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:

- Ralf'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

Measurable spaces

standard, established

Poor type structure

Continuous Probability

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

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 $0 : \mathbb{W}$:

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

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)

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

Axioms for measures

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

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

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

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

$E_{-} : (\mathbb{B}X, \subseteq)^{\omega}, \mu : \mathbb{D}X \vdash$ (Scott Continuity)

$$\mathcal{C}e_{\mu}\left[\bigcup_n E_n\right] = \sup_n \mathcal{C}e_{\mu}[E_n] \quad : \mathbb{W}$$

Axiom for Probability

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

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

Avoid costing:

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

Axioms for measures

Integration:

$$\mu : DX,\ \varphi : W^X \vdash \int \mu\, \varphi \quad : W$$ (Lebesgue integral)

Again, avoid casting:

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

Powerset

$$\mu : \mathbb{W}^X, S : \mathcal{P}X \vdash S \text{ supports } \mu :=$$

$$\forall x : X.\ \mu x > 0 \Rightarrow x \in S \quad : \mathrm{Prop}$$

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

$\mathrm{supp}\,\mu$ 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:

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

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

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

$\mu : DX, E : BX \vdash \mathcal{C}_{\mu}[E] := \sum_{x \in E} \mu x$

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

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

$\mathcal{C}_{\mu}[E] = \sum_{x \in E \cap S} \mu x$

Ex:

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

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

NB: $E : BX \vdash [- \in E] : X \to W$

indicator function

- $E : BX \vdash \underset{\underline{0}}{Ce}[E] = 0$

Validate axioms

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

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

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

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

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

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:

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

(Weak) disintegration problem:

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

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. $\#_{\underline{2}}$

$k : \left(D(\underline{2}^{\underline{n+1}})\right)^{\underline{2}}$

$$k(x; f) := \begin{cases} f(n) = x : & 1 \\ \text{o.w.} : & 0 \end{cases}$$

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

$$\boxed{\text{NB: } \mathrm{supp}(kx) \quad \checkmark\ c+bl}$$

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

Probability measures

$$PX := \{\mu : DX \mid \mathcal{E}_{\mu}[X] = 1\} \overset{\subseteq}{\hookrightarrow} DX$$

Lemma: $\delta\_ : X \to DX$ and $\oint : D\Gamma \times (DX)^{\Gamma} \to DX$

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

$$\begin{tikzcd}
 & PX \arrow[d, hook, "\mathrm{cast}"] \\
X \arrow[ur, dashed, "\delta_"] \arrow[r, "\delta_"'] \arrow[ur, phantom, shift right=3, "="] & DX
\end{tikzcd}
\qquad
\begin{tikzcd}
P\Gamma \times (PX)^{\Gamma} \arrow[r, dashed, "\oint"] \arrow[d, hook, "\mathrm{cast} \times (\mathrm{cast} \circ)"'] \arrow[dr, phantom, "="] & PX \arrow[d, hook, "\mathrm{cast}"] \\
D\Gamma \times (DX)^{\Gamma} \arrow[r, "\oint"'] & DX
\end{tikzcd}$$

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

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

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

$\mu : DX \vdash \oint \mu(dx)\delta_{x} = \mu$ $: DX$

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

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

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

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

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

↳ measure μ assigns to E

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

Foundations for type-driven probabilistic modelling

Ohad Kammar
University of Edinburgh

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

THE UNIVERSITY *of* EDINBURGH informatics lfcs Laboratory for Foundations of Computer Science BayesCentre

supported by: THE ROYAL SOCIETY The Alan Turing Institute Facebook Research NCSC

Languge of distribution & Probability

Recap

X type (= space) of values / outcomes

DX type of distributions / measures over X

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

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

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

type judgment

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

↳ measure μ assigns to E

Axioms for measures/distributions

Recap

$$\mu : DX \vdash \underset{\mu}{\mathrm{Ce}}[\emptyset] = 0 \quad : W$$

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

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

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

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

Kernels & their Kock integral

Recap

Kernel from Γ to X: $k : (DX)^{\Gamma}$ or $k : \Gamma \longrightarrow DX$

Dirac kernel: $\delta_{\_} : X \longrightarrow DX$

Kock integral: $\mu : D\Gamma,\ k : (DX)^{\Gamma} \vdash \oint \mu k \quad : DX$

or $\oint \mu(dx)\, k(x)$ (dx binding occurrence in $k(x)$)

Giry monads: $(D, \delta_{\_}, \oint)$ & $(P, \delta_{\_}, \oint)$.

