URL: http://www.elsevier.nl/locate/entcs/volume22.html 17 pages

The Category of Markov Kernels

Prakash Panangaden [1]

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

Abstract

Markov kernels are fundamental objects in probability theory. One can define a category based on Markov kernels which has many of the formal properties of the ordinary category of relations. In the present paper we will examine the categorical properties of Markov kernels and stress the analogies and differences with the category of relations. We will show that this category has partially-additive structure and, as such, supports basic constructs like iteration. This allows one to give a probabilistic semantics for a language with while loops in the manner of Kozen. The category in question was originally defined by Giry following suggestions of Lawvere.

1 Introduction

As probability theory is used more and more in computer science it becomes important to understand those aspects of the subject that are relevant for computation. In this paper we focus on some issues that are important for programming language semantics. In particular we examine how Markov kernels can be used to give the semantics of an imperative language with while loops.

In this paper we discuss a categorical construction which allows us to unify some of the ideas in probabilistic semantics. The construction is a small variation of a construction due to Giry [11]. The idea is to look for a monad that imitates some of the properties of the powerset monad and goes back originally to unpublished suggestions of Lawvere [16]. We have, however, modified her definition slightly - as was also done by Kozen - and, in doing so, produced an example of a partially-additive category [17]. This connection allows a simple presentation of Kozen's probabilistic semantics for a language of while loops [14,15]. The material in this paper is mostly not original but lies scattered across the literature. What is new is the realization that the category mentioned above gives an interesting example of the notion of partially additive categories.

[1] Research supported in part by NSERC and FCAR.

The main mathematical object that we use is the Markov kernel, also called a stochastic kernel or a *regular conditional probability distribution.* The fundamental concept underlying this is that of conditional probability. Conditional probability is the basic tool for revising one's estimate of probabilities on the basis of given information. Thus it plays a very similar role to that of the ordinary conditional in logic. This concept becomes particularly subtle in the case of continuous spaces like the reals. Why do we care about continuous spaces? If we have a programming language with iteration or recursion and also a binary probabilistic choice then we get into the realm of continuous spaces immediately. As soon as we put such a binary choice inside a possibly nonterminating loop [14] or a recursion [12] we will find that we have to analyze probabilities over sets of infinite computation sequences. Thus we need to understand conditional probability distributions over such spaces.

In an earlier version of this paper [20] I emphasized the analogy between Markov kernels and binary relations. This will be discussed in section 4 below but will not be the main point of the present paper. The idea of thinking about the category of Markov kernels as a partially additive category is due to Samson Abramsky and arose in email discussions around 1994 in the context of our joint work (with Richard Blute) on nuclear ideals [1].

The next section is a brief recapitulation of the definition of conditional probability distributions especially on continuous state spaces. Section 3 defines the category of interest. Section 4 talks about probability monads and is essentially Giry's [11] construction. Section 5 is the main part of the paper. It discusses the partially additive structure of the category. It includes a review of the definitions. The penultimate section contains an application of the results of the preceding section to a simple language of while loops.

2 Conditional Probability Distributions

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

$$P(A|B) := \frac{P(A \cap B)}{P(B)}.$$

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.

There are simple examples (like the infamous problem of the King's sibling or the even more notorious Monty Hall problem) which show 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.

In the continuous case probabilities are often 0, so conditional probabilities must be defined more subtly than in the discrete case. 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

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

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

This formulation is intuitive but difficult to formalize but hints at the right idea. See my lecture notes [19] for a discussion of conditional probability and how it can be constructed in the continuous case. Of course the probability literature is the place to go for a detailed understanding, we recommend the books of Ash [3], Billingsley [6] and Dudley [9].

We define conditional probability distributions in the following way. Suppose that there is a space X together with a σ-field Σ_X of sets defined on it and similarly (Y, Σ_Y). Assume further that X has a probability measure P defined on it and that $f : X \to Y$ is a measurable function.

Definition 2.1 *A* ***regular conditional probability distribution*** *is a function $h : X \times \Sigma_Y \to [0, 1]$ such that*

(i) *for each fixed $B \in \Sigma_Y$ the function $h(\cdot, B)$ is measurable*

(ii) *for each fixed $x \in X$ $h(x, \cdot)$ defines a probability measure*

(iii) *and the following equation is satisfied:*

$$\forall A \in \Sigma_X . \forall B \in \Sigma_Y . \int_A h(x, B) dP(x) = P((f^{-1}(B)) \cap A).$$

Thus we can think of h as the conditional probability which gives the probability that $f(x)$ is in B given that $x \in A$. In the continuous situation this type of distribution replaces the usual discrete notion. The equation with the integral replaces the discrete equation

$$P(A|B) * P(B) = P(A \cap B).$$

The definition given above is a slightly specialized version of the definition found in Billingsley [6]. It is not easy to guarantee the existence of such distributions. In general, they might not exist but if the underlying space has some additional structure the existence of regular conditional probability distributions can be guaranteed. The basic result given in most textbooks [3,6,9] is that if the space has the topological structure of a complete separable metric space (a Polish space) then the existence of regular conditional probability distributions can be guaranteed. In fact regular conditional probability distributions exist in more general situations, see the book by Hoffman-Jørgensen [13] for a thorough discussion.

