Categorical Probability and Statistics

Peter McCullagh

Department of Statistics
University of Chicago

June 5 2020

Categorical Probability and Statistics

Where is this speaker coming from?

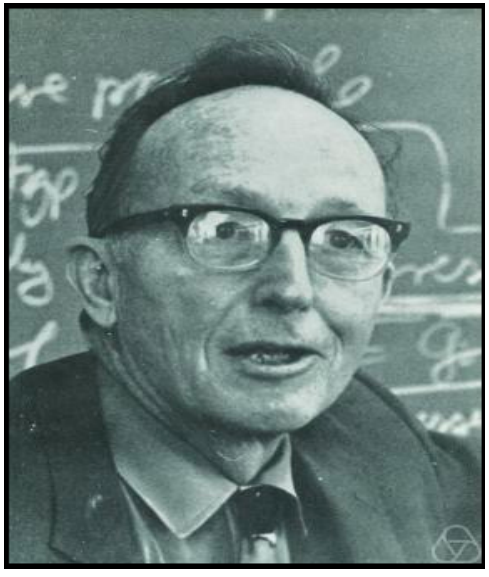
Randomness, repetitive structures, stochastic processes
- Samples and sub-samples; selection
- Simple random samples and sub-samples
- Sample values; symmetric functions;
- cumulants, k-statistics and polykays
- Inheritance under simple random sampling

spectral samples; spectral k-statistics, free cumulants

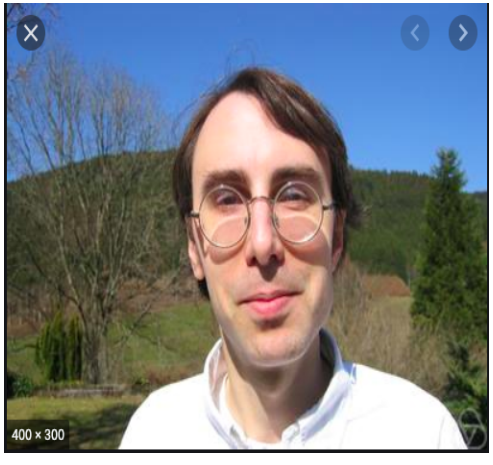
Experimental design and structured samples; Factorial design
- Linear models and factorial subspaces
- Symmetry and group representations
- Marginality and category representations

Kolmogorov consistency
- Projective systems and infinite exchangeability

Recollections of Saunders MacLane 1909–2005



Semi-regular at the Quad-Club lunch
Frequently joined the Stats table
Very strong views on myriad topics
Views freely expressed
Occasionally mentioned category theory
Had no interest in prob or stats
Had no interest in applications of math
Would undoubtedly regard this talk as trivial



Saunders was a curmudgeon, usually friendly
S was an extrovert
He loved debate, argument, controversy

I learned about categories from Burt Totaro
Also representation theory for categories
Burt is the opposite of Saunders

Categorical Probability and Statistics

Samples and sub-samples

Universe: a set $\mathcal{U}$ of observational units a.k.a population
the items (humans/mice/rats/drosophila/...) being studied
the sample $U \subset \mathcal{U}$ actually chosen: $(\#U < \infty)$

process: to each $u \in \mathcal{U}$ there corresponds a value Y_u
observation: to each $u \in U$ there corresponds an obs Y_u
e.g., $Y_u \in \{0, 1\}$ (Covid-19 status)
or $Y_u \in \mathbb{R}$ (height or weight or temp)
or $Y_u \in \mathbb{R}^2$ (systolic, diastolic)

Goal of statistics:
given $Y : U \to \mathbb{R}$ observed on sample
What can we say about Y_u for extra-sample $u \in \mathcal{U} \setminus U$?
—stochastic process

Exchangeability and symmetric functions

Equivalent samples: $\varphi\colon U' \to U$ (bijection)

$n = \#U$ (sample size)

—all samples of the same size are equivalent (same distribution)

Observation $Y\colon U \to \mathbb{R}$; $\quad Y \in \mathbb{R}^U \cong \mathbb{R}^n$

Symmetric function $h\colon \mathbb{R}^n \to \mathbb{R}$ as a statistical summary

$h(y_1, \ldots, y_n) = h(y_{\sigma(1)}, \ldots, y_{\sigma(n)})$

examples $h(y) = y_\bullet = y_1 + \cdots + y_n$

$h(y) = \bar{y}_n = (y_1 + \cdots + y_n)/n$

$h(y) = \sum (y_i - \bar{y}_n)^2/n$

$h(y) = s_n^2 = \sum (y_i - \bar{y}_n)^2/(n-1)$

The statistical problem with symmetric functions ...

