Unbiased Estimator for Deconvolving Human Blood into Antibody Composition

Motivation: We have a HIV antibodies, each with different neutralization profiles against v HIV-1 pseudo-viruses (measured in IC_{50}). Given human blood, tested against the same viruses, we want to know the composition of antibodies, θ.

Caveat: Testing human blood against the panel of viruses gives measurements in ID_{50} not IC_{50}. IC_{50}'s do not add linearly.

Solution: Transform the data. The Gibbs free energy,

$$\Delta G_{ij} = RT \ln \left(X_{ij} - C_{ij} \right)$$

does add linearly across antibodies, where X_{ij} is the IC_{50} value of antibody j on virus i. R, T, and C known. Experimentally determined monotonic calibration curve, f, tells us

$$s \approx f\left((\Delta G)\, \theta \right)$$

where s_i is the ID_{50} measurements of the blood versus virus i.

Unbiased Estimator for Deconvolving Human Blood into Antibody Composition

Model: $f^{-1}(s) \overset{\text{ind}}{\sim} \mathcal{N}((\Delta G)\,\theta, \sigma^2 I_{v\times v})$, unknown $\theta \in \mathbb{R}^a$, σ, known f

Estimand: $g(\theta) = \theta$

Decision Space: $\mathcal{D} = \mathbb{R}^a$ **Loss:** $L(\theta, d) = \|g(\theta) - d\|_2^2$

Initial estimator: Let $r(s)$ = Spearman rank ordering of s. Example: $r\left((2.0, 9.2, 1.0)^T\right) = (2, 3, 1)^T$. To avoid f and ΔG, the authors chose

$$\delta(X) = \underset{\theta \succeq 0}{\operatorname{argmin}} \|r(s) - r(X)\,\theta\|_1$$

- Not UMVUE, Not minimax, Inadmissible

Alternative estimator: $\delta'(X) = \left((\Delta G)^T (\Delta G)\right)^{-1} (\Delta G)^T f^{-1}(s)$

- UMVUE

Preference: I would prefer δ over δ' since $\theta \succeq 0$, but the UMVUE δ' need not satisfy this property. Moreover, δ is sparse.

Minimax estimation of the Scale Parameter of Laplace Distribution under Squared-Log Error Loss Function

Motivation: The classical Laplace distribution with mean zero and variance σ^2 was introduced by Laplace in 1774. This distribution has been used for modeling data that have heavier tails than those of normal disribution. The probability density function of the Laplace distribution with location parameter a and scale parameter θ is given by:

$$f(x|a,\theta) = \frac{1}{2\theta}\exp\left(-\frac{|x-a|}{\theta}\right) \qquad -\infty < x < \infty$$

This paper aims to find a minimax estimation for the scale parameter when the location parameter is known.

Minimax estimation of the Scale Parameter of Laplace Distribution under Squared-Log Error Loss Function

Model: $X_i \overset{\text{ind}}{\sim} \mathcal{L}(a, \theta)$ for unknown $\theta \in \mathbb{R}^+$, known a

Estimand: $g(\theta) = \theta$

Decision Space: $\mathcal{D} = \mathbb{R}^+$ **Loss:** $L(\theta, d) = (\log d - \log \theta)^2 = (\log \frac{d}{\theta})^2$

Initial estimator: $\delta(X) = \frac{\sum_{i=1}^{n} |x_i - a|}{\exp(\Psi(n))}$ where $\Psi(n) = \frac{\Gamma'(n)}{\Gamma(n)}$ the digamma function

- Minimax,Bayes, UMRUE

The prior distribution is Jeffreys prior, that is the prior density of θ is proportional to: $\frac{1}{\theta}\sqrt{\frac{3n}{2}}$

Alternative estimator: $\delta'(X) = \frac{\sum_{i=1}^{n} |x_i - a|}{n}$

- Ubiased

Preference: I would prefer δ because of it's minimaxity. Also for large n the bias is not too much.

Estimating the number of unseen species: How far can one forsee?

Definition: Suppose one draws $\boldsymbol{n}$ samples from a discrete distribution $\mathcal{P}$(unknown) and those n samples contains $S(n)$ unique elements. How many **new unique** elements will one expect to "discover" if $\boldsymbol{m}$ further samples are drawn from $\mathcal{P}$.

Eg:How often species of butterflies are trapped $(n = 639, S(n) = 285)$.

Frequency (i)	1	2	3	4	5
Species (Φ_i)	110	78	44	24	29

Question: How many new species $(U_n(m))$ of butterflies will be "discovered" if we make $m(= 100)$ more observations.

- **Data generating model** $(\mathcal{P})$: Multinomial Distribution, Poisson.
- **Estimand** $g(\theta) : U_n(m)$.
- **Loss Function**: $\left(\frac{U_n^E(m) - U_n(m)}{m}\right)^2$. ($E$ is the estimator)

Estimating the number of unseen species: How far can one forsee?

UMRUE estimator: $U_n^u(m) = -\sum_{i=1}^{n} \left(-\frac{m}{n}\right)^i . \Phi_i$

Estimator proposed: $U_n^h(m) = -\sum_{i=1}^{n} h_i \Phi_i, h_i = -\left(\frac{m}{n}\right)^i \mathbb{P}(poi(r) \geq i)$

- Not UMVUE / UMRUE, Not Bayes, May be admissible
- $\left(U_n^h(m)\right)^+$ dominates $U_n^h(m)$.

Whats good then?

- The maximum loss under Poisson, Multinomial model is bounded
- The maximum loss is close to a lower-bound which is applicable to all the estimators

What is the next step?

- Few "parts" of the estimator aren't well motivated, one could improve it
- Extend this framework to various other Models $\mathcal{P}$.

On improved shrinkage estimators for concave loss

Motivation:
Estimating a location vector of a multivariate model is of interest in many statistical modeling settings. Spherically symmetric distributions are commonly used. For a multivariate normal with square error as loss function, James and Stein famously showed that the sample mean is is inadmissable, and proposed a minimax estimator using shrinkage methods.

Contribution:
This paper develops new minimax estimators for a wider range of distributions and loss functions. For multivariate gaussian scale mixtures and loss functions that are concave functions of squared error, this paper expands the class of known minimax estimators. Similar to the James-Stein estimator, the proposed minimax estimators are of the form:

$$(1 - f(X))X$$

On improved shrinkage estimators for concave loss

Model: X is from a MVN scale mixture: $X|\theta \sim \mathcal{N}(\theta, VI)$ with V a non-neg r.v. with cdf $h(v)$.

Estimand: The location parameter: $g(\theta) = \theta$.

Decision Space: $\mathcal{D} = \mathbb{R}^p$

Loss: $L(\theta, d) = l(\|\theta - d\|^2)$

Initial estimator:

$$\delta_a(X) = (1 - a \times r(\|X\|^2)/\|X\|^2)X$$

Where $r(t)$ and a satisfy: $0 \leq r(t) \leq 1$, $r(t)$ non-decreasing, $r(t)/t$ non-increasing, $0 < a < 2/\mathbb{E}^*[1/\|X\|^2]$

- Minimax, Not UMRUE, Not MRE (location equivariant)

Alternative estimator: $\delta'(X) = X$

- MRE (location equivariant) when l(t) = t

Preference: If the. only goal were to minimize the loss function L, then δ_a is a better estimator because it dominates δ'. However, if each component of X were independently of interest, and the values were not related in some way, then δ' would be more informative.

Parameter Estimation for Hidden Markov Models with Intractable Likelihoods

Definition: A Hidden Markov model (HMM) is a statistical tool used to model data output over time. It consists of a hidden process, $\{X_k\}$, and an observed process $\{Y_k\}$. The model obeys the Markov property: $Y_k|X_k$ is conditionally independent of all preceding random variables.
Motivation: HMMs have a wide range of applications including speech and handwriting recognition, stock market analysis, and population genetics. We consider a simple finance application.
Model: One can model the returns of a set of stocks using an HMM with two hidden states. Specifically, $\{X_k\} \in \{1, -1\}$, and

$$Y_k|X_k \sim S_\alpha(\sigma, 0, X_k + \delta),$$

where $S_\alpha(\sigma, 0, X_k + \delta)$ denotes the α-stable distribution with stability α, skewness σ, scale 0, and location $X_k + \delta$. We assume α is known and that we have an observed sequence $y_1, \ldots y_n$.

Parameter Estimation for Hidden Markov Models with Intractable Likelihoods

Estimand: $g(\theta) = (\sigma, \delta) = \theta$. **Decision Space:** $\mathcal{D} = [-1, 1] \times (-\infty, \infty)$.
Loss: For fixed $\epsilon > 0$,

$$L(\theta, d) = \begin{cases} 0 & \text{if } \|\theta - d\|^2 < \epsilon \\ 1 & \text{otherwise.} \end{cases}$$

Estimator: We cannot express the likelihood analytically, so we approximate it and take

$$\delta(Y) = \arg\sup_{\theta \in \Theta} p_\theta^\epsilon(y_1, \ldots y_n),$$

where p_θ^ϵ is the likelihood function of the perturbed process $\{X_k, Y_k + \epsilon Z_k\}$ and $Z_k \sim U_{B(0,1)}$.
Properties: Not Bayes, not UMRU, Inadmissible.
Alternative estimator: Assume that θ has a prior distribution π. Choose

$$\delta'(Y) = \arg\sup_{\theta \in \Theta} p^\epsilon(\theta | y_1, \ldots, y_n).$$

This is Bayes under prior π.
Preference: Second estimator if reasonable prior exists.

