Foundations for Type-Driven Probabilistic Modelling

Ohad Kammar
University of Edinburgh

Laboratory for Software Science
9 December, 2025
Chair of Programming Languages
Institute of Computer Science
University of Tartu

supported by:

Computational golden era

logic-rich & type-rich computation

statistical computation

Computational golden era

logic-rich & type-rich computation

- Expressive type systems: Haskell, OCaml, Rust, Agda, Idris
- Mechanised mathematics: Agda, Rocq, Isabelle/HOL, Lean
- Verification: SMT-powered real-world systems

statistical computation

Generative modelling with efficient inference: Monte-Carlo simulation or gradient-based optimisation

This course

Typed interface to probability/statistics

Every concept has:

- a type
- associated operations
- properties in terms of these operations.

course page

Two implementations/models

discrete model $\hookrightarrow$ **full model**

discrete model
familiar maths
introductory

full model
supports discrete
and
continuous distributions
same language

Motivation: why foundations?

discrete probability
countably supported distributions
good type-structure
(this course)

measure theory
standard, established
poor type-structure

↘ **well-behaved probability** ↗
s-finite distributions
over standard Borel spaces

continuous probability ↗
Lebesgue measure over $\mathbb{R}^n$

↘ **quasi-Borel spaces**
new, experimental
rich type-structure
(this course)

Takeaway

Use types to abstract away from the model

Motivation: why types?

- **spotlights** meaningful operations

$$\int : (\mathsf{Distribution}\,X) \times (\mathsf{RandomVariable}\,X) \to [0,\infty]$$

- document **intent**:

  probability $(\mathsf{Distribution}\,X)$ vs. density $(X \to [0,\infty])$ vs. random variable

- succinctness: omit and elaborate details
- especially **formal** types, allow using theory correctly without fully understanding it

Lecture plan

Lecture 1: discrete model (today)

- Language of probability and distribution
- Discrete model
- Simply-typed probability
- Dependently-typed probability

Lecture 2: the full model

- Borel sets and measurable spaces
- Quasi-Borel spaces
- Type structure & standard Borel spaces
- Integration & random variables

course page

ask questions on the
Scottish PL Institute
Zulip stream #qbs

Language of **probability** & **distribution**

X	type (=space) of **values**/**outcomes**
$\mathsf{D}X$	type of **distributions**/**measures** over X
$\mathsf{P}X \subseteq \mathsf{D}X$	sub-type of **probability distributions** over X
$\mathcal{B}_X \subseteq \mathcal{P}X$	type of **events**: subsets we wish to measure
$\mathbb{W}$	type of **weights**: values in $[0,\infty]$
$\int, \mathbb{E}$	Lebesgue integration and the expectation operation

Type judgements describe well-formed values/outcomes of a given type, e.g.:

$$\mu : \mathsf{D}X, E : \mathcal{B}_X \vdash \underset{\mu}{\mathrm{Ce}}[E] : \mathbb{W}$$

(measures weight $\mathrm{Ce}_\mu[E]$ of event E according to distribution μ)

Propositions describe properties of well-formed values/outcomes of a given type, e.g.:

$$y_1, y_2 : Y \vdash y_1 \overset{Y}{=} y_2 : \mathsf{Prop} \qquad \mu : \mathsf{P}X, E : \mathcal{B}_X \vdash \underset{\mu}{\Pr}[E] = \underset{\mu}{\mathrm{Ce}}[E]$$

(probability of event according to probability distribution is its measure)

Axioms for events and distributions

Empty event

$$\emptyset : \mathcal{B}_X$$

Empty events weight zero

$$\mu : \mathsf{D}X \vdash \underset{\mu}{\mathrm{Ce}}\,[\emptyset] = 0$$

Axioms for events and distributions

Boolean Sub-algebra of Events

$E : \mathcal{B}_X \vdash E^{\complement} : \mathcal{B}_X \qquad E, F : \mathcal{B}_X \vdash E \cap F : \mathcal{B}_X$ so also: $E, F : \mathcal{B}_X \vdash X, E \cup F : \mathcal{B}_X$

Disjoint additivity

$w, v : \mathbb{W} \vdash w + v : \mathbb{W} \qquad E, C : \mathcal{B}_X, \mu : \mathsf{D}X \vdash \underset{\mu}{\mathrm{Ce}}[E] = \underset{\mu}{\mathrm{Ce}}[E \cap C] + \underset{\mu}{\mathrm{Ce}}\left[E \cap C^{\complement}\right]$

Axioms for events and distributions

Boolean Sub-algebra of Events

$E : \mathcal{B}_X \vdash E^{\complement} : \mathcal{B}_X \qquad E, F : \mathcal{B}_X \vdash E \cap F : \mathcal{B}_X$ so also: $E, F : \mathcal{B}_X \vdash X, E \cup F : \mathcal{B}_X$

Disjoint additivity

$w, v : \mathbb{W} \vdash w + v : \mathbb{W} \qquad E, C : \mathcal{B}_X, \mu : \mathrm{D}X \vdash \underset{\mu}{\mathrm{Ce}}[E] = \underset{\mu}{\mathrm{Ce}}[E \cap C] + \underset{\mu}{\mathrm{Ce}}\left[E \cap C^{\complement}\right]$

Exercise

Derive ‘axiomatically’ that:

- measurement is **monotone**:

$$\mu : \mathrm{D}X, E \subseteq F \vdash \underset{\mu}{\mathrm{Ce}}[E] \leq \underset{\mu}{\mathrm{Ce}}[F]$$

- the **inclusion-exclusion** principle:

$$\mu : \mathrm{D}X, E, F : \mathcal{B}_X \vdash \underset{\mu}{\mathrm{Ce}}[E \cup F] + \underset{\mu}{\mathrm{Ce}}[E \cup F] = \underset{\mu}{\mathrm{Ce}}[E] + \underset{\mu}{\mathrm{Ce}}[F]$$

Axioms for events and distributions

Consider posets:

$$\omega \coloneqq (\mathbb{N}, \leq) \qquad (\mathcal{B}_X, \subseteq) \qquad (\mathbb{W}, \leq)$$

ω-chains in a poset $P = (\underline{P}, \leq)$:

$$P^\omega \coloneqq \left\{ p_{\_} \in \underline{P}^{\mathbb{N}} \middle| p_0 \leq p_1 \leq \cdots \right\}$$

Chain-closure of events and weights

$$E_{\_} : (\mathcal{B}_X, \subseteq)^\omega \vdash \bigcup_n E_n : \mathcal{B}_X \qquad w_{\_} : (\mathbb{W}, \leq)^\omega \vdash \sup_n w_n : \mathbb{W}$$

Scott-continuity of measurement

$$E_{\_} : (\mathcal{B}_X, \subseteq)^\omega, \mu : \mathrm{D}X \vdash \mathrm{Ce}_\mu \left[\bigcup_n E_n\right] = \sup_n \mathrm{Ce}_\mu [E_n]$$

Axiom for probability

Probability distributions have total mass one

$$\mathsf{P}X := \{\mu \in \mathsf{D}X \mid \mathrm{Ce}_{\mu}[X] = 1\} \qquad \mu : \mathsf{P}X \vdash \mathsf{cast}\,\mu : \mathsf{D}X$$

i.e., if we define:

$$\mathbb{I} := [0,1] \qquad \mu : \mathsf{P}X, E : \mathcal{B}_X \vdash \Pr_{\mu}[E] := \mathop{\mathrm{Ce}}_{\mathsf{cast}\,\mu}[E] : \mathbb{I}$$

then:

$$\mu : \mathsf{P}X \vdash \Pr_{\mu}[X] = 1$$

Integration

Lebesgue integration w.r.t. a distribution

$$\mu : \mathsf{D}X, f : \mathbb{W}^X \vdash \int \mu(\mathrm{d}x) f(x) : \mathbb{W}$$

(NB: We succinctly write $\mathbb{W}^X$ for the type of functions $X \to \mathbb{W}$.)

Expectation w.r.t. a probability distribution

$$\mu : \mathsf{P}X, f : \mathbb{W}^X \vdash \mathbb{E}_{x\sim\mu}\left[f(x)\right] \coloneqq \int (\mathsf{cast}\,\mu)(\mathrm{d}x) f(x) : \mathbb{W}$$

We'll use variations on this notation, e.g.:

$$\int \mathrm{d}\mu f, \int f \mathrm{d}\mu, \int f(x)\mu(\mathrm{d}x), \mathbb{E}_\mu\left[f\right]$$

Summary

Have: Language and (some) axioms

Want: Model

Today: **discrete** model

Next week: **full** model

Lecture plan

Lecture 1: discrete model (today)

- Language of probability and distribution
- Discrete model
- Simply-typed probability
- Dependently-typed probability

Lecture 2: the full model

- Borel sets and measurable spaces
- Quasi-Borel spaces
- Type structure & standard Borel spaces
- Integration & random variables

course page

ask questions on the Scottish PL Institute Zulip stream #qbs

Discrete model

X: types denote **sets**

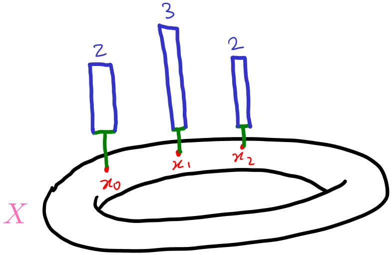
$\mathrm{D}X$: set of **histograms**:

Discrete model

X: types denote **sets**

$\mathrm{D}X$: set of **histograms**:

$$\mathrm{D}X \coloneqq \{\mu : X \to \mathbb{W} | \mu \text{ is } \textbf{countably supported} \text{ (next slide)}\}$$



$$\mu\, x_0 = 2 \quad \mu\, x_1 = 3 \quad \mu\, x_2 = 2$$

Countably supported distributions

Support

A subset S **supports** a weight function $\mu : X \to \mathbb{W}$ when μ is 0 outside S:

$$\mu : \mathbb{W}^X, S : \mathcal{P}X \vdash S \text{ **supports** } \mu \coloneqq (\forall x : X.(\mu\, x > 0) \implies x \in S) : \mathsf{Prop}$$

The subsets supporting a weight function μ are closed under intersections.
$\implies$ There is a smallest supporting subset, called the **support** of μ:

$$\mu : \mathbb{W}^X \vdash \operatorname{supp} \mu \coloneqq \{x \in X | \mu\, x > 0\}$$

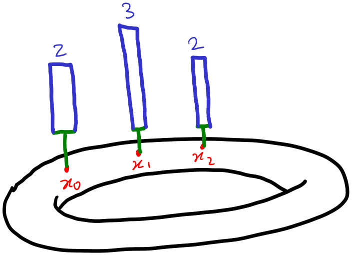
Discrete model

X: types denote **sets**

$\mathsf{D}X$: set of **histograms**:

$$\mathsf{D}X := \{\mu : X \to \mathbb{W} | \mu \text{ is } \textbf{countably supported} \}$$
$$:= \{\mu : X \to \mathbb{W} | \exists S \in \mathcal{P}X . S \text{ is countable}\}$$
$$:= \{\mu : X \to \mathbb{W} | \operatorname{supp} \mu \text{ is countable}\}$$

X



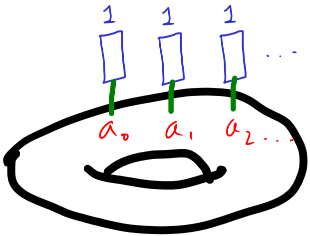
$$\mu\, x_0 = 2 \quad \mu\, x_1 = 3 \quad \mu\, x_2 = 2$$

Example distributions

Counting distribution

Counts the outcomes in a countable subset:

$$S : \mathcal{P}_{\mathrm{ctbl}} X \vdash \#_S \coloneqq \left(\lambda x. \begin{cases} x \in S : & 1 \\ x \notin S : & 0 \end{cases}\right) : \mathrm{D}X$$

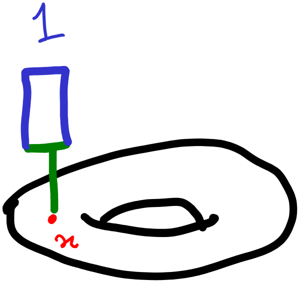


Example distributions

Dirac

A point mass:

$$x : X \vdash \delta_x := \left(\lambda x'.\begin{cases} x' = x : & 1 \\ x' \neq x : & 0 \end{cases}\right) : \mathrm{D}X$$



(NB: $x : X \vdash \delta_x = \#_{\{x\}}$.)

Example distributions

Zero

No mass anywhere:

$$\vdash \mathbf{0} := \underline{0} := (\lambda x.0) : \mathrm{D}X$$

(NB: $\vdash \mathbf{0} = \#_{\emptyset}$.)

Discrete model

X: types denote **sets**

$\mathsf{D}X$: set of **histograms**:

$$\mathsf{D}X := \{\mu : X \to \mathbb{W} | \mu \text{ is } \textbf{countably supported} \}$$

$\mathcal{B}_X$: **every subset** can be measured:

$$\mathcal{B}_X := \mathcal{P}X$$

Measurement: weighted sum of all (supported) outcomes:

$$\mu : \mathsf{D}X, E : \mathcal{B}_X \vdash \underset{\mu}{\mathrm{Ce}}[E] := \sum_{x \in E} \mu\, x$$

$$:= \sum_{x \in E \cap \operatorname{supp} \mu} \mu\, x$$

NB: $\mu : \mathsf{D}X, E : \mathcal{B}_X, S : \mathcal{P}_{\mathrm{ctbl}}X, S$ supports $\mu \vdash \mathrm{Ce}_\mu [E] = \sum_{x \in E \cap S} \mu\, x$.

Example measurements

$$(\text{NB: } \mu : \mathsf{D}X, E : \mathcal{B}_X, S : \mathcal{P}_{\mathrm{ctbl}}X, S \text{ supports } \mu \vdash \mathrm{Ce}_{\mu}[E] = \sum_{x \in E \cap S} \mu\, x.)$$

Counting distribution

counts supported outcomes

$$S : \mathcal{P}_{\mathrm{ctbl}}X, E : \mathcal{B}_X \vdash \underset{\#_S}{\mathrm{Ce}}[E] = |E \cap S| := \begin{cases} E \cap S \text{ has } n \in \mathbb{N} \text{ elements:} & n \\ E \cap S \text{ is infinite:} & \infty \end{cases}$$

Example measurements

(NB: $\mu : \mathrm{D}X, E : \mathcal{B}_X, S : \mathcal{P}_{\mathrm{ctbl}}X, S$ supports $\mu \vdash \mathrm{Ce}_{\mu}[E] = \sum_{x \in E \cap S} \mu\, x$.)

Counting distribution

counts supported outcomes

$$S : \mathcal{P}_{\mathrm{ctbl}}X, E : \mathcal{B}_X \vdash \underset{\#_S}{\mathrm{Ce}}[E] = |E \cap S| := \begin{cases} E \cap S \text{ has } n \in \mathbb{N} \text{ elements:} & n \\ E \cap S \text{ is infinite:} & \infty \end{cases}$$

Dirac

detects given outcome:

$$x : X, E : \mathcal{B}_X \vdash \mathrm{Ce}_{\boldsymbol{\delta}_x}[E] = \begin{cases} x \in E : & 1 \\ x \notin E : & 0 \end{cases}$$

Example measurements

(NB: $\mu : \mathrm{D}X, E : \mathcal{B}_X, S : \mathcal{P}_{\mathrm{ctbl}}X, S$ supports $\mu \vdash \mathrm{Ce}_{\mu}[E] = \sum_{x \in E \cap S} \mu\, x$.)

Counting distribution

counts supported outcomes

$$S : \mathcal{P}_{\mathrm{ctbl}}X, E : \mathcal{B}_X \vdash \underset{\#_S}{\mathrm{Ce}}[E] = |E \cap S| := \begin{cases} E \cap S \text{ has } n \in \mathbb{N} \text{ elements:} & n \\ E \cap S \text{ is infinite:} & \infty \end{cases}$$

Dirac

detects given outcome:

$$x : X, E : \mathcal{B}_X \vdash \mathrm{Ce}_{\boldsymbol{\delta}_x}[E] = \begin{cases} x \in E : & 1 \\ x \notin E : & 0 \end{cases}$$

Zero

measures every event as zero:

$$E : \mathcal{B}_X \vdash \mathrm{Ce}_{\mathbf{0}}[E] = 0$$

The discrete model validates the axioms

Exercise

$$\mu : \mathsf{D} \vdash \underset{\mu}{\mathrm{Ce}}\left[\emptyset\right] = 0$$

$$E, C : \mathcal{B}_X, \mu : \mathsf{D} \vdash \underset{\mu}{\mathrm{Ce}}\left[E\right] = \underset{\mu}{\mathrm{Ce}}\left[E \cap C\right] + \underset{\mu}{\mathrm{Ce}}\left[E \cap C^{\complement}\right]$$

$$E_{-} : (\mathcal{B}_X, \subseteq)^{\omega}, \mu : \mathsf{D}x \vdash \underset{\mu}{\mathrm{Ce}}\left[\bigcup_n E_n\right] = \sup_n \underset{\mu}{\mathrm{Ce}}\left[E_n\right]$$

Parameterised distributions

Kernel

$k : X \rightsquigarrow Y$ from X to Y: function $k : X \to \mathsf{D}Y$.

Kernels are open/parameterised distributions.

Examples

Dirac and the counting distribution form kernels:

$$\boldsymbol{\delta}_{\_} : X \rightsquigarrow \mathsf{D}X \qquad \#_{\_} : \mathcal{P}_{\mathrm{ctbl}}X \rightsquigarrow \mathsf{D}X$$

NB: This definition is **internal**: when we consider the full model, we will define kernels as those functions internal to the model rather than the set-theoretic functions.

Action of kernels on distributions

Kock integral

$$\mu : \mathsf{D}X, k : (\mathsf{D}Y)^X \vdash \oint \mathrm{d}\mu k : \mathsf{D}Y$$

This **distribution-valued** integral is implicit in many probability texts. It corresponds to integrating against an arbitrary weight function or random variable.

Discrete model interpretation

$$\begin{aligned}\oint \mathrm{d}\mu k &:= \lambda y. \sum_{x \in X} \mu\, x \cdot k(x; y) \\ &:= \lambda y. \sum_{x \in \operatorname{supp} \mu} \mu\, x \cdot k(x; y)\end{aligned}$$

NB1: we write $k(x; y) := k(x)(y)$ for the uncurried function.

NB2: $\mu : \mathsf{D}X, k : (\mathsf{D}Y)^X, S : \mathcal{P}_{\mathrm{ctbl}}X, S$ supports $\mu \vdash \oint \mathrm{d}\mu k = \lambda y. \sum_{x \in S} \mu\, x \cdot k(x; y)$

Example

Weak Disintegration Problem (non-standard terminology)

Input: distributions $\mu : \mathsf{D}\Theta$, $\nu : \mathsf{D}X$

Output: kernel $k : \Theta \rightsquigarrow \mathsf{D}X$ such that: $\nu = \oint \mathrm{d}\mu k$.

Such a **weak disintegration** of ν w.r.t. μ provides an 'explanation' of an observed distribution $\nu \in \mathsf{D}X$ in terms of a given distribution on parameters $\mu \in \mathsf{D}\Theta$. I use the term 'explanation' because it explains how the parameters transform into observations.

