Tests multiples et bornes post hoc pour des données hétérogènes

Soutenance de thèse

Guillermo Durand

Sous la direction d'E. Roquain et P. Neuvial
LPSM, Sorbonne Université

26 novembre 2018

Summary of contributions

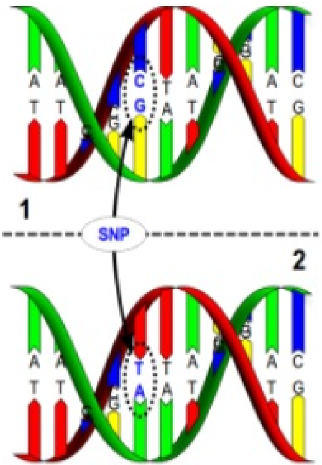
- 2 published papers on previous works [Durand and Lessard (2016)], [Chatelain et al. (2018)]
- 1 paper in revision about optimal weighting arXiv:1710.01094
- 1 published paper about discrete tests [Döhler, Durand, and Roquain (2018)]
- 1 submitted paper about post hoc arXiv:1807.01470
- participation to 2 R packages: DiscreteFDR and sansSouci

Table of contents

Table of contents

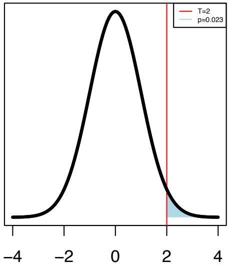
SNP association study

- A SNP = a location on the genome where there is variation
- Study the association of a SNP with a trait: case-control study
- Apply a **statistical test** to a measure $X \sim \mathcal{N}(\mu, 1)$
- Question: is $\mu = 0$ (no association) or > 0 ?



Single testing

- Null hypothesis H_0: "$\mu = 0$" versus alternative H_1: "$\mu > 0$"
- X in the tail of $\mathcal{N}(0,1) \Rightarrow$ unrealistic $\Rightarrow$ reject H_0
- $\Leftrightarrow$ Reject H_0 if the p-value $p(X) = \bar{\Phi}(X)$ is small ($\leq \alpha$)



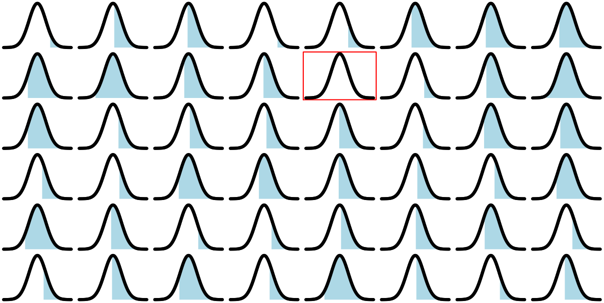
Probability to wrongly reject H_0

$$\mathbb{P}_{H_0}(p(X) \leq \alpha) \leq \alpha \text{ (uniformity under } H_0)$$

$\Longrightarrow$ false positive control

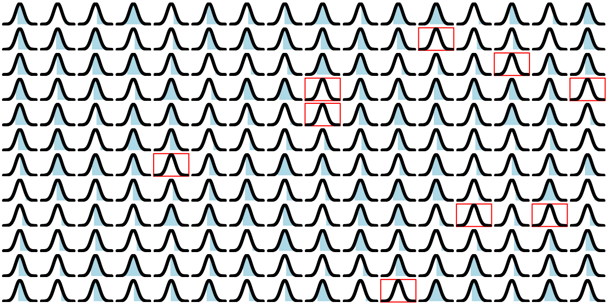
Multiple testing

- Now what if we test m SNPs at the same time?
- m null hypotheses $H_{0,i}$: "$\mu_i = 0$" versus $H_{1,i}$: "$\mu_i > 0$"
- Example if $m = 48$ and only noise (no signal):



Multiple testing

- $m = 192$, and only noise:



- $\mathbb{P}(\text{make at least one FP}) = 1 - (1-\alpha)^m \underset{m\to\infty}{\longrightarrow} 1$

Modern days applications

- $m = 10^4, 10^5, 10^6$
- Too many false positives if we keep using α as threshold

Table of contents

Multiple testing procedures

- What we want: a rejection set R with a control on $V(R)$ the number of false positives in R

Bonferroni procedure

- Use α/m instead of α
- Then $\mathbb{P}(V(R) > 0) \leq \alpha$

This can be too stringent when:

- we want a lot of detections
- we allow some false positives

False Discovery Proportion (FDP) and False Discovery Rate (FDR)

$$\mathrm{FDP}(R) = \frac{V(R)}{|R| \vee 1} \quad ; \quad \mathrm{FDR}(R) = \mathbb{E}\left[\mathrm{FDP}(R)\right]$$

Benjamini-Hochberg (BH) procedure

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

- Sort p-values: $p_{(1)} \leq \cdots \leq p_{(m)}$
- Let $\hat{k} = \max \left\{ k \in [\![1, m]\!] : p_{(k)} \leq \alpha \frac{k}{m} \right\}$ or 0 if empty set
- Reject $H_{0,i}$ if $p_i \leq \frac{\alpha \hat{k}}{m}$
- Theorem: FDR of BH $\leq \alpha$ under independence or PRDS

Useful other formulation

$\frac{\hat{k}}{m} = \max \left\{ u : \widehat{G}(u) \geq u \right\} = \mathcal{I}\left(\widehat{G}\right)$ where

$$\widehat{G} : u \mapsto m^{-1} \sum_{i=1}^{m} \mathbb{1}_{\{p_i \leq \alpha u\}}, u \in [0, 1]$$

- $\widehat{G}$ empirical c.d.f. of p-values (up to α)
- Useful for asymptotics ($m \to \infty$)

An illustration of $\mathcal{I}(\widehat{G})$

Last crossing point between $\widehat{G}$ and the identity function

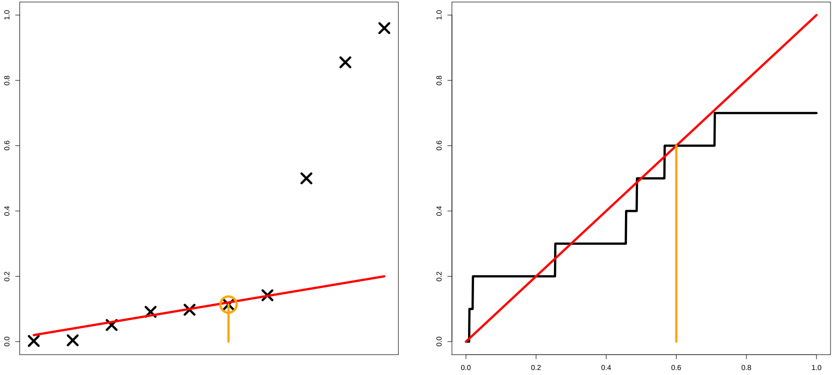


Table of contents

Table of contents

Heterogeneity

Example: GWAS study

- Multiple SNPs tested, with heterogeneous Minor Allele Frequency (MAF)
- Distinguish SNPs with low and large MAF because this changes the power of tests
- $\Longrightarrow$ form two groups of SNPs

Multiple other examples [Cai and Sun (2009)]

- Sociologic studies, fMRI...

Setting

- G groups of sizes m_g with true null proportion $\pi_{g,0}$
- $p_{g,i} \sim \mathcal{U}([0,1])$ under the null (noise); $p_{g,i} \sim F_g$ **strictly concave** under the alternative (signal)
- weak dependence [Storey, Taylor, and Siegmund (2004)] and technical assumptions

Quantities of interest

- we want FDR(R) control
- we study the optimality of the global **power**:

$$\mathrm{Pow}(R) = m^{-1}\mathbb{E}\left[|R| - V(R)\right]$$

Main tool

Attribute **weights** to each group [Holm (1979)],[Genovese, Roeder, and Wasserman (2006)],[Blanchard and Roquain (2008)]...

Table of contents

Weighted BH (WBH)

Take some weights $(w_g)_{1\leq g\leq G}, w_g \geq 0$, apply BH to modified p-values $\tilde{p}_{g,i} = p_{g,i}/w_g$

Interpretation with the $\mathcal{I}$ functional

Define

$$\widehat{G}_w : u \mapsto m^{-1} \sum_{g=1}^{G} \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \leq \alpha u w_g\}}$$

and reject all $p_{g,i} \leq \alpha \hat{u} w_g$ with $\hat{u} = \mathcal{I}\left(\widehat{G}_w\right) = \max\left\{u : \widehat{G}_w(u) \geq u\right\}$.

