Asymptotic Results on Adaptive False Discovery Rate Controlling Procedures Based on Kernel Estimators

Pierre Neuvial

HAL Id: hal-02644679
https://hal.inrae.fr/hal-02644679v1

Submitted on 28 May 2020

Asymptotic Results on Adaptive False Discovery Rate Controlling Procedures Based on Kernel Estimators

Pierre Neuvial[*]   PIERRE.NEUVIAL@GENOPOLE.CNRS.FR
Laboratoire Statistique et Génome
Université d'Évry Val d'Essonne
UMR CNRS 8071 – USC INRA
23 boulevard de France
91 037 Évry, France

Editor: Olivier Teytaud

Abstract

The False Discovery Rate (FDR) is a commonly used type I error rate in multiple testing problems. It is defined as the expected False Discovery Proportion (FDP), that is, the expected fraction of false positives among rejected hypotheses. When the hypotheses are independent, the Benjamini-Hochberg procedure achieves FDR control at any pre-specified level. By construction, FDR control offers no guarantee in terms of power, or type II error. A number of alternative procedures have been developed, including plug-in procedures that aim at gaining power by incorporating an estimate of the proportion of true null hypotheses.

In this paper, we study the asymptotic behavior of a class of plug-in procedures based on kernel estimators of the density of the p-values, as the number m of tested hypotheses grows to infinity. In a setting where the hypotheses tested are independent, we prove that these procedures are asymptotically more powerful in two respects: (i) a tighter asymptotic FDR control for any target FDR level and (ii) a broader range of target levels yielding positive asymptotic power. We also show that this increased asymptotic power comes at the price of slower, non-parametric convergence rates for the FDP. These rates are of the form $m^{-k/(2k+1)}$, where k is determined by the regularity of the density of the p-value distribution, or, equivalently, of the test statistics distribution. These results are applied to one- and two-sided tests statistics for Gaussian and Laplace location models, and for the Student model.

Keywords: multiple testing, false discovery rate, Benjamini Hochberg's procedure, power, criticality, plug-in procedures, adaptive control, test statistics distribution, convergence rates, kernel estimators

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, and DNA microarray or high-throughput sequencing analyses in genomics. Given a set of observations corresponding either to a null hypothesis or an alternative hypothesis, the goal of multiple testing is to infer which of them correspond to true alternatives. This requires the definition of risk measures

---

that are adapted to the large number of tests performed: typically 10^4 to 10^6 in genomics. The False Discovery Rate (FDR) introduced by Benjamini and Hochberg (1995) is one of the most commonly used and one of the most widely studied such risk measure in large-scale multiple testing problems. The FDR is defined as the expected proportion of false positives among rejected hypotheses. A simple procedure called the Benjamini-Hochberg (BH) procedure provides FDR control when the tested hypotheses are independent (Benjamini and Hochberg, 1995) or follow specific types of positive dependence (Benjamini and Yekutieli, 2001).

When the hypotheses tested are independent, applying the BH procedure at level α in fact yields $\text{FDR} = \pi_0 \alpha$, where π_0 is the unknown fraction of true null hypotheses (Benjamini and Yekutieli, 2001). This has motivated the development of a number of "plug-in" procedures, which consist in applying the BH procedure at level $\alpha/\hat{\pi}_0$, where $\hat{\pi}_0$ is an estimator of π_0. A typical example is the Storey-λ procedure (Storey, 2002; Storey et al., 2004) in which $\hat{\pi}_0$ is a function of the empirical cumulative distribution function of the p-values.

In this paper, we consider an asymptotic framework where the number m of tests performed goes to infinity. When $\hat{\pi}_0$ converges in probability to $\pi_{0,\infty} \in [\pi_0, 1)$ as $m \to +\infty$, the corresponding plug-in procedure is by construction asymptotically more powerful than the BH procedure, while still providing $\text{FDR} \leq \alpha$. However, as FDR control only implies that the *expected* FDP is below the target level, it is of interest to study the *fluctuations* of the FDP achieved by such plug-in procedures around their corresponding FDR. This paper studies the influence of the plug-in step on the asymptotic properties of the corresponding procedure for a particular class of estimators of π_0, which may be written as kernel estimators of the density of the p-value distribution at 1.

2. Background and Notation

In this section, we introduce the multiple testing setting considered in this paper, and define two central concepts: plug-in procedures and criticality.

2.1 Multiple Testing Setting

We consider a test statistic X distributed as F_0 under a null hypothesis $\mathcal{H}_0$ and as F_1 under an alternative hypothesis $\mathcal{H}_1$. We assume that for $a \in \{0,1\}$, F_a is continuously differentiable, and that the corresponding density function, which we denote by f_a, is positive. This testing problem may be formulated in terms of p-values instead of test statistics. The p-value function is defined as $p : x \mapsto \mathbb{P}_{\mathcal{H}_0}(X \geq x) = 1 - F_0(x)$ for one-sided tests and $p : x \mapsto \mathbb{P}_{\mathcal{H}_0}(|X| \geq |x|)$ for two-sided tests. As F_0 is continuous, the p-values are uniform on $[0,1]$ under $\mathcal{H}_0$. For consistency we denote by G_0 the corresponding distribution function, that is, the identity function on $[0,1]$. Under $\mathcal{H}_1$, the distribution function and density of the p-values are denoted by G_1 and g_1, respectively. Their expression as functions of the distribution of the test statistics are recalled in Proposition 1 below in the case of one- and two-sided p-values. For two-sided p-values, we assume that the distribution function of the test statistics under $\mathcal{H}_0$ is symmetric (around 0):

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

Assumption (1) is typically satisfied in usual models such as Gaussian or Laplace (double exponential) models. Under Assumption (1), the two-sided p-value satisfies $p(x) = 2(1 - F_0(|x|))$ for any $x \in \mathbb{R}$.

Proposition 1 (One- and two-sided p-values) *For $t \in [0,1]$, let $q_0(t) = F_0^{-1}(1-t)$. The distribution function G_1 and the and density function g_1 of the p-value under $\mathcal{H}_1$ at t satisfy the following:*

1. *for a one-sided p-value, $G_1(t) = 1 - F_1(q_0(t))$ and $g_1(t) = (f_1/f_0)(q_0(t))$;*

2. *for a two-sided p-value, $G_1(t) = 1 - F_1(q_0(t/2)) + F_1(-q_0(t/2))$ and*
   $g_1(t) = 1/2 \left((f_1/f_0)(q_0(t/2)) + (f_1/f_0)(-q_0(t/2)) \right).$

The assumption that f_1 is positive entails that g_1 is positive as well. We further assume that

$$G_1 \text{ is concave.} \tag{2}$$

As g_1 is a function of the likelihood ratio f_1/f_0 and the non-increasing function q_0, Assumption (2) may be characterized as follows:

Lemma 2 (Concavity and likelihood ratios) *1. For a one-sided p-value, Assumption (2) holds if and only if the likelihood ratio f_1/f_0 is non-decreasing.*

2. *For a two-sided p-value under Assumption (1), Assumption (2) holds if and only if $x \mapsto (f_1/f_0)(x) + (f_1/f_0)(-x)$ is non-decreasing on $\mathbb{R}_+$.*

We consider a sequence of independent tests performed as described above and indexed by the set $\mathbb{N}^*$ of positive integers. We assume that either all of them are one-sided tests, or all of them are two-sided tests. This sequence of tests is characterized by a sequence $(\mathbf{H}, \mathbf{p}) = (H_i, p_i)_{i \in \mathbb{N}^*}$, where for each $i \in \mathbb{N}^*$, p_i is a p-value associated to the ith test, and H_i is a binary indicator defined by

$$H_i = \begin{cases} 0 \text{ if } \mathcal{H}_0 \text{ is true for test } i \\ 1 \text{ if } \mathcal{H}_1 \text{ is true for test } i \end{cases}.$$

We also let $m_0(m) = \sum_{i=1}^m (1 - H_i)$, and $\pi_{0,m} = m_0(m)/m$. Following the terminology proposed by Roquain and Villers (2011), we define the *conditional setting* as the situation where $\mathbf{H}$ is deterministic and $\mathbf{p}$ is a sequence of independent random variables such that for $i \in \mathbb{N}^*$, $p_i \sim G_{H_i}$. This is a particular case of the setting originally considered by Benjamini and Hochberg (1995), where no assumption was made on the distribution of p_i when $H_i = 1$. In the present paper, we consider an *unconditional setting* introduced by Efron et al. (2001), which is also known as the "random effects" setting. Specifically, $\mathbf{H}$ is a sequence of random indicators, independently and identically distributed as $\mathcal{B}(1 - \pi_0)$, where $\pi_0 \in (0,1)$, and conditional on $\mathbf{H}$, $\mathbf{p}$ follows the conditional setting, that is, the p-values satisfy $p_i | H_i \sim G_{H_i}$. This unconditional setting has been widely used in the multiple testing literature, see, for example, Storey (2003); Genovese and Wasserman (2004); Chi (2007a). In this setting, the p-values are independently, identically distributed as $G = \pi_0 G_0 + (1 - \pi_0) G_1$, and $m_0(m)$ follows the binomial distribution $Bin(m, \pi_0)$.

Remark 3 *We are assuming that $\pi_0 < 1$, which implies that the proportion $1 - \pi_{0,m}$ of true null alternatives does not vanish as $m \to +\infty$. While this restriction is natural in the unconditional setting considered in this paper, we note that our results do not apply to the "sparse" situation where $\pi_{0,m} \to 1$ as $m \to +\infty$.*

As G_0 is the identity function, the multiple testing model is entirely characterized by the two parameters π_0 and G_1 (or, equivalently, π_0 and G), where G_1 is itself entirely characterized by F_0 and F_1, by Proposition 1. The mixture distribution G is concave if and only if Assumption (2) holds. More generally, we note that making a regularity assumption on G_1 (or g_1) is equivalent to making the same regularity assumption on G (or g):

Remark 4 (Differentiability assumptions) *Throughout the paper, differentiability assumptions on the distribution of the p-values near 1 are expressed in terms of g, the (mixture) p-value density. As $g = \pi_0 + (1 - \pi_0)g_1$, we note that they could equally be written in terms of g_1, the p-value density under the alternative hypothesis.*

2.2 Type I and II Error Rate Control in Multiple Testing

We define a multiple testing procedure $\mathcal{P}$ as a collection of functions $(\mathcal{P}_\alpha)_{\alpha \in [0,1]}$ such that for any $\alpha \in [0,1]$, $\mathcal{P}_\alpha$ takes as input a vector of m p-values, and returns a subset of $\{1, \ldots m\}$ corresponding to the indices of hypotheses to be rejected. For a given procedure $\mathcal{P}$ and a given $\alpha \in [0,1]$, the function $\mathcal{P}_\alpha$ will be called "Procedure $\mathcal{P}$ at (target) level α". In this paper, we focus on *thresholding-based* multiple testing procedures, for which the rejected hypotheses are those with p-values less than a threshold. Each possible value for the threshold corresponds to a trade-off between false positives (type I errors) and false negatives (type II errors). Most risk measures developed for multiple testing procedures are based on type I errors. We focus on one such measure, the False Discovery Rate (FDR), which is one of the most widely used error rate in multiple testing. Denoting by R_m be the total number of rejections of $\mathcal{P}_\alpha$ among m hypotheses tested, and by V_m the number of false rejections, the corresponding False Discovery Proportion is defined as $\text{FDP}_m = V_m/(R_m \vee 1)$, and the False Discovery Rate is the expected FDP, that is:

$$\text{FDR}_m = \mathbb{E}\left[\frac{V_m}{R_m \vee 1}\right].$$

A trivial way to control the FDR, or any risk measure only based on type I errors, is to make no rejection with high probability. Obviously, this is not the best strategy, as it may lead to a high number of type II errors. The performance of multiple testing procedures may be evaluated through their power, which is a function of the number of type II errors. Specifically, the power of a multiple testing procedure at level α is generally defined as the (random) proportion of correct rejections (true positives) among true alternative hypotheses (see, for example, Chi, 2007a):

$$\Pi_m = \frac{R_m - V_m}{(m - m_0(m)) \vee 1}.$$

Remark 5 *All of the quantities defined in this section implicitly depend on the multiple testing procedure considered, $\mathcal{P} = (\mathcal{P}_\alpha)_{\alpha \in [0,1]}$. However, for simplicity, we will write R_m, V_m, FDR_m, and Π_m, instead of $R_m^{\mathcal{P}_\alpha}, V_m^{\mathcal{P}_\alpha}, \text{FDR}_m^{\mathcal{P}_\alpha}$, and $\Pi_m^{\mathcal{P}_\alpha}$ whenever not ambiguous.*

Remark 6 *By definition, the power of a thresholding-based procedure is a non-decreasing function of its threshold. Therefore, among thresholding-based procedures that yield FDR less than a prescribed level, maximizing power is equivalent to maximizing the threshold of the procedure.*

2.3 The Benjamini-Hochberg Procedure

Suppose we wish to control the FDR at level α. Let $p_{(1)} \leq \ldots \leq p_{(m)}$ be the ordered p-values, and denote by $H_{(i)}$ the null hypothesis corresponding to $p_{(i)}$. Define $\widehat{I}_m(\alpha)$ as the largest index $k \geq 0$ such that $p_{(k)} \leq \alpha k/m$. The Benjamini-Hochberg procedure at level α rejects all $H_{(i)}$ such that $i \leq \widehat{I}_m(\alpha)$ (if $\widehat{I}_m(\alpha) = 0$, then no rejection is made). This procedure has been proposed by Benjamini and Hochberg (1995) in the context of FDR control; Seeger (1968) reported that it had previously been used by Eklund (1961–1963) in another multiple testing context. When all true null hypotheses are independent, the BH procedure at level α yields *strong* FDR control, that is, it entails FDR $\leq \alpha$ regardless of the number of true null hypotheses (Benjamini and Hochberg, 1995). The BH procedure also controls the FDR when the p-values satisfy specific forms of positive dependence, see Benjamini and Yekutieli (2001). Figure 1 illustrates the application of the BH 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 BH procedure. An equivalent definition is that the procedure rejects all hypotheses with associated p-value is less than $\widehat{\tau}_m(\alpha) = \alpha \widehat{I}_m(\alpha)/m$. The
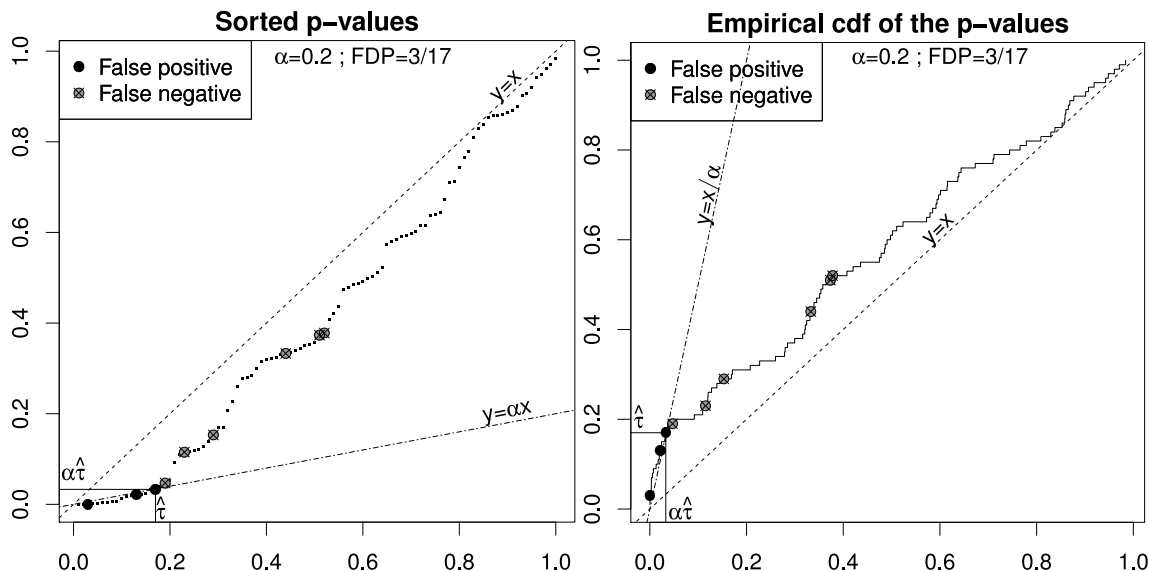

Figure 1: Illustrations of the BH procedure on a simulated example with $m = 100$. Left: sorted p-values: $i/m \mapsto p_{(i)}$. Right: empirical distribution function: $t \mapsto \widehat{\mathbb{G}}_m(t)$.

right panel provides a dual representation of the same information, where the x and y axes have been swapped. It gives a geometrical interpretation of $\widehat{\tau}_m(\alpha)$ as the largest crossing point between the line $y = x/\alpha$ and the empirical distribution function of the p-values, defined for $t \in [0,1]$ by $\widehat{\mathbb{G}}_m(t) = \sum_{i=1}^{m} \mathbf{1}_{P_i \leq t}$:

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

2.4 Plug-in Procedures

In our setting where all of the hypotheses tested are independent, the BH procedure at target level α (henceforth denoted by BH(α) for short) in fact yields FDR control at level $\pi_0 \alpha$ exactly (Benjamini

and Hochberg, 1995; Benjamini and Yekutieli, 2001). This entails that the $BH(\alpha')$ procedure yields $FDR \leq \alpha$ if and only if $\alpha' \leq \alpha/\pi_0$. Therefore, as the threshold of the $BH(\alpha)$ procedure is a non-decreasing function of α and by Remark 6, the $BH(\alpha/\pi_0)$ procedure is optimal in our setting, in the sense that it yields maximum power among procedures of the form $BH(\alpha')$ that control the FDR at level α. As π_0 is unknown, this procedure cannot be implemented; it is generally referred to as the Oracle BH procedure.

