Foundations for type-driven probabilistic modelling

Ohad Kammar
University of Edinburgh

Logic Summer School
Australian National University
4–16 December, 2023
Canberra, ACT, Australia

supported by: The Royal Society, The Alan Turing Institute, Facebook Research, NCSC

Computational golden era of:

logic & type rich computation

Statistical computation

Computational golden era of:

logic & type rich computation

expressive type systems:

Haskell, OCaml, Idris

mechanised mathematics:

Agda, Coq, Isabelle/HoL, Lean

Verification:

SMT-powered, realistic systems

Statistical computation

generative modelling

+

efficient inference:

Monte-Carlo Simulation

or gradient-based optimisation

"AI"

Computational golden era of:

logic & type rich computation

Statistical computation

Clear connection to Foundations:

- Radek's
- John's courses
- Michael's
- Dominik's

- this course

Why foundations?

countably supported measures

Discrete Probability ↪ Well-behaved Probability

s-finite measures over standard Borel spaces

Continuous Probability

Lebesgue measure on $\mathbb{R}^n$

→ Measurable spaces

standard, established

Poor type structure

→ Quasi-Borel spaces

new, experimental

rich type structure

(this course)

Takeaway:

use types to develop smoothly

Plan:

1) type-driven probability: discrete case (Mon + Tue (?))

2) Borel sets & measurable spaces (Tue)

3) Quasi Borel spaces, simple type structure (Wed)

4) Dependent type structure & standard Borel spaces (Thu)

5) Integration & random variables (Fri)

Please ask questions!

Course web page

Scribble

Advertisement

Interested in the computational / mathematical / logical foundations of Computer science?

Hedging your bets on a funded PhD offer from ANU?

Check out our PhD Programs:

THE UNIVERSITY OF EDINBURGH

lfcs | Laboratory for Foundations of Computer Science

LFCS PhD

Dependable AI for Robotics CDT

Languge of distribution & Probability

X type (=space) of values / outcomes

DX type of distributions / measures over X

$PX \subseteq DX$ Sub type of Probability measures (total measure 1)

BX type of measurable events - Subsets of X we wish to measure

W type of weights : $[0, \infty]$

type judgement

$$\mu : DX, E : BX \vdash \mathrm{Ce}_{\mu}[E] : W$$

measure μ assigns to E

Axioms for measures

Empty event: $\emptyset : \mathbb{B}X$

Its measure is $0 : \mathbb{W}$:

$$\mu : \mathbb{D}X \vdash \underset{\mu}{\mathbb{C}e}[\emptyset] = 0 \quad : \mathbb{W}$$

Axioms for measures

BX is a Boolean sub-algebra:

$$E : BX \vdash E^c : BX$$

$$E, F : BX \vdash E \cup F, E \cap F : BX$$

$E, C : BX, \mu : DX \vdash$ (disjoint additivity)

$$\underset{\mu}{Ce}[E] = \underset{\mu}{Ce}[E \cap C] + \underset{\mu}{Ce}[E \cap C^c] : W$$

Axioms for measures

$\omega := (\mathbb{N}, \leq)$ $\quad (\mathbb{B}, \subseteq)$ $\quad (\mathbb{W}, \leq)$ posets

$$(BX, \subseteq)^{\omega} := \{ (E_n)_{n\in\mathbb{N}} \in (BX)^{\mathbb{N}} \mid E_0 \subseteq E_1 \subseteq E_2 \subseteq \cdots \}$$

$(BX, \subseteq)$ and $(\mathbb{W}, \leq)$ are ω-chain-closed:

$E_- : (BX, \subseteq)^{\omega} \vdash \bigcup_n E_n : BX$ $\qquad a_- : (\mathbb{W}, \leq)^{\omega} \vdash \sup_n a_n : \mathbb{W}$

$E_- : (BX, \subseteq)^{\omega}, \mu : DX \vdash$ (Scott continuity)

$$\mathrm{Ce}_{\mu}\left[\bigcup_n E_n\right] = \sup_n \mathrm{Ce}_{\mu}[E_n] \quad : \mathbb{W}$$

