ELSEVIER

Theoretical Computer Science 253 (2001) 287–309

Theoretical
Computer Science

www.elsevier.com/locate/tcs

Measure and probability for concurrency theorists

Prakash Panangaden[1]

School of Computer Science, McGill University, Montreal, Quebec, Canada H3A2A7

Abstract

The concept of conditional probability plays a fundamental role in probability theory. Just as implication plays a fundamental role in logical reasoning so conditional probability plays an analogous role in probabilistic reasoning. The purpose of this article is to give an expository account of conditional probability, in particular conditional probability distributions on continuous spaces. This necessitates background in measure theory which is also reviewed. This is intended for an audience of concurrency theorists interested in using these ideas for probabilistic semantics.
© 2001 Elsevier Science B.V. All rights reserved.

Keywords: Measure theory; Probability theory; Conditional probability; Concurrency; Probabilistic semantics

1. Introduction

The concept of conditional probability plays a fundamental role in probability theory. Just as implication plays a fundamental role in logical reasoning so conditional probability plays an analogous role in probabilistic reasoning. Indeed many probabilistic "logics" overlook this and try to formulate ways of probabilistic reasoning that circumvent the apparatus provided by modern probability theory. In the opinion of the author, modern probability is already a marvelous tool well adapted to the needs of computer scientists. In algorithmics, combinatorics and complexity theory there have already developed [32, 40] deep and rich connections with probability theory.

In areas like concurrency theory, verification and semantics it is fair to say that the connections are much more recent and researchers are only beginning to delve into measure theory and serious probability theory. It is hoped that the present article will bring one concept – that of conditional probability – into the standard lexicon of

[1] Research Supported by NSERC, Canada.
E-mail address: prakash@cs.mcgill.ca (P. Panangaden).

PII: S0304-3975(00)00096-7

workers in concurrency and encourage them to delve deeper into measure theory and probability theory. I do not claim to be an expert but rather one, who has recently learned some of these ideas and found them fascinating and fruitful [7, 10, 11].

This article is not an introduction to probability theory – there are already many excellent introductions – but rather a survey of the basic ideas with emphasis on the notion of conditional probability. I assume that the reader is already familiar with the basic concepts of discrete probability and with notions like random variable and also basic measure theory.

In the area of semantics, one of the earliest serious investigations is due to Kozen [28, 29]. His work uses measure-theoretic ideas in a serious way and, in particular, uses conditional probability distributions or Markov kernels as a central tool. In his work he gives probabilistic semantics of a language of while loops and also develops a "Stone-type" duality using the idea that measurable functions are "probabilistic predicates". This analogue of "predicate transformer" semantics has been extensively developed by a group at Oxford [35]. In a different vein Gupta, Jagadeesan and Saraswat [18] have developed a modeling language for probabilistic systems based on the concurrent constraint programming paradigm. As soon as one adds recursion to the language [17] one is forced into the realm of continuous spaces and the ideas expounded in the present paper are relevant.

The other main starting point for the use of probability theory in semantics was the work of Saheb-Djahromi [37, 38]. His work combined probability theory and domain theory and was the inspiration for probabilistic powerdomains [23, 22] by Jones and Plotkin. Ultimately this led to the enormously fruitful work of Edalat and others [12–14] on integration on domains.

The work on verification of probabilistic systems has exploded in recent years – it is impossible to attempt a survey here. There are approaches based on automata theory, process algebra equivalences, logics and model checking. A very interesting development has been the use of a probabilistic process algebra as a compositional performance evaluation tool [20]. It is hoped that the contents of the present paper might serve as a useful orientation to all this work.

2. Discrete conditional probability

Conditional probabilities relate probabilistic information with definite information and are the key to probabilistic reasoning. In the discrete case the conditional probability can be defined as follows:

$$P(A \mid B) \overset{\text{def}}{=} \frac{P(A \cap B)}{P(B)}.$$

This should be read as "the probability of A being true *given that* B is true". Of course, this makes sense only if $P(B) \neq 0$. If the probability of B is zero and yet B is asserted then the subsequent reasoning cannot be expected to give meaningful answers – or can it?

We consider some simple example scenarios. The first is taken from Feller [15] and is a standard "puzzle" from an introductory probability courses. Suppose that there is a certain community in which the families all have exactly two children and each birth is equally likely to produce a boy or a girl. A salesman arrives at the door of a randomly selected house and notices a Barbie doll in the front yard. Leaving aside issues of political correctness, let us take this as certain information that one of the children in the house is a girl. What is the probability that the other child is a girl? A common erroneous answer is $\frac{1}{2}$. The reasoning is as follows: "Since the births are independent the fact that one child is a girl does not affect the sex of the other child so, since either outcome is equally likely, the probability is a half." Unfortunately this type of reasoning does not take into account the way information is sampled. It also does not use the simple formula for conditional probability given above. The correct answer is $\frac{1}{3}$ and can be arrived at as follows. There are four types of homes, which may be classified as *gg* – both children are girls – or *bb* or *bg* or *gb* with the evident interpretation. These four types of homes are all equally likely. The partial information that has been obtained – namely that one of the children is a girl – rules out *bb*. Thus by the conditional probability formula we get $\frac{1}{3}$. This example is taken from Feller [15].

This simple example shows that there are pitfalls in using one's intuitions. They tend to be incorrect. Formal probability theory was invented and refined over the years by these – and other much more subtle – examples. Using formalisms which shortcut or simplify the basic ideas too much can lead to errors.

The next example shows how to use conditional probability in a familiar process-algebra setting. A probabilistic process can perform three actions a, b and c. The action a takes 1s and then causes termination, the other two processes take 2 and 3 s, respectively, and then end up returning the process to its initial state. Assume that each action is enabled in the initial state and are chosen at random with equal probability. What is the expected time before termination? Clearly one can solve this using high-school techniques, i.e. by summing the obvious infinite series. However there is a nice trick which uses conditional probability, or, to be more precise, conditional expectation.

Let the expected time to termination be written E and the expected time to termination given that the first action selected was a (resp. b, c) be written E_a (E_b, E_c). Now we have the situation

$$E = \tfrac{1}{3}E_a + \tfrac{1}{3}E_b + \tfrac{1}{3}E_c,$$

but we also have that

$$E_b = 2 + E \quad \text{and} \quad E_c = 3 + E.$$

This equation stem from the fact that when, say b, is selected after 2 s we return to the initial state and we are back to the situation described by E. We also have $E_a = 1$

so we get

$$E = \tfrac{1}{3} + \tfrac{1}{3}(2 + E) + \tfrac{1}{3}(3 + E)$$

or, solving for E, we get $E = 6$. The recursion on the expected value takes the place of the usual infinite-series argument.

3. The need for measure theory

Elementary probability theory can be summed up easily. Imagine that one has a process which makes a single step and can end up in any one of a finite set S of final states each with equal likelihood. Then the probability that the final state lies in a subset A – often called an *event* – is given by $|A|/|S|$ where $|\cdot|$ denotes the size of a finite set. From this simple intuition one can define concepts like the probability of more complex processes which might involve several steps or interaction between different observations.

