MARKOV CATEGORIES

A TUTORIAL

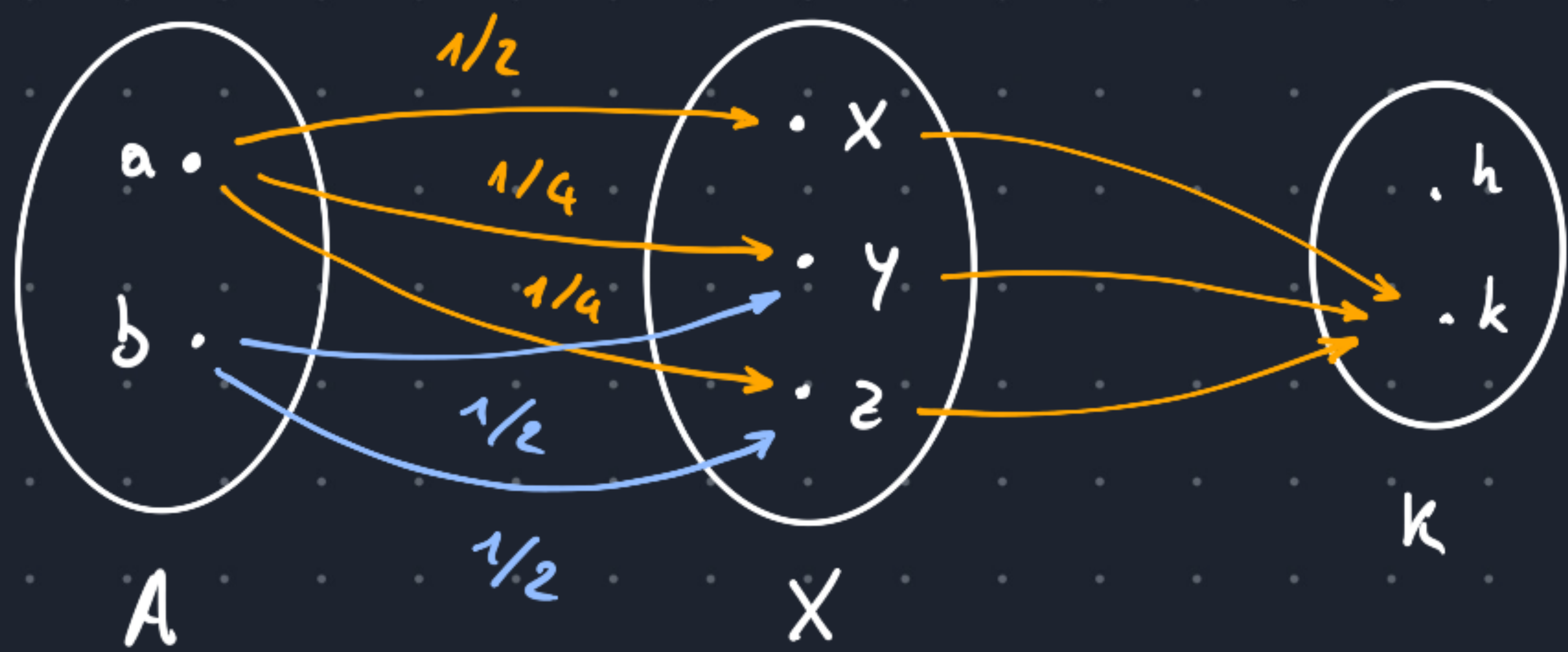
Applied Category Theory 2023

PAOLO PERRONE (University of Oxford)

Basic idea

Morphisms with "randomness"!

Example. The category **FinStoch** whose morphisms are stochastic matrices.



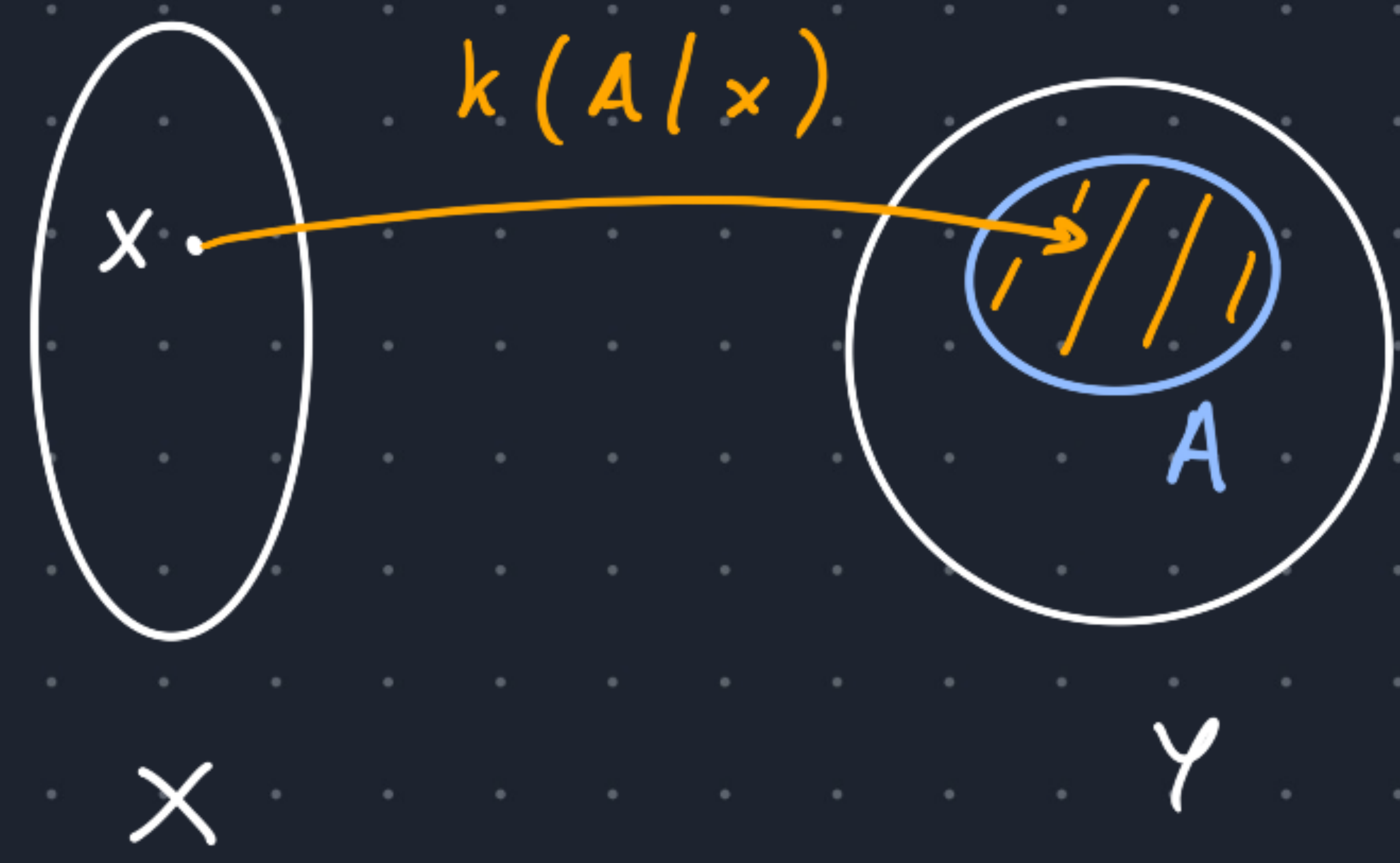
↙	a	b
x	1/2	0
y	1/4	1/2
z	1/4	$\boxed{1/2}$

$f(z|b)$

$$g \circ f(k|a) = \sum_{x \in X} g(k|x)\, f(x|a) \qquad \text{(Chapman-Kolmogorov)}$$

$1 \xrightarrow{p} X$ "states" = prob. measures (of finite support)

Example. The category **Stoch** whose morphisms are **Markov kernels**.



$$X \times \Sigma_Y \longrightarrow [0,1]$$

$$(x, A) \longmapsto k(A|x)$$

prob. measure

measurable

$$h \circ k\,(B|x) = \int_Y h(B|y)\, k(dy|x)$$

$$X \xrightarrow{f} Y \qquad K_f(A|x) := \begin{cases} 1 & f(x) \in A \\ 0 & f(x) \notin A \end{cases} \qquad \mathbf{Meas} \xrightarrow{K} \mathbf{Stoch}$$

Main definition

A *Markov category* is a symmetric monoidal category $(\mathcal{C}, \otimes, I)$, where each object X is equipped with maps

$$X \xrightarrow{\text{copy}} X \otimes X \qquad X \xrightarrow{\text{del}} I$$

such that

Without this: "GS" or "CD" category

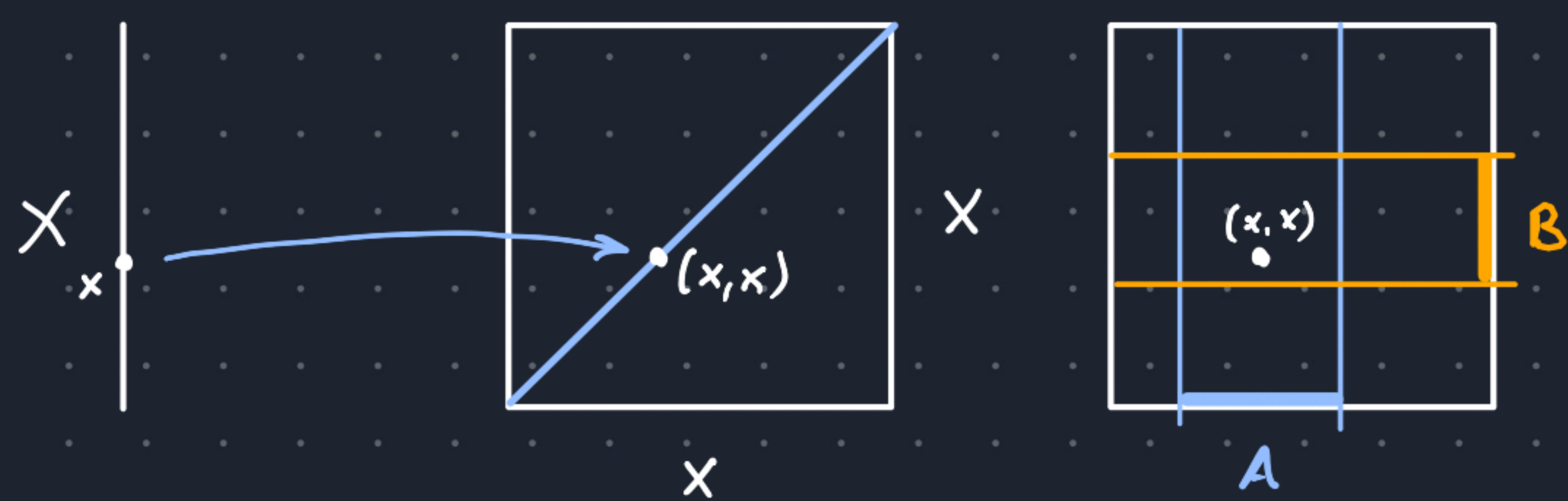
Example. In FinStoch, $X \xrightarrow{\text{copy}} X \otimes X$ for $X = \{x, y\}$ is:

X
x
y

$X \otimes X$	x	y
x	$\boxed{(x,x)}$	(x,y)
y	(y,x)	$\boxed{(y,y)}$

$\swarrow$	x	y
(x,x)	1	0
(x,y)	0	0
(y,x)	0	0
(y,y)	0	1

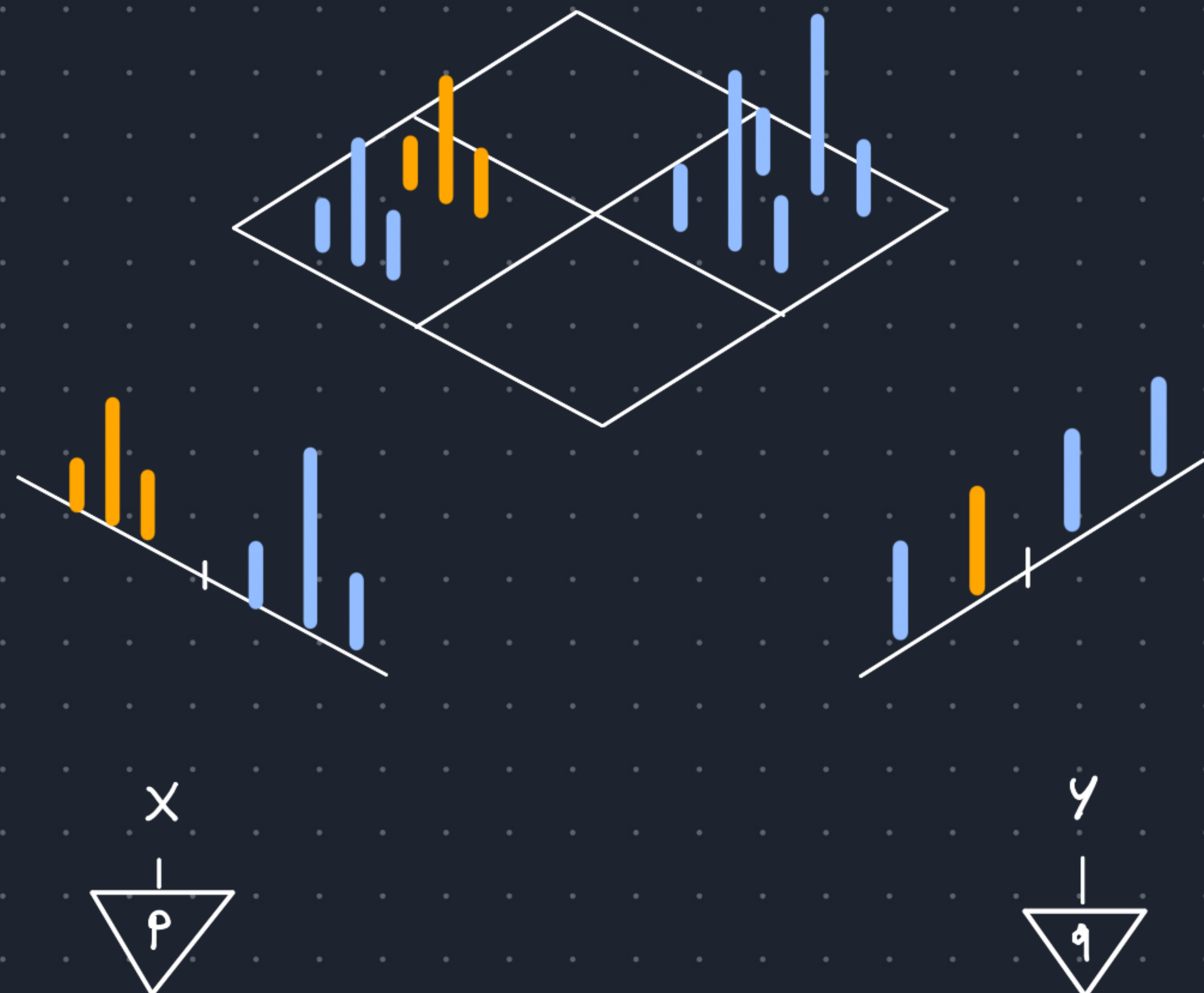
Example. In Stoch, more generally,



$$(x,x) \in A \times B \iff x \in A \cap B$$

$$\text{copy}(A \times B \mid x) = \begin{cases} 1 & x \in A \cap B \\ 0 & x \notin A \cap B \end{cases}$$

Joints & Marginals



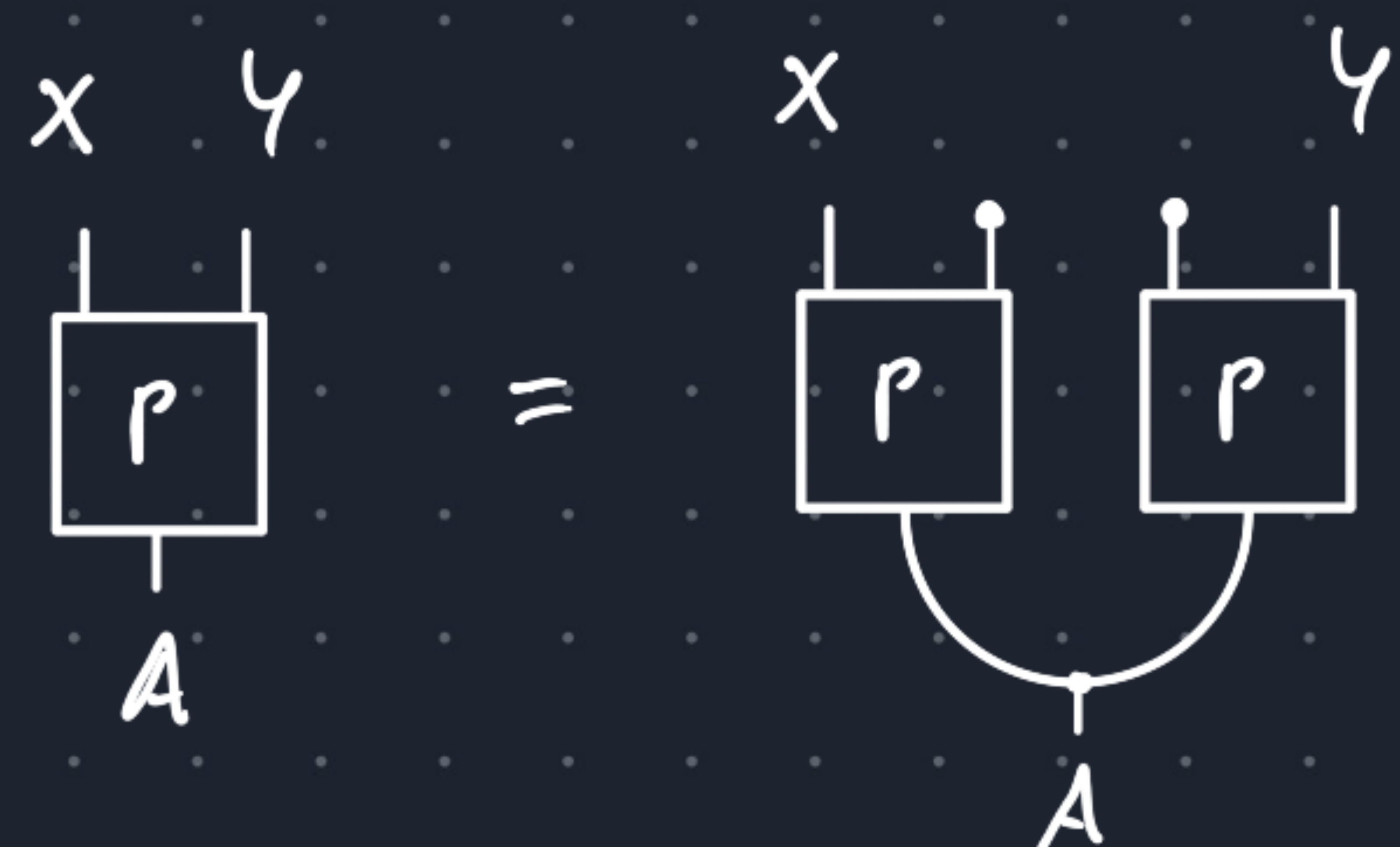
$$\sum_y r(x,y) = p(x)$$

(same for y)

Stochastic independence

p exhibits independence of X, Y

$$p(x,y) = p(x)\,p(y)$$



h exhibits conditional independence of X, Y given A.

