

# Estimating the Proportion of False Null Hypotheses among a Large Number of Independently Tested Hypotheses

Nicolai Meinshausen\*

John Rice†

October 14, 2004

## Abstract

We consider the problem of estimating the number of false null hypotheses among a very large number of independently tested hypotheses, focusing on the situation in which the proportion of false null hypotheses is very small. We propose a family of methods for establishing lower  $100(1-\alpha)\%$  confidence bounds for this proportion, based on the empirical distribution of the p-values of the tests. Methods in this family are then compared in terms of ability to consistently estimate the proportion by letting  $\alpha \rightarrow 0$  as the number of hypothesis tests increases and the proportion decreases. This work is motivated by a signal detection problem occurring in astronomy.

## 1 Introduction

Suppose that a large number of null hypotheses are independently tested and it is expected that only a small fraction of them are false. In some situations, one would like to estimate this fraction. Each hypothesis test might be an attempt to detect an independent realization of a rare signal and the number of actual signals present is of interest.

An example, which motivated our work, is afforded by the Taiwanese-American Occultation Survey (TAOS), which we now briefly describe. TAOS will attempt to detect small objects in the Kuiper Belt, a region of the solar system beyond the orbit of Neptune. The Kuiper Belt contains an unknown number of objects (KBOs), most of which are believed to be so small that they do not reflect enough light back to earth to be directly observed. The purpose of the TAOS project is to estimate the number of these KBOs down to the typical size of cometary nuclei (a few km) by observing occultations. The idea of the occultation technique is simple to describe. One monitors the light from a collection of

---

\*Seminar für Statistik, ETH Zürich, Switzerland

†Department of Statistics, University of California, Berkeley, USA

stars that have angular sizes smaller than the expected angular sizes of comets. An occultation is manifested by detecting the partial or total reduction in the flux from one of the stars for a brief interval when an object in the Kuiper Belt passes between it and the observer. Four dedicated robotic telescopes will automatically monitor 2-3,000 stars every clear night for several years and their combined results will be used to test for an occultation of each star approximately every 0.20 seconds, yielding on the order of  $10^{11}$  tests per year. The number of occultations expected per year ranges from tens to a few thousands, depending on what model of the Kuiper Belt is used. Having conducted a large number of tests, it is then of interest to estimate the number of occultations, or the occultation rate, since this will provide information on the distribution of KBOs. Note that in this context we are not so much interested in which particular null hypotheses are false as in how many are. The TAOS project is further described in Liang et al. (2002) and Chen et al. (2003).

We will base our analysis on the distribution of the p-values of the hypothesis tests. Let  $\{G_\theta, \theta \in \Theta\}$  be some family of distributions, where  $\theta$  is possibly infinite-dimensional and  $G_0(t) = t$  with  $0 \in \Theta$  is the uniform distribution on  $[0, 1]$ . All p-values are assumed to be independently distributed according to

$$P_i \sim G_{\theta_i}, \quad i = 1, \dots, n.$$

If a null hypothesis is true, the distribution of its p-value is uniform on  $[0, 1]$  and  $P_i \sim G_0$ . We suppose that neither the family  $\{G_\theta(t), \theta \in \Theta\}$  nor the parameter vector  $(\theta_1, \dots, \theta_n)$  is known, except from the fact that  $G_0$  corresponds to the uniform distribution.

The proportion of the null hypothesis that are false (the fraction of occultations in the TAOS example) is denoted by  $\lambda$ ,

$$\lambda = n^{-1} \sum_{i=1}^n \mathbf{1}\{\theta_i \neq 0\}. \quad (1)$$

Our goal is to construct a lower bound  $\hat{\lambda}$  with the property

$$P(\hat{\lambda} \leq \lambda) \geq 1 - \alpha \quad (2)$$

for a specified confidence level  $1 - \alpha$ . Such a lower bound would allow one to assert, with a specified level of confidence, that the proportion of false null hypotheses is at least  $\hat{\lambda}$ . The global null hypothesis that there are no false null hypotheses can be tested at level  $\alpha$  by rejecting when  $\hat{\lambda} > 0$ .

Our construction is closely related to that in Meinshausen and Bühlmann (2004), which treats the case of possibly dependent tests, but with an observational structure that allows the use of permutation arguments that are not available in our case. Another estimate is examined in Nettleton and Hwang (2003), but it does not have a property like (2). Our methodology is related to that of controlling the false discovery rate (Benjamini and Hochberg, 1995);

Storey, 2002), but the goals are different—we are not so much interested in which particular hypotheses are false as in how many are. However we note that an estimate of the number of false null hypotheses can be usefully employed in adaptive control of FDR (Benjamini and Hochberg, 2000). In a modification of the original FDR method, Storey (2002), also estimated the proportion of false hypotheses. The empirical distribution of p-values was used in Schweder and Spjøtvoll (1982) to estimate the number of true null hypotheses; the methods used there are different than ours and do not provide explicit lower confidence bounds. The methods in this paper extend a proposal of Genovese and Wasserman (2004). We also relate our results to those of Donoho and Jin (2004).

## 2 Theory and Methodology

The estimate hinges on the definition of bounding functions and bounding sequences.

Let  $U$  be uniform on  $[0, 1]$ . For any real-valued function  $\delta(t)$  on  $[0, 1]$ , which is strictly positive on  $(0, 1)$ , define  $V_{n,\delta}$  as the supremum of the weighted empirical distribution

$$V_{n,\delta} := \sup_{t \in (0,1)} \frac{U_n(t) - t}{\delta(t)}.$$

**Definition 1 (Bounding sequence)** *Let  $\delta(t)$  be a real-valued function on  $[0, 1]$ , strictly positive on  $(0, 1)$ . A series  $\beta_{n,\alpha}$  is called a bounding sequence for a real-valued, positive function  $\delta(t)$  if, for a constant level  $\alpha$ ,*

- (a)  $n\beta_{n,\alpha}$  is monotonically increasing with  $n$ ,
- (b)  $P(V_{n,\delta} > \beta_{n,\alpha}) < \alpha$  for all  $n$ .

The definition of a bounding sequence depends neither on the unknown proportion of false null hypotheses nor on the unknown distribution  $G(t)$  of p-values under the alternative.

Denote the empirical distribution of p-values by  $F_n(t)$ ,

$$F_n(t) := \sum_{i=1}^n 1\{P_i \leq t\}.$$

An estimate of the proportion  $\lambda$  of false null hypothesis can make use of the fact that the contribution of true null hypotheses to the difference between  $F_n(t)$  and  $t$  is essentially bounded by  $\beta_{n,\alpha}\delta(t)$ , simultaneously for all  $t \in (0, 1)$ .

**Definition 2** *Let  $\beta_{n,\alpha}$  be a bounding sequence for  $\delta(t)$  at level  $\alpha$ . An estimate for the proportion  $\lambda$  of false null hypotheses is given by*

$$\hat{\lambda} = \sup_{t \in (0,1)} \frac{F_n(t) - t - \beta_{n,\alpha}\delta(t)}{1 - t}. \quad (3)$$

This estimate is indeed a lower bound for  $\lambda$ , as shown in the following Theorem.

**Theorem 1** *Let  $\beta_{n,\alpha}$  be a bounding sequence for  $\delta(t)$  at level  $\alpha$  and  $\hat{\lambda}$  defined by (3). Then,*

$$P(\hat{\lambda} \leq \lambda) \geq 1 - \alpha. \quad (4)$$

PROOF OF THEOREM 1.

