Quasi-Measurable Spaces

A Convenient Foundation of Probability Theory

Patrick Forré

AI4Science Lab, AMLab, Informatics Institute, University of Amsterdam, The Netherlands

The Category of Measurable Spaces

- Let $\mathscr{X}$ be a set. A **σ-algebra** on $\mathscr{X}$ is a set of subsets $\mathscr{B} \subseteq 2^{\mathscr{X}}$ such that:
  - $\varnothing \in \mathscr{B}$,
  - $A_n \in \mathscr{B}, n \in \mathbb{N} \implies \bigcup_{n \in \mathbb{N}} A_n \in \mathscr{B}$
  - $A \in \mathscr{B} \implies \mathscr{X} \backslash A \in \mathscr{B}$
- A tuple $(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$ of a set $\mathscr{X}$ and a σ-algebra $\mathscr{B}_{\mathscr{X}}$ is called **measurable space**.
- A map $f: \mathscr{X} \to \mathscr{Y}$ between measurable spaces $(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$ and $(\mathscr{Y}, \mathscr{B}_{\mathscr{Y}})$ is called a **measurable map** if: $B \in \mathscr{B}_{\mathscr{Y}} \implies f^{-1}(B) \in \mathscr{B}_{\mathscr{X}}$.
  - Note that the compositions of two measurable maps is a measurable map.
- Meas denotes the **category of measurable spaces** and **measurable maps**.

Kolmogorov's approach to Probability Theory (1933)

- **Kolmogorov Axioms**:
    - A probability distribution is just a normalized measure.

- Probability Theory can thus be viewed as a sub-field of Measure Theory.

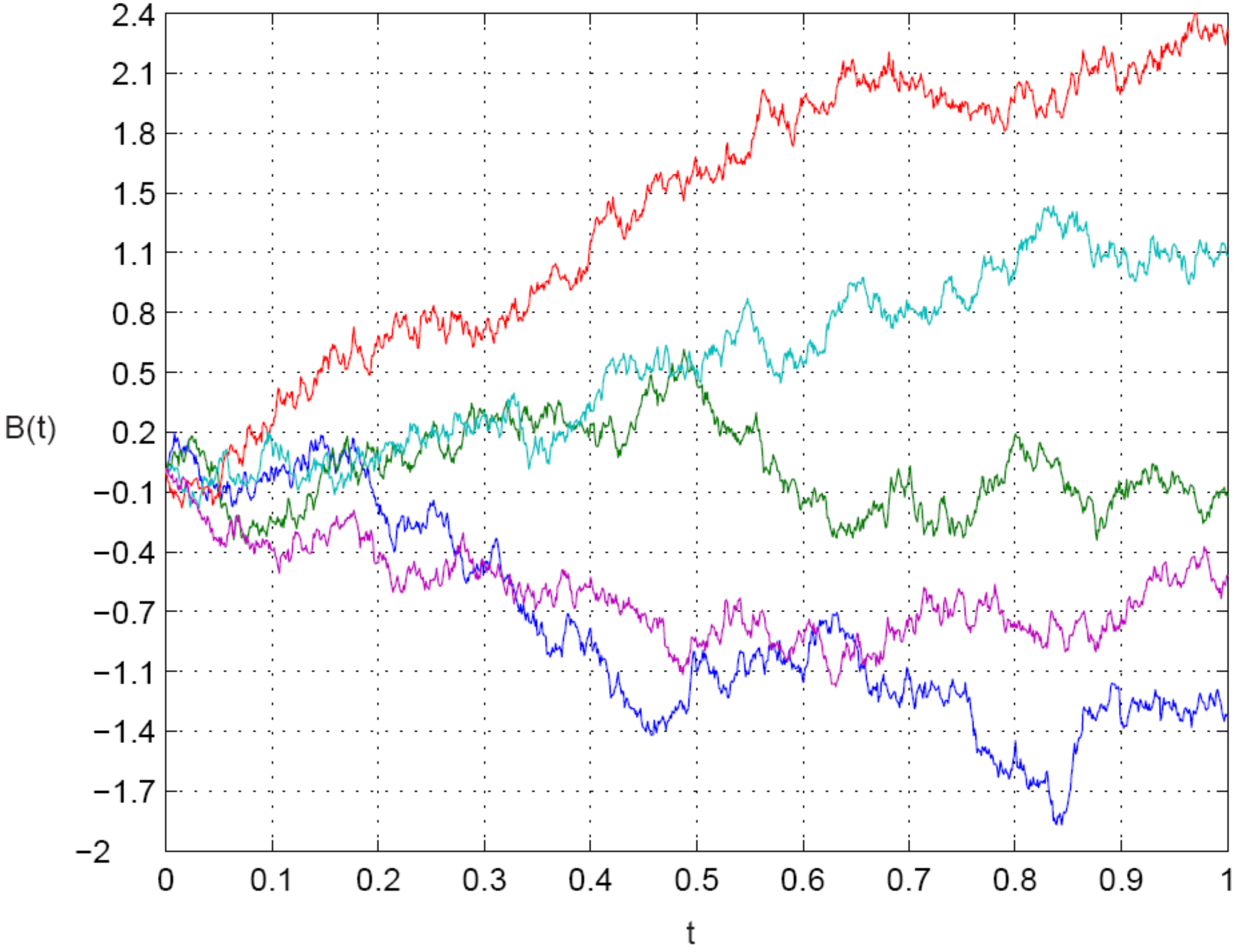
- Andrei Kolmogoroff. *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Ergebnisse der Mathematik und Ihrer Grenzgebiete. 1. Folge, Nr. 2, Springer (1933).

How to formalize Random Variables?

- **Sample space** is a measurable space: $\left(\Omega, \mathscr{B}_\Omega\right)$
  - where $\mathscr{B}_\Omega$ is the σ-algebra/set of admissible outcome events on Ω.
- **State space** is a measurable space: $\left(\mathscr{X}, \mathscr{B}_\mathscr{X}\right)$
  - where $\mathscr{B}_\mathscr{X}$ is another **σ-algebra**/set of admissible events on $\mathscr{X}$.
- Admissible **random variables** are all measurable maps:
  - $X \in \mathrm{Meas}\left((\Omega, \mathscr{B}_\Omega), (\mathscr{X}, \mathscr{B}_\mathscr{X})\right)$
- For fixed **probability measure** P on $(\Omega, \mathscr{B}_\Omega)$ the distribution of X is:
  - push-forward probability measure: $X_* P$ on $\mathscr{B}_\mathscr{X}$ (also written as: $P(X)$).

Motivation
- Problems with that Approach

Problem 1
- Stochastic Processes



Definitions - Stochastic Process

- Let $(\Omega, \mathscr{B}_\Omega)$ be the sample space, $(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$ the state space, $(\mathscr{T}, \mathscr{B}_{\mathscr{T}})$ the time space, e.g. $\mathscr{T} = \mathbb{N}$ or $\mathscr{T} = \mathbb{R}_{\geq 0}$.

- A *stochastic process* is a measurable map:

    - $X = (X_t)_{t \in \mathscr{T}} : \ \Omega \to \prod_{t \in \mathscr{T}} \mathscr{X}, \qquad \omega \mapsto (X_t(\omega))_{t \in \mathscr{T}},$

        - i.e. a random tuple indexed by time

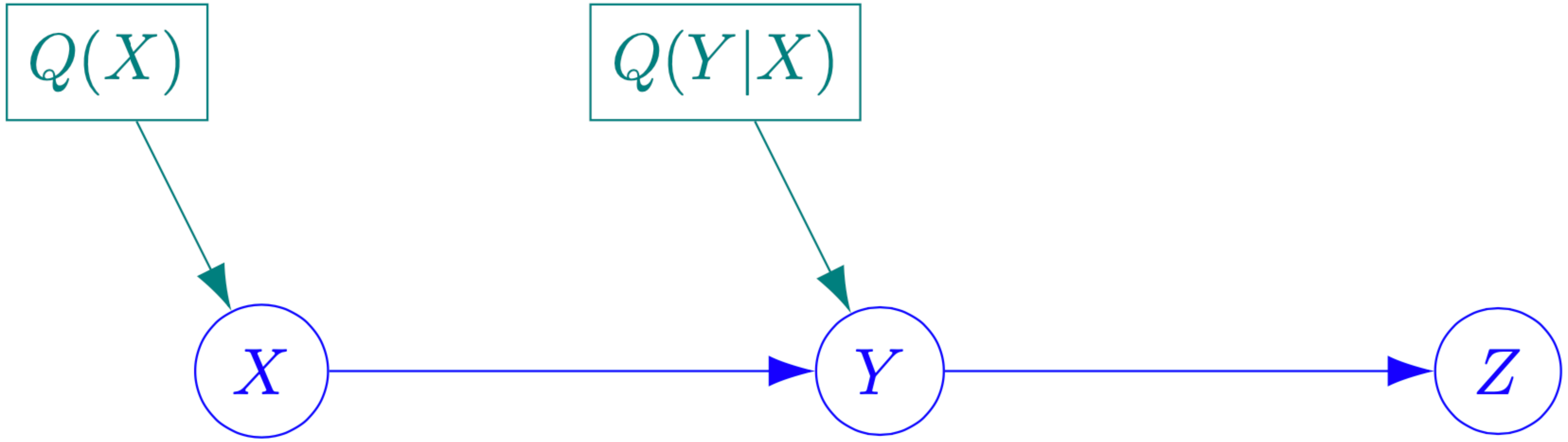
- A *random time* is a measurable map:

