Probabilistic Relations

Prakash Panangaden*
School of Computer Science
McGill University
Montreal, Quebec, Canada

May 17, 1998

Abstract

The notion of binary relation is fundamental in logic. What is the correct analogue of this concept in the probabilistic case? I will argue that the notion of conditional probability distribution (Markov kernel, stochastic kernel) is the correct generalization. One can define a category based on stochastic kernels which has many of the formal properties of the ordinary category of relations. Using this concept I will show how to define iteration in this category and give a simple treatment of Kozen's language of while loops and probabilistic choice. I will use the concept of stochastic relation to introduce some of the ongoing joint work with Edalat and Desharnais on Labeled Markov Processes. In my talk I will assume that people do not know what partially additive categories are but that they do know basic category theory and basic notions like measure and probability. This work is mainly due to Kozen, Giry, Lawvere and others.

1 Introduction

The notion of binary relation and relational composition is fundamental in logical reasoning. Probability theory is intended to be a tool for quantitative reasoning about probabilities. However, when seen from the viewpoint of a computer scientist, probability theory looks like a branch of pure mathematics and one develops "probabilistic logics" for reasoning about probability. In fact probability theory already contains the ingredients of a system of reasoning. In particular the key ingredient - a notion analogous to the conditional in logic - is indeed present; it is nothing but the conditional probability distribution or conditional expectation. This construct is the key tool for *revising one's probability estimates in the presence of new information.*

This is not to deny the importance of work in probabilistic logic. These logics often provide "short-cuts" or "abstractions" or may be paths to algorithmic reasoning. There are, unfortunately, some logics which conceal the subtleties of

In this paper we discuss a categorical construction which allows us to unify some of the ideas in probabilistic semantics. The fundamental idea – to look for a monad which imitates some of the properties of the powerset monad – goes back to Lawvere [Law63] but the detailed development is due to Giry [Gir81]. We have, however, modified her definition

*Research supported in part by NSERC and FCAR.

slightly and, in doing so, produced an example of a partially-additive category [MA86]. This connection allows a simple presentation of Kozen's probabilistic semantics for a language of while loops [Koz81, Koz85]. The material in this paper is not original but lies scattered across the literature. The name **SRel** is an evocation of the analogy with relations and first appears in print in [Abr96].

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

$$P(A|B) \stackrel{\text{def}}{=} \frac{P(A \cap B)}{P(B)}.$$

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.

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. Using logics which are shortcuts or simplifications are dangerous if one does not have a good feel for probabilistic thinking.

In the continuous case most of the probabilities are 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 [Pan97] 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 [Ash72], Billingsley [Bil95] and Dudley [Dud89].

Ultimately we think of conditional probability 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 P defined on it and that $f : X \to Y$ is a measurable function. A regular conditional probability distribution is a function $h : X \times \Sigma_Y \longrightarrow [0,1]$ such that for each fixed $B \in \Sigma_Y$ the function $h(\cdot, B)$ is measurable and for each fixed $x \in X$ $h(x, \cdot)$ defines a probability measure. Furthermore if we compute the integral we get the following equality:

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

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 density replaces the usual discrete notion.

3 The Category SRel

We begin by defining the category which plays the role of stochastic relations. Essentially they are slightly modified conditional probability distributions. The basic existence theory for these objects is beyond the scope of the present paper but we mention in passing that this type of object can be shown to exist in a very general class of spaces called "analytic spaces" [Dud89].

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 stochastic kernels. More precisely, a morphism from (X, Σ_X) to (Y, Σ_Y) is a function $h : X \times \Sigma_Y \longrightarrow [0, 1]$ such that*

1. *$\forall B \in \Sigma_Y . \lambda x \in X . h(x, B)$ is a bounded measurable function,*
2. *$\forall x \in X . \lambda B \in \Sigma_Y . h(x, B)$ is a subprobability measure on Σ_Y.*

The composition rule is as follows. Suppose that h is as above and $k : (Y, \Sigma_Y) \longrightarrow (Z, \Sigma_Z)$. Then we define $k \circ h : (X, \Sigma_X) \longrightarrow (Z, \Sigma_Z)$ 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. 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", $\delta(x, A)$. The fact that it is the identity is simply the equation

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

which we have verified before as our very first computation of a Lebesgue integral.

