Commutative Monads for Probabilistic Programming Languages (Extended Abstract)

Xiaodong Jia[*‡], Bert Lindenhovius[†], Michael Mislove[‡] and Vladimir Zamdzhiev[§]
[*] School of Mathematics, Hunan University, Changsha, 410082, China
[†] Department of Knowledge-Based Mathematical Systems, Johannes Kepler Universität, Linz, Austria
[‡] Department of Computer Science, Tulane University, New Orleans, LA, USA
[§] Université de Lorraine, CNRS, Inria, LORIA, F 54000 Nancy, France

Abstract

A long-standing open problem in the semantics of programming languages supporting probabilistic choice is to find a commutative monad for probability on the category DCPO. In this paper we present three such monads and a general construction for finding even more. We show how to use these monads to provide a sound and adequate denotational semantics for the Probabilistic FixPoint Calculus (PFPC) – a call-by-value simply-typed lambda calculus with mixed-variance recursive types, term recursion and probabilistic choice. We also show that in the special case of continuous dcpo's, all three monads coincide with the valuations monad of Jones, and we fully characterise the induced Eilenberg-Moore categories by showing that they are all isomorphic to the category of continuous Kegelspitzen of Keimel and Plotkin.

This is an extended abstract for our LICS'21 paper [1]. Preprint: https://arxiv.org/abs/2102.00510.

I. Introduction

Probabilistic methods now are a staple of computation. The initial discovery of randomized algorithms [2] was quickly followed by the definition of Probabilistic Turing machines and related complexity classes [3]. There followed advances in a number of areas, including, e.g., process calculi, probabilistic model checking and verification [4]–[6], right through to the recent development of statistical probabilistic programming languages (cf. [7]–[9]), not to mention the crucial role probability plays in quantum programming languages [10], [11].

Domain theory, a staple of denotational semantics, has struggled to keep up with these advances. Domain theory encompasses two broad classes of objects: *directed complete partial orders (dcpo's)*, based on an order-theoretic view of computation, and the smaller class of *(continuous) domains*, those dcpo's that also come equipped with a notion of approximation. However, adding probabilistic choice to the domain-theoretic approach has been a challenge. The canonical model of (sub)probability measures in domain theory is the family of *valuations* – certain maps from the lattice of open subsets of a dcpo to the unit interval. It is well-known that the valuations form a monad $\mathcal{V}$ on $\mathbf{DCPO}$ (the category of dcpo's and Scott-continuous functions) and on $\mathbf{DOM}$ (the full subcategory of $\mathbf{DCPO}$ consisting of domains) [12], [13].

In fact, the monad $\mathcal{V}$ on $\mathbf{DOM}$ is commutative [13], which is important for two reasons: (1) its commutativity is equivalent to the Fubini Theorem [13], a cornerstone of integration theory and (2) computationally, commutativity of a monad together with adequacy can be used to establish contextual equivalences for effectful programs. However, in order to do so, one typically needs a Cartesian closed category for the semantic model, and $\mathbf{DOM}$ is not closed; in fact, despite repeated attempts, it remains unknown whether there is *any* Cartesian closed category of *domains* on which $\mathcal{V}$ is an endofunctor; this is the well-known *Jung-Tix Problem* [14]. On the other hand, it also is unknown if the monad $\mathcal{V}$ is commutative on the larger Cartesian closed category $\mathbf{DCPO}$. In this paper, we offer a solution to this conundrum.

A. Our Contributions

We use *topological methods* to construct a commutative valuations monad $\mathcal{M}$ on $\mathbf{DCPO}$, as follows: it is straightforward to show the family $\mathcal{S}D$ of *simple valuations* on a dcpo D can be equipped with the structure of a commutative monad, but $\mathcal{S}D$ is not a dcpo, in general. So, we complete $\mathcal{S}D$ by taking the smallest subdcpo $\mathcal{M}D \subseteq \mathcal{V}D$ that contains $\mathcal{S}D$. This defines the object-mapping of a monad $\mathcal{M}$ on $\mathbf{DCPO}$. The unit, multiplication and strength of the monad $\mathcal{M}$ at D are given by the restrictions of the same operations of $\mathcal{V}$ to $\mathcal{M}D$. Topological arguments then imply that $\mathcal{M}$ is a commutative valuations monad on $\mathbf{DCPO}$.