Conditional probability distributions arise naturally in the theory of Markov processes and are often called Markov kernels. In a stochastic process one has a probability space $(\Omega, \mathcal{F}, P)$ and a family of random variables X_i defined on it. In general, the index set for the family can be almost anything, but it is usual to consider the index set to be the positive integers and to interpret the indices as representing time steps, or to have them be a subset of the reals and interpret the indices to be instants of continuous time. Let us consider the discrete time case. One can now consider the conditional probability distributions

$$P(X_i \in A_i | X_0 \in A_0, \ldots, X_{i-1} \in A_{i-1}).$$

If these distributions only depend on the last "state" i.e. if

$$P(X_i \in A_i | X_0 \in A_0, \ldots, X_{i-1} \in A_{i-1}) = P(X_i \in A_i | X_{i-1} \in A_{i-1})$$

we have a Markov process. If, in addition, we have

$$\forall n. P(X_{i+n} \in A | X_{i+n-1} \in B) = P(X_i \in A | X_{i-1} \in B)$$

we say that the transitions are time-independent. In this case we can express the behaviour of the system in terms of transition probabilities. This is precisely the conditional probability density for the random variable X_i given the value of X_{i-1}, in other words it is a Markov kernel. One can think of Markov kernels as generalizations of transition probability matrices.

3 The Category SRel

We begin by defining the category which behaves like a stochastic analogue of relations. Essentially its morphisms are slightly modified regular conditional probability distributions.

Definition 3.1 *The precategory* **SRel** *has as objects* (X, Σ_X)*, sets equipped with a* σ*-field. The morphisms are conditional probability densities or Markov kernels. More precisely, a morphism from* (X, Σ_X) *to* (Y, Σ_Y) *is a function* $h : X \times \Sigma_Y \to [0, 1]$ *such that*

(i) $\forall x \in X.\ \lambda B \in \Sigma_Y.\ h(x, B)$ *is a subprobability measure on* Σ_Y*,*

(ii) $\forall B \in \Sigma_Y.\ \lambda x \in X.\ h(x, B)$ *is a measurable function.*

The composition rule is as follows. Suppose that h *is as above and* $k : (Y, \Sigma_Y) \to (Z, \Sigma_Z)$*. Then we define* $k \circ h : (X, \Sigma_X) \to (Z, \Sigma_Z)$ *by the formula* $(k \circ h)(x, C) = \int_Y k(y, C) h(x, dy)$*.*

It is clear that the composition formula really defines an object with the required properties.

This is very close to Giry's definition except that we have a subprobability measure rather than a probability measure. The point of using subprobability measures is that we can view the system as being *partially defined* and we get a nontrivial notion of approximation. With Giry's original definition the partially additive structure trivializes and the treatment of iteration that we have here cannot be done without some different approach.

Henceforth, we write simply X for an object in **SRel** rather than (X, Σ_X) unless we really need to emphasize the σ-field. Before proceeding we prove the

Proposition 3.2 *With composition defined as above* **SRel** *is a category.*

Proof. We use h, k as standing for generic morphisms of type X to Y and Y to Z respectively. We write A, B, C for measurable subsets of X, Y, Z respectively. The identity morphism on X is the Dirac delta "function", $\delta(x, A)$. Recall that the delta "function" is defined by

$$\delta(x, A) = 1 \text{ if } x \in A$$

and is 0 otherwise. With A fixed it is just the characteristic function of A and with x fixed it is the point measure concentrated on x. The fact that it is the identity is simply the equation

$$h(x, B) = \int_X h(x', B)\delta(x, dx')$$

which is a simple computation of a Lebesgue integral.

To verify associativity we use the monotone convergence theorem. Suppose h, k are as above and that $p : Z \to W$ is a morphism and D is a measurable subset of W, we have to show

$$\int_Y [\int_Z p(z, D)k(y, dz)]h(x, dy) = \int_Z p(z, D)[\int_Y k(y, dz)h(x, dy)].$$

The free variables in the above are x and D. Note that this is not just a Fubini type rearrangement of order of integration, the roles of the Markov kernels change. On the left hand side the expression in square brackets produces a measurable function of z, for a fixed D, this measurable function is the integrand for the outer (Y) integration and the measure for this integration is $h(x, dy)$. On the right hand side the expression in square brackets defines a measure on Σ_Z which is used to integrate the measurable function $p(z, D)$ over Z. Now the above equation is just a special instance of the equation

$$\int_Y [\int_Z P(z)k(y, dz)]h(x, dy) = \int_Z P(z)[\int_Y k(y, dz)h(x, dy)]$$

where $P(z)$ is an arbitrary real-valued measurable function on Z. To prove this equation we need only verify it for the very special case of a characteristic function χ_C for some measurable subset C of Z. With $P = \chi_C$ we argue as follows. Recall that whenever we integrate a characteristic function χ_C with respect to any measure ν we get $\nu(C)$. Thus on the left hand side the expression in square brackets becomes $k(y, C)$ and the overall expression is $\int_Y k(y, C)h(x, dy)$. On the right hand side the result is the measure evaluated on C. In other words the expression in square brackets evaluated at C. This is exactly $\int_Y k(y,C)h(x,dy)$. The proof is now routinely completed by first invoking linearity to conclude that the required equation holds for any simple function and then the monotone convergence theorem to conclude that it holds for any measurable function. □

4 Probability Monads

This section is essentially a summary of part of Giry's paper [11] and can be skipped without loss of continuity.