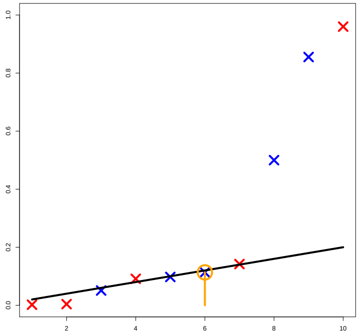
- A constraint is necessary for FDR control, for example:

$$w \in \mathcal{W} = \left\{w \geq 0, \sum_{g=1}^{G} m_g w_g \leq m\right\}$$

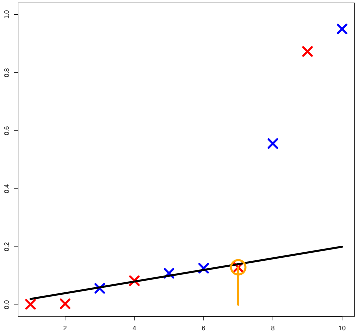
Unweighted vs Weighted BH procedure

Two groups with $w_1 + w_2 = 2$

- $w_1 = w_2 = 1$ (BH):



- $w_1 = 1.1,\ w_2 = 0.9$:



- Weights can increase detections $\Longrightarrow$ increase power ?

Multi-Weighting

[Roquain and van de Wiel (2009)]

- A generalization needed for the search for optimal power

Now weights are a function $u \mapsto W(u)$ for $u \in [0,1]$. If

$$\widehat{G}_W : u \mapsto m^{-1} \sum_{g=1}^{G} \sum_{i=1}^{m_g} \mathbb{1}_{\{p_{g,i} \leq \alpha u W_g(u)\}} \text{ is nondecreasing,}$$

then $\mathrm{MWBH}(W) = \{(g,i) : p_{g,i} \leq \alpha \hat{u} W_g(\hat{u})\}$ with $\hat{u} = \mathcal{I}\left(\widehat{G}_W\right)$.

Optimal weighting

- Fix u and w and define $R_{u,w}$ which rejects all $p_{g,i} \leq \alpha u w_g$
- Maximize its power for all u on the weight space $\mathcal{W}$:

Optimal oracle weights [Roquain and van de Wiel (2009)]

$$W^*_{or}(u) = \arg\max_{w \in \mathcal{W}} \mathrm{Pow}\left(R_{u,w}\right)$$

Theorem [Roquain and van de Wiel (2009)]

Existence and uniqueness of W^*_{or} if regularity assumptions on F_g and $\pi_{g,0} = \pi_0$.
Moreover, MWBH$\left(W^*_{or}\right)$ has asymptotical:

- FDR control at level $\pi_0\alpha$
- Power optimality among all WBH procedures

Issues and consequences

1. F_g's are unknown in practice ! So is W^*_{or}
2. The assumption $\pi_{g,0} = \pi_0$ removes some heterogeneity
3. $\sum_g m_g w_g \leq m \Longrightarrow \pi_0\alpha$-FDR control $\Longrightarrow$ conservativeness

Goal

- Estimate the oracle optimal weights
- enlarge $\mathcal{W}$ in a way that incorporates $\pi_{g,0}$ estimators
- keep asymptotical results on FDR control and power optimality

$\Longrightarrow$ Adaptive Data-Driven Optimal Weighting (ADDOW)

Table of contents

π_0 estimation

- Each $\pi_{g,0}$ is estimated with $\hat{\pi}_{g,0}$
- $\widehat{\mathcal{W}} = \left\{ w : \sum_g m_g \hat{\pi}_{g,0} w_g \leq m \right\}$ allows larger weights than $\mathcal{W}$!
- Asymptotics: $\hat{\pi}_{g,0} \xrightarrow[m \to \infty]{\mathbb{P}} \tilde{\pi}_{g,0} \geq \pi_{g,0} \Longrightarrow$ over-estimation (e.g. Storey)

Multiplicative Estimation (ME) case

There exists $C \geq 1$ such that $\tilde{\pi}_{g,0} = C \pi_{g,0}$ $\forall g$

- Includes the consistent case $\tilde{\pi}_{g,0} = \pi_{g,0}$ $\forall g$ $(C = 1)$
- Includes the case where $\hat{\pi}_{g,0} = 1$ and $\pi_{g,0} = \pi_0$ $\forall g$ $(C = 1/\pi_0)$

Definition of ADDOW

$$\text{ADDOW} = \text{MWBH}\left(\widehat{W}^*\right)$$

where

$$\forall u,\ \widehat{W}^*(u) = \arg\max_{w \in \widehat{\mathcal{W}}} \widehat{G}_w(u)$$

$\Longrightarrow$ $\widehat{W}^*$ maximizes the *number of rejections*

Key idea

Under (ME), maximizing the rejections is the same as maximizing the power

Remark : if $\hat{\pi}_{g,0} = 1\ \forall g$, ADDOW=IHW [Ignatiadis et al. (2016)]

Table of contents

Asymptotic FDR control

Theorem

$$\lim_{m\to\infty} \text{FDR}(\text{ADDOW}) \leq \alpha$$

Moreover if $\alpha \leq \tilde{\pi}_0$ and (ME):

$$\lim_{m\to\infty} \text{FDR}(\text{ADDOW}) = \frac{\alpha}{C}$$

By-product

If $\pi_{g,0} = \pi_0 \; \forall g$,

$$\lim_{m\to\infty} \text{FDR}(\text{IHW}) = \pi_0\alpha.$$

Proofs inspired by [Roquain and van de Wiel (2009)], [Hu, Zhao, and Zhou (2010)] and [Zhao and Zhang (2014)].

Power optimality

Theorem

If (ME),

$$\lim_{m\to\infty} \operatorname{Pow}(\text{ADDOW}) \geq \limsup_{m\to\infty} \operatorname{Pow}\left(\text{MWBH}\left(\widehat{W}\right)\right)$$

for any weight function sequence such that $\widehat{W}(u) \in \widehat{\mathcal{W}}$.

By-product

If $\pi_{g,0} = \pi_0 \ \forall g$,

$$\lim_{m\to\infty} \operatorname{Pow}(\text{IHW}) \geq \limsup_{m\to\infty} \operatorname{Pow}\left(\text{MWBH}\left(\widehat{W}\right)\right)$$

for any weight function sequence such that $\sum_g m_g \widehat{W}_g(u) \leq m$.

Comparison with other methods

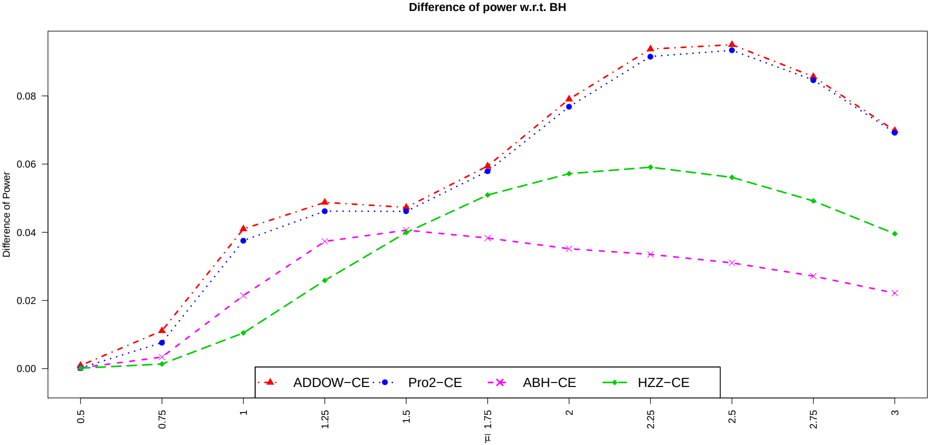
$\alpha = 0.05$, $\pi_{1,0} = 0.7$, $\pi_{2,0} = 0.8$, $m_1 = m_2 = 2000$, $\mu_1 = \bar{\mu}$ and $\mu_2 = 2\bar{\mu}$

4 methods compared with $\hat{\pi}_{g,0} = \pi_{g,0}$ (oracle) and varying signal parameter $\bar{\mu}$:

- ADDOW
- Pro2 [Zhao and Zhang (2014)]
- HZZ [Hu, Zhao, and Zhou (2010)]
- Adaptive BH

