# Quasi-Measurable Spaces

# A Convenient Foundation of Probability Theory

**Patrick Forré**

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

# References

- The talk is based on the following papers:

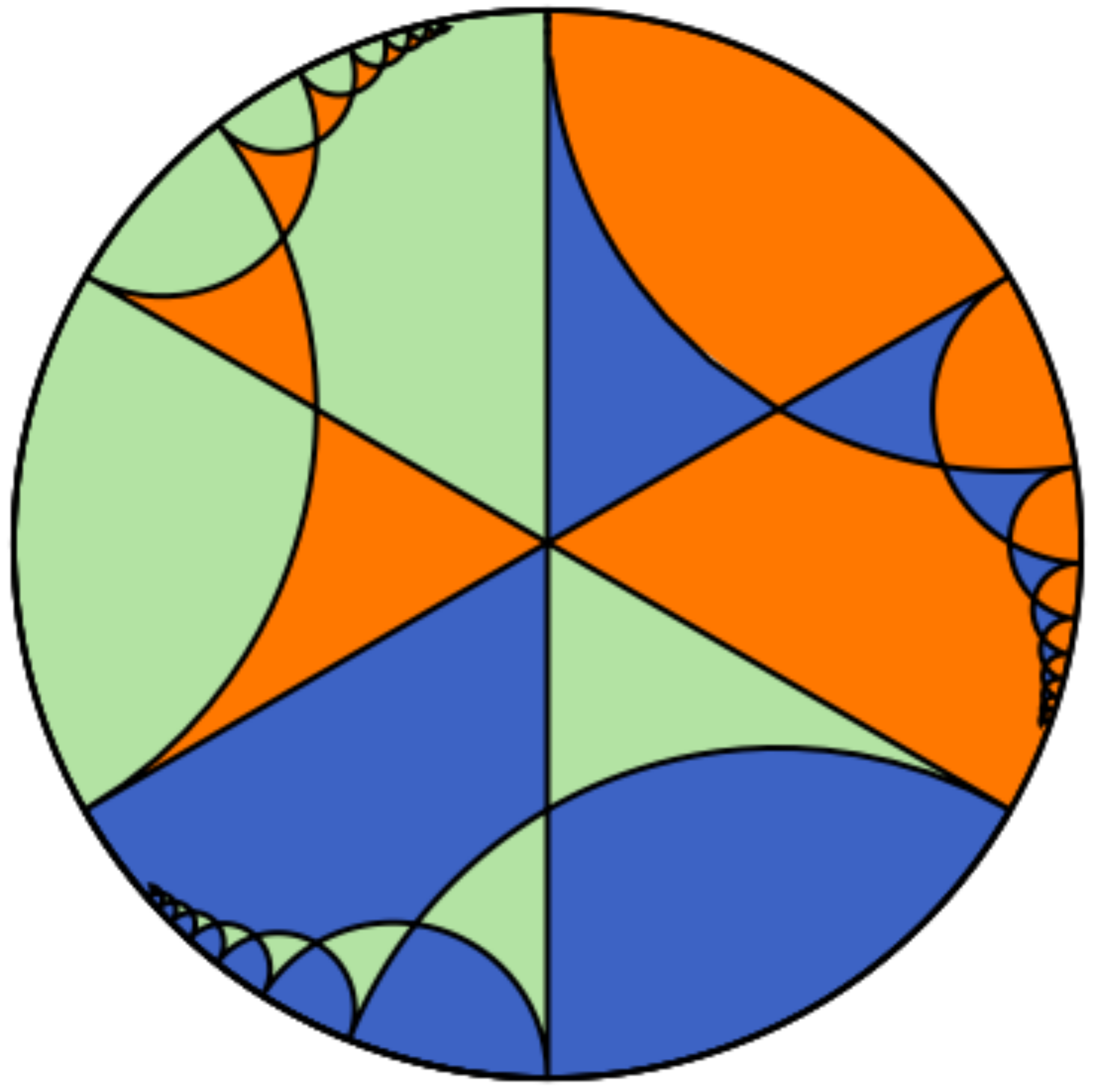
    - Chris Heunen, Ohad Kammar, Sam Staton, Hongseok Yang. ***A convenient category for higher-order probability theory***. 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS), 2017.

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

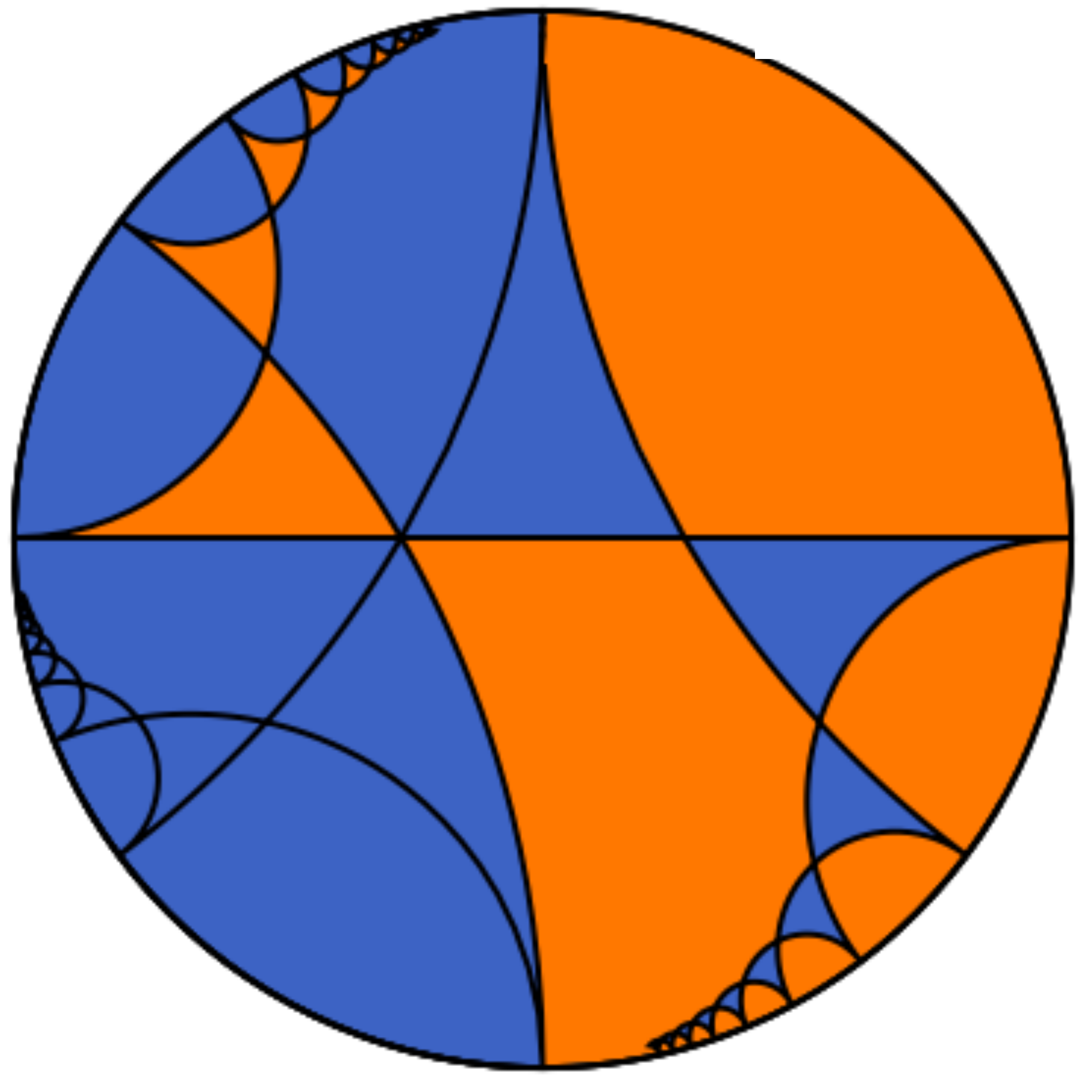
# Why Measure Theory to do Probability Theory?