In what sense are we entitled to think of the category **SRel** as a category of relations? It has a peculiarly asymmetric character and lacks some of the key properties associated with a category of relations, in particular it lacks closed structure. There is, however, one way in which it does resemble the category of relations. Recall that the category of relations is the Kleisli category of the powerset functor over the category of sets. It turns out that **SRel** is the Kleisli category of a functor, which resembles the powerset functor, over the category **Meas** of measurable spaces and measurable functions.

We define the (covariant) functor $\Pi : \mathbf{Meas} \to \mathbf{Meas}$ as follows. On objects

$$\Pi(X) := \{\nu | \nu \text{ is a subprobability measure on } X\}.$$

For any $A \in \Sigma_X$ we get a function $p_A : \Pi(X) \to [0,1]$ given by $p_A(\nu) := \nu(A)$. The σ-field structure on $\Pi(X)$ is the least σ-field such that all the p_A maps are measurable. A measurable function $f : X \to Y$ becomes $\Pi(f)(\nu) = \nu \circ f^{-1}$. Checking that Π is a functor is trivial. One can think of $\Pi(X)$ as the collection of probabilistic subsets of X.

We define the appropriate natural transformations $\eta : I \to \Pi$ and $\mu : \Pi^2 \to \Pi$ (Try not to confuse μ with a measure.) as follows:

$$\eta_X(x) = \delta(x,\cdot), \mu_X(\Omega) = \lambda B \in \Sigma_X . \int_{\Pi(X)} p_B \Omega.$$

The definition of η is clear but the definition of μ needs to be deconstructed. First note that Ω is a measure on $\Pi(X)$. Recall that p_B is the measurable function, defined on $\Pi(X)$, which maps a measure ν to $\nu(B)$. The σ-field on $\Pi(X)$ has been defined precisely to make p_B a measurable function. Now the integral $\int_{\Pi(X)} p_B\Omega$ should be meaningful. Of course one has to verify that $\mu_X(\Omega)$ is a subprobability measure. The only minor subtlety is verifying that countable additivity holds. The proof that Π is a monad is done in Giry's paper [11].

Theorem 4.1 (Giry) *The triple* (Π, η, μ) *is a monad on* ***Meas****.*

Now that we have that Π is a monad we can investigate the Kleisli category. A map, $X \to Y$, in this category would be a map $X \to \Pi Y$ in **Meas**. But if we recall that $\Pi(Y)$ is $\Sigma_Y \to [0,1]$ then by uncurrying we can write a Kleisli map as $X \times \Sigma_Y \to [0,1]$, i.e. precisely the type of the morphisms in **SRel**.

Of course one has to verify that one gets exactly the **SRel** morphisms. In other words we have to check that any Markov kernel comes from a map in the Kleisli category; the point to verify is that the map is measurable with respect to the σ-field defined on $\Pi(Y)$.

One point of this observation is that one can use the monad Π in a way analogous to the use of the powerset. For example, deVink and Rutten [10] define probabilistic processes as coalgebras of a functor constructed from Π just as ordinary transition systems can be defined as coalgebras of a functor constructed from the powerset functor.

5 The Additive Structure of SRel

We will examine the properties of the category **SRel**, especially the **partially additive** structure [17].

We begin by establishing that **SRel** has countable coproducts.

Proposition 5.1 *The category* **SRel** *has countable coproducts.*

Proof. Given a countable family $\{(X_i, \Sigma_i)|i \in I\}$ of objects in **SRel** we define (X, Σ) as follows. As a set X is just the disjoint union of the X_i. We write the pair (x, i) for an element of X, where the second member of the pair is a "tag", i.e. an element of I, which indicates which summand the element x is drawn from. The σ-field on X is generated by the measurable sets of each summand. Thus, a generic measurable set in X will be of the form $\uplus_{i\in I} A_i \times \{i\}$, where each A_i is in Σ_i. We will usually just write $\uplus_{i\in I} A_i$ with the manipulation of tags ignored when we are talking about measurable sets.

