On the Attainable Statistical Error in Permutation Invariant Problems

Asaf Weinstein

Abstract. Consider observing $X_i \sim f_{\theta_i}$, $i = 1, ..., m$, where θ_i are fixed and unknown parameters. This situation is commonly encountered across different fields of application, where the scientist would like to answer questions such as: Is any of the θ_i different from zero? Which of the θ_i can be declared nonzero such that the False discovery rate (FDR) is controlled? Which of the θ_i can be declared positive, such that the directional-FDR is controlled? What are the magnitudes, or the ranks, of the θ_i's corresponding to the largest K observations X_i? And so on. These question are different in various aspects, and are usually treated in separate frameworks in the statistical literature. Still, one common feature to all of these questions is that they give no *a priori* preference to any of the individual components $i = 1, ..., m$. We leverage this fact to show that, if the statistical model is correspondingly indifferent to the labels $i = 1, ..., m$—which is a weaker assumption than independence of the X_i's—then all of these problems can, in a sense, be solved *optimally*. More specifically, we show that, for every fixed $\boldsymbol{\theta} = (\theta_1, ..., \theta_m)$, the procedure that postulates a uniform prior over all permutations of $\boldsymbol{\theta}$, and minimizes the corresponding Bayes risk, attains the minimum frequentist risk among all procedures that respect the permutation invariance in the problem. This gives an attainable lower bound on the risk of effectively any reasonable decision rule in a permutation invariant problem. On the algorithmic side, we discuss connections to De Finetti's theorem and the prospect of asymptotically attaining this bound uniformly in $\boldsymbol{\theta}$ by adopting Robbins's empirical Bayes approach.