The typical concepts that one learns: independence, expectation value, and conditional probability are fairly clear – at least in their intuitive conception – in the "discrete" case described above. These concepts suffice to analyze much of the work in probabilistic process algebra. In some sense the relevant concepts are essentially those of Boolean algebra. However, in the continuous case, the same concepts require different mathematics. In some sense one can say that one has to move from Boolean algebras to σ-Boolean algebras. Measure theory evolved – essentially in Kolmogorov's hands – in an attempt to provide rigourous foundations for probability theory. The need for such extensions to high-school probability theory arose from statistical mechanics and the need to explain physical phenomena like Brownian motion.

For researchers interested in systems like process-control systems, telecommunication systems, networks there are very similar phenomena. There is a uncontrolled physical phenomenon, "noise" or "drift" and some controlling software. Understanding how these interact is essential for the design and analysis of such systems.

In order to see how measure theory is forced we will consider a classical example – an infinite sequence of coin tosses. This is paradigmatic of an infinitely repeated operation and will be relevant for any analysis of recursive or indefinite iteration in a probabilistic setting. Even though the basic actions are discrete we are led to measure theory by the infinite repetition. Now if we asked naive questions such as "what is the probability of the sequence $(HT)^{\infty}$" we would get 0 as the answer. From this alone we can conclude very little. Right away we observe a striking difference from the finite case. Knowing all the singleton probabilities does not tell us the probabilities associated with other sets. The singleton sets are no longer the "atomic building blocks" from which everything else can be built. We want to be able to say things like "the probability of getting infinitely many heads is 1" which we certainly cannot conclude by simple counting arguments.

What we need is a notion that allows us to define the probabilities in a suitable limiting fashion. We expect that there are certain sets that we can easily associate probabilities to, and such that we can define the probabilities associated with other sets by operations performed on these basic sets. But what are the reasonable operations? It seems compelling that the operations of the discrete theory should survive – these are finite union, finite intersection and complementation. Thus we expect that we will have a family of sets closed under these operations. We further expect that

$$Pr(A) \wedge Pr(B) = Pr(A \cap B)$$

with similar formulas for disjoint union and complementation. We have seen that we cannot expect a summation formula for *arbitrary* unions but, if we want limits to be computable, we can demand that *countable* unions behave like finite unions. In other words we demand that the family of sets that we are working with be closed under countable union and complement; intersection is, of course, superfluous. We demand that if we have a pairwise disjoint family of sets A_i then

$$Pr\left(\bigcup_i A_i\right) = \sum_i Pr(A_i) \quad \text{and} \quad Pr(A^c) = 1 - P(A).$$

The axioms of probability theory are almost precisely these.

From this can we compute the probability of having infinitely many heads? The probability of having the first toss be a head followed by an infinite sequence of tails is 0. The probability of exactly one head anywhere is again 0, by considering the countable union. The probability of any fixed finite number of heads is 0, again by taking a countable union and the probability of finitely many heads is again 0. Thus the probability of infinitely many heads is 1. Of course not all answers should be 0 and 1. The probability of a head followed by an *arbitrary* sequence should be $\frac{1}{2}$. The sets which look like initial finite sequences followed by an arbitrary sequence are the sets which serve as the basis from which to compute all probabilities.

This raises the natural questions: can we compute probabilities for all the sets this way? It turns out that the answer is "no"! There are sets for which probability or "measure" cannot be sensibly defined. This never happens when the space of outcomes or states is countable but happens in many common uncountable spaces.

The key point to take away from this is that we expect to work with *countable* operations – finite ones are not enough and arbitrary ones are impossible.

4. Basic measure theory

In this section we discuss the axioms for measure theory from an abstract point of view. Intuitively a measure is a notion of "size" that one wishes to attach to sets. This notion is intended to reflect the geometric notion of size coming from examples like area and volume. Measure turns out to be poorly related to set-theoretic conceptions of size. It would be pleasant if we could take all sets to be measurable; unfortunately this is not possible, even for such common spaces as the real numbers **R**. In situations with a countable set of possible states we can indeed take all sets to be measurable and much of the subtleties of measure theory can be dispensed with. However, results and proofs obtained in the discrete case are not a very reliable guide to the continuous case.

We will not give any proofs in detail but refer to the author's notes available on the web [34] or to the standard literature. Particularly the books of Ash [3], Billingsley [6], Rudin [36], Kingman and Taylor [27] and Malliavin [31].

4.1. Measurable spaces

Definition 1. A *measurable space* (X, Σ) is a set X together with a family of subsets of X, called a *σ-field* or *σ-algebra*, satisfying the following axioms:
(1) $\emptyset \in \Sigma$,
(2) $A \in \Sigma$ implies that $A^c \in \Sigma$, and
(3) if $\{A_i \in \Sigma \mid i \in I\}$ is a countable family then $\bigcup_{i \in I} A_i \in \Sigma$.
If we require only finite additivity rather than countable additivity we get a *field.*

Note that, unlike open sets in a topology, measurable sets are closed under complementation and hence under countable intersections as well. This makes a dramatic difference to the properties of measurable functions, compared with continuous functions, as we shall see below. Note also that singletons may or may not belong to a σ-field. In most σ-fields that we are interested in the singletons will be measurable sets.

We develop some of the basic properties of σ-fields.

Proposition 2. *The intersection of an arbitrary collection of σ-fields on a set X is a σ-field on X.*

Corollary 3. *Given any subset $\mathscr{B}$ of $\mathscr{P}(X)$ there is a least σ-field containing $\mathscr{B}$.*

We often refer to the least σ-field containing $\mathscr{B}$ as the σ-field *generated* by $\mathscr{B}$.

Example 4. Given a set X the powerset $\mathscr{P}(X)$ is a σ-field. The set consisting of just X and $\emptyset$ is another σ-field. If X is a countable set and all singletons are measurable the σ-field is $\mathscr{P}(X)$. This is the situation with most discrete situations.

These are extreme examples of course. A more interesting example and a good source of counter-examples is the following.

Example 5. Let X be an uncountable set. The collection of all countable (finite or infinite) sets and cocountable sets (complements of countable sets) forms a σ-field on X. This is the σ-field generated by the singletons.

The next example is of fundamental importance.

Example 6. Given a topological space $(X, \mathscr{T})$ we define $\mathscr{B}(X)$ to be the σ-field generated by the open sets (or, equivalently, by the closed sets). Strictly speaking we should write $\mathscr{B}((X, \mathscr{T}))$ since the σ-field depends on the topology and not just on the set X but it is customary to write as we have done since the topology is usually clear from context. The sets in the σ-field $\mathscr{B}(X)$ are called the *Borel sets*. The most important instance of this is the collection of Borel sets in $\mathbf{R}$. One often says "Borel sets" to refer to this special case.

There are other ways of going from the topology to a σ-field. For example one could work with the continuous functions rather than with the open sets.

The class of σ-fields can be characterized in terms of monotonicity properties; this will be very useful when we discuss integration. We introduce the following convenient notation. If we have a nested family of sets

$$A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \ldots$$

with $\bigcup_n A_n = A$ we write $A_n \uparrow A$. Similarly if

$$A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \ldots$$

with $\bigcap_n A_n = A$ we write $A_n \downarrow A$.

