Stochastic Analysis

Edited by D.G. Kendall & E.F. Harding

Stochastic Analysis

A Tribute to the Memory
of
Rollo Davidson

Edited by
D. G. KENDALL
and
E. F. HARDING

John Wiley & Sons

London *New York* *Sydney* *Toronto*

Library of Congress catalog card number 72-8605

ISBN 0 471 46890 8

Printed in Great Britain by John Wright & Sons Ltd., at The Stonebridge Press, Bristol, England

Preface

This volume, and the companion work *Stochastic Geometry*, have been compiled by the friends of Rollo Davidson and by one or two others who had not met him but who hoped to become his friends. The two books are very closely linked, and those who wish to explore either one in depth will almost certainly need to consult the other. The degree of unity within each single book is, however, much greater than that which could have been attained by bringing them together, and we believe that most people will find the two-volume arrangement a great convenience.

We have taken some pains to try to avoid the 'miscellaneous' character often to be found in such cooperative works; both books contain much that is new, even to the specialist, but we believe that each will be found valuable as an introduction to the field named in its title, especially by those interested in the possibility of undertaking research in either area and daunted by the scattered nature of the periodical literature and the absence of any comparable synthesis of it.

Each book is furnished with an introductory chapter surveying the field, and setting the stage for the specialist articles which follow it.

This volume contains a very large part of the periodical literature on the theory of Delphic semigroups, and is in fact the first book in any language dealing with that subject. It also presents a large number of new results in the theory of Kingman's p-functions which characterize regenerative phenomena, and thus it complements his recent monograph *Regenerative Phenomena* (John Wiley & Sons, London, 1972). Other papers presented here for the first time discuss various special classes of stochastic processes, and a number of surveys are also included, dealing for example with the sample-path properties of additive processes, and with Doob's 'theory of versions'.

We are especially grateful for the trouble which has been taken by Mr D. S. Griffeath, Professor J. F. C. Kingman, and Professor G. E. H. Reuter in preparing for publication one of Davidson's most remarkable unpublished manuscripts from a rough draft found among his papers. We also wish to thank Dr G. K. Eagleson and Dr D. N. Shanbhag for

their assistance with another Davidson manuscript, and Miss Mary Brooks and Miss Madeleine Wuidart who drew the pictures and made the index.

Chapters 2.1–2.5 were first published by Springer–Verlag; 2.6 by the Cambridge Philosophical Society; and 2.7 by the Academy of Sciences, Paris. To all these bodies we are most grateful for their kindness in making possible the complete coverage attempted in this book.

D. G. K.
E. F. H.

Contents

1.1

An Introduction to Stochastic Analysis

D. G. KENDALL

(1) Stochastic analysis is the field of interest of the members of a loosely knit body called the Stochastic Analysis Group, which was formed in Oxford in December 1961 to promote interest in the analytical aspects of probability theory among mathematicians and statisticians in the United Kingdom. Very much has been achieved since then, and the primary purpose of the Group may be said to have been attained, as the flourishing state of the two magnificent journals (*Journal of Applied Probability*, *Advances in Applied Probability*) published from Sheffield University under the editorship of Professor Gani bear witness. (Most of the published work emanating from the Group has appeared in their pages, or in those of the journal *Zeitschrift für Wahrscheinlichkeitstheorie*, which was founded during the same period, and has been edited with such distinction by Professor Schmetterer.) There are, however, still some who need an occasional reminder that the fashionable trees of statistics and operational research draw some of their nourishment from mathematical roots, as well as from the photosynthetic activities of practical consultation. In this review, intended primarily for mathematicians, we shall therefore take the opportunity not only to say what stochastic analysts do, and why what they do is useful, but also to give some indication of the mathematical foundations of the subject, and of its links with other branches of mathematics.