Discrete model

type : Set $\qquad W := [0,\infty] \qquad \mathcal{B}X := \mathcal{P}X$

$$DX := \{ \mu : X \to W \mid \mathrm{supp}\,\mu \text{ countable} \}$$

$$PX := \{ \mu \in DX \mid \underset{\mu}{\mathrm{Ce}}[X] = 1 \}$$

$$\underset{\mu}{\mathrm{Ce}}[E] := \sum_{x \in E} \mu x \qquad \delta_x := \lambda x'. \begin{cases} x = x' : & 0 \\ x \neq x' : & 1 \end{cases}$$

$$\oint \mu k := \lambda x. \sum_{r \in \Gamma} \mu r \cdot k(r; x)$$

Ex distributions

Counting measure (X ctbl): $\#_X := \lambda x. 1$

Dirac measure δ_x (prev slide)

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

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

Product measures

$$\mu : DX, \nu : DY \vdash \mu \otimes \nu := \oint \mu(\mathrm{d}x) \oint \nu(\mathrm{d}y) \delta_{\langle x, y \rangle} : D(X \times Y)$$

$$= \lambda(x, y).\ \mu x \cdot \nu y$$

$\uparrow$ discrete model

($\otimes$ lifts along $P \hookrightarrow D$)

Ex: $\#_{X \times Y} = \#_X \otimes \#_Y$

build measures compositionally

Indeed:

$$(\# \otimes \#)(x, y) = \# x \cdot \# y = 1 \cdot 1 = 1 = \#(x, y)$$

Notation: $\lambda : D(X \times Y),\ k : (DZ)^{X \times Y} \vdash \oint\!\!\oint \lambda(dx, dy)\, k(x, y)$
$:= \oint \lambda k$

Fubini-Tonelli Thm:

Integrate in any order:

$\mu : DX,\ \nu : DY,\ k : (DZ)^{X \times Y} \vdash$

$$\oint \mu(dx) \oint \nu(dy)\, k(x, y) = \oint\!\!\oint (\mu \otimes \nu)(dx, dy)$$

$$= \oint \nu(dy) \oint \mu(dx)\, k(x, y)$$

Pushing a measure forward

$\mu : D\Omega,\ \alpha : X^{\Omega} \vdash \mu_f := \oint \mu(d\omega)\,\delta_{\alpha\omega} : DX$

$$= \lambda x. \sum_{\substack{\omega \in \Omega \\ \alpha\omega = x}} \mu\,\omega$$

$\alpha : X^{\Omega}$: random element (w.r.t. μ)

$\mu_\alpha : DX$: the law of α

Ex: We can represent configurations of 2 dice

using $\underline{6}\times\underline{6}$

Letting $(+) : \underline{6}^2 \longrightarrow \mathbb{N}^2 \xrightarrow{(+)} \mathbb{N}$

we have that the law of $(+)$:

$$(\#_{\underline{6}} \otimes \#_{\underline{6}})_{(+)} : D\mathbb{N}$$

is the number of rolls whose sum is given

build measures compositionally

Scaling a measure

$$(\cdot) : W \times DX \longrightarrow DX$$

$$a \cdot \mu := \lambda x.\ a \cdot \mu x$$

NB: $\mathrm{supp}(a \cdot \mu) = \begin{cases} a = 0 : & \emptyset \\ a \neq 0 : & \mathrm{supp}\,\mu \end{cases}$

✓ ctbl

$(\cdot) : W \times DX \to DX$ is an action of monoid $(W, (\cdot), 1)$ on DX:

$\mu : DX \vdash$

$$1 \cdot \mu = \mu$$

$a, b : W, \mu : DX \vdash$

$$a \cdot (b \cdot \mu) = (a \cdot b) \cdot \mu$$

Normalisation

unit type $\to$ $\mathbb{1} := \{\langle\rangle\}$ $\leftarrow$ unit value

$$\mu : DX,\quad Ce_{\mu}[X] \neq 0, \infty \vdash$$

$$\|\mu\| := \left(\frac{1}{Ce_{\mu}[X]}\right) \cdot \mu \quad : PX$$

Ex:

$$\emptyset \neq A \subseteq_{fin} X \quad : \mathcal{U}_{A \subseteq X} := \|(\#_A)_{A\subseteq X}\| : PX$$

$$\mathbb{1} \xrightarrow{\#_A} DA \xrightarrow{(-)_{A\subseteq X}} DX \xrightarrow{\|-\|} PX$$

I.e.

$$\mathcal{U}_{A\subseteq X} := \lambda x. \begin{cases} x \in A : & \frac{1}{|A|} \\ x \notin A : & 0 \end{cases} \quad \text{so} \quad \mathcal{U}_{\{x\}\subseteq X} = \delta_x$$

