Sparse multiple testing: can one estimate the null distribution?

Etienne Roquain[1]
Joint work with A. Carpentier[2], S. Delattre[3], N. Verzelen[4],

[1]LPSM, Sorbonne Université, France
[2]Otto-von-Guericke-Universität Magdeburg, Allemagne
[3]LPSM, Université de Paris, France
[4]INRAE, Montpellier, France

MMMS2 Luminy, 02/06/2020

Arxiv 1912.03109. "On using empirical null distributions in Benjamini-Hochberg procedure"

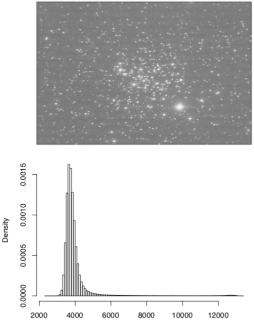
To appear in AoS. "Estimating minimum effect with outlier selection "

ANR “Sanssouci", ANR “BASICS", GDR ISIS “TASTY"

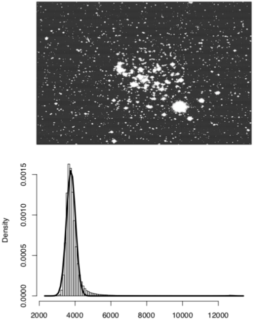
Motivation 1: null distribution unknown

M67 photography, Package `photutils`

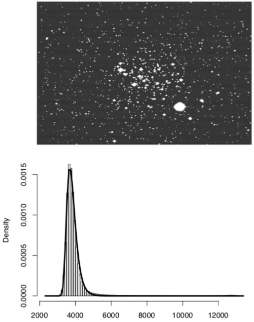
Original



Gaussian fitting



Gumbel fitting

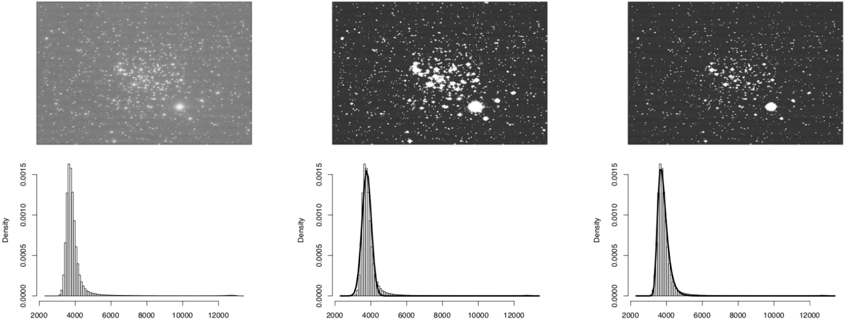


- Naive null distribution fitting
- Impact on the risk?

Motivation 1: null distribution unknown

M67 photography, Package `photutils`

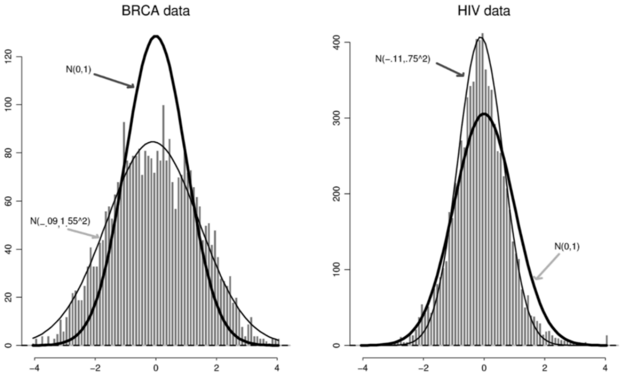
Original | Gaussian fitting | Gumbel fitting



- Naive null distribution fitting
- Impact on the risk?

Motivation 2: null distribution wrong

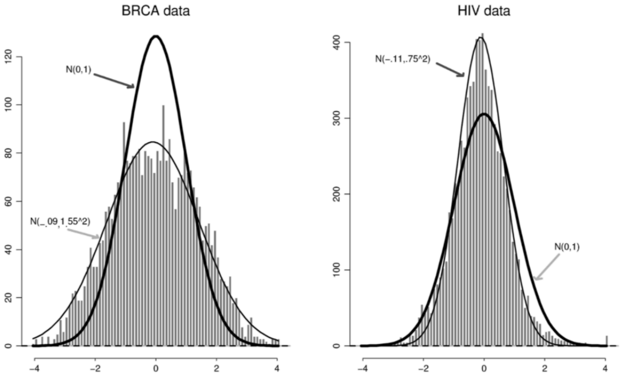
Figure 4 in [Efron (2008)]



- Empirical null [Efron (2004,2007,2008,2009)]
- Impact on the risk?

Motivation 2: null distribution wrong

Figure 4 in [Efron (2008)]



- Empirical null [Efron (2004,2007,2008,2009)]
- Impact on the risk?

Existing work (selection)

Estimation of the null:

- Series of work [Efron (2004,2007,2008,2009)]
- Minimax rate with Fourier analysis: [Jin and Cai (2007)]; [Cai and Jin (2010)]
- Two group mixture model: [Efron et al. (2001)]; [Sun and Cai (2009)]; [Cai and Sun (2009)]; [Padilla and Bickel (2012)]; [Nguyen and Matias (2014)]; [Heller and Yekutieli (2014)]; [Zablocki et al. (2017)]; [Amar et al. (2017)]; [Cai et al. (2019)]; [Rebafka et al. (2019)]
- Estimation in factor model: [Efron (2007a)]; [Leek and Storey (2008)]; [Friguet et al. (2009)]; [Fan et al. (2012)]; [Fan and Han (2017)]

Impact on the risk:

- FDR control in symmetric, centered, one-sided case: [Barber and Candès (2015)]; [Arias-Castro and Chen (2017)]

Lower bounds in multiple testing:

- [Arias-Castro and Chen (2017)]; [Rabinovich et al. (2017)]; [Castillo and R. (2020).]

Existing work (selection)

Estimation of the null:

- Series of work [Efron (2004,2007,2008,2009)]
- Minimax rate with Fourier analysis: [Jin and Cai (2007)]; [Cai and Jin (2010)]
- Two group mixture model: [Efron et al. (2001)]; [Sun and Cai (2009)]; [Cai and Sun (2009)]; [Padilla and Bickel (2012)]; [Nguyen and Matias (2014)]; [Heller and Yekutieli (2014)]; [Zablocki et al. (2017)]; [Amar et al. (2017)]; [Cai et al. (2019)]; [Rebafka et al. (2019)]
- Estimation in factor model: [Efron (2007a)]; [Leek and Storey (2008)]; [Friguet et al. (2009)]; [Fan et al. (2012)]; [Fan and Han (2017)]

Impact on the risk:

- FDR control in symmetric, centered, one-sided case: [Barber and Candès (2015)]; [Arias-Castro and Chen (2017)]

Lower bounds in multiple testing:

- [Arias-Castro and Chen (2017)]; [Rabinovich et al. (2017)]; [Castillo and R. (2020).]