    - $T : \ \Omega \to \mathscr{T}, \qquad \omega \mapsto T(\omega).$

Problem

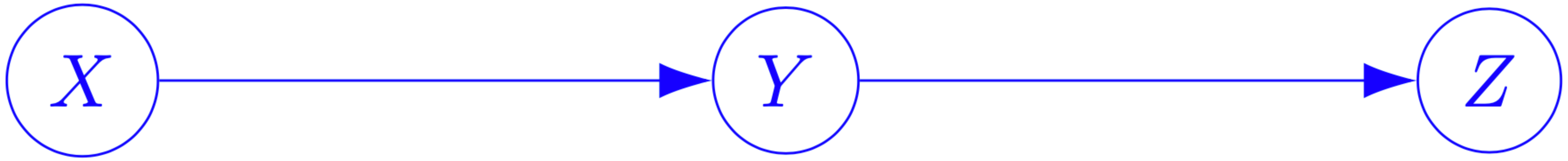
- The *stopped process*:
  - $X_T : \Omega \to \mathcal{X}, \qquad \omega \mapsto X_{T(\omega)}(\omega)$
  - is **not** measurable anymore in general (or only guaranteed under additional, typically topological, assumptions).

Problem 2
- Probabilistic Graphical Models



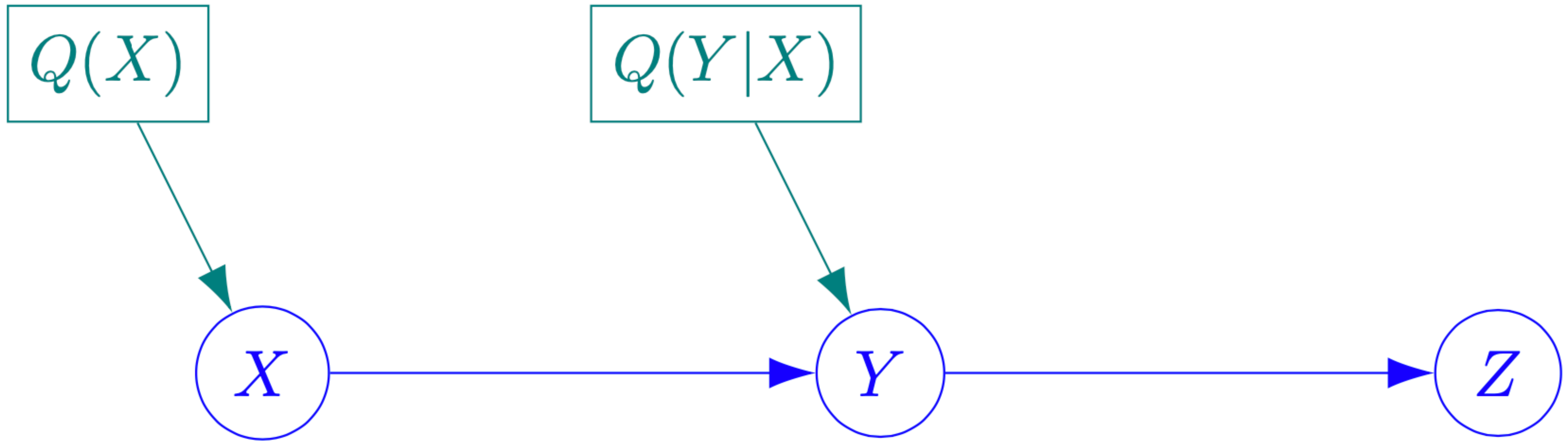
Conditional Independence in Probabilistic Graphical Models

- Consider a **Markov chain**:



- We have:
  - factorization: $P(X, Y, Z) = P(Z|Y) \otimes P(Y|X) \otimes P(X)$
  - tells us that Z is only dependent on Y, and, independent of X when conditioned on Y, but then also of the choice of $P(Y|X)$ and $P(X)$.
- We want to be able to:
  - formalize conditional independence: $Z \perp\!\!\!\perp X, Q(Y|X), Q(X) \,|\, Y$
    - including non-random variables $Q(X)$ and $Q(Y|X)$
  - read this off a graph via d-separation (or similar).

Including Non-Random Variables



- $Q(Y|X)$ is non-random and takes values in $\mathscr{L} := \mathrm{Meas}\left(\mathscr{X}, \mathscr{P}(\mathscr{Y})\right)$
  - Then Y is determined by the new mechanism:
    - $\mathscr{L} \times \mathscr{X} \to \mathscr{P}(\mathscr{Y}), \quad \left(Q(Y|X), x\right) \mapsto Q(Y|X = x).$
  - similarly for X.

Problem

- For the map:
  - $\mathscr{L} \times \mathscr{X} \to \mathscr{P}(\mathscr{Y}), \quad \big(Q(Y|X), x\big) \mapsto Q(Y|X=x).$
- to become measurable we need a well-behaved σ-algebra on:
  - $\mathscr{L} := \mathrm{Meas}\left(\mathscr{X}, \mathscr{P}(\mathscr{Y})\right).$

- This is generally impossible!

Random Functions do not exist in Meas

- <u>Theorem</u> (Aumann, 1961):
  - There is **no** σ-algebra $\mathscr{B}_{\mathscr{L}}$ on $\mathscr{L} := \mathrm{Meas}(\mathbb{R}, \mathbb{R})$ such that the evaluation map is measurable:
    - $\mathrm{ev} : \mathscr{L} \times \mathbb{R} \to \mathbb{R}, \quad (f, x) \mapsto f(x),$
  - where $\mathbb{R}$ carries the Borel-σ-algebra and $\mathscr{L}$ is the space of all measurable maps from $\mathbb{R}$ to $\mathbb{R}$, and the product carries the product-σ-algebra.
- So there is no well-behaved way to define a probability distribution over all measurable functions in a fully non-parametric way.

- Robert J. Aumann. *Borel structures for function spaces*. Illinois Journal of Mathematics 5.4 (1961): 614-630.

Problem 3
- Probabilistic Programs

```
def prog_prob_prog(a):
    return lambda m,s: [Z:=np.random.uniform(), a*m+s*Z][-1]
```

```
for n in range(5):
    print(prog_prob_prog(a=n)(m=5,s=2))
```

```
1.8106662116099772
6.762509413168864
10.365457994333775
15.884402920590935
21.48676872656254
```

Definition - Probabilistic Programs

- A *probabilistic program* with inputs $x \in \mathcal{X}$ and outputs $z \in \mathcal{Z}$ is a measurable map:
    - measurable map: $K: \; \mathcal{X} \to \mathcal{P}(\mathcal{Z})$
        - i.e. a Markov kernel.
- The set of probabilistic programs: $\mathrm{Meas}\left(\mathcal{X}, \mathcal{P}(\mathcal{Z})\right).$

Problem

- *Curry / Uncurry* probabilistic programs would translate to isomorphism:
  - $\mathrm{Meas}\left(\mathscr{X} \times \mathscr{Y}, \mathscr{P}\left(\mathscr{Z}\right)\right) \cong \mathrm{Meas}\left(\mathscr{X}, \mathrm{Meas}\left(\mathscr{Y}, \mathscr{P}(\mathscr{Z})\right)\right)$
    - $K \mapsto \tilde{K}$ with $\tilde{K}(x)(y) = K(x, y)$

- There does **not exists** any σ-algebra on $\mathrm{Meas}\left(\mathscr{Y}, \mathscr{P}(\mathscr{Z})\right)$ that would make that work (similar argument as before).

- Alonzo Church. *A formulation of the simple theory of types*. The Journal of Symbolic Logic 5.2 (1940): 56-68.

Problem 4
- Causal Assumptions

Causal Inference - Estimating Treatment Effects

- For estimating treatment effect, in the typical case, we have the variables:
  - X = observed treatment variable,
  - Y = observed outcome,
  - Y_x = potential outcome variable under (forced) treatment $X = x$,
  - Z = all other relevant features of the patient.
- Estimation is not possible without further assumptions.
- Typical assumptions made are:
  - **Strong Ignorability**: $X \perp\!\!\!\perp (Y_x)_{x \in \mathscr{X}} \,|\, Z$,
  - **Consistency**: $Y = Y_X$ a.s.

Problem

- Here, $(Y_x)_{x \in \mathcal{X}}$ is used as a vector of random variables from which we can pick components: $\left(\tilde{x}, (Y_x)_{x \in \mathcal{X}}\right) \mapsto Y_{\tilde{x}}$.
    - However, the following map is, in general, not measurable:
        - $\mathcal{X} \times \prod_{x \in \mathcal{X}} \mathcal{Y} \to \mathcal{Y}, \quad \left(\tilde{x}, (y_x)_{x \in \mathcal{X}}\right) \mapsto y_{\tilde{x}}$.
- So *Strong Ignorability* does formally not go well with *Consistency*.

