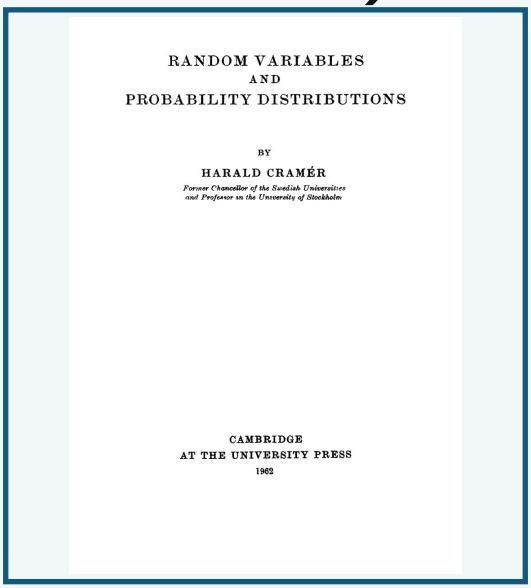
Probability Distributions (Harald Cramér)



RANDOM VARIABLES
AND
PROBABILITY DISTRIBUTIONS

BY

HARALD CRAMÉR

Former Chancellor of the Swedish Universities and Professor in the University of Stockholm

CAMBRIDGE
AT THE UNIVERSITY PRESS
1962

Complete Version Available On

ebookbe.com

r scan the QR code to access the digital study porta

RANDOM VARIABLES
AND
PROBABILITY DISTRIBUTIONS

BY

HARALD CRAMÉR

Former Chancellor of the Swedish Universities
and Professor in the University of Stockholm

CAMBRIDGE
AT THE UNIVERSITY PRESS
1962

PUBLISHED BY
THE SYNDICS OF THE CAMBRIDGE UNIVERSITY PRESS
Bentley House, 200 Euston Road, London, N.W. 1
American Branch: 32 East 57th Street, New York 22, N.Y
West African Office: P.O. Box 33. Ibadan, Nigeria

First printed 1937
Second edition 1962

First printed at the University Press, Cambridge
Reprinted by offset-lithography by Bradford & Dickens, London, W.C.1

CONTENTS

Preface to the First Edition . . . . . *page* vi

Preface to the Second Edition vii

Abbreviations viii

FIRST PART. PRINCIPLES

Chap. I. Introductory remarks 1

II. Axioms and preliminary theorems . . . 9

SECOND PART. DISTRIBUTIONS IN R_1

III. General properties. Mean values . . . 18

IV. Characteristic functions 24

V. Addition of independent variables. Convergence "in probability". Special distributions 36

VI. The normal distribution and the central limit theorem 50

VII. Liapounoff's theorem. Asymptotic expansions 70

VIII. A class of stochastic processes 90

THIRD PART. DISTRIBUTIONS IN R_k

IX. General properties. Characteristic functions 101

X. The normal distribution and the central limit theorem 110

Bibliography 116

Some Recent Works on Mathematical Probability . . 119

PREFACE

The Mathematical Theory of Probability has lately become of growing importance owing to the great variety of its applications, and also to its purely mathematical interest. The subject of this tract is the development of the purely mathematical side of the theory, without any reference to the applications. The axiomatic foundations of the theory have been chosen in agreement with the theory given by A. Kolmogoroff in his work *Grundbegriffe der Wahrscheinlichkeitsrechnung*, to which I am greatly indebted. In accordance with this theory, the subject has been treated as a branch of the theory of completely additive set functions. The method principally used has been that of *characteristic functions* (or *Fourier-Stieltjes transforms*).

The limitation of space has made it necessary to restrict the programme somewhat severely. Thus in the first place it has proved necessary to consider exclusively probability distributions in spaces of a finite number of dimensions. With respect to the advanced part of the theory, I have found it convenient to confine myself almost entirely to problems connected with the so-called *Central Limit Theorem* for sums of independent variables, and with some of its generalizations and modifications in various directions. This limitation permits a certain uniformity of method, but obviously a great number of important and interesting problems will remain unmentioned.

My most sincere thanks are due to my friends W. Feller, O. Lundberg and H. Wold for valuable help with the preparation of this work. In particular the constant assistance and criticism of Dr Feller has been very helpful to me.

