A Large-sample Approach to Controlling False Discovery Rates

Christopher Genovese[1] and Larry Wasserman[2]
Carnegie Mellon University

This paper extends the theory of false discovery rates (FDR) pioneered by Benjamini and Hochberg (1995). We give statistical models underlying multiple testing in the independent, continuous case. We define the realized FDR and False Nondiscovery Rate (FNR) as stochastic processes and characterize their asymptotic behavior. We develop methods for estimating the p-value distribution, even in the non-identifiable case. We also develop a new method for controlling the probability of a large realized FDR.

KEYWORDS: Multiple Testing, p-values, False Discovery Rate, Bootstrap

[1]Research supported by NSF Grant SES 9866147.

[2]Research supported by NIH Grants R01-CA54852-07 and MH57881 and NSF Grants DMS-98-03433 and DMS-0104016.

Contents

Notation Index

The following summarizes the most common recurring notation and indicates where each symbol is defined.

Symbol	Description	Section	Page
m	Total number of tests performed	2.1	5
P^m	Vector of p-values $(P_1, \ldots, P_m)$	2.2	6
H^m	Vector of hypothesis indicators $(H_1, \ldots, H_m)$	2.2	6
$P_{(i)}$	The ith smallest p-value; $P_{(0)} \equiv 0$		
M_0	Number of true null hypotheses	2.2	7
M_1	Number of false null hypotheses	2.2	7
a	Probability of a false null (Mixture Model)	2.2	6
F, f	Alternative p-value distribution (CDF,PDF)	2.2	6
$\widehat{F}_m$	Projection estimator of F	3.3	17
$\mathcal{F}$	class of alternative p-value distributions	2.2	7
$\mathcal{F}_S, \mathcal{F}_C, \mathcal{F}_\Theta$	specific classes	2.2	7
G, g	Marginal distribution (CDF, PDF) of the P_is	2.2	6
$\widehat{G}$	Generic Estimator of G	3	12
$\mathbb{G}_m$	Empirical CDF of P^m	3	12
$\widehat{G}_{\text{EDF}}, \widehat{G}_{\text{LCM}}$	Estimators of G	3,3	12,12
U	Uniform CDF	2.2	6
Γ	FDP process	2.5	10
Ξ	FNP process	2.5	10
γ, ξ	FDR, FNR functions	2.5	10
ϵ_m	Dvoretzky-Kiefer-Wolfowitz nghd. radius	3	12
Q	Asymptotic mean of Γ	2.5	10
$\widetilde{Q}$	Asymptotic mean of Ξ	2.5	10

We use $1\{\ldots\}$ and $\mathsf{P}\{\ldots\}$ to denote respectively the indicator and probability of the event $\{\ldots\}$; subscripts on P specify the underlying distributions when necessary. We also use E to denote expectation, and $X_m \rightsquigarrow X$ to denote that X_m converges in distribution to X. We use z_α to denote the upper α-quantile of a standard normal.

1 Introduction

Among the many challenges raised by the analysis of large data sets is the problem of multiple testing. In some settings, it is not unusual to test thousands or even millions of hypotheses. Examples include function magnetic resonance imaging, microarray analysis in genetics, and source detection in astronomy. Traditional methods that provide strong control of familywise error often have low power and can be unduly conservative in some applications.

Benjamini and Hochberg (BH 1995) pioneered an alternative: controlling the False Discovery Rate (FDR), the expected proportion of false rejections among all rejections. BH (1995) provided a distribution-free, finite sample method for choosing a p-value threshold that guarantees that the FDR is less than a target level α. The same paper demonstrated that the BH procedure is often more powerful than traditional methods that control familywise error.

Recently, there has been much further work on FDR. We shall not attempt a complete review here but mention the following. Benjamini and Yekutieli (2001) extended the BH method to a class of dependent tests. Efron, Tibshirani and Storey (2001) developed an empirical Bayes approach to multiple testing and made interesting connections with FDR. They also showed how these methods are useful Storey (2001a,b) connected the FDR concept with a certain Bayesian quantity and proposed a new FDR method, positive FDR, which has higher power than the original BH method. Genovese and Wasserman (2001) showed that, asymptotically, the BH method corresponds to a fixed threshold method that rejects all p-values less than a threshold u^*, and they characterized u^*. They also introduced the False Nondiscovery Rate (FNR) and found the optimal threshold t^* in the sense of minimizing FNR subject to a bound on FDR. The two thresholds are related by $u^* < t^*$, implying that BH is (asymptotically) conservative. Abramovich, Benjamini, Donoho and Johnstone (2000) established a connection between FDR and minimax point estimation. An interesting open question is whether the asymptotic results obtained in this paper can be extended to the sparse regime in the aforementioned paper where the fraction of alternatives tends to zero.

In this paper, we develop some large-sample theory for false discovery rates and present new methods for controlling quantiles of the false discovery distribution. An essential idea is to view the proportion of false discoveries as a stochastic process indexed by the p-value threshold. The problem of choosing a threshold then becomes a problem of controlling a stochastic process.

The main contributions of the paper include the following:

1. Definition and asymptotics for the FDP and FNP processes;

2. Estimators of the p-value distribution, even in the non-identifiable case;

3. Verification of the asymptotic validity of a class of methods for FDR control;

4. New methods, which we call *confidence thresholds*, for controlling quantiles of the false discovery distribution

2 Preliminaries

2.1 Notation

Consider a multiple testing situation in which m tests are being performed. Suppose M_0 of the null hypotheses are true and $M_1 = m - M_0$ null hypotheses are false. We can categorize the m tests in the following 2×2 table on whether each null hypothesis is rejected and whether each null hypothesis is true:

	H_0 Not Rejected	H_0 Rejected	Total
H_0 True	$M_{0\|0}$	$M_{1\|0}$	M_0
H_0 False	$M_{0\|1}$	$M_{1\|1}$	M_1
Total	$m - R$	R	m

We define the False Discovery Proportion (FDP) and the False Nondiscovery Proportion (FNP) by

$$\mathrm{FDP} = \begin{cases} \frac{M_{1|0}}{R} & \text{if } R > 0 \\ 0, & \text{if } R = 0, \end{cases} \tag{1}$$

and

$$\text{FNP} = \begin{cases} \dfrac{M_{0|1}}{m-R} & \text{if } R < m \\ 0 & \text{if } R = m. \end{cases} \tag{2}$$

The first is the proportion of rejections that are incorrect, and the second – the dual quantity – is the proportion of non-rejections that are incorrect. Notice that FDR $= \mathsf{E}\,(\text{FDP})$, and following Genovese and Wasserman (2001), we define FNR $= \mathsf{E}\,(\text{FNP})$. Storey (2002) considers a different definition of FDR, called pFDR for positive FDR, by conditioning on the event that $R > 0$ and discusses the advantages and disadvantages of this definition.