—The equivalence classes are isolated

—nothing to connect samples of size 5 with samples of size 6

Simple random sampling

A s.r.s. of size n taken from 'population' $[N] = \{1, \ldots, N\}$

(conventional) All subsets of size n have equal probability

(for today) each $\varphi\colon [n] \to [N]$ is 1–1 with probability $1/N^{\downarrow n}$

$$N^{\downarrow n} = N(N-1)\cdots(N-n+1) = \#\,\mathsf{Hom}([n],[N])$$

s.r.s. obs $y\varphi$ by composition $[n] \xrightarrow{\varphi} [N] \xrightarrow{y} \mathbb{R}$

Example: $N = 4$; $n = 3$; $y = (6.2, 4.8, 5.1, 3.2)$

$$y\varphi \triangleq \begin{cases} (6.2, 4.8, 5.1) & \text{w.p. } 1/4^{\downarrow 3}; \quad [3!] \\ (6.2, 4.8, 3.2) & \text{w.p. } 1/4^{\downarrow 3}; \quad [3!] \\ (6.2, 5.1, 3.2) & \text{w.p. } 1/4^{\downarrow 3}; \quad [3!] \\ (4.8, 5.1, 3.2) & \text{w.p. } 1/4^{\downarrow 3}; \quad [3!] \end{cases}$$

Exchangeability and inheritance on the average

Illustration: $N = 4;\ \ n = 3;\quad y = (6.2, 4.8, 5.1, 3.2)$

$$\bar{y}_N = k_{N,1}(y) = \sum y_i/N = 4.825$$

$$k_{N,2}(y) = \sum (y_i - \bar{y}_N)^2/(N-1) = 4.6075$$

$$k_{N,3}(y) = \sum (y_i - \bar{y}_N)^3 \frac{N}{(N-1)(N-2)} = -1.11375$$

$$k_{n,1}(y\varphi) \stackrel{\Delta}{=} \{5.367, 4.373, 4.833, 4.367\} \quad \text{w.p. } 1/4 \text{ each}$$

$$\text{ave}_\varphi\big(k_{n,1}(y\varphi)\big) = 4.825$$

$$\text{ave}_\varphi\big(k_{n,2}(y\varphi)\big) = 4.6075$$

$$\text{ave}_\varphi\big(k_{n,3}(y\varphi)\big) = -1.11375$$

Natural statistics with respect to S.R.S.

A natural statistic T of degree d is
a sequence of functions $T_n \colon \mathbb{R}^n \to \mathbb{R}$
—defined for every $n \geq d \geq 0$

For every $y \in \mathbb{R}^N$ and s.r.s. $\varphi \colon [n] \to [N]$

$$\operatorname*{Ave}_{\varphi \in \mathsf{Hom}([n],[N])} T_n(y\varphi) = T_N(y)$$

In general, called U-statistics
Polynomial functions: k-statistics and polykays

Relation between symmetric functions on different spaces
k-statistics (Fisher 1929); Inheritance (Tukey 1950s)

Statistical theory for spectral sampling

Objects Y are $n \times n$ matrices (symmetric or Hermitian)
Functions $T_n(Y)$ are class functions $T_n(UYU^*) = T_n(Y)$

Statistics: Y is a random $N \times N$ Hermitian matrix

Y is freely randomized if, for each U unitary, $Y \sim UYU^*$

if $H \perp\!\!\!\perp Y$ is a random Haar-distributed matrix, order N
then HYH^* is a freely randomized version of Y

$(HYH^*)_{n\times n}$ is the leading $n \times n$ sub-matrix
then $(HYH^*)_{n\times n}$ is also freely randomized

$\Lambda(Y) = \{\lambda_1, \ldots, \lambda_N\}$
$\Lambda\big((HYH^*)_{n\times n}\big)$ is a spectral sub-sample

Natural statistics for spectral samples

A natural statistic T of degree d is
a sequence of class functions $T_n \colon \mathcal{H}_n \to \mathbb{R}$
—defined for every $n \geq d$.

