

Tests multiples pour données discrètes

Guillermo Durand

Joint works with Romain Périer, Gilles Blanchard, Sebastian Döhler, Florian Junge, Etienne Roquain

Université Paris-Saclay, CNRS, Inria, Laboratoire de Mathématiques d'Orsay

Séminaire de Statistique et Optimisation, Institut Mathématique de Toulouse



2026-05-12

Table of contents

1. Usual MT setting
2. Discrete heterogeneous setting
3. New procedures
4. Post hoc confidence envelopes

General multiple testing setting

- ▶ Random data $X : (\Omega, \mathcal{T}, \mathbb{P}) \rightarrow (\mathcal{X}, \Xi)$ with unknown distribution $\mu \in \mathfrak{F}$ a family of distributions
- ▶ m null hypotheses $H_{0,i} \subset \mathfrak{F}$ about μ
- ▶ $\mathcal{H}_0 = \mathcal{H}_0(\mu) = \{i : \mu \in H_{0,i}\}$: $i \in \mathcal{H}_0 \Leftrightarrow \mu \in H_{0,i} \Leftrightarrow H_{0,i}$ is true
- ▶ m p -values $p_i = p_i(X)$ such that $p_i \succeq \mathcal{U}([0, 1])$ if $i \in \mathcal{H}_0$
- ▶ Notation: for any subset of hypotheses $S \subseteq \mathbb{N}_m$: $V(S) = |S \cap \mathcal{H}_0| =$ false discoveries in S

Classical multiple testing theory

- ▶ Form a rejection procedure $R : \mathcal{X} \rightarrow \mathcal{P}(\mathbb{N}_m)$ with a statistical guarantee on $V(R(X))$ no matter μ
- ▶ $\text{FWER}(R) = \mathbb{P}(V(R(X)) > 0)$
 - ▶ Controlled by the famous Bonferroni procedure:
$$R_{\text{Bonf}}(X) = \{i : p_i(X) \leq \frac{\alpha}{m}\}.$$
- ▶ FWER control too stringent for applications $\Rightarrow \text{FDP}(R, X) = \frac{V(R(X))}{|R(X)| \vee 1}$ and $\text{FDR}(R) = \mathbb{E}[\text{FDP}(R, X)]$ [Benjamini and Hochberg (1995)]
 - ▶ FDP or FDR control \Rightarrow allow for some false positives but in controlled proportion
 - ▶ Controlled under wPRDS (or independence) by the Benjamini-Hochberg procedure [Benjamini and Yekutieli (2001)]

Benjamini-Hochberg procedure

- ▶ Order the p -values $p_{(1)} \leq \dots \leq p_{(m)}$, let $p_{(0)} = 0$
- ▶ Let $\hat{k}_{\text{BH}} = \max \left\{ k : p_{(k)}(X) \leq \frac{\alpha k}{m} \right\}$ and
 $R_{\text{BH}}(X) = \left\{ i : p_i(X) \leq \frac{\alpha \hat{k}_{\text{BH}}}{m} \right\}$.

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

- ▶ If the p -values satisfy the wPRDS dependence condition, then

$$\forall \alpha \in (0, 1), \forall \mu \in \mathfrak{F}, \text{FDR}(R_{\text{BH}}) \leq \alpha \frac{m_0}{m} \leq \alpha,$$

where $m_0 = |\mathcal{H}_0|$

- ▶ If the p -values are independent and if $p_i \sim \mathcal{U}([0, 1])$ for all $i \in \mathcal{H}_0$,

$$\forall \alpha \in (0, 1), \forall \mu \in \mathfrak{F}, \text{FDR}(R_{\text{BH}}) = \alpha \frac{m_0}{m} \leq \alpha$$

Take-home message: the uniform case is the “worst case” for which BH is calibrated \Rightarrow BH is too conservative for super-uniform p -values

Table of contents

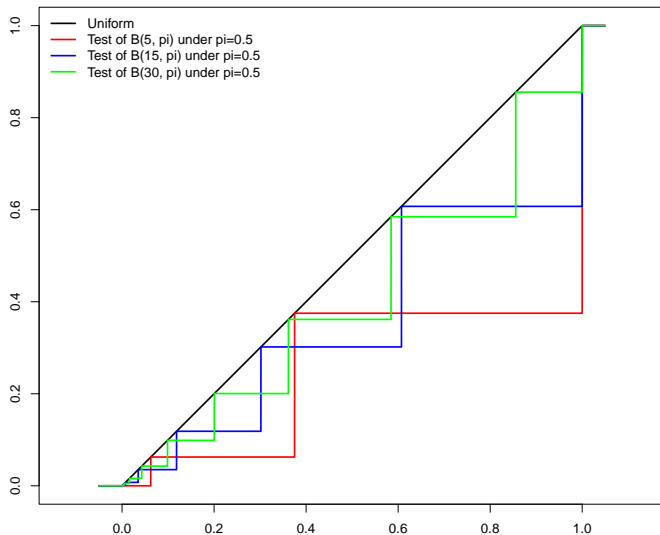
1. Usual MT setting
2. Discrete heterogeneous setting
3. New procedures
4. Post hoc confidence envelopes

