PROBABILITY

The Classical Limit Theorems

Henry McKean

Probability: The Classical Limit Theorems

HENRY MCKEAN
New York University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107053212

First published 2014

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
McKean, Henry P. (Henry Pratt), 1930- author.
Probability : the classical limit theorems / Henry Mckean, New York University.
pages cm
1. Limit theorems (Probability theory) I. Title.
QA273.67.M35 2014
519.2–dc23
2014022165

ISBN 978-1-107-05321-2 Hardback
ISBN 978-1-107-62827-4 Paperback

Probability: The Classical Limit Theorems

The theory of probability has been extraordinarily successful at describing a variety of natural phenomena, from the behavior of gases to the transmission of information, and is a powerful tool with applications throughout mathematics. At its heart are a number of concepts familiar in one guise or another to many: Gauss' bell-shaped curve, the law of averages, and so on, concepts that crop up in so many settings that they are, in some sense, universal. This universality is predicted by probability theory to a remarkable degree. It is the aim of the book to explain the theory, prove classical limit theorems, and investigate their ramifications.

The author assumes a good working knowledge of basic analysis, real and complex. From this, he maps out a route from basic probability, via random walks, Brownian motion, the law of large numbers and the central limit theorem, to aspects of ergodic theorems, equilibrium and nonequilibrium statistical mechanics, communication over a noisy channel, and random matrices. Numerous examples and exercises enrich the text.

HENRY MCKEAN is a professor in the Courant Institute of Mathematical Sciences at New York University. He is a fellow of the American Mathematical Society and in 2007 he received the Leroy P. Steele Prize for his life's work.

Il y a des faussetés déguisées qui représentent si bien la verité, que ce serait mal juger de ne s'y laisser pas tromper.

La Rochefoucauld, Maximes no. 282

DEDICATION

To the memory of M. Kac, W. Feller, K. Itô, N. Levinson, and Gretchen Warren, who taught me so much about everything, and to my dear wife Rasa.

Preface

The goal of this book is to present the elementary facts of classical probability, namely the law of large numbers (LLN) and the central limit theorem (CLT), first in the simplest setting (Bernoulli trials), then in more generality, and finally in some of their ramifications in, e.g. arithmetic, geometry, information and coding, and classical mechanics.

Let's talk about coin-tossing to illustrate the principal themes in the simplest way.

Let the coin be honest so that the probability of heads or tails is the same ($= 1/2$). After a large number (n) of independent trials, you have $\# = \mathbf{e}_1 + \cdots + \mathbf{e}_n$ successes, meaning heads, let's say $\mathbf{e} = 1$ for heads, $\mathbf{e} = 0$ for tails. The law of large numbers states that $\#/n$ tends to $1/2$ as $n \uparrow \infty$ with probability $1 - \mathbb{P}\big[\lim_{n\uparrow\infty} \frac{\#}{n} = \frac{1}{2}\big] = 1$ as it is written. That's only common sense if you like. The central limit theorem is deeper. It says that if you center and scale $\#$ as in $(\# - n/2)$ over $\sqrt{n}/2$, then for large n you will see the celebrated bell-shaped curve of Gauss:

$$\lim_{n\uparrow\infty} \mathbb{P}\left[a \leq \frac{\# - n/2}{\sqrt{n}/2} < b\right] = \int_a^b \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}\, \mathrm{d}x \quad \text{for any } a < b.$$

I say it lies deeper, but it is only the proof that it is so. The phenomenon itself is easily illustrated in Nature. To do this, it is best to make a little change, making new $\mathbf{e}$s from the old by the rule $\mathbf{e} \to 2(\mathbf{e} - \frac{1}{2}) = \pm 1$. Then $\mathbf{x}(n) = \mathbf{e}_1 + \cdots + \mathbf{e}_n$ is the standard random walk, so-called, taking independent steps ± 1 with probabilities $\mathbb{P}(\mathbf{e} = \pm 1) = \frac{1}{2}$, and you have the more symmetrical law:

$$\lim_{n\uparrow\infty} \mathbb{P}\left[a \leq \frac{\mathbf{x}(n)}{\sqrt{n}} < b\right] = \int_a^b \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}\, \mathrm{d}x.$$

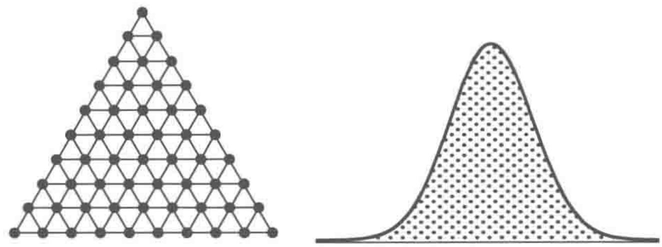


Figure 1

The walk is easily simulated. Take a board studded with nails as in Figure 1 (left) and incline it not too steeply (at 20° say), pour in bird shot from a little funnel at the top, and look to see what piles up at the bottom. You *should* see a scaled approximation to Gauss's celebrated bell-shaped curve $(2\pi)^{-1/2}e^{-x^2/2}$ as in Figure 1 (right) and that is just what happens: the individual shot, running down, hits a nail and is deflected, roughly half the time to the right and half the time to the left, imitating the random walk to produce a bell-shaped heap at the bottom in vindication of CLT.

Of course, that is not *exactly* what happens: the shot is not perfectly round, the nails are not perfectly placed, perfect statistical independence does not exist in Nature, and so on. But it is *approximately* so, which is why I have placed on the title page the maxim of La Rochefoucauld (who surely never thought about the standard random walk but has it just right): That there are certain deceptions, wrong in fact, but which come so close to the real truth that it would be a mistake of judgment not to let oneself be fooled by them. Or, to vary the *mot*: Probability is only a manner of speaking – not the real thing – but is wonderful how well it works.

