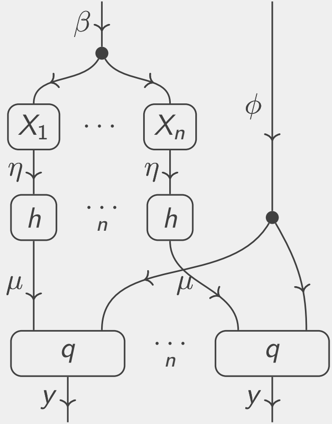
# The Algebra of Statistical Theories and Models

Categorical Probability and Statistics 2020 Workshop

**Evan Patterson**

Statistics Department
Stanford University

June 6, 2020



# Structuralism about statistical models

Statistical models

(1) are *not* black boxes, but have meaningful internal structure
(2) are *not* uniquely determined, but bear meaningful relationships to alternative, competing models
(3) are sometimes purely phenomenological, but are often derived from, or at least motivated by, more general scientific theory

**This project** aims to understand (1) and (2) via categorical logic.

**In this talk**, I assume some knowledge of category theory but as little as I can about statistics.

# Statistical models, classically

A statistical model is a parameterized family $\{P_\theta\}_{\theta \in \Omega}$ of probability distributions on a common space $\mathcal{X}$:

- $\Omega$ is the parameter space
- $\mathcal{X}$ is the sample space

Think of a statistical model as a data-generating mechanism

$$P : \Omega \to \mathcal{X}.$$

Statistical inference aims to approximately invert this mechanism: find an estimator $d : \mathcal{X} \to \Omega$ such that, for any $\theta \in \Omega$,

$$d(X) \approx \theta \qquad \text{given data} \qquad X \sim P_\theta.$$

# Statistical models, classically

This setup goes back to Wald's statistical decision theory (Wald 1939; Wald 1950). Within it, one can already:

- define general concepts like *sufficiency* and *ancillarity*
- establish basic results like the *Neyman-Fisher factorization* and *Basu's theorem*

Recently, Fritz has shown that much of this may be reproduced in a purely synthetic setting (Fritz 2019)

However, the classical definition of statistical model abstracts away a large part of statistics:

(1) formalizes models as black boxes

(2) does not at all formalize relationships between different models

# Logical theories and models

Can logic help formalize the structure of statistical models?

Mathematical logic distinguishes between *theories* and *models*:

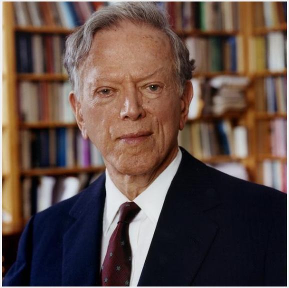
| Logical theory | Model of theory |
|---|---|
| Axiomatic | Constructed |
| Synthetic/qualitative | Analytic/quantitative |
| Formal language | Informal mathematics (usually) |
| Machine representable | Not (directly) representable |

What about *statistical theories* and *statistical models*?

# Models in logic and in science

Not a new idea to draw a connection between models in logic and in statistics (or in science generally).

> *I claim that the concept of model in the sense of Tarski may be used without distortion and as a fundamental concept in all of the disciplines... In this sense I would assert that the meaning of the concept of model is the same in mathematics and the empirical sciences. The difference to be found in these disciplines is to be found in their use of the concept.*
> *(Suppes 1961)*



Patrick Suppes
(1922–2014)

# The semantic view of scientific theories

Suppes initiated the "semantic view" of scientific theories:

- Many different flavors, from different philosophers (van Fraassen, Sneed, Suppe, Suppes, . . . )
- For Suppes, "to axiomatize a theory is to define a set-theoretical predicate" (Suppes 2002)

Difficulties for statistical models and beyond:

- After Suppes, proponents of the semantic view paid little attention to statistics
- Set theory is impractical to implement, esp. with probability
- Hard to make sense of relationships between logical theories

# The algebraization of logic

Beginning with Lawvere's thesis (Lawvere 1963), categorical logic has achieved an *algebraization* of logic:

- Logical theories are replaced by categorical structures
- Obliterates the distinction between *syntax* and *semantics*

Some consequences:

- Theories are *invariant* to presentation
- *Functorial semantics*, especially outside of **Set**
- "Plug-and-play" logical systems, via different categorical gadgets
- Theories have *morphisms*, which formalize relationships

# Dictionary between category theory, logic, and statistics

| Category theory | Mathematical logic | Statistics |
|---|---|---|
| Category $\mathsf{T}$ | Theory | Statistical theory* |
| Functor $M : \mathsf{T} \to \mathsf{S}$ | Model | Statistical model |
| Natural transformation $\alpha : M \to M'$ | Model homomorphism | Morphism of statistical model |

