Optimal Rates and Tradeoffs in Multiple Testing

Maxim Rabinovich, Aaditya Ramdas,
Michael I. Jordan, and Martin J. Wainwright
Department of Electrical Engineering and Computer Science, University of California—Berkeley

Multiple hypothesis testing

Testing problems often occur in large groups:

- Associations between disease and each site in a genome.
- Temperature fluctuations in astronomical data.
- Loci of infection in an epidemic.

Mathematically, we model this by $n \gg 1$ choices between nulls and alternatives:

$$H_{0i} \text{ (null) vs. } H_{1i} \text{ (alt)}, \quad 1 \le i \le n$$

Two error measures:

$$\text{Type I: } \mathrm{FDR} = \frac{\#\text{ false positives}}{\#\text{ rejections}},$$
$$\text{Type II: } \mathrm{FNR} = \frac{\#\text{ false negatives}}{\#\text{ signals}}$$

Sparse generalized Gaussians model

Assumption.

- n observations $X_1, \ldots, X_n$ such that
$$X_i \sim \psi_0(\,) \text{ if } i \text{ is null,}$$
$$X_i - \mu_n \sim \psi_0(\,) \text{ if } i \text{ is non-null,}$$
where
$$\mu_n = (\gamma r \log n)^{1/\gamma}$$
- For some $0 < \beta < 1$, there are $n^{1-\beta} \ll n$ non-nulls.
- For some $\gamma > 1$, tails go like $\exp\left(-\frac{|x|^\gamma}{\gamma}\right)$, so
$$\log \mathbb{P}_0 (X \ge x) = -\frac{|x|^\gamma}{\gamma} \pm O(1)$$

Algorithms

We limit our attention to thresholding algorithms that are given a target FDR level q:

$$\text{reject } i \iff X_i \ge t(X_{1:n},\ q) \text{ for some threshold function } t.$$

Two main choices of threshold:

- **Benjamini-Hochberg (BH).**
$$t_{\mathrm{BH}}(X_{1:n},\ q) = \min\left\{t \colon \mathbb{P}_0(X \ge t) \le \frac{q \cdot \#(X_i \ge t)}{n}\right\}$$
- **Barber-Candès (BC).**
$$t_{\mathrm{BC}}(X_{1:n},\ q) = \min\left\{t \colon \frac{\#(X_i \le -t)}{n} \le \frac{q \cdot \#(X_i \ge t)}{n}\right\}$$

Main results

- **Describe the critical regime** where good FDR implies $0 < \mathrm{FNR} < 1$.
- **Determine the optimal rate for the feasible regime**, which is $n^{-\kappa^*}$, where κ^* solves a fixed point equation.
- **Establish the optimal FDR-FNR tradeoff**, which is described by the two sides of the fixed point equation.
- **Prove optimality of both BC and BH.**

Proof strategy

- Analyze FDR-FNR tradeoff for procedures with **data-independent thresholds.**
- Establish a **comparison principle** showing that *any* thresholding procedure must behave approximately like a data-independent thresholding procedure.
- Show that both BC and BH place their thresholds **near the minimal value for a given FDR.**

Related work

- Asymptotic optimality (without rates) of BH and BC in the feasible regime—that is, $\mathrm{FDR} + \mathrm{FNR} \to 0$ [Arias-Castro and Chen, 2016].
- Asymptotic phase transition for the single testing problem for the global null
$$H_0^* \colon \text{ all } i \text{ are null vs. } H_1^* \colon \exists i \text{ non-null,}$$
which is statistically much easier (better rates, more favorable detection boundary) [Donoho and Jin, 2004].
- Asymptotic Bayes optimality of FDR control procedures for binary classification with class imbalance at the critical point $\mu_n = (\gamma\beta \log n)^{1/\gamma}$ [Neuvial et al., 2012].

Critical regime

In finite samples, there is a lower bound strictly above β such that shifts with r smaller than that lower bound cannot be detected. This lower bound r_n^- depends on the target FDR rate.

Theorem 1. *There exists a constant $c_0 > 0$ such that*

$$r \le r_n^- = \begin{cases} \beta + \frac{\log \frac{1}{q_n}}{\log n} + \frac{\log \frac{1}{c_0}}{\log n}, & \text{if } q_n \ge \frac{8 \log 4}{3} \cdot \frac{1}{n^{1-\beta}}, \\ 1 + \frac{\log \frac{1}{c_0}}{\log n} & \text{o.w.} \end{cases}$$

then

$$\mathrm{FDR} \le q_n \Longrightarrow \mathrm{FNR} \ge \frac{1}{32}.$$

Optimal rate and tradeoff

Here we assume $r \ge r_n^-$, so we are in the regime where the problem is feasible. The crucial quantity in defining the rates is the γ-distance:

$$d_\gamma(a, b) = \left|a^{1/\gamma} - b^{1/\gamma}\right|^\gamma$$

Theorem 2. *Any threshold-based multiple testing procedure satisfies*

$$\mathrm{FDR} + \mathrm{FNR} \gtrsim n^{-\kappa^*},$$

where κ^* is the unique solution to

$$\kappa = d_\gamma(\beta + \kappa, r)$$

Theorem 3. *More precisely, any threshold-based multiple testing procedure satisfies the tradeoff*

$$\mathrm{FDR} \lesssim n^{-\kappa} \Longrightarrow \mathrm{FNR} \gtrsim n^{-d_\gamma(\beta+\kappa, r)}.$$

Dense and barely feasible instances

In addition to addressing the classical problem, we analyze two interesting but challenging regimes:

Example (Dense case)**.** Suppose the number of signals is αn for a sequence α_n. If the shift parameter $0 < r < 1$ is a constant, then

$$\mathrm{FDR} + \mathrm{FNR} \gtrsim n^{-r/2^\gamma}$$

and both BC and BH achieve the optimal rate.

Example (Barely feasible case)**.** Suppose $0 < \beta < 1$ is fixed. If

$$r_n - \beta = \Delta_n \to 0,$$

then the optimal rate is given by

$$\mathrm{FDR}_n + \mathrm{FNR}_n \ge \exp\left(-c_\beta \Delta_n \log n \pm o(\Delta_n \log n)\right),$$

Further, BH and BC achieve this rate up to the constants in the little-o term.

References

Ery Arias-Castro and Shiyun Chen. Distribution-free multiple testing. *arXiv preprint arXiv:1604.07520*, 2016.

David Donoho and Jiashun Jin. Higher criticism for detecting sparse heterogeneous mixtures. *Annals of Statistics*, pages 962–994, 2004.

Pierre Neuvial, Etienne Roquain, et al. On false discovery rate thresholding for classification under sparsity. *The Annals of Statistics*, 40(5):2572–2600, 2012.