The distribution of p-values  $F_n$  is bounded by  $F_n(t) \leq \lambda + (1 - \lambda)U_{n_0}(t)$ , where  $n_0 = (1 - \lambda)n$  and  $U_{n_0}(t)$  is the empirical distribution of  $n_0$  independent Uniform(0,1)-distributed random variables. Thus

$$P(\hat{\lambda} > \lambda) \leq P\left(\sup_{t \in (0,1)} \frac{\lambda + (1 - \lambda)U_{n_0}(t) - \beta_{n,\alpha}\delta(t)}{1 - t} > \lambda\right) \quad (5)$$

$$= P\left(\sup_{t \in (0,1)} (1 - \lambda)(U_{n_0}(t) - t) - \beta_{n,\alpha}\delta(t) > 0\right) \quad (6)$$

$$= P\left(\sup_{t \in (0,1)} U_{n_0}(t) - t - \frac{n}{n_0}\beta_{n,\alpha}\delta(t) > 0\right). \quad (7)$$

As  $n\beta_{n,\alpha}$  is monotonically increasing,  $n\beta_{n,\alpha}/n_0 \geq \beta_{n_0,\alpha}$  and the proof follows by property (b) in Definition 1.

**Asymptotic control** Instead of finite-sample control, it is sometimes more convenient to resort to asymptotic control. A sequence  $\beta_{n,\alpha}$  is said to be an *asymptotic bounding sequence* if  $\beta_{n,\alpha}$  fulfills condition (a) from Definition 1 and, additionally a modified condition (b'),

$$\limsup_{n \rightarrow \infty} P(V_{n,\delta} > \beta_{n,\alpha}) < \alpha.$$

If we suppose that the absolute number of false null hypotheses  $n\lambda$  is growing with  $n$ , that is  $n\lambda \rightarrow \infty$  for  $n \rightarrow \infty$ , it holds for an asymptotic bounding sequence that

$$\limsup_{n \rightarrow \infty} P(\hat{\lambda} \leq \lambda) \geq 1 - \alpha.$$

Asymptotic control is typically useful in the following situation. For a given bounding function  $\delta(t)$  and two sequences  $a_n, b_n$ , consider weak convergence of

$$a_n V_{n,\delta} - b_n \rightarrow_D L, \quad (8)$$

to a distribution  $L$ . Any sequence  $\beta_{n,\alpha}$  which satisfies the monotonicity condition (a) of Definition 1 and, additionally,  $\beta_{n,\alpha} \geq a_n^{-1}(L^{-1}(1 - \alpha) + b_n)$ , is thus an asymptotic bounding sequence at level  $\alpha$ .

As an important example, consider the bounding function  $\delta(t) = \sqrt{t(1 - t)}$ . The following Lemma is due to Jäschke and can be found in Shorack and Wellner (1986), p. 599, Theorem 1 (18).

**Lemma 1** Let  $a_n = \sqrt{2n \log \log n}$  and  $b_n = 2 \log \log n + \frac{1}{2} \log \log \log n - \frac{1}{2} \log 4\pi$ . Then

$$a_n \sup_{t \in (0,1)} \frac{U_n(t) - t}{\sqrt{t(1-t)}} - b_n \rightarrow_D E^2, \quad (9)$$

where  $E$  is the Gumbel distribution  $E(x) = \exp(-\exp(-x))$ .

**Remark 1** The convergence in (9) is in general slow. Nevertheless, the result is of interest here. First, the number of tested hypotheses is potentially very large (e.g.  $10^{12}$  in the TAOS setting described in the introduction). Moreover, the slow convergence is mainly caused by values of  $t$  which are of order  $1/n$ . The expected value of the smallest  $p$ -value of true null hypotheses is at least  $1/n$  and it might be useful to truncate in practice the range over which the supremum is taken in (3) to  $(1/n, 1 - 1/n)$ . Doing so, the following asymptotic results are still valid, while the approximation by the Gumbel distribution is empirically a good fit even for moderate values of  $n$  (Donoho and Jin, 2004).

Similar weak convergence results for other bounding function can be found in Csörgő and Horvath (1993) or Shorack and Wellner (1986).

## 2.1 Bounding functions

The estimate is determined by the choice of the function  $\delta(t)$ , the so-called bounding function, and a suitable bounding sequence.

There are many conceivable bounding functions. Bounding functions of particular interest include

- linear bounding function:  $\delta(t) = t$ .
- constant bounding function:  $\delta(t) = 1$ .
- standard deviation-proportional bounding function:  $\delta(t) = \sqrt{t(1-t)}$ .

The *linear bounding function* is closely related to the False Discovery Rate (FDR), as introduced by Benjamini and Hochberg (1995). In the FDR setting, the empirical distribution of  $p$ -values is compared to the linear function  $t/\alpha$ . The last down-crossing of the empirical distribution over the line  $t/\alpha$  determines the number of rejections that can be made when controlling FDR at level  $\alpha$ . It is interesting to compare this to the current setting. In particular, it follows by a result in Daniels (1945) that the optimal bounding sequence at level  $\alpha$  is given for the linear bounding function by  $\beta_{n,\alpha} = 1/\alpha - 1$ . Let  $\hat{\lambda}$  be the estimate under the linear bounding function. The estimate vanishes hence, that is  $\hat{\lambda} = 0$ , if and only if no rejections can be made under FDR-control at the same level. Note that the bounding sequence is independent of the number of observations. This leads to weak power to detect the full proportion  $\lambda$  of false null hypotheses when the proportion  $\lambda$  is rather high but the distribution of  $p$ -values under the alternative deviates only weakly from the uniform distribution, as shown in an asymptotic analysis below.

An estimate under a *constant bounding function* was already proposed in Genovese and Wasserman (2004). Using the Dvoretzky-Kiefer-Wolfowitz (DKW) inequality, a bounding sequence is given by  $\beta_{n,\alpha}^2 = \frac{1}{2n} \log \frac{2}{\alpha}$ . In contrast to the linear bounding function, this bounding function sequence vanishes for  $n \rightarrow \infty$ . However, the estimate is unable to detect any proportion of false null hypotheses which is of smaller order than  $\sqrt{n}$ . The intuitive reason is that the bounding function  $\delta(t)$  is not vanishing for small values of  $t$ . Any evidence from false null hypotheses, however strong it may be, is hence lost if there are just a few false null hypotheses.

As already argued above, a bounding sequence for the *standard deviation-proportional bounding function* is given by  $\beta_{n,\alpha} = a_n^{-1}(E^{-1}(1-\alpha) + b_n)$ , where  $E$  is the Gumbel distribution and  $a_n, b_n$  are defined as in Lemma 1. Note that the bounding sequence is vanishing at almost the same rate as for the constant bounding function. In contrast to the constant bounding function, however, the standard deviation-proportional bounding function vanishes for small  $t$ . It will be seen that the standard deviation-proportional bounding function possesses optimal properties among a large class of possible bounding functions.

**Asymptotic properties of bounding sequences** Faced with an enormous number of potential bounding functions, it is of interest to look at general properties of bounding functions, especially the asymptotic behaviour of the resulting estimates.

The asymptotic properties turn out to be mainly determined by the behaviour of  $\delta(t)$  close to the origin.

**Definition 3** *Let, for every  $\nu \in [0, 1]$ ,  $Q_\nu$  be a family of real-valued functions on  $[0, 1]$ . In particular,  $\delta(t) \in Q_\nu$  iff*

(a)  $\delta(t)$  is non-negative and finite on  $[0, 1]$  and strictly positive on  $(0, 1)$ ,

(b)  $\delta(1-t) \geq \delta(t)$  for  $t \in (0, \frac{1}{2})$ ,

(c) the function  $\delta(t)$  is regularly varying with power  $\nu$ , that is

$$\lim_{t \rightarrow 0} \frac{\delta(bt)}{\delta(t)} = b^\nu.$$

Most bounding functions of interest are members of  $Q_\nu$  for some value of  $\nu \in [0, 1]$ . The constant bounding function is member of  $Q_0$ , while the linear bounding function is member of  $Q_1$  and the standard deviation-proportional bounding function is a member of  $Q_{1/2}$ .