Problem 5
- Counterfactual Probabilities

Counterfactual Probabilities

- For reasoning about treatment effect we consider the variables:
  - X = observed treatment variable,
  - Y = observed outcome,
  - Y_x = potential outcome variable under (forced) treatment $X = x$.
- **Conditional counterfactual probabilities**:
  - $C(A \mid x, x') := P(Y_x \in A \mid X = x')$
  - "What would have happened (with which probability) under treatment $X = x$ given that the patient was actually treated with $X = x'$?"

Problems

- Not clear if conditional counterfactual probabilities are probability measures in A and/or measurable in x, x' or jointly.
  - $C(A \mid x, x') := P(Y_x \in A \mid X = x')$
- Not clear if conditioning is well-defined here, dependent on how to view $x \mapsto Y_x$.

Quasi-Measurable Spaces

References

- The talk is based on the following papers:

    - Chris Heunen, Ohad Kammar, Sam Staton, Hongseok Yang. ***A convenient category for higher-order probability theory***. 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS), 2017.

    - Patrick Forré, ***Quasi-Measurable Spaces***, 2021, https://arxiv.org/abs/2109.11631.

Main Idea behind Quasi-Measurable Spaces

- Main idea: *Exchange the role of σ-algebras and random variables!!!*
- **Sample space** is a measurable space: $\left(\Omega, \mathscr{B}_{\Omega}\right)$
  - where $\mathscr{B}_{\Omega}$ is the σ-algebra/set of admissible outcome events on Ω.
- **State space** is a "***quasi-measurable space***": $\left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$
  - where $\mathscr{X}^{\Omega}$ is a set of admissible **random variables**.
- **σ-algebra** of admissible events is:
  - $\mathscr{B}_{\mathscr{X}} := \mathscr{B}\left(\mathscr{X}^{\Omega}\right) := \left\{A \subseteq \mathscr{X} \mid \forall X \in \mathscr{X}^{\Omega} . X^{-1}(A) \in \mathscr{B}_{\Omega}\right\}$
- For fixed **probability measure** P on $\left(\Omega, \mathscr{B}_{\Omega}\right)$ the distribution of X is:
  - push-forward probability measure: $X_{*}P$ on $\mathscr{B}_{\mathscr{X}}$ (also written as: $P(X)$).

The Sample Space - Act 1 - Random Variables

- The **Sample Space** (Ω, Ω^Ω) consists of:
  - a set: Ω
  - a set of maps: $\Omega^\Omega \subseteq \{\Phi : \Omega \to \Omega\}$
    - such that:
      - $\mathrm{id}_\Omega \in \Omega^\Omega$,
      - Ω^Ω contains all constant maps,
      - Ω^Ω is closed under composition:
        - $\Phi_1, \Phi_2 \in \Omega^\Omega \implies \Phi_2 \circ \Phi_1 \in \Omega^\Omega$.
- Standard example:
  - $\Omega^\Omega := \mathrm{Meas}\left((\Omega, \mathscr{B}_\Omega), (\Omega, \mathscr{B}_\Omega)\right)$ for some carefully chosen σ-algebra: $\mathscr{B}_\Omega$.

Quasi-Measurable Spaces

- A **Quasi-Measurable Space** $\left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$ w.r.t. sample space $(\Omega, \Omega^{\Omega})$ - per definition - consists of:
  - a set: $\mathscr{X}$
  - a set of **admissible random variables**: $\mathscr{X}^{\Omega}$,
    - i.e. a set of maps: $X: \ \Omega \to \mathscr{X}$, such that:
      - all *constant maps* $\Omega \to \mathscr{X}$ are in $\mathscr{X}^{\Omega}$,
      - $\mathscr{X}^{\Omega}$ is *closed under pre-composition* with Ω^{Ω}:
        - $X \in \mathscr{X}^{\Omega}, \Phi \in \Omega^{\Omega} \implies X \circ \Phi \in \mathscr{X}^{\Omega}$.

Quasi-Measurable Maps

- Let $\left(\mathcal{Z}, \mathcal{Z}^{\Omega}\right)$ and $\left(\mathcal{X}, \mathcal{X}^{\Omega}\right)$ two quasi-measurable spaces.
- A map $g: \mathcal{Z} \to \mathcal{X}$ is called **quasi-measurable** if
  - $Z \in \mathcal{Z}^{\Omega} \implies g(Z) := g \circ Z \in \mathcal{X}^{\Omega}$
- The set of all quasi-measurable maps is abbreviated:
  - $\mathrm{QMS}\left(\left(\mathcal{Z}, \mathcal{Z}^{\Omega}\right), \left(\mathcal{X}, \mathcal{X}^{\Omega}\right)\right)$ or $\mathrm{QMS}\left(\mathcal{Z}, \mathcal{X}\right)$ for short.
- Note that the *composition* of two quasi-measurable maps is again *quasi-measurable*.
- The class of all quasi-measuable spaces (w.r.t. a fixed sample space) together with all **quasi-measurable maps** builds a **category**: QMS.

The Product Space

- Let $\left(\mathscr{X}_i, \mathscr{X}_i^{\Omega}\right)$ be a family of quasi-measurable spaces, $i \in I$.
- Then we turn the product space: $\prod_{i \in I} \mathscr{X}_i$
  - into a quasi-measurable space by putting: $\left(\prod_{i \in I} \mathscr{X}_i\right)^{\Omega} := \prod_{i \in I} \mathscr{X}_i^{\Omega}$
  - product random variables on the product are of the form:
    - $X(\omega) = \left(X_i(\omega)\right)_{i \in I}$ with $X_i \in \mathscr{X}_i^{\Omega}$ for all $i \in I$.

The Function Space

- Let $\left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$ and $\left(\mathscr{Z}, \mathscr{Z}^{\Omega}\right)$ two quasi-measurable spaces. We put:
  - $\mathscr{X}^{\mathscr{Z}} := \mathrm{QMS}\left(\left(\mathscr{Z}, \mathscr{Z}^{\Omega}\right), \left(\mathscr{X}, \mathscr{X}^{\Omega}\right)\right)$
  - $\left(\mathscr{X}^{\mathscr{Z}}\right)^{\Omega} := \left\{X: \ \Omega \to \mathscr{X}^{\mathscr{Z}} \mid \big((\omega, z) \mapsto X(\omega)(z)\big) \in \mathrm{QMS}(\Omega \times \mathscr{Z}, \mathscr{X})\right\}$
    - function-valued random variables are defined via the product structure
- Then $\left(\mathscr{X}^{\mathscr{Z}}, \left(\mathscr{X}^{\mathscr{Z}}\right)^{\Omega}\right)$ is a quasi-measurable space.
- Note that such a construction was not possible for measurable spaces!!!

Currying, Uncurrying and the Evaluation Map

- Let $\left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$, $\left(\mathscr{Y}, \mathscr{Y}^{\Omega}\right)$ and $\left(\mathscr{Z}, \mathscr{Z}^{\Omega}\right)$ be quasi-measurable spaces.
- We can then **curry** and **uncurry**:
  - $\mathrm{QMS}\left(\mathscr{Z} \times \mathscr{Y}, \mathscr{X}\right) \cong \mathrm{QMS}\left(\mathscr{Y}, \mathscr{X}^{\mathscr{Z}}\right) = \mathrm{QMS}\Big(\mathscr{Y}, \mathrm{QMS}\left(\mathscr{Z}, \mathscr{X}\right)\Big)$
- In particular, the **evaluation map** is quasi-measurable:
  - $\mathrm{ev}: \ \mathscr{X}^{\mathscr{Z}} \times \mathscr{Z} \to \mathscr{X}, \quad \mathrm{ev}(g, z) := g(z).$
- Note that this was not possible in Meas for measurable spaces!

Theorem - QMS cartesian closed

- Theorem: The category of quasi-measurable spaces QMS is **cartesian closed**.

- Remark: We can also construct the following in QMS:

  - **coproducts**, **equalizers**, **coequalizers**, thus:

  - all small **limits** and all small **colimits**

- Remark: This means that QMS allows for **simply typed λ-calculus.**

- Remark: Note that most of this is not true for the category of measurable spaces Meas!