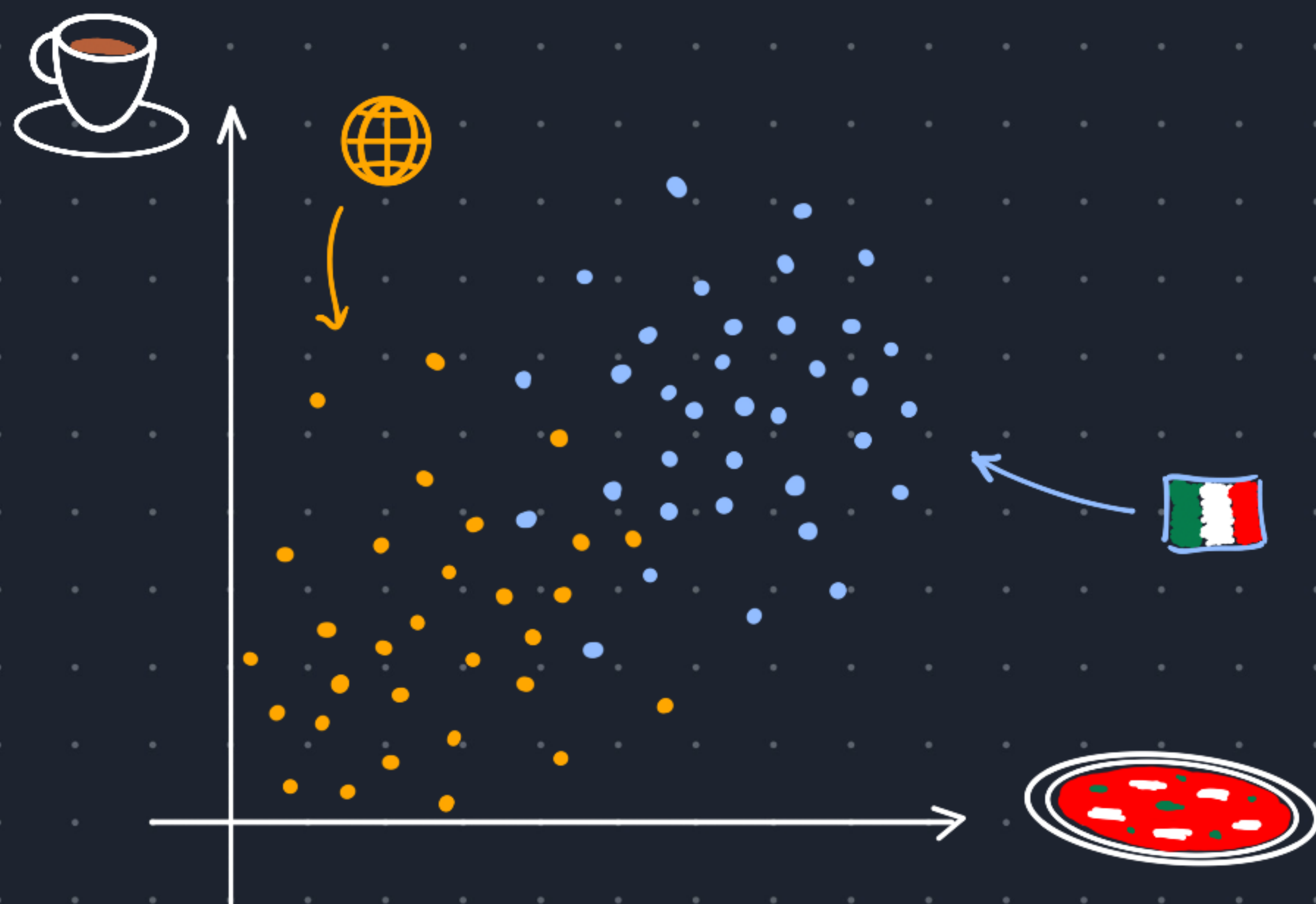
$$p(x,y \mid a) = p(x \mid a)\; p(x \mid a)$$

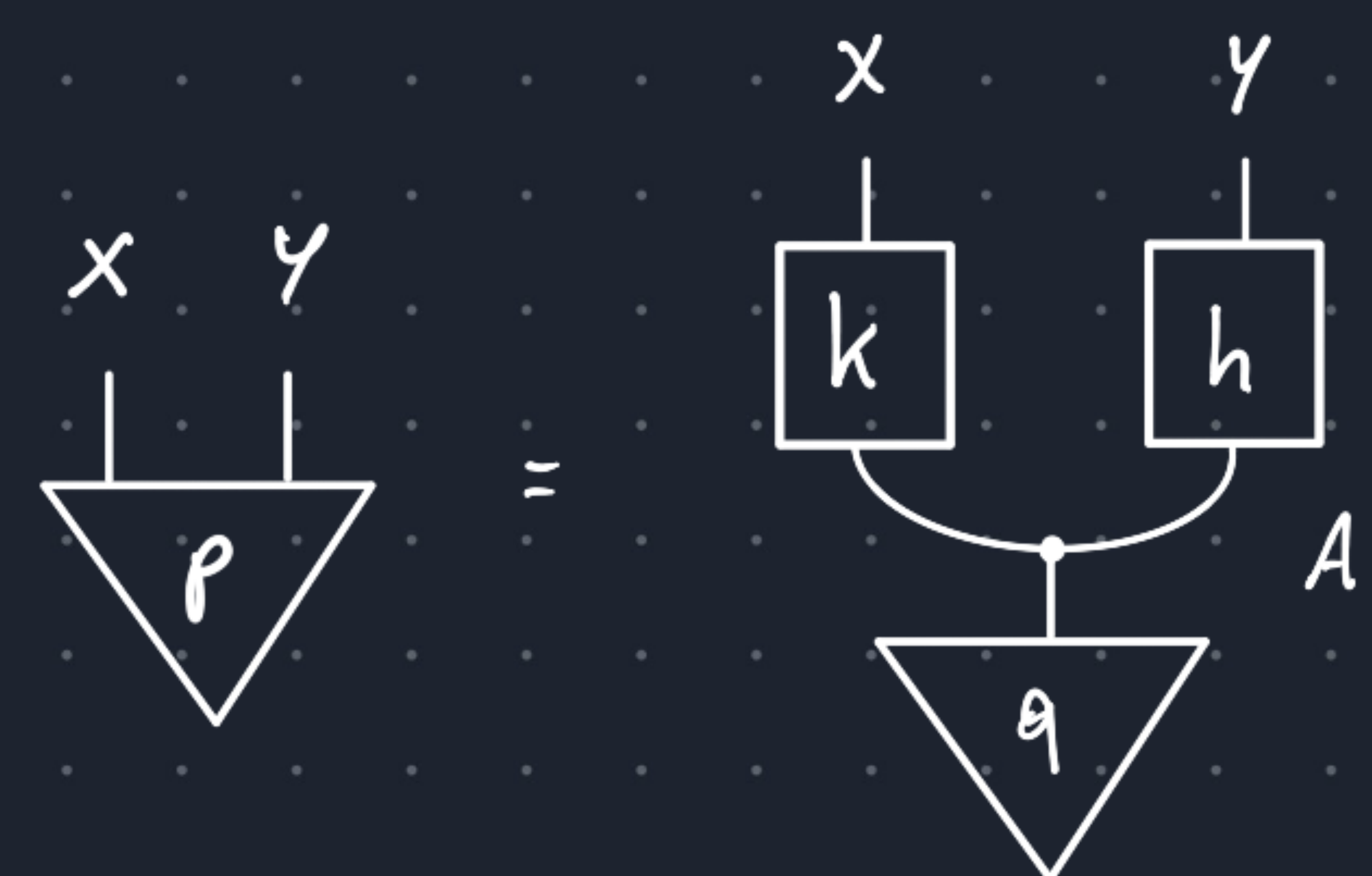
(See Exercise 1.1 later for more.)