In fact, there are several completions of $\mathcal{S}D$ that give rise to commutative valuations monads on $\mathbf{DCPO}$. These completions are determined by so-called $\mathtt{K}$-categories, introduced by Keimel and Lawson [15]. This observation allows us to define two additional commutative valuations monads, $\mathcal{W}$ and $\mathcal{P}$, on $\mathbf{DCPO}$ simply by specifying their corresponding $\mathtt{K}$-categories. Finally, while we have identified three such $\mathtt{K}$-categories, there likely are more that meet our requirements, each of which would define yet another commutative monad of valuations on $\mathbf{DCPO}$ containing $\mathcal{S}$.

With this background, we now summarise our main results.

Commutative monads: A $\mathtt{K}$-category is a full subcategory of the category $\mathbf{T_0}$ of T_0-spaces satisfying properties that imply it determines a *completion* of each T_0-space among the objects of the $\mathtt{K}$-category. For example, each $\mathtt{K}$-category defines a completion of a poset endowed with its Scott topology, among the objects in the $\mathtt{K}$-category. In particular, each $\mathtt{K}$-category determines a completion of the family $\mathcal{S}D$ when considered as a subset of $\mathcal{V}D$, for each dcpo D.

By specifying an additional constraint on $\mathtt{K}$-categories, we can show the corresponding completions of $\mathcal{S}$ define commutative monads on $\mathbf{DCPO}$. We identify three commutative monads concretely: $\mathcal{M}$, $\mathcal{W}$ and $\mathcal{P}$, corresponding to the $\mathtt{K}$-categories of d-spaces, that of well-filtered spaces and that of sober spaces, respectively. As part of our construction, we also prove the most general Fubini Theorem for dcpo's yet available.

Eilenberg-Moore Algebras: All three of $\mathcal{M}, \mathcal{W}$ and $\mathcal{P}$ restrict to monads on $\mathbf{DOM}$, where they coincide with $\mathcal{V}$. We characterize their Eilenberg-Moore categories over $\mathbf{DOM}$ by showing they are isomorphic to the category of continuous Kegelspitzen and Scott-continuous linear maps [16]; this corrects an error in [13].

On the larger category $\mathbf{DCPO}$, we show the Eilenberg-Moore algebras of our monads $\mathcal{M}, W$ and $\mathcal{P}$ are Kegelspitzen. It is unknown if every Kegelspitze is an $\mathcal{M}$-algebra.

Semantics: We consider the *Probabilistic FixPoint Calculus* (PFPC) – a call-by-value simply-typed lambda calculus with mixed-variance recursive types, term recursion and probabilistic choice. We show that each of the Kleisli categories of our three commutative monads is a sound and computationally adequate model of $\mathtt{PFPC}$. Moreover, we show that adequacy holds in a strong sense, i.e., the interpretation of each term is a (potentially infinite) convex sum of the values it reduces to.

II. Commutative Monads for Probability

A. Domain-theoretic and Topological Preliminaries

A nonempty subset A of a partially ordered set (*poset*) D is *directed* if each pair of elements in A has an upper bound in A. A *directed-complete partial order,* (*dcpo*, for short) is a poset in which every directed subset A has a supremum $\sup A$. For example, the unit interval $[0,1]$ is a dcpo in the usual ordering. A function $f\colon D \to E$ between two (posets) dcpo's is *Scott-continuous* if it is monotone and preserves (existing) suprema of directed subsets. The category $\mathbf{DCPO}$ of dcpo's and Scott-continuous functions is complete, cocomplete and cartesian closed [17]. A *domain* is a dcpo which comes equipped with some suitable additional structure for approximation (details omitted here). Domains and Scott-continuous maps form an important subcategory $\mathbf{DOM}$.

The *Scott topology* σD on a dcpo D consists of the upper subsets $U = {\uparrow}U = \{x \in D \mid (\exists u \in U)\, u \leq x\}$ that are *inaccessible by directed suprema:* i.e., if $A \subseteq D$ is directed and $\sup A \in U$, then $A \cap U \neq \emptyset$. The space $(D, \sigma D)$ is also written as ΣD. Scott-continuous functions between dcpo's D and E are exactly the continuous functions between ΣD and ΣE [18, Proposition II-2.1].

