

# Foundations for Type-Driven Probabilistic Modelling

Ohad Kammar  
University of Edinburgh

Logic and Semantics Group  
3 February, 2026  
Department of Computer Science  
University of Aarhus



Laboratory for Foundations  
of Computer Science



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.



Two implementations/models

**discrete model**

familiar maths  
introductory



**full model**

supports discrete  
and  
continuous distributions  
same language

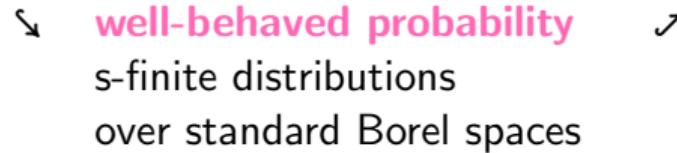
# Motivation: why foundations?

## discrete probability

countably supported distributions  
good type-structure  
**(this course)**

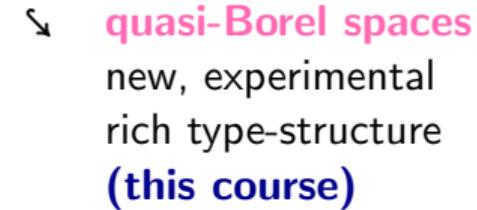
## measure theory

standard, established  
poor type-structure



## continuous probability

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



## Takeaway

Use types to abstract away from the model

# Motivation: why types?

- ▶ **spotlights** meaningful operations

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

- ▶ document **intent**:  
probability (**Distribution**  $X$ ) vs. density ( $X \rightarrow [0, \infty]$ ) vs. random variable
- ▶ succinctness: omit and elaborate details
- ▶ especially **formal** types, allow using theory correctly without fully understanding it

# Lecture plan

## Part 1: the **discrete** model (now)

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



course page

## Part 2: the **full** model

- ▶ Borel sets and measurable spaces
- ▶ Quasi-Borel spaces
- ▶ Type structure & standard Borel spaces
- ▶ Dependently-typed structure
- ▶ Integration



ask questions on the  
Scottish PL Institute  
Zulip stream #qbs

# Language of probability & distribution

$X$  type (=space) of **values/outcomes**

$\text{DX}$  type of **distributions/measures** over  $X$

$\text{PX} \subseteq \text{DX}$  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.:

$$\mu : \text{DX}, E : \mathcal{B}_X \vdash \text{Ce}_{\mu}[E] : \mathbb{W}$$

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

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

$$y_1, y_2 : Y \vdash y_1 \stackrel{Y}{=} y_2 : \text{Prop} \quad \mu : \text{PX}, E : \mathcal{B}_X \vdash \Pr_{\mu}[E] = \text{Ce}_{\mu}[E]$$

(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

$$\mu : \mathsf{DX} \vdash \mathsf{Ce}_{\mu} [\emptyset] = 0$$

# Axioms for events and distributions

Boolean Sub-algebra of Events

$$E : \mathcal{B}_X \vdash E^C : \mathcal{B}_X \quad E, F : \mathcal{B}_X \vdash E \cap F : \mathcal{B}_X \quad \text{so also: } E, F : \mathcal{B}_X \vdash X, E \cup F : \mathcal{B}_X$$

Disjoint additivity

$$w, v : \mathbb{W} \vdash w + v : \mathbb{W} \quad E, C : \mathcal{B}_X, \mu : \mathsf{DX} \vdash \underset{\mu}{\mathsf{Ce}}[E] = \underset{\mu}{\mathsf{Ce}}[E \cap C] + \underset{\mu}{\mathsf{Ce}}[E \cap C^C]$$

# Axioms for events and distributions

Boolean Sub-algebra of Events

$$E : \mathcal{B}_X \vdash E^C : \mathcal{B}_X \quad E, F : \mathcal{B}_X \vdash E \cap F : \mathcal{B}_X \quad \text{so also: } E, F : \mathcal{B}_X \vdash X, E \cup F : \mathcal{B}_X$$

Disjoint additivity

$$w, v : \mathbb{W} \vdash w + v : \mathbb{W} \quad E, C : \mathcal{B}_X, \mu : \mathsf{DX} \vdash \underset{\mu}{\mathsf{Ce}}[E] = \underset{\mu}{\mathsf{Ce}}[E \cap C] + \underset{\mu}{\mathsf{Ce}}[E \cap C^C]$$

Exercise

Derive ‘axiomatically’ that:

- ▶ measurement is **monotone**:

$$\mu : \mathsf{DX}, E \subseteq F \vdash \underset{\mu}{\mathsf{Ce}}[E] \leq \underset{\mu}{\mathsf{Ce}}[F]$$

- ▶ the **inclusion-exclusion** principle:

$$\mu : \mathsf{DX}, E, F : \mathcal{B}_X \vdash \underset{\mu}{\mathsf{Ce}}[E \cup F] + \underset{\mu}{\mathsf{Ce}}[E \cap F] = \underset{\mu}{\mathsf{Ce}}[E] + \underset{\mu}{\mathsf{Ce}}[F]$$

# Axioms for events and distributions

Consider posets:

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

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

$$P^\omega := \{ p_\cdot \in \underline{P}^{\mathbb{N}} \mid p_0 \leq p_1 \leq \dots \}$$

Chain-closure of events and weights

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

Scott-continuity of measurement

$$E_\cdot : (\mathcal{B}_X, \subseteq)^\omega, \mu : \text{DX} \vdash \text{Ce}_\mu [\bigcup_n E_n] = \sup_n \text{Ce}_\mu [E_n]$$

# Axiom for probability

Probability distributions have total mass one

$$\text{PX} := \{\mu \in \text{DX} \mid \text{Ce}_\mu[X] = 1\} \quad \mu : \text{PX} \vdash \text{cast } \mu : \text{DX}$$

i.e., if we define:

$$\mathbb{I} := [0,1] \quad \mu : \text{PX}, E : \mathcal{B}_X \vdash \Pr_\mu[E] := \text{Ce}_{\text{cast } \mu}[E] : \mathbb{I}$$

then:

$$\mu : \text{PX} \vdash \Pr_\mu[X] = 1$$

# Integration

Lebesgue integration w.r.t. a distribution

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

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

Expectation w.r.t. a probability distribution

$$\mu : \mathsf{P}X, f : \mathbb{W}^X \vdash \mathbb{E}_{x \sim \mu} [f(x)] := \int (\mathsf{cast} \mu)(dx) f(x) : \mathbb{W}$$

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

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

# Summary

Have: Language and (some) axioms

Want: Model

Today: **discrete** model

Next week: **full** model

# Lecture plan

## Part 1: the **discrete** model (now)

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



course page

## Part 2: the **full** model

- ▶ Borel sets and measurable spaces
- ▶ Quasi-Borel spaces
- ▶ Type structure & standard Borel spaces
- ▶ Dependently-typed structure
- ▶ Integration



ask questions on the  
Scottish PL Institute  
Zulip stream #qbs

# Discrete model

$X$ : types denote **sets**

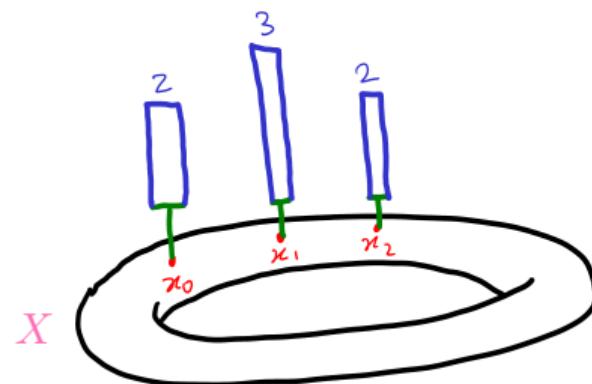
$\mathbf{D}X$ : set of **histograms**:

# Discrete model

$X$ : types denote **sets**

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

$\mathbb{D}X := \{\mu : X \rightarrow \mathbb{W} \mid \mu \text{ is } \mathbf{countably \ supported} \text{ (next slide)}\}$



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

# Countably supported distributions

## Support

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

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

The subsets supporting a weight function  $\mu$  are closed under intersections.

$\implies$  There is a smallest supporting subset, called the **support** of  $\mu$ :

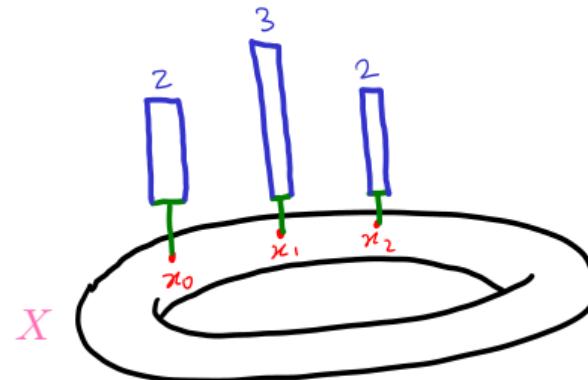
$$\mu : \mathbb{W}^X \vdash \text{supp } \mu := \{x \in X | \mu x > 0\}$$

# Discrete model

$X$ : types denote **sets**

$\mathbf{DX}$ : set of **histograms**:

$$\begin{aligned}\mathbf{DX} &:= \{\mu : X \rightarrow \mathbb{W} \mid \mu \text{ is } \mathbf{countably\ supported}\} \\ &:= \{\mu : X \rightarrow \mathbb{W} \mid \exists S \in \mathcal{P}X. S \text{ is countable}\} \\ &:= \{\mu : X \rightarrow \mathbb{W} \mid \text{supp } \mu \text{ is countable}\}\end{aligned}$$



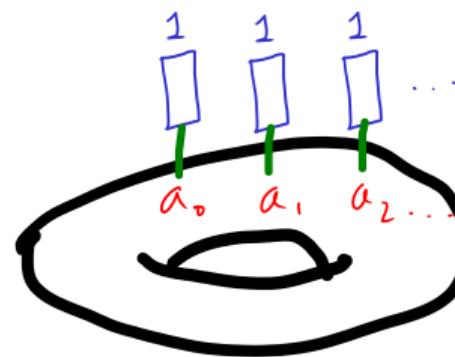
$$\mu x_0 = 2 \quad \mu x_1 = 3 \quad \mu x_2 = 2$$

# Example distributions

## Counting distribution

Counts the outcomes in a countable subset:

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

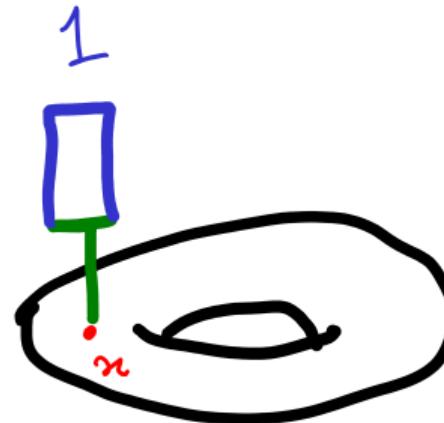


## Example distributions

Dirac

A point mass:

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



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

# Example distributions

## Zero

No mass anywhere:

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

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

# Discrete model

$X$ : types denote **sets**

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

$$\text{D}X := \{\mu : X \rightarrow \mathbb{W} \mid \mu \text{ is } \text{countably supported}\}$$

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

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

Measurement: weighted sum of all (supported) outcomes:

$$\begin{aligned}\mu : \text{D}X, E : \mathcal{B}_X \vdash \text{Ce}_\mu [E] &:= \sum_{x \in E} \mu x \\ &:= \sum_{x \in E \cap \text{supp } \mu} \mu x\end{aligned}$$

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

# Example measurements

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

## Counting distribution

counts supported outcomes

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

## Example measurements

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

### Counting distribution

counts supported outcomes

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

### Dirac

detects given outcome:

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

# Example measurements

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

## Counting distribution

counts supported outcomes

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

## Dirac

detects given outcome:

$$x : X, E : \mathcal{B}_X \vdash \text{Ce}_{\delta_x}[E] = \begin{cases} x \in E : 1 \\ x \notin E : 0 \end{cases}$$

## Zero

measures every event as zero:

$$E : \mathcal{B}_X \vdash \text{Ce}_0[E] = 0$$

# The discrete model validates the axioms

## Exercise

$$\mu : D \vdash \underset{\mu}{\text{Ce}}[\emptyset] = 0$$

$$E, C : \mathcal{B}_X, \mu : D \vdash \underset{\mu}{\text{Ce}}[E] = \underset{\mu}{\text{Ce}}[E \cap C] + \underset{\mu}{\text{Ce}}[E \cap C^c]$$

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

# Parameterised distributions

## Kernel

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

Kernels are open/parameterised distributions.

## Examples

Dirac and the counting distribution form kernels:

$$\delta_- : X \rightsquigarrow DX \quad \#_- : \mathcal{P}_{\text{ctbl}} X \rightsquigarrow DX$$

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

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

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

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

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

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

# Example

## Weak Disintegration Problem (non-standard terminology)

Input: distributions  $\mu : D\Theta$ ,  $\nu : DX$

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

Such a **weak disintegration** of  $\nu$  w.r.t.  $\mu$  provides an ‘explanation’ of an observed distribution  $\nu \in DX$  in terms of a given distribution on parameters  $\mu \in D\Theta$ . I use the term ‘explanation’ because it explains how the parameters transform into observations.

# Example

## Weak Disintegration Problem (non-standard terminology)

Input: distributions  $\mu : D\Theta$ ,  $\nu : DX$

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

### Example disintegration

For  $n \in \mathbb{N}$ , write  $\mathbf{Fin} n := \{0, \dots, n - 1\}$ . For countable  $X$ , write  $\# := \#_X : DX$ .

Here is a disintegration of  $\# \in D((\mathbf{Fin} 2)^{\mathbf{Fin}(n+1)})$  w.r.t.  $\# \in D(\mathbf{Fin} 2)$ :

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

$f : \mathbf{Fin}(n+1) \rightarrow \mathbf{Fin} 2$  function

so can take only one value: 0 or 1

$$\downarrow \\ = 1 = \# f$$

# Sub-type of probability distributions

## Sub-types

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

$$\{x : X | \varphi\} \subseteq X \quad y : \{x : X | \varphi\} \vdash \text{cast } y := y : X$$

we **lift** values in  $X$  that satisfy  $\varphi$  to the sub-type:

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

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

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

i.e., the lifting is unique and elements in the sub-type satisfy  $\varphi$ .

# Sub-type of probability distributions

## Magnitude and probability distributions

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

## Event probability

$$\mu : \mathsf{P}X, E : \mathcal{B}_X \vdash \Pr_{\mu}[E] := \mathsf{lift} \left( \mathsf{Ce}_{\mathsf{cast} \mu}[E] \right) : \mathbb{I}$$

## Stochastic kernel

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

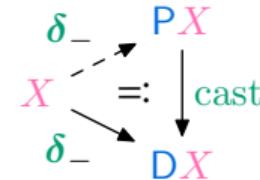
NB: in the **discrete model** these distinctions and rules amount to pure pedantry. This pedantry will pay off in the **full model**.

# Lifting Dirac and Kock

## Lemma

Dirac kernels  $\delta_- : X \rightarrow DX$  lift along `cast`:

$$x : X \vdash \|\delta_x\| = \text{Ce}_{\delta_x}[X] = 1 \quad \text{so we can overload:}$$



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

$$\mu : PX, k : (PY)^X \vdash \text{Ce}_{ʃ(\text{cast } \mu)(dx) \text{ cast}(k x)}[Y] = 1$$

$$\begin{array}{ccc} (PX) \times (PY)^X & \xrightarrow{\text{ʃ}} & PY \\ \text{cast} \times (\text{cast} \circ) \downarrow & =: & \downarrow \text{cast} \\ (DX) \times (DY)^X & \xrightarrow{\text{ʃ}} & DY \end{array}$$

so we can overload:

### Proposition

The triple  $(D, \delta_-, \oint)$  forms a monad over **Set**:

$$x : X, k : (DY)^X$$

$$\vdash \oint d\delta_x k = k x$$

$$\mu : DX$$

$$\vdash \oint \mu(dx) \delta_x = \mu$$

$$\mu : DX, k : (DY)^X, \ell : (DZ)^Y$$

$$\vdash \oint (\oint \mu(dx) k x) (dy) \ell y = \oint \mu(dx) \oint k(x; dy) \ell y$$

### Corollary

The triple  $(P, \delta_-, \oint)$  forms a monad over **Set**.

# Weighted average

Lebesgue integral

Integration is the raison d'être for distributions:

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

In the **discrete model**:

$$\int d\mu f := \sum_{x \in X} (\mu x) \cdot (f x) := \sum_{x \in \text{supp } \mu} (\mu x) \cdot (f x)$$

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

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

# Weighted average

## Expectation

To emphasise that some  $\mu$  is a probability distribution, we will use the notation:

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

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

## Boolean type

The simplest kind of distinguishing outcomes:

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

## 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:

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

# Simplest probabilistic model

## Bernoulli kernel

Single trial succeeding with the given probability:

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

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:

$$\frac{x : X \vdash M : \mathbb{B}}{\{x \in X | M\} : \mathcal{B}_X} \quad \{x \in X | \varphi x\} := \{x \in X | \varphi x = \text{True}\}$$

## 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:

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

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

Notation:

$$\frac{\Gamma \vdash \mu : \mathbf{D}X \quad \Gamma, x : X \vdash M : \mathbb{B}}{\Gamma \vdash \underset{x \sim \mu}{\text{Ce}}[M] := \underset{\mu}{\text{Ce}}[\{x \in X | M\}] : \mathbb{W}}$$

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

# Language of probability & distribution (recap)

$X$  type of **values/outcomes**

$\text{DX}$  type of **distributions/measures** over  $X$

$\text{PX} \subseteq \text{DX}$  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.:

$$\mu : \text{DX}, E : \mathcal{B}_X \vdash \text{Ce}_{\mu}[E] : \mathbb{W}$$

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

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

$$y_1, y_2 : Y \vdash y_1 \stackrel{Y}{=} y_2 : \text{Prop} \quad \mu : \text{PX}, E : \mathcal{B}_X \vdash \text{cast}_{\mu} \text{Pr}[E] = \text{Ce}_{\mu}[E]$$

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

# Lecture plan

## Part 1: the **discrete** model (now)

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



course page

## Part 2: the **full** model

- ▶ Borel sets and measurable spaces
- ▶ Quasi-Borel spaces
- ▶ Type structure & standard Borel spaces
- ▶ Dependently-typed structure
- ▶ Integration



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 ( $\otimes$ ) :  $\mathbf{D}X \times \mathbf{D}Y \rightarrow \mathbf{D}(X \times Y)$
- ▶ Random elements and their laws (push-forward measure):  
 $(\lambda(\mu, \alpha) . \mu_\alpha) : \mathbf{D}\Omega \times X^\Omega \rightarrow \mathbf{D}X$

NB:

- ▶ Dirac kernel  $\delta_- : X \rightarrow \mathbf{D}X$
- ▶ Kock integration  
 $\oint : \mathbf{D}X \times (\mathbf{D}Y)^{\mathbf{D}X} \rightarrow \mathbf{D}Y$

## 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 ( $\otimes$ ) :  $\mathbf{D}X \times \mathbf{D}Y \rightarrow \mathbf{D}(X \times Y)$
- ▶ Random elements and their laws (push-forward measure):  
 $(\lambda(\mu, \alpha) . \mu_\alpha) : \mathbf{D}\Omega \times X^\Omega \rightarrow \mathbf{D}X$

NB:

- ▶ Dirac kernel  $\delta_- : X \rightarrow \mathbf{D}X$
- ▶ Kock integration  
 $\oint : \mathbf{D}X \times (\mathbf{D}Y)^{\mathbf{D}X} \rightarrow \mathbf{D}Y$

## 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

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

In the discrete model:

$$w \cdot \mu := \lambda x. w \cdot \mu x \quad \text{supp}(w \cdot \mu) \subseteq \text{supp } \mu$$

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

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

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

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

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

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

# Affine combinations of distributions: scaling

## Normalisation

$$\mu : \text{DX}, \|\mu\| \neq 0, \infty \vdash \frac{\mu}{\|\mu\|} := \text{lift} \left( \frac{1}{\|\mu\|} \cdot \mu \right) : \text{PX}$$

measurement is homogeneous

$$\text{Indeed: } \left\| \frac{\mu}{\|\mu\|} \right\| = \left\| \frac{1}{\|\mu\|} \cdot \mu \right\| = \frac{1}{\|\mu\|} \cdot \|\mu\| = 1$$

## Discrete uniform / categorical distribution

Random unbiased choice between finitely many options/categories:

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

In the discrete model:

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

so:  $x : X \vdash \mathbf{U}_{\{x\}} = \delta_x$ .