Comparison with other methods

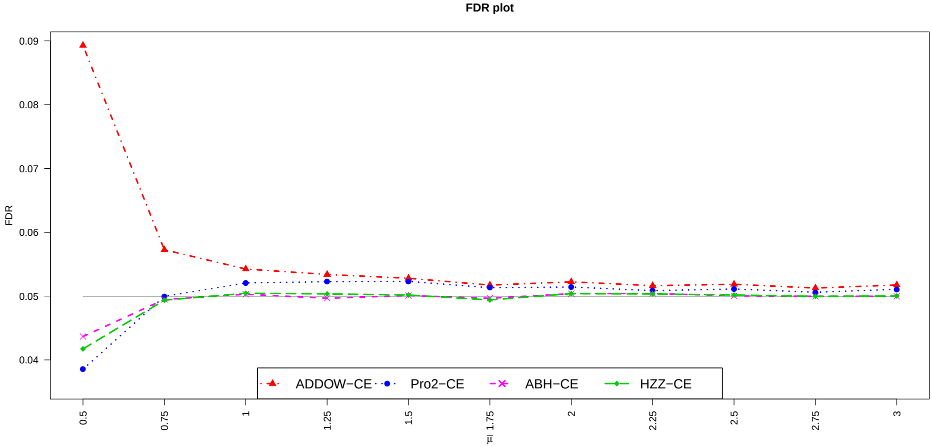
$\alpha = 0.05$, $\pi_{1,0} = 0.7$, $\pi_{2,0} = 0.8$, $m_1 = m_2 = 2000$, $\mu_1 = \bar{\mu}$ and $\mu_2 = 2\bar{\mu}$



- ADDOW > Pro2 > HZZ & ABH

Comparison with other methods

$\alpha = 0.05$, $\pi_{1,0} = 0.7$, $\pi_{2,0} = 0.8$, $m_1 = m_2 = 2000$, $\mu_1 = \bar{\mu}$ and $\mu_2 = 2\bar{\mu}$

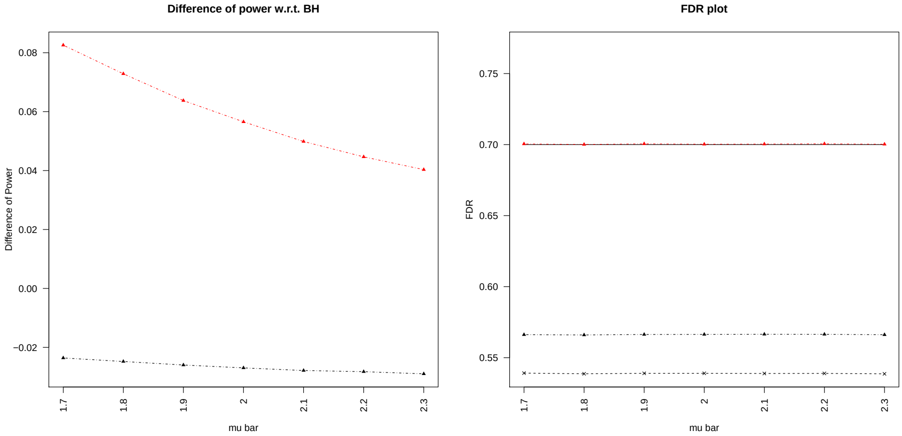


- Overfitting for ADDOW and Pro2

BH better than IHW?

$\alpha = 0.7$, $\pi_{1,0} = 0.05$, $\pi_{2,0} = 0.85$, $m_1 = 1000$, $m_2 = 9000$, $\mu_1 = 2$ and $\mu_2 = \bar{\mu}$

- ADDOW with oracle $\hat{\pi}_{g,0}$ vs IHW vs BH with heterogeneous $\pi_{g,0}$



- Remark: $m = 10^4$ here and no overfitting

Table of contents

Outlook

- Optimal asymptotical properties but with restrictive dependence assumptions and finite sample overfitting
- Incorporate the dependence ?
- Use a better estimator of the rejections than $\widehat{G}_w$? LCM ?
- FDR bound in finite sample ? (done in [Ignatiadis and Huber (2017)] with folds and censoring)
- Convergence rates ? With more regularity assumptions on F_g ?

Table of contents

Table of contents

Replication crisis

- Published results/experiments impossible to reproduce
- 70% fail rate (*Nature* poll)

A possible explanation: *p*-hacking [Wasserstein and Lazar (2016)]

- Pre-selecting variables that seem significant, exclude others from experiment
- Theoretical results no longer hold
- Results poorly interpretable and non reproductible

Toy example

- GWAS study with 10^6 genetic variants
- MTP over only the 10 smallest p-values
- Distribution of 10 smallest $\neq$ distribution of 10 p-values taken at random

Our general problem

Confidence bounds on any set of selected variables [Goeman and Solari (2011)]

A confidence (post hoc) bound is a (random) function $\widehat{V}$ such that

$$\mathbb{P}\left(\forall S \subset \mathbb{N}_m, V(S) \leq \widehat{V}(S)\right) \geq 1 - \alpha$$

- Hence for any selected $\widehat{S}$, $\mathbb{P}\left(V(\widehat{S}) \leq \widehat{V}(\widehat{S})\right) \geq 1 - \alpha$ holds
- Not a classic MTP: a guarantee over any selected set instead of a rejected set
- Originates from [Genovese and Wasserman (2006)],[Meinshausen (2006)]

BNR methodology

[Blanchard, Neuvial, and Roquain (2018)]

Key concept: reference family

- $\mathfrak{R} = (R_k, \zeta_k)$ (random) with Joint Error Rate (JER) control:

$$\mathbb{P}\left(\forall k, V(R_k) \leq \zeta_k\right) \geq 1 - \alpha$$

- Confidence bound only on the members of $\mathfrak{R}$
- $\Longrightarrow$ Derivation of a global confidence bound
- Flexible approach: we choose $\mathfrak{R}$

Two different interpolation bounds

- $V^*_{\mathfrak{R}}(S) = \max\left\{|S \cap A|, \forall k, |R_k \cap A| \leq \zeta_k\right\}$ difficult to compute
- $\overline{V}_{\mathfrak{R}}(S) = \min_k \left(\zeta_k + |S \setminus R_k|\right) \wedge |S|$ less sharp but easy to compute

Simes bound

[Goeman and Solari (2011)],[Blanchard, Neuvial, and Roquain (2018)]

Choice of $\mathfrak{R}$:

- Fix $\zeta_k = k - 1, \forall\, 1 \leq k \leq m$
- $R_k = \{i : p_i < \alpha k/m\}$

Simes bound formula

$$V^*_{\mathfrak{R}}(S) = \overline{V}_{\mathfrak{R}}(S) = \min_k \left(k - 1 + \sum_{i \in S} \mathbb{1}_{\{p_i \geq \alpha k/m\}}\right) \wedge |S|$$

Based on:

- Simes inequality and PRDS for JER control
- Nestedness of R_k's for $V^*_{\mathfrak{R}}(S) = \overline{V}_{\mathfrak{R}}(S)$

New approach

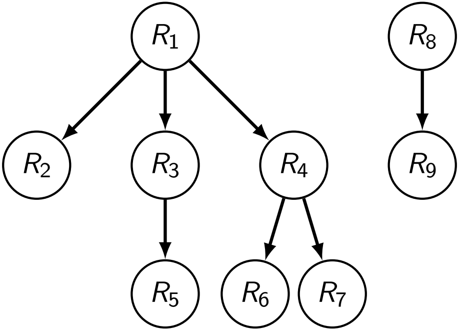
Opposite to Simes approach of BNR

1. Deterministic R_k's with a localized structure (e.g. chromosomes). Requires to compute $V^*_{\mathfrak{R}}$
2. Over-estimate $V(R_k)$ in each R_k with a simple method
   - Example with the DKWM inequality [Dvoretzky, Kiefer, and Wolfowitz (1956)],[Massart (1990)] (independence needed):
   - $\zeta_k =$
     $$|R_k| \wedge \min_{0\leq \ell \leq |R_k|} \left\lfloor \frac{C}{2(1-p_{(\ell)})} + \left(\frac{C^2}{4(1-p_{(\ell)})^2} + \frac{\sum_{i\in R_k} \mathbf{1}\{p_i > p_{(\ell)}\}}{1-p_{(\ell)}} \right)^{1/2} \right\rfloor^2,$$
     where $C = \sqrt{\frac{1}{2}\log\left(\frac{K}{\alpha}\right)}$ (union bound)
   - Other estimators possible

Table of contents

Choice of R_k: Forest structure

- $\forall k, k' \in \mathcal{K}, \;\; R_k \cap R_{k'} \in \{R_k, R_{k'}, \varnothing\}$
- Includes nested families or totally disjoint families



Main points

1. Accommodates to realistic localization structures
2. There is a simple algorithm to compute $V^*_{\mathfrak{R}}(S)$ with this structure

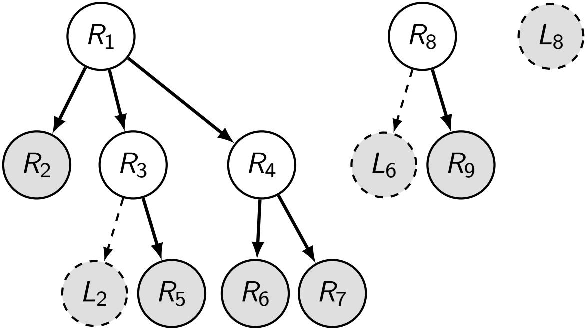
Properties of forest structures

Proposition

There is a partition $(L_n)_{1 \leq n \leq N}$ of $\mathbb{N}_m$ (the leaves) such that each R_k can be written as a union $\bigcup_{i \leq n \leq j} L_n$.

