Categorical Algebra of Conditional Probability

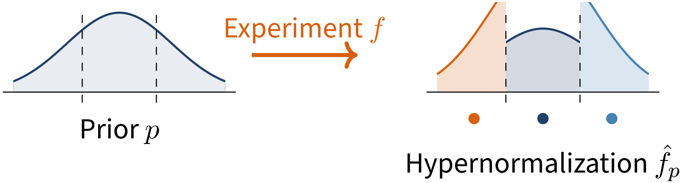
Mika Bohinen & Paolo Perrone

University of Oxford

SYCO 13, London – Thursday 24th April, 2025

Overview

- Every statistical experiment induces a canonical decomposition of a state
- Informativeness of experiments corresponds to algebraic properties of these decompositions
- Current theory depends on Bayesian inverses, which may not always exist
- Our contribution: Breaking this dependence, enabling purely algebraic study



Key Contributions

- We prove the Giry monad is Beck-Chevalley through:
    1. Weakly cartesian multiplication (Theorem 3.1)
    2. Weak pullback preservation (Theorem 3.7)
- We establish a universal property of hypernormalizations (Theorem 3.11)
- We provide an alternative definition without Bayesian inverses (Definition 3.12)

A Synthetic Perspective

Our work reveals the compositional nature of statistical experiments through categorical algebra rather than measure theory.

1

Quick Recap of Markov Categories

Markov Categories

Definition: Markov Category

A **Markov category** is a symmetric monoidal category $(\mathbf{C}, \otimes, I)$ where:

- Every object has a copy operation $\mathtt{copy}_X : X \to X \otimes X$
- Every object has a delete operation $\mathtt{del}_X : X \to I$
- These form a commutative comonoid structure
- The monoidal unit I is terminal

Intuition

Markov categories provide a synthetic framework for probability theory where morphisms represent stochastic processes. This enables us to abstract away from many measure-theoretic subtleties.

Example: Category of Markov Kernels

Example: BorelStoch

- Objects are measurable spaces (X, Σ_X)
- Morphisms are Markov kernels $k : X \to Y$ where:
  - $k(B|-)$ is measurable for all $B \in \Sigma_Y$
  - $k(-|x)$ is a probability measure for all $x \in X$
- Composition is via the Chapman-Kolmogorov equation:

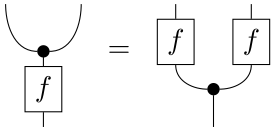
$$(h \circ k)(C|x) = \int_Y h(C|y)k(dy|x)$$

- **BorelStoch** is the full subcategory of standard Borel spaces

Deterministic Morphisms

Definition: Deterministic Morphism

A morphism $f : A \to X$ in a Markov category is **deterministic** if:

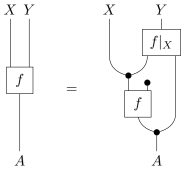


The subcategory of **C** with deterministic morphisms is denoted $\mathbf{C}_{\mathrm{det}}$.

Conditionals

Definition: Conditional

For a morphism $f : A \to X \otimes Y$, a **conditional** of f with respect to X is a morphism $f|_X : X \otimes A \to Y$ such that:



A Markov category **has conditionals** if every morphism admits a conditional with respect to any of its outputs.

Representable Markov Categories

Definition: Representable Markov Categories

A **distribution object** for X is an object PX with a morphism $\mathsf{samp}_X : PX \to X$ such that:

$$\mathbf{C}_{\mathrm{det}}(A, PX) \cong \mathbf{C}(A, X)$$

The inverse of this map is denoted $(-)^\# : \mathbf{C}(A, X) \to \mathbf{C}_{\mathrm{det}}(A, PX)$, and we set $\delta_X = (1_X)^\# : X \to PX$. A Markov category is **representable** if every object has a distribution object.

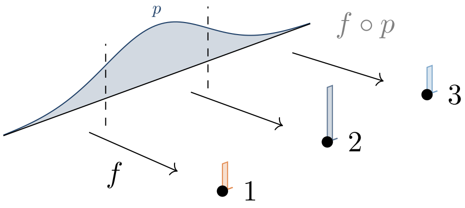
2

Statistical Experiments and their Decompositions

Statistical Experiments

Definition: Statistical Experiment

A **probability space** is a pair (Θ, p) where $p : I \to \Theta$. A **statistical experiment** on (Θ, p) is a morphism $f : \Theta \to X$ (up to p-almost sure equality).