Standard vocabulary

Joint distributions: $\mu : D(X_1 \times X_2)$

Marginal distribution: $X_1 \xleftarrow{\pi_1} X_1 \times X_2 \xrightarrow{\pi_2} X_2$

law of projection

$\mu_{\pi_i} : DX_i$

marginalisation: $\mu_{\pi_1} = \iint \mu(dx, dy)\, \delta_x$

integrate out y

Exercise: $\mu : PX,\ \nu : DX \vdash (\mu \otimes \nu)_{\pi_2} = \nu$

independence

Pairing r.e.s:

$$\alpha : X^{\Omega}, \beta : Y^{\Omega} \vdash$$

$$\langle \alpha, \beta \rangle := \lambda \omega . \langle \alpha \omega, \beta \omega \rangle : (X \times Y)^{\Omega}$$

$$\begin{tikzcd}
& \Omega \arrow[dl, "\alpha"'] \arrow[d, dashed, "\langle \alpha, \beta \rangle"] \arrow[dr, "\beta"] & \\
X & X \times Y \arrow[l, "\pi_1"] \arrow[r, "\pi_2"'] & Y
\end{tikzcd}$$

$$\lambda : D\Omega, \alpha : X^{\Omega}, \beta : Y^{\Omega} \vdash \alpha \perp_{\lambda} \beta := \lambda_{\langle \alpha, \beta \rangle} = \lambda_{\alpha} \otimes \lambda_{\beta}$$

$: \mathrm{Prop}$

α, β independent w.r.t. λ

Ex (Durrett) represent outcomes of 3 coin tosses:

$$C := \{T, H\} \quad \Omega := C \times C \times C \quad \lambda : U_C \otimes U_C \otimes U_C : P\Omega$$

$$\pi_i : \Omega \longrightarrow C \quad \text{outcome of } i^{th} \text{ toss}$$

$$\mathrm{Same}_{ij} : \Omega \xrightarrow{\langle \pi_i, \pi_j \rangle} C \times C \xrightarrow{(\overset{?}{=})} \mathbb{B}$$

where: $(\overset{?}{=}) : C^2 \longrightarrow \mathbb{B} := \{\mathrm{True}, \mathrm{False}\}$

$$x \overset{?}{=} y := \begin{cases} x = y : & \mathrm{True} \\ x \neq y : & \mathrm{False} \end{cases}$$

Ex (Durrett) represent outcomes of 3 coin tosses:

$C := \{T, H\}$ $\Omega := C \times C \times C$ $\lambda : U_C \otimes U_C \otimes U_C : P\Omega$

$\pi_i : \Omega \longrightarrow C$ outcome of i^{th} toss

$\mathrm{Same}_{ij} : \Omega \xrightarrow{\langle \pi_i, \pi_j \rangle} C \times C \xrightarrow{(\overset{?}{=})} \mathbb{B}$

$$\lambda_{\mathrm{Same}_{12}} T \overset{\text{marginalisation}}{=} (U_C \otimes U_C)_{(\overset{?}{=})} T = \underbrace{\frac{1}{4}}_{U_C(H)\cdot U_C(H)} + \overbrace{\frac{1}{4}}^{U_C(T)\cdot U_C(T)} = \frac{1}{2}$$

So $\lambda_{\mathrm{Same}_{12}} F = \frac{1}{2}$ too

$\underline{\text{Ex}}$ (Durrett) represent outcomes of 3 coin tosses:

$C := \{T, H\}$ $\quad \Omega := C \times C \times C$ $\quad \lambda : U_C \otimes U_C \otimes U_C : P\Omega$

$\pi_i : \Omega \longrightarrow C$ outcome of i^{th} toss

$$\mathrm{Same}_{ij} : \Omega \xrightarrow{\langle \pi_i, \pi_j \rangle} C \times C \xrightarrow{(\overset{?}{=})} \mathbb{B}$$

$\underline{i \neq j}$: $\quad \lambda_{\mathrm{Same}_{ij}} = U_{\mathbb{B}}$

$\lambda_{\langle \mathrm{Same}_{12}, \mathrm{Same}_{23} \rangle} :$

$$(T,T) \mapsto \underbrace{\tfrac{1}{2} \cdot \tfrac{1}{2} \cdot \tfrac{1}{2}}_{\lambda(H,H,H)} + \underbrace{\tfrac{1}{2} \cdot \tfrac{1}{2} \cdot \tfrac{1}{2}}_{\lambda(T,T,T)} = \frac{1}{4}$$

$$(T,F) \mapsto \underbrace{\tfrac{1}{2} \cdot \tfrac{1}{2} \cdot \tfrac{1}{2}}_{\lambda(H,H,T)} + \underbrace{\tfrac{1}{2} \cdot \tfrac{1}{2} \cdot \tfrac{1}{2}}_{\lambda(T,T,H)} = \frac{1}{4}$$

