Probability, Random Variables, and Stochastic Processes

McGraw-Hill Series In Systems Science

ATHANASIOS PAPOULIS

Professor of Electrical Engineering
Polytechnic Institute of Brooklyn

Probability, Random Variables, and Stochastic Processes

McGRAW-HILL BOOK COMPANY
new york st. louis san francisco toronto london sydney

Probability, Random Variables, and Stochastic Processes

Library of Congress Catalog Card Number 64-22956

48448

Probability, Random Variables, and Stochastic Processes

McGraw-Hill Series In Systems Science

EDITORIAL CONSULTANTS

George Dantzig
William Linvill
Lotfi Zadeh

Papoulis Probability, Random Variables, and Stochastic Processes
Zadeh and Desoer Linear System Theory: The State Space Approach

Preface

Several years ago I reached the conclusion that the theory of probability should no longer be treated as adjunct to statistics or noise or any other terminal topic, but should be included in the basic training of all engineers and physicists as a separate course. I made then a number of observations concerning the teaching of such a course, and it occurs to me that the following excerpts from my early notes might give you some insight into the factors that guided me in the planning of this book:

"Most students, brought up with a deterministic outlook of physics, find the subject unreliable, vague, difficult. The difficulties persist because of inadequate definition of the first principles, resulting in a constant confusion between assumptions and logical conclusions. Conceptual ambiguities can be removed only if the theory is developed axiomatically. They say that this approach would require measure theory, would reduce the subject to a branch of mathematics, would force the student to doubt his intuition leaving him without convincing alternatives, but I don't think so. I believe that most concepts needed in the applications can be explained with simple mathematics, that probability, like any other theory, should be viewed as a conceptual structure and its conclusions should rely not on intuition but on logic. The various concepts must, of course, be related to the physical world, but such motivating sections should be separated from the deductive part of the theory. Intuition will thus be strengthened, but not at the expense of logical rigor.

"There is an obvious lack of continuity between the elements of probability as presented in introductory courses, and the sophisticated concepts needed in today's applications. How can the average student, equipped only with the probability of cards and dice, understand prediction theory or harmonic analysis? The applied books give at most a brief discussion of background material; their objective is not the use of the applications to strengthen the student's understanding of basic concepts, but rather a detailed discussion of special topics.

"Random variables, transformations, expected values, conditional densities, characteristic functions cannot be mastered with mere exposure. These concepts must be clearly defined and must be developed, one at a time, with sufficient elaboration. Special topics should be used to illustrate the theory, but they must be so presented as to minimize peripheral, descriptive material and to concentrate on probabilistic content. Only then the student can learn a variety of applications with economy and perspective."

I realized that to teach a convincing course, a course that is not a mere presentation of results but a connected theory, I would have to reexamine not only the development of special topics, but also the proofs of many results and the method of introducing the first principles.

"The theory must be mathematical (deductive) in form but without the generality or rigor of mathematics. The philosophical meaning of probability must somehow be discussed. This is necessary to remove the mystery associated with probability and to convince the student of the need for an axiomatic approach and a clear distinction between assumptions and logical conclusions. The axiomatic foundation should not be a mere appendix but should be recognized throughout the theory.

"Random variables must be defined as functions with domain an abstract set of experimental outcomes and not as points on the real line. Only then infinitely dimensional spaces are avoided and the extension to stochastic processes is simplified.

"The inadequacy of averages as definitions and the value of an underlying space is most obvious in the treatment of stochastic processes. Time averages must be introduced as stochastic integrals, and their relationship to the statistical parameters of the process must be established only in the form of ergodicity.

"The emphasis on second-order moments and spectra, utilizing the student's familiarity with systems and transform techniques, is justified by the current needs.

"Mean-square estimation (prediction and filtering), a topic of considerable importance, needs a basic reexamination. It is best understood if it is divorced from the details of integral equations or the calculus of variations, and is presented as an application of the orthogonality principle (linear regression), simply explained in terms of random variables.

"To preserve conceptual order, one must sacrifice continuity of special topics, introducing them as illustrations of the general theory."

These ideas formed the framework of a course that I taught at the Polytechnic Institute of Brooklyn. Encouraged by the students' reaction, I decided to make it into a book. I should point out that I did not view my task as an impersonal presentation of a complete theory, but rather as an effort to explain the essence of this theory to a particular group of students. The book is written neither for the handbook-oriented students nor for the sophisticated few who can learn the subject from advanced mathematical texts. It is written for the majority of engineers and physicists who have sufficient maturity to appreciate and follow a logical presentation, but, because of their limited mathematical background, would find a book such as Doob's too difficult for a beginning text.