Example

Weak Disintegration Problem (non-standard terminology)

Input: distributions $\mu : \mathrm{D}\Theta$, $\nu : \mathrm{D}X$

Output: kernel $k : \Theta \rightsquigarrow \mathrm{D}X$ such that: $\nu = \oint \mathrm{d}\mu k$.

Example disintegration

For $n \in \mathbb{N}$, write $\mathbf{Fin}\, n := \{0, \ldots, n-1\}$. For countable X, write $\# := \#_X : \mathrm{D}X$.
Here is a disintegration of $\# \in \mathrm{D}\left((\mathbf{Fin}\, 2)^{\mathbf{Fin}\,(n+1)}\right)$ w.r.t. $\# \in \mathrm{D}(\mathbf{Fin}\, 2)$:

$$k(x; f) := \begin{cases} f n = x : & 1 \\ \text{otherwise:} & 0 \end{cases} \quad \text{Indeed:} \left(\oint \mathrm{d}\# k\right) f = \sum_{b \in \mathbf{Fin}\, 2} \overbrace{\# b}^{1} \cdot k(b; f) = k(0; f) + k(1; f)$$

$f : \mathbf{Fin}\,(n+1) \to \mathbf{Fin}\, 2$ function
so can take only one value: 0 or 1

$$\downarrow$$

$$= 1 = \# f$$

Sub-type of probability distributions

Sub-types

Given type X and $x : X \vdash \varphi : \mathsf{Prop}$, take the **sub-type** and the **coercion** as follows:

$$\{x : X | \varphi\} \subseteq X \qquad y : \{x : X | \varphi\} \vdash \mathsf{cast}\, y \coloneqq y : X$$

we **lift** values in X that satisfy φ to the sub-type:

$$\frac{\Gamma \vdash M : X \qquad \Gamma \vdash \varphi\,[x \mapsto M]}{\Gamma \vdash \mathsf{lift} M : \{x : X | \varphi\}} \qquad \frac{\Gamma \vdash M : X \qquad \Gamma \vdash \{\varphi\}\, x \mapsto M}{\Gamma \vdash \mathsf{cast}(\mathsf{lift} M) = M}$$

The axiom implies that $\mathsf{lift} M$ lifts M along cast. Moreover:

$$y : \{x \in X | \varphi\} \vdash \mathsf{lift}(\mathsf{cast}\, y) = y \quad y : \{x \in X | \varphi\} \vdash \varphi\,[x \mapsto \mathsf{cast}\, y]$$

i.e., the lifting is unique and elements in the sub-type satisfy φ.

Sub-type of probability distributions

Magnitude and probability distributions

$$\mu : \mathsf{D}X \vdash \|\mu\| \coloneqq \underset{\mu}{\mathrm{Ce}}[X] : \mathbb{W} \qquad \mathsf{P}X \coloneqq \{\mu \in \mathsf{D}X \mid \|\mu\| = 1\} \qquad \mathbb{I} \coloneqq [0,1] \coloneqq \{w \in \mathbb{W} \mid w \leq 1\}$$

Event probability

$$\mu : \mathsf{P}X, E : \mathcal{B}_X \vdash \underset{\mu}{\mathrm{Pr}}[E] \coloneqq \mathsf{lift}\left(\underset{\mathsf{cast}\,\mu}{\mathrm{Ce}}[E]\right) : \mathbb{I}$$

Stochastic kernel

$k : X \rightsquigarrow Y$ from X to Y: function $X \to \mathsf{P}Y$.

NB: in the **discrete model** these distinctions and rules amount to pure pedantry. This pedantry will pay off in the **full model**.

Lifting Dirac and Kock

Lemma

Dirac kernels $\boldsymbol{\delta}_{\_} : X \to \mathsf{D}X$ lift along cast:

$$x : X \vdash \|\boldsymbol{\delta}_x\| = \underset{\boldsymbol{\delta}_x}{\mathrm{Ce}}\,[X] = 1$$

so we can overload:

$$\begin{tikzcd} & \mathsf{P}X \arrow[d, "\mathrm{cast}"] \\ X \arrow[ur, dashed, "\boldsymbol{\delta}_{\_}"] \arrow[r, phantom, "=:"] \arrow[r, "\boldsymbol{\delta}_{\_}"'] & \mathsf{D}X \end{tikzcd}$$

Kock integrals of stochastic kernels by probability distributions lift along cast:

$$\mu : \mathsf{P}X, k : (\mathsf{P}Y)^X \vdash \mathrm{Ce}_{\oint(\mathsf{cast}\,\mu)(\mathrm{d}x)\,\mathsf{cast}(k\,x)}\,[Y] = 1$$

so we can overload:

$$\begin{tikzcd} (\mathsf{P}X) \times (\mathsf{P}Y)^X \arrow[r, dashed, "\oint"] \arrow[d, "\mathrm{cast} \times (\mathrm{cast} \circ)"'] \arrow[dr, phantom, "=:"] & \mathsf{P}Y \arrow[d, "\mathrm{cast}"] \\ (\mathsf{D}X) \times (\mathsf{D}Y)^X \arrow[r, "\oint"'] & \mathsf{D}Y \end{tikzcd}$$

Proposition

The triple $\left(\mathsf{D}, \boldsymbol{\delta}_{-}, \oint\right)$ forms a monad over $\mathbf{Set}$:

$$
\begin{array}{ll}
x : X, k : (\mathsf{D}Y)^X & \vdash \oint \mathrm{d}\boldsymbol{\delta}_x k = k\,x \\
\mu : \mathsf{D}X & \vdash \oint \mu(\mathrm{d}x)\boldsymbol{\delta}_x = \mu \\
\mu : \mathsf{D}X, k : (\mathsf{D}Y)^X, \ell : (\mathsf{D}Z)^Y & \vdash \oint \left(\oint \mu(\mathrm{d}x) k\,x\right)(\mathrm{d}y)\ell\,y = \oint \mu(\mathrm{d}x) \oint k(x; \mathrm{d}y)\ell\,y
\end{array}
$$

Corollary

The triple $\left(\mathsf{P}, \boldsymbol{\delta}_{-}, \oint\right)$ forms a monad over $\mathbf{Set}$.

Weighted average

Lebesgue integral

Integration is the raison d'être for distributions:

$$\mu : \mathsf{D}X, f : \mathbb{W}^X \vdash \int \mathrm{d}\mu f : \mathbb{W}$$

In the **discrete model**:

$$\int \mathrm{d}\mu f \coloneqq \sum_{x \in X} (\mu\, x) \cdot (f\, x) \coloneqq \sum_{x \in \operatorname{supp} \mu} (\mu\, x) \cdot (f\, x)$$

As usual, replace $\operatorname{supp} \mu$ by any countable supporting set:

$$\mu : \mathsf{D}X, f : \mathbb{W}^X, S : \mathcal{P}X, S \text{ supports } \mu \vdash \quad \int \mathrm{d}\mu f = \sum_{x \in S} (\mu\, x) \cdot (f\, x)$$

Weighted average

Expectation

To emphasise that some μ is a probability distribution, we will use the notation:

$$\mu : \mathsf{P}X, f : \mathbb{W}^X \vdash \quad \mathbb{E}_\mu[f] \coloneqq \int \mathrm{d}(\mathsf{cast}\,\mu) f : \mathbb{W}$$

When calculating, however, we will usually use $\int$ and implicitly cast any probability distribution to its corresponding distribution.

Booleans

Boolean type

The simplest kind of distinguishing outcomes:

$$\mathbb{B} := \{\mathbf{True}, \mathbf{False}\} \qquad \frac{\Gamma \vdash M : \mathbb{B} \qquad \Gamma \vdash N_1 : X \qquad \Gamma \vdash N_2 : X}{\Gamma \vdash \text{if } M \text{ then } N_1 \text{ else } N_2 : X}$$

Iverson bracket

Lets us replace Boolean propositions with arithmetic expressions:

$$b : \mathbb{B} \vdash [b] := (\text{if } b \text{ then } 1 \text{ else } 0) : \mathbb{W}$$

For example:

$$b : \mathbb{B}, w, v : \mathbb{W} \vdash \text{if } b \text{ then } w \text{ else } v = [b] \cdot w + (1 - [b]) \cdot w$$

Simplest probabilistic model

Bernoulli kernel

Single trial succeeding with the given probability:

$$\mathbf{B} : \mathbb{I} \rightsquigarrow \mathbb{B} \qquad \mathbf{B}p := \lambda b. \begin{cases} b = \mathbf{True} : & p \\ b = \mathbf{False} : & 1 - p \end{cases}$$

For example, for a payoff of 10 units if the trial succeeds then the expected payoff is:

$$\mathbb{E}_{b \sim \mathbf{B}\frac{1}{4}} [[b] \cdot 10] = \frac{1}{4} \cdot 10 + (1 - \frac{1}{4}) \cdot 0 = \frac{10}{4} + 0 = \frac{5}{2}$$

Events as functions

Proposition

Membership testing induces an isomorphism between events and Boolean propositions:

$$(\in) : \mathcal{B}_X \xrightarrow{\cong} \mathbb{B}^X$$

Its inverse sends each Boolean property to the set of outcomes satisfying it:

$$\frac{x : X \vdash M : \mathbb{B}}{\{x \in X | M\} : \mathcal{B}_X} \qquad \{x \in X | \varphi\, x\} \coloneqq \{x \in X | \varphi\, x = \mathbf{True}\}$$

Characteristic function

represents an event as weight functions: $E : \mathcal{B}_X \vdash [- \in E] : \mathbb{W}^X$

By the above proposition, every (internal) $\{0, 1\}$-valued weight function is the characteristic function of some event, namely, the inverse image of 1.

Measurement through integration

Lemma

We can replace event measurement by integration of characteristic functions:

$$\mu : \mathsf{D}X, E : \mathcal{B}_X \vdash \underset{\mu}{\mathrm{Ce}}[E] = \int \mu(\mathrm{d}x)\,[x \in E]$$

We can deduce properties for $\mathrm{Ce}[-]$ and $\mathrm{Pr}[-]$ from those of the Lebesgue integral.

Notation:

$$\frac{\Gamma \vdash \mu : \mathsf{D}X \qquad \Gamma, x : X \vdash M : \mathbb{B}}{\Gamma \vdash \underset{x\sim\mu}{\mathrm{Ce}}[M] \coloneqq \underset{\mu}{\mathrm{Ce}}[\{x \in X | M\}] : \mathbb{W}}$$

and similarly for $\mathrm{Pr}_{x\sim\mu}[M]$.

Language of probability & distribution (recap)

X type of **values/outcomes**

$\mathsf{D}X$ type of **distributions/measures** over X

$\mathsf{P}X \subseteq \mathsf{D}X$ sub-type of **probability distributions** over X

$\mathcal{B}_X \subseteq \mathcal{P}X$ type of **events**: subsets we wish to measure

$\mathbb{W}$ type of **weights**: values in $[0,\infty]$

$\int, \mathbb{E}$ Lebesgue integration and the expectation operation

Type judgements describe well-formed values/outcomes of a given type, e.g.:

$$\mu : \mathsf{D}X, E : \mathcal{B}_X \vdash \underset{\mu}{\mathrm{Ce}}[E] : \mathbb{W}$$

(measures weight $\mathrm{Ce}_\mu[E]$ of event E according to distribution μ)

Propositions describe properties of well-formed values/outcomes of a given type, e.g.:

$$y_1, y_2 : Y \vdash y_1 \overset{Y}{=} y_2 : \mathsf{Prop} \qquad \mu : \mathsf{P}X, E : \mathcal{B}_X \vdash \mathsf{cast}\, \underset{\mu}{\mathrm{Pr}}[E] = \underset{\mu}{\mathrm{Ce}}[E]$$

(probability of event according to probability distribution is its measure)

Lecture plan

Lecture 1: discrete model (today)

- Language of probability and distribution
- Discrete model
- Simply-typed probability
- Dependently-typed probability

Lecture 2: the full model

- Borel sets and measurable spaces
- Quasi-Borel spaces
- Type structure & standard Borel spaces
- Integration & random variables

course page

ask questions on the Scottish PL Institute Zulip stream #qbs

Simply-typed foundations for probabilistic modelling

Compositional building blocks for modelling

- Affine combinations of distributions
- Product measures $(\otimes) : \mathrm{D}X \times \mathrm{D}Y \to \mathrm{D}(X \times Y)$
- Random elements and their laws (push-forward measure):
  $(\lambda(\mu, \alpha).\mu_\alpha) : \mathrm{D}\Omega \times X^\Omega \to \mathrm{D}X$

NB:

- Dirac kernel $\delta_\_ : X \to \mathrm{D}X$
- Kock integration
  $\oint : \mathrm{D}X \times (\mathrm{D}Y)^{\mathrm{D}X} \to \mathrm{D}Y$

Standard vocabulary

- Joint and marginal distributions
- Independence
- Distribution/probability preservation and invariance
- Density and absolute continuity
- Almost certain/sure properties

Simply-typed foundations for probabilistic modelling

Compositional building blocks for modelling

- Affine combinations of distributions
- Product measures $(\otimes) : \mathrm{D}X \times \mathrm{D}Y \to \mathrm{D}(X \times Y)$
- Random elements and their laws (push-forward measure):
  $(\lambda\,(\mu, \alpha)\,.\mu_\alpha) : \mathrm{D}\Omega \times X^\Omega \to \mathrm{D}X$

NB:

- Dirac kernel $\delta_{\_} : X \to \mathrm{D}X$
- Kock integration
  $\oint : \mathrm{D}X \times (\mathrm{D}Y)^{\mathrm{D}X} \to \mathrm{D}Y$

Standard vocabulary

- Joint and marginal distributions
- Independence
- Distribution/probability preservation and invariance
- Density and absolute continuity
- Almost certain/sure properties

Affine combinations of distributions: scaling

Scaling distributions

$$w : \mathbb{W}, \mu : \mathsf{D}X \vdash w \cdot \mu : \mathsf{D}X$$

In the discrete model:

$$w \cdot \mu := \lambda x. w \cdot \mu\, x \qquad \operatorname{supp}(w \cdot \mu) \subseteq \operatorname{supp} \mu$$

The function $(\cdot) : \mathbb{W} \times \mathsf{D}X \to \mathsf{D}X$ is a **monoid action** for the monoid $(\mathbb{W}, (\cdot), 1)$:

$$\mu : \mathsf{D}X \vdash 1 \cdot \mu = \mu \qquad w, v : \mathbb{W}, \mu : \mathsf{D}X \vdash w \cdot (v \cdot \mu) = (w \cdot v) \cdot \mu$$

Integration and measurement are homogeneous w.r.t. scaling:

$$w : \mathbb{W}, \mu : \mathsf{D}X, k : (\mathsf{D}Y)^X \vdash \oint \mathrm{d}(w \cdot \mu) k = w \cdot \oint \mathrm{d}\mu k$$

$$w : \mathbb{W}, \mu : \mathsf{D}X, f : \mathbb{W}^X \vdash \int \mathrm{d}(w \cdot \mu) f = w \cdot \int \mathrm{d}\mu f$$

$$w : \mathbb{W}, \mu : \mathsf{D}X, E : \mathcal{B}_X \vdash \underset{w \cdot \mu}{\mathrm{Ce}}[f] = w \cdot \underset{\mu}{\mathrm{Ce}}[f]$$

Affine combinations of distributions: scaling

Normalisation

$$\mu : \mathsf{D}X, \|\mu\| \neq 0, \infty \vdash \frac{\mu}{\|\mu\|} := \mathsf{lift}\left(\frac{1}{\|\mu\|} \cdot \mu\right) : \mathsf{P}X$$

$$\text{Indeed: } \left\|\frac{\mu}{\|\mu\|}\right\| = \left\|\frac{1}{\|\mu\|} \cdot \mu\right\| \overset{\substack{\text{measurement is homogeneous}\\ \downarrow}}{=} \frac{1}{\|\mu\|} \cdot \|\mu\| = 1$$

Discrete uniform / categorical distribution

Random unbiased choice between finitely many options/categories:

$$S : \mathcal{P}_{\text{fin}}(X), S \neq \emptyset \vdash \quad \mathbf{U}_S := \frac{\mathsf{lift}\#_S}{\|\mathsf{lift}\#_S\|} : \mathsf{P}X$$

In the discrete model:

$$\mathbf{U}_S = \lambda x. \begin{cases} x \in S : & \frac{1}{|S|} \\ x \notin S : & 0 \end{cases}$$

so: $x : X \vdash \mathbf{U}_{\{x\}} = \boldsymbol{\delta}_x$.

Weights as distributions

Unit type

$$\mathbb{1} \coloneqq \{()\}$$

Proposition

The following two functions are mutually inverse:

$$\begin{tikzcd} \mathrm{D}\mathbb{1} \arrow[r, bend left, "\|-\|"] & \mathbb{W} \arrow[l, bend left, "(\cdot\boldsymbol{\delta}_{()})"] \end{tikzcd}$$

Proof

Calculate: $\mu : \mathrm{D}\mathbb{1} \vdash \mu \mapsto \mu\,() \mapsto \lambda().\mu\,() = \mu$ and $w : \mathbb{W} \vdash w \mapsto \lambda().w \mapsto w$. ■

Internalising Lebesgue integration

Proposition

We can recover Lebesgue integration from Kock integration:

$$\begin{tikzcd}
\mathsf{D}X \times \mathbb{W}^{X} \arrow[r, "\mathrm{id} \times (\cong \circ)"] \arrow[d, "\int"'] \arrow[dr, phantom, "="] & \mathsf{D}X \times (\mathsf{D}\mathbb{1})^{X} \arrow[d, "\oint"] \\
\mathbb{W} & \mathsf{D}\mathbb{1} \arrow[l, "\cong"]
\end{tikzcd}$$

Since measurement also reduced to Lebesgue integration, it usually suffices to prove properties of Kock integration and derive them for Lebesgue integration and for measurement.

Affine combinations of distributions: addition

Summation

$$\mu_{\_} : (\mathsf{D}X)^I, I \text{ countable} \vdash \sum_{i\in I} \mu_i : \mathsf{D}X$$

In the discrete model:

$$\sum_{i\in I} \mu_i := \lambda x. \sum_{i\in I} \mu_i\, x \qquad \operatorname{supp} \sum_{i\in I} \mu_i = \bigcup_{i\in I} \operatorname{supp} \mu_i$$

Affine and convex combinations

An **affine** combination is a countable sequence of weights $w_{\_} : \mathbb{W}^I$.
It is **convex** when $\sum_{i\in I} w_i = 1$.

