An introduction to categorical probability theory

Rob Cornish

Department of Statistics, University of Oxford

February 21, 2023

Motivation

It should be said: for someone trained in formal methods, the area of probability theory can be rather sloppy: everything is called 'P', types are hardly ever used, crucial ingredients (like distributions in expected values) are left implicit, basic notions (like conjugate prior) are introduced only via examples, calculation recipes and algorithms are regularly just given, without explanation, goal or justification, etc. This hurts, especially because there is so much beautiful mathematical structure around. (Jacobs [2019])

Classical probability theory

Start with an underlying probability space $(\Omega, \Sigma_\Omega, \mathbb{P})$

Model phenomena of interest using random variables (i.e. measurable functions) $X : \Omega \to \mathcal{X}$, i.e.

$$
\begin{tikzcd}
 & (\mathcal{X}, \Sigma_{\mathcal{X}}) \\
(\Omega, \Sigma_\Omega, \mathbb{P}) \arrow[ur, "X"] \arrow[dr, "Y"'] & \vdots \\
 & (\mathcal{Y}, \Sigma_{\mathcal{Y}})
\end{tikzcd}
$$

Can consider many distinct $\mathcal{X}$ and $\mathcal{Y}$, but Ω is fixed throughout

Usually study the joint or marginal behaviour of X, Y, etc.

Problems

This picture is quite complex

Many seemingly different components playing different roles:

- The underlying measurable space (Ω, Σ_Ω)
- The probability measure $\mathbb{P}$
- Random variables X, Y, etc.
- Joint and marginal distributions of X, Y, etc.

Also somewhat at odds with how we think intuitively:

- Distributions are secondary objects (cf. Bayesian statistics)
- Random variables are static (can't "sample" from them)
  - OK for fixed datasets, but often ill-suited for describing computation
- Kolmogorov-style conditioning is highly technical

Main point

Like a complex, low-level programming language, this inhibits abstraction and compositionality, and makes it difficult to say simple things

Categorical probability reorganises the existing theory in a way that makes reasoning about higher-level concepts easy and intuitive

Case study

PROBABILISTIC SYMMETRIES AND INVARIANT NEURAL NETWORKS

BY BENJAMIN BLOEM-REDDY[1] AND YEE WHYE TEH[2]

[1]*Department of Statistics*
University of British Columbia
benbr@stat.ubc.ca

[2]*Department of Statistics*
University of Oxford
y.w.teh@stats.ox.ac.uk

Background: group invariance

Often it is desirable for a function $f : \mathcal{X} \to \mathcal{Y}$ to be invariant to the action of a group $\mathcal{G}$

Example:

- $\mathcal{X}$ consists of sequences of profiles of subjects in an i.i.d. population
- $\mathcal{G}$ consists of permutations of the indices of these sequences
- $f : \mathcal{X} \to \mathcal{Y}$ makes some prediction about the population

Important question: for a given group $\mathcal{G}$, characterise the class of $f : \mathcal{X} \to \mathcal{Y}$ such that

$$f(g \cdot x) = f(x) \qquad \text{for all } g \in \mathcal{G} \text{ and } x \in \mathcal{X}$$

Probabilistic Symmetries

Bloem-Reddy and Teh [2020] consider a probabilistic version of this

Setup: $X : \Omega \to \mathcal{X}$ and $Y : \Omega \to \mathcal{Y}$ are random variables representing data and prediction respectively

Aim is to characterise when Y is conditionally $\mathcal{G}$-invariant in the sense that

$$\mathbb{P}(Y \in B \mid X \in A) = \mathbb{P}(Y \in B \mid X \in g \cdot A)$$

for all $g \in \mathcal{G}$, $A \in \Sigma_{\mathcal{X}}$ with $\mathbb{P}(X \in A) > 0$, and $B \in \Sigma_{\mathcal{Y}}$

Main result (on invariance)

