Dilations and Information Flow in Categorical Probability

Tomáš Gonda

Joint work with: T Fritz, P Perrone, NG Houghton-Larsen, and D Stein

20 December 2022

universität innsbruck

References

[1] C Heunen, O Kammar, S Staton, and H Yang,
A convenient category for higher-order probability theory.
LICS 2017, page 12. arXiv:1701.02547.

[2] T Fritz,
A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics.
Adv. Math. 370, 107239 (2020). arXiv:1908.07021.

[3] T Fritz and EF Rischel,
The zero-one laws of Kolmogorov and Hewitt–Savage in categorical probability.
Compositionality 2, 3 (2020). arXiv:1912.02769.

[4] T Fritz, T Gonda, P Perrone, EF Rischel,
Representable Markov Categories and Comparison of Statistical Experiments in Categorical Probability.
arXiv:2010.07416.

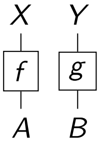
[5] T Fritz, T Gonda, P Perrone,
De Finetti's Theorem in Categorical Probability.
J. Stoch. Anal. 2(4), 6 (2021) arXiv:2105.02639.

[6] T Fritz and A Klingler,
The d-separation criterion in Categorical Probability.
arXiv:2207.05740.

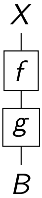
Markov categories

FinStoch

$$\leftrightarrow \quad f(x|a) \in [0,1]$$



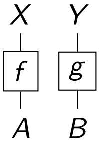
$$\leftrightarrow \quad f \otimes g(x, y|a, b) := f(x|a)\, g(y|b)$$



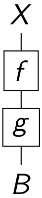
$$\leftrightarrow \quad f \circ g(x|b) := \sum_{y \in Y} f(x|y)\, g(y|b)$$

Kl(D_R) (semiring-valued kernels)

$$\leftrightarrow \qquad f(x|a) \in R$$



$$\leftrightarrow \qquad f \otimes g(x, y|a, b) := f(x|a)\, g(y|b)$$

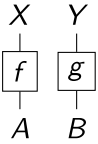


$$\leftrightarrow \qquad f \circ g(x|b) := \sum_{y \in Y} f(x|y)\, g(y|b)$$

Stoch

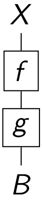
$\leftrightarrow$

$$f : \Sigma_X \times A \to [0, 1]$$



$\leftrightarrow$

$$f \otimes g(S \otimes T|a, b) := f(S|a)\, g(T|b)$$



$\leftrightarrow$

$$f \circ g(S|b) := \int_Y f(S|y)\, g(dy|b)$$

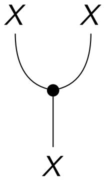
Others

- **BorelStoch** $\hookrightarrow$ **Stoch** has $\mathbb{N}$, $[0, 1]$, ...
- **Kl**(D_R): kernels valued in a semiring $(R, \cdot, +)$
  - **Kl**$(D_{\{0,1\}})$: possibilistic
  - **FinStoch** $\hookrightarrow$ **Kl**$(D_{\mathbb{R}_+})$
  - **Kl**$(D_{\mathbb{R}})$: negative "probabilities"
  - $R \sim$ distributive lattice
  - $R \sim$ ideals of a commutative ring
  - ...
- **QBStoch** [1]: uncertainty about functions
- ...

Copying

Additional structure:

- Every X has a **copying morphism**:



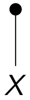
e.g.

$$\mathrm{copy}(x_1, x_2 | x) = \begin{cases} 1 & \text{if } x_1 = x_2 = x \\ 0 & \text{otherwise.} \end{cases}$$

Deletion

Additional structure:

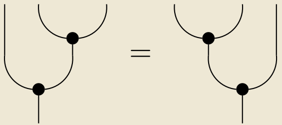
- Every X has a **deletion morphism**:

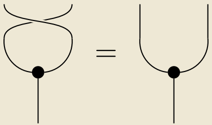


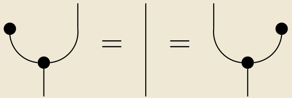
such as $\mathrm{del}(*|x) = 1$.