This object will be "the" coproduct in **SRel**. The injections $\iota_i : X_i \to X$ are $\iota_i(x, \uplus_{k\in I} A_k) = \delta((x, i), \uplus_{k\in I} A_k) = \delta(x, A_i)$. The first equality is by definition and the second is by the defining property of the delta function. Given a family $f_j : X_j \to (Y, \Sigma_Y)$ of **SRel** morphisms we construct the mediating morphism $f : X \to Y$ by $f((x, i), B) = f_i(x, B)$. We check the required commutativity by calculating

$$(f \circ \iota_j)(x_j, B) = \int_X f(x, B)\delta((x_j, j), dx) = \int_{X_j} f_j(x, B)\delta(x_j, dx) = f_j(x_j, B).$$

This is clearly the only way to construct f and satisfies all the required commutativities. □

This is very analogous to the construction in **Rel** but there the coproduct is actually a biproduct (since **Rel** is a self-dual category). The coproduct is **SRel** is not a biproduct. In fact it has a kind of restricted universality property that we will explain after we have discussed the partially additive structure of **SRel**.

It is easy to define a symmetric tensor product. Given (X, Σ_X) and (Y, Σ_Y) we define $(X, \Sigma_X) \otimes (Y, \Sigma_Y)$ as $(X \times Y, \Sigma_X \otimes \Sigma_Y)$ where we mean the tensor product of σ-fields defined as the σ-field generated by sets of the form $A \times B$ - and Cartesian product of the carrier sets. To be brief, we write $X \otimes Y$ for the combination of Cartesian product of the sets and the tensor product of the σ-fields. Given $f : X \to X'$ and $g : Y \to Y'$ we define $f \otimes g : X \otimes Y \to X' \otimes Y'$ by

$$(f \otimes g)((x, y), A' \times B') = f(x, A')g(y, B')$$

where A' and B' are measurable subsets of X' and Y' respectively. Of course this defines the putative measure only on rectangles, but they form a semi-ring and we can extend the measure to all measurable subsets of $X' \times Y'$. It is easy to see that one can define a symmetry for $\otimes$.

In **Rel** we actually have a compact closed category in which the internal hom and the tensor coincide, this is a very special situation. In **SRel**, though the tensor is defined in a manner very analogous to the way that it is defined in **Rel**, we do not even get closed structure. The reader should try to construct what seem at first sight to be the evident evaluation and coevaluation and see what fails. Roughly speaking one gets stuck at the point where one is required to manufacture a canonical measure on a σ-field; the only obvious candidate – the counting measure – miserably fails to satisfy the required equations.

In fact there is a general phenomenon at work here. In situations coming from analysis one finds that one has something that superficially looks like a compact-closed category but in fact turns out to fail that property at some crucial stage. Typically one has no identity morphisms, if one tries to put in the identity morphisms in some way then one loses the algebraic structure that one is looking for. It turns out that these non-categories have a certain structure called a nuclear ideal system; see the recent paper by Abramsky, Blute and Panangaden [1].

5.1 Partially Additive Structure

This subsection is a summary of the definitions of partially additive structure due to Manes and Arbib [17]. Their exposition concentrates on examples like partial functions. The category **SRel** provides a very nice example of their theory. Given $f, g : X \to Y$ in **SRel** we can *sometimes* add them by writing $(f + g)(x, B) = f(x, B) + g(x, B)$. It may happen that the sum exceeds 1 in which case it is not defined[2], but if the sum $f(x, Y) + g(x, Y)$ is bounded by 1 for all x then we get a well-defined subprobability measure and a natural notion of adding morphisms. The only point that requires a mild verification is countable additivity. This is exactly the type of situation axiomatized in the theory of partially additive categories.

Definition 5.2 *A* ***partially additive monoid*** *is a pair $(M, \sum)$ where M is a nonempty set and $\sum$ is a partial function from the collection of all countable*

[2] If we force this sum to be total for example by using $\min(1, \ldots)$ we will lose additivity.

subsets of M to M. We say that $\{x_i | i \in I\}$ is ***summable*** *if $\sum_{i\in I} x_i$ is defined. The following axioms are obeyed.*

(i) ***Partition-Associativity:*** *Suppose that $\{x_i | i \in I\}$ is a countable family and $\{I_j | j \in J\}$ is a countable partition of I. Then $\{x_i | i \in I\}$ is summable iff for every $j \in J$ $\{x_i | i \in I_j\}$ is summable and $\{\sum_{i\in I_j} x_i | j \in J\}$ is summable. In this case we require*

$$\sum_{i\in I} x_i = \sum_{j\in J} \sum_{i\in I_j} x_i.$$

(ii) ***Unary-sum:*** *A singleton family is always summable.*

(iii) ***Limit:*** *If $\{x_i | i \in I\}$ is countable and* every finite subfamily *is summable then the whole family is summable.*

One can think of this as axiomatising an abstract notion of convergence. However the first axiom says, in effect, that we are working with *absolute* convergence and hence rearrangements of any kind are permitted once we know that a sum is defined. Note that one can have some finite sums undefined and some infinite sums defined. The usual notion of complete partial order with sup as sum gives a model of these axioms. A vector space gives a typical nonexample, the limit axiom fails.

We state a simple proposition without proof.

Proposition 5.3 *The sum of the empty family exists, call it 0. It is the identity for $\sum$.*

Though this proposition is easy to prove it has important consequences as we shall see presently.