It holds in general for any bounding function that bounding sequences can not be of smaller order than the inverse square root of  $n$ . In particular, note that by Definition 1 of a bounding sequence it has to hold for any  $t \in (0, 1)$  that  $P(U_n(t) - t - \beta_{n,\alpha}\delta(t) > 0) < \alpha$  for all  $n \in \mathbb{N}$ . As  $nU_n(t) \sim B(n, t)$  is

binomially distributed with mean  $nt$  and variance proportional to  $n$ , it follows indeed that

$$\liminf_{n \rightarrow \infty} n^{\frac{1}{2}} \beta_{n,\alpha} > 0$$

Consider now bounding functions  $\delta(t)$ , which are member of  $Q_\nu$  with some  $\nu \in (\frac{1}{2}, 1]$ . It follows directly from Theorem 1.1 (iii), p. 255, in Csörgő and Horvath (1993) that a more restrictive assumption has to hold in this case, namely

$$\liminf_{n \rightarrow \infty} n^{1-\nu} \beta_{n,\alpha} > 0 \quad (10)$$

For  $\nu = 1$  this amounts to  $\liminf_{n \rightarrow \infty} \beta_{n,\alpha} > 0$ . The linear bounding function is a member of  $Q_1$ , explaining the lack of convergence to zero of the corresponding optimal bounding sequence  $1/\alpha - 1$ .

For bounding functions  $\delta(t) \in Q_\nu$  with  $\nu \in [0, \frac{1}{2}]$ , there exists some constant  $c > 0$ , so that  $c\delta(t)^2 \geq t(1-t)$ . Hence, using Lemma 1, there exist bounding sequences so that

$$\limsup_{n \rightarrow \infty} \left( \frac{n}{\log \log n} \right)^{\frac{1}{2}} \beta_{n,\alpha} < \infty. \quad (11)$$

The different asymptotic behaviour of the bounding sequences influences the asymptotic power to detect false null hypotheses, as will be seen in the following.

### 3 Power

We examine the influence of the bounding function  $\delta(t)$  on the power to detect false null hypothesis. For simplicity of exposition, it is assumed that the p-values of all false null hypotheses follow a common distribution  $G$ , while p-values of true null hypotheses have a uniform distribution on  $[0, 1]$ . For some  $\gamma \in (0, 1)$ , let

$$\lambda \sim n^{-\gamma}.$$

A value of  $\gamma = 0$  corresponds to a fixed proportion of false null hypotheses, while  $\gamma = 1$  corresponds to a fixed absolute number of false null hypotheses. Here all cases between those two extremes are considered.

**Bounding sequences with vanishing level** For the asymptotic analysis, it is convenient to let  $\alpha = \alpha_n$  decrease monotonically for  $n \rightarrow \infty$ , so that  $\alpha_n \rightarrow 0$  for  $n \rightarrow \infty$ . Note that  $\alpha_n \rightarrow 0$  is equivalent to  $P(V_{n,\delta} > \beta_{n,\alpha_n}) \rightarrow 0$  for  $n \rightarrow \infty$ . For notational simplicity, this assumption is strengthened slightly to

$$V_{n,\delta}/\beta_{n,\alpha_n} \rightarrow_p 0 \quad n \rightarrow \infty. \quad (12)$$

In almost all cases of interest, (12) and  $\alpha_n \rightarrow 0$  are equivalent.

To maintain reasonable power, one would like to avoid that the level  $\alpha_n$  is vanishing too fast as  $n \rightarrow \infty$ . For bounding functions  $\delta(t) \in Q_\nu$  with  $\nu \in [0, \frac{1}{2}]$  it is required that

$$\limsup_{n \rightarrow \infty} \left( \frac{n}{\log n} \right)^{\frac{1}{2}} \beta_{n,\alpha_n} < \infty. \quad (13)$$

It follows from (11) that it is always possible to find a sequence  $\alpha_n \rightarrow 0$  so that both (12) and (13) are satisfied. If both (12) and (13) are fulfilled, the sequence  $\alpha_n$  is said to *vanish slowly*. For bounding functions  $\delta(t) \in Q_\nu$  with  $\nu \in (1/2, 1]$ , it will be seen below that the power is poor no matter how slow the sequence  $\alpha_n$  vanishes for  $n \rightarrow 0$ .

### 3.1 Case I: many false null hypotheses, $\gamma \in [0, \frac{1}{2})$

The fluctuations in the empirical distribution function are negligible compared to the signal from false null hypotheses if  $\gamma \in [0, \frac{1}{2})$ . Hence one should be able to detect (asymptotically) the full proportion of false null hypotheses in this first setting.

This is indeed achieved, as long as we look for bounding functions in  $Q_\nu$  with  $\nu \in [0, \frac{1}{2}]$ , as shown below. If on the other hand,  $\nu \in (\frac{1}{2}, 1]$ , one is in general unable to detect the full proportion of false null hypotheses. The proportion of detected false null hypotheses is even converging in probability to zero for large values of  $\gamma$  if  $\nu$  is in the range  $(\frac{1}{2}, 1]$ . This includes in particular the linear FDR-style bounding function  $t \in Q_1$ , which is only able to detect a non-vanishing proportion of false null hypotheses (asymptotically) as long as the proportion  $\lambda$  is bounded from below, which is only fulfilled for  $\gamma = 0$ .

**Theorem 2** *Let  $G$  be continuous and  $\inf_{t \in (0,1)} G'(t) = 0$ . Let  $\hat{\lambda}$  be the estimate under bounding function  $\beta_{n,\alpha}\delta(t)$ , where  $\delta(t) \in Q_\nu$  with  $\nu \in [0, 1]$  and  $\beta_{n,\alpha}$  is a bounding sequence. If  $\nu \in [0, \frac{1}{2}]$  and  $\alpha_n$  vanishes slowly, then, for all  $\gamma \in [0, \frac{1}{2})$ ,*

$$\frac{\hat{\lambda}}{\lambda} \rightarrow_p 1 \quad n \rightarrow \infty.$$

*However, for  $\nu \in (\frac{1}{2}, 1]$  and  $\gamma \in (1 - \nu, \frac{1}{2})$ ,*

$$\frac{\hat{\lambda}}{\lambda} \rightarrow_p 0 \quad n \rightarrow \infty.$$

**Remark 2** *The case  $\inf_{t \in (0,1)} G'(t) = 0$  corresponds to the “pure” case in Genovese and Wasserman (2004). If  $\inf_{t \in (0,1)} G'(t) > 0$ , the results above (and below) hold true if  $\lambda$  is replaced by*

$$\underline{\lambda} = (1 - \inf_{t \in (0,1)} G'(t))\lambda.$$

*Without making parametric assumptions about the distribution  $G$  under the alternative, identifying  $\underline{\lambda}$  is indeed the best one can hope for.*

The message from the Theorem 2 is that one should look for bounding functions in  $Q_\nu$  with  $\nu \in [0, \frac{1}{2}]$ . This guarantees proper behavior of the estimate if the proportion  $\lambda$  of false null hypotheses is vanishing slower than the square root of the number of observations.



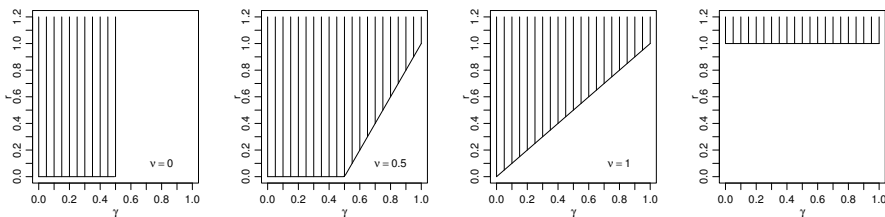


Figure 1: For  $\nu = 0$  (left),  $\nu = 1/2$  (second from left),  $\nu = 1$  (second from right), an illustration of the asymptotic properties of the estimate  $\hat{\lambda}$ . The shaded area marks those areas in the  $(r, \gamma)$ -plane where  $\hat{\lambda}/\lambda \rightarrow_p 1$ , whereas for the white areas  $\hat{\lambda}/\lambda \rightarrow_p 0$ . The choice  $\nu = 1/2$  is seen to be optimal. The corresponding plot for control of the family-wise error rate is shown on the right for comparison.

