Foundations for type-driven probabilistic modelling

Ohad Kammar
University of Edinburgh

TutorialFest
ACM-SIGPLAN 51st Symposium on
Principles of Programming Languages (POPL'24)
15 January, 2024
City of Westminister, England

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

Everything POPL

- LAFI
- some POPL
- this tutorial

Plan

1) Typed setting for Probability & Statistics

2) 2 Implementations

Discrete Model $\hookrightarrow$ Full Model

familiar maths to introduce language

Same langugage

Why foundations?

Discrete Probability

countably supported measures

Continuous Probability

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

Well-behaved Probability

s-finite measures over standard Borel spaces

Measurable spaces

standard, established

Poor type structure

Quasi-Borel spaces

new, experimental

rich type structure

(this course)

Takeaway:

use types to develop smoothly

Why types?

- spotlights meaningful operations

$$\int : (\mathrm{Distribution}\ A) \times (A \to [0,\infty]) \to [0,\infty]$$

- documents intent

Probability Distribution A vs Density $A \to [0,\infty]$

- succinctness: easier to elaborate details
- esp. formal types: use theory without fully understanding it.

Plan:

1) type-driven probability: discrete case
2) Borel sets & measurable spaces
3) Quasi Borel spaces

Lecture 1

4) Type structure & standard Borel spaces
5) Integration & random variables

Lecture 2

Please ask questions!

Course web page

scribble

Advertisement

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

Check out our PhD Programs:

THE UNIVERSITY *of* EDINBURGH

informatics

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 measue 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)

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

Axioms for measures

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

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

Axiom for Probability

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

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

Avoid costing:

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

Axioms for measures

Integration:

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

Again, avoid casting:

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

(Expectation)