Discrete setting

- ▶ As soon as data is discrete, e.g. count data
- ▶ p -value distribution is also discrete but still super-uniform
- ▶ Example 1: binomial distributions
 - ▶ $\mathfrak{F} = \left\{ \bigotimes_{j=1}^m \mathcal{B}(n_j, \pi_j) : \forall j \in \mathbb{N}_m, \pi_j \in [0, 1] \right\}$ for some n_1, \dots, n_m
 - ▶ $H_{0,i} = \left\{ \bigotimes_{j=1}^m \mathcal{B}(n_j, \pi_j) \in \mathfrak{F} : \pi_i = \pi_{i,0} \right\}$ for some $\pi_{1,0}, \dots, \pi_{m,0}$
 - ▶ $p_i(\cdot) : x \mapsto \mathbb{P}(N_i \leq x_i)$ or $p_i(\cdot) : x \mapsto \mathbb{P}(N_i \geq x_i)$ for one-sided tests, where $N_i \sim \mathcal{B}(n_i, \pi_{i,0})$
 - ▶ $p_i(\cdot) : x \mapsto \mathbb{P}(f_i(N_i) \leq f_i(x_i))$ for two-sided tests, where $f_i =$ pmf of $\mathcal{B}(n_i, \pi_{i,0})$

Discrete setting

CDFs of discrete and uniform p-values



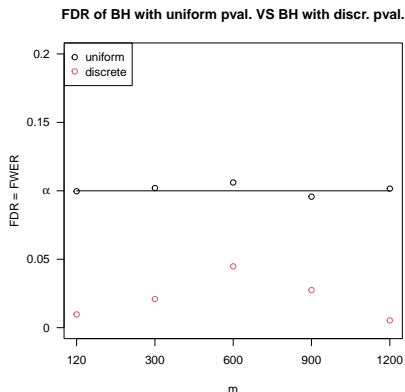
Discrete setting

- ▶ Example 2: Fisher's exact test for 2×2 contingency tables
 - ▶ to test independence or homogeneity of binomials, by conditioning on the marginal counts:
 - ▶ $(X_1, Y_1), \dots, (X_n, Y_n)$ iid $\sim \text{Cat}(\pi_{i,j})_{\substack{1 \leq i \leq 2, \\ 1 \leq j \leq 2}}$, if $X_1 \perp Y_1$ then
$$\mathcal{L}(N_{11} | N_{1\cdot}, N_{\cdot 1}) = \mathcal{H}(n, N_{1\cdot}, N_{\cdot 1})$$
 - ▶ (X_1, \dots, X_{n_X}) iid $\sim \mathcal{B}(\pi_X)$, (Y_1, \dots, Y_{n_Y}) iid $\sim \mathcal{B}(\pi_Y)$, indep samples, if $\pi_X = \pi_Y$ then $\mathcal{L}(N_{1X} | N_{1\cdot}) = \mathcal{H}(n_X + n_Y, N_{1\cdot}, n_X)$
 - ▶ e.g., study association between amnesia and some drugs in a database of reported adverse events from the Medicines and Healthcare products Regulatory Agency in the UK [Heller and Gur (2014)]
- ▶ Other examples exist like Poisson variables modelling read counts of different DNA bases in next-generation sequencing [Jiménez-Otero, Uña-Álvarez, and Pardo-Fernández (2019)]

Discrete setting

The problem

- ▶ As discrete p -values are larger than uniforms, classic (=BH) thresholds are too low \implies too conservative \implies loss of power
- ▶ BH under full null, $m/3$ 2-sided p -values derived from $\mathcal{H}(60, 5, 30)$, $m/3$ from $\mathcal{H}(60, 12, 30)$, $m/3$ from $\mathcal{H}(60, 21, 30)$
- ▶ MC estimation of the FDR with 10^4 replications



Discrete setting

The idea

- ▶ Under the null, the cdf of p_i is known and can be plugged in a new procedure
- ▶ If the p -values are also heterogeneous, such procedure could regain power
- ▶ Formalism for the remainder:
 - 1 $\forall i \in \mathbb{N}_m, \exists \mathcal{A}_i$ finite such that $\mathbb{P}(p_i \in \mathcal{A}_i) = 1, \forall \mu \in H_{0,i}$
 - 2 $\forall \mu \in H_{0,i}$, the cdf of p_i is equal to F_i (or, more generally, upper-bounded by F_i)
 - 3 let $\mathcal{A} = \{0\} \cup \bigcup_{i=1}^m \mathcal{A}_i$

Table of contents

1. Usual MT setting
2. Discrete heterogeneous setting
3. New procedures
4. Post hoc confidence envelopes

Step-up procedures

and step-down procedures

- ▶ Given a nondecreasing nonnegative sequence $\tau = (\tau_1, \dots, \tau_m)$, the respective step-up and step-down procedures associated with τ are:

$$\begin{aligned}R^{\text{SU}}(\tau) &= \{i \in \mathbb{N}_m : p_i \leq \tau_{\hat{k}^{\text{SU}}}\} \\R^{\text{SD}}(\tau) &= \{i \in \mathbb{N}_m : p_i \leq \tau_{\hat{k}^{\text{SD}}}\}\end{aligned}$$