### 3.2 Case II: few false null hypotheses, $\gamma \in [\frac{1}{2}, 1)$

As seen above, bounding functions in  $Q_\nu$  with  $\nu \leq \frac{1}{2}$  detect asymptotically the full proportion  $\lambda$  of false null hypotheses if  $\lambda$  is vanishing not as fast as the square root of the number of observations.

For  $\gamma > \frac{1}{2}$ , no method whatsoever can detect asymptotically the full proportion of false null hypotheses *if the distribution under the alternative is fixed*. It is hence interesting to consider cases where the signal from false null hypotheses is increasing in strength. Let therefore  $G = G^{(n)}$ , the distribution of p-values under the alternative, be a function of the number  $n$  of tests to conduct. The superscript is dropped in the following for notational simplicity.

**Shift-location testing** It is perhaps helpful to think about  $G$  as being induced by some shift-location testing problem. For each test it is assumed that there is a test statistic  $Z_i$ , which follows some distribution  $T_0$  under the null hypothesis  $H_{0,i}$  and some shifted distribution  $T_{\mu_n}$  under the alternative  $H_{1,i}$ ,

$$\begin{aligned} H_{0,i} &: Z_i \sim T_0, \\ H_{1,i} &: Z_i \sim T_{\mu_n}. \end{aligned} \tag{14}$$

In the Gaussian case, this amounts e.g. to  $T_0 = \mathcal{N}(0, 1)$  and  $T_\mu = \mathcal{N}(\mu_n, 1)$ . To have an interesting problem, one needs for  $\gamma \in (\frac{1}{2}, 1)$  in general that the shift  $\mu_n$  between the null and alternative hypothesis is increasing for increasing number of tests, that is  $\mu_n \rightarrow \infty$  for  $n \rightarrow \infty$ . On the other hand, one would like to keep the problem subtle. For the Gaussian case it was shown in Donoho and Jin (2004) that an interesting scaling is given by  $\mu_n = \sqrt{2r \log n}$  with  $r \in (0, 1)$ . In this regime, the smallest p-value stems with high probability from a true null hypothesis. The false null hypotheses have hence little influence on the extremes of the distribution.

Instead of assuming Gaussianity of the test statistics, Donoho and Jin (2004) considered a variety of different distributions. Under a generalized Gaussian

(Subbotin) distribution, the density is for some positive value of  $\kappa$  proportional to

$$T'_\mu(x) \propto \exp\left(-\frac{|x-\mu|^\kappa}{\kappa}\right).$$

The case  $\kappa = 2$  corresponds clearly to a Gaussian distribution, and  $\kappa = 1$  corresponds to the double exponential case. The shift parameter is chosen then as

$$\mu_n = (\kappa r \log n)^{\frac{1}{\kappa}} \quad (15)$$

for some  $r \in (0, 1)$ . Note that the expectation of the smallest p-value from true null hypotheses vanishes like  $n^{-1}$ , whereas under the scaling (15), the median p-value of false null hypotheses vanishes like  $n^{-r}$  for  $n \rightarrow \infty$  with some  $r \in (0, 1)$ . In fact, consider for any member of the generalized Gaussian Subbotin distribution the q-quantile  $G^{-1}(q)$  of the distribution of p-values under the alternative. For some constant  $c_q$ , the q-quantile is proportional to

$$G^{-1}(q) \propto \int_{\mu_n + c_q}^{\infty} \exp\left(-\frac{x^\kappa}{\kappa}\right) dx.$$

Applying l'Hopitals rule twice, it follows for any  $c$  and  $\kappa > 0$  that

$$\lim_{a \rightarrow \infty} \frac{\log \int_{a+c}^{\infty} \exp\left(-\frac{x^\kappa}{\kappa}\right) dx}{-\frac{a^\kappa}{\kappa}} = 1.$$

Thus it holds under the scaling (15) for any every  $q \in (0, 1)$  and positive  $\kappa$  that the scaling of the q-quantile is given by

$$\log G^{-1}(q) \sim -r \log n. \quad (16)$$

With probability converging to 1 for  $n \rightarrow \infty$ , a p-value under a false null hypothesis is hence larger than the smallest p-value from all true null hypotheses as long as  $r \in (0, 1)$ . For  $r > 1$ , the problem gets trivial as the probability that an arbitrarily high proportion of p-values under false null hypotheses is smaller than the smallest p-value from all true null hypotheses converges to 1 for  $n \rightarrow \infty$ .

The point of introducing the shift-location model under generalized Gaussian Subbotin distributions was just to identify (16) with  $r \in (0, 1)$  as the interesting scaling behaviour of quantiles of  $G$ , the p-value distribution for alternative hypotheses. The setting (16) is potentially of interest beyond any shift-location model. We adopt the scaling (16) for the following without making any explicit distributional assumptions about underlying test statistics.

**Theorem 3** *Let  $\lambda \sim n^{-\gamma}$  with  $\gamma \in [\frac{1}{2}, 1)$  and let the distribution  $G$  of p-values under the alternative fulfill (16) for some  $r \in (0, 1)$ . Let  $\hat{\lambda}$  be the estimate of  $\lambda$  under a bounding function  $\beta_{n,\alpha}\delta(t)$ , where  $\delta(t) \in Q_\nu$  with  $\nu \in [0, \frac{1}{2}]$  and  $\beta_{n,\alpha_n}$  is a bounding sequence for  $\delta(t)$ . Let  $\alpha_n$  vanish slowly. If  $r > \frac{1}{\nu}(\gamma - \frac{1}{2})$ , it holds for any  $\epsilon > 0$ ,*

$$\frac{\hat{\lambda}}{\lambda} \rightarrow_p 1. \quad (17)$$

If on the other hand,  $r < \frac{1}{\nu}(\gamma - \frac{1}{2})$ , then

$$\frac{\hat{\lambda}}{\lambda} \rightarrow_p 0. \quad (18)$$

**Remark 3** *The analysis was only carried out for functions with  $\nu \in [0, \frac{1}{2}]$  due to the deficits of the functions with  $\nu \in (\frac{1}{2}, 1]$  discussed in the previous section. Nevertheless, it would be possible to carry out the same analysis here. For  $\nu = 1$ , one obtains e.g. a critical boundary  $r > \gamma$ .*

The message from the last theorem is that among all bounding function in  $Q_\nu$  with  $\nu \in [0, \frac{1}{2}]$ , it is best to choose a member of  $Q_{1/2}$ . Bounding functions in  $Q_{1/2}$  increase the chance to detect the full proportion  $\lambda$  of false null hypotheses, as illustrated for a few special cases in Figure 1. The area in the  $(r, \gamma)$ - plane, where  $\hat{\lambda}/\lambda$  converges in probability to 1 for a bounding function in  $Q_{1/2}$  includes in particular all areas of convergence for bounding function in  $Q_\nu$  with  $\nu \in [0, \frac{1}{2}]$ .

