqrt{x(1-x)}}\, dx - P_{2n}\left(2na \le T'_{2n} < 2n(1-a)\right) \right| < \epsilon$$

for large enough n. Now,

$$\begin{aligned} P_{2n}\left(T'_{2n} < 2na\right) + P_{2n}\left(2na \le T'_{2n} < 2n(1-a)\right) \\ + P_{2n}\left(T'_{2n} \ge 2n(1-a)\right) = 1 \end{aligned}$$

and $\frac{1}{\pi}\int_0^1 \frac{1}{\sqrt{x(1-x)}}\, dx = 1$. We deduce that

$$\begin{aligned} \Bigg| \frac{1}{\pi}\int_0^a \frac{1}{\sqrt{x(1-x)}}\, dx + \frac{1}{\pi}\int_{1-a}^1 \frac{1}{\sqrt{x(1-x)}}\, dx \\ - P_{2n}\left(T'_{2n} < 2na\right) - P_{2n}\left(T'_{2n} \ge 2n(1-a)\right) \Bigg| < \epsilon, \end{aligned}$$

for large enough n, so

$$P_{2n}\left(T'_{2n} < 2na\right) + P_{2n}\left(T'_{2n} \geq 2n(1-a)\right) < 3\epsilon.$$

We have thus shown that there exists an $a > 0$ such that

$$P_{2n}\left(T'_{2n} < 2na\right) < 3\epsilon$$

for large enough n. Since $P_{2n}\left(T'_{2n} < 2na\right)$ is an increasing function of a that approaches 0 as a approaches 0 for each fixed n, we conclude that $\lim_{a\to 0} P_{2n}\left(T'_{2n} < 2na\right) = 0$ uniformly in n.

Using the first step of the proof again, then

$$\lim_{n\to\infty} P_{2n}\left(T'_{2n} \leq 2nb\right) = \frac{1}{\pi}\int_0^b \frac{1}{\sqrt{x(1-x)}}\,dx$$

for all $b \in (0,1)$. The same argument shows that if $0 < a < b < 1$, then

$$\lim_{n\to\infty} P_{2n}\left(2na < T'_{2n} < 2nb\right) = \frac{1}{\pi}\int_a^b \frac{1}{\sqrt{x(1-x)}}\,dx$$

and

$$\lim_{n\to\infty} P_{2n}\left(T'_{2n} < 2na\right) = \frac{1}{\pi}\int_0^a \frac{1}{\sqrt{x(1-x)}}\,dx,$$

and the proposition follows immediately. □

Proof of the Arcsine Law. To complete the proof of the arcsine law (Theorem 10.1), we just need to compare the random variables T_n and T'_n. Since

$$T_{2n} = T'_{2n} - \#\left\{k : 1 \leq k \leq n, M_{2k-1} > 0 \text{ and } M_{2k} = 0\right\},$$

it follows that

$$\left|T_{2n} - T'_{2n}\right| \leq \#\left\{k : 1 \leq k \leq n \text{ and } M_{2k} = 0\right\} = U_{2n}.$$

We will prove the law of returns to the origin, which implies that

$$\lim_{n\to\infty} P_{2n}\left(U_{2n} > 2n\epsilon\right) = 0$$

for every $\epsilon > 0$, in the next section. However, we can also give a direct proof of this result. Since

$$E_{2n}\left[U_{2n}\right] = E_{2n}\left[\sum_{k=1}^{n} \chi_{(M_{2k}=0)}\right]$$
$$= \sum_{k=1}^{n} P_{2n}\left(M_{2k}=0\right) = \sum_{k=1}^{n} 2^{-2k}\binom{2k}{k},$$

Markov's inequality implies that

$$P_{2n}\left(U_{2n} > 2n\epsilon\right) \leq \frac{1}{2n\epsilon}\sum_{k=1}^{n} 2^{-2k}\binom{2k}{k}$$

for every $\epsilon > 0$. Now, it is easy to check that

$$\lim_{k\to\infty} 2^{-2k}\binom{2k}{k} = 0$$

by using Stirling's formula, so Cesàro's principle implies that

$$\lim_{n\to\infty} P_{2n}\left(U_{2n} > 2n\epsilon\right) = 0.$$

Since the event $(T_2 n < 2n\alpha)$ is contained in the union

$$\left(|T_{2n} - T'_{2n}| > 2n\epsilon\right) \cup \left(T'_{2n} \leq 2n(\alpha+\epsilon)\right),$$

we have that

$$P_{2n}\left(T_{2n} < 2n\alpha\right)$$
$$\leq P_{2n}\left(|T_{2n} - T'_{2n}| > 2n\epsilon\right) + P_{2n}\left(T'_{2n} \leq 2n(\alpha+\epsilon)\right).$$

On one hand,

$$P_{2n}\left(|T_{2n} - T'_{2n}| > 2n\epsilon\right) \leq P_{2n}\left(U_{2n} > 2n\epsilon\right) \longrightarrow 0$$

as $n \to \infty$. On the other hand, Proposition 10.9 implies that

$$\lim_{n\to\infty} P_{2n}\left(T'_{2n} \leq 2n(\alpha+\epsilon)\right) = \frac{1}{\pi}\int_0^{\alpha+\epsilon} \frac{1}{\sqrt{x(1-x)}}\,dx$$

and

$$\lim_{\epsilon\to 0} \frac{1}{\pi}\int_0^{\alpha+\epsilon} \frac{1}{\sqrt{x(1-x)}}\,dx = \frac{1}{\pi}\int_0^{\alpha} \frac{1}{\sqrt{x(1-x)}}\,dx;$$

therefore,

$$\limsup_{n\to\infty} P_{2n}\left(T_{2n} < 2n\alpha\right) \leq \frac{1}{\pi}\int_0^{\alpha} \frac{1}{\sqrt{x(1-x)}}\,dx.$$

Since $T_{2n} \leq T'_{2n}$, Proposition 10.9 directly implies that

$$\liminf_{n\to\infty} P_{2n}\left(T_{2n} < 2n\alpha\right) \geq \frac{1}{\pi}\int_0^\alpha \frac{1}{\sqrt{x(1-x)}}\, dx.$$

We conclude that

$$\lim_{n\to\infty} P_{2n}\left(T_{2n} < 2n\alpha\right) = \frac{1}{\pi}\int_0^\alpha \frac{1}{\sqrt{x(1-x)}}\, dx.$$

To complete the proof, we only need to check that

$$\lim_{n\to\infty} P_{2n+1}\left(T_{2n+1} < (2n+1)\alpha\right) = \frac{1}{\pi}\int_0^\alpha \frac{1}{\sqrt{x(1-x)}}\, dx.$$

This follows easily from the fact that

$$P_{2n+1}\left(T_{2n+1} < (2n+1)\alpha\right) \leq P_{2n}\left(T_{2n} < (2n+1)\alpha\right)$$

and

$$P_{2n+1}\left(T_{2n+1} < (2n+1)\alpha\right) \geq P_{2n+2}\left(T_{2n+2} < (2n+1)\alpha\right).$$

We have now proved the arcsine law. □

10.5. Proof of the Law of Returns to the Origin

We will explicitly determine the distribution of the random variable U_{2n}. It is clear that this random variable only takes integer values between 0 and n. Note that for each integer k between 0 and n, the probability $P_{2n}(U_{2n} = k)$ equals 2^{-2n} multiplied by the number of paths with origin $(0,0)$ and length $2n$ that touch the x-axis at exactly $k+1$ points.

Proposition 10.10. *For each k between 0 and n,*

$$P_{2n}\left(U_{2n} = k\right) = \frac{1}{2^{2n-k}}\binom{2n-k}{n}.$$

Note that $P_{2n}\left(U_{2n} = k\right)$ is a decreasing function of k.

Proof. We proved the formula $P_{2n}\left(U_{2n} = 0\right) = 2^{-2n}\binom{2n}{n}$ in Corollary 10.4. We will use induction on n to prove the general formula. If

$U_{2n}(\omega) > 0$, then there exists a minimal $j \geq 1$ such that $M_{2j}(\omega) = 0$. We let $t(\omega)$ be this integer j. Let $k > 0$; then

$$P_{2n}\left(U_{2n} = k\right) = \sum_{j=1}^{n} P_{2n}\left(t = j \text{ and } U_{2n} = k\right).$$

If $j > n - k + 1$, then the event $(t = j$ and $U_{2n} = k)$ is empty. If $j \leq n - k + 1$, then the cardinality of this event equals the number of paths with origin $(0,0)$ and endpoint $(2j, 0)$ that do not touch the x-axis between the origin and endpoint multiplied by the number of paths with origin $(2j, 0)$ and length $2n - 2j$ that touch the x-axis at exactly k points. These numbers are given respectively by Corollary 10.6 and by the induction hypothesis. Thus, we obtain that

$$\#\left(t = j \text{ and } U_{2n} = k\right) = \frac{2}{j}\binom{2j-2}{j-1} 2^{k-1} \binom{2n-2j-k+1}{n-j}.$$

This implies that

$$\#\left(U_{2n} = k\right) = 2^k \sum_{j=1}^{n-k+1} \frac{1}{j}\binom{2j-2}{j-1}\binom{2n-2j-k+1}{n-j},$$

and Lemma 10.7 allows us to conclude that

$$\#\left(U_{2n} = k\right) = 2^k \binom{2n-k}{n}.$$

□

Proof of the Law of Returns to the Origin. Fix a real number $\alpha > 0$. We will use notation from our proof of the central limit theorem: if $(s_{n,k})_{n>0, 0 \leq k < \alpha\sqrt{2n}}$ and $(t_n)_{n>0}$ are two families of real numbers, we write $s_{n,k} = O_u(t_n)$ if the absolute value of $s_{n,k}$ is bounded above by a multiple of t_n, uniformly in k.

Proposition 10.10 and Stirling's formula provide an estimate of $P_{2n}\left(U_{2n} = k\right)$ when $0 \leq k < a\sqrt{2n}$:

$$P_{2n}\left(U_{2n} = k\right) = \frac{1}{\sqrt{2\pi}}\sqrt{\frac{2n-k}{n(n-k)}}\left(1 - \frac{k}{2n}\right)^n \cdot \left(1 + \frac{k}{2(n-k)}\right)^{n-k}\left(1 + O_u\left(\frac{1}{n}\right)\right).$$

Hence,

$$P_{2n}(U_{2n}=k)$$
$$=\frac{1}{\sqrt{2\pi}}\sqrt{\frac{2}{n}}\left(1+O_u\left(\frac{1}{\sqrt{n}}\right)\right)\times\exp\left[n\left(-\frac{k}{2n}-\frac{k^2}{8n^2}\right)+\right.$$
$$\left.(n-k)\left(\frac{k}{2(n-k)}-\frac{k^2}{8(n-k)^2}\right)+O_u\left(\frac{1}{\sqrt{n}}\right)\right],$$

and finally

$$P_{2n}(U_{2n}=k)=\frac{1}{\sqrt{2\pi}}\sqrt{\frac{2}{n}}\exp\left(-\frac{k^2}{4n}\right)\left(1+O_u\left(\frac{1}{\sqrt{n}}\right)\right).$$

This estimate, uniform in k, allows us to write

$$P_{2n}\left(U_{2n}<\alpha\sqrt{2n}\right)$$
$$=\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{2n}}\sum_{0\le k<\alpha\sqrt{2n}}\exp\left(-\frac{1}{2}\left(\frac{k}{\sqrt{2n}}\right)^2\right)\left(1+O\left(\frac{1}{\sqrt{n}}\right)\right).$$

As in Lemma 7.4, viewing this as a Riemann sum yields

$$\lim_{n\to\infty}P_{2n}\left(U_{2n}<\alpha\sqrt{2n}\right)=\sqrt{\frac{2}{\pi}}\int_0^\alpha\exp\left(-\frac{1}{2}x^2\right)\,dx\,.$$

Finally, note that $U_{2n+1}=U_{2n}$ since M_{2n+1} is never zero. Thus,

$$P_{2n+1}\left(U_{2n+1}<\alpha\sqrt{2n+1}\right)=P_{2n}\left(U_{2n}<\alpha\sqrt{2n+1}\right),$$

and it easily follows that

$$\lim_{n\to\infty}\left(P_{2n}\left(U_{2n}<\alpha\sqrt{2n+1}\right)-P_{2n}\left(U_{2n}<\alpha\sqrt{2n}\right)\right)=0.$$

We have thus shown that

$$\lim_{n\to\infty}P_n\left(U_n<\alpha\sqrt{n}\right)=\sqrt{\frac{2}{\pi}}\int_0^\alpha e^{-x^2/2}\,dx.$$

□

http://dx.doi.org/10.1090/stml/028/12

Chapter 11

The Strong Law of Large Numbers

We will now present a new class of asymptotic properties of the game of Heads or Tails that addresses infinite sequences of coin tosses and uses the concept of almost sure events.

The weak law of large numbers tells us that the probability that the empirical probability of success is close to the theoretical probability of success is large if we play our coin-tossing game long enough. The strong law of large numbers tells us that the empirical probability of success S_n/n approaches the probability of success p as the number of tosses n approaches infinity. To be rigorous, however, we need to make this statement more precise. The problem is that there do exist infinite sequences of heads and tails such that the proportion of heads does not converge to p or even such that the proportion of heads does not converge at all. However, we can exclude such sequences by resorting to the concept of an almost sure event. The strong law of large numbers tells us that the sequence (S_n/n) almost surely converges to p. This fundamental result is due to Borel.[1] Following Borel's ideas, we will also illustrate the strong law of large numbers by the concept of normal numbers.

[1]E. Borel, *Sur les probabilités dénombrables et leurs applications arithmétiques*, Rendiconti del Circolo Math. di Palermo, vol. 26, pp. 247–271, 1909.

This chapter is divided into five sections. In the first section, we will define the concept of an almost sure event. In the second section, we will prove the strong law of large numbers as a consequence of the large deviations estimate. In the third and fourth sections, we will discuss normal numbers. In the fifth section, which is essentially independent of the third and fourth sections, we will present the Borel–Cantelli lemmas and new approaches to the strong law of large numbers.

The law of the iterated logarithm, which is the subject of the next chapter, will provide an estimate of the rate of convergence in the strong law of large numbers.

11.1. Almost Sure Events, Independent Events

The experiment we will consider is an infinite sequence of independent trials of an elementary experiment having two possible outcomes (which we call success and failure). Let a fixed parameter p between 0 and 1 be the probability of success; then $1 - p$ is the probability of failure. The space that we consider is the set Ω of infinite sequences of 0's and 1's:

$$\Omega = \{\omega = (\omega_n)_{n\geq 1} \ : \ \omega_n = 0 \text{ or } 1 \text{ for all } n \geq 1\}.$$

Each ω represents a possible outcome of the elementary experiment: the nth coordinate ω_n equals 1 if the outcome of the nth trial is success and 0 if the outcome is failure. We define $S_n(\omega) = \omega_1 + \omega_2 + \cdots + \omega_n$ to represent the number of successes observed after n trials.

We will call any subset of Ω that can be defined by a condition depending on only finitely many coordinates a *finite type event*; in other words, the realization of such an event is determined after a fixed and finite number of elementary trials. Using the notation introduced in previous chapters, we can make the following formal definitions.

A subset A of Ω is a *finite type event* if there exists an integer $n = n(A) \geq 1$ and a subset A' of Ω_n such that

$$A = \{\omega \in \Omega \ : \ \omega^{(n)} \in A'\}, \quad \text{where } \omega^{(n)} := (\omega_1, \omega_2, \ldots, \omega_n).$$

The *probability* of the finite type event A is the number

$$P(A) := P_{n(A)}(A').$$

Thus,

$$P(A) = \sum_{\omega^{(n)} \in A'} p^{S_n(\omega)} q^{n-S_n(\omega)}.$$

In the definition of a finite type event A, the number $n(A)$ is not uniquely determined by A (if n works in the definition, then any number larger than n will work as well). For the definition of the probability of A to make sense, the value given by the above expression must be independent of the choice of $n(A)$. We leave it to the reader to check this easy fact.

The set $\mathcal{E}$ of finite type events contains Ω and $\emptyset$, and it is closed under taking complements and under finite union and intersection. Such a set of subsets of Ω is called a *Boolean algebra* of subsets of Ω. The probability P is a function from $\mathcal{E}$ to the interval $[0, 1]$ such that $P(\Omega) = 1$ and $P(A \cup B) = P(A) + P(B)$ for every A and B in $\mathcal{E}$ satisfying $A \cap B = \emptyset$.

We will now give the main definition of the section.

A subset N of Ω is a *negligible event* if for every $\epsilon > 0$ there exists a countable set $\{A_k : k \geq 1\}$ of finite type events such that

$$N \subset \bigcup_{k \geq 1} A_k \quad \text{and} \quad \sum_{k \geq 1} P(A_k) < \epsilon.$$

A subset of Ω is an *almost sure event* if its complement is negligible. If A is an almost sure event, we say that "ω almost surely belongs to A."

Remark. Since we will only consider finite type, negligible, and almost sure events, we will usually omit the qualifier "finite type."

We will also use the property of *invariance under shifting* of the probability P: if A is an event and k is a positive integer, then the event $\{\omega \in \Omega : (\omega_k, \omega_{k+1}, \omega_{k+2}, \ldots) \in A\}$ has the same probability as the event A.

Proposition 11.1. *Every subset of Ω that is contained in a negligible event is also negligible. Every countable union of negligible events is negligible. If p is not 0 or 1, then every countable subset of Ω is negligible.*

Proof. The first assertion follows immediately from the definition of a negligible event.

To prove the second assertion, let $\{N_n : n \geq 1\}$ be a countable set of negligible events. Let $N = \bigcup_{n\geq 1} N_n$ and fix $\epsilon > 0$. For each n, there exists a countable set $\{A_{n,k} : n, k \geq 1\}$ of finite type events such that

$$N_n \subset \bigcup_{k\geq 1} A_{n,k} \quad \text{and} \quad \sum_{k\geq 1} P(A_{n,k}) < \epsilon 2^{-n}.$$

The set $\{A_{n,k} : n, k \geq 1\}$ of events covers N and the sum of the probabilities of these events is less than $\sum_n \epsilon 2^{-n} = \epsilon$. This proves that N is a negligible event.

To prove the last assertion, suppose that p is not 0 or 1. First we will prove that every singleton subset of Ω is a negligible event. If ω is a fixed element of Ω, then for $n \geq 1$ the singleton $\{\omega\}$ is contained in the event $\{\omega' \in \Omega : \omega'^{(n)} = \omega^{(n)}\}$ that has probability less than $(\max(p, 1-p))^n$. Since this upper bound becomes arbitrarily small as n is made arbitrarily large, the singleton $\{\omega\}$ is a negligible event. Since every countable set is the countable union of singleton sets, then every countable subset of Ω is negligible. □

Let $(A_i)_{i\in I}$ be a family of events. The events A_i are *independent* if

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

for every finite set of distinct indices $i_1, i_2, \ldots, i_n \in I$. We leave it to the reader to prove that the complementary events A_i^c are independent if the events A_i are.