ATTN: Comment on rFDR mentioned by referee

2.2 Models

Let $H_i = 0$ (or 1) if the i^{th} null hypothesis is true (false) and Let P_i denote the i^{th} p-value. Define vectors $P^m = (P_1, \ldots, P_m)$ and $H^m = (H_1, \ldots, H_m)$. Let $P_{(1)} < \cdots < P_{(m)}$ denote the ordered p-values, and define $P_{(0)} \equiv 0$.

In this paper, we use a random effects (or hierarchical) model as in Efron et al (2001). Specifically, we assume the following for $0 \le a \le 1$:

$$\begin{aligned} H_1, \ldots, H_m &\sim \text{Bernoulli}(a) \\ \Xi_1, \ldots, \Xi_m &\sim \mathcal{L}_{\mathcal{F}} \\ P_i | H_i = 0, \Xi_i = \xi_i &\sim \text{Uniform}(0,1) \\ P_i | H_i = 1, \Xi_i = \xi_i &\sim \xi_i \end{aligned}$$

where $\Xi_1, \ldots, \Xi_m$ denote distribution functions and $\mathcal{L}_{\mathcal{F}}$ is an arbitrary probability measure over a class of distribution functions $\mathcal{F}$. It follows that the marginal distribution of the p-values is

$$G = (1-a)U + aF \tag{3}$$

where $U(t)$ denotes the Uniform(0,1) CDF and $F(t) = \int \xi(t) d\mathcal{L}_{\mathcal{F}}(\xi)$. Except where noted, we assume that G is strictly concave.

Remark 2.1. A more common approach in multiple testing is to use a conditional model in which $H_1, \ldots, H_m$ are fixed, unknown binary values.

The results in this paper can be cast in a conditional framework but we find the random effects framework to be more convenient.

The distribution F is assumed to belong to some set of distributions $\mathcal{F}$. Examples of choices for $\mathcal{F}$ include the following:

$$\begin{aligned}
\mathcal{F}_S &= \{\text{CDF } F : F \geq U \text{ and } F \text{ absolutely continuous}\} \\
\mathcal{F}_C &= \{F \in \mathcal{F}_S : F \text{ is concave}\} \\
\mathcal{F}_\Theta &= \{F_\theta : \theta \in \Theta\},
\end{aligned}$$

where Θ is a finite-dimensional parameter space. Here, $\mathcal{F}_S$ is the set of p-value distributions that are stochastically smaller than the Uniform(0,1). Except where otherwise stated, we take $\mathcal{F} = \mathcal{F}_C$, although many of the results directly generalize to $\mathcal{F}_S$. In all cases, we assume that F is absolutely continuous, and we denote its density by f. It follows that G is absolutely continuous, and we denote its density by g.

Define $M_0 = \sum_i (1 - H_i)$ and $M_1 = \sum_i H_i$. Under the mixture model, $M_0 \sim \text{Binomial}(\text{m}, 1 - \text{a})$ and $M_1 = m - M_0$.

2.3 The Benjamini-Hochberg and Plug-in Methods

The Benjamini-Hochberg (BH) procedure is a distribution free method for choosing which null hypotheses to reject while guaranteeing that $\text{FDR} \leq \alpha$ for some pre-selected level α. The procedure rejects all null hypotheses for which $P_i \leq P_{(R_{\text{BH}})}$, where

$$R_{\text{BH}} = \max\left\{0 \leq i \leq m : P_{(i)} \leq \alpha \frac{i}{m}\right\}. \tag{4}$$

BH (1995) proved that this procedure guarantees

$$\text{FDR} \leq \frac{M_0}{m}\,\alpha \leq \alpha, \tag{5}$$

regardless of how many nulls are true and regardless of the distribution of the p-values under the alternatives. (Under the mixture model, the guarantee is that $\text{FDR} \leq (1 - a)\alpha$.)

Genovese and Wasserman (2001) showed that, asymptotically, the BH procedure corresponds to rejecting the null when the p-value is less than u^* where u^* is the solution to the equation $F(u) = \beta u$ and $\beta = (1 - \alpha + \alpha M_1/m)/(\alpha M_1/m)$. Here, F is the (common) distribution of the p-values under the alternative. This u^* satisfies $\alpha/m \leq u^* \leq \alpha$ for large m, which shows that the BH method is intermediate between Bonferroni (corresponding to α/m) and uncorrected testing (corresponding to α). They also showed that u^* is strictly less than the optimal p-value cutoff.

Storey (2002) found an estimator $\widehat{\mathrm{FDR}}(t)$ of FDR for fixed t. One can then define a threshold T by finding the largest t such that $\widehat{\mathrm{FDR}}(t) \leq \alpha$. Indeed, this is suggested in equation (13) of Storey (2002), although he does not explicitly advocate this as a formal procedure. It remains an open question whether $\mathrm{FDR}(T) \leq \alpha$. We address this question asymptotically in Section xxxx.