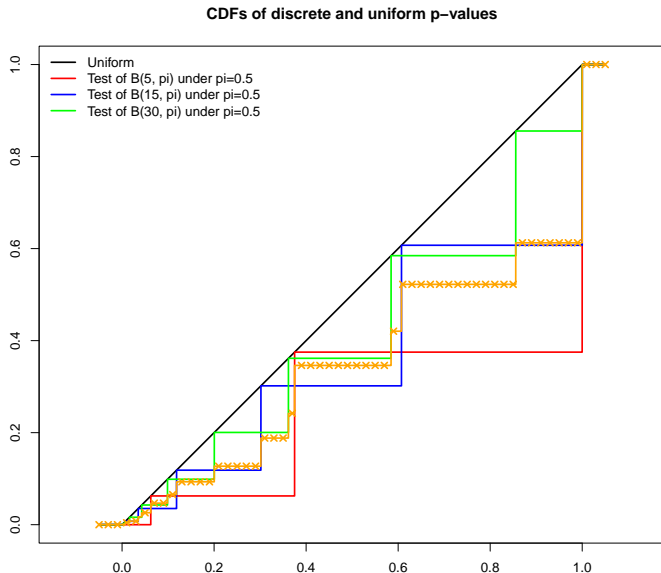
with

$$\begin{aligned}\hat{k}^{\text{SU}} &= \max \{0 \leq k \leq m : p_{(k)} \leq \tau_k\} \\ \hat{k}^{\text{SD}} &= \max \left\{ 0 \leq k \leq m : \forall k' \leq k, p_{(k')} \leq \tau_{k'} \right\}\end{aligned}$$

- ▶ with $\tau_0 = \tau_1$ by convention
- ▶ The τ_k are called the critical values
- ▶ $R_{\text{BH}} = R^{\text{SU}}(\tau)$ with $\tau_k = \alpha \frac{k}{m}$

FDR control with discrete p -values

Heyse procedure



FDR control with discrete p -values

Heyse procedure

- ▶ \Rightarrow idea: “invert” the mean c.d.f. at $\alpha \frac{k}{m}$ and apply a SU procedure [Heyse (2011)]
- ▶ Let $\bar{F}(t) = \frac{1}{m} \sum_{i=1}^m F_i(t)$
- ▶ Let $\tau_k = \max \left\{ t \in \mathcal{A} : \bar{F}(t) \leq \alpha \frac{k}{m} \right\}$
- ▶ $R^{\text{Heyse}} = R^{\text{SU}}(\tau)$
- ▶ BH is also the SU procedure with $\xi_k = \max \left\{ t \in \mathcal{A} : t \leq \alpha \frac{k}{m} \right\}$ (effective critical values), $\bar{F}(\xi_k) \leq \xi_k \leq \alpha \frac{k}{m}$ so $\tau_k \geq \xi_k$: Heyse less conservative than BH, only with heterogeneity though: if $F_i = \bar{F}$ and $F_i(t) = t, \forall t \in \mathcal{A}_i = \mathcal{A}$ then $\bar{F}(t) = t$ for all $t \in \mathcal{A}$ and $\tau_k = \xi_k$
- ▶ Problem: R^{Heyse} doesn't control the FDR! [Döhler, D., and Roquain (2018)]

FDR control with discrete p -values

- ▶ Heyse *almost* works though, it works up to a small rescaling factor
- ▶ Let $\tau_m = \max \left\{ t \in \mathcal{A} : \frac{1}{m} \sum_{i=1}^m \frac{F_i(t)}{1-F_i(t)} \leq \alpha \right\}$
- ▶ For $k < m$, let $\tau_k = \max \left\{ t \in \mathcal{A} : t \leq \tau_m, \sum_{i=1}^m \frac{F_i(t)}{1-F_i(\tau_m)} \leq \alpha k \right\}$
- ▶ Let $R^{\text{HSU}} = R^{\text{SU}}(\tau)$
- ▶ Can be more conservative than BH but not that much, and in practice isn't

Corollary [Döhler, D., and Roquain (2018)]

Assume that for all $\mu \in \mathfrak{F}$, the $(p_i)_{i \in \mathcal{H}_0}$ are independent, and they are independent from the $(p_i)_{i \in \mathcal{H}_1}$.

Then for all $\mu \in \mathfrak{F}$,

$$\text{FDR} \left(R^{\text{HSU}} \right) \leq \alpha$$

FDR control with discrete p -values

- ▶ We can do even better by implicit adaptivity to m_0 :

- ▶ For $k < m$, let

$$\tau_k = \max \left\{ t \in \mathcal{A} : t \leq \tau_m, \max_{\substack{A \subseteq \mathcal{N}_m \\ |A|=m-k+1}} \sum_{i \in A} \frac{F_i(t)}{1 - F_i(\tau_m)} \leq \alpha k \right\}$$

- ▶ Idea: if k “good” rejections, $m_0 \leq m - k + 1$ so the control is only needed for the worst case with $m - k + 1$ kept null hypotheses
- ▶ Let $R^{\text{AHSU}} = R^{\text{SU}}(\tau)$
- ▶ Less conservative than HSU because of larger critical values

Corollary [Döhler, D., and Roquain (2018)]

Assume that for all $\mu \in \mathfrak{F}$, the $(p_i)_{i \in \mathcal{H}_0}$ are independent, and they are independent from the $(p_i)_{i \in \mathcal{H}_1}$.

Then for all $\mu \in \mathfrak{F}$,

$$\text{FDR} \left(R^{\text{AHSU}} \right) \leq \alpha$$

FDR control with discrete p -values

- ▶ Both FDR controls come from the same bound

Theorem [Döhler, D., and Roquain (2018)]

Assume that for all $\mu \in \mathfrak{F}$, the $(p_i)_{i \in \mathcal{H}_0}$ are independent, and they are independent from the $(p_i)_{i \in \mathcal{H}_1}$. Let any critical value sequence τ with $F_i(\tau_m) < 1$ for all $i \in \mathbb{N}_m$.