Proposition 11.2. *Events that are determined by coordinates with disjoint sets of indices are independent.*

We can rephrase this statement in the following way. Let $(A_i)_{i\in I}$ be a family of events. If for each $i \in I$ there is a finite subset E_i of $\mathbb{N}$ and a subset A_i' of $\{0,1\}^{E_i}$ such that

$$E_i \cap E_j = \emptyset \qquad \text{if } i \neq j$$

and

$$A_i = \{\omega \in \Omega \ : \ (\omega_n)_{n \in E_i} \in A'_i\},$$

then the events $(A_i)_{i \in I}$ are independent.

It is easy to verify this proposition using the definition of probability, and we leave the proof to the reader.

We will now provide a few examples of negligible events. In these examples, the parameter p does not equal 0 or 1.

Example. The set of sequences that are periodic after a certain point is negligible. (This is a special case of the following example.)

Example. Let b be a word constructed from the alphabet $\{0, 1\}$; that is, let b be a finite sequence of 0's and 1's. We claim that the set of infinite sequences of 0's and 1's not including the word b is negligible.

Because the set of words is countable, this claim implies by Proposition 11.1 that all possible words almost surely appear in the sequence ω.

Now let us prove the claim. We consider a word b of length $j > 0$. For each $m \geq 0$, let A_m be the event including all ω such that

$$(\omega_{mj+1}, \omega_{mj+2}, \ldots, \omega_{(m+1)j}) \neq b.$$

We know that $P(A_0) < 1$, and the property of invariance under shifting implies that all the events A_m have the same probability. In addition, Proposition 11.2 implies that the events A_m are independent. Therefore, $P(\bigcap_{k \leq m} A_k) = (P(A_0))^{m+1}$. The set of sequences ω such that

$$(\omega_{n+1}, \omega_{n+2}, \ldots, \omega_{n+j}) \neq b$$

for all $n \geq 0$ is contained in $\bigcap_{k \leq m} A_k$. Since the probability of this event can be made arbitrarily small by choosing m to be arbitrarily large, this set is negligible.

Example. The central limit theorem can be used to show that if $(a_n)_{n \geq 1}$ is an unbounded sequence of real positive numbers, then

$$\limsup_{n \to \infty} a_n \sqrt{n} \left| \frac{S_n}{n} - p \right| = +\infty$$

almost surely.

In Chapter 12 we will prove a much more precise version of this theorem; we will now provide only a sketch of the proof.

Let m be a positive number and let (n_k) be a strictly increasing sequence of integers. The set of ω such that

$$\limsup_{n\to\infty} a_n\sqrt{n}\left|\frac{S_n(\omega)}{n}-p\right|<m$$

is contained in the union over k of the finite type events

$$\left(a_{n_k}\sqrt{n_k}\left|\frac{S_{n_k}}{n_k}-p\right|<m\right).$$

The central limit theorem implies that these events have arbitrarily small probability as long as n_k and a_{n_k} are large enough. For each $\epsilon>0$, we can choose a sequence (n_k) satisfying

$$\sum_{k=1}^{+\infty} P\left(a_{n_k}\sqrt{n_k}\left|\frac{S_{n_k}}{n_k}-p\right|<m\right)<\epsilon.$$

We conclude that $\limsup_n a_n\sqrt{n}\left|\frac{S_n}{n}-p\right|\geq m$ almost surely, and applying Proposition 11.1 completes the proof.

11.2. Borel's Strong Law of Large Numbers

Theorem 11.3 (Borel's strong law of large numbers). *Almost surely,*

$$\lim_{n\to+\infty}\frac{1}{n}S_n(\omega)=p. \tag{11.1}$$

Proof. Let $R_n=\frac{1}{n}S_n(\omega)-p$. The sequence $(R_n(\omega))_{n\geq1}$ fails to approach zero if and only if there is an $m\geq1$ such that for each $n\geq1$ there exists a $k\geq n$ satisfying $|R_k(\omega)|>\frac{1}{m}$. In symbols, the set of ω not satisfying (11.1) is

$$\bigcup_{m\geq1}\bigcap_{n\geq1}\bigcup_{k\geq n}\left\{\omega\in\Omega\ :\ |R_k(\omega)|>\frac{1}{m}\right\}.$$

We want to show that this set is a negligible event. By Proposition 11.1, it suffices to show that

$$N_m:=\bigcap_{n\geq1}\bigcup_{k\geq n}\left\{\omega\in\Omega\ :\ |R_k(\omega)|>\frac{1}{m}\right\}$$

is negligible for each $m\geq1$.

For each $k \geq 1$, let $A_{m,k} := \left\{\omega \in \Omega : |R_k(\omega)| > \frac{1}{m}\right\}$. By the large deviations estimate, there exists a constant $c = c(p, m) > 0$ such that $P(A_{m,k}) \leq e^{-ck}$. Since the series $\sum_{k \geq 1} e^{-ck}$ converges, for every $\epsilon > 0$ there exists an $n \geq 1$ such that $\sum_{k \geq n} P(A_{m,k}) < \epsilon$. Because $N_m \subset \bigcup_{k \geq n} A_{m,k}$, this proves that each N_m is a negligible event. □

After discussing the concept of a random variable for infinite sequences of elementary experiments, we will state a generalization of Borel's strong law of large numbers (see Section 11.5). Nonetheless, the theorem that we just proved already has a wide range of applications. This law of large numbers tells us, for example, that the asymptotic proportion of either heads or tails in an infinite sequence of tosses of a fair coin is 1/2. Moreover, it tells us that the asymptotic proportion of any outcome in an infinite sequence of trials is the probability of that outcome for a single trial. Corollary 11.5 will make this notion precise.

Proposition 11.4. *Let $(A_n)_{n \geq 1}$ be a sequence of equiprobable independent random events with probability $P(A)$. The asymptotic empirical probability that these events will occur is almost surely $P(A)$; that is,*

$$\lim_{n \to +\infty} \frac{1}{n} \# \{k \ : \ 1 \leq k \leq n \ \text{ and } \ \omega \in A_k\} = P(A)$$

almost surely.

By setting $A_n := (\omega_n = 1)$, we see that this proposition is a generalization of Theorem 11.3. In fact, the proof of this proposition will reveal that it is simply a different way of writing that theorem.

Proof. For each ω, we create a sequence $\rho = (\rho_n)_{n \geq 1}$ of 0's and 1's by setting $\rho_n = 1$ if $\omega \in A_n$ and $\rho_n = 0$ otherwise. We need to prove that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \rho_k = P(\rho_i = 1) = P(A) \tag{11.2}$$

almost surely.

For each n and each $(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) \in \{0,1\}^n$, let $s := \sum_k \epsilon_k$; then

$$P(\rho_k = \epsilon_k,\ 1 \le k \le n) = P(A)^s (1 - P(A))^{n-s}.$$

The same calculations we used to show that $\frac{1}{n}\sum_{k=1}^{n} \omega_k$ almost surely converges to p can be applied to complete the proof that the convergence (11.2) is almost sure. □

Corollary 11.5. *Let A be a finite type event. For each integer $n \ge 1$ and each $\omega \in \Omega$, let $S(A, n, \omega)$ be the number of integers k between 1 and n such that $(\omega_k, \omega_{k+1}, \omega_{k+2}, \ldots) \in A$. Then*

$$\lim_{n\to\infty} \frac{1}{n} S(A, n, \omega) = P(A)$$

almost surely.

Proof. There exists a positive integer m such that the event A depends only on coordinates with index not greater than m. Equivalently, there exists an $A' \subset \Omega_m$ such that

$$A = \{\omega \ :\ (\omega_1, \omega_2, \ldots, \omega_m) \in A'\}.$$

For each integer j between 1 and m, consider the sequence $(A_{j,n})_{n\ge 0}$ of events defined by

$$\begin{aligned} A_{j,n} :&= \{\omega \ :\ (\omega_{j+nm}, \omega_{j+nm+1}, \omega_{j+nm+2}, \ldots) \in A\} \\ &= \left\{\omega \ :\ (\omega_{j+nm}, \omega_{j+nm+1}, \ldots, \omega_{j+(n+1)m-1}) \in A'\right\}. \end{aligned}$$

For fixed j and varying n, the events $A_{j,n}$ are independent and each has probability $P(A)$.

Let $S(A, j, n, \omega)$ be the number of integers k between 0 and $n-1$ such that $\omega \in A_{j,k}$. Proposition 11.4 implies that

$$\lim_{n\to\infty} \frac{1}{n} S(A, j, n, \omega) = P(A)$$

almost surely. In addition,

$$\frac{1}{nm} S(A, nm, \omega) = \frac{1}{m} \sum_{j=1}^{m} \frac{1}{n} S(A, j, n, \omega).$$

The corollary then follows from the inequality

$$\begin{aligned}\frac{1}{(n+1)m}S(A,nm,\omega) &\leq \frac{1}{nm+k}S(A,nm+k,\omega)\\ &\leq \frac{1}{nm}S(A,(n+1)m,\omega),\end{aligned}$$

which holds for every k from 0 to m. □

If $b = (b_1, b_2, \ldots, b_j)$ is a word constructed from the alphabet $\{0,1\}$ and $s := \sum_{i=1}^{j} b_i$, we say that $p^s q^{j-s}$ is the *probability of the word* b.

The following result is a special case of Corollary 11.5.

Corollary 11.6. *In the sequence ω, every word b almost surely occurs with asymptotic frequency equal to its probability.*

11.3. Random Sequences Taking Several Values

We will slightly enlarge the setting of our study to provide illustrations and applications of the law of large numbers: for the moment we will consider sequences of elementary random experiments with several possible outcomes. Since this extension is purely formal and will not create any new complications, we can explain the new setting quickly.

Consider a probabilisitic experiment with d possible outcomes, labeled from 1 to d. Let the probability of each outcome i be p_i; the numbers p_i are positive and their sum is 1. The probability space naturally associated to the outcome of this elementary experiment is $\Omega_1 = \{1, 2, \ldots, d\}$ and the probability of a subset A of Ω_1 is $P(A) = \sum_{i\in A} p_i$. The sample space for a sequence of n independent trials of this experiment is the product set $\Omega_n = \Omega_1^n$, which is equipped with the probability P_n given by

$$P_n\left(\omega^{(n)}\right) = \prod_{k=1}^{n} p_{\omega_k} = \prod_{i=1}^{d} p_i^{\#\{k\ :\ 1\leq k\leq n,\ \omega_k=i\}},$$

where $\omega^{(n)} = (\omega_1, \omega_2, \ldots, \omega_n) \in \Omega_n$.

The sample space for an infinite sequence of independent trials of this elementary experiment is the set $\Omega = \Omega_1^{\mathbb{N}^*}$ of infinite sequences of elements of Ω_1. As in the case where $d = 2$, we can define the concepts of finite type, negligible, and almost sure events.

The propositions that we proved about sequences of experiments with two possible outcomes generalize easily to the case of d possible outcomes. In particular, the strong law of large numbers says that for each i

$$\lim_{n \to +\infty} \frac{1}{n} \# \{k \ : \ 1 \leq k \leq n \ \text{ and } \ \omega_k = i\} = p_i$$

almost surely, and Corollary 11.5 remains true as stated.

11.4. Normal Numbers

We will now connect the properties of our coin-tossing game to the statistical properties of decimal expansions of real numbers. In fact, we will consider representations in an arbitrary base b, where b is a positive integer. First we will review the concept of the base b representation of a number and then we will define the concept of a Lebesgue-negligible set.

Let b be a positive integer. Then each real number x can be written uniquely as

$$x = x_0 + \sum_{i=1}^{+\infty} \frac{x_i}{b^i},$$

where $x_0 \in \mathbb{N}$, $x_I \in \{0, 1, 2, \ldots, b-1\}$ for $i \geq 1$ and where the x_i's do not all equal $b - 1$ after a certain point.[2]

We let $|I|$ be the length of the real interval I. A subset E of the real line $\mathbb{R}$ is *Lebesgue negligible* (or *of measure zero*) if for every $\epsilon > 0$ there exists a countable family $(I_k)_{k \geq 1}$ such that

$$E \subset \bigcup_{k \geq 1} I_k \quad \text{and} \quad \sum_{k \geq 1} |I_k| < \epsilon \, .$$

If E is negligible, then we say that almost every real number belongs to the complement of E.

Let Ω be the set $\{0, 1, \ldots, b-1\}^{\mathbb{N}^*}$ of sequences of integers between 0 and $b - 1$, and let $\omega = (\omega_n)_{n \geq 1}$ be an element of Ω. Let Ω'

[2]The concept of the *base b expansion* of a number is discussed in many books; for example, see Sections 2.1 and 12.1 of *Elementary Number Theory and its Applications, 4th ed.* by Kenneth H. Rosen (Addison-Wesley, 2000).

be the subset containing sequences that do not constantly equal $b-1$ after a certain point. We equip the finite set $\{0, 1, \ldots, b-1\}$ with the uniform probability $p_0 = p_1 = \cdots = p_{b-1} = \frac{1}{b}$. We can define the concept of a negligible subset of the set Ω by following the construction in Section 11.3. Since the complement of Ω' in Ω is countable, it is a negligible subset of Ω. We consider the function Φ from Ω' to the interval $[0, 1)$ defined by

$$\Phi(\omega) = \sum_{i=1}^{+\infty} \frac{\omega_i}{b^i}.$$

Proposition 11.7. *The function Φ is a bijection and a subset A of Ω' is negligible in Ω if and only if its image $\Phi(A)$ is Lebesgue negligible.*

Before proving this proposition, we will use the following elementary lemma to define the *measure* $|C|$ of a set C that is a finite union of real intervals.

Lemma 11.8. *If C is a finite union of real intervals, then C can be written as a finite union $C = \bigcup_k I_k$, where the I_k are pairwise disjoint intervals. Although this representation of C is not unique, the sum $\sum_k |I_k|$ of the lengths of the intervals depends only on C.*

We set $|C| = \sum_k |I_k|$.

Proof of Proposition 11.7. The fact that the function Φ is a bijection is a well-known result in number theory.

Let $n > 0$ and $(a_1, a_2, \ldots, a_n) \in \Omega_i^n$. Setting $\alpha := \sum_{i=1}^{n} \frac{a_i}{b^i}$, we have that

$$\Phi\left(\{\omega \in \Omega' \ : \ \omega_i = a_i, 1 \leq i \leq n\}\right) = \left[\alpha, \alpha + \frac{1}{b^n}\right).$$

Therefore, the image under Φ of a finite type event in Ω is a finite union of b-adic intervals (which are intervals with boundaries of the form $\frac{a}{b^j}$ with a and j are integers). Moreover, note that

$$P\left(\{\omega \in \Omega' \ : \ \omega_i = a_i, 1 \leq i \leq n\}\right) = \frac{1}{b^n} = \left|\left[\alpha, \alpha + \frac{1}{b^n}\right)\right|.$$

From this we conclude that $P(B) = |\Phi(B)|$ for any finite type event. This shows that the image of a negligible event under Φ is Lebesgue negligible.

To prove the converse, we start by noting that any interval can be approximated by a b-adic interval. For any real interval I and $\epsilon > 0$, there exists a b-adic interval I' closed on the left and open on the right such that $I \subset I'$ and $|I'| < |I| + \epsilon$. Let E be a Lebesgue-negligible subset of $[0, 1)$. By approximating intervals covering E by b-adic intervals, we can show that for any $\epsilon > 0$ there exists a countable family (I'_k) of b-adic intervals closed on the left and open on the right such that $E \subset \bigcup I'_k$ and $\sum |I'_k| < \epsilon$. Set $A_k = \Phi^{-1}(I^I_k)$. For each k, the set A_k is a finite type event and $P(A_k) = |I'_k|$. (This fact is easy to check by writing the interval I'_k as a finite union of pairwise disjoint intervals of the form $[\alpha, \alpha + b^{-j})$, where $b^j\alpha \in \mathbb{N}$.) Noting that $\Phi^{-1}(E) \subset \bigcup A_k$ and $\sum P(A_k) < \epsilon$, we conclude that $\Phi^{-1}(E)$ is negligible in Ω. □

Following Borel, we say that a real number is normal in base b (or b-normal) if every block of digits appears in the base b representation of the number with an asymptotic frequency equal to $b^{-\ell}$, where ℓ is the length of the block. In other words, $x = x_0 + \sum_{i=1}^{+\infty} \frac{x_i}{b^i}$ is normal in base b if and only if

$$\lim_{n\to\infty} \frac{1}{n} \#\{i \;:\; 1 \le i \le n,\; (x_i, x_{i+1}, \dots, x_{i+\ell-1}) = a\} = \frac{1}{b^\ell}$$

for every $\ell > 0$ and $a = (a_1, a_2, \dots, a_\ell) \in \{0, 1, \dots, b-1\}^\ell$.

A real number is *absolutely normal* if it is normal in every base.

Corollary 11.5 and Proposition 11.7 imply that almost every real number is normal in any given base b. Furthermore, we can show that a countable union of Lebesgue-negligible subsets of $\mathbb{R}$ is Lebesgue negligible in the same way that we proved the corresponding result for negligible subsets of Ω. This implies that *almost every real number is absolutely normal*, yet a simple example of an absolutely normal number has never been found! Isn't that troubling? Of course, rational numbers are never normal because their representation is periodic after a certain point. As for irrational numbers, we don't even know whether numbers such as $\sqrt{2}$, $\ln 2$, and π are normal! Such problems seem to form an insurmountable obstacle for contemporary mathematics.

11.5. The Borel–Cantelli Lemmas

Cantelli[3] extended Borel's work on the strong law of large numbers to reach the lemmas that bear these names and that are critical in making the law of large numbers more precise.

We can summarize the argument used to prove Theorem 11.3 in the following way: the large deviations estimate implies the convergence of a series of probabilities of events, which in turn implies that a certain event is negligible. In fact, this principle is used throughout probability theory, and we will give a precise statement of the principle (Proposition 11.9) as well as several applications. For this discussion we use the mathematical model described in Section 11.1.

If $(A_n)_{n\geq 1}$ is a sequence of subsets of Ω, we let A_n *infinitely often* be the set $\bigcap_{n\geq 1}\bigcup_{k\geq n} A_k$. Thus ω is an element of the set A_n infinitely often if and only if ω belongs to A_n for infinitely many indices n.

Proposition 11.9 (first Borel–Cantelli lemma). *Let $(A_n)_{n\geq 1}$ be a sequence of events. If $\sum_{n\geq 1} P(A_n)$ converges, then A_n infinitely often is a negligible event.*

Proof. The set A_n infinitely often is contained in $\bigcup_{k\geq n} A_k$ for each $n \geq 1$. Since the series $\sum_{n\geq 1} P(A_n)$ converges, the sum of the probabilities of the A_k for $k \geq n$ can be made arbitrarily small for large n. The result then follows immediately from the definition of a negligible event. □

We will now illustrate this result with three examples and a statement of a more general law of large numbers (Theorem 11.12).

