Analysis of Probabilistic Systems

Bootcamp Lecture 1: Probabilistic Programming Languages

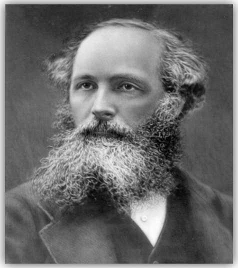
Prakash Panangaden[1]

[1]School of Computer Science
McGill University

Fall 2016, Simons Institute

Outline

The true logic!



The “true” logic of the World.

The actual science of logic is conversant at present only with things either certain, impossible or entirely doubtful; none of which, fortunately, we have to reason on. Therefore the true logic for this world is the calculus of Probabilities which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man’s mind — James Clerk Maxwell (1850)

The age of stochasticity!?

The dawning of the age of stochasticity

For over two millenia, Aristotle's logic has ruled over the thinking of Western intellectuals. All precise theories, all scientific models, even models of the process of thinking itself, have in principle conformed to the straight- jacket of logic. But from its shady beginnings devising gambling strategies and counting corpses in medieval London, probability theory and statistical inference now emerge as better foundations for scientific models,... — David Mumford

Conditioning as inference

- $\dfrac{p \quad p \Rightarrow q}{q}$ or $(p \wedge (p \Rightarrow q)) \Rightarrow q$
- $\text{Pr}(A \mid B) = \dfrac{\text{Pr}(A \cap B)}{\text{Pr}(B)}$ or
- $\text{Pr}(A \wedge B) = \text{Pr}(A \mid B) \cdot \text{Pr}(B)$.
- Conditioning allows you to revise your probability estimates as you gain information.
- (Bayes' theorem) $\text{Pr}(H \mid O) = \dfrac{\text{Pr}(O \mid H) \cdot \text{Pr}(H)}{\text{Pr}(O)}$.
- What if $\text{Pr}(B) = 0$?
- In discrete probability $\text{Pr}(B) > 0$ is a side condition.
- On continuous spaces; one cannot just dismiss this case.
- One needs conditional probability distributions or regular conditional probability densities aka *disintegrations*.

Basic discrete probability

- S: a finite set; the sample space.
- $A, B, \ldots \subseteq S$: events.
- $\text{Pr}(\cdot) : S \longrightarrow [0, 1]$ assigns probabilties to elements of S and hence to events.
- $X : S \longrightarrow \mathbb{R}$ a random variable.
- $\text{Pr}(U \subset \mathbb{R})_X := \text{Pr}(\{s | X(s) \in U\})$; probability distribution on $\mathbb{R}$ induced by X.
- Notation $\text{Pr}(\{X = r\})$ or $\text{Pr}(\{X \in U\})$.

Independence

- Fix $(S, \text{Pr}(\cdot))$.
- Two events A, B are independent if $\text{Pr}(A \wedge B) = \text{Pr}(A) \cdot \text{Pr}(B)$.
- Two random variables X, Y are independent if for all r_1, r_2, $\text{Pr}(\{X = r_1\})$ and $\text{Pr}(\{Y = r_2\})$ are independent.
- Given three random variables X, Y, Z we say X, Y are independent *conditional on* Z if
  $\text{Pr}(X = r_1 \wedge Y = r_2 \mid Z = r_3) =$
  $\text{Pr}(X = r_1 \mid Z = r_3) \cdot \text{Pr}(Y = r_2 \mid Z = r_3)$.
- For 3 or more we can define pairwise independence and mutual independence. They are not the same.
- Consider two tosses of a fair coin. A: first toss is H, B second toss is H and C: the two tosses give the same result.
  $\text{Pr}(A) = \text{Pr}(B) = \text{Pr}(C) = \frac{1}{2}$.
- $\text{Pr}(AB) = \text{Pr}(AC) = \text{Pr}(BC) = \frac{1}{4}$.
- However $\text{Pr}(ABC) = \frac{1}{4}$ not $\frac{1}{8}$.

Probabilistic models

- Reasoning under uncertainty involves the interplay between independence and conditioning: Bayes nets.
- Hugely successful framework: graphical models – a way of representing Bayes nets.
- Conditional independence structure captures the model in a compact way and is useful for computing marginals and conditional probabilities.
- Claim: these models are indeed useful for many situations but they are not compositional and they often “flatten out” the structure inherent in a situation.
- Example: Perhaps a subgraph is repeated many times in a regular pattern.
- A new hope: use the theory of programming languages to give a *structural* way of presenting and reasoning about probabilistic models.

Other developments

- Verification of probabilistic systems: Vardi, de Alfaro, Segala, Baier, Katoen, Kwiatkowska, Hermanns, ...
- Performance analysis: Jane Hillston
- Probabilistic transition systems: Larsen, Skou, van Glabbeek, ...
- Continuous state spaces: Desharnais, Edalat, P.; Doberkat, Rutten and de Vink, Mislove et al.,
- Metrics and approximation theory: Smolka et al.; Desharnais Gupta, Jagadesan, P.; Chaput, Danos, P. and Plotkin; van Breugel and Worrell.
- Combining probability and non-determinism: Mislove, Keimel and Plotkin, Varacca, Winskel, Goubault-Larrecq,...