Proposition

Each forest structure can be completed to include all leaves.



New interpolation bounds

Recall: $\overline{V}_{\mathfrak{R}}(S) = \min_k \left(\zeta_k + |S \setminus R_k|\right) \wedge |S|$

Definition

For any $q \leq K$,

$$\widetilde{V}^q_{\mathfrak{R}}(S) = \min_{Q \subset \mathcal{K}, |Q| \leq q} \left(\sum_{k \in Q} \zeta_k \wedge |S \cap R_k| + \left| S \setminus \bigcup_{k \in Q} R_k \right| \right),$$

and

$$\widetilde{V}_{\mathfrak{R}}(S) = \widetilde{V}^K_{\mathfrak{R}}(S).$$

Property

$$V^*_{\mathfrak{R}}(S) \leq \widetilde{V}_{\mathfrak{R}}(S) \leq \widetilde{V}^{K-1}_{\mathfrak{R}}(S) \leq \cdots \leq \widetilde{V}^2_{\mathfrak{R}}(S) \leq \widetilde{V}^1_{\mathfrak{R}}(S) = \overline{V}_{\mathfrak{R}}(S)$$

LPSM Laboratoire de Probabilités Statistique & Modélisation — Sciences Sorbonne Université

Main results

Theorem

For a reference family with a forest structure and $\ell =$ number of leaves,

$$V^*_{\mathfrak{R}}(S) = \widetilde{V}_{\mathfrak{R}}(S) = \widetilde{V}^{\ell}_{\mathfrak{R}}(S).$$

Lemma

There is a simple algorithm to compute $\widetilde{V}_{\mathfrak{R}}$ if $\mathfrak{R}$ is complete.

Lemma

Completing the family does not change $V^*_{\mathfrak{R}}$ and $\widetilde{V}_{\mathfrak{R}}$.

Corollary

There is a simple algorithm to compute $V^*_{\mathfrak{R}}(S)$ by:

1. Completing the family
2. Travel across the forest from the leaves

Table of contents

Outlook

- Incorporation of local structure $\Longrightarrow$ better power
- DKWM inequality involves independence
- Other over-estimators of true nulls ? [Blanchard, Neuvial, and Roquain (2018)],[Hemerik and Goeman (2018)]
- Other families combining BNR approach and deterministic regions ?
  - $\mathfrak{R} = (R_{k,i_k}, \zeta_{k,i_k})_{\substack{k \in \mathcal{K} \\ 1 \leq i_k \leq |R_k|}}$

Conclusion

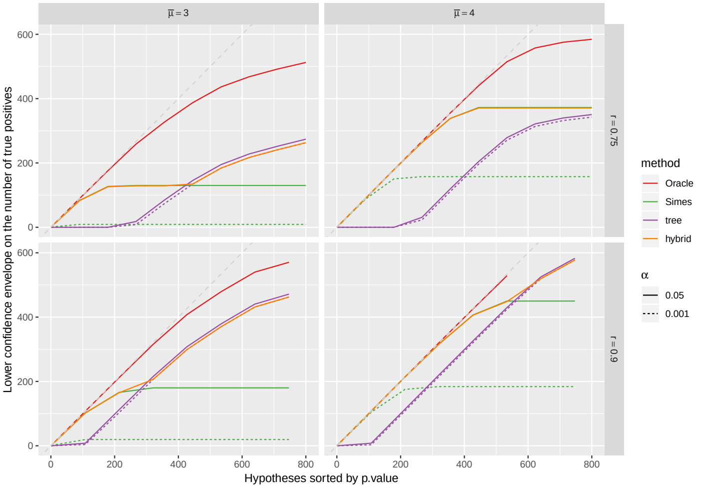
3 different ways to tackle heterogeneity in multiple testing:

- heterogeneity under the alternative with groups and weights
- heterogeneity under the discrete null
- localization heterogeneity with post hoc bounds

Next step: combine them all?

Hybrid bound

- $V^{\gamma}_{\text{hybrid}}(\alpha, S) = \min\left(V_{\text{Simes}}((1-\gamma)\alpha, S), V_{\text{tree}}(\gamma\alpha, S)\right)$
- $\gamma = 0.02$: favors Simes, convenient thanks to V_{tree} relation to α



Bibliography I

Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing". In: *Journal of the royal statistical society. Series B (Methodological)*, pp. 289–300.

Benjamini, Yoav and Daniel Yekutieli (2001). "The control of the false discovery rate in multiple testing under dependency". In: *Annals of statistics*, pp. 1165–1188.

Blanchard, Gilles, Pierre Neuvial, and Etienne Roquain (2018). "A unified approach to post hoc false positive control". In: *arXiv preprint arXiv:1703.02307*.

Blanchard, Gilles and Etienne Roquain (2008). "Two simple sufficient conditions for FDR control". In: *Electronic journal of Statistics* 2, pp. 963–992.

— (2009). "Adaptive false discovery rate control under independence and dependence". In: *J. Mach. Learn. Res.* 10, pp. 2837–2871.

Cai, T. Tony and Wenguang Sun (2009). "Simultaneous testing of grouped hypotheses: Finding needles in multiple haystacks". In: *Journal of the American Statistical Association* 104.488, pp. 1467–1481.

Chatelain, Clément et al. (2018). "Performance of epistasis detection methods in semi-simulated GWAS". In: *BMC Bioinformatics* 19.1, p. 231. ISSN: 1471-2105. DOI: 10.1186/s12859-018-2229-8. URL: https://doi.org/10.1186/s12859-018-2229-8.

Chen, X. and R. Doerge (2015). "A weighted FDR procedure under discrete and heterogeneous null distributions". In: *arXiv preprint arXiv:1502.00973*.

Bibliography II

Döhler, S. (2016). "A discrete modification of the Benjamini—Yekutieli procedure". In: *Econometrics and Statistics*. ISSN: 2452-3062. DOI: http://dx.doi.org/10.1016/j.ecosta.2016.12.002. URL: http://www.sciencedirect.com/science/article/pii/S2452306216300351.

Döhler, Sebastian, Guillermo Durand, and Etienne Roquain (2018). "New FDR bounds for discrete and heterogeneous tests". In: *Electron. J. Statist.* 12.1, pp. 1867–1900. ISSN: 1935-7524. DOI: 10.1214/18-EJS1441.

Durand, Guillermo and Sabin Lessard (2016). "Fixation probability in a two-locus intersexual selection model". In: *Theoretical population biology* 109, pp. 75–87.

Dvoretzky, Aryeh, Jack Kiefer, and Jacob Wolfowitz (1956). "Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator". In: *The Annals of Mathematical Statistics*, pp. 642–669.

Gavrilov, Yulia, Yoav Benjamini, and Sanat K. Sarkar (2009). "An adaptive step-down procedure with proven FDR control under independence". In: *Ann. Statist.* 37.2, pp. 619–629. ISSN: 0090-5364. DOI: 10.1214/07-AOS586. URL: http://dx.doi.org/10.1214/07-AOS586.

Genovese, Christopher R., Kathryn Roeder, and Larry Wasserman (2006). "False discovery control with p-value weighting". In: *Biometrika*, pp. 509–524.

Genovese, Christopher R and Larry Wasserman (2006). "Exceedance control of the false discovery proportion". In: *Journal of the American Statistical Association* 101.476, pp. 1408–1417.

Bibliography III

Goeman, Jelle J and Aldo Solari (2011). "Multiple testing for exploratory research". In: *Statistical Science*, pp. 584–597.

Heller, R. and H. Gur (2011). "False discovery rate controlling procedures for discrete tests". In: *arXiv preprint arXiv:1112.4627*. arXiv: 1112.4627 [stat.ME].

Hemerik, Jesse and Jelle J Goeman (2018). "False discovery proportion estimation by permutations: confidence for significance analysis of microarrays". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 80.1, pp. 137–155.

Heyse, Joseph F. (2011). "A false discovery rate procedure for categorical data". In: *Recent Advances in Bio- statistics: False Discovery Rates, Survival Analysis, and Related Topics*. World Scientific, pp. 43–58.