(2) We shall begin by recalling that a real-valued random variable X is a measurable mapping from a probability-space $(\Omega, \mathscr{F}, \mathrm{pr})$ (Ω any non-vacuous set, $\mathscr{F}$ a σ-algebra of subsets thereof and pr a non-negative measure on $\mathscr{F}$ of total mass 1) into the real line R endowed with Borel sets. When this random variable is considered in isolation from others, what is important is not the mapping X itself but rather the probability

measure P_X defined on the Borel subsets B of the line by

(1) $$P_X(B) = \mathrm{pr}\,(X^{-1}B).$$

This measure P_X is called the 'distribution' of X, and it tells us the probability with which the realized 'value' $X(\omega)$ $(\omega \in \Omega)$ will fall in the generic Borel set B. Obviously many random variables, perhaps defined over different probability-spaces, will have the same distribution, and we can regard the probability-space

$$(R, \mathscr{B}(R), P_X)$$

(with the identity-mapping) as supplying a uniquely defined 'canonical' model for all of them. Here Ω has been replaced by the real line R, $\mathscr{B}(R)$ denotes the σ-algebra of Borel sets in R and P_X is the distribution just defined. The identity-mapping of R into R *is* the random variable. For some purposes an equally well-defined and minor adjustment of the model is convenient, in which $\mathscr{B}(R)$ is replaced by the smallest σ-algebra $\mathscr{B}^+(R)$ containing both all the Borel sets and also all subsets of Borel sets having P_X-measure zero. That there exists a unique consistent extension of P_X from $\mathscr{B}(R)$ to $\mathscr{B}^+(R)$ is a familiar fact, and we shall often use this and similar 'completion' procedures in the pages which follow. (The passage from Borel sets to Lebesgue-measurable sets on the line is an instance of the same technical device, and indeed a special case of the procedure described here if we replace R by the unit segment and P_X by Lebesgue measure.) Notice, however, that in the canonical model,

$$\text{identity}: R \to R,$$

while we can use either $\mathscr{B}(R)$ or $\mathscr{B}^+(R)$ in the measurability condition on the left-hand side, we must always use $\mathscr{B}(R)$ on the right-hand side.

(3) When two random variables X and Y are of interest, defined over the same probability-space, their two distributions P_X and P_Y do not suffice for their study save in the very exceptional case when X and Y are 'independent'; the pair (X, Y) is a measurable mapping from the probability-space into the plane R^2, and we define the 'joint distribution' $P_{X,Y}$ by

(2) $$P_{X,Y}(B) = \mathrm{pr}\,((X, Y)^{-1}B)$$

for all planar Borel sets B. Here (and in the higher but finite-dimensional analogues) it is obvious what the analogous canonical models (using the σ-algebra of Borel sets, or its completion) should be. Nor (as we shall see in more detail below) is there any difficulty in extending these definitions to random variables which take their values in an arbitrary second countable compact Hausdorff space Z, and the specific examples just mentioned can be included within such a generalization by an appropriate compactification of $R, R^2, \ldots$, etc. The Alexandrov one-point compactification is convenient save in the case of R, and here the two-point compactification $R \cup \{-\infty, \infty\}$ can be used instead, if desired, and has certain obvious advantages. (There are also some special problems connected with Markov processes where quite sophisticated compactifications may be appropriate.) We indicate such compactifications by a bar; thus $\bar{R}$ will denote the compactified real line.

(4) Now let us generalize these ideas to *an indexed family* X of component random variables X_α, where α ranges through some arbitrary index-set A, called the 'parameter-set'; each one of the random variables takes its values in R, compactified to $\bar{R}$, or more generally in any fixed second countable compact Hausdorff space Z which is called the 'state-space'. Thus, if ω in Ω is fixed, then $X(\omega)$ maps A into Z, while if α in A is fixed, then X_α maps Ω into Z. In such a situation we speak of the whole family X in association with the probability-space $(\Omega, \mathscr{F}, \text{pr})$ as a 'stochastic process'. The terminology harks back to the days when the parameter-set A was invariably the real line, the non-negative half-line, the integers or the non-negative integers, and was thought of as the time-axis; in those more special situations we can think of $X_\alpha(\omega)$ as specifying the realized 'state' of a randomly developing system as 'time' α 'proceeds' through A, the identification of ω in Ω having fixed all the chance contributions to this development. More generally, when ω is free, we can think of X as a generic element of the space Z^A consisting of all Z-valued functions over A; any one such function is called a 'path' or 'trajectory', and for given ω the particular function $X(\omega)$ is called the 'sample path'. In fact there is no distinction now between a stochastic process and a random function; the domain of the function can be arbitrary, and a very wide range of choices is available for its range Z. A detail which will be obvious, but which requires emphasis, is that we do not have a stochastic process X unless the component random variables X_α are all defined over the same probability-space. The first step will be to seek the proper analogue to the formulae (1) and (2), and the instinctive solution, to accept equation (1) as it stands with $X = \{X_\alpha \colon \alpha \in A\}$ and with B a Borel set, turns out to be *the wrong one* (save for rather special parameter-sets A).

