From Brown’s Ancillarity Paradox to an Empirical Partially Bayes Estimator for Mendelian Randomization

Qingyuan Zhao
(Joint work with Yang Chen, Jingshu Wang, Dylan Small)

Department of Statistics, The Wharton School, University of Pennsylvania

November 30, 2018
THE 1985 WALD MEMORIAL LECTURES

AN ANCILLARITY PARADOX WHICH APPEARS IN MULTIPLE LINEAR REGRESSION\(^1\)

BY LAWRENCE D. BROWN

Cornell University

Consider a multiple linear regression in which \( Y_i, \ i = 1, \ldots, n, \) are independent normal variables with variance \( \sigma^2 \) and \( E(Y_i) = \alpha + V_i'\beta, \) where \( V_i \in \mathbb{R}^r \) and \( \beta \in \mathbb{R}^r. \) Let \( \hat{\alpha} \) denote the usual least squares estimator of \( \alpha. \) Suppose that \( V_i \) are themselves observations of independent multivariate normal random variables with mean 0 and known, nonsingular covariance matrix \( \theta. \) Then \( \hat{\alpha} \) is inadmissible under squared error loss if \( r \geq 2. \)

Several estimators dominating \( \hat{\alpha} \) when \( r \geq 3 \) are presented. Analogous results are presented for the case where \( \sigma^2 \) or \( \theta \) are unknown and some other generalizations are also considered. It is noted that some of these results for \( r \geq 3 \) appear in earlier papers of Baranchik and of Takada.

\( \{V_i\} \) are ancillary statistics in the above setting. Hence admissibility of \( \hat{\alpha} \) depends on the distribution of the ancillary statistics, since if \( \{V_i\} \) is fixed instead of random, then \( \hat{\alpha} \) is admissible. This fact contradicts a widely held notion about ancillary statistics; some interpretations and consequences of this paradox are briefly discussed.
A similar phenomenon around the same time

The Annals of Statistics
1985, Vol. 13, No. 3, 914–931

USING EMPIRICAL PARTIALLY BAYES INFEERENCE FOR INCREASED EFFICIENCY

BY BRUCE G. LINDSAY

The Pennsylvania State University

Empirical partially Bayes methods are considered as a means of improving efficiency in a class of problems in which the number of nuisance parameters increases to infinity. In the method used, the parameter of interest is estimated in an asymptotically unbiased way while James-Stein shrinkage is applied to the nuisance parameter estimates. When the shrinkage estimators are carefully chosen, this yields estimators generally more efficient than maximum likelihood. In the models considered, the conditional structure imposed allows construction of a simple estimator which is broadly consistent and efficient.

Some relations

► Both papers consider regression-type problems.
► MLE is sub-optimal in both papers.
► The construction in both papers relies on shrinkage.
► Main difference: the 2nd paper has growing # nuisance parameters.
An example: Linear regression with measurement error

- Suppose we observe

\[
\left( \begin{array}{c} X \\ Y \end{array} \right) \sim N \left( \left( \begin{array}{c} \mu \\ \beta \mu \end{array} \right), I_{2n} \right)
\]

and we are interested in estimating \( \beta \).

- The log-likelihood function is given by

\[
l(\beta, \mu) = \sum_{i=1}^{n} l_i(\beta, \mu_i) = -\frac{1}{2} \sum_{i=1}^{n}(X_i - \mu_i)^2 + (Y_i - \beta \mu_i)^2.
\]

- It is straightforward to show that

\[
\hat{\mu}_{i, MLE}(\beta) = \frac{X_i + \beta Y_i}{1 + \beta^2}
\]

is a sufficient “statistic” for \( \mu_i \).
The conditional score function (Lindsay, 1985) for $\beta$ is given by

$$s_i(\beta, \mu_i) = \frac{\partial}{\partial \beta} l_i(\beta, \mu_i) - \mathbb{E}\left[\frac{\partial}{\partial \beta} l(\beta, \mu_i) \mid \hat{\mu}_{i, \text{MLE}}(\beta)\right]$$

$$= \frac{\mu_i(Y_i - \beta X_i)}{1 + \beta^2}.$$

Observation 1: $\mu_i$ only appear as “weight” to the residual $Y_i - \beta X_i$.
Observation 2: $Y_i - \beta X_i$ has mean 0 and is independent of $\mu_{i, \text{MLE}}$.

This motivates a general class of estimating equations

$$s(\beta) = \sum_{i=1}^{n} \frac{f_i(\hat{\mu}_{i, \text{MLE}}, \beta)}{1 + \beta^2} \cdot \frac{(Y_i - \beta X_i)}{1 + \beta^2}.$$

The maximum likelihood equation can be recovered by using $f_i(\hat{\mu}_{i, \text{MLE}}, \beta) = \hat{\mu}_{i, \text{MLE}}$. 
Empirical partially Bayes

- Lindsay showed that the optimal weight is given by

\[ f_i(\hat{\mu}_{i, \text{MLE}}, \beta) = \mathbb{E}[\mu_i | \hat{\mu}_{i, \text{MLE}}], \]

where the expectation is taken over the (empirical) distribution of \( \mu_i \).

- In practice, he suggested to use empirical Bayes estimate of \( \mu_i \) as the weight. Hence the name “empirical partially Bayes”.

- The whole approach can be generalized to exponential family, see his paper. In fact, the working example was paired exponentials.

What’s different with the normal distribution

- With normal distribution, this whole approach with James-Stein shrinkage (i.e. normal prior for \( \mu_i \)) won’t make any difference because the shrinkage is multiplicative.

- Instead, it is crucial to selectively shrink \( \hat{\mu}_{i, \text{MLE}} \) (e.g. soft thresholding).
Application: Mendelian randomization

Cause and effect

- Suppose we want to estimate the causal effect of obesity (measured by the body mass index, BMI) on the risk of heart disease.

Observational studies

- Traditionally, this is done by observational studies that control confounders such as demographics, lifestyle, socioeconomic status.
- However, this can still have bias due to unmeasured confounders.

Mendelian randomization

- An increasingly popular method is to use genetic variants as instrumental variables.
- For example, FTO (fat mass and obesity-associated protein) is known to be related to obesity in humans.
- If obesity increases the risk of heart disease, variants near the FTO gene should also be associated with heart disease.
Implementation

Correspondence with the regression problem

- $X$: Variant association with BMI.
- $Y$: Variant association with heart disease.
- $\beta$: Causal effect of obesity (BMI) on heart disease.

Further details

- The empirical partially Bayes estimator allows us to use all the variants, even if they are only weakly associated with BMI (i.e. the corresponding $\mu$ is small).
- We used spike-and-slab Gaussian mixture to model $\mu$.
- We modified the estimating equation to allow for invalid instruments (e.g. variants that affect heart health through other pathways such as LDL cholesterol).
Results

CAD = Coronary Artery Disease; IS = Ischemic Stroke.

Z-score for the effect of BMI on IS
With MLE weights = 3.08; With shrinkage weights = 3.34.
Thank you!