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.
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

\[
P : \Omega \rightarrow \mathcal{X}.
\]

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

\[
d(X) \approx \theta \quad \text{given data} \quad X \sim P_\theta.
\]
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:

<table>
<thead>
<tr>
<th>Logical theory</th>
<th>Model of theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axiomatic</td>
<td>Constructed</td>
</tr>
<tr>
<td>Synthetic/qualitative</td>
<td>Analytic/quantitative</td>
</tr>
<tr>
<td>Formal language</td>
<td>Informal mathematics (usually)</td>
</tr>
<tr>
<td>Machine representable</td>
<td>Not (directly) representable</td>
</tr>
</tbody>
</table>

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)
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

<table>
<thead>
<tr>
<th>Category theory</th>
<th>Mathematical logic</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category $T$</td>
<td>Theory</td>
<td>Statistical theory*</td>
</tr>
<tr>
<td>Functor $M : T \to S$</td>
<td>Model</td>
<td>Statistical model</td>
</tr>
<tr>
<td>Natural transformation $\alpha : M \to M'$</td>
<td>Model homomorphism</td>
<td>Morphism of statistical model</td>
</tr>
</tbody>
</table>

*Statistical theories $(T, p)$ have extra structure, the sampling morphism $p : \theta \to x$*
Family tree of categorical logics

- symmetric monoidal category
  (symmetric monoidal theory)
  ↓
- cartesian category
  (algebraic theory)
  ↓
- regular category
  (regular logic: $\exists, \land, \top$)
  ↓
- coherent category
  (coherent logic: $\exists, \land, \lor, \top, \bot$)
  ↓
- elementary topos
  (first- and higher-order logic)
- cartesian closed category
  (typed $\lambda$-calculus with $\times, 1$)
  ↓
- bicartesian closed category
  (typed $\lambda$-calculus with $\times, +, 1, 0$)
Probability and statistics in the family tree

- symmetric monoidal category
- Markov category
- linear algebraic Markov category (statistical theory)
- cartesian category (algebraic theory)
- linear algebraic cartesian category
Informal example: linear models

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

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

A **theory of a linear model** $(\text{LM}, p)$ 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 $M : \text{LM} \to \text{Stat}$. 
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 $\mathcal{X} \to \text{Prob}(\mathcal{Y})$.

**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

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

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

\[ \mathcal{N}_d : \mathbb{R}^d \times S^d_+ \to \mathbb{R}^d, \quad (\mu, \Sigma) \mapsto \mathcal{N}_d(\mu, \Sigma). \]
Synthetic reasoning about Markov kernels

Two fundamental operations on Markov kernels:

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

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

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:

\[
\begin{array}{c}
\xrightarrow{M} \\
\downarrow \\
\xrightarrow{Y}
\end{array}
\quad \quad \text{=} 
\quad \begin{array}{c}
\xrightarrow{M} \\
\downarrow \\
\xrightarrow{M}
\end{array}
\quad \begin{array}{c}
\xrightarrow{M} \\
\downarrow \\
\xrightarrow{M}
\end{array}.
\]

**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

\[ x \rightarrow y \]

such that every morphism \( f : x \rightarrow y \) preserves deleting:

\[ x \rightarrow y = \bullet \]
**Definition:** A morphism $f : x \to y$ in a Markov category is deterministic if

\[
\begin{array}{c}
x \\
\downarrow \\
f \\
\downarrow \\
y \\
\downarrow \\
y
\end{array}
\quad =
\begin{array}{c}
x \\
\downarrow \\
f \\
\downarrow \\
y \\
\downarrow \\
y
\end{array}.
\]

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)
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 \( S^d_+ \subset \mathbb{R}^{d \times d} \)
- 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:

\[(\text{Cone}, \oplus, 0) \rightarrow (\text{CMon}, \oplus, 0)\]
\[(\text{Vect}_\mathbb{R}, \oplus, 0) \rightarrow (\text{Set}, \times, 1)\]
\[(\text{Aff}_\mathbb{R}, \times, 1) \rightarrow (\text{Conv}, \times, 1)\]

