



HAL
open science

Intrinsic Bounds and False Discovery Rate Control in Multiple Testing Problems

Pierre Neuvial

► **To cite this version:**

Pierre Neuvial. Intrinsic Bounds and False Discovery Rate Control in Multiple Testing Problems. 2010. hal-00460677v1

HAL Id: hal-00460677

<https://hal.science/hal-00460677v1>

Preprint submitted on 2 Mar 2010 (v1), last revised 19 Apr 2013 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Intrinsic Bounds and False Discovery Rate Control in Multiple Testing Problems

Pierre Neuvial

PIERRE.NEUVIAL@CURIE.FR

Laboratoire de Probabilités et Modèles Aléatoires

Université Paris VII-Denis Diderot

175 rue du Chevaleret, 75013 Paris, France

INSERM, U900, Paris, F-75248 France

Ecole des Mines de Paris, ParisTech, Fontainebleau, F-77300 France

Institut Curie, 26 rue d'Ulm, Paris cedex 05, F-75248 France

Editor:

This version: 2010-03-03 08:49

Abstract

When testing a large number of independent hypotheses, three different questions are of interest: are some hypotheses true alternatives? How many of them? Which of them? These questions give rise to a detection, an estimation, and a selection problem. Recent work demonstrates the existence of intrinsic bounds in these problems: detection and estimation boundaries in sparse location models, and criticality for the selection problem. We study consequences of such limitations in terms of power of False Discovery Rate (FDR) controlling procedures. FDR is the expected False Discovery Proportion (FDP), that is, the expected proportion of false rejections among all rejected hypotheses.

For the selection problem, we illustrate the connection between criticality and the regularity of the distribution of the test statistics, and discuss expected and observed consequences of criticality in terms of power of FDR controlling procedures, on both simulated and real data. For the problem of estimating the fraction of true null hypotheses, we make explicit connections between the parameters of the multiple testing problem and consistency and convergence rates of a broad class of non-parametric estimators, and prove that these convergence rates determine that of the FDP achieved by “plug-in” multiple testing procedures, which are incorporating such an estimator in order to yield tighter FDR control.

Keywords: Multiple testing, False Discovery Rate, Benjamini Hochberg’s procedure, power, criticality, proportion of true null hypotheses.

Contents

| | | |
|----------|--|----------|
| 1 | Introduction | 2 |
| 1.1 | Intrinsic bounds in multiple comparison problems | 3 |
| 1.2 | Estimation, selection, and power of FDR controlling procedures | 4 |
| 2 | Background and notation | 4 |
| 2.1 | Model | 4 |
| 2.2 | False Discovery Rate control | 6 |
| 2.3 | Criticality of the selection problem. | 7 |

| | | |
|----------|--|-----------|
| 3 | Criticality and distribution of the test statistics | 8 |
| 3.1 | Interpretation and illustration of criticality | 8 |
| 3.1.1 | Illustration in location models | 9 |
| 3.1.2 | Illustration for Student test statistics | 12 |
| 3.2 | Criticality and power of the BH95 procedure | 15 |
| 3.2.1 | Observed and expected power of the BH95 procedure in location models | 15 |
| 3.2.2 | Longitudinal sample size and criticality: theory and practice for the Student distribution | 16 |
| 4 | Estimation of π_0 | 19 |
| 4.1 | Consistency, purity and criticality | 19 |
| 4.2 | Asymptotic properties of non-parametric estimators of π_0 | 20 |
| 4.2.1 | Non-parametric estimators of π_0 with known convergence rates | 21 |
| 4.2.2 | Asymptotic properties and regularity near 1 | 21 |
| 4.2.3 | Kernel estimators | 22 |
| 4.3 | Convergence rate of plug-in procedures | 23 |
| 4.4 | Regularity of g_1 at 1 | 25 |
| 4.4.1 | Two-sided problems | 25 |
| 4.4.2 | One-sided problems | 26 |
| 4.4.3 | Practical implications | 28 |
| | Appendix: proofs | 28 |
| | A Proofs of section 2 | 29 |
| | B Proofs of section 3 | 29 |
| | C Proofs of section 4 | 33 |
| | References | 39 |

1. Introduction

Multiple simultaneous hypothesis testing has become a major issue for high-dimensional data analysis in a variety of fields, including non-parametric estimation by wavelet methods in image analysis, functional magnetic resonance imaging (fMRI) in medicine, source detection in astronomy, or DNA microarray analysis in genomics. Given a possibly large set of observations corresponding either to a null hypothesis \mathcal{H}_0 , or an alternative hypothesis \mathcal{H}_1 , several questions are of interest:

1. a **detection problem**: are there any true alternatives?
2. an **estimation problem**: how many hypotheses are true alternatives?
3. a **selection problem**: which hypotheses are true alternatives?

These three problems have been studied in the framework of *mixture models*: a p -value of the test of the null hypothesis \mathcal{H}_0 against the alternative \mathcal{H}_1 is associated with each observation, and the distribution of these p -values is modeled as a mixture of a null and an alternative distribution.

The detection and the estimation problem can be viewed as standard testing and estimation problems. The originality of recent contributions (Abramovich et al., 2006; Cai et al., 2007; Donoho and Jin, 2004, 2006; Jin and Cai, 2007; Meinshausen and Rice, 2006; Jin, 2008) comes from the fact that they focus on *sparse* mixture models, in which

the fraction of true alternatives tends to 0, and the dissimilarity between the distributions under \mathcal{H}_0 and \mathcal{H}_1 increases as the number m of tested hypotheses tend to $+\infty$.

The selection problem is by nature less standard: as it involves testing a *large* number of hypotheses, it requires the definition of appropriate risk measures. The False Discovery Rate (FDR) introduced by [Benjamini and Hochberg \(1995\)](#) has become the most popular of these measures. FDR is the expected False Discovery Proportion (FDP), that is, the expected proportion of erroneous rejections among selected hypotheses. A related quantity is the Positive False Discovery rate (pFDR), that is, the conditional expectation of FDP given that at least one discovery has been made. [Benjamini and Hochberg \(1995\)](#) proved that the so-called BH95 procedure controls FDR at any desired level in $[0, 1]$. The selection problem has mostly been studied in *dense* (that is, non sparse) situations where all model parameters remain fixed as $m \rightarrow +\infty$ [Benjamini and Hochberg \(1995\)](#); [Genovese and Wasserman \(2002, 2004\)](#); [Storey \(2002\)](#).

1.1 Intrinsic bounds in multiple comparison problems

A natural question is whether there exist constraints on the performance of a given procedure for the detection, estimation or selection problem, or intrinsic limits to these three problems.

Detection. The detection problem consists in testing the null hypothesis that the proportion of true null hypotheses is 1, against the alternative that it is smaller than 1. Intrinsic bounds to the detection problem have been characterized in the case of sparse Gaussian mixtures ([Ingster, 1997, 1999](#); [Jin, 2002](#); [Ingster and Suslina, 2003](#)) : a sharp *detection boundary* separates situations in which the Likelihood Ratio Test (LRT) asymptotically almost surely correctly detects, from situations in which it asymptotically almost surely fails to detect. [Donoho and Jin \(2004\)](#) proved that a procedure named *higher criticism*, originally proposed by [Tukey \(1976\)](#), achieves quasi-optimal detection boundary for a few specific sparse location models, including Gaussian mixtures.

Estimation. Focusing on sparse Gaussian mixtures, [Cai et al. \(2007\)](#) proved that the region where the detection problem can be solved coincides with the region where the fraction of true alternatives can be consistently estimated. They derived minimax convergence rates in this region, and proposed an estimation procedure that achieve the optimal rate. [Meinshausen and Rice \(2006\)](#) focused on a family of estimators and derived the corresponding estimation boundary; their results are valid for any sparse mixture.

In the case of Gaussian mixtures, [Jin \(2008\)](#) suggested to estimate the fraction of true null hypotheses using the Fourier transform of the characteristic function of the p -values, and proved the consistency of a family of such estimators in the sparse and non sparse situation.

Selection. For the selection problem, [Chi \(2007a\)](#) demonstrated the existence of a possibly positive lower bound below which no multiple testing procedure can control pFDR. In such “critical” situations, the power of the BH95 procedure converges to 0 in probability. Criticality is not specific to pFDR. Other risk measures in multiple testing problems have the same kind of intrinsic limitations, for example the positive False Discovery Excessive Probability (pFDEP), that is, the conditional expectation that the FDP exceeds a given threshold ([Chi and Tan, 2008](#)).

In applications it is common for the p -values to have been generated by a set of longitudinal observations. For example, in genomic studies one typically tests for the equality of gene expression levels across two groups of samples, and one p -value is generated for each of m genes of interest. The influence of longitudinal sample size —the number of data points used to generate each p -value— on criticality has been studied by [Chi \(2007b\)](#).

1.2 Estimation, selection, and power of FDR controlling procedures

Although the concept of FDR control was introduced for the selection problem in the dense mixture model, FDR and the BH95 procedure have also been successfully applied to sparse settings, by [Donoho and Jin \(2004\)](#) for the detection problem, and by [Abramovich and Benjamini \(1995\)](#) and [Donoho and Jin \(2006\)](#) for the selection problem, in which it was demonstrated to satisfy remarkable minimax properties.

In this paper, we study how the settings of a multiple comparison problem induce limitations to the estimation and selection problems, and how these limitations translate in terms of power of FDR controlling procedures. We consider a dense setting, where the proportion of true null hypotheses is positive and fixed, and the distribution of the p -values under the null and alternative are fixed as well.

Organization of the paper. Section 2 provides background and notation. In Section 3 we give theoretical interpretation and illustration of criticality, and discuss expected and observed practical consequences of criticality in terms of power of FDR controlling procedures. In section 4 we analyze the problem of estimating the fraction π_0 of true null hypotheses based on observed p -values near 1. We study how convergence properties of the FDP achieved by plug-in FDR controlling procedures based on an estimator $\hat{\pi}_0$ of π_0 are determined by convergence properties of $\hat{\pi}_0$, which are in turn determined by regularity properties of the distribution of p -values near 1.

2. Background and notation

2.1 Model

Testing one hypothesis. As we are interested in applications such as microarray data analysis in which each observation is the result of a test based on longitudinal data, we explicitly model one observation as a realization from a test statistic X . We assume that X is distributed as F_0 under the null hypothesis \mathcal{H}_0 and as F_1 under the alternative hypothesis \mathcal{H}_1 , and denote by f_0 and f_1 the corresponding density functions. This testing problem may be formulated in terms of p -values than test statistics. The p -value function is defined as $p^+ : x \mapsto \mathbb{P}_{\mathcal{H}_0}(X \geq x)$ for one-sided tests and $p^\pm : x \mapsto \mathbb{P}_{\mathcal{H}_0}(|X| \geq |x|)$ for two-sided tests. By definition the p -values are uniform on $[0, 1]$ under \mathcal{H}_0 ; their distribution functions under \mathcal{H}_1 are derived in the next two Propositions.

Proposition 1 (One-sided p -value) *The one-sided p -value at observation $x \in \mathbb{R}$ may be written as $p^+(x) = 1 - F_0(x)$. The corresponding distribution function G_1^+ and density g_1^+ under the alternative hypothesis \mathcal{H}_1 are respectively given by*

$$G_1^+(u) = 1 - F_1(F_0^{-1}(1 - u)) \quad (1)$$

$$g_1^+(u) = \frac{f_1}{f_0}(F_0^{-1}(1 - u)) . \quad (2)$$

for any $u \in [0, 1]$.

For two sided tests we will assume that F_0 is symmetric, that is, that we have

$$\forall x \in \mathbb{R}, F_0(x) + F_0(-x) = 1 ,$$

or, equivalently that

$$\forall x \in \mathbb{R}, f_0(-x) = f_0(x) .$$

Proposition 2 (Two-sided p -value) *Assume that F_0 is symmetric. Then the two-sided p -value at observation $x \in \mathbb{R}$ may be written as $p^\pm(x) = 2(1 - F_0(|x|))$. The corresponding distribution function G_1^\pm under the alternative hypothesis \mathcal{H}_1 is given by*

$$G_1^\pm(u) = 1 - F_1(F_0^{-1}(1 - u/2)) + F_1(F_0^{-1}(u/2)) \quad (3)$$

$$g_1^\pm(u) = \frac{1}{2} \left(\frac{f_1}{f_0}(F_0^{-1}(1 - u/2)) + \frac{f_1}{f_0}(F_0^{-1}(u/2)) \right). \quad (4)$$

for any $u \in [0, 1]$.

Corollary 3 (One and two-sided p -values) *If F_0 is symmetric, then the distribution function and density function of one- and two-sided p -values under the alternative hypothesis \mathcal{H}_1 are connected by:*

$$G_1^\pm(u) = G_1^+(u/2) + 1 - G_1^+(1 - u/2) \quad (5)$$

$$g_1^\pm(u) = \frac{1}{2} (g_1^+(u/2) + g_1^+(1 - u/2)) \quad (6)$$

Testing several hypotheses: conditional mixture model We assume that m tests are performed as described in the preceding paragraph. For $i \in \{1 \dots m\}$, we let $Y_i = 0$ if hypothesis i is drawn from the null hypothesis \mathcal{H}_0 , and $Y_i = 1$ if it is drawn from the alternative \mathcal{H}_1 ; X_i denotes the corresponding test statistic. Following [Efron et al. \(2001\)](#); [Genovese and Wasserman \(2002, 2004\)](#); [Storey \(2003\)](#), we assume that the random variables $(X_i, Y_i)_{1 \leq i \leq m}$ are identically independently distributed: Y_i is a Bernoulli random variable with success probability $1 - \pi_0$, where π_0 is the unknown proportion of true null hypotheses; the conditional distribution of X_i given Y_i is F_1 if $Y_i = 1$ and F_0 if $Y_i = 0$. The marginal distribution of each X_i is thus

$$F = \pi_0 F_0 + (1 - \pi_0) F_1,$$

and we denote by $f = \pi_0 f_0 + (1 - \pi_0) f_1$ the corresponding density. We denote by G_1 and g_1 the cumulative distribution function and the probability functions of the p -values under \mathcal{H}_1 : we have $G_1 = G_1^+$ and $g_1 = g_1^+$ (as given by [Proposition 1](#)) if one-sided p -values are calculated, and $G_1 = G_1^\pm$ and $g_1 = g_1^\pm$ (as given by [Proposition 2](#)) if two-sided p -values are calculated. The p -values are uniform on $[0, 1]$ under \mathcal{H}_0 ; we denote by G_0 the corresponding cumulative distribution function, which is the identity function. The marginal distribution function and density of the p -values under the mixture are given by $G = \pi_0 G_0 + (1 - \pi_0) G_1$ and $g = \pi_0 + (1 - \pi_0) g_1$.

In this setting, the number $m_0(m)$ of true null hypotheses for a given total number m of hypotheses tested is a random variable, which verifies $\mathbb{E}[m_0(m)/m] = \pi_0$. In order to alleviate notation, we will assume without loss of generality that $m_0(m)/m = \pi_0$ for any m .

Settings. We will assume that G_1 is concave. As G is an affine transform of G_1 , this is equivalent to assuming that G is concave. For one-sided p -values, this is also equivalent to assuming that the likelihood ratio $\frac{f_1}{f_0}$ of the test statistics is non-decreasing (by [Equation \(2\)](#)), that is, that \mathcal{H}_1 dominates \mathcal{H}_0 .

Condition 1 (Concavity) G_1 is concave.

When studying two-sided p -values we will assume that the distribution function of the test statistics under \mathcal{H}_0 is symmetric:

Condition 2 (Symmetry)

$$\forall x \in \mathbb{R}, F_0(x) + F_0(-x) = 1.$$

2.2 False Discovery Rate control

The concept of False Discovery Rate (FDR) has been introduced by [Benjamini and Hochberg \(1995\)](#) in the context of the selection problem. Given a positive rejection threshold t for \mathcal{H}_0 , let R denote the total number of rejections, and $V(t)$ the number of illegitimate rejections at t among m tested hypotheses. The False Discovery Proportion at threshold t is defined by $\text{FDP}(t) = \frac{V(t)}{R(t) \vee 1}$, and the corresponding False Discovery Rate is

$$\text{FDR}(t) = \mathbb{E}[\text{FDP}(t)].$$

A related quantity is the positive false discovery rate (pFDR), that is, the conditional expectation of FDP given that at least one discovery is made:

$$\text{pFDR}(t) = \frac{\pi_0 t}{G(t)}.$$

FDR and pFDR are tightly connected as we have $\text{FDR}(t) = \text{pFDR}(t)\mathbb{P}(R(t) > 0)$. In particular they are asymptotically equivalent for procedures with fixed rejection regions because $\mathbb{P}(R(t) > 0) \rightarrow 1$, as shown by [Storey et al. \(2004\)](#).