Definition

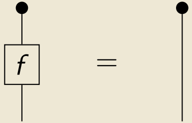
A **Markov category C** is a SMC with copy and del, which are compatible with $\otimes$, satisfy







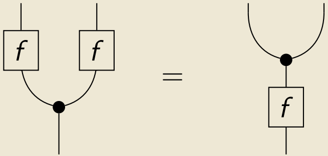
and for any f also



Determinism

Definition

$f : X \to Y$ is **deterministic** if

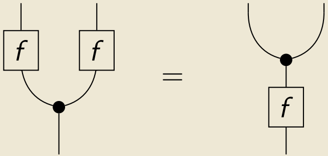


- **Intuition:** Applying f to copies of input = copying output.

Determinism

Definition

$f : X \to Y$ is **deterministic** if

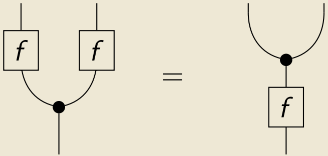


- **Intuition:** Applying f to copies of input = copying output.
- Deterministic morphisms form a subcategory $\mathbf{C}_{\mathrm{det}}$.

Determinism

Definition

$f : X \to Y$ is **deterministic** if

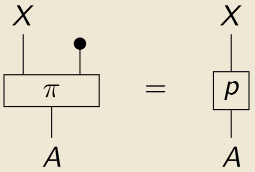


- **Intuition:** Applying f to copies of input = copying output.
- Deterministic morphisms form a subcategory $\mathbf{C}_{\mathrm{det}}$.
- In **BorelStoch** they are measurable maps; in $\mathbf{Kl}(D_R)$ 'functions'.

Dilations

Definition

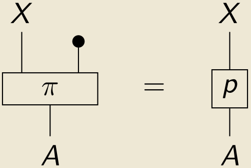
$\pi\colon A \to X \otimes E$ is a **dilation** of $p\colon A \to X$ if



Dilations

Definition

$\pi\colon A \to X \otimes E$ is a **dilation** of $p\colon A \to X$ if



- **Intuition:** Information "leaks" to the environment, e.g.

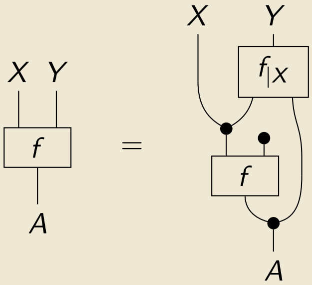
$$X = A = \{0,1\} \quad \text{and} \quad \pi = \frac{1}{2}\mathrm{id}_X \otimes \delta_e + \frac{1}{2}\mathrm{flip}_X \otimes \delta_{e'}$$

Information Flow Axioms

Conditionals

Definition

C has conditionals if for every f there is $f_{|X}$ with

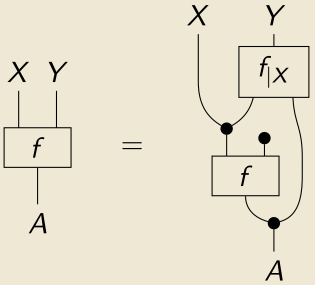


- **Intuition:** The outputs of f can be generated one at a time.

Conditionals

Definition

C has conditionals if for every f there is $f_{|X}$ with

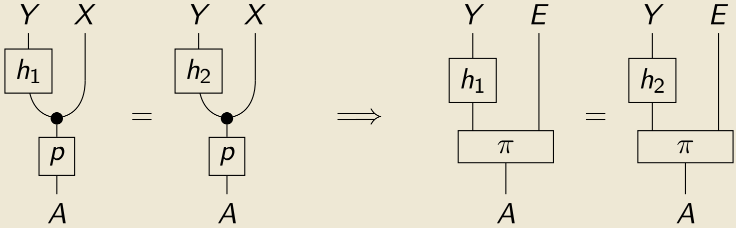


- **Intuition:** The outputs of f can be generated one at a time.
- Used for de Finetti's Thm [5], d-separation [6], BSS Thm [4].

Causality

Definition

C is **causal** if, for every dilation π of p:

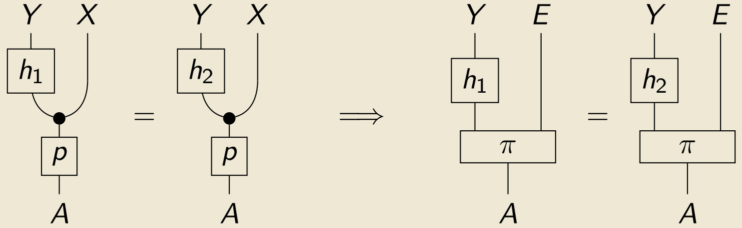


- **Intuition:** Equality almost surely is "local".