SLOPE Adaptive Variable Selection via Convex Optimization

Motivation: A common challenge in many scientific problems is the ability to achieve correct model selection while also allowing for an inferential selection framework. A recent estimator that is closely linked to the LASSO and Benjamini-Hochberg (BH), is Sorted L-One Penalized Estimation:

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^P}{\operatorname{argmin}} \frac{1}{2}||y - X\beta||_2^2 + \sum_{i=1}^{p} \lambda_{BH_i}|b|_i$$

which controls the False Discovery Rate (FDR) at level $q \in [0, 1)$ under orthogonal designs under λ_{BH_i}:

$$\lambda_{BH_i} = \sigma\Phi^{-1}(1 - iq/2p)$$

This is equivalent to the isotonic formulation:

$$\min_{\beta} \frac{1}{2}||y - \lambda_{BH} - \beta||_2^2$$

$$\text{subject to } \beta_1 \geq \ldots \geq \beta_p \geq 0$$

SLOPE Adaptive Variable Selection via Convex Optimization

Model: Gaussian sequence X, $X^T X = I_p$ such that $Y = \beta + z \sim \mathcal{N}(\beta, \sigma^2 I_p)$

Estimand: $g(\beta) = \beta$

Decision Space: $\mathcal{D} = \mathbb{R}^P$ **Loss:** $\ell_2 : L(\beta, d) = (g(\beta) - d)^2$

Initial estimator: Given an orthogonal design:

$$\delta(X) = \text{prox}_\lambda(X^T y) = \underset{\beta}{\text{argmin}} \frac{1}{2}||y - \beta||_2^2 + \sum_{i=1}^{p} \lambda_i |\beta|_{(i)}$$

where $|\beta|_{(i)}$ is the i-th order statistic.

- Not UMVUE / UMRUE, Bayes - Laplacians, Minimax (Asymptotic)

Alternative estimator: $\hat{\beta} = X^T Y$

- Unbiased and UMVUE

Improved Minimax Estimation of a multivariate Normal Mean under heteroscedaticity

Problem and Model: Let

$$\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} = N(\begin{pmatrix} \theta_1 \\ \vdots \\ \theta_p \end{pmatrix}, \Sigma), \qquad \text{where} \begin{cases} \theta = (\theta_1, \cdots, \theta_p)^T \text{ is unknown} \\ \Sigma = D = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ 0 & 0 & \dots & d_p \end{pmatrix} \end{cases}$$

Goal: Find an estimator $\delta(X) \in \mathbb{R}^p$ of θ under squared-error loss:

$$\boxed{L(\delta(X), \theta) = (\delta(X) - \theta)^T (\delta(X) - \theta)}$$

Suggested approach: develop a minimax shrinkage estimator of the form $\delta(X) = (I - \lambda A)X$ where A is known

Sketch of the proof:

- For a fixed A, $\lambda_{opt} = \frac{DA}{\mathbb{E}_\theta(X^T A^T A X)} \implies \delta_{A,c} = (I - \frac{c}{X^T A^T A X} A)X$
- Theory tells us that $\delta_{A,c}$ is minimax if:

  $0 \leq c \leq 2(tr(DA) - 2\lambda_{max}(DA)) \implies$ pick $c*(A, D) = tr(DA) - 2\lambda_{max}(DA)$

- Pick A such that A minimises the upper bound of the previous inequality under Gaussian prior $\Lambda = N(0, \Gamma)$ where Γ is diagonal

Improved Minimax estimator

Initial estimator:

$$\delta_A(X) = X - \frac{\sum_{j=1}^{p}(a_j^*)^2(d_j+\gamma_j)}{\sum_{j=1}^{p}(a_j^*)^2 X_j^2}AX$$

where $A = \text{diag}(a_1^*, a_2^*, \cdots, a_p^*)$ is given by

$$\begin{cases} a_j^* = (\sum_{k=1}^{n_0} \frac{d_k+\gamma_k}{d_k^2})^{-1} \frac{n_0-2}{d_j} & j = 1, \cdots, n_0 \\ a_j^* = \frac{d_j}{d_j+\gamma_j} & j = n_0+1, \cdots, p \end{cases}$$

and $n_0 = \text{argmax}\{i : \quad d_i a_i^* = \max(a_j^* d_j, j \in [1,p])\}$

- Minimax, not admissible, Not Bayes, not UMRUE

Alternative estimator: $\delta_0(X) = X$

- Unbiased, direct

Preference: For estimating θ, the unbiasedness property seems to lead to a far greater risk ($\sim p\sigma^2$ in case of homoscedasticity), whereas δ_A has far smaller risk

$$\text{If } n_0 \geq 4, \quad R(\theta, \delta_A) \leq \sum_{i=1}^{p} d_i - \sum_{j=5}^{p} \frac{d_j^2}{d_j+\gamma_j} = R(\theta, \delta_0(X)) - \sum_{j=5}^{p} \frac{d_j^2}{d_j+\gamma_j}$$

Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation

Motivation: Kernel density estimation (KDE)

- Nonparametric way to estimate probability density function of a random variable
- Essentially a generalization of the histogram
- Uses kernel functions instead of step functions to locally approximate the density

Given a random sample $X_1, \ldots, X_n \in \mathbb{R}^d$ from an unknown density f, the standard KDE of $f(x)$ has the form $\widehat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K_H(x - X_i)$ where $K_H(x - X_i)$ is a kernel function centered at X_i with bandwidth matrix H.

This paper

- Examines locally adaptive KDE where a variable bandwidth matrix H_i is used for the kernel function centered at X_i
- Uses Gaussian kernel functions; each H_i is a covariance matrix
- Develops Bayes estimators for the variable bandwidth (covariance) matrices

Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation

Estimand: Each H_i used in KDE $\widehat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{H_i}(x - X_i)$ where $K_{H_i}(x - X_i) = \frac{1}{(2\pi)^{d/2}(\det H_i)^{1/2}} \exp\left(-\frac{1}{2}(x - X_i)^T H_i^{-1}(x - X_i)\right)$

Model: Likelihood: $\widehat{f}_{-i}(x_i|H_i, x_{j\neq i}) = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} K_{H_i}(x_j - x_i)$
Prior: $W^{-1}(Q, r)$ for each H_i where Q is SPD $d \times d$ matrix and $r \geq d$

Decision Space: SPD $d \times d$ matrices **Loss:** $L(H_i, \widehat{H}_i) = \mathrm{tr}\,(H_i - \widehat{H}_i)^2$

Initial estimator: $\widehat{H}_i(X) = \frac{1}{r-d}\sum_{j=1,j\neq i}^{n} \frac{(\det B_{i,j})^{-(r+1)/2}}{\sum_{j=1,j\neq i}^{n}(\det B_{i,j})^{-(r+1)/2}} B_{i,j}$
where $B_{i,j} = (X_j - X_i)(X_j - X_i)^T + Q$

- Not UMVUE/UMRUE, Bayes wrt inverse-Wishart prior, admissible

Alternative estimator: $\widehat{\widehat{H}}_i(X) = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n}(X_j - X_i)(X_j - X_i)^T$

- MLE wrt likelihood: $f_{x_i}(x_{j\neq i}|H_i) = \prod_{j=1,j\neq i}^{n} K_{H_i}(x_j - x_i)$

Preference: For small sample size, prefer $\widehat{H}_i$ as small sample MLEs of covariance matrices have been shown to have distorted eigenvalue structure. For large sample size, prefer $\widehat{\widehat{H}}_i$ or the sample covariance matrix. MLE and sample covariance matrices are usually consistent.

Kernel Regularized Least Squares

Motivation: When the correct model specification is unknown (as is generally the case in social science), researchers often engage in ad hoc searches across polynomial and interaction terms in GLMs. More flexible methods, such as neural nets and generalized additive models, are often useful for prediction but difficult to interpret. By constructing a new set of features based on the kernel

$$k(x_i, x_j) = \exp\left(-\frac{||x_j - x_i||^2}{\sigma^2}\right)$$

$$K_{i,j} = k(x_i, x_j)$$

and using Tikhonov regularization, it is possible to search systematically over a broader space of smooth, continuous functions. By taking the expectations of the derivative of these functions over the sample, "coefficients" can be obtained. By taking derivatives at particular points, heterogeneous effects can be discovered.

Kernel Regularized Least Squares

Model: $y = Kc$ with $K \in \mathbb{R}^{n \times n}$ observed and $y, c \in \mathbb{R}^n$ unknown.
$y^{\text{obs}} = y + \epsilon$ with $y^{\text{obs}} \in \mathbb{R}^n$ observed and $\epsilon \in \mathbb{R}^n$ unknown, $\mathbb{E}[\epsilon | K] = 0$

Estimand: $g(c^*, \epsilon) = Kc^*$,
where $c^* = \underset{c \in \mathbb{R}^n}{\text{argmin}} (y - Kc)^T (y - Kc) - \lambda c^T K c$

Decision Space: $\mathcal{D} = \mathbb{R}^n$ **Loss:** $L(\theta, d) = (g(c, \epsilon) - d)^2$

Initial estimator: $\delta(y, K) = K(K + \lambda I)^{-1} y^{\text{obs}}$

- Not Bayes, Not location MREE, Not location-scale MREE

Alternative estimator: $\delta'(y, K) = K(I_n + K^T K)^{-1} K^T y$

- Bayes under Λ such that $c \sim MVN(0, I_n)$ and $\epsilon \sim MVN(0, I_n)$.

Preference: I would generally prefer δ to δ', because δ' makes a strong assumption on the prior which will not in general be satisfied.

An Example of an Improvable Rao-Blackwell improvement, inefficient MLE and unbiased generalized Bayes estimator.

Motivation:

- Data compression outside of exponential families can be difficult, especially without complete sufficient statistics.
- The authors want to improve the Rao-Blackwellized estimator $\mathbb{E}[X_1 \mid X_{(1)}, X_{(n)}] = \frac{X_{(1)}+X_{(n)}}{2}$; in the given model, $T = (X_{(1)}, X_{(n)})$ is minimal sufficient but not complete.
- The authors show that this Rao-Blackwell improvement can be uniformly dominated by another unbiased estimator.

Decision Problem:

- Model $\mathcal{P} = \{U((1-k)\theta, (1+k)\theta) : \theta \in \Theta\}$, $k \in (0,1)$ known.
- Estimand $g(\theta) = \theta$.
- Decision space $D = \mathbb{R}$.
- Loss function is squared error loss, $L(\delta) = (\delta - \theta)^2$.

An Example of an Improvable Rao-Blackwell improvement, inefficient MLE and unbiased generalized Bayes estimator.

Initial estimator δ:

$$\delta(X_1, \ldots, X_n) = \frac{(1-k)X_{(1)} + (1+k)X_{(n)}}{2\left(k^2 \frac{n-1}{n+1} + 1\right)}$$

- Not UMRU/UMVU, not Bayes, not MRE, uniformly dominates the Rao-Blackwell improvement

Proposed alternate δ':

$$\delta'(X_1, \ldots, X_n) = \frac{(1+k)X_{(1)} + (1-k)X_{(n)}}{2}$$

- Not UMRU/UMVU, not Bayes, MRE, minimax

Preference: I would prefer to use δ' for this task. Although δ dominates in a large and useful class of unbiased estimators, with δ' we get stronger properties such as minimum risk equivariance and minimality, which I find convincing arguments for using δ' over δ.

