Probabilistic Programming and a Domain Theoretic Approach to Skorohod's Theorem

Michael Mislove[1]

Department of Computer Science
Tulane University, New Orleans, LA 70118

A number of probabilistic programming languages have emerged that can manipulate and simulate probability distributions. Some of the languages, such as BUGS [2] and Infer.NET [5] can express only limited finite distributions, while others such as Stan [8] can support continuous distributions, but again with limitations. Finally, *universal languages* – Church [3], Anglican [9] – allow any computable prior for Bayesian modeling. In most instances, a programming language syntax is provided that allows sampling from distributions, and building more complicated distributions from simpler ones. Some presentations offer semantic models in which the basic distributions are seen as primitive inputs, while others, notably [7] give a semantics in a denotational style. This note describes a result that we believe could be useful both in devising semantic models for these languages, and also for recasting how to view the distributions that are the focus of the work.

These ideas arise from the main result from [6], which uses domain-theoretic techniques to prove Skorohod's Theorem, a staple of stochastic process theory. Loosely speaking, Skorohod's Theorem shows how to convert questions of weak convergence $\mu_n \longrightarrow_w \mu$ of a sequence of probability measures on a Polish space to questions about almost sure convergence $X_n \longrightarrow_{a.s.} X$ of an associated sequence of random variables on a standard Borel space. Since approximation schemes such as MCMC that are prevalent in probabilistic programming applications are stochastic processes, the hope is that a domain-theoretic approach to Skorohod's Theorem will aid in incorporating other stochastic processes into probabilistic programming languages.

To formally state the main result from [6], we need some notation. We let $CT = \{0,1\}^* \cup \{0,1\}^\omega$ denote the Cantor tree. CT is a domain when endowed with the prefix order (treating elements of the tree as words over $\{0,1\}$), and in fact it is

[1] The support of the US AFOSR is gratefully acknowledged.

a *bounded complete domain*, meaning every non-empty subset has a greatest lower bound. The Scott topology exists on any domain; it is always sober, but it is T_1 iff the underlying domain has the flat order. Scott continuous maps figure prominently in domain theory – they are used to define recursion via least fixed points.

The Cantor tree CT endowed with the Borel σ-algebra generated by the Scott topology is a standard Borel space. The maximal elements of CT form a copy of the Cantor set, $\mathcal{C}$. We also use Haar measure, $\nu_{\mathcal{C}}$ on $\mathcal{C} \simeq \{0,1\}^{\mathbb{N}}$ viewed as an infinite product of two-point groups. It's important to note that the canonical map from $\mathcal{C}$ onto the unit interval, and in fact, the canonical map can be used to realize a Borel isomorphism between $\mathcal{C}$ and $[0,1]$ that sends ν_C to Lebesgue measure. Here then is the formal statement of the main results from [6]:

Theorem 1. *(Skorohod's Theorem for Domains) If D be a countably-based bounded complete domain, then for each $\mu \in \mathsf{Prob}\, D$ there is a measurable map $X_\mu \colon \mathcal{C} \longrightarrow D$ satisfying $X_{\mu *}\, \nu_{\mathcal{C}} = \mu$.*[2] *Moreover, if $\mu_n \in \mathsf{Prob}\, D$ with $\mu_n \longrightarrow_w \mu \in \mathsf{Prob}$ in the weak topology, then $X_{\mu_n} \longrightarrow X_\mu$ a.s. wrt $\nu_{\mathcal{C}}$.*

Corollary *(Skorohod's Theorem) If P is a Polish space, then each probability measure μ on P can be realized as $X_\mu\, \lambda$, where $X_\mu \colon [0,1] \longrightarrow P$ is a measurable map and λ denotes Lebesgue measure. Moreover, if $\mu_n \in \mathsf{Prob}\, P$ with $\mu_n \longrightarrow_w \mu$ in the weak topology, then $X_{\mu_n} \longrightarrow X_\mu$ a.s. wrt λ.*

Skorohod's Theorem is a corollary of Theorem 1 as follows. Any Polish space P has a *computational model,* a countably-based bounded complete domain D for which P is homeomorphic to the set $\mathrm{Max}\, D$ of maximal elements endowed with the relative Scott topology. In fact, $\mathrm{Max}\, D$ is a G_δ, hence a Borel subset of D. The final ingredient is the Borel isomorphism we alluded to above between $[0,1]$ and $\mathcal{C}$ that takes Lebesgue measure to $\nu_{\mathcal{C}}$.