Existing work (selection)

Estimation of the null:

- Series of work [Efron (2004,2007,2008,2009)]
- Minimax rate with Fourier analysis: [Jin and Cai (2007)]; [Cai and Jin (2010)]
- Two group mixture model: [Efron et al. (2001)]; [Sun and Cai (2009)]; [Cai and Sun (2009)]; [Padilla and Bickel (2012)]; [Nguyen and Matias (2014)]; [Heller and Yekutieli (2014)]; [Zablocki et al. (2017)]; [Amar et al. (2017)]; [Cai et al. (2019)]; [Rebafka et al. (2019)]
- Estimation in factor model: [Efron (2007a)]; [Leek and Storey (2008)]; [Friguet et al. (2009)]; [Fan et al. (2012)]; [Fan and Han (2017)]

Impact on the risk:

- FDR control in symmetric, centered, one-sided case: [Barber and Candès (2015)]; [Arias-Castro and Chen (2017)]

Lower bounds in multiple testing:

- [Arias-Castro and Chen (2017)]; [Rabinovich et al. (2017)]; [Castillo and R. (2020).]

Setting

Observations

$$Y = (Y_i)_{1\le i\le n} \text{ indep },\quad Y_i \sim P_i,\quad \text{parameter } P = (P_i)_{1\le i\le n} \in \mathcal{P}$$

Gaussian null assumption:

Most of the P_i's equal $\mathcal{N}(\theta, \sigma^2)$, for some unknown θ, σ

Example:

$$P = (P_1, \mathcal{N}(\theta,\sigma^2), P_3, \mathcal{N}(\theta,\sigma^2), \mathcal{N}(\theta,\sigma^2), \mathcal{N}(\theta,\sigma^2), P_7, \mathcal{N}(\theta,\sigma^2))$$

- Ensures $\theta = \theta(P)$ and $\sigma = \sigma(P)$ uniquely defined
- Test $H_{0,i}$: "$P_i = \mathcal{N}(\theta(P), \sigma^2(P))$" against $H_{1,i}$: "$P_i \neq \mathcal{N}(\theta(P), \sigma^2(P))$"

"item i comes from the background" "item i comes from signal"

Setting

Observations

$$Y = (Y_i)_{1\leq i\leq n} \text{ indep }, \quad Y_i \sim P_i, \quad \text{parameter } P = (P_i)_{1\leq i\leq n} \in \mathcal{P}$$

Gaussian null assumption:

Most of the P_i's equal $\mathcal{N}(\theta, \sigma^2)$, for some unknown θ, σ

Example:

$$P = \big(P_1, \mathcal{N}(\theta,\sigma^2), P_3, \mathcal{N}(\theta,\sigma^2), \mathcal{N}(\theta,\sigma^2), \mathcal{N}(\theta,\sigma^2), P_7, \mathcal{N}(\theta,\sigma^2)\big)$$

- Ensures $\theta = \theta(P)$ and $\sigma = \sigma(P)$ uniquely defined
- Test $H_{0,i}$: "$P_i = \mathcal{N}(\theta(P), \sigma^2(P))$" against $H_{1,i}$: "$P_i \neq \mathcal{N}(\theta(P), \sigma^2(P))$"

"item i comes from the background" "item i comes from signal"

Setting

Observations

$$Y = (Y_i)_{1\leq i\leq n} \text{ indep }, \quad Y_i \sim P_i, \quad \text{parameter } P = (P_i)_{1\leq i\leq n} \in \mathcal{P}$$

Gaussian null assumption:

Most of the P_i's equal $\mathcal{N}(\theta, \sigma^2)$, for some unknown θ, σ

Example:

$$P = \big(P_1, \mathcal{N}(\theta,\sigma^2), P_3, \mathcal{N}(\theta,\sigma^2), \mathcal{N}(\theta,\sigma^2), \mathcal{N}(\theta,\sigma^2), P_7, \mathcal{N}(\theta,\sigma^2)\big)$$

- Ensures $\theta = \theta(P)$ and $\sigma = \sigma(P)$ uniquely defined
- Test $H_{0,i}$: "$P_i = \mathcal{N}(\theta(P), \sigma^2(P))$" against $H_{1,i}$: "$P_i \neq \mathcal{N}(\theta(P), \sigma^2(P))$"

  "item i comes from the background" "item i comes from signal"

Criteria

- True null set $\mathcal{H}_0(P) = \{i \,:\, \text{P satisfies } H_{0,i}\},\ n_0(P) = |\mathcal{H}_0(P)|$
- False null set $\mathcal{H}_1(P) = \mathcal{H}_0(P)^c,\ n_1(P) = |\mathcal{H}_1(P)|$
- for a procedure $R(Y) \subset \{1, \dots, n\}$

$$\text{FDP}(P, R(Y)) = \frac{|R(Y) \cap \mathcal{H}_0(P)|}{|R(Y)| \vee 1} \quad \text{'false discovery proportion'}$$

$$\mathbf{E}_P[\text{FDP}(P, R(Y))] = \text{FDR}(P, R) \quad \text{'false discovery rate'}$$

$$\text{TDP}(P, R(Y)) = \frac{|R(Y) \cap \mathcal{H}_1(P)|}{n_1(P) \vee 1} \quad \text{'true discovery proportion'}$$

$$\mathbf{E}_P[\text{TDP}(P, R(Y))] = \text{TDR}(P, R) \quad \text{'true discovery rate'}$$

- Sparse multiple testing (enough background)

$$n_1(P) \le k_n \text{ with } k_n \text{ 'small'}$$

Criteria

- True null set $\mathcal{H}_0(P) = \{i \,:\, \text{P satisfies } H_{0,i}\},\ n_0(P) = |\mathcal{H}_0(P)|$
- False null set $\mathcal{H}_1(P) = \mathcal{H}_0(P)^c,\ n_1(P) = |\mathcal{H}_1(P)|$
- for a procedure $R(Y) \subset \{1, \dots, n\}$

$$\mathrm{FDP}(P, R(Y)) = \frac{|R(Y) \cap \mathcal{H}_0(P)|}{|R(Y)| \vee 1} \quad \text{'false discovery proportion'}$$

$$\mathbf{E}_P[\mathrm{FDP}(P, R(Y))] = \mathrm{FDR}(P, R) \quad \text{'false discovery rate'}$$

$$\mathrm{TDP}(P, R(Y)) = \frac{|R(Y) \cap \mathcal{H}_1(P)|}{n_1(P) \vee 1} \quad \text{'true discovery proportion'}$$

$$\mathbf{E}_P[\mathrm{TDP}(P, R(Y))] = \mathrm{TDR}(P, R) \quad \text{'true discovery rate'}$$

- Sparse multiple testing (enough background)

$$n_1(P) \le k_n \text{ with } k_n \text{ 'small'}$$

Criteria

- True null set $\mathcal{H}_0(P) = \{i \,:\, \text{P satisfies } H_{0,i}\},\; n_0(P) = |\mathcal{H}_0(P)|$
- False null set $\mathcal{H}_1(P) = \mathcal{H}_0(P)^c,\; n_1(P) = |\mathcal{H}_1(P)|$
- for a procedure $R(Y) \subset \{1, \dots, n\}$

$$\text{FDP}(P, R(Y)) = \frac{|R(Y) \cap \mathcal{H}_0(P)|}{|R(Y)| \vee 1} \quad \text{'false discovery proportion'}$$

$$\mathbf{E}_P[\text{FDP}(P, R(Y))] = \text{FDR}(P, R) \quad \text{'false discovery rate'}$$

$$\text{TDP}(P, R(Y)) = \frac{|R(Y) \cap \mathcal{H}_1(P)|}{n_1(P) \vee 1} \quad \text{'true discovery proportion'}$$

$$\mathbf{E}_P[\text{TDP}(P, R(Y))] = \text{TDR}(P, R) \quad \text{'true discovery rate'}$$