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

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

T he free variables in the above are x and D. Note that this is not just a Fubini type rearrangement of order of integration, the role of the stochastic 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 wrt 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

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 as we discuss in the next section. 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 **Mes** of measurable spaces and measurable functions.

We define the functor $\Pi : \mathbf{Mes} \longrightarrow \mathbf{Mes}$ as follows. On objects

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

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

We claim that Π is a monad. We define the appropriate natural transformations $\eta : I \longrightarrow \Pi$ and $\mu : \Pi^2 \longrightarrow \Pi$[1] 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 should be 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

[1] Try not to confuse μ with a measure.

$\Pi(X)$, which maps a measure ν to $\nu(B)$. The σ-field on $\Pi(X)$ has been defined precisely to make this 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 subtlety is verifying that countable additivity holds, we leave this as an exercise.

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

Proof. We omit the verification that η_X and μ_X are morphisms. We begin by stating 4 facts that we need in the proof. Let X and Y be measurable spaces and let x, y denote elements of X and Y respectively. Let $f : X \longrightarrow Y$ be measurable, $\nu \in \Pi(X)$, $\Omega \in \Pi^2(X)$ and P, Q be bounded real-valued measurable functions on X and Y respectively.

1. $\int_Y Q\Pi(f)(\nu) = \int_X (Q \circ f)\nu$.
2. $\int_X P\eta_X(x) = P(x)$.
3. Given any real-valued measurable function P we define $\xi_P : \Pi(X) \longrightarrow [0,1]$ by $\forall \nu \in \Pi(X).\xi_P(\nu) = \int_X P\nu$. We claim that ξ_P is measurable.
4. With ξ_P as above we have
$$\int_X P\mu_X(\Omega) = \int_{\Pi(X)} \xi_P \Omega.$$

The first item was our very first example application of the monotone convergence theorem. The second item is an immediate consequence of the properties of the Dirac delta function. We leave the third item as an exercise and verify the fourth.

First note that when P is χ_B then ξ_P is just p_B. Let P be χ_B for some measurable subset B of X. Now we have

$$\int_X P\mu_X(\Omega) = \int_B \mu_X(\Omega) = \mu_X(\Omega)(B) = \int_{\Pi(X)} p_B \Omega == \int_{\Pi(X)} \xi_P \Omega.$$

Thus we have the result for a characteristic function. By linearity it holds for any simple function. Now assume that there is a family of simple functions $s_i \uparrow P$. We have, by the monotone convergence theorem

$$\int_X P\mu_X(\Omega) = \lim_{i \longrightarrow \infty} \int_X s_i \mu_X(\Omega).$$

But we know that this is equal to

$$\lim_{i \longrightarrow \infty} \int_{\Pi(X)} \xi_i \Omega$$

where ξ_i means xi_{s_i}. Now it is easy to see that $\lim_{i \longrightarrow \infty} \xi_i = \xi_P$ so by the monotone convergence theorem we get the result we want.

Now to prove that we have a monad we need to check the naturality of η and μ. The naturality of η is trivial from fact 2. The naturality of μ follows from an easy calculation with fact 1 used at the evident place. The verification of the triangle identity is a good exercise, it just uses the definitions, no subtleties arise. We check the associativity equation explicitly. Let $\Omega' \in \Pi^3(X)$ and $B \in \Sigma_X$. We calculate

$(\mu_X \circ \Pi(\mu_X))(\Omega')(B)$
$\quad = (\mu_X(\Pi(\mu_X)(\Omega')))(B)$
from the definition of μ_X we get
$\quad = \int_{\Pi(X)} p_B \Pi(\mu_X)(\Omega')$
using fact 1 we get
$\quad = \int_{\Pi^2(X)} p_B \circ \mu_X \Omega'$
from the definition of ξ we get
$\quad = \int_{\Pi^2(X)} \xi_{p_B} \Omega'$.