The threshold T chosen this way can also be viewed as a plug-in estimator. Let $\mathbb{G}_m$ be the empirical CDF of P^m. Since ties occur with probability zero, Storey (2002) showed that the BH threshold is equivalent to

$$T_{\mathrm{BH}}(P^m) = \sup\left\{t : \mathbb{G}_m(t) = \frac{t}{\alpha}\right\} \tag{6}$$

This can be viewed as a plug-in estimator of u^*, where

$$u^*(a, G) \quad = \quad \sup\left\{t : G(t) = \frac{t}{\alpha}\right\}. \tag{7}$$

The BH method guarantees that $\mathrm{FDR} = (1-a)\alpha$ in the continuous case, which is conservative because FDR is controlled at below the target level α. This suggests replacing α in u^* by $\alpha/(1-a)$ and suggests therefore that a better threshold would be

$$t^*(a, G) \quad = \quad \sup\left\{t : G(t) = \frac{(1-a)t}{\alpha}\right\} \tag{8}$$

where $\beta - 1/\alpha$ simplifies to $(1-a)(1-\alpha)/a\alpha$. The plug-in estimator for this threshold is of the form

$$\begin{aligned} T_{\mathrm{PI}}(P^m) \quad &= \quad \sup\left\{t : \mathbb{G}_m(t) = \frac{(1-\widehat{a})t}{\alpha}\right\} & (9) \\ &= \quad \sup\left\{t : \frac{(1-\widehat{a})t}{\mathbb{G}_m(t)} = \alpha\right\}. & (10) \end{aligned}$$

The last expression can also be motivated by the observation that, up to an exponentially small term in m, FDR at a fixed threshold t equals $(1 - a)t/G(t)$. (See Lemma 2.1.) The plug-in procedure is mathematically equivalent to the proposal in Storey (2002): estimate FDR at each threshold t and choose the biggest t for which this estimate is less than or equal to α. Storey (2002) suggested estimating a by $\max(0, (\widehat{G}(t_0) - t_0)/(1 - t_0))$ for a fixed $0 < t_0 < 1$. We describe alternate estimators of a in Section 3.2. Storey (2002) provided simulations to show that the plug-in procedure has good power but did not provide a proof that it controls FDR at level α. We settle this question in Section 5 where we show that under weak conditions on $\widehat{a}$ the procedure asymptotically controls FDR at level α.

2.4 Multiple Testing Procedures

A *multiple testing procedure* T is a mapping taking $[0, 1]^m$ into $[0, 1]$, where it is understood that the null hypotheses corresponding to all p-values less than $T(P^m)$ are rejected. We often call T the *threshold*.

Let $\alpha, t \in [0, 1]$ and $0 \le r \le m$, and recall that $P_{(0)} \equiv 0$. Let $\widehat{G}$ and $\widehat{g}$ be generic estimates of G and $g = G'$, respectively.

Examples of multiple testing procedures include the following:

Uncorrected testing	$T_{\mathrm{U}}(P^m) = \alpha$
Bonferroni	$T_{\mathrm{B}}(P^m) = \alpha/m$
Fixed threshold at t	$T_t(P^m) = t$
Benjamini-Hochberg	$T_{\mathrm{BH}}(P^m) = \sup\{t : \mathbb{G}_m(t) = t/\alpha\} = P_{(R_{\mathrm{BH}})}$
Oracle	$T_{\mathrm{O}}(P^m) = \sup\{t : G(t) = (1 - a)t/\alpha\}$
Plug In	$T_{\mathrm{PI}}(P^m) = \sup\{t : \widehat{G}(t) = (1 - \widehat{a})t/\alpha\}$
First r	$T_{(r)}(P^m) = P_{(r)}$
Bayes' Classifier	$T_{\mathrm{BC}}(P^m) = \sup\{t : \widehat{g}(t) > 1\}$
Regression Classifier	$T_{\mathrm{Reg}}(P^m) = \sup\{t : \widehat{\mathsf{P}}\{H_1 = 1 \mid P_1 = t\} > 1/2\}$

2.5 FDP and FNP as Stochastic Processes

An important idea that we use throughout the paper is that the FDP, regarded as a function of the threshold t, is a stochastic process. This observation is crucial for studying the properties of procedures.

Define the *FDP process*

$$\Gamma(t) \equiv \Gamma(t, P^m, H^m) = \frac{\sum_i 1\{P_i \le t\}\ (1 - H_i)}{\sum_i 1\{P_i \le t\} + 1\{\text{all } P_i > t\}}, \tag{11}$$

where the last term in the denominator makes $\Gamma = 0$ when no p-values are below threshold. Also define the *FNP process*

$$\Xi(t) \equiv \Xi(t, P^m, H^m) = \frac{\sum_i 1\{P_i > t\}\ H_i}{\sum_i 1\{P_i > t\} + 1\{\text{all } P_i \le t\}}. \tag{12}$$

The *FDP* and *FNP* of a procedure T are $\Gamma(T) \equiv \Gamma(T(P^m), P^m, H^m)$ and $\Xi(T) \equiv \Xi(T(P^m), P^m, H^m)$. For brevity, we sometimes write Γ and Ξ for $\Gamma(T)$ and $\Xi(T)$. Define $\gamma(t) = \mathsf{E}\,\Gamma(t)$ and $\xi(t) = \mathsf{E}\,\Xi(t)$ to be the FDR and FNR at the fixed threshold t. For convenience, we also define

$$Q(t) = (1 - a)\,\frac{t}{G(t)} \tag{13}$$

$$\widetilde{Q}(t) = a\,\frac{1 - F(t)}{1 - G(t)}. \tag{14}$$

The following lemma is a corollary of Theorem 1 in Storey (2002). We provide a proof to make this connection explicit.

Lemma 2.1. *Under the mixture model, for $t > 0$,*

$$\begin{aligned} \gamma(t) &= Q(t)\ (1 - (1 - G(t))^m) \\ \xi(t) &= \widetilde{Q}(t)\ (1 - G(t)^m) \end{aligned}$$

The second terms on the right-hand side of both equations differ from 1 by an exponentially small quantity.