THEOREM 7. *Let X and Y be random elements of Borel spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively, and $\mathcal{G}$ a compact group acting measurably on $\mathcal{X}$. Assume that P_X is $\mathcal{G}$-invariant, and pick a maximal invariant $M : \mathcal{X} \to \mathcal{S}$, with $\mathcal{S}$ another Borel space. Then $P_{Y|X}$ is $\mathcal{G}$-invariant if and only if there exists a measurable function $f : [0,1] \times \mathcal{S} \to \mathcal{Y}$ such that*

$$(X, Y) \overset{\text{a.s.}}{=} \big(X, f(\eta, M(X))\big) \quad \textit{with } \eta \sim \mathrm{Unif}[0,1] \textit{ and } \eta \perp\!\!\!\perp X\ . \tag{14}$$

Here a maximal invariant is any measurable function M such that

$$M(x) = M(x') \Leftrightarrow x = g \cdot x' \text{ for some } g \in \mathcal{G}$$

Proof

The proof of this is complex and uses highly technical ideas from advanced probability theory, e.g.

- Measurable cross section
- Normalised Haar measure
- Orbit law
- Conditional independence (of X and Y given $M(X)$)

Also only applies when $\mathcal{G}$ is compact and X has a $\mathcal{G}$-invariant marginal

Thoughts

Why is this so hard to show? (E.g. compare deterministic case)

Is it optimal to model a neural network in terms of random variables (X, Y)? And why must $\mathrm{Law}[X]$ be $\mathcal{G}$-invariant?

With the tools of categorical probability, we can not only generalise this result, but we can prove it in a way that maps directly onto our intuitions

Categorical probability theory

Category theory

Definition

A category consists of a collection of objects and a collection of arrows

Each arrow f has a source and target object, denoted $f : \mathcal{X} \to \mathcal{Y}$

There is a composition operation $\circ$ on arrows such that

$$
\begin{aligned}
g \circ f : \mathcal{X} \to \mathcal{Z} \qquad & \text{whenever } f : \mathcal{X} \to \mathcal{Y} \text{ and } g : \mathcal{Y} \to \mathcal{Z} \\
h \circ (g \circ f) = (h \circ g) \circ f \qquad & \text{when } f, g, h \text{ are appropriately typed}
\end{aligned}
$$

For every object $\mathcal{X}$ there is an identity arrow $\mathrm{id}_{\mathcal{X}} : \mathcal{X} \to \mathcal{X}$ satisfying

$$
\begin{aligned}
f \circ \mathrm{id}_{\mathcal{X}} = f \qquad & \text{whenever } f : \mathcal{X} \to \mathcal{Y} \\
\mathrm{id}_{\mathcal{X}} \circ g = g \qquad & \text{whenever } g : \mathcal{Z} \to \mathcal{X}
\end{aligned}
$$

Philosophy: study structural properties extrinsically in terms of arrows

Examples

Categories are everywhere:

- **Set**, the category of sets and functions
- **Meas**, the category of measurable spaces and measurable functions
- **Stoch**, the category of measurable spaces and Markov kernels
- Even $\mathcal{G}$ can be viewed as a category (with a single object, and inverses)

Functors

The only other definition we will need is as follows:

Definition

Given categories C and D, a functor $F : \mathsf{C} \to \mathsf{D}$ assigns:

- To each object $\mathcal{X}$ in C, an object $F\mathcal{X}$ in D
- To each arrow $f : \mathcal{X} \to \mathcal{Y}$ in C, an arrow $Ff : F\mathcal{X} \to F\mathcal{Y}$

Moreover:

- $F\mathrm{id}_{\mathcal{X}} = \mathrm{id}_{F\mathcal{X}}$ for all objects $\mathcal{X}$ in C
- $F(g \circ f) = Fg \circ Ff$ for all suitably typed arrows f and g in C

The Giry Functor [Giry, 1982]

Denote by $P\mathcal{X}$ the set of probability measures on $\mathcal{X}$ (where $\Sigma_{\mathcal{X}}$ implicit)

It turns out P can be thought of as a functor **Meas** $\to$ **Meas**:

- Equip $P\mathcal{X}$ with the (initial) σ-algebra generated by the functions:

$$\begin{aligned} \mathrm{eval}_A : P\mathcal{X} &\to [0,1] \qquad \text{where } A \in \Sigma_{\mathcal{X}} \\ p &\mapsto p(A) \end{aligned}$$

- For measurable $f : \mathcal{X} \to \mathcal{Y}$, define Pf by the pushforward, i.e.