In the other direction we calculate as follows

$(\mu_X \circ \mu_{\Pi(X)})(\Omega')(B)$
$\quad = \mu_X(\mu_{\Pi(X)}(\Omega'))(B)$
from the definition of μ_X
$\quad = \int_{\Pi(X)} p_B \mu_{\Pi(X)}(\Omega')$
using fact 4 we get
$\quad = \int_{\Pi^2(X)} \xi_{p_B} \Omega'$

which is exactly what we got before. ∎

Now that we have that Π is a monad we can investigate the Kleisli category. A map, $X \longrightarrow Y$, in this category would be a map $X \longrightarrow \Pi Y$ in **Mes**. But if we recall that ΠY is $\Sigma_Y \longrightarrow [0,1]$ then by uncurrying we can write a Kleisli map as $X \times \Sigma_Y \longrightarrow [0,1]$, i.e. precisely the type of the morphisms in **SRel**. Of course one has to verify that one gets exactly the **SRel** morphisms. We leave this as an exercise.

5 The Additive Structure of SRel

We will examine the properties of the category **SRel**, especially the **partially additive** structure [MA86].

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 $\iota_i : X_i \longrightarrow X$ are $\iota_i(x, \uplus_{k \in I} A_k) = \delta((x,i), \uplus_{k \in I} A_k) = \delta(x, A_i)$. Given a family $f_j : X_j \longrightarrow (Y, \Sigma_Y)$ of **SRel** morphisms we construct the mediating morphism $f : X \longrightarrow Y$ by $f((x,i), B) = f_i(x, B)$. We check the required commutativity by calculating

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

This is clearly the only way to construct f and satisfy 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). This coproduct 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 earlier and cartesian product of the sets of course. We write $X \otimes Y$ to be brief. Given $f : X \longrightarrow X'$ and $g : Y \longrightarrow Y'$ we define $f \otimes g : X \otimes Y \longrightarrow X' \otimes Y'$ 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 it only on rectangles, but this is 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.

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 exactly the same as 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 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 structire 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 [ABP98].

5.1 Partially Additive Structure

This subsection is a summary of the definitions of partially additive structure due to Manes and Arbib [MA86]. Their exposition concentrates on examples like partial functions. The category **SRel** provides a very nice example of their theory. Given $f, g : X \longrightarrow Y$ 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, 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. 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 which maps* some *countable 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.*

1. ***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*

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

2. ***Unary-sum:*** *A singleton family is always summable.*

3. ***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 $\{f_i : X \longrightarrow Y|i \in I\}$ is summable, then $\forall W, Z, g : W \longrightarrow X, h : Y \longrightarrow Z$ 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 the distinguished member of $hom(X, Y)$, such that $\forall W, X, Y, Z, f : W \longrightarrow X, g : Z \longrightarrow Y$ 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.

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

Proof. A family $\{h_i : X \longrightarrow Y|i \in I\}$ 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. 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 $\{h_i : X \longrightarrow Y | i \in I\}$ in **SRel** is summable, i.e. we assume that

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

We define the sum by the evident pointwise formula. 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 $\{h_i : X \longrightarrow Y | i \in \mathbf{N}\}$ 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 mantra. ∎

We now define some morphisms which are of great importance in the theory of partially additive categories. They exist as soon as one has coproducts and a family of zero morphisms, thus 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 $\mathcal{C}$.*

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

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

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

$$\begin{array}{ccc} \coprod(X_i | i \in I) & \xrightarrow{\Delta} & I \cdot \coprod(X_i | i \in I) \\ \uparrow{\scriptstyle in_j} & & \uparrow{\scriptstyle in_j} \\ X_j & \xrightarrow{in_j} & \coprod(X_i | i \in I) \end{array}$$

3. *We have a morphism σ from $I \cdot X$ to X given by:*

$$\begin{array}{ccc} I \cdot X & \xrightarrow{\sigma} & X \\ \uparrow{\scriptstyle in_j} & \nearrow{\scriptstyle id_X} & \\ X & & \end{array}$$

