Six Statistical Senses

Radu V. Craiu
Toronto

Ruobin Gong
Rutgers

Xiao-Li Meng
Harvard

NESS, Boston 2023

A statistician’s survival kit

- Everything we do is about understanding and surviving variation
- Variations create both information and uncertainty, signals and noise.
- The ability to learn from “alike” events and identify “differences” ensures survival
- Variation → Data → Inference → Prediction or Understanding
- A statistician thrives within the fluid dynamics between similarity and difference captured by a distribution
- So what senses allow us to navigate Tyche’s land?

So… about those senses

- Developed over time, motivated by various problems
- They set statisticians apart from other scientists
- Every sense is hard to pin down exactly
- Instead it is illustrated with ideas, methods, etc which share common reasoning threads
- Emphasis is on statistical ideas and not mathematical laws
- Establish a general categorization of statistical ideas (not a partition!)

So what?

- Will it be easier to do research? Not really.
- Will this make me a better statistician? Somewhat.
- Categorization of ideas can help:
    - Organize knowledge
    - Simplify comprehension
    - Communication
    - Classify a contribution (Which sense are you enhancing?)
    - Crystallize the differences/similarities with adjacent fields (e.g. ML, Math, etc)

The Senses

- Selective Hearing: Informative Ignorance (SRS, Cox PH)
- Following our Nose: Determinacy through Randomness (DP, Monte Carlo)
- Seeing through: Enlightenment from Dark Data (FTE, MI, EM, DA)
- The Magic Touch: Refinement through Confinement (Conditioning, Bayes)
- Just a Taste: Richness in Frugality (Bootstrap, Propensity matching)
- (Sixth) Sense and Sensibility (BFF)

Informative Ignorance

- Hear the essential (signal) and not the incidental (noise)
- Get to the heart of a problem without getting lost in details
  - Randomized sampling
    - Eliminates systematic biases
  - Cox PH
    - $\lambda(t) = \lambda_0(t)\exp(X^T\beta)$ - eliminates the need for modelling λ_0
    - Allows incorporation of time-dependent covariates and random effects
  - Various non-standard likelihoods (composite, partial, profile, etc)

Determinacy through Randomness

- Purposeful and artful insertion of randomness in data collection or analysis
  - Survey examples: Differential privacy, Multiple Imputation & Randomized Sampling
  - Monte Carlo
  - Randomized response of Warner (1965)
    - Proportion π of a feature in a population: $X \in \{0,1\}$
    - Use randomization mechanism: $R \in \{0,1\}$ with Pr$(R = 1) = p \in (1/2,1)$
    - Report $Y = 1_{\{R=X\}}$ but not X nor R
    - Estimate π using $\hat{\pi} = \left(p - 1 + \bar{Y}_n\right)/\left(2p - 1\right)$ where $\bar{Y}_n = \sum_{i=1}^{n} Y_i$

DP and Randomized Response

- Given data X, a randomized mechanism $Y(X)$ is ϵ-private iff

$$\exp(-\epsilon) \leq \frac{\Pr\left(Y(X) = y \mid X = x\right)}{\Pr\left(Y(X) = y \mid X = x'\right)} \leq \exp(\epsilon) \qquad (*)$$

where $|x - x'| = 1$, and ϵ is the privacy budget

- Suppose X is your department's estimated payroll in 2030 which is DP-ed into $Y(X)$
- After recruiting famous professor B in 2023 with salary A your department payroll in 2030 is estimated to be $Y(X') = Y(X + A)$
- So if Pr$(Y(X) = S) < <$ Pr$(Y(X + A) = S)$ then ϵ in $(*)$ is large implying a significant loss of privacy (budget).