Although I have included many useful results, some of them new, my hope is that the book will be judged not for completeness but for organization and clarity. In this context I would like to anticipate a criticism and explain my approach. Some readers will find the proofs of many important theorems lacking in rigor. I emphasize that it was not out of negligence, but after considerable thought, that I decided to give, in several instances, only plausibility arguments. I realize too well that "a proof is a proof or it is not." However, a rigorous proof must be preceded by a clarification of the new idea and by a plausible explanation of its validity. I felt that, for the purposes of this book, the emphasis should be placed on explanation, facility, and economy. I hope that this approach will give you not only a working knowledge, but also an incentive for a deeper study of this fascinating subject.

Although I have tried to develop a personal point of view in practically every topic, I recognize that I owe much to other authors. In particular, the books "Stochastic Processes" by J. L. Doob and "Théorie des Functions Aléatoires" by A. Blanc-Lapierre and R. Fortet influenced greatly my planning of the chapters on stochastic processes.

Finally, it is my pleasant duty to express my sincere gratitude to Misha Schwartz for his encouragement and valuable comments, to Ray Pickholtz for his many ideas and constructive suggestions, and to all my colleagues and students who guided my efforts and shared my enthusiasm in this challenging project.

Athanasios Papoulis

Contents

I

PROBABILITY AND RANDOM VARIABLES

1 The Meaning of Probability

Scientific theories deal with concepts, never with reality. All theoretical results are derived from certain axioms by deductive logic. In physical sciences the theories are so formulated as to correspond in some useful sense to the real world, whatever that may mean. However, this correspondence is approximate, and the physical justification of all theoretical conclusions is based on some form of inductive reasoning.

The student accepts readily this separation between the *conceptual world* (model) and the *physical world* for the so-called deterministic phenomena, but in probabilistic descriptions he confuses these two worlds. He has been taught that the universe evolves according to deterministic laws that specify exactly its future, and a probabilistic description is necessary only because of our ignorance. This deep-rooted skepticism in the validity of probabilistic results can be overcome only by a proper interpretation of the meaning of probability.

In the following discussion we attempt to show that probability is also a deductive science and must be developed axiomatically. It is of course true that the correspondence between analytical results and the real world is imprecise and cannot be "proved"; however, this is characteristic not only of probabilistic results but of all scientific conclusions.

1-1. Preliminary remarks

The theory of probability deals with averages of mass phenomena occurring sequentially or simultaneously: electron emission, telephone calls, radar detection, quality control, system failure, games of chance, statistical mechanics, turbulence, noise, birth and death rates, heredity.

It has been *observed* that in these and other fields certain averages approach a constant value as the number of observations increases. Furthermore, this limiting value remains the same if the averages are evaluated over any sub-sequence specified before the experiment is performed. Thus, in the coin-tossing game, the percentage of heads approaches 0.5 or some other constant, and the same average is obtained if one considers every fourth, say, tossing. (No betting system can beat the roulette.)

The purpose of the theory is to describe and predict such averages, and this is done by associating probabilities with various events. The probability $P(\mathcal{A})$ of an event $\mathcal{A}$ in a clearly specified experiment $\mathfrak{S}$ could be interpreted in the following sense:

> If the experiment is repeated n times and the event $\mathcal{A}$ occurs $n_{\mathcal{A}}$ times, then, with a *high degree of certainty*, the relative frequency $n_{\mathcal{A}}/n$ of the occurrence of $\mathcal{A}$ is *close* to $P(\mathcal{A})$,
>
> $$P(\mathcal{A}) \simeq \frac{n_{\mathcal{A}}}{n} \tag{1-1}$$
>
> provided n is *sufficiently large.*

This interpretation is obviously imprecise; however, it cannot be essentially improved. One could modify it by giving, for example, a probabilistic content to the "high degree of certainty," but such modifications will only postpone the inevitable fact that probability, like any physical theory, is related to physical phenomena only in inexact terms. Nevertheless, the theory is an exact discipline developed logically from clearly defined axioms, and when it is applied to real problems, it *works.*

In any probabilistic investigation of a physical phenomenon one must distinguish three separate steps:

Step 1 (physical). We determine by a process that is not and cannot be made exact the probabilities $P(\mathcal{A})$ of certain events $\mathcal{A}$ (probabilistic data).