Bernoulli revisited

We can express the Bernoulli distribution as follows:

$$p : \mathbb{I} \vdash \mathbf{B}p = \mathsf{lift}\,(p \cdot \boldsymbol{\delta}_{\mathbf{True}} + (1-p) \cdot \boldsymbol{\delta}_{\mathbf{False}}) : \mathsf{P}\mathbb{B}$$

Affinity of integration and convexity of expectation

Theorem (Multi-linearity)

The Kock and Lebesgue integrals and measurement are affine in each argument:

$$\mu_{\_} : (\mathsf{D}X)^I, w_{\_} : \mathbb{W}^I, k : X \rightsquigarrow Y \vdash \oint \mathrm{d}\Big(\sum_{i\in I} w_i \cdot \mu_i\Big)k = \sum_{i\in I} w_i \cdot \oint \mathrm{d}\mu_i k$$

$$\mu : \mathsf{D}X, w_{\_} : \mathbb{W}^I, k_{\_} : (X \rightsquigarrow B)^I \vdash \oint \mathrm{d}\mu\Big(\sum_{i\in I} w_i \cdot k_i\Big) = \sum_{i\in I} w_i \cdot \oint \mathrm{d}\mu k_i$$

$$\mu_{\_} : (\mathsf{D}X)^I, w_{\_} : \mathbb{W}^I, \varphi : \mathbb{W}^X \vdash \int \mathrm{d}\Big(\sum_{i\in I} w_i \cdot \mu_i\Big)\varphi = \sum_{i\in I} w_i \cdot \int \mathrm{d}\mu_i \varphi$$

$$\mu : \mathsf{D}X, w_{\_} : \mathbb{W}^I, \varphi_{\_} : (\mathbb{W}^X)^I \vdash \int \mathrm{d}\mu\Big(\sum_{i\in I} w_i \cdot \varphi_i\Big) = \sum_{i\in I} w_i \cdot \int \mathrm{d}\mu \varphi_i$$

$$\mu_{\_} : (\mathsf{D}X)^I, w_{\_} : \mathbb{W}^I, E : \mathcal{B}_X \vdash \underset{\sum_{i\in I} w_i \cdot \mu_i}{\mathrm{Ce}}[E] = \sum_{i\in I} w_i \cdot \underset{\mu_i}{\mathrm{Ce}}[E]$$

Weight arithmetic

This theorem, a working horse in probability, has several important consequences:

Proposition

The isomorphism $\mathrm{D}\mathbb{1} \cong \mathbb{W}$ is a σ-semiring isomorphism:

$$\left(\mathrm{D}\mathbb{1}, \sum, (\cdot)\right) \cong \left(\mathbb{W}, \sum, (\cdot)\right)$$

and $(\cdot) : \mathbb{W} \times \mathrm{D}X \to \mathrm{D}X$ makes each $\mathrm{D}X$ into a $\mathbb{W}$-module:

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

Convex combinations of probability distributions

Lemma

Convex combination lifts to probability distributions:

$$w_{\_} : \mathbb{W}^{I}, \mu_{\_} : (\mathsf{P}X)^{I}, I \text{ countable}, \sum_{i \in I} w_i = 1 \vdash$$

$$\sum_{i \in I} w_i \cdot \mu_i := \mathsf{lift} \sum_{i \in I} w_i \cdot (\mathsf{cast}\, \mu_i) : \mathsf{P}X$$

Proof

Calculate: $\left\| \sum_{i \in I} w_i \cdot (\mathsf{cast}\, \mu_i) \right\| = \sum_{i \in I} w_i \cdot \|\mathsf{cast}\, \mu_i\| = \sum_{i \in I} w_i \cdot 1 = 1$ ∎

Convex combinations of probability distributions

Corollary (Multi-convexity)

Stochastic Kock integration, expectation and measurement are convex:

$$\mu_{\_} : (\mathsf{D}X)^I, w_{\_} : \mathbb{W}^I, k : X \rightsquigarrow Y, \sum_{i \in I} w_i = 1 \vdash \oint \mathrm{d}(\sum_{i \in I} w_i \cdot \mu_i) k = \sum_{i \in I} w_i \cdot \oint \mathrm{d}\mu_i k$$

$$\mu : \mathsf{D}X, w_{\_} : \mathbb{W}^I, k_{\_} : (X \rightsquigarrow B)^I, \sum_{i \in I} w_i = 1 \vdash \oint \mathrm{d}\mu (\sum_{i \in I} w_i \cdot k_i) = \sum_{i \in I} w_i \cdot \oint \mathrm{d}\mu k_i$$

$$\mu_{\_} : (\mathsf{D}X)^I, w_{\_} : \mathbb{W}^I, \varphi : \mathbb{W}^X, \sum_{i \in I} w_i = 1 \vdash \mathbb{E}_{\sum_{i \in I} w_i \cdot \mu_i} [\varphi] = \sum_{i \in I} w_i \cdot \mathbb{E}_{\mu_i} [\varphi]$$

$$\mu : \mathsf{D}X, w_{\_} : \mathbb{W}^I, \varphi_{\_} : (\mathbb{W}^X)^I, \sum_{i \in I} w_i = 1 \vdash \mathbb{E}_{\mu} \left[\sum_{i \in I} w_i \cdot \varphi_i\right] = \sum_{i \in I} w_i \cdot \mathbb{E}_{\mu} [\varphi_i]$$

$$\mu_{\_} : (\mathsf{D}X)^I, w_{\_} : \mathbb{W}^I, E : \mathcal{B}_X, \sum_{i \in I} w_i = 1 \vdash \Pr_{\sum_{i \in I} w_i \cdot \mu_i} [E] = \sum_{i \in I} w_i \cdot \Pr_{\mu_i} [E]$$

Products

Product distribution

$$\mu : \mathsf{D}X, \nu : \mathsf{D}Y \vdash \quad \mu \otimes \nu := \oint \mu(\mathrm{d}x) \oint \nu(\mathrm{d}y) \boldsymbol{\delta}_{(x,y)} : \mathsf{D}(X \times Y)$$

In the discrete model:

$$\mu \otimes \nu = \lambda(x, y).(\mu\, x) \cdot (\nu\, y) \qquad \operatorname{supp}(\mu \otimes \nu) = (\operatorname{supp} \mu) \times (\operatorname{supp} \nu)$$

Example: counting distribution on product space

$$S : \mathcal{P}_{\text{fin}}(X), T : \mathcal{P}_{\text{fin}}(Y) \vdash \quad \#_{S \times T} \overset{\mathsf{D}(X \times Y)}{=} \#_S \otimes \#_T$$

Indeed: $\operatorname{supp}(\#_S \otimes \#_T) = S \times T = \operatorname{supp} \#_{S \times T}$ and for $(x, y) \in S \times T$:

$$(\#_S \otimes \#_T)(x, y) = 1 \cdot 1 = 1 = \#_{S \times T}(x, y)$$

Products

Notation:

$$\frac{\Gamma \vdash M : \mathsf{D}(X \times Y) \qquad \Gamma, x : X, y : Y \vdash K : \mathsf{D}Z}{\Gamma \vdash \oiint M(\mathrm{d}x, \mathrm{d}y)K \coloneqq \oint \mathrm{d}M(\lambda\,(x, y)\,.K) : \mathsf{D}Z}$$

Theorem (Fubini-Tonelli)

We can integrate products in any order:

$$\mu : \mathsf{D}X, \nu : \mathsf{D}Y, k : (\mathsf{D}Z)^{X \times Y} \vdash$$

$$\oint \mu(\mathrm{d}x) \oint \nu(\mathrm{d}y) k\,(x, y) = \oiint (\mu \otimes \nu)(\mathrm{d}x, \mathrm{d}y) k\,(x, y) = \oint \nu(\mathrm{d}y) \oint \mu(\mathrm{d}x) k\,(x, y)$$

$$\mu : \mathsf{D}X, \nu : \mathsf{D}Y, \varphi : \mathbb{W}^{X \times Y} \vdash$$

$$\int \mu(\mathrm{d}x) \int \nu(\mathrm{d}y) \varphi\,(x, y) = \iint (\mu \otimes \nu)(\mathrm{d}x, \mathrm{d}y) \varphi\,(x, y) = \int \nu(\mathrm{d}y) \int \mu(\mathrm{d}x) \varphi\,(x, y)$$

Applying Fubini-Tonelli

Theorem (Rule of Product)

We can factor out products:

$$\mu : \mathsf{D}X, f : \mathbb{W}^X, \nu : \mathsf{D}Y, g : \mathbb{W}^Y \vdash \quad \iint (\mu \otimes \nu)(\mathrm{d}x, \mathrm{d}y) f x \cdot g y = \left(\int \mathrm{d}\mu f\right) \cdot \left(\int \mathrm{d}\nu g\right)$$

$$\mu : \mathsf{D}X, E : \mathcal{B}_X, \nu : \mathsf{D}Y, F : \mathcal{B}_Y \vdash \quad \underset{\mu \otimes \nu}{\mathrm{Ce}} [E \times F] = \underset{\mu}{\mathrm{Ce}} [E] \cdot \underset{\nu}{\mathrm{Ce}} [F]$$

Theorem

The product lifts to probability distributions:

$$\mu : \mathsf{P}X, \nu : \mathsf{P}Y \vdash (\mu \otimes \nu) \coloneqq \mathsf{lift}(\mathsf{cast}\,\mu \otimes \mathsf{cast}\,\nu) : \mathsf{P}(X \times Y)$$

Products

Binomial distribution

the number of successful outcomes of n independent Bernoulli trials:

$$\mathbf{B}_n : \mathbb{I} \rightsquigarrow \mathsf{P}(\mathbf{Fin}\,(1+n)) \qquad \mathbf{B}_0 p \coloneqq \boldsymbol{\delta}_0 : \mathsf{P}(\mathbf{Fin}\,1)$$

$$\mathbf{B}_{1+n} p \coloneqq \oiint (\mathbf{B}_n p \otimes \mathbf{B} p)(\mathrm{d}c, \mathrm{d}b)\,(\mathrm{if}\, b \,\mathrm{then}\, \boldsymbol{\delta}_{1+c} \,\mathrm{else}\, \boldsymbol{\delta}_c) : \mathsf{P}(\mathbf{Fin}\,(2+n))$$

We can prove by induction on n, using Fubini-Tonelli and the Iverson bracket that:

$$p : \mathbb{I}, k : \mathbf{Fin}\,(1+n) \vdash \Pr_{c \sim \mathbf{B}_n p} [c = k] = \binom{n}{k}$$

Push-forward distributions

Random element

in X any (internal) function:

$$\mu : \mathsf{D}\Omega \vdash \alpha : \Omega \to X$$

Law

of a random element is the distribution:

$$\mu : \mathsf{D}\Omega, \alpha : X^{\Omega} \vdash \mu_{\alpha} := \oint \mu(\mathrm{d}\omega)\boldsymbol{\delta}_{\alpha\,\omega} : \mathsf{D}X$$

Example

Represent outcomes of die roll by $\mathsf{D6} := \{1, 2, \ldots, 6\}$, and two rolls by $\mathsf{D6} \times \mathsf{D6}$.
The sum of the rolls is a random element:

$$(+) : \mathsf{D6} \times \mathsf{D6} \to \mathbb{N}$$

The law of the distribution $\# \otimes \#$ counts the number of configurations in which the two rolls sum to a given number, e.g.: $(\# \otimes \#)_{(+)} : 1 \mapsto 0, 2 \mapsto 1$.

Push-forward distributions

Theorem (Law of the Unconcious Statistician)

Formulae for reparameterising integration and measurement:

$$\mu : \Omega, \alpha : X^{\Omega}, k : X \rightsquigarrow Y \vdash \oint \mathrm{d}\mu_{\alpha} k = \oint \mathrm{d}\mu(k \circ \alpha)$$

$$\mu : \Omega, \alpha : X^{\Omega}, f : \mathbb{W}^{X} \vdash \int \mathrm{d}\mu_{\alpha} f = \int \mathrm{d}\mu(f \circ \alpha)$$

$$\mu : \Omega, \alpha : X^{\Omega}, E : \mathcal{B}_X \vdash \underset{\mu_\alpha}{\mathrm{Ce}} [E] = \underset{\mu}{\mathrm{Ce}} \left[\alpha^{-1} [E]\right] = \underset{\omega \sim \mu}{\mathrm{Ce}} [\alpha\, \omega \in E]$$

Simply-typed foundations for probabilistic modelling

Compositional building blocks for modelling

- Affine combinations of distributions
- Product measures $(\otimes) : \mathrm{D}X \times \mathrm{D}Y \to \mathrm{D}(X \times Y)$
- Random elements and their laws (push-forward measure): $(\lambda (\mu, \alpha) . \mu_\alpha) : \mathrm{D}\Omega \times X^\Omega \to \mathrm{D}X$

NB:

- Dirac kernel $\delta_{\_} : X \to \mathrm{D}X$
- Kock integration $\oint : \mathrm{D}X \times (\mathrm{D}Y)^{\mathrm{D}X} \to \mathrm{D}Y$

Standard vocabulary

- Joint and marginal distributions
- Independence
- Distribution/probability preservation and invariance
- Density and absolute continuity
- Almost certain/sure properties

Standard vocabulary: concepts concerning products

Let $\pi_i : \prod_{i \in I} X_i \to X_i$ be the i-th projection.

Joint distribution: $\mu : \mathsf{D}(X \times Y)$, $\mu : \mathsf{D}\left(\prod_{i \in I} X_i\right)$

Marginal distribution: the law of a projection:

$$\mu : \mathsf{D}\left(\prod_{i \in I} X_i\right) \vdash \mu_{\pi_i} : \mathsf{D}X_i$$

Sometimes refers to any law of a r.e..

Marginalisation: the action of calculating a marginal distribution by integrating all other components.

Exercise

$$\mu : \mathsf{P}X, \nu : \mathsf{D}X \vdash (\mu \otimes \nu)_{\pi_2} = \nu$$

Independence

Pairing random elements

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

Independent random elements

The joint law is the product of the marginals:

$$\mu : \mathsf{D}\Omega, \alpha : X^{\Omega}, \beta : Y^{\Omega} \vdash \alpha \underset{\mu}{\perp} \beta \coloneqq \left(\mu_{(\alpha,\beta)} \overset{\mathsf{D}(X\times Y)}{=} \mu_{\alpha} \otimes \mu_{\beta}\right)$$

More generally, for finite I:

$$\mu : \mathsf{D}\Omega, \alpha_{\_} : (X^{\Omega})^{I} \vdash \underset{\mu}{\perp}_{i}\alpha_{i} \coloneqq \left(\mu_{(\alpha_i)_i} \overset{\mathsf{D}(\prod_i X_i)}{=} \bigotimes_{i\in I} \mu_{\alpha_i}\right)$$

Example [Durett]

Model 3 independent coin tosses:

$$\mathsf{Toss} \coloneqq \{\mathbf{Head}, \mathbf{Tail}\} \qquad \Omega \coloneqq \mathsf{Toss}^3 \qquad \mu \coloneqq \mathbf{U}_{\mathsf{Toss}} \otimes \mathbf{U}_{\mathsf{Toss}} \otimes \mathbf{U}_{\mathsf{Toss}} : \mathrm{P}\Omega$$

The outcome of the i^{th} coin toss is the random element $\pi_i : \Omega \to \mathsf{Toss}$.
Consider the Boolean proposition in which the i^{th} and j^{th} tosses $(i \neq j)$ agree:

$$\mathsf{Same}_{ij} \coloneqq \lambda\omega.\pi_i\omega = \pi_j\omega : \Omega \to \mathbb{B}$$

Calculate:

$$\Pr_{\mu}[\mathsf{Same}_{12}] \overset{\substack{\text{LOTUS}\\ \downarrow}}{=} \Pr_{(x,y)\sim\mu_{(\pi_1,\pi_2)}}[x = y] \overset{\substack{\text{marginalisation}\\ \downarrow}}{=} \Pr_{(x,y)\sim\mathbf{U}\otimes\mathbf{U}}[x = y] \overset{\substack{\text{Fubini}\\ \downarrow}}{=} \int \mathbf{U}(\mathrm{d}x) \Pr_{y\sim\mathbf{U}}[x = y]$$

$$= \frac{1}{2} \cdot \Pr_{y\sim\mathbf{U}}[\mathbf{Head} = y] + \frac{1}{2} \cdot \Pr_{y\sim\mathbf{U}}[\mathbf{Tail} = y] = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

Independence

Example [Durett]

Model 3 independent coin tosses:

$$\mathsf{Toss} \coloneqq \{\mathbf{Head}, \mathbf{Tail}\} \qquad \Omega \coloneqq \mathsf{Toss}^3 \qquad \mu \coloneqq \mathbf{U}_{\mathsf{Toss}} \otimes \mathbf{U}_{\mathsf{Toss}} \otimes \mathbf{U}_{\mathsf{Toss}} : \mathsf{P}\Omega$$

The outcome of the i^{th} coin toss is the random element $\pi_i : \Omega \to \mathsf{Toss}$.
Consider the Boolean proposition in which the i^{th} and j^{th} tosses $(i \neq j)$ agree:

$$\mathsf{Same}_{ij} \coloneqq \lambda\omega.\pi_i\omega = \pi_j\omega : \Omega \to \mathbb{B}$$

Therefore $\mu_{\mathsf{Same}_{12}} = \mathbf{U}_{\mathbb{B}}$ and similarly $\mu_{\mathsf{Same}_{ij}} = \mathbf{U}_{\mathbb{B}}$ for $i \neq j$.

Independence

π_1, Same_{12}, and Same_{13} determine π_2, π_3, so:

$$\Pr_{\omega\sim\mu}[\mathsf{Same}_{12}\omega = \mathbf{True}, \mathsf{Same}_{13}\omega = \mathbf{True}]$$

Fubini-Tonelli

$$\downarrow$$

$$= \int \mathbf{U}_{\mathsf{Toss}}(\mathrm{d}b_1) \Pr_{(b_2,b_3)\sim(\mathbf{U}\otimes\mathbf{U})}[\mathsf{Same}_{12}(b_1,b_2,b_3) = \mathbf{True}, \mathsf{Same}_{13}(b_1,b_2,b_3) = \mathbf{True}]$$

$$= \frac{1}{2} \Pr_{(b_2,b_3)\sim(\mathbf{U}\otimes\mathbf{U})}[\mathsf{Same}_{12}(\mathbf{Head},b_2,b_3) = \mathbf{True}, \mathsf{Same}_{13}(\mathbf{Head},b_2,b_3) = \mathbf{True}]$$

$$+ \frac{1}{2} \Pr_{(b_2,b_3)\sim(\mathbf{U}\otimes\mathbf{U})}[\mathsf{Same}_{12}(\mathbf{Tail},b_2,b_3) = \mathbf{True}, \mathsf{Same}_{13}(\mathbf{Tail},b_2,b_3) = \mathbf{True}]$$

$$= \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$$

and similarly we get $\frac{1}{4}$ in all other cases.

Independence

Example [Durett]

Model 3 independent coin tosses:

$$\mathsf{Toss} \coloneqq \{\mathbf{Head}, \mathbf{Tail}\} \qquad \Omega \coloneqq \mathsf{Toss}^3 \qquad \mu \coloneqq \mathbf{U}_{\mathsf{Toss}} \otimes \mathbf{U}_{\mathsf{Toss}} \otimes \mathbf{U}_{\mathsf{Toss}} : \mathsf{P}\Omega$$

The outcome of the i^{th} coin toss is the random element $\pi_i : \Omega \to \mathsf{Toss}$.
Consider the Boolean proposition in which the i^{th} and j^{th} tosses $(i \neq j)$ agree:

$$\mathsf{Same}_{ij} \coloneqq \lambda\omega.\pi_i\omega = \pi_j\omega : \Omega \to \mathbb{B}$$

Therefore $\mu_{\mathsf{Same}_{12}} = \mathbf{U}_{\mathbb{B}}$ and similarly $\mu_{\mathsf{Same}_{ij}} = \mathbf{U}_{\mathbb{B}}$ for $i \neq j$. So:

$$\mu_{(\mathsf{Same}_{12},\mathsf{Same}_{13})} = \mathbf{U}_{\mathbb{B}\times\mathbb{B}} = \mathbf{U}_{\mathbb{B}} \otimes \mathbf{U}_{\mathbb{B}} = \mu_{\mathsf{Same}_{12}} \otimes \mu_{\mathsf{Same}_{13}}$$

So $\mathsf{Same}_{12} \underset{\mu}{\perp} \mathsf{Same}_{13}$ even though their values depend on the outcome of the first toss.