Example. Let $(k_n)_{n\geq 1}$ and $(\ell_n)_{n\geq 1}$ be two sequences of positive numbers. For each $n \geq 1$, let A_n be the event consisting of outcomes with ℓ_n consecutive successes starting at the k_nth trial; that is,

$$A_n = \{\omega \in \Omega \ : \ \omega_{k_n} = \omega_{k_n+1} = \cdots = \omega_{k_n+\ell_n-1} = 1\}\,.$$

If $\sum_{n\geq 1} p^{\ell_n}$ converges, then it is almost sure that only finitely many events A_n occur.

---

[3]F. P. Cantelli, *Sulla probabilità come limite della frequenza*, Rendiconti d. r. Acad. d. Lincei, vol. 26, pp. 39–45, 1917.

Example. If three games of Heads or Tails are played simultaneously and independently, then it is almost sure that all the games will be tied at the same time only finitely many times. (A game of Heads or Tails is tied after toss n if there have been as many heads as tails at the end of n tosses.) This follows immediately from the first Borel–Cantelli lemma, and we leave the proof to the reader. This example is connected to the recurrence of random walks, which is the topic of Chapter 13.

Example. This example will use the concepts introduced in Section 11.4. First, here is a natural way to associate a real number to the sequence of integers from its decimal expansion. If x is a real number, let $x = x_0 + \sum_{i=1}^{+\infty} x_i 10^{-i}$ be its decimal expansion and consider the integer $L_n(x) = \sum_{i=1}^{n} x_i 10^{i}$ for each positive integer n.

Now, let (ϵ_n) be a sequence of real numbers such that $\sum_n \epsilon_n$ converges. For almost every real x (that is, for every number not in some fixed Lebesgue-negligible set), there exists an $n(x) \in \mathbb{N}$ such that $L_n(x) \geq \epsilon_n 10^{n+1}$ for every $n \geq n(x)$. This can be proved using the fact that the event $(L_n \leq \epsilon_n 10^{n+1})$ has probability of order ϵ_n in the space Ω of decimal expansions.

The proof of the strong law of large numbers (Theorem 11.3) relied on the fact that the set $\left|\frac{1}{n}S_n - p\right| > \frac{1}{m}$ *infinitely often* is a negligible event for any $m > 0$. The next proposition generalizes this method. Before stating the proposition, however, we will extend the concepts of a random variable, of independence, and of expected value that we introduced in Chapters 2 and 3 and Section 11.1. As before, we consider the space Ω of sequences of 0's and 1's and we fix a parameter p representing the probability of 1.

A *finite type random variable* is a real function on Ω that depends only on a finite number of coordinates. In other words, a function X from Ω to $\mathbb{R}$ is a finite type random variable if it takes only a finite number of values, and the set

$$(X = x) := \{\omega \in \Omega \ : \ X(\omega) = x\}$$

is a finite type event for each $x \in \mathbb{R}$. Every finite type random variable can be written as a finite linear combination of characteristic functions of finite type events.

Remark. Since we will only consider finite type random variables, we will simply use the phrase *random variables* to refer to finite type random variables.

If X and Y are two random variables, then the functions $X+Y$, XY, $\min(X,Y)$, and $\max(X,Y)$ are also random variables. If f is any function from $\mathbb{R}$ to itself and X is a random variable, then $f(X) := f \circ X$ is a random variable as well. For each $n \geq 1$, S_n is a random variable.

Note that a function X from Ω to $\mathbb{R}$ is a (finite type) random variable if and only if there exists an integer $n = n(X)$ and a random variable X' defined on the finite space Ω_n such that

$$X(\omega) = X'\left(\omega^{(n)}\right) \quad \text{where} \quad \omega^{(n)} := (\omega_1, \omega_2, \ldots, \omega_n) \tag{11.3}$$

for every $\omega \in \Omega$.

Of course the integer n is not uniquely determined, and any integer larger than n satisfies the above property if n does.

A family $(X_i)_{i \in I}$ of random variables is called a *family of independent random variables* if the events $(X_i = x_i)$ are independent for any choice of real numbers x_i. If $(X_i)_{i \in I}$ is a finite family of random variables, we can choose an integer $n = n(X_i)$ that satisfies (11.3) for each i and conclude that the random variables $(X_i)_{i \in I}$ are independent if and only if the random variables $(X_i')_{i \in I}$ defined on the finite probability space (Ω_n, P_n) are independent. Proposition 3.1 then implies the following fact: if $(X_i)_{i \in I}$ is a family of independent random variables, J and K are disjoint subsets of the set I of indices, and Y and Z are real functions respectively of $(X_i)_{i \in J}$ and $(X_i)_{i \in K}$, then Y and Z are independent random variables.

If X is a random variable, we let $X(\Omega)$ be the set of values taken by X. The *expected value* of X is defined by

$$E[X] := \sum_{x \in X(\Omega)} x P(X = x).$$

Equation (11.3) allows us to write this as $E[X] = E_n[X']$.

The properties of expected value stated in Chapters 2 and 3 for finite probability spaces also hold true for (finite type) random variables defined on Ω.

The following proposition generalizes the method that we used to prove the strong law of large numbers.

Proposition 11.10. *Let $(X_n)_{n\geq 1}$ be a sequence of random variables. If $\sum_{n=1}^{+\infty} P(X_n > \epsilon)$ converges for all $\epsilon > 0$, then $\lim_{n\to+\infty} X_n = 0$ almost surely.*

Proof. Set $A_n := (X_n > \epsilon)$. Proposition 11.9 implies that for every $\epsilon > 0$ there almost surely exists an $n_0(\omega, \epsilon) \geq 0$ such that $|X_n(\omega)| \leq \epsilon$ for every $n \geq n_0(\omega, \epsilon)$. By considering a countable union of negligible events, we conclude that for every positive integer m there almost surely exists $n_0(\omega, 1/m) \geq 0$ such that $|X_n(\omega)| \leq 1/m$ for all $n \geq n_0(\omega, 1/m)$. This implies that the sequence (X_n) almost surely converges to zero. □

When combined with Markov's inequality, this proposition directly implies the following highly useful convergence criterion.

Corollary 11.11. *Let $(X_n)_{n\geq 0}$ be a sequence of random variables. If $\sum_{n=0}^{+\infty} E[|X_n|]$ converges, then the sequence (X_n) almost surely converges to zero.*

As an illustration of this criterion for almost sure convergence, we will now give Cantelli's proof of Borel's strong law of large numbers. After that, we will present a generalized form of the strong law of large numbers.

For each positive integer n, let X_n be the random variable defined by $X_n(\omega) = \omega_n - p$, where $\omega = (\omega_1, \omega_2, \omega_3, \ldots)$. The expected value of each X_n is 0, and the set $\{X_n : n \geq 1\}$ is a family of independent random variables (we say that $(X_n)_{n\geq 1}$ is a *sequence of independent random variables*).

We have

$$\begin{aligned} E\left[\left(\frac{S_n}{n} - p\right)^4\right] &= \frac{1}{n^4} E\left[\left(\sum_{i=1}^{n} X_i\right)^4\right] \\ &= \frac{1}{n^4} \sum_{1\leq i,j,k,\ell\leq n} E\left[X_i X_j X_k X_\ell\right]. \end{aligned}$$

If i does not equal j, k, or ℓ, then X_i is independent from $X_jX_kX_\ell$, so $E(X_iX_jX_kX_\ell) = E(X_i)E(X_jX_kX_\ell) = 0$. This allows us to eliminate many terms from the above expansion, which yields

$$E\left[\left(\frac{S_n}{n} - p\right)^4\right] = \frac{1}{n^4}\left(\sum_{1\le i\le n} E\left[X_i^4\right] + 6\sum_{1\le i<j\le n} E\left[X_i^2X_j^2\right]\right).$$

Since $|X_i| \le 1$, it follows that $E\left[X_i^4\right] \le 1$ and $E\left[X_i^2X_j^2\right] \le 1$. Thus

$$E\left[\left(\frac{S_n}{n} - p\right)^4\right] \le \frac{1}{n^4}\left(n + 3n(n-1)\right) = O\left(\frac{1}{n^2}\right),$$

which is the general term of a convergent series. Corollary 11.11 then implies that the sequence $\left(\left(\frac{S_n}{n} - p\right)^4\right)$ almost surely converges to 0, which is the conclusion of Borel's strong law of large numbers.

The law of large numbers for the game of Heads or Tails, which we have stated in several forms, is the archetype of a large class of theorems in probability theory. The following theorem is a simple example of a widely applicable law of large numbers. Theorems 5.1 and 11.3 are special cases of this theorem.

Theorem 11.12. *Let $(X_n)_{n\ge1}$ be a sequence of pairwise independent random variables such that $E[X_n] = 0$ for each n and that $\sup_{n\ge1} E[X_n^2]$ is finite. Let $R_n = \sum_{i=1}^n X_i$. Then*

$$\lim_{n\to+\infty} P\left(\left|\frac{R_n}{n}\right| \ge \epsilon\right) = 0 \tag{11.4}$$

for each $\epsilon > 0$ and

$$\lim_{n\to+\infty} \frac{R_n}{n} = 0 \tag{11.5}$$

almost surely.

Note that the random variables X_n are pairwise independent if they form a sequence of independent random variables (but, as we noted in Chapter 3, not every sequence of pairwise independent variables is a sequence of independent variables).

Proof. Let $M := \sup_{n\geq 1} E\left[X_n^2\right]$. If i and j are distinct, then $E[X_iX_j] = E[X_i]E[X_j] = 0$. This implies that

$$E\left[\left(\frac{R_n}{n}\right)^2\right] = \frac{1}{n^2}\sum_{i=1}^{n} E\left[X_i^2\right] \leq \frac{M}{n}, \tag{11.6}$$

and (11.4) follows by the Bienaymé–Chebyshev inequality. The result about almost sure convergence is more subtle. Equation (11.6) implies that

$$\sum_{n\geq 1} E\left[\left(\frac{R_{n^2}}{n^2}\right)^2\right] < \infty.$$

Consequently, by Corollary 11.11,

$$\lim_{n\to+\infty} \frac{R_{n^2}}{n^2} = 0$$

almost surely. Let m be the integer part of $\sqrt{n}$. Then $m^2 \leq n < (m+1)^2$. On one hand,

$$\lim_{n\to+\infty} \frac{R_{m^2}}{n} = 0$$

almost surely, and on the other hand,

$$E\left[\left(\frac{R_n}{n} - \frac{R_{m^2}}{n}\right)^2\right] = \frac{1}{n^2} E\left[\left(\sum_{i=m^2+1}^{n} X_i\right)^2\right]$$

$$= \frac{1}{n^2}\sum_{i=m^2+1}^{n} E\left[X_i^2\right] \leq \frac{n-m^2}{n^2} M = O\left(n^{-3/2}\right)$$

is the general term of a convergent series (with index n). Applying Corollary 11.11 implies that

$$\lim_{n\to+\infty} \frac{R_n}{n} - \frac{R_{m^2}}{n} = 0$$

almost surely. □

Proposition 11.13 (second Borel–Cantelli lemma). *Let $(A_n)_{n\geq 1}$ be a sequence of independent events. If $\sum_{n\geq 1} P(A_n)$ diverges, then A_n infinitely often is an almost sure event.*

Under the strong hypothesis that the events A_n are independent, the second Borel–Cantelli lemma provides two pieces of information: first, it affirms the converse of the first Borel–Cantelli lemma; second, it affirms that the event A_n infinitely often is either negligible or almost sure. This last fact is a special case of Kolmogorov's 0-1 law, which is discussed in more advanced texts.

Proof. The complement of the set A_n infinitely often is the set $\bigcup_{n\geq 1}\bigcap_{k\geq n} A_k^c$. To show that this set is a negligible event, we must show that $\bigcap_{k\geq n} A_k^c$ is a negligible event for any fixed n. For each $m \geq n$, let B_m be the event $B_m := \bigcap_{n\leq k\leq m} A_k^c$. Then $\bigcap_{k\geq n} A_k^c \subset B_m$ and

$$\begin{aligned} P(B_m) = \prod_{k=n}^{m} P(A_k^c) &= \prod_{k=n}^{m} (1 - P(A_k)) \\ &\leq \prod_{k=n}^{m} \exp\big(-P(A_k)\big) = \exp\left(-\sum_{k=n}^{m} P(A_k)\right), \end{aligned}$$

since $(A_n)_{n\geq 1}$ is a sequence of independent events. This quantity can be made arbitrarily small by choosing a large enough m, so $\bigcap_{k\geq n} A_k^c$ is a negligible event. □

Example. Consider the setting of the example immediately following the statement and proof of the first Borel Cantelli lemma (Proposition 11.9). If the sequences (k_n) and (ℓ_n) satisfy $k_n + \ell_n \leq k_{n+1}$ for all n, then the events A_n are independent. If it is also true that $\sum_{n\geq 1} p^{\ell_n}$ diverges, then the event A_n infinitely often is almost sure.

Example. How can we randomly choose a real number between 0 and 1 following a uniform probability distribution? This is a subtle question, and we would need to use the Lebesgue measure to answer it carefully. Nonetheless, we can provide a partial answer using the game of Heads or Tails. Fix the parameter p as 1/2. We associate a real number U in the interval $[0, 1]$ to each sequence $\omega \in \Omega$ by proceeding as in the previous section and set

$$U = \sum_{k=1}^{+\infty} \frac{\omega_k}{2^k}.$$

Now, how can we choose a sequence of independent real numbers? Steinhaus suggested the following way. Consider an infinite family $(E_n)_{n\geq 1}$ of pairwise disjoint infinite subsets of $\mathbb{N}^*$ (for example, E_1 could be the set of even numbers, E_2 the set of odd multiples of 3, E_3 the set of multiples of 5 but not of 2 or 3, and so on). Let $E_n = \{a(n,1), a(n,2), a(n,3), \ldots\}$ and set

$$U_n = \sum_{k=1}^{+\infty} \frac{\omega_{a(n,k)}}{2^k}.$$

This provides a sequence (U_n) of independent, uniformly distributed random real numbers between 0 and 1. Let (ϵ_n) be a sequence of positive real numbers. The first Borel–Cantelli lemma implies that if $\sum_n \epsilon_n$ converges, then it is almost sure that $U_n > \epsilon_n$ for large n. The second Borel–Cantelli lemma implies that if $\sum_n \epsilon_n$ diverges, then it is almost sure that $U_n < \epsilon_n$ for infinitely many n.

(Note that the random variables U are not finite type random variables; however, to remain within the scope of this book, we will not explain why the Borel–Cantelli lemmas can be applied in this situation.)

http://dx.doi.org/10.1090/stml/028/13

Chapter 12

The Law of the Iterated Logarithm

12.1. Introduction

In this chapter, we will study the rate of convergence in the strong law of large numbers. The setting is that of the preceding chapters: we consider the game of Heads or Tails with a coin that may be unfair. The probability of success (or heads) for each trial (one coin toss) is p, and the number of successes in n independent trials is S_n. Borel's strong law of large numbers implies that the sequence (S_n/n) almost surely approaches p as n approaches infinity, and the central limit theorem implies that $S_n - np$ is highly likely to have order at most $\sqrt{n}$ for large n.

Khinchin's law of the iterated logarithm adds considerable information to the strong law of large numbers by providing a very precise estimate of the size of the almost sure fluctuations of the sequence $(S_n - np)$. It states that almost surely for any $\epsilon > 0$,

there exist infinitely many n such that

$$S_n - np > (1-\epsilon)\sqrt{2p(1-p)n \ln \ln n},$$

and

$$\text{for every large enough } n,\ S_n - np < (1+\epsilon)\sqrt{2p(1-p)n \ln \ln n}.$$

These statements involve the *iterated logarithm* function ln ln, the composition of the base e logarithm with itself, which is an excellent example of a function that approaches infinity extremely slowly as n increases (use your calculator to see for yourself!).

Like the central limit theorem, the law of the iterated logarithm illustrates the fascinating fact that even complete randomness obeys precise laws.

We will present the results leading up to the law of the iterated logarithm in chronological order. We will prove the following results.

(1) Hausdorff's estimate:[1] Almost surely, for any $\epsilon > 0$,

$$S_n - np = O\left(n^{\epsilon+1/2}\right)$$

as $n \to +\infty$.

(2) Hardy and Littlewood's estimate:[2]

$$S_n - np = O\left(\sqrt{n \ln n}\right)$$

almost surely as $n \to +\infty$.

(3) Khinchin's law of the iterated logarithm.[3]

Our mathematical model based on the pair (Ω, P) is the one described in Section 11.1. We set

$$X_n(\omega) = \omega_n - p$$

and

$$R_n(\omega) = \sum_{k=1}^{n} X_k(\omega) = S_n(\omega) - np,$$

where $\omega = (\omega_n)_{n \geq 1} \in \Omega$. Thus (X_n) is a sequence of independent, identically distributed, finite type random variables with expected value 0. Along with the large and moderate deviations estimates, this is all we need to prove the results of this chapter.

---

[1] F. Hausdorff, *Grundzüge der Mengenlehre*, Leipzig, 1913.

[2] G. H. Hardy and J. E. Littlewood, *Some problems of Diophantine approximations*, Acta Mathematica, vol. 37, pp. 155–339, 1914.

[3] A. Khinchin, *Über einen Satz der Wahrscheinlichkeitsrechnung*, Fundamenta Mathematicae, vol. 6, pp. 9–20, 1924.

12.2. Hausdorff's Estimate

Proposition 12.1. *Almost surely, for any* $\epsilon > 0$,

$$S_n - np = O\left(n^{\epsilon+1/2}\right)$$

as $n \to +\infty$.

Proof. We will use an extension of Cantelli's method of proving the strong law of law numbers, which we presented after Corollary 11.11.

If $i_1, i_2, \ldots, i_k$ are positive integers, then $E\left[X_{i_1}X_{i_2}\cdots X_{i_k}\right] \leq 1$, and if one of the i_j's is distinct from the others, then $E\left[X_{i_1}X_{i_2}\ldots X_{i_k}\right]$ equals 0.

Let k be a positive integer. Then

$$E\left[R_n^{2k}\right] = \sum_{1\leq i_1, i_2, \ldots, i_{2k}\leq n} E\left[X_{i_1}X_{i_2}\ldots X_{i_{2k}}\right] \leq N(k,n),$$

where $N(k,n)$ is the number of functions from the set $\{1,2,\ldots,2k\}$ to the set $\{1,2,\ldots,n\}$ that take each value at least twice. Let $M(k)$ be the number of partitions of $\{1,2,\ldots,2k\}$ into subsets each containing at least two elements. If P is such a partition, then P contains at most k elements, and the number of functions from $\{1,2,\ldots,2k\}$ to $\{1,2,\ldots,n\}$ that are constant on each element of P is at most n^k. This implies that $N(k,n) \leq n^k M(k)$.

Let $\epsilon > 0$. Then

$$E\left[\left(n^{-\epsilon-1/2}R_n\right)^{2k}\right] \leq n^{-2k\epsilon-k}N(k,n) \leq n^{-2k\epsilon}M(k).$$

If we choose $k > \frac{1}{2\epsilon}$, then

$$\sum_{n\geq 1} E\left[\left(n^{-\epsilon-1/2}R_n\right)^{2k}\right] < \infty.$$

By Corollary 11.11, then the sequence $(n^{-\epsilon-1/2}R_n)$ almost surely approaches 0 as n approaches infinity.