Stochastic independence

p exhibits independence of X, Y

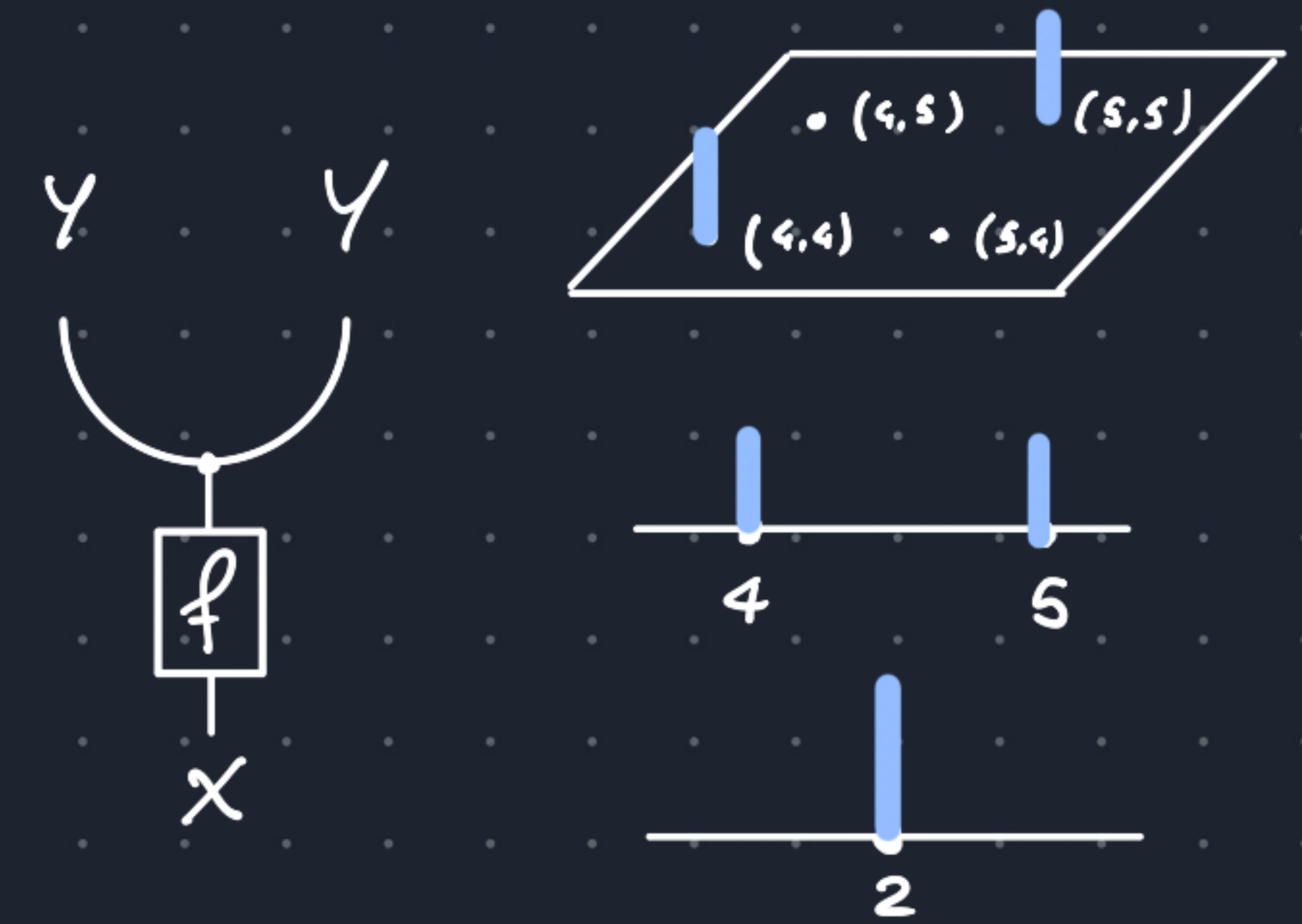
$$p(x,y) = p(x)\,p(y)$$

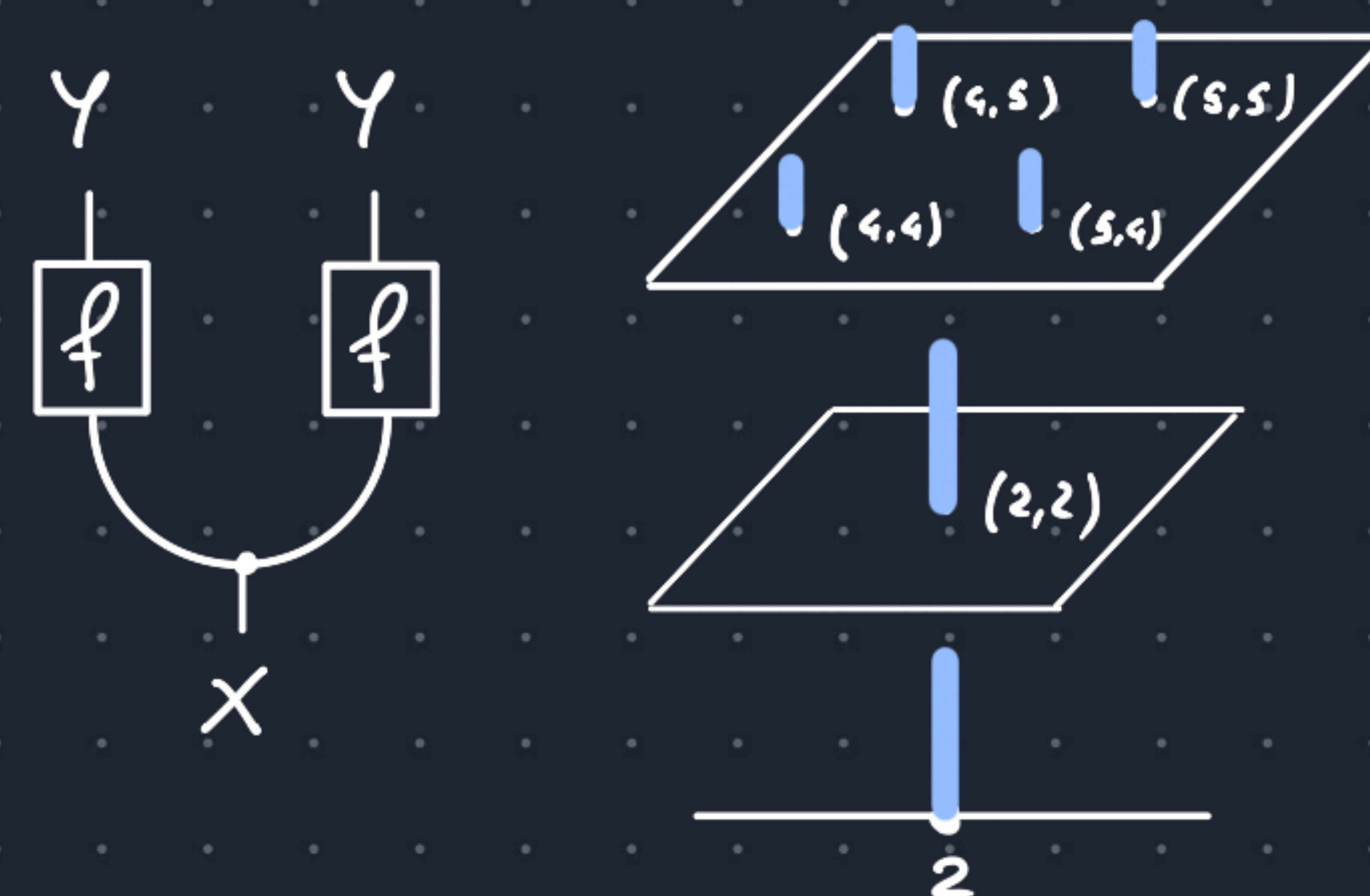




Determinism (a.k.a. copyability)

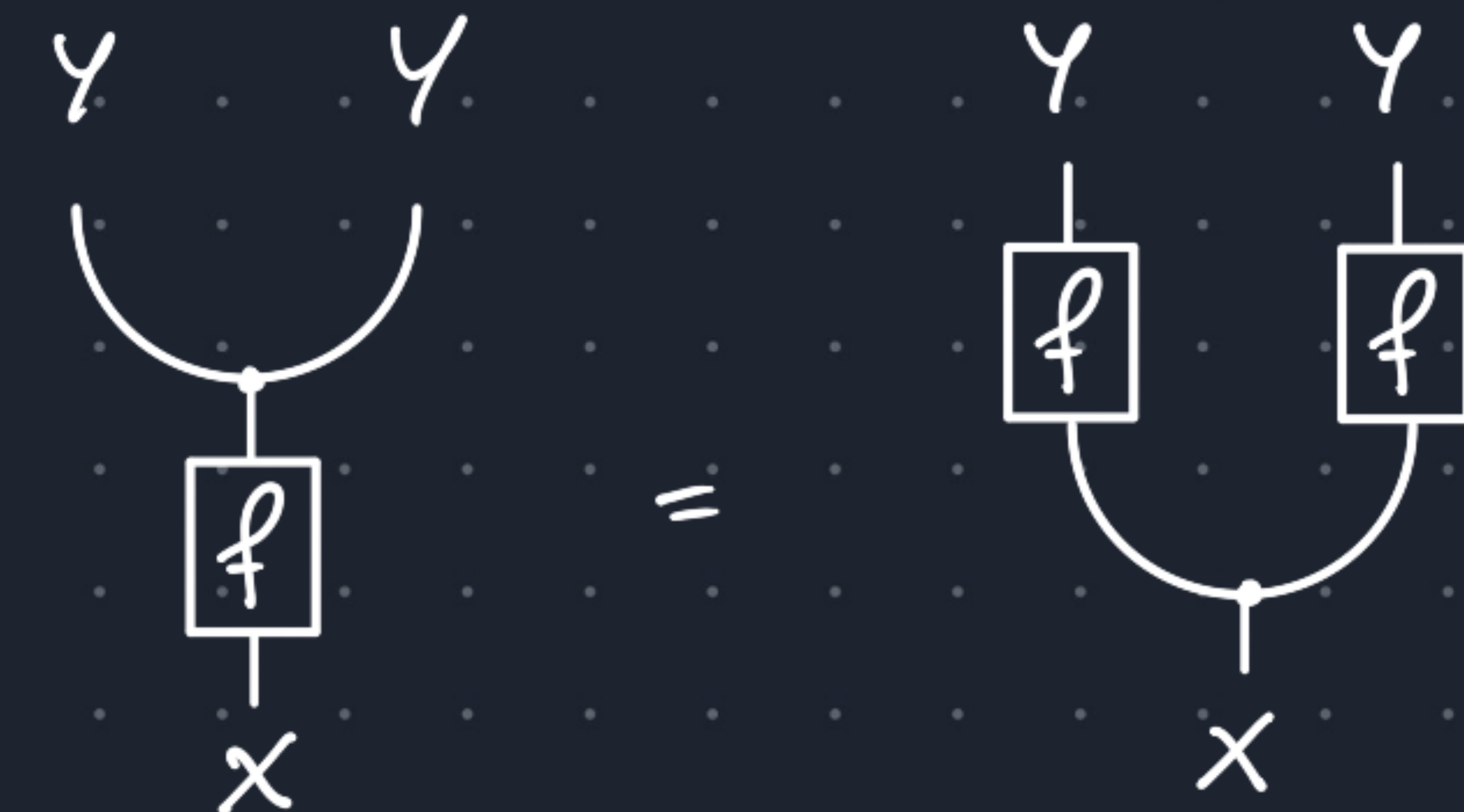
is called *deterministic* if





Determinism (a.k.a. copyability)

