Introduction

There is an order to chaos. Unpredictability is predictable. In fact, randomness itself is so regular that we can assign a number to a random occurrence which tells us in a precise way how likely it is. The number is called its probability.

That is not to say that we can predict the result of a single toss of a fair coin. We cannot. But we can predict that between forty and sixty out of a hundred tosses will be heads. We might—rarely—be wrong about that, but only once or twice in a hundred tries, and if we continue to toss: a thousand times, a million times, and so on, we can be sure that the proportion of heads will approach 1/2.

So randomness has its own patterns. Our aim is to understand them.

Probability is a rather unusual part of mathematics. While its full birth as a mathematical subject can be traced to the correspondence between Fermat and Pascal[1] in the summer of 1654, the subject wasn't put on a rigorous footing until 1934, 270 years later, when A. N. Kolmogorov showed it was properly a part of measure theory[2]. But probability had been around for several centuries before measure theory existed, and it is quite possible to study the subject without it. In fact, probability is taught at many different

[1]Pascal and Fermat were by no means the first to study probabiity, but their work on the "problem of points" was so much deeper than what had gone before that it is properly considered the true beginning of the subject. See Keith Devlin's "The Unfinished Game" [**13**] for an account.

[2]See [**22**] for an English translation of Kolmogorov's landmark paper. It showed that all of probability theory could be regarded as a part measure theory, giving a general existence theorem for stochastic processes (not present, alas, in this book, but see [**12**] or [**9**]) and a rigorous definition of conditional expectations (see Chapter 8), which had previously been confined to special cases. This was quite a change from the more intuitive approach, and it took some time to replace "could be taken" by "is." That was completed by Doob, culminating in his seminal book Stochastic Processes [**12**].

levels, according to the mathematics the students know: in elementary and high school, first year college, third or fourth year college, as well as in graduate school. Certain things are common to all of these courses, but the the more mathematics the student knows, the deeper he or she can go. This particular text is drawn from a two-semester course taught over the years at the University of British Columbia, mainly to fourth-year mathematics honors students. It assumes the student is familiar with calculus and knows some analysis, but not measure theory. Many of the students, but by no means all, take a concurrent course in Lebesgue measure. It is not necessary, but it adds depth, and gives the student some "Aha!" moments, such as the sudden realization: "Aha! The expectation is nothing but a Lebesgue integral[3]!"

We begin with the basic axioms of probability, and the all-important ideas of conditional probability and independence. Then we quickly develop enough machinery to allow the students to solve some interesting problems and to analyze card games and lotteries. Just to show how quickly one can get into non-trivial questions, we work out the problem of the gambler's ruin.

The systematic study of classical probability begins in Chapter Two. Its aim is to prove two of the basic classical theorems of the subject: the law of large numbers and the central limit theorem. Far from being recondite, these theorems are practically part of Western folklore. Who has not heard of the law of averages? That is another name for the law of large numbers. What student has not been subject to "grading on a curve", a direct (and often mistaken) application of the central limit theorem? It is surprising how much of the curriculum is determined by the modest aim of understanding those two results: random variables, their expectations and variances, their distributions, the idea of independence, and the ideas of convergence are needed merely to state the theorems. A number of inequalities, the theory of convergence in distribution, and the machinery of characteristic functions, are necessary to prove them. This, along with enough examples to supply the intuition necessary to understanding, determines the first six chapters.

The second part of the book introduces stochastic processes, and changes the viewpoint. Stochastic processes evolve randomly in time. Instead of limit theorems at infinity, the emphasis is on what the processes actually do; we look at their sample paths, study their dynamics, and see that many interesting things happen between zero and infinity. There is a large selection of stochastic processes to study, and too little time to study them.

[3]On the other hand, students who take probability before measure theory have their "Aha!" moment later, when they realize that the Lebesgue integral is nothing but an expectation.

We want to introduce processes which are major building blocks of the theory, and we aim the course towards Brownian motion and some of its weird and wonderful sample path properties. Once more, this determines much of the curriculum. We introduce the Markov property and stopping times with a study of discrete-parameter Markov chains and random walks, including special cases such as branching processes. Poisson and birth and death processes introduce continuous parameter processes, which prepares for Brownian motion and several related processes.

The one non-obvious choice is martingales. This deserves some explanation. The subject was once considered esoteric, but has since shown itself to be so useful[4] that it deserves inclusion early in the curriculum. There are two obstructions. The first is that its whole setting appears abstract, since it uses sigma-fields to describe information. Experience has shown that it is a mistake to try to work around this; it is better to spend the necessary time to make the abstract concrete by showing how sigma-fields encode information, and, hopefully, make them intuitive. The second obstruction is the lack of a general existence theorem for conditional expectations: that requires mathematics the students will not have seen, so that the only case in which we can actually construct conditional expectations is for discrete sigma-fields, where we can do it by hand. It would be a pity to restrict ourselves to this case, so we do some unashamed bootstrapping. Once we show that our hand-constructed version satisfies the defining properties of the general conditional expectation, we use only these properties to develop the theory. When we have proved the necessary martingale theorems, we can construct the conditional expectation with respect to a general sigma field as the limit of conditional expectations on discrete sigma fields. This gives us the desired existence theorem ... and shows that what we did was valid for general sigma-fields all along. We make free use of martingales in the sequel. In particular, we show how martingale theory connects with a certain part of mathematical finance, the option pricing, or Black-Scholes theory.

The final chapter on Brownian motion uses most of what we have learned to date, and could pull everything together, both mathematically and artistically. It would have done so, had we been able to resist the temptation to spoil any possible finality by showing—or at least hinting at—some of

[4]The tipping point was when engineers started using martingales to solve applied problems, and, in so doing, beat the mathematicians to some very nice theorems. The coup de grâce was struck by the surprising realization that the celebrated Black-Scholes theory of finance, used by all serious option-traders in financial markets was, deeply, martingale theory in disguise. See sections 9.6 and 10.10

the large mathematical territory it opens up: white noise, stochastic integrals, diffusions, financial mathematics, and probabilistic potential theory, for example.

A last word. To teach a course with pleasure, one should learn at the same time. Fair is fair: the students should not be the only learners. This is automatic the first time one teaches a course, less so the third or fourth time. So we tried to include enough sidelights and interesting byways to allow the instructor some choice, a few topics which might be substituted at each repetition. Most of these are starred: *. In fact, we indulged ourselves somewhat, and included personal favorites that we seldom have time to cover in the course, such as the Wiener stochastic integral, the Langevin equation, and the physical model of Brownian motion.