A subset B of a dcpo D is a *sub-dcpo* if every directed subset $A \subseteq B$ satisfies $\sup_D A \in B$. In this case, B is a dcpo in the induced order from D. The *d-topology* on D is the topology whose closed subsets consist of sub-dcpo's of D. Open (closed) sets in the d-topology will be called *d-open (d-closed)*. The *d-closure* of $C \subseteq D$ is the topological closure of C with respect to the d-topology on D, which is the intersection of all sub-dcpo's of D containing C.

The family of open sets of a topological space X, denoted $\mathcal{O}X$, is a complete lattice in the inclusion order. The *specialization order* $\leq_X$ on X is defined as $x \leq_X y$ if and only if x is in the closure of $\{y\}$, for $x, y \in X$. We write ΩX to denote X equipped with the specialization order. It is well-known that X is T_0 if and only if ΩX is a poset. A subset of X is called *saturated* if it is an upper set in ΩX. A space X is called a *d-space* or a *monotone-convergence space* if ΩX is a dcpo and each open set of X is Scott open in ΩX. As an example, ΣD is always a d-space for each dcpo D. The full subcategory of $\mathbf{T_0}$ consisting of d-spaces is denoted by $\mathbf{D}$. There is a functor $\Sigma\colon \mathbf{DCPO} \to \mathbf{D}$ that assigns the space ΣD to each dcpo D, and the map $f\colon \Sigma D \to \Sigma E$ to the Scott-continuous map $f\colon D \to E$. Dually, the functor $\Omega\colon \mathbf{D} \to \mathbf{DCPO}$ assigns ΩX to each d-space X and the map $f\colon \Omega X \to \Omega Y$ to each continuous map $f\colon X \to Y$. In fact, $\Sigma \dashv \Omega$, i.e., Σ is left adjoint to Ω [19].

A T_0 space X is called *sober* if every nonempty closed irreducible subset of X is the closure of some (unique) singleton set, where $A \subseteq X$ is *irreducible* if $A \subseteq B \cup C$ with B and C nonempty closed subsets implies $A \subseteq B$ or $A \subseteq C$. The category of sober spaces and continuous functions is denoted by $\mathbf{SOB}$. Sober spaces are d-spaces, hence $\mathbf{SOB} \subseteq \mathbf{D}$ [15].

B. A Commutative Monad for Probability

To begin, a *subprobability valuation* on a topological space X is a Scott-continuous function $\nu\colon \mathcal{O}X \to [0,1]$ that is strict ($\nu(\emptyset) = 0$), and modular ($\nu(U) + \nu(V) = \nu(U \cup V) + \nu(U \cap V)$). The set of subprobability valuations on X is denoted by $\mathcal{V}X$. The *stochastic order* on $\mathcal{V}X$ is defined pointwise: $\nu_1 \leq \nu_2$ if and only if $\nu_1(U) \leq \nu_2(U)$ for all $U \in \mathcal{O}X$. $\mathcal{V}X$ is a pointed dcpo in the stochastic order, with least element given by the constantly zero valuation $\mathbf{0}_X$ and where the supremum of a directed family $\{\nu_i\}_{i \in I}$ is $\sup_{i \in I} \nu_i \stackrel{\text{def}}{=} \lambda U. \sup_{i \in I} \nu_i(U)$.

The canonical examples of subprobability valuations are the *Dirac valuations* δ_x for $x \in X$, defined by $\delta_x(U) = 1$ if $x \in U$ and $\delta_x(U) = 0$ otherwise. $\mathcal{V}X$ enjoys a convex structure: if $\nu_i \in \mathcal{V}X$ and $r_i \geq 0$, with $\sum_{i=1}^{n} r_i \leq 1$, then the convex sum

$\sum_{i=1}^{n} r_i\nu_i \stackrel{\text{def}}{=} \lambda U. \sum_{i=1}^{n} r_i\nu_i(U)$ also is in $\mathcal{V}X$. The *simple valuations* on D are those of the form $\sum_{i=1}^{n} r_i\delta_{x_i}$, where $x_i \in X$, $r_i > 0, i = 1, \dots, n$ and $\sum_{i=1}^{n} r_i \leq 1$. The set of simple valuations on X is denoted by $\mathcal{S}X$. Clearly, $\mathcal{S}X \subseteq \mathcal{V}X$. Unlike $\mathcal{V}X$, $\mathcal{S}X$ is not directed-complete in the stochastic order in general.