Causality

Definition

C is **causal** if, for every dilation π of p:

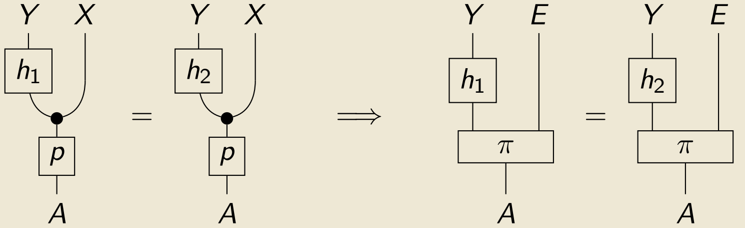


- **Intuition:** Equality almost surely is "local".
- Used for Hewitt–Savage 0-1 laws [3].

Causality

Definition

C is **causal** if, for every dilation π of p:



- **Intuition:** Equality almost surely is "local".
- Used for Hewitt–Savage 0-1 laws [3].

conditionals $\implies$ causality [2]

Causality

Proposition

$\mathbf{Kl}(D_R)$ is causal iff R satisfies

$$s(v+w) = t(v+w) \quad \Longrightarrow \quad sv = tv \ \text{ and } \ sw = tw$$

for all s, t and $v + w$ with complements.

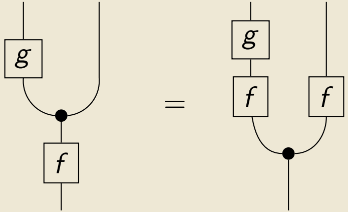
Satisfied

- for every distributive lattice;
- when R is zero-sum-free and has inverses.

Positivity

Definition

C is **positive** if whenever $g \circ f$ is deterministic, then

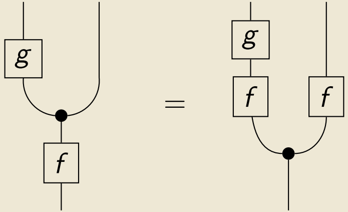


- **Intuition:** Intermediate result of a deterministic process can be produced independently.

Positivity

Definition

C is **positive** if whenever $g \circ f$ is deterministic, then

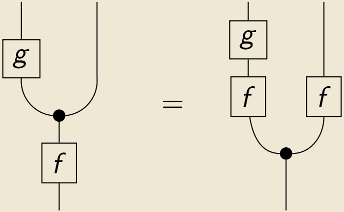


- **Intuition:** Intermediate result of a deterministic process can be produced independently.
- Used for characterizing sufficient statistics [2].

Positivity

Definition

C is **positive** if whenever $g \circ f$ is deterministic, then



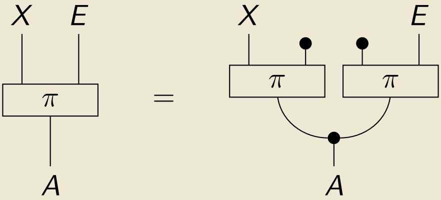
- **Intuition:** Intermediate result of a deterministic process can be produced independently.
- Used for characterizing sufficient statistics [2].

$$\text{conditionals} \implies \text{positivity [2]}$$

Deterministic marginal independence (DMI)

Proposition

Alternatively, **C** is **positive** if every dilation π of a deterministic $p\colon A \to X$ satisfies



- **Intuition:** Deterministic outcomes are independent.

Positivity

Proposition

$\mathbf{Kl}(D_R)$ is representable iff R is **entire**. That is, $R \neq 0$ and

$$rs = 0 \quad \Longrightarrow \quad r = 0 \text{ or } s = 0.$$

Proposition

A representable $\mathbf{Kl}(D_R)$ is positive iff R is **zero-sum-free**. That is,

$$r + s = 0 \quad \Longrightarrow \quad r = s = 0.$$

Positivity

Proposition

$\mathbf{Kl}(D_R)$ is representable iff R is **entire**. That is, $R \neq 0$ and

$$rs = 0 \quad \Longrightarrow \quad r = 0 \text{ or } s = 0.$$

Proposition

A representable $\mathbf{Kl}(D_R)$ is positive iff R is **zero-sum-free**. That is,

$$r + s = 0 \quad \Longrightarrow \quad r = s = 0.$$

- Violated by negative probabilities in $\mathbf{Kl}(D_{\mathbb{R}})$.