For each $\epsilon > 0$, there is a negligible event outside of which $n^{-\epsilon-1/2}R_n$ converges to 0. To complete the proof of Proposition 12.1, we consider a countable family of values of ϵ. Since a countable union of negligible events is negligible, then for each $\epsilon > 0$, $n^{-\epsilon-1/2}R_n$ converges to 0 on the complement of a negligible event. □

12.3. Hardy and Littlewood's Estimate

This result follows easily from the large deviations estimate.

Proposition 12.2.

$$S_n - np = O\left(\sqrt{n \ln n}\right)$$

almost surely as $n \to +\infty$.

Proof. Theorem 6.1 implies that

$$P\left(R_n > \sqrt{n \ln n}\right) \leq \exp\left(-nh_+\left(\sqrt{\frac{\ln n}{n}}\right)\right),$$

where

$$h_+(\epsilon) = \frac{\epsilon^2}{2p(1-p)} + O\left(\epsilon^3\right)$$

as ϵ approaches 0. As n approaches infinity,

$$h_+\left(\sqrt{\frac{\ln n}{n}}\right) = \frac{\ln n}{2p(1-p)n} + o\left(\frac{1}{n}\right)$$

and

$$\exp\left(-nh_+\left(\sqrt{\frac{\ln n}{n}}\right)\right) \sim \exp\left(-\frac{\ln n}{2p(1-p)}\right) = n^{-1/(2p(1-p))},$$

which is the general term of a convergent series since $\frac{1}{2p(1-p)} \geq 2$.

Thus

$$\sum_{n\geq 1} P\left(R_n > \sqrt{n \ln n}\right) < \infty.$$

The first Borel–Cantelli lemma (Proposition 11.9) then implies that, almost surely, $R_n \leq \sqrt{n \ln n}$ for large n. □

12.4. Khinchin's Law of the Iterated Logarithm

Theorem 12.3. *Almost surely,*

$$\limsup_{n\to\infty} \frac{S_n - np}{\sqrt{2p(1-p)n \ln \ln n}} = +1$$

and

$$\liminf_{n\to\infty} \frac{S_n - np}{\sqrt{2p(1-p)n \ln \ln n}} = -1.$$

We will need the following two lemmas to prove this result. The first follows from the large and moderate deviations estimates. The second is a simple example of a *maximal inequality*; such inequalities, which were first studied by Kolmogorov, play a fundamental role in proving almost sure asymptotic results.

For each integer $n > 1$, we set

$$\alpha(n) := \sqrt{2p(1-p)n \ln \ln n}.$$

Lemma 12.4. *For all positive numbers a and δ and large enough n,*

$$(\ln n)^{-a^2(1+\delta)} < P\left(R_n \geq a\alpha(n)\right) < (\ln n)^{-a^2(1-\delta)}.$$

Lemma 12.5. *Suppose that $(Y_n)_{n\geq 1}$ is a sequence of independent random variables with expected value 0 and variance σ^2.*

Let $T_n := Y_1 + Y_2 + \cdots + Y_n$. Then

$$P\left(\max_{1\leq k\leq n} T_k \geq b\right) \leq \frac{4}{3} \cdot P\left(T_n \geq b - 2\sigma\sqrt{n}\right)$$

for every $b \in \mathbb{R}$.

Proof of Lemma 12.4. Theorem 6.1 implies that

$$P\left(R_n \geq a\alpha(n)\right) \leq \exp\left(-nh_+\left(\frac{a\alpha(n)}{n}\right)\right), \tag{12.1}$$

and, since the sequence $(\frac{\alpha(n)}{n})$ approaches zero, Proposition 6.2 implies that

$$h_+\left(\frac{a\alpha(n)}{n}\right) = \frac{a^2}{2p(1-p)}\left(\frac{\alpha(n)}{n}\right)^2 + O\left(\left(\frac{\alpha(n)}{n}\right)^3\right).$$

Thus

$$h_+\left(\frac{a\alpha(n)}{n}\right) = a^2\frac{\ln\ln n}{n} + O\left(\left(\frac{\ln\ln n}{n}\right)^{3/2}\right)$$

and

$$nh_+\left(\frac{a\alpha(n)}{n}\right) \geq a^2(1-\delta)\ln\ln n$$

for large enough n. Combining this with the inequality (12.1), we conclude that

$$P\left(R_n \geq a\alpha(n)\right) \leq (\ln n)^{-a^2(1-\delta)}$$

for large enough n.

Theorem 8.1, which we can apply since $\sqrt{\ln \ln n} = o(n^{1/6})$, implies that

$$P\left(R_n \geq a\alpha(n)\right) = P\left(\frac{S_n}{n} - p \geq \sqrt{\frac{p(1-p)}{n}}\, a\sqrt{2\ln\ln n}\right)$$
$$\sim \frac{1}{\sqrt{2\pi}a\sqrt{2\ln\ln n}} \exp\left(-a^2 \ln\ln n\right)$$
$$= \frac{1}{2a\sqrt{\pi \ln\ln n}} \left(\ln n\right)^{-a^2}.$$

Since $\sqrt{\ln\ln n} = o((\ln n)^{a^2\delta})$, we conclude that

$$P\left(R_n \geq a\alpha(n)\right) \geq \left(\ln n\right)^{-a^2(1+\delta)}$$

for large enough n. □

Proof of Lemma 12.5. Since the random variables Y_n are independent, $\operatorname{var}(T_n - T_k) = (n-k)\sigma^2$ for every $1 \leq k \leq n$. Then the Bienaymé–Chebyshev inequality implies that

$$P\left(|T_n - T_k| \leq 2\sigma\sqrt{n}\right) \geq 1 - \frac{\operatorname{var}(T_n - T_k)}{4\sigma^2 n} = 1 - \frac{n-k}{4n} \geq \frac{3}{4}.$$

Using elementary properties of probability, we can write

$$P\left(\max_{1\leq k\leq n} T_k \geq b\right)$$
$$= \sum_{k=1}^{n} P\left(T_1 < b, T_2 < b, \ldots, T_{k-1} < b \text{ and } T_k \geq b\right),$$

which is bounded above by

$$\frac{4}{3}\sum_{k=1}^{n} P\left(T_1 < b, \ldots, T_{k-1} < b \text{ and } T_k \geq b\right) \cdot P\left(|T_n - T_k| \leq 2\sigma\sqrt{n}\right).$$

Now, the random variable $T_n - T_k$ is independent of the random variables $T_1, T_2, \ldots, T_k$ for every k. Therefore, the above quantity can be written as

$$\frac{4}{3}\sum_{k=1}^{n} P\left(T_1 < b, \ldots, T_{k-1} < b, T_k \geq b \text{ and } |T_n - T_k| \leq 2\sigma\sqrt{n}\right).$$

Because the event $(T_1 < b, \ldots, T_{k-1} < b, T_k \geq b \text{ and } T_n \geq b - 2\sigma\sqrt{n})$ contains $(T_1 < b, \ldots, T_{k-1} < b, T_k \geq b \text{ and } |T_n - T_k| \leq 2\sigma\sqrt{n})$,

$$P\left(\max_{1\leq k\leq n} T_k \geq b\right)$$
$$\leq \frac{4}{3}\sum_{k=1}^{n} P\left(T_1 < b, \ldots, T_{k-1} < b, T_k \geq b \text{ and } T_n \geq b - 2\sigma\sqrt{n}\right).$$

Therefore

$$P\left(\max_{1\leq k\leq n} T_k \geq b\right) \leq \frac{4}{3} P\left(T_n \geq b - 2\sigma\sqrt{n}\right).$$

□

We will now tackle the proof of the law of the iterated logarithm. We just need to prove that

$$\limsup_{n\to\infty} \frac{S_n - np}{\alpha(n)} = +1,$$

since the calculation of the infimum limit follows by replacing S_n by $n - S_n$ and interchanging p with $1 - p$.

The proof will be composed of two steps. For each $\eta > 0$, we will show that

(i) The supremum limit is almost surely less than $1 + \eta$.

(ii) The supremum limit is almost surely greater than $1 - \eta$.

From these steps, we will reach the desired result by letting η approach 0: we associate an almost sure event to each η, consider a sequence of values η approaching 0, and use the fact that a countable union of negligible events is negligible.

As an introduction to the proof, we will provide a simplified argument that yields a partial result. Fix a number $\gamma > 1$ and let $n_k = [\gamma^k]$ be the integer part of γ^k. By Lemma 12.4,

$$P\left(R_{n_k} \geq (1+\eta)\alpha(n_k)\right) < (\ln n_k)^{-(1+\eta)^2(1-\delta)}$$

for every $\delta > 0$ and large enough k. Therefore

$$P\left(R_{n_k} \geq (1+\eta)\alpha(n_k)\right) = O\left(k^{-(1+\eta)^2(1-\delta)}\right).$$

If δ is chosen to be small enough that $(1+\eta)^2(1-\delta) > 1$, then

$$\sum_{k\geq 0} P\left(R_{n_k} \geq (1+\eta)\alpha(n_k)\right) < \infty,$$

and the first Borel–Cantelli lemma (Proposition 11.9) implies that

$$\limsup_{k\to+\infty} \frac{R_{n_k}}{\alpha(n_k)} \leq 1+\eta$$

almost surely.

We have thus proven step (i) for subsequences with exponential growth. It will take some more work to prove the inequality in general. We only need to consider a subsequence to prove step (ii); however, this step will not be easy either because we will need to use the second Borel–Cantelli lemma (Proposition 11.13), which requires a hypothesis of independence.

Proof of the Law of the Iterated Logarithm. Fix $\eta > 0$. For a real number $\gamma > 1$ (to be chosen later) and for each $k \in \mathbb{N}$, set $n_k := [\gamma^k]$. We will show that

$$\sum_{k\geq 0} P\left(\max_{n\leq n_{k+1}} R_n \geq (1+\eta)\alpha(n_k)\right) < \infty. \tag{12.2}$$

Lemma 12.5 implies that

$$P\left(\max_{n\leq n_{k+1}} R_n \geq (1+\eta)\alpha(n_k)\right) \\ \leq \frac{4}{3} P\left(R_{n_{k+1}} \geq (1+\eta)\alpha(n_k) - 2\sqrt{n_{k+1}p(1-p)}\right).$$

Since $\sqrt{n_{k+1}} = o\left(\alpha(n_k)\right)$ as $k \to +\infty$,

$$2\sqrt{n_{k+1}p(1-p)} < \frac{1}{2}\eta\alpha(n_k)$$

for all large enough k. For such a k,

$$P\left(\max_{n\leq n_{k+1}} R_n \geq (1+\eta)\alpha(n_k)\right) \leq \frac{4}{3} P\left(R_{n_{k+1}} \geq (1+\eta/2)\alpha(n_k)\right).$$

We know that $\alpha(n_{k+1}) \sim \sqrt{\gamma}\alpha(n_k)$, and we choose a γ that satisfies

$(1+\eta/2) > (1+\eta/4)\sqrt{\gamma}$. Then, for large enough k, $(1+\eta/2)\alpha(n_k) > (1+\eta/4)\alpha(n_{k+1})$, and therefore

$$P\left(\max_{n \leq n_{k+1}} R_n \geq (1+\eta)\alpha(n_k)\right) \leq \frac{4}{3} P\left(R_{n_{k+1}} \geq (1+\eta/4)\alpha(n_{k+1})\right).$$

Now, we apply Lemma 12.4 (using $a = (1-\delta)^{-1} = (1+\eta/4)$) to find an upper bound for this expression. This yields

$$P\left(\max_{n \leq n_{k+1}} R_n \geq (1+\eta)\alpha(n_k)\right) \leq \frac{4}{3}\left(\ln n_{k+1}\right)^{-(1+\eta/4)}$$

for large enough k. Finally,

$$\left(\ln n_{k+1}\right)^{-(1+\eta/4)} \sim \left(\ln \gamma\right)^{-(1+\eta/4)} k^{-(1+\eta/4)},$$

which is the general term of a convergent series.

We have thus proven (12.2). Then the first Borel–Cantelli lemma implies that

$$\max_{n \leq n_{k+1}} R_n < (1+\eta)\alpha(n_k)$$

and, in particular, we have almost surely that

$$\max_{n_k \leq n < n_{k+1}} R_n < (1+\eta)\alpha(n_k)$$

for large enough k. This implies that, almost surely,

$$R_n < (1+\eta)\alpha(n)$$

for large enough n, which establishes (i).

To prove (ii), it suffices to show that there exists a subsequence (n_k) of the sequence of all integers such that $(R_{n_k} \geq (1-\eta)\alpha(n_k))$ *infinitely often* is an almost sure event. We will choose a subsequence satisfying $n_k = \gamma^k$, where γ is a large enough number (to be chosen later). We will show that

$$\sum_{n \geq 1} P\left(R_{\gamma^n} - R_{\gamma^{n-1}} \geq \left(1 - \frac{\eta}{2}\right)\alpha\left(\gamma^n\right)\right) = \infty \tag{12.3}$$

and that

$$\text{almost surely, } R_{\gamma^{n-1}} \geq -\frac{\eta}{2}\alpha\left(\gamma^n\right) \text{ for large enough } n. \tag{12.4}$$

The random variable $R_{\gamma^n} - R_{\gamma^{n-1}}$ has the same probability distribution as the random variable $R_{\gamma^n - \gamma^{n-1}}$. Thus

$$P\left(R_{\gamma^n} - R_{\gamma^{n-1}} \geq (1 - \tfrac{\eta}{2})\alpha\left(\gamma^n\right)\right) \\ = P\left(R_{\gamma^n - \gamma^{n-1}} \geq (1 - \tfrac{\eta}{2})\alpha\left(\gamma^n\right)\right).$$

As n approaches infinity,

$$\frac{\alpha\left(\gamma^n - \gamma^{n-1}\right)}{\alpha\left(\gamma^n\right)} \sim \sqrt{\frac{\gamma - 1}{\gamma}}.$$

If we choose γ such that $(1 - \eta/2)\sqrt{\gamma} < (1 - \eta/4)\sqrt{\gamma - 1}$, then

$$(1 - \eta/2)\,\alpha\left(\gamma^n\right) < (1 - \eta/4)\,\alpha\left(\gamma^n - \gamma^{n-1}\right)$$

for large enough n, and for such n,

$$P\left(R_{\gamma^n} - R_{\gamma^{n-1}} \geq (1 - \tfrac{\eta}{2})\alpha\left(\gamma^n\right)\right) \\ \geq P\left(R_{\gamma^n - \gamma^{n-1}} \geq (1 - \tfrac{\eta}{4})\alpha\left(\gamma^n - \gamma^{n-1}\right)\right).$$

Now, we use Lemma 12.4 (with $a = (1 + \delta)^{-1} = (1 - \eta/4)$) to find a lower bound for this expression. This yields

$$P\left(R_{\gamma^n} - R_{\gamma^{n-1}} \geq \left(1 - \frac{\eta}{2}\right)\alpha\left(\gamma^n\right)\right) \geq \left(\ln\left(\gamma^n - \gamma^{n-1}\right)\right)^{-1+\eta/4}$$

for every large enough n. Since

$$\left(\ln\left(\gamma^n - \gamma^{n-1}\right)\right)^{-1+\eta/4} \sim (n\ln(\gamma))^{-1+\eta/4}$$

is the general term of a divergent series, this proves (12.3).

As n approaches infinity, $\alpha\left(\gamma^n\right) \sim \sqrt{\gamma}\,\alpha\left(\gamma^{n-1}\right)$. If we choose γ large enough so that $\eta\sqrt{\gamma} > 4$, then $\eta\alpha\left(\gamma^n\right) > 4\alpha\left(\gamma^{n-1}\right)$ for large enough n. For such an n,

$$\left(R_{\gamma^{n-1}} \leq -\frac{\eta}{2}\alpha\left(\gamma^n\right)\right) \subset \left(-R_{\gamma^{n-1}} \geq 2\alpha\left(\gamma^{n-1}\right)\right).$$

Now, applying the upper bound (i) established in the first part of this proof to the sequence $(-R_n)$ implies that, almost surely, $-R_{\gamma^{n-1}} < 2\alpha\left(\gamma^{n-1}\right)$ for large enough n. This proves (12.4).

Since $\left(R_{\gamma^n} - R_{\gamma^{n-1}}\right)$ is a sequence of independent random variables and the sum in (12.3) diverges, the second Borel–Cantelli lemma implies that, almost surely, there exist infinitely many positive integers n such that $R_{\gamma^n} - R_{\gamma^{n-1}} > \left(1 - \frac{\eta}{2}\right)\alpha\left(\gamma^n\right)$. Combined with

(12.4), this implies that there almost surely exist infinitely many positive integers n such that $R_{\gamma^n} > (1 - \eta)\,\alpha\,(\gamma^n)$. This establishes the lower bound (ii). □

http://dx.doi.org/10.1090/stml/028/14

Chapter 13

Recurrence of Random Walks

13.1. Introduction and Definitions

In this chapter we discuss the recurrence of random walks in the discrete space $\mathbb{Z}^N$ with dimension $N \geq 1$. A *random walk* describes the successive positions of a person who steps a random distance in a random direction at each second. We suppose that each of these random steps is independent of the previous steps and that all the random steps follow the same probability distribution. Although we can define a random walk in an arbitrary group, we will limit ourselves to the case of the additive groups $\mathbb{Z}^N$, and we will assume that the length of each step is bounded.

Assuming that the walk consists of infinitely many steps, we ask the following questions: Will the walk return to the starting point? Will the walk reach every point in the space? Will the walk return infinitely many times to the starting point?

We will see that the answers to these questions basically depend on only two parameters: the average displacement at each step and the dimension N of the space in which the walk takes place. We will find the following results: if the expected length of each step is not zero, then the walk almost surely goes to infinity; if the expected length of each step is zero, then the answers to the above questions are affirmative if the walk is in a space of dimension at most two.

The mathematical model that we will use is the one we have described in detail in the previous chapters, because the steps in a random walk form a sequence of identical and independent elementary random experiments.

The *nearest neighbor random walk* on $\mathbb{Z}$ is the walk composed of fixed-length steps, each either forward or backward. Of course, this is the same as our game of Heads or Tails: at each second, the random walker tosses a coin to decide whether to take a step forward or backward. Formally, we consider the space (Ω, P) described in Section 11.1. For $\omega = (\omega_n)_{n\geq 1} \in \Omega = \{0,1\}^{\mathbb{N}^*}$ we set $Y_n(\omega) = 2\omega_n - 1$. Then (Y_n) is a sequence of independent random variables such that

$$P(Y_n = 1) = p \quad \text{and} \quad P(Y_n = -1) = 1 - p.$$

The *random walk* is the sequence of random variables $(M_n)_{n\geq 1}$ defined by

$$M_n := Y_1 + Y_2 + \cdots + Y_n.$$

The parameter p, a real number between 0 and 1, is the probability of taking a step forward. The variable M_n is connected to other variables that we have studied by the equality $M_n = 2S_n - n = 2R_n + (2p-1)n$.