The BH95 procedure. [Benjamini and Hochberg \(1995\)](#), elaborating on previous work by [Simes \(1986\)](#), proposed a simple procedure (henceforth denoted by the BH95 procedure) to control FDR. Suppose we wish to control FDR at level α , and let $P_{(1)} \leq \dots \leq P_{(m)}$ be the sorted p -values. Now let \hat{I}_m be the largest index k such that

$$P_{(k)} \leq \alpha \frac{k}{m}$$

If there is such an index, then all hypotheses with p -values smaller than $\hat{\tau}_m = \alpha \hat{I}_m / m$ are rejected. Otherwise, no rejection is made. The BH95 procedure provides strong control of the FDR ([Benjamini and Hochberg, 1995](#)):

$$\text{FDR}(\hat{\tau}_m) \leq \pi_0 \alpha.$$

Figure 1 illustrates the application of the BH95 procedure with $\alpha = 0.2$ to $m = 100$ simulated hypotheses, among which 20 are true alternatives. The left panel illustrates the above definition of the BH95 procedure, while the right panel gives an interpretation of this definition in terms of crossing point between the line $y = x/\alpha$ and the empirical distribution function of the p -values:

$$\hat{\tau}_m = \sup\{u \in [0, 1], \hat{\mathbb{G}}_m(u) \geq u/\alpha\}.$$

Plug-in procedures. The BH95 procedure controls FDR at level $\pi_0 \alpha$ in the above mixture model ([Benjamini and Hochberg, 1995](#)). Applying this procedure at level α/π_0 would therefore achieve $\text{FDR} = \alpha$ exactly. However, as π_0 is unknown, this is only an Oracle procedure. It is thus natural to try to estimate π_0 using $\hat{\pi}_0 < 1$, and apply the BH95 procedure at $\alpha/\hat{\pi}_0$, yielding a larger number of significant hypotheses for the same target FDR level ([Benjamini and Hochberg, 2000](#)). These “plug-in” procedures therefore have the same geometric interpretation as the BH95 procedure (see Figure 1) in terms of crossing point, with $\alpha/\hat{\pi}_0$ instead of α , and their rejection threshold can be written as

$$\hat{\tau}_m = \sup\{u \in [0, 1], \frac{\hat{\pi}_0 u}{\hat{\mathbb{G}}_m(u)} \leq \alpha\}.$$

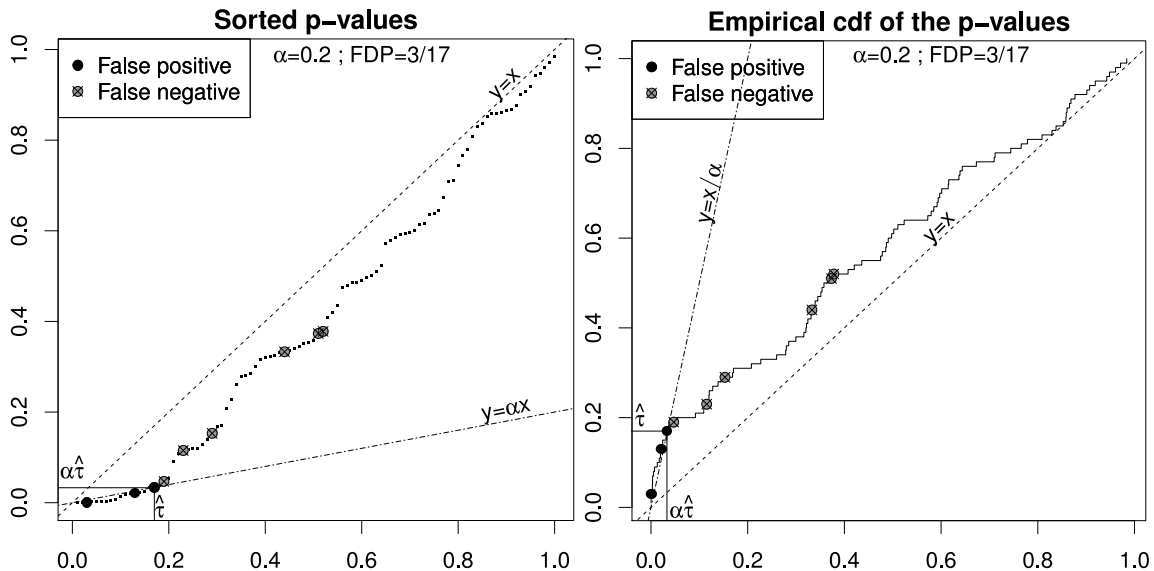


Figure 1: *Illustrations of the BH95 procedure on a simulated example. Left: sorted p-values; each dot correspond to one of 100 tested hypotheses. Right: Empirical distribution function.*

Recalling that $\text{pFDR}(t) = \frac{\pi_0 t}{G(t)}$, the rejection threshold of the plug-in procedure associated with $\hat{\pi}_0$ can therefore be interpreted as the rightmost point for which the corresponding estimate of pFDR is upper bounded by α .

We note that this idea is not specific to FDR control, as it can be (and in fact originally was) applied to the control of Family-wise error rate (FWER), that is, the probability of at least one false rejections: for example, the Bonferroni procedure at level α controls FWER at level $\pi_0 \alpha$, and corresponding plug-in procedures have been developed along similar lines (Hochberg and Benjamini, 1990).

2.3 Criticality of the selection problem.

Chi (2007a) noticed that depending on the distribution function G of the p -values, $\text{pFDR}(t)_{t>0}$ may be bounded away from 0, giving rise to a phenomenon that he called *criticality*: no selection procedure can achieve pFDR smaller than $\underline{\alpha}^* = \inf_{t>0} \text{pFDR}(t)$. Importantly, $\underline{\alpha}^*$ is intrinsic to the selection problem in the sense that it only depends on the parameters of the mixture model:

$$\underline{\alpha}^* = \inf_{t>0} \frac{\pi_0 t}{G(t)}.$$

In particular, $\underline{\alpha}^*$ is defined without a reference to any multiple comparison procedure. Criticality reveals an interesting range of situations in which FDR and pFDR are not asymptotically equivalent anymore (Chi and Tan, 2008): given a multiple comparison problem such that $\underline{\alpha}^* > 0$, any procedure that controls FDR at level $\alpha < \underline{\alpha}^*$ necessarily makes no rejection with positive probability:

$$\mathbb{P}(R = 0) = 1 - \frac{\text{FDR}}{\text{pFDR}} \geq 1 - \frac{\alpha}{\underline{\alpha}^*} > 0.$$

The actual value of $\mathbb{P}(R = 0)$ depends on the FDR controlling procedure. Criticality has been investigated by Chi (2007a) in the context of FDR and pFDR control by the BH95 procedure. The critical value of the BH95 procedure is defined as

$$\alpha^* = \inf_{u \in [0,1]} \frac{u}{G(u)}.$$

Criticality results in a threshold effect in the asymptotic proportion of rejections by the BH95 procedure, which is summarized by Proposition 4.

Proposition 4 (Criticality of the BH95 procedure (Chi, 2007a)) *Let $\rho_m(\alpha) = R_m(\alpha)/m$ be the fraction of rejections by the BH95 procedure.*

1. *If $\alpha > \alpha^*$, then $\rho_m(\alpha)$ converges in probability as $m \rightarrow +\infty$ to a positive value, which we denote by $\rho_\infty(\alpha)$.*
2. *If $\alpha < \alpha^*$, then $\rho_m(\alpha)$ converges to 0 in probability as $m \rightarrow +\infty$.*

Strictly speaking, α^* is not intrinsic to the multiple comparison problem, as it is connected to FDR control by the BH95 procedure. However, we have $\alpha^* = \underline{\alpha}^*/\pi_0$, we have $\alpha^* = 0$ if and only if $\underline{\alpha}^* = 0$. Therefore, the fact that criticality occurs or not—for any procedure—can be characterized in terms of the critical value of the BH95 procedure by the fact that $\alpha^* = 0$ or not.

3. Criticality and distribution of the test statistics

As noted by Chi and Tan (2008), criticality is intrinsic to a multiple testing problem: $\underline{\alpha}^*$ only depends on the characteristics of the model; in particular it does not depend on a multiple testing procedure. However, when a criticality phenomenon occurs, the actual lower bound α^* on the target FDR level that ensures non-trivial FDR control by a given procedure does depend on the procedure.

For simplicity, the results presented in this section are written and illustrated specifically for the BH95 procedure. We begin by providing theoretical interpretation and illustration of criticality by studying how different families of distribution of the test statistics can lead to different behaviors in terms of criticality (Section 3.1). Then we discuss expected and observed practical consequences of criticality by studying the power of the BH95 procedure as a function of the target FDR level α (Section 3.2) in simulations and real data.

We will discuss *location problems*, that is, problems in which the distribution of the test statistic under \mathcal{H}_1 is a shift from that of the test statistic under \mathcal{H}_0 : $F_1 = F_0(\cdot - \theta)$ for some location parameter $\theta > 0$. We will also investigate the case of Student test statistics, which is not a location problem but is widely used in real data analysis.

3.1 Interpretation and illustration of criticality

We begin by recalling a characterization of criticality for the BH95 procedure, in terms of the behavior of the density g_1 under the alternative at 0. Under Condition 1, G is concave and $u \mapsto \frac{u}{G(u)}$ is non-decreasing on $[0, 1]$. The critical value of the BH95 procedure is then given by

$$\alpha^* = \lim_{u \rightarrow 0} \frac{u}{G(u)}. \tag{7}$$

Criticality therefore only depends on the behavior of $\frac{G(u)}{u}$ at 0. This is not surprising if we go back to the interpretation of the BH95 procedure proposed in Figure 1 (right panel):

criticality corresponds to situations where α is so small that there is no positive crossing point between G and the line $y = x/\alpha$. Combining (7) with l'Hôpital's rule, α^* may be written as $\lim_{u \rightarrow 0} 1/g(u)$, which is a function of the likelihood ratio $\frac{f_1}{f_0}$ of the model. Therefore, criticality is governed by the behavior of $\frac{f_1}{f_0}(t)$ as t tends to $+\infty$. These results were established by Chi (2007a) and Chi and Tan (2008) for one-sided p -values, and are summarized by the following Proposition.

Proposition 5 (Criticality and likelihood ratios (Chi (2007a); Chi and Tan (2008))) *Under Conditions 1 and 2, we have:*

1. *If $\frac{f_1}{f_0}$ is bounded as $t \rightarrow +\infty$, then the density g_1 of the p -values under the alternative has a finite limit at 0 (which we denote by $g_1(0)$). Criticality occurs, and the critical value is given by*

$$\alpha^* = \frac{1}{\pi_0 + (1 - \pi_0)g_1(0)};$$

2. *If $\lim_{t \rightarrow +\infty} \frac{f_1}{f_0}(t) = +\infty$, then $\lim_{u \rightarrow 0} \frac{G(u)}{u} = +\infty$, and $\alpha^* = 0$. There is no criticality, and all target FDR levels are attainable.*

Note that Proposition 5 holds for both one- and two-sided p -values. We now illustrate this property in location models (Section 3.1.1) and in the more complicated—but more realistic—situation of Student test statistics (Section 3.1.2)

3.1.1 ILLUSTRATION IN LOCATION MODELS

In location models the behavior of the likelihood ratio $\frac{f_1}{f_0} = \frac{f_0(\cdot - \theta)}{f_0}$ is closely related to the tail behavior of the distribution of the test statistics: for a given non-centrality parameter θ , the heavier tails, the smaller difference between $f_0(\cdot - \theta)$ and f_0 , and the larger critical value. The most well-known location problems are the Gaussian and Laplace (double exponential) location problems, which illustrate distinct behaviors in terms of criticality.

Gaussian test statistics. Assume that the test statistics are distributed as $\mathcal{N}(0, 1)$ under the null hypothesis, and as $\mathcal{N}(\theta, 1)$ under the alternative (with $\theta \neq 0$). The likelihood ratio is thus given by

$$\begin{aligned} \frac{f_1}{f_0}(t) &= \exp\left(-\frac{1}{2}(t - \theta)^2 + \frac{1}{2}t^2\right) \\ &= \exp\left(-\frac{\theta^2}{2} + \theta t\right). \end{aligned}$$

As this likelihood ratio is non decreasing and not bounded as $t \rightarrow +\infty$, Proposition 5 implies that there is no criticality in the Gaussian location problem: $\alpha^* = 0$. Figure 2 illustrates the absence of criticality for the Gaussian location problem: $\alpha^* = 0$ whatever the values of θ and π_0 , as the distribution function has a vertical semi-tangent at the origin. We now investigate the case of Laplace (double exponential) test statistics, which has heavier tails than the Gaussian distribution; this results in a positive critical value.

Laplace test statistics. Assume that the density of the test statistics is $f_0 : t \mapsto \frac{1}{2}e^{-|t|}$ under the null hypothesis, and $f_1 : t \mapsto \frac{1}{2}e^{-|t - \theta|}$ under the alternative, with $\theta > 0$. The corresponding distribution functions under \mathcal{H}_1 are derived in Appendix B for one-sided p -values (Proposition 25). The likelihood ratio of the model is given by $\frac{f_1}{f_0}(t) = e^{|t| - |t - \theta|}$, that is, $\frac{f_1}{f_0}(t) = e^{2t - \theta}$ if $t \leq \theta$, and e^θ if $t > \theta$. The likelihood ratio of this model is therefore

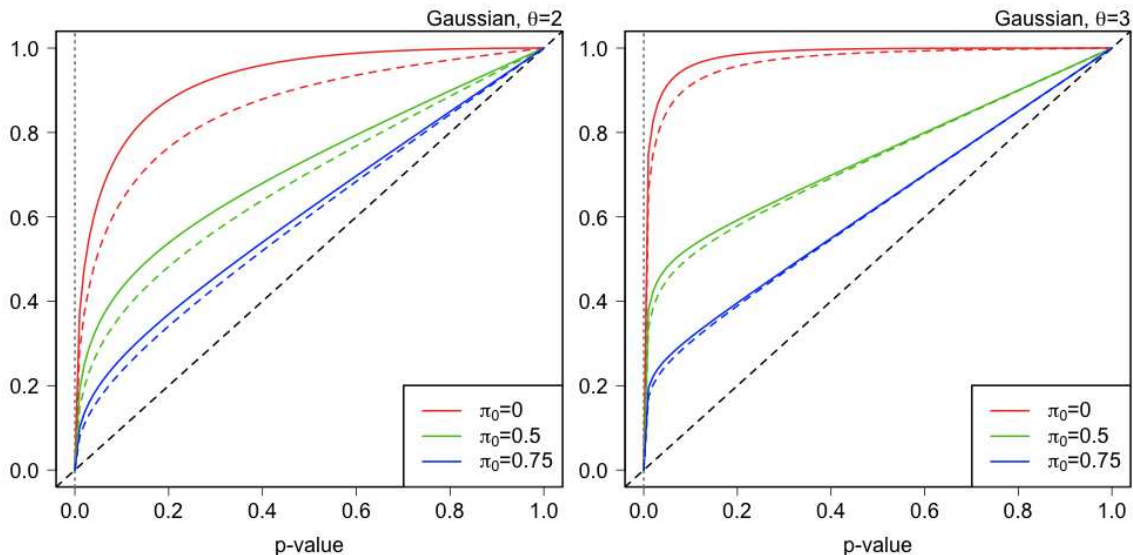


Figure 2: *Non criticality for the Gaussian distribution, illustrated with $\theta = 2$ (left) and $\theta = 3$ (right). Distribution functions of one-sided (solid) and two-sided (dashed) p-values, for $\pi_0 = 0, 0.5$ and 0.75 .*

bounded, which results (by Proposition 5) in a positive critical value. By Proposition 25, we have $\alpha^* = 1/(\pi_0 + (1 - \pi_0)e^\theta)$ for one-sided p -values, and $\alpha^* = 1/(\pi_0 + (1 - \pi_0) \cosh \theta)$ for two-sided p -values. Figure 3 illustrates criticality for the Laplace location problem. The distribution function of one-sided p -values under a Laplace mixture with proportion π_0 of true nulls is linear between 0 and $\frac{e^{-\theta}}{2}$, with slope $\pi_0 + (1 - \pi_0)e^\theta = 1/\alpha^*$ (see Appendix B, Proposition 25). Criticality occurs for any value of θ and π_0 (and would also occur for any FDR controlling procedure). However, the value of α^* depends on both θ and π_0 (and on the procedure), as illustrated in Figure 4.

Numerical example in the one-sided Laplace case. As α^* is a decreasing function of θ and π_0 , the knowledge of a lower bound on θ and $1 - \pi_0$ can be translated into a lower bound on α^* . For example, suppose that we know that $\theta \leq 2$, and $\pi_0 \geq 0.75$ in the one-sided Laplace case. Then $\alpha^* \geq \frac{1}{0.75 + 0.25e^2} = 0.385$, which means that even though π_0 and θ are not exactly known, we know that the BH95 procedure applied in this setting with any target FDR level $\alpha < 0.385$ has asymptotically null power as the number of tested hypotheses grows to $+\infty$. In the case when π_0 is totally unknown, for a given lower bound on θ , there is still a positive minimal α^* , namely $\underline{\alpha}^* = e^{-\theta}$, which corresponds to the limit case when all hypotheses come from the alternative (that is, $\pi_0 = 0$ and $G = G_1$). This limit case is represented in red in Figure 3. For example, with $\theta \leq 2$, then $\underline{\alpha}^* = 0.135$, whatever π_0 .