# Weights as distributions

Unit type

$$\textcolor{blue}{\mathbb{1}} := \{()\}$$

Proposition

The following two functions are mutually inverse:

$$\begin{array}{ccc} & \parallel - \parallel & \\ \textcolor{blue}{D}\textcolor{blue}{\mathbb{1}} & \xrightarrow{\hspace{2cm}} & \textcolor{blue}{\mathbb{W}} \\ & \xleftarrow{\hspace{2cm}} & \\ & (\cdot \delta_0) & \end{array}$$

Proof

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

# Internalising Lebesgue integration

## Proposition

We can recover Lebesgue integration from Kock integration:

$$\begin{array}{ccc} DX \times \mathbb{W}^X & \xrightarrow{\text{id} \times (\cong \circ)} & DX \times (D\mathbb{1})^X \\ \downarrow \int & = & \downarrow \oint \\ \mathbb{W} & \xleftarrow{\cong} & D\mathbb{1} \end{array}$$

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

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

In the discrete model:

$$\sum_{i \in I} \mu_i := \lambda x. \sum_{i \in I} \mu_i x \quad \text{supp } \sum_{i \in I} \mu_i = \bigcup_{i \in I} \text{supp } \mu_i$$

## 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:

$$p : \mathbb{I} \vdash \mathbf{B} p = \text{lift}(p \cdot \delta_{\text{True}} + (1 - p) \cdot \delta_{\text{False}}) : \text{PB}$$

# Affinity of integration and convexity of expectation

## Theorem (Multi-linearity)

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

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

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

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

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

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

# Weight arithmetic

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

## Proposition

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

$$(\mathbf{D}\mathbb{1}, \sum, (\cdot)) \cong (\mathbb{W}, \sum, (\cdot))$$

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

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

# Convex combinations of probability distributions

Lemma

**Convex** combination lifts to probability distributions:

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

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

**Proof**

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

■

# Convex combinations of probability distributions

## Corollary (Multi-convexity)

Stochastic Kock integration, expectation and measurement are convex:

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

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

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

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

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

# Products

## Product distribution

$$\mu : \mathsf{D}X, \nu : \mathsf{D}Y \vdash \mu \otimes \nu := \oint \mu(dx) \oint \nu(dy) \delta_{(x,y)} : \mathsf{D}(X \times Y)$$

In the discrete model:

$$\mu \otimes \nu = \lambda(x, y) \cdot (\mu x) \cdot (\nu y) \quad \text{supp } (\mu \otimes \nu) = (\text{supp } \mu) \times (\text{supp } \nu)$$

Example: counting distribution on product space

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

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

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

# Products

Notation:

$$\frac{\Gamma \vdash M : D(X \times Y) \quad \Gamma, x : X, y : Y \vdash K : DZ}{\Gamma \vdash \iint M(dx, dy) K := \oint dM(\lambda(x, y) . K) : DZ}$$

## Theorem (Fubini-Tonelli)

We can integrate products in any order:

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

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

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

$$\int \mu(dx) \int \nu(dy) \varphi(x, y) = \iint (\mu \otimes \nu)(dx, dy) \varphi(x, y) = \int \nu(dy) \int \mu(dx) \varphi(x, y)$$

# Applying Fubini-Tonelli

## Theorem (Rule of Product)

We can factor out products:

$$\begin{array}{ll} \mu : \mathsf{D}X, f : \mathbb{W}^X, \nu : \mathsf{D}Y, g : \mathbb{W}^Y \vdash & \iint (\mu \otimes \nu)(dx, dy) fx \cdot gy = \left( \int d\mu f \right) \cdot \left( \int d\nu g \right) \\ \mu : \mathsf{D}X, E : \mathcal{B}_X, \nu : \mathsf{D}Y, F : \mathcal{B}_Y \vdash & \underset{\mu \otimes \nu}{\mathsf{Ce}} [E \times F] = \underset{\mu}{\mathsf{Ce}} [E] \cdot \underset{\nu}{\mathsf{Ce}} [F] \end{array}$$

## Theorem

The product lifts to probability distributions:

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

## Binomial distribution

the number of successful outcomes of  $n$  independent Bernoulli trials:

$$\mathbf{B}_n : \mathbb{I} \rightsquigarrow \mathsf{P}(\mathbf{Fin}(1+n)) \quad \mathbf{B}_0 p := \delta_0 : \mathsf{P}(\mathbf{Fin} 1)$$

$$\mathbf{B}_{1+n} p := \iint (\mathbf{B}_n p \otimes \mathbf{B} p)(dc, db) (\text{if } b \text{ then } \delta_{1+c} \text{ else } \delta_c) : \mathsf{P}(\mathbf{Fin}(2+n))$$

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:

$$\mu : D\Omega \vdash \alpha : \Omega \rightarrow X$$

Law

of a random element is the distribution:

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

Example

Represent outcomes of die roll by  $D6 := \{1, 2, \dots, 6\}$ , and two rolls by  $D6 \times D6$ .

The sum of the rolls is a random element:

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

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 Unconscious Statistician)

Formulae for reparameterising integration and measurement:

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

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

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

# Simply-typed foundations for probabilistic modelling

## Compositional building blocks for modelling

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

NB:

- ▶ Dirac kernel  $\delta_- : X \rightarrow \mathbf{D}X$
- ▶ Kock integration  
 $\oint : \mathbf{D}X \times (\mathbf{D}Y)^{\mathbf{D}X} \rightarrow \mathbf{D}Y$

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

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

**Marginal distribution:** the law of a projection:

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

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

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

## Exercise

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

# 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:

$$\mu : D\Omega, \alpha : X^\Omega, \beta : Y^\Omega \vdash \alpha \perp_{\mu} \beta := \left( \mu_{(\alpha, \beta)} \stackrel{D(X \times Y)}{=} \mu_\alpha \otimes \mu_\beta \right)$$

More generally, for finite  $I$ :

$$\mu : D\Omega, \alpha_i : (X^\Omega)^I \vdash \perp_{\mu} \alpha_i := \left( \mu_{(\alpha_i)_i} \stackrel{D(\prod_i X_i)}{=} \bigotimes_{i \in I} \mu_{\alpha_i} \right)$$

# Independence

## Example [Durett]

Model 3 independent coin tosses:

$$\text{Toss} := \{\text{Head}, \text{Tail}\} \quad \Omega := \text{Toss}^3 \quad \mu := \mathbf{U}_{\text{Toss}} \otimes \mathbf{U}_{\text{Toss}} \otimes \mathbf{U}_{\text{Toss}} : P\Omega$$

The outcome of the  $i^{\text{th}}$  coin toss is the random element  $\pi_i : \Omega \rightarrow \text{Toss}$ .

Consider the Boolean proposition in which the  $i^{\text{th}}$  and  $j^{\text{th}}$  tosses ( $i \neq j$ ) agree:

$$\text{Same}_{ij} := \lambda \omega. \pi_i \omega = \pi_j \omega : \Omega \rightarrow \mathbb{B}$$

Calculate:

LOTUS

$$\begin{aligned} \Pr_{\mu} [\text{Same}_{12}] &= \Pr_{(x,y) \sim \mu(\pi_1, \pi_2)} [x = y] = \Pr_{(x,y) \sim \mathbf{U} \otimes \mathbf{U}} [x = y] = \int \mathbf{U}(dx) \Pr_{y \sim \mathbf{U}} [x = y] \\ &= \frac{1}{2} \cdot \Pr_{y \sim \mathbf{U}} [\text{Head} = y] + \frac{1}{2} \cdot \Pr_{y \sim \mathbf{U}} [\text{Tail} = y] = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \end{aligned}$$

marginalisation

Fubini

$$\int \mathbf{U}(dx) \Pr_{y \sim \mathbf{U}} [x = y]$$

# Independence

## Example [Durett]

Model 3 independent coin tosses:

$$\text{Toss} := \{\text{Head}, \text{Tail}\} \quad \Omega := \text{Toss}^3 \quad \mu := \mathbf{U}_{\text{Toss}} \otimes \mathbf{U}_{\text{Toss}} \otimes \mathbf{U}_{\text{Toss}} : P\Omega$$

The outcome of the  $i^{\text{th}}$  coin toss is the random element  $\pi_i : \Omega \rightarrow \text{Toss}$ .

Consider the Boolean proposition in which the  $i^{\text{th}}$  and  $j^{\text{th}}$  tosses ( $i \neq j$ ) agree:

$$\text{Same}_{ij} := \lambda \omega. \pi_i \omega = \pi_j \omega : \Omega \rightarrow \mathbb{B}$$

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

# Independence

$\pi_1$ ,  $\text{Same}_{12}$ , and  $\text{Same}_{13}$  determine  $\pi_2, \pi_3$ , so:

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

Fubini-Tonelli

$$\begin{aligned} & \downarrow \\ &= \int \mathbf{U}_{\text{Toss}}(db_1) \Pr_{(b_2, b_3) \sim (\mathbf{U} \otimes \mathbf{U})} [\text{Same}_{12}(b_1, b_2, b_3) = \text{True}, \text{Same}_{13}(b_1, b_2, b_3) = \text{True}] \\ &= \frac{1}{2} \Pr_{(b_2, b_3) \sim (\mathbf{U} \otimes \mathbf{U})} [\text{Same}_{12}(\text{Head}, b_2, b_3) = \text{True}, \text{Same}_{13}(\text{Head}, b_2, b_3) = \text{True}] \\ &\quad + \frac{1}{2} \Pr_{(b_2, b_3) \sim (\mathbf{U} \otimes \mathbf{U})} [\text{Same}_{12}(\text{Tail}, b_2, b_3) = \text{True}, \text{Same}_{13}(\text{Tail}, b_2, b_3) = \text{True}] \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \end{aligned}$$

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

# Independence

## Example [Durett]

Model 3 independent coin tosses:

$$\text{Toss} := \{\text{Head}, \text{Tail}\} \quad \Omega := \text{Toss}^3 \quad \mu := \mathbf{U}_{\text{Toss}} \otimes \mathbf{U}_{\text{Toss}} \otimes \mathbf{U}_{\text{Toss}} : P\Omega$$

The outcome of the  $i^{\text{th}}$  coin toss is the random element  $\pi_i : \Omega \rightarrow \text{Toss}$ .

Consider the Boolean proposition in which the  $i^{\text{th}}$  and  $j^{\text{th}}$  tosses ( $i \neq j$ ) agree:

$$\text{Same}_{ij} := \lambda \omega. \pi_i \omega = \pi_j \omega : \Omega \rightarrow \mathbb{B}$$

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

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

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

# Distribution preservation

Distribution space  $(\Omega, \mu)$

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

Distribution preserving function

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

$$f : \Omega_1 \rightarrow \Omega_2 \quad (\mu_1)_f = \mu_2$$

$\mu : DX$  is **invariant** under  $f : X \rightarrow X$  when  $f : (X, \mu) \rightarrow (X, \mu)$  is dist. preserving.

Example

Consider the swapping function:  $\text{swap} := (\lambda(x, y). (y, x)) : X \times Y \rightarrow Y \times X$ . Then, for each  $\mu : DX$ ,  $\nu : DY$ , swapping is distribution preserving function:

$$\text{swap} : (X \times Y, \mu \otimes \nu) \rightarrow (Y \times X, \nu \otimes \mu)$$

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

# Density and scaling

## Density

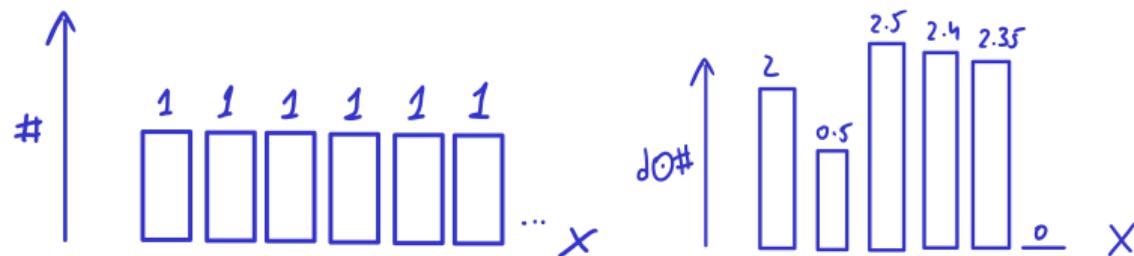
over  $X$  is any weight function  $f : X \rightarrow \mathbb{W}$ .

## Density scaling

We can scale a distribution by a density:

$$f : \mathbb{W}^X, \mu : \mathbf{D}X \vdash f \odot \mu := \int \mu(dx)(f, x) \cdot \delta_x : \mathbf{D}X$$

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  $f : X \rightarrow \mathbb{W}$ .

## Density scaling

We can scale a distribution by a density:

$$f : \mathbb{W}^X, \mu : \mathsf{D}X \vdash f \odot \mu := \oint \mu(dx)(f, x) \cdot \delta_x : \mathsf{D}X$$

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 confuse them. Types help us keep them separate.

# Density and absolute continuity

## Having density

This concept has several names in the literature:

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

- ▶  $f$  is the **density** of  $\mu$  w.r.t.  $\nu$
- ▶  $f$  is a **Radon-Nikodym derivative** of  $\mu$  w.r.t.  $\nu$ .

## Absolute continuity

$\mu$  is **absolutely continuous** w.r.t.  $\nu$  when  $\mu$  has a density w.r.t.  $\nu$ :

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

# 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:

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

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:

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

For probabilities we define:

$$\frac{\Gamma \vdash \mu : \mathbf{P}X \quad \Gamma, x : X \vdash M : \mathbb{B}}{\Gamma \vdash \mu(\mathrm{d}x) \text{ almost surely } M := (\mathbf{cast}\ \mu)(\mathrm{d}x) \text{ almost certainly } M : \mathbf{Prop}}$$

# Existence and almost-sure uniqueness of densities

Theorem (Radon-Nikodym)

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

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

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

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

Example

For all countable  $X$ , we have:

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

Indeed, apply the Radon-Nikodym theorem, since  $\text{supp } \# = X$ .

Constructively, direct calculation shows:  $(\lambda x. \mu x) = \frac{d\mu}{d\#}$ .

# Simply-typed foundations for probabilistic modelling

## Compositional building blocks for modelling

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

NB:

- ▶ Dirac kernel  $\delta_- : X \rightarrow \mathbf{D}X$
- ▶ Kock integration  
 $\oint : \mathbf{D}X \times (\mathbf{D}Y)^{\mathbf{D}X} \rightarrow \mathbf{D}Y$

## Standard vocabulary

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

# Lecture plan

## Part 1: the **discrete** model (now)

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



course page

## Part 2: the **full** model

- ▶ Borel sets and measurable spaces
- ▶ Quasi-Borel spaces
- ▶ Type structure & standard Borel spaces
- ▶ Dependently-typed structure
- ▶ Integration



ask questions on the  
Scottish PL Institute  
Zulip stream #qbs

# Type dependencies

## Example: Binomial kernels

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

$$\vdash \mathbf{B}_n : \mathbb{I} \rightsquigarrow \mathbf{Fin}(1 + n)$$

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

$$n : \mathbb{N} \vdash \mathbf{B}_n : \mathbb{I} \rightsquigarrow \mathbf{Fin}(1 + n)$$

# Family model

Family over an indexing set  $I$

consists of a sequence  $X_ = (X_i)_{i \in I}$  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:

$$\mathbf{Fin} n := \{0, 1, \dots, n - 1\}$$

# Family model

Family over an indexing set  $I$

consists of a sequence  $X_ = (X_i)_{i \in I}$  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

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

$$\theta : I \rightarrow J \quad (f_i : X_i \rightarrow Y_{\theta i})_{i \in I}$$

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:

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

## 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:

Family of distributions  $i : I \vdash X_i \rightarrow Y_i$

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

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

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

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

Both have a **Dirac** distribution:

$$i : I \vdash \delta_- : X_i \rightarrow \mathbf{D}X_i \quad i : I \vdash \delta_- : X_i \rightarrow \mathbf{P}X_i$$

# Extension and dependent pairs

## Extension

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

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

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

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

# Functions and kernels

## Dependent functions

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

$$\frac{i : I \vdash X_i \quad i : I, x : X_i \vdash Y_{i,x}}{i : I \vdash ((x : X) \rightarrow Y_{i,x}) \coloneqq \prod_{x \in X} Y_{i,x}}$$

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

are dependent elements:

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

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

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

# Integration

## Dependent Kock integral

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

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

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

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

$$i : I, \mu : \mathbf{D}X_i, f : (x : X_i) \rightarrow \mathbb{W} \vdash \int d\mu f : \mathbb{W} \quad i : I, \mu : \mathbf{D}X_i, E : \mathcal{B}_{X_i} \vdash \text{Ce}_\mu [E] : \mathbb{W}$$
$$\int d\mu f = \sum_{x \in X} \mu x \cdot f x \quad \text{Ce}_\mu [E] = \sum_{x \in E} \mu x$$

# Random variables

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

## Signed and unsigned random variable

in a probability space  $(\Omega, \mu)$  are random elements  $\alpha : \Omega \rightarrow \overline{\mathbb{R}}$  and  $\alpha : \Omega \rightarrow \mathbb{W}$ .