Remark 7 *If* $\alpha \geq \pi_0$, *then rejecting all null hypotheses is optimal, as it corresponds to the largest possible threshold while still maintaining* $FDR = \pi_0 \leq \alpha$. *Therefore, we will assume that* $\alpha < \pi_0$ *throughout the paper.*

In order to mimic the Oracle procedure, it is natural to apply the BH procedure at level $\alpha/\hat{\pi}_{0,m}$, where $\hat{\pi}_{0,m} \leq 1$ is an estimator of π_0 (Benjamini and Hochberg, 2000). Such plug-in procedures (also known as two-stage adaptive procedures) have the same geometric interpretation as the BH procedure (see Figure 1) in terms of the largest crossing point, with $\alpha/\hat{\pi}_{0,m}$ instead of α. Their rejection threshold can be written as $\hat{\tau}_m^0(\alpha) = \hat{\tau}_m(\alpha/\hat{\pi}_{0,m})$, that is:

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

Note that $\hat{\tau}_m^0$ depends on the observations through both $\widehat{\mathbb{G}}_m$ and $\hat{\pi}_{0,m}$. By construction, a plug-in procedure based on an estimator $\hat{\pi}_{0,m}$ that converges in probability to $\pi_{0,\infty} \in [\pi_0, 1)$ as $m \to +\infty$ is asymptotically more powerful that the original BH procedure.

Adapting a method originally proposed by Schweder and Spjøtvoll (1982), Storey (2002) defined $\hat{\pi}_{0,m}^{\text{Sto}}(\lambda) = \#\{i/P_i \geq \lambda\}/\#\{i \geq \lambda\}$ for $\lambda \in (0,1)$. This estimator is generally referred to as the Storey-λ estimator. It may also be written as a function of the empirical distribution of the p-values:

$$\hat{\pi}_{0,m}^{\text{Sto}}(\lambda) = \frac{1 - \widehat{\mathbb{G}}_m(\lambda)}{1 - \lambda}. \tag{3}$$

The rationale for $\hat{\pi}_{0,m}^{\text{Sto}}(\lambda)$ is that under Assumption (2), larger p-values are more likely to correspond to true null hypotheses than smaller ones. Moreover, $\hat{\pi}_{0,m}^{\text{Sto}}(\lambda)$ converges in probability to $(1 - G(\lambda))/(1 - \lambda)$, where the limit is greater than π_0 as G stochastically dominates the uniform distribution. Several choices of λ have been proposed, including $\lambda = 1/2$ (Storey and Tibshirani, 2003), a data-driven choice based on the bootstrap (Storey et al., 2004), and $\lambda = \alpha$ (Blanchard and Roquain, 2009). In our setting, a slightly modified version of the corresponding plug-in $BH(\alpha/\hat{\pi}_{0,m}^{\text{Sto}}(\lambda))$ procedure where $1/m$ is added to the numerator in (3) achieves strong FDR control at level α (Storey et al., 2004). We note that the Storey-λ estimator $\hat{\pi}_{0,m}^{\text{Sto}}(\lambda)$ can be viewed as a kernel estimator of the density g at 1.

Definition 8 (Kernel of order ℓ and kernel estimator of a density at a point)

1. *A kernel of order $\ell \in \mathbb{N}$ is a function $K : \mathbb{R} \to \mathbb{R}$ such that the functions $u \mapsto u^j K(u)$ are integrable for any $j = 0 \ldots \ell$, and satisfy $\int_{\mathbb{R}} K = 1$, and $\int_{\mathbb{R}} u^j K(u)du = 0$ for $j = 1 \ldots \ell$.*

2. *The kernel estimator of a density g at x_0 based on m independent, identically distributed observations $x_1, \ldots x_m$ from g is defined by*

$$\hat{g}_m(x_0) = \frac{1}{mh}\sum_{i=1}^{m} K\left(\frac{x_i - x_0}{h}\right),$$

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

By Definition 8, $\hat{\pi}_{0,m}^{\text{Sto}}(\lambda)$ is a kernel estimator of the density g at 1 with kernel $K^{\text{Sto}}(t) = \mathbf{1}_{[-1,0]}(t)$ and bandwidth $h = 1 - \lambda$. K^{Sto} is an asymmetric, rectangular kernel of order 0.

2.5 Criticality and Asymptotic Properties of FDR Controlling Procedures

Upper bounds on the asymptotic number of rejections of FDR controlling procedures have been identified and characterized by Chi (2007a) and Chi and Tan (2008), who introduced the notion of *critical value of a multiple testing problem* and that of *critical value of a multiple testing procedure*. Both notions are defined formally below. They are tightly connected, with the important difference that the former only depends on the multiple testing problem, while the latter depends on both the multiple testing problem and a specific multiple testing procedure.

Definition 9 (Critical value of a multiple testing problem) *The critical value of the multiple testing problem parametrized by π_0 and G is defined by*

$$\underline{\alpha}^{\star} = \inf_{t \in (0,1]} \frac{\pi_0 t}{G(t)} \,. \tag{4}$$

Chi and Tan (2008, proof of Proposition 3.2) proved that for any multiple testing procedure, for $\alpha < \underline{\alpha}^{\star}$, there exists a positive constant $c(\alpha)$ such that almost surely, for m large enough, the events $\{V_m/R_m \leq \alpha\}$ and $\{R_m \geq c(\alpha)\log m\}$ are incompatible. This restriction is intrinsic to the multiple testing problem, in the sense that it holds regardless of the considered multiple testing procedure. Obviously, this is not a limitation when $\underline{\alpha}^{\star} = 0$. We introduce the following Condition:

$$\underline{\alpha}^{\star} > 0 \,. \tag{5}$$

Whether Condition (5) is satisfied or not only depends on G. However, the value of $\underline{\alpha}^{\star}$ as defined in (4) depends on both π_0 and G. Under Assumption (2) we have $\underline{\alpha}^{\star} = \lim_{t \to 0} \pi_0 t / G(t) = \pi_0/(\pi_0 + (1 - \pi_0)g_1(0))$, where $g_1(0) \in [0, +\infty]$ is defined by $g_1(0) = \lim_{t \to 0} g_1(t)$. By Proposition 1, $g_1(0)$ only depends on the behavior of the test statistics distribution. In particular, under Assumption (2), Condition (5) is satisfied if and only if the likelihood ratio f_1/f_0 is bounded near $+\infty$.

We now introduce the notion of *critical value of a multiple testing procedure*. Chi (2007a) defined the critical value of the BH procedure as $\alpha_{BH}^{\star} = \inf_{t \in (0,1]} t/G(t)$. Let us denote by

$$\tau_{\infty}(\alpha) = \sup\{t \in [0,1], G(t) \geq t/\alpha\} \tag{6}$$

the rightmost crossing point between G and the line $y = x/\alpha$. Chi (2007a) has proved the following result:

Proposition 10 (Asymptotic properties of the BH procedure) *For $\alpha \in [0,1]$, let $\hat{\tau}_m(\alpha)$ be the threshold of the $BH(\alpha)$ procedure, and let $\tau_{\infty}(\alpha)$ be defined by (6). Let $\alpha_{BH}^{\star} = \inf_{t \in (0,1]} t/G(t)$. As $m \to +\infty$,*

1. If $\alpha < \alpha_{BH}^{\star}$, then $\hat{\tau}_m(\alpha) \overset{a.s.}{\to} 0$;

2. If $\alpha > \alpha_{BH}^{\star}$, then $\hat{\tau}_m(\alpha) \overset{a.s.}{\to} \tau_{\infty}(\alpha)$, where the limit is positive.

A straightforward consequence of Proposition 10 is that the BH(α) procedure has asymptotically null power when $\alpha < \alpha^\star_{BH}$ and positive power when $\alpha > \alpha^\star_{BH}$. The following Definition generalizes the notion of critical value of to a generic multiple testing procedure:

Definition 11 (Critical value of a multiple testing procedure) *Let $\mathcal{P}$ denote a multiple testing procedure. The critical value of $\mathcal{P}$ is defined by*

$$\alpha^\star_{\mathcal{P}} = \sup \left\{ \alpha \in [0,1], \Pi^{\mathcal{P}_\alpha}_m \xrightarrow[m \to +\infty]{a.s.} 0 \right\}.$$

The critical value $\alpha^\star_{\mathcal{P}}$ depends on both the procedure $\mathcal{P}$, and the multiple setting. For the BH procedure, criticality ($\alpha < \alpha^\star_{BH}$) corresponds to situations where the target FDR level α is so small that there is no positive crossing point between G and the line $y = x/\alpha$. Conversely, when $\alpha > \alpha^\star_{BH}$, there is a positive crossing point between G and the line $y = x/\alpha$, as illustrated by Figure 1 (right). The almost sure convergence results of Proposition 10 in the case $\alpha > \alpha^\star_{BH}$ were extended by Neuvial (2008), in the conditional setting. Specifically, the threshold $\widehat{\tau}_m(\alpha)$ of the BH procedure was shown to converge in distribution to $\tau_\infty(\alpha)$ at rate $m^{-1/2}$ as soon as $\alpha > \alpha^\star_{BH}$. Neuvial (2008) also proved that similar central limit theorems hold for a class of thresholding-based FDR controlling procedures that covers some plug-in procedures, including the Storey-λ procedure: the threshold of a procedure $\mathcal{P}$ of this class converges in distribution to a procedure-specific, positive value at rate $m^{-1/2}$ as soon as $\alpha > \alpha^\star_{\mathcal{P}}$.

Remark 12 (Criticality of a multiple testing problem versus criticality of a procedure) *Whether Condition (5) hols or not only depends on the behavior of the test statistics distribution. However, this condition is tightly connected to the critical value of FDR controlling procedures. In order to shed some light on this connection, we note that $\underline{\alpha}^\star = \pi_0 \alpha^\star_{BH}$ may be interpreted as the critical value of the Oracle BH procedure $BH(\alpha/\pi_0)$. Therefore, as the Oracle BH procedure at level α is the most powerful procedure among thresholding-based procedures that control FDR at level α, $\underline{\alpha}^\star$ is a lower bound on the critical values of these procedures. Specifically, multiple problems for which Condition (5) is satisfied or not differ in that:*

- *when Condition (5) is satisfied, all thresholding-based procedures that control FDR have null asymptotic power in a range of levels containing $[0, \underline{\alpha}^\star)$;*

- *when Condition (5) is not satisfied, some procedures (including BH) have positive asymptotic power for any positive level α.*

This paper extends the asymptotic results of Chi (2007a) and Neuvial (2008) to the case of plug-in procedures of the form $BH(\alpha/\widehat{\pi}_{0,m})$, where $\widehat{\pi}_{0,m}$ is a kernel estimator of the p-value distribution g at 1. Specifically, we consider a class of kernel estimators of π_0, which includes a modification of the Storey-λ estimator, where the parameter λ tends to 1 as $m \to \infty$. In Section 3, we prove that this class of estimators of π_0 achieves non-parametric convergence rates of the form $m^{-k/(2k+1)}/\eta_m$, where η_m goes to 0 slowly enough as $m \to +\infty$, and k controls the regularity of g at 1. In Section 4, we characterize the critical value $\alpha^\star_0$ of plug-in procedures based on such estimators, and prove that when the target FDR level α is greater than $\alpha^\star_0$, the convergence rate of these plug-in procedures is $m^{-k/(2k+1)}/\eta_m$, which is slower than the parametric rate achieved by the BH procedure and by the plug-in procedures studied in Neuvial (2008). In Section 5, these results are applied to one and two-sided tests in location and Student models. Practical consequences and possible extensions of this work are discussed in Section 6.

3. Asymptotic Properties of Non-Parametric Estimators of π_0

Let $\lambda \in (0,1)$. The expectation $\overline{\pi_0}(\lambda)$ of the Storey-λ estimator is given by

$$\overline{\pi_0}(\lambda) = \pi_0 + (1-\pi_0)\frac{1-G_1(\lambda)}{1-\lambda}. \tag{7}$$

Moreover, as a regular function of the empirical distribution of the p-values, $\hat{\pi}_{0,m}^{\mathrm{Sto}}(\lambda)$ has the following asymptotic distribution for $\lambda \in (0,1)$ (Genovese and Wasserman, 2004):

$$\sqrt{m}\left(\hat{\pi}_{0,m}^{\mathrm{Sto}}(\lambda) - \overline{\pi_0}(\lambda)\right) \rightsquigarrow \mathcal{N}\left(0, \frac{G(\lambda)(1-G(\lambda))}{(1-\lambda)^2}\right).$$

In our setting, g_1 is positive, as noted in Section 2.1. Therefore, we have $G_1(\lambda) < 1$ for any $\lambda \in (0,1)$, and the bias $\overline{\pi_0}(\lambda) - \pi_0$ is positive: the Storey-λ estimator achieves a parametric convergence rate, but it is not a consistent estimator of π_0. Under Assumption (2), this bias decreases as λ increases (by Equation (7)). In order to mimic the Oracle BH(α/π_0) procedure, it is therefore natural to choose λ close to 1. We consider plug-in procedures where π_0 is estimated by $\hat{\pi}_{0,m}^{\mathrm{Sto}}(1-h_m)$, with $h_m \to 0$ as $m \to +\infty$. As the limit in probability of this estimator is $g(1) = \pi_0 + (1-\pi_0)g_1(1)$, it is consistent if and only if the following "purity" condition, which has been introduced by Genovese and Wasserman (2004), is met:

$$g_1(1) = 0 \tag{8}$$

We note that the Storey-λ estimator is not a consistent estimator of π_0 even in when Condition (8) is met. Moreover, Condition (8) is entirely determined by the shape of the test statistics under the alternative hypothesis. The asymptotic bias and variance of $\hat{\pi}_{0,m}^{\mathrm{Sto}}(1-h_m)$ are characterized by Proposition 13:

Proposition 13 (Asymptotic bias and variance of $\hat{\pi}_{0,m}^{\mathrm{Sto}}(1-h_m)$) *Let* $(h_m)_{m\in\mathbb{N}}$ *be a positive sequence such that* $h_m \to 0$.

1. If $mh_m \to +\infty$ as $m \to +\infty$, then

$$\sqrt{mh_m}\left(\hat{\pi}_{0,m}^{\mathrm{Sto}}(1-h_m) - \mathbb{E}\left[\hat{\pi}_{0,m}^{\mathrm{Sto}}(1-h_m)\right]\right) \rightsquigarrow \mathcal{N}(0, g(1)).$$

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

$$\mathbb{E}\left[\hat{\pi}_{0,m}^{\mathrm{Sto}}(1-h_m)\right] - g(1) \underset{m\to+\infty}{=} \frac{(-1)^k g^{(k)}(1)}{(k+1)!}h_m^k + \mathrm{o}\left(h_m^k\right).$$

Only the bias term in Proposition 13 depends on the regularity k of the distribution near 1: the asymptotic bias is of order h_m^k, while the asymptotic variance of $\hat{\pi}_{0,m}^{\mathrm{Sto}}(1-h_m)$ is of order $(mh_m)^{-1}$, regardless of the regularity of the distribution. The bandwidth h_m in Proposition 13 realizes a trade-off between the asymptotic bias and variance of $\hat{\pi}_{0,m}^{\mathrm{Sto}}(1-h_m)$. When the regularity of the distribution is known, a natural way to resolve this bias/variance trade-off is to calibrate h_m such that the Mean Squared Error (MSE) of the corresponding estimator is asymptotically minimum. This gives rise to an optimal choice of the bandwidth, which is characterized by the following proposition:

Proposition 14 (Asymptotic properties of $\hat{\pi}_{0,m}^{\text{Sto}}(1-h_m)$) *Assume that g is k times differentiable at 1 for $k \geq 1$, with $g^{(l)}(1) = 0$ for $1 \leq l < k$.*

1. *If $g^{(k)}(1) \neq 0$, then the asymptotically optimal bandwidth for $\hat{\pi}_{0,m}^{\text{Sto}}(1-h_m)$ in terms of MSE is of order $m^{-1/(2k+1)}$, and the corresponding MSE is of order $m^{-2k/(2k+1)}$.*

2. *Let η_m be any sequence such that $\eta_m \to 0$ and $m^{k/(2k+1)}\eta_m \to +\infty$ as $m \to +\infty$. Then, letting $h_m(k) = m^{-1/(2k+1)}\eta_m^2$, we have, as $m \to +\infty$:*

$$m^{k/(2k+1)}\eta_m \left(\hat{\pi}_{0,m}^{\text{Sto}}(1-h_m(k)) - g(1) \right) \rightsquigarrow \mathcal{N}(0, g(1)). \tag{9}$$

Proposition 14 is proved in Appendix B. The convergence rate in (9) is a typical convergence rate for non-parametric estimators of a density at a point. However, Proposition 14 cannot be derived from classical results on kernel estimators such as those obtained in Tsybakov (2009). Indeed, such results typically require that the order of the kernel matches the regularity k of the density, whereas the kernel of Storey's estimator, $K^{\text{Sto}}(t) = \mathbf{1}_{[-1,0]}(t)$, is of order 0. The results that can be obtained with kernels of order k are summarized by Proposition 15; we refer to Tsybakov (2009) for a proof of this result.