Proof. Let $I^m = (I_1, \ldots, I_m)$ where $I_i = 1\{P_i \le t\}$. Note that if $i \ne j$, then H_i is independent of I_j given I^m. From Bayes' theorem,

$$\begin{aligned} \mathsf{P}\{1 - H_i = 1 \mid I^m\} &= \mathsf{P}\{H_i = 0 \mid I^m\} \\ &= \frac{\mathsf{P}\{H_i = 0\}\,\mathsf{P}\{P_i \le t \mid H_i = 0\}}{\mathsf{P}\{P_i \le t\}}\, I_i + \end{aligned}$$

$$
\begin{aligned}
& \frac{\mathsf{P}\{H_i = 0\}\,\mathsf{P}\{P_i > t \mid H_i = 0\}}{\mathsf{P}\{P_i > t\}}\,(1 - I_i) \\
&= \frac{(1-a)t}{G(t)}\,I_i + \frac{(1-a)(1-t)}{1-G(t)}\,(1 - I_i) \\
&= Q(t) I_i + \left(1 - \widetilde{Q}(t)\right)\,(1 - I_i).
\end{aligned}
$$

Thus, $\mathsf{E}\left(I_i(1-H_i) \mid I^m\right) = Q(t) I_i$ and $\mathsf{E}\left((1-I_i)H_i \mid I^m\right) = \widetilde{Q}(t)(1-I_i)$. It follows that

$$
\begin{aligned}
\mathsf{E}\left(\Gamma(t) \mid I^m\right) &= Q(t)\,\frac{\sum_i I_i}{\sum_i I_i + \prod_i (1 - I_i)} = Q(t)\,1\{\text{some } P_i \le t\} \\
\mathsf{E}\left(\Xi(t) \mid I^m\right) &= \widetilde{Q}(t)\,\frac{\sum_i 1 - I_i}{\sum_i 1 - I_i + \prod_i I_i} = \widetilde{Q}(t)\,1\{\text{some } P_i > t\}\,.
\end{aligned}
$$

Hence, taking expectations,

$$
\begin{aligned}
\mathsf{E}\,\Gamma(t) &= Q(t)\,\left(1 - (1 - G(t))^m\right) \\
\mathsf{E}\,\Xi(t) &= \widetilde{Q}(t)\,\left(1 - G(t)^m\right),
\end{aligned}
$$

which proves the claim. □

REMARK 2.2. Storey (2001) shows that $Q(t)$ equals the conditional expected value of $\Gamma(t)$ given that at least one null is rejected, that is, the expected pFDR. He also discusses interesting implications of this conditioning.

One of the essential difficulties in studying a procedure T is that $\Gamma(T)$ is the evaluation of the stochastic process $\Gamma(\cdot)$ at a random variable T. Both depend on the observed data, and in general they are correlated. In particular, if $\widehat{Q}(t)$ estimates FDR at a each fixed t and $T = \sup\{t :\ \widehat{Q}(t) \le \alpha\}$, it *does not* follow that $\mathsf{E}\,\Gamma(T) \le \alpha$. Indeed, one might say that we have replaced the original simultaneous inference problem involving many p-values with a new simultaneous inference problem resulting from choosing among all possible thresholds. The stochastic process point of view provides a suitable framework for addressing this problem.

3 Estimating the P-value Distribution

Recall that, under the mixture model, $P_1, \ldots, P_m$ have CDF $G(t) = (1-a)\,t + a\,F(t)$. Let $\mathbb{G}_m$ denote the empirical CDF of P^m. When G is assumed to be concave, a better estimate of G is the least concave majorant (LCM) $\mathbb{G}_{\text{LCM},m}$ defined to be the infimum of the set of all concave CDF's lying above $\mathbb{G}_m$. Most p-value densities in practical problems are decreasing in p which implies that G is concave. We will use the Dvoretzky-Kiefer-Wolfowitz (DKW) inequality: for any $x > 0$,

$$\mathsf{P}\{||\mathbb{G}_m(t) - G(t)||_\infty > x\} \leq 2e^{-2mx^2}$$

where $||F - G||_\infty = \sup_{0 \leq t \leq 1} |F(t) - G(t)|$. Since $G \geq U$, we replace the empirical CDF with $\widehat{G}_{\text{EDF},m}(t) = \max\{\mathbb{G}_m(t), t\}$. The DKW inequality still holds for this estimator. In the concave case, we define $\widehat{G}_{\text{LCM},m}$ analogously using $\mathbb{G}_{\text{LCM},m}$. We typically drop the subscript m when denoting these estimators and use $\widehat{G}_{\text{EDF}}$ and $\widehat{G}_{\text{LCM}}$. We use $\widehat{G}$ to denote the chosen estimator of G. Unless otherwise specified, $\widehat{G} = \widehat{G}_{\text{EDF}}$ if $\mathcal{F} = \mathcal{F}_S$ and $\widehat{G} = \widehat{G}_{\text{LCM}}$ if $\mathcal{F} = \mathcal{F}_C$. Once we obtain estimates $\widehat{a}$ and $\widehat{G}$, we define $\widehat{Q}(t) = (1 - \widehat{a})/\widehat{G}(t)$.

Define $\epsilon_k(\beta)$ to be the DKW bound on the quantile of $||\mathbb{G}_k(t) - G(t)||_\infty$:

$$\epsilon_k(\beta) = \sqrt{\frac{1}{2k} \log\left(\frac{2}{\beta}\right)}.$$

The leading special case is when $k = m$ and $\beta = \alpha$; in this case, we let ϵ_m denote $\epsilon_m(\alpha)$.

3.1 Identifiability and Purity

Before discussing the estimation of a and F, it is helpful to first discuss their identifiability. For example, if a is not identifiable, there is no guarantee that the estimate used in the plug-in method will give good performance. The results in the ensuing subsections show that despite not being completely identified, it is possible to make sensible inferences about a and F.

Say that F is *pure* if $\operatorname{ess\,inf}_t f(t) = 0$ where f is the density of F. Let $\mathcal{O}_F$ be the set of pairs (b, H) such that $b \in [0,1]$, $H \in \mathcal{F}$ and $F = (1-b)U + bH$. F is identifiable if $\mathcal{O}_F = \{(1, F)\}$.