Definition 5.4 *Let $\mathcal{C}$ be a category. A* ***partially additive structure*** *on $\mathcal{C}$ is a partially additive monoid structure on the homsets of $\mathcal{C}$ such that if $\{f_i : X \to Y | i \in I\}$ is summable, then $\forall W, Z, g : W \to X, h : Y \to Z$ we have that $\{h \circ f_i | i \in I\}$ and $\{f_i \circ g | i \in I\}$ are summable and, furthermore, the equations*

$$h \circ \sum_{i\in I} f_i = \sum_{i\in I} h \circ f_i, (\sum_{i\in I} f_i) \circ g = \sum_{i\in I} f_i \circ g$$

hold.

Since any partially additive monoid has a zero element, a category with partially additive structure will have "zero" morphisms.

Definition 5.5 *A category has* ***zero morphisms*** *if there is a distinguished morphism in every homset – we write 0_{XY} for that of $hom(X, Y)$ – such that $\forall W, X, Y, Z, f : W \to X, g : Z \to Y$ we have $g \circ 0_{WZ} = 0_{XY} \circ f$.*

Proposition 5.6 *If a category has a partially additive structure it has zero morphisms.*

This follows immediately from Proposition 5.3. Note that if a category has a partially additive structure, then every homset is nonempty. This immediately rules out, for example, **Set** as a category that could support a partially additive structure.

5.2 **SRel** *as a Partially Additive Category*

Recall the notion of summable family in **SRel**. A family $\{h_i : X \to Y | i \in I\}$ in the homset $hom(X, Y)$ in **SRel** is summable if

$$\forall x \in X. \sum h_i(x, Y) \leq 1.$$

We define the sum by the evident pointwise formula. With this definition of summability.

Proposition 5.7 *The category* **SRel** *has a partially additive structure.*

Proof. Partition associativity follows immediately from the fact that we are dealing with absolute convergence since all the values are nonnegative. The unary sum axiom is immediate. To see the validity of the limit axiom we proceed as follows. Suppose that $\{h_i : X \to Y | i \in I\}$ in **SRel** is indeed summable. It is easy to see that the sum defines a measurable function when the second argument is fixed. Partition associativity follows immediately from the fact that we are dealing with absolute convergence since all the values are nonnegative. The unary sum axiom is immediate. To see the validity of the limit axiom we proceed as follows. Suppose that $\{h_i : X \to Y | i \in \mathbf{N}\}$ is a countable family and that every finite subfamily is summable. The sums $\sum_{i=1}^{n} h_i(x, Y)$ are bounded by 1 for all x. The sum $\sum_{i=1}^{\infty} h_i(x, Y)$ has to converge, being the limit of a bounded monotone sequence of reals and the sum has to be also bounded by 1. Thus the entire family is summable. One has to check that the sum of morphisms defined this way really gives a measure but the verification of countable additivity is easily done by using the fact that each h_i is countably additive and the sums in question can be rearranged since we have only nonnegative terms. The verification of the two distributivity equations is a routine use of the monotone convergence theorem. □

We now define some morphisms which are of great importance in the theory of partially additive categories. They exist as soon as one has countable coproducts and a family of zero morphisms. It will be apparent in a few moments that they always exist in a category with partially additive structure.

Definition 5.8 *Let $\mathcal{C}$ be a category with countable coproducts and zero morphisms and let $\{X_i | i \in I\}$ be a countable family of objects of $\mathcal{C}$.*

(i) *For any $J \subset I$ we define the* ***quasi-projection*** $PR_J : \coprod_{i \in I} X_i \to \coprod_{j \in J} X_j$ *by*

$$PR_J \circ \iota_i = \begin{cases} \iota_i & i \in J \\ 0 & i \notin J \end{cases}$$

(ii) *We write $I \cdot X$ for the coproduct of $|I|$ copies of X. We define the* ***diagonal-injection*** Δ *by couniversality:*

$$\begin{tikzcd} \coprod(X_i|i\in I) \arrow[r, "\Delta"] & I\cdot\coprod(X_i|i\in I) \\ X_j \arrow[u, "in_j"] \arrow[r, "in_j"] & \coprod(X_i|i\in I) \arrow[u, "in_j"'] \end{tikzcd}$$

(iii) *We have a morphism σ from $I \cdot X$ to X given by:*

$$\begin{tikzcd} I\cdot X \arrow[r, "\sigma"] & X \\ X \arrow[u, "in_j"] \arrow[ur, "id_X"'] & \end{tikzcd}$$

These are all very simple maps to describe explicitly. In **Set** we cannot have a map which behaves like PR_J because we do not have zero morphisms. In **SRel** we have

$$PR_J((x,k),\uplus_{j\in J}) = \begin{cases} \delta(x,A_k) & k\in J \\ 0 & k\notin J. \end{cases}$$

The Δ morphism in **SRel** is

$$\Delta((x,k),\uplus_{i\in I}(\uplus_{j\in I}A^i_j)) = \delta(x,A^k_k).$$

The analogous map in **Set** is $\Delta((x,k)) = ((x,k),k)$. Finally

$$\sigma((x,k),A) = \delta(x,A)$$

in **SRel** while in **Set** we have $\sigma((x,k)) = x$.