- Sparse multiple testing (enough background)

$$n_1(P) \le k_n \text{ with } k_n \text{ 'small'}$$

Criteria

- True null set $\mathcal{H}_0(P) = \{i \,:\, \text{P satisfies } H_{0,i}\},\ n_0(P) = |\mathcal{H}_0(P)|$
- False null set $\mathcal{H}_1(P) = \mathcal{H}_0(P)^c,\ n_1(P) = |\mathcal{H}_1(P)|$
- for a procedure $R(Y) \subset \{1, \dots, n\}$

$$\mathrm{FDP}(P, R(Y)) = \frac{|R(Y) \cap \mathcal{H}_0(P)|}{|R(Y)| \vee 1} \quad \text{'false discovery proportion'}$$

$$\mathbf{E}_P[\mathrm{FDP}(P, R(Y))] = \mathrm{FDR}(P, R) \quad \text{'false discovery rate'}$$

$$\mathrm{TDP}(P, R(Y)) = \frac{|R(Y) \cap \mathcal{H}_1(P)|}{n_1(P) \vee 1} \quad \text{'true discovery proportion'}$$

$$\mathbf{E}_P[\mathrm{TDP}(P, R(Y))] = \mathrm{TDR}(P, R) \quad \text{'true discovery rate'}$$

- Sparse multiple testing (enough background)

$$n_1(P) \le k_n \text{ with } k_n \text{ 'small'}$$

Oracle procedure BH^*_α

- Rescaled observation $Z_i = (Y_i - \theta(P))/\sigma(P)$
- Apply the standard BH procedure to the Z_i's:
  - Sorting $|Z|_{(1)} \geq |Z|_{(2)} \geq \cdots \geq |Z|_{(n)}$
  - Quantiles
    $$t_k = \overline{\Phi}^{-1}(\alpha k/(2n))$$
  - Rejection number
    $$\widehat{k} = \max\{k \ : \ |Z|_{(k)} \geq t_k\}$$
  - Select the Z_i's corresponding to $|Z|_{(1)}, |Z|_{(2)}, \dots, |Z|_{(\widehat{k})}$.

Theorem [Benjamini and Hochberg (1995), Benjamini and Yekutieli (2001)]

$$\forall P \in \mathcal{P}, \quad \mathrm{FDR}(P, BH^*_\alpha) = \alpha n_0(P)/n \qquad \simeq \alpha \text{ under sparsity}$$

Oracle procedure BH^*_α

- Rescaled observation $Z_i = (Y_i - \theta(P))/\sigma(P)$
- Apply the standard BH procedure to the Z_i's:
  - Sorting $|Z|_{(1)} \geq |Z|_{(2)} \geq \cdots \geq |Z|_{(n)}$
  - Quantiles
  $$t_k = \overline{\Phi}^{-1}(\alpha k/(2n))$$
  - Rejection number
  $$\widehat{k} = \max\{k \ : \ |Z|_{(k)} \geq t_k\}$$
  - Select the Z_i's corresponding to $|Z|_{(1)}, |Z|_{(2)}, \ldots, |Z|_{(\widehat{k})}$.

Theorem [Benjamini and Hochberg (1995), Benjamini and Yekutieli (2001)]

$$\forall P \in \mathcal{P}, \quad \mathrm{FDR}(P, BH^*_\alpha) = \alpha n_0(P)/n \qquad \simeq \alpha \text{ under sparsity}$$

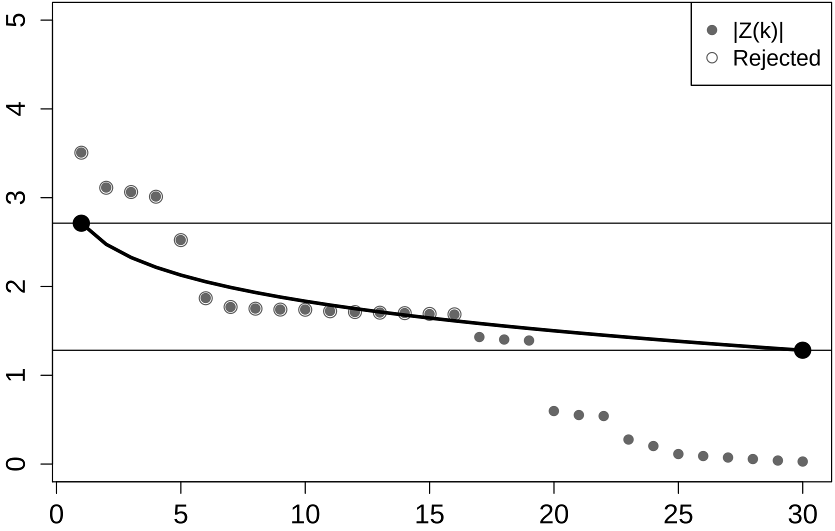
Oracle procedure BH_α^*

- Rescaled observation $Z_i = (Y_i - \theta(P))/\sigma(P)$
- Apply the standard BH procedure to the Z_i's:
  - Sorting $|Z|_{(1)} \geq |Z|_{(2)} \geq \cdots \geq |Z|_{(n)}$
  - Quantiles
$$t_k = \overline{\Phi}^{-1}(\alpha k/(2n))$$
  - Rejection number
$$\widehat{k} = \max\{k \;:\; |Z|_{(k)} \geq t_k\}$$
  - Select the Z_i's corresponding to $|Z|_{(1)}, |Z|_{(2)}, \ldots, |Z|_{(\widehat{k})}$.

Theorem [Benjamini and Hochberg (1995), Benjamini and Yekutieli (2001)]

$$\forall P \in \mathcal{P}, \quad \text{FDR}(P, BH_\alpha^*) = \alpha n_0(P)/n \qquad \simeq \alpha \text{ under sparsity}$$

Oracle procedure BH^*_α



Optimality under a sparsity range

Procedure R optimal: $R \approx \mathrm{BH}^*_\alpha$ **both** for FDP and TDP

Definition

Procedure R optimal for a sparsity k_n: there exists $\eta_n \to 0$, s.t.

$$\textbf{(I)} \quad \limsup_n \sup_{\alpha \in (1/n, 1/2)} \left\{ \sup_{\substack{P \in \mathcal{P} \\ n_1(P) \le k_n}} \{\mathrm{FDR}(P, R)\} - \alpha \right\} \le 0$$

$$\textbf{(II)} \quad \lim_n \sup_{\alpha \in (1/n, 1/2)} \left\{ \sup_{\substack{P \in \mathcal{P} \\ n_1(P) \le k_n}} \left\{ \mathbf{P}_{Y \sim P} \left(\mathrm{TDP}(P, R) < \mathrm{TDP}(P, \mathrm{BH}^\star_{\alpha(1-\eta_n)}) \right) \right\} \right\} = 0$$

- Robust criteria: alternatives arbitrary

Optimality under a sparsity range

Procedure R optimal: $R \approx \mathrm{BH}^*_\alpha$ **both** for FDP and TDP

Definition

Procedure R optimal for a sparsity k_n: there exists $\eta_n \to 0$, s.t.