The **positive** and **negative parts** are unsigned random variables  $\alpha^\pm : \overline{\mathbb{R}}^\Omega \rightarrow \mathbb{W}^\Omega$ :

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

An unsigned r.v.  $\alpha$  is **Lebesgue integrable** when its Lebesgue integral is finite:

$$\int d\mu \alpha < \infty.$$

For a (signed) r.v.  $\alpha$ , when either  $\alpha^+$  or  $\alpha^-$  is Lebesgue integrable, we define:

$$\mu : \mathbf{DX}, \alpha : \overline{\mathbb{R}}^X, \int d\mu \alpha^+, \int d\mu \alpha^- < \infty \vdash \int d\mu \alpha := \int d\mu \alpha^+ - \int d\mu \alpha^-$$

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

# Random variable spaces

Lebesgue integrability is a Boolean property:

$$\mu : \text{DX}, \alpha : X \rightarrow \bar{\mathbb{R}} \vdash \alpha \text{ integrable} := \int d\mu \alpha^+ < \infty \wedge \int d\mu \alpha^- < \infty : \mathbb{B}$$

Lebesgue spaces ensemble

is the family:

$$i : I, p : [1, \infty), \mu : \mathsf{P}X_i \vdash \mathcal{L}_p(X_i, \mu) := \{\alpha : X_i \rightarrow \bar{\mathbb{R}} \mid \alpha^p \text{ integrable}\}$$

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

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

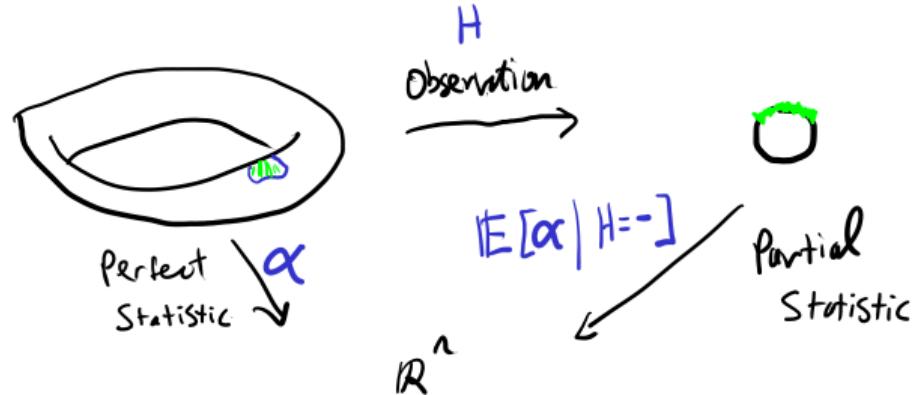
and the fibre 2 has an inner product (almost a Hilbert space!):

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

# Conditioning á la Kolmogorov

Situation:

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



Conditional expectation of  $\alpha$  along  $H$  w.r.t  $\mu$

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

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

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

# Conditioning á la Kolmogorov

## Theorem (Kolmogorov)

Every random variable has a conditional expectation:

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

Therefore:

## Corollary (Internal conditional expectation)

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

$$\mathbb{E}_- [- | - = -] :$$

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

# Conditioning á la Kolmogorov

Conditional probability

of event is a conditional expectation of its characteristic function:

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

Regular conditional probability

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

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

# Conditioning via disintegration

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

## Disintegration Problem (warning: conflicting terminologies in literature)

**Input:** probability distribution  $\mu : P\Omega$ , measurable map  $H : \Omega \rightarrow \Theta$   
induce law  $\nu := \mu_H : P\Theta$

**Output:** probability kernel  $k : \Theta \rightsquigarrow \Omega$  such that:  $\mu = \oint d\nu k$ .

We call  $k$  a **disintegration** of  $\mu$  along  $H$ .

## Proposition

Consider a probability kernel  $k : \Theta \rightsquigarrow \Omega$ . TFAE:

- ▶  $k$  is a disintegration of  $\mu$  along  $H : \Omega \rightarrow \Theta$ ;
- ▶  $k$  is a regular conditional probability kernel of  $\mu$  conditioned on  $H$ .

# Conditioning via disintegration

Fibred disintegration of  $\mu : P(\coprod_{\Theta} \Omega)$  (non-standard terminology and formulation)

a partial dependent kernel  $k : (\theta : \Theta) \rightsquigarrow \Omega_{\perp}$ , defined  $\mu_{\text{dep}}$ -a.s., that disintegrates  $\mu$  along the first projection  $\text{dep} : (\coprod_{\Theta} \Omega) \rightarrow \Theta$ :

$\mu : P\left(\coprod_{\Theta} \Omega\right), k : \Theta \rightsquigarrow \Omega_{\perp} \vdash k \text{ disintegrates fibres of } \mu :=$

$$\mu_{\text{dep}}(\text{Dom}(k)) = 1, \mu = \oint d\mu_{\text{dep}} k : \text{Prop}$$

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

$$-\dagger : \left( \mu : P\left(\coprod_{\Theta} \Omega\right) \right) \rightarrow \{ k : (\theta : \Theta) \rightsquigarrow \Omega_{\perp} \mid k \text{ disintegrates } \mu \text{ along } \text{dep} \}$$

$$\text{Dom}(\mu^{\dagger}) := \{ \theta \mid \mu_{\text{dep}} \theta > 0 \} \quad \mu^{\dagger} := \lambda \theta. \frac{1}{\mu_{\text{dep}} \theta} \odot \mu|_{\text{dep}^{-1}[\theta]}$$

# Bayes's Theorem (adapted from Williams)

Let:

- ▶  $\lambda : P(X \times \Theta)$  be a joint probability distribution.
- ▶  $\mu : D_X$ ,  $\nu : D_\Theta$  be distributions such that  $\lambda \ll \mu \otimes \nu$      $X \xleftarrow{\alpha:=\pi_1} X \times \Theta \xrightarrow{H:=\pi_2} \Theta$
- ▶  $w_{\alpha,H} = \frac{d\lambda}{d\mu \otimes \nu} : X \times \Theta \rightarrow \mathbb{W}$  a Radon-Nikodym derivative

Observation 1

- ▶  $w_\alpha := \lambda x. \int \nu(d\theta) w_{\alpha,H}(x, \theta) : X \rightarrow \mathbb{W}$  then:  $w_\alpha = \frac{d\lambda_\alpha}{d\mu}$
- ▶  $w_H := \lambda \theta. \int \mu(dx) w_{\alpha,H}(x, \theta) : \Theta \rightarrow \mathbb{W}$  then:  $w_H = \frac{d\lambda_H}{d\nu}$

Observation 2

Let:  $w_\alpha(- \mid H = -) : X \times \Theta \rightarrow \mathbb{W}$        $w_\alpha(x \mid H = \theta) := \begin{cases} w_H \theta > 0 : & \frac{w_{\alpha,H}(x, \theta)}{w_H \theta} \\ \text{otherwise:} & 0 \end{cases}$

$\lambda_{\alpha|H=-} : \Theta \rightsquigarrow X$        $\lambda_{\alpha|H=\theta} := \lambda_\alpha(- \mid H = \theta) \odot \nu$ . Then:

$$\lambda_{\alpha|H=-} = \Pr_\lambda [- \mid H = -] \quad (\text{Bayes's formula})$$

# Lecture plan

## Part 1: the **discrete** model

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



course page

## Part 2: the **full** model (now)

- ▶ Borel sets and measurable spaces
- ▶ Quasi-Borel spaces
- ▶ Type structure & standard Borel spaces
- ▶ Dependently-typed structure
- ▶ Integration



ask questions on the  
Scottish PL Institute  
Zulip stream #qbs

# From histograms to measures

The **discrete** model expresses  
**histograms** only.

Also want **continuous** distributions:

- ▶ lengths
- ▶ areas
- ▶ volumes



# Continuous caveat

## Theorem (Vitali 1905)

There is no reasonable generalisation of ‘length’ that measures all subsets of the real line—there is no function  $\lambda : \mathcal{P}\mathbb{R} \rightarrow \mathbb{W}$  satisfying:

$$\begin{array}{lll} \lambda[a, b] = (b - a) & \lambda(s + [E]) = \lambda E & \lambda(\biguplus_{i=0}^{\infty} E_n) = \sum_{i=0}^{\infty} \lambda E_n \\ (\text{generalise length}) & (\text{translation invariance}) & (\sigma\text{-additivity}) \end{array}$$

## Takeaway

$\mathcal{B}_{\mathbb{R}} := \mathcal{P}\mathbb{R}$  as in the **discrete** model excludes **length, area, volume** as distributions.  
⇒ need a different model

# Workaround

Only measure **well-behaved** subsets:

Borel subsets  $\mathcal{B}_{\mathbb{R}} \subseteq \mathcal{P}\mathbb{R}$

smallest  $\sigma$ -field containing all **open intervals**:

$$\overline{\emptyset \in \mathcal{B}_{\mathbb{R}}}$$

(empty set)

$$\overline{E^c \in \mathcal{B}_{\mathbb{R}}}$$

(complements)

$$\overline{\bigcup_{n \in \mathbb{N}} E_n \in \mathcal{B}_{\mathbb{R}}}$$

(countable unions)

$$\overline{(a,b) \in \mathcal{B}_{\mathbb{R}}}$$

(intervals)

## Examples

- ▶ Countable discrete subsets are Borel:

$$\{r\} = \bigcap_{\varepsilon \in \mathbb{Q}_{>0}} (r - \varepsilon, r + \varepsilon) \in \mathcal{B}_{\mathbb{R}} \quad , \quad I \text{ countable} \implies I = \bigcup_{i \in I} \{i\}$$

- ▶ Any interval is Borel, e.g.:  $[a,b] = (a,b) \cup \{a\}$

# Measure theory: generalise the **worst-case** scenario 😊

Measurable space  $M = (\underline{M}, \mathcal{B}_M)$

set of **points**  $a \in \underline{M}$  equipped with a  **$\sigma$ -field**  $\mathcal{B}_M \subseteq \mathcal{P}\underline{M}$ :

$$\overline{\emptyset \in \mathcal{B}_{\mathbb{R}}}$$

(empty set)

$$\overline{E^c \in \mathcal{B}_{\mathbb{R}}}$$

(complements)

$$\overline{\bigcup_{n \in \mathbb{N}} E_n \in \mathcal{B}_{\mathbb{R}}}$$

(countable unions)

## Examples

- ▶ Discrete spaces:  $\overline{I}^{\text{Meas}} := (I, \mathcal{P}I)$
- ▶ Sub-spaces:  $\frac{S \subseteq \underline{M}}{S_M := (S, [\mathcal{B}_M] \cap S)}$  i.e.,  $\mathcal{B}_{S_M} := \{E \cap S | E \in \mathcal{B}_M\}$ , e.g.,  $[0, \infty) \hookrightarrow \mathbb{R}$
- ▶ Products:  $\mathcal{B}_{\prod_{i \in I} M_i} := \sigma \bigcup_{i \in I} \pi_i^{-1} [\mathcal{B}_{M_i}] = \sigma \left\{ \bigtimes_{i \in I} E_i \middle| \begin{array}{l} E_i \in \prod_{i \in I} \mathcal{B}_{M_i}, \\ \exists J \subseteq \text{countable } I. \\ \forall j \notin J. E_j = M_i \end{array} \right\}$ , e.g.:  $\mathbb{R}^n$

Borel measurable function  $f : M \rightarrow K$

function sending points to points and measurable subsets to measurable subsets:

$$f : \underline{M} \rightarrow \underline{K} \quad \mathcal{B}_M \ni f^{-1}[\textcolor{red}{E}] \iff \textcolor{red}{E} \in \mathcal{B}_K$$

## Examples

- ▶  $(+), (\cdot) : \mathbb{R}^2 \rightarrow \mathbb{R}$
- ▶  $|-|, \sin : \mathbb{R} \rightarrow \mathbb{R}$
- ▶ any continuous function  $\mathbb{R}^n \rightarrow \mathbb{R}$
- ▶ any function out of a discrete space: 
$$\frac{f : I \rightarrow \underline{M}}{f : \bar{I} \rightarrow M}$$

## Category $\mathbf{Meas}$

Objects  $M$ : measurable spaces

Arrows  $f : M \rightarrow K$ : Borel measurable functions

$$\frac{}{\text{id} := (\lambda x.x) : M \rightarrow M} \qquad \frac{f : M \rightarrow K \quad g : K \rightarrow L}{g \circ f : (\lambda x.g(f x)) : M \rightarrow L}$$

## Categorical structure

Products, coproducts/disjoint unions, subspaces, projective and injective limits / categorical limits and colimits are all fine.

## Theorem (Aumann'61)

There are no measurable spaces of Borel subsets nor of measurable functions over  $\mathbb{R}$ .

In detail, there are no  $\sigma$ -fields  $\mathcal{B}_{\mathcal{B}_{\mathbb{R}}}$  and  $\mathcal{B}_{\mathbb{R} \rightarrow \mathbb{R}}$  such that, letting  $\mathcal{B}_{\mathbb{R}}$  and  $\mathbb{R} \rightarrow \mathbb{R}$  be the corresponding measurable spaces, the following functions are measurable:

- ▶ Membership testing:

$$(\in) := \left( \lambda r. E. \begin{cases} r \in E : & \text{True} \\ \text{otherwise:} & \text{False} \end{cases} \right) : \mathbb{R} \times \mathcal{B}_{\mathbb{R}} \rightarrow \overline{\{\text{True}, \text{False}\}}$$

- ▶ Evaluation:  $\text{eval} := (\lambda (f, r). fr) : (\mathbb{R} \rightarrow \mathbb{R}) \times \mathbb{R} \rightarrow \mathbb{R}$ .

As a consequence, **Meas** is not Cartesian closed.

## Aumann's Theorem: proof preliminaries

Recall the **Borel hierarchy** over a family of subsets  $\mathcal{U} \subseteq \mathcal{P}X$ , defined by transfinite induction on  $\omega_1 + 1$ , the successor of the first uncountable ordinal:

$$\Sigma_\alpha^\mathcal{U}, \Pi_\alpha^\mathcal{U}, \Delta_\alpha^\mathcal{U} \subseteq \mathcal{P}X \quad (\alpha \in \omega_1)$$

$$\Sigma_1^\mathcal{U} := \mathcal{U}$$

$$\Sigma_{\alpha+1}^\mathcal{U} := \left\{ \bigcup_{i \in I} A_i \middle| I \subseteq \mathbb{N}, A_i \in \mathcal{U} \cup \bigcup_{\beta \leq \alpha} \Pi_\beta^\mathcal{U} \right\} \quad (1 \leq \alpha \in \omega_1)$$

$$\Sigma_\gamma^\mathcal{U} := \bigcup_{\beta < \gamma} \Sigma_\beta^\mathcal{U} \quad (1 \leq \gamma \text{ a limit ordinal in } \omega_1)$$

$$\Pi_\alpha^\mathcal{U} := [\Sigma_\alpha^\mathcal{U}]^C := \left\{ A^C \middle| A \in \Sigma_\alpha^\mathcal{U} \right\} \quad \Delta_\alpha^\mathcal{U} := \Sigma_\alpha^\mathcal{U} \cap \Delta_\alpha^\mathcal{U}$$

# Aumann's Theorem: proof preliminaries

The Borel hierarchy looks like this in general:

$$\begin{array}{ccccccccc} \Sigma_1^{\mathcal{U}} & \subseteq & \Sigma_2^{\mathcal{U}} & \subseteq & \Sigma_3^{\mathcal{U}} & \subseteq & \dots \subseteq & \Sigma_{\omega}^{\mathcal{U}} & \subseteq & \Sigma_{\omega+1}^{\mathcal{U}} & \subseteq & \dots \subseteq & \Delta_{\omega_1}^{\mathcal{U}} & = & \Sigma_{\omega_1}^{\mathcal{U}} \\ \Delta_1^{\mathcal{U}} & \subseteq & \Delta_2^{\mathcal{U}} & \subseteq & \Delta_3^{\mathcal{U}} & \subseteq & \dots \subseteq & \Delta_{\omega}^{\mathcal{U}} & \subseteq & \Delta_{\omega+1}^{\mathcal{U}} & \subseteq & \dots \subseteq & \Delta_{\omega_1}^{\mathcal{U}} & = & \sigma(\mathcal{U}) \\ \Pi_1^{\mathcal{U}} & \subseteq & \Pi_2^{\mathcal{U}} & \subseteq & \Pi_3^{\mathcal{U}} & \subseteq & \dots \subseteq & \Pi_{\omega}^{\mathcal{U}} & \subseteq & \Pi_{\omega+1}^{\mathcal{U}} & \subseteq & \dots \subseteq & \Pi_{\omega_1}^{\mathcal{U}} & = & \Pi_{\omega_1}^{\mathcal{U}} \end{array}$$

For  $\mathcal{U} := \{(a,b) | a, b \in \mathbb{R}\}$ , the hierarchy does not stabilise before  $\omega_1$ :

$$\begin{array}{ccccccccc} \Sigma_1^{\mathcal{U}} & \subset & \Sigma_2^{\mathcal{U}} & \subset & \Sigma_3^{\mathcal{U}} & \subset & \dots \subset & \Sigma_{\omega}^{\mathcal{U}} & \subset & \Sigma_{\omega+1}^{\mathcal{U}} & \subset & \dots \subset & \Delta_{\omega_1}^{\mathcal{U}} & = & \Sigma_{\omega_1}^{\mathcal{U}} \\ \Delta_1^{\mathcal{U}} & \subset & \Delta_2^{\mathcal{U}} & \subset & \Delta_3^{\mathcal{U}} & \subset & \dots \subset & \Delta_{\omega}^{\mathcal{U}} & \subset & \Delta_{\omega+1}^{\mathcal{U}} & \subset & \dots \subset & \Delta_{\omega_1}^{\mathcal{U}} & = & \sigma(\mathcal{U}) = \mathcal{B}_{\mathbb{R}} \\ \Pi_1^{\mathcal{U}} & \subset & \Pi_2^{\mathcal{U}} & \subset & \Pi_3^{\mathcal{U}} & \subset & \dots \subset & \Pi_{\omega}^{\mathcal{U}} & \subset & \Pi_{\omega+1}^{\mathcal{U}} & \subset & \dots \subset & \Pi_{\omega_1}^{\mathcal{U}} & = & \Pi_{\omega_1}^{\mathcal{U}} \end{array}$$

Rank of  $E \in \sigma\mathcal{U}$

first step in which it appears:  $\text{Rank}_E := \min \{\alpha < \omega_1 | A \in \Delta_{\alpha}^{\mathcal{U}}\}$ .