- Θ = real line with normal distribution p
- $X = \{1, 2, 3\}$ = observable outcomes
- f partitions Θ into three regions

Canonical Decompositions (Standard Measures)

Definition: Standard Measure (2.34)

Let (Θ, p) be a probability space in a representable Markov category, and $f : \Theta \to X$ a statistical experiment. The **standard measure** of f is the state $\hat{f}_p$ on $P\Theta$ given by:

$$\hat{f}_p = (f_p^\dagger)^\# \circ f \circ p : I \to P\Theta$$

where $f_p^\dagger$ is the Bayesian inverse of f with respect to p.

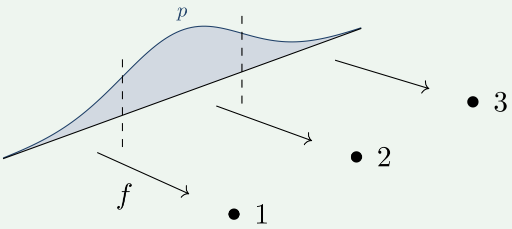
Definition: Hypernormalization (2.35)

For an a.s. deterministic experiment f, the **hypernormalization** of p with respect to f is the standard measure $\hat{f}_p$.

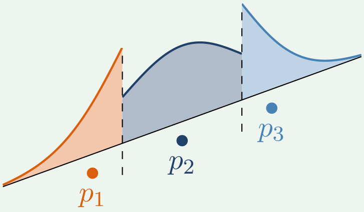
Example: Gaussian Decomposition

Example: Gaussian Experiment

Original distribution p:



Hypernormalization $\hat{f}_p$:



$$\hat{f}_p = w_1 p_1 + w_2 p_2 + w_3 p_3$$

The hypernormalization $\hat{f}_p$ decomposes p into three probability measures

3

Relating Experiment Informativeness to Decompositions

The Blackwell Order on Experiments

Definition: Blackwell Order

Let $f : \Theta \to X$ and $g : \Theta \to Y$ be statistical experiments on (Θ, p). We say that $g \leq f$ in the **Blackwell order** if there exists a morphism $h : X \to Y$ such that $h \circ f =_p g$.

$$\begin{tikzcd} \Theta \arrow[r, "f"] \arrow[dr, "g"'] & X \arrow[d, dashed, "h"] \\ & Y \end{tikzcd}$$

Intuition

f is "more informative" than g if we can recover g by processing the results of f without further access to Θ. Intuitively: f gives us finer-grained information than g.

Stochastic Dominance on Decompositions

Definition: Partial Evaluation Relation (2.23)

Given two states $\pi, \tau : I \to P\Theta$, we say $\pi \leq \tau$ in the **stochastic dominance order** if there exists $\kappa : I \to PP\Theta$ such that $\mathsf{samp} \circ \kappa = \pi$ and $P\mathsf{samp} \circ \kappa = \tau$

$$\begin{tikzcd}
& I \arrow[dl, "\pi"'] \arrow[d, dashed, "\kappa"] \arrow[dr, "\tau"] & \\
P\Theta & PP\Theta \arrow[l, "\mathsf{samp}"] \arrow[r, "P\mathsf{samp}"'] & P\Theta
\end{tikzcd}$$

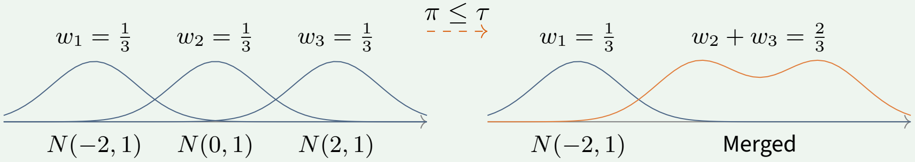
Intuition

- This relation compares how "spread out" probability measures are.
- $\pi \leq \tau$ means τ can be obtained from π by taking "centers of mass" of pieces of π.

Example: Mixture of Gaussians

Example: Mixture of Gaussians

Original Mixture π | Partial Evaluation τ



$$\pi = \tfrac{1}{3}[N(-2,1)] + \tfrac{1}{3}[N(0,1)] + \tfrac{1}{3}[N(2,1)] \qquad \tau = \tfrac{1}{3}[N(-2,1)] + \tfrac{2}{3}\left[\tfrac{1}{2}N(0,1) + \tfrac{1}{2}N(2,1)\right]$$

The partial evaluation τ results from grouping the second and third components of π into a single mixture.