*Statistical theories $(\mathsf{T}, p)$ have extra structure, the sampling morphism $p : \theta \to x$

# Family tree of categorical logics

$$\begin{tikzcd}
\begin{array}{c}\text{symmetric monoidal category}\\ \text{(symmetric monoidal theory)}\end{array} \arrow[d] & \\
\begin{array}{c}\text{cartesian category}\\ \text{(algebraic theory)}\end{array} \arrow[d] \arrow[dr] & \\
\begin{array}{c}\text{regular category}\\ \text{(regular logic: } \exists, \wedge, \top\text{)}\end{array} \arrow[d] & \begin{array}{c}\text{cartesian closed category}\\ \text{(typed } \lambda\text{-calculus with } \times, 1\text{)}\end{array} \arrow[d] \\
\begin{array}{c}\text{coherent category}\\ \text{(coherent logic: } \exists, \wedge, \vee, \top, \bot\text{)}\end{array} \arrow[d] & \begin{array}{c}\text{bicartesian closed category}\\ \text{(typed } \lambda\text{-calculus with } \times, +, 1, 0\text{)}\end{array} \arrow[dl] \\
\begin{array}{c}\text{elementary topos}\\ \text{(first- and higher-order logic)}\end{array} &
\end{tikzcd}$$

# Probability and statistics in the family tree

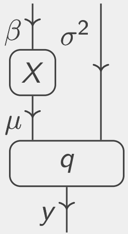
$$
\begin{tikzcd}[column sep=small]
& \text{symmetric monoidal category} \arrow[d] & \\
& \text{Markov category} \arrow[dl] \arrow[dr] & \\
\begin{array}{c}\text{linear algebraic}\\ \text{Markov category}\\ (\textit{statistical theory})\end{array} \arrow[dr] & & \begin{array}{c}\text{cartesian category}\\ \text{(algebraic theory)}\end{array} \arrow[dl] \\
& \begin{array}{c}\text{linear algebraic}\\ \text{cartesian category}\end{array} &
\end{tikzcd}
$$

# Informal example: linear models

A linear model with design matrix $X \in \mathbb{R}^{n \times p}$ has sampling distribution

$$y \sim \mathcal{N}_n(X\beta, \sigma^2 I_n) \quad \text{w/ parameters} \quad \beta \in \mathbb{R}^p,\ \sigma^2 \in \mathbb{R}_+.$$

A theory of a linear model $(\mathrm{LM}, p)$ is generated by objects $y$, $\beta$, $\mu$, $\sigma^2$ and morphisms $X : \beta \to \mu$ and $q : \mu \otimes \sigma^2 \to y$ and has sampling morphism $p$ given by




Then a linear model is a functor $M : \mathrm{LM} \to \mathbf{Stat}$.

# Markov kernels

Statistical theories will have *functorial semantics* in a category of Markov kernels.

**Recall**: A Markov kernel $\mathcal{X} \to \mathcal{Y}$ between measurable spaces $\mathcal{X}, \mathcal{Y}$ is a measurable map $\mathcal{X} \to \mathsf{Prob}(\mathcal{Y})$.

**Examples**:

- A statistical model $(P_\theta)_{\theta \in \Omega}$ is a kernel $P : \Omega \to \mathcal{X}$ (Čencov 1965; Čencov 1982)
- Parameterized distributions, e.g., the normal family

$$\mathcal{N} : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}, \qquad (\mu, \sigma^2) \mapsto \mathcal{N}(\mu, \sigma^2)$$

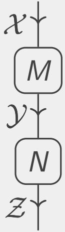
  or, more generally, the $d$-dimensional normal family

$$\mathcal{N}_d : \mathbb{R}^d \times \mathcal{S}_+^d \to \mathbb{R}^d, \qquad (\mu, \Sigma) \mapsto \mathcal{N}_d(\mu, \Sigma).$$

# Synthetic reasoning about Markov kernels

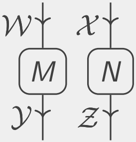
Two fundamental operations on Markov kernels:

1. Composition: For kernels $M : \mathcal{X} \to \mathcal{Y}$ and $N : \mathcal{Y} \to \mathcal{Z}$,




$$(M \cdot N)(dz \,|\, x) := \int_{\mathcal{Y}} N(dz \,|\, y) M(dy \,|\, x)$$

2. Independent product: For kernels $M : \mathcal{W} \to \mathcal{Y}$ and $N : \mathcal{X} \to \mathcal{Z}$,

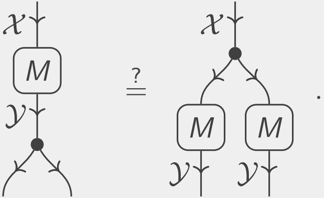



$$(M \otimes N)(w, x) := M(w) \otimes N(x)$$

# Synthetic reasoning about Markov kernels

Also a supply of commutative comonoids, for duplicating and discarding data.