Definition 7. A collection of sets $\mathscr{M}$ is called a *monotone class* if whenever $A_n \uparrow A$ with all $A_n \in \mathscr{M}$ then $A \in \mathscr{M}$ and also if $A_n \downarrow A$ with all $A_n \in \mathscr{M}$ then $A \in \mathscr{M}$.

Clearly any σ-field is a monotone class and, just as for σ-fields, the intersection of monotone classes is a monotone class. Thus we can talk about the monotone class generated by a collection of sets just as we did for σ-fields. Recall that a field is like a σ-field except that we only require finite additivity rather than countable additivity. It is clear that a field that is also a monotone class must be a σ-field.

Proposition 8. *If $\mathscr{F}$ is a field of subsets of X then the monotone class, $\mathscr{M}$, generated by $\mathscr{F}$ is a σ-field on X.*

5. Measurable functions

In analogy with continuous function we define measurable functions in terms of inverse images.

Definition 9. A function f from a σ-field (X, Σ_X) to a σ-field (Y, Σ_Y) is said to be **measurable** if $f^{-1}(B) \in \Sigma_X$ whenever $B \in \Sigma_Y$.

This parallels the definition of continuous function in topology. Traditionally, the phrase "measurable function" is used for a real-valued function but we will use it more generally. Our measurable functions have been sometimes called "measurable transformations". If we consider topological spaces with their Borel σ-field then any continuous function is clearly measurable. However many discontinuous functions are also measurable.

We now discuss a theorem which shows the striking contrast between measure theory and topology. We take (X, Σ) to be a measurable space, (Y, d) to be a metric space with the induced Borel algebra $\mathscr{B}_Y$.

Definition 10. Given a family of functions $\{f_n : X \to Y \mid n \in \mathbb{N}\}$ we say that the family *converges pointwise* to f if $\forall x \in X.\ \lim_{n\to\infty} f_n(x) = f(x)$.

In the next theorem the adjective "measurable" refers to the σ-fields on X (which may be general) and on Y (which is the Borel algebra).

Theorem 11. *If a family of measurable functions $\{f_n : X \to Y \mid n \in \mathbb{N}\}$ converges pointwise to f then f is also measurable.*

6. Measures

Measurable spaces or σ-fields are merely the arenas on which measure theory happens. The key notion of "measure" will now be introduced. Roughly speaking, a measure is an assignment of size to the sets in σ-field. This size is typically a real number but it could be a real number between 1 and 0, a probability measure, or an extended nonnegative real number, i.e. one from $[0, \infty]$, or even a complex number. These theories are all slightly different and play different roles in mathematics. For us the most important case will be probability measure but it is worth seeing what happens when ∞ is admitted as a possible value, this is of particular importance in integration theory. Before we proceed I would like to stress that the symbol ∞ is used in the traditional analysis manner, it is nothing to do with the $\perp$ symbol used in semantics. When we write $x < \infty$ we mean simply that x is finite.

Definition 12. A *measure* (*probability measure*), μ on a measurable space (X, Σ) is a function from Σ (a set function) to $[0, \infty]$ ($[0, 1]$), such that if $\{A_i \mid i \in I\}$ is a countable family of pairwise disjoint sets then

$$\mu\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} \mu(A_i).$$

In particular if I is empty we have

$$\mu(\emptyset) = 0.$$

A set equipped with a σ-field and a measure defined on it is called a *measure space*.

This property is called *countable additivity* or *σ-additivity*. It is possible to develop a theory with just finite additivity but many basic results are counterintuitive.

In the rest of this section we are always talking about a set X equipped with a σ-field Σ and a measure μ. We shall always mean "measurable set" when we just say "set". We use letters like A, B to stand for measurable sets.

Proposition 13 (Monotonicity and Continuity). (1) *If $A \subseteq B$ then $\mu(A) \leqslant \mu(B)$.*

(2) *If $A_i \uparrow A$ then $\lim_{i\to\infty} \mu(A_i) = \mu(A)$.*

(3) *If $A_i \downarrow A$ then $\lim_{i\to\infty} \mu(A_i) = \mu(A)$, if $\mu(A_1)$ is finite.*

The following corollary is immediate.

Corollary 14 (Convexity). *For any countable family of sets B_i we have*

$$\mu\left(\bigcup_i B_i\right) \leqslant \sum_i \mu(B_i).$$

The first example looks natural but is pathological and is an important source of counterexamples.

Example 15. For X an infinite set we define a measure on the powerset of X by setting $\mu(A)$ equal to the number of elements of A if A is finite and ∞ otherwise. This measure is called *counting measure*. Many small variations are possible, such as weighting the points of X differently.

The next example appears artificial but is of central importance.

Example 16. Fix a set X and a point x of X. We define a measure, in fact a probability measure, on the σ-field of all subsets of X as follows. We use the slightly peculiar notation $\delta(x, A)$ to emphasize that x is a parameter in the definition.

$$\delta(x, A) = \begin{cases} 1 \text{ if } x \in A, \\ 0 \text{ if } x \notin A. \end{cases}$$

This measure is called the *Dirac delta measure*. Note that we can fix the set A and view this as the definition of a (measurable) function on X. What we get is the characteristic function of the set A, χ_A.

A set of measure 0 is sometimes called a *negligible* set. A typical example of a negligible set is any countable subset of **R**. A negligible set need not be countable, the Cantor set is an uncountable negligible set. Negligible sets play a very important role in measure theory and one often hears phrases like "almost everywhere" or "almost surely". What they usually mean is that something or other is true except on a negligible set. In fact measure theoretic concepts are usually defined only "almost everywhere". This makes it hard to define, for example, categorical concepts. We cannot just take equivalence classes of functions that agree almost everywhere as the functions. In fact this is a very bad idea since these equivalence classes are not compositional. For example suppose $f, g: X \to Y$ agree almost everywhere. Suppose $h: Z \to X$ is a constant function which lands on one of the points where f and g do not agree. Then $h \circ f$ and $h \circ g$ do not agree anywhere!

To be sure the notion of negligible depends on the measure, so one should be careful in interpreting such phrases. One very annoying feature of negligible sets is that they may contain nonmeasurable subsets, whereas we would certainly like to say that all the subsets of a negligible set are negligible as well. This would be true if we could be certain that all the subsets of a negligible set were measurable but, as we have just observed, this may not be true.

Fortunately this can be fixed by "completing the measure". In order to complete the measure we proceed as follows. Suppose that A and C are two measurable sets and that a measure μ is such that $\mu(A) = \mu(C)$. Then every set B such that $A \subseteq B \subseteq C$ is added to the σ-field and the measure is extended to the new σ-field by assigning $\mu(B) = \mu(A)(= \mu(C))$. Of course, one has to check that this yields a σ-field.