Take for example, Gibbs's statistical mechanics of which you will get a glimpse in Chapter 10. It is based, of course, on ideas from Nature, but the language is probability, and it is successful beyond all dreams. Or again, take Shannon's ideas about the quantity of information and the means (coding) for its faithful communication over a noisy channel – an equally successful statistical picture of the thing, explained, in part, in Chapter 9.

Well, you get the idea of what I want to do and will judge at the end if I succeeded. Einstein said: "Everything should be made as simple as possible, but not more simple". I have tried to follow that.

Acknowledgement

My debt to Anne Boutet de Monvel is very great. She has typed the whole book from my poor writing with such care, both for how it looks and what it says. I do not know how I could have finished it without her friendly help.

Henry McKean
NYC and Essex, MA, December 2013

Guide

Prerequisities

Not much. I need a good *working* knowledge of the vector spaces $\mathbb{R}^d$ and $\mathbb{C}^d$, and of calculus in several variables; also the rudiments of probability and some knowledge of Lebesgue's measure and integral on the unit interval $[0, 1]$. As to the last, I will sketch what I need and ask you to read up on it if you don't know it already, or just to believe what I tell you and use it (with care). Here are some books at about the right level: Munroe [1953], Breiman [1968], and/or the more advanced Billingsley [1979]. Any of these will tell what is wanted, through the little sketch in §1.1 is plenty if you fill it in. Besides, some complex function theory would be nice; it is used sparingly. Ahlfors [1979] is best for this. As to basic probability, Breiman [1968] is excellent and above all Feller [1966] which is full of verve, much information, and a variety of practical examples. In short, only a modest technical machinery is required so as to keep the probability to the fore.

Exercises

Please do these faithfully. It is the only known way to learn the tricks of the trade. Some are marked with a star ($\star$) as being unnecessary to the sequel and/or more difficult. Certain articles and sections are also starred for similar reasons.

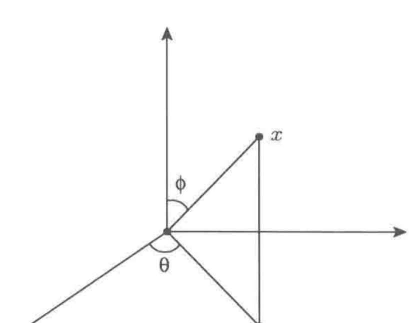
References

References are indicated by a name followed by the year of publication in *square* brackets, as in Feller [1950: 33–37]: 1950 indicates the date of publication, 33–37 gives the paging. These are listed at the end. Things like Gauss (1800) or Jacobi (1820), with the year in *parentheses*, are not precise references, only historical indications: *who* and very roughly *when*.

Notations/Usage

Positive means > 0, non-negative ≥ 0, and similarly for negative and non-positive; x^+/x^- is the positive/negative part of the number x; $x \wedge y$ is the smaller of x and y, $x \vee y$ the larger. The symbol $\simeq$ means approximately equal, up to something really small or with a small percentage error, as the context will indicate; $\lesssim$ is used similarly. $\mathbb{Z}$ is the integers, $\mathbb{Z}^+$ the non-negative integers, $\mathbb{N}$ the whole numbers $(1, 2, 3, \ldots)$, $\mathbb{R}$ the line, $\mathbb{C}$ the complex plane. $\mathbf{X}$ is a (sample) space; A, B, C and the like are (mostly) events, i.e. subsets of $\mathbf{X}$. The symbol $\mathfrak{F}$ denotes a field (of events) meaning that it is closed under complements indicated by a prime ($'$), countable unions ($\cup$), and countable intersections ($\cap$) – not the common usage but more brief. German (fraktur) is reserved for these. Boldface $\mathbf{x}$ and the like is used, not wholly consistently, for random quantities. Italic x and the like is (mostly) for non-random things. $\mathbb{P}$ is probability. $\mathbb{E}$ is expectation as in $\mathbb{E}(\mathbf{x}) = \int \mathbf{x} \, d\mathbb{P}$; the shorthand $\mathbb{E}(\mathbf{x}, A)$ is for $\int_A \mathbf{x} \, d\mathbb{P}$. By $C[0,1]$, $C^2[0,\infty)$, $C^\infty(\mathbb{R})$, are designated various classes of continuous functions. $\mathcal{C}^\infty_\downarrow(\mathbb{R})$ is the class of smooth, rapidly vanishing functions. $L^1(\mathbb{R})$, $L^2(\mathbb{R})$, and so on are the usual Lebesgue spaces. # is for counting as in $\#(p \leq n : p$ a prime number$)$. 2+ means a number a little bit more than 2, say; 2− means a number a little bit less. The natural logarithm to the base $e = 2.718+$ is written log; loglog means log(log). The symbol $\mathbf{1}_K$ is an indicator function: 1 on K, 0 elsewhere. Traces are denoted by tr.

For those not so familiar with spherical polar coordinates in three dimensions, I remind you that $x \in \mathbb{R}^3$ may be written $x = |x|e$ in which $|x|$ is length and e is the (unit) direction $(\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi)$. Here $x_3 = |x|\cos\phi$, $\frac{-\pi}{2} \leq \phi \leq \frac{\pi}{2}$ being co-latitude, and $0 \leq \theta \leq 2\pi$ is the longitude, measured counter-clockwise in the plane $x_3 = 0$ as in the picture.



Contents