Optimal Rates and Tradeoffs in Multiple Testing

Maxim Rabinovich, **Aaditya Ramdas**, Michael I. Jordan, Martin J. Wainwright

June 25, 2017

More tests, more (kinds of) problems

Consider, for now, the Gaussian sequence model :
$X_i \sim N(\mu_i, 1)$, most $\mu_i = 0$, some $\mu_i = \mu > 0$.

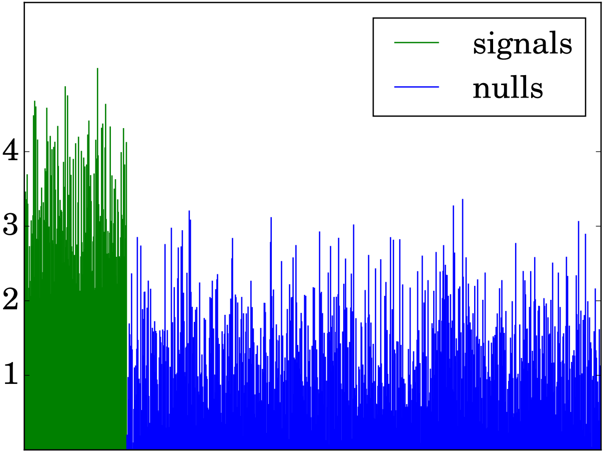
- **Detection.** Is there at least one signal?
- **Localization.** Which observations correspond to signals?
- **Estimation.** What are the values of these signals?

Natural notions of error

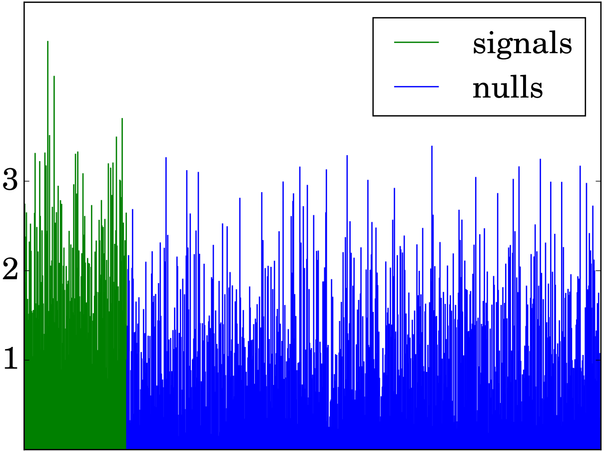
Maxim Rabinovich, **Aaditya Ramdas**, Michael I. Jordan, Martin J. Wainwright

- **Detection.**
  - Type-I error (level) for global null
  - Power against sparse alternatives (eg: higher criticism)
- **Localization.**
  - FDR: proportion of wrong rejections **(this work)**
  - False non-discovery rate (FNR): proportion of missed signals **(this work)**
  - Another option for future work — FWER.

What kind of model?

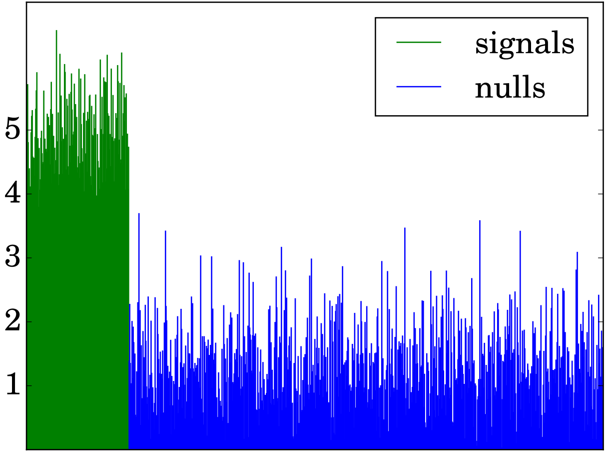


What kind of model?



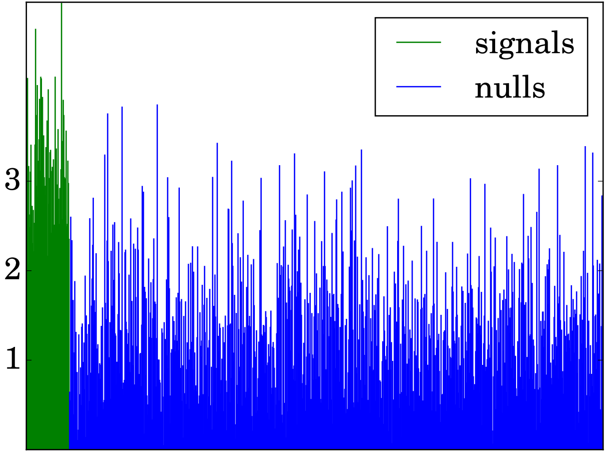
Signal can be weaker

What kind of model?