$$\textbf{(I)}\quad \limsup_n \sup_{\alpha \in (1/n, 1/2)} \left\{ \sup_{\substack{P \in \mathcal{P} \\ n_1(P) \leq k_n}} \{\mathrm{FDR}(P,R)\} - \alpha \right\} \leq 0$$

$$\textbf{(II)}\quad \lim_n \sup_{\alpha \in (1/n, 1/2)} \left\{ \sup_{\substack{P \in \mathcal{P} \\ n_1(P) \leq k_n}} \left\{ \mathbf{P}_{Y \sim P}\left(\mathrm{TDP}(P,R) < \mathrm{TDP}(P, \mathrm{BH}^\star_{\alpha(1-\eta_n)}) \right) \right\} \right\} = 0$$

- Robust criteria: alternatives arbitrary

Optimality under a sparsity range

Procedure R optimal: $R \approx \mathrm{BH}^*_\alpha$ **both** for FDP and TDP

Definition

Procedure R optimal for a sparsity k_n: there exists $\eta_n \to 0$, s.t.

$$\textbf{(I)}\quad \limsup_n \sup_{\alpha\in(1/n,1/2)} \left\{ \sup_{\substack{P\in\mathcal{P}\\ n_1(P)\leq k_n}} \{\mathrm{FDR}(P,R)\} - \alpha \right\} \leq 0$$

$$\textbf{(II)}\quad \lim_n \sup_{\alpha\in(1/n,1/2)} \left\{ \sup_{\substack{P\in\mathcal{P}\\ n_1(P)\leq k_n}} \left\{ \mathbf{P}_{Y\sim P}\left(\mathrm{TDP}(P,R) < \mathrm{TDP}(P, \mathrm{BH}^\star_{\alpha(1-\eta_n)})\right)\right\} \right\} = 0$$

- Robust criteria: alternatives arbitrary

Plugged BH procedure

- Estimation: robust minimax estimator of $\theta(P)$, $\sigma(P)$:

$$\widetilde{\theta} = Y_{(\lceil n/2 \rceil)};\ \ \widetilde{\sigma} = U_{(\lceil n/2 \rceil)} / \overline{\Phi}^{-1}(1/4),\ \ U_i = |Y_i - Y_{(\lceil n/2 \rceil)}|$$

of L^1 max risk $\asymp (k_n/n) \vee n^{-1/2}$ for sparsity k_n [Huber, 1964], [Chen et al. (2018)]

- Plugged BH procedure $BH(\widetilde{\theta}, \widetilde{\sigma})$
  - Rescaled observation $Z_i' = (Y_i - \widetilde{\theta})/\widetilde{\sigma}$
  - Apply the standard BH procedure to the Z_i''s

Plugged BH procedure

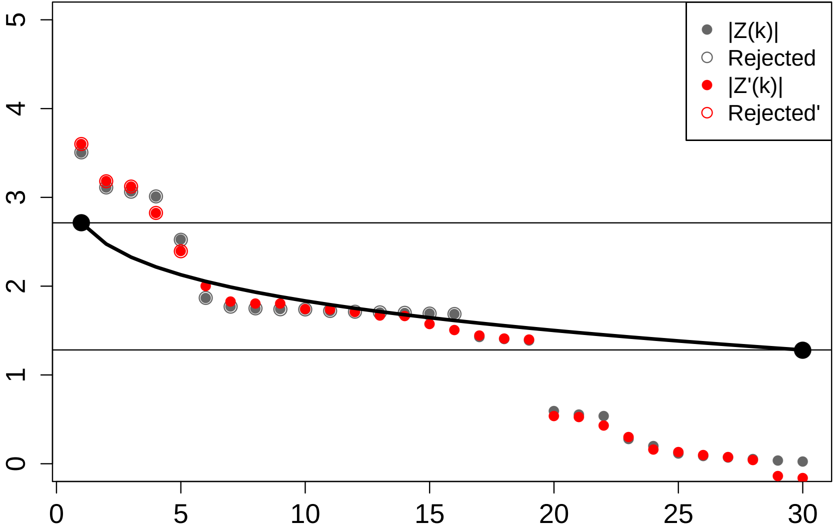
- Estimation: robust minimax estimator of $\theta(P)$, $\sigma(P)$:

$$\widetilde{\theta} = Y_{(\lceil n/2 \rceil)}; \;\; \widetilde{\sigma} = U_{(\lceil n/2 \rceil)} / \overline{\Phi}^{-1}(1/4), \;\; U_i = |Y_i - Y_{(\lceil n/2 \rceil)}|$$

of L^1 max risk $\asymp (k_n/n) \vee n^{-1/2}$ for sparsity k_n [Huber, 1964], [Chen et al. (2018)]

- Plugged BH procedure $BH(\widetilde{\theta}, \widetilde{\sigma})$
  - Rescaled observation $Z_i' = (Y_i - \widetilde{\theta})/\widetilde{\sigma}$
  - Apply the standard BH procedure to the Z_i''s

Plugged BH procedure



Upper bound

Heuristic: $\mathrm{BH}_\alpha(\widetilde{\theta},\widetilde{\sigma}) \approx \mathrm{BH}^*_\alpha$ if

$$|\widetilde{\theta}-\theta(P)| \ll \min_k \left\{\overline{\Phi}^{-1}(\alpha k/n) - \overline{\Phi}^{-1}(\alpha(k+1)/n)\right\} \approx 1/\sqrt{\log n}$$

$$|\widetilde{\sigma}-\sigma(P)| \ll \min_k \left\{(\overline{\Phi}^{-1}(\alpha k/n) - \overline{\Phi}^{-1}(\alpha(k+1)/n))/\overline{\Phi}^{-1}(\alpha k/n)\right\} \approx 1/\log n$$

Suggest $\mathrm{BH}_\alpha(\widetilde{\theta},\widetilde{\sigma}) \approx \mathrm{BH}^*_\alpha$ for $k_n/n \ll 1/\log(n)$.

Proposition 1 [R. and Verzelen (2020)]

$\mathrm{BH}_\alpha(\widetilde{\theta},\widetilde{\sigma})$ is optimal for any sparsity sequence $k_n \ll n/\log(n)$.

Proof: rescaling of p-value process; combining BH procedure and $(\widetilde{\theta},\widetilde{\sigma})$ leave-one-out properties

$$\{p_i(\widetilde{\theta},\widetilde{\sigma}) \leq T_\alpha(Y;\widetilde{\theta},\widetilde{\sigma})\} \subset \{p_i(\widetilde{\theta}^{(i)},\widetilde{\sigma}^{(i)}) \leq T_\alpha(Y^{(i)};\widetilde{\theta}^{(i)},\widetilde{\sigma}^{(i)})\}.$$

Upper bound

Heuristic: $\mathrm{BH}_\alpha(\widetilde{\theta},\widetilde{\sigma}) \approx \mathrm{BH}^*_\alpha$ if

$$|\widetilde{\theta}-\theta(P)| \ll \min_k \left\{\overline{\Phi}^{-1}(\alpha k/n) - \overline{\Phi}^{-1}(\alpha(k+1)/n)\right\} \approx 1/\sqrt{\log n}$$
$$|\widetilde{\sigma}-\sigma(P)| \ll \min_k \left\{(\overline{\Phi}^{-1}(\alpha k/n) - \overline{\Phi}^{-1}(\alpha(k+1)/n))/\overline{\Phi}^{-1}(\alpha k/n)\right\} \approx 1/\log n$$

Suggest $\mathrm{BH}_\alpha(\widetilde{\theta},\widetilde{\sigma}) \approx \mathrm{BH}^*_\alpha$ for $k_n/n \ll 1/\log(n)$.