Given $\nu \in \mathcal{V}X$ and $f\colon X \to [0,1]$ continuous, we can define the integral of f against ν by the Choquet formula

$$\int_{x\in X} f(x)d\nu \stackrel{\text{def}}{=} \int_0^1 \nu(f^{-1}((t,1]))dt,$$

where the right side is a Riemann integral of the bounded antitone function $\lambda t.\nu(f^{-1}((t,1]))$. If no confusion occurs, we simply write $\int_{x\in X} f(x)d\nu$ as $\int f d\nu$. Basic properties of this integral can be found in [13].

For a dcpo D, $\mathcal{V}D$ is defined as $\mathcal{V}(D, \sigma D)$. Using Manes' description of monads (Kleisli triples) [20], Jones proved in her PhD thesis [13] that $\mathcal{V}$ is a monad on $\mathbf{DCPO}$:

- The *unit* of $\mathcal{V}$ at D is $\eta_D^{\mathcal{V}}\colon D \to \mathcal{V}D :: x \mapsto \delta_x$;
- The *Kleisli extension* $f^\dagger$ of a Scott-continuous map $f\colon D \to \mathcal{V}E$ maps $\nu \in \mathcal{V}D$ to $f^\dagger(\nu) \in \mathcal{V}E$ by

$$f^\dagger(\nu) \stackrel{\text{def}}{=} \lambda U \in \sigma E. \textstyle\int_{x\in D} f(x)(U)d\nu.$$

Then the *multiplication* $\mu_D^{\mathcal{V}}\colon \mathcal{V}\mathcal{V}D \to \mathcal{V}D$ is given by $\mathrm{id}_{\mathcal{V}D}^\dagger$; it maps $\varpi \in \mathcal{V}\mathcal{V}D$ to $\lambda U \in \sigma D. \int_{\nu\in\mathcal{V}D} \nu(U)d\varpi \in \mathcal{V}D$. Thus, $\mathcal{V}$ defines an endofunctor on $\mathbf{DCPO}$ that sends a dcpo D to $\mathcal{V}D$, and a Scott-continuous map $h\colon D \to E$ to $\mathcal{V}(h) \stackrel{\text{def}}{=} (\eta_E \circ h)^\dagger$; concretely, $\mathcal{V}(h)$ maps $\nu \in \mathcal{V}D$ to $\lambda U \in \sigma E.\nu(h^{-1}(U))$. Jones [13] also showed that $\mathcal{V}$ is a strong monad over $\mathbf{DCPO}$: its strength at (D, E) is given by

$$\tau_{DE}^{\mathcal{V}}\colon D \times \mathcal{V}E \to \mathcal{V}(D \times E) :: (x, \nu) \mapsto \lambda U. \int_{y\in E} \chi_U(x, y)d\nu,$$

where χ_U is the characteristic function of $U \in \sigma(D \times E)$. Whether $\mathcal{V}$ is a commutative monad on $\mathbf{DCPO}$ has remained an open problem for decades. Proving this to be true requires showing the following Fubini-type equation holds:

$$\int_{x\in D}\int_{y\in E} \chi_U(x, y)d\xi d\nu = \int_{y\in E}\int_{x\in D} \chi_U(x, y)d\nu d\xi, \tag{1}$$

for dcpo's D and E, for $U \in \sigma(D \times E)$ and for $\nu \in \mathcal{V}D, \xi \in \mathcal{V}E$ [12, Section 6]. The difficulty lies in the well-known fact that a Scott open set $U \in \sigma(D \times E)$ might not be open in the product topology $\sigma D \times \sigma E$ in general [18, Exercise II-4.26].

We obtain a commutative monad of valuations on $\mathbf{DCPO}$ by restricting to a suitable completion of $\mathcal{S}D$ inside $\mathcal{V}D$. There are several possibilities (cf. [21]), and we choose the smallest and simplest – the d-closure of $\mathcal{S}D$ in $\mathcal{V}D$.