is called *deterministic* if



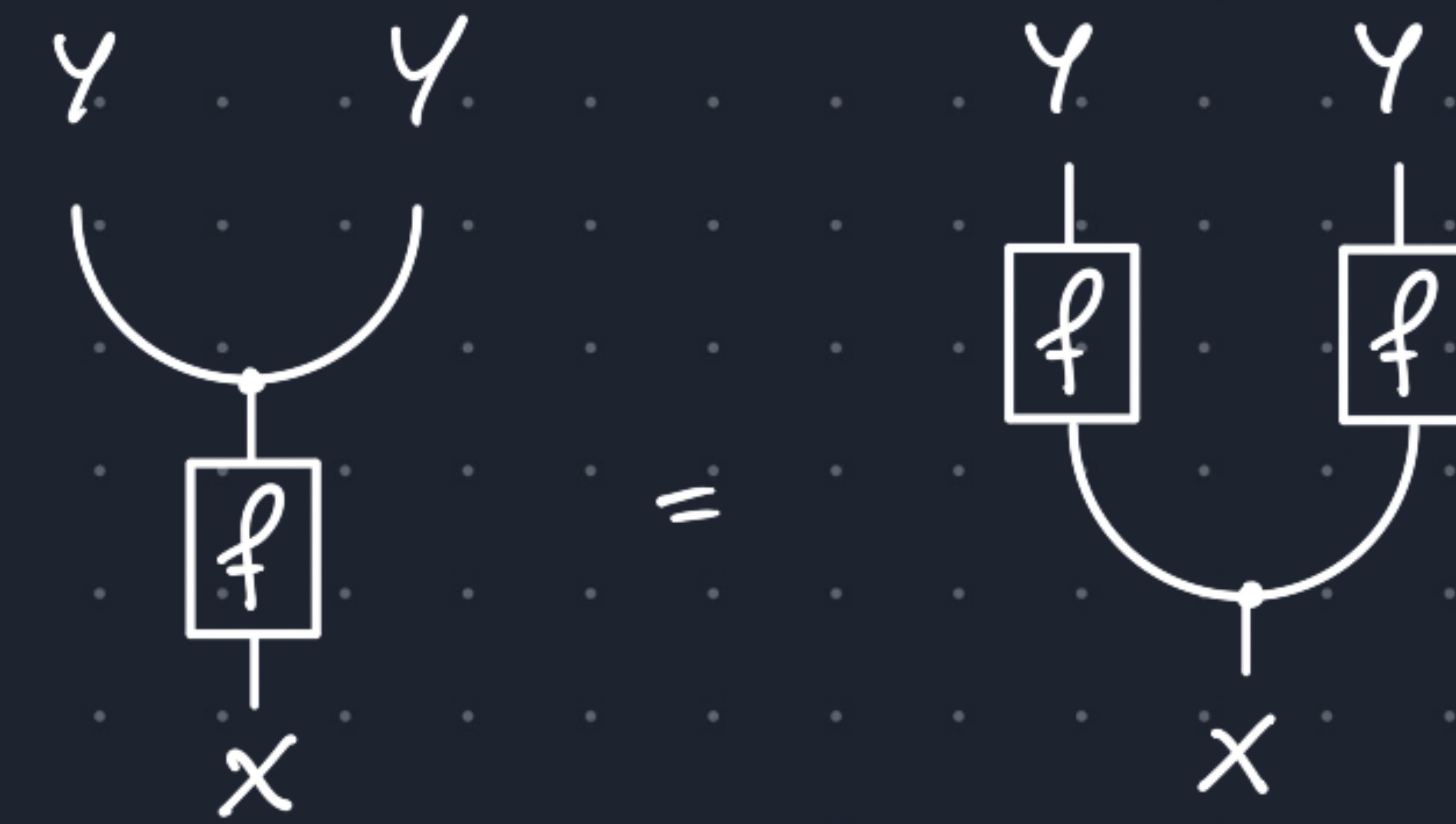
Example.

- In FinStoch, a matrix is deterministic iff all its entries are $f(y|x) = 0$ or 1.
- In Stoch, similarly, $f(A|x) = 0$ or 1.

  (See Exercise 1.2 later.)

Determinism (a.k.a. copyability)

is called *deterministic* if



- The category **BorelStoch** is the subcategory of *standard Borel spaces*.

  (e.g. finite & countable sets, $\mathbb{R} \cong [0,1] \cong \mathbb{R}^n$, etc.)

  In this category, deterministic morphisms are just the measurable functions.

$$K_f(A\,|\,x) := \begin{cases} 1 & f(x) \in A \\ 0 & f(x) \notin A \end{cases}$$

Determinism (a.k.a. copyability)

Proposition. For a Markov category $\mathcal{C}$, TFAE:

1) Every morphism is deterministic;
2) The copy maps are a natural transformation;
3) $\mathcal{C}$ is cartesian monoidal $(\otimes = \times)$

Markov = cartesian + randomness!

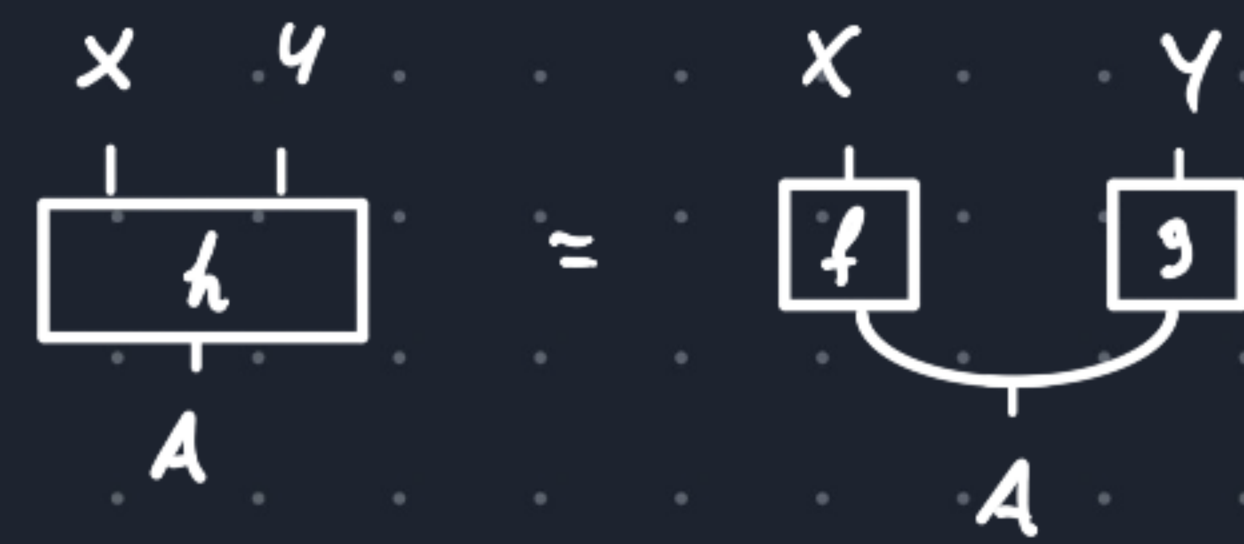
cartesian = Markov + determinism

$$\begin{tikzcd} & A \arrow[dl, bend right] \arrow[d, dashed] \arrow[dr, bend left] & \\ X & X \times Y \arrow[l] \arrow[r] & Y \end{tikzcd}$$

Stochastic interaction is a feature of randomness.

Exercises :

1.1. Show that if a joint morphism decomposes as a product, i.e.



then it is the product of its marginals, i.e.

1.2. Show that the deterministic morphisms of FinStoch are exactly the matrices of entries $\{0, 1\}$.
What's the analogous statement in Stoch ?

1.3. Prove

Proposition. For a Markov category $\mathcal{C}$, TFAE :

1) Every morphism is deterministic ;
2) The copy maps are a natural transformation;
3) $\mathcal{C}$ is cartesian monoidal ($\otimes = \times$)

Hint : show that if a morphism

is deterministic, then it is always making X and Y conditionally independent given A.

Almost-sure equality

Given 

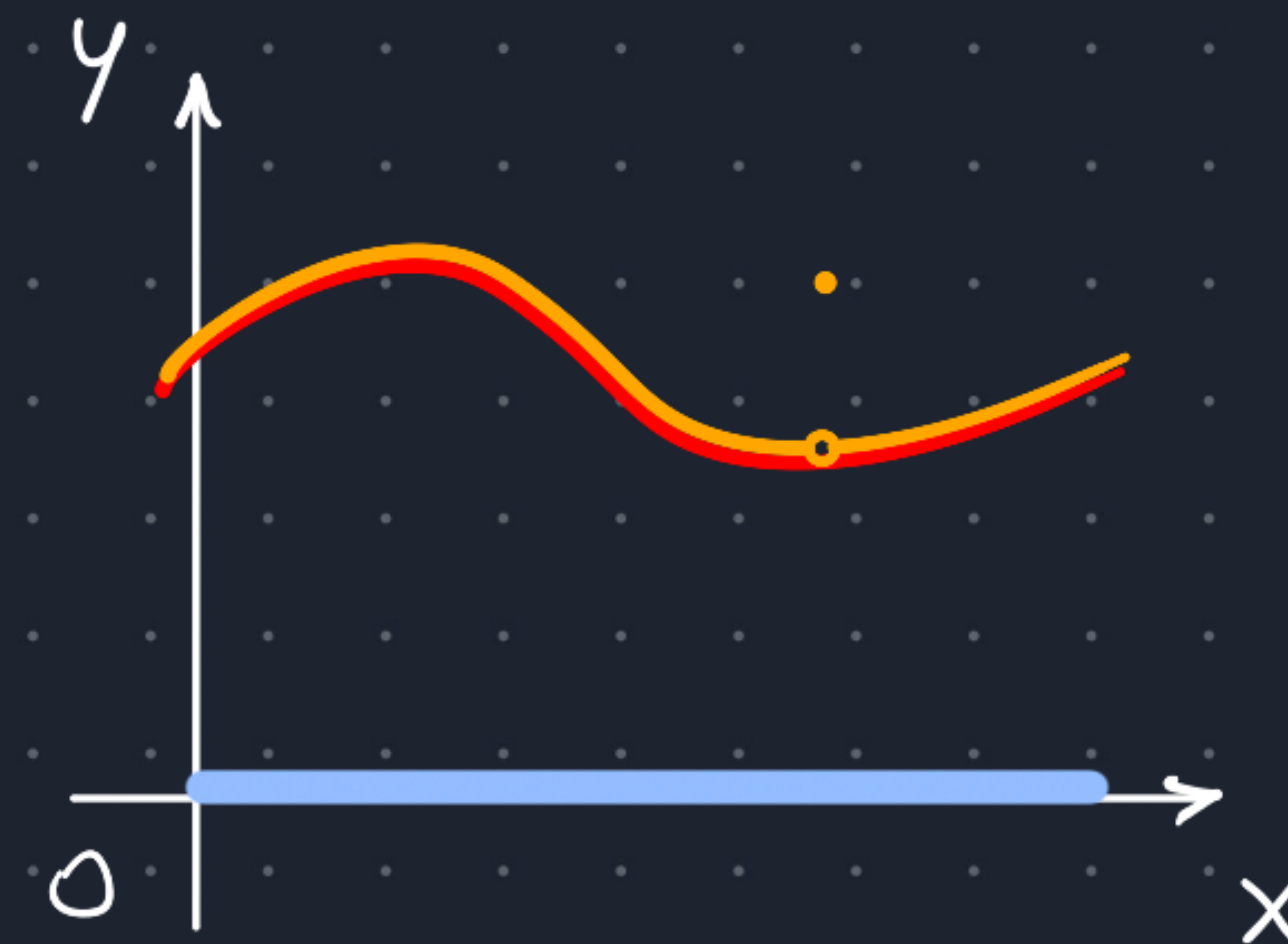
and 

, 

, we say that $f = g$ p-almost surely

if 

$=$ 



Example. In FinStoch, $f(x) = g(x)$ at each x s.t. $p(x) \neq 0$.