Markov kernels obey all laws of a cartesian category, except one:



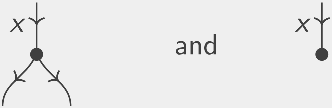

## Proposition

Under regularity conditions, a Markov kernel $M : \mathcal{X} \to \mathcal{Y}$ is **deterministic** if and only if above equation holds.

# Markov categories

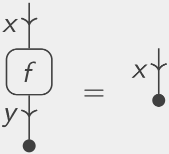
Markov categories are a minimalistic *axiomatization* of categories of Markov kernels (Fong 2012; Cho and Jacobs 2019; Fritz 2019).

**Definition**: A Markov category is a symmetric monoidal category with a supply of commutative comonoids




,

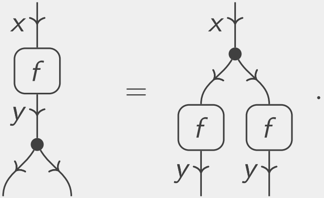
such that every morphism $f : x \to y$ preserves deleting:




.

# Constructions in Markov categories

**Definition**: A morphism $f : x \to y$ in a Markov category is deterministic if




Besides (non)determinism, in a Markov category one can express:

- conditional independence and exchangeability
- disintegration, e.g., for Bayesian inference (Cho and Jacobs 2019)
- many notions of statistical decision theory (Fritz 2019)

# Linear and other spaces for statistical models

In order to specify most statistical models, more structure is needed.

Much statistics happens in Euclidean space or structured subsets thereof:

- real vector spaces
- affine spaces
- convex cones, esp. $\mathbb{R}_+$ or PSD cone $\mathcal{S}_+^d \subset \mathbb{R}^{d \times d}$
- convex sets, esp. $[0, 1]$ or probability simplex $\Delta^d \subset \mathbb{R}^{d+1}$

Also in discrete spaces:

- additive monoids, esp. $\mathbb{N}$ or $\mathbb{N}^k$
- unstructured sets, say $\{1, 2, \ldots, k\}$

# Lattice of linear and other spaces

Such spaces belong to a lattice of symmetric monoidal categories:

$$
\begin{tikzcd}
 & (\mathbf{Cone}, \oplus, 0) \arrow[r] \arrow[ddr] & (\mathbf{CMon}, \oplus, 0) \arrow[dr] & \\
(\mathbf{Vect}_{\mathbb{R}}, \oplus, 0) \arrow[ur] \arrow[dr] & & & (\mathbf{Set}, \times, 1) \\
 & (\mathbf{Aff}_{\mathbb{R}}, \times, 1) \arrow[r] & (\mathbf{Conv}, \times, 1) \arrow[ur] &
\end{tikzcd}
$$

Note:

- **Cone** is category of *conical spaces*, abstracting convex cones
- **Conv** is category of *convex spaces*, abstracting convex sets

# Supplying a lattice of PROPs

Dually, there is a lattice of theories (PROPs):

$$\begin{tikzcd}
 & \mathrm{Th}(\mathbf{CBimon}) \arrow[r, hook] & \mathrm{Th}(\mathbf{Cone}) \arrow[dr, hook] & \\
\mathrm{Th}(\mathbf{CComon}) \arrow[ur, hook] \arrow[dr, hook] & & & \mathrm{Th}(\mathbf{Vect}_{\mathbb{R}}) \\
 & \mathrm{Th}(\mathbf{Conv}) \arrow[r, hook] \arrow[uur, hook] & \mathrm{Th}(\mathbf{Aff}_{\mathbb{R}}) \arrow[ur, hook] &
\end{tikzcd}$$

**Definition**: A supply of a meet-semilattice L of PROPs in a symmetric monoidal category $(\mathsf{C}, \otimes, I)$ consists of a monoid homomorphism

$$\mathsf{P} : (|\mathsf{C}|, \otimes, I) \to (\mathsf{L}, \wedge, \top), \quad x \mapsto \mathsf{P}_x,$$

and for each object $x \in \mathsf{C}$, a strong monoidal functor $s_x : \mathsf{P}_x \to \mathsf{C}$ with $s_x(m) = x^{\otimes m}$, subject to coherence conditions (mildly generalizing Fong and Spivak 2019).

# Linear algebraic Markov categories

**Definition**: A linear algebraic Markov category is a symmetric monoidal category supplying the above lattice of PROPs, such that it is a Markov category.