DP trade-offs

- Warner's randomized response satisfies ϵ-DP criterion since

$$\frac{\Pr\left(Y_i = 1 \mid X_i = 1\right)}{\Pr\left(Y_i = 1 \mid X_i = 0\right)} = \frac{\Pr\left(Y_i = 0 \mid X_i = 0\right)}{\Pr\left(Y_i = 0 \mid X_i = 1\right)} = \frac{p}{1-p} \text{ so}$$

$\epsilon = \text{logit}(p)$

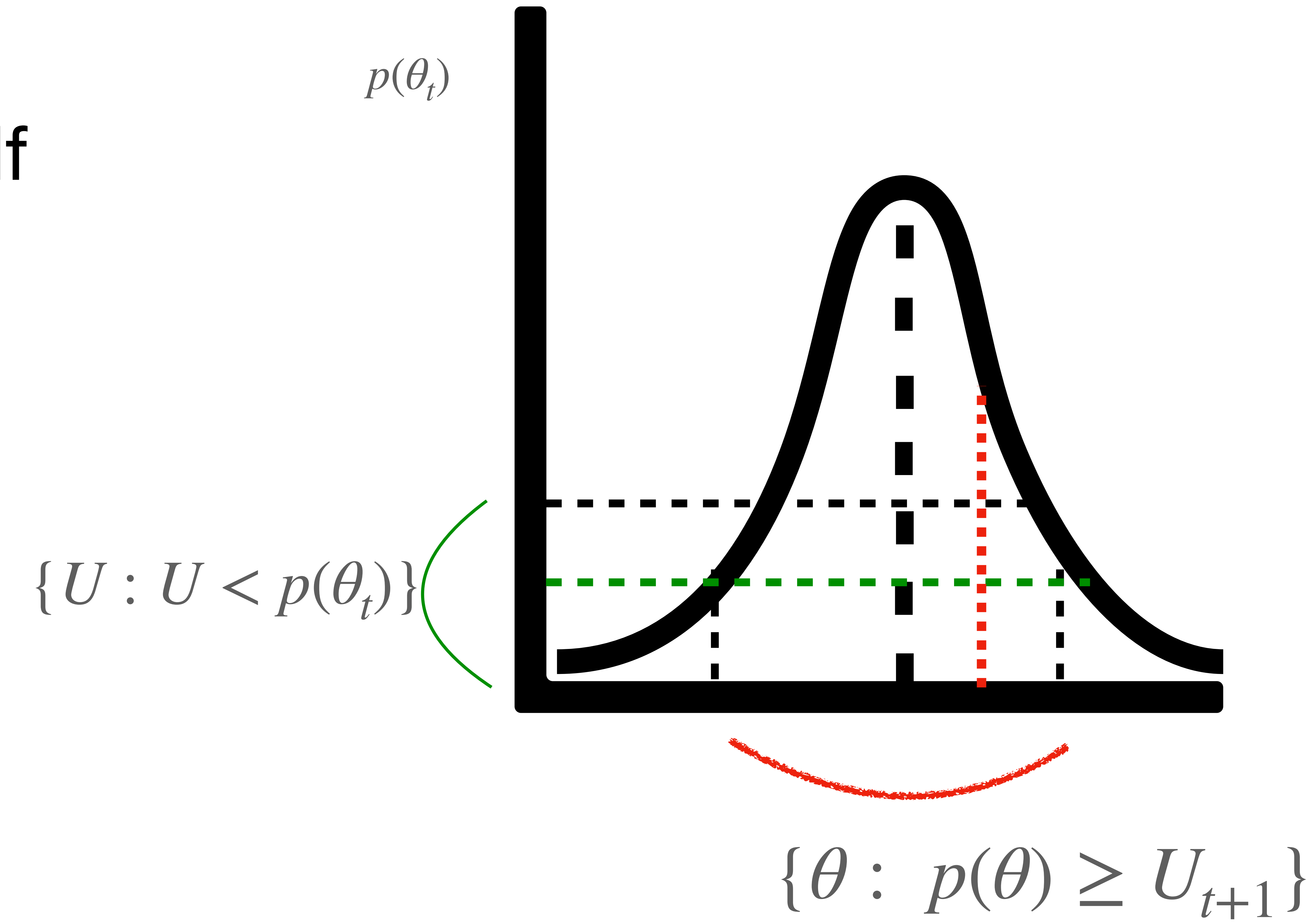
- $p = 0.5$ implies $\epsilon = 0$ - the information in the sample is lost after randomization
- $\text{Var}(\hat{\pi}) = \dfrac{f(1-f)}{n}$ where $f = 1 + 2\pi p - \pi - p$
- It can be shown that $\text{Var}(\hat{\pi})$ increases as ϵ decreases.