Condition-number-regularized covariance estimation

Motivation: Estimation of high-dimensional covariance matrices is known to be a difficult problem,has many applications, and is of current interest to the larger statistics community. In many applications including so-called the "large p small n" setting, the estimate of the covariance matrix is required to be not only invertible, but also well-conditioned. Although many regularization schemes attempt to do this, none of them address the ill-conditioning problem directly.

Model: $X_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu, \Sigma)$ for unknown $\Sigma \in \mathbb{R}^{p \times p}$, known μ

Estimand: Σ

Decision Space: $\mathcal{D} = \{\Sigma : \Sigma \succ 0\}$

Loss: $L(\Sigma, \hat{\Sigma}(X)) = \rho(\hat{\Sigma}) = \begin{cases} -\log \mathbb{P}(X|\hat{\Sigma}) & \text{cond}(\hat{\Sigma}) \leq \kappa_{max} \\ \infty & \text{otherwise} \end{cases}$

In fact, this is a MLE with constraint. But it can be interpreted as Bayes estimator.

Condition-number-regularized covariance estimation

Initial estimator: Because loss function does not depend on Σ, Bayes estimator under any prior distirbution is the same, which is

$$\hat{\Sigma} = \underset{\hat{\Sigma}}{\operatorname{argmin}}\, \mathbb{E}(\rho(\hat{\Sigma})|X) = \underset{\text{cond}(\hat{\Sigma}) \leq \kappa_{max}}{\operatorname{argmin}} \; -\log \mathbb{P}(X|\hat{\Sigma})$$

If we define Q as eigenvector matrix and l_i as eigenvalues of sample covariance matrix. Then solution can be expressed as $\hat{\Sigma} = Q \operatorname{diag}(\lambda_1, \ldots, \lambda_p) Q^T$, where

$$\lambda_i = \begin{cases} \tau^* & l_i \leq \tau^* \\ l_i & \tau^* < l_i < \kappa_{\max}\tau^* \\ \kappa_{\max}\tau^* & \kappa_{\max}\tau^* \leq l_i \end{cases}$$

- biased (even when n goes infinity), Bayes, Admissible

Alternative estimator: $\tilde{\Sigma} = \frac{1}{n-1} \sum_{i=1}^{n} X_i X_i^T$

- UMRUE, ill-conditioned, overfitting

Preference: When p is comparable to n, sample convariance matrix is usually overfitted and ill-conditioned. By enforcing regularization of condition number, we can solve the problem of ill-conditioned directly as well as overfitting. However, when $p \ll n$, I prefer UMRUE $\tilde{\Sigma}$.

Estimation of $\mathbb{P}(X > Y)$ with Topp-Leone Distribution

Motivation: In order to analyze stress-strength reliability of a system, we need to estimate the probability that stress, X, exceeds strength, Y. For instance, if Y indicates the amount of pressure that a bridge can tolerate and X indicates the stress of a flood, the reliability of the bridge is related to the probability $\mathbb{P}(X > Y)$.

Estimation of $\mathbb{P}(X > Y)$ with Topp-Leone Distribution

Model: $X_1, \cdots, X_n \sim TL(\alpha)$ and $Y_1, \cdots, Y_m \sim TL(\beta)$ independent for unknown α and β, where $TL(\alpha)$ has density ($0 < \alpha < 1$ and $0 < x < 1$)

$$f(x;\alpha) = 2\alpha(1-x)x^{\alpha-1}(2-x)^{\alpha-1}$$

Estimand: $g(\alpha, \beta) = \mathbb{P}_{\alpha,\beta}(X_1 > Y_1) = p$

Decision Space: $\mathcal{D} = \mathbb{R}$ **Loss:** $L(\theta, d) = (p-d)^2$

Initial estimator: Let ${}_2F_1(a, b; c; z)$ be the hypergeometric function and define $u := -\sum_{i=1}^{n} \log(2x_i - x_i^2)$ and $v := -\sum_{i=1}^{m} \log(2y_i - y_i^2)$. Then,

$$\delta(X,Y) = \begin{cases} {}_2F_1(1, 1-m; n; u/v), & \text{if } v \geq u, \\ 1 - {}_2F_1(1, 1-n; m; v/u), & \text{otherwise,}. \end{cases}$$

- UMVUE, Not Bayes, Not minimax

Alternative estimator: $\delta'(X) = \frac{1}{nm} \sum_i \sum_j I(X_i > Y_j)$

- UMVUE in non-parametric setting

Preference: The alternative estimator can be computed faster and it is UMVUE in a larger class of densities. If we have a good evidence that X and Y have Topp-Leone densities, it is better to use initial estimator. If not, it is better to use the alternative.

Convergence rate of Bayesian tensor estimator and its minimax optimality

Motivation: Tensor decomposition generalizes matrix factorization and can represent many practical applications, including collaborative filtering, multi-task learning, and spatio-temporal data analysis

Model:

- True tensor $A^* \in \mathbb{R}^{M_1 \times \cdots \times M_K}$ of order K
- Observe n samples $D_n = \{(X_i, Y_i)\}_{i=1}^n$
- Linear model: $Y_i = \langle A^*, X_i \rangle + \epsilon_i$, where $\epsilon_i \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2)$
- A^* is assumed to be "low-rank" ($A_{j_1,\ldots,j_K} = \sum_{r=1}^{d'} U_{r,j_1}^{(1)} U_{r,j_2}^{(2)} \cdots U_{r,j_K}^{(K)}$ for small d')

Estimand: $g(A^*) = A^*$

Decision Space: $\mathcal{D} = \mathbb{R}^{M_1 \times \cdots \times M_K}$

Loss: $L(A^*, \hat{A}) = \|A^* - \hat{A}\|^2$

Convergence rate of Bayesian tensor estimator and its minimax optimality

Inital Estimator: $\hat{A} = \int A\Pi(dA|D_n)$, where Π is the posterior distribution of A given D_n

- Bayes under prior

$$\pi(d') \propto \xi^{d'(M_1+\cdots+M_k)}$$

$$\pi(U^{(1)}, \ldots, U^{(K)}|d') \propto \exp\left\{-\frac{d'}{2\sigma_P^2}\sum_{k=1}^{K}\mathrm{Tr}[U^{(k)^T}U^{(k)}]\right\}$$

- Admissible: Loss is convex $\rightarrow$ Bayes estimator is unique $\rightarrow$ admissible
- Not UMRU: not even unbiased when $A^* \neq 0^{M_1\times\cdots\times M_K}$

Alternative estimator: $\hat{A} = (X^TX)^{-1}X^TY$ (standard linear regression)

- UMVU Estimator
- Initial estimator is still preferred in most cases: the alternative estimator is not even well-defined until n is large and has a large variance

Optimal Shrinkage Estimation of Mean Parameters in Family of Distributions With Quadratic Variance

This paper discusses simultaneous inference of mean parameters in a family of distributions with quadratic variance function. (Normal, Poisson, Gamma, . . .)

Assume we have p indep observations $Y_i,\ i = 1, \cdots, p$ that come from a distribution with $\mathbb{E}(Y_i) = \theta_i \in \Theta$ and $\text{Var}(Y_i) = \frac{\nu_0 + \nu_1\theta_i + \nu_2\theta_i^2}{\tau_i}$. Consider a class of semi-parametric shrinkage estimators of the form

$$\hat{\theta_i}^{b,\mu} = (1 - b_i) \cdot Y_i + b_i \cdot \mu \quad \text{with } b_i \in (0, 1]$$

This is inspired by Bayes estimator as each mean parameter θ_i is the weighted average of Y_i and the prior mean μ.

Under the sum of squared error loss we have

$$l_p(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{b,\mu}) = \frac{1}{p}\sum_{i=1}^{p}(\theta_i - \hat{\theta_i}^{b,\mu})^2$$

however, since it depends on the unknown $\boldsymbol{\theta}$, we need an estimator for the risk $R_p(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{b,\mu}) = \mathbb{E}[l_p(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{b,\mu})]$ in order to find the optimal parameter b and μ.

Optimal Shrinkage Estimation of Mean Parameters in Family of Distributions With Quadratic Variance

Model: Y_i indep with $\mathbb{E}(Y_i) = \theta_i \in \Theta$ unknown and variance quadratic

Estimand: $g(\theta) = \mathbb{E}[l_p(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{b,\mu})]$ for fixed $b, \mu \in \mathbb{R}$

Decision Space: $\mathcal{D} = \mathbb{R}$ **Loss:** $L(\theta, d) = (g(\theta) - d)^2$

Estimator:

$$\delta(Y) = \frac{1}{p}\sum_{i=1}^{p}[b_i^2 \cdot (Y_i - \mu)^2 + (1 - 2b_i) \cdot \frac{V(Y_i)}{\tau_i + \nu_2}]$$

where ν_k are known constants and $V(Y_i) = \nu_0 + \nu_1 Y_i + \nu_2 Y_i^2$.
τ_i is assumed to be known and can be interpreted as the within-group sample size.

- UMVUE, Not Bayes, Inadmissible

Alternative estimator: $\delta'(Y) = (\delta(Y))_+$

- Dominates δ

Preference: If the only goal were to estimate $g(\theta)$, I would prefer the dominating δ' over δ. However, choosing δ' doesn't lead to better selection of b and μ.

Minimax Estimators of Functionals (in particular, Entropy) of Discrete Distributions

Motivation: In a wide variety of fields, it is important to estimate functions of probability distributions. Particularly, estimating entropy is important. For example, in machine learning, we might want to maximize mutual information between input and output, where mutual information is a function of entropy. Thus, finding a good estimator for entropy is of practical importance. It might also be the case that the number of samples available is lower than, or of the order of, the size of the sample space. In this regime, other methods of estimating entropy often underperform, and hence there is a need for a new and improved estimator of entropy in such situations.

Minimax Estimators of Functionals (in particular, Entropy) of Discrete Distributions

Model: The set $\{P : P = (p_1, p_2, ..., p_S)\}$ of discrete probability distributions P of unknown support size S.