Proposition 15 (k^{th} order kernel estimator) *Assume that for $k \geq 1$, g is k times differentiable at 1. Let $\hat{g}_m^k(1)$ be a kernel estimator of $g(1)$ with bandwidth h_m, associated with a kth order kernel.*

1. *The optimal bandwidth for $\hat{g}_m^k(1)$ in terms of MSE is of order $m^{-1/(2k+1)}$, and the corresponding MSE is of order $m^{-2k/(2k+1)}$;*

2. *Let η_m be any sequence such that $\eta_m \to 0$ and $m^{k/(2k+1)}\eta_m \to +\infty$ as $m \to +\infty$. Then letting $h_m(k) = m^{-1/(2k+1)}\eta_m^2$, we have, as $m \to +\infty$:*

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

Propositions 14 and 15 show that the convergence rate of kernel estimators of $g(1)$ with asymptotically optimal bandwidth directly depends on the regularity k of g at 1. The only difference between the two propositions is that the assumption that the first $k-1$ derivatives of g are null at 1 for $\hat{\pi}_{0,m}(1-h_m)$ is not needed for k^{th} order kernel estimators. Importantly, these convergence rates cannot be improved in our setting, in the sense that $m^{-k/(2k+1)}$ is the minimax rate for the estimation of a density at a point where its regularity is of order k (Tsybakov, 2009, Chapter 2).

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). We now briefly review asymptotic properties of these estimators in the context of multiple testing, as stated in Genovese and Wasserman (2004), and show that their convergence rates can essentially be recovered by Propositions 14 and 15.

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, Genovese and Wasserman (2004) obtained an estimator which converged to $g(1)$ at rate $(\ln m)^{1/3}m^{-1/3}$. The same rate of convergence can be achieved by Proposition 14 or 15 (for $\eta_m = (\ln m)^{-1/3}$) if we assume that g is differentiable at 1. This is a slightly stronger assumption than the ones made by Hengartner and Stark (1995), but it still corresponds to a regularity of order 1.

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 $g^{(1)}$ are null, and $g^{(2)}$ 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. The same rate of convergence can be achieved by Proposition 14 or 15 (for $\eta_m = (\ln m)^{-\delta}$) if one assumes that g is twice differentiable at 1 (and additionally that $g^{(1)}(1) = 0$ for Proposition 14). In our setting, the Lipschitz condition for the second derivative is unnecessary: the minimum of g is necessarily achieved at 1 because g is non-increasing (under Assumption (2)), so the first step of the estimation in Swanepoel (1999) may be omitted.

As both estimators are estimators of $g(1)$, the differences in their asymptotic properties are driven by the differences in the regularity assumptions made for g (or g_1) near 1, rather than by their specific form.

4. Consistency, Criticality and Convergence Rates of Plug-in Procedures

The aim of this section is to derive convergence rates for plug-in procedures based on the estimators $\hat{\pi}_{0,m}$ of π_0 studied in Section 3. Specifically, our goal is to establish central limit theorems for the threshold $\hat{\tau}_m^0(\alpha)$ of the plug-in procedure $BH(\alpha/\hat{\pi}_{0,m})$ and the associated False Discovery Proportion, which we denote by $FDP_m(\hat{\tau}_m^0(\alpha))$. The convergence results obtained by Neuvial (2008) cover a broad class of FDR controlling procedures, including the BH procedure and plug-in procedures based on estimators of π_0 that depend on the observations only through the empirical distribution function $\widehat{\mathbb{G}}_m$ of the p-values (Storey, 2002; Storey et al., 2004; Benjamini et al., 2006). Although these results were obtained in the conditional setting of Benjamini and Hochberg (1995), extending them to the unconditional setting considered here is relatively straightforward, because the proof techniques developed in Neuvial (2008) can be adapted to this setting. For completeness, the asymptotic properties of the BH procedure and the plug-in procedure based on the Storey-λ estimator are derived in Appendix C. The problem considered in this section is more challenging, as the kernel estimators introduced in Section 3 depend on m not only through $\widehat{\mathbb{G}}_m$, but also through the bandwidth of the kernel (for example, h_m for $\hat{\pi}_{0,m}^{\mathrm{Sto}}(1 - h_m)$).

Let $\hat{\pi}_{0,m}$ denote a generic estimator of π_0. We assume that $\hat{\pi}_{0,m}$ converges in probability to $\pi_{0,\infty} \leq 1$ as $m \to +\infty$. We do not assume that $\pi_{0,\infty} = \pi_0$. Therefore, $\hat{\pi}_{0,m}$ may or may not be a consistent estimator of π_0. We recall that the $BH(\alpha/\hat{\pi}_{0,m})$ procedure rejects all hypotheses with p-values smaller than

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

We now study the behavior of the $BH(\alpha/\hat{\pi}_{0,m})$ procedure when $\hat{\pi}_{0,m}$ converges at a rate r_m slower than the parametric rate $m^{-1/2}$, that is, $m^{-1/2} = o(r_m)$. We define the asymptotic threshold $\tau_\infty^0(\alpha)$ corresponding to $\hat{\tau}_m^0(\alpha)$ as

$$\tau_\infty^0(\alpha) = \sup\{t \in [0,1], G(t) \geq \pi_{0,\infty} t / \alpha\}.$$

We have $\tau_\infty^0(\alpha) = \tau_\infty(\alpha/\pi_{0,\infty})$, that is, the asymptotic threshold of the BH procedure defined in Equation (6) at level $\alpha/\pi_{0,\infty}$.

Theorem 16 (Asymptotic properties of plug-in procedures) *Let $\hat{\pi}_{0,m}$ be an estimator of π_0 such that $\hat{\pi}_{0,m} \to \pi_{0,\infty}$ in probability as $m \to +\infty$. Let $\alpha_0^\star = \pi_{0,\infty}\alpha_{BH}^\star$. Then:*

1. *$\alpha_0^\star$ is the critical value of the $BH(\alpha/\hat{\pi}_{0,m})$ procedure;*

2. *Further assume that the asymptotic distribution of $\hat{\pi}_{0,m}$ is given by*

$$\sqrt{mh_m}\,(\hat{\pi}_{0,m} - \pi_{0,\infty}) \rightsquigarrow \mathcal{N}(0,s_0^2)$$

for some s_0, with $h_m = o\left(1/\ln\ln m\right)$ and $mh_m \to +\infty$ as $m \to +\infty$. Then, under Assumption (2), for any $\alpha > \alpha_0^\star$,

(a) *The asymptotic distribution of the threshold $\widehat{\tau}_m^0(\alpha)$ is given by*

$$\sqrt{mh_m}\,(\widehat{\tau}_m^0(\alpha) - \tau_\infty^0(\alpha)) \rightsquigarrow \mathcal{N}\left(0, \left(\frac{s_0\tau_\infty^0(\alpha)/\alpha}{\pi_{0,\infty}/\alpha - g(\tau_\infty^0(\alpha))}\right)^2\right);$$

(b) *The asymptotic distribution of the FDP achieved by the $BH(\alpha/\hat{\pi}_{0,m})$ procedure is given by*

$$\sqrt{mh_m}\left(\text{FDP}_m(\widehat{\tau}_m^0(\alpha)) - \frac{\pi_0\alpha}{\pi_{0,\infty}}\right) \rightsquigarrow \mathcal{N}\left(0, \left(\frac{\pi_0\alpha s_0}{\pi_{0,\infty}^2}\right)^2\right).$$

Theorem 16 states that for $\alpha > \alpha_0^\star$, for any estimator $\hat{\pi}_{0,m}$ that converges in distribution at a rate r_m slower than the parametric rate $m^{-1/2}$, the plug-in procedure $BH(\alpha/\hat{\pi}_{0,m})$ converges at rate r_m as well. This is a consequence of the fact that r_m dominates the fluctuations of $\widehat{\mathbb{G}}_m$, which are of parametric order.

We now state the main result of the paper (Corollary 17), that is, the asymptotic properties of plug-in procedures associated with the estimators of π_0 studied in Section 3, for which $s_0^2 = g(1)$. This result can be derived by combining the results of Theorem 16 with those of Propositions 14 and 15.

Corollary 17 *Assume that (2) holds, and that g is k times differentiable at 1 for $k \geq 1$. Define $h_m(k) = m^{-1/(2k+1)}\eta_m^2$, where $\eta_m \to 0$ and $m^{k/(2k+1)}\eta_m \to +\infty$ as $m \to +\infty$. Denote by $\hat{\pi}_{0,m}^k$ one of the following two estimators of π_0:*

- *Storey's estimator $\hat{\pi}_{0,m}^{\text{Sto}}(1 - h_m(k))$; in this case, it is further assumed that $g^{(l)}(1) = 0$ for $1 \leq l < k$;*

- *A kernel estimator of $g(1)$ associated with a kth order kernel with bandwidth $h_m(k)$.*

Then

1. *$\alpha_0^\star = g(1)\alpha_{BH}^\star$ is the critical value of the $BH(\alpha/\hat{\pi}_{0,m}^k)$ procedure;*

2. *For any $\alpha > \alpha_0^\star$,*

(a) *The asymptotic distribution of the threshold $\widehat{\tau}_m^0(\alpha)$ is given by*

$$m^{k/(2k+1)}\eta_m\,(\widehat{\tau}_m^0(\alpha) - \tau_\infty^0(\alpha)) \rightsquigarrow \mathcal{N}\left(0, \left(\frac{\tau_\infty^0(\alpha)/\alpha}{g(1)/\alpha - g(\tau_\infty^0(\alpha))}\right)^2 g(1)\right);$$

(b) *The asymptotic distribution of the FDP achieved by the $BH(\alpha/\hat{\pi}_{0,m})$ procedure is given by*

$$m^{k/(2k+1)}\eta_m\left(\mathrm{FDP}_m(\hat{\tau}_m^0(\alpha)) - \frac{\pi_0\alpha}{g(1)}\right) \rightsquigarrow \mathcal{N}\left(0, \frac{\pi_0^2\alpha^2}{g(1)^3}\right).$$

We note that unlike the modification of the Storey-λ estimator studied here, the estimators of π_0 based on kernels of order k do not require the first $k-1$ derivatives of g at 1 to be null. Therefore, the latter are generally preferable to the former. Corollary 17 has the following consequences, which are also summarized in Table 1:

- Assume that Condition (8) is met. Then the asymptotic threshold of the $BH(\alpha/\hat{\pi}_{0,m})$ procedure is $\tau_\infty(\alpha/\pi_0)$, that is, the asymptotic threshold of the Oracle procedure $BH(\alpha/\pi_0)$. In particular, the asymptotic FDP achieved by the estimators in Corollary 17 is then *exactly* α (and its asymptotic variance is α^2/π_0), whereas the asymptotic FDP of the original BH procedure is $\pi_0\alpha$.

- We have
$$\underline{\alpha}^\star \leq \alpha_0^\star \leq \alpha_{\mathrm{Sto}(\lambda)}^\star \leq \alpha_{BH}^\star.$$

In models where Condition (5) is not satisfied, all the critical values in (4) are null, implying that all the corresponding procedures have positive power for any target FDR level. In models where Condition (5) is satisfied, all the critical values in (4) are positive, and (4) implies that the range of target FDR values α that yield asymptotically positive power is larger for the plug-in procedures studied in this paper than for the BH procedure or the Storey-λ procedure.

- We have $\tau_\infty^0(\alpha) \geq \tau_\infty^{0,\lambda}(\alpha) \geq \tau_\infty(\alpha)$, where $\tau_\infty^{0,\lambda}(\alpha)$ denotes the asymptotic threshold of the Storey-λ procedure, which is formally defined and characterized in Appendix C. Therefore, as the power of a thresholding-based FDR controlling procedure is a non-decreasing function of its threshold (Remark 6), the asymptotic power of the $BH(\alpha/\hat{\pi}_{0,m})$ procedure is greater than that of both the Storey-λ and the original BH procedures, even in the range $\alpha > \alpha_{BH}^\star$ where all of them have positive asymptotic power.

Name	$\hat{\pi}_{0,m}$	FDR$/\alpha$	Rate	(Asy. var. of FDP)/ FDR
BH	1	π_0	$m^{-1/2}$	$(\pi_0\tau_\infty(\alpha))^{-1}-1$
Oracle BH	π_0	1	$m^{-1/2}$	$(\tau_\infty(\alpha/\pi_0))^{-1}-1$
Storey-λ	$\hat{\pi}_{0,m}^{\mathrm{Sto}}(\lambda)$	$\pi_0/\overline{\pi_0}(\lambda)$	$m^{-1/2}$	$(\pi_0\tau_\infty^{0,\lambda}(\alpha))^{-1}+(1-G(\lambda))^{-1}$
Kernel($h_m(k)$)	$\hat{\pi}_{0,m}^k$	$\pi_0/g(1)$	$m^{-k/(2k+1)}$	$g(1)^{-1}$

Table 1: Summary of the asymptotic properties of the FDR controlling procedures considered in this paper, for a target FDR level α greater than the (procedure-specific) critical value. Note that "Storey-λ" denotes the original procedure with a fixed λ, while our extension with $\lambda = 1 - h_m(k)$ is categorized in the table as a particular case of kernel estimator (last row). For Storey-λ, we also assume that $\lambda > \tau_\infty^{0,\lambda}(\alpha)$.

These results characterize the increase in asymptotic power achieved by plug-in procedures based on kernel estimators of π_0. However, this increased asymptotic power comes at the price of a slower

convergence rate. Specifically, the convergence rate of plug-in procedures is the non-parametric rate $m^{-k/(2k+1)}/\eta_m$ (where k controls the regularity of g) for the $\mathrm{BH}(\alpha/\hat{\pi}_{0,m}^k)$ procedure, while the parametric rate $m^{-1/2}$ was achieved by the original BH procedure, the Oracle BH procedure, and the Storey-λ procedure (as proved in Appendix C).

5. Application to Location and Student Models

In Section 4 we proved that the asymptotic behavior of plug-in procedures depends on whether the target FDR level α is above or below the critical value $\alpha_0^\star$ characterized by Theorem 16, and by establishing convergence rates for these procedures when $\alpha > \alpha_0^\star$. Both the critical value $\alpha_0^\star$ and the obtained convergence rates depend on the test statistics distribution. In the present section, these results are applied to Gaussian and Laplace location models, and to the Student model. We begin by defining these models (Section 5.1) and studying criticality in each of them (Section 5.2). Then, we derive convergence rates for plug-in procedures based on the kernel estimators of π_0 considered in Sections 3 and 4, both for two-sided tests (Section 5.3) and one-sided tests (Section 5.4).

5.1 Models for the Test Statistics

In this subsection, we define the location and Student models used throughout Section 5.

5.1.1 LOCATION MODELS

In location models 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 $\theta > 0$. The most widely studied location models are the Gaussian and Laplace (double exponential) location models. Both the Gaussian and the Laplace distribution can be viewed as instances of a more general class of distributions introduced by Subbotin (1923) and given for $\gamma \geq 1$ by

$$f_0^\gamma(x) = \frac{1}{C_\gamma} e^{-|x|^\gamma/\gamma}, \text{ with } C_\gamma = \int_{-\infty}^{+\infty} e^{-|x|^\gamma/\gamma} dx = 2\Gamma(1/\gamma)\gamma^{1/\gamma-1}.$$

Therefore, the likelihood ratio in the γ-Subbotin location model may be written as

$$\frac{f_1^\gamma}{f_0^\gamma}(x) = \exp\left(\frac{|x|^\gamma}{\gamma} - \frac{|x-\theta|^\gamma}{\gamma}\right). \tag{10}$$

The Gaussian case corresponds to $\gamma = 2$ and the Laplace case to $\gamma = 1$. In the Laplace case, the distribution of the p-values under the alternative can be derived explicitly, see Lemma 22 in Appendix. We focus on $1 \leq \gamma \leq 2$ as this corresponds to situations in which Assumption (2) is satisfied. Specifically, for one-sided tests, Assumption (2) holds as soon as $\gamma \geq 1$, because then f_1^γ/f_0^γ is non-decreasing; for two-sided tests, if additionally $\gamma \leq 2$, then Assumption (2) holds (as proved in Appendix A, Proposition 23).