Linear algebraic Markov categories come

- *in the small*, as statistical theories
- *in the large*, as the semantics of statistical theories

# Category of statistical semantics

The linear algebraic Markov category **Stat** has

- as objects, the pairs $(V, A)$, a finite-dimensional real vector space $V$ with a measurable subset $A \subset V$
- as morphisms $(V, A) \to (W, B)$, the Markov kernels $A \to B$
- a symmetric monoidal structure, given by

  $$(V, A) \otimes (V, B) := (V \oplus W, A \times B), \qquad I := (0, \{0\})$$

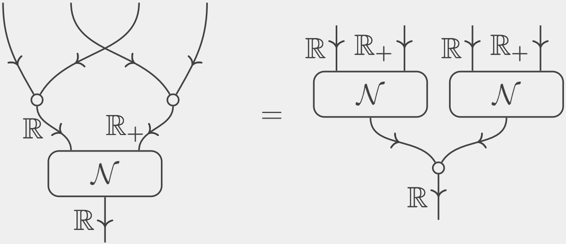
  and by the independent product of Markov kernels
- a supply according to whether the subset $A$ is closed under linear/affine/conical/convex combinations, addition, or nothing.

# Additivity of normal family

Normal family is additive: if $X_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu_i, \sigma_i^2)$, then

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

In **Stat**, additivity is the equation:

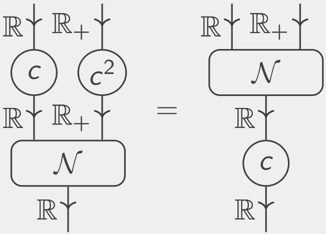


# Homogeneity of normal family

Normal family is homogeneous: if $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$cX \sim \mathcal{N}(c\mu, c^2\sigma^2), \qquad \forall c \in \mathbb{R}.$$

In **Stat**, homogeneity is the equation:




$\forall c \in \mathbb{R}.$

Call both properties together "linear-quadratic".

# Presenting the normal family

Isotropic normal family can be presented by generators and relations.

## Theorem

For any $d \geq 1$, a linear algebraic Markov category $\mathsf{C}$, containing a morphism $f : y^{\otimes d} \otimes s \to y^{\otimes d}$, can presented such that for any supply preserving functor

$$M : \mathsf{C} \to \mathbf{Stat} \quad \text{with} \quad M(y) = \mathbb{R} \text{ and } M(s) = \mathbb{R}_+,$$

the Markov kernel $M(f) : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d$ is the isotropic normal family, up to an absolute scale.

That is, there exists $\sigma_0^2 \in \mathbb{R}_+$ such that

$$M(f)(\mu, \phi) = \mathcal{N}_d(\mu, \phi\sigma_0^2 I_d), \quad \forall \mu \in \mathbb{R}^d,\ \phi \in \mathbb{R}_+.$$

# Characterizations of normal distribution

Main ingredient is the symmetry of the normal distribution.

**Theorem (Maxwell)**

For any $d \geq 2$, a random vector $Y \in \mathbb{R}^d$ has i.i.d. centered normal distribution if and only if $Y$ is spherically symmetric and has independent components.

Or, simpler:

**Theorem (Pólya 1923)**

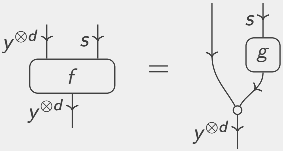
If $X$ and $Y$ are i.i.d. random variables such that

$$X \overset{d}{=} \frac{1}{\sqrt{2}}(X + Y),$$

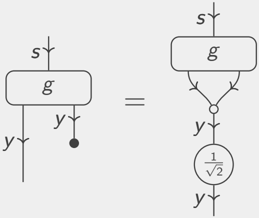
then $X$ is centered normal.

# Sketch of presentation

1. Reduce to centered case via generator $g : s \to y^{\otimes d}$:




2. Assert that $g$ is homogeneous, in above sense
3. Assert that $g$ has independent (or i.i.d.) components
4. Axiomatize Maxwell's or Pólya's theorem, e.g., when $d = 2$,



# Statistical theories and models

A statistical theory $(\mathsf{T}, p)$ consists of

- a small linear algebraic Markov category $\mathsf{T}$
- a morphism $p : \theta \to x$ in $\mathsf{T}$, the *sampling morphism*

A model of a statistical theory $(\mathsf{T}, p)$ is a supply preserving functor $M : \mathsf{T} \to \mathbf{Stat}$.

- $\Omega := M(\theta)$ is the *parameter space*
- $\mathcal{X} := M(x)$ is the *sample space*
- $P := M(p) : \Omega \to \mathcal{X}$ is the *sampling distribution*

**Note**: Statistical theories generally have many different models.

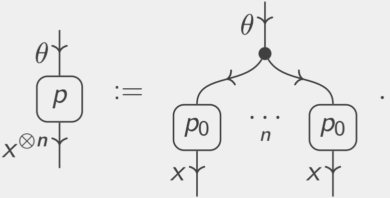
# A few simple statistical theories

**Example**: The initial statistical theory $(\mathsf{T}, p)$ is freely generated by one morphism $p : \theta \to x$ on discrete objects $\theta$ and $x$.