Axiom for Probability

$$\mathrm{Cost} : PX \xhookrightarrow{\subseteq} DX \qquad\qquad 1 : W$$

$$\mu : PX \vdash \underset{\mathrm{Cost}\ \mu}{\mathcal{E}}[X] = 1 \quad : W$$

Avoid casting:

$$E : BX, \mu : PX \vdash \underset{\mu}{\Pr}[E] := \underset{\mathrm{Cost}\ \mu}{\mathcal{E}}[E] \quad : \quad [0,1] \subseteq W$$

Axioms for measures

Integration:

$$\mu : DX,\ \varphi : \mathbb{W}^X \vdash \int \mu\, \varphi \quad : \mathbb{W}$$

(Lebesgue integral)

Again, avoid casting:

$$\mu : PX,\ \varphi : \mathbb{W}^X \vdash \underset{\mu}{\mathbb{E}}[\varphi] := \int (\mathrm{cast}\ \mu)\, \varphi \quad : \mathbb{W}$$

(Expectation)

More structure & notation later (... technical ...)

Have: language + axioms

Want: model

today: discrete measures

rest of course: discrete + continuous

Discrete model

type X : set

$$DX := \{\mu : X \to \mathbb{W} \mid \mu \text{ is countably supported}\}$$

(next slide)

Support

$$\mu : W^X, S : \mathcal{P}X \vdash S \text{ supports } \mu := \forall x : X.\ \mu x > 0 \Rightarrow x \in S \quad : \mathrm{Prop}$$

Powerset

$$\mu : W^X \vdash \mathrm{supp}\,\mu := \{x \in X \mid \mu x > 0\} \quad : \mathcal{P}X$$

$\mathrm{supp}\,\mu$ is the smallest set supporting μ

Discrete model

type X : set

$$DX := \{\mu : X \to \mathbb{W} \mid \mu \text{ is countably supported}\}$$

$$:= \{\mu : X \to \mathbb{W} \mid \operatorname{supp} \mu \text{ is countable}\}$$

Ex. measures

- X ctbl. Counting measure $\#_X : DX$

  $\#_X := \lambda x : X . 1$ (NB: $\mathrm{supp}\,\#_X = X$ ✓ctbl)

- Dirac measure:

  $$x : X \vdash \delta_x := \lambda x' . \begin{cases} x = x' : & 1 \\ \text{o.w.} : & 0 \end{cases} : DX$$

  NB: $\mathrm{supp}\,\delta_x = \{x\}$ ✓ctbl

- Zero measure $\underline{0} := \lambda x . 0 : DX$

  NB: $\mathrm{supp}\,\underline{0} = \emptyset$ ✓ctbl

Discrete model

type X : set

$$DX := \{\mu : X \to W \mid \mu \text{ is countably supported}\}$$

$$\mu : DX, E : BX \vdash \underset{\mu}{Ce}[E] := \sum_{x \in E} \mu x$$

$$:= \sum_{x \in E \cap \mathrm{Supp}\,\mu} \mu x$$

Lemma: $\mu : DX$, $S \in P_{ctbl} X$, S supports μ, $E : BX \vdash$

$$\underset{\mu}{Ce}[E] = \sum_{x \in E \cap S} \mu x$$

Ex:

- $E : \mathbb{B}X \vdash \underset{\#_X}{Ce}[E] = |E| := \begin{cases} E \text{ has } n \text{ elements}: & n \\ E \text{ infinite}: & \infty \end{cases}$

- $E : \mathbb{B}X, x : X \vdash \underset{\delta_x}{Ce}[E] = \begin{cases} x \in E: & 1 \\ x \notin E: & 0 \end{cases} =: [x \in E] : W$