Distribution preservation

Distribution space (Ω, μ)

A type Ω equipped with a distribution $\mu : \mathsf{D}\Omega$. Define **probability space** analogously.

Distribution preserving function

$f : (\Omega_1, \mu_1) \to (\Omega_2, \mu_2)$ is a function whose is the co domain distribution:

$$f : \Omega_1 \to \Omega_2 \qquad (\mu_1)_f = \mu_2$$

$\mu : \mathsf{D}X$ is **invariant** under $f : X \to X$ when $f : (X, \mu) \to (X, \mu)$ is dist. preserving.

Example

Consider the swapping function: $\mathsf{swap} \coloneqq (\lambda\,(x, y)\,.\,(y, x)) : X \times Y \to Y \times X$. Then, for each $\mu : \mathsf{D}X$, $\nu : \mathsf{D}Y$, swapping is distribution preserving function:

$$\mathsf{swap} : (X \times Y, \mu \otimes \nu) \to (Y \times X, \nu \otimes \mu)$$

swap is invariant in the case $X = Y$ and $\mu = \nu$.

Density and scaling

Density

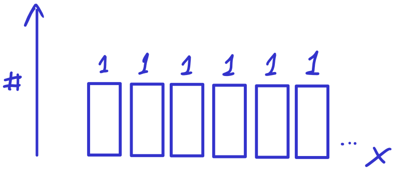
over X is any weight function $f : X \to \mathbb{W}$.

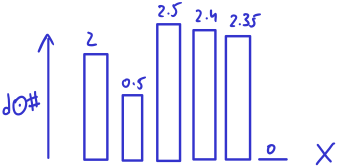
Density scaling

We can scale a distribution by a density:

$$f : \mathbb{W}^X, \mu : \mathsf{D}X \vdash f \odot \mu := \oint \mu(\mathrm{d}x)(f, x) \cdot \boldsymbol{\delta}_x : \mathsf{D}X$$

Scaling does not lift to probability distributions: $\|f \odot \mu\| \neq 1$ even if $\|\mu\| = 1$.





Density and scaling

Density

over X is any weight function $f : X \to \mathbb{W}$.

Density scaling

We can scale a distribution by a density:

$$f : \mathbb{W}^X, \mu : \mathsf{D}X \vdash f \odot \mu := \oint \mu(\mathrm{d}x)(f, x) \cdot \boldsymbol{\delta}_x : \mathsf{D}X$$

Scaling does not lift to probability distributions: $\|f \odot \mu\| \neq 1$ even if $\|\mu\| = 1$.

Warning!

The types of distributions and densities over X in the **discrete** model are close, but still **different**. They coincide on **countable** types, so people often confused them. Types help us keep them separate.

Density and absolute continuity

Having density

This concept has several names in the literature:

$$\mu, \nu : \mathsf{D}X, f : \mathbb{W}^X \vdash \left(f = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right) := (\mu = f \odot \nu) : \mathsf{Prop}$$

- f is the **density** of μ w.r.t. ν
- f is a **Radon-Nikodym derivative** of μ w.r.t. ν.

Absolute continuity

μ is **absolutely continuous** w.r.t. ν when μ has a density w.r.t. ν:

$$\mu, \nu : \mathsf{D}X \vdash (\mu \ll \nu) := \exists f : \mathbb{W}^X . f = \frac{\mathrm{d}\mu}{\mathrm{d}\nu} : \mathsf{Prop}$$

Density and absolute continuity

Example

The **uniform distribution** is absolutely continuous w.r.t. the **counting measure** over the same support. Indeed, it has these two densities:

$$S : \mathcal{P}_{\mathrm{fin}}(X) \vdash \left(\lambda x. \frac{1}{|S|}\right), \left(\lambda x. \begin{cases} x \in S : & \frac{1}{|S|} \\ x \notin S : & 0 \end{cases}\right) = \frac{\mathrm{d}\mathbf{U}_S}{\mathrm{d}\#_S}$$

These two densities are different, but they agree on the support, motivating the following concept.

Almost certain/sure properties

Almost certain event

is one we can assert without changing the distribution:

$$\frac{\Gamma \vdash \mu : \mathsf{D}X \qquad \Gamma, x : X \vdash M : \mathbb{B}}{\Gamma \vdash \mu(\mathrm{d}x) \ \mathsf{almost\ certainly}\ M \coloneqq [M] \odot \mu = \mu : \mathsf{Prop}}$$

For probabilities we define:

$$\frac{\Gamma \vdash \mu : \mathsf{P}X \qquad \Gamma, x : X \vdash M : \mathbb{B}}{\Gamma \vdash \mu(\mathrm{d}x) \ \mathsf{almost\ surely}\ M \coloneqq (\mathsf{cast}\,\mu)(\mathrm{d}x) \ \mathsf{almost\ certainly}\ M : \mathsf{Prop}}$$

Existence and almost-sure uniqueness of densities

Theorem (Radon-Nikodym)

For **probability** distributions, we characterise absolute continuity as follows:

$$\mu, \nu : \mathsf{P}X \vdash (\mu \ll \nu) \iff \forall E : \mathcal{B}_X . \Pr_{\nu}[E] = 0 \implies \Pr_{\mu}[E] = 0$$

In that case, if $f, g = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ then $\nu(\mathrm{d}x)$ almost surely $f\,x = g\,x$.

In the **discrete model**, this characterisation amounts to $\operatorname{supp}\mu \subseteq \operatorname{supp}\nu$.

Example

For all countable X, we have:

$$\forall \mu : \mathsf{D}X . \mu \ll \#_X$$

Indeed, apply the Radon-Nikodym theorem, since $\operatorname{supp}\# = X$.
Constructively, direct calculation shows: $(\lambda x.\mu\,x) = \frac{\mathrm{d}\mu}{\mathrm{d}\#}$.

Simply-typed foundations for probabilistic modelling

Compositional building blocks for modelling

- Affine combinations of distributions
- Product measures $(\otimes) : \mathrm{D}X \times \mathrm{D}Y \to \mathrm{D}(X \times Y)$
- Random elements and their laws (push-forward measure):
  $(\lambda\,(\mu, \alpha)\,.\mu_\alpha) : \mathrm{D}\Omega \times X^\Omega \to \mathrm{D}X$

NB:

- Dirac kernel $\delta_{\_} : X \to \mathrm{D}X$
- Kock integration
  $\oint : \mathrm{D}X \times (\mathrm{D}Y)^{\mathrm{D}X} \to \mathrm{D}Y$

Standard vocabulary

- Joint and marginal distributions
- Independence
- Distribution/probability preservation and invariance
- Density and absolute continuity
- Almost certain/sure properties

Lecture plan

Lecture 1: discrete model (today)

- Language of probability and distribution
- Discrete model
- Simply-typed probability
- Dependently-typed probability

Lecture 2: the full model

- Borel sets and measurable spaces
- Quasi-Borel spaces
- Type structure & standard Borel spaces
- Integration & random variables

course page

ask questions on the Scottish PL Institute Zulip stream #qbs

Example: Binomial kernels

We've defined, for every $n \in \mathbb{N}$, the binomial kernel:

$$\vdash \mathbf{B}_n : \mathbb{1} \mathrel{\circ\!\!\rightsquigarrow} \mathbf{Fin}\,(1 + n)$$

We will now look at **dependent-type** structure which allows us to view these as one kernel internally:

$$n : \mathbb{N} \vdash \mathbf{B}_n : \mathbb{1} \mathrel{\circ\!\!\rightsquigarrow} \mathbf{Fin}\,(1 + n)$$

Family model

Family over an indexing set I

consists of a seqeuence $X_{\_} = (X_i)_{i \in I}$ of sets.
We call each set X_i the **fibre over** i.

Family F

a pair $F = (I, X_{\_})$ consisting of (indexing) set I and a family $X_{\_}$ over it.
Notation: $F = I \vdash X_{\_}$

$$= i : I \vdash X_i.$$

Example

The family $n : \mathbb{N} \vdash \mathbf{Fin}\, n$ has $\mathbb{N}$ as the indexing set. The fibre over $n \in \mathbb{N}$ is:

$$\mathbf{Fin}\, n \coloneqq \{0, 1, \ldots, n - 1\}$$

Family model

Family over an indexing set I

consists of a seqeuence $X_{\_} = (X_i)_{i \in I}$ of sets.
We call each set X_i the **fibre over i**.

Family F

a pair $F = (I, X_{\_})$ consisting of (indexing) set I and a family $X_{\_}$ over it.
Notation: $F = I \vdash X_{\_}$

$$= i : I \vdash X_i.$$

Family map

$(\theta, f_{\_}) : (I \vdash X_{\_}) \to (J \vdash Y_{\_})$ is a pair of a function between the indexing sets and a sequence of functions between the corresponding fibres:

$$\theta : I \to J \qquad (f_i : X_i \to Y_{\theta\, i})_{i \in I}$$

Notation: $\theta \vdash f_{\_}$. We won't use these maps explicitly, but they are the foundation.

Terms in context

Dependent elements $i : I \vdash M : X_i$

in family $i : I \vdash X_i$ are I-indexed sequences of elements from the corresponding fibres:

$$(M \in X_i)_{i \in I}$$

Example

We have the elements:

$$n : \mathbb{N} \vdash 0, \ldots, n - 1 : \mathbf{Fin}\, n$$

Subsumption

Every simple type becomes a family by ignoring the dependency through the constant family, e.g., $i : I \vdash \mathbb{N}$ and $i : I \vdash 42 : \mathbb{N}$.

Simple functions

Fibred exponential

of two families over the same indexing set $i : I \vdash X_i, Y_i$ is the family:

$$i : I \vdash X_i \to Y_i$$

Family of distributions

over a family $i : I \vdash X_i$ is the family:

$$i : I \vdash \mathsf{D}X_i$$

Its sub-family of fibred **probability** distributions:

$$i : I \vdash \mathsf{P}X_i$$

Both have a **Dirac** distribution:

$$i : I \vdash \boldsymbol{\delta}\_ : X_i \to \mathsf{D}X_i \qquad i : I \vdash \boldsymbol{\delta}\_ : X_i \to \mathsf{P}X_i$$

Extension and dependent pairs

Extension

of indexing set I by a **variable** of the family $i : I \vdash X_i$ is the (indexing) set:

$$\coprod_{i \in I} X_i \coloneqq \bigcup_{i \in I} \{i\} \times X_i = \left\{ (i, x) \in I \times \bigcup_{i \in I} X_i \middle| x \in X_i \right\}$$

Notation: $(i : I, x : X_i) \coloneqq \coprod_{i \in I} X_i$ and we'll often write i, x instead of (i, x).

Dependent pairs

$$\frac{i : I \vdash X_i \qquad i : I, x : X_i \vdash Y_{i,x}}{i : I \vdash (x : X_i) \times (Y_{i,x}) \coloneqq \coprod_{x \in X_i} Y_{i,x}}$$

Functions and kernels

Dependent functions

we identify a function f with a tuple $(f\ x)_x$ as usual:

$$\frac{i : I \vdash X_i \qquad i : I, x : X_i \vdash Y_{i,x}}{i : I \vdash ((x : X) \to Y_{i,x}) := \prod_{x \in X} Y_{i,x}}$$

Dependent kernels $i : I \vdash k : (x : X_i) \rightsquigarrow Y_{i,x}$

are dependent elements:

$$i : I \vdash k : (x : X_i) \to \mathsf{D} Y_{i,x}$$

Dependent **stochastic** kernels $i : I \vdash k : (x : X_i) \circ\!\!\rightsquigarrow Y_{i,x}$ are similarly:

$$i : I \vdash k : (x : X_i) \to \mathsf{P} Y_{i,x}$$

Dependent Kock integral

$$i : I, \mu : \mathsf{D}X_i, k : (x : X_i) \rightsquigarrow Y_{i,x} \vdash \oint \mathrm{d}\mu k : \mathsf{D}Y_{i,x}$$

and in the **discrete model** we define it for i, μ, k as in the simply-typed case:

$$(\oint \mathrm{d}\mu k) y \coloneqq \sum_{x \in X_i} \mu\, x \cdot k(x; y) : \mathbb{W}$$

Through the identification $\mathbb{W} \cong \mathsf{D}\mathbb{1}$ and characteristic functions, we reduce dependent Lebesgue integration and measurement to dependent Kock integration:

$$i : I, \mu : \mathsf{D}X_i, f : (x : X_i) \to \mathbb{W} \vdash \int \mathrm{d}\mu f : \mathbb{W} \qquad i : I, \mu : \mathsf{D}X_i, E : \mathcal{B}_{X_i} \vdash \underset{\mu}{\mathrm{Ce}}[E] : \mathbb{W}$$

$$\int \mathrm{d}\mu f = \sum_{x \in X} \mu\, x \cdot f\, x \qquad \mathrm{Ce}_\mu [E] = \sum_{x \in E} \mu\, x$$

Random variables

Let $\overline{\mathbb{R}} := [-\infty, \infty]$ be the extended real line.

Signed and unsigned random variable

in a probability space (Ω, μ) are random elements $\alpha : \Omega \to \overline{\mathbb{R}}$ and $\alpha : \Omega \to \mathbb{W}$.
The **positive** and **negative parts** are unsigned random variables $-^{\pm} : \overline{\mathbb{R}}^{\Omega} \to \mathbb{W}^{\Omega}$:

$$\alpha^{+} := \lambda\omega.\max(\alpha\,\omega, 0) = [\alpha \geq 0] \cdot |\alpha| \qquad \alpha^{-} := \lambda\omega. -\min(\alpha\,\omega, 0) = [\alpha \leq 0] \cdot |\alpha|$$

An unsigned r.v. α is **Lebesgue integrable** when its Lebesgue integral is finite:
$\int \mathrm{d}\mu\alpha < \infty$.
For a (signed) r.v. α, when either α^{+} or α^{-} is Lebesgue integrable, we define:

$$\mu : \mathsf{D}X, \alpha : \overline{\mathbb{R}}^{X}, \int \mathrm{d}\mu\alpha^{+}, \int \mathrm{d}\mu\alpha^{-} < \infty \vdash \qquad \int \mathrm{d}\mu\alpha := \int \mathrm{d}\mu\alpha^{+} - \int \mu\alpha^{-}$$

A signed variable is **Lebesgue integrable** when both its parts are Lebesgue integrable.

Random variable spaces

Lebesgue integrability is a Boolean property:

$$\mu : \mathsf{D}X, \alpha : X \to \overline{\mathbb{R}} \vdash \alpha \text{ integrable} := \int \mathrm{d}\mu\alpha^{+} < \infty \wedge \int \mathrm{d}\mu\alpha^{-} < \infty : \mathbb{B}$$

Lebesgue spaces ensemble

is the family:

$$i : I, p : [1,\infty), \mu : \mathsf{P}X_i \vdash \qquad \mathcal{L}_p(X_i, \mu) := \left\{\alpha : X_i \to \overline{\mathbb{R}} \middle| \alpha^p \text{ integrable}\right\}$$

Every fibre has a vector space structure and a norm (almost a Banach space!):

$$i : I, p : [1,\infty), \mu : \mathsf{P}X_i, \alpha : \mathcal{L}_p(X_i, \mu) \vdash \|\alpha\|_p := \sqrt[p]{\mathbb{E}_\mu\left[|\alpha|^p\right]} : \mathbb{W}$$

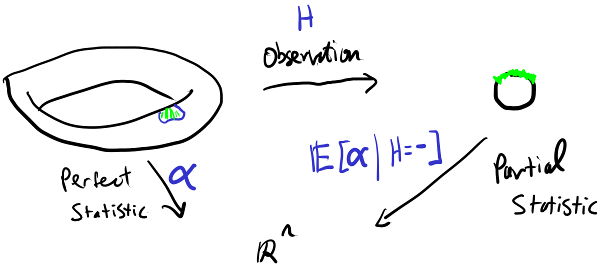
and the fibre 2 has an inner product (almost a Hilbert space!):

$$i : I, \mu : \mathsf{P}X_i, \alpha, \beta : \mathcal{L}_2(X_i, \mu) \vdash (\alpha, \beta) := \sqrt{\mathbb{E}_\mu\left[\alpha \cdot \beta\right]} : \mathbb{W}$$

Conditioning

Situation:

- Statistical model $\mu : \mathsf{D}\Omega$
  (voters in the next election)
- Perfect statistic $\alpha : \Omega \to \mathbb{R}$
  (expected winning candidate)
- Observation $H : \mathsf{D} \to X$
  (poll voting intention)



Conditional expectation of α along H w.r.t μ

Statistic $\beta : X \to \mathbb{R}$ that 'best' approximates $H \circ \alpha$ staistically. Halmos and Doob's definition: any measurement we make of β agrees with measurement of α:

$$\mu : \mathsf{D}\Omega, H : \Omega \to X, \alpha : \mathcal{L}_1(\Omega, \mu), \beta : \mathcal{L}_1(X, \mu_H) \vdash$$

$$\left(\beta = \mathop{\mathbb{E}}_{\mu}[\alpha | H = -]\right) := \left(\forall \varphi : \mathcal{L}_1 X, \mu_H . \int \mathrm{d}\mu_H \beta \cdot \varphi = \int \mathrm{d}\mu \alpha(\varphi \circ H)\right) \quad : \mathsf{Prop}$$

Conditioning

Theorem (Kolmogorov)

Every random variable has a conditional expectation:

$$\mu : \mathsf{D}\Omega, H : \Omega \to X, \alpha : \mathcal{L}_1(\Omega, \mu) \vdash \qquad \exists \beta : \mathcal{L}_1(X, \mu_H).\beta = \mathop{\mathbb{E}}_{\mu}[\alpha | H = -]$$

Therefore:

Corollary (Internal conditional expectation)

In the **discrete model** we have a dependent function:

$$\mathbb{E}_{-}[-|- = -] :$$
$$(\mu : \mathsf{D}\Omega) \to (H : \Omega \to X) \to (\alpha : \mathcal{L}_1(\Omega, \mu)) \to \left\{ \beta : \mathcal{L}_1(X, \mu_H) \middle| \beta = \mathop{\mathbb{E}}_{\mu}[\alpha | H = -] \right\}$$

Conditioning

Conditional probability

of event is a conditional expectation of its characteristic function:

$$\mu : \mathsf{P}\Omega, H : \Omega \to X, E : \mathcal{B}_\Omega, \alpha : \mathcal{L}_1(\Omega, \mu) \vdash$$
$$\left(\alpha = \Pr_{\mu}[E|H = -]\right) \coloneqq \left(\alpha = \mathop{\mathbb{E}}_{\omega\sim\mu}[\omega \in E|H = -]\right) : \mathsf{Prop}$$

Regular conditional probability

a kernel that agrees with the conditional expectation of the characteristic functions:

$$\mu : \mathsf{P}\Omega, H : \Omega \to X, k : X \multimap\!\!\rightsquigarrow \Omega \vdash$$
$$\left(k = \Pr_{\mu}[-|H = -]\right) \coloneqq \left(\forall E \in \mathcal{B}_\Omega . k(-; E) = \mathop{\mathbb{E}}_{\omega\sim\mu}[\omega \in E|H = -]\right) : \mathsf{Prop}$$

Conditioning

Kolmogorov's theorem does **not** ensure the existence of a regular conditional probability, although the constructive, discrete, definition does.