Estimand: Functionals of the form $F(P) = \sum_{i=1}^{S} f(p_i)$, where $f : (0, 1] \to \mathbb{R}$ is continuous. In particular, the entropy, $F(P) = \sum_{i=1}^{S} -p_i \ln(p_i)$

Decision Space: $\mathcal{D} = \mathbb{R}$ **Loss:** $L(F, \hat{F}) = \mathrm{E}_P(F(P) - \hat{F})^2$

Initial estimator: For a Bernoulli dist., $f^c(p) = f(\hat{p}) - \frac{f''(\hat{p})\hat{p}(1-\hat{p})}{2n}$.
When F is the entropy, we get $\hat{p}\ln(\hat{p}) + \frac{(1-\hat{p})}{2n}$

- Not UMVUE / UMRUE (but consistent), Not minimax (but rate-optimal, so approx minimax), Inadmissible

Alternative estimator: Work has been done by others to find a Bayes estimator under a Dirichlet prior for this problem, different from the estimator used by the authors of this paper.

Preference: The Bayes estimator under a Dirichlet prior is only consistent for $n \gg S$. The estimator in our paper is consistent for $n \gg \frac{S}{\ln(S)}$, so it is preferable in this regime.

Estimation of a multivariate normal mean with a bounded signal to noise ratio under scaled squared error loss (Kortbi and Marchand, 2013)

Model: $X \sim N_p(\theta, \sigma^2 I_p)$ and $S \sim \sigma^2 \chi^2_k$ independent
Estimand: θ
Loss: $L((\theta, \sigma), d) = \frac{\|d-\theta\|^2}{\sigma^2}$
Parameter Space: $\Omega_m = \{(\theta, \sigma) \in \mathbb{R}^p \times \mathbb{R}^+ : \frac{\|\theta\|}{\sigma} \leq m\}$
Decision Space: $\mathcal{D} = \mathbb{R}$

Motivation: This problem setup has many applications. One example is normal full rank linear models.

$Y \sim N_p(Z\beta, \sigma^2 I_p)$ with orthogonal design matrix Z, unknown parameter vector β, and constraint $\frac{\|\beta\|}{\sigma} \leq m$.

Proposed Estimator: $\delta_1(X, S) = \frac{m^2}{m^2+p} X$
Minimax under class of estimators linear in X, Not UMVUE/UMRUE, Not MRE
Alternative Estimator: $\delta_2(X, S) = X$
UMVUE/UMRUE
Preference: δ_1 is preferable in this setup because it makes use of the constrained parameter space to shrink the estimate. Hence, unbiasedness is not necessarily a desirable property.

Equivariant Minimax Dominators of the MLE in the Array Normal Model

Motivation: Tensor data is common in multiple fields (signal processing, machine learning, etc.). Most statistical analysis of tensors fit a model $X = \Theta + E$, where E represents some error term. Understanding the residual variation E plays an important role in tasks such as prediction, model-checking, and improved parameter estimation.

Model: $X \sim \mathcal{N}_{p_1 \times ... \times p_K}(0, \Sigma_K \otimes ... \otimes \Sigma_1)$

Estimand: $g(\theta) = (\sigma^2, \Sigma_K, ..., \Sigma_1)$, where $\sigma^2 > 0, \Sigma_i \in \mathcal{S}^+_{p_i}, |\Sigma_i| = 1$

Decision Space: $\mathbb{R}^{1 \times p_K^2 \times ... \times p_1^2}$

Loss: $L(\Sigma, S) = \frac{s^2}{\sigma^2} \sum_{k=1}^{K} \frac{p}{p_k} \operatorname{tr}[S_k \Sigma_k^{-1}] - Kp \log \frac{s^2}{\sigma^2} - Kp$

Equivariant Minimax Dominators of the MLE in the Array Normal Model

Define: $\mathcal{E}_k = \left(\mathbb{E}\left[(\sigma^2\Gamma_k^T\Sigma_k\Gamma_k)^{-1}|X\right]\right)^{-1}$, $\hat{\Sigma}_i(\Gamma, X) = \mathcal{E}_i/|\mathcal{E}_i|^{1/p_i}$,
$\hat{\sigma}^2(\Gamma, X) = \left(1/K\sum_{k=1}^K |\mathcal{E}_k|^{-1/p_k}\right)^{-1}$

Estimator: $\delta(X) = (\hat{\sigma}^2(I, X), \hat{\Sigma}_K(I, X), ..., \hat{\Sigma}_1(I, X))$.

- Minimax, UMREE under Lower Triangular Group, Inadmissible

Define: $S_k(X) = \int_{\mathcal{O}_{p_K}\times...\times\mathcal{O}_{p_1}} \frac{\Gamma_k^T\hat{\Sigma}_k(\Gamma, X)\Gamma_k}{\mathrm{tr}(\hat{\Sigma}_k(\Gamma, X))} d\Gamma_1...d\Gamma_K$
$\widetilde{\sigma}^2 = \int_{\mathcal{O}_{p_K}\times...\times\mathcal{O}_{p_1}} \hat{\sigma}^2(\Gamma, X)d\Gamma_1...d\Gamma_K$
$\widetilde{\Sigma}_k = S_k(X)/|S_k(X)|^{1/p_k}$

Alternative estimator: $\delta'(X) = (\widetilde{\sigma}^2, \widetilde{\Sigma}_K, ..., \widetilde{\Sigma}_1)$.

- Dominates δ, UMREE under Orthogonal Group

Preference: I would pick the first estimator, since it has comparable empirical results compared to the second estimator, and doesn't require approximately evaluating an infeasible integral over the space of orthogonal matrices

Convergence rate of Bayesian tensor estimator and its minimax optimality

Motivation: Low Rank Tensor approximation is useful for many practical applications, such as collaborative filtering, multitask learning, and spatial temporal data analysis. The problem is a NP hard problem in general. The paper suggests a computationally tractable Bayes Estimator which is valid without using convex regularization.

A tensor A $\in \mathbb{R}^{M_1 X M_2 X \ldots X M_k}$ has CP-rank d if there exist matrices U(k) $\in \mathbb{R}^{dX M_k}$ (k = 1; ...; K) such that $A_{j1;\ldots;jK} = \sum_{r=1}^{d} U^{(1)}_{r,j1} U^{(2)}_{r,j2} \ldots U^{(K)}_{r,jK}$ and d is the minimum number to yield this decomposition (we do not require the orthogonality of U(k)).

A general assumption is CP rank for the tensor A^* is small. This method is also adaptable in case the CP rank is not known.

Convergence rate of Bayesian tensor estimator and its minimax optimality

Model: $Y_i =< A^*, X_i > +\epsilon \quad X_i$ is tensor $\in \mathbb{R}^{M_1 X M_2 X \ldots X M_k}$, $<,>$ is the tensor inner product and ϵ_i is i.i.d. Gaussian noise $N(0, \sigma^2)$

Estimand: $g(\theta) = A^*$ a low rank tensor $\in \mathbb{R}^{M_1 X M_2 X \ldots X M_k}$

Decision Space: $\mathcal{D} = \mathbb{R}^{M_1 X M_2 X \ldots X M_k}$

Loss: $L(A, A^*) = ||A - A^*||_{L_2(P(X))}$ generalization error

Properties of the estimator:

- Biased, Bayes under Gaussian prior, Inadmissible, not UMRU, not UMVUE, nearly minimax optimal

Alternative estimator: $A' = \text{argmin}(l(A) + \lambda |A|_{tr})$
$|A|_{tr}$ is tensor trace norm

- Solution of convex regularized tensor problem

Preference: If the goal was to estimate A^* to get a quick convergence rate then I would prefer the convex estimator A' over A. However the Bayesian estimator suggested in the paper is valid in general in the absence of convexity as well and is adaptable to unknown rank as well.

Near Minimax Line Spectral Estimation

Motivation: Determine the locations and magnitudes of spectral lines from noisy temporal samples. It is a fundamental problem in statistical signal processing.
Model: Consider k sinusoids $\{e^{i2\pi j f_l}\}_{l=1}^k$, $f_l \in [0,1]$ unknown.
Let $x^* \in \mathbb{C}^n$ be $n = 2m+1$ equi-spaced samples of their superposition with unknown weights $c_l \in \mathbb{C}$,

$$x_j^* = \sum_{l=1}^{k} c_l \cdot e^{i2\pi j f_l}, \quad j \in \{-m, \ldots, m\}.$$

We observe a noised version $y \in \mathbb{C}^n$

$$y = x^* + \omega, \quad \omega \in N(0, \sigma^2 \mathbf{I}_n).$$

Estimand: $x^* \in \mathbb{C}^n$ and hence $\{c_l\}_{l=1}^k$, $\{f_l\}_{l=1}^k$ and k.
Remark: To render the estimation feasible, we require that $\{f_l\}_{l=1}^k$ satisfy separation condition

$$\min_{p \neq q} |f_p - f_q| > \frac{4}{n}.$$

Near Minimax Line Spectral Estimation

Estimand: $x^* \in \mathbb{C}^n$. **Decision space:** $\mathcal{D} = \mathbb{C}^n$.
Loss: $L(\hat{x}, x^*) = \frac{1}{n}\|\hat{x} - x^*\|_2^2$.
Estimator: (Atomic norm soft thresholding.) Let $\tau > 0$,

$$\hat{x} = \delta_\tau(y) = \operatorname*{argmin}_z \frac{1}{2}\|y - z\|_2^2 + \tau\|z\|_{\mathcal{A}},$$

where $\|\cdot\|_{\mathcal{A}}$ is atomic norm defined by

$$\|z\|_{\mathcal{A}} = \inf\{\sum_a c_a : z = \sum_a c_a a, a \in \mathcal{A}, c_a > 0\}$$

and $\mathcal{A}$ is a set of all sinusoids basis with frequency in a continuous set.
Remark: This problem can be formulated into an SDP problem.
Properties:

1. Like Lasso, not unbiased, not Bayes. (As $\tau > 0$.)
2. Asymptotically minimax. Risk $= O(\sigma^2 \frac{k \log n}{n})$, $\tau = O(\sigma\sqrt{n \log n})$
3. Some insurance for the estimation for $\{c_l\}_{l=1}^k$ and $\{f_l\}_{l=1}^k$.

