Probabilistic Programming and a Domain-theoretic Approach to Skorohod's Theorem

Michael Mislove

Tulane University

Workshop on Probabilistic Programming Semantics
Paris, Jussieu

January 17, 2017

Supported by US AFOSR

Outline

I. Stochastic Processes and Skorohod's Theorem

II. Domains and Probability Measures

III. Proving Skorohod's Theorem

Along the way, compare with the computational Polish space approach

Stochastic Processes and Skorohod's Theorem

A *stochastic process* is a time-indexed family $\{X_t \mid t \in T \subseteq \mathbb{R}_+\}$ of random variables / elements $X_t : \Omega \to S$, where $(\Omega, \Sigma_\Omega, \mu)$ is a probability space, and S is a Polish space.

Fact: If S is Polish, then so is $(Prob(S), d_p)$, where d_p is the *Prokhorov metric*.

Skorohod's Theorem

Let S be a Polish space, let $\nu \in \mathrm{Prob}\, S$, and let λ denote Lebesgue measure on $[0,1]$. Then there is a random variable $X : [0,1] \to S$ satisfying $X_* \lambda = \nu$.

Moreover, if $\nu_n, \nu \in \mathrm{Prob}\, S$ satisfy $\nu_n \to_w \nu$, then the random variables $X_n, X : [0,1] \to S$ with $X_* \lambda = \nu, X_{n*} \lambda = \nu_n$ satisfy $X_n \to X$ λ-a.e.

Fact: We could use any *standard probability space* (S, Σ_S, m) instead of $([0,1], \lambda)$.

Domains are partially ordered sets with additional properties;

Informatic partial order

$p \sqsubseteq q$ if q contains more information than p.

Example: Upper Space $U(X) = (\{K \subseteq X \mid \emptyset \neq K \text{ compact}\} \cup \{X\}, \supseteq)$.

Directed completeness

$\emptyset \neq D \subseteq P$ *directed* if $x, y \in D \Rightarrow (\exists z \in D)\, x, y \leq z$.
P *directed complete:* D directed $\Rightarrow$ $\sup D$ exists.

$$\mathcal{F} \subseteq U(X) \text{ directed} \Rightarrow \sup \mathcal{F} = \bigcap \mathcal{F}.$$

Approximation

$x \ll y$ iff $y \leq \sup D \Rightarrow (\exists d \in D)\, x \leq d$.
Domain: $\Downarrow y = \{x \mid x \ll y\}$ directed and $y = \sup \Downarrow y$

$$K \ll L \text{ iff } L \subseteq K^\circ; \qquad L = \bigcap\{K \mid L \subseteq K^\circ\} = \sup\{K \mid K \ll L\}.$$

Scott Topology

U Scott open if:

- $U = \uparrow U = \{x \in P \mid (\exists u \in U)\, u \leq x\}$ and
- D directed, $\sup D \in U \;\Rightarrow\; D \cap U \neq \emptyset$.

Morphisms

$f\colon P \to Q$ is *Scott continuous* if:

- f is monotone, and
- D directed $\Rightarrow\; f(\sup D) = \sup f(D)$.

$f\colon X \to Y \;\Rightarrow U(f)\colon U(X) \to U(Y)$ by $U(f)(K) = f(K)$

is monotone and $f(\bigcap \mathcal{F}) = \bigcap f(\mathcal{F})$.

Lawson Topology

Basis: $\{\twoheaduparrow x \setminus \uparrow F \mid F \in \mathcal{P}_{<\omega} D\}$

Hausdorff refinement of Scott topology.

All the domains we discuss are Lawson compact.

Embedding X in $U(\overline{X})$

X Polish $\Rightarrow X \hookrightarrow [0,1]^\omega \Rightarrow X \subseteq_{G_\delta} \overline{X}$ compact Polish

$\{K_n \mid K_n \in U(\overline{X}), n > 0\}$: neighborhood basis of compact subsets of $\overline{X}$.
Then:

1° $X \subseteq \overline{X} \hookrightarrow \operatorname{Max} U(\overline{X}) \subseteq U(\overline{X})$ by $x \mapsto \{x\} = \bigcap_n \{K_n \mid x \in K_n^\circ\}$

2° X inherits the Scott topology = Lawson topology on $\operatorname{Max} U(\overline{X})$.

3° Each family $\{K_1, \ldots, K_{n_m}\}$ with $\overline{X} \subseteq \bigcup_{n_m} K_i^\circ$ defines a

Scott-continuous projection $\psi_m \colon U(\overline{X}) \to U(\overline{X})$ with finite image.