Holm, Sture (1979). "A simple sequentially rejective multiple test procedure". In: *Scandinavian journal of statistics*, pp. 65–70.

Hu, James X., Hongyu Zhao, and Harrison H. Zhou (2010). "False discovery rate control with groups". In: *Journal of the American Statistical Association* 105.491.

Ignatiadis, Nikolaos and Wolfgang Huber (2017). "Covariate powered cross-weighted multiple testing". In: *arXiv preprint arXiv:1701.05179*. URL: https://arxiv.org/abs/1701.05179.

Ignatiadis, Nikolaos et al. (2016). "Data-driven hypothesis weighting increases detection power in genome-scale multiple testing". In: *Nature methods* 13.7, pp. 577–580.

Marcus, Ruth, Peritz Eric, and K Ruben Gabriel (1976). "On closed testing procedures with special reference to ordered analysis of variance". In: *Biometrika* 63.3, pp. 655–660.

Bibliography IV

Massart, Pascal (1990). "The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality". In: *The Annals of Probability*, pp. 1269–1283.

Meinshausen, Nicolai (2006). "False discovery control for multiple tests of association under general dependence". In: *Scandinavian Journal of Statistics* 33.2, pp. 227–237.

Roquain, Etienne and Mark A. van de Wiel (2009). "Optimal weighting for false discovery rate control". In: *Electronic Journal of Statistics* 3, pp. 678–711.

Storey, John D, Jonathan E Taylor, and David Siegmund (2004). "Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66.1, pp. 187–205.

Wasserstein, Ronald L and Nicole A Lazar (2016). "The ASA's statement on p-values: context, process, and purpose". In: *The American Statistician* 70.2, pp. 129–133.

Zhao, Haibing and Jiajia Zhang (2014). "Weighted p-value procedures for controlling FDR of grouped hypotheses". In: *Journal of Statistical Planning and Inference* 151, pp. 90–106.

Multi-Weighting

A practical way to compute $\mathcal{I}\left(\widehat{G}_W\right)$

No need to compute $W(u)$ for each u !

$\forall k \in [\![1, m]\!]$, compute all $\frac{p_i}{W_i(\frac{k}{m})}$ and take q_k the k-th smallest.

Let $q_0 = 0$.

Then $\mathcal{I}\left(\widehat{G}_W\right) = m^{-1} \max\{k \in [\![0, m]\!] : q_k \leq \alpha \frac{k}{m}\}$.

About the computation of $\widehat{W}^*$

Key ideas

- Compute only $\widehat{W}^*(u)$ for $u = \frac{1}{m}, \frac{2}{m}, \dots, \frac{m-1}{m}, 1$
- Fixing u, $w \mapsto \widehat{G}_w(u)$ only jumps at the $\frac{p_{g,i}}{\alpha u} \Longrightarrow$ let $\widehat{W}_g^*(u) = \frac{p_{g,i_g}}{\alpha u}$ such that $\sum m_g \hat{\pi}_{g,0} \frac{p_{g,i_g}}{\alpha u} \leq m$ and $\sum_g i_g$ is maximal
- $\widehat{G}_w(u)$ nondecreasing in u AND w : try to reject 1 hyp, then 2, then 3... for $u = \frac{1}{m}$, when fail at k hyp, try to reject k hyp for $u = \frac{2}{m}$, ...

Proof ideas

- $\widehat{G}_{\widehat{W}^*}(\hat{u}) \xrightarrow{\mathbb{P}} G^{\infty}_{W^*}(u^*)$ by LLN and careful use of maximality
- Then $\hat{u} \xrightarrow{\mathbb{P}} u^*$ by continuity of $\mathcal{I}(\cdot)$
- Then $\widehat{W}^*(\hat{u}) \xrightarrow{\mathbb{P}} W^*(u^*)$ by reductio ad absurdum
- $\Longrightarrow \text{FDP} = \dfrac{m^{-1}\sum_g \sum_i \mathbb{1}_{\left\{p_{g,i} \le \alpha \hat{u} \widehat{W}^*_g(\hat{u}) \text{ and } H_{g,i} \text{ true}\right\}}}{\widehat{G}_{\widehat{W}^*}(\hat{u})} \xrightarrow{\mathbb{P}}$
  $\dfrac{\alpha \sum_g \pi_g \pi_{g,0} u^* W^*_g(u^*)}{u^*} = \alpha \sum_g \pi_g \pi_{g,0} W^*_g(u^*) \le \alpha$ (weight space choice)
- under (ME), maximize rejections $\Leftrightarrow$ maximize power because we can write $\sum_g \pi_g \pi_{g,0} U(\alpha u w_g) \le \frac{\alpha u}{C} \sum_g \pi_g \tilde{\pi}_{g,0} w_g \le \frac{\alpha u}{C} \Longrightarrow$ no dependence in w in $G^{\infty}_w - P^{\infty}_w$!

sADDOW$_\beta$

Stabilization for weak signal

- ADDOW overfits so FDR control lost with weak signal in finite sample
- We should prefer BH then
- $\Longrightarrow$ test if there is signal before choosing the procedure, like KS tests

Definition

$$\text{sADDOW}_\beta = \begin{cases} \text{ADDOW} & \text{if } \phi_\beta = \mathbb{1}_{\{Z_m > q_{\beta,m}\}} = 1 \\ \text{BH} & \text{if } \phi_\beta = \mathbb{1}_{\{Z_m > q_{\beta,m}\}} = 0 \end{cases}$$

with $Z_m = \sqrt{m}\sup_{u\in[0,1]}\left(\widehat{G}_{\widehat{W}^*}(u) - \alpha u\right)$ and $q_{\beta,m}$ the $(1-\beta)$ quantile of Z_{0m} (independent copy of Z_m under full null, and independence).

sADDOW$_\beta$

Stabilization for weak signal

Main idea (under independence)

Weak signal $\Longrightarrow$ Z_m close to Z_{0m} in distribution, and

$$\begin{aligned}
\mathrm{FDR}\left(\mathrm{sADDOW}_\beta\right) &= \mathbb{E}\left[\phi_\beta \,\mathrm{FDP}\left(\mathrm{ADDOW}\right) + (1-\phi_\beta)\,\mathrm{FDP}\left(\mathrm{BH}\right)\right] \\
&\leq \mathbb{E}\left[\phi_\beta + \mathrm{FDP}\left(\mathrm{BH}\right)\right] \\
&\leq \mathbb{P}\left(Z_m > q_{\beta,m}\right) + \frac{m_0}{m}\alpha \\
&\lesssim \mathbb{P}\left(Z_{0m} > q_{\beta,m}\right) + \frac{m_0}{m}\alpha \\
&\leq \beta + \frac{m_0}{m}\alpha
\end{aligned}$$

- $\beta \to 0$?

sADDOW$_\beta$ equivalent to ADDOW

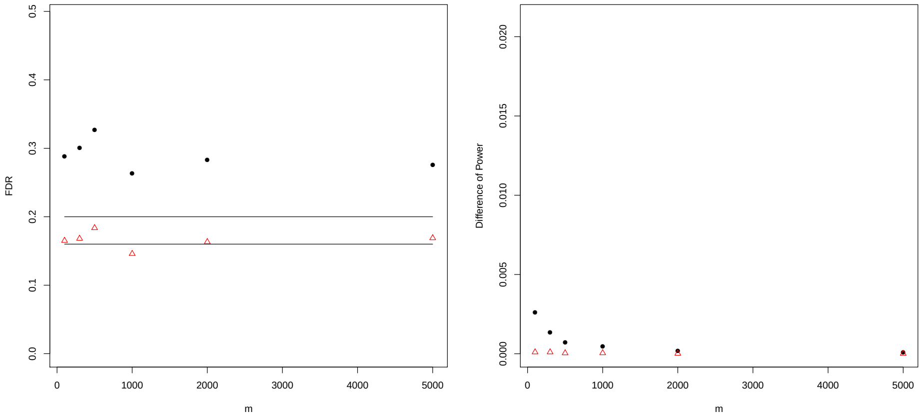
Theorem

sADDOW$_\beta$ is asymptotically equivalent to ADDOW because $\phi_\beta \xrightarrow{a.s.} 1$ when $m \to \infty$, even if $\beta = \beta_m \to 0$ not too slowly ($\beta_m \geq \exp\left(-m^{1-\nu}\right), \nu > 0$).