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

Have: language + axioms

Want: model

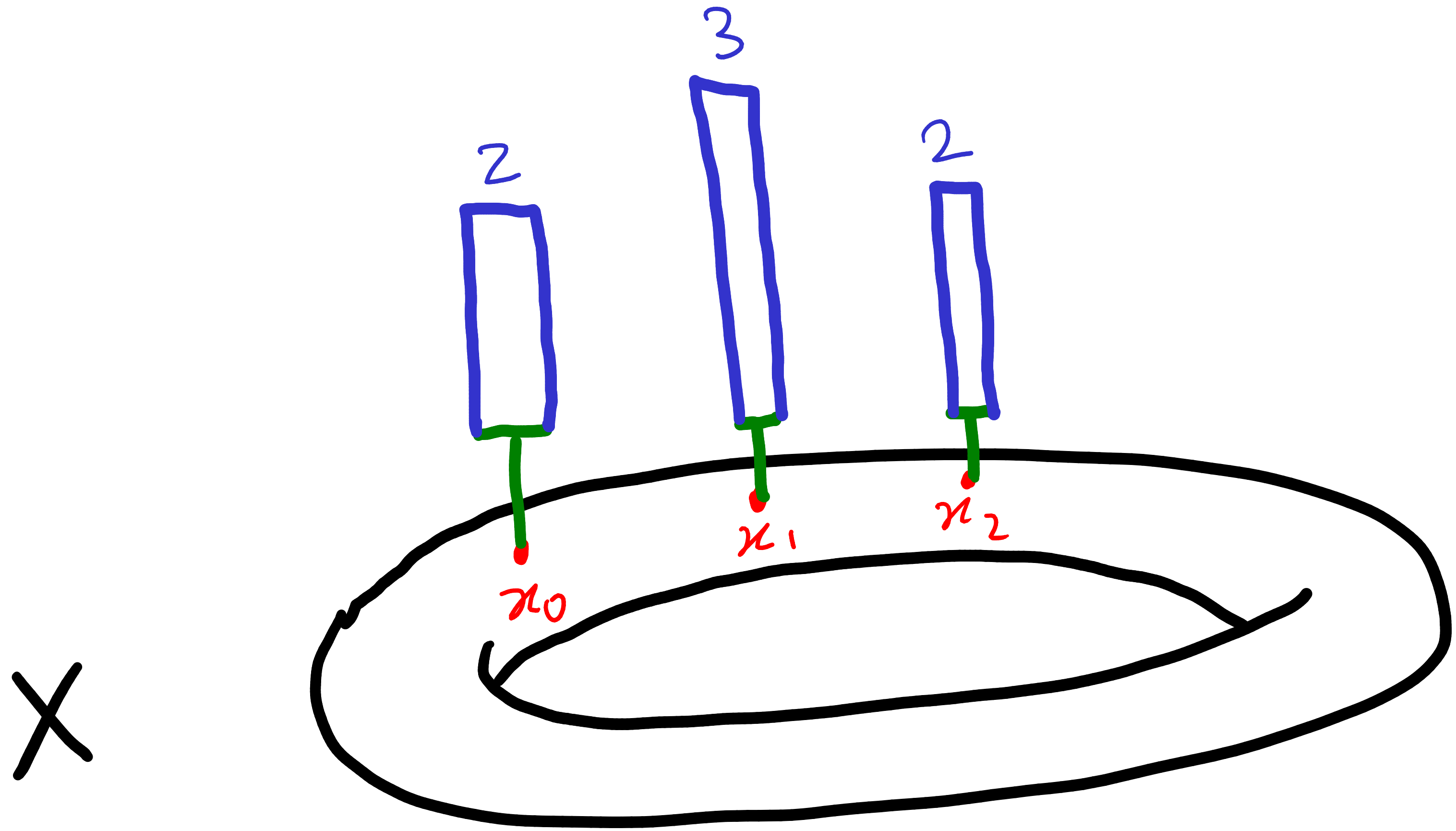
Part 1: discrete measures

Part 2: discrete + continuous

Discrete model

type X : set

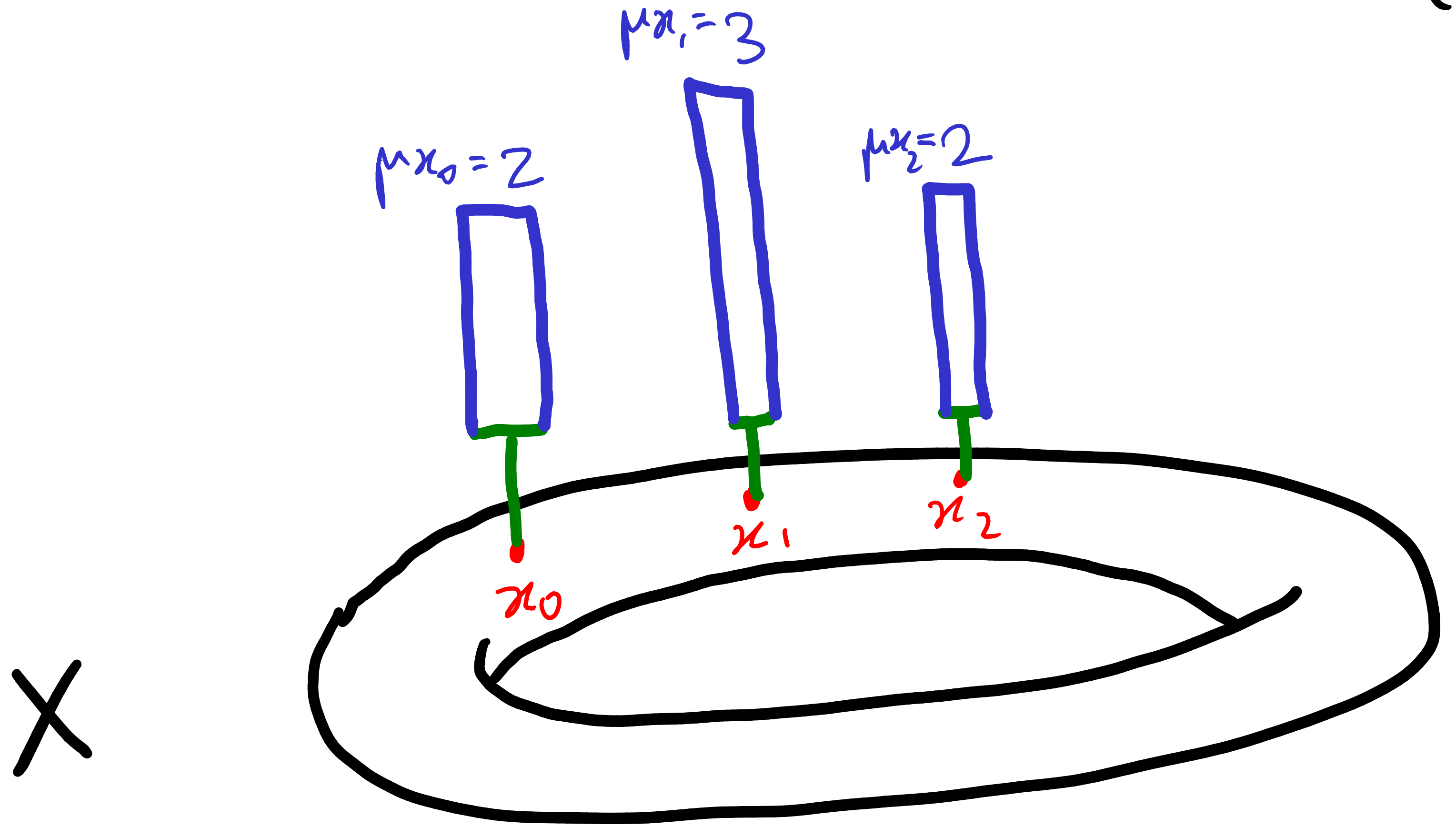
$DX :=$ histograms



Discrete model

type X : set

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



Support