### 3.3 Connection to the Family-wise error rate

A different estimate of  $\lambda$  is obtained by controlling the family-wise error rate (FWER). In particular, let the estimate be the total number of p-values less than the FWER-threshold  $\alpha/n$ , divided by the total number of hypotheses,

$$\hat{\lambda} = F_n\left(\frac{\alpha}{n}\right).$$

This is an estimate of  $\lambda$  with the desired property  $P(\hat{\lambda} > \lambda) < \alpha$ . Controlling the family-wise error rate has often been criticised for lack of power. Indeed, in the asymptotic analysis above, it is straightforward to show that the area in the  $(r, \gamma)$ -plane, where  $\hat{\lambda}/\lambda \rightarrow_p 1$  is restricted to the half-plane  $r > 1$  (neglecting again what happens directly on the border  $r = 1$ ). In comparison to other estimates proposed here, the family-wise error rate is hence particularly bad for estimating  $\lambda$  if there are many false null hypotheses, each with a very weak signal. In addition, the construct requires that p-values can be determined accurately down to precision  $\alpha/n$ , which might be prohibitively small. In contrast, the performance of estimates of the form (3) do not deteriorate significantly if p-values are truncated at larger values.

The drawbacks of the family-wise error rate are a consequence of the stricter inference one is trying to make when controlling the family-wise error rate. In particular, one is trying to infer exactly *which* hypotheses are false nulls as opposed to only *how many* false nulls there are in total. The loss in power is hence the price one is paying for this more ambitious goal.

### 3.4 Connection to Higher Criticism

A connection of the proposed estimate to the Higher Criticism method in Donoho and Jin (2004) for detection of sparse heterogeneous mixtures emerges.

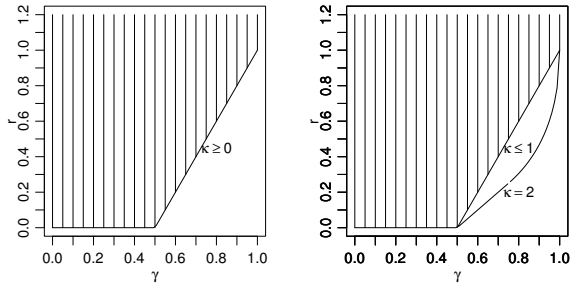


Figure 2: Comparison between the estimate of  $\hat{\lambda}$  and detection regions under Higher Criticism if test statistics follows the location-shift model (14) and are distributed according to the generalized Gaussian Subbotin distribution with shift parameter (15). The shaded area in the left panel shows again the area of convergence in probability of  $\hat{\lambda}/\lambda$  to 1 for a bounding function in the class  $Q_{1/2}$ . The shaded area in the right panel corresponds to the the region where Higher Criticism can reject asymptotically the null hypothesis  $H_0 : \bar{\lambda} = 0$  for  $\kappa \leq 1$ , including the double-exponential case. The line below marks the detection boundary for the Gaussian case ( $\kappa = 2$ ).

In their setup, p-values  $P_i$ ,  $i = 1, \dots, n$  are distributed i.i.d. according to a mixture distribution

$$P_i \sim (1 - \bar{\lambda})H + \bar{\lambda}G,$$

where  $H$  is the uniform distribution and  $G$  the distribution of p-values under the alternative hypothesis. In Donoho and Jin (2004) the focus is on testing the global null hypothesis that there are no false null hypotheses at all,

$$H_0 : \quad \bar{\lambda} = 0.$$

In this current paper we are in contrast interested in quantifying the proportion  $\lambda$  of false null hypotheses. The proportion  $\lambda$  of false null hypotheses, as defined for the current paper in (1), can be viewed as a realization of a random variable with a Binomial distribution  $n\lambda \sim B(n, \bar{\lambda})$ . For the asymptotic considerations of this paper, however, the distinction between  $\bar{\lambda}$  and  $\lambda$  is of little importance as the ratio  $\lambda/\bar{\lambda}$  converges almost surely to 1 for  $n \rightarrow \infty$ .

The two goals of Higher Criticism and the current paper are connected. If there is evidence for a nonnegative proportion of false null hypotheses with the method proposed previously, then the global null  $H_0$  can clearly be rejected. In other words, if one obtains a positive estimate  $\hat{\lambda} > 0$  with  $P(\hat{\lambda} > \lambda) < \alpha$ , then the global null hypothesis  $H_0 : \bar{\lambda} = 0$  can be rejected at level  $\alpha$ . Note that the level is correct even for finite samples and not just asymptotically.

The connection between the two methods works as well in the reverse direction if an optimal bounding function is chosen. It emerged in particular from the analysis above that bounding functions which are member of  $Q_{1/2}$

have optimal asymptotic properties. For the particular choice of a standard deviation-proportional bounding function in  $Q_{1/2}$ , let  $\hat{\lambda}$  be an estimate of  $\lambda$  and let  $\beta_{n,\alpha}$  be a bounding sequence which satisfies

$$\beta_{n,\alpha} = n^{-\frac{1}{2}}(2 \log \log n)^{\frac{1}{2}}(1 + o(1)).$$

Donoho and Jin (2004) are not specific about choice of a critical value for Higher Criticism. However, choosing  $\sqrt{n}\beta_{n,\alpha}$  as a critical value meets their requirements. The Higher criticism procedure rejects in this case if and only if the estimate  $\hat{\lambda}$  of the proportion of false null hypotheses is positive,

$$\{\text{Reject } H_0 : \bar{\lambda} = 0 \text{ with Higher Criticism}\} = \{\hat{\lambda} > 0\}.$$

If both  $\bar{\lambda} \sim n^{-\gamma}$  and  $\lambda \sim n^{-\gamma}$  for some  $\gamma \in [0, 1]$ , the question arises if the area in the  $(\gamma, r)$ -plane, where

$$P(\text{Higher Criticism rejects } H_0) \rightarrow 1 \tag{19}$$

is identical to the area where

$$\hat{\lambda}/\lambda \rightarrow_p 1. \tag{20}$$

Intuitively, it is clear that it is somewhat easier to test for the global null hypothesis  $H_0 : \bar{\lambda} = 0$ , as done in Higher Criticism, than to estimate the precise proportion  $\lambda$  of false null hypotheses, as done in this paper. One would therefore expect that the area of convergence in the  $(\gamma, r)$ -plane of (19) includes the area of convergence of (20).

It is hence maybe surprising that for some cases the areas of convergence in the  $(\gamma, r)$ -plane of (19) and (20) agree. To illustrate the point, consider again the shift-location model (14) under a generalized Gaussian Subbotin distribution with parameter  $\kappa \in (0, 2)$  and a shift (15) of test statistics under the alternative.

The area in the  $(\gamma, r)$ -plane where  $\hat{\lambda}/\lambda \rightarrow_p 1$  is in this setting independent of the parameter  $\kappa$ . The detection boundary for Higher Criticism, however, does depend on  $\kappa$ . For the Gaussian case ( $\kappa = 2$ ) and in general for  $\kappa > 1$ , the detection boundary for Higher Criticism is, for  $\gamma \in (1/2, 1)$ , below the area where  $\hat{\lambda}/\lambda \rightarrow_p 1$ . The reason for this is intuitively clear. The Higher Criticism method looks in these cases for evidence against  $H_0$  in the extreme tails of the distribution  $G$ , see Donoho and Jin (2004). At these points, only a vanishing proportion of all p-values from false null hypotheses can be found. If trying to estimate the full proportion of false null hypotheses, the evidence for a certain amount of false null hypotheses has to be found at less extreme points, where one can expect a significant proportion of p-values from false null hypotheses. This limits the region of convergence in the sense of (20) compared to the area where Higher Criticism can successfully reject the global null hypothesis  $H_0 : \bar{\lambda} = 0$ .

However, for  $\kappa \leq 1$  (including thus the case of a double exponential distribution) the two areas where (19) and (20) hold respectively are identical, as shown in Figure 2. In the white area, both Higher Criticism and the current

method fail to detect (asymptotically) the presence of false null hypotheses and not even the likelihood ratio test is able to reject in these cases (asymptotically) the global null hypotheses  $H_0 : \bar{\lambda} = 0$  that there are only true null hypotheses (Donoho and Jin, 2004). It is hence of interest to see that for  $\kappa \leq 1$  it holds that  $\hat{\lambda}/\lambda \rightarrow_p 1$  whenever the likelihood ratio test succeeds (asymptotically) in rejecting the global null hypothesis.