Then for all $\mu \in \mathfrak{F}$,

$$\text{FDR} \left(R^{\text{SU}}(\tau) \right) \leq \max_{1 \leq k \leq m} \max_{\substack{A \subseteq \mathbb{N}_m \\ |A|=m-k+1}} \frac{1}{k} \sum_{i \in A} \frac{F_i(\tau_k)}{1 - F_i(\tau_m)} \quad (1)$$

- ▶ This bound is very general and can be applied to other (non-discrete, non-super-uniform) settings e.g. weighting
- ▶ Freedom in the choice of τ_m , as soon as $\tau_m \in \left\{ t \in \mathcal{A} : \forall i \in \mathbb{N}_m, F_i(t) < 1, \max_{i \in \mathbb{N}_m} \frac{F_i(t)}{1 - F_i(t)} \leq \alpha m \right\}$
- ▶ \Rightarrow idea: tune it, ongoing exploration by Romain Périer

FDR control with discrete p -values

- ▶ HSD: SD with $\tau_k = \max \left\{ t \in \mathcal{A} : \sum_{i=1}^m \frac{F_i(t)}{1-F_i(t)} \leq \alpha k \right\}$
- ▶ AHSD: SD with $\tau_k = \max \left\{ t \in \mathcal{A} : \max_{\substack{A \subseteq \mathbb{N}_m \\ |A|=m-k+1}} \sum_{i \in A} \frac{F_i(t)}{1-F_i(t)} \leq \alpha k \right\}$
- ▶ Higher critical values than HSU and AHSU, but SD: no procedure generally better than the other
- ▶ No singular choice of τ_m here

Theorem [Döhler, D., and Roquain (2018)]

Assume that for all $\mu \in \mathfrak{F}$, the $(p_i)_{i \in \mathcal{H}_0}$ are independent, and they are independent from the $(p_i)_{i \in \mathcal{H}_1}$. Let any critical value sequence τ with $F_i(\tau_m) < 1$ for all $i \in \mathbb{N}_m$.

Then for all $\mu \in \mathfrak{F}$,

$$\text{FDR} \left(R^{\text{SD}}(\tau) \right) \leq \max_{1 \leq k \leq m} \max_{\substack{A \subseteq \mathbb{N}_m \\ |A|=m-k+1}} \frac{1}{k} \sum_{i \in A} \frac{F_i(\tau_k)}{1-F_i(\tau_k)}$$

New FDR bounds

Computational trick [D., Junge, Döhler, and Roquain (2019)]

- ▶ No need to compute τ_k to execute HSU, AHSU (except τ_m), HSD, AHSD
- ▶ For example, for AHSU:

$$\forall k < m, p_{(k)} \leq \tau_k \iff p_{(k)} \leq \tau_m \text{ and } \max_{\substack{A \subseteq \mathbb{N}_m \\ |A|=m-k+1}} \frac{1}{m} \sum_{i \in A} \frac{F_i(p_{(k)})}{1 - F_i(\tau_m)} \leq \alpha \frac{k}{m}$$

Numerical experiment

► Fisher's exact test of homogeneity, amnesia dataset

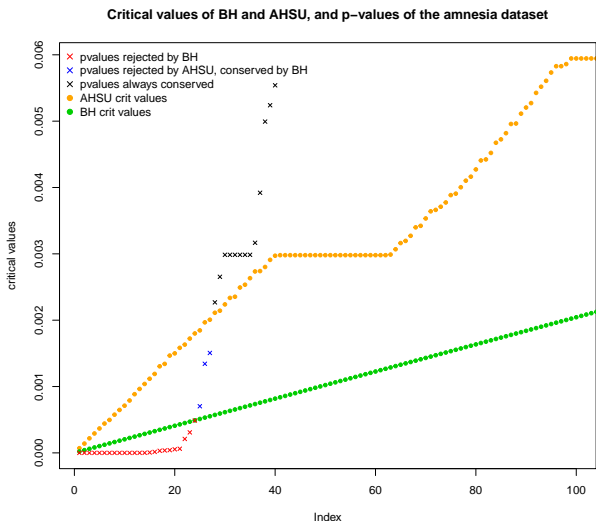


Table of contents

1. Usual MT setting
2. Discrete heterogeneous setting
3. New procedures
4. Post hoc confidence envelopes

Motivation: exploratory analysis in multiple testing

Exploratory analysis: searching interesting hypotheses that will be cautiously investigated after.

Desired properties [Goeman and Solari (2011)]:

- ▶ Mildness: allows some false positives
- ▶ Flexibility: the procedure does not prescribe, but advise
- ▶ Post hoc: take decisions on the procedure after seeing the data

[Goeman and Solari (2011)]

This **reverses the traditional roles** of the user and procedure in multiple testing. Rather than [...] to let the user choose the quality criterion, and to let the procedure return the collection of rejected hypotheses, the **user chooses the collection of rejected hypotheses freely**, and the multiple testing procedure returns the **associated quality criterion**.