Alternative estimator: $\delta_0(y) = y$ is unbiased. But it is not good in that 1) it is dominated by $\delta_\tau(y)$ for $\tau = O(\sigma\sqrt{n \log n})$, in the sparse setting . 2) $\{\hat{c}_l\}$ and $\{\hat{f}_l\}$ are not sparse.

Convex Point Estimation using Undirected Bayesian Transfer Hierarchies

Motivation: Classical hierarchical Bayes models do not cope well with scarcity of instances.($X'X$ may not be invertible!) It can also be computationally expensive due to the requirements of proper priors. It may not lead to efficient estimates due to non-convexity. The proposed undirected transfer learning in this paper leads to convex objectives. The condition of proper priors is no longer needed, which allows for flexible specification of joint distributions over transfer hierarchies.

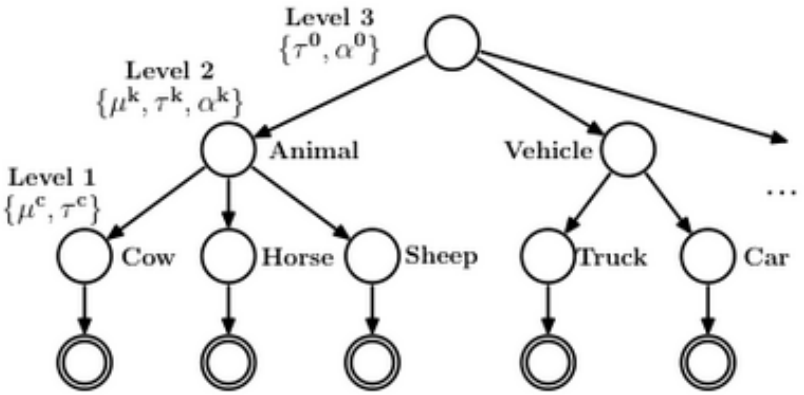

Figure: Hierarchical Bayes Model

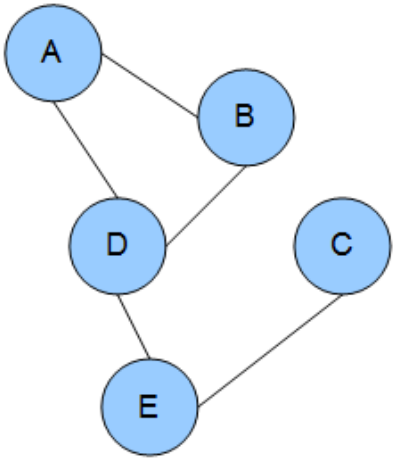

Figure: Undirected Markov Process

Convex Point Estimation using Undirected Bayesian Transfer Hierarchies

Model: $X_{ij} \sim \mathcal{N}(\mu_c, \Sigma_c)$ unknown μ_c, Σ_c, $\theta_c \equiv (\mu_c, \Sigma_c)$.
$\theta_c \sim \text{Div}(\theta_{par})$. (1) θ_{par} fixed; (2) $\theta_{par} \sim$ inverse gamma known.

Loss: $\log P(X, \theta_c, \theta_{par}) = \log \sum P(X_{ij}|\theta_c) - \beta \sum div(\theta_c, \theta_{par})$

$$L = \begin{cases} 1 & \text{if } |\delta - \theta_c| > \epsilon, \epsilon \to 0 \\ 0 & \text{otherwise} \end{cases}$$

Estimand: $\theta_c = (\mu_c, \Sigma_c)$ (estimated separately)

Decision Space: $\mu_c \in \mathbb{R}^{ij}, \Sigma_c \in \mathbb{R}^{ij \times ij}$.

Initial estimator: posterior mode

- Bayes, admissible, Not UMRUE

Alternative estimator: $\delta'(\mu_c) = \bar{X}_{ij}, \delta'(\Sigma_c) = \frac{1}{j(i-1)} \sum (X_{ij} - \bar{X}_{ij})^2$

- UMRUE

Preference: In order to derive δ, we must make sure $n > p$ to avoid singular covariance matrix. But scarcity of observation is always a problem.δ does not invoke such problems. Moreover, the added divergence term $(\beta \sum_i 1/\lambda_i \text{div}(\theta_i c, \theta_i par))$ gives more flexibility

Gaussian Sequence Model

- Consider the Gaussian sequence model with white noise, $y_i = \theta_i + \epsilon z_i$, $i \in \mathbb{N}$ where $\theta \in \Theta(a,c) := \left\{\sum_{i=1}^{\infty} a_i^2 \theta_i^2 \leq c^2\right\}$.

- Let $a_1 = 0$, $a_{2k} = a_{2k+1} = (2k)^m$ for $m > 0$ (this corresponds to the Sobolev space of order m under the trigonometric basis for the model $dX(t) = f(t)dt + \epsilon dW(t)$). Define the minimax risk $R_\epsilon(m,c) := \inf_{\widehat{\theta}} \sup_{\theta \in \Theta(a,c)} \mathbb{E}_\theta \|\theta - \widehat{\theta}\|^2$ and let $R_{\epsilon,L}(m,c)$ be the corresponding minimax risk among linear estimators.

- Pinsker's theorem states that

$$\lim_{\epsilon \to 0} \epsilon^{-\frac{4m}{2m+1}} R_{\epsilon,L}(m,c) = \lim_{\epsilon \to 0} \epsilon^{-\frac{4m}{2m+1}} R_\epsilon(m,c) = P_{m,c}$$

  where $P_{m,c} := \left(\frac{c^2(2m+1)}{\pi^{2m}}\right)^{\frac{1}{2m+1}} \left(\frac{m}{m+1}\right)^{\frac{2m}{2m+1}}$.
  The linear minimax estimator is given by the shrinkage estimator $\widehat{\theta}_i(y) = [1 - a_i/\mu]_+ y_i$ where μ is determined by $\sum_{i=1}^{\infty} \epsilon^2 a_i [\mu - a_i]_+ = c^2$.

Communication/Storage Constraints

- Consider the same setting but where there the estimators are constrained to use storage $C(\widehat{\theta})$ no greater than B_ϵ bits.

- Define the minimax risk in this computationally restrained setting as $R_\epsilon(m, c, B_\epsilon) := \inf_{\widehat{\theta}: C(\widehat{\theta}) \leq B_\epsilon} \sup_{\theta \in \Theta(a,c)} \mathbb{E}\|\widehat{\theta} - \theta\|^2$. Then,

$$R_\epsilon(m, c, B_\epsilon) \simeq \underbrace{P_{m,c}\epsilon^{\frac{4m}{2m+1}}}_{\text{estimation error}} + \underbrace{\frac{c^2 m^{2m}}{\pi^{2m}} B_\epsilon^{-2m}}_{\text{quantization error}}$$

- *Over-sufficient regime*: $B_\epsilon \gg \epsilon^{-\frac{2}{2m+1}}$, number of bits is very large, classical rate of convergence obtains under the same constant $P_{m,c}$.
  Sufficient regime: $B_\epsilon \sim \epsilon^{-\frac{2}{2m+1}}$, minimax rate is still of the same order but with a new constant $P_{m,c} + Q_{m,c}$.
  Insufficient regime: $B_\epsilon \ll \epsilon^{-\frac{2}{2m+1}}$ but with $B_\epsilon \to \infty$, rate deteriorates to $\lim_{\epsilon \to 0} B_\epsilon^{2m} R_\epsilon(m, c, B_\epsilon) = \frac{c^2 m^{2m}}{\pi^{2m}}$.

Note on the Linearity of Bayesian Estimates in the Dependent Case

Motivation: This paper shows the relationship between the MLE and the Bayesian estimator when we assume dependent observations. Especially when they are generated from Markov chains, it proves that the Bayesian estimator is just a linear function of the MLE.

Model: $X = (X_0, \ldots, X_n)$ is the first $(n+1)$ observations of a Markov chain with a finite state space $E = \{1, 2, \ldots, s\}$ and the transition matrix $P = (p_{ij})$. We assume the distribution of X_0 is known.

Estimand: $g(\theta) = (p_{ij}, (i,j) \in E^2, i \neq j) \in [0,1]^{s(s-1)}$

Decision Space: $\mathcal{D} = [0,1]^{s(s-1)}$ **Loss:** $L(\theta, d) = (g(\theta) - d)^2$

Note on the Linearity of Bayesian Estimates in the Dependent Case

Initial estimator: the natural choice of prior distributions is the multivariate beta distribution. Then the Bayesian estimator of p_{ij} is

$$\bar{p}_{ij} = N_n^{ij} + a_{ij} / \sum_{j \in E} N_n^{ij} + a_{ij}, \quad N_n^{ij} \equiv \sum_{t=1}^{n} I(X_{t-1} = i, X_t = j)$$

- UMVUE / UMRUE, Bayes, Inadmissible

Alternative estimator: Under the Jeffrey's prior distribution (assuming $s = 3$)

$$p_1 = \frac{p_{21}p_{32} + p_{31}(p_{21} + p_{23})}{(p_{12} + p_{13})(p_{32} + p_{23}) + p_{31}(p_{12} + p_{21} + p_{23}) + p_{21}(p_{13} + p_{32})}$$

Preference: I prefer the latter one because we can utilize on approximation scheme - the independent Metropolis-Hasting algorithm (IMH).

Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs

Motivation: Regression-discontinuity (RD) design has become one of the most important quasi-experimental tools used by economists and social scientists to study the effect of a particular treatment on a population.

However, they may be prone to sensitivity to the specific bandwidth employed and so this paper proposes a more robust confidence interval estimator using a novel way to estimate standard errors.
Model: $X_i, i = 1, ..., n$ is a random sample with a density with respect to the Lebesgue measure.