(Durrett)

Ex represent outcomes of 3 coin tosses:

$C := \{T, H\}$ $\quad \Omega := C \times C \times C$ $\quad \lambda : U_C \otimes U_C \otimes U_C : P\Omega$

$\pi_i : \Omega \longrightarrow C$ outcome of i^{th} toss

$\mathrm{Same}_{ij} : \Omega \xrightarrow{\langle \pi_i, \pi_j \rangle} C \times C \xrightarrow{(\overset{?}{=})} \mathbb{B}$

i≠j: $\lambda_{\mathrm{Same}_{ij}} = U_{\mathbb{B}}$

$$\lambda_{\langle \mathrm{Same}_{12}, \mathrm{Same}_{23} \rangle} = U_{\mathbb{B} \times \mathbb{B}} = U_{\mathbb{B}} \otimes U_{\mathbb{B}} = \lambda_{\mathrm{Same}_{12}} \otimes \lambda_{\mathrm{Same}_{13}}$$

So $\mathrm{Same}_{12} \perp_{\lambda} \mathrm{Same}_{13}$

independence

Pairing r.v.s:

$$\begin{tikzcd} & \Omega \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] \arrow[d, dashed, "\langle\alpha,\beta\rangle"] & \\ X & X \times Y \arrow[l, "\pi_1"] \arrow[r, "\pi_2"'] & Y \end{tikzcd}$$

$\alpha : X^\Omega, \beta : Y^\Omega \vdash$

$$\langle\alpha,\beta\rangle := \lambda\omega. \langle \alpha\omega, \beta\omega\rangle : (X \times Y)^\Omega$$

$$\lambda : D\Omega, \alpha : X^\Omega, \beta : Y^\Omega \vdash \alpha \perp_\lambda \beta := \lambda_{\langle\alpha,\beta\rangle} = \lambda_\alpha \otimes \lambda_\beta : \mathrm{Prop}$$

α, β independent w.r.t. λ

I-ary version:

$$\lambda : D\Omega, \alpha_- : \prod_{i \in I} X_i^\Omega \vdash \mathop{\perp}_{\lambda}{}_{i \in I} \alpha_i := \forall J \subseteq_{fin} I. \lambda_{\langle \alpha_j \rangle_{j \in J}} = \bigotimes_{j \in J} \lambda_{\alpha_j} : \mathrm{Prop}$$

α_- independent w.r.t. λ

Ex (Durrett) represent outcomes of 3 coin tosses:

$C := \{T, H\}$ $\quad \Omega := C \times C \times C$ $\quad \lambda : U_C \otimes U_C \otimes U_C : P\Omega$

$\pi_i : \Omega \longrightarrow C$ outcome of i^{th} toss

$\mathrm{Same}_{ij} : \Omega \xrightarrow{\langle \pi_i, \pi_j \rangle} C \times C \xrightarrow{(\overset{?}{=})} \mathbb{B}$

$i \neq j$: $\lambda_{\mathrm{Same}_{ij}} = U_{\mathbb{B}}$

$\underset{\substack{* \quad * \\ k}}{i \neq j}$: $\mathrm{Same}_{ij} \perp_{\lambda} \mathrm{Same}_{jk}$ $\qquad \not\perp_{\lambda} \{\mathrm{Same}_{12}, \mathrm{Same}_{23}, \mathrm{Same}_{13}\}$

Intuition: $\mathrm{Same}_{13} = \mathrm{IFF}(\mathrm{Same}_{12}, \mathrm{Same}_{23})$

Calc:

$$\lambda_{\langle \mathrm{Same}_{12}, \mathrm{Same}_{23}, \mathrm{Same}_{13} \rangle}(T,T,T) = \underbrace{\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}}_{\hookrightarrow \lambda(H,H,H)} + \underbrace{\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}}_{\hookrightarrow \lambda(T,T,T)} = \frac{1}{4} \neq \frac{1}{2^3} = \lambda_{\mathrm{Same}_{12}} \otimes \lambda_{\mathrm{Same}_{23}} \otimes \lambda_{\mathrm{Same}_{13}}$$

Vocabulary

(Discrete) Measure space $(X, \mu : DX)$

measure preserving $f : (X, \mu) \to (Y, \nu)$

function $f : X \to Y$ s.t. $\mu_f = \nu$

$\mu : DX, f : X \to Y \vdash \mu$ invariant under f :=

$$f : (X, \mu) \to (X, \mu)$$

Ex:

$\mu : DX, \nu : DY \vdash$

$\mathrm{swap} : (X \times Y, \mu \otimes \nu) \longrightarrow (Y \times X, \nu \otimes \mu)$ so