More generally, a random walk on $\mathbb{Z}^N$ composed of a finite number of possible steps can be described as follows: let $(e_1, e_2, \ldots, e_d)$ be a finite family of elements of $\mathbb{Z}^N$ and $(p_1, p_2, \ldots, p_d)$ be a family of real positive numbers that sum to 1. (Each e_i represents an allowed step, and p_i is the probability of taking the step represented by e_i.) Consider the probability space (Ω, P) described in Section 11.3, and let $Y_n(\omega) = e_{\omega_n}$ for each $\omega = (\omega_n) \in \Omega = \{1, 2, \ldots, d\}^{N^*}$. Then the *random walk* is the sequence of random variables $(M_n)_{n\geq 1}$ defined by

$$M_n := Y_1 + Y_2 + \cdots + Y_n .$$

Before proceeding, we must give a precise definition of the concept of recurrence. A random walk $(M_n)_{n\geq 1}$ is called *recurrent* if the walk almost surely returns to the origin infinitely many times; that is, a random walk is recurrent if the set $(M_n = 0)$ *infinitely often* is an almost sure event. A random walk is *transient* if the walk almost surely returns to the starting point only finitely many times; that is, a random walk is transient if the set $(M_n = 0)$ infinitely often is a negligible event.

We will also use the following definition: a random walk is *centered* if the expected length of a step is zero; that is, if $E[Y_n] := \sum_{i=1}^{d} p_i e_i = 0$.

The rest of this chapter is composed of three sections. In Section 13.2, we will describe the properties of recurrence of the nearest neighbor random walk on $\mathbb{Z}$; since this walk is the same as the game of Heads or Tails, we will be able to use the results of the preceding chapters. In Section 13.3, we will study the properties of random walks in a more general setting. In Section 13.4, we will describe the properties of recurrence of random walks on $\mathbb{Z}^N$.

13.2. Nearest Neighbor Random Walks on $\mathbb{Z}$

Here the random variables Y_n take the values $+1$ and -1 with probability p and $1-p$ respectively, and $M_n = Y_1 + Y_2 + \cdots + Y_n$. If $p > 1/2$ (respectively, $p < 1/2$), the strong law of large numbers implies that M_n almost surely approaches infinity (respectively, negative infinity) as n approaches infinity. Thus the random walk is transient if $p \neq 1/2$.

If $p = 1/2$, the random walk is called the *simple random walk* on $\mathbb{Z}$. This is a centered random walk. We can show that the walk is recurrent is several ways; this follows, for example, from the law of returns to zero (Theorem 10.2) or from the law of the iterated logarithm (Theorem 12.3). The law of the iterated logarithm implies that $\limsup M_n = +\infty$ and $\liminf M_n = -\infty$ almost surely. Since the length of each step is 1, this implies that the random walk almost surely reaches each point in $\mathbb{Z}$ infinitely many times.

Do these arguments generalize to other random walks? Only partially. If the random walk is not centered, then the generalized law of large numbers (Theorem 11.12) immediately implies that the walk is transient. The question of recurrence, however, is more subtle. For example, consider the random walk (M_n) on $\mathbb{Z}$ with probability distribution $P(Y_n = -3) = 2/5$ and $P(Y_n = 2) = 3/5$. This is a centered random walk, and the law of the iterated logarithm implies that $\limsup M_n = +\infty$ and $\liminf M_n = -\infty$ almost surely; however, we cannot conclude only from this information that the walk is recurrent. The same discussion applies to any centered random walk on $\mathbb{Z}^N$ where $N \geq 2$.

13.3. General Results about Random Walks

The most striking general property of random walks is that every random walk is either recurrent or transient. This all-or-nothing principle is the aim of Theorem 13.1. The connection between the number of times that the random walk returns to the starting point and the number of times that the random walk reaches any given point is stated in Proposition 13.4.

Theorem 13.1. *Every random walk $(M_n)_{n\geq 1}$ is either recurrent or transient, and the following three statements are equivalent:*

(i) *The random walk is recurrent.*

(ii) $\lim_{n\to+\infty} P\,(\textit{there exists } k \leq n \textit{ such that } M_k = 0) = 1.$

(iii) $\sum_{n=1}^{+\infty} P\,(M_n = 0) = +\infty.$

It is easy to check that property (ii) implies that the random walk almost surely returns at least once to the starting point. In fact, the converse is also true: property (ii) is equivalent to the almost sure return of the walk to the origin. However, we will not provide a proof of this fact.

In addition, note the parallel between the implication (iii) $\Rightarrow$ (ii) and the second Borel–Cantelli lemma. However, the lemma cannot be directly applied because the random variables M_n are not independent.

If m, s, and t are positive integers with $s \leq t$, let $A^m_{s,t}$ be the event consisting of random walks that return to the starting point at least m times between steps s and t. In symbols,

$$\omega \in A^m_{s,t} \iff \#\{n \ :\ s \leq n \leq t \text{ and } M_n(\omega) = 0\} \geq m.$$

We will use the following two lemmas to prove Theorem 13.1.

Lemma 13.2. $P\left(A_{1,t}^m\right) \leq \left(P\left(A_{1,t}^1\right)\right)^m$ *for every* $m, t > 0$.

Lemma 13.3. $\left(P\left(A_{1,t}^1\right)\right)^m \leq P\left(A_{1,mt}^m\right)$ *for every* $m, t > 0$.

Proof of Lemma 13.2. It is clear that $P\left(A_{1,t}^m\right) = 0$ if $m > t$. The lemma is trivial for $m = 1$, and we proceed by induction on m. To simplify the exposition, we will only show the first induction step; the general step is not difficult once the first step is understood. We will thus prove that the lemma is true for $m = 2$. By considering the first two returns to the origin, we can write the event $A_{1,t}^2$ as a finite union of pairwise disjoint events. Thus

$$P\left(A_{1,t}^2\right) = \sum_{1 \leq j < k \leq t} P\big((M_i \neq 0,\ 1 \leq i < j) \text{ and } (M_j = 0) \\ \text{ and } (M_i \neq 0,\ j < i < k) \text{ and } (M_k = 0)\big),$$

so

$$P\left(A_{1,t}^2\right) = \sum_{1 \leq j < k \leq t} P\big((M_i \neq 0,\ 1 \leq i < j) \text{ and } (M_j = 0) \\ \text{ and } (M_i - M_j \neq 0,\ j < i < k) \text{ and } (M_k - M_j = 0)\big).$$

Since the event $((M_i \neq 0, 1 \leq i < j) \text{ and } (M_j = 0))$ is independent from the event $((M_i - M_j \neq 0, j < i < k) \text{ and } (M_k - M_j = 0))$,

$$P\left(A_{1,t}^2\right) = \sum_{1 \leq j < k \leq t} P\big((M_i \neq 0,\ 1 \leq i < j) \text{ and } (M_j = 0)\big) \\ \times P\big((M_i - M_j \neq 0,\ j < i < k) \text{ and } (M_k - M_j = 0)\big).$$

Then, by the property of invariance under shifting (see Section 11.1),

$$P\big((M_i - M_j \neq 0,\ j < i < k) \text{ and } (M_k - M_j = 0)\big) \\ = P\big((M_{i-j} \neq 0,\ j < i < k) \text{ and } (M_{k-j} = 0)\big),$$

whence

(13.1)

$$P\left(A_{1,t}^2\right) = \sum_{1 \leq j < k \leq t} P\big((M_i \neq 0,\ 1 \leq i < j) \text{ and } (M_j = 0)\big) \\ \times P\big((M_i \neq 0,\ 1 \leq i < k - j) \text{ and } (M_{k-j} = 0)\big).$$

We conclude that

$$P\left(A_{1,t}^2\right) \leq \sum_{j=1}^{t}\sum_{\ell=1}^{t} P\big((M_i \neq 0,\ 1 \leq i < j) \text{ and } (M_j = 0)\big) \\ \times P\big((M_i \neq 0,\ 1 \leq i < \ell) \text{ and } (M_\ell = 0)\big).$$

By using a decomposition similar to the one used in the beginning of this calculation, we also have that

$$P\left(A_{1,t}^1\right) = \sum_{j=1}^{t} P\big((M_i \neq 0,\ 1 \leq i < j) \text{ and } (M_j = 0)\big),$$

and thus $P\left(A_{1,t}^2\right) \leq \left(P\left(A_{1,t}^1\right)\right)^2$. □

Proof of Lemma 13.3. We prove this lemma by induction on m. As for Lemma 13.2, we will only study the case of $m = 2$.

We can write equation (13.1) as

$$P\left(A_{1,2t}^2\right) = \sum_{1 \leq j < k \leq 2t} \Big[P\big((M_i \neq 0,\ 1 \leq i < j) \text{ and } (M_j = 0)\big) \\ \times P\big((M_i \neq 0,\ 1 \leq i < k-j) \text{ and } (M_{k-j} = 0)\big)\Big],$$

which implies that

$$P\left(A_{1,2t}^2\right) \geq \sum_{j=1}^{t}\sum_{k=j+1}^{j+t} P\big((M_i \neq 0,\ 1 \leq i < j) \text{ and } (M_j = 0)\big) \\ \times P\big((M_i \neq 0,\ 1 \leq i < k-j) \text{ and } (M_{k-j} = 0)\big).$$

By setting $\ell = k - j$, this becomes

$$P\left(A_{1,2t}^2\right) \geq \sum_{j=1}^{t} P\big((M_i \neq 0,\ 1 \leq i < j) \text{ and } (M_j = 0)\big) \\ \times \sum_{\ell=1}^{t} P\big((M_i \neq 0,\ 1 \leq i < \ell) \text{ and } (M_\ell = 0)\big),$$

which simplifies to

$$P\left(A_{1,2t}^2\right) \geq \left(P\left(A_{1,t}^1\right)\right)^2.$$

□

Proof of Theorem 13.1. If the series $\sum_{n\geq 1} P(M_n = 0)$ converges, then the set $(M_n = 0)$ infinitely often is a negligible event by the first Borel–Cantelli lemma; thus the random walk is transient, so it is not recurrent. This proves the implication (i)⇒(iii).

Next we will prove the implication (iii)⇒(ii). Since the sequence $\left(P\left(A^1_{1,n}\right)\right)_{n\geq 1}$ is increasing and bounded by 1, we can set

$$\rho = \lim_{n\to\infty} P\left(A^1_{1,n}\right) .$$

Then condition (ii) is simply $\rho = 1$. Since expected value is a linear function,

$$\sum_{k=1}^{n} P(M_k = 0) = E\left[\sum_{k=1}^{n} \chi_{(M_k=0)}\right] ,$$

so

$$\begin{aligned}\sum_{k=1}^{n} P(M_k = 0) &= E\left[\#\{j \ : \ 1 \leq j \leq n \text{ and } M_j = 0\}\right] \\ &= \sum_{j=1}^{n} j\, P\left((M_k)_{1\leq k\leq n} \text{ includes } j \text{ zeros}\right) \\ &= \sum_{j=1}^{n} j\left(P\left(A^j_{1,n}\right) - P\left(A^{j+1}_{1,n}\right)\right) = \sum_{j=1}^{n} P\left(A^j_{1,n}\right).\end{aligned}$$

Then Lemma 13.2 implies that

$$\sum_{k=1}^{n} P(M_k = 0) \leq \sum_{j=1}^{n} \left(P\left(A^1_{1,n}\right)\right)^j \leq \sum_{j=1}^{n} \rho^j ,$$

and thus

$$\sum_{k=1}^{+\infty} P(M_k = 0) = +\infty \Rightarrow \rho = 1 .$$

This proves that (iii)⇒(ii).

To complete the proof, we just need to prove that (ii)⇒(i). Let m be a positive integer, and let $A^m_{1,\infty}$ be the set of ω such that the sequence $(M_n(\omega))_{n\geq 1}$ contains at least m zeros. For each $t > 0$,

$$A^m_{1,\infty} \supset A^m_{1,mt} .$$

Then by Lemma 13.3,

$$P\left(A^m_{1,mt}\right) \geq \left(P\left(A^1_{1,t}\right)\right)^m .$$

Suppose that condition (ii) is satisfied. Then $\lim_{t\to\infty} P\left(A^1_{1,t}\right) = 1$, so the set $A^m_{1,\infty}$ contains events of probability arbitrarily close to 1. Therefore $A^m_{1,\infty}$ is an almost sure event.

The set of ω such that the sequence $\left(M_n(\omega)\right)_{n\geq 1}$ contains infinitely many zeros is the intersection of the sets $A^m_{1,\infty}$ as m ranges over $\mathbb{N}$. Since the countable intersection of almost sure events is an almost sure event (Proposition 11.1), we conclude that the sequence $(M_n)_{n\geq 1}$ almost surely contains infinitely many zeros. Thus the random walk is recurrent.

We have now established the equivalence between the conditions (i), (ii), and (iii). Thus the random walk is either recurrent or transient, according to whether $\sum_n P(M_n = 0)$ diverges or converges. □

Now we will study the number of times that a random walk reaches other points in the space. Let S be the semigroup generated by the set $E := \{e_1, e_2, \ldots, e_d\}$ of allowed steps in the random walk (we suppose that the all the probabilities p_i are positive). Thus S is the set of elements of $\mathbb{Z}^N$ that can be written as a finite sum of (not necessarily distinct) elements of E. Note that S is exactly the set of points attainable by the random walk.

Proposition 13.4. *If the random walk is recurrent, then S is a group and the random walk almost surely reaches every point in S infinitely many times.*

If the random walk is transient, then every point of S is almost surely only reached finitely many times; in other words, it is almost sure that $\lim_{n\to\infty} |M_n| = +\infty$.

Here $|\cdot|$ is any given norm on the space $\mathbb{R}^N$. In the statement of the following lemma, we use the *L-infinity norm* $|(x_1, x_2, \ldots, x_N)| = \max\{|x_1|, |x_2|, \ldots, |x_N|\}$.

We will use the following lemma in our proof of the proposition.

Lemma 13.5. *The random walk (M_n) on $\mathbb{Z}^N$ satisfies*

$$\sum_{n=1}^{+\infty} P(M_n = x) \leq 1 + \sum_{n=1}^{+\infty} P(M_n = 0)$$

for every $x \in \mathbb{Z}^N$ *and*

$$\sum_{n=1}^{+\infty} P\left(|M_n| \leq m\right) \leq (2m+1)^N \left(1 + \sum_{n=1}^{+\infty} P\left(M_n = 0\right)\right)$$

for every $m > 0$.

Proof. By decomposing the event $(M_n = x)$ into a union of pairwise disjoint events based on the first time that the random walk reaches the point x, we obtain

$$\begin{aligned} &P\left(M_n = x\right) \\ &= \sum_{k=1}^{n} P\left((M_i \neq x, 1 \leq i < k) \text{ and } (M_k = x) \text{ and } (M_n - M_k = 0)\right). \end{aligned}$$

Then, by using the independence of the successive steps of the random walk and the invariance of probability under shifting as we did in the proof of Lemma 13.2, we obtain

$$\begin{aligned} &P\left(M_n = x\right) \\ &\quad= \sum_{k=1}^{n} P\left((M_i \neq x, 1 \leq i < k) \text{ and } (M_k = x)\right) \cdot P\left(M_{n-k} = 0\right), \end{aligned}$$

which implies that

$$\begin{aligned} &\sum_{n=1}^{+\infty} P\left(M_n = x\right) \\ &\quad= \sum_{k=1}^{+\infty} \sum_{n=k}^{+\infty} P\left((M_i \neq x, 1 \leq i < k) \text{ and } (M_k = x)\right) \cdot P\left(M_{n-k} = 0\right) \\ &\quad= \sum_{k=1}^{+\infty} P\left((M_i \neq x, 1 \leq i < k) \text{ and } (M_k = x)\right) \cdot \sum_{n=0}^{+\infty} P\left(M_n = 0\right), \end{aligned}$$

where $M_0 = 0$. Since the events $((M_i \neq x, 1 \leq i < k) \text{ and } (M_k = x))$ are pairwise disjoint as k varies, the sum of their probabilities is at most 1. This proves the first stated inequality. The second follows immediately from the fact that

$$P\left(|M_n| \leq m\right) = \sum_{x \in \mathbb{Z}^N, |x| \leq m} P\left(M_n = x\right),$$

and that the cardinality of $\{x \in \mathbb{Z}^N, |x| \leq m\}$ is $(2m+1)^N$ for every positive integer m. □

Proof of Proposition 13.4. Suppose that the random walk (M_n) is transient. Theorem 13.1 implies that the series $\sum_{n\geq 1} P(M_n = 0)$ converges, and the above lemma implies that $\sum_{n\geq 1} P(|M_n| \leq m)$ converges from any $m > 0$. Then the first Borel–Cantelli lemma implies the second assertion of the proposition.

Now, suppose that the random walk (M_n) is recurrent.

Let $x \in S$, and fix a $k \geq 0$ such that $P(M_k = x) > 0$. There almost surely exists an $n > k$ such that $M_n = 0$. Since an almost sure event and a (finite type) event with positive probability cannot be disjoint, there exists an $\omega \in \Omega$ such that $M_k(\omega) = x$ and $M_n(\omega) = 0$, where $k < n$. Consequently, $-x \in S$. Since S is a semigroup by definition, then S is a group.

Now we need to show that if the random walk visits a point x with positive probability, then the random walk will almost surely return to x infinitely many times. The proof goes as follows. The random walk almost surely returns to zero infinitely many times. Each time the walk returns to zero, the situation is the same as if a new walk, independent of the previous one, were starting. This provides an infinite sequence of identical and independent experiments, where each one has a positive probability of reaching the point x. This implies that the random walk almost surely reaches the point x infinitely many times. We will now formalize this proof.