Estimand: Σ, the middle matrix of a generalized Huber-Eicker-White heteroskedasticity-robust standard error

Decision Space: $\mathcal{D} = \mathbb{R}_+^{(p+1)\times(q+1)}$ **Loss:** $L(\Sigma, d) = ||\Sigma - d||_F^2$

Initial estimator:

$$\Psi_{UV+,p,q}(h_n, b_n) = \frac{1}{n}\sum_{i=1}^{n} 1(X_i \geq 0) K_{h_n}(X_i) K_{b_n}(X_i)$$

$$\times r_p(X_i/h_n) r_q(X_i/b_n)' \sigma^2_{UV+}(X_i)$$

Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs

- Admissible, generalized Bayes, Not unbiased

Alternative estimator: Based on nearest-neighbor estimators with a fixed tuning parameter.

$$\hat{\Psi}_{UV+,p,q}(h_n, b_n) = \frac{1}{n}\sum_{i=1}^{n} 1(X_i \geq 0) K_{h_n}(X_i) K_{b_n}(X_i)$$

$$\times r_p(X_i/h_n) r_q(X_i/b_n)' \hat{\sigma}^2_{UV+}(X_i)$$

Here

$$\hat{\sigma}^2_{UV+}(X_i) = 1(X_i \geq 0)\frac{J}{J+1}$$

$$\times \left(U_i - \sum_{j=1}^{J} U_{l_{-,j}(i)}/J\right)\left(V_i - \sum_{j=1}^{J} V_{l_{-,j}(i)}/J\right)$$

- Asymptotically valid, more robust in finite samples

Preference: Alternative, if constrained by sample size, otherwise the initial estimator for its simplicity.

Ridge regression and asymptotic minimax estimation over spheres of growing dimension

Lee Dicker, Rutgers University

Motivation: In the regression setting, goal is to derive minimax estimators for the parameter vector β when β is d-dimensional and constrained to $S^{d-1}(\tau)$, the sphere of radius τ. Take $d \to \infty$.

Focus is on Ridge Regression estimator:

$$\hat{\beta}_r(t) = (X^T X + \frac{d}{t^2} I_d)^{-1} X^T y$$

which estimates the coefficients and constrains them via a regularization parameter d/t^2.

It turns out that the optimal regularization parameter is $t = \tau$, but this is typically unknown and must be estimated from the data.

Ridge regression and asymptotic minimax estimation over spheres of growing dimension

Model: $x_1, \ldots x_n \sim N(0, I_d)$; $\varepsilon \sim N(0, I_n)$; $y = X\beta + \varepsilon$

Estimand: Want to estimate **β**, which lies in $S^{d-1}(\tau)$

Risk Function: $R(\beta, \beta^*) = E(\| \beta - \beta^* \|^2)$

Decision Space: R^d

Estimator: Ridge regression estimator, with estimator of τ:

$$\hat{\beta}_r(\hat{\tau}) = (X^T X + \frac{d}{\hat{\tau}^2} I_d)^{-1} X^T y \quad where \quad \hat{\tau}^2 = max(\frac{1}{n}||y||^2 - 1, 0)$$

- **Not UMVUE (biased)**, **Bayes** **(under β ~** $N(0, \frac{\hat{\tau}^2}{d} I_d)$**)**
- **Not minimax**, **but asymptotically minimax** (as $d \to \infty$)

Alternative: Standard OLS estimate $\beta^\star = (X^T X)^{-1} X^T y$ is **UMVUE** by the Gauss-Markov Theorem

Preference: Ridge estimator is preferred as it is more robust to overfit as $d \to \infty$.

Introduction and motivation: In the estimation of the covariance matrix of a multivariate normal distribution, the best equivariant estimator under the group transformation of lower triangular matrices with positive diagonal elements, which is also called the James-Stein estimator, is known to be minimax by the Hunt-Stein theorem from the invariance approach.

However, finding a least favorable sequence of prior distributions has been an open question for a long time. This paper addresses this classical problem and accomplishes the specification of such a sequence.

This is an interesting issue in statistical decision theory. Moreover, in the case that the parameter space is restricted, it is not clear whether the best equivariant estimator maintains the minimax property.

Model: $\mathbf{S} = \mathbf{X}^t \mathbf{X}$, where $\mathbf{X} \sim N_n(0, \mathbf{\Sigma})$ for unknown $\mathbf{\Sigma} \in \mathbb{R}^{p \times p}$.

Estimand: $g(\mathbf{\Sigma}) = \mathbf{\Sigma}$.

Decision Space: $\mathcal{D} = \mathbb{R}^{p \times p}$ positive-definite

Loss: (Stein) $L_S(\mathbf{\Sigma}, \boldsymbol{\delta}) = \text{tr}(\mathbf{\Sigma}^{-1}\boldsymbol{\delta}) - \log|\mathbf{\Sigma}^{-1}\boldsymbol{\delta}| - p$.

Group of transforms: Cholesky decomposition: $\mathbf{S} = \mathbf{T}\,\mathbf{T}^t$, for $\mathbf{T} \in \Gamma_p^+$.

$$\mathcal{G}(p): \ (\mathbf{T}, \mathbf{\Sigma}) \to (\mathbf{AT}, \mathbf{A\Sigma A}^t), \quad \mathbf{A} \in \Gamma_p^+.$$

Estimator:

$$\boldsymbol{\delta}^{JS}(\mathbf{S}) = \mathbf{T}\,\mathbf{D}^{JS}\,\mathbf{T}^t$$

$$\mathbf{D}^{JS}(\mathbf{S}) = \text{diag}(d_1, \ldots, d_p), \qquad d_i = \frac{1}{n + p - 2i + 1}.$$

- Minimax, MREE, Inadmissible.

Dominating estimator:

$$\boldsymbol{\delta}^*(\mathbf{S}) = \mathbf{R}\,\text{diag}\left(\frac{\lambda_i}{n + p - 2i + 1}\right)\mathbf{R}^t, \quad \mathbf{S} = \mathbf{R}\,\text{diag}\,(\lambda_i)\,\mathbf{R}^t$$

- Dominates $\boldsymbol{\delta}^{JS}$, Minimax.

Preference: I would prefer $\boldsymbol{\delta}^*$ over $\boldsymbol{\delta}^{JS}$. However $\boldsymbol{\delta}^*$ is also Inadmissible.

A Comparison of the Classical Estimators with the Bayes Estimators of One Parameter Inverse Rayleigh Distribution

Motivation: The inverse Rayleigh Distribution has many applications in the area of reliability and survival studies. The object of this paper is to evaluate the Bayes estimator under different priors and loss.

Model: $f(x;\Theta) = \frac{2\Theta}{x^3} \times e^{-\frac{\Theta}{x^2}} \quad x > 0, \Theta > 0$

Loss: $L(\widehat{\Theta}, \Theta) = (\Theta - \widehat{\Theta})^2$

Initial estimator:

Bayes* : $\widehat{\Theta}_j = \frac{\sum_{j=0}^{k} a_j \frac{\Gamma(n+1+j)}{T^{j+1}\Gamma(n)}}{\sum_{j=0}^{k} a_j \frac{\Gamma(n+j)}{T^{j}\Gamma(n)}}$ $\quad (T = \sum \frac{1}{x_i^2})$

$$\widehat{\Theta}_{E1} = [a_0\left(\frac{n+1}{p}\right) + a_1(\frac{(n+2)(n+1)}{p^2})]/[a_0 + a_1(\frac{n+1}{p})]$$

Presenter: Yixin Tang Author, Year: Huda A.Rasheed Siba Zaki Ismail 2015

A Comparison of the Classical Estimators with the Bayes Estimators of One Parameter Inverse Rayleigh Distribution

Under $L(\widehat{\Theta}, \Theta) = (\Theta - \widehat{\Theta})^2$: not admissible; biased;
$\widehat{\Theta}_j$: $Jeffrey's\ prior$ $\widehat{\Theta}_{E1}$: $Exponential\ prior$

Alternative estimator:

$\widehat{\Theta_1'} = [n-2]/\sum \frac{1}{x_i^2}$ Min MSE admissible

$\widehat{\Theta_2'} = [n-1]/\sum \frac{1}{x_i^2}$ UMVUE; inadmissible; dominated by $\widehat{\Theta_1'}$

The choice of Min MSE or Bayes is dependent on which loss function is applied. When we choose mean square loss and bias is not so concerned, it is better to use Min MSE .

Presenter: Yixin Tang Author, Year: Huda A.Rasheed Siba Zaki Ismail 2015

Bayes Minimax Estimation Under Power Priors of Location Parameters for a Wide Class of Spherically Symmetric Distributions

Motivation: In 1956, Charles Stein showed that, when estimating the mean vector θ of a p-dimensional random vector with a normal distribution with identity covariance matrix, estimators of the form

$$(1 - a/(\|X\|^2 + b))X$$

dominate the usual estimator X. Since then, a significant portion of research in statistical theory has been focused on extending this result to other families of distributions. This particular paper focuses on spherically symmetric distributions, i.e. distributions that can be expressed as $f(\|x - \theta\|^2)$, where f is some real-valued function.

Bayes Minimax Estimation Under Power Priors of Location Parameters for a Wide Class of Spherically Symmetric Distributions

Model: $X \overset{\text{ind}}{\sim} f(\|x-\theta\|^2)$ for fixed $f : \mathbb{R} \to \mathbb{R}$ and unknown $\theta \in \mathbb{R}^p$,

Estimand: Location parameter θ

Decision Space: $\mathcal{D} = \mathbb{R}^p$ **Loss:** $L(\theta, d) = \|\theta - d\|^2$

Initial estimator:

$$\delta(X) = X + \frac{\nabla M(\|X\|^2)}{m(\|X\|)^2},$$

where ∇ denotes the gradient operator and m, M are the marginal densities with respect to $f(t)$ and $F(t) = \int_t^\infty f(t)dt$, respectively.

- Bayes, Minimax, Biased

Alternative estimator: $\delta'(X) = X$

- Unbiased

Preference: It is shown in the paper that δ dominates δ' at the cost of being biased. Therefore, I would only prefer δ' if unbiasedness were extremely important.

Estimating the Accuracies of Multiple Classifiers without Labeled Data

Motivation: Consider a Machine Learning setup, with a twist: instead of having labeled test data, you only have the predictions of multiple classifiers over unlabeled test data. This happens when, for example, you receive predictions from multiple experts, and you don't have labels because they're expensive, or will be available only after making the decision.
The paper tries to estimate the label of each data point (basically, to come up with a meta-classifier) by estimating the accuracy of each classifier, and combining their predictions accordingly.

Model: Consider a binary classification problem: $\mathcal{X}$ is an instance space with output space $\mathcal{Y} = \{-1, 1\}$. Now, let $\{f_i\}_{i=1}^{m}$ be $m \geq 3$ classifiers operating on $\mathcal{X}$. We have n unlabeled test instances, and the $m \times n$ matrix of predictions $\mathbf{Z}$, given by $\mathbf{Z}_{ij} = f_i(x_j)$.