The Slice Categories

- Let $\left(\mathscr{T}, \mathscr{T}^{\Omega}\right)$ be a quasi-measurable space.
- The slice category $\mathrm{QMS}_{\mathscr{T}}$ of all *quasi-measurable spaces over* $\mathscr{T}$ is given by:
  - objects: quasi-measurable maps: $T_{\mathscr{X}} : \mathscr{X} \to \mathscr{T}$
  - morphisms: quasi-measurable maps: $f : \mathscr{X} \to \mathscr{Y}$
    - s.t.: $T_{\mathscr{Y}} \circ f = T_{\mathscr{X}}$.
- By abuse of notation we will just write $\mathscr{X}$ instead of $\left(\mathscr{X}, \mathscr{X}^{\Omega}, T_{\mathscr{X}}\right)$

The Fibre Product

- Let $\mathscr{T} \in \mathrm{QMS}$ and $\mathscr{X}, \mathscr{Y} \in \mathrm{QMS}_{\mathscr{T}}$.
- The *fibre product* of $\mathscr{X}$ and $\mathscr{Y}$ over $\mathscr{T}$ is given as follows:
  - $\mathscr{X} \times_{\mathscr{T}} \mathscr{Y} := \left\{ (x, y) \in \mathscr{X} \times \mathscr{Y} \mid T_{\mathscr{X}}(x) = T_{\mathscr{Y}}(y) \right\}$
  - $\left(\mathscr{X} \times_{\mathscr{T}} \mathscr{Y}\right)^{\Omega} := \left\{ (X, Y) \in \mathscr{X}^{\Omega} \times \mathscr{Y}^{\Omega} \mid T_{\mathscr{X}} \circ X = T_{\mathscr{Y}} \circ Y \right\}$
  - $T: \mathscr{X} \times_{\mathscr{T}} \mathscr{Y} \to \mathscr{T}, \qquad T(x, y) := T_{\mathscr{X}}(x) = T_{\mathscr{Y}}(y)$
- This makes $\mathscr{X} \times_{\mathscr{T}} \mathscr{Y}$ a quasi-measurable space over $\mathscr{T}$.
- In fact, the fibre product is the categorical product of $\mathrm{QMS}_{\mathscr{T}}$.

The Internal Hom

- Let $\mathscr{T} \in \mathrm{QMS}$ and $\mathscr{X}, \mathscr{Y}, \mathscr{Z} \in \mathrm{QMS}_{\mathscr{T}}$.
- The *internal hom* from $\mathscr{Y}$ to $\mathscr{Z}$ over $\mathscr{T}$ is given as follows:
  - $\mathscr{Q}_{\mathscr{T}}(\mathscr{Y}, \mathscr{Z}) := \coprod_{t \in \mathscr{T}} \mathrm{QMS}(\mathscr{Y}_t, \mathscr{Z}_t)$ (coproduct taken in Sets)
  - $T_{\mathscr{Q}} : \mathscr{Q}_{\mathscr{T}}(\mathscr{Y}, \mathscr{Z}) \to \mathscr{T}, \quad T_{\mathscr{Q}}(t, g) := t.$
  - $\mathscr{Q}_{\mathscr{T}}(\mathscr{Y}, \mathscr{Z})^{\Omega} := \Big\{ (T, G) : \Omega \to \mathscr{Q}_{\mathscr{T}}(\mathscr{Y}, \mathscr{Z}) \mid T \in \mathscr{T}^{\Omega}$ and

    $\forall \Phi \in \Omega^{\Omega}, \forall Y \in \mathscr{Y}^{\Omega}$ s.t. $T \circ \Phi = T_{\mathscr{Y}} \circ Y.$

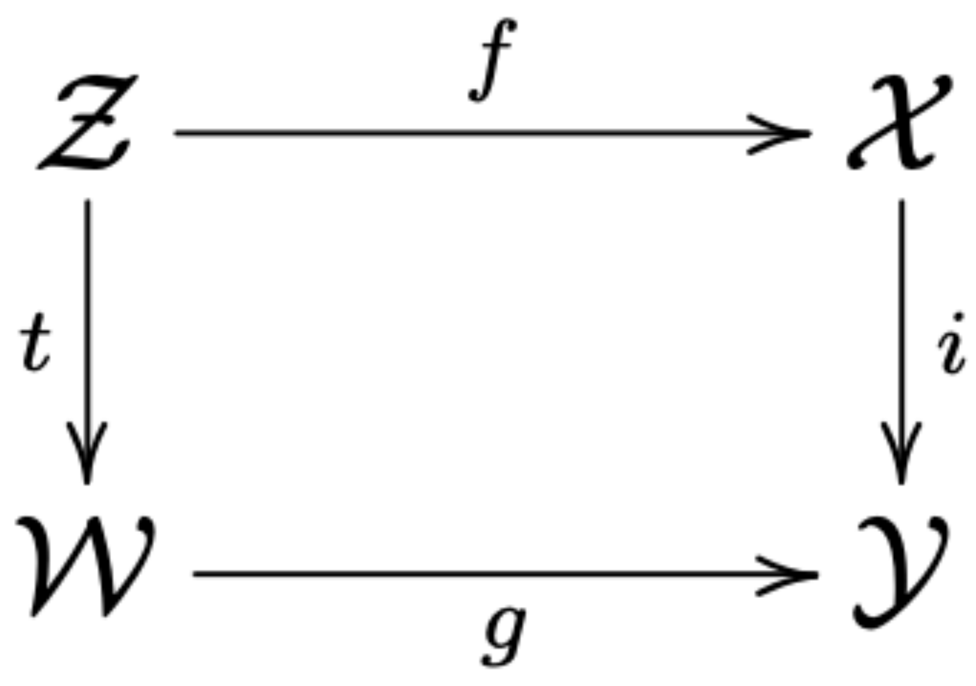
    $G(\Phi)(Y) \in \mathscr{Z}^{\Omega} \Big\}$
    - with $G(\Phi)(Y) : \Omega \to \mathscr{Z}, \quad \omega \mapsto G(\Phi(\omega))(Y(\omega))$

The Internal Hom - Theorem

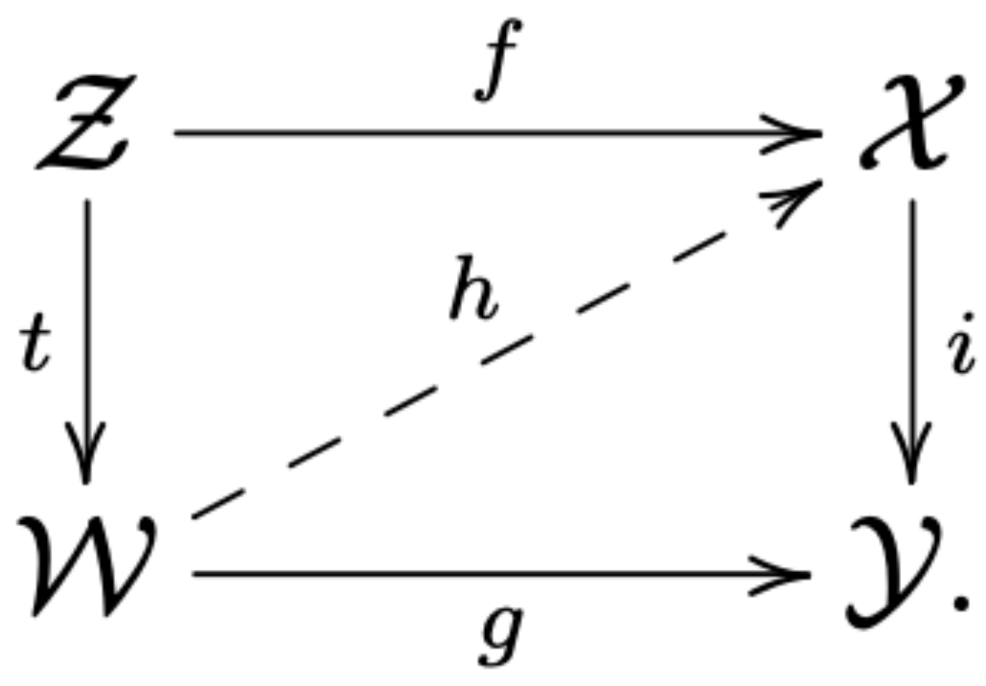
- Let $\mathscr{T} \in \mathrm{QMS}$ and $\mathscr{X}, \mathscr{Y}, \mathscr{Z} \in \mathrm{QMS}_{\mathscr{T}}$.
- The *internal hom* $\mathscr{Q}_{\mathscr{T}}(\mathscr{Y}, \mathscr{Z})$ is a quasi-measurable space over $\mathscr{T}$.
- The internal hom defines an exponential object in the slice category $\mathrm{QMS}_{\mathscr{T}}$:
  - The curry map is well-defined and a canonical bijection:
    - $\mathrm{QMS}_{\mathscr{T}}\left(\mathscr{X} \times_{\mathscr{T}} \mathscr{Y}, \mathscr{Z}\right) \cong \mathrm{QMS}_{\mathscr{T}}\left(\mathscr{X}, \mathscr{Q}_{\mathscr{T}}(\mathscr{Y}, \mathscr{Z})\right)$
    - $g \mapsto \tilde{g}$ with $\tilde{g}(x) = (T_{\mathscr{X}}(x), g_x)$ with $g_x(y) := g(x, y)$.
- Furthermore, it induces an isomorphism in $\mathrm{QMS}_{\mathscr{T}}$:
    - $\mathscr{Q}_{\mathscr{T}}\left(\mathscr{X} \times_{\mathscr{T}} \mathscr{Y}, \mathscr{Z}\right) \cong \mathscr{Q}_{\mathscr{T}}\left(\mathscr{X}, \mathscr{Q}_{\mathscr{T}}(\mathscr{Y}, \mathscr{Z})\right)$.