Now the system of Borel sets on the compactified line $\bar{R}$ can usefully be thought of as the smallest σ-algebra containing the open intervals and at least one compactification-point, or alternatively we can describe it as the smallest σ-algebra containing all the half-open intervals $(x', x'']$, where x' and x'' are extended real numbers and $x' \leqslant x''$. It is entirely reasonable to use this σ-algebra on $\bar{R}$ because

(3) $$x' < X \leqslant x''$$

typifies a practically possible observation on a real-valued random variable X. If we follow up this idea we see that a typical practically possible observation on a real-valued *stochastic process* X will take the form

(4) $$x'_{\alpha_j} < X_{\alpha_j} \leqslant x''_{\alpha_j} \quad \text{for } j = 1, 2, \ldots, n,$$

where n is any positive integer, the αs are in A, and the xs are extended real numbers. If we ask what is the smallest σ-algebra of subsets of $\bar{R}^A$ which contains all sets of the form (4), where now X_α is thought of as the αth coordinate of a generic point in that function-space, the answer is *the σ-algebra $\mathscr{B}_0$ of Baire sets*. This is *not* the same thing as the σ-algebra $\mathscr{B}$ of Borel sets, and to make this clear it will suffice to say that here (and also more generally) the σ-algebra of Borel sets is that which contains all the compact sets, and is otherwise minimal, while the σ-algebra of Baire sets is that which contains all those compact sets *which happen to be expressible as countable intersections of open sets*, and is otherwise minimal; these definitions work for any compact Hausdorff space, and show that in general $\mathscr{B}$ is bigger than $\mathscr{B}_0$. Of course $\bar{R}^A$ is compact Hausdorff, and so we can use these definitions in our work. (A good general reference for Baire and Borel sets is Halmos [29], but the reader should be prepared to find his account considerably more complicated; this is due to the fact that he does not confine his attention to the compact case.)

(5) It is obviously important to know when the Borel and Baire σ-algebras coincide, and there are just two facts concerning this matter which we shall need to use here:

(*i*) If the space is not merely compact Hausdorff, but is in addition *second countable*, then the two σ-algebras will be identical. We shall assume throughout that our *component* random variables range through state-spaces Z of this sort, and so in the definition of such a random variable, where we require the mapping

(5) $$X_\alpha : \Omega \to Z$$

to be measurable, we can use the Borel or Baire σ-algebras indifferently on the right-hand side, for they are the same.

(*ii*) If the space is merely compact Hausdorff, and not necessarily second countable, then in general $\mathscr{B}$ can be properly bigger than $\mathscr{B}_0$. In particular this can happen if we are talking about the space Z^A; Z^A is a Cartesian product of copies of the space Z, and so inherits both its compact and Hausdorff properties, but it will inherit its second countable character if and only if the parameter-set A is countable. We can in fact be more precise than this: *in Z^A the systems of Borel and Baire sets coincide if and only if A is countable.* (A useful fact to bear in mind is that the singleton sets in Z^A are always Borel sets, because they are compact, but they are Baire sets if and only if A is countable.)

(6) Let us now apply (*i*) and (*ii*) above to the most general situation we shall want to consider; suppose in fact that we have a probability-space $(\Omega, \mathscr{F}, \mathrm{pr})$ and that we have associated with it a family of component mappings (for $\alpha \in A$)

(6) $$X_\alpha: \Omega \to Z,$$

which we can also think of as a combined mapping

(7) $$X: \Omega \to Z^A$$

into the Cartesian-product space. In order to call $X = \{X_\alpha: \alpha \in A\}$ a stochastic process, we want each mapping (6) to be measurable (it being understood that Ω carries the σ-algebra $\mathscr{F}$). We have already remarked at (*i*) in the preceding section that we can use the Baire or Borel σ-algebra indifferently in Z, and so the question of what we mean by the random-variable status of X_α is not in question; we simply mean that $X_\alpha^{-1} B \in \mathscr{F}$ whenever B is a Baire (= Borel) set in Z.