$$\begin{aligned} Pf : P\mathcal{X} &\to P\mathcal{Y} \\ Pf(p) &\mapsto f\#p \end{aligned}$$

- Check functor axioms hold

This reduces already the complexity of our original picture (since $\mathbb{P} \in P\Omega$)

Markov kernels

Consider a measurable function $k : \mathcal{X} \to P\mathcal{Y}$

By definition of P:

- $k(x)(-)$ is a probability measure for all $x \in \mathcal{X}$
- $k(-)(B) = \mathrm{eval}_B \circ k$ is measurable for all $B \in \Sigma_{\mathcal{Y}}$

Hence k is a Markov kernel: can think of as $k : \mathcal{X} \times \Sigma_{\mathcal{Y}} \to [0,1]$ such that

- $k(x, -)$ is a probability measure for all $x \in \mathcal{X}$
- $k(-, B)$ is measurable for all $B \in \Sigma_{\mathcal{Y}}$

(Precisely: write $k : \mathcal{X} \to P\mathcal{Y}$ as $k : \mathcal{X} \to (\Sigma_{\mathcal{Y}} \mapsto [0,1])$ and uncurry)

Giry monad

We can consider Markov kernels to be generalised measurable functions:

- Every "normal" $f : \mathcal{X} \to \mathcal{Y}$ can be canonically identified with $\delta_{\mathcal{Y}} \circ f : \mathcal{X} \to P\mathcal{Y}$, where

$$\begin{aligned} \delta_{\mathcal{Y}} : \mathcal{Y} &\to P\mathcal{Y} \\ y &\mapsto \mathrm{Dirac}(y) \end{aligned}$$

- Every "generalised generalised" function $k : \mathcal{X} \to PP\mathcal{Y}$ can be canonically identified with $E_{\mathcal{Y}} \circ k : \mathcal{X} \to P\mathcal{Y}$, where

$$\begin{aligned} E_{\mathcal{Y}} : PP\mathcal{Y} &\to P\mathcal{Y} \\ p &\mapsto \int_{P\mathcal{Y}} p(\mathrm{d}q)\, q(-) \end{aligned}$$

P, $\delta_{\mathcal{Y}}$, and $E_{\mathcal{Y}}$ moreover satisfy coherence conditions and so give rise to a monad structure on **Meas**

Kleisli composition

The monad structure on **Meas** yields a canonical notion of composition of generalised functions (i.e. Markov kernels)

Given $k : \mathcal{X} \to P\mathcal{Y}$ and $\ell : \mathcal{Y} \to P\mathcal{Z}$, define $\ell \circ_{\mathsf{kl}} k : \mathcal{X} \to P\mathcal{Z}$ via the following composition:

$$\mathcal{X} \xrightarrow{k} P\mathcal{Y} \xrightarrow{P\ell} PP\mathcal{Z} \xrightarrow{E_{\mathcal{Z}}} P\mathcal{Z}$$

Can show this is the usual Chapman-Kolmogorov equation:

$$(\ell \circ_{\mathsf{kl}} k)(x)(A) = \int_{\mathcal{Y}} k(x)(\mathrm{d}y)\, \ell(y)(A) \qquad \text{where } A \in \Sigma_{\mathcal{Z}}$$

Dirac maps $\delta_{\mathcal{X}} : \mathcal{X} \to P\mathcal{X}, x \mapsto \mathrm{Dirac}(x)$ behave like identities

Kleisli category

This gives rise to the Kleisli category of **Meas**, known as **Stoch**:

	Meas	**Stoch**
Objects	Measurable spaces	Measurable spaces
Arrows	Measurable functions	Markov kernels
Composition	Composition of functions	Chapman-Kolmogorov
Identities	Identity functions	Dirac maps

Kleisli adjunction

We have a bijective correspondence (in fact an adjunction):

$$\text{Markov kernels } \mathcal{X} \to \mathcal{Y} \qquad \leftrightsquigarrow \qquad \text{Measurable functions } \mathcal{X} \to P\mathcal{Y}$$

We saw that identity kernels correspond to Dirac maps, i.e.

