PROBABILITIES ON ALGEBRAIC STRUCTURES

Ulf Grenander

PROBABILITIES ON ALGEBRAIC STRUCTURES

Ulf Grenander

PROBABILITIES ON ALGEBRAIC STRUCTURES

BY

ULF GRENANDER

UNIVERSITY OF STOCKHOLM

ALMQVIST & WIKSELL
STOCKHOLM
GÖTEBORG · UPPSALA

JOHN WILEY & SONS INC.
NEW YORK · LONDON

PRINTED IN SWEDEN BY
Almqvist & Wiksells
BOKTRYCKERI AKTIEBOLAG
UPPSALA 1963

INTRODUCTION

This book attempts to present a unified and coherent theory of the calculus of probability on algebraic structures. Among such structures we shall consider some that merit special attention because of the role they play in the applications of the theory as well as because of their intrinsic interest. These are topological semi-groups and groups, topological vector spaces and algebras.

The author became interested in the subject through a number of practical problems, some of which are described in chapter 1. Therefore the emphasis will be on concrete results that can actually be used and that lead, at least in principle, to solutions that can be determined exactly or approximately, analytically or numerically. We do not try to reach results of the most general character or to develop the theory to its most refined form. Therefore we have not hesitated to impose conditions like separability, Borel measurability etc. in situations where this may not be necessary but where such conditions result in simplifications. Some reader may feel inclined to remove such imperfections, to reach necessary and sufficient conditions and so on. The author hopes that this will be done in the future.

To avoid obscuring the main ideas of the theory by lengthy computations and technical arguments such manipulations have sometimes been sketched only. This has been done especially when references could be cited to the literature where a complete treatment has been given.

To make the presentation as concrete as possible we will discuss a number of special cases at the end of every chapter. The author believes that some of these cases are as important as the theorems that they exemplify. They may sometimes indicate how to extend the theory.

The reader should observe that all the references as well as supplementary information are given in the Notes at the end of the book. Chapter 1 contains an outline of the history of the subject and also some remarks concerning the practical background to the theory. The arguments in this chapter are of a heuristic nature and will reappear in a rigorous form in the later chapters.

At a first reading the reader may prefer to skip sections 4.2 (where the very technical proofs are only sketched), 5.4 and 6.6 (which are of special nature).

It is difficult to specify the necessary prerequisites for reading the book, but it seems clear that the standard "calculus and mathematical maturity" would scarcely be adequate. Since the book is written for probabilists it will be assumed that the reader is well acquainted with probability and measure theory, say corresponding to the content of Loève: *Probability Theory* and Halmos: *Measure Theory*. Many of the measurability arguments used are of standard type and will only be hinted at. The reader should also have some knowledge of basic topological algebra. Neumark: *Normierte Algebren* is warmly recommended as a lucid and up to date presentation. Since this topic may be less well known among probabilists it seemed appropriate to be more complete when discussing related questions and the reader will find a number of the fundamental definitions and logical relations in the Notes. We must also ask of the reader that he know the elements of functional analysis. A suitable book would be Hille-Phillips: *Functional Analysis and Semi-groups*, which also contains some highly relevant information about semi-groups; this is almost indispensable when studying homogeneous processes.

The object of our study is the probability distribution on a structure in which at least one binary algebraic operation has been defined in such a way that it is continuous in a suitable topology. We can then talk of a stochastic element drawn at random from the structure. Composing two such independent stochastic elements via a binary operation we are led to convolutions (this is the key word!) of two probability distributions. To a considerable extent we shall study how convolutions behave, especially when we perform many of them. Our approach will be analytical, the main tool being Fourier analysis. There is no doubt that this is the correct approach if we want definite results, algorithms etc. There is however another way, more algebraic or perhaps probabilistic, in manner. We then consider the set of probability distributions in question as forming a topological semi-group. Applying semi-group theory we can reach certain results of considerable interest. Actually this is quite an attractive application of the general theory of semi-groups, but it will not be discussed more than occasionally in the text.