$\mu : DX \vdash \mu \otimes \mu$ invariant under swap

Weights as measures

NB: unit type $\to$ $\mathbb{1} := \{\langle\rangle\}$ $\leftarrow$ unit value

Observation:

$$\begin{tikzcd} D\mathbb{1} \arrow[r, bend left, "\lambda\mu.\, ce_{\mu}[\mathbb{1}]"] & W \arrow[l, bend left, "(-\odot\delta_{\langle\rangle})"] \end{tikzcd}$$

mutually inverse

Proof:

$$\mu \longmapsto \mu(\langle\rangle) \qquad \& \qquad \lambda\langle\rangle.a \longmapsfrom a$$

$$\mu = \lambda\langle\rangle.\mu\langle\rangle \qquad\qquad a = (\lambda\langle\rangle.a)\langle\rangle$$

$\blacksquare$

Integration

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

(Lebesgue integral)

$$:= \sum_{x \in X} \mu x \cdot \varphi x$$

Can derive it:

$$\begin{tikzcd} DX \times W^X \arrow[rr, "DX \times (\cong \circ -)"] \arrow[d, "\int"'] & & DX \times (D\mathbb{1})^X \arrow[d, "\oint"] \\ W & := & D\mathbb{1} \arrow[ll, "\cong"] \end{tikzcd}$$

Additivity:

$$I \text{ ctbl}, \mu_{-} : (DX)^I \vdash \sum_{i \in I} \mu_i : DX$$

$$:= \lambda x. \sum_{i \in I} \mu_i x$$

NB: $\operatorname{supp} \sum_i \mu_i \subseteq \bigcup_i \operatorname{supp} \mu_i$ ✓ctbl

Ex: Bernoulli distribution

$$p : [0,1] \vdash B(p) := p \cdot \delta_{\mathrm{True}} + (1-p) \cdot \delta_{\mathrm{False}} : P\mathbb{B}$$

i.e. Bp: $\mathrm{True} \mapsto p$
$\mathrm{False} \mapsto 1-p$

Thm (affine-linearity):

$\oint$ is affine-linear in each argument:

$$\frac{I \text{ ctbl} \quad \mu_{-} : (D\Gamma)^{I}, k : (DX)^{\Gamma}_{,} \quad a_{-} : \mathbb{W}^{I}}{} \vdash \oint\left(\sum_{i \in I} a_i \cdot \mu_i\right) k = \sum_{i \in I} a_i \cdot \oint \mu_i k$$

$$I \text{ ctbl}, \mu : D\Gamma, a_{-} : \mathbb{W}^{I}, k_{-} : DX^{I} \vdash$$

$$\int \mu(dx)\left(\sum_{i \in I} a_i \cdot k_i(x)\right) = \sum_{i \in I} a_i \cdot \oint \mu k_i$$

Prop: $\mathbb{W} \cong D\mathbb{1}$ is a σ-semi-ring isomorphism:

$$(\mathbb{W}, \Sigma, (\cdot), 1) \cong (D\mathbb{1}, \Sigma, (\cdot), \delta_{\langle\rangle})$$

and $(\cdot) : \mathbb{W} \times DX \to DX$ makes DX into a module:

$$\left(\sum_{i\in I} a_i\right)\cdot \mu = \sum_{i\in I}(a_i \cdot \mu) \qquad a \cdot \sum_{i\in I} \mu_i = \sum_{i\in I} a \cdot \mu_i$$

Corollary: $\int$ is affine-linear in each argument.

Random variable :

NB : $\overline{\mathbb{R}} := [-\infty, \infty]$

A random element $\alpha : \overline{\mathbb{R}}^{\Omega}$ (wrt some $\mu : D\Omega$)

Can add, multiply r.v.'s.

To integrate r.v.'s:

$$(-)^{+} : \overline{\mathbb{R}}^{\Omega} \longrightarrow W^{\Omega}$$

$$\alpha^{+} := \lambda w. \begin{cases} \alpha w \geq 0 : & \alpha w \\ \text{o.w} : & 0 \end{cases} = [\alpha - \geq 0] \cdot |\alpha|$$

$$\alpha^{-} := \lambda w. \begin{cases} \alpha w \leq 0 : & |\alpha w| \\ \text{o.w.} : & 0 \end{cases} = [\alpha - \leq 0] \cdot |\alpha|$$

So $\alpha = \alpha^{+} - \alpha^{-}$

$\mu : D\Omega, \alpha : \overline{\mathbb{R}}^{\Omega}, \int \mu\alpha^+ < \infty$ or $\int \mu \alpha^- < \infty \vdash$

$$\int \mu \alpha := \int \mu \alpha^+ - \int \mu \alpha^- \quad : \overline{\mathbb{R}}$$