NB: $E : \mathbb{B}X \vdash [- \in E] : X \to W$

indicator function

- $E : \mathbb{B}X \vdash \underset{\underline{0}}{Ce}[E] = 0$

Validate axioms

$$\mu : DX \vdash \underset{\mu}{\mathcal{C}e}[\emptyset] = 0 \quad : W$$

$$E, C : BX, \mu : DX \vdash$$

$$\underset{\mu}{\mathcal{C}e}[E] = \underset{\mu}{\mathcal{C}e}[E \cap C] + \underset{\mu}{\mathcal{C}e}[E \cap C^c] \quad : W$$

$$E_{\_} : (BX, \subseteq)^{\omega}, \mu : DX \vdash$$

$$\underset{\mu}{\mathcal{C}e}\Big[\bigcup_n E_n\Big] = \sup_n \underset{\mu}{\mathcal{C}e}[E_n] \quad : W$$

Kernels κ from Γ to X:

$$\kappa : (DX)^{\Gamma}$$

kernels are "open/parameterised" measures

Ex: Dirac kernel: $\delta_{\_} : (DX)^{X}$

Kock Integral

$$\mu : D\Gamma, \kappa : DX^{\Gamma} \vdash \oint \mu \kappa : DX$$

In discrete model:

$$\oint \mu \kappa := \lambda x{:}X. \sum_{\gamma \in \Gamma} \mu\gamma \cdot \overbrace{k(\gamma; x)}^{:= \kappa \gamma x}$$

(Weak) disintegration problem:

Input: $\mu : D\Gamma$ $\nu : DX$

Output: a kernel $k : (DX)^\Gamma$ s.t.

$$\oint \mu k = \nu$$

Call such k $\underline{\underline{a}}$ (weak) disintegration of ν w.r.t. μ.

(non-standard terminology)

Ex disintegration:

$\underline{n} := \{0, 1, 2, \dots, n-1\}$

disintegrate $\#_{\underline{2}^{\underline{n+1}}}$ w.r.t. $\#_{\underline{2}}$

$k : \left(D(\underline{2}^{\underline{n+1}})\right)^{\underline{2}}$ $\quad k(x; f) := \begin{cases} f(n) = x : & 1 \\ \text{o.w.} : & 0 \end{cases}$

$$\left(\oint \#_{\underline{2}} k\right) f = \sum_{x \in \underline{2}} \overbrace{\#_{\underline{2}} x}^{1} \cdot k(x; f)$$

$\boxed{\text{NB: } \mathrm{supp}(kx) \ \checkmark \ \text{ctbl}}$

$$= k(0; f) + k(1; f) = k(fn; f) = 1 = \#_{\underline{2}^{\underline{n+1}}}(f)$$

Probability measures

$$P X := \{\mu : DX \mid \oint_\mu [X] = 1\} \overset{\subseteq}{\hookrightarrow} DX$$

Lemma: $\delta_- : X \to DX$ and $\oint : D\Gamma \times (DX)^\Gamma \to DX$

lift along the inclusion $\mathrm{cast} : P \overset{\subseteq}{\hookrightarrow} D$:

$$\begin{tikzcd}
 & PX \arrow[d, "\mathrm{cast}"] \\
X \arrow[ur, dashed, "\delta_-"] \arrow[r, "\delta_-"'] \arrow[ur, phantom, "="] & DX
\end{tikzcd}
\qquad
\begin{tikzcd}
P\Gamma \times (PX)^\Gamma \arrow[r, dashed, "\oint"] \arrow[d, "\mathrm{cast} \times (\mathrm{cast}^\Gamma)"'] \arrow[dr, phantom, "="] & PX \arrow[d, "\mathrm{cast}"] \\
D\Gamma \times (DX)^\Gamma \arrow[r, "\oint"'] & DX
\end{tikzcd}$$

Prop (discrete Giry): (Michèle Giry '82)

$(D, \delta_{-}, \oint)$ is a monad i.e.

$$\mu : \Gamma, k : (DX)^{\Gamma} \vdash \oint \delta_{\mu} k = k\mu$$

$$\mu : DX \vdash \oint \mu(dx)\delta_x = \mu \quad : DX$$

$$\mu : D\Gamma, k : (DX)^{\Gamma}, t : (DY)^{X} \vdash$$

$$\oint \mu(d\gamma)\left(\oint (k\gamma) t\right) = \oint \left(\oint \mu k\right)(dx)\, t(x)$$

Corollary: $(P, \delta_{-}, \oint)$ is a monad.