Post hoc and replication crisis

Post hoc done wrong: p -hacking

- ▶ Pre-selecting variables that seem significant, exclude others
- ▶ Theoretical results no longer hold because the selection step is random
- ▶ Example: selecting the 1000 smallest p -values in a genetic study with 10^6 variants

- ▶ p -hacking may be one of the causes of the replication crisis (many published results non reproducible)

⇒ need for exploratory analysis MT procedures with the above properties

Post hoc inference

Or simultaneous inference

Upper bounds for any set of selected variables

A (post hoc) confidence envelope is a random function

$$\widehat{V} : \mathcal{P}(\mathbb{N}_m) \rightarrow \llbracket 0, m \rrbracket$$

such that:

$$\mathbb{P} \left(\forall S \subset \mathbb{N}_m, V(S) \leq \widehat{V}(S) \right) \geq 1 - \alpha. \quad (2)$$

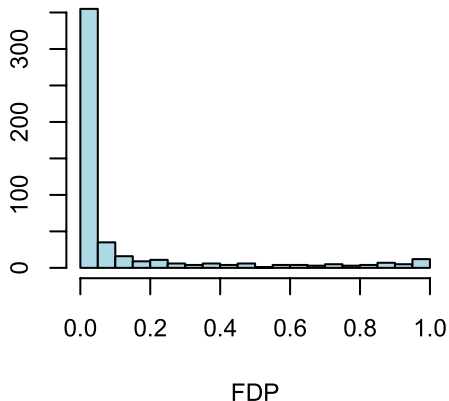
\widehat{V} depends on (X, α) and (2) has to be true for all μ .

- ▶ Hence for any selected \widehat{S} , $\mathbb{P} \left(V(\widehat{S}) \leq \widehat{V}(\widehat{S}) \right) \geq 1 - \alpha$ holds
- ▶ Also an FDP bound: $\mathbb{P} \left(\forall S \subset \mathbb{N}_m, \text{FDP}(S) \leq \widehat{V}(S)/|S| \right) \geq 1 - \alpha$
- ▶ \implies allows construction of sets with bounded FDP (FDX control)
- ▶ First proposed by [\[Genovese and Wasserman \(2006\)\]](#)
- ▶ A guarantee over any selected set instead of a rejected set, advise some \widehat{S} instead of prescribe one R : the MT paradigm is reversed

FDX control is more relevant than FDR control

[Bogdan, Berg, Sabatti, et al. (2015)]

- ▶ A bad example of BH with positive dependence (hence FDR control)



BNR technology

[Blanchard, Neuvial, and Roquain (2020)]

Key concept: reference family

- ▶ $\mathfrak{R} = (R_k, \zeta_k)_{k \in \mathcal{K}}$ (random) such that Joint Error Rate (JER) control:

$$\text{JER}(\mathfrak{R}) = \mathbb{P}(\exists k, |R_k \cap \mathcal{H}_0| > \zeta_k) \leq \alpha. \quad (3)$$

\mathfrak{R} depends on X and (3) has to be true no matter $\mathcal{L}(X)$.

- ▶ Conversely, $\mathbb{P}(\forall k, V(R_k) \leq \zeta_k) \geq 1 - \alpha$
- ▶ Confidence bound only on the members of \mathfrak{R}
- ▶ \implies Derivation of a global confidence bound by interpolation

BNR technology

[Blanchard, Neuvial, and Roquain (2020)]

Idea: we get the following info on \mathcal{H}_0 :

$$\mathcal{H}_0 \in \mathcal{A}(\mathfrak{R}) = \{A \in \mathcal{P}(\mathbb{N}_m) : \forall k, |R_k \cap A| \leq \zeta_k\}.$$

Two different bounds

- ▶ $V_{\mathfrak{R}}^*(S) = \max \{|S \cap A| : A \in \mathcal{A}(\mathfrak{R})\}$ optimal but hard to compute
- ▶ $\overline{V}_{\mathfrak{R}}(S) = \min_k (\zeta_k + |S \setminus R_k|) \wedge |S|$ easier to compute, $\geq V_{\mathfrak{R}}^*(S)$
- ▶ Property: if (R_k) is nested, then $\overline{V}_{\mathfrak{R}} = V_{\mathfrak{R}}^*$

BNR technology

Practical construction

- ▶ In [Blanchard, Neuvial, and Roquain (2020)], constrain $\zeta_k = k - 1$, $R_k = \{i \in \mathbb{N}_m, p_i \leq t_k\}$, $k \in \mathbb{N}_m$, $t_k \nearrow$ and search for $(t_k)_{1 \leq k \leq m}$ such that JER control holds
- ▶ \Rightarrow JER control becomes “simultaneous k -FWER control”:
$$\text{JER}(\mathfrak{R}) = \mathbb{P}(\exists k, |R_k \cap \mathcal{H}_0| \geq k)$$
- ▶ $t_k \nearrow \Rightarrow$ nested R_k 's
- ▶ In the end, $V_{\mathfrak{R}}^*(S) = \min_{1 \leq k \leq m} (k - 1 + \sum_{i \in S} \mathbb{1}_{\{p_i > t_k\}}) \wedge |S|$

BNR technology

Practical construction, Simes inequality

Theorem (Simes inequality [Simes (1986)])

Assume that for all $\mu \in \mathfrak{F}$, the $(p_i)_{i \in \mathcal{H}_0}$ are independent or wPRDS.
Then for all $\mu \in \mathfrak{F}$,

$$\mathbb{P} \left(\exists i \leq m_0, p_{(i:\mathcal{H}_0)} \leq \frac{\alpha i}{m_0} \right) \leq \alpha,$$

where $p_{(1:\mathcal{H}_0)} \leq \dots \leq p_{(m_0:\mathcal{H}_0)}$ is the ordering of $(p_i)_{i \in \mathcal{H}_0}$.