Positivity

Proposition

$\mathbf{Kl}(D_R)$ is representable iff R is **entire**. That is, $R \neq 0$ and

$$rs = 0 \quad \Longrightarrow \quad r = 0 \text{ or } s = 0.$$

Proposition

A representable $\mathbf{Kl}(D_R)$ is positive iff R is **zero-sum-free**. That is,

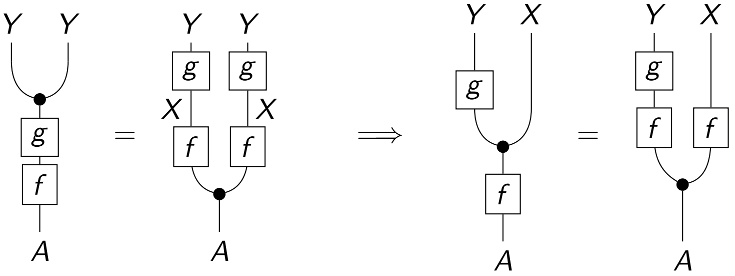
$$r + s = 0 \quad \Longrightarrow \quad r = s = 0.$$

- Violated by negative probabilities in $\mathbf{Kl}(D_{\mathbb{R}})$.
- Causality condition for $s = 1$ and $t = 0$ is:

$$1 \cdot (v + w) = 0 \cdot (v + w) \quad \Longrightarrow \quad 1 \cdot v = 0 \cdot v \text{ and } 1 \cdot w = 0 \cdot w$$