5.1.2 STUDENT MODEL

Student's t distribution is widely used in applications, as it naturally arises when testing equality of means of Gaussian random variables with unknown variance. In the Student model with parameter $\nu > 0$, F_0 is the (central) t distribution with ν degrees of freedom, and F_1 is the non-central t distribution with ν degrees of freedom and non-centrality parameter $\theta > 0$. The Student model is not

a location model, as F_1 cannot be written as a translation of F_0. Following Chi (2007a, Equation (3.5)), we note that the likelihood ratio of the Student model may be written as

$$\frac{f_1}{f_0}(t) = \sum_{j=0}^{+\infty} a_j(\nu, \theta) \psi_{(j,\nu)}(t), \tag{11}$$

where $\psi_{(j,\nu)}(t) = (t/\sqrt{t^2+\nu})^j = \operatorname{sgn}(t)^j \left(1+\nu/t^2\right)^{-j/2}$ for $t \in \mathbb{R}$ and

$$a_j(\nu, \theta) = e^{-\theta^2/2} \frac{\Gamma((\nu+j+1)/2)}{\Gamma((\nu+1)/2)} \frac{(\sqrt{2}\theta)^j}{j!}.$$

Remark 18 *The sequence $a_j(\nu, \theta)$ is positive, and it is not hard to see that $(\sum_j a_j(\nu, \theta))$ is a convergent series using Stirling's formula. Therefore, as $\psi_{(j,\nu)}(t) \in [-1,1]$, the dominated convergence theorem ensures that the right-hand side of Equation* (11) *is well-defined for any $t \in \mathbb{R}$.*

Another useful expression for the Student likelihood ratio may be derived from the integral expression of the density of a non-central t distribution given by Johnson and Welch (1940):

$$\frac{f_1}{f_0}(t) = \exp\left[-\frac{\theta^2}{2} \frac{1}{1+\frac{t^2}{\nu}}\right] \frac{Hh_\nu\left(-\frac{\theta t}{\sqrt{\nu+t^2}}\right)}{Hh_\nu(0)}, \tag{12}$$

where $Hh_\nu(z) = \int_0^{+\infty} \frac{u^\nu}{\nu!} e^{-\frac{1}{2}(u+z)^2} dx$. As noted by Chi (2007a, Section 3.1), the likelihood ratio of Student test statistics is non-decreasing, which implies that Assumption (2) holds for one-sided tests. It also holds for two-sided tests, as proved in Appendix A, Proposition 26.

The location models and the Student model considered here are parametrized by two parameters: (i) a non-centrality parameter θ, which encodes a notion of distance between $\mathcal{H}_0$ and $\mathcal{H}_1$; (ii) a parameter which controls the (common) tails of the distribution under $\mathcal{H}_0$ and $\mathcal{H}_1$: γ for the γ-Subbotin model, and ν for the Student model with ν degrees of freedom.

5.2 Criticality

As the asymptotic behavior of plug-in procedures crucially depends on whether the target FDR level is above or below the critical value $\alpha_0^\star$ characterized by Theorem 16, it is of primary importance to study criticality in the models we are interested in. Noting that $\alpha_0^\star = \pi_{0,\infty} \alpha_{BH}^\star = \pi_{0,\infty} \underline{\alpha}^\star / \pi_0$, we have $\alpha_0^\star > 0$ if and only if Condition (5) is satisfied, that is, if and only if the likelihood ratio f_1/f_0 is bounded near $+\infty$. In this section, we study Condition (5) in location and Student models.

5.2.1 LOCATION MODELS

In location models, where $f_1 = f_0(\cdot - \theta)$ with $\theta > 0$, the behavior of the likelihood ratio is closely related to the tail behavior of the distribution of the test statistics: for a given non-centrality parameter θ, the heavier the tails, the smaller the difference between f_1 and f_0. In a γ-Subbotin location model, Equation (10) yields $|1 - \theta/x|^\gamma \sim 1 - \gamma\theta/x$ as $x \to +\infty$. Thus $|x|^\gamma \left(1 - |1 - \theta/x|^\gamma\right) \sim \gamma\theta x^{\gamma-1}$, and the behavior of the likelihood ratio f_1^γ/f_0^γ is driven by the value of γ, as illustrated by Figure 2 for the Gaussian and Laplace location models with location parameter $\theta \in \{1, 2\}$.

If $\gamma > 1$, then $\lim_{+\infty} f_1^\gamma/f_0^\gamma = +\infty$. Therefore, the slope of the cumulative distribution function of the p-values is infinite at 0, and Condition (5) is not satisfied for the Subbotin model: $\underline{\alpha}^\star = 0$ for
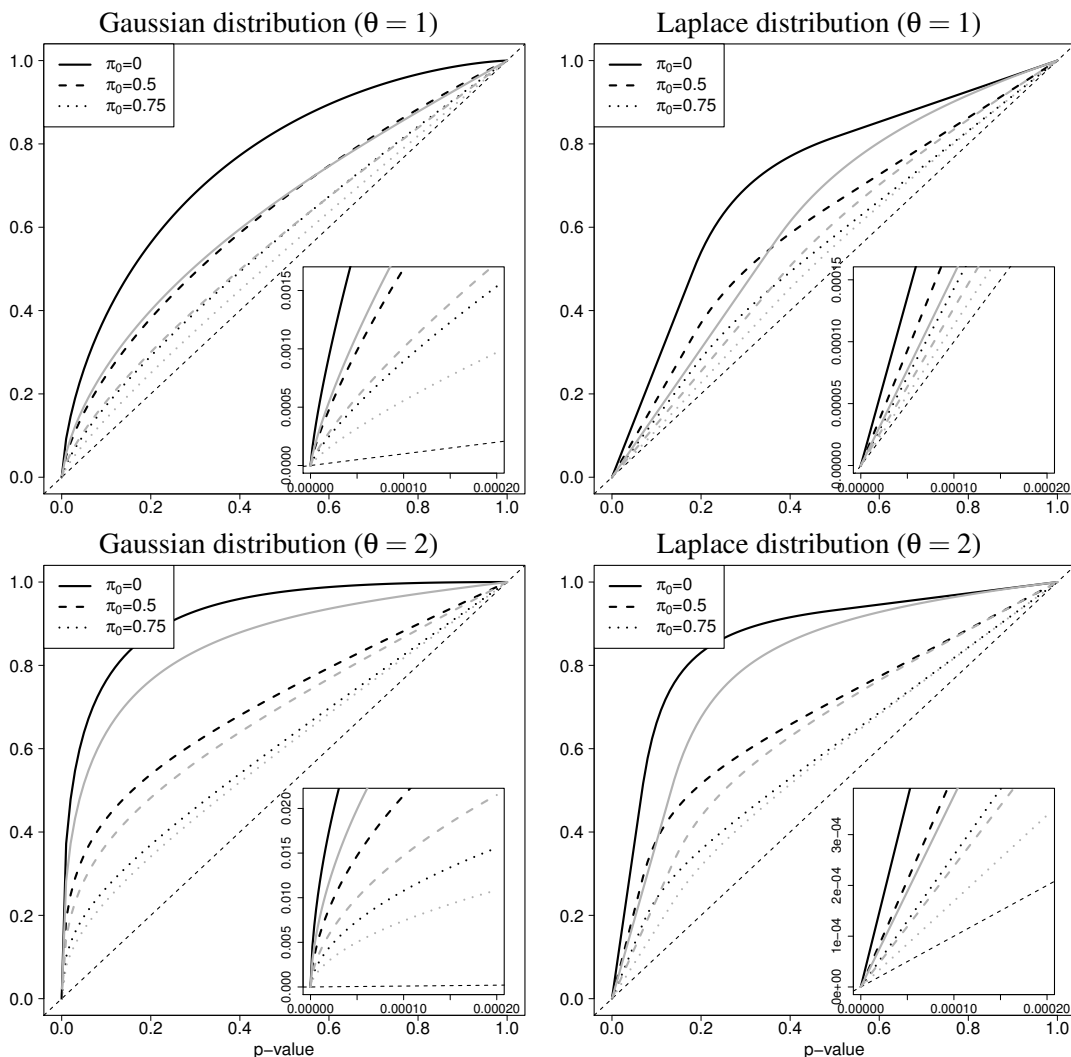
Figure 2: Distribution functions G for one-sided p-values (black) and two-sided p-values (gray), in Gaussian location models (left: Condition (5) is not satisfied), and Laplace location models (right: Condition (5) is satisfied) for $\pi_0 = 0$, 0.5 and 0.75. The location parameter θ is set to 1 in top panels and 2 in bottom panels. Inserted plot: zoom in the region $p < 2.10^{-4}$.

any θ and π_0. This situation is illustrated by Figure 2 (left panels) for the Gaussian model ($\gamma = 2$). In such a situation, for any target FDR level α, the asymptotic fraction of rejections by the BH(α) procedure or by a plug-in procedure of the form BH($\alpha/\hat{\pi}_{0,m}$), where $\hat{\pi}_{0,m} \to \pi_{0,\infty}$ in probability as $m \to +\infty$, is positive by Lemma 27.

If $\gamma = 1$ (Laplace model, as illustrated by Figure 2, right panels), then the likelihood ratio of the model is $f_1^\gamma / f_0^\gamma(x) = \exp(|x| - |x - \theta|)$. It is bounded as $x \to +\infty$, with $\lim_{x \to +\infty} f_1^\gamma / f_0^\gamma(x) = e^\theta$. Therefore, Condition (5) is satisfied for the Laplace location model. Specifically, we have $\underline{\alpha}^\star =$

$\pi_0/(\pi_0 + (1 - \pi_0)g_1(0))$, with $g_1(0) = e^\theta$ for one-sided p-values, and $g_1(0) = \cosh\theta$ for two-sided p-values. Laplace-distributed test statistics appear as a limit situation in terms of criticality: within the family of γ-Subbotin location models with $\gamma \in [1,2]$, the Laplace model ($\gamma = 1$) is the only one for which Condition (5) is satisfied.

5.2.2 STUDENT MODEL

For the Student model, Equation (12) yields that $(f_1/f_0)(t)$ converges to $s_v(\theta)$ as $t \to +\infty$ and $s_v(-\theta)$ as $t \to -\infty$, where $s_v(\theta) = Hh_v(-\theta)/Hh_v(0)$ is positive for any θ. Therefore, Condition (5) is satisfied for one-sided and two-sided tests in the Student model (this had already been noted by Chi (2007a) for one-sided tests). Figure 3 gives the distribution function of one- and two-sided p-values in the Student model with parameters $\theta \in \{1,2\}$ and $v \in \{10,50\}$, for $\pi_0 \in \{0,0.5,0.75\}$. Although criticality is much less obvious than for the Laplace model, the inserted plots which zoom into a region where the p-values are very small ($p < 2.10^{-4}$) do suggest for $v = 10$ that the slope of the distribution function has a finite limit at 0 for the Student model. As an illustration, we calculated that the critical values for one-sided tests in the Student model for $\pi_0 = 0.75$ for $\theta \in \{1,2\}$ are respectively 0.173 and 0.015 for $v = 10$, and 4.10^{-3} and 7.10^{-6} for $v = 50$.

5.3 Two-Sided Tests

In this section we study consistency and convergence rates for two-sided tests.

5.3.1 CONSISTENCY

Let us first recall that by Proposition 1.2, we have for two-sided tests under a model satisfying Assumption (1):

$$g_1(t) = \frac{1}{2}\left(\frac{f_1}{f_0}(q_0(t/2)) + \frac{f_1}{f_0}(-q_0(t/2))\right), \tag{13}$$

where $q_0 : t \mapsto F_0^{-1}(1-t)$ tends to 0 as $t \to 1/2$. A straightforward consequence of (13) is that $g_1(1) = (f_1/f_0)(0)$. As $f_1 > 0$, we have $g(1) = \pi_0 + (1 - \pi_0)g_1(1) > \pi_0$. Therefore, Condition (8) is not met, and the kernel estimators of π_0 studied in Section 3 are not consistent for the estimation of π_0. Specifically, we have $g_1(1) = e^{-\theta^2/2}$ for Gaussian and Student test statistics, and $g_1(1) = e^{-\theta}$ for Laplace test statistics.

5.3.2 CONVERGENCE RATES

Another consequence of (13) is that if for $k \geq 1$ the likelihood ratio f_1/f_0 is k times semi-differentiable at 0, then g is k times (left-)differentiable at 1. In particular, this holds *for any k* in the γ-Subbotin location model with $\gamma \in [1,2]$, which covers the Gaussian and Laplace cases. It also holds for the Student model (as proved in Proposition 25). For these models, Corollary 17 entails that for any $k > 0$, if $\hat{\pi}_{0,m}$ is a kernel estimator of g associated with a kth order kernel with bandwidth $h_m(k) = m^{-1/(2k+1)}\eta_m^2$ (where $\eta_m \to 0$ and $m\eta_m \to +\infty$ as $m \to +\infty$), then the corresponding plug-in procedure $\text{BH}(\alpha/\hat{\pi}_{0,m})$ converges in distribution at rate $m^{-k/(2k+1)}/\eta_m$ for any α greater than $\alpha_0^\star = g(1)\alpha_{BH}^\star$. These results are summarized in the last column of Table 2.

Let us now consider the modification of the Storey-λ estimator introduced in Section 3: $\hat{\pi}_{0,m} = \hat{\pi}_{0,m}^{\text{Sto}}(1 - h_m)$, with $h_m \to 0$ as $m \to +\infty$. By Corollary 17, the optimal convergence rate of the
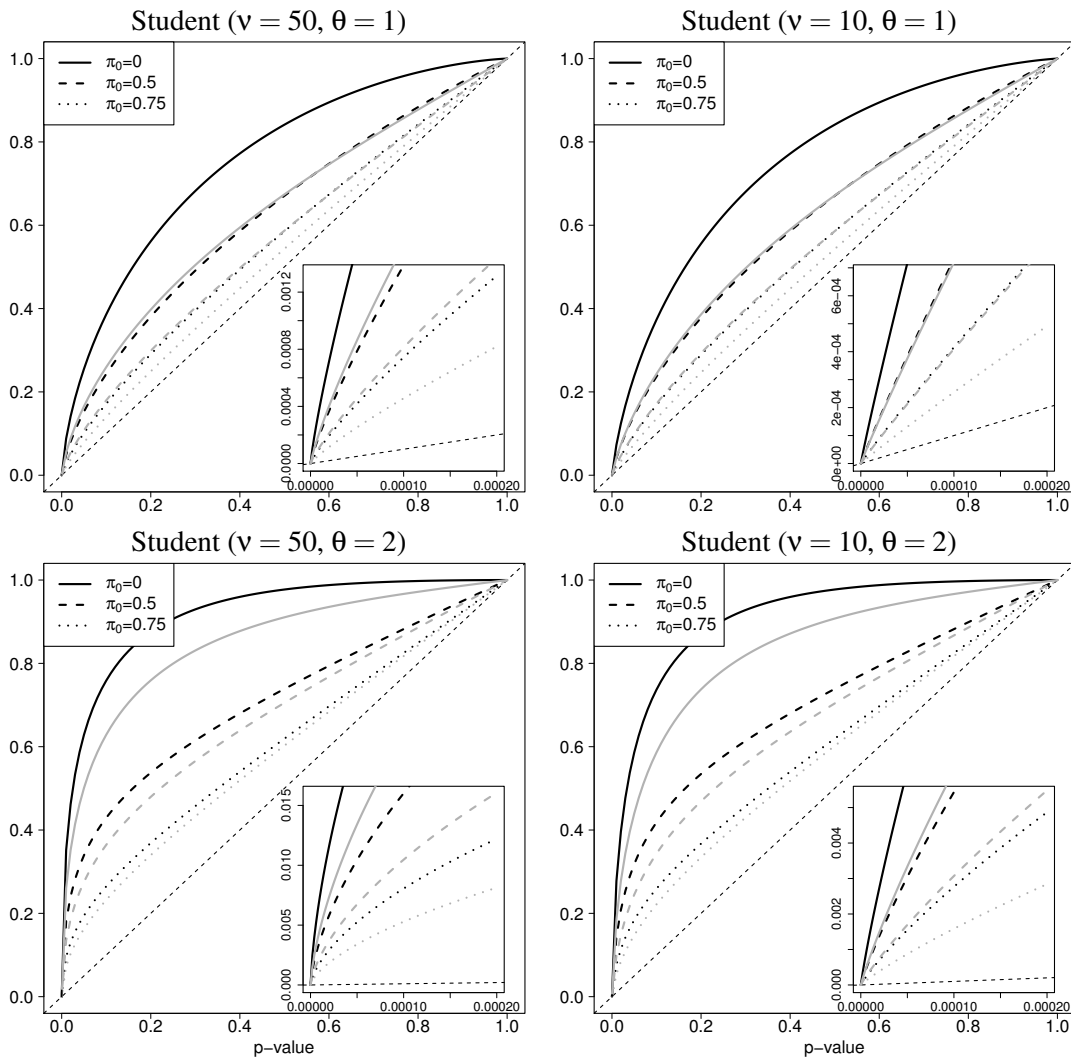
Figure 3: Distribution functions G for one-sided p-values (black) and two-sided p-values (gray) in Student models with $\nu = 50$ degrees of freedom (left) and $\nu = 10$ (right). The location parameter θ is set to 1 in top panels and 2 in bottom panels. Any Student model satisfies Condition (5). Inserted plots: zoom in the region $p < 2.10^{-4}$.