- ▶ Simple proof given the FDR control of R_0^{BH} : BH applied to $(p_i)_{i \in \mathcal{H}_0}$
- ▶ Notice that FWER = FDR when all null hypotheses are true
- ▶ Hence

$$\begin{aligned} \mathbb{P} \left(\exists i \leq m_0, p_{(i:\mathcal{H}_0)} \leq \frac{\alpha i}{m_0} \right) &= \text{FWER}(R_0^{\text{BH}}) \\ &= \text{FDR}(R_0^{\text{BH}}) \\ &\leq \alpha \end{aligned}$$

BNR technology

Practical construction, Simes inequality

- ▶ Consequence: $\forall \mu \in \mathfrak{F}, \mathbb{P} \left(\exists i \leq m_0, p_{(i:\mathcal{H}_0)} \leq \frac{\alpha i}{m} \right) \leq \alpha$ (because $m \geq m_0$)
- ▶ $\implies t_k = \frac{\alpha k}{m}$ induces JER control
- ▶ Proof:

$$\begin{aligned} \exists k \leq K : |R_k \cap \mathcal{H}_0| \geq k &\Leftrightarrow \exists k \leq m_0 : \left| \left\{ i \in \mathbb{N}_m : p_i \leq \frac{\alpha k}{m} \right\} \cap \mathcal{H}_0 \right| \geq k \\ &\Leftrightarrow \exists k \leq m_0 : \left| \left\{ i \in \mathcal{H}_0 : p_i \leq \frac{\alpha k}{m} \right\} \right| \geq k \\ &\Leftrightarrow \exists k \leq m_0 : p_{(k:\mathcal{H}_0)} \leq \frac{\alpha k}{m} \end{aligned}$$

BNR technology

Practical construction with heterogeneity

- ▶ Core idea: we can derive a similar “Simes-like” inequality thanks to the FDR control of HSU and AHSU
- ▶ And then build $\mathfrak{R} = (\{i \in \mathbb{N}_m, p_i \leq \tau_k\}, k - 1)_{k \in \mathbb{N}_m}$ with the τ_k critical values of HSU or AHSU
- ▶ Going from the “full null” case to the general case is slightly more subtle though, the “ $m \geq m_0$ ” argument is still key but not that straightforward

BNR technology

Practical construction with heterogeneity

Theorem (Heterogeneous Simes inequality) [Périer, Blanchard, Döhler, et al. (2026)]

Assume that for all $\mu \in \mathfrak{F}$, the $(p_i)_{i \in \mathcal{H}_0}$ are independent. Let (τ_k) a sequence of critical values such that

$$\max_{1 \leq k \leq m} \max_{\substack{A \subseteq \mathbb{N}_m \\ |A|=m-k+1}} \frac{1}{k} \sum_{i \in A} \frac{F_i(\tau_k)}{1 - F_i(\tau_m)} \leq \alpha.$$

Then for all $\mu \in \mathfrak{F}$,

$$\mathbb{P} \left(\exists i \leq m_0, p_{(i:\mathcal{H}_0)} \leq \tau_i \right) \leq \alpha.$$

- ▶ we recognize the bound in (1)

BNR technology

Practical construction with heterogeneity

- ▶ Proof: note that $\frac{1}{1-F_i(\tau_{m_0})} \leq \frac{1}{1-F_i(\tau_m)}$, $\forall i$, then apply the SU bound (1) to the full null SU procedure:

$$\begin{aligned}\mathbb{P}\left(\exists i \leq m_0, p_{(i:\mathcal{H}_0)} \leq \tau_i\right) &= \text{FWER}\left(R_0^{\text{SU}}\left((\tau_k)_{k \leq m_0}\right)\right) \\ &= \text{FDR}\left(R_0^{\text{SU}}\left((\tau_k)_{k \leq m_0}\right)\right) \\ &\leq \max_{1 \leq k \leq m_0} \max_{\substack{A \subseteq \mathcal{H}_0 \\ |A|=m_0-k+1}} \frac{1}{k} \sum_{i \in A} \frac{F_i(\tau_k)}{1 - F_i(\tau_{m_0})} \\ &\leq \max_{1 \leq k \leq m} \max_{\substack{A \subseteq \mathcal{N}_m \\ |A|=m-k+1}} \frac{1}{k} \sum_{i \in A} \frac{F_i(\tau_k)}{1 - F_i(\tau_m)} \\ &\leq \alpha\end{aligned}$$

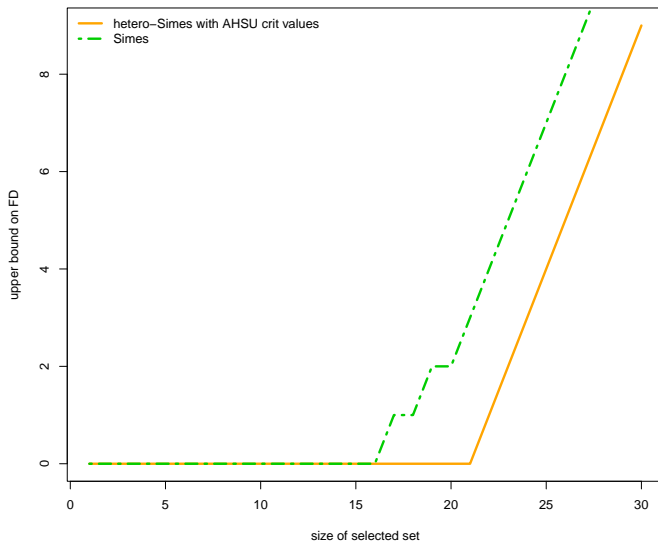
Application to the amnesia dataset

- ▶ Plot of $t \mapsto V_{\mathfrak{A}}^*(S_t)$ where $S_t =$ the indexes of the t smallest p -values
- ▶ Smaller is better
- ▶ Fast computation of $(V_{\mathfrak{A}}^*(S_t))_{t \in \mathbb{N}_m}$ with Algorithm 1 of [Enjalbert-Courrech and Neuvial (2022)]