The proof of the results above employs 'abstract' domain theory:

- Every domain has two intrinsic topologies:
  · The Scott topology consists of those upper sets $U = \uparrow U = \{x \in D \mid (\exists y \in U)\, y \leq x\}$ that are *inaccessible by directed suprema*: $\sup S \in U \;\Rightarrow\; S \cap U \neq \emptyset$ if S is directed. Scott-continuous maps are monotone maps that preserve directed suprema; equivalently, they are the mappings between domains that are continuous wrt to the Scott topologies.
  · There is a Hausdorff refinement of the Scott topology, called the *Lawson topology*, for which the family $\{U \setminus \uparrow F \mid U \text{ Scott open } \& \ F \text{ finite}\}$ is a basis. The Scott and Lawson topologies generate the same Borel σ-algebra.
- For a large class of domains – bounded complete domains among them – the Lawson topology is compact. Such domains are called *coherent*.
- A *basis* for a domain is a subset $B \subseteq D$ satisfying $\{\uparrow x \mid x \in B\}$ is a basis for the Scott topology.
- For any domain D, the family $\mathsf{SProb}(D)$ of subprobability measures on D is

[2] If $f \colon X \longrightarrow Y$ and $\mu \in \mathsf{SProb}(X)$, then $f_*\mu \in \mathsf{SProb}(Y)$ is the push forward of μ under f.

again a domain, when endowed with the *valuation order:* If $\mu \in \mathsf{SProb}(D)$, then $\mu\colon \Sigma(D) \longrightarrow [0,1]$ by $\mu(U) = \int \chi_U d\mu$ is a valuation on $\Sigma(D)$ (the lattice of Scott-open sets), and $\mu \leq \nu \in \mathsf{SProb}(D)$ iff $\mu(U) \leq \nu(U)$ for all $U \in \Sigma(D)$.

- If D is a coherent domain, then $\mathsf{SProb}(D)$ also is coherent, and the weak topology on $\mathsf{SProb}(D)$ agrees with the Lawson topology. Moreover, if D has a least element, then the family of probability measures $\mathsf{Prob}(D)$ also is a coherent domain.
- If D is a countably based bounded complete domain, then $\mathsf{Max}\, D$, the family of maximal elements of D, is a Polish space in the inherited Scott topology. $\mathsf{Max}\, D$ also is a G_δ in D.
  Moreover, $\mathsf{Prob}(\mathsf{Max}\, D) = \mathsf{Max}\,\mathsf{SProb}(D)$, and the weak topology on $\mathsf{Prob}(\mathsf{Max}\, D)$ is the inherited Scott topology, which agrees with the inherited Lawson topology.
- An enumeration of the basis B leads to an abstract notion of computability in the domain D, and by extension, in $\mathsf{SProb}(D)$ by enumerating simple measures with rational coefficients concentrated on finite subsets of B.
- For any countably based bounded complete domain D, there is a countable chain $\phi_n\colon \mathsf{SProb}(D) \longrightarrow \mathsf{SProb}(D)$ of Scott-continuous maps satisfying $\phi_n \leq \phi_{n+1}$, $\sup_n \phi_n = 1_{\mathsf{SProb}(D)}$ and $\phi_n(D) \subseteq B$ has finite image in B.
  This realizes $\mu = \sup_n \phi_{n*}\mu$ as the supremum of a countable chain of simple measures, each supported on a finite subset of B, for each $\mu \in \mathsf{SProb}(D)$.

The last bullet is where the proof of Theorem 1 starts: given $\mu \in \mathsf{Prob}(D)$,

(i) We write $\mu = \sup_n \phi_{n*}\mu$ as the supremum of a chain of simple measures.