## 4 Numerical Examples

The behaviour of the estimate  $\hat{\lambda}$  is illustrated by some numerical examples. The distribution  $G$  of p-values under the alternative is induced by shift-location testing as in (14). The results are shown for the double-exponential distribution ( $\kappa = 1$ ) in Figure 3 and for the Gaussian distribution ( $\kappa = 2$ ) in Figure 4. Test statistics under false null hypotheses are shifted by  $\mu_n = r \log n$  for  $\kappa = 1$  and  $\mu_n = \sqrt{2r \log n}$  for  $\kappa = 2$ .

The proportion  $\lambda$  of false null hypotheses is given by  $n\lambda = \lfloor n^{1-\gamma} \rfloor$ , which implies  $\lambda \sim n^{-\gamma}$  for  $n \rightarrow \infty$ . The estimate  $\hat{\lambda}$  is simulated at level  $\alpha = 0.05$  for both  $\kappa = 1$  and  $\kappa = 2$  at 100 evenly spaced grid-points in the  $(r, \gamma)$ - plane 100 times each. The average proportion  $\hat{\lambda}/\lambda$  is calculated and a contour plot is shown in Figures 3 and 4 respectively for the constant bounding function ( $\nu = 0$ ), the standard deviation-proportional bounding function ( $\nu = 1/2$ ) and the linear, FDR-style bounding function ( $\nu = 1$ ).

The areas where  $\hat{\lambda}/\lambda$  converges to 1 or 0 respectively are separated in both Figure 3 and Figure 4 by an unbroken line. The asymptotic behaviour seems to be a good indicator of the performance for finite but reasonably large  $n$ .

## 5 Acknowledgements

This research was supported in part by a grant from the U.S. National Science Foundation and by FIM, Institute for Mathematical Research, ETH Zürich.

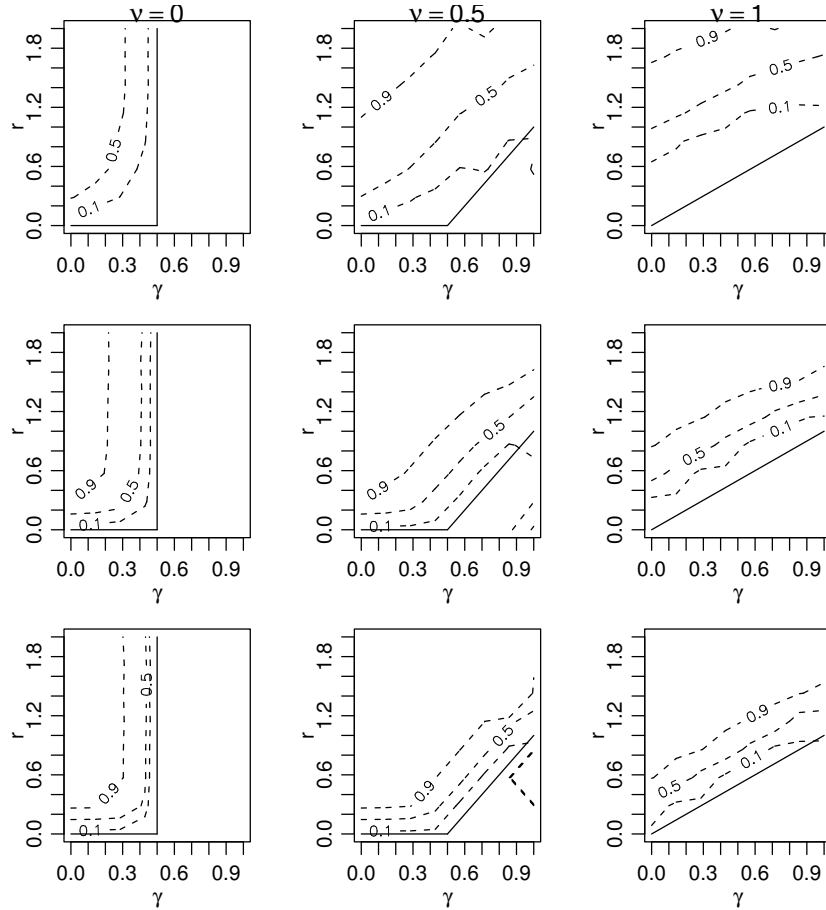


Figure 3: For  $\nu = 0$  (left column),  $\nu = 1/2$  (middle) and  $\nu = 1$  (right), a contour plot of the average detected proportion  $\hat{\lambda}/\lambda$  of all false null hypotheses over 100 simulations for  $n = 10^2$  tests (top row),  $n = 10^4$  (middle row) and  $n = 10^6$  tests (bottom row). The unbroken line separates the area where  $\hat{\lambda}/\lambda$  converges in probability to 1 from the area where the ratio  $\hat{\lambda}/\lambda$  converges to 0.

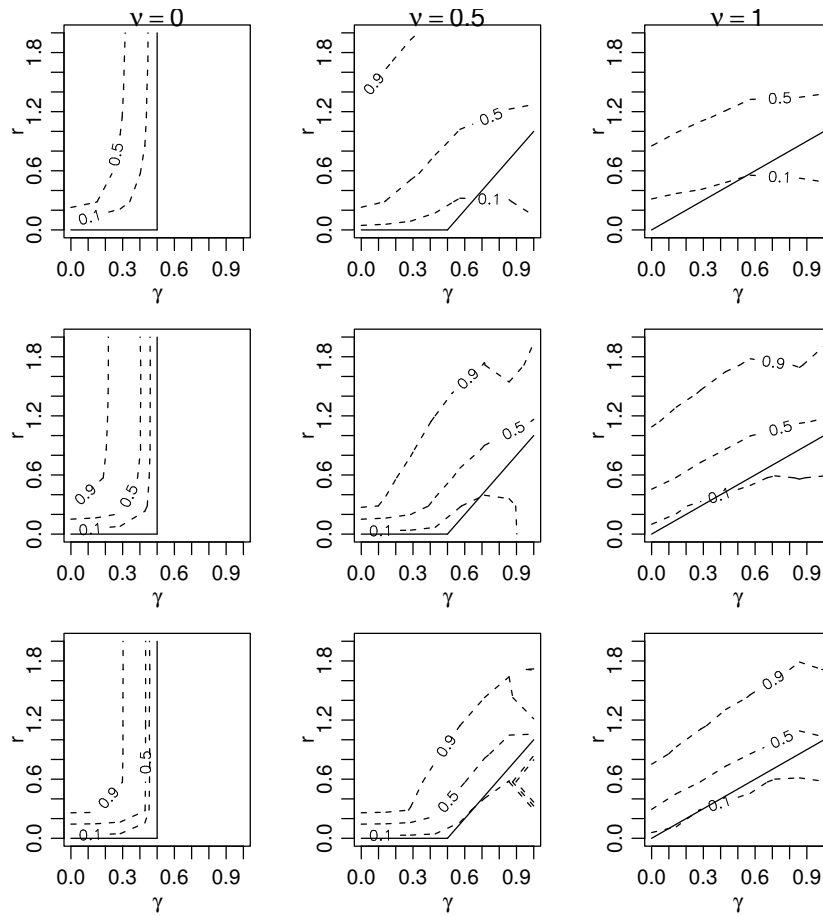


Figure 4: The same as in Figure 3 for a Gaussian distribution ( $\kappa = 2$ ) of the test statistics.