Note:
- **Cone** is category of *conical spaces*, abstracting convex cones
- **Conv** is category of *convex spaces*, abstracting convex sets
Dually, there is a lattice of theories (PROP): 

\[
\begin{align*}
\text{Th}(\text{CBimon}) & \leftrightarrow \text{Th}(\text{Cone}) \\
\text{Th}(\text{CComon}) & \leftrightarrow \text{Th}(\text{Vect}_R) \\
\text{Th}(\text{Conv}) & \leftrightarrow \text{Th}(\text{Aff}_R)
\end{align*}
\]

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

\[
P : (|C|, \otimes, I) \to (L, \wedge, \top), \quad x \mapsto P_x,
\]

and for each object $x \in C$, a strong monoidal functor $s_x : P_x \to C$ with $s_x(m) = x^{\otimes m}$, subject to coherence conditions (mildly generalizing Fong and Spivak 2019).
**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
The linear algebraic Markov category \textbf{Stat} has

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

\[(V, A) \otimes (V, B) := (V \oplus W, A \times B), \quad I := (0, \{0\})\]

and by the \textbf{independent product} of Markov kernels

- a supply according to whether the subset \(A\) is \textbf{closed} under linear/affine/conical/convex combinations, addition, or nothing.
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

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

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

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 $C$, containing a morphism $f : y \otimes^d \otimes s \to y \otimes^d$, can presented such that for any supply preserving functor

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

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

\[
\begin{align*}
  y \otimes d & \downarrow s \\
  y \otimes d & \downarrow y \\
  f & \\
  y \otimes d & \downarrow y \\
  = & \\
  y \otimes d & \downarrow y
\end{align*}
\]

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 $d = 2$,
A statistical theory \((T, p)\) consists of

- a small linear algebraic Markov category \(T\)
- a morphism \(p : \theta \to \times\) in \(T\), the sampling morphism

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

- \(\Omega := M(\theta)\) is the parameter space
- \(\times := M(\times)\) is the sample space
- \(P := M(p) : \Omega \to \times\) is the sampling distribution

Note: Statistical theories generally have many different models.
A few simple statistical theories

**Example:** The initial statistical theory \((T, p)\) 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 \((T, p)\) is freely generated by one morphism \(p_0 : \theta \to x\) on discrete objects \(\theta\) and \(x\), with

\[
p \quad := \quad p_0 \quad \cdots \quad p_0
\]
The **theory of a linear model** (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 $M : \text{LM} \rightarrow \text{Stat}$ are linear models:

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

The sampling distribution is then

$$M(p) : \mathbb{R}^p \times \mathbb{R}_+ \rightarrow \mathbb{R}^n$$

$$(\beta, \sigma^2) \mapsto \mathcal{N}_n(X_M\beta, \sigma^2 I_n)$$

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

$$M(p) : (\beta, \sigma^2) \mapsto \mathcal{N}_n(X_M\beta, \sigma^2 V_M).$$
A **Bayesian statistical theory** $\langle T, p, \pi \rangle$ consists of

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

A **model** of a Bayesian theory $\langle T, p, \pi \rangle$ is a model $M : T \rightarrow \text{Stat}$ 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*
A model homomorphism between models $M$ and $M'$ of a statistical theory $(T, p)$ is a monoidal natural transformation 

\[ T \xrightarrow{\alpha} \text{Stat} \]

**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 $(LM, p)$, with designs

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

**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

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

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

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 $\alpha : M \rightarrow M'$ 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

$$\hat{\beta} \in \arg\min_{\beta \in \mathbb{R}^p} \|X_M \beta - y\| \quad \text{implies} \quad \hat{\beta}' \in \arg\min_{\beta' \in \mathbb{R}^{p'}} \|X_{M'} \beta' - y'\|. $$
Another theory of a linear model

