Post hoc inference via multiple testing

Pierre Neuvial

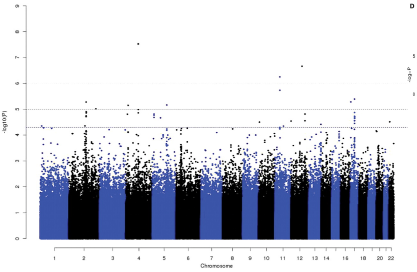
Institut de Mathématiques de Toulouse

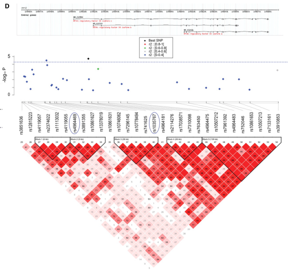
Joint work with Gilles Blanchard and Etienne Roquain

Banff, 2017-03-27

Outline

Genome-Wide Association Studies





Typical analysis steps

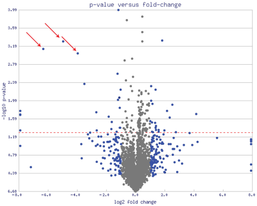
1. define a list of candidates using a multiple testing procedure
2. refine this list based on prior knowledge (genome regions)

Limitations

- Initial selection does not take advantage of available prior knowledge
- No formal risk assessment can be made on the resulting candidate sets

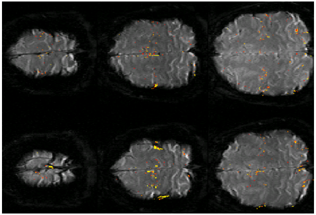
Other motivating examples

Cancer studies



Differential gene expression analyses

Neuroimaging



Activation of brain regions

Typical analysis steps

1. define a list of candidates using a multiple testing procedure
2. refine this list based on prior knowledge (gene pathways, brain regions)

Limitations

- Initial selection does not take advantage of available prior knowledge
- No formal risk assessment can be made on the resulting candidate sets

Multiple testing: notations

- $\mathcal{H} = \{1, \dots m\}$ m null hypotheses to be tested
- $\mathcal{H}_0 \subset \mathcal{H}$: true null hypotheses, $\mathcal{H}_1 = \mathcal{H} \setminus \mathcal{H}_0$
- $(p_i)_{1 \leq i \leq m}$: p-values

Multiple testing procedures

Aim at building from the data a set R of rejected hypotheses satisfying a statistical guarantee, e.g. controlling:

- (k-)Family-Wise Error Rate: $k\text{-FWER} = \mathbb{P}(|R \cap \mathcal{H}_0| > k - 1)$
- False Discovery Rate: $\text{FDR} = \mathbb{E}\left(\frac{|R \cap \mathcal{H}_0|}{|R| \vee 1}\right)$

Most procedures used in applications are thresholding procedures:

$$R = \{i \in \mathcal{H}, p_i \leq \hat{t}\}$$

Post hoc inference

Goal: Confidence statements for the number of true/false positives on any number of arbitrary rejection sets, possibly selected after data analysis

Formal objective

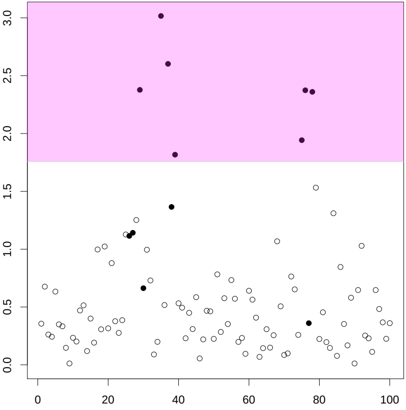
Find $\overline{V}_\alpha, \overline{S}_\alpha$ such that $\forall R \subset \{1 \dots m\}$,

$$\mathbb{P}(|R \cap \mathcal{H}_0| \geq \overline{V}_\alpha(R)) \geq 1 - \alpha$$

$$\mathbb{P}(|R \cap \mathcal{H}_1| \leq \overline{S}_\alpha(R)) \geq 1 - \alpha$$

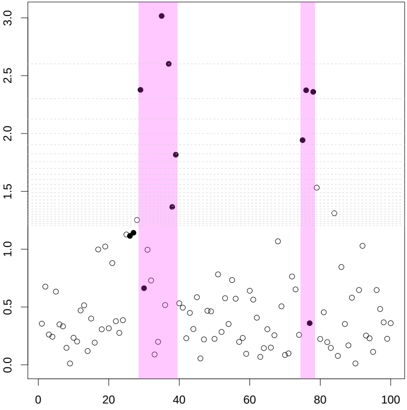
Post hoc inference in a nutshell

Classical multiple testing



FDR $\leq 25\%$

Post hoc inference



With probability $\geq 75\%$

$|R \cap \mathcal{H}_1| \geq 2$ and $|R' \cap \mathcal{H}_1| \geq 1$

State of the art: Goeman and Solari (2011)

Existing post hoc procedures[1] are based on *closed testing*[2]

- Require testing all $2^m - 1$ possible intersections between the m original hypotheses!
- Not feasible for $m \geq 20$ or 30.

