Cambridge University Press
978-0-521-71821-9 — Trends in Stochastic Analysis
Edited by Jochen Blath , Peter Mörters , Michael Scheutzow
Excerpt
More Information

I. Foundations and techniques in stochastic analysis

1

Random variables – without basic space

Götz Kersting
Fachbereich Informatik und Mathematik
Universität Frankfurt
D-60054 Frankfurt am Main, Germany

Abstract

The common definition of a random variable as a measurable function works well 'in practice', but has conceptual shortcomings, as was pointed out by several authors. Here we treat random variables not as derived quantities but as mathematical objects, whose basic properties are given by intuitive axioms. This requires that their target spaces fulfil a minimal regularity condition saying that the diagonal in the product space is measurable. From the axioms we deduce the basic properties of random variables and events.

1.1 Introduction

In this paper we define the concept of a *stochastic ensemble*. It is our intention thereby to give an intuitive axiomatic approach to the concept of a random variable. The primary ingredient is a sufficiently rich collection of random variables (with 'good' target spaces). The set of observable events will be derived from it.

Among the notions of probability it is the *random variable* which in our view constitutes the fundamental object of modern probability theory. Albeit in the history of mathematical probability *events* came first, random variables are closer to the roots of understanding nondeterministic phenomena. Nowadays events typically refer to random variables and are no longer studied for their own sake, and for *distributions* the situation is not much different. Moreover, random variables turn out to be flexible mathematical objects. They can be handled in other ways than events or distributions (think of couplings), and these ways often conform to intuition. 'Probabilistic', 'pathwise' methods gain importance and combinatorial constructions with random variables can substitute (or nicely prepare) analytic methods. It was a common belief that first

Trends in Stochastic Analysis, ed. J. Blath, P. Mörters and M. Scheutzow.
Published by Cambridge University Press. ©Cambridge University Press 2008.

of all the distributions of random variables matter in probability, but this belief is outdated.

Today it is customary to adapt random variables to a context from measure theory. Yet the feeling has persisted that random variables are objects in their own right. This was manifest, when measure theory took over in probability: According to J. Doob (interviewed by Snell [9]) ‘it was a shock for probabilists to realize that a function is glorified into a random variable as soon as its domain is assigned a probability distribution with respect to which the function is measurable’. Later the experts insisted that it is the idea of random variables which conforms to intuition. Legendary is L. Breiman’s [2] statement: ‘Probability theory has a right and a left hand. On the right is the rigorous foundational work using the tools of measure theory. The left hand “thinks probabilistically,” reduces problems to gambling situations, coin-tossing, and motions of a physical particle’. In applications of probability the concept of a random variable never lost its appeal. We may quote D. Mumford [8]: ‘There are two approaches to developing the basic theory of probability. One is to use wherever possible the reduction to measure theory, eliminating the probabilistical language ... The other is to put the concept of “random variable” on center stage and work with manipulations of random variables wherever possible’. And, ‘for my part, I find the second way ... infinitely clearer’.

Example To illustrate this assertion let us consider different proofs of the central limit theorem saying that $(X_1 + \ldots + X_n)/\sqrt{n}$ is asymptotically normal for iid random variables $X_1, X_2, \ldots$ with mean 0 and variance 1. There is the established analytic approach via characteristic functions. In contrast let us recall a coupling method taken from [2], which essentially consists in replacing $X_1, \ldots, X_n$ one after the other by independent standard normal random variables $Y_1, \ldots, Y_n$. In more detail this looks as follows: Let $f : \mathbb{R} \to \mathbb{R}$ be thrice differentiable, bounded and with bounded derivatives. Then it is sufficient to show that

$$\mathbf{E}\Big[f\Big(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\Big) - f\Big(\frac{Y_1 + \cdots + Y_n}{\sqrt{n}}\Big)\Big]$$

converges to zero. The integrand may be expanded into

$$\sum_{i=1}^{n} \Big[f\Big(\frac{Z_i}{\sqrt{n}}\Big) - f\Big(\frac{Z_{i-1}}{\sqrt{n}}\Big)\Big]$$

with $Z_i := X_1 + \cdots + X_i + Y_{i+1} + \cdots + Y_n$. By means of two Taylor expansions around $U_i := X_1 + \cdots + X_{i-1} + Y_{i+1} + \cdots + Y_n$ the summands turn into

$$\frac{X_i - Y_i}{\sqrt{n}} f'(U_i) + \frac{X_i^2 - Y_i^2}{2n} f''(U_i) + O_P(n^{-3/2})$$