Blackwell-Sherman-Stein Theorem

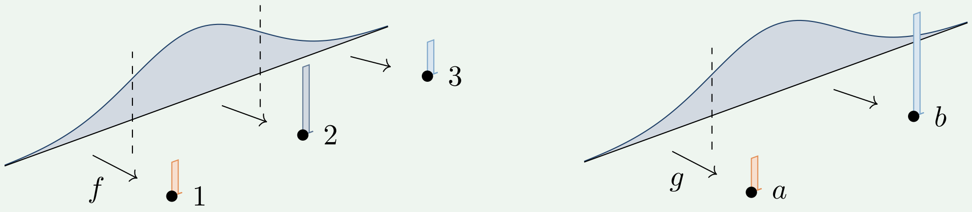
Theorem: Blackwell-Sherman-Stein (2.38)

Let **C** be an a.s.-compatibly representable Markov category. Let $f : \Theta \to X$ and $g : \Theta \to Y$ be statistical experiments on (Θ, p). Then $g \leq f$ in the Blackwell order if and only if $\hat{g}_p \geq \hat{f}_p$ in the stochastic dominance order.

Example: Coarser vs. Finer Experiments

Example: Coarse-Graining

Consider $Y = \{a, b\}$ and $h : X \to Y$ with $h(1) = a$, $h(2) = h(3) = b$. Then $g = h \circ f$ is less informative than f:



$$g \leq f \quad \text{in Blackwell order} \quad \Longleftrightarrow \quad \hat{g}_p \geq \hat{f}_p \quad \text{in stochastic dominance}$$

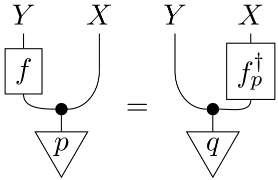
4

The Problem: Dependence on Bayesian Inverses

Bayesian Inverses

Definition: Bayesian Inverse

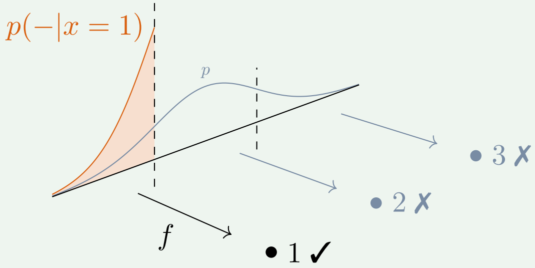
For a statistical experiment $f : \Theta \to X$ and state $p : I \to \Theta$, a **Bayesian inverse** is a morphism $f_p^\dagger : X \to \Theta$ such that:



where $q = f \circ p$. Note that this is a special case of a conditional, i.e., having conditionals implies that we have Bayesian inverses.

Example: Bayesian Inverse

Example: Bayesian Updating



- f partitions Θ into 3 regions
- After observing outcome $x = 1$:
  - Regions 2 and 3 get probability zero
  - Region 1 now has probability one
  - Within region 1, probability is proportional to prior p
- Mathematically:

$$p(A|x=1) = \frac{p(A \cap f^{-1}(1))}{p(f^{-1}(1))}$$

Revisiting Standard Measures

- Recall that the standard measure is defined as:

$$\hat{f}_p = (f_p^\dagger)^\# \circ f \circ p : I \to P\Theta$$

- This definition depends critically on the existence of the Bayesian inverse $f_p^\dagger$.

Existence Conditions for Bayesian Inverses

- Bayesian inverses are a special case of conditionals
- In BorelStoch, conditionals exist as *regular conditional probability distributions*
- But in general categorical settings:
  - Not all Markov categories have conditionals
  - Some categories have conditionals only in special cases
  - Existence can depend on measure-theoretic properties

Why This Matters

- The dependence on Bayesian inverses ties our algebraic characterization of experiment informativeness back to probabilistic notions.
- This undermines our goal of a purely algebraic understanding.

The Need for a Purely Algebraic Approach

The Core Problem

The Blackwell-Sherman-Stein theorem links:

- Blackwell order on experiments (probabilistic notion)
- Stochastic dominance on standard measures (algebraic notion)

But the definition of standard measures still depends on Bayesian inverses!