Subbotin test statistics. Gaussian and Laplace distributions can be viewed as instances of a more general class of distribution introduced by Subbotin (1923). Let us define the density of the test statistics by

$$f_0^\gamma(t) = \frac{1}{C_\gamma} \exp\left(-\frac{|t|^\gamma}{\gamma}\right)$$

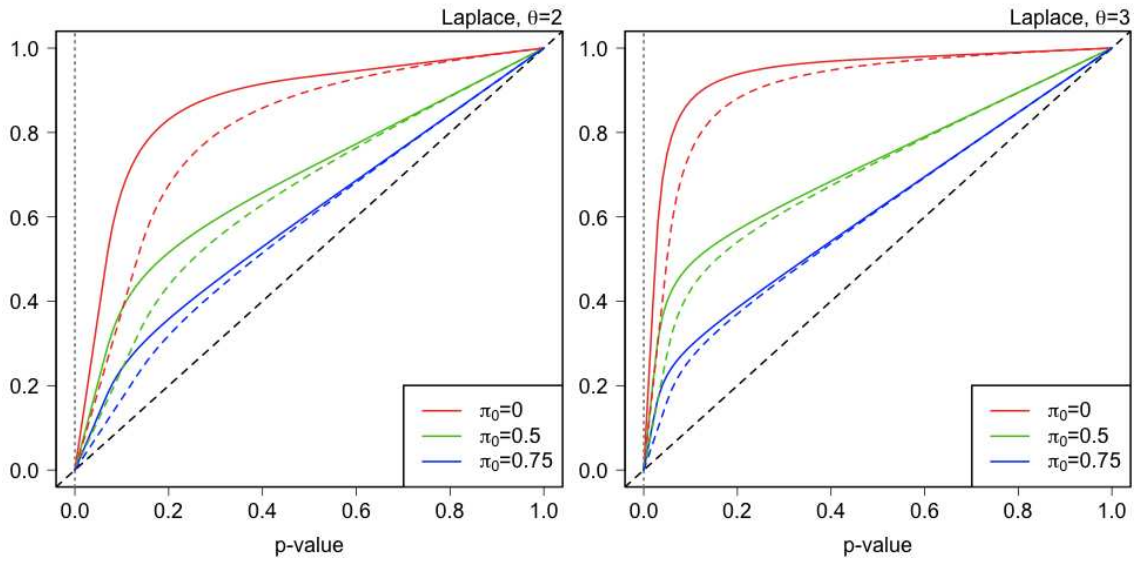


Figure 3: *Criticality for the Laplace distribution, illustrated with $\theta = 2$ (left) and $\theta = 3$ (right), respectively. Distribution functions of one-sided (solid) and two-sided (dashed) p-values for $\pi_0 = 0, 0.5$ and 0.75 .*

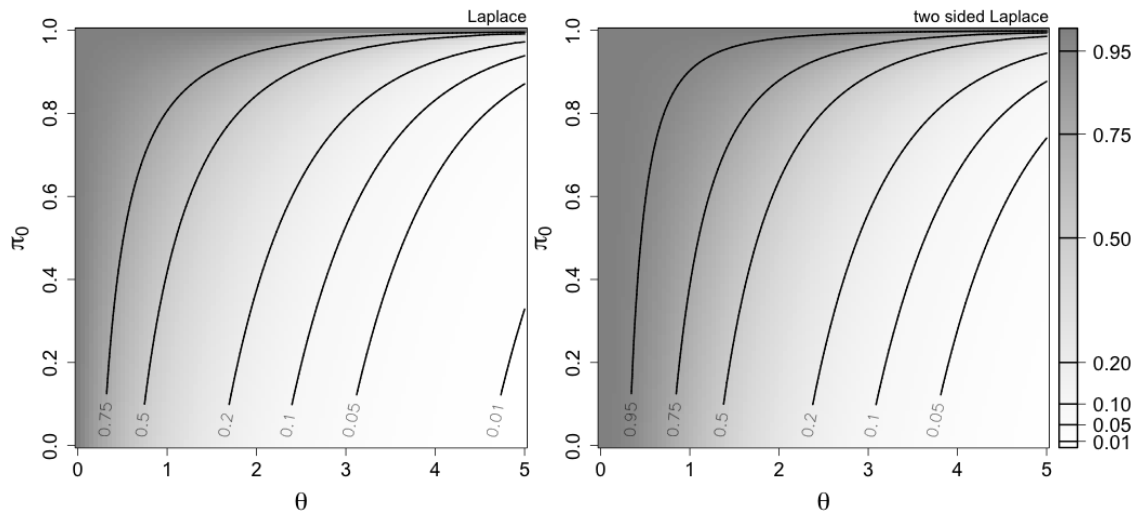


Figure 4: *Critical values α^* for one-sided (left) and two-sided (right) Laplace distributions, as a function of θ and π_0 . Solid black lines represent level curves for a few values of α^* , which are also marked in the color scale.*

under \mathcal{H}_0 , and $f_1^\gamma(t) = f_0^\gamma(t - \theta)$ under \mathcal{H}_1 , where C_γ is a normalizing constant that makes f_0^γ a density. The Gaussian case corresponds to $\gamma = 2$ and the Laplace case to $\gamma = 1$. The

likelihood ratio for this location problem is given by

$$\begin{aligned} \frac{f_1^\gamma}{f_0^\gamma}(t) &= \exp\left(\frac{|t|^\gamma}{\gamma} - \frac{|t-\theta|^\gamma}{\gamma}\right) \\ &= \exp\left(\frac{|t|^\gamma}{\gamma} \left(1 - \left|1 - \frac{\theta}{t}\right|^\gamma\right)\right). \end{aligned}$$

We focus on $\gamma \geq 1$ because it corresponds to situations in which $\frac{f_1^\gamma}{f_0^\gamma}$ is non-decreasing. As $t \rightarrow +\infty$, $|1 - \frac{\theta}{t}|^\gamma \sim 1 - \frac{\gamma\theta}{t}$, so that $\frac{|t|^\gamma}{\gamma} \left(1 - |1 - \frac{\theta}{t}|^\gamma\right) \sim \theta t^{\gamma-1}$, and the behavior of $\frac{f_1^\gamma}{f_0^\gamma}(t)$ is driven by the value of γ : if $\gamma = 1$, $\lim_{t \rightarrow +\infty} \frac{f_1^\gamma}{f_0^\gamma}(t) = e^\theta$ and there is a positive critical value, as noted above. If $\gamma > 1$ (for example in the Gaussian case), $\lim_{t \rightarrow +\infty} \frac{f_1^\gamma}{f_0^\gamma}(t) = +\infty$ and there is no criticality. Laplace-distributed test statistics appear as a limit case in terms of criticality: within the family of Subbotin location models, there is no criticality if and only if the tails of the test statistics are lighter than exponential.

3.1.2 ILLUSTRATION FOR STUDENT TEST STATISTICS

The study of location models in the preceding section provides insight into the connection between tail behavior of the test statistics and criticality. In practice however, test statistics are typically assumed to follow a Student distribution, because they have been generated from longitudinal observations that can be assumed to be Gaussian with unknown variance.

Definition 6 (Student distribution) *Let Z_θ be normally distributed with mean θ and variance 1, and Y independently distributed as central χ^2 with k degrees of freedom. Then the random variable $T_{k,\theta} = \frac{Z_\theta}{\sqrt{Y/k}}$ is said to follow a t distribution (Student distribution) with k degrees of freedom and non-centrality parameter θ . If $\theta = 0$, $T_{k,0}$ is denoted by T_k and is said to follow a (central) t distribution with k degrees of freedom.*

Note the Student multiple testing problem is not a location model, as a non-central Student random variable is not a translation from a central Student random variable. As a practical illustration for a Student multiple comparison problem, we will consider a microarray data set (Golub et al., 1999) which consists of the measured expression level of $m = 3051$ genes in blood samples from 38 patients suffering from two types of leukemia: acute lymphoblastic leukemia (ALL, 27 cases) and acute myeloid leukemia (AML, 11 cases). The goal of the original study was to find genes that are significantly over- or under-expressed in one class of patients with respect to the other class. We have used the data from the R package `multtest` available from Bioconductor (Gentleman et al., 2004). This data was preprocessed as described in Dudoit et al. (2002). For each gene, we performed a *two-sided* Student test of the null hypothesis that this gene is equally expressed in the two classes of patients.

We assume that we are observing (X_1, \dots, X_{n_X}) independent observations distributed as $\mathcal{N}(\mu_X, \sigma^2)$ and (Y_1, \dots, Y_{n_Y}) independent observations distributed as $\mathcal{N}(\mu_Y, \sigma^2)$. We also assume that $(X_i)_{1 \leq i \leq n_X}$ and $(Y_i)_{1 \leq i \leq n_Y}$ are independent. We focus on the (two-sided) problem of testing $\mathcal{H}_0 : \mu_X = \mu_Y$ against $\mathcal{H}_1 : \mu_X \neq \mu_Y$.

Proposition 7 *Letting $\bar{X}_{n_X} = \frac{1}{n_X} \sum_{i=1}^{n_X} X_i$ and $\bar{Y}_{n_Y} = \frac{1}{n_Y} \sum_{i=1}^{n_Y} Y_i$, define the Student test statistic of \mathcal{H}_0 against \mathcal{H}_1 as*

$$T_{n_X+n_Y} = \frac{\bar{Y}_{n_Y} - \bar{X}_{n_X}}{S_{XY} \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}},$$

where

$$S_{XY} = \frac{1}{n_X + n_Y - 2} \left(\sum_{i=1}^{n_X} (X_i - \bar{X}_{n_X})^2 + \sum_{i=1}^{n_Y} (Y_i - \bar{Y}_{n_Y})^2 \right)$$

is an estimator of the common standard deviation of the two samples. Under \mathcal{H}_0 , $T_{n_X+n_Y}$ follows a central t distribution with $n_X + n_Y - 2$ degrees of freedom; under \mathcal{H}_1 , $T_{n_X+n_Y}$ follows a non-central t distribution with $n_X + n_Y - 2$ degrees of freedom and non-centrality parameter

$$\theta = \frac{\mu_Y - \mu_X}{\sigma \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}. \quad (8)$$

Remark 8 (effect size) The non-centrality parameter may be written as $\theta = \delta / \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}$, where the effect size

$$\delta = \frac{\mu_Y - \mu_X}{\sigma} \quad (9)$$

does not depend on the number n_X and n_Y of longitudinal observations in each group. δ characterizes the distributions $\mathcal{N}(\mu_X, \sigma^2)$ and $\mathcal{N}(\mu_Y, \sigma^2)$ of the observations.

The following Proposition gives an expression of the likelihood ratio for Student test statistics.

Proposition 9 (Likelihood ratio for Student test statistics) The likelihood ratio between a central Student distribution with k degrees of freedom and a non-central Student distribution with k degrees of freedom and non-centrality parameter θ may be written as

$$\frac{f_1}{f_0}(t) = \exp \left[-\frac{\theta^2}{2} \frac{1}{1 + \frac{t^2}{k}} \right] \frac{Hh_k \left(-\frac{\theta t}{\sqrt{k+t^2}} \right)}{Hh_k(0)}.$$

where

$$Hh_k(z) = \int_0^{+\infty} \frac{x^k}{k!} e^{-\frac{1}{2}(x+z)^2} dx.$$

Proposition 10 (Criticality — Student multiple comparison problem) Consider the Student multiple comparison problem where test statistics are distributed as central Student with k degrees of freedom under \mathcal{H}_0 and non-central Student with k degrees of freedom and non-centrality parameter δ under \mathcal{H}_1 . The corresponding likelihood ratio is non-decreasing, and bounded as $|t| \rightarrow +\infty$: therefore, there is a positive critical value, which is given by $\alpha^* = \left(\pi_0 + (1 - \pi_0) \frac{Hh_k(-\theta)}{Hh_k(0)} \right)^{-1}$ for one-sided p -values and by $\alpha^* = \left(\pi_0 + (1 - \pi_0) \frac{Hh_k(\theta) + Hh_k(-\theta)}{2Hh_k(0)} \right)^{-1}$ for two-sided p -values.

The fact that $\alpha^* > 0$ is consistent with the fact that the Student distribution has polynomial tails, that is, heavier than for the Laplace location problem, in which criticality already occurred. This result is illustrated by Figure 5. The parameters of the Student distribution functions have been chosen as follows. First, the number of degrees of freedom in the right panel has been chosen to match the actual number of observations in the Golub data set: $n_X + n_Y - 2 = 27 + 11 - 2 = 36$ degrees of freedom. The value of $\theta = 2.5$ has been chosen empirically to maximize the fit between observed (solid black line) and expected distributions of two-sided Student p -values: for $\pi_0 = 0.5$ (dashed green line in the right panel), the fit is quite good.

As noted above (Remark 8), the effect size $\delta = (\mu_Y - \mu_X)/\sigma$ characterizes the distributions of the longitudinal observations, regardless of the number of longitudinal observations. Using Equation (9), δ can be estimated as $\delta \approx 0.9$ for this data set. The left panel illustrates the distribution functions for the same multiple comparison problem —as characterized by its effect size δ and fraction of true null hypotheses π_0 — with a smaller number of observations: we have chosen $n_X = 8$ and $n_Y = 3$, which corresponds to approximately 30% of the original number of observations. The parameters of the associated Student distribution are therefore $n_x + n_Y - 2 = 9$ degrees of freedom, and $\theta \approx 1.3$ for $\delta = 0.9$.

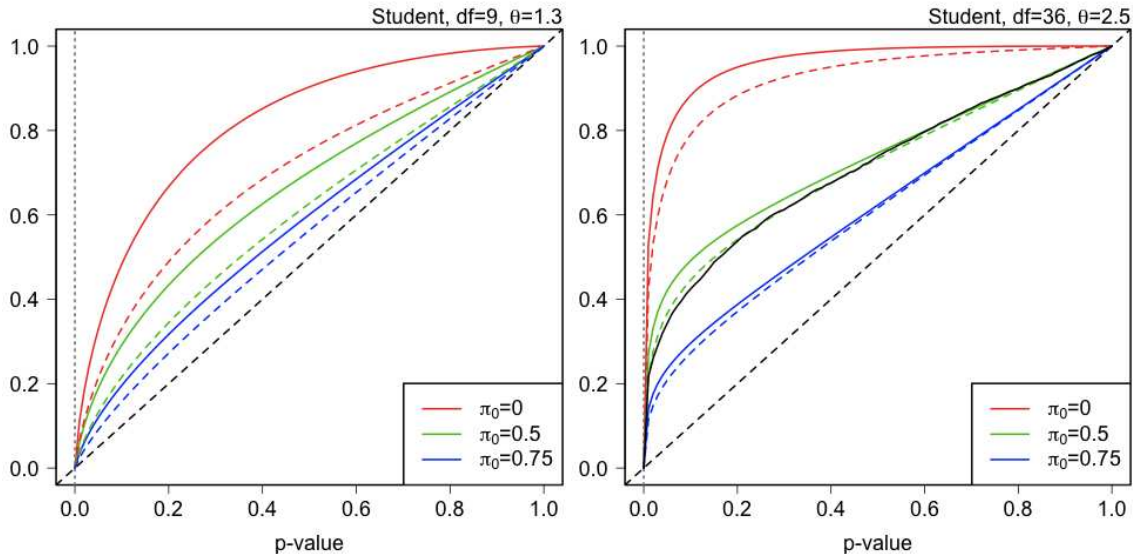


Figure 5: *Criticality for the Student distribution with 9 degrees of freedom and $\theta = 1.3$ (left), and (right) 36 degrees of freedom and $\theta = 2.5$. The effect size is the same for both panels: $\delta = 0.9$. Distribution functions of one-sided (solid) and two-sided (dashed) p-values for $\pi_0 = 0, 0.5$ and 0.75 . Solid black line in the right panel is the observed distribution of two-sided p-values in the Golub data set.*

In the mixture model, we assumed that each gene is either non-differentially expressed: $\mu_X = \mu_Y$ or differentially expressed with a common, positive effect size, $\delta = (\mu_X - \mu_Y)/\sigma$. In practice, the distinction between differentially expressed and non differentially expressed genes is not as clear cut. We believe that the closeness between the Student distribution with 36 degrees of freedom and non-centrality parameter $\theta = 2.5$ to the distribution of the observed p -values (dashed green and solid black lines in the left panel of Figure 5) indicates that our model is relevant to real data analysis.

The comparison between the left and the right panel of Figure 5 illustrates the influence of longitudinal sample size on criticality: although the effect size is the same in both panels ($\delta = 0.9$), criticality is much more serious when sample sizes n_X and n_Y is small, because both the number of degrees of freedom $n_X + n_Y - 2$ and the non-centrality parameter $\theta = \delta/\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}$ are increasing functions of n_X and n_Y .

The influence of longitudinal sample size on criticality has been studied by Chi (2007b). Although the supremum of the likelihood ratio of the Student multiple comparison problem is bounded for a given number of degrees of freedom k and effect size δ by Proposition 9,

this likelihood ratio grows to $+\infty$ whenever $\delta \rightarrow 0$ and $k \rightarrow +\infty$, provided that $k\delta \rightarrow +\infty$ (Chi, 2007b, Lemma 2.1.3). This implies that criticality is canceled out by the supply of longitudinal observations. The influence of longitudinal sample size on power in real data analyses will be discussed in the next section.

3.2 Criticality and power of the BH95 procedure

In the preceding section we have seen that different families of distribution of the test statistics can lead to different behaviors in terms of criticality, and that when criticality occurs the critical value α^* depends on the parameters of this distribution. However it is still unclear how FDR controlling procedures behave in practice depending on α and α^* , for two main reasons. First, criticality is an asymptotic notion: it characterizes the *asymptotic* behavior of the proportion of rejected hypotheses as the number of hypotheses grows to infinity (see Proposition 4). Second, criticality only gives a *binary* interpretation of the situation: either criticality occurs or does not occur, and we have little indication on how serious criticality is in a given setting.

In this section, we study the power of the BH95 procedure as a function of the target FDR level α , in order to gain insight into the practical consequences of criticality. The power of a FDR controlling procedure at level α can be defined as the proportion of true positives among rejections:

$$\Pi_m(\alpha) = \frac{R_m(\alpha) - V_m(\alpha)}{m - m_0}.$$

The corresponding asymptotic power $\Pi_\infty(\alpha)$ is defined as the limit in probability of $\Pi_m(\alpha)$ as $m \rightarrow +\infty$. Proposition 11 demonstrates that criticality results in a thresholding effect in the asymptotic power achieved by the BH95 procedure, similarly to what we observed in Proposition 4.

Proposition 11 (Asymptotic power of the BH95 procedure, Chi 2007a) *Let α^* be the critical value of the BH95 procedure.*

1. If $\alpha < \alpha^*$, $\Pi_\infty = 0$;
2. If $\alpha > \alpha^*$, then

$$\Pi_\infty = \frac{1 - \pi_0 \alpha}{1 - \pi_0} \rho_\infty(\alpha).$$

Proposition 11 motivates the following two questions:

- when $\alpha < \alpha^*$, the BH95 procedure has asymptotically null power; how does the power of the BH95 procedure behave for a finite number of hypotheses ?
- when $\alpha > \alpha^*$, the BH95 procedure has asymptotically *positive* power; how large is this power, both asymptotically and for a finite number of hypotheses ?

In order to address these questions, we compare the expected and observed power of the BH95 procedure for different location models using a simulation study (Section 3.2.1). Then we illustrate the influence of longitudinal sample size on power and criticality in a real microarray data set, and discuss the connection with asymptotic results (Section 3.2.2).