# Aumann's Theorem

## Proof

Assume to the contrary there was some  $\sigma$ -field providing a measurable space of Borel subsets  $\mathcal{B}_{\mathbb{R}}$  such that membership testing is measurable:

$$(\in) : \mathbb{R} \times \mathcal{B}_{\mathbb{R}} \rightarrow \overline{\{\text{True}, \text{False}\}} \quad \text{NB: } \mathcal{B}_{\mathbb{R} \times \mathcal{B}_{\mathbb{R}}} = \sigma([\mathcal{B}_{\mathbb{R}}] \times [\mathcal{B}_{\mathcal{B}_{\mathbb{R}}}] )$$

Let  $\alpha := \text{Rank } (\in)^{-1} [\text{True}] < \omega_1$ , and find  $E \in \mathcal{B}_{\mathbb{R}}$  with  $\text{Rank}_E > \alpha$ . Then:

$$\begin{aligned} \alpha &< \text{Rank } E = \text{Rank} \left( ((\in) \circ (-, E))^{-1} [\text{True}] \right) = \text{Rank} \left( (-, E)^{-1} \left( (\in)^{-1} [\text{True}] \right) \right) \\ &\leq \text{Rank} \left( (\in)^{-1} [\text{True}] \right) = \alpha \end{aligned}$$

So  $\alpha < \alpha$ , a contradiction, and the postulated  $\sigma$ -field cannot exist. A similar proof replacing  $E$  with its characteristic function proves eval cannot be measurable. ■

# Some higher-order structure in Meas

## Sequences

By generalities,  $(\bar{I} \rightarrow M) = \prod_{i \in I} M$ . For countable  $I$ , we use  $\bar{I} \rightarrow M$  for sequences.

## Example

A sequence  $a_- : \mathbb{N} \rightarrow \mathbb{R}$  is **Cauchy** when its tail elements tend infinitesimally close:

$$\forall \varepsilon > 0. \exists N \in \mathbb{N}. \forall m, n > N. |a_n - a_m| < \varepsilon$$

The Cauchy property characterises convergence to a finite limit. We can define the Cauchy property through quantification over countable sets:

$$\text{Cauchy} \in \mathcal{B}_{\mathbb{N} \rightarrow \mathbb{R}} \quad \text{Cauchy} := \bigcap_{\varepsilon \in \mathbb{Q}_{>0}} \bigcup_{N \in \mathbb{N}} \bigcap_{m, n \in \mathbb{N}} \{a_- \in \underline{\mathbb{N} \rightarrow \mathbb{R}} \mid |a_n - a_m| < \varepsilon\}$$

measurability through **type-checking**:

$$= \left\{ a_- \in \underline{\mathbb{N} \rightarrow \mathbb{R}} \mid \begin{array}{l} \forall \varepsilon : \mathbb{Q}_{>0}. \exists N : \mathbb{N}. \forall m, n : \mathbb{N}. \\ m, n > N \implies |a_n - a_m| < \varepsilon \end{array} \right\}$$

# Measurability through type-checking

With a few simple building blocks:

$$\limsup : (\mathbb{N} \rightarrow \mathbb{R}) \rightarrow [-\infty, \infty]$$

$$\limsup^{-1}[b, \infty] = \{a \in \mathbb{N} \rightarrow \mathbb{R} \mid \forall n : \mathbb{N}. \exists m : \mathbb{N}. m > n \wedge a_n \geq b\} \in \mathcal{B}_{\mathbb{N} \rightarrow \mathbb{R}}$$

we can discharge measurability through type-checking:

$$\lim : \mathbf{Cauchy} \rightarrow \mathbb{R} \quad \lim a_- := \limsup a_-$$

$$\mathbf{Vanishing} := \left\{ r_- : \mathbb{N} \rightarrow \mathbb{R}_{>0} \mid \lim_{n \rightarrow \infty} r_n = 0 \right\} \in \mathcal{B}_{\mathbb{N} \rightarrow \mathbb{R}}$$

$$\text{approx}_- : \mathbf{Vanishing} \times \mathbb{R} \rightarrow (\mathbb{N} \rightarrow \mathbb{Q}) \quad \text{such that: } |r - \text{approx}_{\Delta_-} r n| < \Delta_n$$

However, not all operations of interest support this technique:

✖  $\limsup : (\mathbb{N} \rightarrow \mathbb{R} \rightarrow [-\infty, \infty]) \rightarrow (\mathbb{R} \rightarrow [-\infty, \infty])$        $\limsup f_- := \lambda x. \limsup_{n \rightarrow \infty} f_n x$

as they are intrinsically higher-order.

# Agenda

## Goal

Measurability by type-checking! Want to extend the model without sacrificing/compromising the language of probability we developed.

## Challenge: compositionality

For **higher-order** building blocks, classical measure theory **defers** measurability proofs until we resume 1<sup>st</sup>-order fragment.

Some probabilistic concepts are inherently higher-order, and classical measure theory makes them 2<sup>nd</sup>-class.

# Measure theory's **best-case** scenario

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

A measurable space  $S \in \mathbf{Meas}$  is **standard Borel** when there is a measurable isomorphism  $S \cong E$  for some  $E \in \mathcal{B}_{\mathbb{R}}$ .

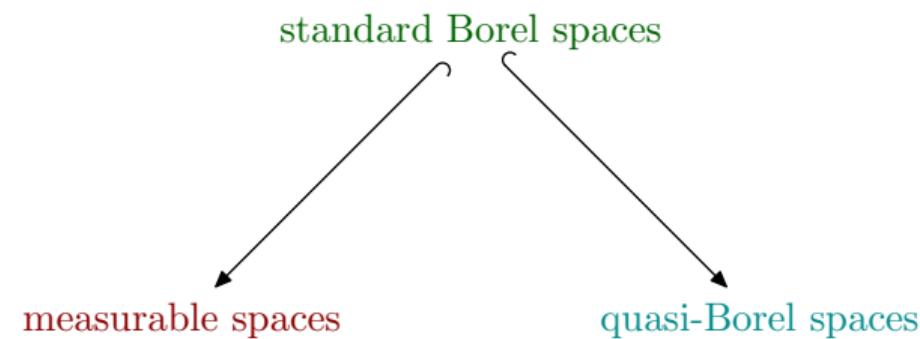
### Concrete spaces are standard

- ▶ Discrete countable spaces  $\bar{I}$
- ▶ Countable products of standard spaces are standard:  $\mathbb{R}^n$ ,  $\mathbb{N} \rightarrow \mathbb{R}$ ,  $\mathbb{N} \rightarrow \mathbb{B}$ ,  $\mathbb{N} \rightarrow \mathbb{N}$ .
- ▶ Borel subspaces of standard spaces are standard:  $[0,1]$ ,  $\mathbb{R}_{>0}$ ,  $\mathbb{R}_{\geq 0}$ .
- ▶ Countable coproducts of standard spaces are standard:  $[0,\infty]$ ,  $[-\infty,\infty]$ .

# Plan: use a different conservative extension

Concrete spaces:

Abstract spaces:



# Lecture plan

## Part 1: the **discrete** model

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



course page

## Part 2: the **full** model

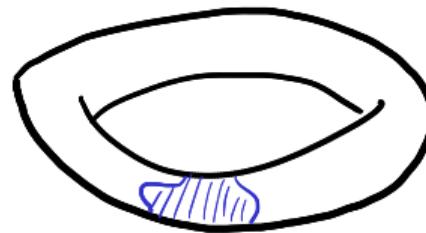
- ▶ Borel sets and measurable spaces
- ▶ Quasi-Borel spaces
- ▶ Type structure & standard Borel spaces
- ▶ Dependently-typed structure
- ▶ Integration



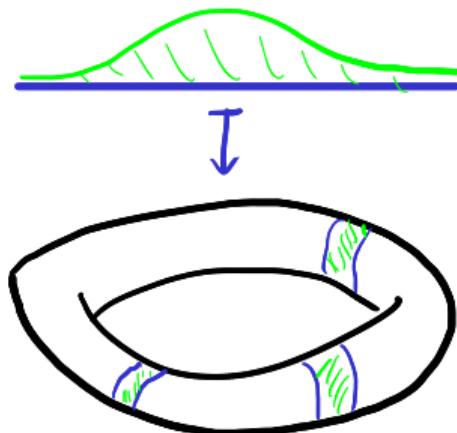
ask questions on the  
Scottish PL Institute  
Zulip stream #qbs

# Core idea

measurable spaces



quasi Borel spaces



primitive notions: points & measurable subsets

points & random elements

derived notions: random elements  $\alpha : \mathbb{R} \rightarrow M$

measurable subsets  $E \in \mathcal{B}_X$

# Core definitions

## Metaphorology<sup>1</sup> $\mathcal{R}$

over set  $\underline{X}$  of **points**: subset of functions  $\alpha : \mathbb{R} \rightarrow \underline{X}$  closed under:

**constants:**

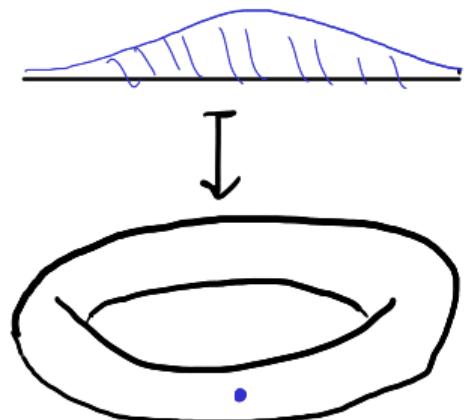
$$\frac{x \in \underline{X}}{x := (\lambda r.x) \in \mathcal{R}}$$

**precomposition:**

$$\frac{\alpha \in \mathcal{R} \quad \varphi \in \text{Meas}(\mathbb{R}, \mathbb{R})}{(\alpha \circ \varphi) \in \mathcal{R}}$$

**recombination:**

$$\frac{I \subseteq \mathbb{N} \quad E_- : I \rightarrow \mathcal{B}_{\mathbb{R}} \quad \mathbb{R} = \biguplus_{i \in I} E_i \quad \alpha_- : I \rightarrow \mathcal{R}}{[E_i \cdot \alpha_i]_{i \in I} := (\lambda r \in E_i \cdot \alpha_i r) \in \mathcal{R}}$$



<sup>1</sup> $\mu\varepsilon\tau\alpha$  ('meta', across) and  $\varphi\varepsilon\rho\omega$  ('phero', to carry).

# Core definitions

## Metaphorology<sup>1</sup> $\mathcal{R}$

over set  $\underline{X}$  of **points**: subset of functions  $\alpha : \mathbb{R} \rightarrow \underline{X}$  closed under:

constants:

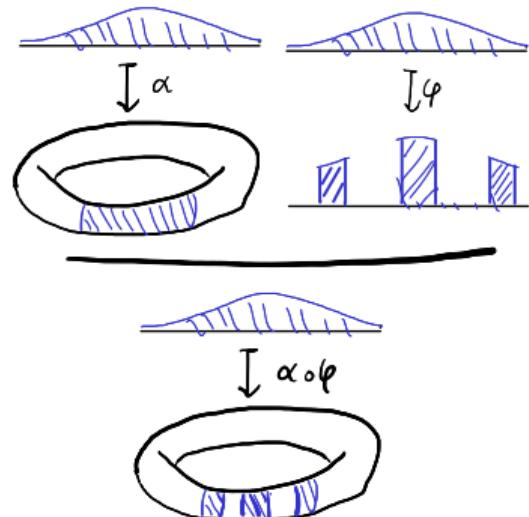
$$\frac{x \in \underline{X}}{x := (\lambda r.x) \in \mathcal{R}}$$

precomposition:

$$\frac{\alpha \in \mathcal{R} \quad \varphi \in \text{Meas}(\mathbb{R}, \mathbb{R})}{(\alpha \circ \varphi) \in \mathcal{R}}$$

recombination:

$$\frac{I \subseteq \mathbb{N} \quad E_- : I \rightarrow \mathcal{B}_{\mathbb{R}} \quad \mathbb{R} = \biguplus_{i \in I} E_i \quad \alpha_- : I \rightarrow \mathcal{R}}{[E_i \cdot \alpha_i]_{i \in I} := (\lambda r \in E_i \cdot \alpha_i r) \in \mathcal{R}}$$



<sup>1</sup> $\mu\varepsilon\tau\alpha$  ('meta', across) and  $\varphi\varepsilon\rho\omega$  ('phero', to carry).

# Core definitions

## Metaphorology<sup>1</sup> $\mathcal{R}$

over set  $\underline{X}$  of **points**: subset of functions  $\alpha : \mathbb{R} \rightarrow \underline{X}$  closed under:

constants:

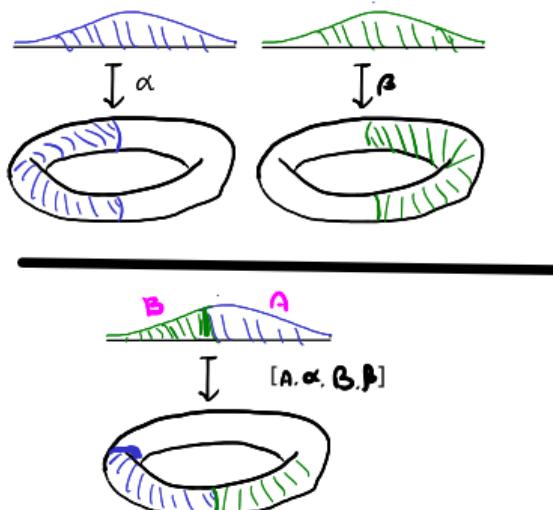
$$\frac{x \in \underline{X}}{x := (\lambda r.x) \in \mathcal{R}}$$

precomposition:

$$\frac{\alpha \in \mathcal{R} \quad \varphi \in \text{Meas}(\mathbb{R}, \mathbb{R})}{(\alpha \circ \varphi) \in \mathcal{R}}$$

recombination:

$$\frac{I \subseteq \mathbb{N} \quad E_- : I \rightarrow \mathcal{B}_{\mathbb{R}} \quad \mathbb{R} = \biguplus_{i \in I} E_i \quad \alpha_- : I \rightarrow \mathcal{R}}{[E_i \cdot \alpha_i]_{i \in I} := (\lambda r \in E_i \cdot \alpha_i r) \in \mathcal{R}}$$



<sup>1</sup> $\mu\varepsilon\tau\alpha$  ('meta', across) and  $\varphi\varepsilon\rho\omega$  ('phero', to carry).

# Core definitions

Metaphorology<sup>1</sup>  $\mathcal{R}$

over set  $\underline{X}$  of **points**: subset of functions  $\alpha : \mathbb{R} \rightarrow \underline{X}$  closed under:

**constants:**

$$\frac{x \in \underline{X}}{\underline{x} := (\lambda r.x) \in \mathcal{R}}$$

**precomposition:**

$$\frac{\alpha \in \mathcal{R} \quad \varphi \in \text{Meas}(\mathbb{R}, \mathbb{R})}{(\alpha \circ \varphi) \in \mathcal{R}}$$

**recombination:**

$$\frac{I \subseteq \mathbb{N} \quad E_- : I \rightarrow \mathcal{B}_{\mathbb{R}} \quad \mathbb{R} = \biguplus_{i \in I} E_i \quad \alpha_- : I \rightarrow \mathcal{R}}{[E_i \cdot \alpha_i]_{i \in I} := (\lambda r \in E_i \cdot \alpha_i r) \in \mathcal{R}}$$

Quasi Borel space (qbs)  $\underline{X}$   
set of points  $\underline{X}$  equipped with  
a metaphorology  $\mathcal{R}_{\underline{X}}$  over it

---

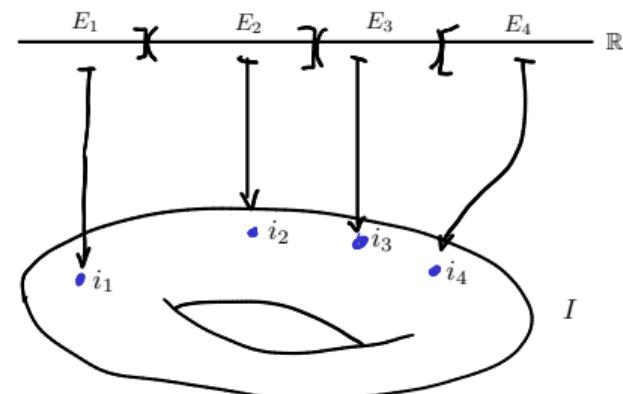
<sup>1</sup> $\mu\varepsilon\tau\alpha$  ('meta', across) and  $\varphi\varepsilon\rho\omega$  ('phero', to carry).