In BorelStoch, the set $\{x \in X \,;\, f(x) \neq g(x)\}$ has p-measure zero.

(See Exercise 2.1 later.)

Conditioning & Bayesian inverses

Given

, a conditional distribution of p given X

is a morphism*

such that

(* see Exercise 2.2 later)

$$p(x,y) = p(x)\, p(y|x)$$

Theorem. FinStoch has all conditional distributions.

BorelStoch too. ← Traditionally the chosen category for "classical" prob. theory.

Stoch, in general, does not.

Conditioning & Bayesian inverses

Given 

, a conditional distribution of p given X

is a morphism 

such that 

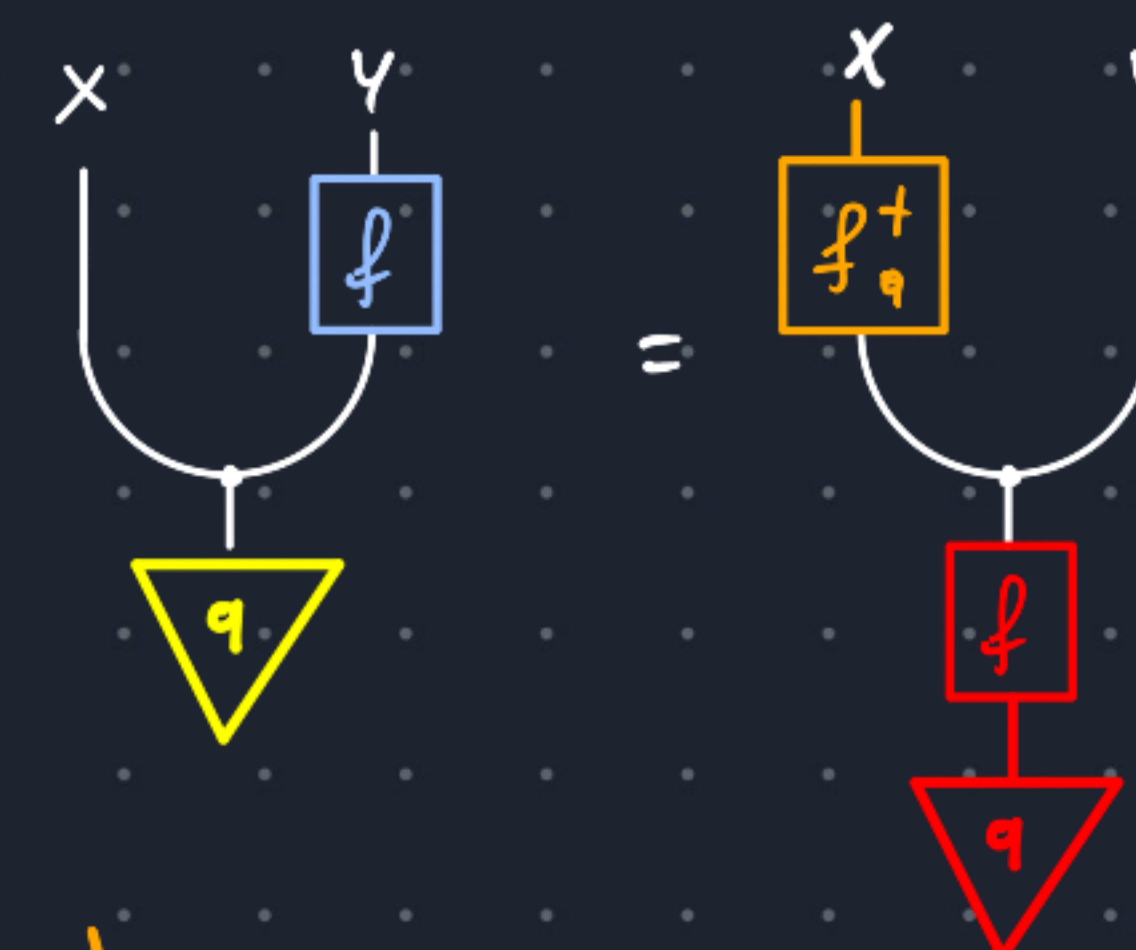
Given 

and 

, a Bayesian inverse of f relative to q, is a conditional dist.

for the joint 

, ie. a morphism 

such that 

$p(x)\,p(y|x) = p(y)\,p(x|y)$

The ProbStoch construction

"Cat. of probability spaces & transport plans"

Definition. Let $\mathcal{C}$ be a Markov category with all conditional distributions.

The category ProbStoch $(\mathcal{C})$ has:

- As objects, pairs (X, ) where $X \in \mathcal{C}$, $p : I \to X$

  (e.g. prob. spaces)

- As morphisms (X, ) $\longrightarrow$ (Y, ), equivalence classes

  under p-a.s. equality of morphisms  such that  $=$  .

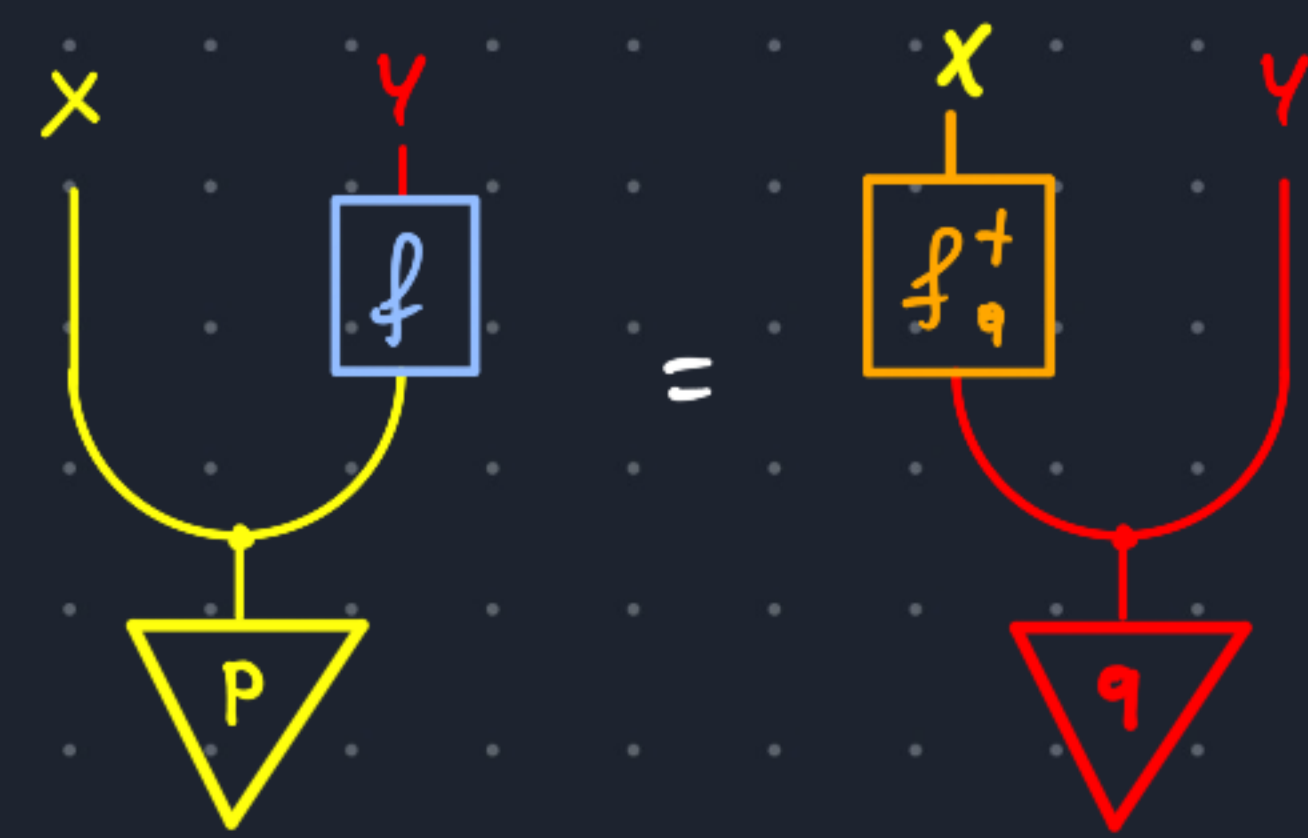
  (measure-preserving)

(See Exercise 2.4 later.)

The ProbStoch construction

"Cat. of probability spaces & transport plans"

Theorem. $\mathrm{ProbStoch}(\mathcal{C})$ is a dagger category, where $\dagger$ = Bayesian inversion.



$$
\begin{tikzcd}
(X, p) \arrow[r, shift left, "f"] & (Y, q) \arrow[l, shift left, "f^\dagger_q"]
\end{tikzcd}
$$

Definition. A dagger structure on a category $\mathcal{C}$ is a "self-duality" functor $\mathcal{C} \xrightarrow[\simeq]{\dagger} \mathcal{C}^{op}$ which is

- Identity on objects: $X^\dagger = X$
- Involutive: $f^{\dagger\dagger} = f$.