3.2.1 OBSERVED AND EXPECTED POWER OF THE BH95 PROCEDURE IN LOCATION MODELS

Figure 6 displays power as a function of α in the same settings as Figure 2 and Figure 3: Gaussian (top) and Laplace (bottom) distributions, with non-centrality parameter $\theta = 1$

(left) or 2 (right), with $\pi_0 \in \{0, 0.75, 0.9\}$. Asymptotic power (Π_∞) is represented by solid lines. This figure also summarizes the results of a simulation study that we performed in order to compare Π_∞ with the power Π_m achieved for a finite number of observations. For each simulation unit, $m = 1000$ p -values were generated according to the above settings. Dashed and dash-dotted lines correspond to the median, 5% and 95% quantiles of Π_m over $B = 1000$ repetitions the simulation study. For the Laplace problem, where criticality occurs, critical regions for each value of π_0 are identified by colored rectangles.

The asymptotic power curves (dashed lines) of the Laplace distributions in Figure 6 illustrate the singularity in the power function at $\alpha = \alpha^*$ characterized by Proposition 11: the asymptotic power is identically 0 for $\alpha < \alpha^*$ and positive for $\alpha > \alpha^*$, with a vertical semi-tangent at $\alpha = \alpha^*$. As there is no criticality for the Gaussian distribution ($\alpha^* = 0$), there is no such singularity for this distribution, and the asymptotic power is a smooth function of the target level α .

The curves Π_m and Π_∞ are quite similar for $\alpha > \alpha^*$, while Π_m tends to be slightly larger than Π_∞ for $\alpha < \alpha^*$: the singularity at $\alpha = \alpha^*$ is smoothed in the observed power function. This suggests that the effect of criticality in real data analysis is less dichotomous than suggested by the mere distinction of $\alpha < \alpha^*$ versus $\alpha > \alpha^*$. This conclusion is reinforced by the comparison of power functions across parameters within a given family of distributions. For the Gaussian distribution when the non-centrality parameter θ is small (top left panel), no or very few rejections are made for small values of α , even though there is no criticality ($\alpha^* = 0$). For any positive target level α , the associated power will eventually be positive for m large enough, but it can still be quite small if θ is small and π_0 is large. For the Laplace distribution (bottom row), the observed (finite sample) power can be positive even for $\alpha < \alpha^*$, although it is generally quite small.

3.2.2 LONGITUDINAL SAMPLE SIZE AND CRITICALITY: THEORY AND PRACTICE FOR THE STUDENT DISTRIBUTION

In order to study how the results presented in the preceding section translate in real data analysis, we go back to the analysis of the Golub et al. (1999) microarray data set described in Section 3.1.2. As we do not know which genes are truly differentially expressed in this study, we cannot calculate the power $\Pi_m(\alpha) = (R_m(\alpha) - V_m(\alpha))/(m - m_0)$ at threshold α as we did in the preceding section. Indeed, both the total number of true null hypotheses m_0 and the number of $V_m(\alpha)$ of false positives at threshold α are unobserved. Therefore, we focus on the fraction $\rho_m(\alpha)$ of hypotheses rejected by the BH95 procedure at threshold α :

$$\rho_m(\alpha) = R_m(\alpha)/m.$$

We emphasize that asymptotic power Π_∞ and asymptotic fraction of rejected hypotheses ρ_∞ are expected to have the same type of behavior: both are null for $\alpha < \alpha^*$, and they are connected by $\Pi_\infty = \frac{1-\pi_0\alpha}{1-\pi_0}\rho_\infty(\alpha)$ when $\alpha > \alpha^*$ (Propositions 4 and 11).

Figure 7 compares $\rho_\infty(\alpha)$ (dashed curves) for Student test statistics to the observed fraction $\rho_m(\alpha)$ of genes declared differentially expressed by the BH95 procedure at level α in the Golub et al. (1999) data set (solid curves). Red curves correspond to the entire data set (38 samples), and green and blue curves correspond to subsets of 60 and 30% of the original data set, respectively. For each sampling rate s , 100 resamplings of the original data set were performed as follows: $\lfloor s \cdot n_X \rfloor$ and $\lfloor s \cdot n_Y \rfloor$ samples were chosen randomly among ALL and AML samples, respectively, and the BH95 procedure was applied to Student two-sided p -values of differential expression between the two groups.

For ρ_∞ (dashed lines), the parameters of the Student multiple comparison problem were chosen as described in the preceding section (p. 14): the signal no noise ratio was set to $\delta = 0.9$ and the proportion of true null hypotheses to $\pi_0 = 0.5$. These parameters

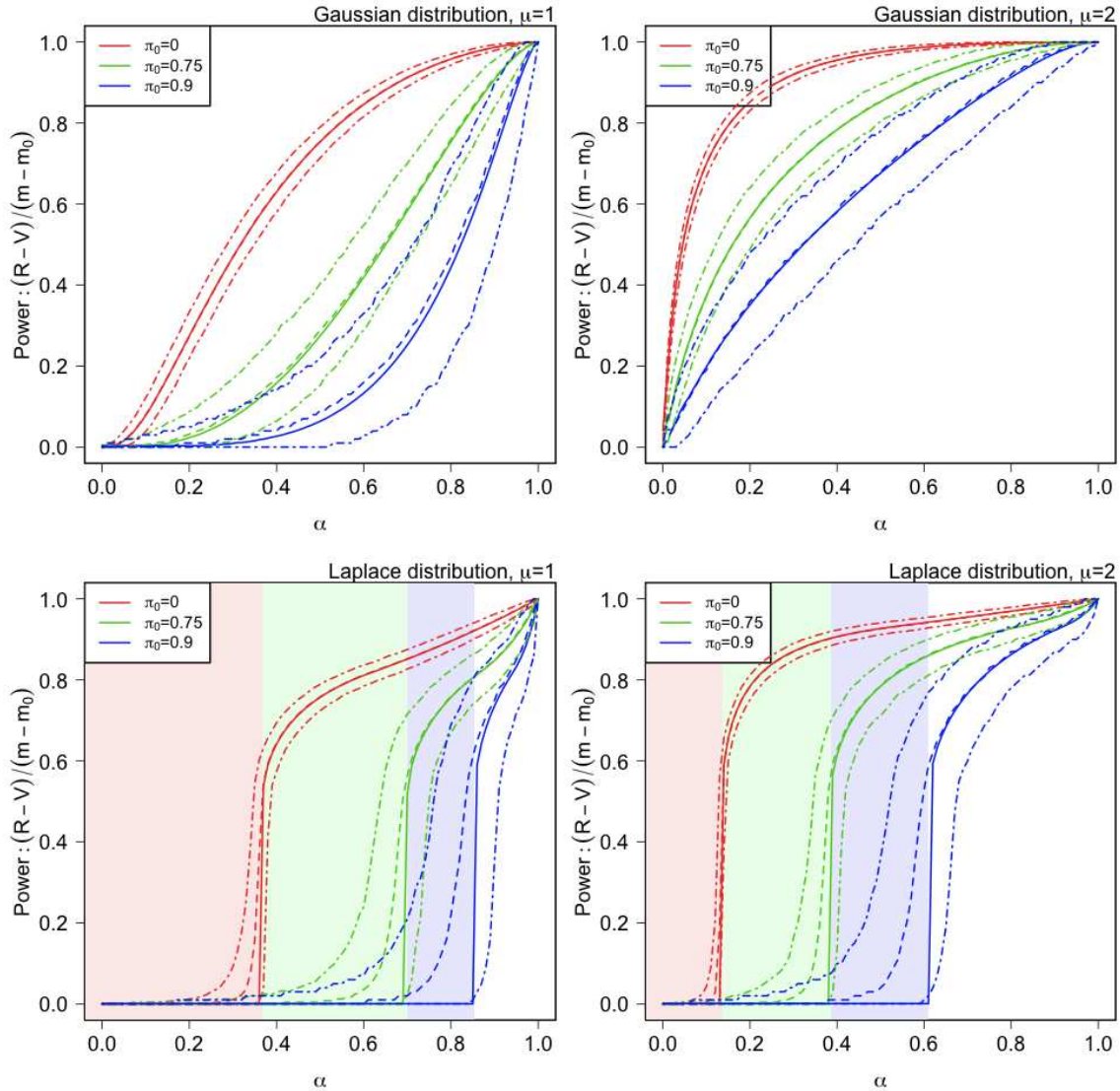


Figure 6: Power Π_m achieved by the BH95 procedure as a function of target FDR level α , for different flavors of one-sided Gaussian and Laplace problems: $\theta = 1$ (left) and $\theta = 2$ (right). Top: Gaussian distribution (no criticality); bottom: Laplace distribution (criticality). Solid lines represent $\Pi_\infty(\alpha)$, the limit in probability of $\Pi_m(\alpha)$ as $m \rightarrow +\infty$. Dashed and dash-dotted lines correspond to the median, 5% and 95% quantiles of $\Pi_m(\alpha)$ over $B = 1000$ simulations. Colored regions in the bottom plot illustrate critical regions for each value of π_0 ; in areas where the critical regions overlap, the color corresponding to the smallest α^* has been used.

were kept constant across sampling rates, and the number of degrees of freedom and non

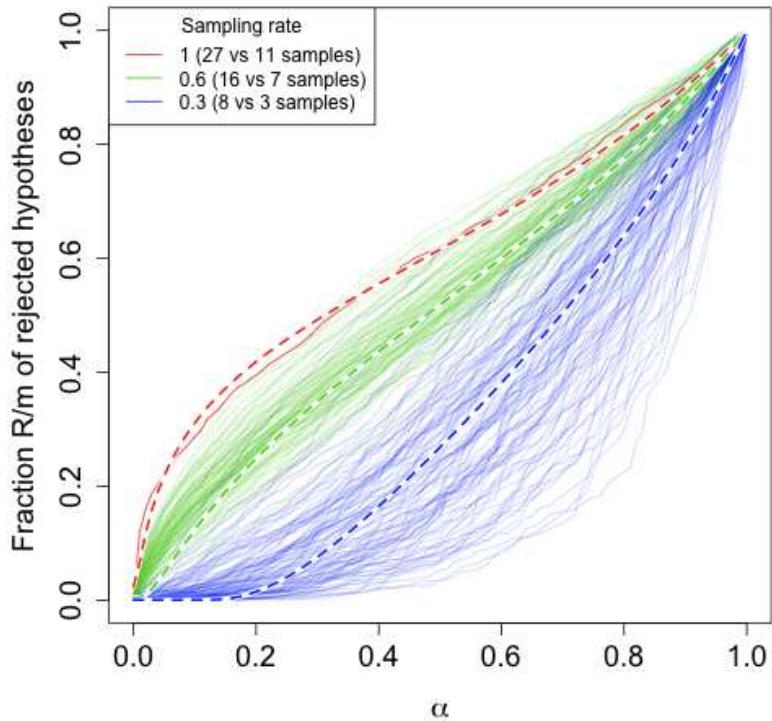


Figure 7: *The fraction of hypotheses rejected by the BH95 procedure as a function of target FDR level α illustrates the influence of longitudinal sample size on power and criticality on the Golub data set. Solid lines correspond to random samplings of the original data set, each color corresponding to a different sampling rate: full data set (red, 38 samples), sampling rate 0.6 (green, 23 samples), sampling rate 0.3 (blue, 11 samples). For sampling rates 0.3 and 0.6, 100 random resamplings of the original data set have been performed. Dashed lines represent the asymptotic fraction of hypotheses rejected assuming a mixture model of Student distributions (see main text for details).*

centrality parameters were adjusted accordingly for each sampling fraction, as described in Table 1.

| Sampling rate | | 100% | 60% | 30% |
|--------------------------|--|------|-----|-----|
| ALL sample size | n_X | 27 | 16 | 7 |
| AML sample size | n_Y | 11 | 7 | 3 |
| Degrees of freedom | $k = n_X + n_Y$ | 36 | 21 | 8 |
| non centrality parameter | $\theta = \delta / \sqrt{1/n_X + 1/n_Y}$ | 2.5 | 2 | 1.3 |

Table 1: *Parameters used for the resampling study in Figure 7. δ was set to 0.9.*

The solid red curve in Figure 7 corresponds to the observed fraction of rejections in the Golub data set, while the dashed red curve corresponds to the *asymptotic* fraction of rejected hypotheses of a Student multiple comparison problem with $\delta = 0.9$, $\pi_0 = 0.5$, $n_X = 27$ and $n_Y = 11$. We interpret the closeness of these two curves as the combination of two elements. First, the parameters δ and π_0 adequately describe the data set. This is consistent with the fact that δ and π_0 were chosen to maximize the fit between the expected and observed distribution function of the two-sided Student p -values (Figure 5, right panel). Second, the asymptotic fraction of rejected hypotheses is a good approximation of the observed fraction of rejected hypotheses.

There is non-negligible variability in the distribution of observed fractions of rejected hypotheses in each of the two sampling scenarios (solid green and blue curves). The asymptotic fraction of rejections (dashed curves) are consistent with the observed fractions (solid curves), although for a sampling rate of 60% (green) the asymptotic fraction seems to underestimate the observed fractions for a small target FDR.

4. Estimation of π_0

In this section we focus on the estimation of the fraction π_0 of true null hypotheses in the settings described in Section 2.1, where the density g_1 of the p -values under the alternative hypothesis is assumed to be decreasing (Condition 1). As g_1 is unknown, a natural approach to estimate π_0 from observed (one- or two-sided) p -values drawn from a mixture with density $g = \pi_0 + (1 - \pi_0)g_1$ is to focus on p -values near 1. In this section, $\hat{\pi}_0$ denotes a generic estimator of π_0 based on this idea. A number of such estimators have been studied in this context (Efron et al., 2001; Genovese and Wasserman, 2004; Meinshausen and Bühlmann, 2005; Meinshausen and Rice, 2006; Schweder and Spjøtvoll, 1982; Langaas et al., 2005; Storey et al., 2004; Benjamini et al., 2006).

The problem of estimating π_0 is not only of interest in itself, it is also motivated by power consideration in multiple testing problems: for example, using the plug-in procedure $\text{BH95}(\alpha/\hat{\pi}_0)$, where $\hat{\pi}_0$ is an estimator of π_0 , yields tighter FDR control than the standard BH95 procedure (Benjamini and Hochberg, 2000). The goals of this section are to understand what drives the regularity properties of these estimators of π_0 in our setting, and investigate the consequences of these regularity properties in terms of FDR controlling capabilities of associated plug-in BH95 procedures.

We begin by pointing out a connexion between criticality in one-sided, symmetric location models and a necessary condition to achieve consistency of $\hat{\pi}_0$ (Section 4.1). Then we show how convergence rates of $\hat{\pi}_0$ are connected to regularity properties of g_1 near 1 (Section 4.2). Finally, we prove that the convergence rate of the False Discovery Proportion (FDP) achieved by plug-in procedures of the form $\text{BH95}(\alpha/\hat{\pi}_0)$ are determined by the convergence rate of $\hat{\pi}_0$ (Section 4.3) and conclude by studying convergence rates in Gaussian and Laplace location models (Section 4.4).

4.1 Consistency, purity and criticality

Since we are focusing on estimators of $g(1) = \pi_0 + (1 - \pi_0)g_1(1)$, a necessary condition for such an estimator to be consistent for π_0 is

$$g_1(1) = 0.$$

This is the *purity* condition introduced by Genovese and Wasserman (2004). Criticality is related to the behavior of g_1 at 0, and purity is related to the behavior of g_1 at 1. We begin by identifying a connection between purity and criticality in one-sided symmetric location

models, where we have $f_1(x) = f_0(x - \theta)$ (location model) and $f_0(-x) = f_0(x)$ (symmetry, i.e. Condition 2).

Lemma 12 (Likelihood ratios in symmetric location models) *Consider the multiple location problem in which test statistics have densities f_0 under \mathcal{H}_0 , and $f_1 = f_0(\cdot - \theta)$ under \mathcal{H}_1 for some $\theta \neq 0$. Under Condition 2 (symmetry), we have*

$$\lim_{x \rightarrow -\infty} \frac{f_0(x)}{f_1(x)} = \lim_{x \rightarrow +\infty} \frac{f_1(x)}{f_0(x)}.$$

Proposition 13 (Purity and criticality in one-sided symmetric location models) *Let g_1^+ be the density of one-sided p -values under the alternative hypothesis, and α^* the critical value of the BH95 procedure for the corresponding multiple comparison problem. Under Conditions 1 and 2,*

1. *Non criticality and purity are equivalent: $\alpha^* = 0$ if and only if $g_1^+(1) = 1$;*
2. *If $\lim_{x \rightarrow +\infty} \frac{f_1(x)}{f_0(x)}$ is finite, then $\alpha^* = (\pi_0 + (1 - \pi_0)g_1^+(0))^{-1}$ and $g^+(1) = \pi_0 + (1 - \pi_0)g_1^+(1)$ are connected by $g_1^+(0)g_1^+(1) = 1$.*

Going back to the examples of Section 3, as there is no criticality in the Gaussian case, the purity condition is verified and π_0 can be consistently estimated using an estimator of $g_1^+(1)$. In the Laplace case, there is a positive critical value, given by

$$\alpha^* = \frac{1}{\pi_0 + (1 - \pi_0)e^\theta}$$

for one-sided p -values, and π_0 cannot be consistently estimated based on the behavior of the p -values at 1 because $g^+(1) = \pi_0 + (1 - \pi_0)e^{-\theta} > \pi_0$. The situation is markedly different for two-sided p -values, as $g_1(1)$ is not determined by the behavior of the likelihood ratio at $+\infty$ but at 0:

Proposition 14 (Two-sided symmetric multiple testing problems are generally impure) *Let g_1^\pm be the density of the two-sided p -values under the alternative hypothesis, and α^* the critical value of this multiple comparison problem. Under Condition 2 (symmetry), we have:*

$$g_1^\pm(1) = \frac{f_1}{f_0}(0).$$

Proposition 14 directly follows from equation (6), combined with the fact that $F_0^{-1}(1/2) = 0$ in symmetric models. As a consequence, *criticality and purity are not equivalent for two-sided p -values*. For example, the two-sided Gaussian location problem is always impure: $g^\pm(1) = e^{-\theta^2/2}$ but has no criticality: $\lim_{u \rightarrow 0} 1/g_1^\pm(u) = 0$.