H. C.

Department of Mathematical Statistics
University of Stockholm
December 1936

PREFACE TO THE SECOND EDITION

This Tract has now been out of print for a number of years. Since there still seems to be some demand for it, the Syndics of the Cambridge University Press have judged it desirable to publish a new edition.

However, owing to the vigorous development of Mathematical Probability Theory since 1937, any attempt to bring the book up to date would have meant rewriting it completely, a task that would have been utterly beyond my possibilities under present conditions. Thus I have had to restrict myself in the main to a number of minor corrections, otherwise leaving the work—including the Bibliography—where it was in 1937.

Besides the minor corrections, most of which are concerned with questions of terminology, there are, in fact, only two major alterations. In the first place, a serious error in the statement and proof of Theorem 11 has been put right. Further, the contents of Chapter IV, § 4, which are fundamental for the theory of asymptotic expansions, etc.. developed in Chapter VII, have been revised and simplified. This permits a new formulation of the important Lemma 4, on which the proofs of Theorems 24–26 are based. Finally a brief list of recent works on the subject in the English language has been added.

H.C.

University Chancellor's Office
Stockholm
March 1960

ABBREVIATIONS AND NOTATIONS

Symbol	Signification	Explanation
d.f.	Distribution function	*page* 11
pr.f.	Probability function	11
s.d.	Standard deviation	21
$E(X)$	Mean value (or mathematical expectation) of X	20
$D(X)$	Standard deviation of X	21
c.f.	Characteristic function	24
$F(x)=F_1(x)*F_2(x)$	$F(x)=\int_{-\infty}^{\infty} F_1(x-t)\,dF_2(t)$	37
convergence i.pr.	Convergence in probability	39
$(F(x))^{n*}$	$F(x)*F(x)*\dots(n$ times$)$	53

The *union* or *sum* of any finite or enumerable sequence of sets $S_1, S_2, \dots$ is denoted by

$$S=S_1+S_2+\dots.$$

The *intersection* or *product* of the sets $S_1, S_2, \dots$ is denoted by

$$S=S_1S_2\dots.$$

The inclusion sign $\subset$ is used in relations of the type $S_1 \subset S$ indicating that S_1 is a subset of S, and also in relations of the type $x \subset S$ to express the fact that x is an element of the set S.

FIRST PART

PRINCIPLES

CHAPTER I

INTRODUCTORY REMARKS

1. In the most varied fields of practical and scientific experience, cases occur where certain observations or trials may be repeated a large number of times under similar circumstances. Our attention is then directed to a certain quantity, which may assume different numerical values at successive observations. In many cases each observation yields not only one, but a certain number of quantities, say k, so that generally we may say that the result of each observation is a definite point X in a space of k dimensions ($k \geqslant 1$), while the result of the whole series of observations is a sequence of points: $X_1, X_2, \ldots$.

Thus if we make a series of throws with a given number of dice, we may observe the sum of the points obtained at each throw. We are then concerned with a variable quantity, which may assume every integral value between m and $6m$ (both limits inclusive), where m is the number of dice. On the other hand, in a series of measurements of the state of some physical system, or of the size of certain organs in a number of individuals belonging to the same biological species, each observation furnishes a certain number of numerical values, i.e. a definite point in a space of a fixed number of dimensions.

In certain cases, the observed characteristic is only indirectly expressed as a number. Thus if, in a mortality investigation, we observe during one year a large number of persons, we may at each observation (i.e. for each person) note the *number of deaths* which take place during the year, so that in this case the observed

quantity assumes the value 0 or 1 according as the corresponding person is alive at the end of the year or not.

In a given class of observations, let R denote the set of points which are *a priori* possible positions of our variable point X, and let S be a sub-set of R. Further, let a series of n observations be made, and count the number ν of those observations, where the following *event* takes place: *the point X determined by the observation belongs to S*. Then the ratio ν/n is called the *frequency* of that event or, as we may shortly put it, *the frequency of the relation (or event) $X \subset S$*. Obviously any such frequency always lies between 0 and 1, both limits inclusive. If $S = S_1 + S_2$, where S_1 and S_2 have no common point, and if ν_1/n and ν_2/n are the frequencies corresponding to S_1 and S_2, we obviously have $\nu = \nu_1 + \nu_2$ and thus

$$\nu/n = \nu_1/n + \nu_2/n. \tag{1}$$

When we are dealing with such frequencies, a certain peculiar kind of regularity very often presents itself. This regularity may be roughly described by saying that, for any given sub-set S, the frequency of the relation (or event) $X \subset S$ *tends to become more or less constant as n increases*. In certain cases, such as e.g. cases of biological measurements, our observations may be regarded as samples from a very large or even infinite population, so that for indefinitely increasing n the frequency would ultimately reach an ideal value, characteristic of the total population.