Given a measure space (X, Σ, μ) one often implicitly talks about the completion with respect to μ rather than the given σ-field. It is conventional when talking about the reals to use the phrase "Borel field" or "Borel sets" to refer to the σ-field generated by the open intervals and the phrase "Lebesgue measurable sets" to talk about the sets that arise from the completion process with respect to the standard Lebesgue measure.

A very useful type of theorem constructs a measure on a σ-field by starting with data on a restricted family of sets that generate the σ-field. We state a typical theorem of this kind.

Definition 17. A family $\mathscr{F}$ of subsets of X is called a **semi-ring** if
(1) $\emptyset \in \mathscr{F}$,
(2) $A, B \in \mathscr{F} \Rightarrow A \cap B \in \mathscr{F}$ and,
(3) if $A \subseteq B$ are in $\mathscr{F}$ then there are *finitely* many *pairwise disjoint* subsets $C_1, \ldots, C_k \in \mathscr{F}$ such that $B - A = \bigcup_{i=1}^{k} C_i$.

This is not the form of the definition that one is used to in algebra because of the strange last condition but this is precisely the property that holds for "hyperrectangles" in $\mathbf{R}^n$.

Theorem 18. *Suppose that $\mathscr{F}$ is a semi-ring on X and $\mu: \mathscr{F} \to [0, \infty]$ satisfies*
(1) $\mu(\emptyset) = 0$,
(2) *μ is finitely additive and*
(3) *μ is countably subadditive.*
Then μ extends to a measure on the σ-field generated by $\mathscr{F}$.

The proof of this theorem may be found in a standard text on probability and measure, for example the book by Kingman and Taylor [27] or the one by Billingsley [6] or the one by Ash [3]. The intervals on the reals form a typical example of a semi-ring and the length function satisfies the conditions of the theorem so this extension theorem applied to this situation gives the well-known Lebesgue measure.

7. Integration

In this brief section we introduce the basic definitions of abstract integration theory. When integrating a function, say f, from the reals to the reals, one divides up the domain into "small" intervals $\{I_i \mid i \in \mathscr{I}\}$ (or rectangles in two dimensions) such that the function is varying only slowly in each interval and then computes the limit of the sum

$$\lim \sum_i f(x_i) * \mathit{length}(I_i),$$

where x_i is chosen from I_i and the limit is taken as the intervals get more refined. This is the rough idea behind the Riemann-style integration taught in elementary courses. However it requires the function being integrated to be "well behaved"; i.e. it cannot have too many discontinuities.

The result is that this integral behaved badly with respect to limiting operations. In particular naive interchange of sums and integrals is usually not justified. Lebesgue's breakthrough was to realize that a much better integral – at least from the point of view of convergence properties – could be defined by dividing up the range rather than the domain. It is this approach that we sketch in this section.

We first define the *simple functions*. These are measurable functions whose range is a finite set; thus they serve to define a finite partition of the range space. They will be the functions that we start our definition with and from there we extend to the general measurable functions by a limiting process. What makes all this possible is the fact that *all* measurable functions are limits of sequences of simple functions.

Suppose that we have a simple function from the reals to the reals, say s, whose range is the set $\{a_1, \ldots, a_n\}$. We define $\forall i \in \{1, \ldots, n\}. A_i \overset{\text{def}}{=} s^{-1}(a_i)$. The A_i are measurable sets if s is a measurable function but they could be quite complicated otherwise; far more complicated than the intervals that arise in Riemann integration. The natural definition for the integral of s is

$$\int s \, \mathrm{d}\mu = \sum_i a_i \mu(A_i),$$

where μ is Lebesgue measure. This pleasant picture is complicated by questions of well definedness immediately. What if $\mu(A_1)$ is infinite and $a_1 > 0$? It is reasonable to assign the value ∞ to the integral in this case, but what if, in addition, $\mu(A_2) = \infty$ and $a_2 < 0$? Thus it is entirely possible for a function to be measurable but not have a sensible integral.

Definition 19. We say that a simple function s is *integrable* if whenever a is in the range of s, $a \neq 0 \Rightarrow \mu(s^{-1}(a)) < \infty$.

It is possible to define integrable to mean that the sums arising in the definition of the integral are well-defined but the present definition is the usual one in the analysis literature. We can now make the proposed definition above official.

Definition 20. Suppose that (X, Σ, μ) is a measure space and that $s : X \rightarrow \mathbf{R}$ is an integrable, simple function with range $\{a_1 \ldots, a_n\}$. We say that the *integral* of s over X with respect to the measure μ is $\int_X s\mu = \sum_{i=1}^{n} a_i \mu(s^{-1}(a_i))$.

For the rest of this section we fix a measure space (X, Σ, μ). When we say "real-valued function" we will mean measurable real-valued function defined on X. Suppose f is a real-valued function defined on X. We write $f_+(x) = \max(f(x), 0)$ and $f_-(x) = \max(-f(x), 0)$; clearly both f_+ and f_- are measurable if f is.

Definition 21. Suppose that f is an everywhere nonnegative real-valued function. We say that f is *integrable* if it the everywhere nonnegative simple functions less than f are integrable and their integrals are bounded. If f is integrable we define

$$\int_X f\mu = \bigsqcup \int_X s\mu,$$

where the sup is over all nonnegative simple functions below f. If we have a measurable function g which takes on both positive and negative values we say that g is integrable if both g_+ and g_- are integrable and we set

$$\int_X g\mu = \int_X g_+\mu - \int_X g_-\mu.$$

Example 22. We take as our measure space (X, Σ, δ_x) where δ_x is the Dirac measure concentrated at the point x of X. Let f be any nonnegative real-valued function. We claim that

$$\int_X \mathrm{d}\delta_x = f(x).$$

Note that the simple function $s(x) = f(x)$ and 0 everywhere else is a simple function below f. The integral of s with respect to δ_x is $f(x)$. Furthermore any simple function t below f has the integral $t(x) \leqslant f(x)$. Thus the sup of the integrals of all the simple functions below f is precisely $f(x)$.

The next example is the standard advertisement for the superior generality of the Lebesgue integral.

Example 23. Let $f : [0, 1] \rightarrow \mathbf{R}$ be given by $f(x) = 0$ if x is rational and $f(x) = 1$ if x is irrational. This f is in fact a simple function, in fact it is even the characteristic function of a measurable set. Thus its integral is just the measure of the irrationals between 0 and 1 which is 1.

One has to be careful about how the word "integrable" is used. Its use suggests that a function that is not integrable cannot have a sensible integral assigned to it. The definition is rather conservative and often people would like to say that certain integrals are defined but "divergent". Thus, for example, it is common to say that $\int_{-\infty}^{\infty} 1 \, \mathrm{d}x = \infty$. One uses the phrase "has a definite integral" for the less stringent condition. Thus one says that the function $\lambda x.1$ is not integrable but has definite integral between $-\infty$ and ∞ of ∞.

Example 24. The identity function on the reals is measurable (even continuous) but not integrable. This is a more troublesome example. A so-called pragmatic view of this is that the integral is 0; for example one can argue by symmetry. While this is what physicists and engineers usually say, the correct statement is that this function is not integrable and does not have a definite integral over the given range.