Taking expectations the first two terms vanish because of independence, and a closer look at the remainder gives the assertion (see [2], page 168). □

From an architectural point of view these considerations and statements suggest that we try to start from random variables in the presentation of probability theory and therewith to bring intuition and methods closer together – rather than to gain random variables as derived quantities in the accustomed measure-theoretic manner. We like to show that this can be accomplished without much technical effort. For this purpose we may leave aside distributions in this paper.

Let us comment on the difference of our approach to the customary one of choosing a certain σ-field $\mathcal{E}$ on some basic set Ω as a starting point and then indentifying events and random variables with measurable sets and measurable functions. In our view this is a set-theoretic *model* of the probabilistic notions.

To explain this first by analogy let us recall how natural numbers are treated in mathematics. There are two ways: Either one starts from the well-known Peano axioms. Then the *set* of natural numbers is the object of study, and a natural number is nothing more than an element of this structured set. Or natural numbers are introduced by a set-theoretic construction, e.g. $0 := \emptyset, 1 := \{0\}, \ldots, n+1 := n \cup \{n\}, \ldots$ (see [7]). This setting exhibits aspects which are completely irrelevant for natural numbers (such as $n \in n+1$ and $n \subset n+1$) and which stress that we are dealing with a *model* of the natural numbers. Different models may each have additional properties and structures which are actually irrelevant (and to some extent misleading) for the study of natural numbers. It is not important that they are 'isomorphic' in any respect.

Analoguous observations can be made in our context, if events and random variables are represented by a measure space $(\Omega, \mathcal{E})$ and associated measurable functions. Note the following: There are subsets of Ω not belonging to $\mathcal{E}$, which are totally irrelevant. To some extent this is also true for the elements ω of Ω (as also Mumford [8] pointed out;

already Caratheodory considered integration on spaces without points in his theory of soma [3]). The 'small omegas' do not show up in any relevant result of probability theory, and one could do without them, if they were not needed to *define* measurable functions. Next the notion of a random variable is ambiguous: There are random variables and a.s. defined random variables, represented by measurable functions and equivalence classes of measurable functions. This distinction, though unavoidable in the traditional setting, is somewhat annoying. Finally note that probabilists leave aside the question of isomorphy of measurable spaces.

All these observations indicate that measurable spaces and mappings indeed make up a model of events and random variables. This is not to say that such models should be avoided, but one should not overlook that they might mislead. Aspects like the construction of non-Borel-sets are of no relevance in probability and may distract beginners. Also one should be cautious in giving the elements of Ω some undue relevance ('state of the world'), which may create misconceptions.

Example This example of possible misconception is taken from the textbook [1] (Example 4.6 and 33.11). Let $\Omega = [0, 1]$, endowed with the Borel-σ-field and Lebesgue-measure λ. Let $\mathcal{F}$ be the sub-σ-field of sets B with $\lambda(B) = 0$ or 1. Then $\mathcal{F}$ presents an observer, who lacks information. It is mistaken to argue that $\mathcal{F}$ presents full information, because it contains all one-point sets such that the observer can recognize which event $\{\omega\}$ takes place and which 'state' ω is valid. Therefore for any Borel-set $E \subset \Omega$ the conditional probability $\lambda(E|\mathcal{F})$ is $\lambda(E)$ a.s., and in general not $\mathbf{1}_E$ a.s. □

The eminent geometer H. Coxeter pinpoints such delusion due to models in stating: 'When using models, it is desirable to have two rather than one, so as to avoid the temptation to give either of them undue prominence. Our ... reasoning should all depend on the axioms. The models, having served their purpose of establishing relative consistency, are no more essential than diagrams' (see Section 16.2 in [4]). Coxeter has the circle and halfplane models of hyperbolic geometry in mind, but certainly his remark applies more generally.