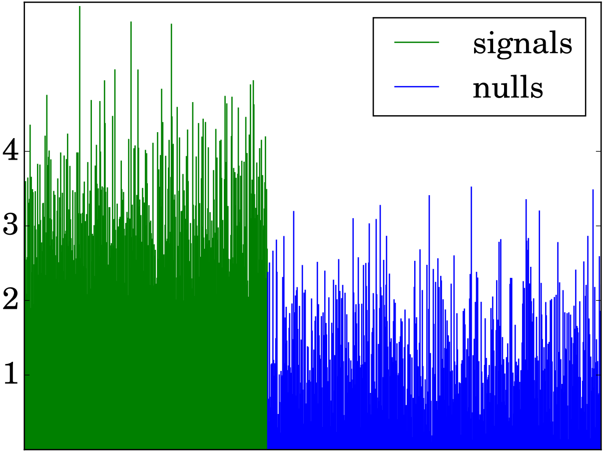
...or stronger

What kind of model?



Signals can be more sparse

What kind of model?



...or less sparse

Sparse (generalized) Gaussians model

- **Sparsity.** $\#\text{signals} = n^{1-\beta} \ll n, \quad 0 < \beta < 1.$
- **Distributions.** Tails $\sim \exp\left(-\frac{|x|^\gamma}{\gamma}\right), \quad \gamma \geq 1$

$$\gamma = 2 \implies \text{Gaussian-like tails } \textbf{(this talk)}$$

- **Signal strength.** Signals shifted positively by

$$\begin{aligned} \mu_n &= (\gamma r \log n)^{1/\gamma} \\ &= \sqrt{2r \log n} \quad \textbf{(this talk)} \end{aligned}$$

Sparse (generalized) Gaussians model

- **Sparsity.** $\#\text{signals} = n^{1-\beta} \ll n, \quad 0 < \beta < 1.$
- **Distributions.** Tails $\sim \exp\left(-\frac{|x|^\gamma}{\gamma}\right), \quad \gamma \geq 1$

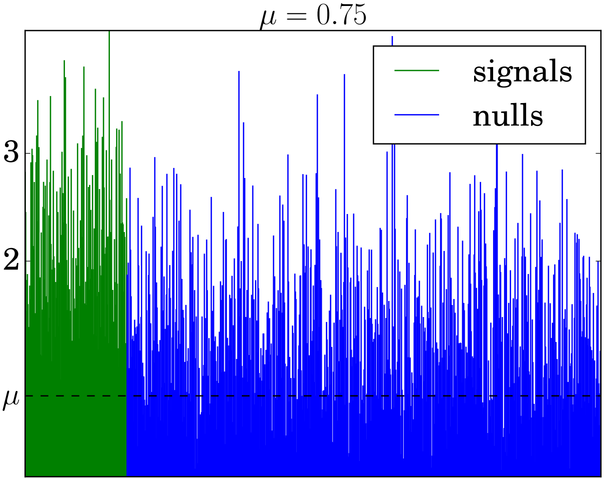
$$\gamma = 2 \Longrightarrow \text{Gaussian-like tails } \textbf{(this talk)}$$

- **Signal strength.** Signals shifted positively by

$$\begin{aligned}\mu_n &= (\gamma r \log n)^{1/\gamma} \\ &= \sqrt{2r \log n} \quad \textbf{(this talk)}\end{aligned}$$

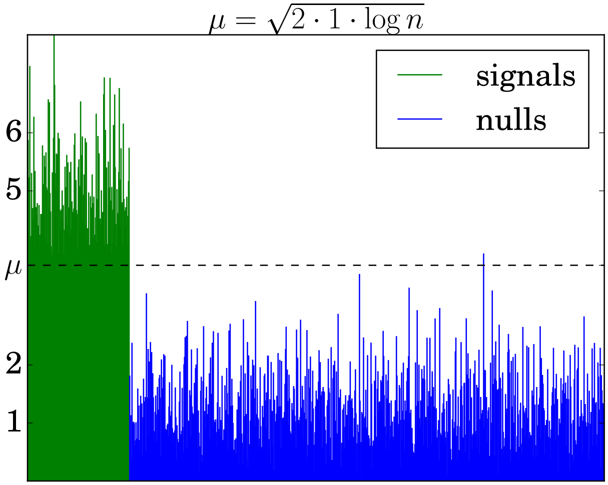
Rate depends on four parameters : n (#signals), β (sparsity level), r (signal strength), γ (tail decay)

Why this parameterization?