In practice: "shortcuts"

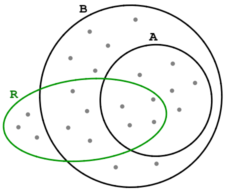
- computationally efficient procedures (complexity $\sim m \log(m)$)
- increased conservativeness and/or narrower applicability:
- Simes' shortcut: valid under positive dependence between hypotheses (PRDS)

[1] Multiple testing for exploratory research. *Stat. Science* (2011)

[2] Marcus, Peritz and Gabriel, *Biometrika* (1976).

Joint Family-Wise Error Rate (JER)

Intuition



Given A and B such that:

- $|A \cap \mathcal{H}_0(P)| \leq 5$
- $|B \cap \mathcal{H}_0(P)| \leq 7$

Then we can guarantee:

$$|R \cap \mathcal{H}_1(P)| \geq 1$$

Definition

Let $\mathfrak{R} = (R_k)_{k=1\ldots m}$ be a *reference family* of rejection sets. Then

$$\mathrm{JER}(\mathfrak{R}) = \mathbb{P}(\exists k \in \{1, \ldots, m\}, |R_k \cap \mathcal{H}_0| \geq k)$$

Consequently, $\mathfrak{R}$ is said to control JER at level $\alpha \in [0, 1]$ if:

$$\mathbb{P}(\forall k \in \{1, \ldots, m\}, |R_k \cap \mathcal{H}_0| \leq k - 1) \geq 1 - \alpha$$

Post hoc inference through JER control

JER control

$$\mathbb{P}\big(\forall k \in \{1, \dots, m\}, |R_k \cap \mathcal{H}_0| \le k - 1\big) \ge 1 - \alpha$$

Upper bound on the number of false positives

Given a JER controlling family $(R_k)_{k=1\dots m}$, with probability larger than $1-\alpha$, for any rejection set R,

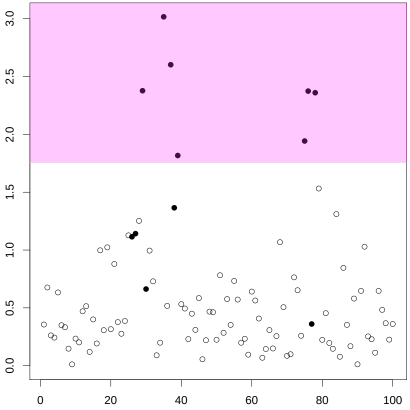
$$|R \cap \mathcal{H}_0| \le |R| \wedge \min_{1 \le k \le |R|} \{|R \cap (R_k)^c| + k - 1\}$$

Applicable to

- data-driven rejection sets
- any number of rejection sets

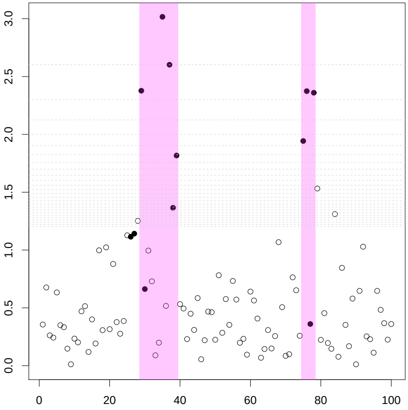
Illustration

Classical multiple testing



FDR $\leq 25\%$

Post hoc inference



With probability $\geq 75\%$

$|R \cap \mathcal{H}_1| \geq 2$ and $|R' \cap \mathcal{H}_1| \geq 1$

How can JER control be achieved?

Simes' inequality [3]

If the p-values (p_i), $1 \leq i \leq m$, are PRDS then

$$\mathbb{P}(\exists k \in \{1, \ldots, m_0\} \ : \ q_{(k)} \leq \alpha k / m_0) \leq \alpha,$$

where $q_{(1)} \leq \cdots \leq q_{(m_0)}$ denote the ordered p-values under H_0

[3] R. J. Simes. *Biometrika* 73.3 (1986), pp. 751–754.