$\text{BH}(\alpha/\hat{\pi}_{0,m})$ procedure is then determined by the order of the first non null derivative of g at 1. In order to calculate this order, we use the following lemma:

Lemma 19 *Under Assumption* (1)*, the density function g_1 of two-sided p-values under the alternative hypothesis satisfies:*

1. If f_1/f_0 is semi-differentiable at 0, with left-derivative ℓ_- and right-derivative ℓ_+, then $g_1^{(1)}$ is semi-differentiable at 1 and we have:

$$g_1^{(1)}(1) = -\frac{\ell_+ - \ell_-}{4f_0(0)}. \tag{14}$$

   In particular, $g_1^{(1)}(1) = 0$ if and only if f_1/f_0 is differentiable at 0.

2. If f_1/f_0 is twice differentiable at 0, then $g_1^{(1)}$ is twice differentiable at 1 and we have:

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

Lemma 19 may be applied to two-sided tests for γ-Subbotin location models, and for the Student model. For the two-sided Gaussian model, f_1/f_0 is C^∞ near 0 and $(f_1/f_0)^{(2)}(0) \neq 0$. The same holds for the two-sided Student model, as shown in Appendix A.2 (Proposition 25). For both models, Lemma 19 entails that $g^{(1)}(1) = 0$ and $g^{(2)}(1) > 0$. For two-sided Laplace test statistics, the likelihood ratio $f_1/f_0 : t \mapsto \exp(|t - \theta| - |t|)$ has a singularity at $t = 0$ but it is semi-differentiable at 0 (and differentiable on $(-\infty, \theta) \setminus \{0\}$), with left and right derivatives at 0 given by $\ell_- = 0$ and $\ell_+ = e^{-\theta}$. Lemma 19 yields that $g^{(1)}(1) = -(1 - \pi_0)e^{-\theta}/2$. In particular, letting $k = 1$ for the Laplace model and $k = 2$ for the Gaussian and Student models, Corollary 17 yields that if $\hat{\pi}_{0,m} = \hat{\pi}_{0,m}^{Sto}(1 - m^{-1/(2k+1)}\eta_m^2)$, where $\eta_m \to 0$, then for any $\alpha > \alpha_0^\star = g(1)\alpha_{BH}^\star$, the FDP of the $BH(\alpha/\hat{\pi}_{0,m})$ procedure converges in distribution at rate $m^{-k/(2k+1)}/\eta_m$ toward $\pi_0\alpha/g(1)$, where $g(1) = \pi_0 + (1 - \pi_0)e^{-\theta^2/2}$ in the Gaussian and Student models, and $g(1) = \pi_0 + (1 - \pi_0)e^{-\theta}/2$ in the Laplace model. These rates are slower than those obtained at the beginning of this section for k^{th} order kernels because the latter do not require the derivatives of g of order $l < k$ to be null at 1, which implied that any $k > 0$ could be chosen (see Table 2 for a comparison).

5.4 One-Sided Tests

In this section we study consistency and convergence rates for one-sided tests.

5.4.1 CONSISTENCY

For one-sided tests, we have $g_1(t) = (f_1/f_0)(q_0(t))$. As $\lim_{t \to 1} q_0(t) = -\infty$, Condition (8) is met if and only if the likelihood ratio $(f_1/f_0)(t)$ tends to 0 as $t \to -\infty$. For the Student model, f_1/f_0 tends to $s_\nu(-\theta) > 0$ as $t \to -\infty$. This implies that Condition (8) is not satisfied in that model: π_0 cannot be consistently estimated using a consistent estimator of $g(1)$, because $g(1) = \pi_0 + (1 - \pi_0)e^{-\theta} > \pi_0$. For location models, we begin by establishing a connection between purity and criticality (Proposition 21), which is a consequence of the following symmetry property:

Lemma 20 (Likelihood ratios in symmetric location models) *Consider a location model in which the 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 Assumption* (1)*, we have*

$$\lim_{-\infty}\frac{f_0}{f_1} = \lim_{+\infty}\frac{f_1}{f_0}.$$

For one-sided tests in symmetric location models, Lemma 20 implies the following result:

Proposition 21 (Purity and criticality) *Let g_1 be the density of one-sided p-values under the alternative hypothesis, and $\underline{\alpha}^\star$ the critical value of the multiple testing problem. Under Assumption (1) and Assumption (2),*

1. *Condition (5) and Condition (8) are complementary events, in the sense that $\underline{\alpha}^\star = 0$ if and only if $g_1(1) = 0$;*

2. *If $\lim_{+\infty} f_1/f_0$ is finite, then $\underline{\alpha}^\star = \pi_0/(\pi_0 + (1-\pi_0)g_1(0))$ and $g(1) = \pi_0 + (1-\pi_0)g_1(1)$ are connected by $g_1(0)g_1(1) = 1$.*

Proposition 21 implies that contrary to two-sided location models, in which we always have $g_1(1) > 0$, consistency *may* be achieved in one-sided location models using kernel estimators such as those considered here, depending on model parameters. In particular, there is no criticality in the one-sided Gaussian model, implying that Condition (8) is satisfied in that model: we have $g(1) = \pi_0$, and π_0 can be consistently estimated using the kernel estimators of $g(1)$ introduced in Section 3. In the one-sided Laplace model, Condition (5) is satisfied, implying that Condition (8) is not satisfied in that model: π_0 cannot be consistently estimated using these kernel estimators of $g(1)$.

5.4.2 CONVERGENCE RATES

For the one-sided Student model, Proposition 25 entails that g_1 is C^∞, and all its derivatives or order greater than 1 are null at 1. Therefore, any $k > 0$, if $\hat{\pi}_{0,m}^k$ denotes any of the two estimators studied in Corollary 17 for a kth order kernel with bandwidth $h_m(k) = m^{-1/(2k+1)}\eta_m^2$ (where $\eta_m \to 0$ and $m\eta_m \to +\infty$ as $m \to +\infty$), then the corresponding plug-in procedure $\text{BH}(\alpha/\hat{\pi}_{0,m}^k)$ converges in distribution at rate $m^{-k/(2k+1)}/\eta_m$ for any α greater than $\alpha_0^\star = g(1)\alpha_{BH}^\star$. These results are summarized in the first row of Table 2.

For the one-sided Laplace model, the distribution of the p-values satisfies $G_1(t) = 1 - (1-t)e^{-\theta}$ for $t \geq 1/2$, see Lemma 22 in Appendix A. Therefore, for $t \geq 1/2$, $(1 - G(t))/(1-t)$ is constant, equal to $g(1) = \pi_0 + (1-\pi_0)e^{-\theta}$, as illustrated by the solid curves in the right panels of Figure 2. Therefore, for any fixed $\lambda \geq 1/2$, the Storey-λ estimator is an unbiased estimator of $g(1)$, which converges to $g(1)$ at rate $m^{-1/2}$. The same property holds for any kernel estimator of $g(1)$ with a fixed bandwidth. These results are summarized in the third row of Table 2.

In the Gaussian model however, the regularity of g_1 near 1 is poor for one-sided tests: we have

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

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

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

Therefore, g_1 is not differentiable at 1, and the convergence rates of the kernel estimators of π_0 studied in Section 3 are slower than $m^{-1/3}$ in our setting. These results are summarized in the second row of Table 2.

The difference between one- and two-sided tests in the Gaussian location model is illustrated by Figure 4 for $\theta = 1$, that is when testing $\mathcal{N}(0,1)$ against $\mathcal{N}(1,1)$. 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 $g(1) =$

Model	$\lim_0 1/g_1$	$g_1(1)$	Convergence rates $\hat{\pi}_{0,m}^{\text{Sto}}(1-h_m(k))$	$\hat{g}_m^k(1)/\eta_m$
1-sided Student	$s_v(\theta)$	$s_v(-\theta)$	$\ll m^{-k/(2k+1)}/\eta_m$	$\ll m^{-k/(2k+1)}/\eta_m$
1-sided Gaussian	0	0	$\gg m^{-1/3}$	$\gg m^{-1/3}$
1-sided Laplace	$e^{-\theta}$	$e^{-\theta}$	$m^{-1/2}$	$m^{-1/2}$
2-sided Student	$(s_v(\theta)+s_v(-\theta))/2$	$e^{-\theta^2/2}$	$m^{-2/5}/\eta_m$	$\ll m^{-k/(2k+1)}/\eta_m$
2-sided Gaussian	0	$e^{-\theta^2/2}$	$m^{-2/5}/\eta_m$	$\ll m^{-k/(2k+1)}/\eta_m$
2-sided Laplace	$\cosh\theta$	$e^{-\theta}$	$m^{-1/3}/\eta_m$	$\ll m^{-k/(2k+1)}/\eta_m$

Table 2: Properties of one- and two-sided test statistics distributions in Student, Gaussian, and Laplace models, and convergence rates of the kernel estimators studied. When the rate depends on k, the value of k may be chosen arbitrarily large. η_m is a sequence such that $\eta_m \to 0$ and $m\eta_m \to +\infty$ as $m \to +\infty$.

$\pi_0 + (1-\pi_0)e^{-\theta^2/2}$ at rate $m^{-2/5}$, by Corollary 17. Conversely, the density of one-sided p-values tends to 0 at 1, but is not differentiable: the true π_0 can be estimated consistently, but the convergence rate is slower.

6. Concluding Remarks

This paper studies asymptotic properties of a family of plug-in procedures based on the BH procedure. When compared to the BH procedure or to the Storey-λ procedure, the results for general models obtained in Section 4 show that incorporating the proposed estimators of π_0 into the BH procedure asymptotically yields (i) tighter FDR control (or, equivalently, greater power) and (ii) smaller critical values, thereby increasing the range of situations in which the resulting procedure has positive asymptotic power. These improvements come at the price of a reduction in the convergence rate from the parametric rate $m^{-1/2}$ to a non-parametric rate $m^{-k/(2k+1)}$, where k is connected to the order of differentiability of the test statistics distribution. As the results obtained for the proposed modification of the Storey-λ estimator $\hat{\pi}_{0,m}^{\text{Sto}}(1-h_m)$ require stronger conditions (null derivatives of g_1) than for kernel estimators with a kernel of order k, we conclude that it is generally better to use the latter class of estimators.

Our application of these results to specific models for the test statistics sheds some light on the influence of the test statistics distribution on convergence rates of plug-in procedures:

- When the test statistics distribution is C^∞ (which is the case for two-sided Gaussian test statistics, and for Laplace and Student tests statistics), the obtained convergence rates are slower than the parametric rate, but may be arbitrarily close to it by choosing a kernel of sufficiently high order. The resulting estimators are not consistent estimators of π_0, although the bias decreases as the non-centrality parameter θ increases.

- When the regularity of the test statistics distribution is poor (such as in the one-sided Gaussian model), the convergence rate of the FDP achieved by the plug-in procedures studied in this paper is slower. The plug-in procedures studied are still asymptotically more powerful than the BH procedure or the Storey-λ procedure, but the FDP actually achieved by the plug-in procedures studied for a given m may be far from the target FDR level.
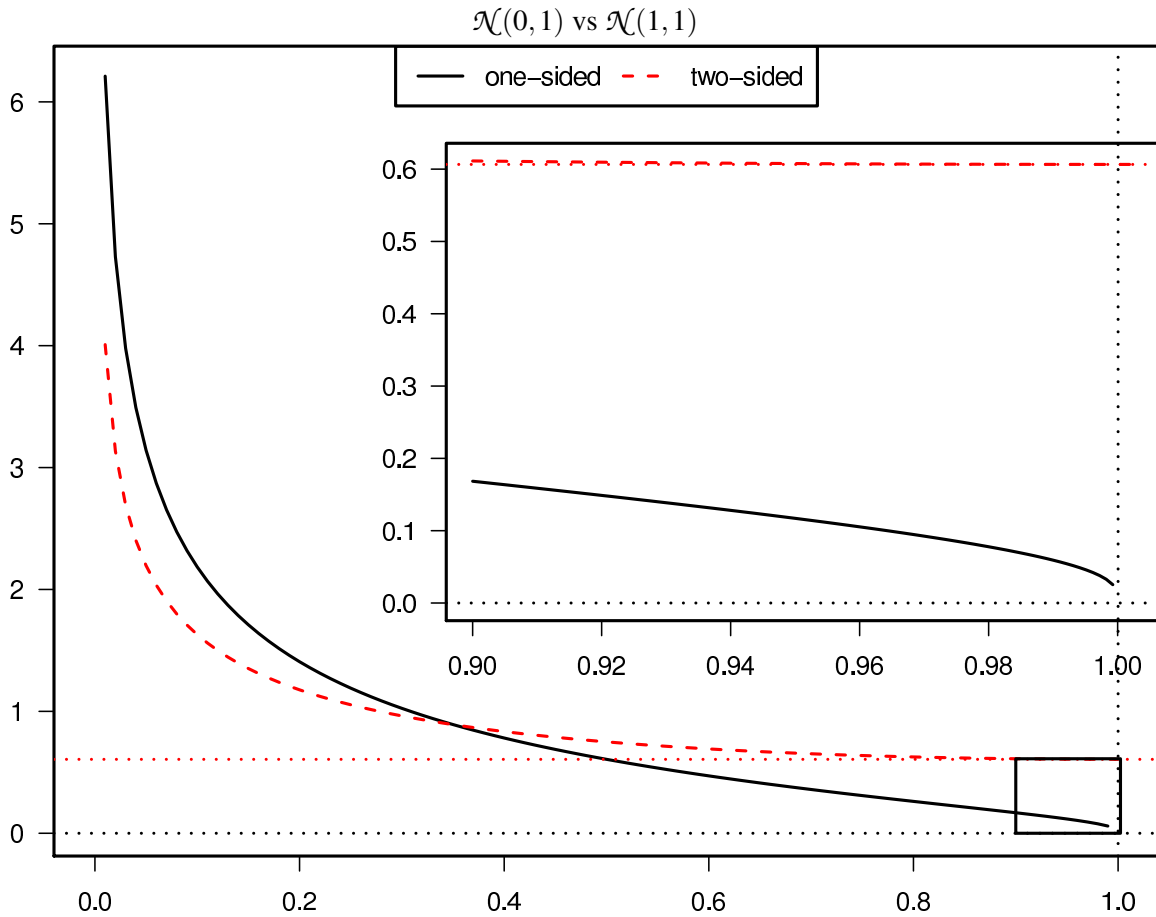
Figure 4: Density of one- and two-sided p-values under the alternative hypothesis for the location model $\mathcal{N}(0,1)$ versus $\mathcal{N}(1,1)$. Inserted plot: zoom in the region $[0.9, 1]$, which is highlighted by a black box in the main plot.

Obtaining more precise conclusions in the context of a specific data set or application exceeds the scope of the present paper, as it would require extending the obtained results to more realistic settings such as the ones that are now described.