## Observation

Every statistical model $P : \Omega \to \mathcal{X}$ is a model of the initial theory.

**Example**: The theory of $n$ i.i.d. samples $(\mathsf{T}, p)$ is freely generated by one morphism $p_0 : \theta \to x$ on discrete objects $\theta$ and $x$, with

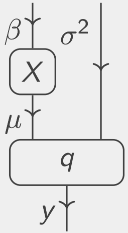


# Theory of a linear model

The theory of a linear model $(\mathsf{LM}, p)$ is generated by

- vector space objects $\beta$, $\mu$, and $y$
- conical space object $\sigma^2$
- linear map $X : \beta \to \mu$, i.e., morphism $X : \beta \to \mu$ subject to equations of linearity and determinism
- linear-quadratic morphism $q : \mu \otimes \sigma^2 \to y$

with sampling morphism $p : \beta \otimes \sigma^2 \to y$ given by



# Linear models, as models of a theory

The standard models $M : \mathrm{LM} \to \mathbf{Stat}$ are linear models:

- $M(y) = M(\mu) = \mathbb{R}^n$ for some dimension $n$
- $M(\beta) = \mathbb{R}^p$ for some dimension $p$
- $M(\sigma^2) = \mathbb{R}_+$
- $M(q) : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^n$ is isotropic normal family
- $X_M := M(X) : \mathbb{R}^p \to \mathbb{R}^n$ is arbitrary linear map

The sampling distribution is then

$$M(p) : \begin{array}{l} \mathbb{R}^p \times \mathbb{R}_+ \to \mathbb{R}^n \\ (\beta, \sigma^2) \mapsto \mathcal{N}_n(X_M \beta, \sigma^2 I_n) \end{array}$$

Another model is a weighted linear model where, for fixed $V_M \in \mathcal{S}_+^p$,

$$M(p) : (\beta, \sigma^2) \mapsto \mathcal{N}_n(X_M \beta, \sigma^2 V_M).$$

# Bayesian statistical theories and models

A Bayesian statistical theory $(\mathsf{T}, p, \pi)$ consists of

- a statistical theory $(\mathsf{T}, \theta \xrightarrow{p} x)$
- a morphism $I \xrightarrow{\pi} \theta$, the *prior morphism*

A model of a Bayesian theory $(\mathsf{T}, p, \pi)$ is a model $M : \mathsf{T} \to \mathbf{Stat}$ of the underlying statistical theory.

- $M(p)$ is the *sampling distribution*
- $M(\pi)$ is the *prior*
- $M(\pi \cdot p)$ is the *marginal* or *prior predictive distribution*

# Morphisms of statistical models

A model homomorphism between models $M$ and $M'$ of a statistical theory $(\mathsf{T}, p)$ is a monoidal natural transformation

$$\begin{tikzcd} \mathsf{T} \arrow[r, bend left, "M"{name=U}] \arrow[r, bend right, "M'"'{name=D}] & \mathbf{Stat} \arrow[from=U, to=D, Rightarrow, "\alpha"'] \end{tikzcd}$$

## Proposition

The components $\alpha_x : M(x) \to M'(x)$ of model homomorphism are supply homomorphisms. In particular, they are **deterministic**.

# Morphisms of linear models

Let $M, M'$ be linear models, as models of $(\mathsf{LM}, p)$, with designs

$$X_M := M(X) \in \mathbb{R}^{n \times p}, \qquad X_{M'} := M'(X) \in \mathbb{R}^{n' \times p'}.$$

## Proposition

A model homomorphism $\alpha : M \to M'$ is uniquely determined by linear maps $A := \alpha_y \in \mathbb{R}^{n' \times n}$ and $B := \alpha_\beta \in \mathbb{R}^{p' \times p}$ such that

$$AA^\top \propto I_{n'} \qquad \text{and} \qquad AX_M = X_{M'} B.$$

# Symmetries of statistical models

## Corollary

An isomorphism of linear models $\alpha : M \cong M'$, with $n = n'$ and $p = p'$, is uniquely determined by linear maps $A := \alpha_y \in \mathrm{CO}(n)$ and $B := \alpha_\beta \in \mathrm{GL}(p, \mathbb{R})$ such that $X_{M'} = A X_M B^{-1}$.