Simes-based JER control

Corollary of Simes' inequality

Under PRDS, JER control at level α is achieved by the Simes reference family:

$$R_k = \{1 \leq i \leq m \,:\, p_i \leq \alpha k/m\}, 1 \leq k \leq m$$

Proposition (Post hoc bound for the Simes family)

Under PRDS, with probability larger than $1-\alpha$, for any R,

$$|R \cap \mathcal{H}_0| \leq |R| \wedge \min_{1 \leq k \leq |R|} \left\{ \sum_{i \in R} \mathbf{1}\{p_i > \alpha k/m\} + k - 1 \right\}.$$

- We recover the bound obtained by Goeman and Solari (2011)
- Easier to interpret: no closed testing or shortcuts
- JER: a generic device to build post hoc bounds

Dependence-free JER control?

Under arbitrary dependence, with probability larger than $1-\alpha$, for any R,

$$|R \cap \mathcal{H}_0| \leq |R| \wedge \min_{1 \leq k \leq |R|} \left\{ \sum_{i \in R} \mathbf{1}\left\{p_i > \alpha / C_m k/m\right\} + k - 1 \right\},$$

$C_m = \sum_{k=1}^{m} k^{-1} \sim \log(m)$: Hommel's correction factor for dependency[4]

Dependence-free adjustment is not a sensible objective

- implies adjusting to a worst case dependency
- very conservative (cf Benjamini-Yekutieli for FDR control)

We want to be adaptive to dependency

[4] G Hommel. "Tests of the overall hypothesis for arbitrary dependence structures". Biometrische Zeitschrift 25.5 (1983), pp. 423–430.

Sharpness and conservativeness of the Simes family

Simes' equality is sharp under independence, but conservative under positive dependence.

Conservativeness of JFWER control under PRDS

Toy example: Gaussian equi-correlation, white setting ($m_0 = m = 1,000$):
Test statistics $\sim \mathcal{N}(0, \Sigma)$ with $\Sigma_{ii} = 1$ and $\Sigma_{ij} = \rho$ for $i \neq j$.

Equi-correlation level: ρ	0	0.1	0.2	0.4	0.8
Achieved JFWER $\times \alpha^{-1}$	0.99	0.85	0.72	0.42	0.39

Can we build a family achieving sharper JFWER control?

We want to be adaptive to dependency

JER control with λ adjustment

Rejection kernel

Consider the reference family:

$$R_k = \{1 \leq i \leq m \,:\, p_i \leq t_k(\alpha)\}, 1 \leq k \leq m\,,$$

where $t_k(0) = 0$ and $t_k(\cdot)$ is non-decreasing and left-continuous on $[0, 1]$

- Example (Simes family): $t_k(\alpha) = \alpha k/m$

The associated *rejection kernel* is the collection of $(t_k(\lambda)_{k=1\ldots m})$ for all $0 \leq \lambda \leq 1$

Single-step λ adjustment

$$\lambda(\alpha) = \max\left\{\lambda \geq 0 \,:\, \mathbb{P}\left(\min_{1\leq k\leq K}\left\{t_k^{-1}\left(p_{(k:\mathcal{H})}\right)\right\} \leq \lambda\right) \leq \alpha\right\}.$$

The family $\mathfrak{R}_{\lambda(\alpha)}$ controls JER at level α.

Calculating the adjustment factor $\lambda(\alpha)$

$$\lambda(\alpha) = \max\left\{\lambda \geq 0 \,:\, \mathbb{P}\left(\min_{1 \leq k \leq K}\left\{t_k^{-1}\left(p_{(k:\mathcal{H})}\right)\right\} \leq \lambda\right) \leq \alpha\right\}$$

Calculating $\lambda(\alpha)$ requires the knowledge of the distribution of $(p_{(k:\mathcal{H})})_k$!

Using Monte-Carlo approximation if the joint null distribution is known

- see below example of Gaussian equi-correlation
- more in G. Blanchard, P. Neuvial, E. Roquain (2017), arxiv:1703.02307

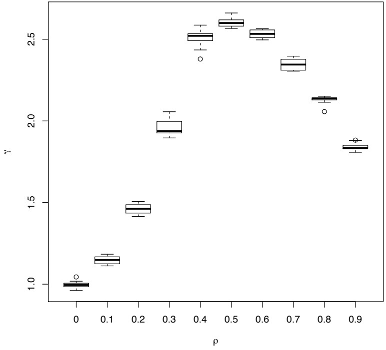
Permutation testing is justified in some applications, including:

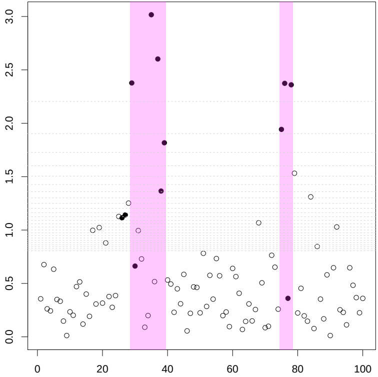
- differential expression analyses
- GWAS with discrete (case/control) or quantitative phenotype

(restriction: the reference thresholds t must be deterministic)

JER control with λ adjustment for the linear kernel

Example under positive dependency (Gaussian equi-correlation)

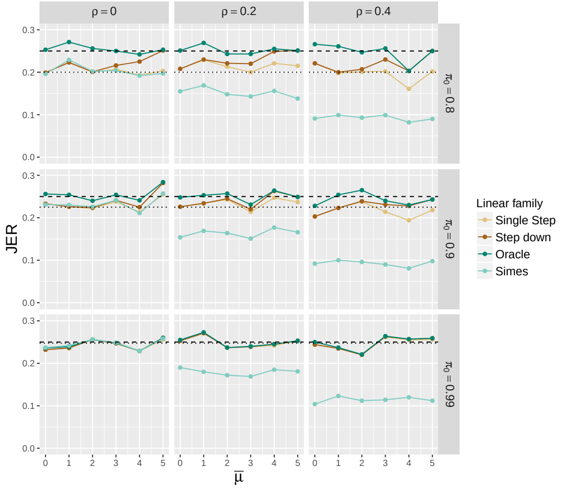




With probability $\geq 1 - \alpha = 75\%$:

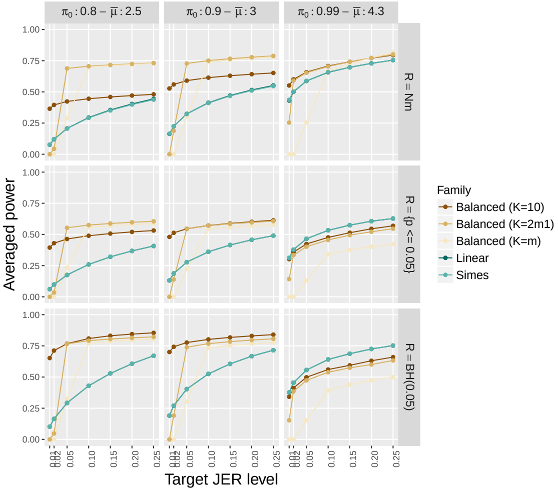
$t_k(\alpha)$	Lower bound on $\lvert R \cap \mathcal{H}_1 \rvert$
$\alpha k/m$	$\lvert R \cap \mathcal{H}_1 \rvert \geq 2$ and $\lvert R' \cap \mathcal{H}_1 \rvert \geq 1$
$\alpha\lambda(\alpha)k/m$	$\lvert R \cap \mathcal{H}_1 \rvert \geq 3$ and $\lvert R' \cap \mathcal{H}_1 \rvert \geq 2$

JER control under Gaussian equi-correlation



- $X_i \sim \mathcal{N}(0,1)$ under H_0
- $X_i \sim \mathcal{N}(\bar{\mu},1)$ under H_1
- $\mathrm{cor}(X_i, X_j) = \rho$ for $i \neq j$
- $\alpha = 0.25$

Estimation power for under Gaussian equi-correlation



- $X_i \sim \mathcal{N}(0,1)$ under H_0
- $X_i \sim \mathcal{N}(\bar{\mu},1)$ under H_1
- $\mathrm{cor}(X_i, X_j) = \rho$ for $i \neq j$
- $\bar{\mu} = 2$
- Estimation power: $E(\overline{S}(\mathcal{H}_1))/m_1$

Conclusions

Summary

- JER: a new risk measure for multiple testing
- generalizes existing post hoc procedures
- can be used to build post hoc inference procedures

Results not discussed here

- Other choices for the kernel
- Step-down procedures
- Control of $\mathbb{P}(\forall k \in \{1, \dots, K\}, |R_k \cap \mathcal{H}_0| \leq \zeta_k)$
- Detection power: connection to "higher criticism" in a sparse setting

Ongoing/future works

- Applications to GWAS, differential expression and neuro-imaging
- Structured rejection sets: algorithms and statistical results
- Software and visualization tools

Thanks

- Etienne Roquain, Paris 6 University, France
- Gilles Blanchard, Potsdam University, Germany

Reference

G. Blanchard, P. Neuvial, E. Roquain (2017), *Post hoc inference via joint family-wise error rate control* Arxiv preprint:1703.02307

Funding

- CNRS PEPS FaSciDo (2015-2016)
- ANR project JCJC SansSouci (2016-2019)

We are hiring a postdoc!