Proposition 1 [R. and Verzelen (2020)]

$\mathrm{BH}_\alpha(\widetilde{\theta},\widetilde{\sigma})$ is optimal for any sparsity sequence $k_n \ll n/\log(n)$.

Proof: rescaling of p-value process; combining BH procedure and $(\widetilde{\theta},\widetilde{\sigma})$ leave-one-out properties

$$\{p_i(\widetilde{\theta},\widetilde{\sigma}) \leq T_\alpha(Y;\widetilde{\theta},\widetilde{\sigma})\} \subset \{p_i(\widetilde{\theta}^{(i)},\widetilde{\sigma}^{(i)}) \leq T_\alpha(Y^{(i)};\widetilde{\theta}^{(i)},\widetilde{\sigma}^{(i)})\}.$$

Upper bound

Heuristic: $\mathrm{BH}_\alpha(\widetilde{\theta},\widetilde{\sigma}) \approx \mathrm{BH}^*_\alpha$ if

$$|\widetilde{\theta}-\theta(P)| \ll \min_k \left\{\overline{\Phi}^{-1}(\alpha k/n) - \overline{\Phi}^{-1}(\alpha(k+1)/n)\right\} \approx 1/\sqrt{\log n}$$
$$|\widetilde{\sigma}-\sigma(P)| \ll \min_k \left\{(\overline{\Phi}^{-1}(\alpha k/n) - \overline{\Phi}^{-1}(\alpha(k+1)/n))/\overline{\Phi}^{-1}(\alpha k/n)\right\} \approx 1/\log n$$

Suggest $\mathrm{BH}_\alpha(\widetilde{\theta},\widetilde{\sigma}) \approx \mathrm{BH}^*_\alpha$ for $k_n/n \ll 1/\log(n)$.

Proposition 1 [R. and Verzelen (2020)]

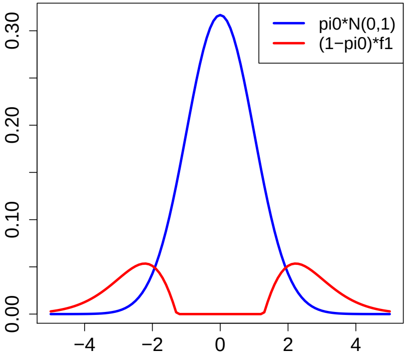
$\mathrm{BH}_\alpha(\widetilde{\theta},\widetilde{\sigma})$ is optimal for any sparsity sequence $k_n \ll n/\log(n)$.

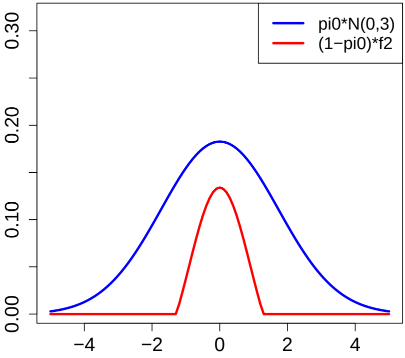
Proof: rescaling of p-value process; combining BH procedure and $(\widetilde{\theta},\widetilde{\sigma})$ leave-one-out properties

$$\{p_i(\widetilde{\theta},\widetilde{\sigma}) \leq T_\alpha(Y;\widetilde{\theta},\widetilde{\sigma})\} \subset \{p_i(\widetilde{\theta}^{(i)},\widetilde{\sigma}^{(i)}) \leq T_\alpha(Y^{(i)};\widetilde{\theta}^{(i)},\widetilde{\sigma}^{(i)})\}.$$

Idea

Procedure BH*

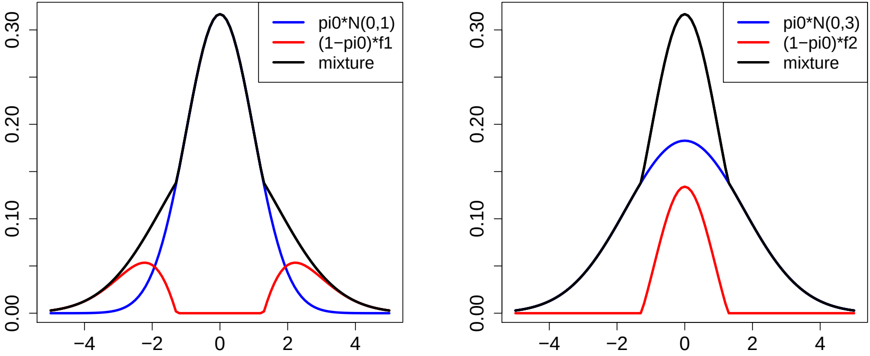




- rejects something
- does not reject anything

Idea

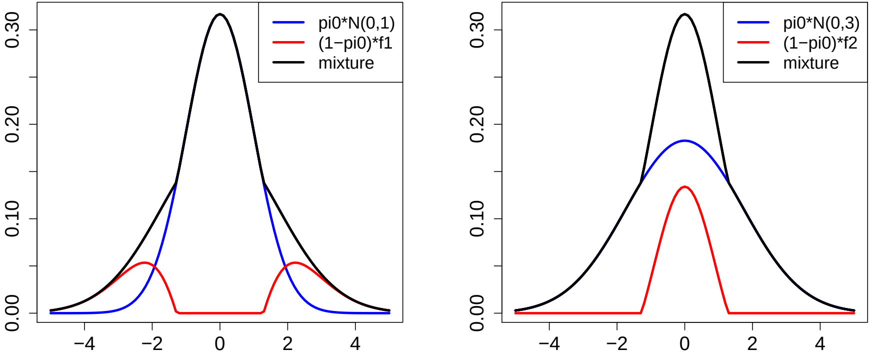
Any procedure $R = R(Y)$



Does not distinguish between the two!

Idea

Any procedure $R = R(Y)$



Does not distinguish between the two!
Not able to mimic BH^*

Lower bound

Proposition 2 [R. and Verzelen (2020)]

For a sparsity $k_n \gg n/\log(n)$, there exists no optimal procedure.

Proof : Le Cam's two-point reduction scheme with the above configuration.

- for all $n \geq c_1$, any $\alpha \in (0,1)$, any k with $c_2 \frac{n \log(2/\alpha)}{\log(n)} \leq k < n/2$
- For any multiple testing procedure R such that

$$\mathrm{FDR}(P,R) \leq c_3 \ , \text{ for any } P \in \mathcal{P} \text{ with } n_1(P) \leq k \ ,$$

- Then there exists some $P \in \mathcal{P}$ with $n_1(P) \leq k$ such that we have

$$|R(Y) \cap \mathcal{H}_1(P)| = 0 \text{ with } P\text{-proba } \geq 2/5$$
$$|\mathrm{BH}^{\star}_{\alpha/2} \cap \mathcal{H}_1(P)| \geq c_4 \alpha^{-1} n^{1/2}/\log^{1/2} n \text{ with } P\text{-proba } \geq 4/5.$$

Lower bound

Proposition 2 [R. and Verzelen (2020)]

For a sparsity $k_n \gg n/\log(n)$, there exists no optimal procedure.

Proof : Le Cam's two-point reduction scheme with the above configuration.