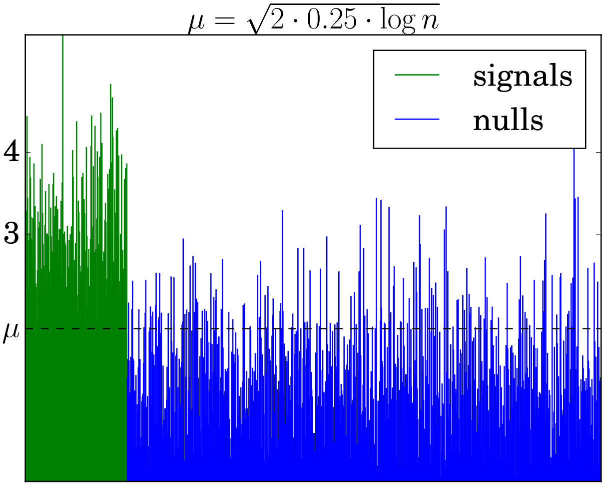
Constant μ — undetectable

Why this parameterization?



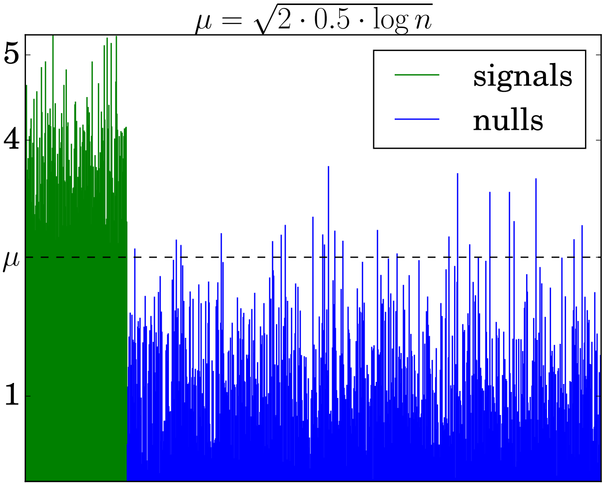
$\mu \sim \sqrt{2 \cdot 1 \cdot \log n}$ — easy

Why this parameterization?



$\mu \sim \sqrt{2 \cdot r \cdot \log n}$, $0 < r < 1$ — “right” parameterization

Why this parameterization?



$\mu \sim \sqrt{2 \cdot r \cdot \log n}, \;\; 0 < r < 1$ — "right" parameterization

Known results (Detection)

- Feasible region:

$$r > \begin{cases} \beta - \frac{1}{2} & \text{if } \frac{1}{2} < \beta \leq \frac{3}{4}, \\ \left(1 - \sqrt{1-\beta}\right)^2 & \text{if } \frac{3}{4} < \beta < 1. \end{cases}$$

- Higher-criticism (HC) thresholding asymptotically consistent (**?**) for *detection*.

Known results (localization)

- Combined FDR + FNR : Castro and Chen (**?**)
  - $r > \beta$, known algorithms asymptotically consistent
  - $r < \beta$, no algorithm can be consistent
- Complementary work: asymptotic Bayes optimality under sparsity (ABOS) (**??**)
  - Weighted probability of false positive + false negative by Bogdan-Chakrabarti-Fromelet-Ghosh
  - Bayes classification risk by Neuvial-Roquain
- Complementary work: asymptotic minimax optimality of BH-derived thresholding (**?**)
  - In the estimation context, for denoising an approximately-sparse high-dimensional vector, by Abramovich-Benjamini-Donoho-Johnstone.

The open problems

- FDR-FNR tradeoff (non-asymptotic, finite sample)
- Minimax rates for FDR + FNR
- Finite-sample optimality of known procedures

We resolve all three.