4.2 Asymptotic properties of non-parametric estimators of π_0

In this section we consider non-parametric estimators of π_0 based on the distribution of the p -values near 1, and show how asymptotic properties of such estimators are driven by the regularity of g near 1. As discussed in Section 4.1, such estimators may or may not achieve consistency, depending on whether the purity condition $g_1(1) = 0$ is met. We let $\overline{\pi_0} = g(1)$, that is,

$$\overline{\pi_0} = \pi_0 + (1 - \pi_0)g_1(1).$$

4.2.1 NON-PARAMETRIC ESTIMATORS OF π_0 WITH KNOWN CONVERGENCE RATES

To the best of our knowledge, the only non-parametric estimators of π_0 for which convergence rates have been established in our setting are those proposed by Storey (2002), Swanepoel (1999) and Hengartner and Stark (1995). The use of these estimators in the context of multiple testing problems was discussed by Genovese and Wasserman (2004). We briefly review the asymptotic properties of these estimators stated in Genovese and Wasserman (2004).

Storey's estimator. Adapting a method originally proposed by Schweder and Spjøtvoll (1982), Storey (2002) defined $\hat{\pi}_0(\lambda) = \frac{1 - \hat{G}_m(\lambda)}{1 - \lambda}$ for $0 \leq \lambda < 1$. As a smooth functional of the empirical distribution of the p -values, this estimator has the following asymptotic distribution, provided that $G(\lambda) < 1$ (Genovese and Wasserman, 2004):

$$\sqrt{m} \left(\hat{\pi}_0(\lambda) - \frac{1 - G(\lambda)}{1 - \lambda} \right) \rightsquigarrow \mathcal{N} \left(0, \frac{G(\lambda)(1 - G(\lambda))}{(1 - \lambda)^2} \right)$$

This estimator converges at the parametric rate $1/\sqrt{m}$, and it is asymptotically biased (even if the purity condition is met) because $\frac{1 - G(\lambda)}{1 - \lambda} > \bar{\pi}_0$ for $\lambda < 1$.

Confidence envelopes for the density. Hengartner and Stark (1995) derived a finite sample confidence envelope for a monotone density. Assuming that G is concave and that g is Lipschitz in a neighborhood of 1, the resulting estimator, which we denote by $\hat{\pi}_0^{\text{HS}}$, converges to $\bar{\pi}_0$ at rate $(\ln m)^{1/3} m^{-1/3}$.

Spacings-based estimator. Swanepoel (1999) proposed a two-step estimator of the minimum of an unknown density based on the distribution of the spacings between observations: first, the location of the minimum is estimated, and then the density at this point is itself estimated. Assuming that at the value at which the density g achieves its minimum, g and \dot{g} are null, and \ddot{g} is bounded away from 0 and $+\infty$ and Lipschitz, then for any $\delta > 0$, there exists an estimator converging at rate $(\ln m)^\delta m^{-2/5}$ to the true minimum.

In our setting, the Lipschitz condition on \ddot{g} is unnecessary: the minimum of g is necessarily achieved at 1 because g is non-increasing (under Condition 1), so the first step of the estimation may be omitted. The corresponding estimator is denoted by $\hat{\pi}_0^{\text{sw}}$.

4.2.2 ASYMPTOTIC PROPERTIES AND REGULARITY NEAR 1

We will show in this section that the differences in the asymptotic properties of these estimators of π_0 in our context are in fact driven by the differences in the regularity assumptions that were made, rather than by the specific form of the estimators. As these estimators are essentially estimators of $g(1)$, their asymptotic properties are driven by the regularity of g near 1.

Storey's estimator is asymptotically biased (even if the purity condition is met) because $\frac{1 - G(\lambda)}{1 - \lambda} > \bar{\pi}_0$ for $\lambda < 1$. In order to make this estimator consistent for the estimation of $\bar{\pi}_0$, we let $h = 1 - \lambda$ go to 0 as the number m of tested hypotheses goes to $+\infty$. The asymptotic bias and variance of the corresponding estimator are derived in Proposition 15:

Proposition 15 (Asymptotic bias and variance of $\hat{\pi}_0(1 - h_m)$) *Let*

$$\hat{\pi}_0(\lambda) = \frac{1 - \hat{G}_m(\lambda)}{1 - \lambda}$$

for $0 < \lambda < 1$. Let h_m be a positive sequence such that $h_m \rightarrow 0$.

1. if $mh_m \rightarrow +\infty$ as $m \rightarrow +\infty$, then

$$\sqrt{mh_m} (\hat{\pi}_0(1 - h_m) - \mathbb{E}[\hat{\pi}_0(1 - h_m)]) \rightsquigarrow \mathcal{N}(0, \overline{\pi}_0).$$

2. Assume that g_1 is k times differentiable at 1, with $g_1^{(l)}(1) = 0$ for $1 \leq l < k$ and $g_1^{(k)}(1) \neq 0$. Then

$$\mathbb{E}[\hat{\pi}_0(1 - h_m)] - \overline{\pi}_0 \underset{m \rightarrow +\infty}{\sim} (1 - \pi_0) \frac{(-1)^k g_1^{(k)}(1)}{(k + 1)!} h_m^k.$$

Proposition 15 shows that the bias of Storey's original estimator, $\overline{\pi}_0 - \frac{1-G(\lambda)}{1-\lambda}$, can be canceled out asymptotically by letting $h_m = 1 - \lambda \rightarrow 0$, at the price of a reduction of the convergence rate: the bandwidth h_m balances bias and variance of Storey's estimator. Moreover, if the purity condition is met, then we have $\overline{\pi}_0 = \pi_0$ and $\hat{\pi}_0(1 - h_m)$ is a consistent estimator of π_0 .

Note that only the bias term in Proposition 15 depends on the regularity of the distribution: the asymptotic bias is of order h_m^k , while the asymptotic variance of $\hat{\pi}_0(1 - h_m)$ is of order $(mh_m)^{-1}$, regardless of the regularity of the distribution. A natural way to resolve this bias/variance trade-off when the regularity of the distribution is known is to calibrate h_m such that the Mean Squared Error (MSE) of the corresponding estimator is minimum.

Proposition 16 (Optimal bandwidth — Storey's estimator) *Assume that g_1 is k times differentiable at 1, with $g_1^{(l)}(1) = 0$ for $1 \leq l < k$, and $g_1^{(k)}(1) \neq 0$, for some positive integer k .*

1. The optimal bandwidth in terms of MSE is of order $h_m^*(k) = m^{-\frac{1}{2k+1}}$, and the corresponding MSE is of order $m^{-\frac{2k}{2k+1}}$.
2. Taking $h_m(k) = h_m^*(k)\eta_m^2$, where $\eta_m \rightarrow 0$ as $m \rightarrow +\infty$, we have

$$m^{\frac{k}{2k+1}} \eta_m (\hat{\pi}_0(1 - h_m(k)) - \overline{\pi}_0) \rightsquigarrow \mathcal{N}(0, \overline{\pi}_0).$$

As a consequence, if we allow the parameter λ of Storey's estimator to go to 1 as $m \rightarrow +\infty$, the resulting estimator essentially achieves the same convergence rates as $\hat{\pi}_0^{\text{Sw}}$ and $\hat{\pi}_0^{\text{HS}}$:

$\hat{\pi}_0^{\text{HS}}$: Assume that g_1 is differentiable at 1. This is a slightly stronger assumption than made by [Hengartner and Stark \(1995\)](#). Then Proposition 16 with $k = 1$ ensures that the convergence rate of Storey's estimator with bandwidth $m^{-1/3}/\eta_m^2$, where $\eta_m = (\ln m)^{-1/3}$ goes to 0 as $m \rightarrow +\infty$, is $(\ln m)^{1/3} m^{-1/3}$. This is the convergence rate of Hengartner and Stark's estimator.

$\hat{\pi}_0^{\text{Sw}}$: Assume that g_1 is twice differentiable at 1. Then Proposition 16 with $k = 2$ ensures that the convergence rate of Storey's estimator with bandwidth $m^{-1/5}\eta_m^2$, where $\eta_m = (\ln m)^{-\delta}$ goes to 0 as $m \rightarrow +\infty$ is $(\ln m)^\delta m^{-2/5}$ for any fixed $\delta > 0$. This is the convergence rate of Swanepoel's estimator.

4.2.3 KERNEL ESTIMATORS

The examples developed so far in this section illustrate the fact that the convergence rates of $\hat{\pi}_0$ are determined by the regularity of g near 1. The convergence rates we obtained are typical convergence rates for non-parametric estimators. We prove that the same type of result holds for a broad class of kernel estimators of $g(1)$.

Definition 17 (Kernel of order ℓ) A kernel of order $\ell \in \mathbb{N}$ is a function $K : \mathbb{R} \rightarrow \mathbb{R}$ such that the functions $u \mapsto u^j K(u)$ are integrable for any $j = 0 \dots \ell$, and verify $\int_{\mathbb{R}} K = 1$, and $\int_{\mathbb{R}} u^j K(u) = 0$ for $j = 1 \dots \ell$.

Definition 18 (Kernel estimator of a density) The kernel estimator of a density g at p_0 based on m independent, identically distributed observations P_1, \dots, P_m from g is defined by

$$\hat{g}(p_0) = \frac{1}{mh} \sum_{i=1}^m K\left(\frac{P_i - p_0}{h}\right),$$

where $h > 0$ is called the bandwidth of the estimator, and K is a kernel.

The estimator proposed by Storey (2002) is a kernel estimator with asymmetric rectangular kernel of order 0, and bandwidth $1 - \lambda$. We generalize the results obtained so far in Section 4.2 to more general kernels. Tsybakov (2009) established lower bounds on the convergence rate of kernel estimators of $g(1)$, depending on the regularity of g at 1. If g is k times differentiable at 1, with $g^{(k)}(1) \neq 0$, considering a kernel estimator $\hat{g}(1)$ associated with a k^{th} order kernel and fixed bandwidth h , the asymptotic variance of $\hat{g}(1)$ is of order $\frac{1}{mh}$, and the asymptotic bias of $\hat{g}(1)$ is of order h^k .

As for the special case of Storey's estimator, optimal convergence results for kernel estimators may be obtained by letting h to go to 0 as $m \rightarrow +\infty$ and balancing asymptotic bias and variance in order to minimize MSE.

Proposition 19 (Optimal bandwidth — k^{th} order kernel estimator Tsybakov (2009)) Assume that g is k times differentiable at 1, with $g^{(k)}(1) \neq 0$. Let $\hat{g}(1)$ be a kernel estimator with bandwidth h_m , associated with a k^{th} order kernel.

1. The optimal bandwidth for $\hat{g}(1)$ in terms of MSE is of order $h_m^*(k) = m^{-\frac{1}{2k+1}}$, and the corresponding MSE is of order $m^{-\frac{2k}{2k+1}}$.
2. Taking $h_m(k) = h_m^*(k)\eta_m^2$, where $\eta_m \rightarrow 0$ as $m \rightarrow +\infty$, we have

$$m^{\frac{k}{2k+1}} \eta_m (\hat{g}(1) - g(1)) \rightsquigarrow \mathcal{N}(0, g(1)).$$

As a consequence, the convergence rate of the optimal kernel estimator of $g(1)$ directly depends on the regularity k of g at 1. Using this class of kernel estimators, we obtain the same convergence rates as in the special cases of Storey's, Hengartner and Stark's, or Swanepoel's estimators, under essentially the same regularity conditions, that is, k^{th} order differentiability of the distribution g_1 of the p -values under the alternative. The only difference is that the assumption that the k first derivatives of g_1 are null at 1 for Storey's estimator is not needed for the kernel estimators used here, as they are k^{th} order kernels.

4.3 Convergence rate of plug-in procedures

We illustrate a connection between the selection and the estimation problem, which can be viewed as a motivation for the estimation of π_0 in multiple testing problems. As noted in Section 2.2, the BH95 procedure at level α controls FDR at level $\pi_0\alpha$; this triggered the development of plug-in versions of this procedure (Benjamini and Hochberg, 2000), which estimate π_0 by $\hat{\pi}_0 < 1$, and apply the BH95 procedure at $\alpha/\hat{\pi}_0$, yielding a larger number of significant hypotheses for the same target FDR level. We elaborate on this connection between selection and estimation by showing that the convergence rate of a given estimator

$\hat{\pi}_0$ of π_0 determines the asymptotic FDR controlling capabilities of the corresponding plug-in procedure.

The False Discovery Proportion (FDP) achieved by a broad class of FDR controlling procedures including plug-in procedures based on the BH95 procedure has been studied by [Neuvial \(2008\)](#), for estimators of π_0 that depend on the observations only through the empirical distribution function of the p -values ([Storey, 2002](#); [Storey et al., 2004](#); [Benjamini et al., 2006](#)). As a consequence, their convergence rate is $1/\sqrt{m}$ and the FDP achieved by the corresponding plug-in procedures also converges at rate $1/\sqrt{m}$ to their asymptotic FDR in the sub-critical case ([Neuvial, 2008](#)). However, these estimators are not consistent for $\overline{\pi}_0 = g(1)$, as they are essentially estimators of $\frac{1-G(u_0)}{1-u_0}$, at some $u_0 < 1$ (for example, $u_0 = \lambda$ for Storey's estimator). Conversely, the estimators studied in [Section 4.2](#) achieve consistent estimation of $\overline{\pi}_0$, at the price of a slower convergence rate. Here we show how these results translate in terms of asymptotic FDR control: the plug-in procedure $\text{BH95}(\alpha/\hat{\pi}_0)$ asymptotically controls FDR at level $\pi_0\alpha/\overline{\pi}_0$, and the convergence rate of its FDP is the same as that of $\hat{\pi}_0$.

We use the same notation as in [Neuvial \(2008\)](#) and refer to this paper for more detailed explanation. In particular, we assume without loss of generality that π_0 does not depend on m , although strictly speaking we should write $\pi_0 = m_0(m)/m$, where the number $m_0(m)$ of true null hypotheses depends on the total number m of hypotheses tested, and π_0 is only the limit of $m_0(m)/m$ as $m \rightarrow +\infty$. We study the BH95 procedure at level $\alpha/\hat{\pi}_0$, where $\hat{\pi}_0$ is an estimator of π_0 that converges to π_0 at rate $\sqrt{mh_m}$, where $h_m \rightarrow 0$. This procedure rejects all hypotheses with p -values smaller than

$$\hat{\tau} = \sup \left\{ t \in [0, 1], \widehat{\mathbb{G}}_m(t) \geq \hat{\pi}_0 t / \alpha \right\} .$$

The associated proportion of rejections and proportion of incorrect rejections are given by $\hat{\rho} = \widehat{\mathbb{G}}_m(\hat{\tau}) = \hat{\tau}\hat{\pi}_0/\alpha$, and $\hat{\nu} = \pi_0 \widehat{\mathbb{G}}_{0,m}(\hat{\tau})$, respectively, where $\widehat{\mathbb{G}}_{0,m}$ denotes the empirical distribution function of p -values that correspond to true null hypotheses. We define the corresponding asymptotic threshold as the threshold of the $\text{BH95}(\alpha/\overline{\pi}_0)$ procedure:

$$\tau^* = \sup \left\{ t \in [0, 1], G(t) \geq \overline{\pi}_0 t / \alpha \right\} .$$

Note that by the definition of $\hat{\tau}$ and τ^* , we have $\widehat{\mathbb{G}}_m(\hat{\tau}) = \hat{\pi}_0 \hat{\tau} / \alpha$ and $G(\tau^*) = \overline{\pi}_0 \tau^* / \alpha$. The following Proposition shows that the convergence rate of $(\hat{\tau}, \hat{\nu}, \hat{\rho})$ is driven by the convergence rate of $\hat{\pi}_0$.

Proposition 20 *Let α^* be the critical value of the BH95 procedure. Let $\alpha > \overline{\pi}_0 \alpha^*$, and $\hat{\pi}_0$ be an estimator of π_0 with asymptotic distribution given by $\sqrt{mh_m}(\hat{\pi}_0 - \pi_0) \rightsquigarrow \mathcal{N}(0, v(\overline{\pi}_0))$ for some function v , where $h_m = o(1/\ln \ln m)$. Then, as $m \rightarrow +\infty$,*

$$\begin{pmatrix} \hat{\tau} \\ \hat{\nu} \\ \hat{\rho} \end{pmatrix} - \begin{pmatrix} \tau^* \\ \pi_0 \tau^* \\ \overline{\pi}_0 \tau^* / \alpha \end{pmatrix} = \frac{\tau^* / \alpha}{g(\tau^*) - \overline{\pi}_0 / \alpha} \begin{pmatrix} 1 \\ \pi_0 \\ g(\tau^*) \end{pmatrix} (\hat{\pi}_0 - \overline{\pi}_0) (1 + o_P(1)) .$$

Note that $\overline{\pi}_0 \alpha^*$ is the critical value of the $\text{BH95}(\alpha/\hat{\pi}_0)$ as long as $\hat{\pi}_0$ converges in probability to $\overline{\pi}_0$. Therefore the condition $\alpha > \overline{\pi}_0 \alpha^*$ simply ensures that we are not in a critical situation for the $\text{BH95}(\alpha/\hat{\pi}_0)$ procedure. Combined with the fact that the FDP achieved by a multiple testing procedure is a smooth function of the proportion of rejections $\hat{\rho}$ and the proportion of incorrect rejections $\hat{\nu}$, [Theorem 21](#) implies that the convergence rate of $\hat{\pi}_0$ determines the convergence rate of the FDP of the associated plug-in procedure:

Theorem 21 (Asymptotic FDP for plug-in procedures) *Let $\hat{\pi}_0$ be an estimator of π_0 with asymptotic distribution given by*

$$\sqrt{mh_m}(\hat{\pi}_0 - \pi_0) \rightsquigarrow \mathcal{N}(0, v(\pi_0))$$

for some function v , with $h_m = o(1/\ln \ln m)$. Consider the plug-in procedure based on $\hat{\pi}_0$, which applies the BH95 procedure at level $\alpha/\hat{\pi}_0$. Let α^* be the critical value of the standard BH95(α) procedure. Under Condition 1 (concavity of G), for any $\alpha > \pi_0\alpha^*$, the asymptotic distribution of the FDP achieved by the BH95($\alpha/\hat{\pi}_0$) procedure is given by

$$\sqrt{mh_m} \left(\text{FDP} - \frac{\pi_0\alpha}{\pi_0} \right) \rightsquigarrow \mathcal{N} \left(0, \left(\frac{\pi_0\alpha}{\pi_0} \right)^2 \frac{v(\pi_0)}{\pi_0^2} \right).$$

As a consequence, the FDP of plug-in BH95 procedures associated with the estimators of π_0 studied in Section 4.2 can be derived by combining the results of Proposition 16, Proposition 19 and Theorem 21.