Equivalently, morphisms of $\mathrm{ProbStoch}(\mathcal{C})$ are couplings: joint distributions

with

Exercises :

2.1. Show that, in FinStoch, BorelStoch, given $p : I \to X$ and $f, g : X \to Y$, we have that $f = g$ p-almost surely iff $p(\{x \in X ; f(x) \neq g(x)\}) = 0$.

2.2. Suppose the following cond. dist. exists:

Show that any other morphism 

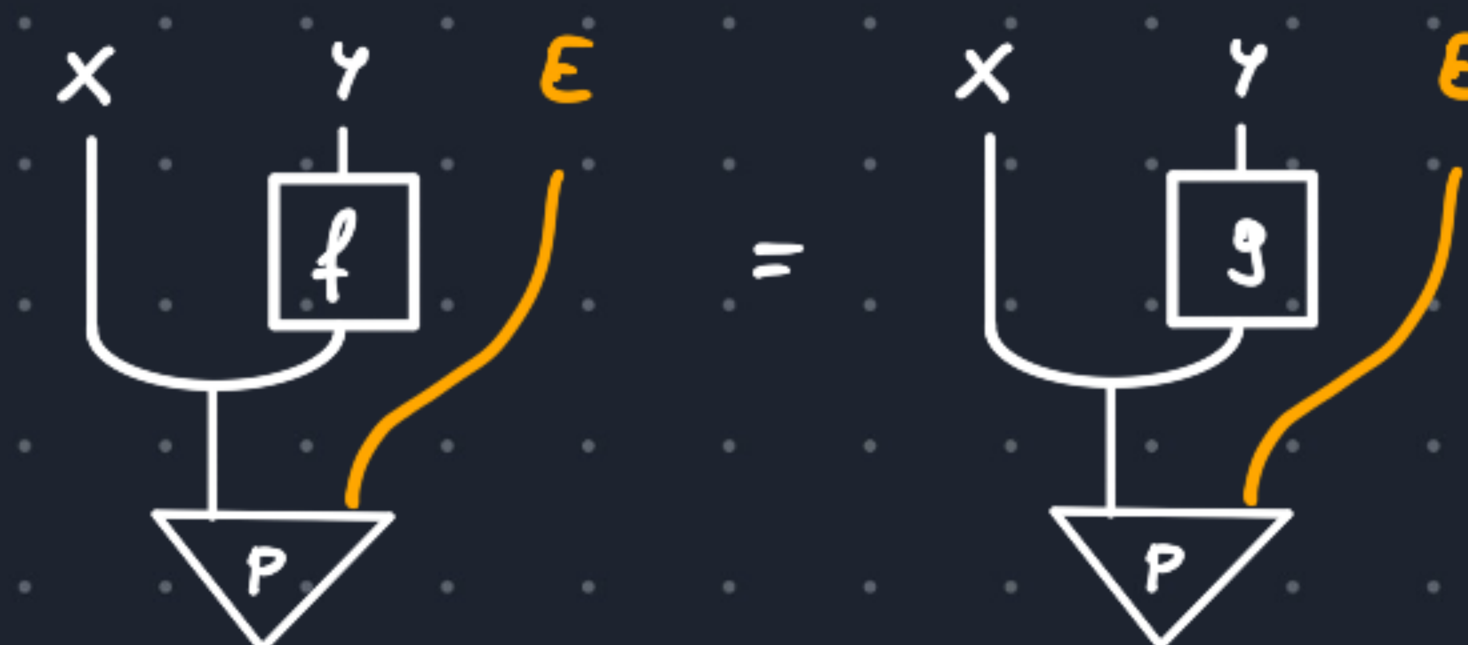
is a conditional iff it is p.a.s. equal to $\boxed{P|_X}$.

2.3. Suppose $\mathcal{C}$ has conditional distributions.

Prove the equality strenghtening property:

If

then also



2.4. Use Ex. 2.3 to show that composition in ProbStoch($\mathcal{C}$) is well defined.

Markov categories and monads

Proposition. Let $\mathcal{D}$ be cartesian monoidal.

Let (P, μ, η) be a monad on $\mathcal{D}$ which is

- Affine : $P1 \cong 1$
- Monoidal (= commutative) : $PA \times PB \xrightarrow{\nabla} P(A \times B)$

Then Kleisli(P) is a Markov category.

Example. Stoch is the Kleisli category of the Giry monad on Meas. (Exercise 3.2)

FinStoch is almost the Kleisli cat. of the distribution monad on Set.

Markov categories and monads

$$\mathcal{D}(A, PB) \cong \mathcal{D}_p(A, B)$$

$$\mathcal{D}(PB, PB) \cong \mathcal{D}_p(PB, B)$$

$$\mathrm{id} \longmapsto \mathrm{samp}$$

In basic probability theory:

$$\mathrm{Bernoulli}(p) = \begin{cases} 1 & \text{with prob. } p \\ 0 & \text{with prob. } 1-p \end{cases}$$

$$[0,1] \xrightarrow{\text{Bernoulli}} \{0,1\}$$

$$[0,1] \cong P(\{0,1\})$$

The unit of the adjunction is

$$\mathcal{D}(A, PA) \cong \mathcal{D}_p(A, A)$$

$$\delta \longleftarrow\!\!\shortmid\; \mathrm{id}$$

$$X \xrightarrow{\delta} PX$$

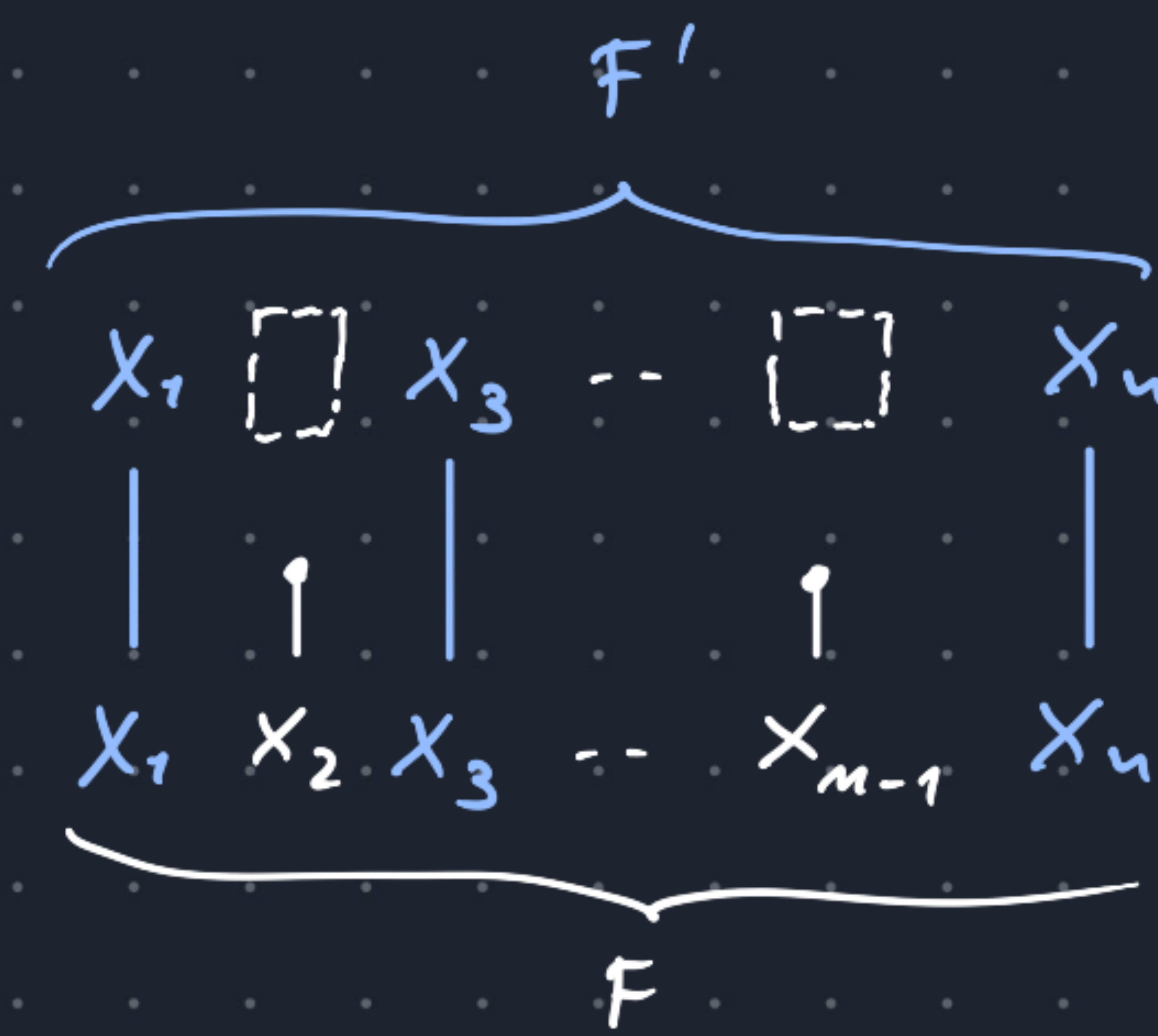
Kolmogorov products

- Probability theory is all about stochastic processes in infinite time, which need objects in the form $X^{\mathbb{N}}$ (at least!)

- While cartesian products can be infinite, we need infinite monoidal products.

- For finite sets F, we have $\bigotimes_{i \in F} X_i$.

- Given subsets $F' \subseteq F$, we can marginalize $\bigotimes_{i \in F} X_i \longrightarrow \bigotimes_{i \in F'} X_i$ by discarding.



Kolmogorov products

Definition. Let $\mathcal{C}$ be a Markov category. Let I be an infinite set, $\{X_i\}_{i\in I}$.