An axiomatic concept of random variables should avoid the asserted flaws. The reader may judge our approach from this viewpoint. This paper owes a lot to discussions with Hermann Dinges, who put forward related ideas already in [5] (jointly with H. Rost). For further discussion we refer to H. Dinges [6] and D. Mumford [8].

The paper is organized as follows. In Section 1.2 we have a look at those properties of events and random variables independent of a measure-theoretic representation (this section may be skipped). In Section 1.3 we discuss the class of measurable spaces which are suitable to serve as target spaces of random variables. Section 1.4 contains the axioms for general systems of random variables, which we call stochastic ensembles. In Section 1.5 we derive events and deduce their properties from these axioms. In Section 1.6 we discuss equality and a.s. equality of random variables. In Section 1.7 we address convergence of random variables in order to exemplify how to work within our framework of axioms.

1.2 Events and random variables – an outline

Random variables and events rely on each other. Random variables can be examined from the perspective of events, and vice versa. In this introductory section we describe this interplay in an non-systematic manner and detached from the measure-theoretic model.

The field $\mathcal{E}$ of events is a σ-complete Boolean lattice. In particular:

- Each event E possesses a complementary event E^c.
- For any finite or infinite sequence $E_1, E_2, \dots$ of events there exists its union $\bigcup_n E_n$ and its intersection $\bigcap_n E_n$.
- There are the sure and the impossible events E_{sure} and E_{imp}.

Also $E_1 \subset E_2$, iff $E_1 \cap E_2 = E_1$. Since events are no longer considered as subsets of some space, unions and intersections have to be interpreted here in the lattice-theoretic manner.

A random variable X first of all has a *target space* S equipped with a σ-field $\mathcal{B}$. Intuitively S is the set, where X may take its values. Collections of random variables obey the following simple rules:

- To each random variable X with target space S and to each measurable $\varphi : S \to S'$ a random variable with target space S' is uniquely associated, denoted by $\varphi(X)$.
- To each sequence $X_1, X_2, \dots$ of random variables with target spaces $S_1, S_2, \dots$ a random variable with target space $S_1 \times S_2 \times \cdots$ equipped with the product-σ-field is uniquely associated, denoted by $(X_1, X_2, \dots)$.

The corresponding calculation rules are obvious; we will come back to them. We point out that not every measurable space is suitable as a target space – a minimal condition will be given in the next section. Uncountable products $\otimes_{i\in I}(S_i,\mathcal{B}_i)$ of measurable spaces are in general not admissible target spaces. This conforms to the fact that in probability an uncountable family of random variables $(X_i)_{i\in I}$ is at most provisionally considered as a single random variable with values in the product space, before proceeding to a better suited target space.

The connection between random variables and events is established by the remark that to any random variable X and to any measurable subset B of its target space S an event

$$\{X \in B\}$$

is uniquely associated. The events $\{X \in B\}$ uniquely determine X, where B runs through the measurable subsets of S. The calculation rules are

$$\Big\{X \in \bigcup_n B_n\Big\} \;=\; \bigcup_n \{X \in B_n\},\;\; \Big\{X \in \bigcap_n B_n\Big\} \;=\; \bigcap_n \{X \in B_n\},$$
$$\{X \in B^c\} \;=\; \{X \in B\}^c,\;\; \{X \in S\} \;=\; E_{\text{sure}},\quad \{X \in \emptyset\} \;=\; E_{\text{imp}},$$

where $B, B_1, B_2, \ldots$ are measurable subsets of the target space of X. If these properties hold, the mapping $B \mapsto \{X \in B\}$ is called a σ-*homomorphism*. Moreover

$$\{\varphi(X) \in B'\} \;=\; \{X \in B\},\quad \text{where } B = \varphi^{-1}(B')$$
$$\{(X_1, X_2, \ldots) \in B_1 \times B_2 \times \cdots\} \;=\; \bigcap_n \{X_n \in B_n\}.$$