6.1 Extensions of the Multiple Testing Setting Considered

An interesting research direction would be to extend the multiple testing setting considered here to more realistic assumptions. A typical example of application is the case of differential expression analyses in genomics, which aim at identifying those genes whose expression level differs between two known populations of samples. First, we have assumed that all null hypotheses are independent, and that all true alternative hypotheses follow the same distribution. The independence assumption is not realistic, as genes are known to interact with each other, in particular through transcriptional regulation networks. Moreover, the degree of differential expression needs not be identical for all genes under $\mathcal{H}_1$. As for the results on criticality that have been used in this paper, the proof given

in Chi (2007a) essentially relies on the assumption that the *p*-values are independently and identically distributed. Therefore, it seems that these results could be extended to composite distributions under $\mathcal{H}_1$, provided that the corresponding marginal distributions are still independently and identically distributed. Extending these results to settings where the independence assumption is relaxed seems a more challenging question. As for the convergence results established in Section 4, their proofs rely on the formalism laid down by Neuvial (2008). Therefore, these results could be extended to other dependency assumptions, or to composite distributions under $\mathcal{H}_1$ provided that the convergence in distribution of the empirical distribution functions $(\widehat{\mathbb{G}}_{0,m}, \widehat{\mathbb{G}}_{1,m})$ holds under these assumptions. In that spirit, the results of Neuvial (2008) have been extended to an equi-correlated Gaussian model (Delattre and Roquain, 2011) and to a more general Gaussian model where the covariance matrix is supposed to be close enough to the identity as the number of tests grows to infinity (Delattre and Roquain, 2012).

Second, we have shown that the asymptotic properties of FDR controlling procedures are driven by the shape and regularity of the test statistics distribution. In practice, the test statistics distribution depends on the size of the sample used to generate them. In differential expression analyses, a natural test statistic is Student's *t*, whose distribution depends on sample size through both the number of degrees of freedom ν and a non-centrality parameter θ. In the spirit of the results of Chi (2007b) on the influence of sample size on criticality, it would be interesting to study the convergence rates of plug-in procedures when both the sample size and the number of hypotheses tested grow to infinity.

6.2 Alternative Strategies to Estimate π_0

The estimators of π_0 considered in this paper are kernel estimators of the density *g* at 1. Therefore, they achieve non-parametric convergence rates of the form $m^{-k/(2k+1)}/\eta_m$, where *k* controls the regularity of *g* near 1 and $\eta_m \to 0$ slowly enough. An interesting open question is whether these non-parametric rates may be improved. Other strategies for estimating π_0 may be considered to achieve faster convergence rates, including the following two:

- One-stage adaptive procedures as proposed by Blanchard and Roquain (2009) and Finner et al. (2009) allow more powerful FDR control than the standard BH 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 characteristic function of the *Z*-scores associated to the *p*-values. This estimator does not focus on the behavior of the density near 1, and might not suffer from the same limitations as the estimators studied here. This estimator was shown to be consistent for the estimation of π_0 when the *Z*-scores follow a Gaussian location mixture, but no convergence rates were established.

In a general semi-parametric framework where g_1 is not necessarily decreasing, and its regularity is not specified, Nguyen and Matias (2013) have recently proved that if the Lebesgue measure of the set on which g_1 achieves its minimum is null, then no consistent estimator of $\min_t g(t)$ with a finite asymptotic variance can reach the parametric convergence rate $m^{-1/2}$. In our setting where g_1 is decreasing, the measure of the set on which g_1 is minimum is indeed null, except if g_1 is constant on an interval of the form $[t_0, 1]$. For one-sided tests where $g_1(t) = (f_1/f_0)(F_0^{-1}(1-t))$, this extreme situation arises if and only if the likelihood ratio is constant on an interval of the form $[x_0, +\infty)$. Among all models studied in Section 5, the only case in which this occurs is the one-sided

Laplace model, where $(f_1/f_0)(x) = \exp(|x| - |x - \theta|) = e^\theta$ for $x \geq \theta > 0$. The kernel estimators that we have studied here do reach the rate $m^{-1/2}$ in this case.

In the more common situation in which the measure of the set on which g_1 vanishes (or achieves its minimum) is null, the above negative result of Nguyen and Matias (2013) suggests that there is little room for improving on the non-parametric convergence rates obtained in Propositions 14 and 15. We conjecture that it is not possible for consistent estimators of $g(1)$ to reach a parametric convergence rate in this setting.

Acknowledgments

Appendix A. Calculations in Specific Models

In this section we perform calculations in location and Student models.

A.1 Location Models

Lemma 22 gives the distribution of the p-value under the alternative hypothesis for one-sided tests in the Laplace model. The proof is straightforward, so it is omitted.

Lemma 22 (One-sided Laplace location model) *Assume that the probability distribution function 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*

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

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

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

4. *The probability distribution function of one-sided* p-*values under* $\mathcal{H}_1$ *is*

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

Proposition 23 (Concavity in two-sided γ-Subbotin models) *If the test statistics follow a γ-Subbotin distribution with $\gamma \in [1,2]$, then the distribution function of the two-sided p-values under the alternative G_1 is concave.*

Proof [Proof of Proposition 23] Assumption (1) holds for Subbotin models. By Lemma 2, we need to prove that the likelihood ratio $f_1^{\gamma}/f_0^{\gamma}$ of the γ-Subbotin model with γ is such that $h : x \mapsto (f_1^{\gamma}/f_0^{\gamma})(x) + (f_1^{\gamma}/f_0^{\gamma})(-x)$ is non-decreasing on $\mathbb{R}_+$. The function h is differentiable on $(0,+\infty) \setminus \{\theta\}$, and its derivative is given by

$$h'(x) = \left(\frac{f_1^{\gamma}}{f_0^{\gamma}}\right)'(x) - \left(\frac{f_1^{\gamma}}{f_0^{\gamma}}\right)'(-x),$$

where

$$\left(\frac{f_1^{\gamma}}{f_0^{\gamma}}\right)'(y) = \left(\text{sgn}(y)|y|^{\gamma-1} - \text{sgn}(y-\theta)|y-\theta|^{\gamma-1}\right)\frac{f_1^{\gamma}}{f_0^{\gamma}}(y) \tag{15}$$

for any $y \in \mathbb{R} \setminus \{0,\theta\}$. Let $x > 0$ such that $x \neq \theta$, we are going to prove that $h'(x) \geq 0$. As $f_1^{\gamma}/f_0^{\gamma}$ is non-decreasing, both $(f_1^{\gamma}/f_0^{\gamma})'(x)$ and $(f_1^{\gamma}/f_0^{\gamma})'(-x)$ are non-negative. If $(f_1^{\gamma}/f_0^{\gamma})'(-x) = 0$, then $h'(x) \geq 0$ as desired. From now on, we assume that $(f_1^{\gamma}/f_0^{\gamma})'(-x)$ is positive. As $\theta > 0$, (15) entails that

$$\frac{(f_1^{\gamma}/f_0^{\gamma})'(x)}{(f_1^{\gamma}/f_0^{\gamma})'(-x)} = \frac{x^{\gamma-1} - \text{sgn}(x-\theta)|x-\theta|^{\gamma-1}}{(x+\theta)^{\gamma-1} - x^{\gamma-1}} \frac{f_1(x)^{\gamma}}{f_1(-x)^{\gamma}},$$

where $f_1(x)^{\gamma} > f_1(-x)^{\gamma}$ because $-|x-\theta| + |x+\theta| > 0$. As $(f_1^{\gamma}/f_0^{\gamma})'(-x) > 0$, it is enough to show that

$$x^{\gamma-1} - \text{sgn}(x-\theta)|x-\theta|^{\gamma-1} \geq (x+\theta)^{\gamma-1} - x^{\gamma-1} \tag{16}$$

in order to prove that $h'(x) \geq 0$. By the concavity of $x \mapsto x^{\gamma-1}$ on $\mathbb{R}_+$ for $1 \leq \gamma \leq 2$, $\phi : x \mapsto \theta^{-1}(x^{\gamma-1} - (x-\theta)^{\gamma-1})$ is non-increasing on $[\theta,+\infty]$. Therefore, if $x > \theta$ we have $\phi(x) \geq \phi(x+\theta)$ and (16) holds. If $x < \theta$, then noting that for any $a, b > 0$ and $\zeta \in [0,1]$, $a^{\zeta} + b^{\zeta} \geq (a+b)^{\zeta}$, we have, for $1 \leq \gamma \leq 2$, $x^{\gamma-1} + (\theta-x)^{\gamma-1} \geq \theta^{\gamma-1} \geq (x+\theta)^{\gamma-1} - x^{\gamma-1}$, and (16) holds as well. ∎

A.2 Student Model

In this section, we use the function series expansion of the Student likelihood ratio given in (11) to study the regularity of this likelihood ratio.

Lemma 24 (Derivative of the Student likelihood ratio) *Let $\nu \in \mathbb{N}^*$ and $\theta > 0$. The likelihood ratio f_1/f_0 of the Student model with ν degrees of freedom and non-centrality parameter θ is C^1 on $\mathbb{R}$, and for any $t \in \mathbb{R}$,*

$$\left(\frac{f_1}{f_0}\right)'(t) = \nu(\nu+t^2)^{-3/2}\sum_{j=0}^{+\infty} a_j^1(\nu,\theta)\psi_{(j,\nu)}(t), \tag{17}$$

where $a_j^1(\nu,\theta) = (j+1)a_{j+1}(\nu,\theta)$ is such that $(\sum_j a_j^1(\nu,\theta))$ converges absolutely.

Proof [Proof of Lemma 24] As $(\sum_j a_j(\nu,\theta))$ converges absolutely and as $\psi_{(j,\nu)}$ is differentiable on $\mathbb{R}$ for any $j \geq 0$ and bounded (by [-1,1]), the dominated convergence theorem ensures that f_1/f_0 is differentiable on $\mathbb{R}$ and that its derivative is given by:

$$\left(\frac{f_1}{f_0}\right)'(t) = \sum_{j=1}^{+\infty} a_j(\nu,\theta)\psi'_{(j,\nu)}(t).$$

For $t \neq 0$, we have $\log\left(\text{sgn}(t)^j \psi_{(j,\nu)}(t)\right) = -j/2\left(\log(1+\nu/t^2)\right)$, whose derivative is given by $j\nu/(\nu t + t^3)$, so that

$$\psi'_{(j,\nu)}(t) = \psi_{(j,\nu)}(t)\frac{j\nu}{t(\nu+t^2)}.$$

As $\psi_{(j,\nu)}(t) \underset{t\to 0}{\sim} (t/\sqrt{\nu})^j$, we have $\psi_{(j,\nu)}(0) = 0$, $\psi'_{(j,\nu)}(0) = 0$, and $\psi'_{(j,\nu)}$ is continuous at 0. Equation (17) follows by noting that $\psi_{(j+1,\nu)}(t)/\psi_{(j,\nu)}(t) = t/\sqrt{t^2+\nu}$, and that $(\sum_j a_j^1(\nu,\theta))$ converges absolutely by Stirling's formula. ∎

Lemma 24 entails the following result:

Proposition 25 (Regularity of the Student likelihood ratio) *Let $\nu \in \mathbb{N}^*$ and $\theta > 0$. The likelihood ratio f_1/f_0 of the Student model with ν degrees of freedom and non-centrality parameter θ is has the following properties:*

1. *f_1/f_0 is C^∞ on $\mathbb{R}$;*

2. *For any $k \in \mathbb{N}^*$, we have $(f_1/f_0)^{(k)}(t) \to 0$ as $|t| \to +\infty$;*

3. *$(f_1/f_0)^{(2)}(0) \neq 0$.*

Proof [Proof of Proposition 25]

1. By (24), the function series in $(f_1/f_0)'$ has the same form as f_1/f_0; therefore, the result easily follows by induction.

2. By (24), Leibniz formula entails that the successive derivatives of f_1/f_0 are linear combinations of products of function series of the same form as f_1/f_0 by derivatives of $t \mapsto (\nu+t^2)^{-3/2}$. The result follows by the dominated convergence theorem, as all the derivatives of $t \mapsto (\nu+t^2)^{-3/2}$ tend to 0 as $|t| \to +\infty$;

3. The result follows by differentiating (17) at 0.

■

Proposition 26 (Concavity in the two-sided Student model) *The distribution function G_1 of two-sided p-values in the Student model satisfies (2).*

Proof [Proof of Proposition 26] By Lemma 2, we need to prove that the likelihood ratio f_1/f_0 of the Student model is such that $t \mapsto (f_1/f_0)(t) + (f_1/f_0)(-t)$ is non-decreasing. Equation (17) yields for $t \in \mathbb{R}$

$$\left(\frac{f_1}{f_0}\right)'(t) + \left(\frac{f_1}{f_0}\right)'(-t) = \nu(\nu+t^2)^{-3/2} \sum_{j=0}^{+\infty} a_j^1(\nu,\theta) \left(\psi_{(j,\nu)}(t) - \psi_{(j,\nu)}(-t)\right), \quad (18)$$

with $\psi_{(j,\nu)}(t) - \psi_{(j,\nu)}(-t) = (1 - (-1)^j)(t/\sqrt{\nu+t^2})^{-j}$. Therefore, as $a_j^1(\nu,\theta) > 0$, (18) yields $(f_1/f_0)'(t) + (f_1/f_0)'(-t) \geq 0$, which concludes the proof. ■

Appendix B. Asymptotics of a Modification of the Storey-λ Estimator

This section gathers the proofs of the asymptotic properties of the estimator $\hat{\pi}_{0,m}^{\text{Sto}}(1 - h_m)$ stated in Section 3.

Proof [Proof of Proposition 13]

1. We demonstrate that $\hat{\pi}_{0,m}^{\text{Sto}}(1 - h_m)$ 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 $1 - G(1 - h_m)$. Letting

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

we have $\sum_{i=1}^m Y_i^m = \sqrt{mh_m} \left(\hat{\pi}_{0,m}^{\text{Sto}}(1 - h_m) - \mathbb{E}\left[\hat{\pi}_{0,m}^{\text{Sto}}(1 - h_m)\right]\right)$. The $(Y_i^m)_{1 \leq i \leq m}$ are centered, independent random variables, with $\text{Var}\, Y_i^m = \text{Var}\, Z_i^m/(mh_m) = G(1-h_m)(1-G(1-h_m))/(mh_m)$, which is equivalent to $g(1)/m$ as $m \to +\infty$. Therefore,

$$\lim_{m \to +\infty} \sum_{i=1}^m \mathbb{E}\left[(Y_i^m)^2\right] = g(1).$$

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

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

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

$$\sum_{i=1}^{m} \mathbb{E}\left[(Y_i^m)^2 \mathbf{1}_{|Y_i^m|>\varepsilon}\right] \leq \frac{1}{h_m}\mathbb{E}\left[\mathbf{1}_{|Y_1^m|>\varepsilon}\right]$$

$$= \frac{1}{h_m}\mathbb{P}(|Y_1^m| > \varepsilon)$$

$$\leq \frac{1}{h_m}\frac{\operatorname{Var} Y_1^m}{\varepsilon^2}$$

by Chebycheff's inequality. As $mh_m \to +\infty$ and $\operatorname{Var} Y_1^m \sim g(1)/m$ as $m \to +\infty$, the above sum therefore goes to 0 as $mh_m \to +\infty$. The Lindeberg-Feller conditions for the Central Limit Theorem are thus fulfilled, and we have

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

which concludes the proof.

2. As $G(\lambda) = \pi_0\lambda + (1-\pi_0)G_1(\lambda)$, we have, for any $\lambda < 1$,

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

Therefore, the bias is given by

$$\mathbb{E}\left[\hat{\pi}_{0,m}^{\mathrm{Sto}}(\lambda)\right] - \pi_0 = (1-\pi_0)\frac{1-G_1(\lambda)}{1-\lambda}.$$

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

$$1 - G_1(\lambda) = \sum_{l=0}^{k} \frac{(-1)^l g_1^{(l)}(1)}{(l+1)!}(1-\lambda)^{l+1} + \mathrm{o}\left((1-\lambda)^{l+1}\right)$$

$$= (1-\lambda)g_1(1) + \frac{(-1)^k g_1^{(k)}(1)}{(k+1)!}(1-\lambda)^{k+1} + \mathrm{o}\left((1-\lambda)^{k+1}\right)$$

as $g_1^{(l)}(1) = (1-\pi_0)^{-1}g^{(l)}(1) = 0$ for $1 \leq l < k$. Therefore, if $h_m \to 0$ as $m \to +\infty$, we have

$$\mathbb{E}\left[\hat{\pi}_{0,m}^{\mathrm{Sto}}(1-h_m)\right] - g(1) = (1-\pi_0)\frac{(-1)^k g_1^{(k)}(1)}{(k+1)!}h_m^k + \mathrm{o}\left(h_m^k\right),$$

which concludes the proof, as $(1-\pi_0)g_1^{(k)}(1) = g^{(k)}(1)$. ∎

Proof [Proof of Proposition 14] By Proposition 13, the asymptotic variance of $\hat{\pi}_{0,m}(1-h_m)$ is equivalent to $g(1)/(mh_m)$, and the bias is of order h_m^k. The optimal bandwidth is obtained for h_m proportional to $m^{-1/(2k+1)}$, because this choice balances variance and squared bias. The proportionality constant is an explicit function of k, π_0, $g_1(1)$, and $g_1^{(k)}(1)$. By definition, the MSE that corresponds

to this optimal choice is twice the corresponding squared bias, that is, of order $m^{-2k/(2k+1)}$, which completes the proof of (1). To prove (2), we note that

$$\sqrt{mh_m}\left(\hat{\pi}_{0,m}-g(1)\right) = \sqrt{mh_m}\left(\hat{\pi}_{0,m}-\mathbb{E}\left[\hat{\pi}_{0,m}\right]\right) + \sqrt{mh_m}\left(\mathbb{E}\left[\hat{\pi}_{0,m}\right]-g(1)\right),$$

where $\hat{\pi}_{0,m}$ denotes $\hat{\pi}_{0,m}(1-h_m)$ to alleviate notation. The first term (variance) converges in distribution to $\mathcal{N}(0,g(1))$ by Proposition 13 (1) as soon as $\sqrt{mh_m} \to +\infty$. The second term (bias) is of the order of $\sqrt{mh_m}h_m^k = \sqrt{mh_m^{2k+1}}$ by Proposition 13 (2). Taking $h_m(k) = h_m^\star(k)\eta_m^2$, where $\eta_m \to 0$, we have $mh_m^{2k+1} \to 0$, which ensures that the bias term converges in probability to 0. ∎

Appendix C. Extension of Neuvial (2008) to the Unconditional Setting

In this section, we show that the results obtained by Neuvial (2008) in the original (conditional) setting of Benjamini and Hochberg (1995) also hold in the unconditional setting considered here, at the price of an additional term in the asymptotic variance due to the fluctuations of the random variable $\pi_{0,m}$. We start by stating a lemma which provides a lower bound on the critical value of plug-in procedures. It is is a consequence of Proposition 10(1).