A **Kolmogorov product** is a cofiltered limit

$$X^I := \lim_{F \subseteq I} \left(\bigotimes_{i \in F} X_i \right) \longrightarrow \cdots$$

$$\begin{tikzcd} & & X_i \otimes X_j \arrow[r] \arrow[dr] & X_i \\ \cdots \arrow[urr] \arrow[drr] & & & X_j \\ & & X_j \otimes X_k \arrow[ur] \arrow[dr] & \\ & & & X_k \end{tikzcd}$$

such that
- It is preserved by $Y \otimes -$
- The arrows $X^I \longrightarrow X^F$ are deterministic.

Theorem (Kolmogorov extension). BorelStoch has countable Kolmogorov products.

Kolmogorov products

When $\mathcal{C}$ is the Kleisli category of a probability monad,

a Kolmogorov products encodes the absence of infinitary stochastic interactions.

- "No products": P does not preserve finite products

$$P(X \times Y) \xrightarrow{\ncong} PX \times PY$$

- Kolmogorov ext. thm:

$$P(X^{\mathbb{N}}) = P\Big(\lim_{F \subseteq \mathbb{N}} X^F\Big) \xrightarrow{\cong} \lim_{F \subseteq \mathbb{N}} P(X^F) \xrightarrow{\ncong} \lim_{F \subseteq \mathbb{N}} (PX)^F = (PX)^{\mathbb{N}}$$

Exercises

3.1. Prove

> **Proposition.** Let $\mathcal{D}$ be cartesian monoidal.
>
> Let (P, μ, η) be a monad on $\mathcal{D}$ which is
>
> - Affine : $P1 \cong 1$
> - Monoidal : $PA \times PB \xrightarrow{\nabla} P(A \times B)$
>
> Then Kleisli(P) is a Markov category.

3.2. (For people who know some measure theory.)

Prove that Stoch $\simeq$ Kleisli (Giry monad)

Hint: why is a kernel a Kleisli morphism?

3.3. Prove that sampling from a product distribution is the same as sampling the factors independently:

$$\begin{tikzcd} PX \otimes PY \arrow[r, "\nabla"] \arrow[dr, "\mathrm{samp} \otimes \mathrm{samp}"'] & P(X \otimes Y) \arrow[d, "\mathrm{samp}"] \\ & X \otimes Y \end{tikzcd}$$

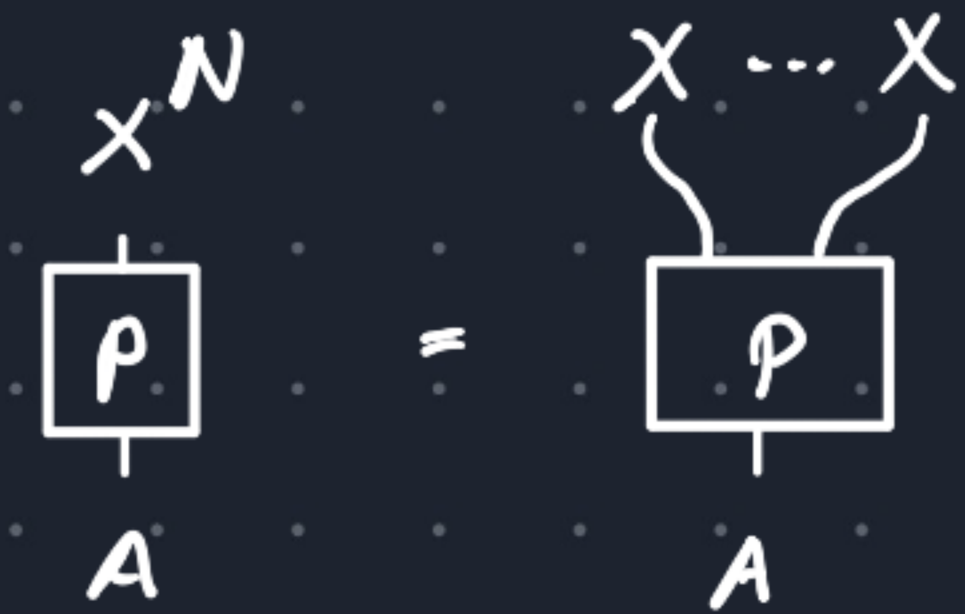
3.4. Prove that the Kolmogorov product

$\bigotimes_{i \in I} X_i$ is the cartesian product

$\prod_{i \in I} X_i$ in the subcategory $\mathcal{C}_{det}$.

The de Finetti theorem

A morphism

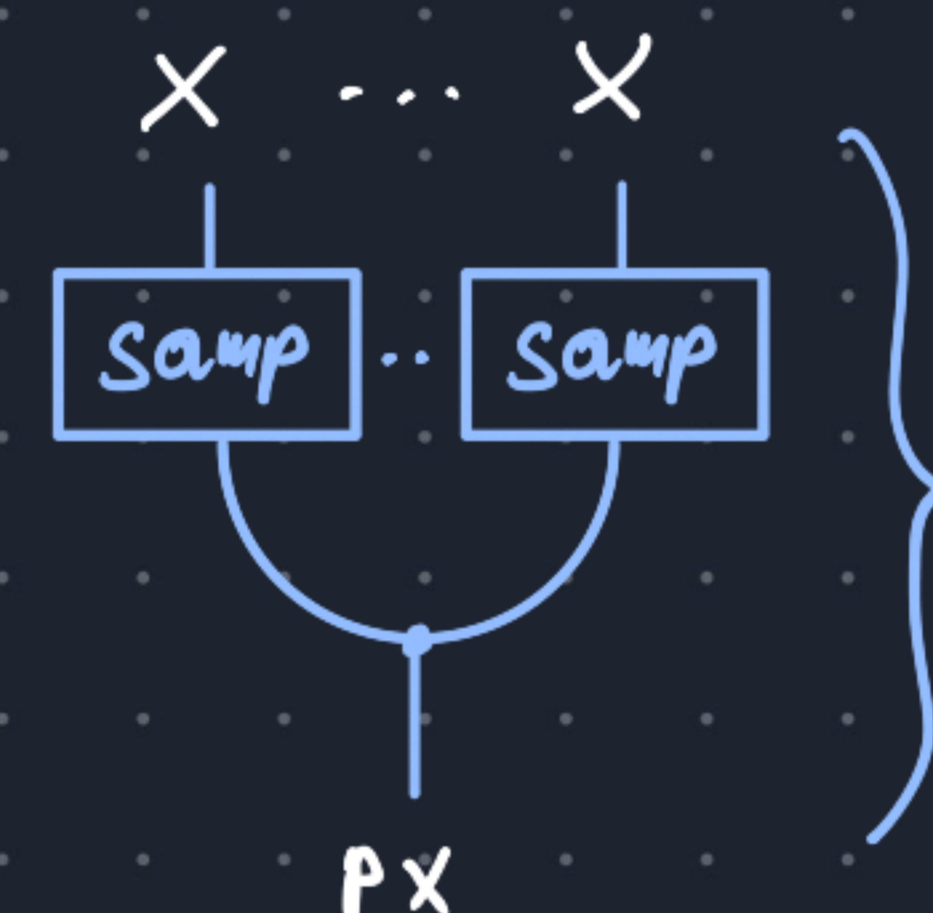


is called exchangeable if it commutes with finite permutations (in the result).

$$\begin{tikzcd} & X^{\mathbb{N}} \arrow[d, "\sigma"] \\ A \arrow[ur, "p"] \arrow[r, "p"'] & X^{\mathbb{N}} \end{tikzcd}$$

Theorem. In BorelStoch, for every A and X, there is a natural bijection

$$\begin{tikzcd} & & X^{\mathbb{N}} \arrow[dd, "\sigma"] \\ A \arrow[urr, "p"] \arrow[r, dashed] \arrow[drr, "p"'] & PX \arrow[ur] \arrow[dr] & \\ & & X^{\mathbb{N}} \end{tikzcd}$$



taking i.i.d. samples!

(limiting cone)

The de Finetti theorem

The X are conditionally independent given PX (the distribution from which they are sampled - independently).

Example. Let $X = \{\text{Heads}, \text{Tails}\}$. Flip a coin repeatedly (exchangeable).

Suppose you see Heads, Heads, Heads, Heads.

What do you expect to see next?

$\boxed{\text{"Same coin!"}}$

What if you <u>know</u> that the coin is fair?

Results so far

Classical probability:

- De Finetti theorem (Fritz-Gonda-Perrone '21)
- d-separation criterion (Fritz-Klingler '22)
- Kolmogorov extension theorem (Fritz-Rischel '19)
- Kolmogorov, H-S 0-1 laws (Fritz-Rischel '19)
- Multinomial, hypergeometric distributions (Jacobs '21)

Statistics:

- Theorems on sufficient statistics (Fritz '19)
- Comparison of experiments (Fritz-Gonda-Perrone-Rischel '20)

Ergodic theory, information theory:

- Ergodic decomposition thm (Moss-Perrone '22)
- Entropy, data processing inequalities (Perrone '22)

Theoretical computer science:

- Privacy eqn (Sabok et. al '20, Fritz et al. '22)
- Observational monads (Moss-Perrone '22)

Quantum probability:

- Quantum Markov categories (Parzygnat '20, '21)

\+ more in progress!

Some references:

- K. Cho, B. Jacobs, *Disintegration and Bayesian inversion via string diagrams*. Mathematical Structures in Computer Science. arXiv: 1709.00322
- T. Fritz, *A synthetic approach to Markov kernels, conditional independence, and theorems on sufficient statistics*. Advances in Mathematics. arXiv: 1908.07021
- T. Fritz, T. Gonda, P. Perrone, E. F. Rischel, *Representable Markov categories and comparison of statistical experiments in categorical probability*. arXiv: 2010.07416
- T. Fritz, T. Gonda, P. Perrone, *The de Finetti theorem in categorical probability*. Journal of Stochastic Analysis. arXiv: 2105.02639
- T. Fritz, T. Gonda, N. Gauguin Houghton-Larsen, P. Perrone, D. Stein, *Dilations and information flow axioms in categorical probability*. arXiv: 2211.02507
- P. Perrone, *Markov Categories and Entropy*. arXiv: 2212.11719

... & more