I have had a number of valuable discussions with colleagues whom I would like to thank for their suggestions: R. Fortet, E. Mourier, L. Schmetterer, Z. Šidak, and A. Špaček. I am very grateful to D. Wehn for putting a manuscript at my disposal before its publication. Sections 4.2–4.4. lean heavily on Wehn's results and exposition. M. Rosenblatt and W. Freibeiger read the entire manuscript and suggested many changes. R. Loynes and P. Martin-Löf scrutinized the book in detail and I am very grateful to them for their work. They spotted a number of mistakes or obscurities and also contributed some essential results to the theory.

I would also like to thank Försäkringsaktiebolaget Skandia and Statens Naturvetenskapliga Forskningsråd (Swedish Natural Science Research Council) for their financial support.

ULF GRENANDER

CONTENTS

CHAPTER 1

HISTORICAL BACKGROUND AND PRACTICAL MOTIVATION OF THE PROBLEM

1.1. Why study probabilities on general structures?

The domain of classical probability theory is the real line R^1. All the wellknown results like the central limit theorem, the law of the iterated logarithm or the law of large numbers concern real valued (or possibly vector valued in R^k) stochastic variables. The real line is so rich in structure that it can support the intricate but beautiful logical construction consisting of the probabilistic concepts and relations.

In mathematics in general there has been a trend towards generalization, abstraction, axiomatics. This is evident also in probability theory. Since Kolmogorov published his epoch-making *Grundbegriffe der Wahrscheinlichkeitsrechnung* we can use probabilistic arguments in completely general spaces without loosing anything in rigor. It is true that in such a generality we cannot always expect results of real mathematical substance, but the general framework is indispensable also for such, more concrete, work that is possible if the probability space is given more structure. The classical results indicate that such advance should be possible by defining *algebraic relations* in the space and studying their probabilistic implications. This leads us automatically to think of notions like groups, topological vector spaces and algebras. It is hard to resist posing the problem: does the classical probability theory have any counterpart in these more general algebraic structures? In the following chapters we shall see that sometimes this extension is made by an immediate and trivial generalization, sometimes a stronger effort is required leading to more profound discoveries, and sometimes we meet challenging problems to which the answers are known only partially if at all.

The fact that probability theory has grown so fast in recent years is due no doubt partly to its intrinsic value and its direct appeal to the mathematician but at least as much to its usefulness, exploited or potential. And so it is also with the present subject. Its motivation does not consist only of a wish to extend the theory to its natural boundaries. There are also a number of seemingly unrelated problems from physics, communication engineering, statistics and so on, that lead us to consider probabilistic relations in algebraic structures not equivalent to the real line (or the R^k-spaces). When more of these problems and results become widely known, we can expect an increased research activity and more rapid advance on both the practical and theoretical side of this subject.

We shall not jump directly into *medias res*. Instead let us start by recalling some fundamental facts and techniques from the classical theory (in 1.2.), then go on to review some problems posed by applications (in 1.3.), and finally (in 1.4.) sketch the outline of the historical development of the theory as it exists at present.

1.2. Classical methods and results

1.2.1. Granting the reader's permission we will take as our point of departure a brief sketch of some facts from elementary probability theory.

On the real line R^1 there are a number of basic probabilistic definitions that we will assume known as well as the corresponding fundamental relations,

probability measure
Borel measure
stochastic variable
independence
various modes of convergence of stochastic variables
(weak) convergence of probability distributions.