Define

$$\begin{aligned}
\zeta_F &= \inf\{b :\ (b,H) \in \mathcal{O}_F\}, \\
\underline{F} &= \frac{F - (1-\zeta_F)U}{\zeta_F}, \\
\underline{a}_F &= a\,\zeta_F.
\end{aligned}$$

We will often drop the subscript F on $\underline{a}_F$ and ζ_F. Note that G can be decomposed as

$$\begin{aligned}
G &= (1-a)U + a\,F \\
&= (1-a)U + a[(1-\zeta)U + \zeta\underline{F}] \\
&= (1-a\zeta)U + a\,\zeta\underline{F} \\
&= (1-\underline{a})U + \underline{a}\,\underline{F}.
\end{aligned}$$

Purity implies identifiability but not vice versa. Consider the following example. Let $\mathcal{F}$ be the Normal $(\theta,1)$ family and consider testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$. The density of the p-value is

$$f_\theta(p) = \frac{1}{2} e^{-n\theta^2/2}\left[e^{-\sqrt{n}\theta\Phi^{-1}(1-p/2)} + e^{\sqrt{n}\theta\Phi^{-1}(1-p/2)}\right].$$

Now, $f_\theta(1) = ae^{-n\theta^2/2} > 0$ so this test is impure. However, the parametric assumption makes a and θ identifiable when the null is false. It is worth noting that $f_\theta(1)$ is exponentially small in n. Hence, the difference between a and $\underline{a}$ is small. Even when X has a t-distribution with ν degrees of freedom $f_\theta(1) = (1+n\theta^2/\nu)^{-(\nu+1)/2}$. Thus, in practical cases, $a-\underline{a}$ will be quite small. On the other hand, one sided tests for continuous exponential families are pure and identifiable.

The problem of estimating a and F has been considered by Efron et al (2001) and Storey (2001b) who also discuss the identifiability issue. In particular, Storey notes that $G(t) = (1-a)t + aF(t) \leq (1-a)t + a$ for all t. It then follows that

$$a \geq \underline{a} \geq \max_t \frac{G(t)-t}{1-t}.$$

Thus a lower bound on a is $\frac{G(t_0)-t_0}{1-t_0}$ for any t_0 and this lower bound is estimable. The following result gives precise information about the best bounds that are possible.

PROPOSITION 3.1. *If $F \in \mathcal{F}_S$ then*

$$\zeta \geq 1 - \inf_t F'(t) \quad \textit{and} \quad \underline{a} \geq 1 - \inf_t G'(t).$$

If $\mathcal{F} = \mathcal{F}_S$ then these inequalities become equalities. If F is concave then the infima are achieved at $t = 1$. For any $b \in [\zeta, 1]$ we can write $G = (1 - ab)U + abF_b$ where $F_b = (F - (1 - b)U)/b$ is a CDF *and $F \leq F_b$.*

3.2 Estimating a

We begin with a uniform confidence interval for $\underline{a}$.

THEOREM 3.1. *Let*

$$a_* = \max_t \frac{\widehat{G}(t) - t - \epsilon_m}{1 - t}.$$

Then $[a_*, 1]$ is a $1 - \alpha$ confidence interval for $\underline{a}$. In fact,

$$\inf_{a,F} \mathsf{P}_{a,F}\{\underline{a} \in [a_*, 1]\} \geq 1 - \alpha,$$

and for each (a, F) pair

$$\mathsf{P}_{a,F}\{\underline{a} \in [a_*, 1]\} \leq 1 - \alpha + \sum_{j=2}^{\infty} (-1)^j \frac{\alpha^{j^2}}{2^{j^2-1}} + O\left(\frac{(\log m)^2}{\sqrt{m}}\right),$$

where the remainder term may depend on a and F. Because $a \geq \underline{a}$, $[a_*, 1]$ is a valid, finite-sample $1 - \alpha$ confidence interval for a as well.

PROOF. The left-hand inequality follows immediately from DKW because $G(t) \geq \widehat{G}(t) - \epsilon_m$ for all t with probability at least $1 - \alpha$. The sum on the right-hand side follows from the closed-form limiting distribution of the Kolmogorov-Smirnov statistic, and the order of the error follows from the Hungarian embedding. To see this, note that

$$\underline{a} < a_* \implies \underline{a}\sqrt{m} < \max_t \sqrt{m}\frac{\mathbb{G}_m(t) - G(t)}{1 - t} + \sqrt{m}\frac{G(t) - t}{1 - t} - \frac{\epsilon_m \sqrt{m}}{1 - t}$$

$$
\begin{aligned}
&\implies \quad \underline{a}\sqrt{m} < \max_t \sqrt{m}\frac{\mathbb{G}_m(t) - G(t)}{1-t} + \sqrt{m}\underline{a} - \frac{\epsilon_m\sqrt{m}}{1-t} \\
&\implies \quad 0 < \max_t \sqrt{m}\frac{\mathbb{G}_m(t) - G(t)}{1-t} - \frac{\epsilon_m\sqrt{m}}{1-t} \\
&\implies \quad 0 < \max_t \sqrt{m}\,(\mathbb{G}_m(t) - G(t)) - \epsilon_m\sqrt{m} \\
&\implies \quad \|\sqrt{m}\,(\mathbb{G}_m(t) - G(t))\,\|_\infty > \epsilon_m\sqrt{m}.
\end{aligned}
$$

Hence,

$$
\mathsf{P}\{\underline{a} < a_*\} \le \mathsf{P}\Big\{\,\|\sqrt{m}\,(\mathbb{G}(t) - G(t))\,\|_\infty > \epsilon_m\sqrt{m}\Big\}\,. \tag{15}
$$

Next, apply the Hungarian embedding (van der Vaart 1998, p. 269)

$$
\limsup_{n\to\infty} \frac{\sqrt{n}}{(\log n)^2}\,\|\sqrt{m}\,(\mathbb{G}_m - G) - \widetilde{\mathbb{G}}_m\|_\infty < \infty \text{ a.s.,}
$$

for a sequence of Brownian bridges $\widetilde{\mathbb{G}}_m$. Then, using the large-sample distribution of the Kolmogorov-Smirnov statistic,

$$
\mathsf{P}\Big\{\,\|\widetilde{\mathbb{G}}\|_\infty > x\Big\} = 2\sum_{j=1}^{\infty}(-1)^{j+1}e^{-2j^2x^2},
$$

for a generic Brownian bridge $\widetilde{\mathbb{G}}$. Taking $x = \epsilon_m$, the probability of the complement of this event gives the formula stated in the theorem.