Estimating the Accuracies of Multiple Classifiers without Labeled Data

Estimand: (i) Estimate the sensitivity (ψ) and specificity (η) of each classifier. (ii) Use this to estimate the true label of every data point.
Decision Space: (i) $\mathcal{D} = [0, 1]$ (ii) $\mathcal{D} = \{0, 1\}$ for each instance.
Loss: $L(\theta, d) = (g(\theta) - d)^{2*}$

Initial estimator: (i) $\hat{\psi} = \frac{1}{2}\left(1 + \hat{u} + \hat{v}\sqrt{\frac{1-b}{1+b}}\right)$ and $\hat{\eta} = \frac{1}{2}\left(1 - \hat{u} + \hat{v}\sqrt{\frac{1-b}{1+b}}\right)$ where b is the class imbalance, and $\hat{u}$ and $\hat{v}$ are unbiased estimates based on mean and covariance. (ii) $\hat{y}^{\mathsf{ML}} = \text{sign}(\sum_i f_i(x) \ln \alpha_i + \beta_i)$ where $\alpha_i = \frac{\psi_i \eta_i}{(1-\psi_i)(1-\eta_i)}$, $\beta_i = \frac{\psi_i(1-\psi_i)}{\eta_i(1-\eta_i)}$

- UMRUE, Admissible, Not Bayes

Alternative estimator: MAP estimate with prior probability p of $+1$.
Same $\hat{y}$, but with $\alpha_i = \frac{\psi_i^p \eta_i^{(1-p)}}{(1-\psi_i)^p (1-\eta_i)^{(1-p)}}$, $\beta_i = \frac{\psi_i^p (1-\psi_i)^p}{\eta_i^{(1-p)} (1-\eta_i)^{(1-p)}}$

Preference: If I had prior information that I wanted to incorporate, then I would pick my alternative. The initial estimator is neutral a priori.

*The analysis is general enough to accommodate other reasonable loss functions.

Estimation of Functionals of Sparse Covariance Matrices

Motivation: High-dimensional statistical tests often leads to null distributions that depend on functionals of correlation matrices. Take a two-sample hypothesis testing in high-dimensions as example. Suppose that we observe two independent samples $X_1^{(1)}, ..., X_{n1}^{(1)} \in \mathbb{R}^p \overset{\text{iid}}{\sim} \mathcal{N}(\mu_1, \Sigma)$ and $X_1^{(2)}, ..., X_{n2}^{(2)} \in \mathbb{R}^p \overset{\text{iid}}{\sim} \mathcal{N}(\mu_2, \Sigma)$. Let $n = n_1 + n_2$. The goal is to test $H_0 : \mu_1 = \mu_2$ vs $H_2 : \mu_1 \neq \mu_2$. A test for this problem is based on a statistic

$$M = (\bar{X}^{(1)} - \bar{X}^{(2)})^T (\bar{X}^{(1)} - \bar{X}^{(2)}) - \frac{n}{n_1 n_2} \text{tr}(\hat{\Sigma})$$

which is asymptotically normal under the null hypothesis with

$$\text{var}(M) = 2 \frac{n(n-1)}{(n_1 n_2)^2} \|\Sigma\|_F^2 (1 + o(1)).$$

In order to compute the critical value of the test, we need estimate the quadratic functional $\|\Sigma\|_F^2$. The paper investigates the optimal rate of estimating $\|\Sigma\|_F^2$ for a class of sparse covariance matrices (indeed correlation matrices).

Estimation of Functionals of Sparse Covariance Matrices

Model: $X_i, ..., X_n \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma)$ for unknown $\Sigma = \{\sigma_{ij}\}_{ij} \in \mathcal{F}_q(R)$, where $\mathcal{F}_q(R) = \{\Sigma \in \mathbb{S}_p^+ : \sum_{i \neq j} |\sigma_{ij}|^q \leq R, \text{diag}(\Sigma) = I_p\}$ for any $q \in [0, 2)$, $R > 0$

Estimand: $g(\Sigma) = \sum_{i \neq j} \sigma_{ij}^2$

Decision Space: $\mathcal{D} = \mathbb{R}_+$ **Loss:** $L(\Sigma, d) = (g(\Sigma) - d)^2$

Initial estimator: Let $\hat{\Sigma} = \{\hat{\sigma}_{ij}\}_{ij} = \frac{1}{n} \sum_{k=1}^n X_k X_k^T$. Then,

$$\delta(X) = \sum_{i \neq j} \hat{\sigma}_{ij}^2 \mathbb{I}\left(|\hat{\sigma}_{ij}| > \tau\right)$$

where $\tau = 2C_0 \sqrt{\frac{\gamma \log p}{n}}$.

- Inadmissible, Not UMVUE / UMRUE, Minimax to a constant rate

Alternative estimator: $\delta'(X) = \frac{\sum_{i \neq j} (\hat{\sigma}_{ij}^2 - \frac{1}{n})}{1 + \frac{1}{n}}$

- UMVUE / UMRUE

Preference: If my goal were to use an unbiased estimator, then I would prefer δ' over δ. Otherwise, if Σ is believed to be sparse, then I would prefer δ because it leverages the sparsity of Σ, and attains smaller variance while not introducing much bias.

Estimating the shape parameter of a Pareto distribution under restrictions

Motivation: Pareto distribution has found widespread applications in studies of economic data. This distribution can be used to adequately model quantities such as individual incomes. Economists have shown that the shape parameter of a Pareto distribution can be used to represent inequalities in income distributions. Thus deriving efficient estimates for the unknown shape parameter is of interest under these situations. For the existence of higher order moments of a Pareto distribution, the shape parameter has to be bounded below by a specified constant.
Now let $X_1, X_2, ..., X_n$ denotes a random sample from a Pareto distribution $P(\alpha, \beta)$ where α is a known scale parameter, and β is the shape parameter. We have priori knowledge that $\beta \geq \beta_0$, our task is to estimate β.

Estimating the shape parameter of a Pareto distribution under restrictions

Model: $X_1 \cdots X_n$ i.i.d with density $f(x, \alpha, \beta) = \frac{\beta \alpha^\beta}{x^{\beta+1}}$
$\alpha \leq x < \infty, 0 < \alpha < \infty, \beta \geq \beta_0$, known α

Estimand: $g(\beta) = n\beta$

Decision Space: $\mathcal{D} = [n\beta_0, \infty)$ **Loss:** $L(g(\beta), d) = (\frac{d}{n\beta} - 1)^2$

Initial estimator: $\delta_{g_0}(T) = \frac{n-2}{T} g_0(T)$, where $T = \frac{1}{n} \sum_{i=1}^n \ln \frac{X_i}{\alpha}$,
$g_0(y) = \frac{n\beta_0}{n-2} \frac{\int_y^\infty u^{n-2} e^{-n\beta_0 u} du}{\int_y^\infty u^{n-3} e^{-n\beta_0 u} du}$

- Not UMVUE / UMRUE, Minimax, Admissible

Alternative estimator: $\frac{n-1}{T}$ UMVUE

Preference: If my only goal was to derive an unbiased estimation, then I would prefer the dominating $\frac{n-1}{T}$ which is the UMVUE. However, $\delta_{g_0}(T)$ is minimax, admissible and generalized bayes, therefore we may prefer $\delta_{g_0}(T)$ if we do not require unbiasedness.

Learning with Noisy Labels

Motivation: Designing supervised learning algorithms that can learn from data sets with noisy labels is a problem of great practical importance. General results have not been obtained before.

Model: $(X, Y) \in \mathcal{X} \times \{\pm 1\} \overset{\text{iid}}{\sim} D$, w/ CCN $\Longrightarrow (X, \tilde{Y})$

Estimand: bounded loss function $l(t, y)$

Decision Space: measurable functions

Loss: $\mathbb{E}_{(X,Y)\sim D}[l(\text{argmin}_{f\in\mathcal{F}}[\frac{1}{n}\sum_{i=1}^{n}\tilde{l}(f(X_i), \tilde{Y}_i)], Y)]$

Initial estimator:

$$\tilde{l}(t, y) \equiv \frac{(1 - \rho_{-y})l(t, y) - \rho_y l(t, -y)}{1 - \rho_{+1} - \rho_{-1}}$$

where

$$\rho_{+1} = \mathbb{P}(\tilde{Y} = -1|Y = +1), \quad \rho_{-1} = \mathbb{P}(\tilde{Y} = +1|Y = -1), \quad \rho_{+1}+\rho_{-1} < 1$$

- Admissible, Not UMRUE, Not Bayes

Alternative estimator: No unique UMRUE exists

Preference: $\tilde{l}(t, y)$ for empirical risk minimization under CCN

Toward Minimax Off-Policy Value Estimation

Motivation: In reinforcement learning, we often want to find the estimated value of a new policy. If π is a distribution over actions and r_a is the reward for action a, we wish to calculate

$$E_\pi[R] = \sum_a \pi_a r_a$$

The reward function is unknown though, and since experiments can get expensive, it is often not feasible to run the policy, so we instead have to approximate the value off of previous observations. This paper examines several estimators for doing so. The one we'll look at here is called the importance sampling estimator.

Toward Minimax Off-Policy Value Estimation

Model: Let $\mathcal{A}$ be a finite set of actions. We are given n actions $A_1, ..., A_n$, $A_i \overset{\text{iid}}{\sim} \pi_D$, where π_D is a known distribution. We are also given real valued rewards $R_1, ..., R_n$, where $R_i \sim \Phi(\cdot|A_i)$ for some unknown distribution Φ. Let $r_\Phi(a) = E_{R\sim\Phi(\cdot|a)}[R]$. In this paper, we consider $\Phi \in \Psi_{\sigma^2, R_{\max}}$, where $\text{Var}(\Phi) \leq \sigma^2$ and $0 \leq r_\Phi(a) \leq R_{\max}$.

Estimand: Given another distribution π, we wish to estimate the value $v_\Phi^\pi = E_{A\sim\pi, R\sim\Phi(\cdot|A)}[R] = \sum_{a\in\mathcal{A}} \pi_a r_\phi(a)$.