Now we can prove some basic properties of the integral. It is customary to introduce the notation $\int_A f \mu$ for the integral of f restricted to the measurable subset A of X with the induced measure. In the next proposition functions and integrals are always on X and f, g are used for integrable functions.

Proposition 25. (1) *If* $0 \leqslant f \leqslant g$ *then* $\int f \mu \leqslant \int g \mu$.
(2) *If* $0 \leqslant f$ *and* $0 \leqslant c$ *is a constant then* $\int c f \mu = c \int f \mu$.
(3) $\int_A f \mu = \int_X f \chi_A \mu$ *where* χ_A *is the characteristic function of* A.

One of the most important properties of the integral is linearity.

Proposition 26. *If* f *and* g *are measurable functions then*

$$\int (f + g)\mu = \int f \mu + \int g \mu.$$

The promised "superiority" of the Lebesgue integral is exemplified by the following theorem.

Theorem 27 (Monotone Convergence Theorem). *If* $\{f_i \mid i \in I\}$ *is a sequence of measurable functions with* $\forall i. f_i \leqslant f_{i+1}$ *and* $f = sup_{i \in I} f_i$ *then*

$$\int f \, \mathrm{d}\mu = \lim_{i \to \infty} \int f_i \, \mathrm{d}\mu.$$

There are much stronger convergence theorems available – for example, the dominated convergence theorem – but the monotone convergence theorem is already very useful.

We close this section with a recapitulation of Kozen's [29] idea of relating measure theory to logic. Imagine a state space equipped with a σ-field. Normally we view a predicate as a function that assigns "true" or "false" to the states. The basic idea is that a measurable function plays the role of a predicate, instead of assigning true or false it assigns a real number. Now instead of evaluating a function at a state we evaluate it over a distribution of possible states. In other words we consider distributions to be generalized states. Now normally we have $s \models \phi$ where s is a state and ϕ is a formula. In Kozen's view we would instead consider $\int f \mu$ as the pairing between the "formula" (measurable function) f and the "state" (distribution) μ.

8. The Radon–Nikodym Theorem

One often needs some notion of "differentiation" of measures. By differentiation we mean an infinitesimal notion of quotient. The precise motivation we have in mind will become clearer when we get to conditional probability distributions. The Radon–Nikodym theorem serves precisely the role of providing a concept of differentiation.

To arrive at a plausible statement we can proceed as follows. Let us think about measures defined on the reals. Consider the function $F(x)$ defined by the integral $F(x) = \int_{\infty}^{x} f(x)\,\mathrm{d}x$, where $\mathrm{d}x$ refers to Lebesgue measure. If f has just a few finite jumps then F is well behaved. If F has a finite jump f will have a singularity. Now a typical singularity can be thought of a Dirac delta "function", which we know can be rigourously defined as a measure concentrated at a point. Now we might think of measures which assign nonzero weight to a single point as being singular but those which do not should be essentially given by a formula like the above for F. More precisely one might conjecture

Conjecture 28. *If λ is a measure on* $\mathbf{R}$ *that has the property $\lambda(\{x\}) = 0$ for any x, then there is some measurable function f such that*

$$\lambda(B) = \int_B f(x)\,\mathrm{d}x$$

for any measurable set B.

Lebesgue showed that this is false but if the hypothesis is strengthened to $\lambda(B) = 0$ whenever B has Lebesgue measure 0 it is true. The Radon–Nikodym theorem generalizes this to the abstract setting. This is precisely the notion of differentiation we need to make sense of conditional probability.

We study the relation of absolute continuity between measure as this is the key assumption in the Radon–Nikodym theorem. There are actually two closely related concepts.

Definition 29. Two measures, μ, v on a measurable space (X, Σ) are *mutually singular*, written as $\mu \perp v$, if there are disjoint measurable sets A, B with $\mu(X \setminus A) = 0$ and $v(X \setminus B) = 0$.

Definition 30. Suppose that μ and v are measures defined on a measurable space (X,Σ). We say that v is *absolutely continuous* with respect to μ, written as $v \ll \mu$, if $\forall A \in \Sigma.\mu(A)=0 \Rightarrow v(A)=0$.

Clearly if f is a measurable function and we define v by $v(A)=\int_A f\mu$ we get a measure such that $v \ll \mu$. The Radon–Nikodym theorem essentially goes in the opposite direction.

Theorem 31 (Radon–Nikodym–Lebesgue). *If μ and v are both σ-finite measures on a measurable space (X,Σ) then*:

(1) *v can be written as $v_a + v_s$ where $v_a \ll \mu$ and $v_s \perp \mu$.*

(2) *there is a non-negative measurable function f such that*

$$\forall A \in \Sigma. \int_A f\mu = v_a(A).$$

If g is another function satisfying the same property as f then the set of points where f and g differ have μ measure 0.

Remark 32. Part (1) of the theorem is usually called the Lebesgue decomposition while part (2) is usually called the Radon–Nikodym theorem. The function f (unique μ-almost everywhere) is often called the Radon–Nikodym derivative and is written as $\mathrm{d}v_a/\mathrm{d}\mu$.

9. Some basic concepts of probability theory

In this section we use the mathematical tools that we have developed to formalize the basic ideas of probability theory in a way suitable for use in situations with continuous state spaces. For example, the concept of conditional probability density will be formalized using the Radon–Nikodym theorem. We will not cover all the basic ideas, for example, the central limit theorem or the theory of large deviations which are both important in any quantitative study of probabilistic processes. The topics that we need will certainly expand as the study of probabilistic semantics grows.

The basic arena for the study of probability is the *probability space*. We give the formal definition immediately.

Definition 33. A *probability space* is a triple $(\Omega,\mathscr{F},P)$ where Ω is a set called the *sample space*, $\mathscr{F}$ is a σ-field on Ω and P is a probability measure on $\mathscr{F}$.

In the discrete case, where Ω is finite or countable, we usually take $\mathscr{F}$ to be the powerset of Ω and we never encounter the subtleties of measure theory.

The intended meaning of a probability space is that one has a (one-step) process operating which ends up in a state or one is carrying out an experiment with the outcomes governed by some probabilistic process. The set Ω is the set of possible states or possible results. A member of $\mathscr{F}$ is called an *event* in probabilistic jargon. The members of $\mathscr{F}$ play the role of *observables*. The idea is that we cannot always tell, with our limited observational powers, exactly which point in Ω occurs; at best we may only be able to identify or specify some larger measurable set. We speak of an event σ *occurring* if the result is in the set σ. In continuous situations it typically happens that the singletons are measurable sets but that P ascribes 0 probability to them. Then we need to work with other measurable sets to say something quantitative.

Random variables are the main objects of interest on probability space. Mathematically they are just measurable functions. Associated with the process described by a probability space are some measurable quantities – the random variables. For example, the probability space may be the state space of a chemical mixture and associated with it are some measurable physical quantities such as temperature and pressure – these are typical random variables. In most textbooks random variables are defined to take values in the real numbers. Conceptually, the theory is affected very little by defining a random variable to take values in any measure space but important quantities, such as the expected value of a random variables, which rely on the arithmetic of the reals, may not make sense.