- for all $n \geq c_1$, any $\alpha \in (0,1)$, any k with $c_2 \frac{n \log(2/\alpha)}{\log(n)} \leq k < n/2$
- For any multiple testing procedure R such that

$$\mathrm{FDR}(P,R) \leq c_3 \text{ , for any } P \in \mathcal{P} \text{ with } n_1(P) \leq k \text{ ,}$$

- Then there exists some $P \in \mathcal{P}$ with $n_1(P) \leq k$ such that we have

$$|R(Y) \cap \mathcal{H}_1(P)| = 0 \text{ with } P\text{-proba } \geq 2/5$$
$$|\mathrm{BH}^\star_{\alpha/2} \cap \mathcal{H}_1(P)| \geq c_4 \alpha^{-1} n^{1/2} / \log^{1/2} n \text{ with } P\text{-proba } \geq 4/5.$$

Main result

Theorem 1 [R. and Verzelen (2020)]

(i) for a sparsity $k_n \gg n/\log(n)$, there exists no optimal procedure (of any kind);

(ii) for a sparsity $k_n \ll n/\log(n)$, $\mathrm{BH}_\alpha(\widetilde{\theta}, \widetilde{\sigma})$ is optimal with $(\widetilde{\theta}, \widetilde{\sigma})$ above.

Procedure R optimal for a sparsity k_n: there exists $\eta_n \to 0$, s.t.

$$\textbf{(I)} \quad \limsup_n \sup_{\alpha \in (1/n, 1/2)} \left\{ \sup_{\substack{P \in \mathcal{P} \\ n_1(P) \le k_n}} \{\mathrm{FDR}(P, R)\} - \alpha \right\} \le 0$$

$$\textbf{(II)} \quad \lim_n \sup_{\alpha \in (1/n, 1/2)} \left\{ \sup_{\substack{P \in \mathcal{P} \\ n_1(P) \le k_n}} \left\{ \mathbf{P}_{Y \sim P} \left(\mathrm{TDP}(P, R) < \mathrm{TDP}(P, \mathrm{BH}^\star_{\alpha(1-\eta_n)}) \right) \right\} \right\} = 0$$

No adaptation across boundary

Remark: always possible to achieve **(I)** by rejecting no null

Reformulation Theorem 1:

(i) if $k_n \gg n/\log(n)$, possible to achieve **(I)** but not with **(II)**;
(ii) if $k_n \ll n/\log(n)$, possible to achieve optimality (both **(I)** and **(II)**).

Procedure achieving (i) and (ii)?

NO !

Theorem 2 [R. and Verzelen (2020)]

- Any procedure achieving **(I)** for a sparsity $k_n \gg n/\log(n)$ will fail to achieve optimality for a sparsity $k_n \ll n/\log(n)$.
- Any procedure achieving optimality for a sparsity $k_n \ll n/\log(n)$ will fail to achieve **(I)** for some regime $k_n \gg n/\log(n)$.

For instance, this is the case for $\mathrm{BH}_\alpha(\widetilde{\theta}, \widetilde{\sigma})$.

No adaptation across boundary

Remark: always possible to achieve **(I)** by rejecting no null

Reformulation Theorem 1:

(i) if $k_n \gg n/\log(n)$, possible to achieve **(I)** but not with **(II)**;

(ii) if $k_n \ll n/\log(n)$, possible to achieve optimality (both **(I)** and **(II)**).

Procedure achieving (i) and (ii)?

NO !

Theorem 2 [R. and Verzelen (2020)]

- Any procedure achieving **(I)** for a sparsity $k_n \gg n/\log(n)$ will fail to achieve optimality for a sparsity $k_n \ll n/\log(n)$.
- Any procedure achieving optimality for a sparsity $k_n \ll n/\log(n)$ will fail to achieve **(I)** for some regime $k_n \gg n/\log(n)$.

For instance, this is the case for $\mathrm{BH}_\alpha(\widetilde{\theta}, \widetilde{\sigma})$.

Location model

Case where $\sigma(P)$ is known

- only estimating $\theta(P)$
- the sparsity boundary becomes $n/\log^{1/2}(n)$

Extension to non-Gaussian null $g(\cdot - \theta)$

- g known, **symmetric**, continuous and non-increasing on $\mathbb{R}_+$
- lower-bound and upper-bound matching up to some term
- Subbotin case: $g(x) = L_\zeta^{-1}\, e^{-|x|^\zeta/\zeta}$, $\zeta > 1$
  The sparsity boundary becomes $n/(\log(n))^{1-1/\zeta}$.

Location model

Case where $\sigma(P)$ is known

- only estimating $\theta(P)$
- the sparsity boundary becomes $n/\log^{1/2}(n)$

Extension to non-Gaussian null $g(\cdot - \theta)$

- g known, **symmetric**, continuous and non-increasing on $\mathbb{R}_+$
- lower-bound and upper-bound matching up to some term
- Subbotin case: $g(x) = L_\zeta^{-1}\, e^{-|x|^\zeta/\zeta}$, $\zeta > 1$
  The sparsity boundary becomes $n/(\log(n))^{1-1/\zeta}$.

One-sided setting

One sided assumption:

- the P_i's under the alternative are assumed $\succeq \mathcal{N}(\theta, \sigma^2)$
- easier problem

Proposition [Carpentier, Dellatre, R., Verzelen (2020)]

Estimation of θ:

- Identifiable as soon as $k \leq n - 1$
- Minimax rate $\frac{k/n}{\log^{1/2}(e \vee (k^2/n))}$ for sparsity $1 \leq k \leq 0.9n$
- In particular, minimax rate $\lesssim 1/\log^{1/2}(n)$

Estimation of σ: same with $\log^{1/2}$ replaced by $\log$

Remark: extra $\log$ in the convergence rate, useful for mimicking the oracle!

One-sided setting

One sided assumption:

- the P_i's under the alternative are assumed $\succeq \mathcal{N}(\theta, \sigma^2)$
- easier problem

Proposition [Carpentier, Dellatre, R., Verzelen (2020)]

Estimation of θ:

- Identifiable as soon as $k \leq n - 1$
- Minimax rate $\frac{k/n}{\log^{1/2}(e \vee (k^2/n))}$ for sparsity $1 \leq k \leq 0.9n$
- In particular, minimax rate $\lesssim 1/\log^{1/2}(n)$

Estimation of σ: same with $\log^{1/2}$ replaced by $\log$

Remark: extra $\log$ in the convergence rate, useful for mimicking the oracle!

Upper bound in one-sided case

Plugged $\text{BH}_\alpha(\widetilde{\theta}, \widetilde{\sigma})$ (one-sided version) with new estimators:

$$\begin{cases} \widetilde{\theta} = Y_{(q_n)} + \widetilde{\sigma}\, \overline{\Phi}^{-1}\left(\frac{q_n}{n}\right) \; ; \\ \widetilde{\sigma} = \frac{Y_{(q_n)} - Y_{(q'_n)}}{\overline{\Phi}^{-1}(q'_n/(n-\ell_0)) - \overline{\Phi}^{-1}(q_n/n)} \end{cases},$$

for $\ell_0 \leq \lfloor 0.9n \rfloor$, $q_n = \lfloor n^{3/4} \rfloor$ and $q'_n = \lfloor n^{1/4} \rfloor$.

Theorem [Carpentier, Dellatre, R., Verzelen (2020)]

$\text{BH}_\alpha(\widetilde{\theta}, \widetilde{\sigma})$ is optimal for any sparsity sequence $k_n \leq \lfloor 0.9n \rfloor$ in the following sense:

- FDR control at level α as before.
- Mimics BH^*_α in terms of $\text{TDR} = \mathbf{E}(\text{TDP})$ provided that $\ell_0/n \asymp n_1(P)/n$

- Optimality even without sparsity!