# Core definitions

## Examples

- ▶ **real line** qbs:  $\mathbb{R} = (\mathbb{R}, \text{Meas}(\mathbb{R}, \mathbb{R}))$
- ▶ **underlying qbs** of a measurable space:  $M = (\underline{M}, \text{Meas}(\mathbb{R}, M))$
- ▶ **indiscrete** qbs over a set: Indiscrete  $I = (I, \mathbb{R} \rightarrow I)$
- ▶ **discrete** qbs over a set:  $\bar{I} = (I, \mathcal{R}_{\bar{I}})$  with the  **$\sigma$ -simple** metaphorology, consisting of recombinations of constants:

$$\mathcal{R}_{\bar{I}} := \left\{ [E_j \cdot i_j]_{j \in J} \middle| \begin{array}{l} J \subseteq \mathbb{N}, E_- : J \rightarrow \mathcal{B}_{\mathbb{R}}, \\ \mathbb{R} = \bigcup_{j \in J} E_j, i_- : J \rightarrow I \end{array} \right\}$$



# Core definitions

Example validation: qbs underlying  $\mathbb{W}$

elements:  $\underline{\mathbb{W}} = [0, \infty]$ ; random elements:  $\mathcal{R}_{\mathbb{W}} = \mathbf{Meas}(\mathbb{R}, \mathbb{W})$ .

constants:  $\underline{w}^{-1}[\underline{E}] = \begin{cases} w \in E : \quad \mathbb{R} \in \mathcal{B}_{\mathbb{R}} \\ w \notin E : \quad \emptyset \in \mathcal{B}_{\mathbb{R}} \end{cases} \implies \underline{w} \in \mathbf{Meas}(\mathbb{R}, \mathbb{W})$ .

precomposition: 
$$\frac{\alpha \in \mathbf{Meas}(\mathbb{R}, \mathbb{W}) \quad \varphi \in \mathbf{Meas}(\mathbb{R}, \mathbb{R})}{(\varphi \circ \alpha) \in \mathbf{Meas}(\mathbb{R}, \mathbb{W})}$$

since composition in  $\mathbf{Meas}$  is function composition

recombination: 
$$\frac{I \subseteq \mathbb{N} \quad \underline{E}_- : I \rightarrow \mathcal{B}_{\mathbb{R}} \quad \alpha_- : I \rightarrow \mathbf{Meas}(\mathbb{R}, \mathbb{W}) \quad \mathbb{R} = \biguplus_{i \in I} E_i}{\forall F \in \mathcal{B}_{\mathbb{W}} : [\underline{E}_i \cdot \alpha_i]_{i \in I}^{-1}[F] = \bigcup_{i \in I} \alpha_i^{-1}[F] \cap E_i \in \mathcal{B}_{\mathbb{R}}}$$

Indeed:

$$r \in \text{LHS} \iff [\underline{E}_i \cdot \alpha_i]_{i \in I} r \in F \iff \exists i \in I. r \in E_i \wedge \alpha_i r \in F \iff r \in \text{RHS}$$

Example validation: discrete qbs over  $I$

elements:  $I$ ; random elements:  $\sigma$ -simple functions.

constants:  $\underline{i} = [\mathbb{R}.\underline{i}]_{j=1}^1 \in \mathcal{R}_{\bar{I}}$

precomposition:  $[\underline{E}_i.\underline{i}_j]_j \circ \varphi = [\varphi^{-1}[\underline{E}_i].\underline{i}_j]_{j, \varphi^{-1}[\underline{E}_j] \neq \emptyset} \in \mathcal{R}_{\bar{I}}$

recombination: slightly more fiddly, but similar:

$$\left[ \underline{F}_\ell \cdot [\underline{E}_{\ell,j}.\underline{i}_{\ell,j}]_{j \in J_\ell} \right]_{\ell \in L} = [\underline{F}_\ell \cap \underline{E}_{\ell,j}.\underline{i}_{\ell,j}]_{\substack{(\ell,j) \in \coprod_{\ell \in L} J_\ell \\ \underline{F}_\ell \cap \underline{E}_{\ell,j} \neq \emptyset}} \in \mathcal{R}_{\bar{I}}$$

# The category of quasi Borel spaces

Quasi-Borel measurable function  $f : \underline{X} \rightarrow \underline{Y}$

function sending points to points and random elements to random elements:

$$f : \underline{X} \rightarrow \underline{Y} \quad \mathcal{R}_{\underline{X}} \ni \alpha \implies f \circ \alpha \in \mathcal{R}_{\underline{Y}}$$

## Examples

- ▶ **constant** functions  $\underline{b} : \underline{X} \rightarrow \underline{Y}$ , as send any r.e. to a  $\sigma$ -simple function.
- ▶  **$\sigma$ -simple** functions  $\alpha : \mathbb{R} \rightarrow \underline{X}$ , as send any r.e. to a  $\sigma$ -simple function.
- ▶ **Borel measurable** functions are quasi-Borel measurable, by precomposition:

$$f \in \mathbf{Meas}(\underline{M}, \underline{K}) \implies f : M \rightarrow K$$

## Category Qbs

consists of quasi Borel spaces and quasi-Borel measurable functions, with identity functions and function composition.

# Language of probability & distribution (full model)

$X$  type of **values/outcomes**: quasi Borel space

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

$\mathbf{P}X \subseteq \mathbf{D}X$  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.:

$$\mu : \mathbf{D}X, E : \mathcal{B}_X \vdash \text{Ce}_{\mu}[E] : \mathbb{W}$$

are measurable functions

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

$$y_1, y_2 : Y \vdash y_1 \stackrel{Y}{=} y_2 : \text{Prop} \quad \mu : \mathbf{P}X, E : \mathcal{B}_X \vdash \text{cast}_{\mu} \text{Pr}[E] = \text{Ce}_{\mu}[E]$$

are arbitrary set-theoretic propositions

# Lecture plan

## Part 1: the **discrete** model

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



course page

## Part 2: the **full** model

- ▶ Borel sets and measurable spaces
- ▶ Quasi-Borel spaces
- ▶ Type structure & standard Borel spaces
- ▶ Dependently-typed structure
- ▶ Integration



ask questions on the  
Scottish PL Institute  
Zulip stream #qbs

# Expectation management

## Proposition

Take  $I$  set and  $\underline{X}$  qbs:

- ▶ every function  $f : I \rightarrow \underline{X}$  is measurable as  $f : \bar{I} \rightarrow \underline{X}$ .
- ▶ every function  $\underline{X} \rightarrow I$  is measurable as  $f : \underline{X} \rightarrow \text{Indiscrete } I$ .

# Expectation management

## Proposition

Take  $I$  set and  $\underline{X}$  qbs:

- ▶ every function  $f : I \rightarrow \underline{X}$  is measurable as  $f : \bar{I} \rightarrow \underline{X}$ .
- ▶ every function  $\underline{X} \rightarrow I$  is measurable as  $f : \underline{X} \rightarrow \text{Indiscrete } I$ .

## Proof

Every function  $f : I \rightarrow \underline{X}$  sends  $\sigma$ -simple functions to  $\sigma$ -simple functions:

$$f \circ [E_j \cdot i_j]_{j \in J} = [E_j \cdot \underline{f}i_j]_{j \in J} \in \mathcal{R}_{\underline{X}}$$

The function  $f \circ \alpha$  is always a random element in Indiscrete  $I$ . ■

# Expectation management

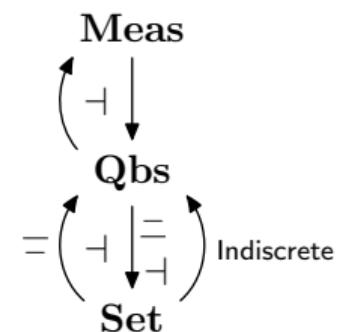
## Proposition

Take  $I$  set and  $X$  qbs:

- ▶ every function  $f : I \rightarrow \underline{X}$  is measurable as  $f : \bar{I} \rightarrow X$ .
- ▶ every function  $\underline{X} \rightarrow I$  is measurable as  $f : X \rightarrow \text{Indiscrete } I$ .

## Adjoint situation

- ▶ limits and colimits are as in **Set**
- ▶ slogan: every measurable space is carried by a qbs.



# Simple type structure

Product  $\prod_{i \in I} X_i \xrightarrow{\pi_j} X_j$

**correlated** random elements metaphorology:

$$\mathcal{R}_{\prod_{i \in I} X_i} := \left\{ (\alpha_i)_{i \in I} : \mathbb{R} \rightarrow \prod_{i \in I} X_i \middle| \forall i : I. \alpha_i \in \mathcal{R}_{X_i} \right\}$$

Coproduct  $\coprod_{i \in I} X_i \xleftarrow{\iota_j} X_j$

metaphorology generated from component metaphorologies:

$$\mathcal{R}_{\coprod_{i \in I} X_i} := \mathcal{R} \left( \bigcup_{i \in I} (\iota_i \circ) [\mathcal{R}_{X_i}] \right)$$

$$= \left\{ [E_j \cdot \alpha_j]_{j \in J} : \mathbb{R} \rightarrow \coprod_{i \in I} X_i \middle| \begin{array}{l} J \subseteq \mathbb{N}, i_- : J \rightarrow I, E_- : J \rightarrow \mathcal{B}_{\mathbb{R}}, \\ \mathbb{R} = \biguplus_{j \in J} E_j, \forall j. \alpha_j \in \mathcal{R}_{X_{i_j}} \end{array} \right\}$$

# Simple type structure

Function space  $(X \rightarrow Y) \times X \xrightarrow{\text{eval}} X$

elements: measurable functions; random elements: curried measurable functions;  
evaluation as in **Set**:

$$\underline{X \rightarrow Y} := \mathbf{Qbs}(X, Y) \quad \mathcal{R}_{X \rightarrow Y} := \mathbf{curry}_X [\mathbf{Qbs}(\mathbb{R} \times X, Y)]$$

$$\text{eval}(f, a) := f a \qquad \qquad = \left\{ \varphi : \mathbb{R} \rightarrow \mathbf{Qbs}(X, Y) \middle| \begin{array}{l} \mathbf{uncurry} \varphi := (\lambda(r, a).(\varphi r)a) \\ : \mathbb{R} \times X \rightarrow Y \end{array} \right\}$$

# Simple type structure

Function space  $(X \rightarrow Y) \times X \xrightarrow{\text{eval}} X$

elements: measurable functions; random elements: curried measurable functions;  
evaluation as in **Set**:

$$\underline{X \rightarrow Y} := \mathbf{Qbs}(X, Y) \quad \mathcal{R}_{X \rightarrow Y} := \mathbf{curry}_X [\mathbf{Qbs}(\mathbb{R} \times X, Y)]$$

$$\text{eval}(f, a) := f a \qquad = \left\{ \varphi : \mathbb{R} \rightarrow \mathbf{Qbs}(X, Y) \middle| \begin{array}{l} \mathbf{uncurry} \varphi := (\lambda(r, a).(\varphi r)a) \\ : \mathbb{R} \times X \rightarrow Y \end{array} \right\}$$

## Meas vs. Qbs

Using the adjunction (to be established), we have:

$$L \begin{array}{c} \dashv \\ \uparrow \end{array} U \downarrow \mathbf{Qbs} \qquad \text{Meas}$$

$$\begin{array}{ccc} (L(\mathbb{R} \rightarrow \mathbb{R})) \times \mathbb{R} & \xrightarrow{\text{Aumann's Thm}} & \mathbb{R} \\ (L\pi_1, L\pi_2) = (\lambda(f, a). (f, a)) \uparrow & = & \parallel \\ L((\mathbb{R} \rightarrow \mathbb{R}) \times \mathbb{R}) & \xrightarrow{\text{Leval}} & L\mathbb{R} \end{array}$$

# Simple type structure

These allow us to interpret a **simple type theory**. We use a **shallow embedding**<sup>2</sup>.

## Simple contexts

- ▶  $x_1 : X_1, \dots, x_n : X_n := \prod_{i=1}^n X_i$
- ▶  $(\Gamma \vdash M : X) := \Gamma \xrightarrow{M} X$

## Measurability by type-checking

Use simple type structure as well as primitive space constructions to define measurable **building blocks**. Instead of proving that a function is measurable, build it from smaller measurable functions using the shallow embedding.

---

<sup>2</sup>Ścibior et al.[POPL'2018] use a deep ambedding, e.g.

## Proposition

For  $X \in \mathbf{Qbs}$ :

$$\underline{\mathbb{R} \rightarrow X} = \mathcal{R}_X$$

## Proof

( $\supseteq$ ). Take  $\alpha \in \mathcal{R}_X$ . By precomposition,  $\alpha : \mathbb{R} \rightarrow X$ . So  $\alpha \in \underline{\mathbb{R} \rightarrow X}$ .

( $\subseteq$ ). Take  $\alpha \in \underline{\mathbb{R} \rightarrow X}$ . Since  $\text{id} \in \mathbf{Qbs}(\mathbb{R}, \mathbb{R}) = \mathcal{R}_{\mathbb{R}}$ ,  $\alpha = \alpha \circ \text{id} \in \mathcal{R}_X$ .

■

Random element space  $\mathcal{R}_X$

the function space  $\mathbb{R} \rightarrow X$ .

## Proposition

For a measurable function  $e : \underline{X} \rightarrow \underline{Y}$ , TFAE, and we then write  $e : \underline{X} \hookrightarrow \underline{Y}$ :

- ▶  $e$  is a **subspace embedding**: a random element  $\beta \in \mathcal{R}_{\underline{Y}}$  that lifts pointwise along  $e$  extends a unique random element along  $e$ :

$$(\forall r : \mathbb{R}. \exists a \in \underline{X}. e a = \beta r) \implies \exists! \alpha \in \mathcal{R}_{\underline{X}}. e \circ \alpha = \beta$$

- ▶  $e$  is injective, and  $\underline{Y}$ 's metaphorology determines  $\underline{X}$ 's:

$$\mathcal{R}_{\underline{X}} = \{\alpha : \mathbb{R} \rightarrow \underline{X} | e \circ \alpha \in \mathcal{R}_{\underline{Y}}\}$$

- ▶  $e$  is a **strong monomorphism**, i.e., right-orthogonal to all epis.

The subspace embedding  $\text{True} : \mathbb{1} \hookrightarrow \text{Indiscrete}\{\text{True}, \text{False}\} =: \text{Prop}$  is a **strong subobject classifier**.

# Internalisation

## Subspaces

We internalise any set-theoretic property  $\varphi : \underline{X} \rightarrow \{\text{True}, \text{False}\}$  into a measurable function  $x : X \vdash \varphi : \text{Prop}$ , forming the subspace over  $\{a \in \underline{X} \mid \varphi a = \text{True}\}$ :

$$\frac{x : X \vdash \varphi : \text{Prop}}{\text{coerce}_\varphi : \{x : X \mid \varphi\} \hookrightarrow X} \quad \frac{\Gamma \vdash M : X \quad \Gamma \vdash \varphi [x \mapsto M]}{\Gamma \vdash \text{lift}_\varphi M : \{x : X \mid \varphi\}}$$

## Examples

- ▶  $\mathbb{N} \hookrightarrow \mathbb{Q} \hookrightarrow \mathbb{R}_{\geq 0} \hookrightarrow \mathbb{R}$
- ▶  $\mathbb{W} \hookrightarrow [-\infty, \infty]$
- ▶ Iverson bracket  $[-] : \mathbb{B} := \overline{\{\text{True}, \text{False}\}} \hookrightarrow \mathbb{R}$
- ▶  $\mathbb{B} \not\hookrightarrow \text{Prop}$ .

# Internalisation

## Proposition

Let  $e : X \hookrightarrow Y$  be a subspace embedding. TFAE, and we say  $e$  is **Borel** and write  $e : X \leftrightarrow Y$ :

- ▶ for every  $\alpha \in \mathcal{R}_Y$ ,  $\alpha^{-1}[e[X]] \in \mathcal{B}_{\mathbb{R}}$ ;
- ▶ the characteristic function is measurable:  $[- \in e[X]] : Y \rightarrow \mathbb{B}$ .

The Borel subspace embedding True :  $\mathbb{1} \leftrightarrow \mathbb{B}$  is a **Borel subspace embedding classifier**.

## Space of Borel subsets $\mathcal{B}_X$

the function space  $X \rightarrow \mathbb{B}$ . It is an internal  $\sigma$ -field:

$$\emptyset : \mathcal{B}_X \quad -^{\complement} : \mathcal{B}_X \rightarrow \mathcal{B}_X \quad \bigcup, \bigcap : (I \rightarrow \mathcal{B}_X) \rightarrow \mathcal{B}_X \quad (I \subseteq \mathbb{N})$$

precisely because  $\mathbb{B}$  is an internal  $\sigma$ -algebra in **Sbs**.

# Internalisation

## Examples

- ▶ Borel subsets of  $\mathbb{R}$  as a quasi Borel space are the usual Borel subsets
- ▶ As we've seen in **Meas**, the Cauchy sequences  $\text{Cauchy} \leftrightarrow (\mathbb{N} \rightarrow \mathbb{R})$ .
- ▶ For discrete spaces,  $\underline{\mathcal{B}_I} = \mathcal{P}I$ .
- ▶ For indiscrete spaces  $\underline{\mathcal{B}_{\text{Indiscrete } I}} = \{\emptyset, I\}$
- ▶ The left adjoint to **Meas**  $\rightarrow$  **Qbs** is given by  $X \mapsto (\underline{X}, \mathcal{B}_X)$ .
- ▶ For measurable space  $M$ , the subsets  $\mathcal{B}_{\mathcal{B}_M}$  are the **Borel-on-Borel** subsets from **classical descriptive set theory** [Sabok et al. '21].

## Non-Examples [Sabok et al. '21]

The following subspace embeddings are not Borel:

$$\{E \in \mathcal{B}_{\mathbb{R}} | E \neq \emptyset\}, \{E \in \mathcal{B}_{\mathbb{R}} | E \text{ is open}\} \hookrightarrow \mathcal{B}_{\mathbb{R}} \quad \{(E, F) \in \mathcal{B}_{\mathbb{R}}^2 | E \subseteq F\} \hookrightarrow \mathcal{B}_{\mathbb{R}}^2$$

Partial map  $f : X \multimap Y$

measurable function  $X \rightarrow Y \amalg \{\perp\} =: Y_\perp$ .

We internalise the partial map space as  $(X \multimap Y) := (X \rightarrow Y_\perp)$ .

The **domain of definition** of such a partial map is a Borel subset  $\text{Dom}(f) \hookrightarrow X$ .

It internalises  $\text{Dom}(-) : (X \multimap Y) \rightarrow \mathcal{B}_X$ .

We use internal partial maps to define the space of distributions.

# Lecture plan

## Part 1: the **discrete** model

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



course page

## Part 2: the **full** model

- ▶ Borel sets and measurable spaces
- ▶ Quasi-Borel spaces
- ▶ Type structure & standard Borel spaces
- ▶ Dependently-typed structure
- ▶ Integration



ask questions on the  
Scottish PL Institute  
Zulip stream #qbs

# Concrete spaces

$\mathbf{Sbs} \hookrightarrow \mathbf{Qbs}$

A quasi Borel space space  $S \in \mathbf{Qbs}$  is **standard Borel** when there is a measurable isomorphism  $S \cong E$  for some  $E \in \mathcal{B}_{\mathbb{R}}$ .

**Proposition**

The adjunction between  $\mathbf{Meas}$  and  $\mathbf{Qbs}$  restricts to an isomorphism between the corresponding categories of standard Borel spaces:

$$\begin{array}{ccccc} & \mathbf{Sbs} & \xrightarrow{\quad} & \mathbf{Meas} & \\ \mathbf{Sbs} & \xrightarrow{L} & = & \xleftarrow{L} & \\ & \parallel & & \parallel & \\ & \mathbf{Qbs} & \xrightarrow{\quad} & \mathbf{Meas} & \\ & \xrightarrow{R} & = & \xleftarrow{R} & \\ & \parallel & & \parallel & \\ & \mathbf{Sbs} & \xleftarrow{\quad} & \xrightarrow{\text{unit}} & \\ & \parallel & & \parallel & \\ & \mathbf{Sbs} & \xrightarrow{\quad} & \xleftarrow{L} & \end{array}$$

# Concrete spaces

Standard Borel spaces have good properties.

E.g.: they have disintegrations along arbitrary maps.