Conditioning

Disintegration of $H : \Omega \to X$ w.r.t. $\mu : \mathsf{D}\Omega$

is a kernel $k : X \rightsquigarrow \Omega$ that weakly disintegrates the law of H and, moreover, $\mathrm{d}x$-almost surely each probability distribution $k(x)$ is in the fibre:

$$\mu : \mathsf{D}\Omega, H : \Omega \to X, k : X \rightsquigarrow \Omega \vdash k \text{ disintegrates } H \text{ w.r.t. } \mu$$
$$:= \mu = \oint \mathrm{d}\mu_H k,$$
$$\mu_H(\mathrm{d}x) \text{ almost surely } k(x; \mathrm{d}\omega) \text{ almost surely } H\omega = x \qquad : \mathsf{Prop}$$

In the **discrete model** we have an internal disintegration:

$$-^{\dagger -} : (H : \Omega \to X) \to (\mu : \mathsf{P}\Omega) \to \{k : X \rightsquigarrow \Omega | k \text{ disintegrates } H \text{ w.r.t. } \mu\}$$
$$H^{\dagger\mu} := \lambda x. \begin{cases} \mu\, x > 0 : & \frac{[H - = x]}{\mu_H x} \odot \mu \\ \mu\, x = 0 : & \mu \end{cases}$$

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):

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

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

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:

$$\mathbb{P}\mathrm{r}_{\lambda}[-|H=-] = \lambda_{X|H=-}$$

Lecture plan

Lecture 1: discrete model

- Motivation
- Language of probability and distribution
- Discrete model
- Simply-typed probability
- Dependently-typed probability

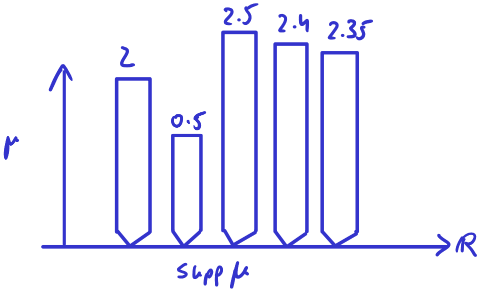
course page

Lecture 2: the full model

- Borel sets and measurable spaces
- Quasi-Borel spaces
- Type structure & standard Borel spaces
- Integration & random variables

ask questions on the Scottish PL Institute Zulip stream #qbs

discrete model measure only histograms:



Want:

- lengths
- areas
- volumes.

Continuous Caveat:

Thm: no $\lambda : \mathcal{P}\mathbb{R} \to [0,\infty]$:

$\lambda(a,b) = b-a$ (generalises length)

$\lambda(r+A) = \lambda A$ (translation invariant)

$\lambda\left(\biguplus_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} \lambda A_n$ σ-additive

Takeaway: taking $\mathcal{B}\mathbb{R} := \mathcal{P}\mathbb{R}$

excludes measures such as:

length, area, volume

Workaround: only measure well-behaved subsets

Def: The Borel subsets $B_{\mathbb{R}} \subseteq P\,\mathbb{R}$:

- open intervals $(a,b) \in B_{\mathbb{R}}$

Closure under σ-algebra operations:

$$\frac{}{\emptyset \in B_{\mathbb{R}}} \qquad \frac{A \in B_{\mathbb{R}}}{A^c := \mathbb{R} \setminus A \in B} \qquad \frac{\vec{A} \in B_{\mathbb{R}}^{\mathbb{N}}}{\bigcup_{n=0}^{\infty} A_n \in B_{\mathbb{R}}}$$

empty set   complements   countable unions

Examples

discrete Countable: $\{r\} = \bigcap_{\varepsilon \in \mathbb{Q}^+} (r-\varepsilon, r+\varepsilon) \in \mathcal{B}_{\mathbb{R}}$

I countable $\Rightarrow$ $I = \bigcup_{r \in I} \{r\} \in \mathcal{B}_{\mathbb{R}}$

closed intervals: $[a,b] = (a,b) \cup \{a,b\}$

Non-examples?

More complicated: analytic, Lebesgue

Def: Measurable space $V = (\lfloor V \rfloor, \mathcal{B}_V)$

Set (carrier) — $\lfloor V \rfloor$

family of subsets $\mathcal{B}_V \subseteq \mathcal{P}\lfloor V \rfloor$

Closed under σ-algebra operations:

$$\frac{}{\emptyset \in \mathcal{B}_V} \qquad \frac{A \in \mathcal{B}_V}{A^c := \lfloor V \rfloor \setminus A \in \mathcal{B}_V} \qquad \frac{\vec{A} \in \mathcal{B}_V^{\mathbb{N}}}{\bigcup_{n=0}^{\infty} A_n \in \mathcal{B}_V}$$

empty set — complements — countable unions

Idea: structure all spaces after the worst-case scenario

Examples

- Discrete spaces $\ulcorner X^{\text{meas}} \urcorner = (X, PX)$

- Euclidean spaces $\mathbb{R}^n$ — replace intervals with chests $\prod_{i=1}^{n} (a_i, b_i)$

  $\mathbb{R}^{\mathbb{N}}$ similarly

  $\{C \cap A \mid C \in B_V\}$

- Subspaces: $A \in P \lfloor V \rfloor$ $\quad A := (A, [B_V] \cap A)$

- Products: $A \times B := (\lfloor A \rfloor \times \lfloor B \rfloor, \sigma([B_A] \times [B_B]))$

Def: Borel measurable functions $f : V_1 \to V_2$

- functions $f : \lfloor V_1 \rfloor \longrightarrow \lfloor V_2 \rfloor$
- inverse image preserves measurability:

$$f^{-1}[A] \in \mathcal{B}_{V_1} \Longleftarrow A \in \mathcal{B}_{V_2}$$

Examples

- $(+), (\cdot) : \mathbb{R}^2 \longrightarrow \mathbb{R}$
- $|-|$, $\sin : \mathbb{R} \to \mathbb{R}$
- any continuous function $f : \mathbb{R}^n \to \mathbb{R}^m$
- any function $f : \ulcorner X \urcorner \to V$

Category **Meas**

Objects: Measurable spaces

Morphisms: Measurable functions

Identities:

$id : V \to V$

Composition:

$$\frac{f : V_2 \to V_3 \qquad g : V_1 \to V_2}{f \circ g : V_1 \to V_3}$$

Meas Category

Products, coproducts/disjoint union, Subspaces

Categorical limits, colimits, but:

Thm [Aumann '61] No σ-algebras $B_{B_{\mathbb{R}}}$, $B_{\mathbb{R}^{\mathbb{R}}}$ for measurable

$$(\ni) : (B_{\mathbb{R}}, B_{B_{\mathbb{R}}}) \times \mathbb{R} \longrightarrow \mathrm{Bool}$$

membership predicate

$$(U, r) \longmapsto [r \in U]$$

$$\mathrm{eval} : (\mathrm{Meas}(\mathbb{R}, \mathbb{R}), B_{\mathbb{R}^{\mathbb{R}}}) \times \mathbb{R} \to \mathbb{R}$$

$$(f, r) \mapsto f(r)$$

Sequential Higher-order structure:

I Countable : $V^{\ulcorner I \urcorner} = \prod_{i \in I} V$

$\Longrightarrow$ Some higher-order structure in **Meas**:

Cauchy $\in B_{[-\infty,\infty]^{\mathbb{N}}}$

$$\mathrm{Cauchy} := \bigcap_{\varepsilon \in \mathbb{Q}^+} \bigcup_{k \in \mathbb{N}} \bigcap_{\substack{m,n \in \mathbb{N} \\ m,n \geq k}} \{ \vec{y} \in [-\infty,\infty]^{\mathbb{N}} \mid |y_m - y_n| < \varepsilon \}$$

lim sup $: [-\infty,\infty]^{\mathbb{N}} \to [-\infty,\infty]$ lim: Cauchy $\to \mathbb{R}$

Compose higher-order building blocks:

lim is measurable!

$$\mathrm{VanishingSeq}(\mathbb{R}) := \left\{ \vec{r} \in \mathbb{R}^{\mathbb{N}} \,\middle|\, \lim_{n\to\infty} r_n = 0 \right\} \in \mathcal{B}_{\mathbb{R}^{\mathbb{N}}}$$

$$\mathrm{approx}\_ : \mathrm{VanishingSeq}(\mathbb{R}^+) \times \mathbb{R} \longrightarrow \mathbb{Q}^{\mathbb{N}}$$

$$\text{s.t.:} \quad \left| (\mathrm{approx}_{\vec{\Delta}}\, r)_n - r \right| < \Delta_n$$

Slogan: Measurable by Type!

Not all operations of interest fit:

$$\limsup : \left([-\infty,\infty]^{\mathbb{R}}\right)^{\mathbb{N}} \longrightarrow [-\infty,\infty]^{\mathbb{R}}$$

$$\limsup := \lambda \vec{f}.\lambda x.\limsup_{n\to\infty} f_n\, x$$

Intrinsically higher-order!

Want

Slogan: measurability by type!

But

For higher-order building blocks

defer measurability proofs until

we resume 1st order fragment ⇒ non compositional

Plan

Def: $V \in \mathbf{Meas}$ is Standard Borel when

$$V \cong A \quad \text{for some } A \in \mathcal{B}_{\mathbb{R}}$$

the "good part" of **Meas** — the subcategory

$$\mathbf{Sbs} \hookrightarrow \mathbf{Meas}$$

Sbs include

- Discrete $\ulcorner I \urcorner$, I countable
- Countable products of Sbs:

$$\mathbb{R}^n, \mathbb{R}^{\mathbb{N}}, 2^{\mathbb{N}}, \mathbb{N}^{\mathbb{N}}$$

- Borel subspaces of Sbs:

$$\mathbb{I} := [0,1] \qquad \mathbb{R}^+ := (0,\infty) \quad \mathbb{R}_{\geq 0} := [0,\infty]$$

- Countable coproducts of Sbs:

$$\mathbb{W} := [0,\infty] \qquad \overline{\mathbb{R}} := [-\infty,\infty]$$

Conservative extensions:

concrete spaces

we "observe"

Standard Borel spaces

Meas

Qbs

abstract (auxiliary?)

spaces

Plan:

1) type-driven probability: discrete case ✓

2) Borel sets & measurable spaces ✓

Lecture 1

3) Quasi Borel spaces

4) Type structure & standard Borel spaces

5) Integration & random variables

Lecture 2

Please ask questions!

Course web page

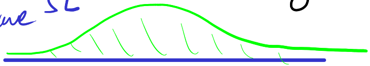
scribble

Core idea

Measure Theory

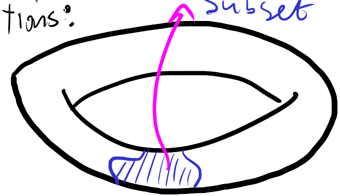
Qbs Theory

sample space Ω



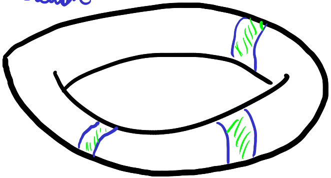
Primitive notions:

measurable subset



random element

$\downarrow \alpha$



Derived notions:

measure

random elements $\alpha : \Omega \to \mathrm{Space}$

Events $E \in B_X$

Def: Quasi-Borel space $X = (\lfloor X \rfloor, R_X)$

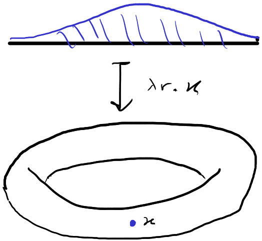
set "carrier"

set of functions $\alpha : \mathbb{R} \to \lfloor X \rfloor$ "random elements"

$R_X \subseteq \lfloor X \rfloor^{\lfloor \mathbb{R} \rfloor}$ closed under:

- Constants:

$$\frac{x \in \lfloor X \rfloor}{(\lambda r.x) \in R_X}$$



- Precomposition:

- recombination

Def: Quasi-Borel space $X = (\lfloor X \rfloor, R_X)$

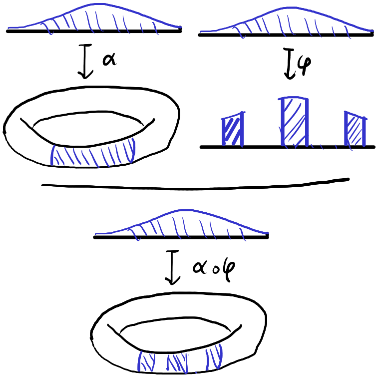
set "carrier"

set of functions $\alpha : \mathbb{R} \to \lfloor X \rfloor$ "random elements"

$R_X \subseteq \lfloor X \rfloor^{\lfloor \mathbb{R} \rfloor}$ closed under:

- precomposition:

$$\frac{\alpha \in R_X \qquad \varphi : \mathbb{R} \to \mathbb{R} \text{ in } \mathbf{Sbs}}{\varphi \circ \alpha : \mathbb{R} \xrightarrow{\varphi} \mathbb{R} \xrightarrow{\alpha} \lfloor X \rfloor \in R_X}$$



Def: Quasi-Borel space $X = (\lfloor X \rfloor, R_X)$

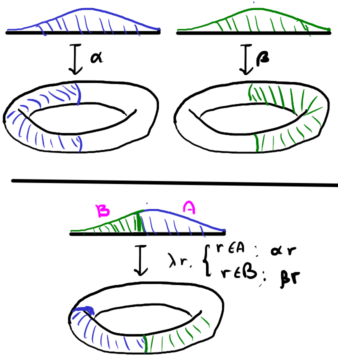
Set "carrier"

Set of functions $\alpha : \mathbb{R} \to \lfloor X \rfloor$ "random elements"

$R_X \subseteq \lfloor X \rfloor^{\lfloor \mathbb{R} \rfloor}$ closed under:

- recombination

$$\frac{\vec{\alpha} \in R_X^{\mathbb{N}} \qquad \mathbb{R} = \biguplus_{n=0}^{\infty} \underbrace{A_n}_{\in B_{\mathbb{R}}}}{\lambda r. \begin{cases} \vdots \\ r \in A_n : \alpha_n r \\ \vdots \end{cases}}$$

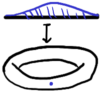


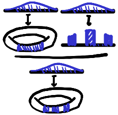
Def: Quasi-Borel space $X = (\lfloor X \rfloor, R_X)$

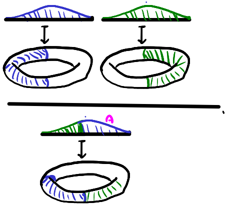
Set "carrier"

Set of functions $\alpha : \mathbb{R} \to \lfloor X \rfloor$ "random elements"

$R_X \subseteq \lfloor X \rfloor^{\lfloor \mathbb{R} \rfloor}$ closed under:

- Constants: 

- precomposition: 

- recombination 

Examples

- $\mathbb{R} = (\lfloor\mathbb{R}\rfloor, \mathrm{Meas}(\mathbb{R},\mathbb{R}))$

  qbs underlying $\mathbb{R}$

- $X \in \mathrm{Set}$, $\ulcorner X \urcorner^{\mathrm{Qbs}} := (X, \sigma\text{-simple}(\mathbb{R}, X))$

  recombination of constants

  $$\lambda r. \begin{cases} \vdots \\ r \in A_n : x_n \\ \vdots \end{cases}$$

  discrete qbs on X

- " $\lfloor X \rfloor_{\mathrm{Qbs}} := (X, X^{\lfloor\mathbb{R}\rfloor})$

  all funtions

  Indiscrete qbs on X

Qbs **morphism** $f : X \to Y$

- function $f : \lfloor X \rfloor \to \lfloor Y \rfloor$
- $$\frac{\left(\mathbb{R} \xrightarrow{\alpha} \lfloor X \rfloor\right) \in R_X}{\left(\mathbb{R} \xrightarrow{\alpha} \lfloor X \rfloor \xrightarrow{f} \lfloor Y \rfloor\right) \in R_Y}$$

Example

- **Constant** functions are qbs morphisms
- **σ-simple** functions are qbs morphisms

Category **Qbs** $\Leftarrow$ - identity, composition

Full model

$\mathrm{type} : \mathbf{Qbs} \qquad W := [0,\infty] \qquad BX := \boxed{}$

$DX := \boxed{}$

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

$\underset{\mu}{\mathrm{Ce}}[E] := \boxed{} \qquad \delta_x := \boxed{}$

$\oint \mu k := \boxed{}$

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

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

Full model

type : $\mathbf{Qbs}$ $\qquad \mathbb{W} := [0,\infty] \qquad \mathcal{B}X := \boxed{\quad}$

$DX := \boxed{\quad}$

$PX := \{\mu \in DX \mid \mathcal{C}e_\mu[X] = 1\}$

$\mathcal{C}e_\mu[E] := \boxed{\quad} \qquad \delta_x := \boxed{\quad}$

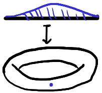
$\oint \mu k := \boxed{\quad}$

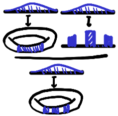
Def: Quasi-Borel space $X = (\lfloor X \rfloor, R_X)$

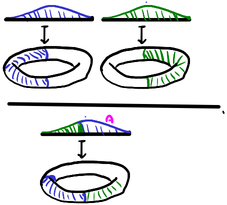
Set "carrier"

Set of functions $\alpha : \mathbb{R} \to \lfloor X \rfloor$ "random elements"

$R_X \subseteq \lfloor X \rfloor^{\lfloor \mathbb{R} \rfloor}$ closed under:

- Constants: 

- precomposition: 

- recombination 

Examples

- $\mathbb{R} = (\lfloor\mathbb{R}\rfloor, \mathrm{Meas}(\mathbb{R},\mathbb{R}))$

  qbs underlying $\mathbb{R}$

- $X \in \mathrm{Set}$, $\ulcorner X \urcorner^{\mathrm{Qbs}} := (X, \sigma\text{-simple}(\mathbb{R}, X))$

  recombination of constants

  $$\lambda r. \begin{cases} \vdots \\ r \in A_n : x_n \\ \vdots \end{cases}$$

  discrete qbs on X

- " $\llcorner X \lrcorner_{\mathrm{Qbs}} := (X, X^{\lfloor\mathbb{R}\rfloor})$

  all functions

  Indiscrete qbs on X

Validate qbs axioms for: $\mathbb{W} := ([0,\infty], \mathrm{Meas}(\mathbb{R}, \mathbb{W}))$

- Constants:

$$E : \mathsf{B}_{\mathbb{W}},\ x : \mathbb{W} \vdash$$

$$(\lambda r : \mathbb{R}.\, x)^{-1}[E] = \begin{cases} x \in E : & \mathbb{R} \\ x \notin E : & \emptyset \end{cases} \in \mathsf{B}_{\mathbb{R}} \quad \checkmark$$

Validate qbs axioms for: $W := ([0,\infty], \mathrm{Meas}(\mathbb{R}, W))$

- Precomposition:

$\alpha : \mathrm{Meas}(\mathbb{R}, W)$, $\varphi : \mathrm{Meas}(\mathbb{R}, \mathbb{R}) \vdash$

$$\mathbb{R} \xrightarrow{\varphi} \mathbb{R} \xrightarrow{\alpha} W \quad \in \mathrm{Meas}(\mathbb{R}, W)$$

Meas is a cat.