Our Goal

- Provide an alternative characterization of standard measures/hypernormalizations
- Remove the dependence on Bayesian inverses and conditionals
- Enable purely algebraic comparisons of experiment informativeness
- Maintain the connection to the traditional probabilistic understanding

5

Algebraic Foundations

Weak Pullbacks and Beck-Chevalley Condition

Definition: Weak Pullback

A commutative square is a **weak pullback** if for any compatible morphisms, there exists a morphism making all triangles commute:

$$\begin{tikzcd}
S \arrow[dr, dashed, "r"] \arrow[drr, bend left, "q"] \arrow[ddr, bend right, "p"'] & & \\
& A \arrow[r, "f"] \arrow[d, "g"'] & B \arrow[d, "m"] \\
& C \arrow[r, "n"'] & D
\end{tikzcd}$$

Weak Pullbacks and Beck-Chevalley Condition

Definition: Beck-Chevalley Monad (2.24)

A monad (T, μ, η) is **Beck-Chevalley** (BC) if:

1. The functor T preserves weak pullbacks
2. The multiplication μ is weakly cartesian (naturality squares are weak pullbacks)

Connection to Probability Theory

Theorem: Proposition 2.25

For a Beck-Chevalley monad, the partial evaluation relation is transitive.

Why This Matters for Our Goal

- The partial evaluation relation corresponds to stochastic dominance
- Beck-Chevalley condition ensures stochastic dominance is well-behaved
- This provides the mathematical foundation for comparing decompositions

The Algebraic Path Forward

Current Approach

Statistical Experiment f

$\Downarrow$

Bayesian Inverse $f^\dagger$

$\Downarrow$

Standard Measure $\hat{f}_p$

$\Downarrow$

Experiment Comparison

Our Approach

Statistical Experiment f

$\Downarrow$

Algebraic Properties

$\Downarrow$

Universal Characterization

$\Downarrow$

Experiment Comparison

6

Breaking the Dependence

Universal Property of Hypernormalizations

Theorem: Universal Property (3.11)

Let (Θ, p) be a probability space, $f : \Theta \to X$ a p-a.s. deterministic experiment with Bayesian inverse, and $\hat{f}_p$ its standard measure. For every $\pi : I \to P\Theta$ such that:

1. $\mathrm{samp}_\Theta \circ \pi = p$ (i.e., π is a decomposition of p)
2. $Pf \circ \pi = \delta \circ q$ (i.e., π is compatible with the partition f)

we have a partial evaluation from π to $\hat{f}_p$.

Intuition

- The hypernormalization $\hat{f}_p$ is the "most refined" decomposition of p that respects the partition structure of f.
- Any other compatible decomposition can be "coarse-grained" to obtain $\hat{f}_p$.

Alternative Definition Without Bayesian Inverses

Definition: Alternative Definition (3.12)

Let (Θ, p) be a probability space and $f : \Theta \to X$ be a.s. deterministic. The **hypernormalization** of p with respect to f is the state $\pi : I \to P\Theta$ which:

1. satisfies $\mathsf{samp} \circ \pi = p$ (i.e., it's a decomposition of p)
2. satisfies $Pf \circ \pi = \delta \circ q$ (i.e., it respects the partition of f)
3. is maximal in the stochastic dominance order among states satisfying the above

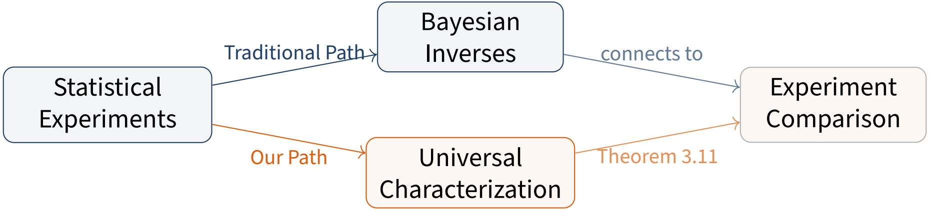
From Probability to Algebra

Traditional Approach

- Starts with probability spaces
- Uses measure theory
- Relies on conditional probability
- Requires Bayesian inverses

Our Algebraic Approach

- Starts with Markov categories
- Uses categorical algebra
- Relies on universal properties
- Based on stochastic dominance



7

Key Technical Results

Theorem 1: Weakly Cartesian Multiplication

Theorem: Multiplication is Weakly Cartesian (3.1)

Let **C** be an a.s.-compatibly representable Markov category with monad (P, μ, δ). If **C** has conditionals, then μ is weakly Cartesian.