In the concave case, the LCM can be substituted for $\widehat{G}$ and the result still holds since, by Marshall's lemma, $\|\widehat{G}_{\text{LCM},m} - G\|_\infty \le \|\widehat{G}_m - G\|_\infty$. $\square$

PROPOSITION 3.2 (STOREY'S ESTIMATOR). *Fix $t_0 \in (0,1)$ and let*

$$
\widehat{a} = \left(\frac{\mathbb{G}_m(t_0) - t_0}{1 - t_0}\right)_+ .
$$

If $G(t_0) > t_0$,

$$
\widehat{a} \xrightarrow{P} \frac{G(t_0) - t_0}{1 - t_0} \le \underline{a},
$$

and

$$
\sqrt{m}\left(\widehat{a} - \frac{G(t_0) - t_0}{1 - t_0}\right) \rightsquigarrow N\left(0, \frac{G(t_0)(1 - G(t_0))}{(1 - t_0)^2}\right).
$$

If $G(t_0) = t_0$,

$$\sqrt{m}\widehat{a} \rightsquigarrow \frac{1}{2}\delta_0 + \frac{1}{2}N^+\left(0, \frac{t_0}{1-t_0}\right),$$

where δ_0 *is a point-mass at zero and* N^+ *is a positive-truncated Normal.*

A consistent estimate of $\underline{a}$ is available if we assume weak smoothness conditions on g. We use the spacings estimator of Swanepoel (1999) which is of the form $2r_m/(mV_m)$ where $r_m = m^{4/5}(\log m)^{-2\delta}$ and V_m is a selected spacing in the order statistics of the p-values.

THEOREM 3.2. *Assume that* g'' *is bounded away from 0 and* ∞ *and is Lipschitz of order* $\lambda > 0$ *at the value* t *where* g *achieves its minimum. For every* $\delta > 0$, *there exists an estimator* $\widehat{a}$ *such that*

$$\frac{m^{(2/5)}}{(\log m)^{\delta}}(\widehat{a} - \underline{a}) \rightsquigarrow N(0, (1-\underline{a})^2).$$

PROOF. Let $\widehat{a}$ be the estimator defined in Swanepoel (1999) with $r_m = m^{4/5}(\log m)^{-2\delta}$ and $s_m = m^{4/5}(\log m)^{4\delta}$. The result follows from Swanepoel (1999, Theorem 1.3). □

REMARK 3.1. An alternative estimator is $\widehat{a} = 1 - \min_t \widehat{g}(t)$ where $\widehat{g}$ is a kernel estimator.

Under the assumption that G is concave, a consistent estimator that does not require as much smoothness, $\widehat{a}_{\text{HS}}$, is derived from the Hengartner and Stark (1995) finite-sample confidence envelope $[\gamma^-(\cdot), \gamma^+(\cdot)]$ for a monotone density. We define

$$\widehat{a}_{\text{HS}} = 1 - \min\left\{h(1) : \gamma^- \leq h \leq \gamma^+\right\}.$$

THEOREM 3.3. *If* G *is concave and* $g = G'$ *is Lipschitz of order 1,*

$$\left(\frac{n}{\log n}\right)^{1/3}(\widehat{a}_{\text{HS}} - \underline{a}) \xrightarrow{P} 0,$$

and $[1-\gamma^+(1), 1-\gamma^-(1)]$ *is a* $1-\alpha$ *confidence interval for* $\underline{a}$ *for* $0 \leq \alpha \leq 1$ *and all* m. *Extending the right endpoint to* 1 *yields* $[1-\gamma^+(1), 1]$, *which is a* $1-\alpha$ *confidence interval for* a.

PROOF. Follows from Hengartner and Stark (1995). $\square$

If a is non-identifiable, then the best we can hope is to consistently estimate $\underline{a}$. The estimators of Theorems 3.2 and 3.3 achieve this. We say that an estimator $\widehat{a}$ is *conservatively consistent* under (a, F) if $\widehat{a} \xrightarrow{P} b < \underline{a}_F$. Storey's estimator is conservatively consistent. With our approach to estimating F in the next section, it is not sufficient to use a conservatively consistent estimator of a.

3.3 Estimating F

In the bootstrap methods that we introduce in section 6, we will need to estimate F. There are many possible methods; we consider here projection estimators defined by

$$\widehat{F}_m = \arg\min_{H\in\mathcal{F}} ||\widehat{G} - (1-\widehat{a})U - \widehat{a}H||_\infty, \tag{16}$$

where $\widehat{a}$ is an estimate of a. The appendix gives an algorithm to find $\widehat{F}_m$.

It is helpful to consider first the case where a is known, and here we substitute a for $\widehat{a}$ in the definition of $\widehat{F}_m$.

THEOREM 3.4. *Let*

$$\widehat{F}_m = \arg\min_{H\in\mathcal{F}} ||\widehat{G} - (1-a)U - aH||_\infty.$$

Then,

$$||F - \widehat{F}_m||_\infty \leq \frac{2||G - \widehat{G}||_\infty}{a} \xrightarrow{a.s.} 0.$$

PROOF.

$$\begin{aligned}
a||F - \widehat{F}_m||_\infty &= ||aF - a\widehat{F}_m||_\infty \\
&= ||(1-a)U + aF - (1-a)U - a\widehat{F}_m||_\infty \\
&= ||G - (1-a)U - a\widehat{F}_m||_\infty \\
&= ||G - \widehat{G} + \widehat{G} - (1-a)U - a\widehat{F}_m||_\infty \\
&\leq ||\widehat{G} - G||_\infty + ||\widehat{G} - (1-a)U - a\widehat{F}_m||_\infty \\
&\leq ||\widehat{G} - G||_\infty + ||\widehat{G} - (1-a)U - aF||_\infty \\
&= ||\widehat{G} - G||_\infty + ||\widehat{G} - G||_\infty.
\end{aligned}$$

The last statement follows from the uniform consistency of $\widehat{G}$. $\square$

When a is unknown, the projection estimator $\widehat{F}$ is consistent whenever we have a consistent estimator of $\underline{a}$. Recall that in the identifiable case, $a = \underline{a}$ and $F = \underline{F}$.