4° Ordering the covers by refinement yields $\mathbf{1}_{U(\overline{X})} = \sup_m \psi_m$.

Prob(D) is a Domain

D (Lawson compact) domain $\Rightarrow$ Prob(D) (Lawson compact) domain:

1° $\mu \leq \nu$ iff $\int f d\mu \leq \int f d\nu$ ($\forall f\colon D \to \mathbb{R}_+$ Scott continuous)

2° D Lawson compact $\Rightarrow$ $(\mathsf{Prob}(D), \mathit{weak}) = (\mathsf{Prob}(D), \mathit{Lawson})$.

3° $D = U(\overline{X}), X$ Polish $\Rightarrow$ $\mathbf{1}_{\mathsf{Prob}(D)} = \sup_n \psi_{n*}$, so

$\mu = \sup_n \psi_{n*}\,\mu$, with $\psi_{n*}\,\mu = \sum_{i \leq m_n} \mu(K_i)\delta_{K_i}$ $(\forall \mu)$.

4° Choosing $r_i < \mu(K_i)$ & $L_i \ll K_i$ and $\epsilon = 1 - (\sum_i r_i)$ produces

$\nu_n \stackrel{\mathsf{def}}{=} \epsilon\delta_X + \sum_{i \leq_{m_n}} r_i\delta_{L_i} \ll \psi_{n*}\mu$.

5° Construct $\nu_1 \ll \cdots \ll \nu_n \ll \nu_{n+1} \ll \cdots$ with $\mu = \sup_n \nu_n$.

Prob(D) is a Domain

D (Lawson compact) domain $\Rightarrow$ Prob(D) (Lawson compact) domain:

1° $\mu \leq \nu$ iff $\int f d\mu \leq \int f d\nu$ ($\forall f : D \to \mathbb{R}_+$ Scott continuous)

2° D Lawson compact $\Rightarrow$ $(\text{Prob}(D), \textit{weak}) = (\text{Prob}(D), \textit{Lawson})$.

3° $D = U(\overline{X}), X$ Polish $\Rightarrow$ $\mathbf{1}_{\text{Prob}(D)} = \sup_n \psi_{n*}$, so

$\mu = \sup_n \psi_{n*}\, \mu$, with $\psi_{n*}\, \mu = \sum_{i \leq m_n} \mu(K_i)\delta_{K_i}$ $(\forall \mu)$.

4° Choosing $r_i < \mu(K_i)$ & $L_i \ll K_i$ and $\epsilon = 1 - (\sum_i r_i)$ produces

$\nu_n \overset{\text{def}}{=} \epsilon\delta_X + \sum_{i \leq m_n} r_i \delta_{L_i} \ll \psi_{n*}\mu$.

5° Construct $\nu_1 \ll \cdots \ll \nu_n \ll \nu_{n+1} \ll \cdots$ with $\mu = \sup_n \nu_n$.

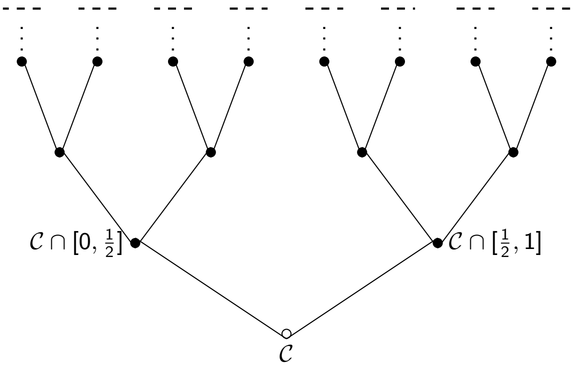
By comparison:

(X, S, d) a *computational Polish space* implies

$\mu = \lim_n \sum_{i \leq m_n} d_i \delta_{x_i} \in \text{Max}\, U(\overline{X})$ with $x_i \in S$.

Then $\nu_n \ll \psi_{n*}\mu \leq \mu \Rightarrow (\exists N)\ n \geq N \Rightarrow \nu_n \ll \sum_{i \leq m_n} d_i \delta_{x_i}$.