The previous theorem implies that $\lim_{t\to+\infty} P\left(A_{1,t}^1\right) = 1$. Then Lemma 13.3 implies that $\lim_{t\to+\infty} P\left(A_{1,mt}^m\right) = 1$ for any $m > 0$, and thus, by monotonicity,

$$\lim_{t\to+\infty} P\left(A_{1,t}^m\right) = 1.$$

If m, s, and t are positive integers, then $A_{s,t}^m \supset A_{1,t}^{m+s}$, so

$$\lim_{t\to+\infty} P\left(A_{s,t}^m\right) = 1.$$

Let $x \in S$ and $\epsilon > 0$. Fix $k \geq 1$ such that $P(M_k = x) > 0$ and set $\delta := 1 - P(M_k = x) < 1$. Define a sequence $(n_j)_{j \geq 0}$ by the recurrence

$$n_0 = 1; \quad n_j - n_{j-1} > k \quad \text{and} \quad P\left(A^1_{n_{j-1}, n_j - k}\right) > 1 - 2^{-j}\epsilon.$$

For $j \geq 1$, let B_j be the event *the random walk returns to the origin at least once between steps n_{j-1} and $n_j - k$ and does not visit x between steps n_{j-1} and n_j*; that is,

$$B_j := A^1_{n_{j-1}, n_j - k} \cap (M_n \neq x, n_{j-1} \leq n \leq n_j).$$

Let J and K be two positive integers such that $0 < J < K$. We will find an upper bound for the probability of the event *the random walk does not reach the point x between steps n_{J-1} and n_K*. First,

$$P(M_n \neq x, n_{J-1} \leq n \leq n_K) \\ \leq P\left(\bigcup_{j=J}^{K} (A^1_{n_{j-1}, n_j - k})^c\right) + P\left(\bigcap_{j=J}^{K} B_j\right)$$

and

$$P\left(\bigcup_{j=J}^{K} \left(A^1_{n_{j-1}, n_j - k}\right)^c\right) \leq \sum_{j=J}^{K} 2^{-j}\epsilon \leq 2^{1-J}\epsilon.$$

Let ℓ_j be a positive integer allowed to vary between n_{j-1} and $n_j - k$. If the event B_j occurs, then there is a unique integer ℓ_j such that

$$(M_\ell \neq 0, n_{j-1} \leq \ell < \ell_j), \quad M_{\ell_j} = 0 \quad \text{and} \quad M_{\ell_j + k} \neq x.$$

This implies that

$$P\left(\bigcap_{j=J}^{K} B_j\right) \\ \leq \sum_{\ell_J, \ldots, \ell_K} P\left(\bigcap_{j=J}^{K} \left(\left(A^1_{n_{j-1}, l_j - 1}\right)^c \cap (M_{\ell_j} = 0) \cap (M_{\ell_j + k} \neq x)\right)\right).$$

Since the random movement $M_{l_K+k} - M_{l_K}$ is independent of the previous movements in the random walk,

$$P\left(\bigcap_{j=J}^{K}\left(\left(A^1_{n_{j-1},l_j-1}\right)^c \cap (M_{\ell_j}=0)\cap(M_{\ell_j+k}\neq x)\right)\right)$$

$$= P\Bigg(\bigcap_{j=J}^{K-1}\left(\left(A^1_{n_{j-1},l_j-1}\right)^c \cap (M_{\ell_j}=0)\cap(M_{\ell_j+k}\neq x)\right)$$

$$\cap\left(A^1_{n_{K-1},l_K-1}\right)^c \cap (M_{\ell_K}=0)\Bigg)\times P\left(M_{\ell_K+k}-M_{\ell_K}\neq x\right),$$

and the property of invariance under shifting implies that

$$P\left(M_{\ell_K+k}-M_{\ell_K}\neq x\right) = P\left(M_k\neq x\right)=\delta.$$

We thus obtain the inequality

$$\sum_{\ell_J,\ldots,\ell_K} P\left(\bigcap_{j=J}^{K}\left(\left(A^1_{n_{j-1},l_j-1}\right)^c \cap (M_{\ell_j}=0)\cap(M_{\ell_j+k}\neq x)\right)\right)$$

$$\leq \delta \sum_{\ell_J,\ldots,\ell_{K-1}} P\left(\bigcap_{j=J}^{K-1}\left(\left(A^1_{n_{j-1},l_j-1}\right)^c \cap (M_{\ell_j}=0)\cap(M_{\ell_j+k}\neq x)\right)\right),$$

which implies, by recurrence, that

$$\sum_{\ell_J,\ldots,\ell_K} P\left(\bigcap_{j=J}^{K}\left(\left(A^1_{n_{j-1},l_j-1}\right)^c \cap (M_{\ell_j}=0)\cap(M_{\ell_j+k}\neq x)\right)\right) \leq \delta^{K-J+1}.$$

Combining the above inequalities yields

$$P\left(M_n\neq x, n_{J-1}\leq n\leq n_K\right)\leq 2^{1-J}\epsilon+\delta^{K-J+1}.$$

For each $J>0$, fix $K=K(J)>0$ such that $\delta^{K-J+1}\leq 2^{1-J}\epsilon$. The set E of ω such that $M_n(\omega)$ equals x for only finitely many n is contained in the union

$$\bigcup_{J>0}\left(M_n\neq x, n_{J-1}\leq n\leq n_{K(J)}\right)$$

and

$$\sum_{J>0} P\left(M_n \neq x, n_{J-1} \leq n \leq n_{K(J)}\right) < 4\epsilon$$

This proves that E is a negligible event.

We have thus proven that the random walk almost surely reaches the point x infinitely many times for any point $x \in S$. Since the set S is countable, we conclude that the random walk almost surely reaches every point in S. This proves the first assertion of the proposition. □

13.4. Recurrence of Random Walks on $\mathbb{Z}^N$

Proposition 13.6. *A random walk on $\mathbb{Z}$ is recurrent if and only if it is centered.*

Proof. The generalized law of large numbers (Theorem 11.12) implies that

$$\lim_{n\to+\infty} \frac{M_n}{n} = E\left[Y_1\right]$$

almost surely. If the walk is not centered, then $\lim_{n\to+\infty} |M_n| = +\infty$, so the walk is transient.

Now suppose that the walk is centered, and exclude the trivial case of all the random variables Y_n being identically equal to 0. For random variable Y_n taking only two values, we can use a variety of arguments to establish the proposition. For example, the de Moivre–Laplace theorem (Theorem 7.3) implies that

$$P\left(M_{2n} = 0\right) \sim \frac{1}{\sqrt{n}}$$

to within a constant multiple, which implies that the series with general term $P\left(M_n = 0\right)$ diverges. Theorem 13.1 then completes the proof. Another argument is based on the law of the iterated logarithm, which implies that

$$\limsup_{n\to+\infty} M_n = +\infty \quad \text{and} \quad \liminf_{n\to+\infty} M_n = -\infty$$

almost surely. It is impossible that $\lim_{n\to+\infty} |M_n| = +\infty$ since each step in the random walk has bounded length, so Theorem 13.1 and Proposition 13.4 together imply that the random walk is recurrent.

We will now give a clever argument[1] using the weak law of large numbers that applies to any centered random walk on $\mathbb{Z}$.

Lemma 13.5 implies that

$$1+\sum_{n=1}^{+\infty} P\left(M_n=0\right) \geq \frac{1}{2m+1} \sum_{n=1}^{+\infty} P\left(\left|M_n\right| \leq m\right),$$

so

$$1+\sum_{n=1}^{+\infty} P\left(M_n=0\right) \geq \frac{1}{2m+1} \sum_{n=1}^{ma} P\left(\left|M_n\right| \leq m\right),$$

and

$$1+\sum_{n=1}^{+\infty} P\left(M_n=0\right) \geq \frac{1}{2m+1} \sum_{n=1}^{ma} P\left(\left|M_n\right| \leq \frac{n}{a}\right)$$

for any positive integer a. Now, the generalized weak law of large numbers (Equation (11.4) of Theorem 11.12) implies that

$$\lim_{n \rightarrow+\infty} P\left(\left|M_n\right| \leq \frac{n}{a}\right)=1,$$

so

$$\lim_{m \rightarrow+\infty} \frac{1}{2m+1} \sum_{n=1}^{ma} P\left(\left|M_n\right| \leq \frac{n}{a}\right)=\frac{a}{2}.$$

Finally, this implies that

$$1+\sum_{n=1}^{+\infty} P\left(M_n=0\right) \geq \frac{a}{2}.$$

Since a is an arbitrary integer, this means that the series with general term $P\left(M_n=0\right)$ diverges, and thus the random walk is recurrent. □

We will now study random walks in dimension $N>1$. Let $Y_n=\left(Y_n^1, Y_n^2, \ldots, Y_n^N\right) \in \mathbb{Z}^N$ be the steps of this walk and let $\sum_{k=1}^n Y_k=M_n=\left(M_n^1, M_n^2, \ldots, M_n^N\right)$ be the walk itself.

If the random walk is not centered, there exists an i between 1 and N such that $E\left[Y_n^i\right] \neq 0$; then the strong law of large numbers implies that $\lim_{n \rightarrow+\infty}\left|M_n^i\right|=+\infty$ almost surely. Thus the walk is transient.

---

[1] K. L. Chung and D. Ornstein, *On the recurrence of sums of random variables*, Bulletin of the American Mathematical Society, vol. 103, pp. 20–32, 1962.

The question of what centered random walks are recurrent was resolved in the middle of the twentieth century.[2] Whether a centered random walk is recurrent depends on its *dimension*, which is the dimension of the subspace of $\mathbb{R}^N$ generated by the set $\{e_1, e_2, \ldots, e_d\}$ of possible steps. Every centered random walk in dimension at most 2 is recurrent; on the other hand, every random walk in dimension greater than 2 is transient.

We proved the result for the one-dimensional case in the previous proposition. The proofs of the above results in their full generality requires the tools of Fourier analysis, so we will limit our study to nearest neighbor random walks.

In the space $\mathbb{Z}^N$, each point has $2N$ nearest neighbors. Let $(b_1, b_2, \ldots, b_N)$ be the standard basis for $\mathbb{R}^N$. The nearest neighbor random walk on $\mathbb{Z}^N$ is defined by $d = 2N$ and

$$(e_1, e_2, \ldots, e_d) = (b_1, -b_1, b_2, -b_2, \ldots, b_N, -b_N).$$

This walk is centered if and only if $p_{2i-1} = p_{2i}$ for every i between 1 and n. We suppose that p_{2n-1} and p_{2i} are nonzero for every i, which ensures that the dimension of the walk is N. In the special case where each p_i equals $1/2N$, we say that this walk is the simple random walk in dimension N. Simple random walks were studied in the foundational work of George Pólya.[3]

Proposition 13.7. *Suppose $(M_n)_{n\geq 1}$ is a nearest neighbor random walk in the space $\mathbb{Z}^N$. Using the notation introduced above, then*

$$P(M_{2n-1} = 0) = 0$$

and

$$P(M_{2n} = 0) = \sum_{k_1+k_2+\cdots+k_N=n} \frac{(2n)!}{(k_1!k_2!\cdots k_N!)^2} \prod_{i=1}^{N} (p_{2i-1}\, p_{2i})^{k_i},$$

where each k_i is a positive integer.

Proposition 13.8. *Every centered nearest neighbor random walk on $\mathbb{Z}^2$ is recurrent.*

[2] K. L. Chung and W. H. J. Fuchs, *On the distribution of values of sums of random variables*, Memoirs of the American Mathematical Society, No. 6, 1951.

[3] G. Pólya, *Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Strassennetz*, Mathematische Annalen, vol. 84, pp. 149–160, 1921.

Proposition 13.9. *Every nearest neighbor random walk on $\mathbb{Z}^N$ for $N \geq 3$ is transient.*

Proof of Proposition 13.7. If ω is an element of Ω that satisfies $M_m(\omega) = 0$, then there exist integers $k_1, k_2, \ldots, k_N$ such that the finite sequence $(Y_1(\omega), Y_2(\omega), \ldots, Y_m(\omega))$ takes the value b_i exactly k_i times and the value $-b_i$ exactly k_i times for $1 \leq i \leq N$. For such an ω to exist, m must be even; setting $m = 2n$, we would then have $k_1 + k_2 + \cdots + k_N = n$.

Let $(E_1, F_1, E_2, F_2, \ldots, E_N, F_N)$ be a partition of the set of integers between 1 and $2n$ such that $\#E_i = \#F_i = k_i$ for each i between 1 and N. By independence, the event *the indices j in E_i are those satisfying $Y_j = b_j$ and the indices j in F_i are those satisfying $Y_j = -b_j$* has probability $\prod_{i=1}^N (p_{2i-1}\, p_{2i})^{k_i}$.

For fixed $(k_1, k_2, \ldots, k_N)$, there are exactly

$$\binom{2n}{k_1} \cdot \binom{2n-k_1}{k_1} \cdot \binom{2(n-k_1)}{k_2} \cdot \binom{2(n-k_1)-k_2}{k_2} \cdots \binom{2(n-k_1-k_2-\cdots-k_{N-1})}{k_N}$$

such partitions. This is because $\binom{2n}{k_1}$ is the number of ways of choosing E_1, $\binom{2n-k_1}{k_1}$ is the number of ways of choosing F_1 given a choice of E_1, $\binom{2(n-k_1)}{k_2}$ is the number of ways of choosing E_2 given a choice of E_1 and F_1, and so forth. It is easy to check that the product of these binomial coefficients simplifies to

$$\frac{(2n)!}{(k_1!k_2!\cdots k_N!)^2}.$$

Hence the event *the sequence $(Y_1(\omega), Y_2(\omega), \ldots, Y_n(\omega))$ takes the value b_i exactly k_i times and the value $-b_i$ exactly k_i times for* $1 \leq i \leq N$ has probability

$$\frac{(2n)!}{(k_1!k_2!\cdots k_N!)^2} \prod_{i=1}^N (p_{2i-1}\, p_{2i})^{k_i},$$

and the stated result follows immediately. □

For two-dimensional random walks, we can give another explicit formula for the probability of returning to zero based on classical results about Legendre polynomials. Since the properties of these polynomials are discussed in detail in other mathematical texts, we will limit ourselves to briefly presenting the formula we will use. For each $n \in \mathbb{N}$, let

$$g_n := \frac{(2n)!}{(2^n n!)^2}\,.$$

We will use the following definition, known as Rodrigues' formula. For each $n \in \mathbb{N}$, the Legendre polynomial of degree n, written L_n, is the nth derivative of the polynomial $(x^2-1)^n$:

$$L_n(x) = \frac{d^n}{dx^n}\left(x^2-1\right)^n.$$

Proposition 13.10. *Let $n \in \mathbb{N}$. For every nonzero real y,*

$$L_n\left(\frac{1}{2}\left(y+\frac{1}{y}\right)\right) = 2^n n! \sum_{k=0}^{n} g_k g_{n-k} y^{2k-n}.$$

Note that all the coefficients in this expansion are positive.

Proof. We will just outline the proof, leaving the calculations as an exercise for the reader.

The binomial theorem yields

$$\left(x^2-1\right)^n = \sum_{k=0}^{n}(-1)^k \binom{n}{k} x^{2(n-k)},$$

By differentiating this term by term and letting m be the integer part of $n/2$, we see that

$$L_n(x) = n! \sum_{k=0}^{m}(-1)^k \frac{(2(n-k))!}{k!(n-k)!(n-2k)!} x^{n-2k}. \tag{13.2}$$

The Taylor series for $u \mapsto (1-u)^{-1/2}$ about zero is

$$\frac{1}{\sqrt{1-u}} = \sum_{n=0}^{+\infty} g_n u^n.$$

By setting $u = 2tx - t^2$, a simple calculation using (13.2) implies that the Taylor series for $t \mapsto (1-2xt+t^2)^{-1/2}$ about zero is

$$\frac{1}{\sqrt{1-2xt+t^2}} = \sum_{n=0}^{+\infty} \frac{1}{2^n n!} L_n(x)\, t^n. \tag{13.3}$$

Let $y \neq 0$ and $x = \frac{1}{2}(y + \frac{1}{y})$. Then

$$1 - 2xt + t^2 = (1 - yt)\left(1 - \frac{1}{y}t\right),$$

so

$$\frac{1}{\sqrt{1 - 2xt + t^2}} = \left(\sum_{n=0}^{+\infty} g_n y^n t^n\right)\left(\sum_{n=0}^{+\infty} g_n y^{-n} t^n\right).$$

By comparing the expansion of this product to (13.3), we obtain

$$\frac{2^n n!}{L_n(x)} = \sum_{k=0}^{n} g_k g_{n-k} y^{2k-n}.$$

□

Now we will study the centered nearest neighbor random walk in dimension $N = 2$. Its probability distribution is determined by the parameter p_1 because $p_2 = p_1$ and $p_3 = p_4 = \frac{1}{2} - p_1$. We set $p := p_1$ and $q := p_3$.

Proposition 13.11. *The random walk $(M_n)_{n \geq 1}$ satisfies*

$$P(M_{2n} = 0) = g_n \sum_{k=0}^{n} g_k g_{n-k} (4p - 1)^{2k}$$

for every n.

Proof. Proposition 13.7 yields

$$\begin{aligned} P(M_{2n} = 0) &= \sum_{k=0}^{n} \frac{(2n)!}{(k!(n-k)!)^2} p^{2k} q^{2(n-k)} \\ &= 2^{2n} g_n \sum_{k=0}^{n} \binom{n}{k}^2 p^{2k} q^{2(n-k)}, \end{aligned}$$

and the two formulas

$$(p^2 + x)^n = \sum_{k=0}^{n} \binom{n}{k} p^{2k} x^{n-k}$$

and

$$(q^2 + x)^n = \sum_{k=0}^{n} \binom{n}{k} q^{2(n-k)} x^k,$$

show that $P(M_{2n} = 0)$ is the coefficient of x^n in the polynomial

$$Q(x) := 2^{2n} g_n (p^2 + x)^n (q^2 + x)^n.$$

In the case where $p = q(= \frac{1}{4})$, we immediately conclude that

$$P(M_{2n} = 0) = 2^{2n} g_n \binom{2n}{n} 16^{-n} = g_n^2 .$$

Now suppose that $p \neq q$. Then $Q(x) = 2^{2n} g_n (p^2q^2 + (p^2+q^2)x + x^2)^n$, and

$$P(M_{2n} = 0) = \frac{1}{n!} Q^{(n)}(0).$$

By rewriting the trinomial $p^2q^2 + (p^2+q^2)x + x^2$, we can rewrite $Q(x)$ as

$$Q(x) = 2^{2n} g_n \left(\frac{p^2 - q^2}{2}\right)^{2n} \left(\left(\frac{2x + p^2 + q^2}{p^2 - q^2}\right)^2 - 1\right)^n .$$

The change of variables $x' = \frac{2x+p^2+q^2}{p^2-q^2}$ allows us to apply the definition of Legendre polynomials, and we obtain

$$Q^{(n)}(0) = 2^{2n} g_n \left(\frac{p^2 - q^2}{2}\right)^n L_n \left(\frac{p^2 + q^2}{p^2 - q^2}\right).$$

Since $q = \frac{1}{2} - p$, we can also write this as

$$Q^{(n)}(0) = 2^{2n} g_n \left(\frac{4p - 1}{8}\right)^n L_n \left(\frac{1}{2}\left(4p - 1 + \frac{1}{4p - 1}\right)\right).$$

Finally, Proposition 13.10 implies that

$$Q^{(n)}(0) = n! g_n \sum_{k=0}^{n} g_k g_{n-k} (4p - 1)^{2k}.$$

□

Now we can prove that centered random walks in two dimensions are recurrent. We will use the formula for the probability of the simple random walk ($p = 1/4$) returning to the origin to show that the simple random walk is recurrent. Since the probability of returning when $p \neq 1/4$ is higher than when $p = 1/4$, this implies that all centered random walks in two dimensions are recurrent.

Proof of Proposition 13.8. By Proposition 13.11,

$$P(M_{2n}=0) \geq g_n^2 = \frac{((2n)!)^2}{(2^n n!)^4}.$$

Note that g_n is the probability that a simple one-dimensional random walk returns to the origin after $2n$ steps. This quantity is estimated by Stirling's formula: $g_n \sim 1/\sqrt{n\pi}$, so the series with general term g_n^2 diverges. Thus

$$\sum_{n=1}^{+\infty} P(M_{2n}=0) = \infty,$$

and the walk is recurrent by Theorem 13.1. □

Proof of Proposition 13.9. Let (M_n) be a centered nearest neighbor random walk in N dimensions, where $N \geq 3$. We will show that

$$P(M_{2n}=0) = O\left(n^{-N/2}\right).$$

This implies that $\sum_n P(M_n=0)$ converges, and thus that the walk is transient.