We are finally ready to define a partially additive category.

Definition 5.9 *A* ***partially additive category****, $\mathcal{C}$, is a category with countable coproducts and a partially additive structure satisfying the following two axioms.*

(i) ***Compatible sum axiom****: If $\{f_i|i\in I\}$ is a countable set of morphisms in $\mathcal{C}(X,Y)$ and there is a morphism $f: X\to I\cdot Y$ with $PR_i\circ f = f_i$ then $\{f_i|i\in I\}$ is summable.*

(ii) ***Untying axiom****: If $f,g: X\to Y$ are summable then $\iota_1\circ f$ and $\iota_2\circ g$ are summable as morphisms from X to $Y+Y$.*

The first axiom says that if a family of morphisms can be "bundled together as a morphism into the copower" then the family is summable. The reverse direction is an easy consequence of the definition of partially additive structure so this is really an if and only if statement in a partially additive category.

Proposition 5.10 *The category* **SRel** *is a partially additive category.*

Proof. We already know that **SRel** has a partially additive structure and has countable coproducts. Suppose that we have the morphisms f_i and f as described in the compatible sum axiom. We verify that the f_i form a summable family. For fixed $x \in X$ and $B \in \Sigma_Y$ we have

$$\begin{aligned}
\textstyle\sum_{i\in I} f_i(x,B) &= \textstyle\sum_{i\in I}(PR_i \circ f)(x,B)\\
&= \textstyle\sum_{i\in I} \int_{I\cdot Y} PR_i(u,B) f(x,du)\\
&= \textstyle\sum_{i\in I} \int_{Y} \chi_B(u) f(x,du)\\
&\quad \text{(in the previous line the integral is over the } i\text{th summand}\\
&\quad \text{of the disjoint union only)}\\
&= \textstyle\sum_{i\in I} f(x, \iota_i(B)) = f(x, I\cdot B).
\end{aligned}$$

In the last line $I \cdot B$ means the disjoint union of $|I|$ many copies of B. From this calculation and the fact that f is a morphism in **SRel** we see that the sum is indeed defined. To verify untying is a very easy exercise. □

6 Kozen's semantics and duality

In this short section we explain the point of defining partially additive categories. Briefly, the point is to support a notion of iteration. We give a simple presentation of Kozen's probabilistic semantics for a language of while loops using the fact that **SRel** supports iteration simply by being a partially additive category. We first prove that there is an iteration operation whenever we have a partially additive category and then give the semantics. Kozen's first presentation was much more elaborate, but in a later paper he sketched essentially this semantics and described a very nice duality theory which gives a notion of probabilistic predicate transformer.

Theorem 6.1 (Arbib-Manes) *Given $f : X \to X+Y$ in a partially additive category, we can find unique $f_1 : X \to X$ and $f_2 : X \to Y$ such that $f = \iota_1 \circ f_1 + \iota_2 \circ f_2$. Furthermore there is a morphism $\dagger f := \sum_{n=0}^{\infty} f_2 \circ f_1^n : X \to Y$. The morphism $\dagger f$ is called the **iterate** of f.*

More can be said about the iteration construct, in fact Bloom and Esik have written a monumental treatise on this topic and compared various axiomatisations of iteration. Iteration is closely linked to the notion of trace and is also the dual of a fixed-point combinator. We will not discuss the various equational properties of iteration except to note the fixed point property: given any $g : X \to X$ we have $\dagger([g, I_Y] \circ f) = \dagger(f \circ g)$ [3].

[3] The notation $[\cdot,\cdot]$ is used to denote maps constructed by use of couniversality.

6.1 While Loops in a Probabilistic Framework

We define the syntax of programs as follows:

$$S ::= x_i := f(\vec{x}) \mid S_1; S_2 \mid \mathit{if}\ \mathbf{B}\ \mathit{then}\ S_1\ \mathit{else}\ S_2 \mid \mathit{while}\ \mathbf{B}\ \mathit{do}\ S.$$

We use the following conventions. We assume that the program has a fixed set of variables $\vec{x}$, say there are n distinct variables, and that they each take values in some measure space (X, Σ). The space (X^n, Σ^n) is the product space where the vector of variables takes its values. We assume that the function f is a measurable function of type $(X^n, \Sigma^n) \to (X, \Sigma)$ and that $\mathbf{B}$ defines a measurable subset of (X^n, Σ^n). We can thus suppress syntactic details about expressions and boolean expressions. It is easy to extend what follows to cover variables of different sorts and to add random assignment.