A Domain Environment for $\mathcal{C}$



$$\mathcal{C} \hookrightarrow \mathsf{Max}\,\mathbb{CT} \subseteq \mathbb{CT} ::= \{0,1\}^* \cup \{0,1\}^\omega$$

$$\mathcal{C}_n = \{0,1\}^n \Rightarrow \exists \pi_n \colon \mathcal{C} \to \mathcal{C}_n \text{ retraction}$$

$$\mathbf{1}_{\mathcal{C}} = \sup_n \iota_n \circ \pi_n|_{\mathcal{C}}, \text{ where } \iota_n \colon \mathcal{C}_n \hookrightarrow \mathcal{C} \text{ lower semicontinuous.}$$

A Domain Environment for $\mathcal{C}$

- Given $\mu \in \mathrm{Prob}U(\overline{X})$ and $\psi_{n*}\mu = \sum_{i \le m_n} \mu(K_i)\delta_{K_i}$, fix p_n and choose $r_i \in D_{p_n} = \{\frac{s}{2^{p_n}} \mid 0 \le s \le 2^{p_n}\}$ and $L_i \ll K_i$.
- Then $\nu_n \stackrel{\text{def}}{=} (1 - \sum_i r_i)\delta_X + \sum_{i \le m_n} r_i \delta_{L_i} \ll \sum_{i \le m_n} \mu(K_i)\delta_{K_i}$.
- Define $f_n \colon \mathcal{C}_{p_n} \to U(\overline{X})$ by $f_n(j) = \begin{cases} L_1 & \text{if } 1 \le j \le r_1 \\ L_2 & \text{if } r_1 < j \le r_1 + r_2 \\ \vdots & \\ X & \text{if } \sum_i r_i < j \end{cases}$

Then $f_{n*}(\frac{1}{2^{p_n}} \sum_{i \le 2^{p_n}} \delta_{\frac{i}{2^{p_n}}}) = (1 - \sum_i r_i)\delta_X + \sum_{i \le m_n} r_i \delta_{L_i} = \nu_n$.

- So, $f_n \circ \pi_n \colon \mathcal{C} \to U(\overline{X})$ is Lawson continuous and $(f_n \circ \pi_n)_*(\mu_{\mathcal{C}}) = \nu_n = (1 - \sum_i r_i)\delta_X + \sum_{i \le m_n} r_i \delta_{L_i}$.

A Domain Environment for $\mathcal{C}$

- Given $\mu \in \mathrm{Prob}\, U(\overline{X})$ and $\psi_{n*}\mu = \sum_{i\le m_n} \mu(K_i)\delta_{K_i}$, fix p_n and choose $r_i \in D_{p_n} = \{\frac{s}{2^{p_n}} \mid 0 \le s \le 2^{p_n}\}$ and $L_i \ll K_i$.
- Then $\nu_n \stackrel{\text{def}}{=} (1 - \sum_i r_i)\delta_X + \sum_{i\le m_n} r_i\delta_{L_i} \ll \sum_{i\le m_n} \mu(K_i)\delta_{K_i}$.

Then $f_{n*}(\frac{1}{2^{p_n}} \sum_{i\le 2^{p_n}} \delta_{\frac{i}{2^{p_n}}}) = (1 - \sum_i r_i)\delta_X + \sum_{i\le m_n} r_i\delta_{L_i} = \nu_n$.

- So, $f_n \circ \pi_n \colon \mathcal{C} \to U(\overline{X})$ is Lawson continuous and $(f_n \circ \pi_n)_*(\mu_{\mathcal{C}}) = \nu_n = (1 - \sum_i r_i)\delta_X + \sum_{i\le m_n} r_i\delta_{L_i}$.
- By construction $f_n \colon {\downarrow}\mathcal{C}_{P_n} \to U(\overline{X})$ satisfies $f_n \le f_{n+1}$, and then

  $X = (\sup_n f_n \circ \pi_n)|_{\mathcal{C}} \colon \mathcal{C} \to U(\overline{X})$ measurable with $X_*\mu_{\mathcal{C}} = \mu$.

A Domain Environment for $\mathcal{C}$

- By construction $f_n \colon {\downarrow}\mathcal{C}_{P_n} \to U(\overline{X})$ satisfies $f_n \leq f_{n+1}$, and then
  $X = (\sup_n f_n \circ \pi_n)|_{\mathcal{C}} \colon \mathcal{C} \to U(\overline{X})$ measurable with $X_* \mu_{\mathcal{C}} = \mu$.
- If $\mu_m \to_w \mu \in \mathrm{Prob}(X)$, define
  $\rho_{m,n} \ll \psi_{n*}\, \mu_m$ and $f_{m,n} \colon {\downarrow}\mathcal{C}_{p_{m,n}} \to U(\overline{X})$ as above.
  Then $X_m = (\sup_n f_{m,n})|_{\mathcal{C}}$ satisfies $X_m \colon \mathcal{C} \to U(\overline{X})$ measurable with $X_{m*}\, \lambda = \mu_m$.
- Argue directly that $X_m \to X$ a.s. λ.

Questions?