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

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

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

supp μ is the smallest set supporting μ

Discrete model

type X : set

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

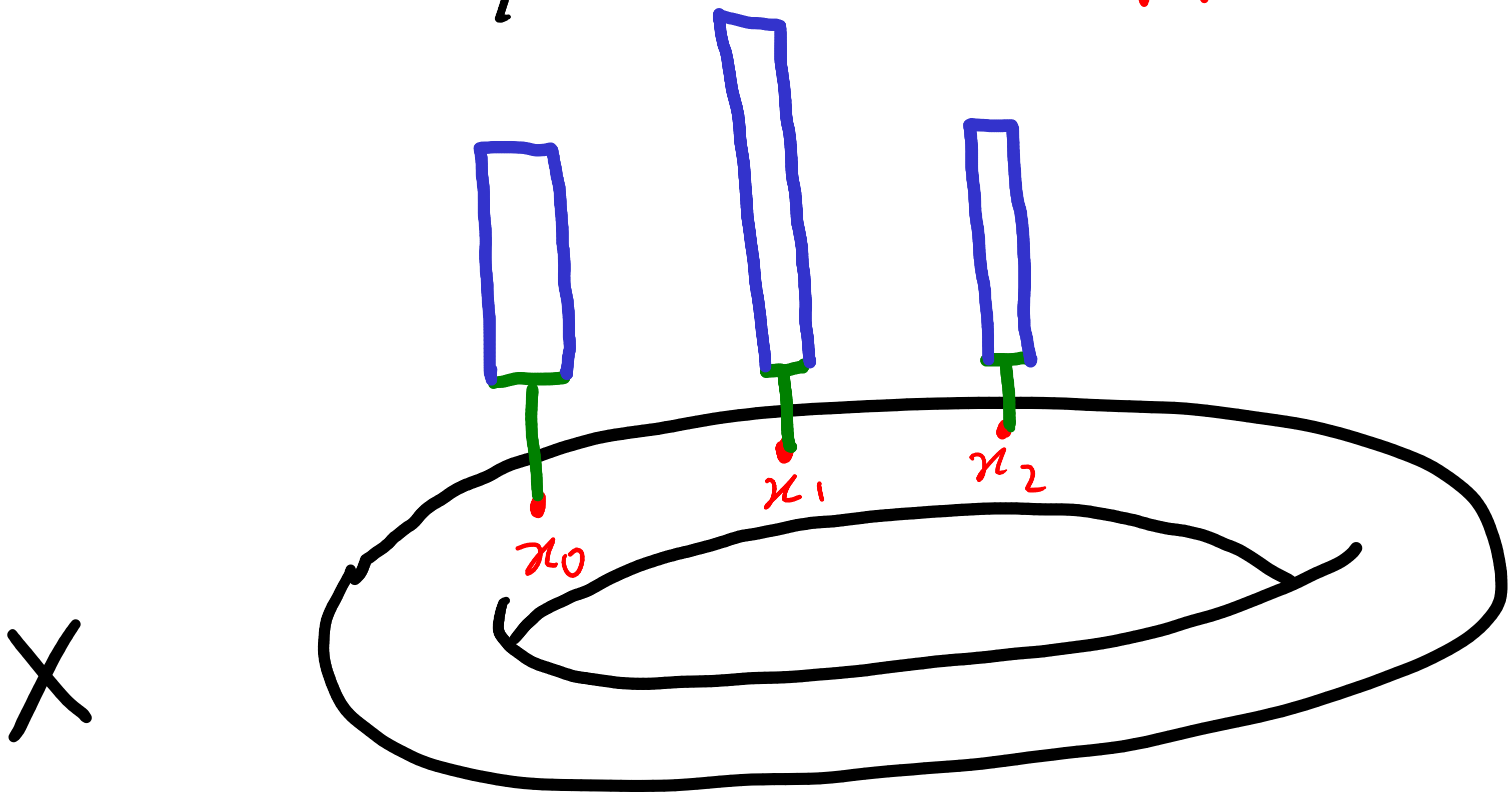
$$:= \{\mu : X \to \mathbb{W} \mid \operatorname{supp} \mu \text{ is countable}\}$$