$$\mathrm{id}_{\mathcal{X}} : \mathcal{X} \to \mathcal{X} \qquad \leftrightsquigarrow \qquad \delta_{\mathcal{X}} : \mathcal{X} \to P\mathcal{X}$$

Interesting question: what Markov kernel corresponds to the measurable function $\mathrm{id}_{P\mathcal{Y}} : P\mathcal{Y} \to P\mathcal{Y}$?

$$\mathsf{samp}_{\mathcal{Y}} : P\mathcal{Y} \to \mathcal{Y} \qquad \leftrightsquigarrow \qquad \mathrm{id}_{P\mathcal{Y}} : P\mathcal{Y} \to P\mathcal{Y}$$

Here $\mathsf{samp}_{\mathcal{Y}}(p)(B) = p(B)$, i.e. $\mathsf{samp}_{\mathcal{Y}}$ draws a sample from its input

New picture

Stoch unifies and generalises the elements in our original picture:

$$\begin{tikzcd}
 & & (\mathcal{X}, \Sigma_{\mathcal{X}}) \\
(\{*\}, \Sigma_{\{*\}}) \arrow[r, "\mathbb{P}"] & (\Omega, \Sigma_{\Omega}) \arrow[ur, "\delta_{\mathcal{X}} \circ X"] \arrow[r, "\delta_{\mathcal{Y}} \circ Y"] \arrow[d, "\delta_{P\mathcal{Z}} \circ k"'] \arrow[dr, "k"] & (\mathcal{Y}, \Sigma_{\mathcal{Y}}) \\
 & (P\mathcal{Z}, \Sigma_{\mathcal{Z}}) \arrow[r, "\mathsf{samp}_{\mathcal{Z}}"'] & (\mathcal{Z}, \Sigma_{\mathcal{Z}})
\end{tikzcd}$$

Although to some extent $(\Omega, \Sigma_{\Omega})$ is redundant now ...

Recap

We have described two categories of interest for probability theory:

- **Meas**, i.e. measurable spaces and measurable functions
- **Stoch**, i.e. measurable spaces and Markov kernels

We have a functor $P : \mathbf{Meas} \to \mathbf{Meas}$ mapping measurable spaces to their space of probability measures

Stoch arises as the Kleisli category of the Giry monad (or can be shown directly to satisfy the axioms of a category)

Nothing surprising, but structural properties like these are powerful

Return to case study

Recap

THEOREM 7. *Let X and Y be random elements of Borel spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively, and $\mathcal{G}$ a compact group acting measurably on $\mathcal{X}$. Assume that P_X is $\mathcal{G}$-invariant, and pick a maximal invariant $M : \mathcal{X} \to \mathcal{S}$, with $\mathcal{S}$ another Borel space. Then $P_{Y|X}$ is $\mathcal{G}$-invariant if and only if there exists a measurable function $f : [0,1] \times \mathcal{S} \to \mathcal{Y}$ such that*

$$(X, Y) \overset{\text{a.s.}}{=} \big(X, f(\eta, M(X))\big) \quad \textit{with } \eta \sim \mathrm{Unif}[0,1] \textit{ and } \eta \perp\!\!\!\perp X \ . \tag{14}$$

Conditional distributions/disintegrations

Proposition

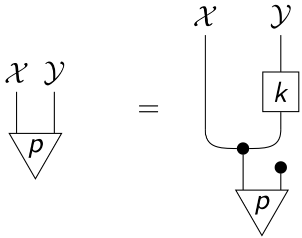
If $\mathcal{Y}$ is standard Borel, then for any distribution p on $\mathcal{X} \otimes \mathcal{Y}$, there exists a Markov kernel $k : \mathcal{X} \to \mathcal{Y}$ such that

$$p(A \times B) = \int_A \mathrm{proj}_{\mathcal{X}}(p)(\mathrm{d}x)\, k(x)(B) \qquad \textit{for all } A \in \Sigma_{\mathcal{X}} \textit{ and } B \in \Sigma_{\mathcal{Y}}.$$