## Theorem (well-known)

A measurable space is standard iff its  $\sigma$ -field is generated by a **Polish topology**: completely metrizable separable topology.

Quasi Borel spaces offer an alternative methodology for showing a space is standard:

### classical theory

### quasi Borel spaces

- ▶ Equip the space with a topology or a  $\sigma$ -field.
- ▶ Show it is standard, typically by metrizing it.
- ▶ Show compatibility with operations of interest, e.g., evaluation is measurable.
- ▶ Use a suitable space, e.g., for evaluation.
- ▶ Exhibit an isomorphism to a standard space.
- ▶ Optionally, study a compatible metric.

## Example (well-known)

Let  $\mathbf{C}_0[a,b]$  be the quasi-Borel space of continuous functions over  $[a,b]$ :

$$\mathbf{C}_0[a,b] := \{f : [a,b] \rightarrow \mathbb{R} \mid f \text{ continuous}\} \hookrightarrow ([a,b] \rightarrow \mathbb{R})$$

Immediately, without need of further proof, we have a quasi-Borel space that supports measurable evaluation, after coercion, and measurable abstraction, subject to lifting.

# Concrete spaces

## Example (well-known)

Let  $\mathbf{C}_0[a,b]$  be the quasi-Borel space of continuous functions over  $[a,b]$ :

$$\mathbf{C}_0[a,b] := \{f : [a,b] \rightarrow \mathbb{R} \mid f \text{ continuous}\} \hookrightarrow ([a,b] \rightarrow \mathbb{R})$$

## Representation as standard space

A continuous function is determined uniquely by its values on rationals, a countably-long vector of real numbers:

$$-|_{\mathbb{Q} \cap [a,b]} : \mathbf{C}_0 \rightarrow (\mathbb{Q} \cap [a,b] \rightarrow \mathbb{R})$$

# Concrete spaces

## Example (well-known)

Let  $\mathbf{C}_0[a,b]$  be the quasi-Borel space of continuous functions over  $[a,b]$ :

$$\mathbf{C}_0[a,b] := \{f : [a,b] \rightarrow \mathbb{R} \mid f \text{ continuous}\} \hookrightarrow ([a,b] \rightarrow \mathbb{R})$$

## Representation as standard space

The key is to characterise the image as a Borel subset:

### Theorem (Cantor)

A function  $f : [a,b] \rightarrow \mathbb{R}$  is continuous iff it is uniformly continuous.

So lift  $-|$  to:

$$\begin{aligned} -|_{\mathbb{Q} \cap [a,b]} &:= (\lambda f. \text{lift } \lambda q. f q) : \mathbf{C}_0 \rightarrow \left\{ r_- : \mathbb{Q} \cap [a,b] \rightarrow \mathbb{R} \mid \begin{array}{l} \forall \varepsilon : \mathbb{Q}_{>0}. \exists \delta : \mathbb{Q}_{>0}. \forall p, q : \mathbb{Q} \cap [a,b]. \\ |p - q| \leq \delta \implies |r_p - r_q| \leq \varepsilon \end{array} \right\} \\ &=: D \leftrightarrow (\mathbb{Q} \cap [a,b] \rightarrow \mathbb{R}) \in \mathbf{Sbs} \end{aligned}$$

# Concrete spaces

## Example (well-known)

Let  $\mathbf{C}_0[a,b]$  be the quasi-Borel space of continuous functions over  $[a,b]$ :

$$\mathbf{C}_0[a,b] := \{f : [a,b] \rightarrow \mathbb{R} \mid f \text{ continuous}\} \hookrightarrow ([a,b] \rightarrow \mathbb{R})$$

## Representation as standard space

Conversely, define the inverse by taking the limit<sup>3</sup>:

$$[-] := \lambda r_-. \lambda x. \lim_{n \rightarrow \infty} r_{\text{approx}} \lambda m. \frac{1}{1+m} xn : D \rightarrow \mathbf{C}_0$$

We now have a standard Borel space equipped with operations of interests, such as abstraction and evaluation.

---

<sup>3</sup>I'm glossing over the edge case  $x = b$ , but a case-split will deal with it.

# Concrete spaces

To fully reproduce the classical account, exhibiting a separable, complete metric that generates the  $\sigma$ -field, we typically need to do more or less the same calculations. To that end, the following two concepts break the task down.

## Compatible metric

over a qbs  $X$  is a metric that is also measurable  $d : X \times X \rightarrow \mathbb{W}$ . It has **measurable limits** when the limit function is a measurable partial function  $\lim : (\mathbb{N} \rightarrow X) \rightarrow X_{\perp}$ .

## Theorem

Let  $d$  be a compatible metric with measurable limits over a qbs  $X$ .

If  $d$  is separable, then  $\mathcal{B}_d = \mathcal{B}_X$ .

## Example

The **uniform convergence metric**  $d(f, g) := \sup_{x \in [a, b]} |f x - g x|$  is compatible over  $C_0[a, b]$ , since we can define it equivalently as  $d(f, g) := \sup_{x \in \mathbb{Q} \cap [a, b]} |f x - g x|$ . It has measurable limits, since it is complete and its limits are taken pointwise. It is separable by Weierstrass's approximation theorem.

# Lecture plan

## Part 1: the **discrete** model

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



course page

## Part 2: the **full** model

- ▶ Borel sets and measurable spaces
- ▶ Quasi-Borel spaces
- ▶ Type structure & standard Borel spaces
- ▶ Dependently-typed structure
- ▶ Integration



ask questions on the  
Scottish PL Institute  
Zulip stream #qbs

# Core idea

## Core advantages

A **convenient** setting for dependently-typed probability:

- ▶ possesses enough type constructions.
- ▶ strictly preserves substitution.
- ▶ smooth internalisation and externalisation.

## Main ideas

- ▶ Use equivalent structure to locally Cartesian-closed structure, guaranteeing enough type constructions.
- ▶ Use families, deferring most issues involving type dependency to meta-level, and focus on measurability structure and requirements.

# Core definitions

$\Gamma$ -fibred metaphorology  $\mathcal{R}^-$  over  $\gamma : \underline{\Gamma} \vdash \underline{X}_\gamma$

$\mathcal{R}_\Gamma$ -indexed family of subsets  $v : \mathcal{R}_\Gamma \vdash \mathcal{R}^v \subseteq (r : \mathbb{R}) \rightarrow \underline{X}_{vr}$  closed under:

**fibred constants:**

$$\frac{\gamma \in \underline{\Gamma} \quad a \in \underline{X}_\gamma}{\underline{a} := (\lambda r.a) \in \mathcal{R}^\gamma}$$

**fibred precomposition:**

$$\frac{\alpha \in \mathcal{R}^v \quad \varphi \in \text{Qbs}(\mathbb{R}, \mathbb{R})}{(\alpha \circ \varphi) \in \mathcal{R}^{v \circ \varphi}}$$

**fibred recombination:**

$$\frac{I \subseteq \mathbb{N} \quad E_- : I \rightarrow \mathcal{B}_\mathbb{R} \quad \mathbb{R} = \biguplus_{i \in I} E_i \quad \alpha_- : (i : I) \rightarrow \mathcal{R}^{v_i}}{[E_i \cdot \alpha_i]_{i \in I} \in \mathcal{R}^{[E_i \cdot v_i]_{i \in I}}}$$

# Core definitions

$\Gamma$ -fibred metaphorology  $\mathcal{R}^-$  over  $\gamma : \underline{\Gamma} \vdash \underline{X}_\gamma$

$\mathcal{R}_\Gamma$ -indexed family of subsets  $v : \mathcal{R}_\Gamma \vdash \mathcal{R}^v \subseteq (r : \mathbb{R}) \rightarrow \underline{X}_{vr}$  closed under:

**fibred constants:**

$$\frac{\gamma \in \underline{\Gamma} \quad a \in \underline{X}_\gamma}{\underline{a} := (\lambda r.a) \in \mathcal{R}^\gamma}$$

**fibred precomposition:**

$$\frac{\alpha \in \mathcal{R}^v \quad \varphi \in \mathbf{Qbs}(\mathbb{R}, \mathbb{R})}{(\alpha \circ \varphi) \in \mathcal{R}^{v \circ \varphi}}$$

**fibred recombination:**

$$\frac{I \subseteq \mathbb{N} \quad E_- : I \rightarrow \mathcal{B}_\mathbb{R} \quad \mathbb{R} = \biguplus_{i \in I} E_i \quad \alpha_- : (i : I) \rightarrow \mathcal{R}^{v_i}}{[E_i \cdot \alpha_i]_{i \in I} \in \mathcal{R}^{[E_i \cdot v_i]_{i \in I}}}$$

# Core definitions

$\Gamma$ -fibred metaphorology  $\mathcal{R}^-$  over  $\gamma : \underline{\Gamma} \vdash \underline{X}_\gamma$

$\mathcal{R}_\Gamma$ -indexed family of subsets  $v : \mathcal{R}_\Gamma \vdash \mathcal{R}^v \subseteq (r : \mathbb{R}) \rightarrow \underline{X}_{vr}$  closed under:

**fibred constants:**

$$\frac{\gamma \in \underline{\Gamma} \quad a \in \underline{X}_\gamma}{\underline{a} := (\lambda r.a) \in \mathcal{R}^\gamma}$$

**fibred precomposition:**

$$\frac{\alpha \in \mathcal{R}^v \quad \varphi \in \text{Qbs}(\mathbb{R}, \mathbb{R})}{(\alpha \circ \varphi) \in \mathcal{R}^{v \circ \varphi}}$$

**fibred recombination:**

$$\frac{I \subseteq \mathbb{N} \quad E_- : I \rightarrow \mathcal{B}_\mathbb{R} \quad \mathbb{R} = \biguplus_{i \in I} E_i \quad \alpha_- : (i : I) \rightarrow \mathcal{R}^{v_i}}{[E_i \cdot \alpha_i]_{i \in I} \in \mathcal{R}^{[E_i \cdot v_i]_{i \in I}}}$$

# Core definitions

$\Gamma$ -fibred metaphorology  $\mathcal{R}^-$  over  $\gamma : \underline{\Gamma} \vdash \underline{X}_\gamma$

$\mathcal{R}_\Gamma$ -indexed family of subsets  $v : \mathcal{R}_\Gamma \vdash \mathcal{R}^v \subseteq (r : \mathbb{R}) \rightarrow \underline{X}_{vr}$  closed under:

**fibred constants:**

$$\frac{\gamma \in \underline{\Gamma} \quad a \in \underline{X}_\gamma}{\underline{a} := (\lambda r.a) \in \mathcal{R}^\gamma}$$

**fibred precomposition:**

$$\frac{\alpha \in \mathcal{R}^v \quad \varphi \in \mathbf{Qbs}(\mathbb{R}, \mathbb{R})}{(\alpha \circ \varphi) \in \mathcal{R}^{v \circ \varphi}}$$

**fibred recombination:**

$$\frac{I \subseteq \mathbb{N} \quad E_- : I \rightarrow \mathcal{B}_\mathbb{R} \quad \mathbb{R} = \biguplus_{i \in I} E_i \quad \alpha_- : (i : I) \rightarrow \mathcal{R}^{v_i}}{[E_i \cdot \alpha_i]_{i \in I} \in \mathcal{R}^{[E_i \cdot v_i]_{i \in I}}}$$

Quasi Borel family (qbf)  $\Gamma \vdash X_-$

a family  $\gamma : \underline{\Gamma} \vdash \underline{X}_\gamma$  equipped with a  $\Gamma$ -fibred metaphorology  $\mathcal{R}_X^-$  over  $\underline{X}_-$ .

## Examples

- ▶ Every qbs  $X$  is a qbf  $\Gamma \vdash X$  with  $(\underline{X})_\gamma := X$  and  $\mathcal{R}_X^v := \mathcal{R}_X$ .
- ▶ We will later see that intervals are qbfs  $a, b : [-\infty, \infty] \vdash [a, b], [a, b), (a, b], (a, b)$  with  $\mathcal{R}^{a \mapsto \alpha, b \mapsto \beta}$  those measurable functions  $\varphi \in \mathbf{Qbs}(\mathbb{R}, [-\infty, \infty])$  such that, for all  $r \in \mathbb{R}$ ,  $\varphi r$  is in the interval delimited by  $\alpha r$  and  $\beta r$ .
- ▶ For measurable map  $\frac{X}{\Gamma} \downarrow d$ , the **preimage** qbf:

$$\gamma : \Gamma \vdash d^{-1}[\gamma] \quad \mathcal{R}_{d^{-1}[-]}^v := \left( \frac{\mathcal{R}_X}{\mathcal{R}_\Gamma} \right)^{-1} [v] \subseteq \mathcal{R}_X$$

# Core definitions

Qbf map  $(\theta \vdash f_-) : (\gamma : \Gamma \vdash X_\gamma) \rightarrow (\delta : \Delta \vdash Y_\delta)$

a family map  $(\theta \vdash f_-) : (\gamma : \underline{\Gamma} \vdash \underline{X}_\gamma) \rightarrow (\delta : \underline{\Delta} \vdash \underline{Y}_\delta)$  such that:

$$\mathcal{R}_X^v \ni \alpha \quad \implies \quad (f_- \stackrel{\theta}{\circ} \alpha) := (\lambda r. f_{\theta r}(\alpha r)) \in \mathcal{R}_Y^{\theta \circ v}$$

A **vertical map**  $\Gamma \vdash f_- : X_- \rightarrow Y_-$  is a qbf map of the form:

$$(\text{id}_\Gamma \vdash f_-) : (\Gamma \vdash X_-) \rightarrow (\Gamma \vdash Y_-)$$

# Core definitions

Qbf map  $(\theta \vdash f_-) : (\gamma : \Gamma \vdash X_\gamma) \rightarrow (\delta : \Delta \vdash Y_\delta)$

a family map  $(\theta \vdash f_-) : (\gamma : \underline{\Gamma} \vdash \underline{X}_\gamma) \rightarrow (\delta : \underline{\Delta} \vdash \underline{Y}_\delta)$  such that:

$$\mathcal{R}_X^v \ni \alpha \quad \Rightarrow \quad (f_- \stackrel{\theta}{\circ} \alpha) := (\lambda r. f_{\theta r}(\alpha r)) \in \mathcal{R}_Y^{\theta \circ v}$$

A **vertical map**  $\Gamma \vdash f_- : X_- \rightarrow Y_-$  is a qbf map of the form:

$$(\text{id}_\Gamma \vdash f_-) : (\Gamma \vdash X_-) \rightarrow (\Gamma \vdash Y_-)$$

We use only vertical maps explicitly. The totality maps forms the foundation though:

## Theorem

The category of qbfs and their maps, **Qbf**, equipped with the functor  $(\Gamma \vdash X_-) \mapsto \Gamma : \mathbf{Qbf} \rightarrow \mathbf{Qbs}$  form a split fibred category.

The Grothendieck construction and the preimage form a fibred adjoint equivalence between this fibred category and the codomain fibration over **Qbs**.

# Contextual structure

Substitution  $\Gamma \vdash -[\theta]$

measurable function  $\theta : \Gamma \rightarrow \Delta$  operates on qbfs and their vertical maps:

$$\frac{\delta : \Delta \vdash X_-}{\gamma : \Gamma \vdash X_-[\theta]_\gamma} \quad \underline{X_-[\theta]}_\gamma := X_{\theta\gamma} \quad \mathcal{R}_{X_-[\theta]}^v := \mathcal{R}_{X_-}^{\theta \circ v}$$
$$\frac{\delta : \Delta \vdash f_\delta : X_\delta \rightarrow Y_\delta}{\gamma : \Gamma \vdash f_-[\theta]_\gamma := f_{\theta\gamma} : X_{\theta\gamma} \rightarrow Y_{\theta\gamma}}$$

Fibred terminal qbf  $\Gamma \vdash \mathbb{1}$

given by the terminal qbf considered as a family. NB:  $\mathbb{1}[\theta] = \mathbb{1}$ .

Terms  $\gamma : \Gamma \vdash M_\gamma : X_\gamma$

vertical maps  $\Gamma \vdash M_- : \mathbb{1} \rightarrow X_-$ . NB:  $\frac{\delta : \Delta \vdash M_\delta : X_\delta}{\Gamma \vdash M_-[\theta]_\gamma = M_{\theta\gamma} : X_{\theta\gamma}}$

# Contextual structure

Context extension  $\gamma : \Gamma, x : X_\gamma$

given by the Grothendieck construction:

$$\underline{\gamma : \Gamma, x : X_\gamma} := \{\gamma, (x \mapsto a) | \gamma \in \underline{\Gamma}, a \in \underline{X_a}\} = \coprod_{\gamma \in \underline{\Gamma}} \underline{X_\gamma}$$

$$\mathcal{R}_{\gamma : \Gamma, x : X_\gamma} := \{(v, \alpha) | v \in \mathcal{R}_\Gamma, \alpha \in \mathcal{R}_X^v\}$$

and we have weakening substitution and variable term:

$$(\gamma : \Gamma, x : X_\gamma) \xrightarrow{\text{weaken}^{x:X}} \Gamma \quad \gamma : \Gamma, x : X_\gamma \vdash x : X_\gamma = X - [\text{weaken}]_{\gamma, x \mapsto x}$$

# Contextual structure

Extensional propositional equality  $\frac{\gamma : \Gamma \vdash M_\gamma, N_\gamma : X_\gamma}{\gamma : \Gamma \vdash M_\gamma =:= K_\gamma}$

fibred sub-terminal, non-empty only in fibres where  $M_\gamma = K_\gamma$ :

$$\underline{M =:= K}_\gamma := \begin{cases} M_\gamma() = K_\gamma() : & \mathbb{1} \\ \text{otherwise :} & \emptyset \end{cases} \quad \mathcal{R}_{M =:= K}^v := \begin{cases} \forall r : \mathbb{R}. M_{vr}() = K_{vr}() : & \{ \underline{() \ } \} \\ \text{otherwise :} & \emptyset \end{cases}$$
$$\frac{\gamma : \Gamma \vdash M_\gamma : X_\gamma}{\gamma : \Gamma \vdash (\text{refl } M)_\gamma := () : (M_- =:= M_-)_\gamma}$$

# Contextual structure

## Extensional propositional fording

For, e.g., transport, generalise prop. equality to package under equality assumption:

$$\frac{\gamma : \Gamma \vdash M_\gamma, K_\gamma : X_\gamma \quad \gamma : \Gamma, x : X_\gamma \vdash Y_{\gamma, x \mapsto x}}{\gamma : \Gamma \vdash ((M =: x := K) \times Y_{-, x \mapsto x})_\gamma}$$

$$\frac{}{((M =: x := K) \times Y_{-, x \mapsto x})_\gamma} := \begin{cases} M_\gamma() =: a := K_\gamma() : Y_{\gamma, x \mapsto a} \\ \text{otherwise : } \emptyset \end{cases}$$

$$\mathcal{R}_{(M =: x := K) \times Y_{-, x \mapsto x}}^v := \begin{cases} M_v() =: \alpha := K_v() : \mathcal{R}_{Y_-}^{v, x \mapsto \alpha} \\ \text{otherwise : } \emptyset \end{cases}$$

$$\frac{\Gamma \vdash M : X_- \quad \Gamma \vdash N : Y_{-, x \mapsto M}}{\Gamma \vdash \text{refl } M, N := N_\gamma() : (M =: x := M) \times Y_{-, x \mapsto x}}$$

$$\frac{\Gamma, x : X_- \vdash Y_- \quad \Gamma, y : X_- \vdash Z_- \quad \Gamma \vdash M, K : X_- \quad \Gamma \vdash f_- : Y_{-, x \mapsto M} \rightarrow Z_{-, y \mapsto M}}{\Gamma \vdash \text{match } - \text{ with } \{f_-\} := (\lambda b. f_{\gamma, x \mapsto M_\gamma()} b) : (M =: x := K) \times Y_- \rightarrow Z_{-, y \mapsto K}}$$

## Dependent Pairs

$$\boxed{\prod_{a \in A} A} := \boxed{[\Gamma, a:A \vdash B]}$$
$$\frac{\Gamma, a:A \vdash B}{\Gamma \vdash \prod_{a \in A} B}$$

:=

$$\begin{array}{c} \downarrow \text{dep}_B \\ \boxed{[\Gamma, a:A]} \\ \downarrow \\ \boxed{[\Gamma]} \end{array}$$

$\text{dep}_{\prod}$

## Dependent products

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

aha:  $(a:A) \rightarrow B$

$$\prod_{a:A} B$$

$$[\Gamma \vdash \prod_{a:A} B] :=$$

$$\left\{ (m_0, f : \{ a \in [\Gamma A] \mid \text{dep } a = m_0 \} \rightarrow [\Gamma, a:A \vdash B]) \middle| \right. \\ \left. \forall a \in [\Gamma, a:A]. \text{dep } a = m_0 \Rightarrow \text{dep}(f a) = a \right\}$$

Exercise: Find the random elements.