(ii) For each n, choose $m_n \geq n$ and define $g_{m_n}\colon \mathcal{C}_{m_n} = \{0,1\}^{2^{m_n}} \longrightarrow D$ so that
  - $g_{m_n*}\nu_{m_n} \ll \phi_{n*}\mu$, and $||g_{m_n*}\nu_{m_n} - \phi_{n*}\mu|| < 2^{-n}$ where ν_{m_n} is normalized counting measure on $\mathcal{C}_{m_n}$,
  - $m_n \leq m_{n+1}$ and $g_{m_n*}\nu_{m_n} \ll g_{m_{n+1}*}\nu_{m_{n+1}}$,
  - and finally $\sup_n g_{m_n*}\nu_{m_n} = \mu$.

(iii) The canonical projection $\pi_{m_n}\colon \mathcal{C} \longrightarrow \mathcal{C}_{m_n}$ satisfies $\pi_{m_n*}\nu_C = \nu_{m_n}$, so $(g_{m_n} \circ \pi_{m_n})_*\nu_C = g_{m_n*}\nu_{m_n}$.

(iv) The Splitting Lemma shows that the sequence $\{g_{m_n} \circ \pi_{m_n}\}_n$ forms an increasing chain of Scott continuous maps whose pointwise supremum $X_\mu\colon \mathcal{C} \longrightarrow D$ is measurable and satisfies $X_{\mu*}(\nu_C) = \mu$.

(v) Finally, the domain structure of $\mathsf{SProb}(D)$ is used to show that if $\mu_n \longrightarrow_w \mu$ in $\mathsf{Prob}(D)$, then $X_{\mu_n} \longrightarrow_{a.s} X_\mu$ in $\mathcal{C}$.

The question is how to combine this 'abstract' domain-theoretic notion of computability with the standard approach to computable metric spaces. In the latter, as exemplified e.g., in [4], one starts with a *computable metric space* $(P, \mathcal{S}, d)$, where (P, d) is a Polish space, $\mathcal{S} \subseteq P$ is a countable dense subset, and $d\colon \mathcal{S} \times \mathcal{S} \longrightarrow \mathbb{R}_+$ is computable. Then the computability structure on $\mathsf{Prob}(P)$ is defined via simple measures with rational coefficients that are concentrated on finite subsets of $\mathcal{S}$.

On the other hand, the simple measures $\phi_{n*}\mu$ that approximate an arbitrary measure μ is the domain setting above are concentrated on the sets $\phi_n(D)$, which are finite subsets of the basis B for the computational model D of the Polish space P. That computational model is built from topological subsets of P – e.g., in the case of a locally compact space, D consists of compact subsets of X under reverse inclusion and in the case the space has is second countable, the basis B consists of a neighborhood basis for P.

The question we will explore concerns how to relate the computability structure on a computable Polish space $(P, \mathcal{S}, d)$ and the computability structure on a computational domain model D for P.

References

[1] Abramsky, S. and A. Jung, Domain Theory, in: Handbook of Logic in Computer Science, Clarendon Press (1994), pp. 1–168.

[2] Gilks, W. R., A. Thomas and D. Spiegelhalter, A language and program for complex Bayesian modeling, The Statisitician **43** (1994), pp. 169–178.

[3] Goodman, N. D., V. K. Mansinghka, D. Roy, K. Bonawitz and J. B. Tenenbaum, Church: a language for generative models, In Proc. 24th Conf. Uncertainty in Artificial Intelligence (2008), pp. 220–229, URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.470.7468&rep=rep1&type=pdf

[4] Hoyrup, M. and C. Rojas, Computability of probability measures and Martin-Löf randomness over metric spaces, Information and Computation **207** (2009), pp. 830–847.

[5] Minka, T. J. Winn, J. Guiver and A. Kannan, Infer.NET 2.3, Nov. 2–9, Software available from http://research.microsoft.com/infernet.

[6] Mislove, M. W., Domains and random variables, arXiv:1607.07698, URL: https://arxiv.org/abs/1607.07698

[7] Ścibior, A., Z. Ghahramani and A. Gordon, Practical probabilistic programming with monads, In: Proceedings of Haskell '15, ACM Publishers, pp. 165–176.

[8] Stan Development Team. Stan: A C++ library for probability and sampling, version 2.2, 2014, URL: http://mc-stan.org/.

[9] Wood, F., J. W. van de Meent and V. Mansinghka, A new approach to probabilistic programming inference, In: Proceedings of 17th International Conference on Artificial Intelligence and Statistics, 2014.