From the present point of view we are most interested in such relations that make use of the *additive* properties of R^1. Let P_1, P_2 be two probability measures, say defined through their distribution functions

$$F_i(y) = P_i\{x \mid x \leqslant y\}; \quad i = 1, 2.$$

If to each P_i there corresponds a stochastic variable x_i; $i = 1, 2$; then the sum $x = x_1 + x_2$ has a probability distribution P given by a distribution function $F(x)$. If x_1 and x_2 are independent F is given as the *convolution* of F_1 and F_2

$$F(x) = \int_{-\infty}^{\infty} F_1(x-y) F_2(dy),$$

or shorter $F = F_1 * F_2$.—We also have the wellknown modifications of this formula if the distributions are absolutely continuous with respect to Lebesgue measure or with respect to counting measure on the set of integers; say the densities are

$$f(x) = \frac{F(dx)}{m(dx)} \text{ and } f_i(x) = \frac{F_i(dx)}{m(dx)},$$

then
$$f(x) = \int f_1(x-y) f_2(y) m(dy).$$

Note that m is in both of these cases a translation invariant measure.— The convolution operation is commutative, $F_1 * F_2 = F_2 * F_1$, and associative $(F_1 * F_2) * F_3 = F_1 * (F_2 * F_3)$. By iteration we can define $F_1 * F_2 * \ldots * F_n(x)$. The study of such convolutions, especially for large values of n, is one of the main tasks of probability theory.

There are some elementary relations. If the mean values exist

$$m_i = \int_{-\infty}^{\infty} x F_i(dx) = Ex_i, \quad i = 1, 2, \ldots n,$$

then the mean value operation is additive

$$m = Ex = m_1 + m_2 + \ldots + m_n.$$

If second moments exist then the variances

$$\text{Var}(x_i) = E(x_i - m_i)^2 = \sigma_i^2$$

satisfy
$$\text{Var}(x) = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2,$$

if the stochastic variables x_i are independent. Under this hypothesis the variance is additive, and since a variance is non-negative the variance does not decrease when we add an independent stochastic element to a given one. In this sense (and in many others) convolution flattens out distributions. It also smoothes distributions, e.g. in the sense that if one of x_i has a continuous distribution then the same is true for the sum $x = x_1 + x_2 + \ldots + x_n$.

The most important analytical tool in this context is the *Fourier transform* or *characteristic function*

$$\hat{P}(z) = \varphi(z) = \int_{-\infty}^{\infty} e^{ixz} F(dx) = E \exp ixz, \; z \text{ real}.$$

The importance of the characteristic function originates in its three properties

a) the characteristic function determines the probability measure uniquely

b) convolution corresponds to ordinary multiplication:

$$\text{if } P = P_1 * P_2 \text{ then } \hat{P} = \hat{P}_1 \cdot \hat{P}_2$$

c) weak convergence of a sequence of probability distributions to a limit distribution corresponds to convergence of the characteristic functions to a continuous limit function.

The *moments* of P are defined by

$$\alpha_k = Ex^k$$

if these integrals exist. Then they can be expressed as derivatives of the characteristic function

$$\alpha_k = 1/i^k \varphi^{(k)}(0).$$

Related concepts are the *semi-invariants* (or *cumulants*)

$$\gamma_k = 1/i^k \left(\frac{d^k}{dz^k} \log \varphi(z)\right)_{z=0}$$

which are of course linear combinations of moments of the same and lower order. The cumulants are additive for independent distributions. It is sometimes said that moments (and cumulants) are clumsy to work with and less generally defined than the characteristic function. This may be so in general investigations, but they are very useful in many cases, also when limit theorems are concerned.

1.2.2. Now let us go ahead to some limit theorems. One of the oldest is the *theorem of Bernoulli* going back to antiquity of probability theory: if x is a binomial variable $\nu = B(n,p)$ the relative frequency

$$p^* = \frac{\nu}{n}$$

converges in probability to the constant p. Or if we write ν as a sum of independent indicator variables $\nu = x_1 + x_2 + \dots + x_n$, where $x_i = 1$ or 0 with probabilities p and $q = 1 - p$ respectively, then

$$p^* = \frac{1}{n}(x_1 + x_2 + \dots + x_n) \to p = Ex_i.$$

Now we know of course that the above formula holds much more generally. This can be phrased in many different ways but the most attractive version