## Full model

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

$\Delta X := (\text{Fri})$

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

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

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

## Plan:

- 1) Type-driven probability: discrete case ✓
  - 2) Borel sets & measurable spaces ✓
  - 3) Quasi Borel spaces ✓
  - 4) Type structure & standard Borel spaces ✓
  - 5) Integration & random variables
- Lecture 1
- Lecture 2

please ask questions!

Scanneable



← Course  
web  
page

## Plan:

- 1) Type-driven probability: discrete case ✓
  - 2) Borel sets & measurable spaces ✓
  - 3) Quasi Borel spaces ✓
  - 4) Type structure & standard Borel spaces ✓
  - 5) Integration & random variables
- Lecture 1
- Lecture 2

please ask questions!

Scalable



← Course  
web  
page

## Full model

type : Qbs    W := [0, ∞]    B<sup>X</sup> := B<sup>X</sup>

D<sup>X</sup> :=

P<sup>X</sup> := {μ ∈ D<sup>X</sup> | Ce<sub>μ</sub>[X] = 1}

Ce<sub>μ</sub>[E] :=       S<sub>μ</sub> :=

Φ<sup>μ</sup> k :=

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

$$\mu : \mathcal{B}_{\mathbb{R}} \rightarrow \mathbb{W} := [0, \infty]$$

Satisfying the measure axioms:

$$\mathbb{E} : \mathcal{B}^{\omega} \rightarrow$$

$$\mu \phi = 0, \quad \mu E = \mu(E \cap F) + \mu(E \cap F^c), \quad \mu(\cup_{n=1}^{\infty} E_n) = \sup_n \mu E_n$$

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

We write  $\mathcal{M}_V$  for the set of measures on  $V$

For qbs  $X$ , take  $\mathcal{M}_X^{\text{meas}}$

Thm (Lebesgue measure):

There is a unique measure  $\lambda \in \mathcal{M}(R)$ , s.t.:

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

## The unrestricted Giry spaces

Equip  $\llbracket GV \rrbracket$  with two gbs structures:

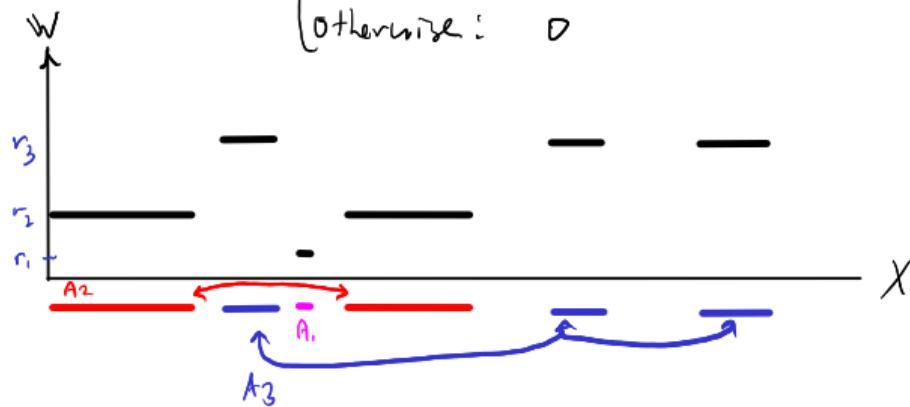
X  $R_{GV} := \{ \alpha: R \rightarrow GV \mid \forall A \in B_V, \lambda r. \alpha(r, A): R \rightarrow W \}$

- ✓  $GV \hookrightarrow W^{B_X}$
- $\hookrightarrow \alpha$  is a kernel.
  - Fewer random elnts
  - $R_{GV} \subseteq R_{G'V}$
  - Lebesgue integral  
measurable in  
both arguments.  
(upcoming)

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

$\exists n \in \mathbb{N}, \vec{A} \in \mathcal{B}_X^n, A_i \cap A_j = \emptyset, r_i \in W$  s.t.  
 $(i \neq j)$

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



Encode into a space:

$$\text{SimpleCode} := \prod_{n \in \mathbb{N}} \mathcal{B}_X^n \times \mathcal{W}^n$$

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

and define an interpretation:

$$[\![ - ]\!]: \text{SimpleCode} \longrightarrow \text{Simple}$$

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

↳ characteristic function  
for  $A_i$

Lemma:  $f: X \rightarrow W$  is measurable → remember!  
gloss morphism!

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

$f_n \in \text{Simple}$ .

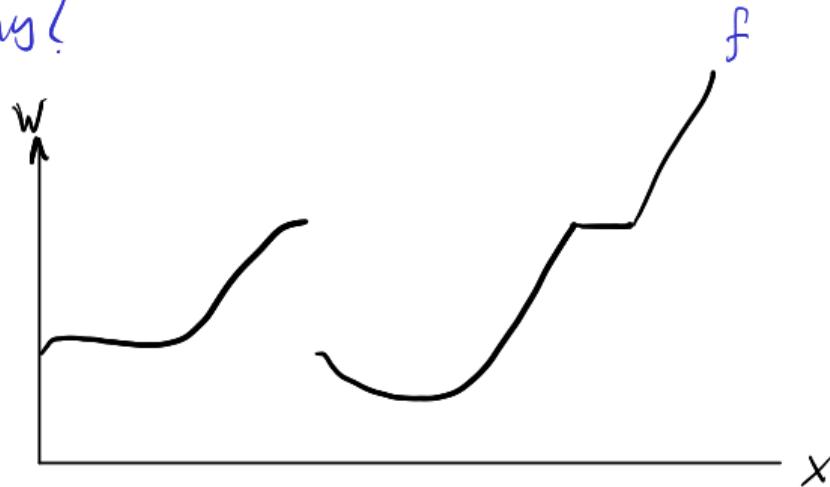
Moreover, we have measurable such choice:

Simple Approx:

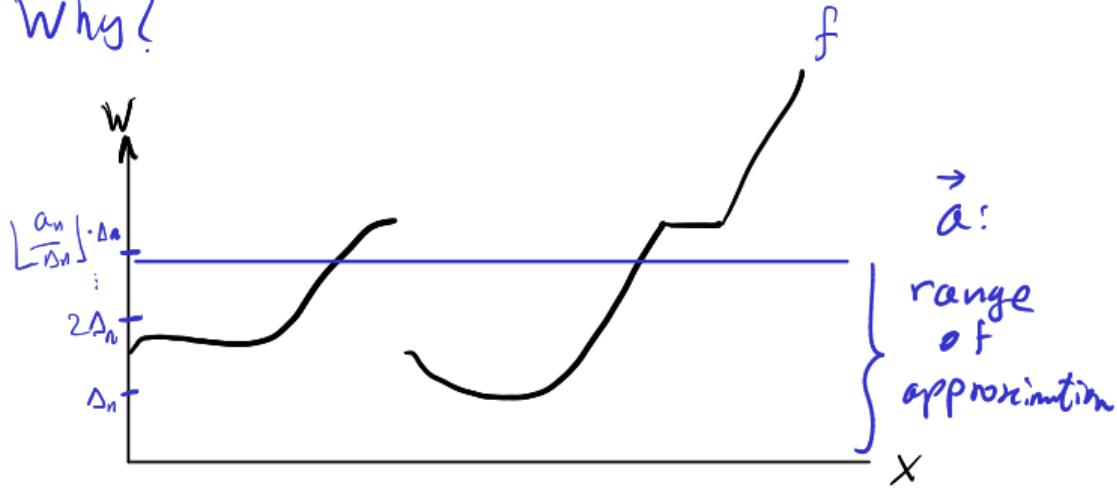
$$\left\{ \vec{\alpha} \in \mathbb{R}^+ \mid \Delta_n \rightarrow 0 \right\} \times \left\{ \vec{\alpha}' \in W^N \mid \begin{array}{l} \vec{\alpha}' \text{ monotone} \\ a_n \rightarrow \infty \end{array} \right\} \times W \xrightarrow{X} \text{SimpleCode}$$

↑  
rate of  
convergence      ↑  
range of  
approximation

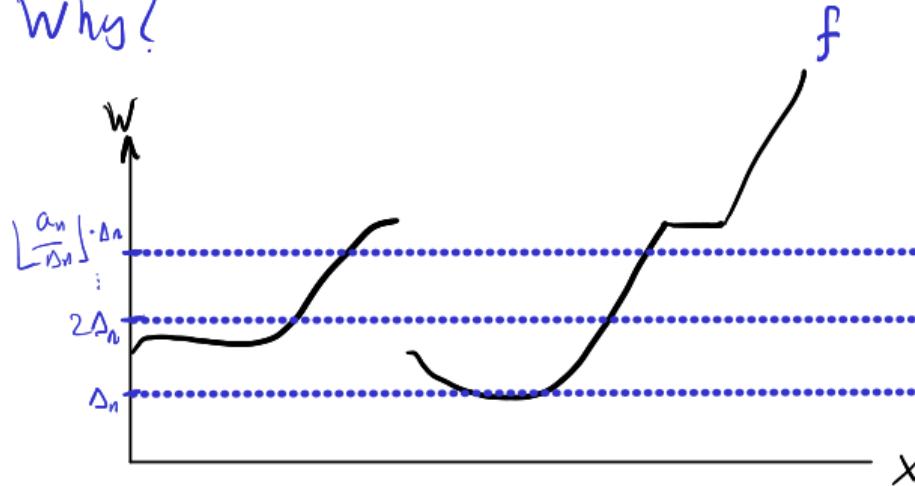
Why?



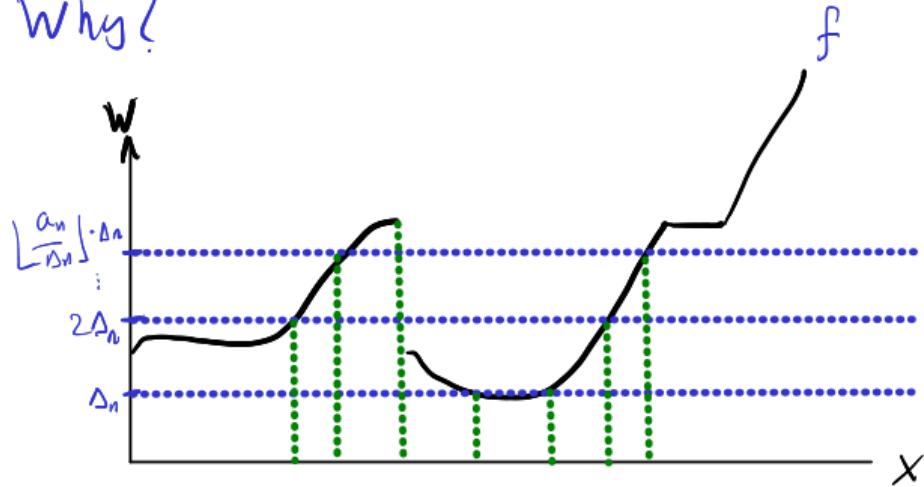
Why?



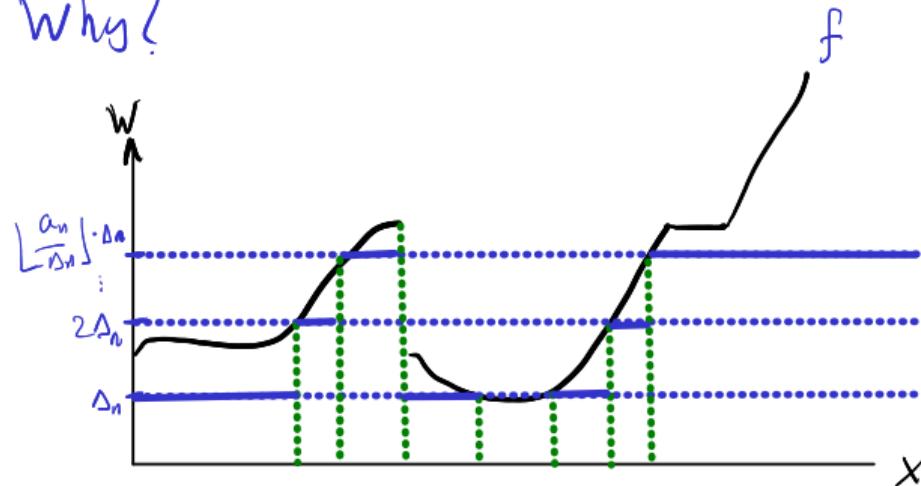
Why?



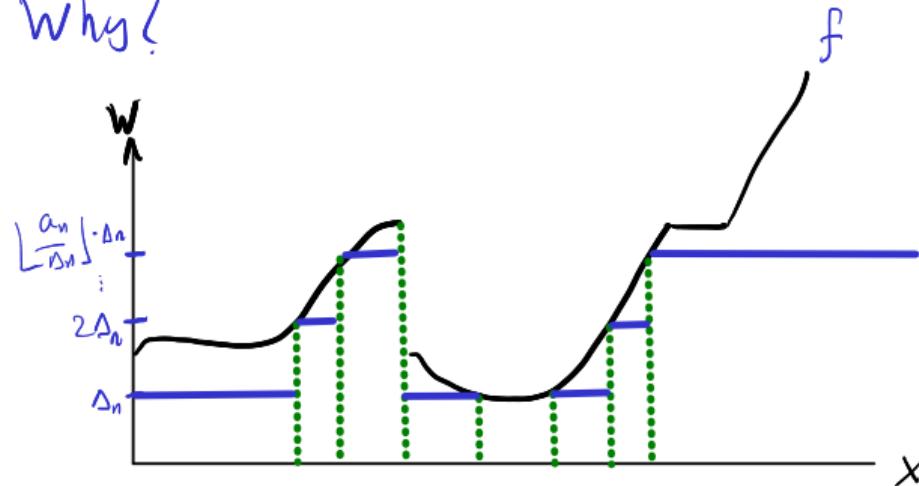
Why?



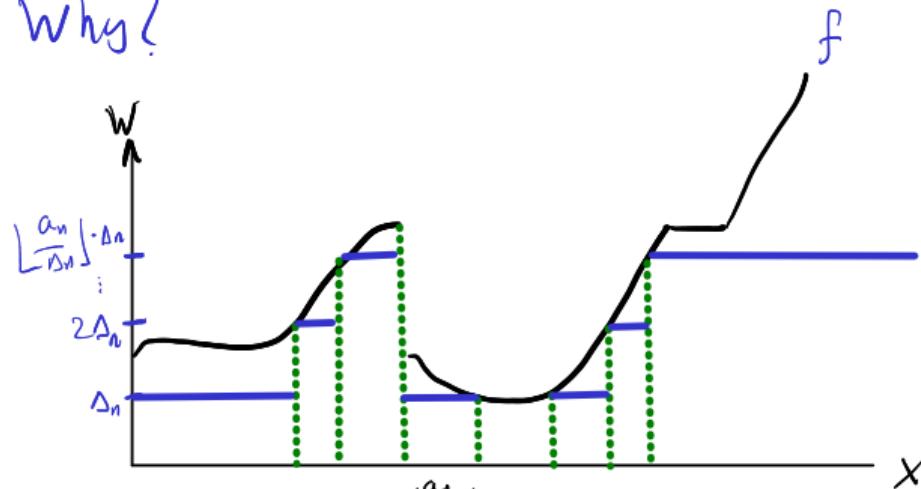
Why?



Why?