Definition 1. For each dcpo D, we define $\mathcal{M}D$ to be the intersection of all sub-dcpo's of $\mathcal{V}D$ that contain $\mathcal{S}D$.

Theorem 2. *$\mathcal{M}$ is a commutative monad on $\mathbf{DCPO}$. The monad operations are (co)restrictions of those of $\mathcal{V}$.*

Remark 3. We note that $\mathcal{M}D$ is the first example of a commutative valuations monad on $\mathbf{DCPO}$ that contains the simple valuations. And, since every valuation on a domain D is a directed supremum of simple valuations [13, Theorem 5.2], it follows that $\mathcal{M} = \mathcal{V}$ on the category $\mathbf{DOM}$.

In fact, the construction of the monad $\mathcal{M}$ is a special case of a more general construction based on $\mathtt{K}$-categories.

Theorem 4. *Any $\mathtt{K}$-category $\mathbf{K}$ with $\mathbf{K} \subseteq \mathbf{D}$ determines a commutative valuations monad $\mathcal{V}_{\mathbf{K}}$ on $\mathbf{DCPO}$. In particular, $\mathcal{M} = \mathcal{V}_{\mathbf{D}}$. The categories of sober spaces $\mathbf{SOB}$ and that of* well-filtered *spaces $\mathbf{WF}$ are both $\mathtt{K}$-categories and subcategories of $\mathbf{D}$ and so they determine two additional commutative monads on $\mathbf{DCPO}$, denoted $\mathcal{P}$ and $\mathcal{W}$, respectively. These monads satisfy the following relationship $\mathcal{S}D \subseteq \mathcal{M}D \subseteq \mathcal{W}D \subseteq \mathcal{P}D \subseteq \mathcal{V}D$ for each dcpo D.*

Remark 5. All subsequent results stated here hold for all three monads $\mathcal{M}$, $\mathcal{W}$ and $\mathcal{P}$.

Kegelspitzen [16] are dcpo's equipped with a convex structure. A Kegelspitze which is also a domain is called *continuous*. We show every continuous Kegelspitze K has a *linear barycenter map* $\beta\colon \mathcal{M}K \to K$ making (K, β) an $\mathcal{M}$-algebra and conversely, every $\mathcal{M}$-algebra (K, β) on $\mathbf{DCPO}$ admits a Kegelspitze structure on K making $\beta\colon \mathcal{M}K \to K$ a linear map.

Theorem 6. *The Eilenberg-Moore category $\mathbf{DOM}^{\mathcal{M}}$ of $\mathcal{M}$ over $\mathbf{DOM}$ is isomorphic to the category of continuous Kegelspitzen and Scott-continuous linear maps.*

Since $\mathbf{DOM}^{\mathcal{M}} = \mathbf{DOM}^{\mathcal{V}}$, the above theorem also characterises the algebras of $\mathcal{V}$ on domains. This corrects an error in [13], where it is claimed that *continuous abstract probabilistic domains* and linear maps are isomorphic to $\mathbf{DOM}^{\mathcal{V}}$. A separating example is the extended non-negative reals $[0, \infty]$, which is a continuous Kegelspitze but not an abstract probabilistic domain.

Our final contribution is to show that our monads can be used to study the semantics of probabilistic programming languages.

Theorem 7. *The Kleisli category $\mathbf{DCPO}_{\mathcal{M}}$ of $\mathcal{M}$ is a sound and (strongly) adequate denotational model of* PFPC.

REFERENCES

[1] X. Jia, B. Lindenhovius, M. W. Mislove, and V. Zamdzhiev, "Commutative monads for probabilistic programming languages," in *Logic in Computer Science (LICS 2021)*, 2021, to appear. [Online]. Available: https://arxiv.org/abs/2102.00510

[2] M. O. Rabin, "Probabilistic algorithms," in *Algorithms and complexity, recent results and new direction*, J. F. Traub, Ed. Academic Press, 1976, pp. 21–40.

[3] J. Gill, "Computational complexity of probabilistic Turing machines," *SIAM Journal on Computing*, pp. 675–695, 1977.