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

- vector spaces \( \beta, \mu, \) and \( y \) and a conical space \( \sigma^2 \)
- linear maps \( X_1, \ldots, X_n : \beta \rightarrow \mu, \)
- a linear-quadratic morphism \( q : \mu \otimes \sigma^2 \rightarrow y \)

with sampling morphism \( p_n : \beta \otimes \sigma^2 \rightarrow y \otimes \overset{n}{\cdots} \) given by
A linear model $M : LM_n \rightarrow \text{Stat}$ now assigns

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

Let $M, M'$ be linear models on $n$ samples, as models of $(LM_n, p_n)$, 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 $\alpha : M \rightarrow M'$ 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

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

\[ F : (T, p) \rightarrow (T', p') \]

is a supply preserving functor \( F : T \rightarrow T' \) such that \( F(p) = p' \).

The theory morphism induces a model migration functor

\[ F^* : \text{Mod}(T') \rightarrow \text{Mod}(T) \]

(cf. Spivak 2012) by pre-composition:

\[
\begin{array}{ccc}
T' & \xrightarrow{F^*} & T \\
M \downarrow & & \downarrow F \\
\text{Stat} & \xrightarrow{F^*} & \text{Stat} \\
\end{array}
\]

\[
\begin{array}{ccc}
T & \xrightarrow{F} & T' \\
\downarrow & & \downarrow \\
\text{Stat} & \xrightarrow{F^*} & \text{Stat} \\
\end{array}
\]
Different theories of linear models are related by theory morphisms:

\[ \text{LM with } n \text{ observations} \]
\[ \text{LM with } p \text{ predictors} \]

\[ \text{LM with } n \text{ observations and } p \text{ predictors} \]
Morphism between two theories of linear model

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

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

\[
F_n : \begin{array}{ccc}
\vcenter{\hbox{\includegraphics[width=.2\textwidth]{design_matrix}}}& \rightarrow & \vcenter{\hbox{\includegraphics[width=.2\textwidth]{design_matrix_split}}}
\end{array}
\]

- splits the morphism $q$ accordingly:

\[
F_n : \begin{array}{ccc}
\vcenter{\hbox{\includegraphics[width=.2\textwidth]{variance_matrix}}}& \rightarrow & \vcenter{\hbox{\includegraphics[width=.2\textwidth]{variance_matrix_split}}}
\end{array}
\]

- preserves the other generators
Theory of a generalized linear model

A theory of a GLM on $n$ samples $(\text{GLM}_n, p_n)$ 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:

\[
\begin{align*}
g \quad \mu & \quad = \quad \mu \\
h \quad \eta & \quad = \quad \eta
\end{align*}
\]

- linear maps $X_1, \ldots, X_n : \beta \to \eta$
- a morphism $q : \mu \otimes \phi \to y$
Theory of a generalized linear model

The sampling morphism $p_n : \beta \otimes \phi \rightarrow y \otimes^n$ 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 $G_n : (\text{GLM}_n, p_n) \rightarrow (\text{LM}_n, p_n)$

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

$$G_n : \begin{array}{c}
\mu \\
\eta
\end{array} \rightarrow \begin{array}{c}
\eta \\
\mu
\end{array}$$

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

Induces a model migration functor $G_n^* : \text{Mod}(\text{LM}_n) \rightarrow \text{Mod}(\text{GLM}_n)$. 
Lax morphisms of statistical theories


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

- a functor \(F : T \rightarrow T'\)
- a morphism \(f_0 : \theta' \rightarrow F(\theta)\) in \(T'\)
- a morphism \(f_1 : x' \rightarrow F(x)\) in \(T'\)

such that the diagram commutes:

\[
\begin{array}{ccc}
\theta' & \xrightarrow{p'} & x' \\
\downarrow{f_0} & & \downarrow{f_1} \\
F\theta & \xrightarrow{Fp} & Fx
\end{array}
\]

*Called “colax”, not “lax,” in (Patterson 2020)
Samples of different sizes as lax theory morphisms

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

\[(1_T, 1_{\theta}, \pi_{m, n-m}) : (T, p_m) \to (T, p_n),\]

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
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 \right]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/](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
  - ...