Explicitly:

$$(\alpha \circ \varphi)^{-1}[E] \in \mathcal{B}\mathbb{R} \xleftarrow{\varphi^{-1}} \alpha^{-1}[E] \in \mathcal{B}\mathbb{R} \xleftarrow{\alpha^{-1}} E \in \mathcal{B}W \quad \checkmark$$

Validate qbs axioms for: $\mathbb{W} := ([0,\infty], \mathrm{Meas}(\mathbb{R}, \mathbb{W}))$

- recombination

I ctbl, $\underline{\alpha} : \mathrm{Meas}(\mathbb{R}, \mathbb{W})^I$, $E_- : B_{\mathbb{R}}^I$, $\mathbb{R} = \biguplus_{i \in I} E_i$, $F : B_{\mathbb{W}^+}$

$$\beta := \left(\lambda r. \begin{cases} \vdots \\ r \in E_i : & \alpha_i r \\ \vdots \end{cases}\right)^{-1}[F]$$

$$= \bigcup_{i \in I} \alpha_i^{-1}[F] \cap E_i \in B_{\mathbb{R}}$$

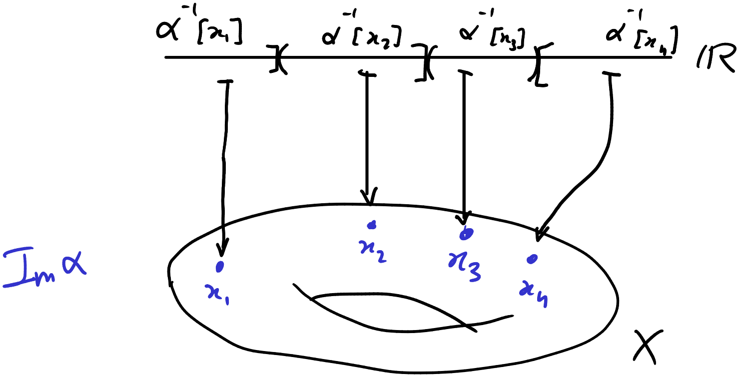
Indeed:

$r \in LHS \iff \beta r \in F \iff \exists i \in I.\, r \in E_i \wedge \alpha_i r \in F \iff r \in RHS$ ✓

σ-simple function

$\alpha : \mathbb{R} \to X$ s.t. $\mathrm{Im}\,\alpha := \alpha[\mathbb{R}]$ is ctbl $\wedge$

$\forall x \in \mathrm{Im}\,\alpha .\ \alpha^{-1}[x] \in \mathcal{B}_{\mathbb{R}}$



Validate Qbs axioms for: $\ulcorner X^{\mathbf{Qbs}} \urcorner := (X, \sigma\text{-Simple}(X))$

NB: f σ-Simple: $\mathrm{Im}\, f$ ctbl $\wedge$ $f^{-1}[x] \in B_{\mathbb{R}}$

- Constants

$\mathrm{Im}(\lambda r.x) = \{x\}$ ctbl ✓

$$y : X \vdash (\lambda r.x)^{-1}[y] = \begin{cases} x = y : & \mathbb{R} \\ x \neq y : & \emptyset \end{cases} \in B_{\mathbb{R}} \checkmark$$

Validate $\mathbf{Qbs}$ axioms for: $\ulcorner X^{\mathbf{Qbs}} \urcorner := (X, \sigma\text{-Simple}(X))$

NB: f σ-Simple: $\operatorname{Im} f$ ctbl $\wedge$ $f^{-1}[x] \in B_{\mathbb{R}}$

- Precomposition:

$\alpha : \sigma\text{-simple}(X), \ \varphi : \mathbf{Meas}(\mathbb{R}, \mathbb{R}) \vdash$

$\operatorname{Im}(\alpha \circ \varphi) \subseteq \operatorname{Im} \alpha$ ctbl ✓

$x : X \vdash$

$(\alpha \circ \varphi)^{-1}[x] = \varphi^{-1}[\alpha^{-1}(x)] \in B_{\mathbb{R}}$ ✓

$\alpha^{-1}(x) \in B_{\mathbb{R}}$

$\varphi : \mathbb{R} \to \mathbb{R}$ measurable

Validate qbs axioms for: $\ulcorner X^{\text{Qbs}} \urcorner := (X, \sigma\text{-Simple}(X))$

- recombination:

$\alpha_{-} : (\sigma\text{-Simple}(X))^{I}, \; E_{-} : B_{\mathbb{R}}^{I}, \; \mathbb{R} = \biguplus_{i \in I} E_i \vdash$

NB: f σ-Simple: $\operatorname{Im} f$ ctbl $\wedge$ $f^{-1}[x] \in B_{\mathbb{R}}$

$\operatorname{Im}[E_i . \alpha_i]_{i \in I} \subseteq \bigcup_{i \in I} \operatorname{Im} \alpha_i$ ctbl ✓

$x : X \vdash$

$[E_i . \alpha_i]_{i \in I}^{-1}(x) = \bigcup_{i \in I} \alpha_i^{-1}[x] \cap E_i \in B_{\mathbb{R}}$ ✓

Prop: $X : \mathrm{Set},\ A : \mathrm{Qbs} \vdash$

- $\forall f : X \to \lfloor A \rfloor .\ f : \lceil X \rceil^{\mathrm{Qbs}} \longrightarrow A$

- $\forall f : \lfloor A \rfloor \longrightarrow X .\ f : A \longrightarrow \lfloor X \rfloor_{\mathrm{Qbs}}$

Prop: $X : \mathrm{Set},\ A : \mathrm{Qbs} \vdash$

- $\forall f : X \to \lfloor A \rfloor .\ f : \ulcorner X \urcorner^{\mathrm{Qbs}} \longrightarrow A$

Pf: $\alpha : R_{\ulcorner X \urcorner^{\mathrm{Qbs}}} \vdash$ α σ-simple $\Rightarrow$

$\alpha = [\alpha^{-1}[x].\lambda r.\, x]_{x \in \mathrm{Im}\,\alpha} \Longrightarrow$

$(f \circ \alpha) = [\alpha^{-1}[x].\lambda r.\, fx]_{x \in \mathrm{Im}\,\alpha} \in R_A$ ✓

recombination

Borel

constant $\in B_A$

ctbl

Prop: $X : \mathrm{Set},\ A : \mathrm{Qbs} \vdash$

- $\forall f : X \to \lfloor A \rfloor .\ f : \ulcorner X \urcorner^{\mathrm{Qbs}} \longrightarrow A$
- $\forall f : \lfloor A \rfloor \longrightarrow X .\ f : A \longrightarrow \lfloor X \rfloor_{\mathrm{Qbs}}$

Prf: $\alpha : R_A \vdash (f \circ \alpha : \mathbb{R} \to X) \in R_{\lfloor X \rfloor_{\mathrm{Qbs}}}$ always. ✓

$\blacksquare$

Useful adjunctions:

$$
\begin{tikzcd}
\mathbf{Meas} \arrow[d, "\lfloor - \rfloor_{Qbs}"] \arrow[d, phantom, shift right=4, "\dashv"] \\
\mathbf{Qbs} \arrow[u, bend left=60, "\ulcorner - \urcorner^{Meas}"] \arrow[d, "\lfloor - \rfloor"] \arrow[d, phantom, shift right=4, "\dashv"] \arrow[d, phantom, shift left=4, "\dashv"] \\
\mathbf{Set} \arrow[u, bend left=60, "\ulcorner - \urcorner^{Qbs}"] \arrow[u, bend right=60, "\lfloor - \rfloor_{Qbs}"']
\end{tikzcd}
$$

$\lfloor V \rfloor_{Qbs} := (\lfloor V \rfloor, \mathrm{Meas}(\mathbb{R}, V))$

$(V \in \mathrm{Meas})$

$\ulcorner X \urcorner^{Meas} := \left\{ A \subseteq \lfloor X \rfloor \;\middle|\; \forall \alpha \in R_X .\ \alpha^{-1}[A] \in \mathcal{B}_{\mathbb{R}} \right\}$

- limits (products, subspaces) and colimits (coproducts, quotients) as in Set
- Slogan: every measurable space is carried by a qbs

Example

Product $(X \times Y, \pi_1, \pi_2)$:

necessarily!

- $\lfloor X \times Y \rfloor = \lfloor X \rfloor \times \lfloor Y \rfloor$
- $R_{X \times Y} = \{ \lambda r. (\alpha r, \beta r) \mid \alpha \in R_X, \beta \in R_Y \}$

correlated random elements

rest of structure as in **Set**.

Function Spaces

Straightforward!

- $\lfloor Y^X \rfloor := \mathbf{Qbs}(X, Y)$
- $R_{Y^X} := \mathrm{uncurry}[\mathbf{Qbs}(\mathbb{R} \times X, Y)]$

  $= \{ \alpha : \mathbb{R} \to \lfloor Y^X \rfloor \mid \lambda(r, x).\, \alpha\, r\, x : \mathbb{R} \times X \to Y \}$
- $\mathrm{eval} : Y^X \times X \to Y$

  $\mathrm{eval}(f, x) := f\,x$

Meas vs Qbs

By generalities:

σ-algebra on $\mathbf{Meas}(\mathbb{R}, \mathbb{R})$

$$\ulcorner \mathbb{R}^{\mathbb{R}}_{\mathbf{Meas}} \urcorner \times \mathbb{R} \longrightarrow \ulcorner \mathbb{R}^{\mathbb{R}}_{\mathbf{Meas}} \urcorner \times \mathbb{R} \dashrightarrow \ulcorner \mathbb{R} \urcorner = \mathbb{R}$$

$\ulcorner \mathrm{eval}_{\mathbf{Meas}} \urcorner$

$$\begin{tikzcd} \mathbf{Meas} \arrow[d, shift left, "\lfloor - \rfloor_{\mathbf{Qbs}}"] \arrow[d, phantom, "\dashv"] \\ \mathbf{Qbs} \arrow[u, bend left, "\ulcorner - \urcorner_{\mathbf{Meas}}"] \end{tikzcd}$$

No factorisation by Aumann's Theorem.

(So: $\ulcorner \mathbb{R}^{\mathbb{R}} \times \mathbb{R} \urcorner \not\cong \ulcorner \mathbb{R}^{\mathbb{R}} \urcorner \times \ulcorner \mathbb{R} \urcorner$)

Simple Type Structure

"Simple" because:

- Simply-typed λ-calculus
- types are simple: $A, B : \mathrm{Type} \vdash B^A : \mathrm{Type}$
  - no polymorphism
  - no term dependency
- Contexts for terms: $\Gamma \vdash t : A$ are simple: $\Gamma = x_1 : A_1, \ldots, x_n : A_n$

  i.e. $\mathrm{List}(\mathrm{Type})$

Simple Type Structure

"Simple" because:

- interpretation is simple:

$$[\![x_1 : A_1, \ldots, x_n : A_n]\!] := \prod_{i=1}^{n} A_i$$

$$[\![\Gamma \vdash t : A]\!] : [\![\Gamma]\!] \longrightarrow A$$

in Qbs

Simple Type Structure Curry-Howard-Lambek

$$\frac{\Gamma \vdash t : A \qquad \Gamma \vdash s : B}{\Gamma \vdash \langle t, s\rangle : A \times B} \rightsquigarrow [\![\Gamma]\!] \xrightarrow{\lambda r.\langle tr, sr\rangle} A \times B$$

is measurable

measurability by type!

$$\frac{\Gamma \vdash t : A \times B \qquad \Gamma, x : A, y : B \vdash s : C}{\Gamma \vdash \mathrm{let}\ (x, y) = t\ \mathrm{in}\ s : C} \rightsquigarrow$$

$$[\![\Gamma]\!] \xrightarrow{\lambda r.\ \mathrm{let}\ (a, b) = tr\ \mathrm{in}\ s\, r[x \mapsto a, y \mapsto b]} C$$

is measurable. etc.

Random element Space

$$R_X := X^{\mathbb{R}} \quad \text{since} \quad \lfloor X^{\mathbb{R}} \rfloor = R_X \text{ as sets.}$$

Why?

($\subseteq$) $\alpha \in \lfloor X \rfloor^{\mathbb{R}} \implies \alpha : \mathbb{R} \to X$ in Qbs.

$id_{\mathbb{R}} : \mathbb{R} \to \mathbb{R}$ measurable $\implies id \in R_{\mathbb{R}}$

$\implies \alpha = \alpha \circ id \in R_X$

Precomposition

($\supseteq$) $\alpha \in R_X \implies \forall \varphi \in R_{\mathbb{R}} = \mathrm{Meas}(\mathbb{R}, \mathbb{R}).\ \alpha \circ \varphi \in R_X \implies \alpha : \mathbb{R} \to X$

$\implies \alpha \in \lfloor X^{\mathbb{R}} \rfloor$

Subspaces

For $X \in \mathbf{Qbs}$, $A \subseteq \lfloor X \rfloor$ set:

$$R_A := \{ \alpha : \mathbb{R} \to A \mid \alpha \in R_X \}$$

Then $A = (A, R_A)$ is the subspace qbs

We write $A \hookrightarrow X$

Borel subspaces ensemble

The σ-algebra $\mathcal{B}_X := \{ A \subseteq \lfloor X \rfloor \mid \forall \alpha \in R_X .\ \alpha^{-1}[A] \in \mathcal{B}_{\mathbb{R}} \}$

internalises as $\mathcal{B}_X = 2^X$, the qbs of

Borel subsets.

$\left(\lfloor \mathcal{B}_{(\mathcal{B}_{\mathbb{R}})} \rfloor \right.$ are the Borel-on-Borel sets from descriptive set theory.

Cf. [Sabok et al. '21] $\left. \right)$

Standard Borel spaces

Def: A qbs S is standard Borel when

$$S \cong A \quad \text{for some } A \in \mathcal{B}_{\mathbb{R}}$$

$$\begin{tikzcd}
\mathbf{Meas} \arrow[rr, bend left] \arrow[rr, phantom, "\top"] & & \mathbf{Qbs} \arrow[ll, bend left] \\
\mathbf{Sbs} \arrow[u, hook] \arrow[rr, bend left] \arrow[rr, phantom, "\cong"] & & \mathbf{Sbs} \arrow[u, hook] \arrow[ll, bend left]
\end{tikzcd}$$

Slogan: Qbs conservative extension of Sbs

Example $C_0 := \{ f : \mathbb{R} \to \mathbb{R} \mid f \text{ continuous} \} \hookrightarrow \mathbb{R}^{\mathbb{R}}$

C_0 is sbs. (Well-known!)

Proof:

$C_0' \in B_{\mathbb{R}^{\mathbb{Q}}}$ sbs!

$$C_0' := \left\{ g \in \mathbb{R}^{\mathbb{Q}} \;\middle|\; \begin{array}{l} \forall a, b \in \mathbb{Q}, \varepsilon \in \mathbb{Q}^{+} \\ \exists \delta \in \mathbb{Q}^{+}\ \forall p, q \in \mathbb{Q}^{+} \cap [a,b], \\ |p - q| < \delta \Rightarrow |g\,p - g\,q| < \varepsilon \end{array} \right\}$$

on closed intervals (= compact intervals)

Continuity $\Updownarrow$ uniform continuity

Borel measurable } by type checks

then $C_0 \cong C_0' \in B_{\mathbb{R}^{\mathbb{Q}}}$:

$$\begin{array}{ll} C_0 \to C_0' & C_0' \to C_0 \\ \varphi \mapsto \varphi|_{\mathbb{Q}} & \varphi \mapsto \lambda r.\ \lim\limits_{n\to\infty} g(\mathrm{approx}_{(\frac{1}{m})_{m\in\mathbb{N}}}\, r)_n \end{array}$$

Example (ctd)

C_0 is sbs, and $eval : C_0 \times \mathbb{R} \to \mathbb{R}$

is measurable.

Avoids:

- constructing complete separable metrics
- proving that evaluation is measurable w.r.t. metric σ-algebra.

Non-examples

$\rightsquigarrow$ [Sabok et al. '21]

- $\{ A \in \mathcal{B}_{\mathbb{R}} \mid A \neq \emptyset \} \hookrightarrow \mathcal{B}_{\mathbb{R}}$

- $\{ (A_1, A_2) \in \mathcal{B}_{\mathbb{R}}^2 \mid A \subseteq B \} \hookrightarrow \mathcal{B}_{\mathbb{R}}^2$

- $\{ A \in \mathcal{B}_{\mathbb{R}} \mid A \text{ open} \} \hookrightarrow \mathcal{B}_{\mathbb{R}}$

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

Dependent Type Structure

types can contain terms:

a type referring to a term

$$X : \mathrm{Type},\ E : B_X \vdash \{x \in X \mid x \in E\} : \mathrm{Type}$$

a type, just like STLC

a term!

Dependent Type Structure

types can contain terms:

a type referring to a term

$$X : \mathrm{Type},\ E : B_X \vdash \{x \in X \mid x \in E\} : \mathrm{Type}$$

a type, just like STLC

a term!