Discrete model

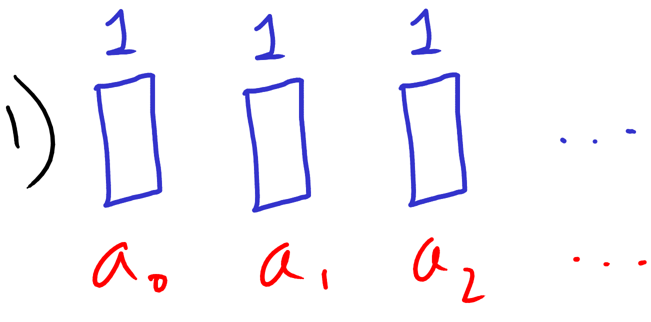
type X : set

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

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



Ex. measures

1) 

. . . Counting measure

. . . (X ctbl)

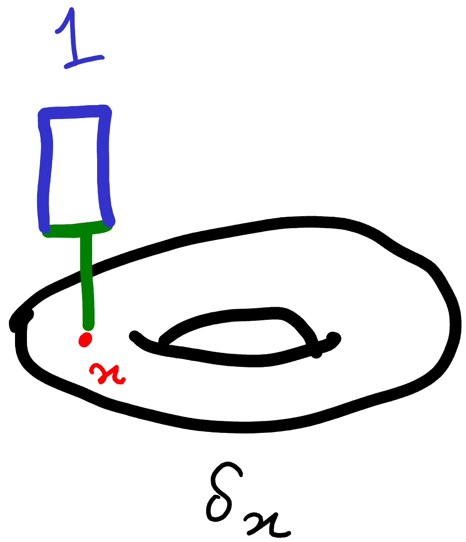
$\#_X : DX$

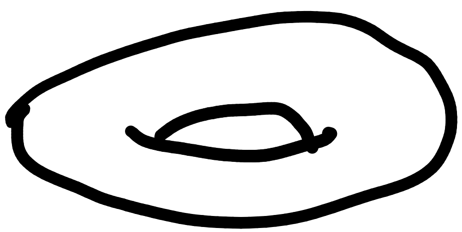
$\#_X := \lambda x : X . 1$

2) Dirac measure:

$x : X \vdash \delta_x : DX$

$$:= \lambda x' . \begin{cases} x = x' : & 1 \\ o.w. : & 0 \end{cases}$$



3) 

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

Discrete model

type X : set

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

$BX := \mathcal{P}X$

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

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

Ex:

- Counting measure counts outcomes:

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

- Dirac measure detects outcomes:

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

- Zero measure is zero:

$$E : BX \vdash \underset{\underline{0}}{\mathcal{C}}[E] = 0$$

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

indicator function

Validate axioms

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

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

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

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

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

Kernels k from Γ to X:

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

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:

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

(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:

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

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

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

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

check:

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

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

Probability measures

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

Probability measures

$$P X := \left\{ \mu : D X \;\middle|\; \oint_{\mu} [X] = 1 \right\} \overset{\subseteq}{\hookrightarrow} D X$$

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

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

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

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

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

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

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

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

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

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 judgment

$$\mu : DX, E : BX \vdash \mathcal{C}e_{\mu}[E] : W$$

↳ measure μ assigns to E

Plan:

1) type-driven probability: discrete case
2) Borel sets & measurable spaces
3) Quasi Borel spaces

Lecture 1

4) Type structure & standard Borel spaces
5) Integration & random variables

Lecture 2

Please ask questions!

Course web page

scribble