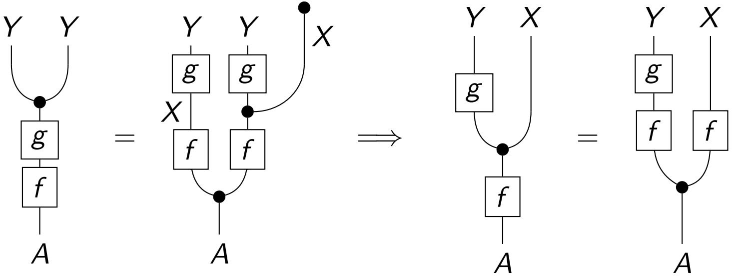
Causality $\implies$ positivity

C is positive if

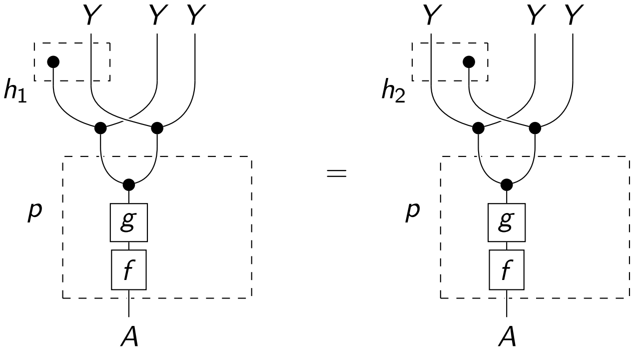


Causality $\implies$ positivity

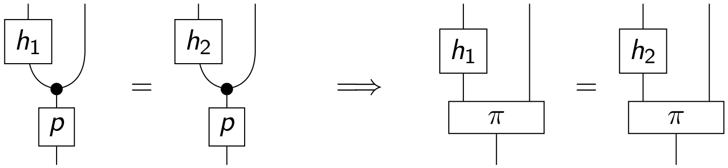
C is positive if

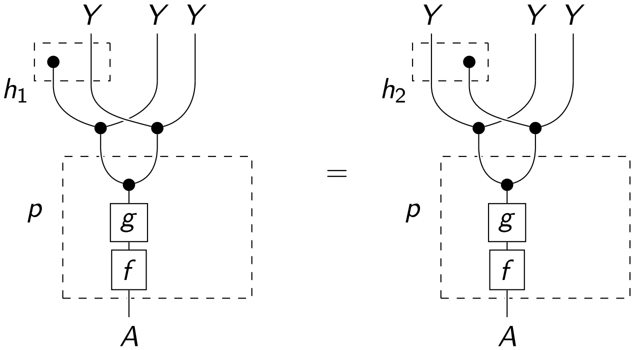


Causality $\implies$ positivity

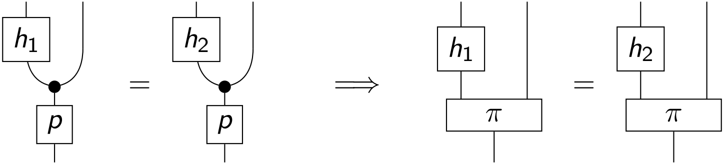


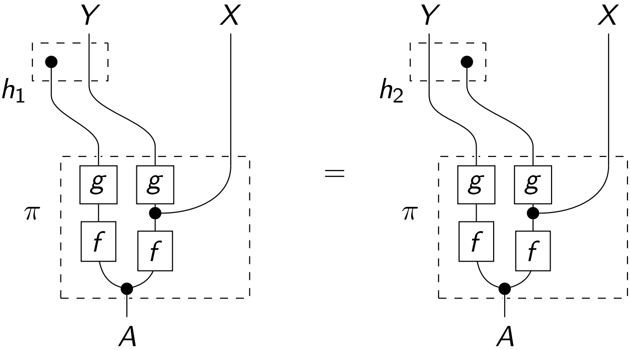
Causality $\implies$ positivity



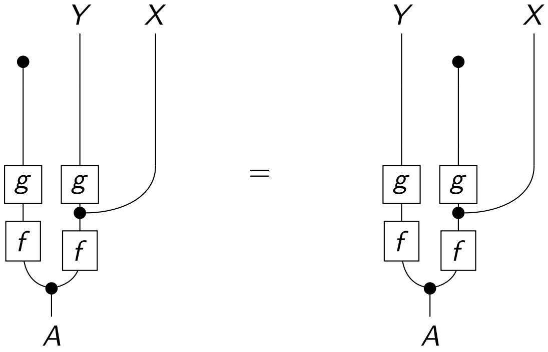


Causality $\implies$ positivity

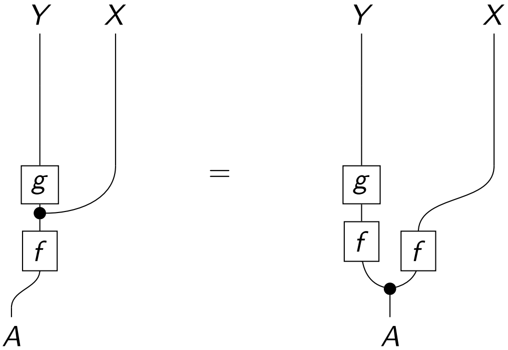




Causality $\implies$ positivity



Causality $\implies$ positivity



Examples

Conditionals	**FinStoch**, **BorelStoch**
$\Downarrow$	
Causality	**Stoch**
$\Downarrow$	
Positivity	
~~Positivity~~	$\mathbf{Kl}(D_{\mathbb{R}})$

Examples

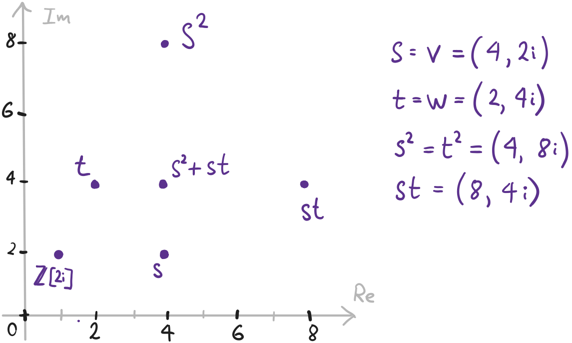
Conditionals	**FinStoch**, **BorelStoch**
$\Downarrow$	
Causality	**Stoch**, **Kl**(D_R) for R a distributive lattice
$\Downarrow$	
Positivity	
~~Positivity~~	**Kl**($D_{\mathbb{R}}$), **QBStoch**

Examples

Conditionals	**FinStoch**, **BorelStoch**
$\Downarrow$	
Causality	**Stoch**, **Kl**(D_R) for R a distributive lattice
$\Downarrow$	
Positivity	**Kl**(D_R) for $R = \mathcal{I}(\mathbb{Z}[2i])$
~~Positivity~~	**Kl**$(D_{\mathbb{R}})$, **QBStoch**