It is convenient to have a graphical way to denote this. Standard commutative diagrams get complex, but string diagrams work:

$$\begin{tikzcd}
\{*\} \arrow[rr, "p"] \arrow[d, "p"'] & & \mathcal{X} \otimes \mathcal{Y} \\
\mathcal{X} \otimes \mathcal{Y} \arrow[r, "\mathrm{proj}_{\mathcal{X}}"] & \mathcal{X} \arrow[r, "\mathrm{copy}_{\mathcal{X}}"] & \mathcal{X} \otimes \mathcal{X} \arrow[u, "\mathrm{id}_{\mathcal{X}} \otimes k"']
\end{tikzcd} \qquad \Leftrightarrow$$

(Compositions along both paths are equal)



(Read from bottom to top)

Invariance under an equivalence relation

Suppose $\sim$ is an arbitrary equivalence relation on $\mathcal{X}$

Definition

A distribution p on $\mathcal{X} \otimes \mathcal{Y}$ is conditionally $\sim$-invariant if p admits a disintegration $k : \mathcal{X} \to \mathcal{Y}$ that is $\sim$-invariant, i.e. $k(x) = k(x')$ if $x \sim x'$.

For $p = \mathrm{Law}[X, Y]$, equivalent to conditional invariance in sense of Bloem-Reddy and Teh [2020] under their setup, i.e. $\mathcal{G}$ is compact, $\mathrm{Law}[X]$ is $\mathcal{G}$-invariant, $\mathcal{Y}$ standard Borel, and

$$x \sim x' \Leftrightarrow x = g \cdot x' \text{ for some } g \in \mathcal{G},$$

Makes sense more generally – could even start with k as the definition of a (probabilistic) neural network

Quotient spaces

Given any measurable space $\mathcal{X}$ and an equivalence relation $\sim$ on $\mathcal{X}$, we can form the quotient space $\mathcal{X}/{\sim}$ of equivalence classes under $\sim$

The σ-algebra is final with respect to the quotient map $q : \mathcal{X} \to \mathcal{X}/{\sim}$

Explicitly, $\Sigma_{\mathcal{X}/\sim} \coloneqq \{B \subseteq \mathcal{X}/{\sim} \mid q^{-1}(B) \in \Sigma_{\mathcal{X}}\}$.

Universal property of the quotient

Proposition

A measurable function $g : \mathcal{X} \to \mathcal{Z}$ is $\sim$-invariant iff there exists a (necessarily unique) measurable function $\tilde{g} : \mathcal{X}/\sim \to \mathcal{Z}$ such that $\tilde{g} \circ q = g$, i.e. the following diagram commutes:

$$\begin{tikzcd} \mathcal{X} \arrow[r, "g"] \arrow[d, "q"'] & \mathcal{Z} \\ \mathcal{X}/\sim \arrow[ur, dashed, "\tilde{g}"'] & \end{tikzcd}$$

Requires proof, but can do so via only elementary definitions

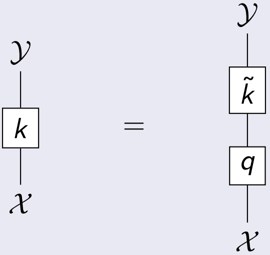
A very natural result in the context of category theory

Invariant kernels via the quotient

Now take $\mathcal{Z} = P\mathcal{Y}$ and interpret within **Stoch**

Corollary

A Markov kernel $k : \mathcal{X} \to \mathcal{Y}$ is $\sim$-invariant iff there exists a Markov kernel $\tilde{k} : \mathcal{X}/\sim \to \mathcal{Y}$ with

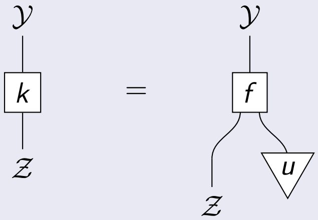


(Note that we are identifying q with its lifted version $\delta_{\mathcal{X}/\sim} \circ q$)

Noise outsourcing

Proposition

For any Markov kernel $k : \mathcal{Z} \to \mathcal{Y}$ with $\mathcal{Y}$ standard Borel, there exists a measurable function $f : \mathcal{Z} \otimes [0,1] \to \mathcal{Y}$ such that



where $u = \mathrm{Uniform}(0,1)$.