<u>Ex.</u> The (discrete) Lebesgue p-space:

$$p : [1, \infty), \mu : P\Omega \vdash \mathcal{L}_p(\Omega, \mu) :=$$

$$\left\{ \alpha : \overline{\mathbb{R}}^{\Omega} \;\middle|\; \mathbb{E}_{\mu}[|\alpha|^p] < \infty \right\}$$

$\mathcal{L}_p(\Omega, \mu)$ has a norm $\|\alpha\| := \sqrt[p]{\mathbb{E}_{\mu}[|\alpha|^p]}$ almost Banach

$\mathcal{L}_2(\Omega, \mu)$ has an inner product $\langle \alpha, \beta \rangle := \mathbb{E}_{\mu}[\alpha \cdot \beta]$ almost Hilbert

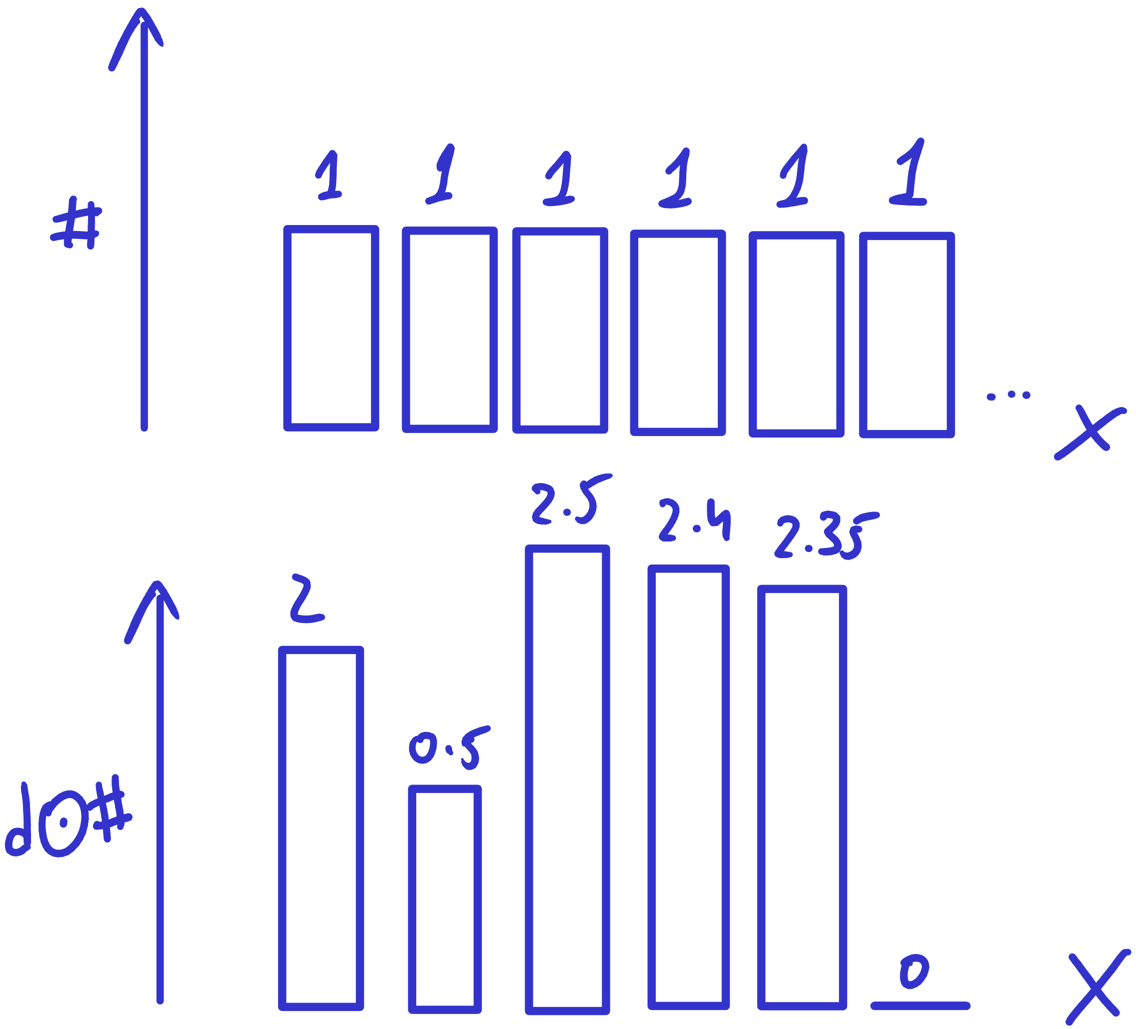
Density

a density over X : $d : X \to W$

$$d : W^X, \mu : DX \vdash d \odot \mu \quad : DX$$
$$:= \oint \mu(dx)(d\,x \cdot \delta_x)$$

Warning The types of measures & densities in the discrete model are close, but still different. They coincide on countable sets, so people often confuse them. Types help us keep them separate.