Corollary 22 *Assume that Condition 1 (concavity) holds and that g_1 is k times differentiable at 1 with $g_1^{(k)}(1) \neq 0$. Let $h_m(k) = m^{-\frac{1}{2k+1}}\eta_m^2$, where $\eta_m \rightarrow 0$ as $m \rightarrow +\infty$. Further assume that we are in one of the following two situations:*

1. $\hat{\pi}_0 = \frac{1 - \hat{G}_m(1 - h_m(k))}{h_m(k)}$, and $g_1^{(l)}(1) = 0$ for $1 \leq l < k$;
2. $\hat{\pi}_0$ is a kernel estimator of g associated with a k^{th} order kernel with bandwidth $h_m(k)$.

Let α^* be the critical value of the BH95 procedure. Then, for any $\alpha > \pi_0\alpha^*$, the asymptotic distribution of the FDP achieved by the BH95($\alpha/\hat{\pi}_0$) procedure is given by

$$m^{\frac{k}{2k+1}}\eta_m \left(\text{FDP} - \frac{\pi_0\alpha}{\pi_0} \right) \rightsquigarrow \mathcal{N} \left(0, \frac{\pi_0^2\alpha^2}{\pi_0^3} \right).$$

In particular, if the purity condition is met, the asymptotic FDP achieved by the estimators in Corollary 22 is exactly α (and the asymptotic variance is α^2/π_0).

4.4 Regularity of g_1 at 1

We conclude this section by studying convergence rates in one- and two-sided Gaussian and Laplace multiple comparison problems.

4.4.1 TWO-SIDED PROBLEMS

Proposition 23 (Behavior of g_1 at 1 in two-sided symmetric models) *Under Condition 2, we have :*

1.

$$g_1^\pm(1) = \frac{f_1}{f_0}(0).$$

2. *If the likelihood ratio $\frac{f_1}{f_0}$ is differentiable in a neighborhood of 0, the density g_1^\pm of the two-sided p -values under the alternative hypothesis verifies*

$$g_1^{\pm(1)}(1) = 0.$$

3. If $\frac{f_1}{f_0}$ is twice differentiable in a neighborhood of 0, then we have

$$g_1^{\pm(2)}(1) = \frac{1}{4f_0(0)^2} \left(\frac{f_1}{f_0} \right)^{(2)}(0).$$

For illustration we apply this result to Storey's estimator in Gaussian two-sided location models, where the likelihood ratio is given by $\frac{f_1}{f_0}(t) = \exp(-\theta^2/2 + \theta t)$. By Proposition 23, we have $g_1^{\pm}(1) = e^{-\theta^2/2}$, $g_1^{\pm(1)}(1) = 0$, and $g_1^{\pm(2)}(1) > 0$. Therefore, we can use Corollary 22 with $k = 2$ to derive the following result:

Proposition 24 (Two-sided Gaussian test statistics) *Assume that test statistics are distributed as $\mathcal{N}(0, 1)$ under the null hypothesis and as $\mathcal{N}(\theta, 1)$ under the alternative. Let $\hat{\pi}_0$ be Storey's estimator with bandwidth $h_m = m^{-1/5}\eta_m^2$, where $\eta_m \rightarrow 0$ as $m \rightarrow +\infty$, that is,*

$$\hat{\pi}_0 = \frac{1 - \widehat{\mathbb{G}}_m(1 - h_m)}{h_m}.$$

Then, letting $\overline{\pi}_0 = \pi_0 + (1 - \pi_0)e^{-\theta^2/2}$, we have:

1.

$$m^{2/5}\eta_m(\hat{\pi}_0 - \overline{\pi}_0) \rightsquigarrow \mathcal{N}(0, \overline{\pi}_0).$$

2. Let α^* be the critical value of the BH95 procedure. For any $\alpha > \overline{\pi}_0\alpha^*$, the asymptotic distribution of the FDP achieved by the BH95($\alpha/\hat{\pi}_0$) procedure is given by

$$m^{2/5}\eta_m \left(\text{FDP} - \frac{\pi_0\alpha}{\overline{\pi}_0} \right) \rightsquigarrow \mathcal{N} \left(0, \frac{\pi_0^2\alpha^2}{\overline{\pi}_0^3} \right).$$

Note that Proposition 23 cannot be applied to two-sided Laplace statistics as the likelihood ratio $\frac{f_1}{f_0}(t) = \exp(|t - \theta| - |t|)$ has a singularity at $t = 0$; in this particular case the distribution of the two-sided p -values can be calculated directly by combining Corollary 3 with Proposition 25 (4):

$$g_1^{\pm}(u) = \frac{e^{-\theta}}{2} \left(1 + \frac{1}{4u^2} \right).$$

Therefore, we have $g_1^{\pm(1)}(1) \neq 0$, and the optimal bandwidth in Corollary 22 is of order $m^{-1/3}$ instead of $m^{-1/5}$.

4.4.2 ONE-SIDED PROBLEMS

For one-sided p -values in symmetric models, we have proved that using the class of estimators of π_0 studied here, consistency can be achieved if and only if there is no criticality (Proposition 13). For Laplace test statistics, criticality occurs, and the distribution of one-sided p -values satisfies $G_1^+(u) = e^{-\theta}u$ for $u \geq 1/2$. Therefore, for $u \geq 1/2$, $\frac{1-G^+(u)}{1-u}$ is constant, equal to $\overline{\pi}_0 = \pi_0 + (1 - \pi_0)e^{-\theta}$, as illustrated by the solid curves in Figure 3. Therefore, Storey's estimator with any $\lambda \geq 1/2$ is an unbiased estimator of $\overline{\pi}_0$, which converges to $\overline{\pi}_0$ at rate $1/\sqrt{m}$.

In the Gaussian case, there is no criticality, so π_0 can be consistently estimated from one-sided p -values, but the regularity of g_1^+ at 1 is poor: we have

$$g_1^+(x) = \exp \left(-\frac{\theta^2}{2} - \theta\Phi^{-1}(x) \right),$$

where $\Phi(=F_0)$ denotes the standard Gaussian distribution function. As $h \rightarrow 0$, $\Phi^{-1}(1-h) \leq \sqrt{2 \ln(1/h)}$, so that

$$g_1^+(1-h) \geq \exp\left(-\frac{\theta^2}{2} - \theta\sqrt{2 \ln(1/h)}\right).$$

This implies that g_1^+ is not differentiable at 1, meaning that consistent estimators of π_0 based on the p -values close to 1 have convergence rates slower than $m^{-1/3}$ in our setting. This difference of behavior between the one- and two-sided Gaussian multiple testing problems is illustrated by Figure 8 for the simplest location model: $\mathcal{N}(0, 1)$ against $\mathcal{N}(1, 1)$.

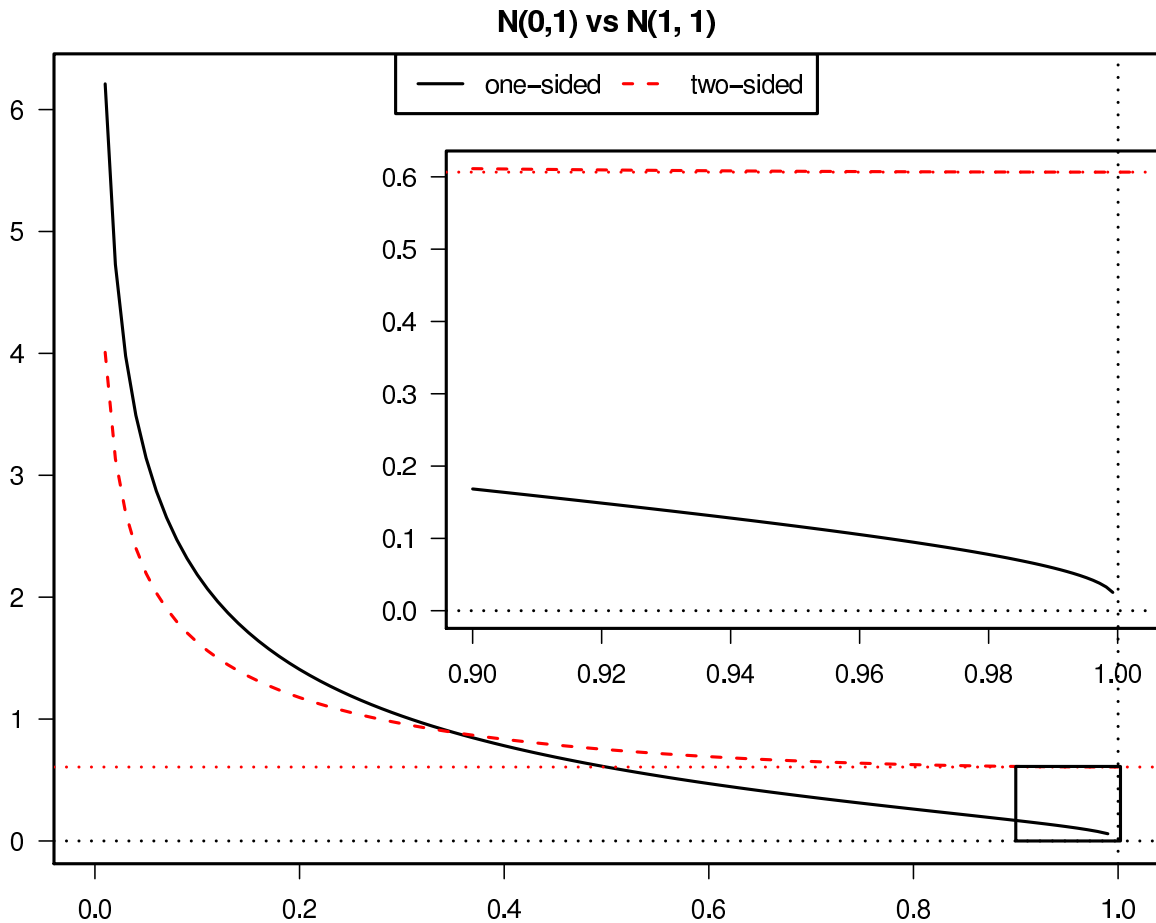


Figure 8: *Density of one- and two-sided p -values under the alternative hypothesis for the location model $\mathcal{N}(0, 1)$ versus $\mathcal{N}(1, 1)$. Inside plot: zoom in the region $[0.9, 1]$, which is highlighted by a black box in the main plot.*

The density of two-sided p -values has a positive limit at 1, and its derivative at 1 is 0, making it possible to estimate the $\pi_0 + (1 - \pi_0)e^{-\theta^2/2}$ at rate $m^{-2/5}$. Conversely, the density of one-sided p -values goes to 0 at 1, but is not differentiable: so that the true π_0 can be estimated but the convergence rate is much slower.

4.4.3 PRACTICAL IMPLICATIONS

In practice, when $g_1(1) > 0$, its value depends on the settings of the multiple comparison problem: for example, in the Student two-sided problem, we have $g_1^\pm(1) = \frac{f_1}{f_0}(0) = e^{-\theta^2/2}$ by Proposition 9, that is, with notation of Section 3,

$$g_1^\pm(1) = \exp\left(-\frac{\delta^2}{1/n_X + 1/n_Y}\right).$$

In particular, $g_1^\pm(1)$ considerably depends on the longitudinal sample size $n_X + n_Y$. For the Golub data set, we have $n_X = 27$, $n_Y = 11$, and we estimated $\theta = 2.5$, corresponding to an effect size $\delta = 0.9$. These parameters yield $g_1^\pm(1) \approx 0.04$, but if we consider a situation where the sample size is twice bigger ($n_X = 54$, $n_Y = 22$), we get $g_1^\pm(1) \approx 0.002$ for the same effect size, and the bias $\overline{\pi_0} - \pi_0$ is negligible in practice. These remarks suggest that it would be interesting to conduct a study of the bias/variance trade-off in the estimation of π_0 that would explicitly take longitudinal sample size into account.

Recent work suggest two alternative research directions for estimating π_0 :

- one-stage adaptive procedures as proposed by [Blanchard and Roquain \(to appear\)](#) and [Finner et al. \(2009\)](#) allow more powerful FDR control than the standard BH95 procedure without explicitly incorporating an estimate of π_0 : they are not plug-in procedures.
- [Jin \(2008\)](#) proposed an estimator of π_0 based on the Fourier transform of the empirical function of the p -values, which does not focus on the behavior of the density near 1, and might not suffer from the same limitations as the estimators studied here. In particular this estimator was shown to be consistent for the estimation of π_0 when the p -values (not the test statistics) follow a Gaussian location mixture.

Appendix: proofs

Appendix A. Proofs of section 2

Proof [Proof of Proposition 1] Let $x \in \mathbb{R}$. We have $p^+(x) = \mathbb{P}_{\mathcal{H}_0}(X \geq x) = 1 - F_0(x)$ by the definition of F_0 . Then, for any $u \in [0, 1]$, we have

$$\begin{aligned} G_1^+(u) &= \mathbb{P}_{\mathcal{H}_1}(p^+(X) \leq u) \\ &= \mathbb{P}_{\mathcal{H}_1}(F_0(X) \geq 1 - u) . \\ &= \mathbb{P}_{\mathcal{H}_1}(X \geq q_0(u)) , \end{aligned}$$

where $q_0 : u \mapsto F_0^{-1}(1 - u)$. Therefore $G_1^+(u) = 1 - F_1(q_0(u))$, which proves Equation (1). Equation (2) then follows from the fact that $q_0'(u) = -1/f_0(q_0(u))$. ■

Proof [Proof of Proposition 2] Let $x \in \mathbb{R}$. We have

$$\begin{aligned} p^\pm(x) &= \mathbb{P}_{\mathcal{H}_0}(|X| \geq |x|) \\ &= \mathbb{P}_{\mathcal{H}_0}(X \geq |x|) + \mathbb{P}_{\mathcal{H}_0}(X \leq -|x|) \\ &= 1 - F_0(|x|) + F_0(-|x|) \\ &= 2(1 - F_0(|x|)) , \end{aligned}$$

as F_0 is assumed to be symmetric. Then, for any $u \in [0, 1]$, we have

$$\begin{aligned} G_1^\pm(u) &= \mathbb{P}_{\mathcal{H}_1}(p^\pm(X) \leq u) \\ &= \mathbb{P}_{\mathcal{H}_1}(F_0(|X|) \geq 1 - u/2) . \\ &= \mathbb{P}_{\mathcal{H}_1}(X \geq q_0(u/2)) + \mathbb{P}_{\mathcal{H}_1}(X \leq -q_0(1 - u/2)) , \end{aligned}$$

where $q_0 : u \mapsto F_0^{-1}(1 - u)$. Therefore $G_1^\pm(u) = 1 - F_1(q_0(u/2)) + F_1(q_0(1 - u/2))$, which proves Equation (3). Equation (4) then follows from the fact that $q_0'(u) = -1/f_0(q_0(u))$. ■

Appendix B. Proofs of section 3

Laplace distribution

Proposition 25 (One-sided Laplace problem) *Assume that the pdf of the test statistics is $f_0 : x \mapsto \frac{1}{2}e^{-|x|}$ under the null hypothesis, and $f_1 : x \mapsto \frac{1}{2}e^{-|x-\theta|}$ under the alternative, with $\theta > 0$ (one-sided test). Then*

1. The one-sided p-value function is

$$1 - F_0(x) = \begin{cases} \frac{1}{2}e^{-|x|} & \text{if } x \geq 0 \\ 1 - \frac{1}{2}e^{-|x|} & \text{if } x < 0 \end{cases}$$

2. The inverse one-sided p-value function is

$$(1 - F_0)^{-1}(u) = \begin{cases} \ln\left(\frac{1}{2u}\right) & \text{if } 0 \leq u \leq \frac{1}{2} \\ \ln(2(1 - u)) & \text{if } \frac{1}{2} < u < 1 \end{cases}$$

3. The cdf of one-sided p-values under \mathcal{H}_1 is

$$G_1^+(u) = \begin{cases} ue^\theta & \text{if } 0 \leq u \leq \frac{e^{-\theta}}{2} \\ 1 - \frac{1}{4u}e^{-\theta} & \text{if } \frac{e^{-\theta}}{2} \leq u \leq \frac{1}{2} \\ 1 - (1 - u)e^{-\theta} & \text{if } u \geq \frac{1}{2} \end{cases}$$

4. The pdf of one-sided p -values under \mathcal{H}_1 is

$$g_1^+(u) = \begin{cases} e^\theta & \text{if } 0 \leq u \leq \frac{e^{-\theta}}{2} \\ \frac{1}{4u^2}e^{-\theta} & \text{if } \frac{e^{-\theta}}{2} \leq u \leq \frac{1}{2} \\ e^{-\theta} & \text{if } u \geq \frac{1}{2} \end{cases}$$

Proof [Proof of Proposition 25] The inverse p -value function directly follows from the p -value function and the pdf of the p -values follows from the cdf, so we only prove (1) (p -value function), and (3) (cdf of the p -values).

Proof of (1). We have $f_0(x) = \frac{1}{2}e^{-|x|}$. Therefore, for $x < 0$, $F_0(x) = \int_{-\infty}^x \frac{1}{2}e^t dt = \frac{1}{2}e^{-|x|}$. For $x > 0$, $F_0(x) = \int_{-\infty}^0 \frac{1}{2}e^t dt + \int_0^x \frac{1}{2}e^{-t} dt = 1 - \frac{1}{2}e^{-|x|}$.