Now let us look at the mapping (7); the only kind of observation we can imagine actually making on such a process would be of the form

(8) $$X_{\alpha_j} \in B_j \quad \text{for } j = 1, 2, \ldots, n,$$

where n is a positive integer, the αs are in A and the Bs are Baire (= Borel) sets in Z. It is obvious, therefore, that we must require (8) to determine a measurable subset of Z^A for every choice of n, the αs and the Bs, and the smallest σ-algebra which contains all these sets as members is *the Baire σ-algebra $\mathscr{B}_0$ on Z^A*. We can therefore call X, defined over $(\Omega, \mathscr{F}, \mathrm{pr})$, a stochastic process when and only when

$$X^{-1} B \in \mathscr{F} \quad \text{for every Baire set } B \text{ in } Z^A.$$

There is another approach which has some interest and leads to the same conclusion. Let Φ denote any continuous real-valued function over Z^A. Then as a minimal desideratum for the stochastic process (7) we might reasonably demand that the composed mapping

$$\Phi \circ X: \Omega \to R$$

should determine a real-valued random variable, and this immediately suggests that we should use, as the σ-algebra on Z^A, the smallest one which makes the mapping

$$\Phi : Z^A \to R \tag{9}$$

measurable for every continuous Φ (R as usual carrying the Borel sets, i.e. the sets in $\mathscr{B}(R)$). But this, once again, turns out to be the Baire σ-algebra on Z^A.

(7) So much being agreed, it will be seen that the correct analogue to equations (1) and (2) will be as follows: the *distribution* (or, as I shall prefer to call it, the *name*) of the stochastic process

$$(\Omega, \mathscr{F}, \mathrm{pr}; Z, A; X) \tag{10}$$

is the probability measure P_X defined by

$$P_X(B) = \mathrm{pr}(X^{-1} B) \tag{11}$$

on the Baire σ-algebra $\mathscr{B}_0$ for Z^A. Our terminology here is dictated by the fact that, when Z and A have been given, the only property of the structure (10) which is of the slightest practical importance is the measure P_X on $\mathscr{B}_0$. If two variants of the structure (10) have the same Z, the same A and the same P_X, then there is absolutely no reason for using one rather than the other apart from questions of aesthetics or analytical expediency. Thus, Z and A being fixed, P_X characterizes an equivalence class of structures (10) which there can be no *practical* reason to refine further. If we write $\mu = P_X$, then for practical purposes (Z, A, μ) defines the process, and this is why (Z and A being normally fixed in any such discussion) we call μ the 'name' of the process. Any structure (10) having μ as its name will here be called a *version* of the process; the collection of all versions with name μ will be called the *name-class* of μ. The methodology of a good deal of stochastic analysis consists in the shrewd choice of an appropriate version for the particular analytical exercise one has in view. As a glance at the literature will show, versions proliferate at an alarming rate, and we shall make a serious attempt here to 'platonize' the situation by making precise what we mean by those versions, which we shall then call *models*, which are *canonical*.

(8) Before proceeding to this matter, let us note one immediate corollary to the discussion so far. We have observed that we are at liberty to confound the Borel and Baire sets in Z^A if and only if A is countable, and the resulting simplifications in the interaction between the measure theory and the topology are so useful that there is an unavoidable methodological gulf between countable-A problems and uncountable-A problems. For

example, this is why in the more classical parts of the theory 'discrete-time' stochastic processes (with $A = \{0, 1, 2, \ldots\}$) are so much more easy to deal with than 'continuous-time' stochastic processes (with $A = \{\alpha\colon 0 \leqslant \alpha < \infty\}$). It will be obvious that one would find it intolerable to have to sacrifice continuous-time stochastic processes altogether, so that the Baire–Borel contrast cannot in general be shirked.

It might perhaps be thought that a drastic simplification in the nature of the state-space Z would have a healing effect on the breach, but this is not so; the difficulties are fully present even in the apparently trivial case when Z contains just two points.