Strong Monomorphism = Embeddings

- A quasi-measurable map $i : \mathscr{X} \to \mathscr{Y}$ is called an *embedding* iff
  - i is injective, and:
  - $\mathscr{X}^\Omega = \left\{X : \Omega \to \mathscr{X} \mid i \circ X \in \mathscr{Y}^\Omega\right\}$
- <u>Lemma</u>: Embeddings are exactly the *strong monomorphisms* in QMS.

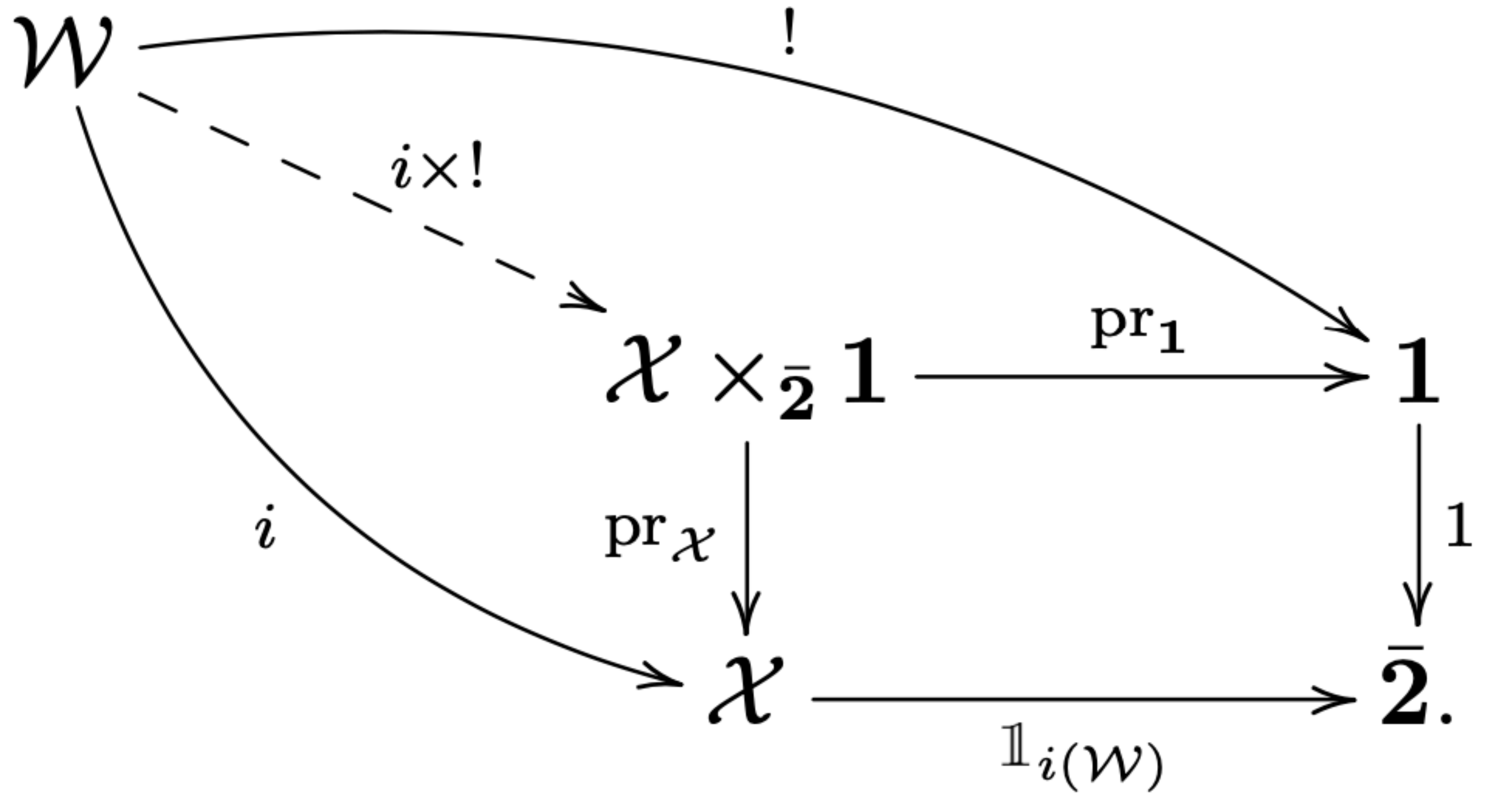


(for t epimorphisms)



Subobject Classifier for Strong Monomorphisms

- Let $\mathbf{2} = \{0,1\}$ be the quasi-measurable space with $\mathbf{2}^\Omega := \mathrm{Sets}(\Omega, \mathbf{2})$.
- Lemma: Then $\mathbf{2}$ together with the constant-1-map $1 : \mathbf{1} \to \mathbf{2}$ is a subobject classifier for strong monomorphisms in QMS.



Definition - Quasitopos

- A **quasitopos** is a category that:
  - has all finite limits,
  - has all finite colimits,
  - is locally cartesian closed,
  - has a subobject classifier for strong monomorphisms.

- Peter T. Johnstone. *Sketches of an Elephant: A Topos Theory Compendium*. Oxford University Press, 2002.

Main Theorems

- Theorem: The category of quasi-measurable spaces QMS forms a **quasitopos**, and, is in particular, **locally cartesian closed**.
- Remark: This means that, besides **simply typed λ-calculus**, we get a **dependent type theory** for QMS.
  - Roughly speaking, this means that we can model programs that can vary the output type/space dependent on the input.
- Remark: Note that most of this is not true for the category of measurable spaces Meas!

The Sample Space - Act 2 - The σ-Algebra

- We now endow the Sample Space (Ω, Ω^Ω) with an additional **σ-algebra** $\mathscr{B}_\Omega$ such that:
    - $\Omega^\Omega \subseteq \mathrm{Meas}\left((\Omega, \mathscr{B}_\Omega), (\Omega, \mathscr{B}_\Omega)\right)$.

- The **Sample Space** is now the triple: $(\Omega, \Omega^\Omega, \mathscr{B}_\Omega)$.

- Standard example:
    - $\Omega^\Omega = \mathrm{Meas}\left((\Omega, \mathscr{B}_\Omega), (\Omega, \mathscr{B}_\Omega)\right)$

Topological and Measurable Spaces as Quasi-Measurable Spaces

- If $\left(\mathcal{X}, \mathcal{E}_{\mathcal{X}}\right)$ is a measurable space or a topological space, etc., then we can turn this into a quasi-measurable space via allowing for the following random variables:
  - $\mathcal{X}^{\Omega} := \mathcal{F}(\mathcal{E}_{\mathcal{X}}) := \left\{X : \Omega \to \mathcal{X} \mid \forall A \in \mathcal{E}_{\mathcal{X}} . X^{-1}(A) \in \mathcal{B}_{\Omega}\right\}$

- Note that the later introduced σ-algebra $\mathcal{B}_{\mathcal{X}}$ might be strictly bigger than the one we started with to turn $\left(\mathcal{X}, \mathcal{E}_{\mathcal{X}}\right)$ into quasi-measurable space $(\mathcal{X}, \mathcal{X}^{\Omega})$:
  - $\mathcal{E}_{\mathcal{X}} \subsetneq \mathcal{B}_{\mathcal{X}} := \mathcal{B}\left(\mathcal{X}^{\Omega}\right)$

The σ-Algebra

- Let $\left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$ be a quasi-measurable space.
- Then the **induced σ-algebra** is:
  - $\mathscr{B}_{\mathscr{X}} := \left\{A \subseteq \mathscr{X} \mid \forall X \in \mathscr{X}^{\Omega} \,.\, X^{-1}(A) \in \mathscr{B}_{\Omega}\right\}$
- We can then define the set of admissible random variables with values in $\mathscr{B}_{\mathscr{X}}$ via:
  - $\left(\mathscr{B}_{\mathscr{X}}\right)^{\Omega} := \left\{\Psi : \Omega \to \mathscr{B}_{\mathscr{X}} \mid \exists D \in \mathscr{B}_{\Omega \times \mathscr{X}} \forall \omega \in \Omega \,.\, \Psi(\omega) = D_{\omega}\right\} \cong \mathscr{B}_{\Omega \times \mathscr{X}}$
    - where $D_{\omega} := \left\{x \in \mathscr{X} \mid (\omega, x) \in D\right\}$
- Then $\left(\mathscr{B}_{\mathscr{X}}, \left(\mathscr{B}_{\mathscr{X}}\right)^{\Omega}\right)$ is a quasi-measurable space.
- Note that this was not possible in the category of measurable spaces!!!