For every $Y \in \mathcal{H}_N$

$$\operatorname*{Ave}_{H \in \mathrm{Haar}_N} T_n\big((HYH^*)_{n\times n}\big) = T_N(Y)$$

Simplest examples:

$$k^{\dagger}_{(1)}(Y) = n^{-1}\operatorname{tr}(Y) = k_{(1)}(\lambda)$$

$$k^{\dagger}_{(2)}(Y) = \frac{1}{n^2-1}\sum(\lambda_i - \bar{\lambda})^2 = \frac{k_{(2)}(\lambda)}{n+1}$$

Examples of natural spectral statistics (Di N. et al 2013)

$$k^{\dagger}_{(2)} = \frac{nS_2 - S_1^2}{n(n^2-1)} = \frac{1}{n^2-1}\sum(\lambda_i - \bar{\lambda})^2 = \frac{k_{(2)}}{n+1}$$

$$k^{\dagger}_{(1^2)} = \frac{nS_1^2 - S_2}{n(n^2-1)} = k_{(1^2)} + \frac{k_{(2)}}{n+1}$$

$$k^{\dagger}_{(3)} = 2\,\frac{2S_1^3 - 3nS_1S_2 + n^2S_3}{n(n^2-1)(n^2-4)} = \frac{2k_{(3)}}{(n+1)(n+2)}$$

$$k^{\dagger}_{(4)} = 6\,\frac{S_4(n^3+n) - 4S_1S_3(n^2+1) + S_2^2(3-2n^2) + 10nS_1^2S_2 - 5S_1^4}{n(n^2-1)(n^2-4)(n^2-9)}$$

$$= 6\,\frac{k_{(4)} + k_{(2^2)}}{(n+1)(n+2)(n+3)}$$

$$k^{\dagger}_{(2^2)} = \frac{k_{(4)} + (n^2+6n+6)k_{(2^2)}/n}{(n+1)(n+2)(n+3)}$$

Limiting behaviour as $n \to \infty$

Theorem (Di Nardo, McC and Senato (2013))

The normalized limit of $k^{\dagger}_{(r)}(Y)$ *as* $n \to \infty$ *is the* r*th free cumulant.*

The normalized limit of $k^{\dagger}_{(r,s)}$ *is the product of two free cumulants*

Categorical interpretation: random embeddings

Simple random samples	:	Spectral random samples
Inj: $[n] \xrightarrow{\varphi} [N]$	:	Euclidean isometries $\mathbb{R}^n \xrightarrow{L} \mathbb{R}^N$
SRS: $[n] \rightsquigarrow [N]$	:	Haar: $\mathbb{R}^n \rightsquigarrow \mathbb{R}^N$
pullback by composition	:	pullback by conjugation

$\#\,\mathrm{Inj}(n, N) = N^{\downarrow n}$; $\quad \#\mathrm{SRS}(n, N) = 1_{n \leq N}$;
Natural statistic is a natural transformation on functors

Categorical Probability and Statistics

The category of injective maps (Inj)

Objects(Inj): finite sets $\Omega, \Omega', \ldots$
Arrows(Inj): 1–1 maps (injective maps $\varphi\colon \Omega' \to \Omega$)

Inj includes symmetric group(s): $[n] \xrightarrow{\varphi} [n]$

$\#\,\mathrm{Hom}([m],[n]) = n^{\downarrow m}$ for $m \leq n$; 0 for $m > n$

Representation of Inj: homomorphism Inj $\to$ Lin(Vect)

$$\begin{array}{ccc} \mathsf{Inj} & \mathsf{Lin} & \mathsf{Lin} \\ \Omega & \mathbb{R}^{\Omega} & \mathbb{R}^{\Omega\times\Omega} \\ {\scriptstyle\varphi}\uparrow & \downarrow{\scriptstyle\varphi^*} & \downarrow{\scriptstyle\varphi^*} \\ \Omega' & \mathbb{R}^{\Omega'} & \mathbb{R}^{\Omega'\times\Omega'} \end{array}$$

$$\Omega' \xrightarrow{\varphi} \Omega \xrightarrow{x} \mathbb{R}; \qquad x \xmapsto{\varphi^*} x\varphi \in \mathbb{R}^{\Omega'}$$

Sub-representations of Inj

Given a representation $\Omega \mapsto T_\Omega$ in which $\varphi\colon \Omega' \to \Omega$ is sent to $T\varphi\colon T_\Omega \to T_{\Omega'}$, a sub-representation is a sequence of subspaces $V_\Omega \subset T_\Omega$ that is preserved by the maps $T\varphi$.

Split by group reps for each Ω

$$\begin{array}{ccccccc} \mathsf{Inj} & & \mathsf{Lin} & & & & \\ \Omega & & \mathbb{R}^\Omega & \cong & \mathbf{1}_\Omega & \oplus & \mathbf{1}_\Omega^\perp \\ {\scriptstyle\varphi}\uparrow & & {\scriptstyle\varphi^*}\downarrow & & {\scriptstyle\varphi^*}\downarrow & & \downarrow\!\!\!\!\times \\ \Omega' & & \mathbb{R}^{\Omega'} & \cong & \mathbf{1}_{\Omega'} & \oplus & \mathbf{1}_{\Omega'}^\perp \end{array}$$

$\mathbf{1}_\Omega \subset \mathbb{R}^\Omega$ is a sub-rep;
no complementary rep, but $\mathbb{R}^\Omega/\mathbf{1}_\Omega$ is a quotient rep.