## References

- [1] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* **57**, 289-300.
- [2] Benjamini, Y. and Hochberg, Y. (2000). The adaptive control of the False Discovery Rate in multiple hypothesis testing with independent statistics. *Journal of Educational and Behavioral Statistics* **25**, 60-83.
- [3] Csörgő, M. and Horvath, L. (1993). *Weighted Approximations in Probability and Statistics*. Wiley
- [4] Chen, W. P.; Zhang, Z. W.; King, S. K.; Alcock, C.; Byun, Y. I.; Cook, K. H.; Dave, R.; Giammarco, J.; Lee T.; Lehner, M.; Liang, C.; Lissauer, J.; Marshall, S.; de Pater, I.; Porrata, R.; Rice, J.; Wang, A.; Wang, S. Y.; and Wen, C. Y. (2003) Fast CCD Photometry in the Taiwanese-American Occultation Survey. *Baltic Astronomy* **12**, 568.
- [5] Daniels, H. E. (1945). The statistical theory of the strength of bundles of thread. *Proc. Roy. Soc. A* **183**, 405-435.
- [6] Donoho, D. and Jin, J. (2004). Higher Criticism for detecting sparse heterogeneous mixtures. *Annals of Statistics* **32**, 992-994
- [7] Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *Annals of Statistics* **32**, 1035-1061
- [8] Liang, C-L, Rice, J.A., de Pater, I., Alcock, C., Axelrod, T., Wang, A. and Marshall, S. (2002). Statistical Methods for Detecting Stellar Occultations by Kuiper Belt Objects: the Taiwanese-American Occultation Survey. *Statistical Science*, to appear.
- [9] Meinshausen, N. and Bühlmann, P. (2004). Upper bounds for the number of true null hypotheses and novel estimates for error rates in multiple testing. Manuscript
- [10] Nettleton, D. and Hwang, G. (2003). Estimating the number of false null hypotheses when conducting many tests. Technical Report, Department of Statistics, Iowa State University.
- [11] Schweder, T. and Spjøtvoll, E. (1983). Plots of p-values to evaluate many tests simultaneously. *Biometrika* **69**, 493-502.
- [12] Shorack, G. and Wellner, J. (1986). *Empirical Processes with Applications to Statistics*. Wiley.
- [13] Storey, J. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B* **64**, 479-408.

## 6 Appendix

PROOF OF THEOREM 2. First it is shown that, as long as  $\gamma \in (0, \frac{1}{2})$  and  $\nu \leq \frac{1}{2}$ , for any given  $\epsilon > 0$ ,

$$P(\hat{\lambda} < (1 - \epsilon)\lambda) \rightarrow 0 \quad n \rightarrow \infty. \quad (21)$$

For any  $t < 1$ ,

$$\hat{\lambda} = \sup_{t \in (0,1)} \frac{F_n(t) - t - \beta_{n,\alpha_n} \delta(t)}{1 - t} \quad (22)$$

$$\geq \frac{F(t) - t}{1 - t} + \frac{F_n(t) - F(t) - \beta_{n,\alpha_n} \delta(t)}{1 - t} \quad (23)$$

$$= \lambda \frac{G(t) - t}{1 - t} + \frac{F_n(t) - F(t) - \beta_{n,\alpha_n} \delta(t)}{1 - t}. \quad (24)$$

As  $\inf_{t \in (0,1)} G'(t) = 0$ , and hence  $\sup_{t \in (0,1)} (G(t) - t)/(1 - t) = 1$ , there exists by continuity of  $G(t)$  some  $t_1$  so that  $(G(t_1) - t_1)/(1 - t_1) > (1 - \epsilon/2)$ . Setting  $\tilde{\epsilon} = \frac{1}{2}(1 - t_1)\epsilon$ , it suffices to show that for every  $\epsilon > 0$ ,

$$P(\beta_{n,\alpha_n} \delta(t_1) + F(t_1) - F_n(t_1) > \epsilon\lambda) \rightarrow 0 \quad n \rightarrow \infty.$$

As  $F_n(t_1) - F(t_1) = O_P(n^{-\frac{1}{2}})$  and  $\lambda \sim n^{-\gamma}$  with  $\gamma < \frac{1}{2}$ , this follows from the finiteness of  $\delta(t)$  and, as  $\alpha_n$  vanishes slowly, from (13). This completes the first part of the proof of Theorem 2.

For the second part, it suffices to show that for  $\nu \in (\frac{1}{2}, 1]$  and  $\gamma \in (1 - \nu, \frac{1}{2})$ , and any  $\epsilon > 0$ ,

$$P(\hat{\lambda} > \epsilon\lambda) \rightarrow 0 \quad n \rightarrow \infty. \quad (25)$$

In this regime, the penalty  $\beta_{n,\alpha_n} \delta(t)$  is asymptotically larger than the signal from false null hypotheses. Using the definition of  $\hat{\lambda}$ , the notation  $n_0 = (1 - \lambda)n$  and  $n_1 = \lambda n$ , and  $F_n(t) = \lambda G_{n_1}(t) + (1 - \lambda)U_{n_0}(t)$ , it follows that

$$\begin{aligned} P(\hat{\lambda} > \epsilon\lambda) &= P\left(\sup_{t \in (0,1)} \frac{F_n(t) - t - \beta_{n,\alpha_n} \delta(t)}{1 - t} > \epsilon\lambda\right) \\ &= P\left(\sup_{t \in (0,1)} \lambda \frac{G_{n_1}(t) - t}{1 - t} - \epsilon\lambda + (1 - \lambda) \frac{U_{n_0}(t) - t}{1 - t} - \beta_{n,\alpha_n} \frac{\delta(t)}{1 - t} > 0\right) \\ &\leq P\left(\sup_{t \in (0,1)} \lambda \frac{G_{n_1}(t) - t}{1 - t} - \epsilon\lambda - \frac{\beta_{n,\alpha_n}}{2} \frac{\delta(t)}{1 - t} > 0\right) + \end{aligned} \quad (26)$$

$$P\left(\sup_{t \in (0,1)} (1 - \lambda) \frac{U_{n_0}(t) - t}{1 - t} - \frac{\beta_{n,\alpha_n}}{2} \frac{\delta(t)}{1 - t} > 0\right). \quad (27)$$

Observe in (27) that  $(1 - \lambda)^{-1} \beta_{n,\alpha_n} = n \beta_{n,\alpha_n} / n_0 \geq \beta_{n_0,\alpha_n} \geq \beta_{n_0,\alpha_{n_0}}$ . Thus (27) can be bounded by  $P(V_{n_0,\delta} > \beta_{n_0,\alpha_{n_0}}/2)$ . By (12) and  $n_0 \rightarrow \infty$  it follows that (27) vanishes for  $n \rightarrow \infty$ . It remains to show that (26) vanishes as well.

Let  $t_2 = \sup\{t \in (0, 1) : G(t) \leq \epsilon/2\}$ . Using Bonferroni's inequality, (26) is bounded by

$$P\left(\sup_{t \in (0, t_2]} \lambda \frac{G_{n_1}(t) - t}{1 - t} - \epsilon\lambda > 0\right) + \quad (28)$$

$$P\left(\sup_{t \in (t_2, 1)} \lambda \frac{G_{n_1}(t) - t}{1 - t} - \frac{\beta_{n, \alpha_n}}{2} \frac{\delta(t)}{1 - t} > 0\right) \quad (29)$$

The first term (28) is bounded by  $P(G_{n_1}(t_2) > \epsilon)$ , which vanishes for  $n \rightarrow \infty$  as  $n_1 = \lambda n \rightarrow \infty$  and, by definition of  $t_2$ ,  $G(t_2) \leq \epsilon/2$ . Using  $G_{n_1}(t) \leq 1$ , the second term (29) equals zero if  $\beta_{n, \alpha_n} \inf_{t \in (t_2, 1)} \delta(t)/(1 - t) > 2\lambda$ . By conditions (a) and (b) in Definition 3, it holds that  $\inf_{t \in (t_2, 1)} \delta(t)/(1 - t) > 0$ . By (10), it follows furthermore that  $\beta_{n, \alpha_n}/\lambda \rightarrow \infty$  for  $n \rightarrow \infty$ , which completes the proof.