# Why Measure Theory in the first place?

- The existence of the Lebesgue measure:
  - does not exist on whole power set $2^{\mathbb{R}}$.
  - but does exist on Borel $\sigma$-algebra $\mathscr{B}_{\mathbb{R}}$.



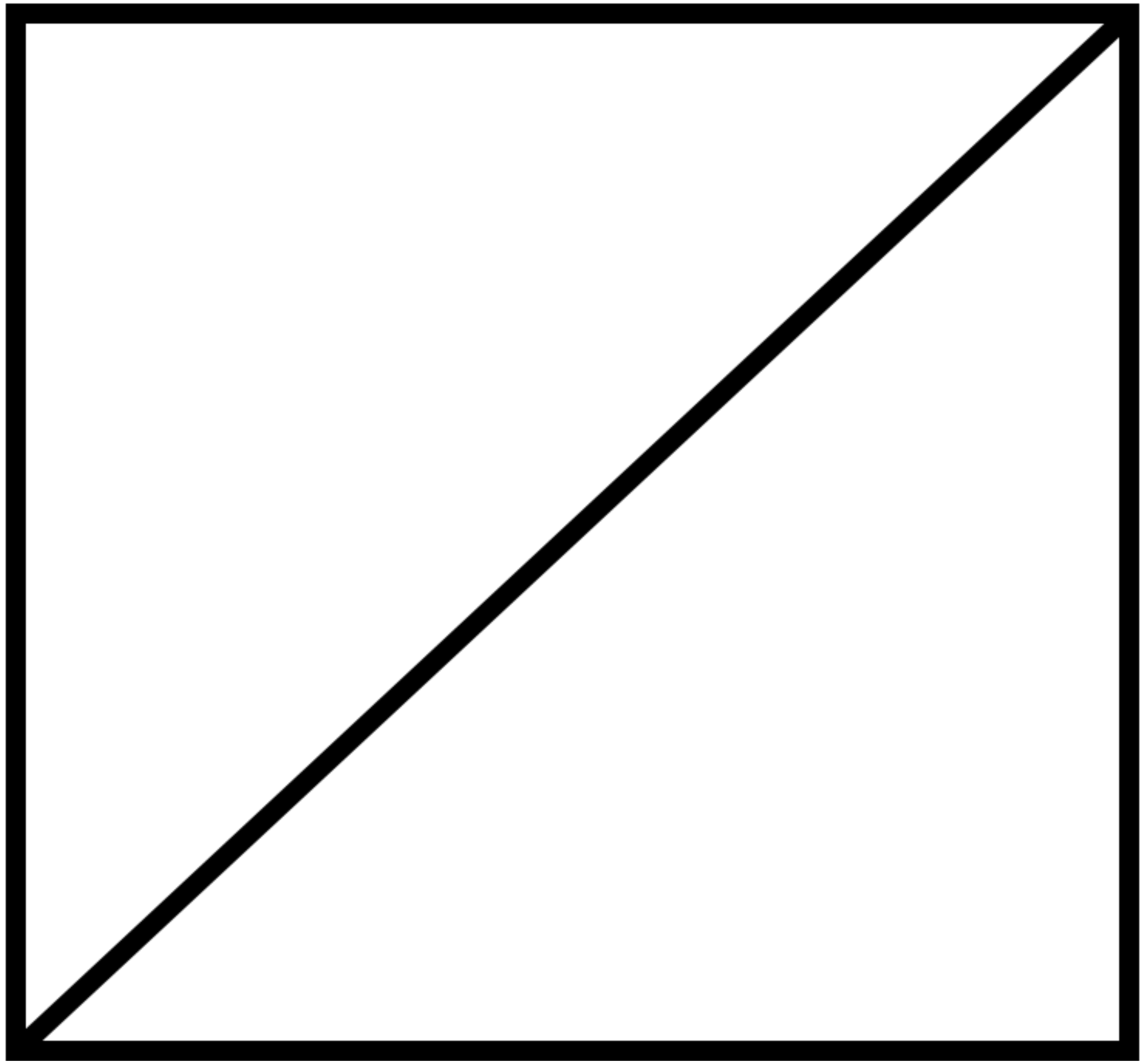
- To prevent set-theoretic paradoxa like Banach-Tarski:
  - the orange is both a third and a half of the Poincaré disk / hyperbolic plane.




- Stan Wagon. The Banach Tarski Paradox. CUP 1985. https://demonstrations.wolfram.com/TheBanachTarskiParadox/

# Discrete and continuous distributions are not expressive enough

- The uniform distribution on the diagonal $\Delta \subseteq [0,1]^2$
  - is neither discrete nor absolute continuous w.r.t. $\lambda^2$.
  - so it can not be described with a probability mass function nor with a probability density w.r.t. $\lambda^2$.