(9) A large number of important questions now pose themselves almost automatically. One, to which we shall return later, is of considerable difficulty. Suppose that we want to talk about random functions having some special *property* Γ; for example we may have in mind the Bachelier–Lévy–Wiener theory of Brownian motion, where $Z = \bar{R}$ and $A = R$, and we may want to be able to speak of this random motion as a random *continuous* motion. Expressed in the notation we are using here, this means that we are dealing with a random path X which (for any fixed ω) is a point in $Z^A = \bar{R}^R$, and we want to be able to say, perhaps with probability one, that X lies in the subset $C(A) = C(R)$ of $Z^A = \bar{R}^R$. Many of the 'nice properties' one would like a sample path to have are of this character; they amount to a requirement of the form $X(\omega) \in \Gamma \subset Z^A$, but often, as here, the portion Γ of path-space to which we should like to restrict ourselves is not a Baire set, and so immediately we are in trouble. We shall see in due course how this difficulty can be turned, but for the moment we record only the following very useful 'rule-of-thumb'. If the specification of the subset Γ involves *essential* reference to uncountably many values of the parameter α, then Γ cannot possibly be a Baire set. If, on the other hand, only countably many αs are involved in the specification of Γ, then Γ *may* be a Baire set. This is perhaps also a good point at which to mention that not all 'nice properties' can be thrown into the form '$X(\omega) \in \Gamma$', and those which cannot be so expressed (e.g. process-measurability) raise further difficulties of a different kind.

We now leave this matter for the moment, and turn to two other problems which have, happily, been fully solved. We have chosen to characterize a name-class of stochastic processes by a rather complicated object: a probability measure μ on the Baire sets of Z^A. Can we simplify this characterization, and can we do so in such a way that the new formulation preserves both the unicity and existence properties? We want to be sure that, when we have learned how to 'spell' the 'name' more simply, it still characterizes the name-class of processes *uniquely*, and also that a name-class always *exists* for it to characterize.

(10) The simplification of the 'name' can be carried out in various ways, and it would be undesirable to go into much detail on this matter here, beyond saying that all amount to specifying in one way or another the system of *finite-dimensional distributions*,

(12) $$\mathrm{pr}_{X_{\alpha_1}, X_{\alpha_2}, \ldots, X_{\alpha_n}} \quad \text{on } \mathscr{B}(Z^n),$$

where n runs through the positive integers and the αs through A. (Note that $\mathscr{B}(Z^n)$ in the expression (12) is the same as $\mathscr{B}_0(Z^n)$.) Clearly the 'name' determines each of the finite-dimensional distributions (which are in fact just the 'names' of all the finite subprocesses). It is quite easy to prove (by monotone-class or by Dynkin's π/λ arguments—see [21] for the latter) that if two stochastic processes determine the same finite-dimensional distributions (12), then they must belong to the same name-class, and conversely. So we are merely left with the question, which are the systems of finite-dimensional distributions that can be associated with name-classes of stochastic processes? This question of *existence* lies a lot deeper than that of *unicity*, but fortunately (with the topological assumption on Z which we assume throughout) the answer is fully known, and is agreeably simple. It was given by P. J. Daniell in 1918 [14, 15] and later, but more definitively, by A. N. Kolmogorov in 1933 [52], and it can be expressed thus: the finite-dimensional distributions correspond to a name-class of stochastic processes when, and only when, they satisfy two trivially obvious consistency conditions associated with (*a*) dropping one of the αs and (*b*) permuting the αs, respectively. Though the theorem is easy to state, it is not so easy to prove. The modern proof uses functional analysis (Stone–Weierstrass plus Riesz) and is associated with the names of Bourbaki and Nelson. (See Nelson [63] or Meyer [61].)

The Daniell–Kolmogorov theorem proves existence by constructing a special member of the name-class which is deservedly called *the* canonical model,

(13) $$(Z^A, \mathscr{B}_0(Z^A), \mu; Z, A; X).$$

Here we have written out $\mathscr{B}_0(Z^A)$ in full, but in future we shall just call this $\mathscr{B}_0$ (and similarly for $\mathscr{B}_0^+$, $\mathscr{B}$, etc.). In (13) μ denotes the 'name' of the name-class of which the model (13) is to be the canonical representative, and X denotes, here and henceforth, the family of 'coordinate-mappings' from Z^A into Z:

(14) $$(.)_\alpha : Z^A \to Z \quad (\alpha \in A).$$