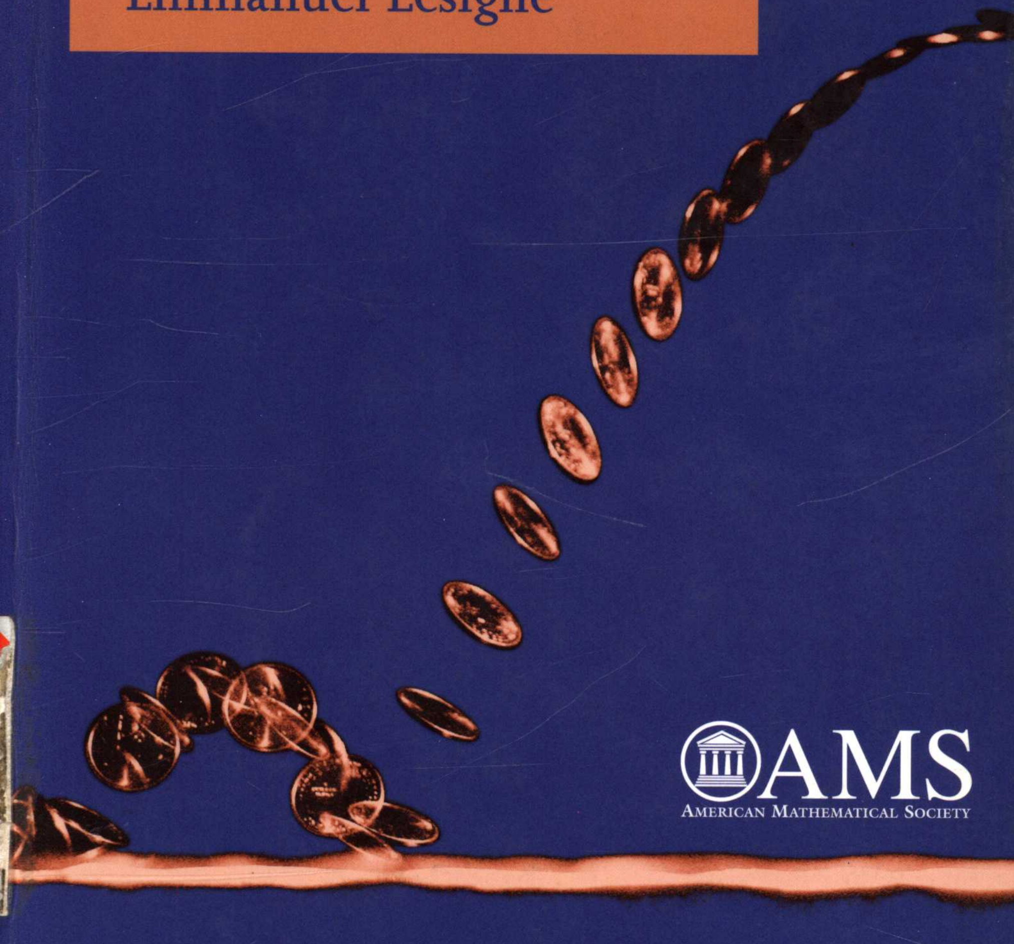
STUDENT MATHEMATICAL LIBRARY
Volume 28

Heads or Tails

An Introduction to Limit Theorems in Probability

Emmanuel Lesigne



AMS
American Mathematical Society

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

Heads or Tails

An Introduction to Limit Theorems in Probability

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.

Contents

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.

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 $\Omega = \{1, 2, 3, \dots, 49\}$ 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

for each $(e_1, e_2, \ldots, e_i) \in \{0, 1\}^i$,

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

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.

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 non-negative 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.