Lemma 27 *Let α_m be a sequence of (possibly data-dependent) levels that converges in probability to $\alpha_\infty \in (0,1)$ as $m \to +\infty$. If $\alpha_\infty < \alpha_{BH}^\star$, then the threshold $\widehat{\tau}_m(\alpha_m)$ of the $BH(\alpha_m)$ procedure converges in probability to 0 as $m \to +\infty$. If the convergence of α_m to α_∞ holds almost surely, then the convergence of $\widehat{\tau}_m(\alpha_m)$ to 0 holds almost surely as well.*

Proof [Proof of Lemma 27] Assume that α_m converges to α_∞ in probability, with $\alpha_\infty < \alpha_{BH}^\star$. Let $\varepsilon > 0$, we are going to show that there exists an integer $N > 0$ such that for a large enough m, the number of rejections of the $BH(\alpha_m)$ procedure is less than N with probability greater than $1-\varepsilon$. Let $\bar{\alpha} = (\alpha_\infty + \alpha_{BH}^\star)/2$. As $\alpha_m \xrightarrow{P} \alpha_\infty < \bar{\alpha}$, there exists an integer M such that for any $m \geq M$, $\alpha_m \leq \bar{\alpha}$ with probability greater than $1-\varepsilon/2$. As $\bar{\alpha} < \alpha_{BH}^\star$, Proposition 10(1) entails that the number of rejections by the $BH(\bar{\alpha})$ procedure is bounded in probability as $m \to +\infty$; that is, there exist two integers N and M' such that for $m \geq M'$, the number of rejections of the $BH(\bar{\alpha})$ procedure is less than N with probability greater that $1-\varepsilon/2$. Thus, for any $m \geq \max(M,M')$, the number of rejections of the $BH(\alpha_m)$ procedure is less than N with probability greater that $1-\varepsilon$. The proof for the almost sure convergence in the case when α_m converges to α_∞ almost surely is similar. ∎

We follow the proof technique introduced by Neuvial (2008), by writing the empirical threshold of a given FDR controlling procedure (and its associated FDP) as the result of the application of a *threshold function* of the empirical distribution of the observed p-values. As the regularity of the threshold functions involved has already been established by Neuvial (2008), the result is a consequence of the fact that the p-value distributions under the null and the alternative hypotheses (as defined below) satisfy Donsker's theorem in the current unconditional setting. This Donsker's theorem has been established by Genovese and Wasserman (2004, Theorem 4.1). For $a \in \{0,1\}$ and $t \in [0,1]$, we let $\widehat{\Gamma}_{a,m}(t) = m^{-1}\sum_{i=1}^m \mathbf{1}_{H_a \text{ true and } P_i \leq t}$.

Proposition 28 (Genovese and Wasserman 2004) *As $m \to +\infty$, we have:*

1.

$$\sqrt{m}\left(\begin{pmatrix}\widehat{\Gamma}_{0,m}(t)\\\widehat{\Gamma}_{1,m}(t)\end{pmatrix}-\begin{pmatrix}\pi_0 t\\(1-\pi_0)g_1(t)\end{pmatrix}\right)\rightsquigarrow\begin{pmatrix}\mathbb{W}_0\\\mathbb{W}_1\end{pmatrix},$$

where $(\mathbb{W}_0,\mathbb{W}_1)$ *is a two-dimensional, centered Gaussian process with covariance function* $\gamma(s,t)$ *defined for any* $(s,t)\in[0,1]^2$ *by*

$$\gamma(s,t)=\begin{pmatrix}\pi_0 s\wedge t-\pi_0^2 st & -\pi_0 s(1-\pi_0)G_1(t)\\-\pi_0 t(1-\pi_0)G_1(s) & (1-\pi_0)G_1(s\wedge t)-(1-\pi_0)^2 G_1(s)G_1(t)\end{pmatrix};$$

2.

$$\sqrt{m}\left(\widehat{\mathbb{G}}_m-G\right)\rightsquigarrow\mathbb{W},$$

where $\mathbb{W}\overset{(d)}{=}\mathbb{W}_0+\mathbb{W}_1$ *is a one-dimensional, centered Gaussian process with covariance function* $(s,t)\mapsto G(s\wedge t)-G(s)G(t)$.

Note that $\widehat{\Gamma}_{0,m}=\pi_{0,m}\widehat{\mathbb{G}}_{0,m}$ and $\widehat{\Gamma}_{1,m}=(1-\pi_{0,m})\widehat{\mathbb{G}}_{1,m}$, where $(\widehat{\mathbb{G}}_{0,m},\widehat{\mathbb{G}}_{1,m})$ are the empirical distribution functions of the *p*-values under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. The results of Neuvial (2008) have been obtained by directly considering the convergence of the process $(\widehat{\mathbb{G}}_{0,m},\widehat{\mathbb{G}}_{1,m})$ instead of $(\widehat{\Gamma}_{0,m},\widehat{\Gamma}_{1,m})$, because $\pi_{0,m}$ was deterministic in the conditional setting (see Neuvial 2009, Theorem 3.1). The results established in Neuvial (2008) (in particular Theorem 3.2) can be translated to the unconditional setting just by replacing the processes $\pi_0\mathbb{Z}_0$ and $\pi_1\mathbb{Z}_1$ in Neuvial (2008) by the processes $\mathbb{W}_0$ and $\mathbb{W}_1$ defined in Proposition 28, and consequently, the process $\mathbb{Z}=\pi_0\mathbb{Z}_0+\pi_1\mathbb{Z}_1$ by $\mathbb{W}=\mathbb{W}_0+\mathbb{W}_1$.

Therefore, the asymptotic properties of the BH procedure and Storey's procedure (that is, $\mathrm{BH}(\cdot/\hat{\pi}_{0,m}^{\mathrm{Sto}}(\lambda))$ in the unconditional setting can be obtained by adapting the proof of the corresponding theorems (Theorems 4.2 and 4.15) in Neuvial (2008):

Corollary 29 (BH procedure, unconditional setting) *For any* $\alpha\geq\alpha_{BH}^{\star}$, *we have*

1. The asymptotic distribution of the threshold $\widehat{\tau}_m(\alpha)$ *is given by*

$$\sqrt{m}\left(\widehat{\tau}_m(\alpha)-\tau_\infty(\alpha)\right)\rightsquigarrow\mathcal{N}\left(0,\frac{G(\tau_\infty(\alpha))(1-G(\tau_\infty(\alpha)))}{(1/\alpha-g(\tau_\infty(\alpha)))^2}\right);$$

2. The asymptotic distribution of the associated FDPs is given by

$$\sqrt{m}\left(\mathrm{FDP}_m(\widehat{\tau}_m(\alpha))-\pi_0\alpha\right)\rightsquigarrow\mathcal{N}\left(0,(\pi_0\alpha)^2\left(\frac{1}{\pi_0\tau_\infty(\alpha)}-1\right)\right).$$

The asymptotic properties of the BH Oracle procedure are simply obtained by applying Corollary 29 at level α/π_0.

Corollary 30 (Storey-λ procedure, unconditional model) *For any* $\lambda\in[0,1)$, *and* $\alpha\in[0,1]$, *let* $\widehat{\tau}_m^{0,\lambda}(\alpha)=\mathcal{T}^{\mathrm{Sto}(\lambda)}(\widehat{\mathbb{G}}_m)$ *be the empirical threshold* $\widehat{\tau}_m^{0,\lambda}(\alpha)$ *of Storey's procedure at level* α, *and* $\tau_\infty^{0,\lambda}(\alpha)=\mathcal{T}^{\mathrm{Sto}(\lambda)}(G)$ *be the corresponding asymptotic threshold. Then,*

1. $\alpha_{\mathrm{Sto}(\lambda)}^{\star}=\overline{\pi}_0(\lambda)\alpha_{BH}^{\star}$ *is the critical value of Storey's procedure;*

2. *For any $\alpha > \alpha^\star_{\text{Sto}(\lambda)}$:*

   (a) *The asymptotic distribution of the threshold $\widehat{\tau}^{0,\lambda}_m(\alpha)$ is given by*

   $$\sqrt{m}\left(\widehat{\tau}^{0,\lambda}_m(\alpha) - \tau^{0,\lambda}_\infty(\alpha)\right) \leadsto \frac{\tau^{0,\lambda}_\infty(\alpha)}{\overline{\pi_0}(\lambda)/\alpha - g(\tau^{0,\lambda}_\infty(\alpha))}\left\{\frac{\mathbb{W}(\tau^{0,\lambda}_\infty(\alpha))}{\tau^{0,\lambda}_\infty(\alpha)} + \frac{1}{\alpha}\frac{\mathbb{W}(\lambda)}{1-\lambda}\right\},$$

   where $\mathbb{W}$ is a centered Gaussian process with covariance function $(s,t) \mapsto G(s \wedge t) - G(s)G(t)$;

   (b) *The asymptotic distribution of the associated FDPs is given by*

   $$\sqrt{m}\left(\text{FDP}_m(\widehat{\tau}^{0,\lambda}_m(\alpha)) - \pi_0\alpha/\overline{\pi_0}(\lambda)\right) \leadsto \mathcal{N}\left(0, \sigma^2_\lambda\right),$$

   where

   $$\sigma^2_\lambda = \left(\frac{\pi_0\alpha}{\overline{\pi_0}(\lambda)}\right)^2\left\{\frac{1}{\pi_0\tau^{0,\lambda}_\infty(\alpha)} + 2\frac{\tau^{0,\lambda}_\infty(\alpha) \wedge \lambda}{\tau^{0,\lambda}_\infty(\alpha)(1 - G(\lambda))} - \frac{1}{1 - G(\lambda)}\right\}.$$

Note that Corollary 30 with $\lambda = 0$ recovers Corollary 29.

Appendix D. Asymptotic Properties of Plug-in Procedures

This section gathers the proofs of Section 4.

D.1 Proof of Theorem 16

We denote by $\widehat{\rho}^0_m(\alpha)$ the proportion of rejections, and by $\widehat{v}^0_m(\alpha)$ the proportion of incorrect rejections by the plug-in procedure $\text{BH}(\alpha/\widehat{\pi}_{0,m})$ (among all m hypotheses tested). They may be written as $\widehat{\rho}^0_m(\alpha) = \widehat{\mathbb{G}}_m(\widehat{\tau}^0_m(\alpha)) = \widehat{\tau}^0_m(\alpha)\widehat{\pi}_{0,m}/\alpha$ and $\widehat{v}^0_m(\alpha) = \pi_{0,m}\widehat{\mathbb{G}}_{0,m}(\widehat{\tau}^0_m(\alpha))$, respectively. The following Lemma shows that the convergence rate of $(\widehat{\tau}^0_m(\alpha), \widehat{v}^0_m(\alpha), \widehat{\rho}^0_m(\alpha))$ for a large enough α is driven by the convergence rate of $\widehat{\pi}_{0,m}$. In order to alleviate notation, we omit the "(α)" in $\widehat{\tau}^0_m, \widehat{\rho}^0_m, \widehat{v}^0_m, \tau^0_\infty, \rho^0_\infty$, v^0_∞ in the remainder of this section. Moreover, $\text{FDP}_m(\widehat{\tau}^0_m(\alpha))$ will simply be denoted by $\widehat{\text{FDP}}^0_m$.

Lemma 31 *Let $\widehat{\pi}_{0,m}$ be an estimator of π_0 such that $\widehat{\pi}_{0,m} \to \pi_{0,\infty}$ in probability as $m \to +\infty$. Define $\alpha^\star_0 = \pi_{0,\infty}\alpha^\star_{BH}$, and let $\alpha > \alpha^\star_0$. Then, under Assumption (2), we have, as $m \to +\infty$:*

1. *$\widehat{\tau}^0_m$ converges in probability to τ^0_∞ as $m \to +\infty$, with $g(\tau^0_\infty) < \pi_{0,\infty}/\alpha$. If the convergence of $\widehat{\pi}_{0,m}$ to $\pi_{0,\infty}$ holds almost surely, then that of $\widehat{\tau}^0_m$ to τ^0_∞ holds almost surely as well;*

2. *Further assume that $\sqrt{mh_m}(\widehat{\pi}_{0,m} - \pi_{0,\infty})$ converges in distribution for some h_m such that $h_m = o(1/\ln\ln m)$ and $mh_m \to +\infty$ as $m \to +\infty$. Then $(\widehat{\tau}^0_m, \widehat{v}^0_m, \widehat{\rho}^0_m)$ converges at in distribution at rate $1/\sqrt{mh_m}$, with*

   $$\begin{pmatrix} \widehat{\tau}^0_m \\ \widehat{v}^0_m \\ \widehat{\rho}^0_m \end{pmatrix} - \begin{pmatrix} \tau^0_\infty \\ v^0_\infty \\ \rho^0_\infty \end{pmatrix} = \frac{\tau^0_\infty/\alpha}{\pi_{0,\infty}/\alpha - g(\tau^0_\infty)}\begin{pmatrix} 1 \\ \pi_0 \\ g(\tau^0_\infty) \end{pmatrix}(\pi_{0,\infty} - \widehat{\pi}_{0,m})(1 + o_P(1)),$$

   where $v^0_\infty = \pi_0\tau^0_\infty$ and $\rho^0_\infty = G(\tau^0_\infty) = \pi_{0,\infty}\tau^0_\infty/\alpha$.

1453

Proof [Proof of Lemma 31] For 1., we assume that the convergence of $\hat{\pi}_{0,m}$ to $\pi_{0,\infty}$ holds in probability. If it also holds almost surely, then the convergence of $\hat{\tau}_m^0$ to τ_∞^0 is almost sure as well. The sketch of the proof is inspired by van der Vaart (1998, Lemma 21.3). Let $\psi_{F,\zeta} : t \mapsto t/\zeta - F(t)$ for any distribution function F and any $\zeta \in (0,1]$. As $\widehat{\mathbb{G}}_m(\hat{\tau}_m^0) = \hat{\pi}_{0,m}\hat{\tau}_m^0/\alpha$ and $G(\tau_\infty^0) = \pi_{0,\infty}\tau_\infty^0/\alpha$, we have $\psi_{G,\alpha/\pi_{0,\infty}}(\tau_\infty^0) = 0$ and $\psi_{\widehat{\mathbb{G}}_m,\alpha/\hat{\pi}_{0,m}}(\hat{\tau}_m^0) = 0$. The proof relies on the following property:

(a) $\psi_{G,\alpha/\pi_{0,\infty}}(\hat{\tau}_m^0)$ converges in probability to $0 = \psi_{G,\alpha/\pi_{0,\infty}}(\tau_\infty^0)$;

(b) $\psi_{G,\alpha/\pi_{0,\infty}}$ is locally invertible in a neighborhood of τ_∞^0, with $\dot{\psi}_{G,\alpha/\pi_{0,\infty}}(\tau_\infty^0) > 0$.