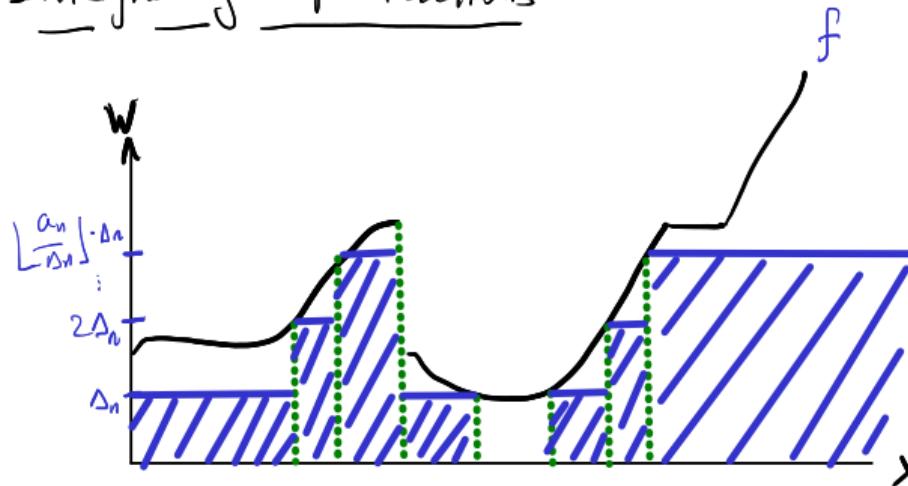


Why?



$$\| \text{Simple Approx}_{\Delta_n} f \| := \sum_{i=1}^{\lfloor \frac{\alpha_n}{\Delta_n} \rfloor} i \cdot \Delta_n [ i \cdot \Delta_n \leq f < (i+1) \Delta_n ] + \left[ \frac{\alpha_n}{\Delta_n} \Delta_n \cdot [ f \geq \lfloor \frac{\alpha_n}{\Delta_n} \rfloor \cdot \Delta_n ] \right] \in \text{Simple}$$

## Integrating Simple Functions



$\int : G \times \text{Simple Code} \rightarrow W$

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

# Integration

Proper higher-order  
operation

$$\int : G \times W^X \rightarrow W$$

$$\int \mu f := \sup \left\{ \int \mu \varphi \mid \varphi \in \text{Simple}, \varphi \leq f \right\}$$

we also  
write

$$= \lim_{n \rightarrow \infty} \int \mu(\text{Simple Approx}_{\vec{\Delta}, \vec{a}} f)_n \sim \begin{matrix} \text{measurable} \\ \text{by} \\ \text{type} \end{matrix}$$

$$\int \mu(\delta x) t$$

$$\text{for } \int \mu(x, t)$$

$$\text{for } \frac{a_n}{\Delta_n} \rightarrow 0, \text{ e.g. } \Delta_n = \frac{1}{2^n}, a_n = n.$$

resolution

The unrestricted Giry Strong Monad

Dirac:

$$\delta: X \rightarrow Gx$$

$$x \mapsto \lambda A. \begin{cases} x \in A : 1 \\ x \notin A : 0 \end{cases}$$

Unlike the unrestricted  
Giry on Meas.

but: non-commutative

Kleisli extension/ Kock integral:

$$\oint: Gx \times Gy^x \rightarrow Gy$$

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

(Fubini Rules,  
just like in  
Meas)

Fubini-Tonelli; fails

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

$\lambda \in G/R$

lebesgue

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

$$\int \#(dx) \underbrace{\int \lambda(dx) k(x,y)}_{y \in \mathbb{R} \setminus \{x\}} = \int \# \Omega = \infty \stackrel{?}{=} 0$$

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

$\neq$

$$\int \lambda(dx) \underbrace{\#(dr) k(x,r)}_{x \in \mathbb{R} \setminus \{r\}} = \int \lambda(dx) \stackrel{?}{=} \infty$$

$\neq 1$

## Randomisable measures monad

$D \rightarrow G$

$$\lambda A. \int_{\text{Dom } \alpha} \lambda(D \alpha)$$

$$LDX := \left\{ \lambda \alpha \mid \alpha: \mathbb{R} \rightarrow X \right\}$$

$$R_{Dx} := \left\{ \lambda x. \lambda_{\alpha x} \mid \alpha: \mathbb{R} \times \mathbb{R} \rightarrow X \right\}$$

$$\delta: x \rightarrow Dx \quad \wp: D^{\mathbb{P}^*(Dx)} \rightarrow Dx \quad \text{lift along } D \rightarrow G.$$

$D$  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  $Meas$ .

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

see [Staton'16]

$$R_{DS} = \{ K : \mathbb{R} \rightarrow G0 \mid K \text{ s-finite kernel} \}$$

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 gbs  $\Omega$  with  
 $\mu \in D\chi$ .

Similarly:  
- finite measure space  
- (Sub) probability space.

$$P\chi := \left\{ \mu \in D\chi \mid \mu \chi = 1 \right\}$$

$$P_{\leq 1}\chi := \left\{ \mu \in D\chi \mid \mu \chi \leq 1 \right\}$$

$$P_{<\infty}\chi := \left\{ \mu \in D\chi \mid \mu \chi < \infty \right\}$$

$$D\chi$$

## Full model

$$\text{type} : \mathbb{Q}_{\text{obs}} \quad W := [0, \infty] \quad \mathcal{B}^X \cong \mathbb{B}^X$$

$$DX := (\{\lambda_\alpha | \alpha : R \rightarrow X\}, \{\lambda_{r,\alpha} | \alpha : R \times R \rightarrow X\})$$

$$Px := \{\mu \in \Omega_X | \underset{\mu}{\text{Ce}}[X] = 1\}$$

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

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

## Plan:

- 1) Type-driven probability: discrete case ✓
  - 2) Borel sets & measurable spaces ✓
  - 3) Quasi Borel spaces ✓
  - 4) Type structure & standard Borel spaces ✓
  - 5) Integration & random variables ✓
- Lecture 1
- Lecture 2

Please ask questions!

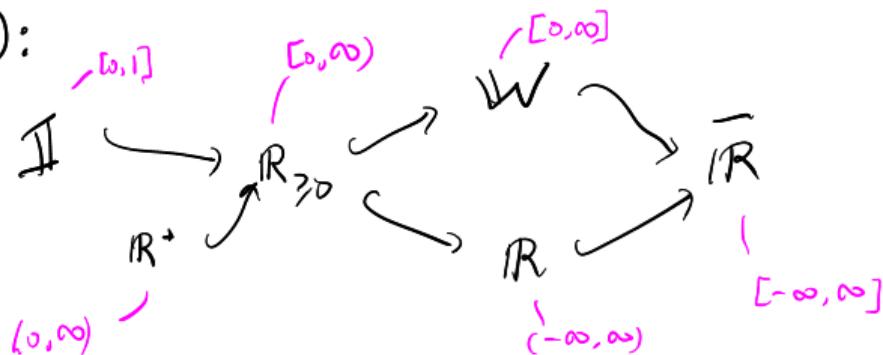
Scalable



← Course  
web  
page

Random variable:  $\xi : \Omega \rightarrow \mathbb{H} \hookrightarrow \bar{\mathbb{R}}$

$\mathbb{H}:$



-  $\mathbb{H}^\Omega$  is a space

-  $W^\Omega$  measurable  $\sigma$ -Semi-module  
for  $W$ :

$$\sum_{n=0}^{\infty} \alpha_n \xi_n :=$$

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

$$\alpha \xi + \zeta := \lambda w \cdot \alpha \cdot \xi w + \zeta w$$

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

$$\Pr_x : P_{\Omega} \times \mathcal{B}_{\Omega} \rightarrow \mathbb{W}$$

$$\Pr_x A := \text{eval}(x, A) = \lambda A$$

Probability Space  $\mathcal{R} = (\Omega, \lambda_{\Omega})$

$P : P_{\Omega} \vdash$  "  $P_{\Omega}$  holds  $\lambda(\Omega)$ -almost surely "

for some  $Q \rightarrowtail \Omega$ ,  $P \models Q$ ,  $[-\epsilon Q] \cdot \lambda = \lambda$

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

$\xi = \zeta$  a.s., when  $\Pr_{\omega \sim \Omega} [\xi_{\omega} \neq \zeta_{\omega}] = 0$

Integrating Random Variables (as discretely)

$(-)_{+}, (-)_{-} : \bar{\mathbb{R}}^n \rightarrow \mathbb{W}^n$  in Obs!

$$\xi_{+} := \max(\xi, 0) \quad \xi_{-} := \max(-\xi, 0)$$

$$\text{So: } \xi = \xi_{+} - \xi_{-}$$

$$\int : \mathcal{P}\Omega \times \mathbb{W}^n \longrightarrow \mathbb{W} \quad \int \text{ respects}$$

$$\int \lambda \xi := \int \lambda \xi_{+} - \int \lambda \xi_{-} \quad \xi = \xi_{+} + \xi_{-} \text{ a.s. equality}$$

$$\Rightarrow \int \lambda \xi = \int \xi_{+}$$

Example

$$\lambda : P\Omega \vdash \text{ASConverge}(\overline{\mathbb{R}})^{\Omega} : \mathcal{B}(\overline{\mathbb{R}}^{N \times \Omega})$$
$$:= \left\{ \vec{\zeta} \in \overline{\mathbb{R}}^{N \times \Omega} \mid \Pr_{w \sim \omega} [\lim_{n \rightarrow \infty} \zeta_n w \neq \perp] \right\}$$

So:

$$\lim^{\text{as}}_m : \overline{\mathbb{R}}^{N \times \Omega} \longrightarrow \overline{\mathbb{R}}^{\Omega} \quad \text{Dom } \lim^{\text{as}} := \text{ASConverge}(\overline{\mathbb{R}})^{\Omega}$$

$$\lim^{\text{as}} \vec{\zeta} := \lambda \omega. \limsup_{n \rightarrow \infty} f_n \omega$$

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

Then (monotone convergence):

let  $\sum \in \mathbb{W}^{N \times \mathbb{N}}$   $\lambda$ -a.s. monotone.

$$\xi = \lim_{n \rightarrow \infty} \xi_n \quad (\text{a.s.})$$



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

Lebesgue Space  $(\Omega \text{ prob. space}, \mathcal{P} \in [1, \infty))$

$$\rho: [1, \infty), \lambda: \mathbb{R}^2 \rightarrow L_{(\Omega, \mathcal{P})}^{\rho}: \mathcal{B}(\mathbb{R}^{\Omega})$$

$$:= \left\{ \xi \in \mathbb{R}^{\Omega} \mid \int |\xi|^{\rho} < \infty \right\} \hookrightarrow \mathbb{R}^{\Omega}$$

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

$$L \quad \rho \leq q \Rightarrow L_{\Omega}^{\rho} \supseteq L_{\Omega}^q$$

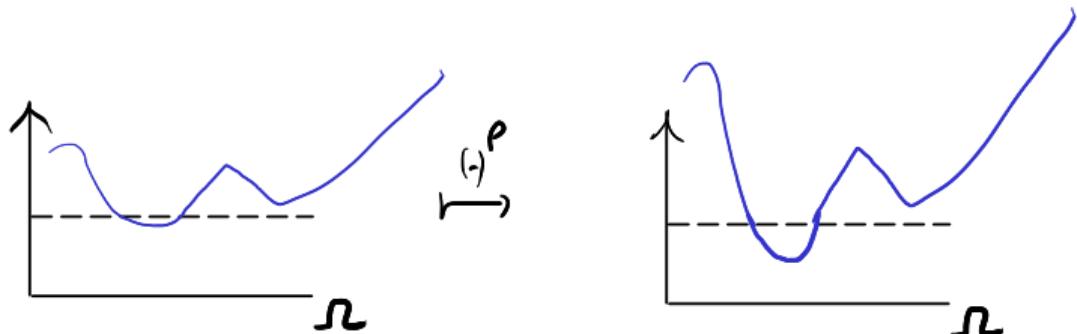
$\int^P$  semi norms

$$\|\cdot\| : \bigcup_{p,\lambda} L^p_{(2,\lambda)} \rightarrow \mathbb{R}_{\geq 0} \quad \|\xi\|_p := \sqrt[p]{\int \lambda |\xi|^p}$$

$L^2$  inner product

$$\langle \cdot, \cdot \rangle : \bigcup_{p,\lambda} L^p_{(2,\lambda)} \times L^p_{(2,\lambda)} \rightarrow \mathbb{R}$$

$$\langle \xi, \eta \rangle := \int \lambda \xi \eta$$



## Statistics

Expectation

$$\mathbb{E} : \bigcup_{\lambda} \mathcal{L}^1 \rightarrow \mathbb{R}$$

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

Covariance and Correlation

$$\text{Cov}, \text{Corr} : \bigcup_{\lambda} \mathcal{L}^2 \rightarrow \mathbb{R}$$

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

$$\text{Corr}(\xi, \zeta) := \frac{\langle \xi, \zeta \rangle}{\|\xi\|_2 \|\zeta\|_2} = \cos(\text{angle}(\xi, \zeta))$$

## Sequential limits

$$\rho: [1, \infty), \lambda: \mathbb{P}X \mapsto \text{Cauchy } L_{\lambda}^{\rho} := \left\{ \vec{\Xi} \mid \forall \varepsilon \in \mathbb{Q}^+ \exists N \in \mathbb{N} \forall m, n \geq N \quad \| \Xi_{n+m} - \Xi_n \|_{\rho} < \varepsilon \right\}$$

Thm:  $L_{\lambda}^{\rho}$  is Cauchy-complete

$\lim: \text{Cauchy } L^{\rho} \rightarrow L^{\rho}$  (convergence in mean)

Why?

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

## Example

Theorem (dominated convergence)

For  $\vec{z}_n, \vec{z} \in \mathbb{L}^1$  s.t.  $\vec{z}_n \leq \vec{z}$  a.s.:

$$1. \lim^{as} \vec{z}_n \rightarrow \vec{z} \in \mathbb{L}^1$$

$$2. \lim^1 \vec{z}_n = \lim^{as} \vec{z}_n \rightarrow \vec{z}$$

$$3. \lim_{n \rightarrow \infty} \int \lambda \vec{z}_n = \int \lambda \lim_{n \rightarrow \infty} \vec{z}_n$$

## Separability

Def:  $L^P$  separable: has countable dense subset

Fact: Separability is property of  $\lambda_2$ :

TFAE:

- $\exists P \ni L^P$  separable
- $\forall P \ni L^P$  separable

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

$$\vec{\beta} : \prod_{(\lambda, p) \in I} L_{(p, \lambda)}^p \quad \text{S.t.}$$

$$\left\{ \vec{\beta}_n^{(p)} \mid n \in \mathbb{N} \right\} \text{ dense in } L_{(p, \lambda)}^p$$

Prop. - Every Sbs  $S$  measurable separable in

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

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

$$\Rightarrow \exists \vec{\beta} \in \prod_{\lambda \in I} L_{(\lambda, \lambda)}^2 \quad \text{orthonormal system}$$

$$\begin{aligned} \langle \beta_n, \beta_m \rangle &= 0 \\ \|\beta_n\|_2 &= 1 \\ (\beta_n) &\text{ dense} \end{aligned}$$

Example

Let  $S \hookrightarrow L^2$  closed Vector subspace.

Orthogonal decomposition linear in fact.

$$\langle P, P^\perp \rangle : L^2 \rightarrow S \times S^\perp$$

When  $S$  is separable with orthonormal system  $\beta$

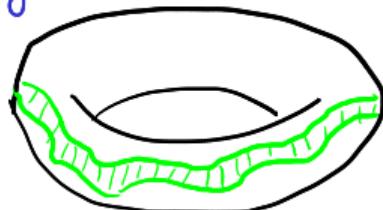
We have a measurable version of

$$\langle P, P^\perp \rangle : L^2 \rightarrow S \times S^\perp$$

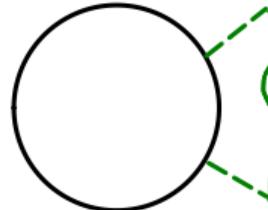
$$P\xi := \sum_{n=0}^{\infty} \langle \xi, \beta_n \rangle \beta_n \quad P^\perp := I_d - P.$$

# Kolmogorov's Conditional Expectation

ground truth space



(H) Sample space



H  
observation

$\xi$   
Statistic  
of interest

!

R

conditional expectation

$$\mathbb{E}[\xi | H = -]$$

Observed  
statistic

# Kolmogorov's Conditional Expectation

A Conditional expectation

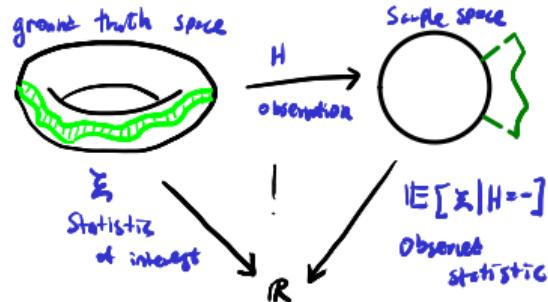
of  $\zeta \in \mathcal{L}_\Omega$  wrt

$H: \Omega \rightarrow \mathbb{H}$  is

$\zeta \in \mathcal{L}_{\mathbb{H}}$  s.t. for all  $A \in \mathcal{B}_{\mathbb{H}}$ :

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

where  $\mu := \lambda_H$

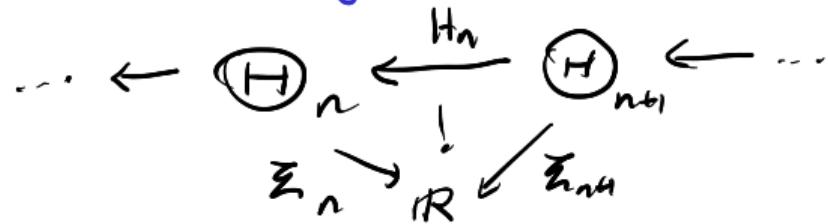


# Conditional expectations

1. unique a.s.

2. fundamental to modern Probability, e.g.:

a martingale



$$\text{st. } \xi_n = \mathbb{E}[\xi_{n+1} | H_n = -]$$

Then (Existence)

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

- When  $(\Omega, \lambda)$  is **separable**

$$\mathbb{E}[-|H=-]: \mathcal{L}_{(\Omega, \lambda)}^1 \rightarrow \mathcal{L}_{(\Theta, \mu)}^1$$

- When  $H$  is  $\mathcal{I}$ -measurably separable

$$\mathbb{E}[-|-=-]: \prod_{\substack{H \in \Theta \\ \lambda \in H^{-1}[\mathcal{I}]}} \mathcal{L}_{(\Omega, \lambda)}^1 \rightarrow \mathcal{L}_{(\Theta, \mu)}^1$$

## Plan:

- 1) Type-driven probability: discrete case ✓
  - 2) Borel sets & measurable spaces ✓
  - 3) Quasi Borel spaces ✓
  - 4) Type structure & standard Borel spaces ✓
  - 5) Integration & random variables ✓
- Lecture 1
- Lecture 2

please ask questions!

Scalable



← Course  
web  
page

## Discrete model

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

$\mathcal{D}X := \{\mu : X \rightarrow \mathbb{W} \mid \text{Supp } \mu \text{ countable}\}$

$\mathcal{P}X := \{\mu \in \mathcal{D}X \mid \underset{x}{\text{Ce}}[\mu] = 1\}$

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

$\phi \mu k := \lambda x. \sum_{m \in \Gamma} \mu^m \cdot k(m; x)$

## Full model

type: Qbs    W := [0, ∞]     $\mathcal{B}^X \cong \mathcal{B}^X$

$D_X := (\{\lambda_\alpha | \alpha: R \rightarrow X\}, \{\lambda_{r,\cdot} | \alpha: R \times R \rightarrow X\})$

$P_X := \{\mu \in \Omega_X | \underset{\mu}{\text{Ce}}[X] = 1\}$

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

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

Enough!

Lunch.