Intuition:



Almost certain properties

$$E : \mathcal{B}X, \mu : DX \vdash \mu(dx)\text{-almost certainly } x \in E \quad : \mathrm{Prop}$$

$$:= [- \in E] \odot \mu = \mu$$

NB: $[x \in E] = \begin{cases} x \in E : & 1 \\ x \notin E : & 0 \end{cases} \quad : \mathbb{W}$

When $\mu : PX$ we say instead

$$\mu(dx)\text{-almost surely } x \in E$$

<u>Exercise</u> Look up the def. of a normed space and modify the definition so that $\mathcal{L}_p(\Omega, \mu)$ is a normed space up-to almost sure equality.

d is a density of μ w.r.t. ν or

d is a Radon-Nikodym derivative w.r.t. ν

Absolute continuity

$\mu, \nu : DX,\ d : W^X \vdash\ d = \frac{d\mu}{d\nu}\ : \mathrm{Prop}$

$:= \mu = d \odot \nu$

$\mu, \nu : DX \vdash\ \mu \ll \nu :=$ μ is absolutely continuous w.r.t. ν $: \mathrm{Prop}$

$:= \exists d : W^X.\ d = \frac{d\mu}{d\nu}.$

$=:$ μ has a density w.r.t. ν

Lemma: $\mu, \nu : DX,$ $\mu \ll \nu,$ $k : (DY)^X$

$$\oint \nu(dx)\, \frac{d\mu}{d\nu}(x) \cdot k x = \oint \mu(dx)\, k x$$

Ex: $U_{A \subseteq X} \ll (\#_A)_{\mathrm{Cust}} : A \subseteq X$

$$\frac{dU_{A \subseteq X}}{d(\#_A)_{\mathrm{Cust}}} = \lambda x. \begin{cases} x \in A: & \frac{1}{|A|} \\ \text{o.w.}: & 0 \end{cases}$$

but also:

$$\frac{dU_{A \subseteq X}}{d(\#_A)_{\mathrm{Cust}}} = \lambda x. \frac{1}{|A|}$$

Radon-Nikodym Thm: (discrete version)

$\mu, \nu : PX \vdash \mu \ll \nu$ iff $\forall x.\ \nu x = 0 \Rightarrow \mu x = 0$

i.e. $\mathrm{supp}\,\mu \subseteq \mathrm{supp}\,\nu$

In that case, if $d_1, d_2 = \frac{d\mu}{d\nu}$ then

$\nu(dx)$-a.s. $d_1 x = d_2 x$

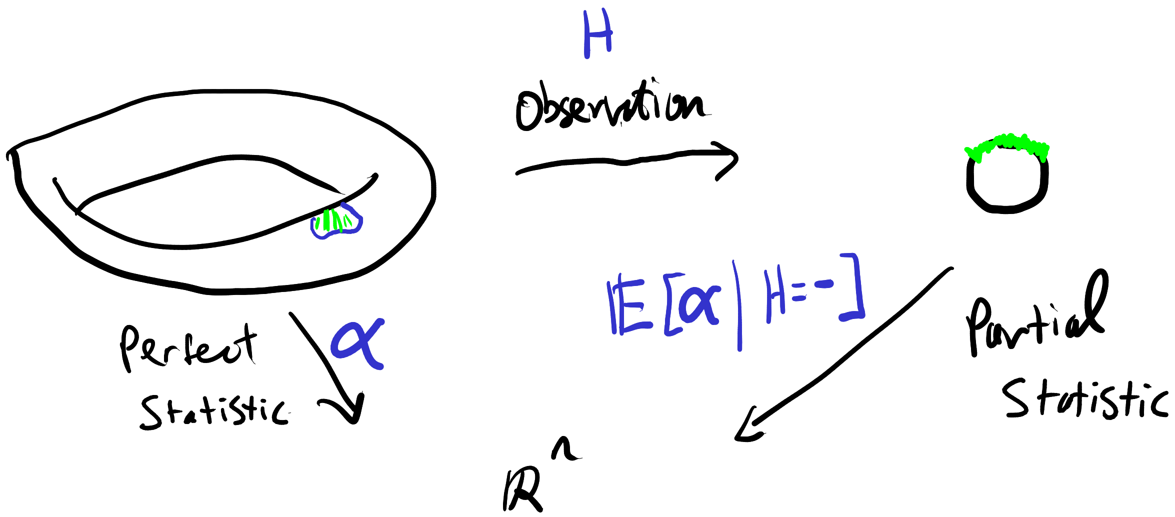
<u>Ex</u>: for ctbl X, $\forall \mu : DX.\ \mu \ll \#_X$. Proof: vacuously, as $\#_X x \neq 0$.