# The Category of Measurable Spaces

- Let $\mathscr{X}$ be a set. A $\sigma$**-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 $\sigma$-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.
- $\mathrm{Meas}$ 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.
    - now allows for Lebesgue’s theory of integration (measure integrals, etc.)

- Measure Theory as an expressive “safe space” of Probability Theory.

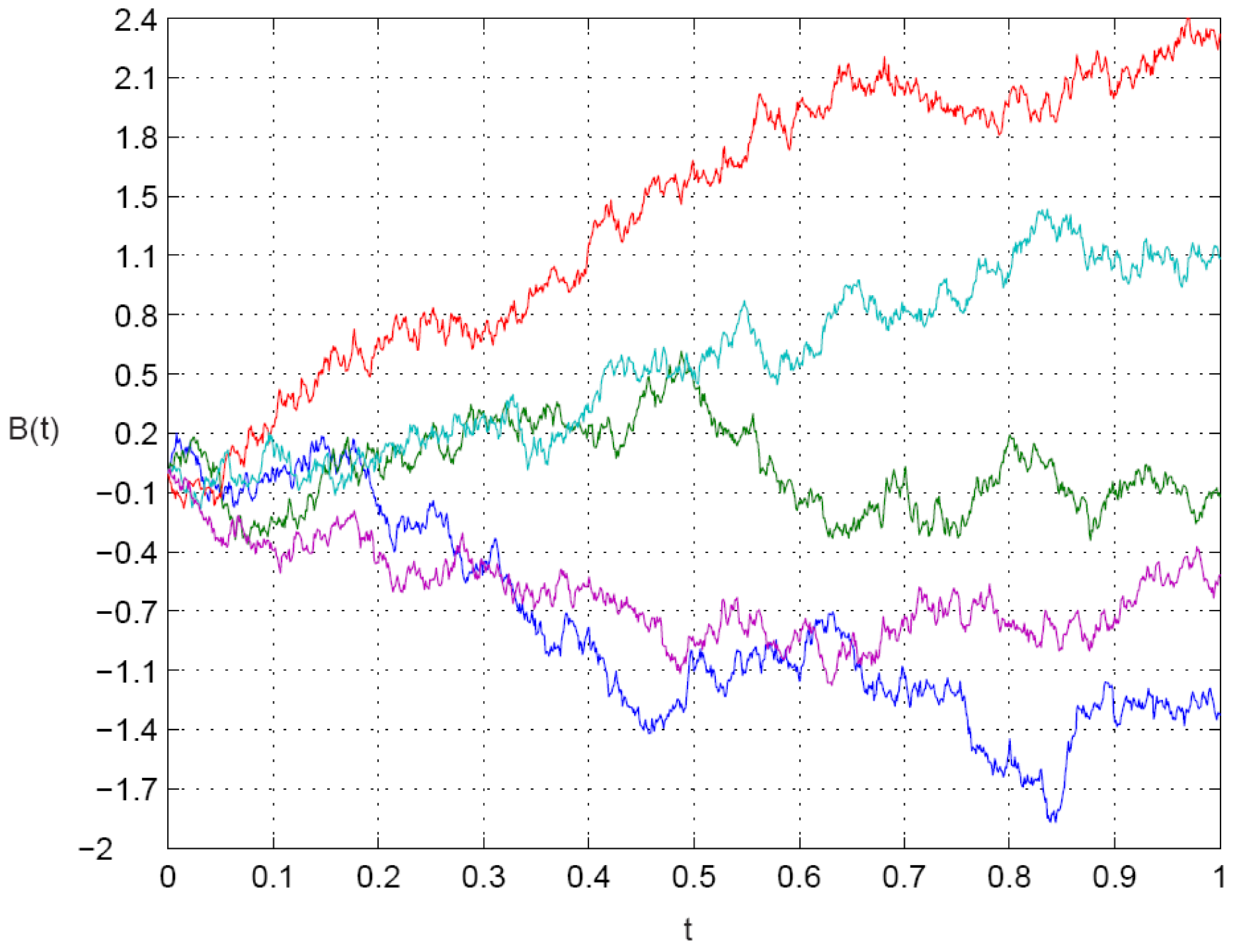

- Andrei Kolmogoroff. *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Ergebnisse der Mathematik und Ihrer Grenzgebiete. 1. Folge, Nr. 2, Springer (1933).

# How to formalize Random Variables?

- **Sample space** is a measurable space: $\left(\Omega, \mathscr{B}_\Omega\right)$
    - where $\mathscr{B}_\Omega$ is the $\sigma$-algebra/set of admissible outcome events on $\Omega$.
- **State space** is a measurable space: $\left(\mathscr{X}, \mathscr{B}_\mathscr{X}\right)$
    - where $\mathscr{B}_\mathscr{X}$ is another **$\sigma$-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: $P(X)$).

# What is a Stochastic Process?

# Different realizations of one stochastic process



# Text book definitions - stochastic process

- *D. Revus, M. Yor - Continuous Martingales and Brownian Motion*:

**(1.1) Definition.** *Let $T$ be a set, $(E, \mathscr{E})$ a measurable space. A* stochastic process *indexed by $T$, taking its values in $(E, \mathscr{E})$, is a family of measurable mappings $X_t$, $t \in T$, from a probability space $(\Omega, \mathscr{F}, P)$ into $(E, \mathscr{E})$. The space $(E, \mathscr{E})$ is called the* state space.

For every $\omega \in \Omega$, the mapping $t \to X_t(\omega)$ is a "curve" in $E$ which is referred to as a *trajectory* or a *path* of $X$. We may think of a path as a point chosen randomly in the space $\mathscr{F}(T, E)$ of all functions from $T$ into $E$, or, as we shall see later, in a reasonable subset of this space.