Theorem - The Adjunction

- A **map** $g : \mathscr{X} \to \mathscr{Y}$ from a quasi-measurable space $(\mathscr{X}, \mathscr{X}^{\Omega})$ to a measurable space $(\mathscr{Y}, \mathscr{B}_{\mathscr{Y}})$ is
    - **measurable** *if and only if* it is **quasi-measurable**,
  - *provided* we use the corresponding choices:
    - $\mathscr{B}_{\mathscr{X}} := \mathscr{B}(\mathscr{X}^{\Omega}) := \{A \subseteq \mathscr{X} \mid \forall X \in \mathscr{X}^{\Omega} \,.\, X^{-1}(A) \in \mathscr{B}_{\Omega}\}$,
    - $\mathscr{Y}^{\Omega} := \mathscr{F}(\mathscr{B}_{\mathscr{Y}}) := \{Y : \Omega \to \mathscr{Y} \mid \forall B \in \mathscr{B}_{\mathscr{Y}} \,.\, Y^{-1}(B) \in \mathscr{B}_{\Omega}\}$.
- In other words, we have the natural identification of sets of maps:
  - $\mathrm{Meas}\left((\mathscr{X}, \mathscr{B}(\mathscr{X}^{\Omega})), (\mathscr{Y}, \mathscr{B}_{\mathscr{Y}})\right) = \mathrm{QMS}\left((\mathscr{X}, \mathscr{X}^{\Omega}), (\mathscr{Y}, \mathscr{F}(\mathscr{B}_{\mathscr{Y}}))\right)$.

The Sample Space - Act 3 - Probability Measures

- We now endow the Sample Space $(\Omega, \Omega^\Omega, \mathscr{B}_\Omega)$ with some additional set of **product compatible probability measures** $\mathscr{P}$ on $\mathscr{B}_\Omega$, i.e. such that:
  - for all $P \in \mathscr{P}$ and $D \in \mathscr{B}_{\Omega \times \Omega}$ the map:
    - $\Omega \to [0,1], \quad \omega \mapsto P(D^\omega), \quad$ is (quasi-)measurable,
      - where $D^\omega := \{\tilde{\omega} \in \Omega \,|\, (\tilde{\omega}, \omega) \in D\}$,
  - for all $P_1, P_2 \in \mathscr{P}$ there exist $\Phi_1, \Phi_2 \in \Omega^\Omega$ and $P \in \mathscr{P}$ such that:
    - $P_1 \otimes P_2 = P(\Phi_1, \Phi_2)$ on $\mathscr{B}_{\Omega \times \Omega}$, i.e. for all $D \in \mathscr{B}_{\Omega \times \Omega}$ we have:
      - $(P_1 \otimes P_2)(D) := \int P_1(D^\omega)\, P_2(d\omega) = P(\{\omega \in \Omega \,|\, (\Phi_1(\omega), \Phi_2(\omega)) \in D\})$.
- The **Sample Space** is now the quadruple: $(\Omega, \Omega^\Omega, \mathscr{B}_\Omega, \mathscr{P})$.

The Space of Push-forward Probability Measures

- Let $\left(\mathcal{X}, \mathcal{X}^{\Omega}\right)$ be a quasi-measurable space. Define:
  - $\mathcal{P}(\mathcal{X}) := \mathcal{P}(\mathcal{X}, \mathcal{X}^{\Omega}) := \left\{P(X) : \mathcal{B}_{\mathcal{X}} \to [0,1] \,|\, X \in \mathcal{X}^{\Omega}, P \in \mathcal{P}\right\}$
  - $\mathcal{P}(\mathcal{X})^{\Omega} := \mathcal{P}(\mathcal{X}, \mathcal{X}^{\Omega})^{\Omega} := \left\{P(X|I) \,|\, X \in (\mathcal{X}^{\Omega})^{\Omega}, P \in \mathcal{P}\right\}$
  - $P(X \in A \,|\, I = \omega) := P\left(\left\{\tilde{\omega} \in \Omega \,|\, X(\omega)(\tilde{\omega}) \in A\right\}\right)$ for $A \in \mathcal{B}_{\mathcal{X}}$

- Lemma: $\left(\mathcal{P}(\mathcal{X}), \mathcal{P}(\mathcal{X})^{\Omega}\right)$ is also a quasi-measurable space.

The Spaces of Markov Kernels and Random Functions

- Let $\left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$ and $\left(\mathscr{Z}, \mathscr{Z}^{\Omega}\right)$ be quasi-measurable spaces.
- Then the **space of Markov kernels** from $\left(\mathscr{Z}, \mathscr{Z}^{\Omega}\right)$ to $\left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$:
  - $\mathscr{P}(\mathscr{X})^{\mathscr{Z}} = \mathrm{QMS}\left(\left(\mathscr{Z}, \mathscr{Z}^{\Omega}\right), \left(\mathscr{P}(\mathscr{X}), \mathscr{P}(\mathscr{X})^{\Omega}\right)\right)$
  - is again a quasi-measurable space.
- Also the space of probability distribution over functions:
  - $\mathscr{P}\left(\mathscr{X}^{\mathscr{Z}}\right)$ is again a quasi-measurable space.
- Note that these construction were not possible in the category of measurable spaces!!!

Some surprising Lemmata

- Let $\left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$ and $\left(\mathscr{Y}, \mathscr{Y}^{\Omega}\right)$ be quasi-measurable spaces.
- Then the following maps are all quasi-measurable:
  - $\mathscr{Y}^{\mathscr{X}} \times \mathscr{B}_{\mathscr{Y}} \to \mathscr{B}_{\mathscr{X}}, \qquad (f, B) \mapsto f^{-1}(B).$
  - $\mathscr{P}(\mathscr{X}) \times \mathscr{B}_{\mathscr{X}} \to [0,1], \qquad (P, A) \mapsto P(A).$
  - $\mathscr{Y}^{\mathscr{X}} \times \mathscr{P}(\mathscr{X}) \to \mathscr{P}(\mathscr{Y}), \qquad (f, P) \mapsto f_* P.$
  - $[0,\infty]^{\mathscr{X}} \times \mathscr{P}(\mathscr{X}) \to [0,\infty], \quad (h, P) \mapsto \int h(x)\, P(dx).$
- Note that such statements were not known or even possible in the category of measurable spaces!!!

Theorem: The Product of Markov Kernels

- Assume that there exists an isomorphism of quasi-measurable spaces:
    - $\Omega \times \Omega \cong \Omega$.
- Then for all quasi-measurable spaces $(\mathcal{X}, \mathcal{X}^\Omega)$, $(\mathcal{Y}, \mathcal{Y}^\Omega)$, $(\mathcal{Z}, \mathcal{Z}^\Omega)$ the **product of Markov kernels**:
    - $\otimes: \ \mathcal{P}(\mathcal{X})^{\mathcal{Y} \times \mathcal{Z}} \times \mathcal{P}(\mathcal{Y})^{\mathcal{Z}} \to \mathcal{P}(\mathcal{X} \times \mathcal{Y})^{\mathcal{Z}}$

$$\big(P(X \,|\, Y, Z) \otimes Q(Y \,|\, Z)\big)(D \,|\, z) := \int P(X \in D^y \,|\, Y = y, Z = z)\, Q(Y \in dy \,|\, Z = z)$$

- is a well-defined quasi-measurable map.

Theorem: Strong Probability Monad

- If $\Omega \times \Omega \cong \Omega$ then the triple $(\mathscr{P}, \delta, \mathbb{M})$ is a **strong probability monad** on the cartesian closed category QMS, where:
  - $\delta : \mathscr{X} \to \mathscr{P}(\mathscr{X}), \qquad \delta_x(A) := \mathbb{1}_A(x),$
  - $\mathbb{M} : \mathscr{P}\left(\mathscr{P}(\mathscr{X})\right) \to \mathscr{P}(\mathscr{X}), \qquad \mathbb{M}(\Pi)(A) := \int P(A)\, d\Pi(P).$
- This thus allows for a notion of computation of monadic type and simply typed λ-calculus.
- We thus get semantics for higher-order probability theory for probabilistic programming language.

Construction of well-behaved Sample Spaces

- Theorem: Let Ω_0 be a set, and:
  - $\mathscr{E}_0$ a countable set of subsets of Ω_0 that separates the points of Ω_0.
  - $\Omega := \prod_{n \in \mathbb{N}} \Omega_0$, and $\mathscr{E} := \{\mathrm{pr}_n^{-1}(A) \mid A \in \mathscr{E}_0, n \in \mathbb{N}\}$,
  - $\tilde{\mathscr{P}} := \{P$ complete perfect probability measure on $\Omega, \mathscr{E} \subseteq \mathscr{B}_P\}$,
  - $\mathscr{B}_\Omega := \bigcap_{P \in \tilde{\mathscr{P}}} \mathscr{B}_P$, the perfect-universal completion of $\mathscr{E}$,
  - $\Omega^\Omega := \mathrm{Meas}\big((\Omega, \mathscr{B}_\Omega), (\Omega, \mathscr{B}_\Omega)\big)$, $\mathscr{P} := \tilde{\mathscr{P}}|_{\mathscr{B}_\Omega}$
- Then $(\Omega, \Omega^\Omega, \mathscr{B}_\Omega, \mathscr{P})$ satisfies all points of act 1-3 and $\Omega \times \Omega \cong \Omega$.