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.*

1. ***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 $f : X \longrightarrow I \cdot Y$ with $PR_i \circ f = f_i$ then $\{f_i | i \in I\}$ is summable.*
2. ***Untying axiom****: If $f, g : X \longrightarrow Y$ 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

$$\begin{aligned} \textstyle\sum_{i\in I} f_i(x,B) &= \textstyle\sum_{i\in I}(PR_i \circ f)(x,B) \\ &= \textstyle\sum_{i\in I} \int_{I\cdot Y} PR_i(u,B) f(x,du) \\ &= \textstyle\sum_{i\in I} \int_Y \chi_B(u) f(x,du) \end{aligned}$$

(in the previous line the integral is over the ith summand of the disjoint union only)

$$= \textstyle\sum_{i\in I} f(x, \iota_i(B)) = f(x, I \cdot B).$$

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 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 $f : X \longrightarrow X + Y$ in a partially additive category, we can find a unique $f_1 : X \longrightarrow X$ and $f_2 : X \longrightarrow Y$ such that $f = \iota_1 \circ f_1 + \iota_2 \circ f_2$. Furthermore there is a morphism $\dagger f =_{df} \sum_{n=0}^{\infty} f_2 \circ f_1^n : X \longrightarrow Y$. The morphism $\dagger f$ is called the* ***iterate*** *of f.*

Proof. The first assertion is trivial. We have $f_1 = PR_X \circ f$ and $f_2 = PR_Y \circ f$ where the PR maps are the ones associated with the coproduct $X+Y$. The second assertion is about the specific family $\{f_2 \circ f_1^n | n \geq 0\}$ being summable. We first prove by induction on k that the finite families $\{f_2 \circ f_1^n | k \geq n \geq 0\}$ are summable and the result then follows from the limit axiom. The base case is just the unary sum axiom applied to f_2. For the inductive step we claim that if $g : X \longrightarrow Y$ is any morphism then $g \circ f_1$ and f_2 are summable. The induction step then follows immediately from the claim by using $\sum_{n=0}^{k} f_2 \circ f_1^n$ for g. To prove the claim we note

$$\begin{aligned}[g, I_Y] \circ f &= [g, I_Y] \circ (\iota_1 \circ f_1 + \iota_2 \circ f_2) \\ &= [g, I_Y] \circ \iota_1 \circ f_1 + [g, I_Y] \circ \iota_2 \circ f_2 \\ &= g \circ f_1 + f_2\end{aligned}$$

Thus the claim is proved. ∎

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 $g : X \longrightarrow X$ we have $\dagger([g, I_Y] \circ f) = \dagger(f \circ g)$.

6.1 While Loops in a Probabilistic Framework

We define the syntax as follows:

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

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 $(X^n, \Sigma^n) \longrightarrow (X, \Sigma)$ 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 $(X^n, \Sigma^n) \longrightarrow (X^n, \Sigma^n)$. We write $\vec{A}$ for the product $A_1 \times \ldots \times A_n$

Assignment: $x := f(\vec{x})$

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

Sequential Composition: $S_1; S_2$

$$[\![S_1; S_2]\!] = [\![S_2]\!] \circ [\![S_1]\!]$$

where the composition on the right-hand side is the composition in **SRel**.

Conditionals: *if* **B** *then* S_1 *else* S_2

$$[\![\mathit{if}\ \mathbf{B}\ \mathit{then}\ S_1\ \mathit{else}\ S_2]\!](\vec{x}, \vec{A}) = \delta(\vec{x}, \mathbf{B}) [\![S_1]\!](\vec{x}, \vec{A}) + \delta(\vec{x}, \mathbf{B}) [\![S_2]\!](\vec{x}, \vec{A})$$

where **B** denotes the complement of **B**.