Decision Space: $\mathcal{D} = \mathbb{R}$ **Loss:** $L(\theta, d) = (v_\Phi^\pi - d)^2$

Initial estimator: $\hat{v} = \frac{1}{n}\sum_{i=1}^n \frac{\pi(A_i)}{\pi_D(A_i)} R_i$

- Not Minimax, Not Bayes, Not UMRUE

Other estimators: There is no UMRUE estimator for this loss.
Sketch of Proof: In the class of models we considered, we can show that our estimator is UMRUE for the subclass where $\Phi(\cdot|a)$ is distributed $\mathcal{N}(\theta_a, \sigma^2)$ for $0 \leq \theta_a \leq R_{\max}$, for unknown parameters θ_a. However, we can also show that it is not UMRUE when $\Phi(\cdot|a)$ is uniform; thus there is no UMRUE for the entire model.

Iterative hard thresholding methods for l_0 regularized convex cone programming

Motivation: Sparse approximation of signals is of great interest for the signal processing community. Often, imaging data comes from a simple structure embedded in a noisy enviroment.
In these settings, images are often believed to be sparse (zero in many entries) in some tranformed domain from the original signal, such as a wavelet basis.
Model: Let $\mathcal{B} = \{y \in \mathbb{R}^m : l \leq y_i \leq u, i = 1, \ldots, m\}$ for known l, u.
$X \overset{\text{ind}}{\sim} \mathcal{N}(\beta\theta, \tau^2 I)$ for unknown $\theta \in \mathcal{B}$ such that $\text{supp}(\theta) = K$, known τ, $\beta \in \mathbb{R}^{n \times m}$, K.

Iterative hard thresholding methods for l_0 regularized convex cone programming

Estimand: $g(\theta) = \theta$.

Decision Space: $\mathcal{D} = \{\theta \in \mathcal{B} : \text{supp}(\theta) = K\}$.

Loss: $L(\theta, d) = \|g(\theta) - d\|_2^2$

Initial estimator: Let $\{\theta^t\}_{t\in\mathbb{N}}$ be the sequence defined recursively as

$$\theta_i^{t+1} = T_K\left(\Pi_{\mathcal{B}}(x^t - \frac{1}{L}(\beta^\intercal(\beta\theta^t - x)))\right)$$

, where

$$T_K(v)_i = \begin{cases} v_i & |v_i| > |v_j| \text{for at least } m - K \text{ distinct } j \\ 0 & \text{o.w..} \end{cases}$$

and let $\delta(x)$ be a fixed point of this sequence.

- Not UMVUE / UMRUE, Not Bayes, Inadmissible

Alternative estimator: $\delta'(X) = \underset{\theta\in\mathcal{B}}{\text{argmin}}\|\beta\theta - X\|_2 + \lambda\|\theta\|_0$, for λ chosen such that $\|\theta^*\|_0 = K$. Solve by testing every support pattern of θ, and solving the constrained least squares problem.

Presenter: Jun Yan

Consistent estimation of the fixed effects ordered logit model

Authors: Gregori Baetschmann,Kevin E. Staub, Rainer Winkelmann

Motivation: When estimating life satisfaction equations, or analysing determinants of job satisfaction or self-assessed health, researchers are often concerned about unobserved heterogeneity. Such heterogeneity can result from omitted variables or from subjective differences in anchoring of responses on the ordered response scale. Ordered logit model is a mathematical model to describe above problems.

Data-generating Model:

$$y_{it}^* = x_{it}\beta + \alpha_i + \varepsilon_{it}, \quad i = 1, .., N, \quad t = 1, .., T,$$

and

$$y_{it} = k, \text{ if } \tau_{ik} < y_{it}^* \leq \tau_{ik+1}, \quad k = 1, .., K,$$

where x_{it} are known. α_i are unknown but fixed $(i = 1, .., N)$. β is the parameter that we want to estimate. τ_{ik} are unknown and increasing, and $\tau_{i1} = -\infty, \tau_{iK+1} = +\infty$. Also, ε_{it} are independent and identically distributed with logistic cumulative distribution function

$$P(\varepsilon_{it} \leq z) = F(z) = \frac{1}{1 + \exp(-z)} \triangleq \Lambda(z). \tag{0.1}$$

Estimand, decision space and loss function:

The estimand is β. The decision space is R.

We can regard the loss function as $L(b) = |b - \beta|$.

Presenter: Jun Yan

The estimator in the paper:

The estimator is constructed in this way: Let $d_{it}^k = 1(y_{it} \geq k)$, then we have $P(d_{it}^k = 0) = \Lambda(\tau_{ik} - x_{it}\beta - \alpha_i)$, and

$$P_i^k(\beta) \triangleq P(d_{it}^k = j_{it}, t = 1, ..T \mid d_{it}^k = g_i) = \frac{\exp(j_{it}x_{it}\beta)}{\sum_{h \in B_i} \exp(h_{it}x_{it}\beta)},$$

where

$$B_i = \left\{h \in \{0,1\}^T \mid h_{it} = g_i\right\}.$$

Then maximize the likelihood we get the estimator:

$$\delta = \arg\max_b(L(b)),$$

$$L(b) = \sum_{k=2}^{K}\sum_{i=1}^{N} \log(P_i^k(b)).$$

Properties: It is not UMRU. It is Bayes estimator. It is minimax.

My statement: No unbiased estimator exists in this case. So no UMRU exists.

My alternative estimator:

The process of constructing δ is a kind of MLE method, in which y_{it} is truncated at k to get $d_{it}^k = 1(y_{it} \geq k)$. Obviously, when truncating y_{it}, some useful information is ingored. So an alternative approach is to do the MLE directly without truncating y_{it}, and we can use the MLE estimate of β as δ'.

Preference: If the data is very large and the number of parameters is small, then I prefer my alternative estimator. Otherwise I prefer the original one.

Unbiased Estimation for the General Half-Normal Distribution

Motivation: Half-normal distribution, a special case of folded normal and truncated normal distributions, appears in noise magnitude models and fatigue crack lifetime models. This paper studies unbiased estimation of location and scale parameters of general half-normal distribution $\mathcal{HN}(\xi, \eta)$, where the mean is $\xi + \eta\sqrt{2/\pi}$.

$$f(x) = \frac{\sqrt{2}}{\eta\sqrt{\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\xi}{\eta}\right)^2\right\} \mathbb{I}_{[\xi,+\infty)}(x)$$

In this presentation, we consider the unbiased estimator for the location parameter of $\mathcal{HN}(\xi, \eta)$ proposed in the paper.

Model: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{HN}(\xi, \eta)$ for unknown $\xi \in \mathbb{R}$ and known $\eta > 0$.

Estimand: ξ **Decision Space:** $\mathcal{D} = \mathbb{R}$ **Loss:** $L(\xi, d) = \left(\frac{\xi - d}{\eta}\right)^2$

Unbiased Estimation for the General Half-Normal Distribution

Initial estimator: Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $X_{(1)} = \min(X_1, \ldots, X_n)$.

Also, X_i can be written as $X_i = \xi + \eta Y_i$ where $Y_i = |Z_i|, Z_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$.

$$\hat{\xi} = \frac{\sqrt{\frac{2}{\pi}}X_{(1)} - \mathbb{E}[Y_{(1)}]\overline{X}}{\sqrt{\frac{2}{\pi}} - \mathbb{E}[Y_{(1)}]}, \mathbb{E}[\hat{\xi}] = \frac{\sqrt{\frac{2}{\pi}}(\xi + \eta\mathbb{E}[Y_{(1)}]) - \mathbb{E}[Y_{(1)}](\xi + \eta\sqrt{\frac{2}{\pi}})}{\sqrt{\frac{2}{\pi}} - \mathbb{E}[Y_{(1)}]} = \xi$$

- Minimum Risk Location Equivariant
- Minimum Risk Location-Scale Equivariant (If we don't know η)
- Not Bayes (unbiased with nonzero MSE)

Alternative estimator: The posterior w.r.t. prior $\Lambda(\xi) = \mathcal{N}(\mu, \sigma)$ is a truncated normal distribution. The Bayes estimator is the posterior mean.

$$\alpha = \frac{(n/\eta^2)\overline{X} + (1/\sigma^2)\mu}{n/\eta^2 + 1/\sigma^2}, \beta = \frac{1}{n/\eta^2 + 1/\sigma^2}, \tilde{\xi} = \alpha - \sqrt{\beta}\frac{\phi\left((X_{(1)} - \alpha)/\beta\right)}{\Phi\left((X_{(1)} - \alpha)/\beta\right)}$$

Preference: If we don't have prior knowledge about ξ, $\hat{\xi}$ is simpler, unbiased, and is a MREE. I would prefer $\hat{\xi}$.

Chebyshev Polynomials, Moment Matching, and Optimal Estimation of the Unseen

The Story: (from prof. Bradley Efron)
In 1940's, the naturalist Corbet had spent two years trapping butterflies in Malaya.

Table 1: Corbet's data on how often species of butterflies were trapped

Frequency	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Species	118	74	44	24	29	22	20	19	20	15	12	14	6	12	6

Corbet asked R. A. Fisher: **how many new species he would see, if he returned to Malaya for another two years of trapping?** Fisher gave his answer:

$$118 - 74 + 44 - 24 + 29 - 22... = 75 \pm 20.9. \quad (1)$$

The "butterflies" can be people in daily life, words in a book, genes of human... **Is this method optimal?**

Chebyshev Polynomials, Moment Matching, and Optimal Estimation of the Unseen

Model: $P = (p_1, p_2, ..., p_k)$ unknown, $p_i \geq 1/k$; $X = (X_1, ..., X_n) \overset{iid}{\sim} P$.

Estimand: $g(P) = \sum_i \mathbb{I}_{\{p_i > 0\}}$.

Decision Space: $\mathcal{D} = \{0, 1, ..., k\}$ **Loss:** $L(P, d) = (g(P) - d)^2$

Initial estimator: Let f_j be the number of elements that occur j times. Then

$$\tilde{g}(P) = \sum_{j=1}^{\infty} \alpha_L(j) f_j, \tag{2}$$

where

$$\alpha_L(j) = \begin{cases} a_j j!/n^j + 1, & j \leq L \\ 1, & j > L. \end{cases} \tag{3}$$

- No UMVUE, Bayes, Inadmissible, Minimax (Sample order).

No Unbiased Estimator Exists

Preference: Stable and accurate. If Sir Fisher were asked how many new species would be trapped for another 10 years, his estimator would explode :)