Definition 34. A *random variable* on a probability space $(\Omega, \mathscr{F}, P)$, is a real-valued, *Borel measurable* function defined on Ω. A random variable which takes on values in the extended reals is called an *extended random variable*. A *random object* is a measurable function on $(\Omega, \mathscr{F})$ which takes values in another measurable space, $(\Xi, \mathscr{G})$.

It is conventional to use uppercase letters like X for random variables rather than letters like f more suggestive of their role as functions.

The most important fact about random variables is that all the probabilistic information is captured in one function from $\mathbf{R}$ to $\mathbf{R}$ called the *distribution function*. Let X be a random variable on a probability space $(\Omega, \mathscr{F}, P)$, fixed for the rest of the paragraph. Now given X we can define a probability measure, P_X on $\mathbf{R}$[2] by the formula

$$P_X(A) = P(\{\omega\colon\ X(\omega) \in A\}),$$

where A is a Borel set. Knowing this measure gives us all the information about the random variable. Now we can define a function $F_X : \mathbf{R} \to [0,1]$ by $F_X(x) = P_X$ by $F_X(x) = P(\{\omega\colon\ X(\omega) \leqslant x\}$, which captures all the information in the measure P_X. This function is increasing, right-continuous and satisfies

$$\lim_{x\to\infty} F_X(x) = 1 \quad \text{and} \quad \lim_{x\to-\infty} F_X(x) = 0.$$

The notion of random variable and of distribution function generalizes in the obvious way to $\mathbf{R}^n$ but of course the computations are more intricate.

[2] We mean on the Borel sets of the reals.

Definition 35. We say a finite set $\{X_1,\ldots,X_n\}$ of random variables defined on $(\Omega,\mathscr{F},P)$ are *independent* if for all Borel sets $B_1,\ldots,B_n$ we have

$$P(\{\omega\colon\ X_1(\omega)\in B_1,\ldots,X_n(\omega)\in B_n\}) = \prod_{i=1}^{n} P(\{\omega\colon\ X_i(\omega)\in B_i\}).$$

This definition does not depend on the random variables taking values in the reals, thus it may be used for arbitrary measurable functions.

The most basic theorem about independence is the fact that the distribution function factorizes.

Theorem 36. *Let $\{X_1\ldots,X_n\}$ be random variables on $(\Omega,\mathscr{F},P)$ and let X be the ($\mathbf{R}^n$-valued) random variable $(X_1,\ldots,X_n)$. Let the distribution functions be F_i and F, respectively. Then*

$$F(x_1,\ldots,x_n) = F_1(x_1)\ldots F_n(x_n).$$

10. Conditional probability in continuous spaces

In the continuous case most of the probabilities are 0, so conditional probabilities must be defined more subtly than in the discrete case. We will present the formal concept in three stages but we begin with an informal argument. Suppose that we have a situation where we wish to define the conditional probability of A given B but B has probability 0 according to our probability measure P. What we do is to consider a family of sets "converging" on B from above. In other words

$$B_1 \supseteq \ldots B_i \supseteq \ldots \text{ with } \bigcap_i B_i = B.$$

Now, we suppose that the conditional probabilities $P(A \mid B_i)$ are well defined. We define the required conditional probability as the "limit" of the $P(A \mid B_i)$ as i tends to infinity.

This formulation is intuitive but difficult to formalize, however the argument hints at the role of a derivative concept. This is just what the Radon–Nikodym theorem provides.

Suppose that we have a probability space (X,Σ,P). Recall that this describes a random mechanism which produces results $x\in X$, distributed according to P. Now fix some measurable set B. Suppose that the observer *knows* that x lies in B. Then, from the point of view of this observer, the probability that x lies in some other measurable set, say A, is $P(A \mid B)$. The observer's knowledge has restricted the sample space but the probability distribution has not changed in any absolute sense. This is the first stage in our presentation and essentially reviews the discrete case.

For the second stage we generalize the above situation slightly. Let $\{B_i \mid i\in I\}$ be a partition of X by sets in Σ, and let Σ' be the σ-field generated by this partition. Now imagine that the observer performs an experiment which allows him to determine to which member of the collection $\{B_i \mid i\in I\}$ the point x belongs. This is the same as determining the members of Σ' to which x belongs. Now after performing the experiment, the conditional probability estimates for x being in A are given by $Q(x) \underset{=}{\text{def}} P(A \mid B_i)$ if $x \in B_i$. We write $P(A \mid \Sigma')(\cdot) : X \to [0,1]$ as the conditional probability of A given Σ' and defined it to be equal to Q if $P(B_i) \neq 0$. If $P(B_i)$ is 0 we give an arbitrary value to $P(A \mid \Sigma')(x)$ for $x \in B_i$. Thus there may be different *versions* of the conditional probability density but they differ on sets of zero probability.

In the final stage we define conditional probability density by viewing Σ' as an arbitrary sub σ-field of Σ. We wish to know the following kind of information. Suppose that we do an experiment and find out in which subset of Σ' a point lies; now we want to know how to estimate probabilities that the point lies in A. We can define a new probability measure V by $V(B) = P(A \cap B)$ for $B \in \Sigma'$. Now using the Radon–Nikodym theorem we can define a conditional probability density function $P(A \mid \Sigma') : X \to [0,1]$ with the properties:

(1) $P(A \mid \Sigma')(\cdot)$ is measurable with respect to Σ' and integrable as well;
(2) for a set B in Σ' we have

$$\int_B P(A \mid \Sigma')(x)\, \mathrm{d}P(x) = P(A \cap B).$$

This density replaces the naive conditional probabilities of the discrete case.

11. Stochastic processes and markov processes

Roughly speaking a *stochastic process* is dynamical probabilistic system. The word "dynamic" is supposed to convey the idea that there is some sort of temporal evolution. The mathematical theory is, however, stated rather more generally.

Definition 37. A *stochastic process* is an indexed family of random variables $X_t : \Omega \to \mathbf{R}$ where $(\Omega, \mathscr{B}, P)$ is a probability space and $t \in T$ is the indexing set.

One usually thinks of T as "time," so it can be viewed as an ordered subset of the reals. In principle, the probability space can vary too but, for simplicity, we shall assume a fixed probability space.

Given this view, we can define the joint distribution $P_{t_1 \ldots t_n}$ of the variables $X_{t_1}, \ldots, X_{t_n}$ as a measure on $\mathbf{R}^n$,

$$P_{t_1 \ldots t_n}(B) = P(\{x \mid (X_{t_1}(x), \ldots, X_{t_n}(x)) \in B\}).$$

This satisfies the obvious consistency requirement – called the *Kolmogorov consistency requirement* – below:

$$P_{t_1 \ldots t_n t_{n+1}}(B \times \mathbf{R}) = P_{t_1 \ldots t_n}(B).$$

This says that the last variable can be integrated out to give the prior distribution. Note that we do not intend the "time" to be discrete here. The second Kolmogorov consistency requirement states that if the variables are permuted then the distributions are altered in the obvious way. The fundamental theorem of the subject says that any family of finite dimensional probability distributions satisfying these two conditions can be realized as a stochastic process.