Proof relies on the DKWM inequality [Massart (1990)]

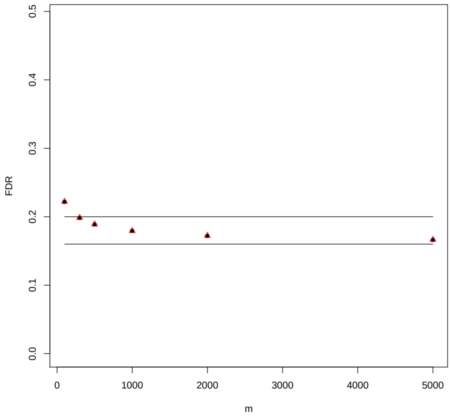
Stabilization for weak signal : $\bar{\mu} = 0.01$

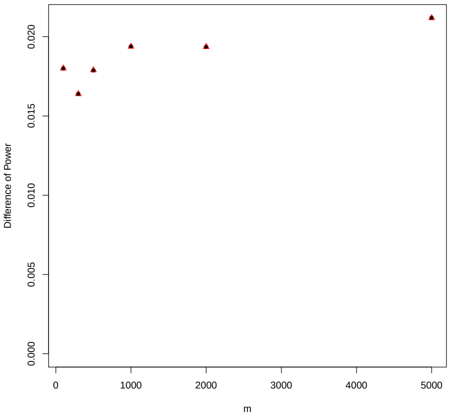
$\pi_1 = \pi_2 = 0.5$, $\pi_0 = 0.8$, $\mu_1 = \bar{\mu}$, $\mu_2 = 2\bar{\mu}$, 1000 replications



Stabilization for strong signal : $\bar{\mu} = 3$

$\pi_1 = \pi_2 = 0.5$, $\mu_1 = \bar{\mu}$, $\mu_2 = 2\bar{\mu}$, 1000 replications





Closed testing for post hoc inference

Designed for FWER control [Marcus, Eric, and Gabriel (1976)]

- Form $H_{0,I} = \bigcap_{i \in I} H_{0,i}$ all intersection hypotheses
- Have a collection of α level local test ϕ_I
- Examples:
  - Bonferroni test $\phi_I = 1$ if $\exists i \in I : p_i \leq \alpha/|I|$
  - Simes test $\phi_I = 1$ if $\exists i \in I : p_{(i:I)} \leq \alpha i/|I|$ (under PRDS)
- Test $H_{0,I}$ only if all $H_{0,J}$, $J \supseteq I$, are rejected
- Reject the individual hypotheses $H_{0,i}$ such that $H_{0,\{i\}}$ has been rejected that way
- Then FWER(Closed testing) $\leq \alpha$

Closed testing for post hoc inference

[Goeman and Solari (2011)]

Main idea

The closed testing provides more information than just the individual rejects:

- Let $\mathcal{X}$ the set of all I such that we rejected $H_{0,I}$
- Simultaneous guarantee over all $H_{0,I}$, $I \in \mathcal{X}$:

$$\mathbb{P}\left(\forall I \in \mathcal{X}, H_{0,I} \text{ is false}\right) \geq 1 - \alpha$$

Confidence bound derivation:

- $V_{\text{GS}}(S) = \max_{\substack{I \subseteq S \\ I \notin \mathcal{X}}} |I|$ is a confidence bound because

$$\begin{aligned} \exists S, |S \cap \mathcal{H}_0| > V_{\text{GS}}(S) &\Longrightarrow \exists S, S \cap \mathcal{H}_0 \in \mathcal{X} \\ &\Longrightarrow \exists I \in \mathcal{X}, H_{0,I} \text{ is true} \end{aligned}$$

- $V_{\text{GS}}(S) = V^*_{\mathfrak{R}}(S)$ with $\mathfrak{R} = (I, |I| - 1)_{I \in \mathcal{X}}$

DKWM use

- Let $S \subset \mathbb{N}_m$
- $N_t(S) = \sum_{i \in S} \mathbf{1}\{p_i(X) > t\}$
- $v = |S \cap \mathcal{H}_0|$

$$v \leq \min_{t \in [0,1)} \left(\frac{\sqrt{\log(1/\lambda)/2}}{2(1-t)} + \left\{ \frac{\log(1/\lambda)/2}{4(1-t)^2} + \frac{N_t(S)}{1-t} \right\}^{1/2} \right)^2$$

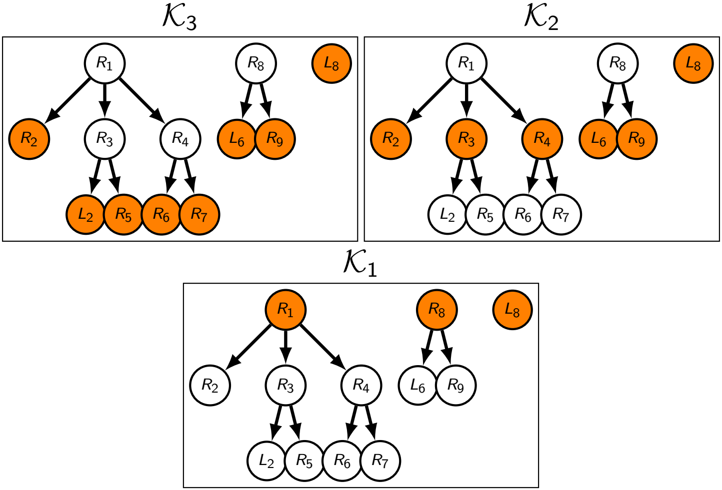
comes from

$$v^{-1} \sum_{i=1}^{v} \mathbf{1}\{U_i > t\} - (1-t) \geq -\sqrt{\log(1/\lambda)/(2v)}, \ \forall t \in [0,1],$$

with probability at least $1-\lambda$ ($U_1, \dots, U_v$ i.i.d. uniform, $N_t(S)$ dominates $\sum_{i=1}^{v} \mathbf{1}\{U_i > t\}$ by independence)

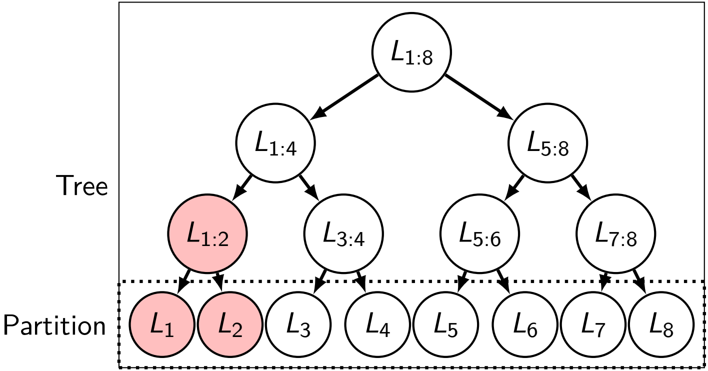
- $S = R_k$ and $\lambda = \alpha/K$ (union bound)

Forest algorithm



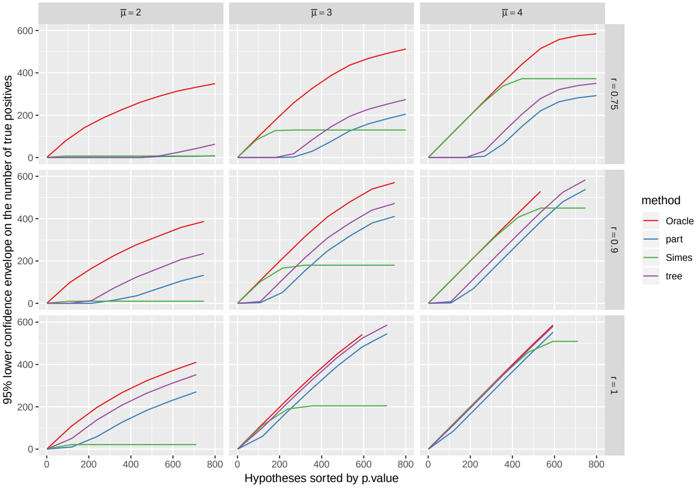
Comparison of Simes vs 2 new bounds

- V_{tree} and V_{part}: complete binary tree or only the leaves partition
- Signal in adjacent leaves, to test V_{tree} w.r.t V_{part}
- Parameters: signal $\bar{\mu}$ and signal proportion in active leaves r