Numerical experiment

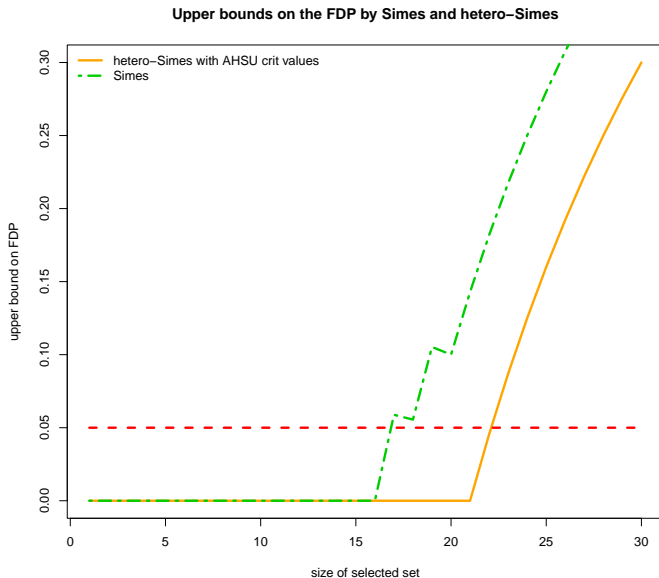
Amnesia dataset

Upper bounds on the number of false discoveries by Simes and hetero-Simes



Numerical experiment

Amnesia dataset



Adaptivity to m_0 by plug-in

- ▶ We can go further by estimating m_0 with our bound and plugging it in
- ▶ A step-down refinement akin to the Holm-Bonferroni procedure
- ① Refine (1) with

$$\text{FDR} \left(R^{\text{SU}}(\tau) \right) \leq \max_{1 \leq k \leq m} \max_{\substack{A \subseteq \mathbb{N}_m \\ |A|=m-k+1}} \frac{1}{k} \sum_{i \in A \cap \mathcal{H}_0} \frac{F_i(\tau_k)}{1 - F_i(\tau_m)} \quad (4)$$

- ② Define “oracle AHSU” critical values with $\tilde{\tau}_m = \tau_m$ and

$$\tilde{\tau}_k = \max \left\{ t \in \mathcal{A} : t \leq \tilde{\tau}_m, \max_{\substack{A \subseteq \mathcal{H}_0 \\ |A|=(m-k+1) \wedge m_0}} \sum_{i \in A} \frac{F_i(t)}{1 - F_i(\tilde{\tau}_m)} \leq \alpha k \right\}$$

Adaptivity to m_0 by plug-in

- 3 Let $B = \{\forall i \leq m_0, p_{(i:\mathcal{H}_0)} > \tilde{\tau}_i\}$, then

$$\mathbb{P}(B^c) = \mathbb{P}(\exists i \leq m_0, p_{(i:\mathcal{H}_0)} \leq \tilde{\tau}_i) \leq \alpha \text{ by (4)}$$
- 4 On B , we also have $\forall i \leq m_0, p_{(i:\mathcal{H}_0)} > \tau_i$ because $\tilde{\tau}_k \geq \tau_k \forall k$, hence

$$m_0 \leq V_{\mathfrak{R}}^*(\mathbb{N}_m) = \hat{m}_0$$
- 5 Define adaptive AHSU critical values $\hat{\tau}_m = \tau_m$ and

$$\hat{\tau}_k = \max \left\{ t \in \mathcal{A} : t \leq \hat{\tau}_m, \max_{\substack{A \subseteq \mathbb{N}_m \\ |A|=(m-k+1) \wedge \hat{m}_0}} \sum_{i \in A} \frac{F_i(t)}{1 - F_i(\hat{\tau}_m)} \leq \alpha k \right\}$$
- 6 On B , we also have $\tilde{\tau}_k \geq \hat{\tau}_k \forall k$ (from $m_0 \leq \hat{m}_0$ and the def of $\tau_k, \hat{\tau}_k$)
- 7 Finally, on B , $\forall i \leq m_0, p_{(i:\mathcal{H}_0)} > \hat{\tau}_i$
 - ▶ JER control with $(\{i \in \mathbb{N}_m, p_i \leq \hat{\tau}_k\}, k-1)_{k \in \mathbb{N}_m}$ and $\hat{\tau}_k \geq \tau_k \forall k$
 - ▶ Get a new \hat{m}_0 and start again, everything stems from $\mathbb{P}(B^c) \leq \alpha$

Conclusion

- ▶ (Many) new procedures controlling the FDR under heterogeneity, well suited to discrete tests
 - ▶ Also, new FD confidence envelopes with adaptive refinements
 - ▶ Challenges: independence assumption [Döhler (2018)]
- ▶ One published paper [Döhler, D., and Roquain (2018)] for FDR control
 - ▶ A suite of R packages [Kihn, Döhler, and Junge (2025), Junge and Kihn (2024), and Döhler, Junge, and D. (2024)] for discrete tests and FDR control
 - ▶ One preprint just out about heterogeneous confidence envelopes [Périer, Blanchard, Döhler, et al. (2026)]

Bibliography I



Benjamini, Yoav and Yosef Hochberg (1995). “Controlling the false discovery rate: a practical and powerful approach to multiple testing”. In: *J. Roy. Statist. Soc. Ser. B* 57.1, pp. 289–300. ISSN: 0035-9246. URL: <https://www.jstor.org/stable/2346101>.