Enlightenment from Dark Data

- Data are never completely observed and analyses can be challenged by various selection biases
- One must detect the latter and find remedies:
- The use of binary recording indicator variable R_i and the use of Pr$(R = 1 \,|\, Y_{obs}, Y_{mis})$ to model missing mechanisms (Rubin 1976)
- Formula of total error in estimating the mean of a finite population (Meng 2018, *AOAS*)
- Adding dark data to enhance understanding or boost computation: EM algorithm and Data Augmentation

Slice Sampling (Neal, 2003)

- Of interest is computing $I = \int h(\theta)p(\theta)dx$ where $p(\theta)$ is a pdf (e.g., a posterior)
- $I = \int h(\theta)g(\theta, U)dUd\theta$ where $g(\theta, U) = 1_{\{U \leq p(\theta)\}}$
  - Alternate $U \sim$ Unif$(0, p(\theta))$ and $\theta \sim$ Unif$\{\theta : \; p(\theta) \geq U\}$



Refinement by confinement

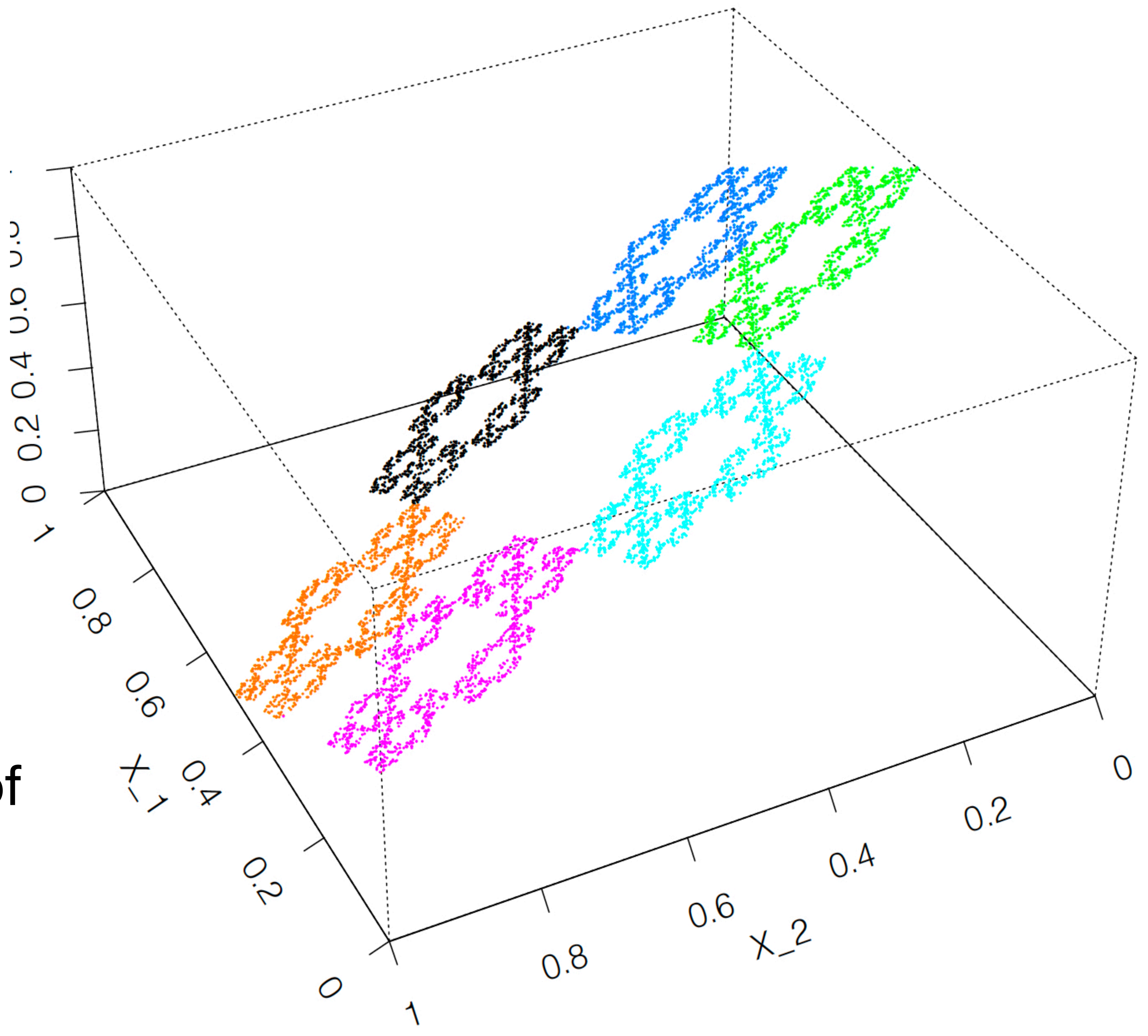
- Creating ingenious model constraints and links, be they conceptual or functional, between different parameters.
- Bayesian pooling and shrinkage tether the parameters in the model and allow the transfer of information between groups of observations.
- This results in more information being available for each parameter estimate.
- Experimental swindles for sampling: antithetic, stratified and blocking
- Other types of conditioning shrinks the sample space to relevant subspaces, and allows the insertion of subject-matter knowledge into the mathematical aspects of the statistical analysis.