THEOREM 3.5. *Let $\widehat{a}$ be a consistent estimator of $\underline{a}$. Then,*

$$||\widehat{F}_m - \underline{F}||_\infty \leq \frac{||\widehat{G} - G||_\infty + |\widehat{a} - \underline{a}|}{\underline{a}} \xrightarrow{P} 0.$$

PROOF. Let $\delta_m = ||\widehat{G} - (1-\widehat{a})U - \widehat{a}\widehat{F}||_\infty$. Since $\widehat{F}$ is the minimizer,

$$\begin{aligned}
\delta_m &\leq ||\widehat{G} - (1-\widehat{a})U - \widehat{a}\underline{F}||_\infty \\
&= ||\widehat{G} - G + (\widehat{a} - \underline{a})U - (\widehat{a} - \underline{a})\underline{F}||_\infty \\
&\leq ||\widehat{G} - G||_\infty + |\widehat{a} - \underline{a}| \\
&\xrightarrow{P} 0.
\end{aligned}$$

We also have that

$$\delta_m \geq \Big| ||\widehat{G} - (1-\widehat{a})U - \widehat{a}\underline{F}||_\infty - \widehat{a}||\underline{F} - \widehat{F}||_\infty \Big|.$$

Since δ_m and $||\widehat{G} - (1-\widehat{a})U - \widehat{a}\underline{F}||_\infty \xrightarrow{P} 0$ by the above and $\widehat{a} \xrightarrow{P} \underline{a}$, it follows that $||\underline{F} - \widehat{F}||_\infty \xrightarrow{P} 0$. Moreover,

$$||\underline{F} - \widehat{F}||_\infty \leq \frac{||\widehat{G} - G||_\infty + |\widehat{a} - \underline{a}|}{\underline{a}}.$$

$\square$

4 Limiting Distributions

In this section we discuss the limiting distribution of Γ and $\widehat{Q}$. Let

$$\Lambda_0(t) = \frac{1}{m}\sum_{i=1}^{m}(1 - H_i)1\{P_i \leq t\} \quad \text{and} \quad \Lambda_1(t) = \frac{1}{m}\sum_{i=1}^{m} H_i 1\{P_i \leq t\}.$$

and, for each $c \in (0,1)$ define

$$\Omega_c(t) = (1-c)\Lambda_0(t) - c\Lambda_1(t) = \frac{1}{m}\sum_i D_i(t)$$

where $D_i(t) = 1\{P_i \le t\}\,(1 - H_i - c)$. Let

$$\mu_c(t) = \mathsf{E}\, D_1(t) = (1-a)t - cG(t).$$

Let $H_1^* \ldots, H_m^* \sim \text{Bernoulli}(a)$ and let $P_i^* \mid H_i^* \sim \widehat{F}_m$ where $\widehat{F}$ is the projection estimator of F defined in 3.3. Let Λ_0^* and Λ_1^* analogously to Λ_0 and Λ_1 but replacing H_i and P_i by H_i^* and P_i^*. Define $\Omega_c^*(t) = (1-c)\Lambda_0^*(t) - c\Lambda_1^*(t)$.

Let (W_0, W_1) be a continuous, two-dimensional, mean zero Gaussian process with covariance kernel $R_{ij}(s,t) = \text{Cov}(W_i(s), W_j(t))$ given by

$$R(s,t) = \begin{bmatrix} (1-a)(s\wedge t) - (1-a)^2 st & -(1-a)s\, aF(t) \\ -(1-a)t\, aF(s) & aF(s\wedge t) - a^2F(s)F(t) \end{bmatrix}. \tag{17}$$

Theorem 4.1. *Let W be a continuous, mean zero Gaussian process with covariance*

$$\begin{aligned} K_\Omega(s,t) &= (1-a)(1-c)\left[(1-c)(s\wedge t - (1-a)st) + ac(tF(s) + sF(t))\right] + \\ &\qquad ac\left[cF(s\wedge t) - acF(s)F(t)\right]. \end{aligned} \tag{18}$$

Then

$$\sqrt{m}(\Omega_c - \mu_c) \rightsquigarrow W.$$

Also,

$$\sqrt{m}(\Omega_c^* - \widehat{\mu}_c) \rightsquigarrow W$$

conditionally, along almost all sequences $P_1, P_2, \ldots$. Furthermore,

$$\sup_{0\le t\le 1}\sup_u \left|\mathsf{P}_G\left\{\sqrt{m}(\Omega_c(t) - \mu_c(t)) \le u\right\} - \mathsf{P}_{\widehat{G}}\left\{\sqrt{m}(\Omega_c^*(t) - \widehat{\mu}_c(t)) \le u\right\}\right| = O(m^{-1/2})$$

almost surely.

Proof. Let

$$Z_m(t) = \sqrt{m}(\Omega_c(t) - \mu_c(t)) \quad \text{and} \quad Z_m^*(t) = \sqrt{m}(\Omega_c^*(t) - \widehat{\mu}_c(t))$$

for $t \in [0,1]$. Let

$$(W_{m,0}(t), W_{m,1}(t)) \equiv (\sqrt{m}(\Lambda_0(t) - (1-a)t),\ \sqrt{m}(\Lambda_1(t) - aF(t))).$$

By standard empirical process theory, $(W_{m,0}(t), W_{m,1}(t))$ converges to (W_0, W_1). The covariance kernel R stated in equation (17) follows by direct calculation. The result for Ω_c is immediate since Ω_c is a linear combination of Λ_0 and Λ_1.

The argument for $Z_m^*(t)$ with $\widehat{G}$, $\widehat{\mu}_c$ and $\widehat{F}_m$ replacing G, μ and F and the last statement in the theorem follow from van der Vaart (1998) Theorem 23.7. This leads to the conclusion that

$$\sqrt{m}(\Omega_c^* - \widehat{\mu}_c) \rightsquigarrow W$$

where W is a mean zero, Gaussian process with covariance $K_\Omega(s,t)$. $\square$

THEOREM 4.2 (LIMITING DISTRIBUTION OF FDP PROCESS). *For $t \in [\delta, 1]$ for any $\delta > 0$, let*

$$Z_m(t) = \sqrt{m}\ (\Gamma_m(t) - Q(t))\,.$$

Let Z be a Gaussian process on $(0,1]$ with mean 0 and covariance kernel

$$K_\Gamma(s,t) = a(1-a)\,\frac{(1-a)stF(s\wedge t)\ +\ aF(s)F(t)(s\wedge t)}{G^2(s)\,G^2(t)}.$$

Then $Z_m \rightsquigarrow Z$ on $[\delta, 1]$.

REMARK 4.1. The reason for restricting the theorem to $[\delta, 1]$ is that the variance of the process is infinite at zero.