Context formation:

$$\frac{\Gamma \vdash A : \mathrm{Type}}{\Gamma, x : A \vdash}$$

Dependent Type Structure

types denote spaces-in-context

$$\begin{tikzcd} [\![\Gamma \vdash A]\!] \arrow[d, "\mathrm{dep}"] \\ [\![\Gamma]\!] \end{tikzcd}$$

Dependent types denote spaces-in-Context

$\Gamma \vdash$ Context

$\Gamma \vdash A$

type in context

E.g.:

A

$\downarrow$

$\mathbb{1}$

simple types

$[\![\Gamma \vdash A]\!]$ Space in context

$\downarrow$ dep

assigns environment

$[\![\Gamma]\!]$ context space

$$[\![E : B_A \vdash \{x \in A \mid x \in E\}]\!]$$

$$\text{ii}$$

$$\{ (E, a) \in B_A \times A \mid a \in E \}$$

$$\downarrow \pi_1$$

$$B_A$$

decoder

Content extension

$$\frac{\Gamma \vdash A}{\Gamma, a : A \vdash}$$

$$\llbracket \Gamma \vdash A \rrbracket \xrightarrow{\mathrm{dep}} \llbracket \Gamma \rrbracket \qquad \llbracket \Gamma, a : A \rrbracket := \llbracket \Gamma \vdash A \rrbracket$$

Substitution

$\Gamma \vdash \sigma : \Delta$ $\qquad \llbracket \sigma \rrbracket : \llbracket \Gamma \rrbracket \longrightarrow \llbracket \Delta \rrbracket$

E.g. weakening $\qquad \Gamma, a : A \vdash \mathrm{wkn} : \Gamma$

$$\llbracket \Gamma, a : A \rrbracket := \llbracket \Gamma \vdash A \rrbracket \xrightarrow[\mathrm{dep}]{\mathrm{wkn}} \llbracket \Gamma \rrbracket$$

Action of substitution on types

$$
\begin{tikzcd}
\llbracket \Gamma \vdash A[\sigma] \rrbracket \\[-2em]
\rotatebox{90}{$=$} \\[-2em]
\{(\gamma, a) \in \llbracket \Gamma \rrbracket \times \llbracket A \rrbracket \mid \mathrm{dep}\, a = \gamma\} \arrow[r, "\pi_2"] \arrow[d, "\mathrm{dep} := \pi_1"'] \arrow[dr, phantom, "\lrcorner", very near start] & \llbracket \Delta \vdash A \rrbracket \arrow[d, "\mathrm{dep}"] \\
\llbracket \Gamma \rrbracket \arrow[r, "\sigma"] & \Delta
\end{tikzcd}
$$

E.g.

$$
\begin{tikzcd}
\llbracket x : A \vdash B[\mathrm{wkn}] \rrbracket := A \times B \arrow[r, "\pi_2"] \arrow[d, "\mathrm{dep} := \pi_1"'] \arrow[dr, phantom, "\lrcorner", very near start] & B \arrow[d] \\
x : A \arrow[r, "\langle\rangle"'] & \mathbb{1}
\end{tikzcd}
$$

simple type

Terms : sections

$$
\begin{tikzcd}
{[\![\Gamma]\!]} \arrow[r, "{[\![\Gamma \vdash M : A]\!]}"] \arrow[dr, equal, "="'] & {[\![\Gamma \vdash A]\!]} \arrow[d, "\mathrm{dep}"] \\
& {[\![\Gamma]\!]}
\end{tikzcd}
$$

e.g.

$$
\begin{tikzcd}
\mathbb{R} \arrow[r, "{[\![x : \mathbb{R} \vdash [x, \infty) : \mathcal{B}_{\mathbb{R}}[\mathrm{wkn}]]\!]}"] \arrow[dr, equal, "="'] & \mathbb{R} \times \mathcal{B}_{\mathbb{R}} \arrow[d, "\pi_1"] \\
& \mathbb{R}
\end{tikzcd}
$$

E.g. Variables:

$$
\begin{tikzcd}
{[\![\Gamma, a : A]\!]} \arrow[r, "{[\![\Gamma, a : A \vdash a : A]\!]}", "{\langle \mathrm{id}, \mathrm{dep}_{\Gamma \vdash A} \rangle}"'] \arrow[dr, equal, "="'] & {[\![\Gamma, a : A \vdash A[\mathrm{wkn}]]\!]} \arrow[d, "\mathrm{dep}"] \\
& {[\![\Gamma, a : A]\!]}
\end{tikzcd}
$$

Exercise:
action of substitution
$M[\sigma]$

Dependent Pairs

$$\frac{\Gamma, a : A \vdash B}{\Gamma \vdash \coprod_{a \in A} B}$$

$$\begin{tikzcd}
\llbracket \coprod_{a \in A} A \rrbracket \arrow[ddr, bend right, "\mathrm{dep}_{\coprod}"'] \arrow[ddr, phantom, ":="] & := \llbracket \Gamma, a : A \vdash B \rrbracket \arrow[d, "\mathrm{dep}_B"] \\
& \llbracket \Gamma, a : A \rrbracket = \llbracket \Gamma \vdash A \rrbracket \arrow[d] \\
& \llbracket \Gamma \rrbracket
\end{tikzcd}$$

Dependent products

$$\frac{\Gamma, a : A \vdash B}{\Gamma \vdash \prod_{a \in A} B}$$

aka: $(a : A) \to B$

$$\lfloor [\![\Gamma \vdash \prod_{a \in A} B]\!] \rfloor :=$$

$$\left\{ (r_0, f : \{ a \in [\![A]\!] \mid \mathrm{dep}\, a = r_0 \} \to [\![\Gamma, a : A \vdash B]\!]) \;\middle|\; \forall a \in [\![\Gamma, a : A]\!] .\ \mathrm{dep}\, a = r_0 \Rightarrow \mathrm{dep}(f a) = a \right\}$$

Exercise: find the random elements.

Full model

type : Qbs $\quad W := [0,\infty] \quad \mathbb{B}X \cong \mathbb{B}^X$

$DX := (\text{Fri})$

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

$\underset{\mu}{\text{Ce}}[E] := (\text{Fri}) \qquad \delta_x := (\text{Fri})$

$\oint \mu k := (\text{Fri})$

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

Partiality cf. [Vákár et al. '19]

A **Borel embedding** $e : X \hookrightarrow Y$

- injective function $e : \lfloor X \rfloor \rightarrowtail \lfloor Y \rfloor$
- its image is Borel: $e[\lfloor X \rfloor] \in \mathcal{B}_Y$
- e is **Strong**: $\alpha \in R_X \iff e \circ \alpha \in R_Y$

Examples

- $\mathbb{1} \hookrightarrow \mathbb{2}$
- S is sbs $\iff \exists\, S \hookrightarrow \mathbb{R}$

Def: A Partial map $f : X \rightharpoonup Y$ is a morphism

$$f : X \to Y \amalg \{\bot\}$$

Its domain of definition

$$f : (Y \amalg \{\bot\})^X \vdash \mathrm{Dom}\, f := \{x \in X \mid f x \neq \bot\} : \mathrm{Type}$$

Depent-type interpretation

$$\begin{tikzcd}
\llbracket \mathrm{Dom}\, f \rrbracket \arrow[rr] \arrow[d, "\mathrm{dep}"] \arrow[drr, phantom, "\lrcorner", very near start] & & \{y \in Y \mid y \in E\} \arrow[d, "\mathrm{dep}"] \\
\llbracket f : (Y \amalg \{\bot\})^X \rrbracket \arrow[rr, "{\llbracket E \mapsto \lambda x.\, f x \neq \bot \rrbracket}"] & & \llbracket E : \mathbb{B}_Y \rrbracket
\end{tikzcd}$$

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

Full model

$\text{type} : \mathbf{Qbs} \qquad \mathbb{W} := [0,\infty] \qquad \mathbb{B}X := \mathbb{B}^X$

$DX := \square$

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

$\underset{\mu}{Ce}[E] := \square \qquad \delta_x := \square$

$\oint \mu k := \square$

Def: A measure μ over $\mathbb{R}$ is a function

$$\mu : B_{\mathbb{R}} \to W := [0, \infty]$$

satisfying the measure axioms:

$$E : B^{\omega} \vdash$$

$$\mu \emptyset = 0, \quad \mu E = \mu(E \cap F) + \mu(E \cap F^{c}), \quad \mu\left(\bigcup_{n} E_n\right) = \sup_{n} \mu E_n$$

For measurable spaces, replace $\mathbb{R}$ with V

We write $\lfloor G V \rfloor$ for the set of measures on V

For qbs X, take $\lfloor G \ulcorner X \urcorner^{\mathrm{Meas}} \rfloor$

Thm (Lebesgue measure):

There is a unique measure $\lambda \in \lfloor G\mathbb{R} \rfloor$ s.t.:

$$\lambda(a, b) = b - a$$

The unrestricted Giry spaces

Equip $\lfloor GV \rfloor$ with two qbs structures:

X $\quad R_{GV} := \{\alpha : \mathbb{R} \to GV \mid \forall A \in B_V . \lambda r. \alpha(r, A) : \mathbb{R} \to \mathbb{W}\}$

↳ α is a kernel.

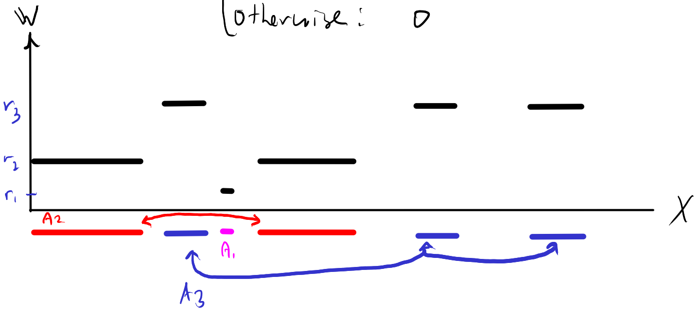
✓ $\quad GV \hookrightarrow \mathbb{W}^{B_X}$

- Fewer random elements $R_{GV} \subseteq R_{G'V}$
- Lebesgue integral measurable in both arguments. (upcoming)

Def: Simple function $\varphi : X \to W$ when

$\exists\, n \in \mathbb{N},\ \vec{A} \in B_X^n,\ A_i \cap A_j = \phi\ (i \neq j),\ \vec{r} \in W$ s.t.

$$\varphi\, x = \begin{cases} \vdots \\ x \in A_i : & r_i \\ \vdots \\ \text{otherwise} : & 0 \end{cases}$$



Encode into a space:

$$\mathrm{SimpleCode} := \coprod_{n \in \mathbb{N}} B_X^n \times W^n$$

$$\mathrm{Simple} := \{ f \in W^X \mid f \text{ simple} \} \hookrightarrow W^X$$

and define an interpretation:

$$[\![-]\!] : \mathrm{SimpleCode} \twoheadrightarrow \mathrm{Simple}$$

$$[\![(n, \vec{A}, \vec{r})]\!] := \sum_{i=1}^{n} r_i \cdot [- \in A_i]$$

↳ characteristic function for A_i

Lemma: $f : X \to W$ is measurable → remember! $\mathbf{Qbs}$ morphism!

iff $f = \lim_{n\to\infty} f_n$ for some monotone sequence

$f_n \in \mathrm{Simple}$.

Moreover, we have measurable such choice:

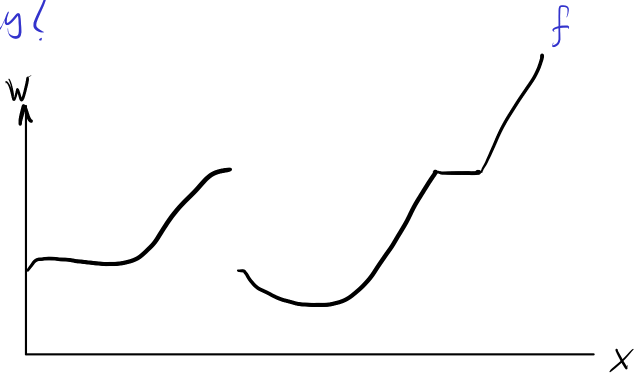
Simple Approx:

$$\{\vec{\Delta} \in \mathbb{R}^+ \mid \Delta_n \to 0\} \times \{\vec{a} \in W^{\mathbb{N}} \mid \vec{a} \text{ monotone}, a_n \to \infty\} \times W^X \longrightarrow \mathrm{SimpleCode}$$

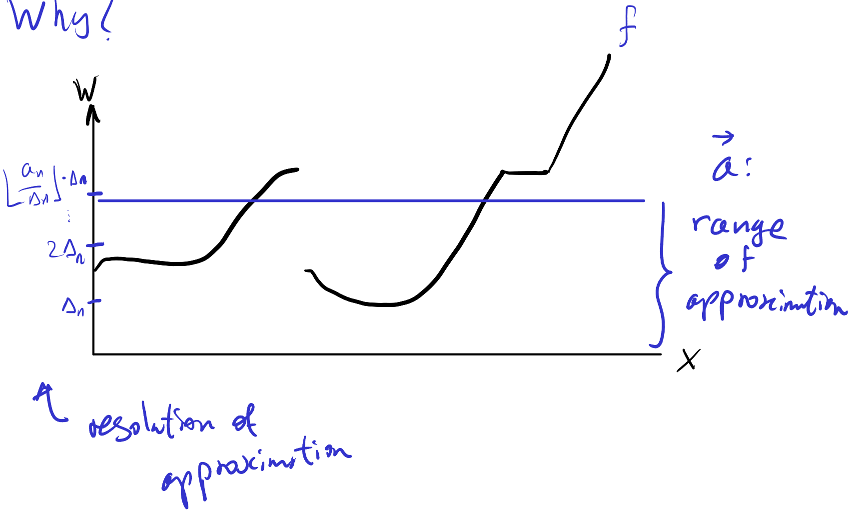
rate of convergence

range of approximation

Why?

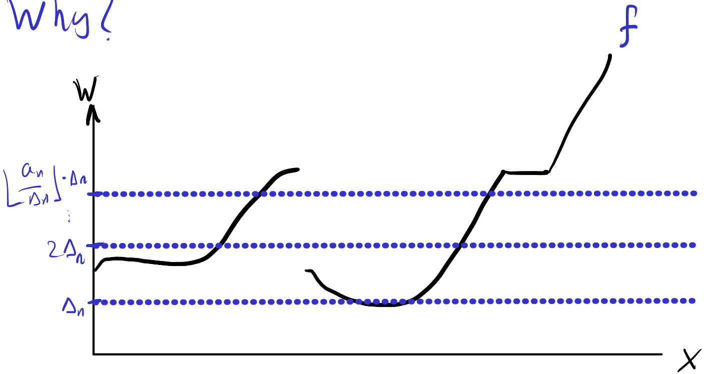


Why?



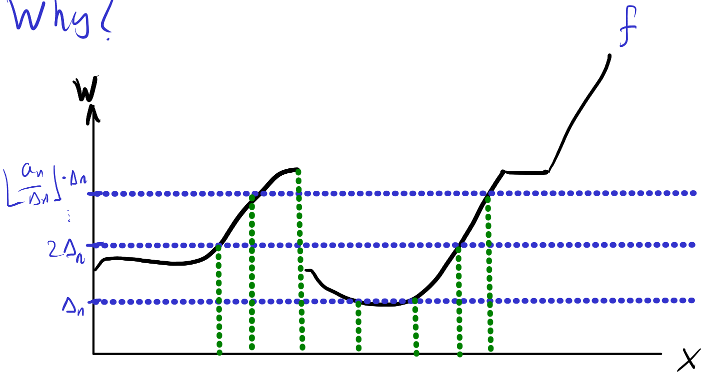
f
W
$\lfloor \frac{a_n}{\Delta_n} \rfloor \cdot \Delta_n$
$2\Delta_n$
Δ_n
$\vec{a}$:
range of f approximation
X
resolution of approximation

Why?



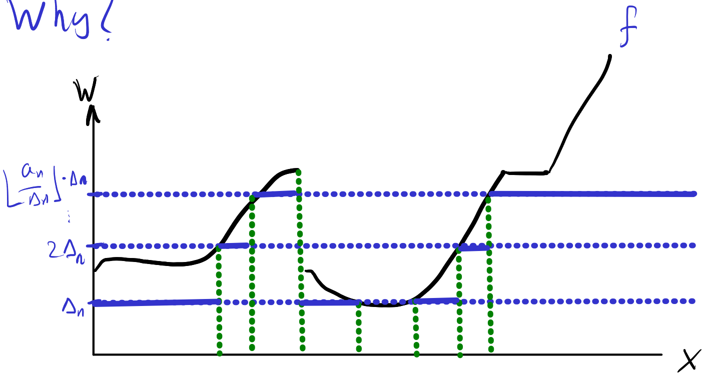
W
f
$\left\lfloor \frac{a_n}{\Delta_n} \right\rfloor \cdot \Delta_n$
$2\Delta_n$
Δ_n
X

Why?



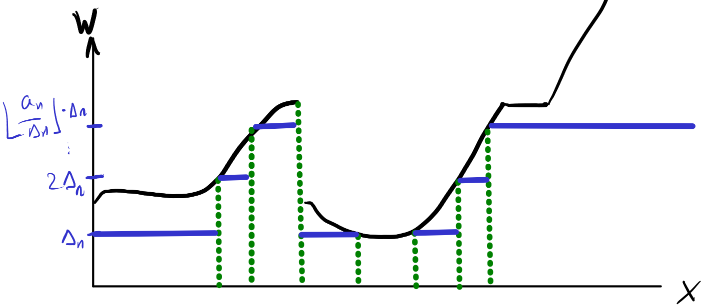
f
w
$\lfloor \frac{a_n}{\Delta_n} \rfloor \cdot \Delta_n$
⋮
$2\Delta_n$
Δ_n
X

Why?



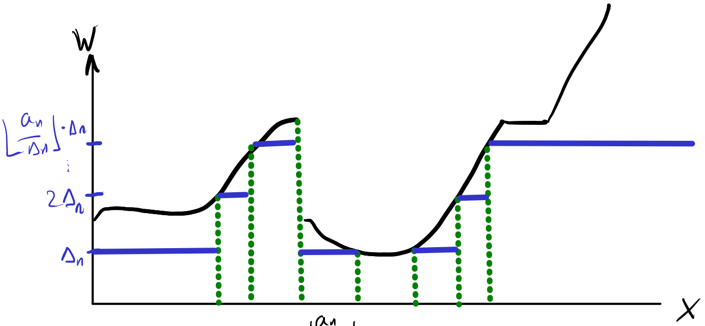
f
W
$\lfloor \frac{a_n}{\Delta_n} \rfloor \cdot \Delta_n$
$\vdots$
$2\Delta_n$
Δ_n
X

Why?



w
$\lfloor \frac{a_n}{\Delta_n} \rfloor \cdot \Delta_n$
$2\Delta_n$
Δ_n
f
X

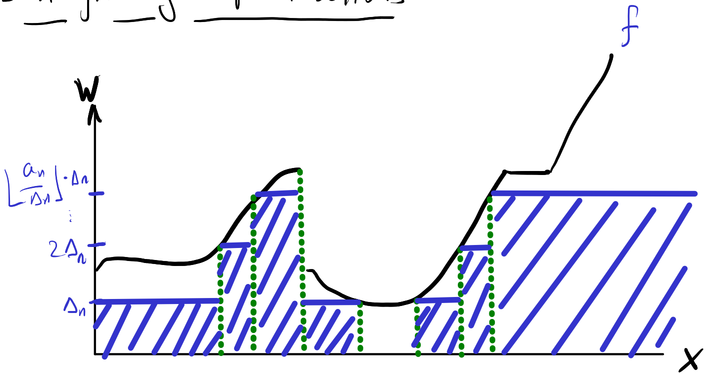
Why?



$$[\![\mathrm{SimpleApprox}_{\vec{\Delta},\vec{a}}\, f]\!] := \sum_{i=1}^{\lfloor \frac{a_n}{\Delta_n} \rfloor} i \cdot \Delta_n \left[i \cdot \Delta_n \leq f < (i+1)\Delta_n \right]$$

$$+ \left\lfloor \frac{a_n}{\Delta_n} \right\rfloor \Delta_n \cdot \left[f \geq \left\lfloor \frac{a_n}{\Delta_n} \right\rfloor \cdot \Delta_n \right] \in \mathrm{Simple}$$

Integrating Simple Functions



$$\int : GX \times \mathrm{SimpleCode} \longrightarrow W$$

$$\int \mu\,(n, \vec{A}, \vec{r}) := \sum_{I \subseteq \{1,\ldots,n\}} \Big(\sum_{i \in I} r_i\Big) \cdot \mu\Big(\bigcap_{i \in I} A_i \setminus \bigcup_{i \notin I} A_i\Big)$$

Integration

Proper higher-order operation

$$\int : GX \times \mathbb{W}^X \longrightarrow \mathbb{W}$$

$$\int \mu f := \sup \left\{ \int \mu \varphi \;\middle|\; \varphi \in \mathrm{Simple},\ \varphi \le f \right\}$$

$$= \lim_{n \to \infty} \int \mu \left(\mathrm{SimpleApprox}_{\vec{\Delta}, \vec{a}}\, f\right)_n$$

measurable by type

we also write $\int \mu(\mathrm{d}x)\, t$ for $\int \mu\, (\lambda x. t)$

for $\frac{a_n}{\Delta_n} \longrightarrow 0$, eg. $\Delta_n = \frac{1}{2^n}$ $a_n = n$.

resolution

The unrestricted Giry Strong Monad

Unlike the unrestricted Giry on Meas.