Symmetry and invariance is a classical topic in statistics. Advantages of our account:

- it does not assume identifiability of model (or loss function)
- it is not restricted to automorphisms or even isomorphisms
- it ensures that transformations preserve *all* structure specified by the theory, not just parameter and sample spaces
- it makes symmetry a property of the theory and model, not an extra structure added arbitrarily

# Equivariance of linear regression

Ordinary-least squares (OLS) linear regression is

- equivariant under model isomorphism (a classical result)
- "laxly" equivariant under model homomorphism

## Theorem

Let $\alpha : M \to M'$ be a homomorphism of linear models. For any $y \in \mathbb{R}^n$ and $\beta \in \mathbb{R}^p$, if $y' := \alpha_y(y)$ and $\beta' := \alpha_\beta(\beta)$, then

$$\|X_{M'}\beta' - y'\| \leq a\|X_M\beta - y\|,$$

where $a := \sqrt{\alpha_{\sigma^2}} \in \mathbb{R}_+$. In particular, if $\alpha$ is an isomorphism, then

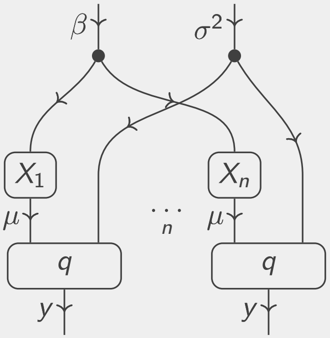
$$\hat{\beta} \in \operatorname*{argmin}_{\beta \in \mathbb{R}^p} \|X_M\beta - y\| \qquad \text{implies} \qquad \hat{\beta}' \in \operatorname*{argmin}_{\beta' \in \mathbb{R}^{p'}} \|X_{M'}\beta' - y'\|.$$

# Another theory of a linear model

The theory of a linear model on $n$ samples $(\mathsf{LM}_n, p_n)$ is generated by

- vector spaces $\beta$, $\mu$, and $y$ and a conical space $\sigma^2$
- linear maps $X_1, \dots, X_n : \beta \to \mu$,
- a linear-quadratic morphism $q : \mu \otimes \sigma^2 \to y$

with sampling morphism $p_n : \beta \otimes \sigma^2 \to y^{\otimes n}$ given by



# Linear models on $n$ samples

A linear model $M : \mathrm{LM}_n \to \mathbf{Stat}$ now assigns

- $M(y) = M(\mu) = \mathbb{R}$
- $M(q) = \mathcal{N} : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$, the univariate normal family

Let $M, M'$ be linear models on $n$ samples, as models of $(\mathrm{LM}_n, p_n)$, with designs $(X_{M,i})_{i=1}^n \in \mathbb{R}^{n \times p}$ and $(X_{M',i})_{i=1}^n \in \mathbb{R}^{n \times p'}$.

## Proposition

A model homomorphism $\alpha : M \to M'$ is uniquely determined by a scalar $a := \alpha_y \in \mathbb{R}$ and a matrix $B := \alpha_\beta \in \mathbb{R}^{p' \times p}$ such that

$$aX_{M,i} = X_{M',i}B, \quad \forall i = 1, \ldots, n.$$

# More theories of a linear model

Theories of a linear model include

- $(\mathsf{LM}, p)$, of a general linear model
- $(\mathsf{LM}_n, p_n)$, of a LM on $n$ observations
- $(\mathsf{LM}_p, q_p)$, of a LM on $p$ predictors
- $(\mathsf{LM}_{n,p}, q_{n,p})$, of a LM on $n$ observations and $p$ predictors

Which theory is the right one? *Wrong question*.

- Different theories allow different models and model homomorphisms
- Yet they are related by morphisms of theories

# Morphisms of statistical theories

**Definition**: A (strict) morphism of statistical theories

$$F : (\mathsf{T}, p) \to (\mathsf{T}', p')$$

is a supply preserving functor $F : \mathsf{T} \to \mathsf{T}'$ such that $F(p) = p'$.

The theory morphism induces a model migration functor

$$F^* : \mathrm{Mod}(\mathsf{T}') \to \mathrm{Mod}(\mathsf{T})$$

(cf. Spivak 2012) by *pre-composition*:

$$\begin{tikzcd} \mathsf{T}' \arrow[d, "M"'] \\ \mathbf{Stat} \end{tikzcd} \quad \overset{F^*}{\longmapsto} \quad \begin{tikzcd} \mathsf{T} \arrow[r, "F"] \arrow[dr, dashed, "F^*(M)"'] & \mathsf{T}' \arrow[d, "M"] \\ & \mathbf{Stat} \end{tikzcd}$$

# Morphisms between theories of linear model

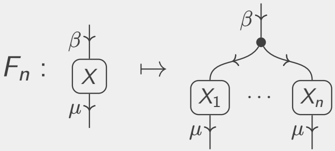
Different theories of linear models are related by theory morphisms:

$$\begin{tikzcd}
 & \underset{\text{general LM}}{(\mathrm{LM}, p)} \arrow[dl, "F_n"'] \arrow[dr, "G_p"] & \\
\underset{\text{LM with } n \text{ observations}}{(\mathrm{LM}_n, p_n)} \arrow[dr, "F_{n,p}"'] & & \underset{\text{LM with } p \text{ predictors}}{(\mathrm{LM}_p, q_p)} \arrow[dl, "G_{n,p}"] \\
 & \underset{\text{LM with } n \text{ observations and } p \text{ predictors}}{(\mathrm{LM}_{n,p}, q_{n,p})} &
\end{tikzcd}$$