From the perspective of events the connection to random variables is as follows: For any event E there is a random variable I_E with values in $\{0, 1\}$, the indicator variable of E, fulfilling

$$\{I_E = 1\} \;=\; E,\; \{I_E = 0\} \;=\; E^c.$$

For any infinite sequence $E_1, E_2, \ldots$ of disjoint events there is a random variable $N \;=\; \min\{n : E_n \text{ occurs}\}$ with values in $\{1, 2, \ldots, \infty\}$ such that

$$\{N = n\} \;=\; E_n,\; \{N = \infty\} \;=\; \bigcap_n E_n^c.$$

For any infinite sequence $E_1, E_2, \ldots$ of events (disjoint or not) there is a random variable X and measurable subsets $B_1, B_2, \dots$ of its target space such that

$$\{X \in B_n\} \;=\; E_n$$

for all n (see Section 1.5).

This is roughly all that mathematically can be stated about events and random variables. A systematic treatment requires an axiomatic approach. There are two possibilities, namely to start from events or to start from random variables.

Either the starting point is the field of events, which is assumed to be a σ-complete Boolean lattice $\mathcal{E}$. Then a random variable X with target space $(S, \mathcal{B})$ is nothing else but a σ-homomorphism from $\mathcal{B}$ to $\mathcal{E}$. It is convenient to denote it as $B \mapsto \{X \in B\}$ again. In this approach some technical efforts are required to show that any sequence $X_1, X_2, \dots$ of random variables may be combined to a single random variable $(X_1, X_2, \dots)$.

Starting from random variables instead is closer to intuition in our view. Also it circumvents the technical efforts just mentioned. This approach will be put forward in the following sections.

1.3 Spaces with denumerable separation

Not every measurable space qualifies as a possible target space. We require that there exists a denumerable system of measurable sets separating points.

Definition *A measurable space* $(S, \mathcal{B})$ *is called a* measurable space with denumerable separation (mSdS)*, if there is a denumerable* $\mathcal{C} \subset \mathcal{B}$ *such that for any pair* $x \neq y$ *of elements in* S *there is a* $C \in \mathcal{C}$ *such that* $x \in C$ *and* $y \notin C$*.*

Examples

(i) Any separable metric space together with its Borel-σ-algebra is an mSdS. This includes the case of denumerable S and in fact any relevant target space of random variables considered in probability.

(ii) If $(S_1, \mathcal{B}_1), (S_2, \mathcal{B}_2), \dots$ is a sequence of mSdS, then also the product space $\otimes_n (S_n, \mathcal{B}_n)$ is an mSdS. Indeed, if $\mathcal{C}_1, \mathcal{C}_2, \dots$ are the separating systems, then

$$\mathcal{C} \;:=\; \bigcup_n \{S_1 \times \cdots \times S_{n-1} \times C_n \times S_{n+1} \times \cdots \;:\; C_n \in \mathcal{C}_n\}$$

is denumerable and separating in the product space.

(iii) An uncountable product of measurable spaces is not an mSdS (up to trivial cases). The reason is that this product-σ-field does not contain one-point sets (see below). □

An mSdS $(S, \mathcal{B})$ has two important properties. Firstly one point subsets $\{x\}$ are measurable, since

$$\{x\} \;=\; \bigcap_{C \in \mathcal{C}\,,\, x \in C} C$$

for all $x \in S$. Secondly the 'diagonal'

$$D \;:=\; \{(x, y) \in S^2 \;:\; x = y\}$$

is measurable in the product space $(S^2, \mathcal{B}^2)$, since

$$D \;=\; \bigcap_{C \in \mathcal{C}} C \times C \cup C^c \times C^c. \tag{1.1}$$

These properties are crucial for target spaces of random variables. Remarkably the second one is characteristic for mSdS.