# Text book definitions - stochastic process

- *D. Revus, M. Yor - Continuous Martingales and Brownian Motion*:

Let $(\mathscr{F}_t)$ be a filtration on $(\Omega, \mathscr{F})$ and $T$ a stopping time. For a process $X$, we define a new mapping $X_T$ on the set $\{\omega : T(\omega) < \infty\}$ by

$$X_T(\omega) = X_t(\omega) \qquad \text{if} \qquad T(\omega) = t.$$

This is the position of the process $X$ at time $T$, but it is not clear that $X_T$ is a random variable on $\{T < \infty\}$. Moreover if $X$ is adapted, we would like $X_T$

**(4.7) Definition.** *A process $X$ is* progressively measurable *or simply* progressive *(with respect to the filtration $(\mathscr{F}_t)$) if for every $t$ the map $(s, \omega) \to X_s(\omega)$ from $[0, t] \times \Omega$ into $(E, \mathscr{E})$ is $\mathscr{B}([0, t]) \otimes \mathscr{F}_t$-measurable. A subset $\Gamma$ of $\mathbb{R}_+ \times \Omega$ is* progressive *if the process $X = 1_\Gamma$ is progressive.*

**(4.8) Proposition.** *An adapted process with right or left continuous paths is progressively measurable.*

# Three Definitions of Stochastic Processes

- 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 what kind of random variable / measurable map?

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

2. $X : \ \Omega \to \mathrm{Meas}(\mathscr{T}, \mathscr{X}), \qquad \omega \mapsto \big(t \mapsto X(\omega)(t)\big),$

3. $X : \ \Omega \times \mathscr{T} \to \mathscr{X}, \qquad (\omega, t) \mapsto X(\omega, t).$

# Problems

- Three different (partially inconsistent) definitions of stochastic processes
  - mismatch between formalization and meaning.
- Not clear how to turn $\mathrm{Meas}(\mathcal{T}, \mathcal{X})$ into a measurable space in itself?
  - existence of well-behaved $\sigma$-algebra unclear
- The measurability of $\omega \mapsto X_{T(\omega)}(\omega)$ only guaranteed under additional assumptions.

# Are Probabilistic Programs functional?

# A program that outputs a probabilistic program

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

- uncurried version:

```
print(prog_prob_prog(a=1))
```

```
<function prog_prob_prog.<locals>.<lambda>
```

```
def prob_prog(a,m,s):
    Z=np.random.uniform()
    return a*m+s*Z
```

```
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
```

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

```
1.5413066059310134
6.248544376268809
10.923467140491365
16.829697838821 6
20.7212242524388 4
```

# Another probabilistic program that outputs a probabilistic program

```
import numpy as np

def prob_prog_1(m=0,s=1):
    Z = np.random.normal()
    X = m+s*Z
    return(X)
```

```
def prob_prog_2(m=0,s=1,b=0):
    U = np.random.uniform()
    if U <= 0.33:
        return 'det fct', lambda x: x**2
    elif U <= 0.66:
        return 'output', b+prob_prog_1(m,s)
    else:
        return 'prob_prog_1', prob_prog_1
```

- Output:

```
for n in range(10):
    print(prob_prog_2(7,3,2))
```

```
('output', 13.334218167868446)
('output', 11.471183953689039)
('det fct', <function prob_prog_2.<locals>.<lambda> at 0x7
('prob_prog_1', <function prob_prog_1 at 0x7fd4f2820550>)
('prob_prog_1', <function prob_prog_1 at 0x7fd4f2820550>)
('prob_prog_1', <function prob_prog_1 at 0x7fd4f2820550>)
('prob_prog_1', <function prob_prog_1 at 0x7fd4f2820550>)
('output', 8.377835031430754)
('prob_prog_1', <function prob_prog_1 at 0x7fd4f2820550>)
('output', 11.669097814447499)
```

- How can one mathematically describe such a (probabilistic) program that outputs a probabilistic program, possibly of different types?

# How to formalize Probabilistic Programs?

- Probabilistic programs:
  - take input $x$,
  - sample internal random number $\omega$,
  - determine (stochastic) output $z$,
- so either:
  - measurable map: $K: \ \Omega \times \mathcal{X} \to \mathcal{Z}$,
    - however, $\omega \in \Omega$ not really an input, rather internal
  - Markov kernels from input space $\mathcal{X}$ to output space $\mathcal{Z}$,
    - measurable map $K: \ \mathcal{X} \to \mathcal{P}(\mathcal{Z})$
    - set of probabilistic programs: $\mathrm{Meas}\left(\mathcal{X}, \mathcal{P}(\mathcal{Z})\right)$.

# Church's Simply Typed $\lambda$-Calculus (1940)

- **Functional Programming** should satisfy "**curry**" / "**uncurry**" operations:
  - $(x, y) \mapsto f(x, y)$ corresponds 1:1 to: $x \mapsto (y \mapsto f(x, y))$
  - $(x, y) \mapsto g(x)(y)$ corresponds 1:1 to: $x \mapsto g(x) = (y \mapsto g(x)(y))$
- This mean a program in two (or more) variables $f(x, y)$ can be expressed as iteratively defining a functions in one variable $g(x)(y)$ and vice versa.
- Requires program-valued programs / function-valued functions.
- Realized in functional programming language Haskell.
- Mathematically corresponds to cartesian closed categories.


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

# Can we Curry / Uncurry Probabilistic Programs?

- Curry / Uncurry operations would translate to isomorphism:
  - $\mathrm{Meas}\left(\mathcal{X} \times \mathcal{Y}, \mathcal{P}\left(\mathcal{Z}\right)\right) \cong \mathrm{Meas}\left(\mathcal{X}, \mathrm{Meas}\left(\mathcal{Y}, \mathcal{P}(\mathcal{Z})\right)\right).$
- This means we need to be able to mathematically describe programs whose outputs are probabilistic programs.
- Furthermore, we need the operation $\mathcal{P}$ to be well-behaved:
  - functorial, respects product structure
    - strong probability monad

# Remark - Monad

- **Monad**
  - theory of functional programming with side effects
  - equivalent to category-theoretical construction in mathematics
- **strong monad**:
  - well-behaved w.r.t. products

# Definition - Monad

- A **Monad** on a category $\mathscr{C}$ is a triple $(\mathscr{P}, \delta, \mathbb{M})$ consisting of:
    - a functor $\mathscr{P}: \mathscr{C} \to \mathscr{C}$,
    - a natural transformation $\delta: \mathrm{id}_{\mathscr{C}} \to \mathscr{P}$,
    - a natural transformation $\mathbb{M}: \mathscr{P}^2 := \mathscr{P} \circ \mathscr{P} \to \mathscr{P}$,
  - such that:
    - $\mathbb{M} \circ \mathscr{P}\mathbb{M} = \mathbb{M} \circ \mathbb{M}\mathscr{P}$ as natural transformations $\mathscr{P}^3 \to \mathscr{P}$,
    - $\mathbb{M} \circ \mathscr{P}\delta = \mathbb{M} \circ \delta\mathscr{P} = \mathrm{id}_{\mathscr{P}}$ as natural transformations $\mathscr{P} \to \mathscr{P}$.
- A monad is called **strong**, if it is also "well-behaved" w.r.t. finite products $\times$.