Then $\lambda x.\ \mu x = \frac{d\mu}{d\#}$.

Conditional expectation

β is a conditional expectation of α w.r.t. μ along H

$$\mu : D\Omega,\ H : X^{\Omega},\ \alpha : \mathcal{L}_1(\Omega, \mu),\ \beta : \mathcal{L}_1(X, \mu_H)$$
$$\vdash \beta = \mathbb{E}[\alpha \mid H = -] \quad : \mathrm{Prop}$$
$$:= \forall \varphi : \mathcal{L}_1(Y, \mu_H^{M}).\ \int \mu_H(dx)\, \beta(x) \cdot \varphi(x) = \int \mu(d\omega)\, \alpha(\omega) \cdot \varphi(H\omega)$$



Thm (Kolmogorov): (discrete version)

There is a funtion

$$\mathbb{E}_{-}[- \mid - = -] \in \prod_{\mu : P\Omega} \prod_{H : X^{\Omega}} \mathcal{L}_1(\Omega, \mu) \to \mathcal{L}_1(X, \mu_H)$$

s.t. $\mathbb{E}_{\mu}[\alpha \mid H = -]$ is a conditional expectation of α w.r.t. μ along H.

Conditional Probability (discrete version):

$$H : X^{\Omega},\ \mu : PX \vdash \mathbb{P}_{\mu}\left[- \mid H = -\right] : (P\Omega)^{X}$$

$$:= \lambda x_0 : X. \lambda \omega_0 : \Omega. \mathbb{E}_{\omega \sim \mu}\left[[\omega_0 = \omega] \mid H\omega = x_0 \right]$$

Bayes's Thm (discrete version, adapted from Williams):

$$\begin{tikzcd} & X \times \Theta \arrow[dl, "\alpha := \pi_1"'] \arrow[dr, "H := \pi_2"] & \\ X & & \Theta \end{tikzcd}$$

observation space.

Let $\lambda : P(X \times \Theta)$ joint probability distribution.

Assume $\mu : DX$, $\nu : D\Theta$ s.t. $\lambda \ll \mu \otimes \nu$.

with $d_{X,H} = \dfrac{d\lambda}{d(\mu \otimes \nu)}$.

Obs 1: $d_X : W^X$

$d_X := \lambda x. \int \nu(d\theta)\, d_{X,H}(x,\theta)$ then $d_X = \dfrac{d\lambda_\alpha}{d\mu}$

& similarly $(d_H : W^{\Theta}) := \lambda\theta. \int \mu(dx)\, d_{X,H}(x,\theta) = \dfrac{d\lambda_H}{d\nu}$

Bayes's Thm (discrete version, adapted from Williams):

Let $\lambda : P(X \times \Theta)$ joint probability distribution.

Assume $\mu : DX$, $\nu : D\Theta$ s.t. $\lambda \ll \mu \otimes \nu$.

with $d_{X,H} = \dfrac{d\lambda}{d(\mu \otimes \nu)}$. $d_X = \dfrac{d\lambda_\alpha}{d\mu}$ $d_\Theta = \dfrac{d\lambda_H}{d\nu}$

$$\begin{array}{c} X \times \Theta \\ \alpha := \pi_1 \swarrow \quad \searrow H := \pi_2 \\ X \qquad \Theta \end{array}$$

observation space.

Let $d_{X|H}(-|-) : X \times \Theta \to W$

$$d_{X|H}(x|\theta) := \begin{cases} d_H\theta \neq 0 : & \dfrac{d_{X,H}(x,\theta)}{d_H\theta} \\ \text{o.w.} : & 0 \end{cases}$$

$\lambda_{X|H=-} : \Theta \to PX$

$\lambda_{X|H=\theta} := d_{X|H}(-|\theta) \odot \mu$

Bayes's formula:

$\mathbb{P}r_\lambda[-|H=-] = \lambda_{X|H=-}$

Summary

$\mu \otimes \nu$ Product measures & Fubini-Tonelli

μ_H push-forward / law

$(Dx, \Sigma, (\cdot))$ module structure and affine linearity of $\oint$

$\int$ Lebesgue integration

Standard vocabulary: joint dist., marginalisation, independence, invariance

density & Radon-Nikodym derivatives (heed the warning)

almost certain properties

Conditional expectation & Probability

with Bayes's Thm.

Plan:

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

2) Borel sets & measurable spaces (~~Tue~~ Wed)

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