Proposition 1.1 *A measurable space $(S, \mathcal{B})$ is an mSdS, if and only if $D \in \mathcal{B}^2$.*

Proof. It remains to prove that $D \in \mathcal{B}^2$ implies the existence of a denumerable separating system $\mathcal{C}$. Let

$$\mathcal{F} \;:=\; \bigcup_{\mathcal{C}} \sigma(\mathcal{C}) \otimes \sigma(\mathcal{C}),$$

where $\sigma(\mathcal{C})$ is the σ-field generated by $\mathcal{C}$ and the union is taken over all denumerable $\mathcal{C} \subset \mathcal{B}$. $\mathcal{F}$ is a sub-σ-field of $\mathcal{B} \otimes \mathcal{B}$ containing all $B_1 \times B_2$ with $B_1, B_2 \in \mathcal{B}$, thus

$$\mathcal{B} \otimes \mathcal{B} \;=\; \bigcup_{\mathcal{C}} \sigma(\mathcal{C}) \otimes \sigma(\mathcal{C}).$$

By assumption it follows that $D \in \sigma(\mathcal{C}) \otimes \sigma(\mathcal{C})$ for some denumerable $\mathcal{C} \subset \mathcal{B}$. We show that $\mathcal{C} \cup \{C^c \;:\; C \in \mathcal{C}\}$ is a separating system. Let $x, y \in S, x \neq y$. Then D does not belong to the σ-field

$$\mathcal{G} \;:=\; \{B \in \sigma(\mathcal{C}) \otimes \sigma(\mathcal{C}) \;:\; \{(x, x), (y, x)\} \subset B \text{ or } \{(x, x), (y, x)\} \subset B^c\}.$$

It follows that $\mathcal{G} \neq \sigma(\mathcal{C}) \otimes \sigma(\mathcal{C})$, thus there are $B_1, B_2 \in \sigma(\mathcal{C})$ such that $B_1 \times B_2 \notin \mathcal{G}$. Thus B_1 contains x or y, but not both, and consequently is not an element of the σ-field

$$\mathcal{H} \ := \ \{B \in \sigma(\mathcal{C}) : \{x, y\} \subset B \text{ or } \{x, y\} \subset B^c\}.$$

Thus $\mathcal{H} \neq \sigma(\mathcal{C})$, therefore there is a $C \in \mathcal{C}$ such that x or y are elements of C, but not both. This finishes the proof. □

The property of denumerable separation proves useful also in the study of σ-homomorphisms between measurable spaces.

Proposition 1.2 *Let $(S, \mathcal{B})$ be a mSdS, let $(\Omega, \mathcal{E})$ be a measurable space and let $h : \mathcal{B} \to \mathcal{E}$ be a σ-homomorphism. Then there is a unique measurable function $\eta : \Omega \to S$ such that $\eta^{-1}(B) = h(B)$ for all $B \in \mathcal{B}$.*

Proof. First we prove that h is not only a σ-homomorphism but a τ-homomorphism, that is

$$h(B) \ = \ \bigcup_{x \in B} h(\{x\}) \tag{1.2}$$

for all $B \in \mathcal{B}$. For the proof let $\{C_1, C_2, \ldots\}$ be a separating system of $\mathcal{B}$. Because h is a σ-homomorphism,

$$h(B) \ = \ \bigcap_n h(B \cap C_n) \cup h(B \cap C_n^c).$$

Since we consider sets here, this expression may be further transformed by general distributivity: Denoting $C_n^+ := C_n$ and $C_n^- := C_n^c$

$$h(B) \ = \ \bigcup_\chi \bigcap_n h(B \cap C_n^{\chi(n)}) \ = \ \bigcup_\chi h\Big(B \cap \bigcap_n C_n^{\chi(n)}\Big),$$

where the union is taken over all mappings $\chi : \mathbb{N} \to \{+, -\}$. Since $\{C_1, C_2, \ldots\}$ is a separating system, $\bigcap_n C_n^{\chi(n)}$ contains at most one element, and for each $x \in S$ there is exactly one χ such that $\{x\} = \bigcap_n C_n^{\chi(n)}$. Therefore (1.2) follows.

In particular $\Omega = \bigcup_{x \in S} h(\{x\})$. This enables us to define η by means of

$$\eta(\omega) = x \quad :\Leftrightarrow \quad \omega \in h(\{x\}),$$

that is $\eta^{-1}(\{x\}) = h(\{x\})$. From (1.2)

$$h(B) \ = \ \bigcup_{x \in B} \eta^{-1}(\{x\}) \ = \ \eta^{-1}(B).$$