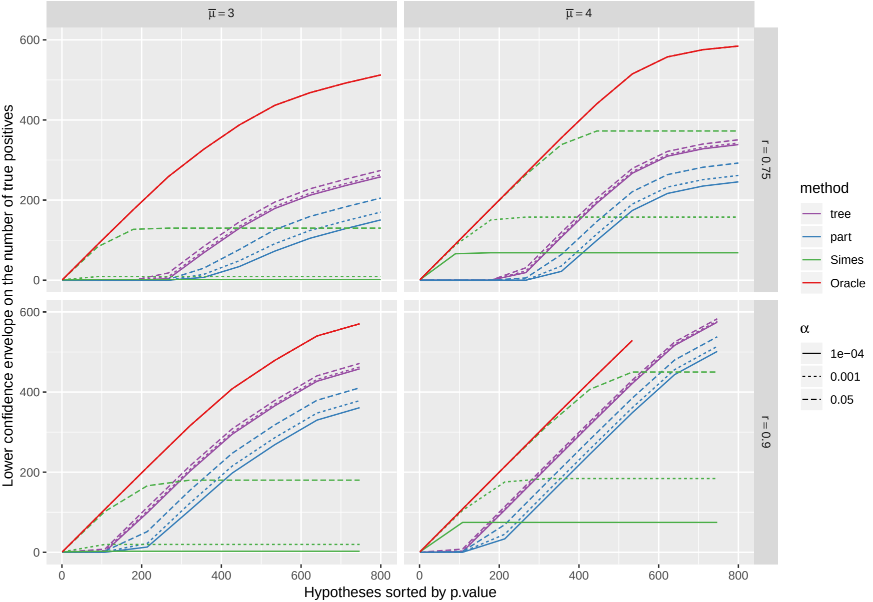
Comparison of Simes vs 2 new bounds

- The choice of S favors the Simes bound of BNR
- Simes better with $\bar{\mu}$, new bounds better with r
- V_{tree} better than V_{part}, despite union bound penalty



Comparison of 3 bounds

Influence of α



An example of discrete test

Fisher's exact test

- GWAS study
- Testing association between allele A and a phenotype (1) of interest

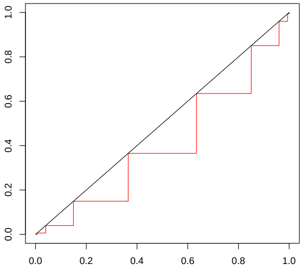
	Phenotype 1	Phenotype 2	Total
Allele A	$n_{1,\mathrm{A}}$	$n_{2,\mathrm{A}}$	n_A
Allele a	$n_{1,\mathrm{a}}$	$n_{2,\mathrm{a}}$	n_a
Total	n_1	n_2	N

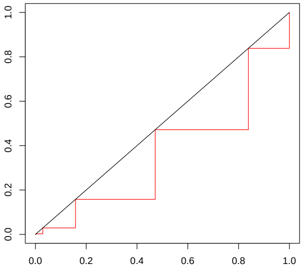
- For large samples, χ^2 approximation: $E_{1,\mathrm{A}} = \frac{n_1 n_\mathrm{A}}{N}$, $\frac{(n_{1,\mathrm{A}} - E_{1,\mathrm{A}})^2}{E_{1,\mathrm{A}}} + \dots$ follows χ^2 distribution under H_0
- What if we want an exact test ?
- Under H_0, conditionally to n_1 and n_A, $n_{1,\mathrm{A}} \sim \mathcal{H}(n_1, n_2, n_\mathrm{A})$, hypergeometric hence discrete

Issue of discrete p-values

Super-uniformity rather than uniformity under H_0

Let's see the c.d.f. of $\mathcal{H}(30, 30, 10)$ and $\mathcal{H}(14, 42, 6)$.





Super-uniformity under H_0

$$\mathbb{P}\left(p \leq u\right) \leq \mathbb{P}\left(U \leq u\right) = u$$

i.e under the null, our p-values are usually larger than uniforms

Issue of discrete p-values

Super-uniformity rather than uniformity under H_0

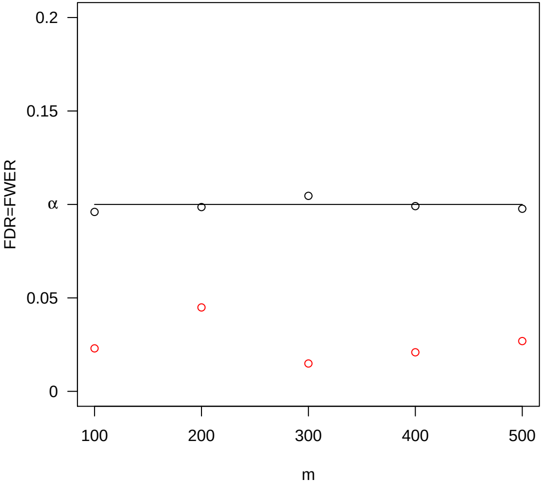
Problem

Usual MT procedures designed for uniformity

- As discrete p-values are larger than uniforms, classic thresholds are too conservative $\Longrightarrow$ loss of power

Issue of discrete p-values

Toy example: BH under full null, $m/2$ p-values $\sim \mathcal{H}(30, 30, 10)$, and $m/2$ p-values $\sim \mathcal{H}(14, 42, 6)$



In this section

- New procedures that use heterogeneous discrete distributions
- New FDR bounds and FDR control of our procedures
- Numerical illustrations

Back to BH (again)

- Reject all $p_i \leq \alpha \frac{\hat{k}}{m}$ where $\hat{k} = \max\{k : p_{(k)} \leq \alpha k/m\}$

Step-up procedure, critical constants

- Take a nondecreasing sequence (τ_k), the critical constants
- Reject all $p_i \leq \tau_{\hat{k}}$ where $\hat{k} = \max\{k : p_{(k)} \leq \tau_k\}$

Examples:

- BH: $\tau_k = \alpha k/m$
- BY [Benjamini and Yekutieli (2001)]: $\tau_k = \alpha k/(m \times \sum_{i=1}^{m} i^{-1})$

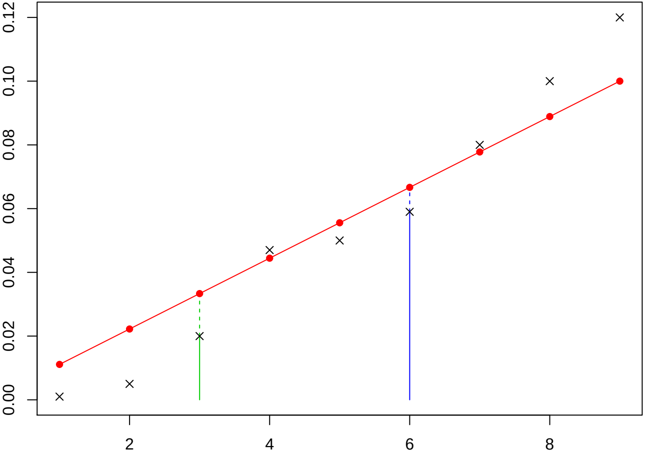
Step-down procedure

- Reject all $p_i \leq \tau_{\hat{k}}$ where $\hat{k} = \max\{k : \forall k' \leq k, p_{(k')} \leq \tau_{k'}\}$

Example:

- HB [Holm (1979)]: $\tau_k = \alpha/(m + k - 1)$

Step-up, step-down



Heyse procedure

[Heyse (2011)]

A step-up procedure

With $\tau_k = \max\{t \in \mathcal{A} \,:\, \overline{F}(t) \leq \alpha k/m\}$, where $\overline{F}(t) = \frac{1}{m}\sum_{i=1}^{m} F_i(t)$, F_i (known) c.d.f. under the null, $\mathcal{A}$ discrete support of p-values

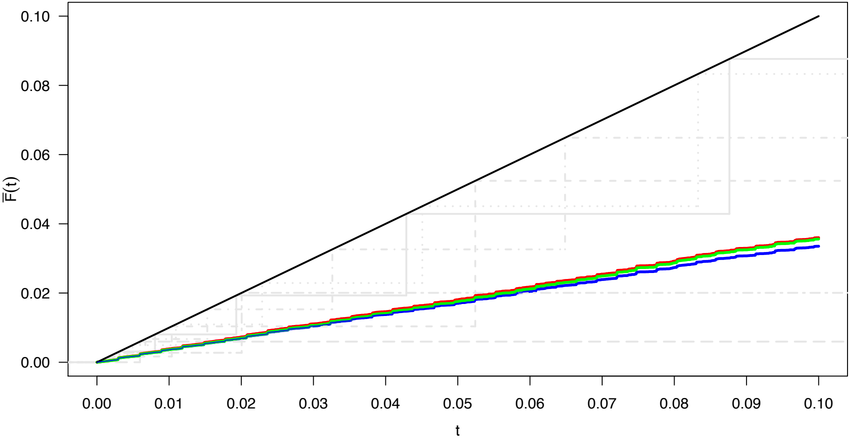
Main ideas:

- "Invert" $\overline{F}$ at $\alpha k/m$
- As $\overline{F}(t) \leq t$, this yields larger critical values than BH
- The more $\overline{F}(t)$ is small compared to t, the larger are the τ_k
- But heterogeneity also needed, otherwise Heyse $=$ BH (because $F_i(t) = t$ when $t \in \mathcal{A}_i$)