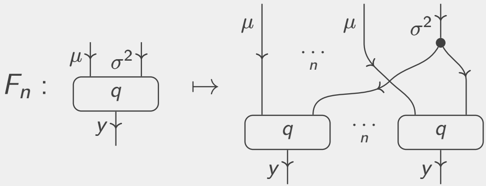
# Morphism between two theories of linear model

A theory morphism $F_n : (\mathrm{LM}, p) \to (\mathrm{LM}_n, p_n)$

- sends $\mu$ to $\mu^{\otimes n}$ and $y$ to $y^{\otimes n}$
- splits the design matrix by rows:




- splits the morphism $q$ accordingly:

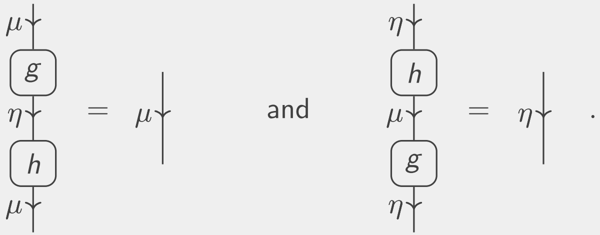



- preserves the other generators

# Theory of a generalized linear model

A theory of a GLM on $n$ samples $(\mathrm{GLM}_n, p_n)$ is generated by

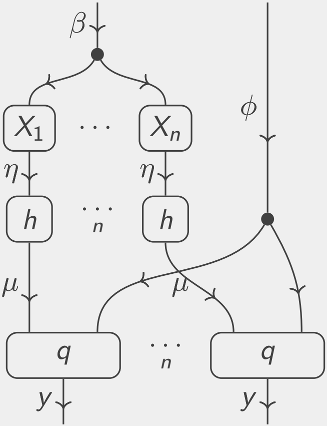
- vector spaces $\beta$ and $\eta$, a convex space $\mu$, and a conical space $\phi$
- a discrete object $y$
- maps $g : \mu \to \eta$ (link function) and $h : \eta \to \mu$ (mean function), which are mutually inverse:




- linear maps $X_1, \dots, X_n : \beta \to \eta$
- a morphism $q : \mu \otimes \phi \to y$

# Theory of a generalized linear model

The sampling morphism $p_n : \beta \otimes \phi \to y^{\otimes n}$ is



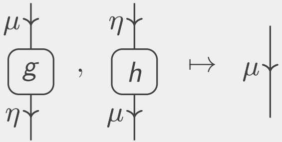
# Morphism between theories of GLM and LM

**Fact**: "A linear model is a special case of a generalized linear model."

Formally, a theory morphism $G_n : (\mathrm{GLM}_n, p_n) \to (\mathrm{LM}_n, p_n)$

- sends both $\mu$ and $\eta$ to $\mu$,
- sends both $g$ and $h$ to the identity $1_\mu$:

$G_n :$




- sends $\phi$ to $\sigma^2$
- preserves the other generators

Induces a model migration functor $G_n^* : \mathrm{Mod}(\mathrm{LM}_n) \to \mathrm{Mod}(\mathrm{GLM}_n)$.

# Lax morphisms of statistical theories

A weaker notion of theory morphism allows for *expansion* of parameter and sample spaces (cf. McCullagh 2002).

A lax* morphism of statistical theories $(\mathsf{T}, \theta \xrightarrow{p} x)$ and $(\mathsf{T}', \theta' \xrightarrow{p'} x')$ consists of

- a functor $F : \mathsf{T} \to \mathsf{T}'$
- a morphism $f_0 : \theta' \to F(\theta)$ in $\mathsf{T}'$
- a morphism $f_1 : x' \to F(x)$ in $\mathsf{T}'$

such that the diagram commutes:

$$\begin{tikzcd}
\theta' \arrow[r, "p'"] \arrow[d, "f_0"'] & x' \arrow[d, "f_1"] \\
F\theta \arrow[r, "Fp"] & Fx
\end{tikzcd}$$