Upper bound in one-sided case

Plugged $\text{BH}_\alpha(\widetilde{\theta}, \widetilde{\sigma})$ (one-sided version) with new estimators:

$$\begin{cases} \widetilde{\theta} = Y_{(q_n)} + \widetilde{\sigma}\, \overline{\Phi}^{-1}\left(\frac{q_n}{n}\right) ; \\ \widetilde{\sigma} = \frac{Y_{(q_n)} - Y_{(q'_n)}}{\overline{\Phi}^{-1}(q'_n/(n-\ell_0)) - \overline{\Phi}^{-1}(q_n/n)} \end{cases},$$

for $\ell_0 \leq \lfloor 0.9n \rfloor$, $q_n = \lfloor n^{3/4} \rfloor$ and $q'_n = \lfloor n^{1/4} \rfloor$.

Theorem [Carpentier, Dellatre, R., Verzelen (2020)]

$\text{BH}_\alpha(\widetilde{\theta}, \widetilde{\sigma})$ is optimal for any sparsity sequence $k_n \leq \lfloor 0.9n \rfloor$ in the following sense:

- FDR control at level α as before.
- Mimics BH^*_α in terms of TDR $= \mathbf{E}(\text{TDP})$ provided that $\ell_0/n \asymp n_1(P)/n$

- Optimality even without sparsity!

Upper bound in one-sided case

Plugged $\text{BH}_\alpha(\widetilde{\theta}, \widetilde{\sigma})$ (one-sided version) with new estimators:

$$\begin{cases} \widetilde{\theta} = Y_{(q_n)} + \widetilde{\sigma}\, \overline{\Phi}^{-1}\left(\frac{q_n}{n}\right) \; ; \\ \widetilde{\sigma} = \frac{Y_{(q_n)} - Y_{(q'_n)}}{\overline{\Phi}^{-1}(q'_n/(n-\ell_0)) - \overline{\Phi}^{-1}(q_n/n)} \end{cases},$$

for $\ell_0 \leq \lfloor 0.9n \rfloor$, $q_n = \lfloor n^{3/4} \rfloor$ and $q'_n = \lfloor n^{1/4} \rfloor$.

Theorem [Carpentier, Dellatre, R., Verzelen (2020)]

$\text{BH}_\alpha(\widetilde{\theta}, \widetilde{\sigma})$ is optimal for any sparsity sequence $k_n \leq \lfloor 0.9n \rfloor$ in the following sense:

- FDR control at level α as before.
- Mimics BH^*_α in terms of $\text{TDR} = \mathbf{E}(\text{TDP})$ provided that $\ell_0/n \asymp n_1(P)/n$

- Optimality even without sparsity!

Outlook

Take home message

- Challenging and useful direction of research
- First results on the feasibility of using empirical null in BH procedure
- Good news: weak sparsity $k_n \ll n/\log(n)$ enough to mimic the oracle
- Bad news: it is needed

Comments

- Robust minimax angle, so quite 'pessimistic'
- One-sided structure on the alternatives makes the problem easier

Future work

- More structured alternatives
- Less structured nulls

Outlook

Take home message

- Challenging and useful direction of research
- First results on the feasibility of using empirical null in BH procedure
- Good news: weak sparsity $k_n \ll n/\log(n)$ enough to mimic the oracle
- Bad news: it is needed

Comments

- Robust minimax angle, so quite 'pessimistic'
- One-sided structure on the alternatives makes the problem easier

Future work

- More structured alternatives
- Less structured nulls

Outlook

Take home message

- Challenging and useful direction of research
- First results on the feasibility of using empirical null in BH procedure
- Good news: weak sparsity $k_n \ll n/\log(n)$ enough to mimic the oracle
- Bad news: it is needed

Comments

- Robust minimax angle, so quite 'pessimistic'
- One-sided structure on the alternatives makes the problem easier

Future work

- More structured alternatives
- Less structured nulls

Some references

- Arias-Castro, E. and Chen, S. (2017). Distribution-free multiple testing. *Electron. J. Stat.*, 11(1):1983–2001.
- Barber, R. F. and Candès, E. J. (2015). Controlling the false discovery rate via knockoffs. *Ann. Statist.*, 43(5):2055–2085.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57(1):289–300.
- Carpentier, A., Delattre, S., Roquain, E., and Verzelen, N. (2018). Estimating minimum effect with outlier selection. *arXiv e-prints*, page arXiv:1809.08330.
- Efron, B. (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *J. Am. Stat. Assoc.*, 99(465):96–104.
- Efron, B. (2008). Microarrays, empirical Bayes and the two-groups model. *Statist. Sci.*, 23(1):1–22.
- Ghosh, D. (2012). Incorporating the empirical null hypothesis into the Benjamini-Hochberg procedure. *Stat. Appl. Genet. Mol. Biol.*, 11(4):Art. 11, front matter+19.
- Huber, P. J. (1964). Robust estimation of a location parameter. *The annals of mathematical statistics*, pages 73–101.