One can think of the probability distribution at time t, i.e.

$$P_t(A) = P(\{x \mid X_t(x) \in A\})$$

as representing the state of a transition system. The passage from P_t to P_s with $t<s$ may be thought of as a "transition" (discrete) or "evolution" of a system. In general, stochastic processes allow one to consider the possibility of the steps depending on the entire past history of the processes. A very important special class of stochastic processes are *Markov processes*. These are processes in which the transitions depend only on the current state.

More precisely we proceed as follows. We write $P(A_{n+1} \mid x_1,\dots,x_n)$ for the conditional probability that the system is in the set A_{n+1} given that at time t_1 it was at x_1, etc. Now in a Markov process we have

$$P(A_{n+1} \mid x_1,\dots,x_n) = P(A_{n+1} \mid x_n).$$

In other words, only the latest time matters. This definition applies equally well to discrete and continuous-time systems. This is a restriction, but a large number of systems are indeed found to be Markovian. Furthermore many apparently non Markovian processes can be redefined to be Markovian by changing the state space. Thus if the transitions depend on a bounded number of past states the state space can be redefined to make it a Markov process by making the states tuples of former states.

The key feature of a Markov process is that one can think of the transitions as being governed by a transition matrix (discrete state space) or Markov kernel (continuous state space). These Markov kernels are precisely the conditional probability distributions that we have been talking about. The situation can be described as follows. Suppose that we have a probability distribution P_i describing the possible state at time i and P_{i+1} describing the possible state at time $i+1$. Then we can use the Radon–Nikodym theorem to obtain the probability distribution at time $i+1$ *given* that the system was at state s at time i. This would yield a Markov kernel $k_i(s,A)$ from which one could recover the conditional probabilities by integration

$$P(A_{i+1} \mid A_i) = \int_{A_i} k_i(s, A_{i+1})\,\mathrm{d}P_i.$$

Typical examples of Markov processes are probabilistic automata, branching processes, random walks, arrival processes and a multitude of others of practical importance. The literature is vast and of varied levels of accessibility. The standard probability texts contain references to the literature on this topic. Very reasonable starting points are the books by Kingman and Taylor [27] or Billingsley [6].

The following example is taken from our earlier work [7]. It shows a process algebra example of an interacting Markov process. There are three labels $\{a,b,c\}$. Suppose that the state space is **R**. The state gives the pressure of a gaseous mixture in a tank in a chemical plant. The environment can interact by (a) simply measuring the pressure, or (b) it can inject some gas into the tank, or (c) it can pump some gas from the tank. The pressure fluctuates according to some thermodynamic laws depending on the reactions taking place in the tank. With each interaction, the pressure changes according to three different probability density functions, say $f(p_0, p), g(p_0, p)$ and $h(p_0, p)$, respectively, with nontrivial dependence on p_0. We interpret these functions as follows. If the initial state is p_0 and suppose that an a transition occurred then the probability that the final state is between p_1 and p_2 is given by the integral

$$\int_{p_1}^{p_2} f(p_0, p)\,\mathrm{d}p.$$

The measure used for the integration is the ordinary Lesbegue measure. Thus we really have defined a Markov kernel for this system (actually a Markov kernel for each label).

In addition, there are two threshold values p_h and p_l. When the pressure rises above p_h the interaction labelled b is disabled, and when the pressure drops below p_l the interaction labelled c is disabled. It is tempting to model this as a three state system, with the continuous state space partitioned by the threshold values. Unfortunately one cannot assign unique transition probabilities to these sets of states for arbitrary choices of f, g and h; only if very implausible uniformity conditions are obeyed can one do this. These conditions require, for example, that for any pressure value, p say, between p_l and p_h the probability of jumping to a pressure value above p_h is independent of the actual value of p. This is very implausible given the intuition that if the value of p is close to p_h the pressure fluctuations in the system are much more likely to carry the value of the pressure above p_h than if the initial pressure p is far below p_h.

The theory of such labelled Markov processes is the subject of the paper [11], which examines the notion of bisimulation in this context.

12. Conclusions

In this brief survey we have highlighted the role of conditional probability. In order to go further the recommended sources are “Real Analysis and Probability” by Ash [3], “Probability and Measure” by Billingsley [6] and “Probability” by Breiman [8]. The next step is to understand the theory of stochastic processes; these are the objects whose behaviour we are trying to reason about. Conditional probability is a fundamental tool in formulating and reasoning about stochastic processes. The books cited will give an entry point into the vast literature on stochastic processes. An excellent introductory text on Markov processes is “Markov Chains” by Norris [33]. The theory of *interacting* Markov processes is the subject of recent work by Desharnais et al. [11] which builds on the fundamental work (in the discrete setting) of Larsen and Skou [30].

From the point of view of applications there have been a number of very interesting results. The most interesting work, in our opinion, is the work of Hillston [20] on developing a process algebra for performance evaluation. Her work is not comparable to ours, because she works with temporal delay in discrete space Markov chains. The main point of her work is a compositional approach to performance evaluation. In her framework, she address continuous time in the following way. The systems being modelled are described by a probabilistic process algebra called PEPA. The semantics of PEPA are given in terms of labelled transition systems.

There are several papers now on probabilistic analysis, modeling and verification. There are even several papers on probabilistic process algebra analyzing notions of testing and simulation, investigating model checking and exploring various other ideas [39, 41, 24–26, 9, 4, 21]. There are several interesting practical developments which are worthy of attention. In particular telecommunication [2], real-time systems [5] and modeling physical systems [19] are areas where probabilistic systems are very important. It is particularly for the last type of application that we expect that the continuous space formalism developed here will be useful. In a recent paper [17] a programming language with probabilistic choice and recursion was developed. This immediately puts the work in the realm of continuous spaces. The semantics of such systems involved the basic ideas of measure theory that we described in the present work.

From a more logical point of view there is a very interesting paper by Giry [16]. She develops a categorical approach to probability theory based on suggestions of Lawvere. In this framework she shows that there is a category where the objects are measurable spaces and the morphisms are conditional probability distributions and – more importantly – that this category has very striking analogies with the category of relations. In a different expository paper I will give an account of this work. The bottom line is that one can fruitfully think of conditional probability as being the generalization of the notion of relations and of composition as a generalization of relational composition. The analogy is not perfect but is surprisingly close. In a recent paper with Abramsky and Blute we explore just how close [1].

Acknowledgements

This paper is a condensation of part of the notes for a course taught at Aarhus in the Fall of 1996 and at an EATCS Summer school in the Fall of 1997 in Udine. I am very grateful to Glynn Winskel, Mogens Nielsen and BRICS for their hospitality during the year 1996–1997 and to Catuscia Palamidessi for inviting me to the EATCS Summer school. The author of this article has benefited by conversations with Luc Devroye, Josée Desharnais and Devdatt Dubhashi. The author is funded by a grant from NSERC (Canada).

References

[1] S. Abramsky, R. Blute, P. Panangaden, Nuclear and trace ideals in tensor-∗ categories, J. Pure Appl. Algebra (1999), in press, Available from `www-acaps.cs.mcgill.ca/~prakash/`.