PROOF OF THEOREM 3. First it is shown that for  $r > \frac{1}{\nu}(\gamma - \frac{1}{2})$ ,

$$P(\hat{\lambda} < (1 - \epsilon)\lambda) \rightarrow 0 \quad n \rightarrow \infty. \quad (30)$$

Here, the penalty is again asymptotically larger than the signal from false null hypotheses for a fixed point  $t \in (0, 1)$ . However, as the signal from false null hypotheses is increasing in strength for larger  $n$ , the evidence for a certain amount of false null hypotheses can be found at decreasing values of  $t$ . Using the definition of  $\hat{\lambda}$ , for any  $t \in (0, 1)$ ,  $\hat{\lambda} \geq F_n(t) - t - \beta_{n, \alpha_n} \delta(t)$  and hence, for any  $t \in (0, 1)$ ,

$$\hat{\lambda}/\lambda - 1 \geq (1 - G_{n_1}(t)) - t - \frac{1 - \lambda}{\lambda}(t - U_{n_0}(t)) - \frac{1}{\lambda}\beta_{n, \alpha_n}\delta(t),$$

where again  $n_1 = \lambda n$  and  $n_0 = (1 - \lambda)n$ . Choosing  $t_{n, \tau} = n^{-r+\tau}$  for some  $0 < \tau < r - \frac{1}{\nu}(\gamma - \frac{1}{2})$ , observe that by (16) it follows that  $1 - G(n^{-r+\tau}) = o(1)$ . Hence

$$\begin{aligned} \hat{\lambda}/\lambda - 1 &\geq (1 - G(t_{n, \tau})) - |G(t_{n, \tau}) - G_{n_1}(t_{n, \tau})| \\ &\quad - t_{n, \tau} - \lambda^{-1}|t_{n, \tau} - U_{n_0}(t_{n, \tau})| - \lambda^{-1}\beta_{n, \alpha_n}\delta(t_{n, \tau}) \\ &= o(1) - o_p(1) - o(1) - O_p(n^{\gamma - (\frac{1}{2} + \frac{r - \tau}{2})}) - O(n^{\gamma - (\frac{1}{2} + (r - \tau)\nu)}) \log n \end{aligned}$$

The proof of (30) follows as  $\gamma < \frac{1}{2} + \nu(r - \tau) \leq \frac{1}{2} + \frac{r - \tau}{2}$ .

Second, it has to be shown that  $P(\hat{\lambda} > \epsilon\lambda) \rightarrow 0$  if  $r < \frac{1}{\nu}(\gamma - \frac{1}{2})$ . Again, the evidence for a certain amount of false null hypotheses would have to be found at decreasing values of  $t$ . However, the decrease has to be so fast in this regime that the signal from false null hypotheses is not captured. Using again the notation  $n_1 = \lambda n$  and  $n_0 = (1 - \lambda)n$ , it holds that  $\hat{\lambda} = \sup_{t \in (0, 1)} D_{n, \lambda}(t)$ , where

$$D_{n, \lambda}(t) := \frac{\lambda(G_{n_1}(t) - t) + (1 - \lambda)(U_{n_0}(t) - t) - \beta_{n, \alpha_n}\delta(t)}{1 - t} \quad (31)$$

Choose a sequence  $t_{n, \rho} = n^{-r-\rho}$  for some  $0 < \rho < \frac{1}{\nu}(\gamma - \frac{1}{2}) - r$ . The regions  $(0, t_{n, \rho}]$  and  $(t_{n, \rho}, 1)$  are considered separately for the following. In particular

it is shown that both  $P(\sup_{t \in (0, t_{n, \rho}]} D_{n, \lambda}(t) > \epsilon \lambda)$  and  $P(\sup_{t \in (t_{n, \rho}, 1)} D_{n, \lambda}(t) > \epsilon \lambda)$  vanish for  $n \rightarrow \infty$ . For  $t > t_{n, \rho}$ , it holds that

$$\begin{aligned}
P\left(\sup_{t \in (t_{n, \rho}, 1)} D_{n, \lambda}(t) > \epsilon \lambda\right) &\leq P\left(\sup_{t \in (t_{n, \rho}, 1)} \lambda + (1 - \lambda) \frac{U_{n_0}(t) - t}{1 - t} - \beta_{n, \alpha_n} \frac{\delta(t)}{1 - t} > 0\right) \\
&\leq P\left(\sup_{t \in (t_{n, \rho}, 1)} (1 - \lambda) \frac{U_{n_0}(t) - t}{1 - t} - \frac{\beta_{n, \alpha_n}}{2} \frac{\delta(t)}{1 - t} > 0\right) \\
&\quad + 1\left\{\sup_{t \in (t_{n, \rho}, 1)} \lambda - \frac{\beta_{n, \alpha_n}}{2} \frac{\delta(t)}{1 - t} > 0\right\} \\
&= P\left(\sup_{t \in (t_{n, \rho}, 1)} (U_{n_0}(t) - t) - \frac{n}{n_0} \frac{\beta_{n, \alpha_n}}{2} \delta(t) > 0\right) \quad (32) \\
&\quad + 1\left\{\inf_{t \in (t_{n, \rho}, 1)} \frac{\beta_{n, \alpha_n}}{2} \frac{\delta(t)}{1 - t} < \lambda\right\}. \quad (33)
\end{aligned}$$

By (12) and as  $n\beta_{n, \alpha_n}$  is monotonically increasing, (32) vanishes for  $n \rightarrow \infty$ . For (33), as  $\delta \in Q_\nu$ , there exists some constant  $c$  so that  $\inf_{t \in (t_{n, \rho}, 1)} \delta(t_{n, \rho}) \geq cn^{-\nu(r+\rho)}$ . It follows by  $r + \rho < \frac{1}{\nu}(\gamma - \frac{1}{2})$  that  $\inf_{t \in (t_{n, \rho}, 1)} \beta_{n, \alpha_n} \delta(t_{n, \rho}) / \lambda \rightarrow \infty$  for  $n \rightarrow \infty$ , which completes the first part of the proof.

It remains to show that  $P(\sup_{t \in (0, t_{n, \rho}]} D_{n, \lambda}(t) > \epsilon \lambda) \rightarrow 0$  for  $n \rightarrow \infty$ . It holds that

$$\begin{aligned}
P\left(\sup_{t \in (0, t_{n, \rho}]} D_{n, \lambda}(t) > \epsilon \lambda\right) &\leq \\
&P\left(\sup_{t \in (0, t_{n, \rho}]} (1 - \lambda) \frac{U_{n_0}(t) - t}{1 - t} - \beta_{n, \alpha_n} \frac{\delta(t)}{1 - t} > \frac{\epsilon}{3} \lambda\right) \quad (34)
\end{aligned}$$

$$+ P\left(\sup_{t \in (0, t_{n, \rho}]} \lambda \frac{G_{n_0}(t) - G(t)}{1 - t} > \frac{\epsilon}{3} \lambda\right) \quad (35)$$

$$+ 1\left\{\sup_{t \in (0, t_{n, \rho}]} \lambda \frac{G(t) - t}{1 - t} > \frac{\epsilon}{3} \lambda\right\} \quad (36)$$

As already argued above, the probability on the right hand side of (34) vanishes for  $n \rightarrow \infty$ . The probability (35) vanishes clearly likewise and it remains to show that (36) vanishes as well for  $n \rightarrow \infty$ . As  $t_{n, \rho} \rightarrow 0$ , it holds that  $(1 - t)^{-1} \leq 2$  for  $t \in (0, t_{n, \rho}]$  and large enough value of  $n$ . The term (36) vanishes hence if  $G(t_{n, \rho}) < \epsilon/6$ . This is equivalent to  $\log G^{-1}(\frac{\epsilon}{6}) < -(\tau + \rho) \log n$  and the claim follows from property (16).