Fubini Theorem

- Let $(\Omega, \Omega^{\Omega}, \mathscr{B}_{\Omega}, \mathscr{P})$ be the sample space from the last slide.
- Let $\left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$ and $\left(\mathscr{Y}, \mathscr{Y}^{\Omega}\right)$ be quasi-measurable spaces and:
  - $f \in [0,\infty]^{\mathscr{X}\times\mathscr{Y}}$, $P \in \mathscr{P}(\mathscr{X})$ and $Q \in \mathscr{P}(\mathscr{Y})$.
- Then we have the equality:
  - $$\int\int f(x, y)\, P(dx)\, Q(dy) = \int\int f(x, y)\, Q(dy)\, P(dx).$$

The Sample Space - Act 4 - The Universal Hilbert Cube

- $\Omega = [0,1]^{\mathbb{N}} = \prod_{n \in \mathbb{N}} [0,1]$, the **Hilbert Cube**,
- $\mathscr{B}_\Omega =$ set of all *universally measurable* subsets of Ω.
  - Note that this is bigger than the Borel σ-algebra on Ω.
- $\mathscr{P} =$ all probability measures on $\mathscr{B}_\Omega$, $\Omega^\Omega = \mathrm{Meas}\left((\Omega, \mathscr{B}_\Omega), (\Omega, \mathscr{B}_\Omega)\right)$.
- We call this Sample Space $(\Omega, \Omega^\Omega, \mathscr{B}_\Omega, \mathscr{P})$ the **Universal Hilbert Cube**.
- Interpretation: Countably infinite sequence of uniformly distributed samples (e.g. from a (pseudo-)random number generator).
- Note that it satisfies act 1-3 and the iso: $\Omega \times \Omega \cong \Omega$ (via "Hilbert's Hotel").

The Category of Quasi-Universal Spaces

- Definition: A **quasi-universal space** $(\mathscr{X}, \mathscr{X}^\Omega)$ is - per definition - just a quasi-measurable space where the sample space Ω is the **universal Hilbert cube**.
- We abbreviate the category of quasi-universal spaces as QUS.

Countably Separated and Standard Quasi-Measurable Spaces

- Definition: A quasi-measurable space $\left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$ is called:
  - **countably separated** if there exists a countable subset $\mathscr{E} \subseteq \mathscr{B}_{\mathscr{X}}$ that separates the points of $\mathscr{X}$.
  - **standard quasi-measurable space** if there are quasi-measurable maps:
    - $\iota : \left(\mathscr{X}, \mathscr{X}^{\Omega}\right) \to \left(\Omega, \Omega^{\Omega}\right)$ and $r : \left(\Omega, \Omega^{\Omega}\right) \to \left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$ s.t.:
      - $r \circ \iota = \mathrm{id}_{\mathscr{X}}$.

Theorem: Disintegration of Markov Kernels

- Let $\left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$ and $\left(\mathscr{Y}, \mathscr{Y}^{\Omega}\right)$ and $\left(\mathscr{Z}, \mathscr{Z}^{\Omega}\right)$ be quasi-universal spaces.
  - Let $\left(\mathscr{Y}, \mathscr{Y}^{\Omega}\right)$ be *countably separated.* and:
  - either $\left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$ or $\left(\mathscr{Z}, \mathscr{Z}^{\Omega}\right)$ be a *standard* quasi-universal space.
- Then the product of Markov kernels:
  - $\otimes: \ \mathscr{P}(\mathscr{X})^{\mathscr{Y} \times \mathscr{Z}} \times \mathscr{P}(\mathscr{Y})^{\mathscr{Z}} \to \mathscr{P}(\mathscr{X} \times \mathscr{Y})^{\mathscr{Z}}$
  - is a (surjective) quotient map of quasi-universal spaces.
- More concretely, for every $P(X, Y | Z) \in \mathscr{P}(\mathscr{X} \times \mathscr{Y})^{\mathscr{Z}}$ there exists $P(X | Y, Z) \in \mathscr{P}(\mathscr{Y})^{\mathscr{Y} \times \mathscr{Z}}$ such that: $P(X, Y | Z) = P(X | Y, Z) \otimes P(Y | Z)$.

Conditional Kolmogorov Extension Theorem

- Let $\left(\mathscr{X}_n, \mathscr{X}_n^\Omega\right)$, $n \in \mathbb{N}$, a sequence of *standard* quasi-universal spaces and $\left(\mathscr{Z}, \mathscr{Z}^\Omega\right)$ be any quasi-universal space.
  - Assume we have $Q_n(X_{0:n} \mid Z) \in \mathscr{P}\left(\mathscr{X}_{0:n}\right)^{\mathscr{Z}}$ such that for every $n \in \mathbb{N}$:
    - $\mathrm{pr}_{0:n,*} Q_{n+1}(X_{0:n+1} \mid Z) = Q_n(X_{0:n} \mid Z)$.
- Then there exists a unique $Q(X_{\mathbb{N}} \mid Z) \in \mathscr{P}\left(\mathscr{X}_{\mathbb{N}}\right)^{\mathscr{Z}}$ such that:
    - $\mathrm{pr}_{0:n,*} Q(X_{0:n+1} \mid Z) = Q_n(X_{0:n} \mid Z)$ for all $n \in \mathbb{N}$,
  - where $\mathscr{X}_{\mathbb{N}} := \prod_{n \in \mathbb{N}} \mathscr{X}_n$.

Conditional De Finetti Theorem

- $\left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$ *standard* quasi-universal spaces, $\left(\mathscr{Z}, \mathscr{Z}^{\Omega}\right)$ any quasi-universal space.
- For a Markov kernel $Q(X_{\mathbb{N}} \mid Z) \in \mathscr{P}\left(\mathscr{X}^{\mathbb{N}}\right)^{\mathscr{Z}}$ the following is equivalent:
  - $Q(X_{\mathbb{N}} \mid Z)$ is **exchangable**, i.e. invariant under all finite permuations: $\rho: \mathbb{N} \cong \mathbb{N}$.
  - There exists a quasi-universal space $\mathscr{Y}$ and $K(X \mid Y) \in \mathscr{P}(\mathscr{X})^{\mathscr{Y}}$ and $P(Y \mid Z) \in \mathscr{P}(\mathscr{Y})^{\mathscr{Z}}$ such that :
    - $$Q(X_{\mathbb{N}} \mid Z) = \left(\bigotimes_{n \in \mathbb{N}} K(X_n \mid Y) \right) \circ P(Y \mid Z).$$
- In this case we can w.l.o.g. take: $\mathscr{Y} = \mathscr{P}(\mathscr{X})$ and $K(X \in A \mid Y = P) := P(A)$.

Tobias Fritz, Tomáš Gonda, Paolo Perrone, *De Finetti's Theorem in Categorical Probability*, 2021, https://arxiv.org/abs/2105.02639.

Transitional Conditional Independence

- Consider a Markov kernel: $P(X, Y, Z | T) \in \mathscr{P}(\mathscr{X} \times \mathscr{Y} \times \mathscr{Z})^{\mathscr{T}}$.
- We say that X is **conditional independent** of Y given Z w.r.t. $P(X, Y, Z | T)$,
  - in symbols: $X \perp\!\!\!\perp Y | Z$ if:
  - there *exists a Markov kernel* $Q(X | Z) \in \mathscr{P}(\mathscr{X})^{\mathscr{Z}}$ such that:
    - $P(X, Y, Z | T) = Q(X | Z) \otimes P(Y, Z | T)$.

Patrick Forré, *Transitional Conditional Independence*, 2021, https://arxiv.org/abs/2104.11547.

Partially Generic Causal Bayesian Networks

- A **partially generic causal Baysian network** - per definition - consists of:
  - a **conditional directed acyclic graph** (CDAG): $G = (J, V, E)$,
  - an input variable X_j on a quasi-universal space $\mathcal{X}_j$ for each $j \in J$,
  - an output variable X_v on a *standard* quasi-universal space $\mathcal{X}_v$ for each $v \in V$,
  - an **exceptional set**: $W \subseteq V$,
  - a Markov kernel: $P_v(X_v | X_{\mathrm{Pa}^G(v)}) \in \mathcal{P}(\mathcal{X}_v)^{\mathcal{X}_{\mathrm{Pa}^G(v)}}$ for $v \in V \setminus W$.

Partially Generic Causal Bayesian Networks

- For a partially generic causal Baysian network with exceptional set W we introduce for $w \in W$:
  - an **indicator variable**: $I_w \to w$,
  - a quasi-universal space: $\mathcal{X}_{I_w} := \mathcal{P}(\mathcal{X}_w)^{\mathcal{X}_{\mathrm{Pa}^G(w)}}$,
  - a **“generic” Markov kernel**:
    - $P_w\left(X_w \in A \,\middle|\, X_{\mathrm{Pa}^G(w)} = x, X_{I_w} = Q\right) := Q\left(X_w \in A \,\middle|\, X_{\mathrm{Pa}^G(w)} = x\right)$.
- So we get a joint Markov kernel: $P(X_V, X_J, X_{I_W} \mid X_J, X_{I_W})$.