We will provide the argument for $N=3$ as it is easy to extend this proof to higher dimensions. We will need the following two estimates:

(13.4) for every $n \in \mathbb{N}$, $\dfrac{1}{(n!)^2} \leq \dfrac{2^{2n}}{(2n)!}$,

and

(13.5)

there is a $c>0$ such that $\dfrac{1}{(n!)^2} \leq c\dfrac{2^{2n}}{\sqrt{n}(2n)!}$ for every $n \in \mathbb{N}^*$.

The inequality (13.4) follows easily from the fact that $(1+1)^{2n} \geq \binom{2n}{n}$. The inequality (13.5) follows from Stirling's formula, which implies that $\frac{(2n)!}{2^{2n}(n!)^2} \sim \frac{1}{\sqrt{n\pi}}$ as n approaches infinity.

Let the parameters of the centered random walk in three dimensions be $p=p_1=p_2$, $q=p_3=p_4$, and $r=p_5=p_6$. These parameters satisfy $p,q,r>0$ and $p+q+r=\frac{1}{2}$ and

$$P(M_{2n}=0) = \sum_{i+j+k=n} \frac{(2n)!}{(i!j!k!)^2} p^{2i} q^{2j} r^{2k}.$$

Let

$$I(p) := \left\{ i \in \mathbb{N} \ : \ 0 \le i \le n \text{ and } \left| p - \frac{i}{2n} \right| < \frac{p}{2} \right\},$$

$$A := \sum_{i+j+k=n, i \notin I(p)} \frac{(2n)!}{(i!j!k!)^2} p^{2i} q^{2j} r^{2k},$$

and

$$B := \sum_{i+j+k=n, i\in I(p), j \in I(q), k \in I(r)} \frac{(2n)!}{(i!j!k!)^2} p^{2i} q^{2j} r^{2k}.$$

By inequality (13.4),

$$A \le \sum_{i+j+k=n, i \notin I(p)} \frac{(2n)! 2^{2n}}{(2i)!(2j)!(2k)!} p^{2i} q^{2j} r^{2k},$$

which can be written as

$$A \le \sum_{0 \le i \le n, i \notin I(p)} \frac{(2n)! 2^{2n}}{(2i)!} p^{2i} \sum_{j=0}^{n-i} \frac{1}{(2j)!(2(n-i-j))!} q^{2j} r^{2(n-i-j)}.$$

By the binomial theorem,

$$\sum_{j=0}^{n-i} \frac{1}{(2j)!(2(n-i-j))!} q^{2j} r^{2(n-i-j)} \le \frac{1}{(2(n-i))!} (q+r)^{2(n-i)},$$

and thus

$$A \le \sum_{0 \le i \le n, i \notin I(p)} \binom{2n}{2i} (2p)^{2i} (2(q+r))^{2(n-i)}. \tag{13.6}$$

When $i \notin I(p)$, we have that $\left|\frac{2i}{2n} - 2p\right| \ge p$. Therefore this upper bound for A is a probability controlled by the large deviations estimate: letting S_n' be a random variable following the binomial distribution with parameters n and $2p$, the upper bound (13.6) implies that

$$A \le P\left(\left| 2p - \frac{S_n'}{2n} \right| \ge p \right).$$

By Theorem 6.1, this quantity approaches zero exponentially as n approaches infinity: there exists a constant $d > 0$ such that $A \le e^{-nd}$. The sums of $\frac{(2n)!}{(i!j!k!)^2} p^{2i} q^{2j} r^{2k}$ for $j \ne I(q)$ or $k \notin I(r)$ can be dealt with in the same way, so we are just left with estimating B.

In the sum defining B, the indices i, j, and k satisfy $i > np$, $j > nq$, and $k > nr$. By (13.5), we have

$$B \leq \sum_{i+j+k=n} (2n)! \frac{c2^{2i}}{\sqrt{i}(2i)!} \frac{c2^{2j}}{\sqrt{j}(2j)!} \frac{c2^{2k}}{\sqrt{k}(2k)!} p^{2i} q^{2j} r^{2k},$$

so

$$B \leq \frac{c^3}{\sqrt{pqr} n^{3/2}} \sum_{i+j+k=n} \frac{(2n)!}{(2i)!(2j)!(2k)!} (2p)^{2i} (2q)^{2j} (2r)^{2k}.$$

By an easy generalization of Newton's binomial theorem,

$$\sum_{i'+j'+k'=2n} \frac{(2n)!}{i'!j'!k'!} (2p)^{i'} (2q)^{j'} (2r)^{k'} = (2p + 2q + 2r)^{2n} = 1;$$

hence,

$$B \leq \frac{c^3}{\sqrt{pqr} n^{3/2}}.$$

□

http://dx.doi.org/10.1090/stml/028/15

Chapter 14

Epilogue

14.1. A Few More General Results

Throughout the majority of this book, we dealt only with sequences of independent random variables taking two values. We extended the range of our study in Chapter 11 by introducing the concept of sequences of finite type random variables and in Chapter 13 by studying sequences of multidimensional random variables.

The limit theorems we stated extend to sums of independent, identically distributed random variables taking values in $\mathbb{R}$. Here are some typical results about these general sums. Proofs of these results can be found in Breiman's book, which is cited in the bibliography.

Let $(X_n)_{n\geq 1}$ be a sequence of independent, identically distributed real random variables (that are finite type random variables and, therefore, bounded). Let m be the expected value of these random variables, and let σ be their standard deviation: $m = E[X_n]$ and $\sigma = E[(X_n - m)^2] = E[X_n^2] - m^2$. Let $S_n = \sum_{k=1}^{n} X_k$.

The Law of Large Numbers. In various forms, the law of large numbers asserts that the sequence $\left(\frac{1}{n}S_n\right)$ converges to m. We have that

$$\text{for every } \epsilon > 0, \quad \lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - m\right| \geq \epsilon\right) \to 0$$

(convergence in probability);

$$\lim_{n\to\infty} E\left[\left(\frac{S_n}{n} - m\right)^2\right] \to 0$$

(convergence in quadratic mean);

$$\lim_{n\to\infty} \frac{S_n}{n} \to m \quad \text{almost surely}$$

(pointwise convergence). We proved these results in Chapter 11.

Large Deviations. The large deviations estimate asserts the existence of a positive function h, depending on the distribution of X_n, such that

$$P\left(\frac{S_n}{n} \geq m + \epsilon\right) \leq e^{-nh(\epsilon)} \quad \text{for every } \epsilon > 0.$$

The proof of this result, which is not hard, is based on the inequality

$$P\left(\frac{S_n}{n} \geq m + \epsilon\right) \leq E\left[\exp\left(t(S_n - n(m+\epsilon))\right)\right] \quad \text{for every } t \geq 0.$$

Central Limit Theorem. From now on we will assume that the standard deviation σ is nonzero.

The central limit theorem states that

$$P\left(a < \frac{S_n - nm}{\sigma\sqrt{n}} < b\right) \xrightarrow{n\to\infty} \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-x^2/2\right)\, \mathrm{d}x$$

uniformly in a and b satisfying $-\infty \leq a < b \leq +\infty$.

Following the work of Paul Lévy and Harald Cramér, the fundamental tool in the proof of this result for the general case is the Fourier transform of probability measures.

Local Limit Theorem, First Form. Suppose that the random variables X_n are centered, which means that $m = 0$. Let G be the additive subgroup of $\mathbb{R}$ generated by the set of values taken by the random variables X_n. The statement of the local limit theorem depends on whether or not G is dense in $\mathbb{R}$. Here we will only consider the first case. Thus, suppose that G is dense in $\mathbb{R}$, which means that the set of values taken by the random variable X_n is not contained in a set of the form $\alpha\mathbb{Z}$ with $\alpha \in \mathbb{R}$. Then, for all real a and b such that $a < b$,

$$P(a < S_n < b) \sim \frac{1}{\sqrt{n}} \frac{1}{\sigma\sqrt{2\pi}} (b-a),$$

as $n \to \infty$.

Local Limit Theorem, Second Form. Now suppose that the random variables X_n are not centered, which means that $m \neq 0$, and that the group G generated by the values taken by the random variables is dense in $\mathbb{R}$. In addition, suppose that the random variables X_n take both positive and negative values (that is, $P(X_n > 0) > 0$ and $P(X_n < 0) > 0$). Consider the *Laplace transform* L defined on $\mathbb{R}$ by

$$L(t) := E[e^{tX_n}].$$

The function L attains its minimum at a unique nonzero point t_0. Then $0 < L(t_0) < 1$, and we set

$$\sigma_0 := \frac{1}{L(t_0)} E\left[X_n^2 e^{t_0 X_n}\right] .$$

A version of the local limit theorem states that, for all real a and b such that $a < b$,

$$P(a < S_n < b) \sim \frac{(L(t_0))^n}{\sqrt{n}} \frac{1}{\sigma_0\sqrt{2\pi}} \int_a^b e^{-t_0 x}\, dx$$

as $n \to \infty$.

Arcsine Law. In the general case, the arcsine law has the same form as in the very special case of the game of Heads or Tails that we described in Chapter 10. However, this similarity is deceptive, because the arguments used to prove the general result are much harder than the combinatorial arguments that we used. (These general arguments use the *Donsker invariance principle*, which is a profound strengthening of the central limit theorem. This principle describes the convergence of the sequence (S_n), suitably normalized, to Brownian motion.)

The arcsine law states that

$$P\left(\#\{k \;:\; 0 \le k \le n \text{ and } S_k > km\} > n\alpha\right) \xrightarrow{n\to\infty} \frac{2}{\pi} \arcsin\sqrt{\alpha}$$

for every α between 0 and 1.

Law of the Iterated Logarithm. The law of the iterated logarithm states that

$$\limsup_{n\to\infty} \frac{S_n - nm}{\sqrt{2n \ln \ln n)}} = \sigma$$

almost surely.

Local Limit Theorem, Multidimensional Case. These results extend to multidimensional random variables, that is, to variables taking values in $\mathbb{R}^N$. The statements are very similar to those in the one-dimensional case, with the most difference appearing in the local limit theorem. Here we will describe the first form of the local limit theorem.

Let (X_n) be a sequence of identical, independently distributed (finite type) random variables taking values in $\mathbb{R}^N$. We write $X_n = (X_{1,n}, X_{2,n}, \ldots, X_{N,n})$. Suppose that these random variables are centered, which means that the expected value of each component $(X_{i,n})$ is zero.

In addition, let $S_n = \sum_{k=1}^n X_k$ and let Γ be the N by N symmetric matrix with coefficients $\gamma_{i,j}$ given by

$$\gamma_{i,j} := \operatorname{cov}(X_{i,n}, X_{j,n}) := E\left[X_{i,n} \cdot X_{j,n}\right].$$

The matrix Γ is positive (as a symmetric matrix). It is invertible if and only if X_n does not take all its values in a proper subspace of $\mathbb{R}^N$. In this nondegenerate case, the local limit theorem states that for all real numbers $a_1, b_1, a_2, b_2, \ldots, a_N, b_n$ such that $a_i < b_i$,

$$P\left(S_n \in \prod_{i=1}^N (a_i, b_i)\right) \sim (2n\pi)^{-N/2} \frac{1}{\det \Gamma} \prod_{i=1}^N (b_i - a_i)$$

as $n \to \infty$.

The most significant part of the above expression is the exponent of n. Note that we get the general term of a convergent series if and only if $N \geq 3$. This yields what we proved in a special case in Chapter 13: centered random walks are recurrent in dimensions one and two and transient in any higher dimension.

14.2. Closing Remarks

Some of the mathematical laws described in this book can be verified experimentally. In particular, this is the case for the weak law of large numbers, the central limit theorem, and the arcsine law. But how can we carry out such experiments? It would be tedious to flip a real coin the large number of times needed to verify these laws. But we can make a computer play Heads or Tails: there are computer programs that simulate choosing random numbers to produce sequences of *pseudorandom variables*. These *pseudorandom number generators* and their uses are presented in the books by Donald Knuth and Nicolas Bouleau that are cited in the bibliography.

We have reached the end of this book. Of course, this book was only one step toward learning what probability theory has to offer. To go further, one must adopt the formalism of measure theory. Since the fundamental work of Andrey Kolmogorov,[1] measure theory has been the basis of probability theory. Once this formalism has been adopted, the concept of a random variable can be extended and the content of this book can be generalized. For example, here are some extensions of the concepts and settings found in this book.

- The limit theorems presented in this book extend to the general case of independent, identically distributed random variables.
- There are many other useful theorems: Kolmogorov's 0-1 law, the rate of convergence of the central limit theorem, the convergence to Brownian motion, renewal theory, the almost sure central limit theorem, etc.
- These theorems extend past the case of sequences of independent, identically distributed random variables: by replacing the hypothesis of independence with a condition of weak dependence or with a martingale condition, we can often get the same law as for the condition of strict independence.

[1]A. Kolmogorov, *Grundbegriffe de Wahrscheinlichkeitsrechnung*, Berlin, 1933.

- The theory of Markov chains, which plays a fundamental role in applications of probability theory, is another setting in which limit theorems are developed and used.
- The study of the game of Heads or Tails and of the random walks associated to it is a first step toward the study of Brownian motion, which is the continuous-time analogue of discrete-time simple random walks.
- Limit theorems are at the heart of the mathematical theory of statistics and, in particular, of the theories of estimation and of statistical tests.

These extensions and applications are described in many books on probability theory.

In conclusion, we note that probability theory (and particularly the theory of convergence theorems) is still an area of productive research today. Because of the work of mathematicians from around the world, this theory is always gaining new discoveries—new concepts, results, and refinements.

Biographies

The biographies are ordered by date of birth; their location in this section is indicated by the number following each name. We present the following authors, who are cited throughout the book: Bernoulli (5), Bernstein (26), Bienaymé (14), Boole (16), Borel (22), Cantelli (23), Cesàro (20), Cramér (31), Fourier (10), Gauss (11), Hardy (25), Hausdorff (21), Huygens (3), Khinchin (32), Kolmogorov (33), Laplace (8), Lebesgue (24), Legendre (9), Lévy (28), Littlewood (27), Markov (19), de Moivre (6), Newton (4), Pascal (2), Poisson (12), Pólya (30), Riemann (18), Rodrigues (13), Steinhaus (29), Stirling (7), Chebyshev (17), Wallis (1), Weierstrass (15).

1. WALLIS, JOHN [Ashford (England), 1616 – Oxford, 1703]: Wallis studied religion and mathematics at Emmanuel College in Cambridge. He was ordained a priest in 1640 and became a professor at Oxford in 1649. His mathematical work set the foundations for Newton's development of infinitesimal calculus. He was the first to correctly define negative and fractional exponents. The integrals that bear his name today appear in his work on the area of the circle. He developed an analytic approach for the study of conic sections. See page 38.

2. PASCAL, BLAISE [Clermont-Ferrand (France), 1623 – Paris, 1662]: Early encounters with the greatest intellectuals of his time helped Pascal develop into a precocious mathematical genius. He wrote his *Essay pour les coniques* in 1640, and in 1642 he developed the first

model for a calculating machine, a precursor to modern calculators. He also created experiments to study the nature of the vacuum, and in 1642 he discovered the triangle that bears his name. He exhibited great literary and philosophical gifts as well: *Les Pensées* and *Les Provinciales* are two masterpieces of French literature. A problem posed by his friend the Chevalier de Méré initiated his study of probability, and he developed the premises of probability theory in his correspondence with Fermat. See page 15.

3. HUYGENS, CHRISTIAAN [The Hague (Netherlands), 1629 – The Hague, 1695]: An astronomer, physician, and engineer, Huygens discovered the rings of Saturn, studied pendulums and falling bodies, and invented the spiral spring used in watches. He was the first to state a theory about the wave properties of light. His mathematical work is also important, especially his work on curves. In this area, he created the theory of envelopes of families of lines and the theory of evolutes and studied the *tractrix* and the *catenary* in detail. His exchanges with Pascal and Fermat contributed to the foundation of probability theory, and he wrote the first treatise on the subject (*De ratiociniis in ludo aleae*). In Paris in 1666, he was one of the founders of the Academy of Sciences. See page 7.

4. NEWTON, ISAAC [Woolsthorpe (England), 1642 – Kensington, 1727]: A precocious scientist, Newton made his main mathematical discovery before his twenty-fifth birthday. In parallel with Leibnitz, he founded differential and integral calculus. His research on functions and curves, and especially on cubics and conic sections, is also very important. His results in these subjects were published years after their discovery: *Philosophiae naturalibus principia mathematica* was published in 1687. Of course, his research in physics is also fundamental: in optics, he described the nature of white light; in mechanics, he stated the laws of universal gravitation. See pages 16, 17, 24.

5. BERNOULLI, JACOB [Basel (Switzerland), 1654 – Basel, 1705]: Along with his younger brother, nephews, and great-nephews, Jacob Bernoulli was a member of a large family of mathematicians. After traveling across Europe to meet the eminent mathematicians of his time, he became a professor at the University of Basel in 1687. He was

particularly interested in infinitesimal calculus and its application to the study of curves. He introduced mathematical rigor in the study of convergence and provided the first proof of the weak law of large numbers for the game of heads or tails. His name is attached to differential equations, a curve (the *lemniscate*), polynomials, and a law of probability. See pages 16, 17, 20.

6. DE MOIVRE, ABRAHAM [Vitry-le-François (France), 1667 – London (England), 1754]: Born in France, de Moivre moved to London at the age of eighteen. There he met Newton and quickly became a famous mathematician. His main works are in probability theory. He characterized the independence of two events by the fact that the probability of both events occurring is the product of the probabilities of each occurring. He discovered Stirling's formula and attained from it the limit form of the binomial law described in this book. He developed the use of polar coordinates to describe complex numbers (and the theorem in this area which bears his name). In addition, he studied applications of mathematics to finance and demography. See pages 29, 32, 36, 39, 41, 46, 121.

7. STIRLING, JAMES [Garden (Scotland), 1692 – Edinburgh, 1770]: After studying at Oxford, Stirling taught in Venice between 1715 and 1725 and in London after 1725. He achieved several results about algebraic curves (for example, that a polynomial of degree n is determined by $n(n+3)/2$ points) and about asymptotic series, in particular the one for $\ln(n!)$. See pages 26, 27, 36, 40, 61, 69, 74, 128.

8. LAPLACE, PIERRE-SIMON [Beaumont-en-Auge (France), 1749 - Paris, 1827]: A mathematician and physicist, Laplace helped found l'École Polytechnique of Paris. He made considerable contributions to astronomy and the theory of gravitation; in particular, he studied the origin and the stability of the solar system. His main mathematical work was in the theory of differential equations and partial derivatives and in probability theory. He developed the use of continuous densities in probability theory and helped create statistical methods. The monumental treatise *Théorie analytique des probabilités*, written in 1812, was an important influence on the mathematics for the remainder of the century. Laplace also played an important public role

during the First Empire and the Restoration. See pages 29, 32, 36, 39, 41, 46, 121, 133.