Example: Antithetic sampling

- The Monte Carlo method approximates $I = \int f(\theta)p(\theta)d\theta$ using $\hat{I} = M^{-1}\sum_{i=1}^{M} f(\theta_i)$ where $\theta_i \sim p$ (iid or via MCMC)
- Var$(\hat{I}) = \frac{\sigma_f^2}{M}$ for iid samples with $\sigma_f^2 =$ Var$_p(f)$
- Savings can be obtained when we can generate K-tuples $\{(\theta_1^{(j)}, \ldots, \theta_K^{(j)} : \; 1 \leq j \leq M/K\}$ so that corr$(f(\theta_i^{(j)}), f(\theta_{i'}^{(j)}) = \rho_f \leq 0 \; \forall i \neq i', j$
- The estimator $\tilde{I} = \sum_{j=1}^{M/K}\sum_{i=1}^{K} f(\theta_i^{j)})$ has variance Var$(\tilde{I}) = \frac{\sigma_f^2}{M}[1 + (K-1)\rho_f]$

Beneficial Tethering

- Craiu and Meng (2005) propose the iterative Latin Hypercube Sampling
- Start with $U_i^{(0)} \sim Unif(0,1)$ iid
- At iteration r $U_i^{(r)} = \dfrac{\tau(i-1) + U_i^{(r-1)}}{K}$ $\forall 1 \le i \le K$, where τ is a random permutation of $\{0,1,2\}$ independent of past.
- Casarin et al (2023, *Stat. Sci.*): sampling on segments unifies many disparate methods for producing antithetic variates



Support of the limiting distribution when K=3

Richness in Frugality

- Parsimony is a time honoured principle in Statistics
- Frugality extends beyond our fundamental (misplaced?) belief that truth is elegant and simple
- Examples: bootstrap, propensity matching, latent variable models, etc
- A parametric model is frugal but can yield substantive understanding
- Frugality: linear regression; super-frugality: sparsity
- Compounding frugality yields abundance: multiresolution models, hierarchical models, etc.

Sixth Sense

- The one that brings it all together
- Bridging what we know with the future (prediction) is a leap (of faith)
- One cannot test the bridge without walking on it
- Multiple paths
    - Bayesian , Frequentist, Fiducial and …. ?
    - More can be found by participating in the BFF annual conference
    - Web: bff-stat.org/

Where else to go

- Stigler's "Seven Pillars of Statistical Wisdom" - accentuate essence of statistics using broad framing
- Gelman and Vehtari (JASA, 2021) "What are the Most Important Statistical Ideas of the Past 50 Years?"
- Wasserman's Blog
- Gelman's Blog
- Robert's Blog

Goodbye gift

PHRONESIS

Wisdom in determining ends and the means of attaining them, practical understanding, sound judgment.