FDR-FNR tradeoff

Maxim Rabinovich, **Aaditya Ramdas**, Michael I. Jordan, Martin J. Wainwright

Theorem

The following trade-off holds must necessarily hold for all threshold-based multiple testing procedures:

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

γ-"distance": $D_\gamma\left(a,\ b\right) = \left|b^{1/\gamma} - a^{1/\gamma}\right|^\gamma = \left(\sqrt{b} - \sqrt{a}\right)^2$

Rate depends on four parameters : n (#signals), β (sparsity level), r (signal strength), γ (tail decay)

Optimal FDR + FNR scaling

Corollary

Any threshold-based multiple testing procedure must satisfy:

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

where κ_* *is the unique solution to the fixed point equation*

$$\kappa = D_\gamma\big(\beta + \kappa,\ r\big).$$

Rate depends on four parameters : n (#signals), β (sparsity level), r (signal strength), γ (tail decay)

Explicitly — $\gamma = 2 \Longrightarrow \kappa_* = \frac{1}{r}\left(\frac{r-\beta}{2}\right)^2$

Known procedures

- Reject based on threshold:

$$t_n(X_1, \ldots, X_n) = \min\left\{t \in \{X_1, \ldots, X_n\} : \widehat{\mathrm{FDP}}(t) \leq q_n\right\}$$

- Two popular choices:

$$\widehat{\mathrm{FDP}}^{\mathrm{BH}}(t) = \frac{\mathbb{P}_0(X \geq t)}{\#(X_i \geq t)/n}$$

$$\widehat{\mathrm{FDP}}^{\mathrm{BC}}(t) = \frac{\left[\#(X_i \leq -t) + 1\right]/n}{\#(X_i \geq t)/n}$$

- Barber-Candès (BC) works if $\mathbb{P}_0$ unknown but symmetric (**??**)

Optimality guarantee

Maxim Rabinovich, **Aaditya Ramdas**, Michael I. Jordan, Martin J. Wainwright

Theorem

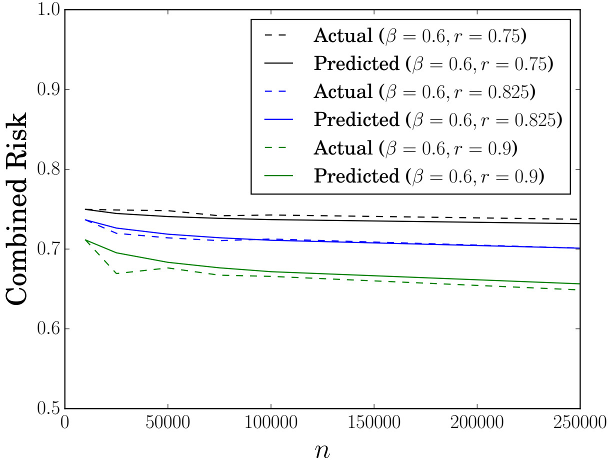
BH and BC with target $\mathrm{FDR} \propto n^{-\kappa}$ *satisfy:*

BH: $\mathrm{FDR} \lesssim n^{-\kappa}$ *and* $\mathrm{FNR} \lesssim n^{-D_\gamma\left(\beta+\kappa,\ r\right)}$

BC: $\mathrm{FDR} \lesssim n^{-\kappa}$ *and* $\mathrm{FNR} \lesssim \max\left\{n^{-\kappa},\ n^{-D_\gamma\left(\beta+\kappa,\ r\right)}\right\}.$

BH always achieves the optimal tradeoff.
BC achieves it if $\mathrm{FDR} \leq \mathrm{FNR}$.

Bound vs. reality for BH



Proof overview

For lower bounds:

- **Challenge:** Need uniform control over all thresholds
- **Idea:** Reduce to data-oblivious procedures with deterministic thresholds
  - Construct optimal data-oblivious threshold
  - Show it's about as good as the best data-aware threshold

For optimality guarantees:

- **Challenge:** BH/BC thresholds are complicated functions of inputs
- **Idea:** bound complicated thresholds by simple data-oblivious thresholds

Summary

Maxim Rabinovich, **Aaditya Ramdas**, Michael I. Jordan, Martin J. Wainwright

We prove non-asymptotic lower+upper bounds for FDR+FNR.

Summary

We prove non-asymptotic lower+upper bounds for FDR+FNR.

We allow the problem parameters to vary with n.

Summary

We prove non-asymptotic lower+upper bounds for FDR+FNR.

We allow the problem parameters to vary with n.

We find that BH and BC procedures are actually finite-sample minimax optimal.

Summary

We prove non-asymptotic lower+upper bounds for FDR+FNR.

We allow the problem parameters to vary with n.

We find that BH and BC procedures are actually finite-sample minimax optimal.

Optimal rates and tradeoffs in multiple testing (on Arxiv)
— M. Rabinovich, A. Ramdas, M. Jordan, M. Wainwright