Discussion:
Inference for Losers

Roger Koenker

University College London

Selective Inference Seminar: 17 February 2022

Inference on the Best (or Nearly Best)

- We have independent, noisy measurements of performance for K treatments,

$$X_k = \mu_k + u_k, \quad u_k \sim \mathcal{N}(0, \sigma_k^2), \quad k = 1, \cdots, K.$$

- Let's consider the σ_k's known constants.
- A k^* is selected as best (or 3rd best) from the $1, \cdots, K$.
- We would like to construct a confidence interval for μ_{k^*}.
- Ignoring the selection choice yields biased intervals.
- Bias correction based on truncated Gaussian representation of X_{k^*}.
- $\mathcal{O}(K \log K)$ algorithm for construction of truncation set.
- Question: Suppose $\mu_k \equiv 0$ and $\sigma_k \equiv 1$ what would the confidence interval look like for μ_{k^*} with $k^* = \{k | X_k = \max\{X_j \; j = 1, \cdots, K\}\}$.

Bayes-time, and the Livin' is Easy

Imagine the Bayesian:

- Given a prior on the μ's,
- Guilt free posterior credible intervals are constructed
- From a strict Bayesian perspective: No bias, no cry.
- If the prior were the usual improper, $\pi(\mu) \propto 1$, our Bayesian has committed the same sin Dillon went to all that trouble to correct.
- Beware the casual uninformative prior!
- Dawid (1994) is highly recommended.

Better Living through Better Priors

Suppose we now consider the conjugate prior, $\pi(\mu) \sim \mathcal{N}(0, \tau^2 I_K)$

- Then the posterior for μ_{k^*} is

$$\mu_{k^*} \mid (X_{k^*} = x_{k^*}) \sim \mathcal{N}\left(\frac{\tau^2}{\tau^2 + \sigma_{k^*}} x_{k^*}, \left(\frac{1}{\tau^2} + \frac{1}{\sigma_{k^*}^2}\right)^{-1}\right)$$

- Rather than accepting x_{k^*} at face-value it is shrunken toward 0 by an amount depending upon τ^2 and $\sigma_{k^*}^2$.
- Posterior credible intervals can be easily constructed as well.
- Beware the casual conjugate prior!
- When K is large, a prior G for the μ_k's can be estimated:

$$\hat{G} = \text{argmax}_{G \in \mathcal{G}} \sum_{k=1}^{K} \log \int \varphi_{\sigma_k}(x_k - \mu) dG(\mu)$$

- Nonlinear shrinkage with this empirical Bayes prior converges to optimal Bayes rule based on $G_n(\mu) = n^{-1} \sum \mathbb{1}(\mu_k \leqslant \mu)$ provided that the μ distribution isn't too heavy tailed.
- Comparisons with these posterior intervals might be interesting.

Some References

ANDREWS, I., D. BOWEN, T. KITAGAWA, AND A. MCCLOSKEY (2022): "Inference for Losers," *American Economic Association Papers and Proceedings*, forthcoming.

DAWID, A. P. (1994): "Selection paradoxes of Bayesian inference," in *Multivariate Analysis and Its Applications*, ed. by T. W. Anderson, K. T. Fang, and I. Olkin, pp. 211–220. IMS.

EFRON, B. (2019): "Bayes, Oracle Bayes and Empirical Bayes," *Statistical Science*, 34, 177–201.

GU, J., AND R. KOENKER (2022): "Invidious Comparisons: Ranking and Selection as Compound Decisions," *Econometrica*, forthcoming.

KIEFER, J., AND J. WOLFOWITZ (1956): "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters," *The Annals of Mathematical Statistics*, 27, 887–906.

KOENKER, R., AND I. MIZERA (2014): "Convex Optimization, Shape Constraints, Compound Decisions and Empirical Bayes Rules," *J. of Am. Stat. Assoc.*, 109, 674–685.

ROBBINS, H. (1956): "An Empirical Bayes Approach to Statistics," in *Proceedings of the Third Berkeley Symposium*, vol. I, pp. 157–163. University of California Press: Berkeley.

SOLOFF, J. A., A. GUNTUBOYINA, AND B. SEN (2021): "Multivariate, Heteroscedastic Empirical Bayes via Nonparametric Maximum Likelihood," Available from `https://arxiv.org/abs/2109.03466`.