We model statements in this programming language as **SRel** morphisms of type $(X^n, \Sigma^n) \to (X^n, \Sigma^n)$. We write $\vec{A}$ for the product $A_1 \times \ldots \times A_n$
Assignment: $x := f(\vec{x})$

$$\llbracket x_i := f(\vec{x}) \rrbracket(\vec{x}, \vec{A}) = \\ \delta(x_1, A_1) \ldots \delta(x_{i-1}, A_{i-1}) \delta(f(\vec{x}), A_i) \delta(x_{i+1}, A_{i+1}) \ldots \delta(x_n, A_n)$$

Sequential Composition: $S_1; S_2$

$$\llbracket S_1; S_2 \rrbracket = \llbracket S_2 \rrbracket \circ \llbracket S_1 \rrbracket$$

where the composition on the right-hand side is the composition in **SRel**.
Conditionals: $\mathit{if}\ \mathbf{B}\ \mathit{then}\ S_1\ \mathit{else}\ S_2$

$$\llbracket \mathit{if}\ \mathbf{B}\ \mathit{then}\ S_1\ \mathit{else}\ S_2 \rrbracket(\vec{x}, \vec{A}) = \delta(\vec{x}, \mathbf{B}) \llbracket S_1 \rrbracket(\vec{x}, \vec{A}) + \delta(\vec{x}, \mathbf{B}^c) \llbracket S_2 \rrbracket(\vec{x}, \vec{A})$$

where $\mathbf{B}^c$ denotes the complement of $\mathbf{B}$.
While Loops: $\mathit{while}\ \mathbf{B}\ \mathit{do}\ S$

$$\llbracket \mathit{while}\ \mathbf{B}\ \mathit{do}\ S \rrbracket = h^\dagger$$

where we are using the $\dagger$ in **SRel** and the morphism $h : (X^n, \Sigma^n) \to (X^n, \Sigma^n) + (X^n, \Sigma^n)$ is given by

$$h(\vec{x}, \vec{A}_1 \uplus \vec{A}_2) = \delta(\vec{x}, \mathbf{B}) \llbracket S \rrbracket(\vec{x}, \vec{A}_1) + \delta(\vec{x}, \mathbf{B}^c) \delta(\vec{x}, \vec{A}_2).$$

The opposite category of **SRel** can be used as the basis for a "predicate transformer" semantics. We sketch the ideas briefly, a detailed exposition would require an excursion into Banach spaces and the topology of these spaces. This part is not self-contained but the reader can still get a good idea of how the construction works without following the details about Banach spaces.

Definition 6.2 *The category* **SPT** *has as objects sets equipped with a σ-field. Given a σ-field we obtain the Banach space of bounded, real-valued, measurable functions defined on X and denoted $\mathcal{F}(X)$. The norm used is the sup norm, i.e. the norm of f is $\sup f$. A morphism $\alpha : X \to Y$ in the category is a linear, continuous function $\alpha : \mathcal{F}(X) \to \mathcal{F}(Y)$.*

Theorem 6.3 (Kozen) $\mathbf{SRel}^{\mathrm{op}} \equiv \mathbf{SPT}$.

Proof. (sketch) Given $h : X \to Y$ in **SRel** we construct $\alpha_h : \mathcal{F}(Y) \to \mathcal{F}(X)$ as follows:

$$\alpha_h = \lambda g \in \mathcal{F}(Y).\ \lambda x \in X. \int_Y g(y)h(x, dy).$$

One has to check that this is linear (clear) and continuous.

Given $\alpha : X \to Y$ in **SPT** we construct $h : Y \to X$ in **SRel** as follows: $h(y, A) = \alpha(\chi_A)(y)$.

We check that these maps are really inverses. Suppose that we start with an **SRel** morphism $h : X \to Y$ and we construct α_h and then go back to **SRel** obtaining a stochastic kernel k. We have $k(x, B) = \alpha_h(\chi_B)(x)$ but by definition of α_h this is $\int_Y \chi_B(y)h(x, dy) = h(x, B)$. Thus we get back our original morphism. The other direction is not quite so trivial. Suppose that we start with an α, construct an h and then α_h. We have to show that for any $f \in \mathcal{F}(X)$ that $\alpha(f) = \alpha_h(f)$. Now we take the special case of a characteristic function χ_A for f. We have then $\alpha_h(\chi_A)(y) = \int_X \chi_A h(y, dx) = h(y, A) = \alpha(\chi_A)(y)$. Thus the required equality holds for characteristic functions. Now we invoke the monotone convergence theorem in the usual way to establish the result for any measurable function. □

In the dual view being adopted here, a bounded, measurable function is the analogue of a predicate on the set of states. An **SRel** morphism is a state transformer while an **SPT** morphism is a predicate transformer. The role of a state is played by a measure on the set of traditional states. The satisfaction relation of ordinary predicates and states is replaced by the integral. Thus the measurable function (predicate) f (ϕ) is "satisfied" by the measure (state) μ (s) written $\int f\mu$ ($s \models \phi$) giving a value in $[0, 1]$ ($\{0, 1\}$).

7 Conclusions

In this note we have given an exposition of (a part of) the work of Giry and have expounded the view that conditional probability distributions play the role of probabilistic relations. This lends some justification to the idea that one can view the Kozen semantics [14] as a state-transformation semantics and its dual [15] as a "predicate-transformer" semantics. The predicate-transformer viewpoint has been pushed to a great extent by the Oxford group [21].