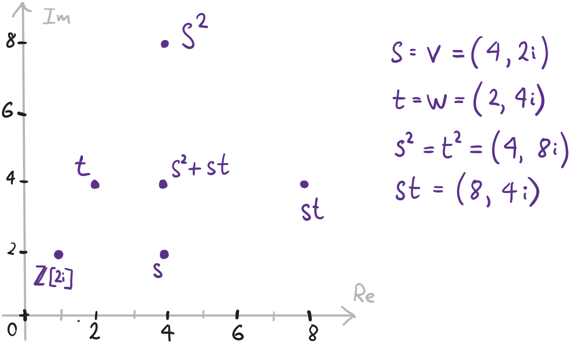
$$R \coloneqq \mathbb{Z}[2i] = \mathbb{Z} \oplus 2i\mathbb{Z} \quad \text{and} \quad (m, 2ki) \coloneqq m\mathbb{Z} \oplus 2ki\mathbb{Z} \subseteq R$$

$$R := \mathbb{Z}[2i] = \mathbb{Z} \oplus 2i\mathbb{Z} \quad \text{and} \quad (m, 2ki) := m\mathbb{Z} \oplus 2ki\mathbb{Z} \subseteq R$$

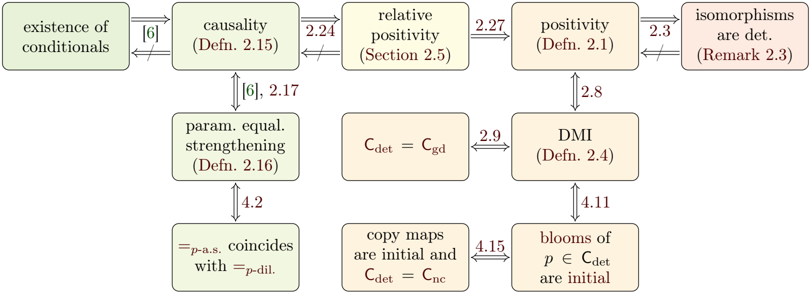


$\mathbf{Kl}(D_R)$ causal only if $s(v+w) = t(v+w) \implies sv = tv$.

$$R := \mathbb{Z}[2i] = \mathbb{Z} \oplus 2i\mathbb{Z} \quad \text{and} \quad (m, 2ki) := m\mathbb{Z} \oplus 2ki\mathbb{Z} \subseteq R$$



$$s(v+w) = s^2 + st = st + t^2 = t(v+w)$$
$$sv = s^2 = (8, 4i) \neq (4, 8i) = st = tv$$


existence of conditionals
[6]
causality (Defn. 2.15)
2.24
relative positivity (Section 2.5)
2.27
positivity (Defn. 2.1)
2.3
isomorphisms are det. (Remark 2.3)
[6], 2.17
2.8
param. equal. strengthening (Defn. 2.16)
$\mathsf{C}_{\mathrm{det}} = \mathsf{C}_{\mathrm{gd}}$
2.9
DMI (Defn. 2.4)
4.2
4.11
$=_{p\text{-a.s.}}$ coincides with $=_{p\text{-dil.}}$
copy maps are initial and $\mathsf{C}_{\mathrm{det}} = \mathsf{C}_{\mathrm{nc}}$
4.15
blooms of $p \in \mathsf{C}_{\mathrm{det}}$ are initial

Bonus slides

Kleisli categories are Markov categories

Proposition

Let

- $\mathbf{D}$ be a category with finite products,
- P a commutative monad on $\mathbf{D}$ with $P(1) \cong 1$.

Then the Kleisli category $\mathbf{Kl}(P)$ is a Markov category.

Examples:

- Kleisli category of the Giry monad: **Stoch**.
- Kleisli category of the non-empty power set monad: **Rel**.
- Kleisli category of the distribution monad: $\mathbf{Kl}(D_R)$.

Proposition

$\mathbf{Kl}(D_R)$ is representable iff R is **entire**. That is, $R \neq 0$ and

$$rs = 0 \implies r = 0 \text{ or } s = 0.$$

Proposition

A representable $\mathbf{Kl}(D_R)$ is positive iff R is **zerosumfree**. That is,

$$r + s = 0 \implies r = s = 0.$$

Proposition

$\mathbf{Kl}(D_R)$ is causal iff R satisfies

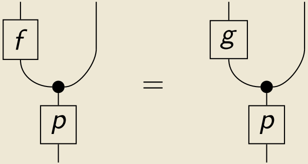
$$s(v + w) = t(v + w) \implies sv = tv \text{ and } sw = tw$$

for all s, t and $v + w$ with complements.