Proof of (3). Let $u \in [0, 1]$. The distribution function of one-sided p -values is given by

$$\begin{aligned} G_1(u) &= \mathbb{P}_\theta(1 - F_0(x) \leq u) \\ &= \mathbb{P}_\theta\left(X \geq (1 - F_0)^{-1}(u)\right) \\ &= \int_{1-F_0(u)}^\theta f_1(x)dx + \int_\theta^{+\infty} f_1(x)dx \\ &= \int_{1-F_0(u)}^\theta \frac{1}{2}e^{-|x-\theta|}dx + \frac{1}{2} \end{aligned}$$

For $u < \frac{1}{2}$, $(1 - F_0)^{-1}(u) = \ln \frac{1}{2u}$ and $(1 - F_0)^{-1}(u) \geq \theta \iff u \leq \frac{e^{-\theta}}{2}$.

Hence if $u \leq \frac{e^{-\theta}}{2}$,

$$\begin{aligned} G_1(u) &= \frac{1}{2} - \frac{1}{2} \int_\theta^{\ln \frac{1}{2u}} e^{-(x-\theta)} dx \\ &= \frac{1}{2} - \frac{1}{2} \left(-e^{-(\ln \frac{1}{2u} - \theta)} - (-1) \right) \\ &= ue^\theta \end{aligned}$$

If $\frac{e^{-\theta}}{2} < u < \frac{1}{2}$,

$$\begin{aligned} G_1(u) &= \frac{1}{2} + \int_{\ln \frac{1}{2u}}^\theta \frac{1}{2}e^{(x-\theta)} dx \\ &= \frac{1}{2} + \frac{1}{2} \left(1 - e^{\ln \frac{1}{2u} - \theta} \right) \\ &= 1 - \frac{1}{4u}e^{-\theta} \end{aligned}$$

Finally, for $u \geq \frac{1}{2}$, $(1 - F_0)^{-1}(u) = \ln 2(1 - u)$. Thus $(1 - F_0)^{-1}(u) \leq \theta \iff u \geq 1 - \frac{e^\theta}{2}$, which always holds for $u \geq \frac{1}{2}$ because $\theta > 0$.

Hence for $u \geq \frac{1}{2}$,

$$\begin{aligned} G_1(u) &= \frac{1}{2} + \int_{\ln 2(1-u)}^{\theta} \frac{1}{2} e^{(x-\theta)} dx \\ &= \frac{1}{2} + \frac{1}{2} \left(1 - e^{\ln 2(1-u) - \theta} \right) \\ &= 1 - (1-u)e^{-\theta} \end{aligned}$$

■

Student distribution

Lemma 26 (Non-central Student distribution) *The density function of a non-central Student distribution with k degrees of freedom and non centrality parameter θ may be written as*

$$f_1(t) = \frac{\Gamma(k+1)}{2^{\frac{k-1}{2}} \Gamma(\frac{k}{2}) \sqrt{k\pi}} \frac{1}{\left(1 + \frac{t^2}{k}\right)^{\frac{k+1}{2}}} \exp\left[-\frac{\theta^2}{2} \frac{1}{1 + \frac{t^2}{k}}\right] Hh_k\left(-\frac{\theta t}{\sqrt{k+t^2}}\right),$$

where

$$Hh_k(z) = \int_0^{+\infty} \frac{x^k}{k!} e^{-\frac{1}{2}(x+z)^2} dx.$$

Proof Let $T_{k,\theta} = \frac{Z_\theta}{\sqrt{U/k}}$, where $Z_\theta \sim \mathcal{N}(\theta, 1)$ and $U \sim \chi^2(k)$, with Z_θ and U independent.

We have $f_1(t) = \frac{d}{dt} (\mathbb{P}(T_{k,\theta} \leq t)) = \frac{d}{dt} (\mathbb{P}(Z_\theta \leq t\sqrt{U/k}))$. As Z_θ and U are independent, we have

$$\begin{aligned} \mathbb{P}(Z_\theta \leq t\sqrt{U/k}) &= \int_{\mathbb{R}} \mathbb{P}(Z_\theta \leq t\sqrt{u/k}) f_U(u) du \\ &= \int_{\mathbb{R}} \Phi(t\sqrt{u/k} - \theta) f_U(u) du \end{aligned}$$

Thus, inverting $\int_{\mathbb{R}}$ and $\frac{d}{dt}$,

$$\begin{aligned} f_1(t) &= \int_{\mathbb{R}} \frac{d}{dt} (\Phi(t\sqrt{u/k} - \theta)) f_U(u) du \\ &= \int_{\mathbb{R}} \sqrt{u/k} \phi(t\sqrt{u/k} - \theta) f_U(u) du \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} (t\sqrt{u/k} - \theta)^2\right] \frac{1}{2^{k/2} \Gamma(k/2)} u^{k/2-1} e^{-\frac{u}{2}} \mathbf{1}_{u>0} du \\ &= \frac{1}{\sqrt{2k\pi}} \frac{1}{2^{\frac{k}{2}} \Gamma(k/2)} \int_{\mathbb{R}_+} \exp\left[-\frac{1}{2} \left(\left(1 + \frac{t^2}{k}\right) u + \theta^2 - 2\theta t\sqrt{u/k}\right)\right] u^{\frac{k-1}{2}} du \end{aligned}$$

Then, using the transformation $v = \sqrt{(1 + \frac{t^2}{k})}u$, we note that

$$\begin{aligned}
 & \int_{\mathbb{R}_+} \exp \left[-\frac{1}{2} \left(\left(1 + \frac{t^2}{k}\right) u + \theta^2 - 2\theta t \sqrt{u/k} \right) \right] u^{\frac{k-1}{2}} du \\
 = & \int_{\mathbb{R}_+} \exp \left[-\frac{1}{2} \left(v^2 + \theta^2 - 2\theta t v \frac{1/\sqrt{k}}{\sqrt{1+t^2/k}} \right) \right] \frac{v^{k-1}}{\left(1 + \frac{t^2}{k}\right)^{\frac{k-1}{2}}} \frac{2v}{1 + \frac{t^2}{k}} dv \\
 = & \int_{\mathbb{R}_+} \exp \left[-\frac{1}{2} \left(\left(v - \theta t \frac{1/\sqrt{k}}{\sqrt{1+t^2/k}} \right)^2 + \theta^2 \left(1 - \frac{t^2/k}{1+t^2/k} \right) \right) \right] \times \frac{2v^k}{\left(1 + \frac{t^2}{k}\right)^{\frac{k+1}{2}}} dv \\
 = & \exp \left[-\frac{\theta^2}{2} \frac{1}{1 + \frac{t^2}{k}} \right] \frac{2}{\left(1 + \frac{t^2}{k}\right)^{\frac{k+1}{2}}} \int_{\mathbb{R}_+} \exp \left[-\frac{1}{2} \left(v - \frac{\theta t}{\sqrt{k+t^2}} \right)^2 \right] v^k dv \\
 = & \exp \left[-\frac{\theta^2}{2} \frac{1}{1 + \frac{t^2}{k}} \right] \frac{2}{\left(1 + \frac{t^2}{k}\right)^{\frac{k+1}{2}}} k! Hh_k \left(\frac{-\theta t}{\sqrt{k+t^2}} \right)
 \end{aligned}$$

Therefore

$$\begin{aligned}
 f_1(t) &= \frac{1}{\sqrt{2k\pi}} \frac{1}{2^{\frac{k}{2}} \Gamma(k/2)} \exp \left[-\frac{\theta^2}{2} \frac{1}{1 + \frac{t^2}{k}} \right] \\
 &\quad \times \frac{2}{\left(1 + \frac{t^2}{k}\right)^{\frac{k+1}{2}}} k! Hh_k \left(\frac{-\theta t}{\sqrt{k+t^2}} \right)
 \end{aligned}$$

which completes the proof because $\Gamma(k+1) = k!$ ■

Proof [Proof of Proposition 9] With notation of Lemma 26, the probability distribution function of central t with k degrees of freedom is given by

$$f_0(t) = \frac{\Gamma(k+1)}{2^{\frac{k-1}{2}} \Gamma(\frac{k}{2}) \sqrt{k\pi}} \frac{1}{\left(1 + \frac{t^2}{k}\right)^{\frac{k+1}{2}}} Hh_k(0),$$

and the likelihood ratio of the model is given by

$$\frac{f_1}{f_0}(t) = \exp \left[-\frac{\theta^2}{2} \frac{1}{1 + \frac{t^2}{k}} \right] \frac{Hh_k \left(-\frac{\theta t}{\sqrt{k+t^2}} \right)}{Hh_k(0)}.$$

The following property of Hh_k is useful to prove that $\frac{f_1}{f_0}$ is non-decreasing ■

Lemma 27

$$H'_{k+1}(z) = -Hh_k(z)$$

Proof [Proof of Lemma 27] Let $k \in \mathbb{N}$. As $H_{k+1}(z) = \int_0^{+\infty} \frac{x^{k+1}}{(k+1)!} e^{-\frac{1}{2}(x+z)^2} dx$, we have

$$\begin{aligned} H'_{k+1}(z) &= \int_0^{+\infty} \frac{x^{k+1}}{(k+1)!} (-(x+z)) e^{-\frac{1}{2}(x+z)^2} dx \\ &= \left[\frac{x^{k+1}}{(k+1)!} e^{-\frac{1}{2}(x+z)^2} \right]_0^{+\infty} - \int_0^{+\infty} \frac{(k+1)x^k}{(k+1)!} e^{-\frac{1}{2}(x+z)^2} dx \\ &= 0 - Hh_k(z) \end{aligned}$$

■

Proof [Proof of Proposition 10]

1. As $t \mapsto \exp\left[-\frac{\theta^2}{2} \frac{1}{1+\frac{t^2}{k}}\right]$ is non-decreasing and $t \mapsto -\frac{\theta t}{\sqrt{k+t^2}}$ is non-increasing, it is sufficient to prove that Hh_k is non-increasing, which follows from lemma 27 because H_{k-1} is positive.
2. by proposition 5 it suffices to note that

$$\lim_{t \rightarrow +\infty} \frac{f_1}{f_0}(t) = \frac{Hh_k(-\theta)}{Hh_k(0)}$$

■

Proof of Proposition 11

Proof [Proof of Proposition 11] By the definition of Π_m we have, for ant $\alpha > 0$,

$$\Pi_m(\alpha) = \frac{1 - V_m(\alpha)/R_m(\alpha)}{1 - \pi_0} R_m(\alpha)/m.$$

When $\alpha > \alpha^*$, the limit in probability of $V_m(\alpha)/R_m(\alpha)$ is the asymptotic FDR achieved by the BH95 procedure, that is, $\pi_0 \alpha$ (Chi, 2007a). When $\alpha < \alpha^*$, the proportion of rejections by the BH95 procedure is asymptotically bounded (Chi, 2007a), so that both Π_∞ and ρ_∞ converge to 0 in probability. ■

Appendix C. Proofs of section 4

Consistency, purity and criticality

Proof [Proof of Lemma 12] We note that

$$\begin{aligned} \frac{f_1(x)}{f_0(x)} &= \frac{f_0(x-\theta)}{f_0(x)} \text{ by definition of a location model} \\ &= \frac{f_0(-x+\theta)}{f_0(-x)} \text{ by symmetry of } f \\ &= \frac{f_0(-x+\theta)}{f_1(-x+\theta)}, \end{aligned}$$

which concludes the proof, as θ is a fixed scalar. ■

Proof [Proof of Proposition 13] We have $\alpha^* = \lim_{u \rightarrow 0} \frac{1}{g(u)}$, where $g = \pi_0 + (1 - \pi_0)g_1$ and

$$g_1(u) = \frac{f_1}{f_0}(-F_0^{-1}(u)).$$

Therefore, as $\lim_{u \rightarrow 0} F_0^{-1}(u) = +\infty$, the result is a consequence of Lemma 12. \blacksquare

Storey's estimator with $\lambda \rightarrow 1$

Proof [Proof of Proposition 15] First, we note that As $G(\lambda) = \pi_0\lambda + (1 - \pi_0)G_1(\lambda)$, we have, for any $\lambda < 1$,

$$\frac{1 - G(\lambda)}{1 - \lambda} = \pi_0 + (1 - \pi_0)\frac{1 - G_1(\lambda)}{1 - \lambda}. \quad (10)$$

1. We demonstrate that $\hat{\pi}_0(\lambda_m)$ may be written as a sum of m independent random variables that satisfy the Lindeberg-Feller conditions for the Central Limit Theorem (Pollard, 1984). Let $Z_i^m = \mathbf{1}_{P_i \geq 1 - h_m}$, where the P_i are the p -values. Z_i^m follows a Bernoulli distribution with parameter $p_m = 1 - G(1 - h_m)$. Denoting

$$Y_i^m = \frac{Z_i^m - \mathbb{E}[Z_i^m]}{\sqrt{mh_m}}$$

we have

$$\sum_{i=1}^m Y_i^m = \sqrt{mh_m}(\hat{\pi}_0(1 - h_m) - \mathbb{E}[\hat{\pi}_0(1 - h_m)]).$$

$(Y_i^m)_{1 \leq i \leq m}$ are centered, independent random variables, with $\text{Var } Y_i^m = \frac{\text{Var } Z_i^m}{mh_m} = \frac{G(1 - h_m)(1 - G(1 - h_m))}{mh_m}$, which, by (10), is equivalent to $\frac{\pi_0}{m}$ as $m \rightarrow +\infty$. Therefore,

$$\lim_{m \rightarrow +\infty} \sum_{i=1}^m \mathbb{E}[(Y_i^m)^2] = \pi_0.$$

Finally we prove that for any $\varepsilon > 0$,

$$\lim_{m \rightarrow +\infty} \sum_{i=1}^m \mathbb{E}[(Y_i^m)^2 \mathbf{1}_{|Y_i^m| > \varepsilon}] = 0.$$

As $Z_i^m \in \{0, 1\}$ and $\mathbb{E}[Z_i^m] \in [0, 1]$, we have $(Y_i^m)^2 \leq \frac{1}{h_m}$, and

$$\begin{aligned} \sum_{i=1}^m \mathbb{E}[(Y_i^m)^2 \mathbf{1}_{|Y_i^m| > \varepsilon}] &\leq \frac{1}{h_m} \mathbb{E}[\mathbf{1}_{|Y_i^m| > \varepsilon}] \\ &= \frac{1}{h_m} \mathbb{P}(|Y_i^m| > \varepsilon) \\ &\leq \frac{1}{h_m} \frac{\text{Var } Y_i^m}{\varepsilon^2} \end{aligned}$$

by Chebycheff's inequality. As $mh_m \rightarrow +\infty$ and $\text{Var } Y_i^m \sim \frac{\pi_0}{m}$ as $m \rightarrow +\infty$, the above sum therefore goes to 0 as $mh_m \rightarrow +\infty$. The Lindeberg-Feller conditions for the Central Limit

Theorem are thus fulfilled, and we have

$$\sum_{i=1}^m Y_i^m \rightsquigarrow \mathcal{N}(0, \bar{\pi}_0),$$

which concludes the proof because $\sum_{i=1}^m Y_i^m = \sqrt{mh_m}(\hat{\pi}_0(1-h_m) - \mathbb{E}[\hat{\pi}_0(1-h_m)])$.

2. By (10), the bias is given by

$$\mathbb{E}[\hat{\pi}_0(\lambda)] - \pi_0 = (1 - \pi_0) \frac{1 - G_1(\lambda)}{1 - \lambda}.$$

A Taylor expansion as $\lambda \rightarrow 1$ yields

$$\begin{aligned} 1 - G_1(\lambda) &= \sum_{l=0}^k \frac{(-1)^l g_1^{(l)}(1)}{(l+1)!} (1-\lambda)^{l+1} + o((1-\lambda)^{l+1}) \\ &= (1-\lambda)g_1(1) + \frac{(-1)^k g_1^{(k)}(1)}{(k+1)!} (1-\lambda)^{k+1} + o((1-\lambda)^{k+1}) \end{aligned}$$

as we assumed that $g_1^{(l)}(1) = 0$ for $1 \leq l \leq k$ and $g_1^{(k)}(1) \neq 0$. Therefore, if $h_m \rightarrow 0$ as $m \rightarrow +\infty$, we have

$$\mathbb{E}[\hat{\pi}_0(1-h_m)] - \bar{\pi}_0 = (1 - \pi_0) \frac{(-1)^k g_1^{(k)}(1)}{(k+1)!} h_m^k + o(h_m^k),$$

which concludes the proof. ■

Proof [Proof of Proposition 16] For (1) we begin by noting that by Proposition 15, the asymptotic variance of $\hat{\pi}_0(1-h_m)$ is equivalent to $\frac{\bar{\pi}_0}{mh_m}$. As we assumed that the first $k-1$ derivatives of g_1 at 1 are null, and that $g_1^{(k)}(1) \neq 0$, a Taylor expansion of $\hat{\pi}_0(1-h_m) - \bar{\pi}_0$ ensures that the bias is of order h_m^k . The optimal bandwidth is obtained for h_m proportional to $m^{-\frac{1}{2k+1}}$, because this choice balances variance and squared bias. The proportionality constant is an explicit function of k , $\bar{\pi}_0$, and $g_1^{(k)}(1)$. By definition, the MSE that corresponds to this optimal choice is twice the corresponding squared bias, i.e. of order $m^{-\frac{k}{2k+1}}$, which completes the proof of (1).