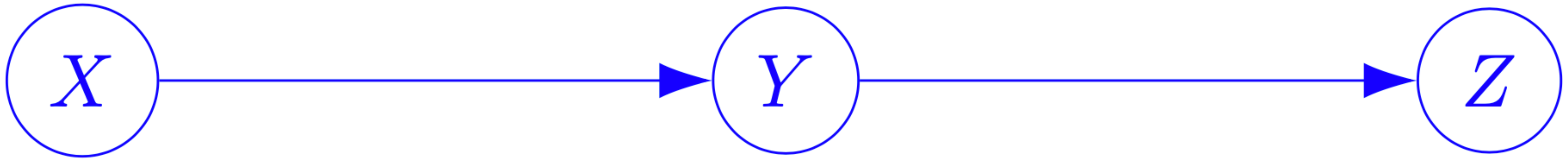
# Problems

- Define strong probability monad $(\mathscr{P}, \delta, \mathbb{M})$
  - Giry monad defined on category of measurable spaces $\mathrm{Meas}$.
- The set of all programs that output probabilistic programs should be:
  - $\mathrm{Meas}\left(\mathscr{X}, \mathrm{Meas}\left(\mathscr{Y}, \mathscr{P}(\mathscr{Z})\right)\right)$
- Not clear how to turn $\mathrm{Meas}\left(\mathscr{Y}, \mathscr{P}(\mathscr{Z})\right)$ into a measurable space in itself
  - existence of well-behaved $\sigma$-algebra unclear
- Is is possible to do more complicated constructions, e.g. dependent products, etc,?
  - Can we get a dependent type theory together with higher-order probability theory?

# Can we do Graphical Reasoning between Random Variables and Mechanisms?

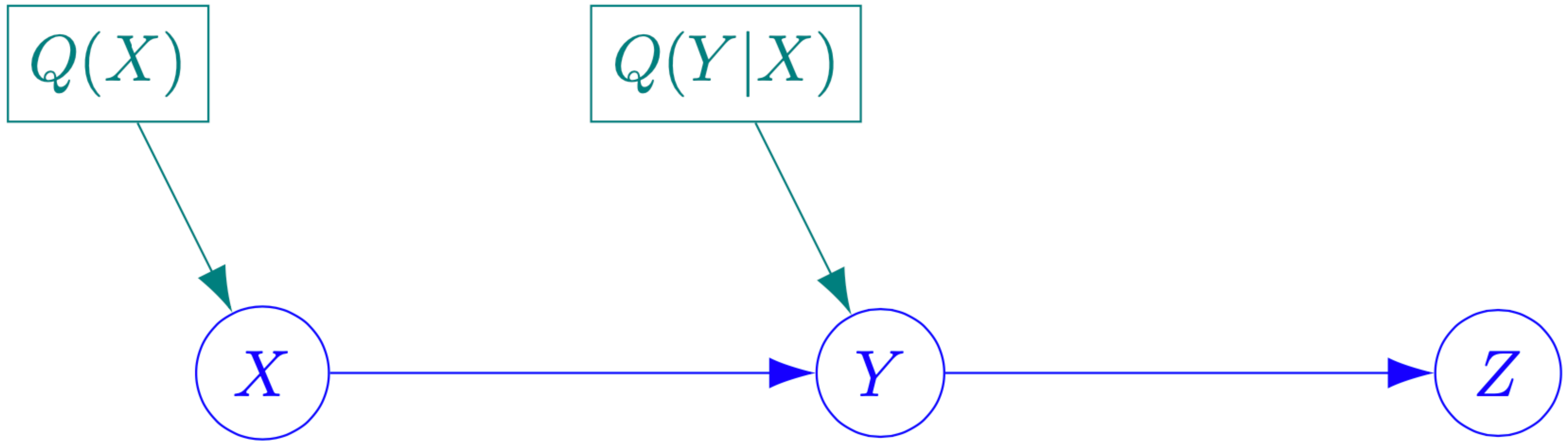
# Conditional Independence in Probabilistic Graphical Models

- Consider a **Markov chain**:

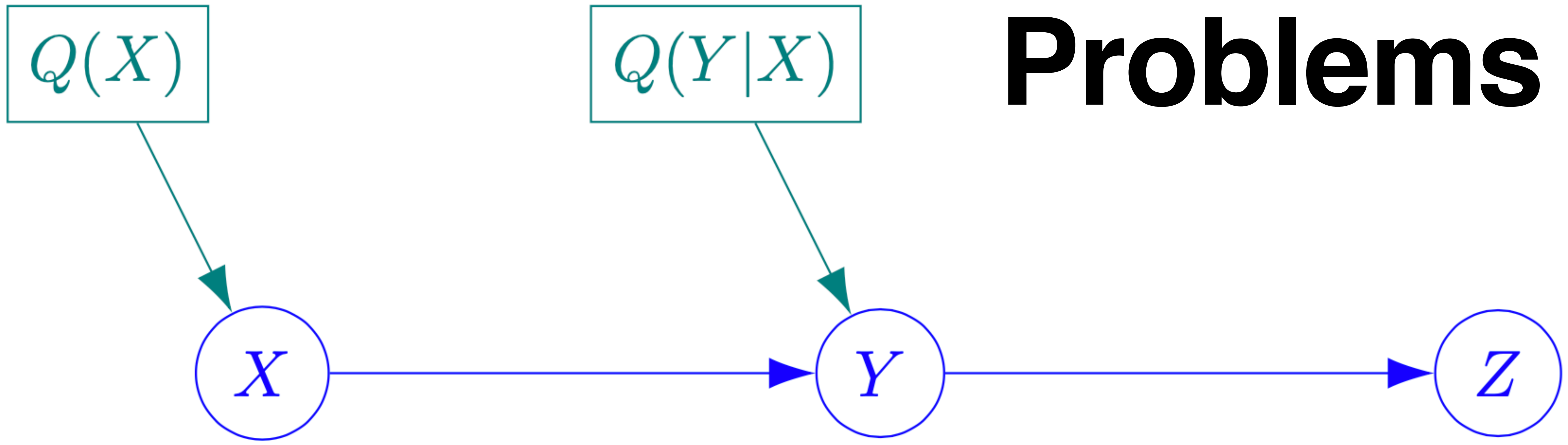



- We have:
  - factorization: $P(X, Y, Z) = P(Z \mid Y) \otimes P(Y \mid X) \otimes P(X)$
  - tells us that $Z$ is only dependent on $Y$, and, independent of $X$ when conditioned on $Y$, but then also of the choice of $P(Y \mid X)$ and $P(X)$.
- We want to be able to:
  - formalize conditional independence: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




# Problems

- Problems:
  - Not clear how to deal with non-random variables.
  - Not clear how to turn $\mathscr{L} := \mathrm{Meas}\left(\mathscr{X}, \mathscr{P}(\mathscr{Y})\right)$ into a measurable space
  - Not clear how to define conditional independence with the two problems above.
  - Not clear if this corresponds to graphical conditional independence criteria.