Benjamini, Yoav and Daniel Yekutieli (2001). “The control of the false discovery rate in multiple testing under dependency”. In: *Ann. Statist.* 29.4, pp. 1165–1188. ISSN: 0090-5364. DOI: 10.1214/aos/1013699998. URL: <https://doi.org/10.1214/aos/1013699998>.



Blanchard, Gilles, Pierre Neuvial, and Etienne Roquain (2020). “Post hoc confidence bounds on false positives using reference families”. In: *Ann. Statist.* 48.3, pp. 1281–1303. ISSN: 0090-5364. DOI: 10.1214/19-AOS1847. URL: <https://doi.org/10.1214/19-AOS1847>.



Bogdan, Małgorzata et al. (2015). “SLOPE—adaptive variable selection via convex optimization”. In: *Ann. Appl. Stat.* 9.3, pp. 1103–1140. ISSN: 1932-6157,1941-7330. DOI: 10.1214/15-AOAS842. URL: <https://doi.org/10.1214/15-AOAS842>.



D., Guillermo et al. (2019). *DiscreteFDR: An R package for controlling the false discovery rate for discrete test statistics*. arXiv: 1904.02054 [stat.CO]. URL: <https://arxiv.org/abs/1904.02054>.



Döhler, Sebastian (2018). “A discrete modification of the Benjamini-Yekutieli procedure”. In: *Econom. Stat.* 5, pp. 137–147. ISSN: 2468-0389,2452-3062. DOI: 10.1016/j.ecosta.2016.12.002. URL: <https://doi.org/10.1016/j.ecosta.2016.12.002>.

Bibliography II



Döhler, Sebastian, Guillermo D., and Etienne Roquain (2018). “New FDR bounds for discrete and heterogeneous tests”. In: *Electron. J. Stat.* 12.1, pp. 1867–1900. DOI: 10.1214/18-EJS1441. URL: <https://doi.org/10.1214/18-EJS1441>.



Döhler, Sebastian, Florian Junge, and Guillermo D. (2024). *DiscreteFDR: FDR Based Multiple Testing Procedures with Adaptation for Discrete Tests*. R package version 2.1.0. DOI: 10.32614/CRAN.package.DiscreteFDR. URL: <https://CRAN.R-project.org/package=DiscreteFDR>.



Enjalbert-Courrech, Nicolas and Pierre Neuvial (Oct. 2022). “Powerful and interpretable control of false discoveries in two-group differential expression studies”. In: *Bioinformatics* 38.23, pp. 5214–5221. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btac693. eprint: https://academic.oup.com/bioinformatics/article-pdf/38/23/5214/47465958/btac693_supplementary_data.pdf. URL: <https://doi.org/10.1093/bioinformatics/btac693>.



Genovese, Christopher R. and Larry Wasserman (2006). “Exceedance control of the false discovery proportion”. In: *J. Amer. Statist. Assoc.* 101.476, pp. 1408–1417. ISSN: 0162-1459. DOI: 10.1198/016214506000000339. URL: <https://doi.org/10.1198/016214506000000339>.



Goeman, Jelle J. and Aldo Solari (2011). “Multiple testing for exploratory research”. In: *Statist. Sci.* 26.4, pp. 584–597. ISSN: 0883-4237. DOI: 10.1214/11-STS356. URL: <https://doi.org/10.1214/11-STS356>.

Bibliography III



Heller, Ruth and Hadas Gur (2014). *False discovery rate controlling procedures for discrete tests*. arXiv: 1112.4627 [stat.ME]. URL: <https://arxiv.org/abs/1112.4627>.



Heyse, Joseph F (2011). "A false discovery rate procedure for categorical data". In: *Recent advances in biostatistics: False discovery rates, survival analysis, and related topics*. World Scientific, pp. 43–58.



Jiménez-Otero, Norman, Jacobo de Uña-Álvarez, and Juan Carlos Pardo-Fernández (2019). "Goodness-of-fit tests for disorder detection in NGS experiments". In: *Biometrical Journal* 61.2, pp. 424–441. DOI: <https://doi.org/10.1002/bimj.201700284>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/bimj.201700284>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.201700284>.



Junge, Florian and Christina Kihn (2024). *DiscreteTests: Vectorised Computation of P-Values and Their Supports for Several Discrete Statistical Tests*. R package version 0.2.1. DOI: [10.32614/CRAN.package.DiscreteTests](https://doi.org/10.32614/CRAN.package.DiscreteTests). URL: <https://CRAN.R-project.org/package=DiscreteTests>.



Kihn, Christina, Sebastian Döhler, and Florian Junge (2025). *DiscreteDatasets: Example Data Sets for Use with Discrete Statistical Tests*. R package version 0.1.2. DOI: [10.32614/CRAN.package.DiscreteDatasets](https://doi.org/10.32614/CRAN.package.DiscreteDatasets). URL: <https://CRAN.R-project.org/package=DiscreteDatasets>.



Périer, Romain et al. (2026). *Confidence envelopes for the false discoveries with heterogeneous data*. arXiv: 2604.11812 [math.ST]. URL: <https://arxiv.org/abs/2604.11812>.

Bibliography IV



Simes, R. J. (1986). “An improved Bonferroni procedure for multiple tests of significance”. In: *Biometrika* 73.3, pp. 751–754. ISSN: 0006-3444. DOI: [10.1093/biomet/73.3.751](https://doi.org/10.1093/biomet/73.3.751). URL: <https://doi.org/10.1093/biomet/73.3.751>.



Tarone, Robert E (1990). “A modified Bonferroni method for discrete data”. In: *Biometrics*, pp. 515–522.