This step could be based on (1-1): $P(\mathcal{A})$ is equated to the experimentally determined ratio $n_{\mathcal{A}}/n$ (relative-frequency approach). For example, if a loaded die is rolled 1,000 times and five shows 203 times, then the probability of five equals about 0.2. In some cases, $P(\mathcal{A})$ is found a priori by pure reasoning without any experimentation (classical approach to be soon discussed). Given a "fair" die, one "reasons" that, because of its symmetry, the probability of five equals $\frac{1}{6}$.

Step 2 (conceptual). We assume that probabilities satisfy certain axioms, and by deductive reasoning we determine from the probabilities $P(\mathcal{A})$ of certain events $\mathcal{A}$ the probabilities $P(\mathcal{B})$ of other events $\mathcal{B}$.

For example, in the rolling of a fair die we deduce that the probability of the event "even" equals $\frac{3}{6}$. Our statement is of the following form:

$$\text{If} \qquad P(1) = P(2) = \cdots = P(6) = \tfrac{1}{6} \qquad \text{then} \qquad P(\text{even}) = \tfrac{3}{6}$$

Step 3 (physical). We make a physical *prediction* using the numbers $P(\mathcal{B})$ determined in step 2.

This step is again imprecise and could be based on (1-1). If, for example, we roll a fair die 1,000 times, we expect that an even number will show in about one-half of these rollings.

The theory of probability deals only with step 2; i.e., from certain assumed probabilities it tells us how to derive other probabilities. One might say that such derivations are mere tautologies because the results are contained in the assumptions. This is true in the same sense that the intricate equations of motion of a satellite are included in Newton's laws. But no one denies the value of the science of mechanics.

We could not emphasize too strongly the need for separating clearly the above three steps in the solution of a problem. The student must distinguish between the data that are either assumed or determined by inexact reasoning and the results that are obtained logically.

Steps 1 and 3 are the subject of investigation of statistics. However, even in statistics all results are given in terms of probabilities, with the difference that final experimental testing is applied to events whose probability is *almost* 1. In this case the relative-frequency interpretation takes the following form:

> If the probability of an event is *almost* 1, then, with a *high degree of certainty*, this event will occur in a single trial.

Even so, the difficulty of assigning probabilities to real events is not eliminated, and one might wonder how to proceed in a physical problem. To determine the probability of heads, should he toss a coin one hundred or one thousand times? Suppose that, after one thousand tossings, the average number of heads settled to the value 0.48. How can he make a *prediction* on the basis of this observation? Out of what logical necessity must he deduce that, at the next thousand tossings, the average will be about 0.48? This question can be answered only by some form of inductive reasoning. However, such reasoning is used, not only in probabilistic statements, but in all conclusions drawn from experience, even in the so-called deterministic sciences. Consider, for example, the development of classical mechanics. It was *observed* that bodies fell according to certain rules, and on the basis of this observation Newton's laws were formulated and used *successfully* to predict future events. If one wants to "prove" that the future will evolve in the predicted manner, he will have to invoke a metaphysical cause like "regularity in nature." The physicist bases his conclusions on inductive reasoning, and he is content that his predictions are correct. This point of view gives him also the flexibility to abandon any theory if subsequent evidence contradicts it. If he finds it useful to describe certain phenomena probabilistically, he does so without the need for a deterministic explanation.

To conclude, we repeat that the probability $P(\mathcal{A})$ of an event $\mathcal{A}$ must be interpreted as a number assigned to this event, as mass is assigned to a body. In the development of the theory one should not worry about the "physical meaning" of $P(\mathcal{A})$. This is what is done in all theories.

Consider, for example, circuit analysis. One assumes that a resistor is a two-terminal device whose voltage is proportional to the current

$$R = \frac{v(t)}{i(t)} \tag{1-2}$$

But a physical resistor does not obey (1-2). It is a complicated concept without obvious terminals, with distributed inductance and capacitance; it generates thermal noise; and only within certain errors in certain frequency ranges and with many other qualifications can a relationship of the form (1-2) be claimed. Nevertheless, in the development of the theory one ignores all these uncertainties. He assumes that a resistor has a value R satisfying (1-2), and with the help of certain laws, he develops circuit analysis. It would, indeed, be very confusing if at each stage of this development he were concerned with the *true* meaning of R.

Like circuit analysis, or electromagnetic theory, or any other scientific discipline, probability must be presented axiomatically. These theories would, of course, be of no value to physics unless they could help us solve real problems. We should be able to assign specific, if only approximate, values to real resistors or probabilities to certain events (step 1); we should also be able to give physical meaning to the quantities that were derived from the theory (step 3). This link between idealized concepts and the physical world is essential, but must be separated from the purely logical structure of each theory (step 2).