Standard result (e.g. Lemma 3.22 of Kallenberg [2002])

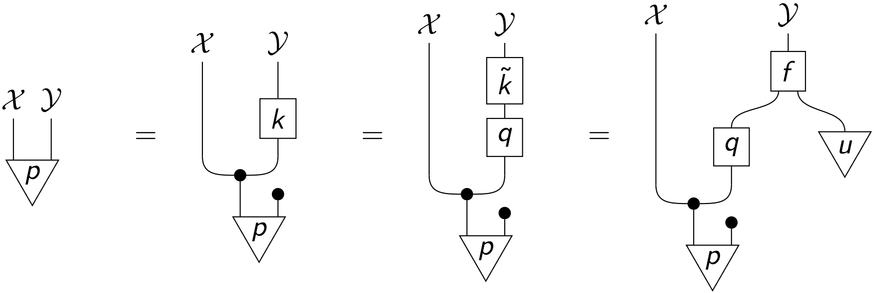
Combining these results

Proposition

If $\mathcal{Y}$ is standard Borel, then $\mathrm{Law}[X, Y]$ is conditionally $\sim$-invariant iff there exists a measurable function $f : \mathcal{X}/{\sim} \otimes [0, 1] \to \mathcal{Y}$ such that

$$(X, Y) \overset{\mathrm{d}}{=} (X, f(q(X), \eta)) \qquad \textit{where } \eta \sim \mathrm{Uniform}(0, 1),\ \eta \perp\!\!\!\perp X$$

Proof: writing $p := \mathrm{Law}[X, Y]$, conditional $\sim$-invariance implies



Conversely, right-hand side is conditionally $\sim$-invariant since q is.

Comparison with original result

THEOREM 7. *Let X and Y be random elements of Borel spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively, and $\mathcal{G}$ a compact group acting measurably on $\mathcal{X}$. Assume that P_X is $\mathcal{G}$-invariant, and pick a maximal invariant $M : \mathcal{X} \to \mathcal{S}$, with $\mathcal{S}$ another Borel space. Then $P_{Y|X}$ is $\mathcal{G}$-invariant if and only if there exists a measurable function $f : [0,1] \times \mathcal{S} \to \mathcal{Y}$ such that*

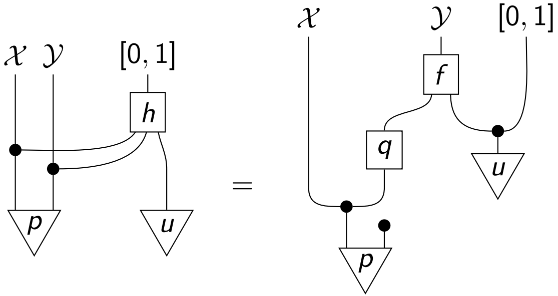
$$(X, Y) \overset{\text{a.s.}}{=} \big(X, f(\eta, M(X))\big) \quad \textit{with } \eta \sim \text{Unif}[0,1] \textit{ and } \eta \perp\!\!\!\perp X \,. \tag{14}$$

Not quite done:

$$(X, Y) \overset{\text{d}}{=} (X, f(q(X), \eta)) \qquad \not\Rightarrow \qquad Y \overset{\text{a.s.}}{=} f(q(X), \eta)$$

Completing the proof

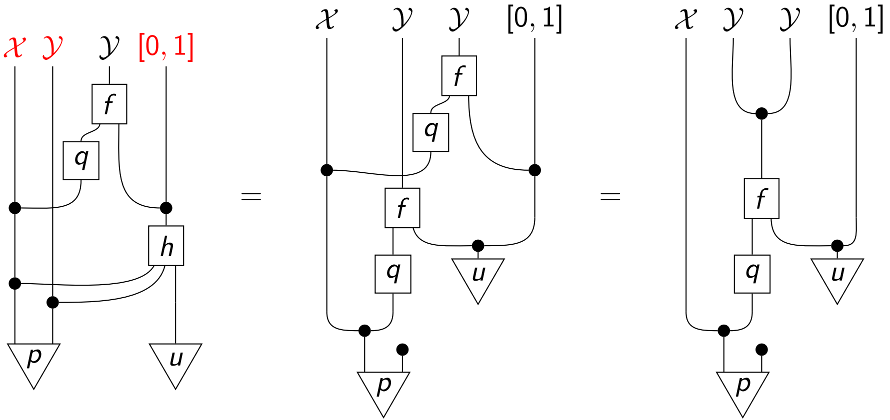
Choose $h : \mathcal{X} \otimes \mathcal{Y} \otimes [0,1] \to [0,1]$ such that