Corollary: 3.2

The Giry monad on standard Borel spaces has weakly cartesian multiplication.

Theorem 1: Weakly Cartesian Multiplication

Proof Idea.

We need to show that for deterministic $f : X \to Y$, the following square is a weak pullback:

$$\begin{tikzcd} PX \arrow[r, "Pf"] \arrow[d, "\mathtt{samp}"'] & PY \arrow[d, "\mathtt{samp}"] \\ X \arrow[r, "f"'] & Y \end{tikzcd}$$

Given morphisms $p : A \to X$ and $q : A \to PY$ with $\mathsf{samp}_Y \circ q = f \circ p$, we construct:

$$r = P(f_p^\dagger) \circ \sigma_{Y,A} \circ q : A \to PX$$

where $\sigma_{Y,A}$ is the strength and $f_p^\dagger$ is the Bayesian inverse. $\square$

Theorem 2: Preservation of Weak Pullbacks

Definition: Equalizer Principle (3.5)

A Markov category satisfies the **equalizer principle** if:

- Equalizers in $\mathbf{C}_{\text{det}}$ exist
- For every equalizer diagram, every morphism p with $f =_p g$ factors uniquely across the equalizer

Theorem: Functor Preserves Weak Pullbacks (3.7)

Let $\mathbf{C}$ be a representable Markov category with monad (P, μ, δ). If $\mathbf{C}$ has conditionals and satisfies the equalizer principle, then P preserves weak pullbacks.

Theorem 2: Preservation of Weak Pullbacks

Proof Idea.

- We show that P turns pullbacks into weak pullbacks
- For a pullback in $\mathbf{C}_{\mathrm{det}}$ and compatible morphisms p, q, form their conditional product ρ
- Use the equalizer principle to show ρ factors through the pullback

$\square$

The Giry Monad is Beck-Chevalley

Corollary: Beck-Chevalley Monad (3.8)

If **C** is a.s.-compatibly representable with conditionals and satisfies the equalizer principle, then the monad $(P, P\mathtt{samp}, \delta)$ is Beck-Chevalley.

Theorem: Giry Monad is Beck-Chevalley (3.9)

The Giry monad on standard Borel spaces is Beck-Chevalley.

Connecting the Pieces

Weakly cartesian multiplication → Canonical decompositions of a state → Pure algebraic characterization

enables

justifies

Beck-Chevalley Property

Universal Property of Hypernor-malizations

Bayesian-Free Definition

Breaking the Dependence!

8

Conclusions & Future Work

Summary of Contributions

Main Results

1. **The Giry Monad is Beck-Chevalley**
   - Multiplication is weakly cartesian (Theorem 3.1)
   - Monad preserves weak pullbacks (Theorem 3.7)
2. **Universal Property of Hypernormalizations** (Theorem 3.11)
   - Hypernormalizations are maximal in the stochastic dominance order
   - They are the canonical decompositions respecting a partition
3. **Alternative Definition** (Definition 3.12)
   - Purely algebraic characterization
   - No dependence on Bayesian inverses or conditionals

Open Questions

Immediate Questions

- Suppose we have a representable Markov category without conditionals and whose underlying monad is Beck-Chevalley. Does the Blackwell-Sherman-Stein theorem still hold when we substitute hypernormalization for standard measures?
- Does the existence of hypernormalization imply that the stochastic dominance order is transitive?

Future Directions

Completing the Equivalence

- Prove full equivalence between algebraic and probabilistic approaches
- Establish necessary algebraic conditions for informativeness
- Formalize universality beyond deterministic experiments

Connections to Other Areas

- Explore links to fibrations and descent theory
- Connect partitions to categorical descent
- Develop applications in causal inference

The Big Picture: What other aspects of probability can be algebraicized?

Thank You

Questions?

Key References

- Bohinen & Perrone (2025): Categorical Algebra of Conditional Probability
- Fritz (2020): A synthetic approach to Markov kernels, conditional independence, and theorems on sufficient statistics
- Cho & Jacobs (2019): Disintegration and Bayesian inversion via string diagrams
- Fritz et al. (2023): Representable Markov categories and comparison of statistical experiments in categorical probability
- Constantin et al. (2020): Partial evaluations and the compositional structure of the bar construction