Probability and domains

- Saheb-Djahromi (1978) developed probabilistic LCF.
- (1980) Developed probability distributions on domains. Showed that dcpo structure is preserved.
- Higher-order language but probabilities only at ground types.
- Claire Jones and Gordon Plotkin (1989) developed a probabilistic power domain.
- Tix, Jung: showed that this was a “vexing” construction. No known cartesian-closed categories suitable for programming language semantics in which this lives.
- Edalat: integration on domains.

Kozen's language (1981)

Syntax

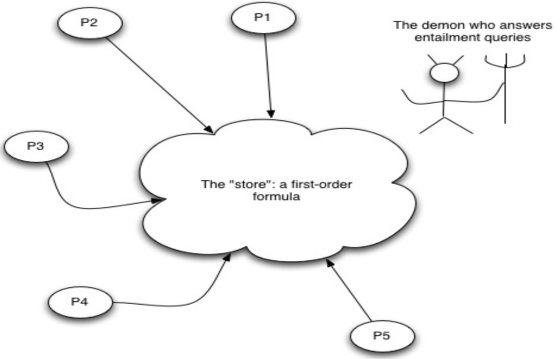
$$S ::== x_i := f(\vec{x}) | S_1 ; S_2 | \textit{if } \mathbf{B} \textit{ then } S_1 \textit{ else } S_2 | \textit{while } \mathbf{B} \textit{ do } S.$$

- Semantics: programs transform input distributions to output distributions.
- Backward semantics: expectation value transformers.
- Many contributions from the "Oxford group": Hoare, He, Seidel, Morgan, McIver, ...

Probabilistic ccp

- Saraswat, Gupta, Jagadeesan (1997): Combined probability with the concurrent constraint programming paradigm.
- Gupta, Jagadeesan, P. (1999) added recursion, showed interesting computational effects with no pure measure-theory counterpart.
- Explicit use of conditioning.
- Largely ignored.

The ask/tell model



CCP processes

$\mathsf{ask}(\phi)$: does the current store (σ) entail ϕ?

$\mathsf{tell}(\phi)$: add ϕ to the current store.

$P_1 || P_2$: run P_1 and P_2 in parallel.

$\mathsf{new}\ X\ \mathsf{in}\ P$: fresh local variable; $\nu X.P$.

recursive procedures

Probabalistic CCP

New ingredient: **choose** X **from** Dom **in** P

X: local variable, scope is P

Dom is a finite set

- Random variables are hidden
- Each random variable has its own independent probability distribution.

Constraints and conditioning

choose X **from** $\{0, 1, 2, 3\}$ **in**

$\mathsf{tell}(X \leq 2) \mid\mid [\mathsf{ask}((X = 0) \vee (X = 1)) \rightarrow \mathsf{tell}(a)] \mid\mid$
$[\mathsf{ask}(X = 2) \rightarrow \mathsf{tell}(b)]$

Produces a with probability 0.5 and b with probability 0.25,

however, it cannot produce true because of the constraint on X.

Inconsistent stores are discarded and the probabilities are renormalized.

Modelling probabilistic systems

- Systems feature probabilistic components: protocols using randomization, bluetooth, controllers, cells,...
- Probabilistic transition systems as abstractions of real systems
- Logics for reasonong about probability

Labelled Transition Systems

- A set of states S,
- a set of *labels* or *actions*, L or $\mathcal{A}$ and
- a transition *relation* $\subseteq S \times \mathcal{A} \times S$, usually written

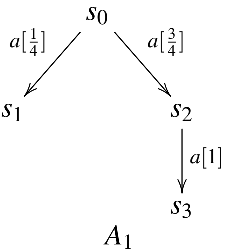
$$\rightarrow_a \subseteq S \times S.$$

  The transitions could be indeterminate (nondeterministic).
- We write $s \xrightarrow{a} s'$ for $(s, s') \in \rightarrow_a$.

Discrete probabilistic transition systems

- Just like a labelled transition system with probabilities associated with the transitions.
- $$(S, \mathsf{L}, \forall a \in \mathsf{L}\ T_a : S \times S \longrightarrow [0, 1])$$
- The model is *reactive*: All probabilistic data is *internal* - no probabilities associated with environment behaviour.

Examples of PTSs



Probability at higher type

- Conditioning as a basic primitive: CCP+prob, CHURCH, Tabular
- Domain theory in conjunction with probability: Saheb-Djahromi, Ramsey & Pfeffer, Ugo Dal Lago, Varacca & Goubault-Larrecq,...
- Prof. Scott's work showing that random variables can be introduced in a model of λ-calculus. Application of a random variable to another produces a random variable.
- Recent work (2016) by Staton et al. on semantics and reasoning principles for a higher-order probabilistic programming language.
- Recent work (2016) on approximate equational reasoning by Mardare, P. and Plotkin.

The Shock

Ackerman, Freer, Roy

There is a pair of computable probability random variables such that the conditional probability distribution is not computable.

We need to control how we allow conditioning.

Four more lectures

- Review of measure and integration
- Lawvere-Gìry monad
- Logical characterization of bisimulation
- Metrics measuring behavioural similarity of Markov processes
- Guest lecture: Mislove on probability theory and domain theory