Almost sure equality

Definition

f and g are **equal p-almost surely**, $f =_{p\text{-a.s.}} g$, if

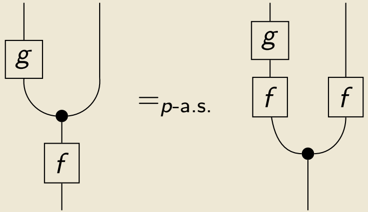


- **Intuition:** f and g are the same on all inputs in p's support.
- Other concepts (e.g. positivity) relativize w.r.t. $=_{p\text{-a.s.}}$.

Relative positivity

Definition

C is **relatively positive** if whenever $g \circ f$ is -a.s.p deterministic, then

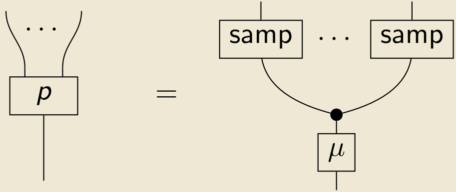


Abstract de Finetti's theorem

Theorem

Let $\mathbf{C}$ be an a.s.-compatibly representable Markov category with conditionals and countable Kolmogorov products.

For every exchangeable $p\colon A \to X^{\mathbb{N}}$, there is $\mu\colon A \to PX$ such that



- **BorelStoch** satisfies these assumptions.

The classical Hewitt–Savage zero-one law

Theorem

Let $(x_n)_{n\in\mathbb{N}}$ *be independent and identically distributed random variables, and* S *any event depending only on the* x_n *and invariant under finite permutations.*

Then $\mathbf{P}(S) \in \{0,1\}$.

The synthetic Hewitt–Savage zero-one law

Theorem

Let J be an infinite set and $\mathbf{C}$ a causal Markov category. Suppose that:

- The Kolmogorov power $X^{\otimes J} := \lim_{F \subseteq J \text{ finite}} X^{\otimes F}$ exists.
- $p \colon A \to X^{\otimes J}$ displays the conditional independence $\perp_{i \in J} X_i \parallel A$.
- $s \colon X^J \to T$ is deterministic.
- For every finite permutation $\sigma \colon J \to J$, permuting the factors $\tilde{\sigma} \colon X^{\otimes J} \to X^{\otimes J}$ satisfies

$$\tilde{\sigma} p = p, \qquad s \tilde{\sigma} = s.$$

Then sp is deterministic.

Proof is by string diagrams, but far from trivial!

Diagram categories and ergodic theory

Proposition

Let $\mathbf{D}$ be any category and $\mathbf{C}$ a Markov category. The category in which

- Objects are functors $\mathbf{D} \to \mathbf{C}_{\mathrm{det}}$,
- Morphisms are natural transformations with components in $\mathbf{C}$.

With the poset $\mathbf{D} = \mathbb{Z}$, we get a category of **discrete-time stochastic processes**.

This generalizes an observation going back to (Lawvere, 1962).

We can also take $\mathbf{D} = \mathbf{B}G$ for a group G, resulting in categories of dynamical systems with deterministic dynamics but stochastic morphisms.

Dilations

A dilation $t\colon P\Theta \to P\Theta$ preserves barycenters.

$$\begin{array}{ccc} P\Theta & \xrightarrow{\;t\;} & P\Theta \\ {\scriptstyle \text{samp}}\downarrow & & \downarrow{\scriptstyle \text{samp}} \\ \Theta & = & \Theta \end{array}$$

If it commutes μ-a.s., then t is a μ-dilation.

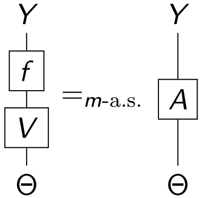
Synthetic BSS Theorem

If a Markov category

- is representable, and
- has Bayesian inverses,

then for any $m\colon I \to \Theta$

$\exists$ a morphism $f\colon X \to Y$ $\qquad$ $\exists$ an $(\hat{A}\, m)$-dilation $t\colon P\Theta \to P\Theta$



$\iff$