[4] J. Baeten, J. Bergstra, and S. Smolka, "Axiomatizing probabilistic processes: ACP with generative probabilities," *Information and Computation*, vol. 121, pp. 234–255, 1995.

[5] K. G. Larsen and A. Skou, "Bisimulation through probabilistic testing," in *Proceedings of the 16th Annual ACM Symposium on Principles of Programming Languages*. IEEE Press, 1989.

[6] C. Morgan, A. McIver, and K. Seidel, "Probabilistic predicate transformers," *ACM Transactions on Programming Languages and Systems*, vol. 18, pp. 325–353, 1996.

[7] "Archive of workshops on probabilistic programming languages." [Online]. Available: https://pps2017.luddy.indiana.edu/2017/01/

[8] S. Staton, F. Wood, H. Yang, C. Heunen, and O. Kammar, "Semantics for probabilistic programming: higher-order functions, continuous distributions, and soft constraints," in *Proceedings of 2016 31st Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*. ACM Press, 2016.

[9] M. Vákár, O. Kammar, and S. Staton, "A domain theory for statistical probabilistic programming," *Proc. ACM Program. Lang.*, vol. 3, no. POPL, pp. 36:1–36:29, 2019. [Online]. Available: https://doi.org/10.1145/3290349

[10] M. Pagani, P. Selinger, and B. Valiron, "Applying quantitative semantics to higher-order quantum computing," in *The 41st Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL '14, San Diego, CA, USA, January 20-21, 2014*, S. Jagannathan and P. Sewell, Eds. ACM, 2014, pp. 647–658. [Online]. Available: https://doi.org/10.1145/2535838.2535879

[11] K. Cho, "Semantics for a quantum programming language by operator algebras," *New Generation Computing*, vol. 34, pp. 25–68, 2016.

[12] C. Jones and G. D. Plotkin, "A probabilistic powerdomain of evaluations," in *Proceedings of the Fourth Annual Symposium on Logic in Computer Science (LICS '89), Pacific Grove, California, USA, June 5-8, 1989*. IEEE Computer Society, 1989, pp. 186–195. [Online]. Available: https://doi.org/10.1109/LICS.1989.39173

[13] C. Jones, "Probabilistic Non-determinism," Ph.D. dissertation, University of Edinburgh, UK, 1990. [Online]. Available: http://hdl.handle.net/1842/413

[14] A. Jung and R. Tix, "The troublesome probabilistic power domain," in *Comprox III, Third Workshop on Computation and Approximation*, vol. 13, 1998, pp. 70 – 91.

[15] K. Keimel and J. D. Lawson, "D-completions and the d-topology," *Ann. Pure Appl. Log.*, vol. 159, no. 3, pp. 292–306, 2009. [Online]. Available: https://doi.org/10.1016/j.apal.2008.06.019

[16] K. Keimel and G. D. Plotkin, "Mixed powerdomains for probability and nondeterminism," *Logical Methods in Computer Science*, vol. 13, Issue 1, Jan. 2017. [Online]. Available: https://lmcs.episciences.org/2665

[17] S. Abramsky and A. Jung, "Domain Theory," in *Handbook of Logic in Computer Science (Vol. 3)*. Oxford, UK: Oxford University Press, 1994, pp. 1–168. [Online]. Available: http://dl.acm.org/citation.cfm?id=218742.218744

[18] G. Gierz, K. H. Hofmann, K. Keimel, J. D. Lawson, M. W. Mislove, and D. S. Scott, *Continuous Lattices and Domains*, ser. Encyclopedia of Mathematics and Its Applications 93. Cambridge University Press, 2003.

[19] W. K. Ho, J. Goubault-Larrecq, A. Jung, and X. Xi, "The Ho-Zhao problem," *Log. Methods Comput. Sci.*, vol. 14, no. 1, 2018. [Online]. Available: https://doi.org/10.23638/LMCS-14(1:7)2018

[20] E. G. Manes, *Algebraic Theories*, ser. Graduate Texts in Mathematics. Springer-Verlag, 1976, vol. 26. [Online]. Available: https://doi.org/10.1007/978-1-4612-9860-1

[21] X. Jia and M. Mislove, "Completing Simple Valuations in K-categories," 2020, preprint. [Online]. Available: https://arxiv.org/abs/2002.01865