9. Legendre, Adrien-Marie [Toulouse (France), 1752 – Paris, 1833]: Legendre was the student of the Abbé Marie at the Collège Mazarin in Paris and then a professor at l'École Militaire from 1775 to 1780. In 1782, he won a prize from the Berlin Academy for his calculations on the trajectory of projectiles that took air resistance into account. He started teaching at l'École Normale Supérieure in 1795. After working on mechanics, he made fundamental contributions to number theory (see Hardy and Wright, *An Introduction to the Theory of Numbers*). In addition, he reworked Euclid's *Elements* into a more modern text. The English translation of his work replaced Euclid's text as the geometry textbook used in American schools of the time. Being a man of great honesty, he acknowledged that Abel's and Jacobi's work on elliptic functions was better than his own. His pension was revoked because he opposed political pressure on the scientific world, and he finished his life in misery and loneliness. See pages 124, 125, 127.

10. Fourier, Jean-Baptiste Joseph [Auxerre (France), 1768 - Paris, 1830]: "Fourier is a product of the French Revolution, and his life is a sort of film of French history from 1770 to 1830" (Jean-Pierre Kahane, *Series de Fourier*, Cassini, 1998). From a modest background that prevented him from pursuing the military career that he coveted, Fourier participated in the Revolution and became a student at l'École Normale Supérieure when it opened in 1794. The following year, he taught probability theory at l'École Polytechnique, which had just opened. He participated in Napoleon's Egyptian campaign and held political positions under the Consulate and the Empire. His most important mathematical work deals with solutions to differential equations describing the propagation of heat through a solid. In this work, Fourier introduced the decomposition of a periodic function as the sum of trigonometric functions. Since then, Fourier analysis has become an important branch of mathematics. See pages 36, 56.

11. Gauss, Carl Friedrich [Brunswick (Germany), 1777-Göttingen, 1855]: One of the greatest mathematicians of all time and a scholar whose stature is comparable to Newton's, Gauss was the father of

modern mathematics in Germany. In addition to doing fundamental work on astronomy and physics, he made significant contributions to almost every branch of mathematics. He gave the first complete proof of the *fundamental theorem of algebra.* His principal books, *Disquisitiones arithmeticae* and *Disquisitiones generales circa superficies curvas* are about number theory and differential geometry. See pages 29, 30, 31.

12. POISSON, SIMÉON DENIS [Pithiviers (France), 1781 – Sceaux (France), 1840]: A graduate of l'École Polytechnique, where he was noticed by Lagrange and Laplace, Poisson is considered to be one of the founders of mathematical physics. Having mastered contemporary techniques in mathematical analysis, he applied these to various problems, including fluid mechanics, movement of planets, elasticity, theory of heat, electrostatics, and probability. He showed that, under certain conditions, the binomial distribution can be approximated by the Poisson distribution: $P(S_n = k) = e^{-np}(np)^k/k!$ (see Chapter 4). He received both academic and political honors: a professor at l'École Polytechnique from 1806 and later at the Faculté des Sciences, he was admitted to the Academy of Sciences in 1812, and he was made a baron by Louis XVIII. See page 20.

13. RODRIGUES, BENJAMIN OLINDE [Bordeaux (France), 1794 – Paris, 1851]: A graduate of l'École Normale Supérieure, Rodrigues made important contributions to geometry; in particular, he studied the composition of successive rotations of the plane and the theory of surfaces. His formula describing Legendre polynomials generalizes to other families of orthogonal polynomials. See page 125.

14. BIENAYMÉ, IRÉNÉE-JULES [Paris (France), 1796 – Paris, 1878]: A statistician and an inspector general in the Administration of Finances, Bienaymé applied probability theory to financial calculations and was admitted to the Academy of Sciences in 1852. See pages 9, 20, 23, 94, 102.

15. WEIERSTRASS, KARL THEODOR WILHELM [Ostenfelde (Westphalia, now Germany), 1815 – Berlin (Germany), 1897]: The influence of Weierstrass, a great teacher, was considerable. He introduced modern language and rigor to mathematical analysis. His early works

are on elliptic integrals and abelian functions. He studied the construction of the real numbers, developed the definition of uniform convergence, and proved his famous theorem of approximation by polynomials. He constructed new functions of a real or complex variable as sums of infinite series or products. In linear algebra, he founded the modern theory of determinants. See pages 20, 21.

16. BOOLE, GEORGE [Lincoln (England), 1815 – Cork (Ireland), 1864]: The son of a British shopkeeper, Boole was largely self-taught. At twenty, he read works by Laplace and Lagrange and studied differential equations. In 1849, he became a professor at Queens College, Cork. He is the founder of logic as a branch of mathematics independent of philosophy, and he discovered a connection between his formal theory of logic and probability theory. See page 79.

17. CHEBYSHEV, PAFNUTY LVOVICH [Okatovo (Russia), 1821 – Saint Petersburg, 1894]: Chebyshev's thesis, which he defended in 1846, was on probability theory. He became a professor at the University of Saint Petersburg in 1847, where he principally studied number theory. He made important contributions to the study of the distribution of prime numbers in the integers; in particular, he proved Bertrand's Postulate: for every integer $n \geq 2$, there exists a prime number between n and $2n$. He also contributed to approximation theory by studying the polynomials that bear his name today. See pages 9, 20, 23, 94, 102.

18. RIEMANN, GEORG FRIEDRICH BERNHARD [Breselenz (Hanover, now Germany), 1826 – Selasca (Italy), 1866]: A mathematical genius who made lasting contributions, Riemann studied in Berlin and Göttingen, where he defended his thesis on the foundations of complex analysis in 1851. His work in analysis and number theory is fundamental, and his conjecture, called the *Riemann hypothesis*, is the most famous problem of twenty-first century mathematics. He was the first to define integrals for noncontinuous functions. His work in geometry founded an important part of modern mathematics and theoretical physics, particularly topology and relativity. See pages 1, 36, 42, 49, 61, 75.

19. MARKOV, ANDREI ANDREYEVICH [Riazin (Russia), 1856 – Petrograd (now Saint Petersburg), 1922]: The student of Chebyshev,

Markov worked on number theory and analysis before tackling probability theory. He invented the concept of *Markov chains*, which are sequences of random variables subject to a condition that generalizes independence. This concept gained fundamental importance in the applications of probability. He is considered to have founded the theory of stochastic processes. See pages 8, 24, 72, 136.

20. CESÀRO, ERNESTO [Naples (Italy), 1859 – Torre Annunziata, 1906]: An Italian mathematician with diverse interests, Cesàro was a professor at the University of Naples. He studied the connection between arithmetic and probability: he proved that the probability that two random integers are relatively prime is $6/\pi^2$ (try to prove it yourself!). In the field of convergence of entire series, he introduced the concept of convergence that bears his name today. See page 72.

21. HAUSDORFF, FELIX [Breslau (Germany; now Wroclaw, Poland), 1868 – Bonn (Germany), 1942]: A German mathematician who taught at the Universities of Leipzig, Greisswald, and Bonn, Hausdorff made significant contributions to set theory, functional analysis, topology, and measure theory. His name is attached to the axioms of a topological space, to a concept of the dimension of a subset of a metric space, and to a concept of measure. He met a tragic end: threatened with being moved to a Nazi internment camp for Jews, he committed suicide with his wife and sister-in-law. See pages 98, 99.

22. BOREL, ÉMILE [Saint Affrique (France), 1871 – Paris, 1956]: A graduate of l'École Normale Supérieure, Borel was one of the founders of measure theory, which allowed Lebesgue to found the modern theory of integration. He initiated the study of probability based on measure theory that is used universally today. With the help of many collaborators, he was the editor of *Traité du Calcul des Probabilités et ses applications* (Gauthier-Villars, Paris, 1920s), and important work on probability theory and its applications. He made significant contributions to the theory of real functions and the summation of numerical series, and he is one of the founders of game theory. He also wrote on philosophy, pedagogy, political economics, and the history of science. As a deputy, a minister during the Third Republic, and an opposer of the Vichy regime, Borel also participated in important

political activities. See pages viii, 77, 78, 82, 83, 88, 89, 90, 93, 94, 97, 100, 104, 105, 106, 112, 115, 118.

23. CANTELLI, FRANCESCO [Palermo (Italy), 1875 – Rome, 1966]: An Italian mathematician, Cantelli taught at the School of Economics and Commerce in Rome. He worked on probability theory and statistics. He participated actively in the debate about the mathematical foundations of probability between the Bayesian and frequentist points of view. See pages 78, 89, 90, 92, 94, 96, 99, 100, 104, 105, 106, 112, 115, 118.

24. LEBESGUE, HENRI [Beauvais (France), 1875 – Paris, 1941]: At l'École Normale Supérieure, Lebesgue was taught by Borel. In his thesis, (*Intégrale, longueur, aire*, defended in 1901), he founded the modern theory of integration, which allowed Kolmogorov to create rigorous axioms of probability. Lebesgue's contributions to the theory of trigonometric series (for example, under what conditions is a periodic function a sum of Fourier series?) are also important. Teaching at Nancy, Rennes, Poitiers, and at the Collège de France starting in 1912, Lebesgue is recognized for his pedagogical influence. See pages 86, 87, 88, 90.

25. HARDY, GODFREY HAROLD [Cranleigh (England), 1877 – Cambridge, 1947]: A professor at Cambridge, Hardy favored analytic number theory, that is, the theory of applying real and complex analysis to number theory. His collaboration with Littlewood produced many significant contributions to analytic number theory. An advocate of pure mathematics, he wrote several mathematical texts as well as an interesting autobiography, *A Mathematician's Apology* (Cambridge, 1940). See pages 98, 100.

26. BERNSTEIN, SERGEI NATANOVICH [Odessa (Ukraine), 1880 – Moscow (Russia), 1968]: Bernstein first taught at the Universities of Kharkov and of Paris, and then at the University of Leningrad (Saint Petersburg) until 1941. He worked on probability theory as well as on differential equations and functional analysis. See pages 12, 21, 23.

27. LITTLEWOOD, JOHN EDENSOR [Rochester (England), 1885 – Cambridge, 1977]: A professor at Cambridge, Littlewood mainly

worked on analytic number theory and trigonometric series. His collaboration with Hardy was particularly fruitful. See pages 98, 100.

28. LÉVY, PAUL [Paris (France), 1886 - Paris, 1971]: From a family of mathematicians, Lévy was a student and subsequently a professor at l'École Polytechnique. He was a major player in French mathematics of the twentieth century. He work on functional analysis and the theory of differential equations, but he is best known for his formidable contributions to probability theory. He systematized the use of the characteristic function (that is, the Fourier transform of a probability measure) in the study of convergences like the central limit theorem or the arcsine law. He studied stochastic processes and introduced the concept of *martingales.* With Wiener, Lévy is the father of the modern theory of Brownian motion. His books include *Leçons d'analyse fonctionnelle* (1922), *Calcul des probabilités* (1925–1951), *Théorie de l'addition des variables aléatoires* (1937–1954), and *Processus stochastiques et mouvement brownien* (1948). His autobiographical reflections, *Quelques aspects de la pensées d'un mathématicien*, are also interesting. See pages 59, 132.

29. STEINHAUS, HUGO DYONIZY [Jaslo (Austria, now Poland), 1887 – Wroclaw (Poland), 1972]: Steinhaus was taught by Hilbert and was very influenced by Lebesgue's ideas. With Banach at the end of World War I, he helped create the fruitful school of Polish mathematics, which founded the theory of functional analysis. He worked on trigonometric series, measure theory, topology, and game theory. In 1923, he used measure theory to provide the first rigorous model of the infinite game of heads or tails. See page 96.

30. PÓLYA, GEORGE [Budapest (Hungary), 1887 – Palo Alto (California), 1985]: Of Hungarian origin, Pólya studied in Budapest and taught in Zurich from 1914 to 1940 before emigrating to the United States. He worked mainly on number theory and combinatorics, which are specialties of Hungarian mathematics, functional analysis, and probability theory, including random walks and characteristic functions in particular. Many of his books are very successful, including *Inequalities* (written with Hardy and Littlewood), exercise books, and *How to Solve It*, a study of the problem solving process. See page 123.

31. CRAMÉR, CARL HARALD [Stockholm (Sweden), 1893 – Stockholm, 1985]: Cramér was a mathematician and an actuary for a life insurance company. He worked on probability, statistics, and game theory. He is considered to be the founder of the mathematical theory of statistics, which is built rigorously from probability theory. See pages 46, 56, 132.

32. KHINCHIN, ALEXANDR YAKOVLEVICH [Kondrovo (Russia), 1894 – Moscow, 1959]: One of the masters of the Soviet school of probability, Khinchin stated the law of the iterated logarithm and developed the theory of stochastic processes, often in competition with Lévy. He also worked on measure theory and number theory, in particular on Diophantine approximation and continued fractions. See pages 97, 98, 100.

33. KOLMOGOROV, ANDREI NIKOLAEVICH [Tambov (Russia), 1903 – Moscow, 1987]: One of the great Soviet mathematicians, Kolmogorov became famous in 1922 for his results on trigonometric series. He was interested in a vast range of mathematics. His contributions are fundamental, particular in the theory of dynamical systems (the stability of the solar system) and in complexity theory (the theory of computation). He is doubtless the main player in twentieth century probability theory: he gave the theory a solid mathematical foundation and developed the theory of stochastic processes, particularly Markov processes. He was also very interested in the teaching of mathematics and played an important role in educational and scientific politics in the USSR. He received high honors from his country and was admitted to the most important international scientific academies. See pages 95, 101, 135.

Bibliography

This book's rigorous but elementary approach to probability theory is also found in the first volume of William Feller's book and throughout Yakov Sinai's monograph. In addition, Feller's book contains an extensive bibliography.

There are also many modern works at a more advanced level, including the books by John Lamberti, William Feller, Leo Breiman, Patrick Billingsley, and Richard Durrett. Those who read French can add the books by Rényi and by Dacunha-Castelle and Duflo to this list. All these books provide an introduction to probability theory, a detailed discussion of an extensive portion of the theory, and bibliographical references for further reading.

A short and pleasant survey of probability theory is offered by Deheuvels' little book.

Billingsley, P., *Probability and Measure.* Wiley, New York, 1979.

Bouleau, N., *Probabilités de l'ingénieur, Variables aléatoires et simulation.* Hermann, Paris, 1986.

Breiman, L., *Probability.* Addison-Wesley, Reading, Mass., 1968.

Dacunha-Castelle, D. and M. Duflo, *Probabilités et Statistiques.* Masson, Paris, 1982.

Deheuvels, P., *La probabilité, le hasard et la certitude. Que sais-je?,*

Presses Universitaires de France, Paris, 1982.
Durrett, R., *Probability: Theory and Examples.* Wadsworth&Brooks, Pacific Grove, California, 1991.
Feller, W., *An Introduction to Probability Theory and Its Applications.* Wiley, New York, 1957.
Knuth, D., *The Art of Computer Programming*, Vol. 2, *Seminumerical algorithms.* Addison-Wesley, Reading, Mass., 1969.
Lamperti, J., *Probability, A Survey of the Mathematical Theory.* Benjamin, New York, 1966.
Rényi, A., *Calcul des Probabilités.* Dunod, Paris, 1966.
Sinai, Y., *Probability Theory: An Introduction Course.* Springer Textbook, Springer-Verlag, Berlin, 1992.

Index

Titles in This Series

28 **Emmanuel Lesigne,** Heads or tails: An introduction to limit theorems in probability, 2005

27 **Reinhard Illner, C. Sean Bohun, Samantha McCollum, and Thea van Roode,** Mathematical modelling: A case studies approach, 2005

26 **Robert Hardt, Editor,** Six themes on variation, 2004

25 **S. V. Duzhin and B. D. Chebotarevsky,** Transformation groups for beginners, 2004

24 **Bruce M. Landman and Aaron Robertson,** Ramsey theory on the integers, 2004

23 **S. K. Lando,** Lectures on generating functions, 2003

22 **Andreas Arvanitoyeorgos,** An introduction to Lie groups and the geometry of homogeneous spaces, 2003

21 **W. J. Kaczor and M. T. Nowak,** Problems in mathematical analysis III: Integration, 2003

20 **Klaus Hulek,** Elementary algebraic geometry, 2003

19 **A. Shen and N. K. Vereshchagin,** Computable functions, 2003

18 **V. V. Yaschenko, Editor,** Cryptography: An introduction, 2002

17 **A. Shen and N. K. Vereshchagin,** Basic set theory, 2002

16 **Wolfgang Kühnel,** Differential geometry: curves - surfaces - manifolds, 2002

15 **Gerd Fischer,** Plane algebraic curves, 2001

14 **V. A. Vassiliev,** Introduction to topology, 2001

13 **Frederick J. Almgren, Jr.,** Plateau's problem: An invitation to varifold geometry, 2001

12 **W. J. Kaczor and M. T. Nowak,** Problems in mathematical analysis II: Continuity and differentiation, 2001

11 **Michael Mesterton-Gibbons,** An introduction to game-theoretic modelling, 2000

10 **John Oprea,** The mathematics of soap films: Explorations with Maple®, 2000

9 **David E. Blair,** Inversion theory and conformal mapping, 2000

8 **Edward B. Burger,** Exploring the number jungle: A journey into diophantine analysis, 2000

7 **Judy L. Walker,** Codes and curves, 2000

6 **Gérald Tenenbaum and Michel Mendès France,** The prime numbers and their distribution, 2000

5 **Alexander Mehlmann,** The game's afoot! Game theory in myth and paradox, 2000

4 **W. J. Kaczor and M. T. Nowak,** Problems in mathematical analysis I: Real numbers, sequences and series, 2000

3 **Roger Knobel,** An introduction to the mathematical theory of waves, 2000

TITLES IN THIS SERIES

2 **Gregory F. Lawler and Lester N. Coyle,** Lectures on contemporary probability, 1999

1 **Charles Radin,** Miles of tiles, 1999

Everyone knows some of the basics of probability, perhaps enough to play cards. Beyond the introductory ideas, there are many wonderful results that are unfamiliar to the layman, but which are well within our grasp to understand and appreciate. Some of the most remarkable results in probability are those that are related to limit theorems—statements about what happens when the trial is repeated many times. The most famous of these is the Law of Large Numbers, which mathematicians, engineers, economists, and many others use every day.

In this book, Lesigne has made these limit theorems accessible by stating everything in terms of a game of tossing of a coin: heads or tails. In this way, the analysis becomes much clearer, helping establish the reader's intuition about probability. Moreover, very little generality is lost, as many situations can be modelled from combinations of coin tosses.

This book is suitable for anyone who would like to learn more about mathematical probability and has had a one-year undergraduate course in analysis.

For additional information and updates on this book, visit

www.ams.org/bookpages/stml-28

ISBN 0-8218-3714-1

9 780821 837146

STML/28

AMS on the Web
www.ams.org