While Loops: *while* **B** *do* S

$$[\![\mathit{while}\ \mathbf{B}\ \mathit{do}\ S]\!] = h^{\dagger}$$

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

$$h(\vec{x}, \vec{A_1} \uplus \vec{A_2}) = \delta(\vec{x}, \mathbf{B}) [\![S]\!](\vec{x}, \vec{A_1}) + \delta(\vec{x}, \mathbf{B}^c) \delta(\vec{x}, \vec{A_2}).$$

The opposite category can be used as the basis for a "predicate transformer" semantics. We sketch the ideas briefly, a detailed exposition would require an excusion 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 sup defines the norm. A morphism $\alpha : X \longrightarrow Y$ in the category is a linear, continuous function $\alpha : \mathcal{F}(X) \longrightarrow \mathcal{F}(Y)$.*

Theorem 6.3 (Kozen)

$$\boldsymbol{SRel}^{op} \equiv \mathbf{SPT}.$$

Proof(sketch). Given $h : X \longrightarrow Y$ in **SRel** we construct $\alpha_h : \mathcal{F}(Y) \longrightarrow \mathcal{F}(X)$ 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 $\alpha : X \longrightarrow Y$ in SPT we construct $h : Y \longrightarrow X$ 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 $h : X \longrightarrow Y$ and we construct α_h and then go back to **SRel** obtaining a stochastc 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 mantra and see that it works 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 survey 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 [Koz81] as a state-transformation semantics and its dual [Koz85] as a "predicate-transformer" semantics. The predicate-transformer viewpoint has been pushed to a great extent by the Oxford group [Pro].

In going to continuous state spaces [BDEP97, JDP98] one needs a generalization of the notion of probabilistic transition relation and the concept of conditional probability distribution serves ideally for this purpose. I hope that the exposition of the present paper brings out why this is the correct generalization.

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. I have benefited from conversations with Samson Abramsky, Richard Blute, Luca Cattani, Luc Devroye, Josee Desharnais, Devdatt Dubhashi, Abbas Edalat, Ian Stark and Glynn Winskel. The author is funded by a grant from NSERC (Canada).

A Compact Closed Categories

We assume the reader is familiar with the notion of a *symmetric monoidal* category. A suitable reference is [Lan71]. We now review some of the different closed structures such a category could have.

Definition A.1 A symmetric monoidal category is *closed* or *autonomous* if, for all objects A and B, there is an object $A \multimap B$ and an adjointness relation:

$$Hom(A \otimes B, C) \cong Hom(B, A \multimap C)$$

The unit and counit of this adjunction are the familiar morphisms:

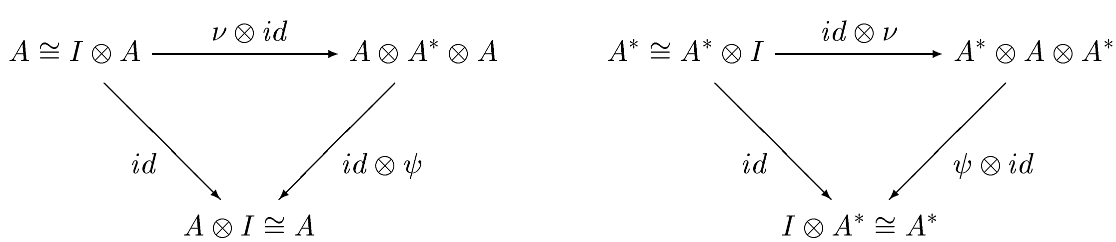
$$ev\colon A \otimes (A \multimap B) \Rightarrow B \qquad coev\colon A \Rightarrow B \multimap (A \otimes B)$$

Examples of autonomous categories include the category of vector spaces and the category of relations.

Definition A.2 A *compact closed category* is a symmetric monoidal category such that for each object A there exists a dual object A^*, and canonical morphisms:

$$\nu\colon I \to A \otimes A^*$$
$$\psi\colon A^* \otimes A \to I$$

such that the usual adjunction triangles commute:



It is easy to see that a compact closed category is indeed closed and that $A \multimap B \cong A^* \otimes B$.

Compact categories could also be defined as $*$-autonomous categories [Bar80, RBS93] with the additional isomorphism $A^* \otimes B^* \cong (A \otimes B)^*$.

We briefly describe the prototypical example, the *category of relations*.