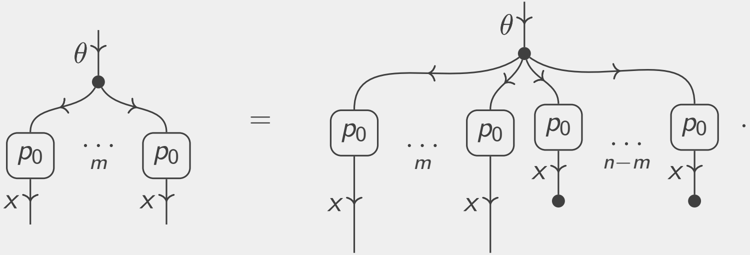
*Called "colax", not "lax," in (Patterson 2020)

# Samples of different sizes as lax theory morphisms

Recall the theory of $n$ i.i.d. samples $(\mathsf{T}, p_n)$. For any numbers $m \leq n$, projection gives a lax theory morphism

$$(1_{\mathsf{T}}, 1_\theta, \pi_{m,n-m}) : (\mathsf{T}, p_m) \to (\mathsf{T}, p_n),$$

where laxness condition is



# Conclusion

**Summary**:

1. introduced *statistical theories* in style of categorical logic
2. recovered *statistical models* as models of statistical theories
3. obtained notion of *statistical model homomorphism*
4. formalized relationships using *morphisms of statistical theories*
5. accompanied by *model migration functors*

**Future work**: lots!

- mathematical investigation of linear algebraic Markov categories
- *compositionality* of statistical theories and models
- *software* and integration with probabilistic programming

# Outlook

How can statistics support scientific theories and models broadly?

- Traditionally, statistics has emphasized the formal testing of null hypotheses, as if they exist in isolation
- Rather, science involves an intricate web of interconnected theories, models, experiments, and data

Again, a long precedent in philosophy of science:

> *[E]xact analysis of the relation between empirical theories and relevant data calls for a hierarchy of models of different logical type.* (Suppes 1966)

Suppes' hierarchy of models:

1. theoretical model
2. model of the experiment
3. data model [roughly, a statistical model]

How to make mathematics and statistics out of such ideas?

# Thanks!

**Main reference** is my PhD thesis (Patterson 2020)

- Available at `https://www.epatters.org/papers/`
- Many more examples of statistical theories and models:
  - contingency tables
  - simple Bayesian and hierarchical models
  - linear mixed models
  - generalized linear (mixed) models
  - ...

# References I


Čencov, N. N. (1965). "The categories of mathematical statistics". *Dokl. Akad. Nauk SSSR* 164.3. In Russian, pp. 511–514.

– (1982). *Statistical decision rules and optimal inference*. Translations of Mathematical Monographs 53. American Mathematical Society.

Cho, Kenta and Bart Jacobs (2019). "Disintegration and Bayesian inversion via string diagrams". *Mathematical Structures in Computer Science* 29.7, pp. 938–971. DOI: 10.1017/S0960129518000488.

Fong, Brendan (2012). "Causal theories: A categorical perspective on Bayesian networks". MSc thesis. University of Oxford. arXiv: 1301.6201.

Fong, Brendan and David I. Spivak (2019). "Supplying bells and whistles in symmetric monoidal categories". arXiv: 1908.02633.

Fritz, Tobias (2019). "A synthetic approach to Markov kernels, conditional independence, and theorems on sufficient statistics". arXiv: 1908.07021.

# References II


Lawvere, F. William (1963). “Functorial semantics of algebraic theories”. Republished in *Reprints in Theory and Applications of Categories*, No. 5 (2004), pp. 1–121. PhD thesis. Columbia University.

McCullagh, Peter (2002). “What is a statistical model?” *Annals of Statistics* 30.5, pp. 1225–1267. DOI: `10.1214/aos/1035844977`.

Patterson, Evan (2020). “The algebra and machine representation of statistical models”. PhD thesis. Stanford University.

Pólya, Georg (1923). “Herleitung des Gaußschen Fehlergesetzes aus einer Funktionalgleichung”. *Mathematische Zeitschrift* 18.1, pp. 96–108. DOI: `10.1007/BF01192398`.

Spivak, David I. (2012). “Functorial data migration”. *Information and Computation* 217, pp. 31–51. DOI: `10.1016/j.ic.2012.05.001`. arXiv: `1009.1166`.

# References III


Suppes, Patrick (1961). "A comparison of the meaning and uses of models in mathematics and the empirical sciences". *The concept and the role of the model in mathematics and natural and social sciences*, pp. 163–177. DOI: 10.1007/978-94-010-3667-2_16.

– (1966). "Models of data". *Studies in logic and the foundations of mathematics*, pp. 252–261. DOI: 10.1016/S0049-237X(09)70592-0.

– (2002). *Representation and invariance of scientific structures*. Online edition. CSLI Publications.

Wald, Abraham (1939). "Contributions to the theory of statistical estimation and testing hypotheses". *The Annals of Mathematical Statistics* 10.4, pp. 299–326. DOI: 10.1214/aoms/1177732144.

– (1950). *Statistical decision functions*. Wiley.