Sub-representations of $\mathrm{Inj}^2, \mathrm{Inj}^3, \ldots$

Objects in Inj^2, Inj^3 are Cartesian products (rectangles,...)
Morphisms: ordered pairs (φ, ψ)
Given the tensor product representation what are the sub-reps?

Revert to statistical terminology for a factorial design:

A is a factor $u \mapsto A_u$ on $\mathcal{U}$ (row)
B is a factor $u \mapsto C_u$ on $\mathcal{U}$ (col)
C is a factor $u \mapsto C_u$ on $\mathcal{U}$ (treat)
Response Y is a function $\mathcal{U} \to \mathbb{R}$; $\mu_u = E(Y_u)$

$$\begin{array}{ccc} \mathcal{U} & \xrightarrow{Y} & \mathbb{R} \\ \mathcal{U} & \xrightarrow{(A,B,C)} & \Omega_A \times \Omega_B \times \Omega_C \xrightarrow{\mu} \mathbb{R} \end{array}$$

Q. What are the Inj^3-sub-reps in $\mathbb{R}^{\Omega_A} \otimes \mathbb{R}^{\Omega_B} \otimes \mathbb{R}^{\Omega_C}$?
—called factorial subspaces

Sub-representations of $\mathrm{Inj}^2, \mathrm{Inj}^3, \ldots$: Factorial subspaces

Q. What are the Inj^3-sub-reps in $\mathbb{R}^{\Omega_A} \otimes \mathbb{R}^{\Omega_B} \otimes \mathbb{R}^{\Omega_C}$?

Statistical notation: $A \equiv \mathbb{R}^{\Omega_A}, \ldots$

Sub-reps in $(\mathbf{1}_A \subset A) \otimes (\mathbf{1}_B \subset B) \otimes (\mathbf{1}_C \subset C)$

$\mathbf{1}_A \otimes \mathbf{1}_B \otimes \mathbf{1}_C \equiv \mathbf{1}, \quad A \otimes \mathbf{1}_B \otimes \mathbf{1}_C \equiv A, \ldots$

2^3 indecomposables $\mathbf{1}, A, B, C, AB, AC, BC, ABC$

$\mathbf{1} \subset A \subset AB \subset ABC$

... plus vector spans $A + B, \quad A + BC, \quad AC + BC, \ldots$

How many sub-reps?

free distributive lattice; monotone subsets; simplicial complexes; hereditary hypergraphs; Dedekind numbers;

k	0	1	2	3	4	5	6
D_k	2	3	6	20	168	7581	7828354

What does Inj and Inj^k-representation give us?

The answer (intuition) is not new
—factorial subspaces integrated into software 50 years ago

The formulation of the question is new:

It offers insight into why certain group reps are unacceptable:
it offers an explanation for marginality principle

It enables us to formulate and answer related questions

Inj-Sub-representations in $\mathbb{R}^{\Omega\times\Omega}$

Inj-Sub-representations in $\mathbb{R}^{\Omega^3}$

Summary:

Three areas in which categorical ideas play a role

(i) Inheritance and k-statistics (reverse martingale)
—relation to symmetric functions, moments and cumulants

(ii) Inheritance and spectral k-statistics
—relation to class functions, spectral moments and free cumulants

(ii) Representation theory for Inj, $\mathrm{Inj}^2, \ldots$
—understanding of factorial subspaces as projective systems

Random isometric embeddings

$$\begin{array}{ccccc} \text{SRS} & V/S & \xrightarrow{k} & \mathbb{R} \\ [m] & \mathbb{R}^m/S_m & \xrightarrow{k_m} & \mathbb{R} \\ \uparrow & \downarrow & & \| \\ [n] & \mathbb{R}^n/S_n & \xrightarrow{k_n} & \mathbb{R} \end{array} \qquad \begin{array}{ccccc} \text{Haar} & \mathrm{FR}(\mathcal{H}) & \xrightarrow{k} & \mathbb{R} \\ \mathbb{R}^m & \mathrm{FR}(\mathcal{H}_m) & \xrightarrow{k_m} & \mathbb{R} \\ \uparrow & \downarrow & & \| \\ \mathbb{R}^n & \mathrm{FR}(\mathcal{H}_n) & \xrightarrow{k_n} & \mathbb{R} \end{array}$$