Existence of h follows by disintegrating right-hand side along $\mathcal{X} \times \mathcal{Y}$ and then applying noise outsourcing result

Completing the proof

Now affix the same (q, f) construction to both sides:



$\Rightarrow$ If $\xi \sim \mathrm{Uniform}(0,1)$ with $\xi \perp\!\!\!\perp (X, Y)$, then letting $\eta := h(X, Y, \xi)$, have $(X, Y, f(q(X), \eta), \eta) \overset{\mathrm{d}}{=} (X, f(q(X), \xi), f(q(X), \xi), \xi)$

$\Rightarrow Y \overset{\text{a.s.}}{=} f(q(X), \eta)$ and $\eta \overset{\mathrm{d}}{=} \xi \sim \mathrm{Uniform}(0,1)$ with $\eta \perp\!\!\!\perp X$

Combining these results

THEOREM 7. *Let X and Y be random elements of Borel spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively, and $\mathcal{G}$ a compact group acting measurably on $\mathcal{X}$. Assume that P_X is $\mathcal{G}$-invariant, and pick a maximal invariant $M : \mathcal{X} \to \mathcal{S}$, with $\mathcal{S}$ another Borel space. Then $P_{Y|X}$ is $\mathcal{G}$-invariant if and only if there exists a measurable function $f : [0,1] \times \mathcal{S} \to \mathcal{Y}$ such that*

$$(X, Y) \stackrel{\text{a.s.}}{=} \big(X, f(\eta, M(X))\big) \quad \textit{with } \eta \sim \mathrm{Unif}[0,1] \textit{ and } \eta \perp\!\!\!\perp X \, . \tag{14}$$

Theorem (Our version)

If $\mathcal{Y}$ is Borel, then $\mathrm{Law}[X, Y]$ is conditionally $\sim$-invariant iff there exists a measurable function $f : \mathcal{X}/\!\sim \otimes [0,1] \to \mathcal{Y}$ such that

$$(X, Y) \stackrel{\text{a.s.}}{=} (X, f(q(X), \eta)) \qquad \textit{where } \eta \sim \mathrm{Uniform}(0,1) \textit{ with } \eta \perp\!\!\!\perp X$$

(More precisely, both statements should refer to the existence of an extension of the underlying probability space that admits suitable choices of η and f)

Possibly better to express entirely via Markov kernels

Conclusion

Summary and Outlook

Categorical probability offers a high-level perspective on the classical theory that makes abstraction easier and helps theory follow intuition

The outlook is very positive:

- Lots of activity in categorical probability, e.g. Perrone [2018], Cho and Jacobs [2019], Jacobs [2019], Fritz [2020]
- Category theory has been hugely successful elsewhere, e.g. pure maths, computer science, quantum mechanics

Also not difficult to learn! Just a new way of thinking about things you already know...

References I

Bart Jacobs. Structured probabilistic reasoning. 2019.

Benjamin Bloem-Reddy and Yee Whye Teh. Probabilistic symmetries and invariant neural networks. *J. Mach. Learn. Res.*, 21:90–1, 2020.

Michele Giry. A categorical approach to probability theory. In *Categorical aspects of topology and analysis*, pages 68–85. Springer, 1982.

Olav Kallenberg. *Foundations of modern probability*. Springer, 2 edition, 2002.

Paolo Perrone. Categorical probability and stochastic dominance in metric spaces, 2018.

Kenta Cho and Bart Jacobs. Disintegration and bayesian inversion via string diagrams. *Mathematical Structures in Computer Science*, 29(7):938–971, mar 2019. doi: 10.1017/s0960129518000488. URL `https://doi.org/10.1017%2Fs0960129518000488`.

Tobias Fritz. A synthetic approach to markov kernels, conditional independence and theorems on sufficient statistics. *Advances in Mathematics*, 370:107239, aug 2020. doi: 10.1016/j.aim.2020.107239. URL `https://doi.org/10.1016%2Fj.aim.2020.107239`.