Asymptotic Minimaxity of False Discovery Rate Thresholding for Sparse Exponential Data

September 23, 2004

Abstract

Control of the *False Discovery Rate* (FDR) is a recent innovation in multiple hypothesis testing, allowing the user to limit the fraction of rejected null hypotheses which correspond to false rejections (i.e. false discoveries). The FDR principle also can be used in multiparameter estimation problems to set thresholds for separating signal from noise when the signal is sparse. Success has been proven when the noise is Gaussian; see [1].

In this paper, we consider the application of FDR thresholding to a non-Gaussian setting, in hopes of learning whether the good asymptotic properties of FDR thresholding as an estimation tool hold more broadly than just at the standard Gaussian model. We consider a vector X_i, $i = 1, \ldots, n$ whose coordinates are independent exponential with individual means μ_i. The vector μ is thought to be sparse, with most coordinates 1 and a small fraction significantly larger than 1. This models a situation where most coordinates are simply 'noise', but a small fraction of the coordinates contain 'signal'.

We develop an estimation theory working with $\log(\mu_i)$ as the estimand, and use the per-coordinate mean-squared error in recovering $\log(\mu_i)$ to measure risk. We consider minimax estimation over parameter spaces defined by constraints on the per-coordinate ℓ^p norm of $\log(\mu_i)$: $Ave_i \log^p(\mu_i) \leq \eta^p$. Members of such spaces are vectors (μ_i) which are sparsely heterogeneous.

We find that, for large n and small η, FDR thresholding is nearly minimax, increasingly so as η decreases. The FDR control parameter $0 < q < 1$ plays an important role: when $q \leq \frac{1}{2}$, the FDR estimator is nearly minimax, while choosing a fixed $q > \frac{1}{2}$ prevents near minimaxity. These conclusions mirror those found by Abramovich et al in the Gaussian case.

The techniques developed here seem applicable to a wide range of other distributional assumptions, other loss measures, and non-i.i.d. dependency structures.

We will also compare our results with work in the Gaussian setting [1].

This is joint work with David Donoho.

References

[1] Abramovich, F. and Benjamini, Y. and Donoho, D. and Johnstone, I. (2000). Adapting to Unkown Sparsity by Controlling the False Discovery Rate. *Accepted for publication pending revision, Ann. Statist..*