Illustration - Gaussian alternative

$$k = n^{1/2}$$

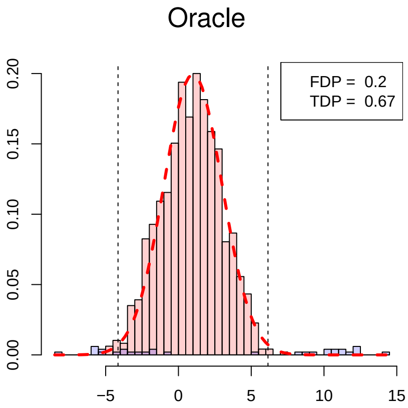


Illustration - Gaussian alternative

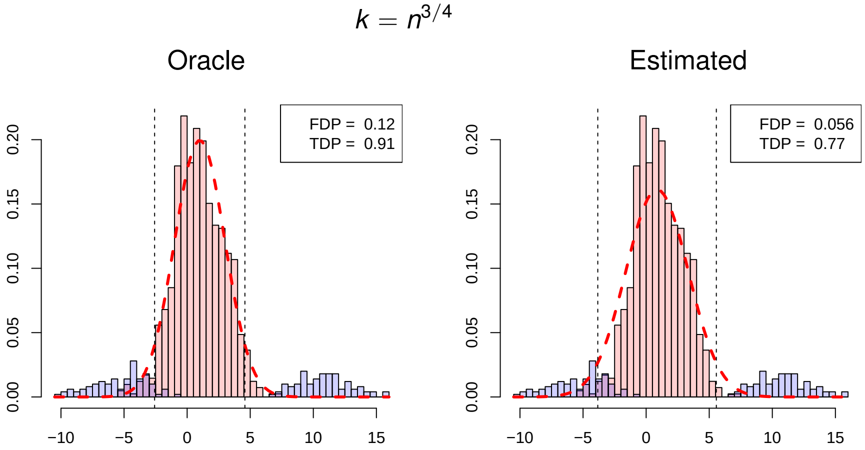


Illustration - Gaussian alternative

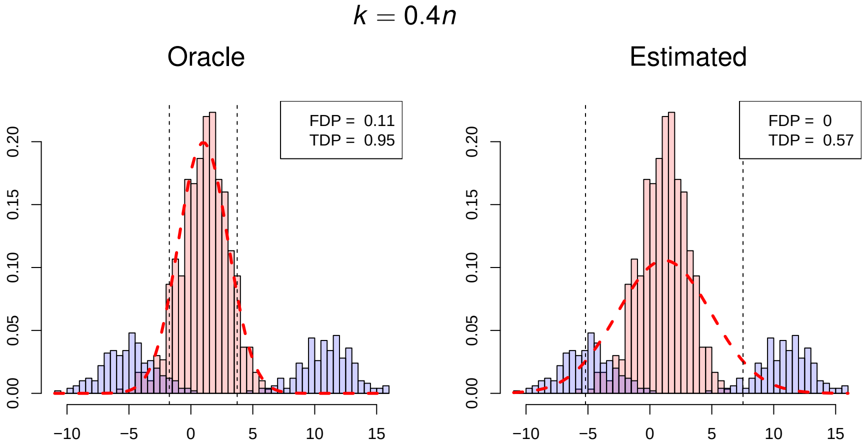


Illustration - f_1 alternative

$$k = n^{1/2}$$

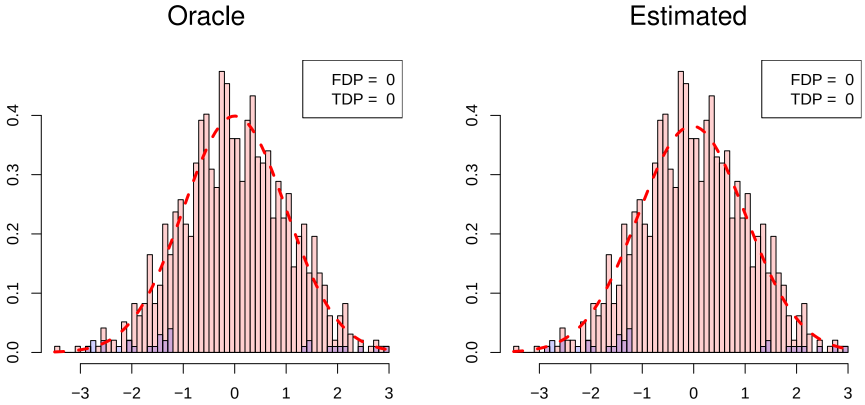


Illustration - f_1 alternative

$$k = n^{3/4}$$

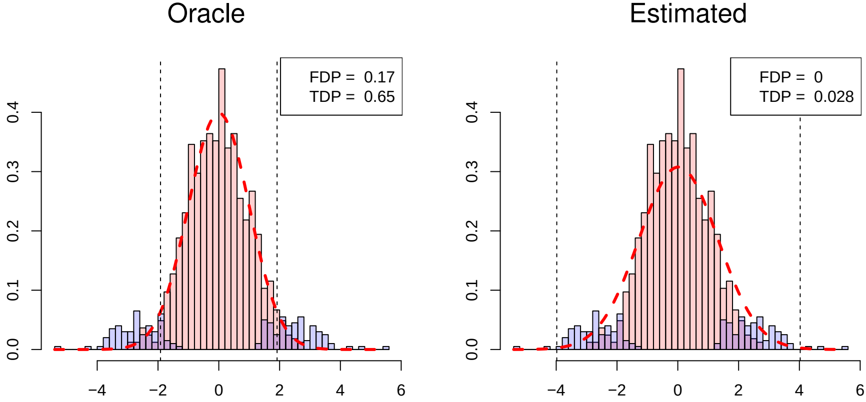


Illustration - f_1 alternative

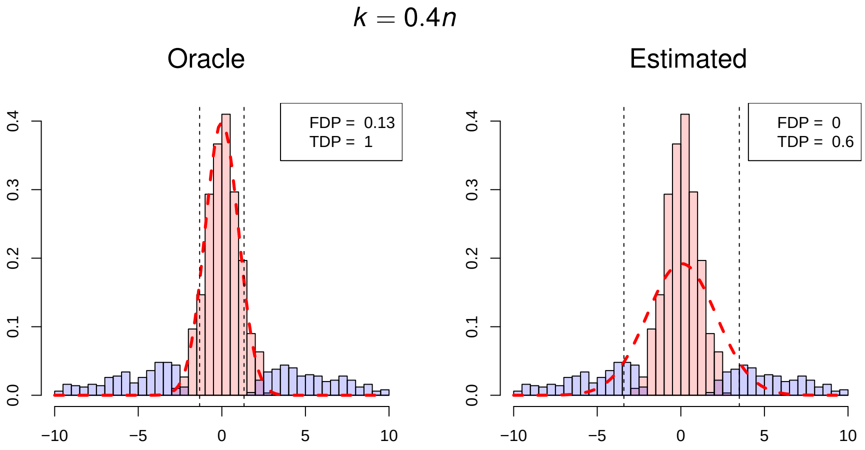


Illustration - f_2 alternative

$$k = n^{1/2}$$

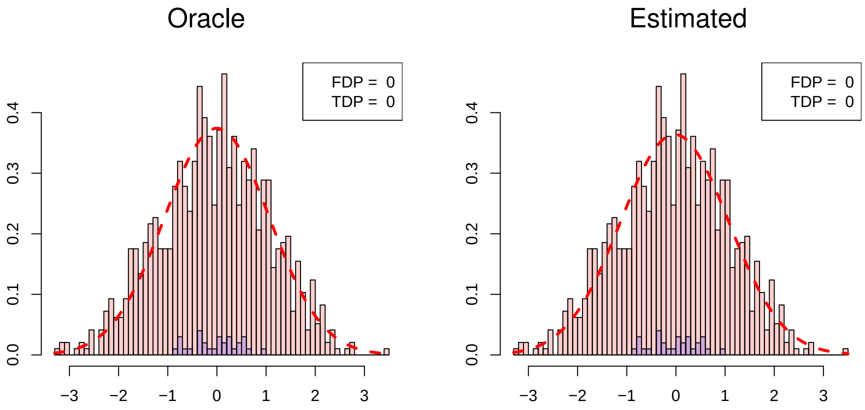


Illustration - f_2 alternative

$$k = n^{3/4}$$

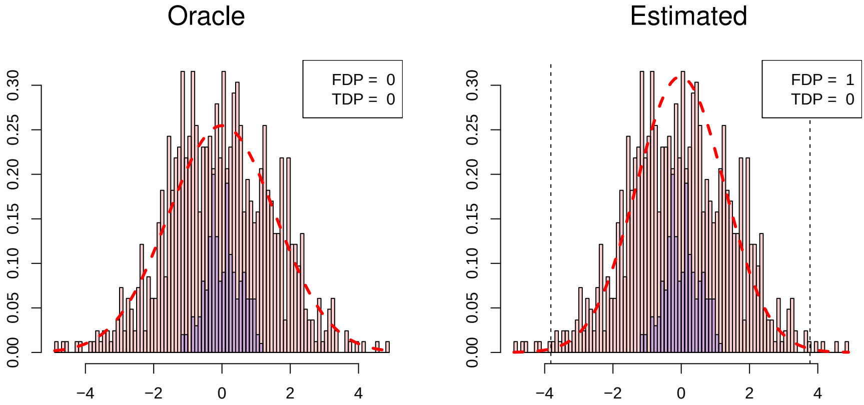


Illustration - f_2 alternative

$k = 0.4n$