# How to formalize standard 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: $X \perp\!\!\!\perp (Y_x)_{x \in \mathcal{X}} \,|\, Z$,
  - Consistency: $Y = Y_X$ a.s.

# Problems

- 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.
- Not immediate clear how to fix this.

# Explanation

- Let $\mathcal{X} = \mathbb{R}$, $\mathcal{Y}_x := \mathcal{Y} := \{0,1\}$, then the following map is **not measurable**:
  - $e : \mathcal{X} \times \prod_{x \in \mathcal{X}} \mathcal{Y} \to \mathcal{Y}, \quad \left(\tilde{x}, (y_x)_{x \in \mathcal{X}}\right) \mapsto y_{\tilde{x}}.$
- Otherwise, we had: $D := e^{-1}(1) \in \mathcal{B}_{\mathcal{X}} \otimes \bigotimes_{x \in \mathcal{X}} \mathcal{B}_{\mathcal{Y}_x}.$
- Then $D$ lies in a sub-$\sigma$-algebra generated by only countably many cylinder sets.
- So there exists countable subset: $\mathcal{C} \subseteq \mathcal{X}$ s.t.: $D = B \times \prod_{x \in \mathcal{X} \setminus \mathcal{C}} \mathcal{Y}_x$ with $B \subseteq \mathcal{X} \times \prod_{x \in \mathcal{C}} \mathcal{Y}_x.$
- For $x \in \mathcal{X} \setminus \mathcal{C}$: $e(x, 0_{\mathcal{C}}, 0,.,0, 1_x, 0,.,0) = 1$, so $(x, 0_{\mathcal{C}}, 0,.,0, 1_x, 0,.,0) \in D$, so $(x, 0_{\mathcal{C}}) \in B$.
- But then: $(x, 0_{\mathcal{C}}, 0,.,0, 0_x, 0,.,0) \in D = e^{-1}(1)$ and thus: $e(x, 0_{\mathcal{C}}, 0,.,0, 0_x, 0,.,0) = 1$.
- but clearly: $e(x, 0_{\mathcal{C}}, 0,.,0, 0_x, 0,.,0) = 0$, which is a contradiction.

# How to formalize 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 $x \mapsto Y_x$.

# If we are going to change all of this are we still able to do standard thing in Probability Theory and (Bayesian) Statistics?

# Bad News

# Random Functions do not exist in $\mathrm{Meas}$

- <u>Theorem</u> (Aumann, 1961):
  - There is **no** $\sigma$-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-$\sigma$-algebra and $\mathscr{L}$ is the space of all measurable maps from $\mathbb{R}$ to $\mathbb{R}$, and the product carries the product-$\sigma$-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.

# Quasi-Measurable Spaces

# Recall: Usual measure-theoretic approach

- **Sample space** is a measurable space: $(\Omega, \mathscr{B}_\Omega)$
    - where $\mathscr{B}_\Omega$ is the $\sigma$-algebra/set of admissible outcome events on $\Omega$.
- **State space** is a **measurable space**: $(\mathscr{X}, \mathscr{B}_\mathscr{X})$
    - where $\mathscr{B}_\mathscr{X}$ is another **$\sigma$-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: $P(X)$).

# Main Idea behind Quasi-Measurable Spaces

- <u>Main idea</u>: *Exchange the role of $\sigma$-algebras and random variables!!!*
- **Sample space** is a measurable space: $\left(\Omega, \mathscr{B}_\Omega\right)$
  - where $\mathscr{B}_\Omega$ is the $\sigma$-algebra/set of admissible outcome events on $\Omega$.
- **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**.
- **$\sigma$-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** $(\Omega, \Omega^\Omega)$ consists of:
    - a set: $\Omega$
    - a set of maps: $\Omega^\Omega \subseteq \{\Phi : \Omega \to \Omega\}$
        - such that:
            - $\mathrm{id}_\Omega \in \Omega^\Omega$,
            - $\Omega^\Omega$ contains all constant maps,
            - $\Omega^\Omega$ 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 $\sigma$-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 $\Omega^{\Omega}$:
        - $X \in \mathscr{X}^{\Omega}, \Phi \in \Omega^{\Omega} \implies X \circ \Phi \in \mathscr{X}^{\Omega}$.

# Quasi-Measurable Maps

- Let $\left(\mathscr{Z}, \mathscr{Z}^{\Omega}\right)$ and $\left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$ two quasi-measurable spaces.
- A map $g: \mathscr{Z} \to \mathscr{X}$ is called **quasi-measurable** if
  - $Z \in \mathscr{Z}^{\Omega} \implies g(Z) := g \circ Z \in \mathscr{X}^{\Omega}$
- The set of all quasi-measurable maps is abbreviated:
  - $\mathrm{QMS}\left(\left(\mathscr{Z}, \mathscr{Z}^{\Omega}\right), \left(\mathscr{X}, \mathscr{X}^{\Omega}\right)\right)$ or $\mathrm{QMS}\left(\mathscr{Z}, \mathscr{X}\right)$ 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**: $\mathrm{QMS}$.