Heyse procedure

An illustration

- Compensation effect from heterogeneity



Heyse procedure

Problem

Heyse procedure does not control the FDR

Counter-examples exist

In the following

Build upon Heyse ideas but with FDR control

HSU and HSD

Slight modifications of $\overline{F}$

$$\overline{F}_{\mathrm{SU}}(t) = \frac{1}{m}\sum_{i=1}^{m}\frac{F_i(t)}{1 - F_i(\tau_m)}; \quad \overline{F}_{\mathrm{SD}}(t) = \frac{1}{m}\sum_{i=1}^{m}\frac{F_i(t)}{1 - F_i(t)}$$

where

$$\tau_m = \max\{t \in \mathcal{A} \, : \, \overline{F}_{\mathrm{SD}}(t) \leq \alpha\}$$

HSU

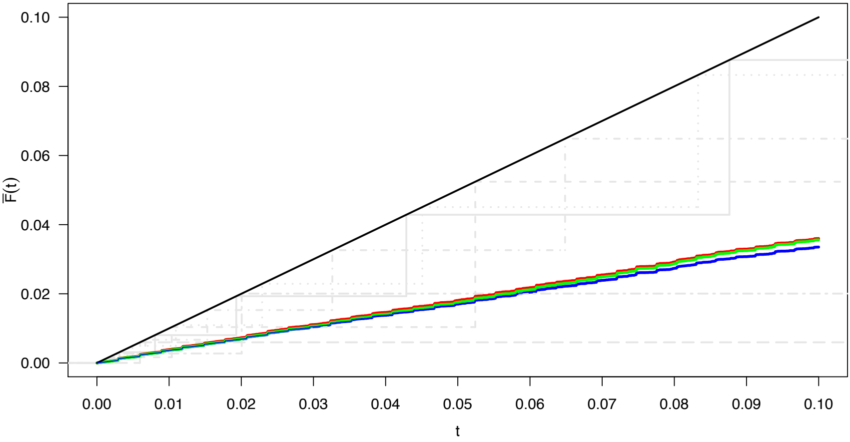
SU with $\tau_k = \max\{t \in \mathcal{A} \, : \, t \leq \tau_m, \, \overline{F}_{\mathrm{SU}}(t) \leq \alpha k/m\}$, $k \leq m - 1$

HSD

SD with $\tau_k = \max\{t \in \mathcal{A} \, : \, \overline{F}_{\mathrm{SD}}(t) \leq \alpha k/m\}$, $k \leq m$

HSU and HSD

- $\overline{F}_{\mathrm{SU}}, \overline{F}_{\mathrm{SD}} \geq \overline{F}$
- But not that much!



AHSU ans AHSD

Adaptive procedures

AHSU

SU with $\tau_k =$
$\max\left\{t \in \mathcal{A} : t \leq \tau_m, \left(\frac{F(t)}{1-F(\tau_m)}\right)_{(1)} + \cdots + \left(\frac{F(t)}{1-F(\tau_m)}\right)_{(m-k+1)} \leq \alpha k\right\}$,
$k \leq m-1$

AHSD

SD with
$\tau_k = \max\left\{t \in \mathcal{A} \,:\, \left(\frac{F(t)}{1-F(t)}\right)_{(1)} + \cdots + \left(\frac{F(t)}{1-F(t)}\right)_{(m-k+1)} \leq \alpha k\right\}$, $k \leq m$

AHSU ans AHSD

Why "adaptive" ?

Back to HB

- SD with $\tau_k = \alpha/(m+k-1)$
- Sequential point of view: if $p_{(1)} \leq \alpha/m$, then at most $m-1$ true nulls, let's see if $p_{(2)} \leq \alpha/(m-1)$...
- Adapts to the quantity of signal
- Controls the FWER

- HSD is the discrete version of GBS [Gavrilov, Benjamini, and Sarkar (2009)]: $\tau_k = \frac{\alpha k}{m-(1-\alpha)k+1}$
- GBS itself is the FDR version of HB

New FDR bounds

Under independence

Theorem

$$\mathrm{FDR}(\mathbf{SU}(\tau)) \leq \min\left(\sum_{i=1}^{m} \max_{k} \frac{F_i(\tau_k)}{k},\right.$$
$$\left.\max_{k} \frac{1}{k}\left(\left(\frac{F(\tau_k)}{1-F(\tau_m)}\right)_{(1)} + \cdots + \left(\frac{F(\tau_k)}{1-F(\tau_m)}\right)_{(m-k+1)}\right)\right)$$

New FDR bounds

Under independence

Theorem

$$\mathrm{FDR}(\mathbf{SD}(\tau)) \leq \min\left(\sum_{i=1}^{m} \max_k \frac{F_i(\tau_k)}{k},\right.$$
$$\left.\max_k \frac{1}{k}\left(\left(\frac{F(\tau_k)}{1-F(\tau_k)}\right)_{(1)} + \cdots + \left(\frac{F(\tau_k)}{1-F(\tau_k)}\right)_{(m-k+1)}\right)\right)$$

Direct corollaries

Corollary

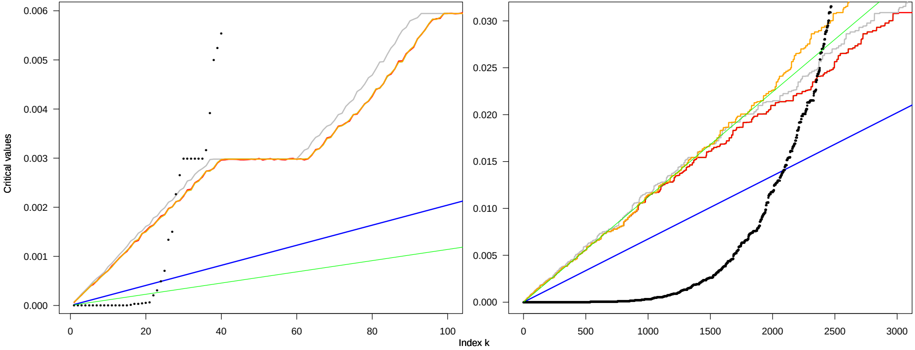
HSU, HSD, AHSU, AHSD all control the FDR under indepence

Also recovery of other results

- FDR control under independence of BH and GBS
- [Blanchard and Roquain (2009)] recovered in a special case
- [Roquain and van de Wiel (2009)] recovered in finite sample $\Longrightarrow$ connection with weighting

Real data analysis

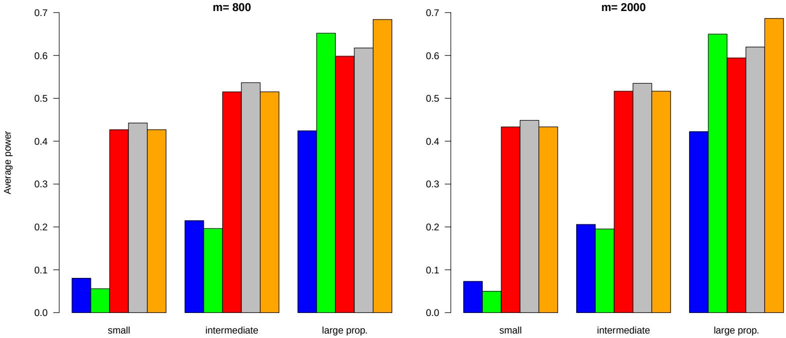
Pharmacovigilance data [Heller and Gur (2011)] and methylation data [Chen and Doerge (2015)]



- Blue: BH
- Green: Storey with $\lambda = 1/2$ but without censoring [Storey, Taylor, and Siegmund (2004)]
- Grey: Heyse
- Red: HSU
- Orange: AHSU

Simulations

Fisher's exact test on m contingency tables



$$\pi_0 = 0.9, 0.7, 0.2$$

Conclusion

- New powerful procedures well-suited for discrete tests under independence
- New bounds also relevant in other contexts (i.e. weighting)
- Positive dependence ? Any dependence ? [Döhler (2016)]
- π_0 estimation ?

A discrete procedure always better than BH: RBH

SU procedure

With $\tau_k = \lambda_\alpha k/m$, $\lambda_\alpha = \max\{\lambda \in [0,1] : \Psi(\lambda_\alpha) \leq \alpha\}$, and

$$\Psi(\lambda) = \min\left(\lambda, \max_{1\leq k\leq m}\left(\frac{1}{k}\sum_{i=1}^{m}\frac{F_i(\lambda k/m)}{1-F_i(\lambda)}\right)\right)$$

- FDR control by the new FDR bounds
- If $\Psi(\lambda_\alpha) = \alpha$ then RBH always better than BH because $\alpha \leq \lambda_\alpha$