In going to continuous state spaces [5,8] one needs a generalization of the notion of probabilistic transition relation and the concept of conditional probability distribution serves this purpose. In fact, as I have tried to argue, the category **SRel** of Markov kernels has many of the formal properties of the category of relations. It strikes one immediately that the morphisms are directional whereas binary relations have a very simple operation – relational converse – which allows one to reverse the direction of a morphism. In fact this is not a major point. In some sense **SRel** is "one-half" of a category of probabilistic relations. One can define a category using joint distributions on the product spaces as the relations. The Markov kernels then emerge by use of the Radon-Nikodym theorem, one gets a pair of them. This is no different from thinking of a relation as a pair of multifunctions. However, one crucial missing ingredient for a close analogy with relations is the compact closed structure. In joint work with Richard Blute and Samson Abramsky, we tried hard to define a suitable category that had this closed structure. In the end we realized that a new type of closed structure was needed and we axiomatized this under the name of nuclear ideals [1].

The use of subprobability relations and the resulting partially additive structure shows that there are interesting mixtures of ideas from probability theory and domain theory. Of course the theory of partially additive categories is not domain theory but it can be viewed as an alternative to that subject. In a recent paper on probabilistic concurrent constraint programming [12] we developed ideas very much in the same spirit though with a different technical development.

Acknowledgments

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 1996-97 and to Catuscia Palamidessi for inviting me to the EATCS Summer school. The ideas reported here were evolved in the context of my joint work with Samson Abramsky and Richard Blute on nuclear ideals and I am very grateful to them for that collaboration. I have benefited from conversations with Josée Desharnais, Esfan Haghverdi, Phil Scott, Jan Rutten Erik de Vink and especially Dexter Kozen. The author is funded by a grant from NSERC (Canada) and FCAR (Québec).

References

[1] S. Abramsky, R. Blute, and P. Panangaden. Nuclear and trace ideals in tensor-$*$ categories. *Journal of Pure and Applied Algebra*, 1999. in press, available from `www-acaps.cs.mcgill.ca/~prakash/`.

[2] S. Abramsky. Retracing some paths in process algebra. In Montanari and Sassone, editors, *Proceedings of CONCUR 96*, number 1119 in Lecture Notes In Computer Science, pages 1–17. Springer-Verlag, 1996.

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

[4] M. Barr. $*$-*Autonomous Categories.* Number 752 in Lecture Notes in Mathematics. Springer-Verlag, 1980.

[5] R. Blute, J. Desharnais, A. Edalat, and P. Panangaden. Bisimulation for labelled Markov processes. In *Proceedings of the Twelfth IEEE Symposium On Logic In Computer Science, Warsaw, Poland*, 1997.

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

[7] R. Blute. Linear logic, coherence and dinaturality. *Theoretical Computer Science*, 115:3–41, 1993.

[8] J. Desharnais, A. Edalat, and P. Panangaden. A logical characterization of bisimulation for labeled Markov processes. In *proceedings of the 13th IEEE Symposium On Logic In Computer Science, Indianapolis*, pages 478–489. IEEE Press, June 1998.

[9] R. M. Dudley. *Real Analysis and Probability.* Wadsworth and Brookes/Cole, 1989.

[10] E. de Vink and J. J. M. M. Rutten. Bisimulation for probabilistic transition systems: A coalgebraic approach. In *Proceedings of the 24th International Colloquium On Automata Languages And Programming*, 1997.

[11] M. Giry. A categorical approach to probability theory. In B. Banaschewski, editor, *Categorical Aspects of Topology and Analysis*, number 915 in Lecture Notes In Mathematics, pages 68–85. Springer-Verlag, 1981.

[12] V. Gupta, R. Jagadeesan, and P. Panangaden. Stochastic processes as concurrent constraint programs. In *Proceedings of the 26th Proceedings Of The Annual ACM Symposium On Principles Of Programming Languages*, 1999.

[13] J. Hoffman-Jørgenson. *Probability With a View Towards Applications - 2 volumes.* Chapman and Hall, 1994.

[14] D. Kozen. Semantics of probabilistic programs. *Journal of Computer and Systems Sciences*, 22:328–350, 1981.

[15] D. Kozen. A probabilistic PDL. *Journal of Computer and Systems Sciences*, 30(2):162–178, 1985.

[16] F. W. Lawvere. Functorial semantics of algebraic theories. *Proc. Nat. Acad. Sci. U.S.A.*, 50:869–872, 1963.

[17] E. Manes and M. Arbib. *Algebraic Approaches to Program Semantics.* Springer-Verlag, 1986.

[18] Saunders Mac Lane. *Categories for the Working Mathematician*, volume 5 of *Graduate texts in Mathematics.* Springer-Verlag, New York, 1971.

[19] P. Panangaden. Stochastic techniques in concurrency. `www.sable.mcgill.ca/~prakash`, 1997.

[20] P. Panangaden. Probabilistic relations. In C. Baier, M. Huth, M. Kwiatkowska, and M. Ryan, editors, *Preliminary Proceedings of PROBMIV'98*, pages 59–74, 1998.

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