# 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 $\left(\mathcal{X}, \mathcal{X}^{\Omega}\right)$ and $\left(\mathcal{Z}, \mathcal{Z}^{\Omega}\right)$ two quasi-measurable spaces. We put:
  - $\mathcal{X}^{\mathcal{Z}} := \mathrm{QMS}\left(\left(\mathcal{Z}, \mathcal{Z}^{\Omega}\right), \left(\mathcal{X}, \mathcal{X}^{\Omega}\right)\right)$
  - $\left(\mathcal{X}^{\mathcal{Z}}\right)^{\Omega} := \left\{X :\ \Omega \to \mathcal{X}^{\mathcal{Z}} \mid \left((\omega, z) \mapsto X(\omega)(z)\right) \in \mathrm{QMS}(\Omega \times \mathcal{Z}, \mathcal{X})\right\}$
    - function-valued random variables are defined via the product structure
- Then $\left(\mathcal{X}^{\mathcal{Z}}, \left(\mathcal{X}^{\mathcal{Z}}\right)^{\Omega}\right)$ 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 $\mathrm{Meas}$ for measurable spaces!!!

# More Category-theoretical Constructions

- Similary, we can define the following in $\mathrm{QMS}$:
  - **coproducts**, **equalizers**, **coequalizers**, thus:
  - all small **limits** and all small **colimits**
- even more, we get in $\mathrm{QMS}$:
  - **fibre products**: $\mathscr{X} \times_{\mathscr{S}} \mathscr{Y} \to \mathscr{S}$,
  - **internal homs**: $\mathscr{Q}_{\mathscr{S}}(\mathscr{X}, \mathscr{Y}) \to \mathscr{S}$
    - $\mathrm{QMS}_{\mathscr{S}}(\mathscr{X} \times_{\mathscr{S}} \mathscr{Y}, \mathscr{Z}) = \mathrm{QMS}_{\mathscr{S}}(\mathscr{X}, \mathscr{Q}_{\mathscr{S}}(\mathscr{Y}, \mathscr{Z}))$.
- Note that the latter was not possible in $\mathrm{Meas}$ for measurable spaces!!!

# Main Theorems

- Theorem: The category of quasi-measurable spaces $\mathrm{QMS}$ forms a **quasitopos**, and, is in particular, **locally cartesian closed**.
- Remark: This means that, besides **simply typed $\lambda$-calculus**, we get a **dependent type theory** for $\mathrm{QMS}$. Roughly speaking, this means that we can model programs that can vary the output type/space dependent on the input. This makes it easy to implement all result obtained inside $\mathrm{QMS}$ in a theorem prover like Lean, Agda or Coq, etc.

- Theorem: The category of quasi-measurable spaces $\mathrm{QMS}$ forms a **Heyting category**.
- Remark: $\mathrm{QMS}$ has thus an internal logic of a (typed) **intuitionistic first-order logic**.

- Remark: Note that most of this is not true for the category of measurable spaces $\mathrm{Meas}$!!!

# 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.

# The Sample Space - Act 2 - The $\sigma$-Algebra

- We now endow the Sample Space $(\Omega, \Omega^\Omega)$ with an additional **$\sigma$-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 $\sigma$-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 $\sigma$-Algebra

- Let $\left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$ be a quasi-measurable space.
- Then the **induced $\sigma$-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:
    - $\Omega \to [0,1], \quad \omega \mapsto P(D^\omega),$ 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 $\left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$ be a quasi-measurable space. Define:
  - $\mathscr{P}(\mathscr{X}) := \mathscr{P}(\mathscr{X}, \mathscr{X}^{\Omega}) := \left\{P(X) : \mathscr{B}_{\mathscr{X}} \to [0,1] \mid X \in \mathscr{X}^{\Omega}, P \in \mathscr{P}\right\}$
  - $\mathscr{P}(\mathscr{X})^{\Omega} := \mathscr{P}(\mathscr{X}, \mathscr{X}^{\Omega})^{\Omega} := \left\{P(X \mid I) \mid X \in (\mathscr{X}^{\Omega})^{\Omega}, P \in \mathscr{P}\right\}$
  - $P(X \in A \mid I = \omega) := P\left(\left\{\tilde{\omega} \in \Omega \mid X(\omega)(\tilde{\omega}) \in A\right\}\right)$ for $A \in \mathscr{B}_{\mathscr{X}}$

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

# The Spaces of Markov Kernels and Random Functions

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

# Some surprising Lemmata

- Let $\left(\mathcal{X}, \mathcal{X}^{\Omega}\right)$ and $\left(\mathcal{Y}, \mathcal{Y}^{\Omega}\right)$ be quasi-measurable spaces.
- Then the following maps are all quasi-measurable:
  - $\mathcal{Y}^{\mathcal{X}} \times \mathcal{B}_{\mathcal{Y}} \to \mathcal{B}_{\mathcal{X}}, \qquad (f, B) \mapsto f^{-1}(B).$
  - $\mathcal{P}(\mathcal{X}) \times \mathcal{B}_{\mathcal{X}} \to [0,1], \qquad (P, A) \mapsto P(A).$
  - $\mathcal{Y}^{\mathcal{X}} \times \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{Y}), \qquad (f, P) \mapsto f_* P.$
  - $[0,\infty]^{\mathcal{X}} \times \mathcal{P}(\mathcal{X}) \to [0,\infty], \quad (h, P) \mapsto \int h(x)\, P(dx).$
- 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 $\mathrm{QMS}$, 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 $\lambda$-calculus.
- We thus get semantics for higher-order probability theory for probabilistic programming language.

# Construction of well-behaved Sample Spaces

- Theorem: Let $\Omega_0$ be a set, and:
  - $\mathscr{E}_0$ a countable set of subsets of $\Omega_0$ that separates the points of $\Omega_0$.
  - $\Omega := \prod_{n \in \mathbb{N}} \Omega_0,$ 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 $\mathscr{E}$,
  - $\Omega^\Omega := \mathrm{Meas}\big((\Omega, \mathscr{B}_\Omega), (\Omega, \mathscr{B}_\Omega)\big), \quad \mathscr{P} := \tilde{\mathscr{P}}|_{\mathscr{B}_\Omega}$
- 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 $\Omega$.
  - Note that this is bigger than the Borel $\sigma$-algebra on $\Omega$.
- $\mathscr{P} =$ all probability measures on $\mathscr{B}_\Omega$, $\quad \Omega^\Omega = \mathrm{Meas}\left((\Omega, \mathscr{B}_\Omega), (\Omega, \mathscr{B}_\Omega)\right)$.
- 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** $\left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$ is - per definition - just a quasi-measurable space where the sample space $\Omega$ is the **universal Hilbert cube**.
- We abbreviate the category of quasi-universal spaces as $\mathrm{QUS}$.