To prove (2), we note that

$$\sqrt{mh_m}(\hat{\pi}_0 - \bar{\pi}_0) = \sqrt{mh_m}(\hat{\pi}_0 - \mathbb{E}[\hat{\pi}_0]) + \sqrt{mh_m}(\mathbb{E}[\hat{\pi}_0] - \bar{\pi}_0),$$

where $\hat{\pi}_0$ denotes $\hat{\pi}_0(1-h_m)$ to alleviate notation. The first term (variance) converges in distribution to $\mathcal{N}(0, \bar{\pi}_0)$ by Proposition 15 (1) as soon as $\sqrt{mh_m} \rightarrow +\infty$. The second term (bias) is of the order of $\sqrt{mh_m}h_m^k = \sqrt{mh_m^{2k+1}}$ by Proposition 15 (2). Taking $h_m(k) = h_m^*(k)\eta_m^2$, where $\eta_m \rightarrow 0$, we have $mh_m^{2k+1} \rightarrow 0$, which ensures that the bias term converges in probability to 0. ■

Asymptotic FDP for plug-in procedures

Lemma 28 *With assumptions of Proposition 20, $\hat{\tau}$ converges almost surely to τ^* as $m \rightarrow +\infty$.*

Proof [Proof of Lemma 28] Let $\psi_{F,\gamma} : u \mapsto F(u) - u/\gamma$ for any distribution function F and any $\gamma \in (0, 1]$. As $\widehat{\mathbb{G}}_m(\widehat{\tau}) = \widehat{\pi}_0 \widehat{\tau}/\alpha$ and $G(\tau^*) = \overline{\pi}_0 \tau^*/\alpha$, we have $\psi_{G,\alpha/\overline{\pi}_0}(\tau^*) = 0$ and $\psi_{\widehat{\mathbb{G}}_m,\alpha/\widehat{\pi}_0}(\widehat{\tau}) = 0$. The idea of the proof is to note that

1. $\psi_{G,\alpha/\overline{\pi}_0}(\widehat{\tau})$ converges almost surely to $0 = \psi_{G,\alpha/\overline{\pi}_0}(\tau^*)$
2. $\psi_{G,\alpha/\overline{\pi}_0}$ is locally invertible in a neighborhood of τ^* .

To prove (1), we note that

$$\begin{aligned} \psi_{G,\alpha/\overline{\pi}_0}(\widehat{\tau}) &= G(\widehat{\tau}) - \overline{\pi}_0 \widehat{\tau}/\alpha \\ &= (G - \widehat{\mathbb{G}}_m)(\widehat{\tau}) + (\widehat{\mathbb{G}}_m(\widehat{\tau}) - \widehat{\pi}_0 \widehat{\tau}/\alpha) + (\widehat{\pi}_0 - \overline{\pi}_0) \widehat{\tau}/\alpha. \end{aligned}$$

The first terms converges to 0 almost surely, the second is identically null, and the third converges almost surely to 0 because $\widehat{\pi}_0$ converges in probability to π_0 , and $\widehat{\tau} \in [0, 1]$. Item (2) holds because we are in a subcritical situation: $\alpha > \overline{\pi}_0 \alpha^*$, with $\alpha^* = \lim_{u \rightarrow 0} u/G(u)$ (see (Neuvial, 2008, Lemma 7.6 page 1097) for a proof of the invertibility). Combining (1) and (2), $\widehat{\tau}$ converges almost surely to τ^* . ■

Proof [Proof of Proposition 20] We only give the proof for $\widehat{\tau}$, as the proofs for $\widehat{\nu}$ and $\widehat{\rho}$ are quite similar. The idea is that the fluctuations of $\widehat{\mathbb{G}}_m - G$, which are of order $1/\sqrt{m}$ by Donsker's theorem (Donsker, 1951), are negligible with respect to the fluctuations of $\widehat{\pi}_0 - \overline{\pi}_0$, because these are assumed to be of order $1/\sqrt{mh_m}$ with $h_m \rightarrow 0$. Letting $\bar{\mathbb{G}}_m = \widehat{\mathbb{G}}_m - G$ be the centered empirical process associated with G , we have

$$\begin{aligned} G(\widehat{\tau}) - G(\tau^*) &= (G(\widehat{\tau}) - \widehat{\mathbb{G}}_m(\widehat{\tau})) + (\widehat{\mathbb{G}}_m(\widehat{\tau}) - G(\tau^*)) \\ &= \bar{\mathbb{G}}_m(\widehat{\tau}) + (\widehat{\pi}_0 \widehat{\tau}/\alpha - \overline{\pi}_0 \tau^*/\alpha) \end{aligned}$$

because $\widehat{\mathbb{G}}_m(\widehat{\tau}) = \widehat{\pi}_0 \widehat{\tau}/\alpha$ and $G(\tau^*) = \overline{\pi}_0 \tau^*/\alpha$. Therefore,

$$G(\widehat{\tau}) - G(\tau^*) = \bar{\mathbb{G}}_m(\widehat{\tau}) + \frac{\widehat{\pi}_0}{\alpha}(\widehat{\tau} - \tau^*) + \frac{\widehat{\pi}_0 - \overline{\pi}_0}{\alpha} \tau^*.$$

As $\|\bar{\mathbb{G}}_m\|_\infty \sim c\sqrt{\ln \ln m/m}$ and $h_m = o(1/\ln \ln m)$, we have $\bar{\mathbb{G}}_m(\widehat{\tau}) = o_P(1/\sqrt{mh_m})$. Since $\widehat{\tau} \xrightarrow{a.s.} \tau^*$ as $m \rightarrow +\infty$, we also have $G(\widehat{\tau}) - G(\tau^*) = (\widehat{\tau} - \tau^*)(g(\tau^*) + o_P(1))$ by Taylor's formula. Hence we have

$$\left(g(\tau^*) - \frac{\widehat{\pi}_0}{\alpha}\right)(\widehat{\tau} - \tau^*) = (\widehat{\pi}_0 - \overline{\pi}_0) \frac{\tau^*}{\alpha} + o_P\left(1/\sqrt{mh_m}\right).$$

Finally, as

$$\sqrt{mh_m}(\widehat{\pi}_0 - \overline{\pi}_0) \rightsquigarrow \mathcal{N}(0, v(\overline{\pi}_0)),$$

we have

$$\widehat{\tau} - \tau^* = \frac{\tau^*/\alpha}{g(\tau^*) - \overline{\pi}_0/\alpha}(\widehat{\pi}_0 - \overline{\pi}_0)(1 + o_P(1)),$$

which concludes the proof for $\widehat{\tau}$. ■

Proof [Proof of Theorem 21] By Proposition 20, we have

$$\sqrt{mh_m} \left(\begin{pmatrix} \widehat{\nu} \\ \widehat{\rho} \end{pmatrix} - \begin{pmatrix} \pi_0 \tau^* \\ \overline{\pi}_0 \tau^*/\alpha \end{pmatrix} \right) \rightsquigarrow \mathcal{N}(0, V),$$

with

$$V = v(\bar{\pi}_0) \left(\frac{\tau^*/\alpha}{g(\tau^*) - \bar{\pi}_0/\alpha} \right)^2 \left(\frac{\pi_0}{g(\tau^*)} \right) (\pi_0 - g(\tau^*)).$$

We write $\text{FDP} = \gamma(\hat{\nu}, \hat{\rho})$, where $\gamma : (x, y) \mapsto x/y$ for any $x \geq 0$ and $y > 0$. γ is differentiable at $(\pi_0\tau^*, \bar{\pi}_0\tau^*/\alpha)$, with derivative

$$\begin{aligned} \dot{\gamma}_{\pi_0\tau^*, \bar{\pi}_0\tau^*/\alpha} &= (1/(\bar{\pi}_0\tau^*/\alpha), -\pi_0\tau^*/(\bar{\pi}_0\tau^*/\alpha)^2) \\ &= \frac{\alpha}{\bar{\pi}_0\tau^*} (1, -\frac{\pi_0\alpha}{\bar{\pi}_0}). \end{aligned}$$

As $\gamma(\pi_0\tau^*, \bar{\pi}_0\tau^*/\alpha) = \frac{\pi_0\alpha}{\bar{\pi}_0}$, the Delta method yields

$$\sqrt{mh_m} \left(\text{FDP}_m - \frac{\pi_0\alpha}{\bar{\pi}_0} \right) \rightsquigarrow \mathcal{N}(0, w),$$

with

$$\begin{aligned} w &= v(\bar{\pi}_0) \left(\frac{\tau^*/\alpha}{g(\tau^*) - \bar{\pi}_0/\alpha} \right)^2 \dot{\gamma}_{\pi_0\tau^*, \bar{\pi}_0\tau^*/\alpha} \left(\frac{\pi_0}{g(\tau^*)} \right) (\pi_0 - g(\tau^*)) \dot{\gamma}'_{\pi_0\tau^*, \bar{\pi}_0\tau^*/\alpha} \\ &= v(\bar{\pi}_0) \left(\frac{\tau^*/\alpha}{g(\tau^*) - \bar{\pi}_0/\alpha} \right)^2 \left(\frac{\alpha}{\bar{\pi}_0\tau^*} (1 - \alpha) \left(\frac{\pi_0}{g(\tau^*)} \right) \right)^2 \\ &= v(\bar{\pi}_0)/\bar{\pi}_0^2 \left(\frac{\pi_0 - \frac{\pi_0\alpha}{\bar{\pi}_0} g(\tau^*)}{g(\tau^*) - \bar{\pi}_0/\alpha} \right)^2 \\ &= v(\bar{\pi}_0) \frac{\pi_0\alpha^2}{\bar{\pi}_0} / \bar{\pi}_0^2. \end{aligned}$$

■

Proof of Proposition 23

Lemma 29 (Regularity of g_1 in two-sided symmetric models) *Under Condition 2, if the likelihood ratio $\frac{f_1}{f_0}$ is differentiable on \mathbb{R} , then the density g_1^\pm of the p -values under the alternative hypothesis is differentiable on $(0, 1]$ and verifies, for any $u \in (0, 1]$,*

$$g_1^{\pm(1)}(u) = \frac{1}{4f_0(F_0^{-1}(1-u/2))} \left(\left(\frac{f_1}{f_0} \right)' (F_0^{-1}(1-u/2)) - \left(\frac{f_1}{f_0} \right)' (F_0^{-1}(u/2)) \right)$$

Proof [Proof of Lemma 29] By Proposition 2, we have

$$g_1^{\pm(1)}(u) = \frac{1}{2} \left(\frac{f_1}{f_0}(q_0(u/2)) + \frac{f_1}{f_0}(-q_0(u/2)) \right),$$

where $q_0 : u \mapsto F_0^{-1}(1-u)$. Therefore,

$$g_1^{\pm(1)}(u) = \frac{1}{2} \left\{ \left(\frac{f_1}{f_0} \right)' (q_0(u/2)) - \left(\frac{f_1}{f_0} \right)' (-q_0(u/2)) \right\} \times \frac{1}{2} q_0'(u/2)$$

which concludes the proof as q_0 satisfies $q_0(1-u) = -q_0(u)$ and $q_0'(u) = -1/f_0(q_0(u))$. ■

Proof [Proof of Proposition 23] (1) was already stated in Proposition 14. By Lemma 29, we have $g_1^{\pm(1)}(1) = 0$, as $q_0(1/2) = F_0^{-1}(1/2) = 0$, which proves (2). For (3), note that if $\frac{f_1}{f_0}$ is twice differentiable in a neighborhood of 1, then by Lemma 29 g_1^{\pm} is itself twice differentiable. Writing $g_1^{\pm(1)}(u) = a(u)b(u)$, with

$$\begin{cases} a(u) &= 1/(4f_0(F_0^{-1}(1-u/2))) \\ b(u) &= \left(\frac{f_1}{f_0}\right)'(F_0^{-1}(1-u/2)) - \left(\frac{f_1}{f_0}\right)'(F_0^{-1}(u/2)) \end{cases},$$

we have $g_1^{\pm(2)}(u) = a'(u)b(u) + a(u)b'(u)$. As $q_0(1/2) = F_0^{-1}(1/2) = 0$, we have $b(1) = 0$, so that $g_1^{\pm(2)}(1) = a(1)b'(1)$, where

$$b'(u) = \frac{1}{2f_0(F_0^{-1}(1-u/2))} \left(\left(\frac{f_1}{f_0}\right)^{(2)}(F_0^{-1}(1-u/2)) + \left(\frac{f_1}{f_0}\right)^{(2)}(F_0^{-1}(u/2)) \right).$$

Thus, $a(1) = 1/(4f_0(0))$ and $b'(1) = 1/(2f_0(0)) \times 2\frac{f_1}{f_0}(0)$, which concludes the proof. ■

Acknowledgments

I would like to thank Stéphane Boucheron for insightful advice and comments, and Etienne Roquain for useful discussions. I am also grateful to two anonymous referees for constructive remarks and suggestions that helped improve the paper.

This work was supported by the association Courir pour la vie, courir pour Curie, and by the French ANR project TAMIS.

References

- Felix Abramovich and Yoav Benjamini. Thresholding of wavelet coefficients as multiple hypotheses testing procedure. In Antoniadis A. and Oppenheim G., editors, *Wavelets and Statistics*, volume 103, pages 5–14. Springer-Verlag, 1995.
- Felix Abramovich, Yoav Benjamini, David L. Donoho, and Ian M. Johnstone. Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.*, 34(2):584–653, 2006.
- Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 57(1):289–300, 1995.
- Yoav Benjamini and Yosef Hochberg. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics*, 25(1):60–83, 2000.
- Yoav Benjamini, Abba M. Krieger, and Daniel Yekutieli. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507, 2006.
- Gilles Blanchard and Etienne Roquain. Adaptive FDR control under independence and dependence. *Journal of Machine Learning Research*, to appear. arXiv preprint math.ST/0707.0536v2.
- T. Tony Cai, Jiashun Jin, and Mark G Low. Estimation and confidence sets for sparse normal mixtures. *Ann. Statist.*, 35(6):2421–2449, 2007.
- Zhiyi Chi. On the performance of FDR control: constraints and a partial solution. *Ann. Statist.*, 35(4):1409–1431, 2007a.
- Zhiyi Chi. Sample size and positive false discovery rate control for multiple testing. *Electronic Journal of Statistics*, 1:77–118, 2007b.
- Zhiyi Chi and Zhiqiang Tan. Positive false discovery proportions: intrinsic bounds and adaptive control. *Statistica Sinica*, 18(3):837–860, 2008.
- David L. Donoho and Jiashun Jin. Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.*, 32(3):962–994, 2004.
- David L. Donoho and Jiashun Jin. Asymptotic minimaxity of false discovery rate thresholding for sparse exponential data. *Ann. Statist.*, 34(6):2980–3018, 2006.
- Monroe D. Donsker. An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.*, 6:12, 1951.
- Sandrine Dudoit, Jane Fridlyand, and Terence P. Speed. Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association*, 97(457):77–87, 2002.

- Bradley Efron, Robert Tibshirani, John D. Storey, and Virginia Tusher. Empirical bayes analysis of a microarray experiment. *Journal of the American Statistical Association*, 96(465):1151–1160, 2001.
- H Finner, T Dickhaus, and M Roters. On the false discovery rate and an asymptotically optimal rejection curve. *The Annals of Statistics*, 2009. Ann. Statist. (to appear).
- Christopher R. Genovese and Larry Wasserman. Operating characteristics and extensions of the false discovery rate. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64(3):499–517, 2002. doi: 10.1111/1467-9868.00347.
- Christopher R. Genovese and Larry Wasserman. A Stochastic Process Approach to False Discovery Control. *Ann. Statist.*, 32(3):1035–1061, 2004.
- R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. H. Yang, and J. Zhang. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology*, 5:R80, 2004.
- T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286(5439):531–537, Oct 1999.
- Nicolas W. Hengartner and Philip B. Stark. Finite-Sample Confidence Envelopes for Shape-Restricted Densities. *Ann. Statist.*, 23(2):525–550, 1995.
- Yosef Hochberg and Yoav Benjamini. More powerful procedures for multiple significance testing. *Statistics in Medicine*, 9(7):811–818, 1990.
- Yuri I. Ingster. Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.*, 1997.
- Yuri I. Ingster. Minimax detection of a signal for l n -balls. *Math. Methods Statist.*, 7(4):401–428, 1999.
- Yuri I. Ingster and Irina A. Suslina. Nonparametric Goodness-of-Fit Testing under Gaussian Models. *Lecture Notes in Statistics*, 169, 2003.
- Jiashun Jin. Detection boundary for sparse mixtures. Unpublished manuscript, 2002.
- Jiashun Jin. Proportion of non-zero normal means: universal oracle equivalences and uniformly consistent estimators. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(3):461–493, 2008.
- Jiashun Jin and T. Tony Cai. Estimating the Null and the Proportion of non-Null Effects in Large-Scale Multiple Comparisons. *Journal of the American Statistical Association*, 102(478):495–506, 2007.
- Mette Langaas, Bo H. Lindqvist, and Egil Ferkingstad. Estimating the proportion of true null hypotheses, with application to DNA microarray data. *J Bioinform Comput Biol*, 67(4):555–572, 2005.

- Nicolai Meinshausen and Peter Bühlmann. Lower bounds for the number of false null hypotheses for multiple testing of associations. *Biometrika*, 92:893–907, 2005.
- Nicolai Meinshausen and John Rice. Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.*, 34(1):373–393, 2006.
- Pierre Neuvial. Asymptotic properties of false discovery rate controlling procedures under independence. *Electron. J. Statist.*, 2:1065–1110, 2008. ISSN 1935-7524. doi: 10.1214/08-EJS207.
- David B. Pollard. *Convergence of Stochastic Processes*. Springer-Verlag, 1984.
- Tore Schweder and Emil Spjøtvoll. Plots of p-values to evaluate many tests simultaneously. *Biometrika*, 69(3):493–502, 1982.
- R. J. Simes. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754, 1986.
- John D. Storey. A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64(3):479–498, 2002.
- John D. Storey. The positive false discovery rate: A bayesian interpretation and the q-value. *Ann. Statist.*, 31(6):2013–2035, 2003.
- John D. Storey, Jonathan E. Taylor, and David Siegmund. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66(1):187–205, 2004.
- M.T. Subbotin. On the law of frequency of errors. *Matematicheskii Sbornik*, 31:296–301, 1923.
- Jan W. H. Swanepoel. The limiting behavior of a modified maximal symmetric 2s-spacing with applications. *Ann. Statist.*, 27(1):24–35, 1999.
- Alexandre B Tsybakov. *Introduction to non-parametric estimation*. Springer, 2009.
- John W. Tukey. T13 N: The Higher Criticism. *Course notes, Princeton University*, 1976.