It is thus suggested that in cases where the above-mentioned type of regularity appears, we should try to introduce a number $P(S)$ to represent such an ideal value of the frequency ν/n corresponding to the sub-set S. The number $P(S)$ is then called *the probability of the sub-set S, or of the event $X \subset S$*. It follows from (1) that we should obviously choose $P(S)$ such that

$$P(S_1 + S_2) = P(S_1) + P(S_2) \tag{2}$$

for any two sub-sets S_1 and S_2 of R which have no common point. Further, it is obvious that we should always have $P(S) \geqq 0$ and that for the particular set $S = R$ we should have $P(R) = 1$.

The investigation of *set functions* of the type $P(S)$ and their mutual relations is the object of the Mathematical Theory of Probability. This theory should be considered as a branch of Pure Mathematics, founded on an axiomatic basis, in the same sense as Geometry or Theoretical Mechanics.[1] Once the fundamental conceptions have been introduced and the axioms have been laid down (and in this procedure we are, of course, guided by empirical considerations), the whole body of the theory should be constructed by purely mathematical deductions from the axioms. The practical value of the theory will then have to be tested by experience, just in the same way as a theorem in euclidean geometry, which is intrinsically a purely mathematical proposition, obtains a practical value because experience shows that euclidean geometry really conforms with sufficient accuracy to a large group of empirical facts.

2. The axiomatic basis of a theory may, of course, always be constructed in many different ways, and it is well known that, with respect to the foundations of the Theory of Probability, there has been a great diversity of opinions.

The type of statistical regularity indicated above was first observed in connection with ordinary games of chance with cards, dice, etc., and this gave occasion to the origin and early development of the theory.[2] In every game of this character, all the results that are *a priori* possible may be arranged in a finite number of cases which are supposed to be perfectly symmetrical. This led to the famous *principle of equally possible cases* which, after having been more or less tacitly assumed by earlier writers, was explicitly framed by Laplace [1], as the fundamental principle of the whole theory. Throughout the whole century following the publication of Laplace's classical treatise, a large amount of work has been spent on the discussion of this principle.

During the course of this discussion, it has been maintained

[1] This view seems to have been first explicitly expressed by v. Mises [2].
[2] Cf. Todhunter [1].

by various authors that the validity of the principle of equally possible cases is necessarily restricted to the field of games of chance, so that it is wholly incapable of serving as the basic principle of the theory. Attempts have been made[1] to establish the theory on an essentially different basis, the probabilities being directly defined as ideal values of statistical frequencies. The most successful attempt on this line is due to v. Mises [2, 3], who endeavours to reach in this way an axiomatic foundation of the theory in the modern sense.

The fundamental conception of the v. Mises theory is that of a "*Kollektiv*", by which is meant an unlimited sequence K of similar observations, each furnishing a definite point belonging to an *a priori* given space R of a finite number of dimensions. The first axiom of v. Mises then postulates the existence of the limit

$$\lim_{n\to\infty} \nu/n = P(S) \tag{3}$$

for every simple sub-set $S \subset R$, while the second axiom requires that the analogous limit should still exist and have the same value $P(S)$ for every sub-sequence K' that can be formed from K according to a rule such that it can always be decided whether the nth observation of K should belong to K' or not, *without knowing the result of this particular observation.*[2] It does, however, seem difficult to give a precise mathematical meaning to the condition printed in italics, and the attempts to express the second axiom in a more rigorous way do not, so far, seem to have reached satisfactory and easily applicable results. Though fully recognizing the value of a system of axioms based on the properties of statistical frequencies, I think that these difficulties must be considered sufficiently grave to justify, at least for the time being, the choice of a fundamentally different system.

The underlying idea of the system that will be adopted here

[1] For the history of these attempts, cf. Keynes [1], chaps. VII–VIII.

[2] The second axiom as given by v. Mises [3], p. 18, is somewhat more complicated. It can, however, be shown that this is equivalent to the simpler statement given above.