# 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:
    - $\imath : \left(\mathscr{X}, \mathscr{X}^{\Omega}\right) \to \left(\Omega, \Omega^{\Omega}\right)$ and $r : \left(\Omega, \Omega^{\Omega}\right) \to \left(\mathscr{X}, \mathscr{X}^{\Omega}\right)$ s.t.:
      - $r \circ \imath = \mathrm{id}_{\mathscr{X}}$.

# 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 $Q_n(X_{0:n} \,|\, Z) \in \mathscr{P}\left(\mathscr{X}_{0:n}\right)^{\mathscr{Z}}$ such that for every $n \in \mathbb{N}$:
    - $\mathrm{pr}_{0:n,*} Q_{n+1}(X_{0:n+1} \,|\, Z) = Q_n(X_{0:n} \,|\, Z)$.
- Then there exists a unique $Q(X_{\mathbb{N}} \,|\, Z) \in \mathscr{P}\left(\mathscr{X}_{\mathbb{N}}\right)^{\mathscr{Z}}$ such that:
    - $\mathrm{pr}_{0:n,*} Q(X_{0:n+1} \,|\, Z) = Q_n(X_{0:n} \,|\, Z) \quad$ 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 $Q(X_{\mathbb{N}} \,|\, Z) \in \mathscr{P}\left(\mathscr{X}^{\mathbb{N}}\right)^{\mathscr{Z}}$ the following is equivalent:
  - $Q(X_{\mathbb{N}} \,|\, Z)$ is **exchangable**, i.e. invariant under all finite permuations: $\rho : \mathbb{N} \cong \mathbb{N}$.
  - There exists a quasi-universal space $\mathscr{Y}$ and $K(X \,|\, Y) \in \mathscr{P}(\mathscr{X})^{\mathscr{Y}}$ and $P(Y \,|\, Z) \in \mathscr{P}(\mathscr{Y})^{\mathscr{Z}}$ such that :
    - $$Q(X_{\mathbb{N}} \,|\, Z) = \left( \bigotimes_{n \in \mathbb{N}} K(X_n \,|\, Y) \right) \circ P(Y \,|\, Z).$$
- In this case we can w.l.o.g. take: $\mathscr{Y} = \mathscr{P}(\mathscr{X})$ and $K(X \in A \,|\, Y = P) := P(A)$.

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

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: $P_v(X_v | X_{\mathrm{Pa}^G(v)}) \in \mathscr{P}(\mathcal{X}_v)^{\mathcal{X}_{\mathrm{Pa}^G(v)}}$ for $v \in V \backslash W$.

# 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: $P(X_V, X_J, X_{I_W} \,|\, X_J, X_{I_W})$.

# 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

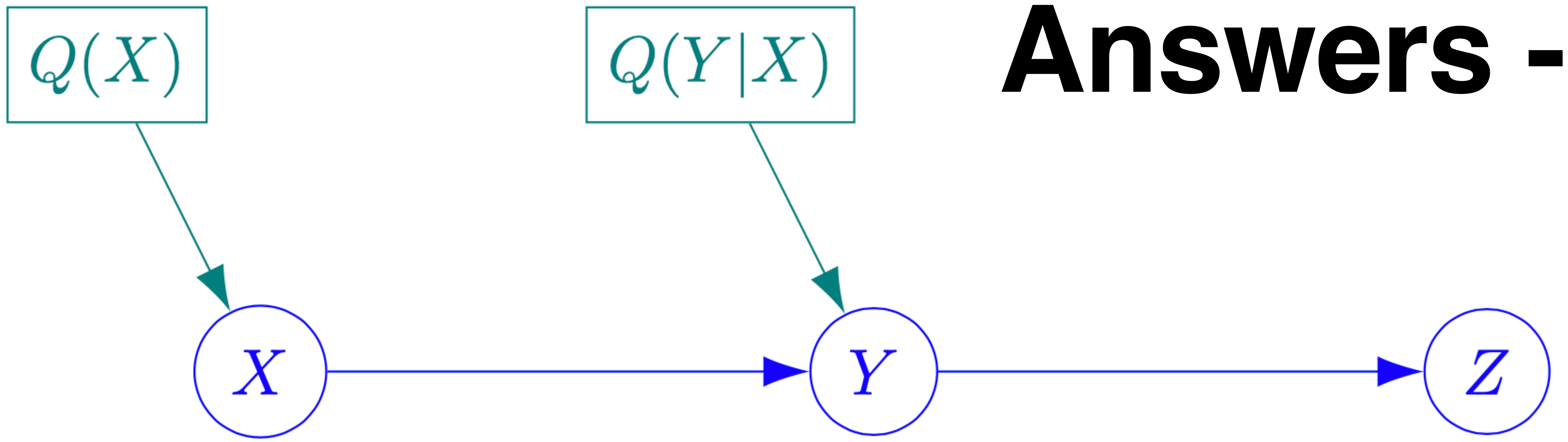
# Answers - Stochastic Process

- <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 $x \in \mathscr{X}$ and output $z \in \mathscr{Z}$ is quasi-measurable map: $\mathscr{X} \to \mathscr{P}(\mathscr{Z})$.
- Theorem: We have the natural **curry / uncurry isomorphism**:
  - $\mathrm{QMS}\left(\mathscr{X} \times \mathscr{Y}, \mathscr{P}\left(\mathscr{Z}\right)\right) \cong \mathrm{QMS}\left(\mathscr{X}, \mathrm{QMS}\left(\mathscr{Y}, \mathscr{P}(\mathscr{Z})\right)\right)$
- Theorem: $\mathrm{QMS}$ is a **quasitopos**, thus allows for **dependent type theory**.
- Theorem: The triple $(\mathscr{P}, \delta, \mathbb{M})$ forms a **strong probability monad** on the category of quasi-measurable spaces $\mathrm{QMS}$ (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.
- <u>Theorem</u>: **Global Markov Property**: For $A, B, C \subseteq V \cup I_W \cup J$ we have:
  - $A \perp B \,|\, C \implies X_A \perp\!\!\!\perp X_B \,|\, X_C$.
- <u>Example</u>: 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:
    - $P(G \mid X) \in \mathscr{P}(\mathscr{G})^{\mathscr{X}}$ such that $P(G, X) = P(G \mid X) \otimes P(X)$.
- Evaluation maps and push-forwards are quasi-measurable, which implies:
    - $C(A \mid x, x') := P(G(x) \in A \mid X = x')$ 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 $\mathrm{Meas}$ —> categorical construction in $\mathrm{QMS}$
- 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!