[2] R. Alur, L. Jagadeesan, J.J. Kott, J.E. von Olnhausen, Model-checking of real-time systems: a telecommunications application, Proc. 19th Internat. Conf. on Software Engineering, 1997.

[3] R.B. Ash, Real Analysis and Probability, Academic Press, New York, 1972.

[4] C. Baier, M. Kwiatkowska, Domain equations for probabilistic processes, Electronic Notes in Theoretical Computer Science, Vol. 7, July 1997. Extended Abstract.

[5] A. Benveniste, B.C. Levy, E. Fabre, P. Le Guernic, A calculus of stochastic systems for the specification, simulation and hidden state estimation of mixed stochastic/nonstochastic systems, Theoret. Comput. Sci. 152 (2) (1995) 171–217.

[6] P. Billingsley, Probability and Measure, Wiley-Interscience, New York, 1995.

[7] R. Blute, J. Desharnais, A. Edalat, P. Panangaden, in: Bisimulation for labelled Markov processes, Proc. 12th IEEE Symp. On Logic In Computer Science, Warsaw, Poland, 1997.

[8] L. Breiman, Probability, Addison-Wesley, Reading, MA, 1968.

[9] R. Cleaveland, S. Smolka, A. Zwarico, Testing preorders for probabilistic processes, Proc. Internat. Colloquium On Automata Languages And Programming 1992, Lecture Notes In Computer Science, Vol. 623, Springer, Berlin, 1992.

[10] J. Desharnais, A. Edalat, P. Panangaden, A logical characterization of bisimulation for labeled Markov processes, Proc. 13th IEEE Symp. On Logic In Computer Science, Indianapolis, IEEE Press, New York, June 1998, pp. 478–489.

[11] J. Desharnais, A. Edalat, P. Panangaden, Bisimulation for labeled Markov processes, Inform. and Comput. in press.

[12] A. Edalat, Domain theory and integration, Theoret. Comput. Sci. 151 (1995) 163–193.

[13] A. Edalat, Domain theory in stochastic processes, Proc. 10th Annual IEEE Symp. On Logic In Computer Science, IEEE Computer Society Press, Silverspring, MD, 1995.

[14] A. Edalat, Dynamical systems, measures and fractals via domain theory, Inform. and Comput. 120 (1) (1995) 32–48.

[15] W. Feller, An Introduction to Probability Theory and its Applications I, 3rd Edition, Wiley, New York, 1968.

[16] M. Giry, A categorical approach to probability theory, in: B. Banaschewski (Ed.), Categorical Aspects of Topology and Analysis, Lecture Notes In Mathematics, Vol. 915, Springer, Berlin, 1981, pp. 68–85.

[17] V. Gupta, R. Jagadeesan, P. Panangaden, Stochastic processes as concurrent constraint programs, Proc. 26th Proc. Of The Annual ACM Symp. On Principles Of Programming Languages, 1999.

[18] V. Gupta, R. Jagadeesan, V. Saraswat, Probabilistic concurrent constraint programming, in: A. Mazurkiewicz, J. Winkowski (Eds.), Proc. CONCUR97, Lecture Notes In Computer Science, Vol. 1243, Springer, Berlin, 1997.

[19] V. Gupta, V. Saraswat, P. Struss, A model of a photocopier paper path, Proc. 2nd IJCAI Workshop on Engineering Problems for Qualitative Reasoning, 1995.

[20] J. Hillston, A compositional approach to performance modelling, Ph.D. Thesis, University of Edinburgh, 1994. To be published as a Distinguished Dissertation by Cambridge University Press.

[21] M. Huth, M. Kwiatkowska, On probabilistic model checking, Tech. Rep. CSR-96-15, University of Birmingham, 1996. Available from http://www.cs.bham.ac.uk/mzk/.

[22] C. Jones, Probabilistic non-determinism, Ph.D. Thesis, University of Edinburgh, 1990. CST-63-90.

[23] C. Jones, G.D. Plotkin, A probabilistic powerdomain of evaluations, Proc. 4th Annual IEEE Symp. On Logic In Computer Science, 1989, pp. 186–195.

[24] B. Jonsson, K. Larsen, Specification and refinement of probabilistic processes, In Proc. 6th Ann. IEEE Symp. On Logic In Computer Science, 1991.

[25] B. Jonsson, W. Yi, Compositional testing preorders for probabilistic processes, Proc. 10th Ann. IEEE Symp. On Logic In Computer Science, 1995, pp. 431–441.

[26] C.-C. Jou, S.A. Smolka, Equivalences, congruences, and complete axiomatizations for probabilistic processes, in: J.C.M. Baeten, J.W. Klop (Eds.), CONCUR 90 First Internat. Conf. on Concurrency Theory, Lecture Notes In Computer Science, Vol. 458, Springer, Berlin, 1990.

[27] J.F.C. Kingman, S.J. Taylor, Introduction to Measure and Probability, Cambridge University Press, Cambridge, 1966.

[28] D. Kozen, Semantics of probabilistic programs, J. Comput. Systems Sci. 22 (1981) 328–350.

[29] D. Kozen, A probabilistic PDL, J. Comp. Systems Sci. 30 (2) (1985) 162–178.

[30] K.G. Larsen, A. Skou, Bisimulation through probablistic testing, Inform. and Comput. 94 (1991) 1–28.

[31] P. Malliavin, Integration and Probability, Graduate Texts in Mathematics, vol. 157, Springer, Berlin, 1995. French edition appeared in 1993.

[32] R. Motwani, P. Raghavan, Randomized Algorithms, Cambridge University Press, Cambridge, 1995.

[33] J.R. Norris, Markov Chains, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 1997.

[34] P. Panangaden, Stochastic techniques in concurrency, Unpublished Lecture Notes from a course given at BRICS. Available from `www.sable.mcgill.ca/~prakash`, 1997.

[35] Probabilistic systems group, collected reports, Available from `www.comlab.ox.ac.uk` in the directory `/oucl/groups/probs/bibliography.html`.

[36] W. Rudin, Real and Complex Analysis, McGraw-Hill, New York, 1966.

[37] N. Saheb-Djahromi, Probabilistic LCF, in: Mathematical Foundations Of Computer Science, Lecture Notes In Computer Science, Vol. 64, Springer, Berlin, 1978.

[38] N. Saheb-Djahromi, Cpos of measures for nondeterminism, Theoret. Comput. Sci. 12 (1) (1980) 19–37.

[39] R. Segala, N. Lynch, Probabilistic simulations for probabilistic processes, in: B. Jonsson, J. Parrow (Eds.), Proc. CONCUR94, Lecture Notes In Computer Science, Vol. 836, Springer, Berlin, 1994, pp. 481–496.

[40] J. Spencer, N. Alon, P. Erdos, The Probabilistic Method, Wiley, New York, 1992.

[41] R. van Glabbeek, S. Smolka, B. Steffen, C. Tofts, Reactive generative and stratified models for probabilistic processes, Proc. 5th Annual IEEE Symp. On Logic In Computer Science, 1990.