Theorem: Global Markov Property

- For every partially generic causal Bayesian network with exceptional set W

  and any subsets: $A, B, C \subseteq V \cup I_W \cup J$ we have the implication:

  - $A \perp B \,|\, C \implies X_A \perp\!\!\!\perp X_B \,|\, X_C$.

(Proposed) Answers

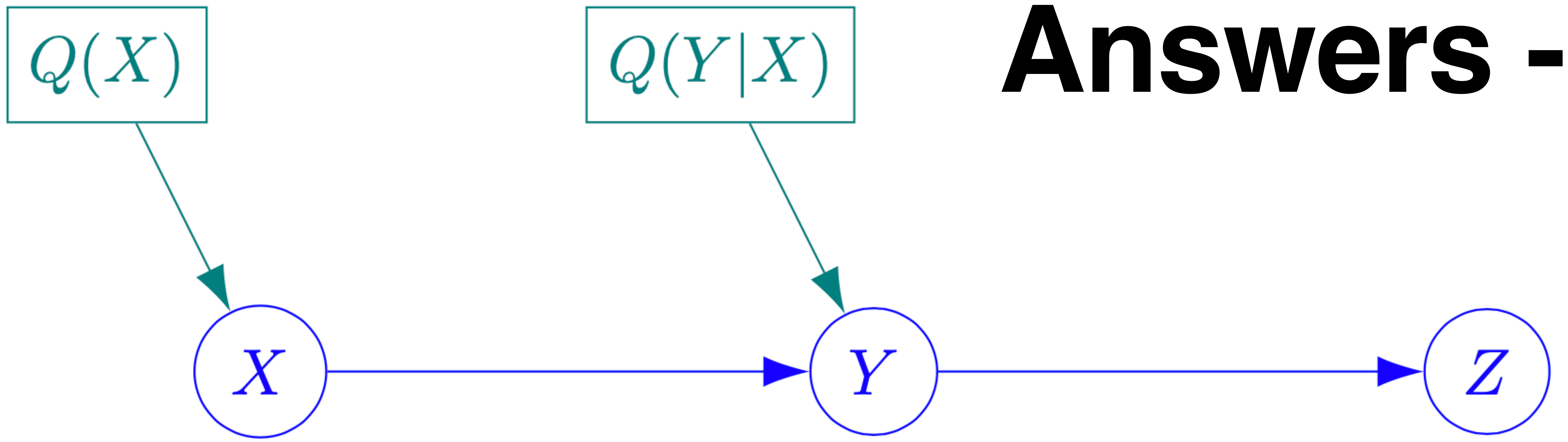
Answers - Stochastic Process

- <u>Definition</u>: A **stochastic process** is a quasi-measurable map:
    - $X: \ \Omega \to \mathcal{X}^{\mathcal{T}}, \qquad \omega \mapsto \big(t \mapsto X(\omega)(t)\big).$
- <u>Lemma</u>: This is *equivalent* to a quasi-measurable map: $X: \ \Omega \times \mathcal{T} \to \mathcal{X}, \ \ (\omega, t) \mapsto X(\omega, t).$
- <u>Lemma</u>: The map: $\mathcal{X}^{\mathcal{T}} \to \prod_{t \in \mathcal{T}} \mathcal{X}, \quad X \mapsto \big(X(t)\big)_{t \in \mathcal{T}},$ is quasi-measurable.
- <u>Lemma</u>: If $T: \ \Omega \to \mathcal{T}$ is quasi-measurable (random time) then the map:
    - $\Omega \to \mathcal{X}, \ \ \omega \mapsto X(\omega)(T(\omega))$ is again quasi-measurable.

Answers - Probabilistic Programs

- Definition: A **probabilistic program** with input $x \in \mathcal{X}$ and output $z \in \mathcal{Z}$ is quasi-measurable map: $\mathcal{X} \to \mathcal{P}(\mathcal{Z})$.
- Theorem: We have the natural **curry / uncurry isomorphism**:
  - $\mathrm{QMS}\left(\mathcal{X} \times \mathcal{Y}, \mathcal{P}\left(\mathcal{Z}\right)\right) \cong \mathrm{QMS}\left(\mathcal{X}, \mathrm{QMS}\left(\mathcal{Y}, \mathcal{P}(\mathcal{Z})\right)\right)$
- Theorem: QMS is a **quasitopos**, thus allows for **dependent type theory**.
- Theorem: The triple $(\mathcal{P}, \delta, \mathbb{M})$ forms a **strong probability monad** on the category of quasi-measurable spaces QMS (for certain sample spaces, e.g. the universal Hilbert cube). Thus allows for **higher-order probabilistic programs**.

Answers - Graphical Models



- **Partially generic causal Bayesian networks** can model graphical models with non-random input variables.
- **Transitional conditional independence** also works with non-random input variables.
- Theorem: **Global Markov Property**: For $A, B, C \subseteq V \cup I_W \cup J$ we have:
    - $A \perp B \,|\, C \implies X_A \perp\!\!\!\perp X_B \,|\, X_C$.
- Example: Here $Q(Y|X)$ is a non-random input variable with values in $\mathscr{L} := \mathrm{QUS}\left(\mathscr{X}, \mathscr{P}(\mathscr{Y})\right)$
    - Then Y is determined by the new quasi-measurable mechanism:
        - $\mathscr{L} \times \mathscr{X} \to \mathscr{P}(\mathscr{Y}), \quad \left(Q(Y|X), x\right) \mapsto Q(Y|X=x)$.
    - We can now read off the graph: $Z \perp\!\!\!\perp X, Q(Y|X), Q(X) \,|\, Y$.

Answers - Causal Assumptions

- Model **potential outcome** as quasi-measurable map / random function:
    - $G: \ \Omega \to \mathscr{Y}^{\mathscr{X}}$
- Potential outcome under treatment $X = x$ then: $Y_x := G(x)$.
- Rephrase causal assumptions:
    - Strong Ignorability: $X \perp\!\!\!\perp G \,|\, Z$,
    - Consistency: $Y = G(X)$.
- Everything is well-defined and quasi-measurable.

Answers - Counterfactual Probabilities

- Theorem: Disintegration of Markov kernels.
- Model potential outcome as: $G \in (\mathscr{Y}^{\mathscr{X}})^{\Omega}$
- Assume that $\mathscr{X}$ to countably separated quasi-universal space.
- Then via the disintegration theorem there exists conditional:
    - $P(G \,|\, X) \in \mathscr{P}(\mathscr{G})^{\mathscr{X}}$ such that $P(G, X) = P(G \,|\, X) \otimes P(X)$.
- Evaluation maps and push-forwards are quasi-measurable, which implies:
    - $C(A \,|\, x, x') := P(G(x) \in A \,|\, X = x')$ defines:
    - well-defined and quasi-measurable $C \in \mathscr{P}(\mathscr{Y})^{\mathscr{X} \times \mathscr{X}}$
- So, conditional counterfactual probabilities are well-defined and quasi-measurable.

Answers - Statistics and Probability Theory

- For (standard) quasi-universal spaces we at least can do the following:
  - Theorem: Disintegration of Markov kernels.
    - Remark: This allows for Bayes' Rule and thus Bayesian Statistics.
  - Theorem: Fubini Theorem.
  - Theorem: Conditional de Finetti Theorem.
  - Theorem: Kolmogorov Extension Theorem.
  - Theorem: Global Markov Property for graphical models like partially generic causal Bayesian networks.

Recommendation

- For probabilistic programming, graphical models, causality, statistics, etc.
  - use for:
    - sample space —> the universal Hilbert cube
  - replace:
    - measurable spaces —> quasi-measurable spaces
    - measurable maps —> quasi-measurable maps
    - categorical construction in Meas —> categorical construction in QMS
- study more of the (classical) theory in this framework (e.g. martingales).

- Patrick Forré, *Quasi-Measurable Spaces*, 2021, https://arxiv.org/abs/2109.11631.

More about Convenient Categories

- **Probability Theory**
    - **Quasi-Borel Spaces** - by Chris Heunen, Ohad Kammar, Sam Staton, Hongseok Yang
    - **Quasi-Measurable Spaces** - by Patrick Forré, https://arxiv.org/abs/2109.11631
- **Topology**
    - **Compactly Generated Weakly Hausdorff Spaces** (CGWH) - by Witold Hurewicz, David Gale, Norman Steenrod, John C. Moore, Michael C. McCord, Neil Strickland, et al (script)
    - **Condensed Sets** - by Peter Scholz, Dustin Clausen (script)
- **Differential Geometry**
    - **Diffeological Spaces** - by Kuo Tsai Chen, Jean-Marie Souriau, Patrick Iglesias-Zemmour, John Baez, Alexander Hoffnung, Andrew Stacey, et al.

Thank you for your attention!