Dirac:

$$\delta : X \to GX$$
$$x \mapsto \lambda A. \begin{cases} x \in A : & 1 \\ x \notin A : & 0 \end{cases}$$

but: non-commutative

(Fubini fails, just like in Meas)

Kleisli extension/Kock integral:

$$\oint : GX \times GY^X \longrightarrow GY$$

$$\oint \mu f := \lambda A. \int \mu(dx)\, f(x; A)$$

Fubini-Tonelli fails

$\# \in G\mathbb{R}$ $\quad \#E := \begin{cases} E \text{ finite}: & |E| \\ \text{o.w.}: & \infty \end{cases}$

$\lambda \in G\mathbb{R}$ lebesgue

$k : \mathbb{R}\times\mathbb{R} \to W \cong G\mathbb{1}$

$k(x,y) := [x=y]$

$$\oint \#(dr) \oint \underbrace{\lambda(dx)\, k(x,y)} = \oint \#\, \underline{0} = \underline{0} \cong 0$$

$$y : \mathbb{R} \vdash \{\langle\rangle\} \mapsto \lambda\{y\}\cdot 1 + \lambda\{y\}^{c}\cdot 0 = 0$$

$\#$

$$\oint \lambda(dx) \oint \underbrace{\#(dr)\;\; k(x,y)} = \oint \lambda(dx)\, \delta_{\langle\rangle} \cong \infty$$

$$x : \mathbb{R} \vdash \{\langle\rangle\} \mapsto \#\{x\}\cdot 1 + 0 = 1$$

Randomisable measures monad

$D \rightarrowtail G$

$\lambda A. \int_{\mathrm{Dom}\,\alpha} \lambda(\mathrm{Dom}\,\alpha)$

$\lfloor DX \rfloor := \{ \lambda_\alpha \mid \alpha : \mathbb{R} \rightharpoonup X \}$

Lebesgue measure

$R_{DX} := \{ \lambda x. \lambda_{dx} \mid \alpha : \mathbb{R} \times \mathbb{R} \rightharpoonup X \}$

$\delta : X \to DX \qquad \oint : D\Gamma \times (DX)^{\Gamma} \longrightarrow DX$ lift along $D \rightarrowtail G$.

D validates our measure axioms including Fubini-Tonelli:

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

$$\oint \mu(dx) \oint \nu(dy)\, \delta_{(x,y)} = \oint \nu(dy) \oint \mu(dx)\, \delta_{(x,y)} =: \mu \otimes \nu$$

Thm: For sbs S, $PS, D_{\leq 1}S, D_{<\infty}S \in Sbs$

and agree with their counterparts on **Meas**.

$DS_J = \{\mu \mid \mu \text{ s-finite}\}$

$R_{DS} = \{k : \mathbb{R} \to GD \mid k \text{ s-finite kernel}\}$

See [Staton'16]

Open: Is there a counterpart to D in **Meas**?

More modestly, is $DS \in Sbs$?

(Hypothesis: no)

Distribution Submonads

A measure space

$\Omega = (\Omega, \mu)$

is a qbs Ω with

$\mu \in DX$.

Similarly: finite measure space

– (sub) probability space.

$$
\begin{array}{l}
PX := \{\mu \in DX \mid \mu X = 1\} \\
\quad \downarrow \\
P_{\leq 1} X := \{\mu \in DX \mid \mu X \leq 1\} \\
\quad \downarrow \\
P_{<\infty} X := \{\mu \in DX \mid \mu X < \infty\} \\
\quad \downarrow \\
DX
\end{array}
$$

Full model

type : Qbs $\quad W := [0,\infty] \quad \mathbb{B}X \cong \mathbb{B}^X$

$$DX := \left(\{\lambda_\alpha \mid \alpha : \mathbb{R} \to X\}, \{\lambda r.\lambda_{\alpha(r,-)} \mid \alpha : \mathbb{R}\times\mathbb{R} \to X\} \right)$$

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

$$\underset{\mu}{\mathcal{E}}[E] := \mu E \qquad \delta_x := E \mapsto \begin{cases} x \in E : 1 \\ x \notin E : 0 \end{cases}$$

$$\oint \mu k := \lambda E. \int \mu(dx) k(x; E)$$

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

Random variable: $\xi : \Omega \longrightarrow \Theta \hookrightarrow \overline{\mathbb{R}}$

Θ:

$$
\begin{tikzcd}
 & & \underset{[0,\infty]}{\mathbb{W}} \arrow[dr, hook] & \\
\underset{[0,1]}{\mathbb{I}} \arrow[r, hook] & \underset{[0,\infty)}{\mathbb{R}_{\geq 0}} \arrow[ur, hook] \arrow[dr, hook] & & \underset{[-\infty,\infty]}{\overline{\mathbb{R}}} \\
\underset{(0,\infty)}{\mathbb{R}^{+}} \arrow[ur, hook] & & \underset{(-\infty,\infty)}{\mathbb{R}} \arrow[ur, hook] &
\end{tikzcd}
$$

- Θ^{Ω} is a space

- $\mathbb{R}^{\Omega}$ measurable vector space:

$$\alpha\xi + \zeta := \lambda\omega.\ \alpha\cdot\xi\omega + \zeta\omega$$

- $\mathbb{W}^{\Omega}$ measurable σ-semi-module for $\mathbb{W}$: $\sum_{n=0}^{\infty} \alpha_n \xi_n :=$

$$\lambda\omega.\ \sum_{n=0}^{\infty} \alpha_n \cdot \xi_n$$

$$\Pr_{-} : P\Omega \times B_\Omega \to W$$

$$\Pr_\lambda A := \mathrm{eval}(\lambda, A) = \lambda A$$

Probability Space $\Omega = (\Omega, \lambda_\Omega)$

$P : P\Omega \vdash$ "$P x$ holds $\lambda(dx)$-almost surely"

for some $Q \hookrightarrow \Omega$, $P \supseteq Q$, $[- \in Q] \cdot \lambda = \lambda$

Example $(\xi, \zeta \in \mathbb{H}^\Omega)$

$\xi = \zeta$ a.s. when $\Pr_{\omega \sim \lambda} [\xi \omega \neq \zeta \omega] = 0$

Integrating random variables (as discretely) in Qbs!

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

$\xi_+ := \max(\xi, 0) \qquad \xi_- := \max(-\xi, 0)$

So: $\xi = \xi_+ - \xi_-$

$\int : P\Omega \times W^\Omega \longrightarrow W$

$\int \lambda\, \xi := \int \lambda\, \xi_+ - \int \lambda\, \xi_-$

$\int$ respects a.s. equality:

$\xi = \zeta \ (\lambda\text{-a.s})$

$\Rightarrow \int \lambda\, \xi = \int \zeta.$

Example

$$\lambda : P\Omega \vdash \mathrm{ASConverge}(\overline{\mathbb{R}})^{\Omega} : \mathcal{B}\left(\overline{\mathbb{R}}^{\mathbb{N}\times\Omega}\right)$$

$$:= \left\{ \vec{z} \in \overline{\mathbb{R}}^{\mathbb{N}\times\Omega} \;\middle|\; \Pr_{\omega\sim\lambda}\left[\lim z_n \omega \neq \bot\right] \right\}$$

So:

$$\lim^{as} : \overline{\mathbb{R}}^{\mathbb{N}\times\Omega} \longrightarrow \overline{\mathbb{R}}^{\Omega} \qquad \mathrm{Dom}\,\lim^{as} := \mathrm{ASConverge}(\overline{\mathbb{R}})^{\Omega}$$

$$\lim^{as} \vec{z} := \lambda\omega.\ \limsup_{n\to\infty} f_n\,\omega$$

$\lim^{as}$ respects a.s. equality.

Thm (monotone convergence):

Let $\vec{\xi} \in W^{\mathbb{N}\times\Omega}$ λ-a.s. monotone.

$$\xi = \lim_{n\to\infty}^{as} \xi_n \qquad (a.s.)$$

$$\Downarrow$$

$$\int \lambda\, \xi = \lim_{n\to\infty} \int \lambda\, \xi_n$$

Lebesgue space

$\left(\begin{array}{l} \Omega \text{ prob. space,} \\ \quad p \in [1, \infty) \end{array} \right)$

$p : [1, \infty), \lambda : P\Omega \vdash \mathcal{L}^{p}_{(\Omega, \lambda)} : \mathcal{B}(\mathbb{R}^{\Omega})$

$$:= \left\{ \xi \in \mathbb{R}^{\Omega} \,\middle|\, \int \lambda |\xi|^{p} < \infty \right\} \hookrightarrow \mathbb{R}^{\Omega}$$

Ensemble $\mathcal{L}_{\Omega} := \prod_{\substack{\lambda \in P\Omega \\ p \in [1, \infty)}} \mathcal{L}^{p}_{(\Omega, \lambda)}$

$\llcorner \quad p \leq q \Rightarrow \mathcal{L}^{p}_{\Omega} \supseteq \mathcal{L}^{q}_{\Omega}$

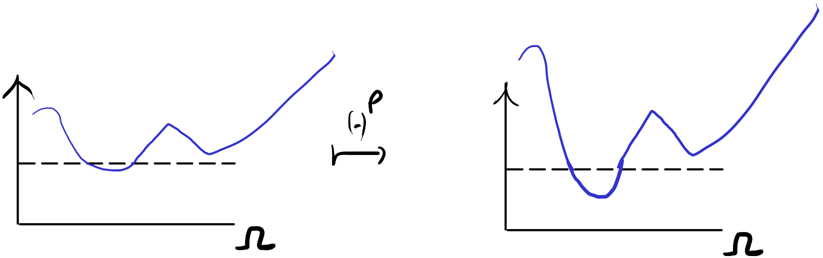
$\mathcal{L}^p$ semi norms

$\|-\| : \coprod_{p,\lambda} \mathcal{L}^p_{(\Omega,\lambda)} \longrightarrow \mathbb{R}_{\geq 0}$ $\quad \|\xi\|_p := \sqrt[p]{\int \lambda |\xi|^p}$

$\mathcal{L}^2$ inner product

$\langle -, - \rangle : \coprod_{p,\lambda} \mathcal{L}^p_{(\Omega,\lambda)} \times \mathcal{L}^p_{(\Omega,\lambda)} \longrightarrow \mathbb{R}$

$\langle \xi, \zeta \rangle_{p,\lambda} := \int \lambda \, \xi \zeta$



Statistics

Expectation

$\mathbb{E} : \coprod_\lambda \mathcal{L}^1 \longrightarrow \mathbb{R}$

$\mathbb{E}_\lambda \xi := \int \lambda \xi$

Covariance and Correlation

$\mathrm{Cov}, \mathrm{Corr} : \coprod_\lambda \mathcal{L}^2 \longrightarrow \mathbb{R}$

$\mathrm{Cov}(\xi, \zeta) := \langle \xi - \mathbb{E}\xi, \zeta - \mathbb{E}\zeta \rangle$

$\mathrm{Corr}(\xi, \zeta) := \dfrac{\langle \xi, \zeta \rangle}{\|\xi\|_2 \cdot \|\zeta\|_2} = \cos(\mathrm{angle}(\xi, \zeta))$

Sequential limits

$p : [1,\infty)$, $+ : P \times +$ Cauchy $\mathcal{L}^p_{(\Omega,\mu)} \subseteq \mathcal{B}\left(\mathcal{L}^p_{(\mathbb{R},\cdot)}\right)^{\mathbb{N}}$

$$:= \left\{ \vec{\xi} \;\middle|\; \forall \varepsilon \in \mathbb{Q}^+ \,\exists k \in \mathbb{N}\, \forall m,n \geq k.\; \|\xi_{k+n} - \xi_{k+m}\|_p < \varepsilon \right\}$$

Thm: $\mathcal{L}^p_\Omega$ is Cauchy-Complete

$$\lim^p : \text{Cauchy}\,\mathcal{L}^p \longrightarrow \mathcal{L}^p$$ (convergence in mean)

Why?

1. Every Cauchy sequence has an a.s. converging subseq.
2. We can find it measurably

Example

Thm (dominated convergence)

For $\vec{\xi}, \zeta \in \mathcal{L}^1$ s.t. $\xi_n \leq \zeta$ a.s.:

1. $\lim^{as} \vec{\xi} \in \mathcal{L}^1$
2. $\lim^1 \vec{\xi} = \lim^{as} \vec{\xi}$
3. $\lim_{n \to \infty} \int \lambda\, \xi_n = \int \lambda \lim_{n \to \infty} \xi_n$

Separability

Def: $\mathcal{L}^p$ separable: has contable dense subset

Fact: Separability is property of λ_Ω:

TFAE:

- $\exists p \geq 1.\ \mathcal{L}^p$ separable
- $\forall p \geq 1.\ \mathcal{L}^p$ separable

Measurable separability in $I \hookrightarrow P\Omega \times [1,\infty)$

$$\vec{\beta} : \prod_{(\lambda,p)\in I} \mathcal{L}^{p}_{(\Omega,\lambda)}{}^{\mathbb{N}} \quad \text{s.t.}$$

$$\left\{ \vec{\beta}^{\lambda,p}_{n} \mid n \in \mathbb{N} \right\} \text{ dense in } \mathcal{L}^{p}_{(\Omega,\lambda)}$$

Prop. - every sbs S measurably separable in

$$PS \times [1,\infty)$$

- $I \hookrightarrow P\Omega \times \{2\}$ measurably separable

$$\Rightarrow \exists\, \vec{\beta} \in \prod_{\lambda\in I} \mathcal{L}^{2}_{(\Omega,\lambda)} \text{ orthonormal system}$$

$\langle \beta_n, \beta_m \rangle = 0$

$\|\beta_n\|_2 = 1$

(β_n) dense

Example

Let $S \hookrightarrow \mathcal{L}^2$ closed vector subspace.

orthogonal decomposition — linear in fact.

$$\langle P, P^{\perp}\rangle : \lfloor \mathcal{L}^P \rfloor \to \lfloor S \rfloor \times \lfloor S^{\perp} \rfloor$$

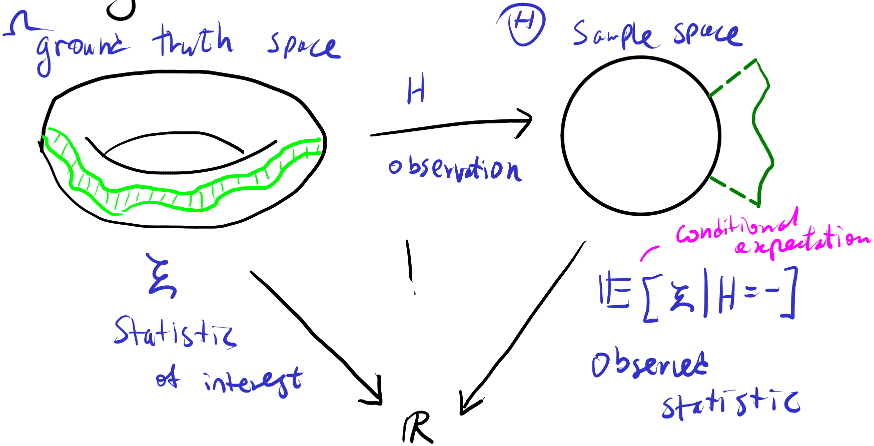
When S is separable with orthonormal system β

We have a measurable version of

$$\langle P, P^{\perp}\rangle : \mathcal{L}^2 \to S \times S^{\perp}$$

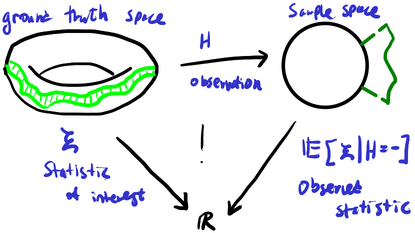
$$P\xi := \sum_{n=0}^{\infty} \langle \xi, \beta_n\rangle \beta_n \qquad P^{\perp} := \mathrm{Id} - P.$$

Kolmogorov's Conditional Expectation



Kolmogorov's Conditional Expectation

A conditional expectation of $\xi \in \mathcal{L}^1_\Omega$ wrt $H : \Omega \to \Theta$ is $Z \in \mathcal{L}^1_{\Theta}$ s.t. for all $A \in \mathcal{B}_{\Theta}$:



$$\int_A \mu Z = \int_{H^{-1}[A]} \lambda \xi$$

where $\mu := \lambda_H$

Conditional expectations

1. unique a.s.

2. fundamental to modern probability, eg:

   a martingale

$$
\begin{tikzcd}
\cdots & \Theta_n \arrow[l] \arrow[dr, "Z_n"'] & & \Theta_{n+1} \arrow[ll, "H_n"'] \arrow[dl, "Z_{n+1}"] \arrow[dl, phantom, shift right=8, "!"] & \cdots \arrow[l] \\
& & \mathbb{R} & &
\end{tikzcd}
$$

s.t. $Z_n = \mathbb{E}[Z_{n+1} \mid H_n = -]$

Thm (Existence)

- $\exists$ $\mathbb{E}[- \mid H = -] : \lfloor \mathcal{L}^1_{(\Omega,\lambda)} \rfloor \to \lfloor \mathcal{L}^1_{(\Theta,\mu)} \rfloor$

- When (Ω,λ) is separable
  $\mathbb{E}[- \mid H = -] : \mathcal{L}^1_{(\Omega,\lambda)} \to \mathcal{L}^1_{(\Theta,\mu)}$

- When Θ is $\mathcal{I}^1$-measurably separable
  $\mathbb{E}[- \mid - = -] : \prod_{\substack{H \in \Theta^\Omega \\ \lambda \in H_*^{-1}[\mathcal{I}]}} \mathcal{L}^1_{(\Omega,\lambda)} \to \mathcal{L}^1_{(\Theta,\mu)}$

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

Discrete model

type : Set $\qquad W := [0,\infty] \qquad BX := PX$

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

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

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

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

Full model

type : Qbs $\quad W := [0,\infty] \quad \mathbb{B}X \cong \mathbb{B}^X$

$$DX := \left(\{\lambda_\alpha \mid \alpha : \mathbb{R} \to X\}, \{\lambda r. \lambda_{\alpha(r,-)} \mid \alpha : \mathbb{R}\times\mathbb{R} \to X\} \right)$$

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

$$\underset{\mu}{Ce}[E] := \mu E \qquad \delta_x := E \mapsto \begin{cases} x \in E : 1 \\ x \notin E : 0 \end{cases}$$

$$\oint \mu k := \lambda E. \int \mu(dx) k(x;E)$$

Enough!

Lunch.