To prove (a), we note that

$$
\begin{aligned}
-\psi_{G,\alpha/\pi_{0,\infty}}(\hat{\tau}_m^0) &= G(\hat{\tau}_m^0) - \pi_{0,\infty}\hat{\tau}_m^0/\alpha \\
&= (G - \widehat{\mathbb{G}}_m)(\hat{\tau}_m^0) + (\widehat{\mathbb{G}}_m(\hat{\tau}_m^0) - \hat{\pi}_{0,m}\hat{\tau}_m^0/\alpha) + (\hat{\pi}_{0,m} - \pi_{0,\infty})\hat{\tau}_m^0/\alpha .
\end{aligned}
$$

The first term converges to 0 almost surely, the second one is identically null, and the third one converges in probability to 0 as $\hat{\pi}_{0,m}$ converges in probability to $\pi_{0,\infty}$, and $\hat{\tau}_m^0 \in [0,1]$. Item (b) holds as G in concave (by Assumption (2)) and $\alpha/\pi_{0,\infty} > \alpha_{BH}^\star$, where $\alpha_{BH}^\star = \lim_{u \to 0} u/G(u)$ is the critical value of the BH procedure (see Neuvial 2008, Lemma 7.6 page 1097 for a proof of the invertibility).

Combining (a) and (b), $\hat{\tau}_m^0$ converges in probability to τ_∞^0, and $\dot{\psi}_{G,\alpha/\pi_{0,\infty}}(\tau_\infty^0) = \pi_{0,\infty}/\alpha - g(\tau_\infty^0)$ is positive, which proves the first item. For the second item, we only give the proof for $\hat{\tau}_m^0$, as the proofs for $\hat{v}_m^0$ and $\hat{\rho}_m^0$ are similar. The idea of the proof is that the fluctuations of $\bar{\mathbb{G}}_m = \widehat{\mathbb{G}}_m - G$, the centered empirical process associated with G, are of order $1/\sqrt{m}$ by Donsker's theorem (Donsker, 1951); thus, these fluctuations are negligible with respect to the fluctuations of $\hat{\pi}_{0,m} - \pi_{0,\infty}$, which are assumed to be of order $1/\sqrt{mh_m}$ with $h_m \to 0$. We have

$$
\begin{aligned}
G(\hat{\tau}_m^0) - G(\tau_\infty^0) &= (G(\hat{\tau}_m^0) - \widehat{\mathbb{G}}_m(\hat{\tau}_m^0)) + (\widehat{\mathbb{G}}_m(\hat{\tau}_m^0) - G(\tau_\infty^0)) \\
&= -\bar{\mathbb{G}}_m(\hat{\tau}_m^0) + (\hat{\pi}_{0,m}\hat{\tau}_m^0/\alpha - \pi_{0,\infty}\tau_\infty^0/\alpha)
\end{aligned}
$$

because $\widehat{\mathbb{G}}_m(\hat{\tau}_m^0) = \hat{\pi}_{0,m}\hat{\tau}_m^0/\alpha$ and $G(\tau_\infty^0) = \pi_{0,\infty}\tau_\infty^0/\alpha$. Therefore,

$$
G(\hat{\tau}_m^0) - G(\tau_\infty^0) = -\bar{\mathbb{G}}_m(\hat{\tau}_m^0) + \frac{\hat{\pi}_{0,m}}{\alpha}(\hat{\tau}_m^0 - \tau_\infty^0) + \frac{\hat{\pi}_{0,m} - \pi_{0,\infty}}{\alpha}\tau_\infty^0 .
$$

As $\hat{\tau}_m^0 \xrightarrow{P} \tau_\infty^0$ as $m \to +\infty$, we also have $G(\hat{\tau}_m^0) - G(\tau_\infty^0) = (\hat{\tau}_m^0 - \tau_\infty^0)(g(\tau_\infty^0) + o_P(1))$ by Taylor's formula. Hence we have

$$
\left(g(\tau_\infty^0) - \hat{\pi}_{0,m}/\alpha + o_P(1)\right)(\hat{\tau}_m^0 - \tau_\infty^0) = -\bar{\mathbb{G}}_m(\hat{\tau}_m^0) + (\hat{\pi}_{0,m} - \pi_{0,\infty})\tau_\infty^0/\alpha .
$$

Now because $\hat{\pi}_{0,m}$ converges in probability to $\pi_{0,\infty}$, we have

$$
g(\tau_\infty^0) - \hat{\pi}_{0,m}/\alpha = (g(\tau_\infty^0) - \pi_{0,\infty}/\alpha)(1 + o_P(1)).
$$

By 1, we have $\pi_{0,\infty}/\alpha > g(\tau_\infty^0)$, so that for sufficiently large m:

$$
\hat{\tau}_m^0 - \tau_\infty^0 = \frac{\bar{\mathbb{G}}_m(\hat{\tau}_m^0)}{g(\tau_\infty^0) - \pi_{0,\infty}/\alpha}(1 + o_P(1)) + \frac{\tau_\infty^0/\alpha}{g(\tau_\infty^0) - \pi_{0,\infty}/\alpha}(\hat{\pi}_{0,m} - \pi_{0,\infty}).
$$

Finally, we note that as $\|\bar{\mathbb{G}}_m\|_\infty \sim c\sqrt{\ln\ln m/m}$ (by the Law of the Iterated Logarithm) and $h_m = o\left(1/\ln\ln m\right)$, we have $\bar{\mathbb{G}}_m(\hat{\tau}_m^0) = o_P\left(1/\sqrt{mh_m}\right)$. Moreover, $\sqrt{mh_m}\left(\hat{\pi}_{0,m} - \pi_{0,\infty}\right)$ converges in distribution, so that the term $(\hat{\pi}_{0,m} - \pi_{0,\infty})\tau_\infty^0/\alpha$ dominates the right-hand side. Finally, we have

$$\hat{\tau}_m^0 - \tau_\infty^0 = \frac{\tau_\infty^0/\alpha}{g(\tau_\infty^0) - \pi_{0,\infty}/\alpha}(\hat{\pi}_{0,m} - \pi_{0,\infty})(1 + o_P(1)),$$

which concludes the proof for $\hat{\tau}_m^0$. ∎

Proof [Proof of Theorem 16] 1. is a consequence of Lemma 27 combined with Lemma 31(1); 2.(a) is a consequence of Lemma 27(2). Let us prove 2.(b). By Lemma 31, we have

$$\sqrt{mh_m}\left(\begin{pmatrix}\hat{\nu}_m^0\\\hat{\rho}_m^0\end{pmatrix} - \begin{pmatrix}\nu_\infty^0\\\rho_\infty^0\end{pmatrix}\right) \rightsquigarrow \xi_\infty\begin{pmatrix}\pi_0\\g(\tau_\infty^0)\end{pmatrix}X,$$

where $X \sim \mathcal{N}(0, s_0^2)$ and

$$\xi_\infty = \frac{\tau_\infty^0/\alpha}{\pi_{0,\infty}/\alpha - g(\tau_\infty^0)}.$$

Recall that $\widehat{FDP}_m^0 = \hat{\nu}_m^0/(\hat{\rho}_m^0 \vee m^{-1})$. We begin by noting that for a large enough m, we have $\hat{\rho}_m^0 > 1/m$ almost surely. This is a consequence of the fact that (i) $\hat{\rho}_m^0 = \hat{\mathbb{G}}_m(\hat{\tau}_m^0) = \hat{\pi}_{0,m}\hat{\tau}_m^0/\alpha$, with $\hat{\tau}_m^0$ bounded away from 0 (by 1.), and (ii) $\hat{\pi}_{0,m}$ converges to $\pi_{0,\infty} \geq \pi_0 > \alpha$. As a consequence, the factor m^{-1} may be omitted in $\widehat{FDP}_m^0$ for a large enough m; the FDP may then be written as $\widehat{FDP}_m^0 = \gamma(\hat{\nu}_m^0, \hat{\rho}_m^0)$, where $\gamma : (u,v) \mapsto u/v$ for any $u \geq 0$ and $v > 0$. γ is differentiable for any such (u,v), with derivative $\dot{\gamma}_{u,v} = (1/v, -u/v^2) = 1/v(1, -u/v)$. In particular, recalling that $\nu_\infty^0 = \pi_0\tau_\infty^0$ and $\rho_\infty^0 = G(\tau_\infty^0) = \pi_{0,\infty}\tau_\infty^0/\alpha$, we have

$$\dot{\gamma}_{\nu_\infty^0, \rho_\infty^0} = \frac{\alpha}{\tau_\infty^0\pi_{0,\infty}}\left(1, -\frac{\pi_0\alpha}{\pi_{0,\infty}}\right). \tag{19}$$

As $\gamma(\nu_\infty^0, \rho_\infty^0) = \pi_0\alpha/\pi_{0,\infty}$, the Delta method yields

$$\sqrt{mh_m}\left(\widehat{FDP}_m^0 - \frac{\pi_0\alpha}{\pi_{0,\infty}}\right) \rightsquigarrow \mathcal{N}\left(0, w^2\right),$$

with $w = s_0\xi_\infty\,\dot{\gamma}_{\nu_\infty^0, \rho_\infty^0}\begin{pmatrix}\pi_0\\g(\tau_\infty^0)\end{pmatrix}$.

By (19), we have $\dot{\gamma}_{\nu_\infty^0, \rho_\infty^0}\begin{pmatrix}\pi_0\\g(\tau_\infty^0)\end{pmatrix} = \frac{\alpha^2\pi_0}{\tau_\infty^0\pi_{0,\infty}^2}(\pi_{0,\infty}/\alpha - g(\tau_\infty^0))$, which implies that $w = s_0\pi_0\alpha/\pi_{0,\infty}^2$. ∎

D.2 Consistency, Purity and Criticality

Proof [Proof of Lemma 20] We note that

$$
\begin{aligned}
\frac{f_1(x)}{f_0(x)} &= \frac{f_0(x-\theta)}{f_0(x)} \text{ by definition of a location model} \\
&= \frac{f_0(-x+\theta)}{f_0(-x)} \text{ by Assumption (1)} \\
&= \frac{f_0(-x+\theta)}{f_1(-x+\theta)},
\end{aligned}
$$

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

Proof [Proof of Proposition 21] We have $\alpha_{BH}^\star = \lim_{t\to 0} 1/g(t)$, where $g = \pi_0 + (1-\pi_0)g_1$ and

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

Therefore, as $\lim_{t\to 0} F_0^{-1}(t) = +\infty$, the result is a consequence of Lemma 20. ∎

D.3 Regularity of g_1 for Two-Sided Tests in Symmetric Models

Proof [Proof of Lemma 19]

1. We make the additional assumption that there exists $\eta > 0$ such that f_1/f_0 is differentiable on $V_\eta = [-\eta, \eta] \setminus \{0\}$, and that its derivative tends to ℓ_- as $u \to 0^-$ and ℓ_+ as $u \to 0^+$. This assumption makes the proof simpler, and it holds in the models considered in this paper. However, the result still holds (and is simpler to state) without this extra assumption. By Proposition 1, we have under Assumption (1)

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

where $q_0(t/2) = F_0^{-1}(1 - t/2)$ maps $Q_\eta = [2(1 - F_0(\eta)), 1)$ onto $(0, \eta]$. Therefore, g_1 is differentiable on Q_η and satisfies, for any t in Q_η:

$$
\begin{aligned}
g_1^{(1)}(t) &= \frac{1}{2}\left\{\left(\frac{f_1}{f_0}\right)'(q_0(t/2)) - \left(\frac{f_1}{f_0}\right)'(-q_0(t/2))\right\} \times \frac{1}{2}q_0'(t/2) \\
&= -\frac{1}{4f_0(q_0(t/2))}\left(\left(\frac{f_1}{f_0}\right)'(q_0(t/2)) - \left(\frac{f_1}{f_0}\right)'(-q_0(t/2))\right). \qquad (20)
\end{aligned}
$$

As $t \to 1$, $q_0(t/2) \to 0^+$, (20) implies that g_1 is differentiable at 1 with derivative given by (14).

2. Similarly, we prove the result with the extra assumption that f_1/f_0 is twice differentiable in a neighborhood of 0. Then (20) entails that g_1 is itself twice differentiable in a neighborhood

of 1. Writing $g_1^{(1)}(t) = a(t)b(t)$, with

$$\begin{cases} a(t) & = 1/(4f_0(q_0(t/2))) \\ b(t) & = -(f_1/f_0)'(q_0(t/2)) + (f_1/f_0)'(-q_0(t/2)) \end{cases},$$

we have $g_1^{(2)}(t) = a'(t)b(t) + a(t)b'(t)$. As $q_0(1/2) = F_0^{-1}(1/2) = 0$, we have $b(1) = 0$, so that $g_1^{(2)}(1) = a(1)b'(1)$, where $a(1) = 1/(4f_0(0))$ and

$$b'(t) = \frac{1}{2f_0(q_0(t/2))} \left(\left(\frac{f_1}{f_0}\right)^{(2)}(q_0(t/2)) + \left(\frac{f_1}{f_0}\right)^{(2)}(-q_0(t/2)) \right).$$

Thus $b'(1) = 1/(2f_0(0)) \times 2(f_1/f_0)^{(2)}(0)$, which concludes the proof.

$\blacksquare$

References

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 57(1):289–300, 1995.

Y. Benjamini and Y. 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.

Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188, 2001.

Y. Benjamini, A. M. Krieger, and D. Yekutieli. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507, 2006.

G. Blanchard and E. Roquain. Adaptive FDR control under independence and dependence. *Journal of Machine Learning Research*, 10:2837–2871, 2009.

Z. Chi. On the performance of FDR control: constraints and a partial solution. *The Annals of Statistics*, 35(4):1409–1431, 2007a.

Z. Chi. Sample size and positive false discovery rate control for multiple testing. *Electronic Journal of Statistics*, 1:77–118, 2007b.

Z. Chi and Z. Tan. Positive false discovery proportions: intrinsic bounds and adaptive control. *Statistica Sinica*, 18(3):837–860, 2008.

S. Delattre and E. Roquain. On the false discovery proportion convergence under Gaussian equi-correlation. *Statistics & Probability Letters*, 81(1):111–115, January 2011.

S. Delattre and E. Roquain. Asymptotics of empirical distribution function for Gaussian subordinated arrays with an application to multiple testing. HAL preprint, 2012. URL `http://hal.archives-ouvertes.fr/hal-00739749`.

M. D. Donsker. An invariance principle for certain probability limit theorems. *Memoirs of the American Mathematical Society*, 6:12, 1951.

B. Efron, R. J. Tibshirani, J. D. Storey, and V. G. Tusher. Empirical bayes analysis of a microarray experiment. *Journal of the American Statistical Association*, 96(456):1151–1160, December 2001.

G. Eklund. Massignifikansproblemet. Unpublished seminar papers, Uppsala University Institute of Statistics, 1961–1963.

H. Finner, T. Dickhaus, and M. Roters. On the false discovery rate and an asymptotically optimal rejection curve. *The Annals of Statistics*, 37(2):596–618, April 2009.

C. R. Genovese and L. Wasserman. A stochastic process approach to false discovery control. *The Annals of Statistics*, 32(3):1035–1061, 2004.

N. W. Hengartner and P. B. Stark. Finite-sample confidence envelopes for shape-restricted densities. *The Annals of Statistics*, 23(2):525–550, 1995.

J. 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.

N. L. Johnson and B. L. Welch. Applications of the non-central t-distribution. *Biometrika*, 31(3): 362–389, 1940.

P. Neuvial. Asymptotic properties of false discovery rate controlling procedures under independence. *Electronic Journal of Statistics*, 2:1065–1110, 2008.

P. Neuvial. Corrigendum to "Asymptotic properties of false discovery rate controlling procedures under independence". *Electronic Journal of Statistics*, 3:1083–1083, 2009.

V. H. Nguyen and C. Matias. On efficient estimators of the proportion of true null hypotheses in a multiple testing setup. Hal preprint, 2013. URL http://hal.archives-ouvertes.fr/hal-00647082.

D. B. Pollard. *Convergence of Stochastic Processes*. Springer-Verlag, 1984.

E. Roquain and F. Villers. Exact calculations for false discovery proportion with application to least favorable configurations. *The Annals of Statistics*, 39(1):584–612, 2011.

T. Schweder and E. Spjøtvoll. Plots of p-values to evaluate many tests simultaneously. *Biometrika*, 69(3):493–502, 1982.

P. Seeger. A note on a method for the analysis of significances en masse. *Technometrics*, pages 586–593, 1968.

J. D. Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):479–498, 2002.

J. D. Storey. The positive false discovery rate: A bayesian interpretation and the q-value. *The Annals of Statistics*, 31(6):2013–2035, 2003.

J. D. Storey and R. J. Tibshirani. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, 100(16):9440–5, August 2003.

J. D. Storey, J. E. Taylor, and D. O. Siegmund. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(1):187–205, 2004.

M. T. Subbotin. On the law of frequency of errors. *Matematicheskii Sbornik*, 31:296–301, 1923.

J. W. H. Swanepoel. The limiting behavior of a modified maximal symmetric 2s-spacing with applications. *The Annals of Statistics*, 27(1):24–35, 1999.

A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2009.

A. W. van der Vaart. *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.