Definition A.3 The *category of relations*, **Rel**, has sets as objects, a morphism from X to Y will be a relation on $X \times Y$, with the usual relational composition.

In what follows, X, Y, Z will denote sets, and x, y, z will denote elements. A binary relation on $X \times Y$ will be denoted $x\mathcal{R}y$. The identity relation will be denoted $\mathcal{ID}$, and is defined as $x\mathcal{ID}x$, for all $x \in X$. Given a relation $\mathcal{R}\colon X \Rightarrow Y$, we let $\overline{\mathcal{R}}\colon Y \Rightarrow X$ denote the converse relation.

We verify that **Rel** is compact. The tensor product $\otimes$ is given by taking the products of sets, and on morphisms, we have:

$$\mathcal{R}\colon X \Rightarrow Y \qquad \mathcal{S}\colon X' \Rightarrow Y'$$

$$(x, x')\mathcal{R} \otimes \mathcal{S}(y, y') \text{ if and only if } x\mathcal{R}y \text{ and } x'\mathcal{S}y'$$

The unit for the tensor is given by any one point set. We define the functor $(\ )^- \colon \mathbf{Rel} \Rightarrow \mathbf{Rel}$ by:

$$X^- = X \quad \mathcal{R}^- = \overline{\mathcal{R}}$$

The relation $\nu \colon I \to X \otimes X^-$ is given by $*\nu(x, x)$ for all $x \in X$ and similarly for ψ.

References

[ABP98] S. Abramsky, R. Blute, and P. Panangaden. Nuclear and trace ideals in tensor-$*$ categories. *Journal of Pure and Applied Algebra*, 1998. in press, Available from `www.daimi.aau.dk/~prakash/` or from `www-acaps.cs.mcgill.ca/~prakash/`.

[Abr96] 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.

[Ash72] R. B. Ash. *Real Analysis and Probability.* Academic Press, 1972.

[Bar80] M. Barr. *$*$-Autonomous Categories.* Number 752 in Lecture Notes in Mathematics. Apringer-Verlag, 1980.

[BDEP97] Richard Blute, Josée Desharnais, Abbas Edalat, and Prakash Panangaden. Bisimulation for labelled markov processes. In *Proceedings of the Twelfth IEEE Symposium On Logic In Computer Science, Warsaw, Poland.*, 1997.

[Bil95] P. Billingsley. *Probability and Measure.* Wiley-Interscience, 1995.

[Dud89] R. M. Dudley. *Real Analysis and Probability.* Wadsworth and Brookes/Cole, 1989.

[Gir81] Michèle 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.

[JDP98] Abbas Edalat Josee Desharnais and Prakash Panangaden. A logical characterization of bisimulation for labeled markov processes. In *proceedings of the 13th IEEE Symposium On Logic In Computer Science, Indianapolis.* IEEE Press, June 1998.

[Koz81] D. Kozen. Semantics of probabilistic programs. *Journal of Computer and Systems Sciences*, 22:328–350, 1981.

[Koz85] D. Kozen. A probabilistic PDL. *Journal of Computer and Systems Sciences*, 30(2):162–178, 1985.

[Lan71] Saunders Mac Lane. *Categories for the Working Mathematician*, volume 5 of *Graduate texts in Mathematics.* Springer-Verlag, New York, 1971.

[Law63] F. W. Lawvere. Functorial semantics of algebraic theories. *Proc. Nat. Acad. Sci. U.S.A.*, 50:869–872, 1963.

[MA86] E. Manes and M. Arbib. *Algebraic Approaches to Program Semantics*. Springer-Verlag, 1986.

[Pan97] Prakash Panangaden. Stochastic techniques in concurrency. `www.sable.mcgill.ca/~prakash`, 1997.

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

[RBS93] P. Panangaden R. Blute and R. Seely. Holomorphic functions as a model of exponential types in linear logic. In S. Brookes et. al., editor, *Proceedings of the Ninth International Conference on Mathematical Foundations of Programming Semantics*, volume 802 of *Lecture Notes In Computer Science*. Springer-Verlag, 1993.