PROOF. Note that $\Gamma_m(t) = \Lambda_0(t)/(\Lambda_0(t) + \Lambda_1(t)) \equiv r(\Lambda_0, \Lambda_1)$ where Λ_0 and Λ_1 are defined as before, $r(\cdot,\cdot)$ maps $\ell^\infty \times \ell^\infty \to \ell^\infty$ where ℓ^∞ is the set of bounded functions on $(\delta, 1]$ endowed with the sup norm. Note that $r((1-a)U, aF) = Q$. It can be verified that $r(\cdot,\cdot)$ is Fréchet differentiable at $((1-a)U, aF)$ with derivative

$$r'_{((1-a)U,aF)}(V) = \frac{aFV_0 - (1-a)UV_1}{G^2}$$

where $U(t) = t$, $V = (V_0, V_1)$. Hence, by the functional delta method (van der Vaart 1998, Theorem 20.8),

$$Z_m \rightsquigarrow r'_{((1-a)U, aF)}(W) = \frac{aFW_0 - (1-a)UW_1}{G^2},$$

where (W_0, W_1) is the process defined just before equation (17). The covariance kernel of the latter expression is $K_\Gamma(s,t)$. $\square$

Remark 4.2. A Gaussian limiting process can be obtained for FNP (i.e., $\Xi(t)$) along similar lines.

The next two theorems follow from the previous results with an application of the functional delta method.

Theorem 4.3. *For any $\delta > 0$,*

$$\sqrt{m}(\widehat{Q}(t) - Q(t)) \rightsquigarrow W$$

on $[\delta, 1]$, where W is a mean 0 Gaussian process on $(0,1]$ with covariance kernel

$$K_Q(s,t) = Q(s)\, Q(t)\, \frac{G(s \wedge t) - G(s)G(t)}{G(s)\, G(t)}.$$

Theorem 4.4. *We have*

$$\sqrt{m}(\widehat{Q}^{-1}(v) - Q^{-1}(v)) \rightsquigarrow W$$

where W is a mean 0 Gaussian process with covariance kernel

$$\begin{aligned} K_{Q^{-1}}(u,v) &= \frac{K_Q(s,t)}{Q'(s)Q'(t)} \\ &= (1-a)^2\, u\, v\, \frac{G(s \wedge t) - G(s)G(t)}{[1-a-ug(s)]\ [1-a-vg(t)]} \end{aligned}$$

with $s = Q^{-1}(u)$ and $t = Q^{-1}(v)$.

5 Asymptotic Validity of Plug-in Procedures

Let $\widehat{Q}^{-1}(c) = \sup\{0 \le t \le 1 : \widehat{Q}(t) \le c\}$. Then, the plug-in threshold T_{PI} defined earlier can be written $T_{\text{PI}}(P^m) = \widehat{Q}^{-1}(\alpha)$. Here we establish the asymptotic validity of T_{PI} in the sense that $\mathsf{E}\,\Gamma(T) = \alpha + O(m^{-1/2})$.

We first tackle the case where a is known. Define $\widehat{Q}_a(t) = (1-a)t/\widehat{G}(t)$ to be the estimator of Q when a is known.

Theorem 5.1. *Assume that a is known and let $\widehat{Q} = \widehat{Q}_a$. Let $t_0 = Q^{-1}(\alpha)$ and assume $G \neq U$. Then,*

$$\begin{aligned}
\sqrt{m}(T_{\text{PI}} - t_0) &\rightsquigarrow N(0, K_{Q^{-1}}(t_0, t_0)) \\
\sqrt{m}(Q(T_{\text{PI}}) - \alpha) &\rightsquigarrow N(0, (Q'(t_0))^2 K_{Q^{-1}}(t_0, t_0)),
\end{aligned}$$

and

$$\mathsf{E}\,\Gamma(T_{\text{PI}}) = \alpha + o(1).$$

Proof.
The first two statements follow from Theorem 4.4 and the delta method.
For the last claim, let $0 < \delta < t_0$ and note the following.

$$\begin{aligned}
|\Gamma_m(T) - \alpha| &\le |\Gamma_m(T) - Q(T)| + |Q(T) - \alpha| \\
&\le \sup_t |\Gamma_m(t) - Q(t)| 1\{T < \delta\} + \\
&\quad \sup_t |\Gamma_m(t) - Q(t)| 1\{T \ge \delta\} + |Q(T) - \alpha| \\
&\le 1\{T < \delta\} + \sup_{t \ge \delta} |\Gamma_m(t) - Q(t)| + |Q(T) - \alpha| \\
&= 1\{T < \delta\} + \frac{1}{\sqrt{m}} \sup_{t \ge \delta} |\sqrt{m}(\Gamma_m(t) - Q(t))| + |Q(T) - \alpha| \\
&= O_P(m^{-1/2}).
\end{aligned}$$

Because $0 \le \Gamma_m \le 1$, the sequence is uniformly integrable, and the result follows. $\square$

Next, we consider the case where a is unknown and possibly non-identifiable.

Theorem 5.2 (Asymptotic Validity of Plug-in Method). *Let $\widehat{a}$ be a consistent estimator of $\underline{a}$. Then,*

$$\mathsf{E}\,\Gamma(T_{\text{PI}}) \leq \alpha + o(1).$$

Proof. Let $\widehat{G} = \widehat{G}_{\text{LCM}}$. By assumption, $\widehat{a} \xrightarrow{P} \underline{a}$. Then,

$$\widehat{Q}(t) = \frac{1-\widehat{a}}{1-\underline{a}} \frac{1-\underline{a}}{1-a} \widehat{Q}_a(t).$$

Let $1+\delta = (1-\underline{a})/(1-a)$.

By Slutzky's Lemma, $\widehat{Q}(t)$ has the same limiting distribution, when properly centered and scaled, as $(1+\delta)\widehat{Q}_a(t)$. Hence,

$$\begin{aligned}
\widehat{Q}^{-1}(\alpha) &= \widehat{Q}_a^{-1}\left(\frac{\alpha}{1+\delta} + o_P(1)\right) \\
&\leq \widehat{Q}_a^{-1}\left(\alpha + o_P(1)\right),
\end{aligned}$$

by the concavity of $\widehat{G}_{\text{LCM}}$.

Because $\widehat{Q}^{-1} \xrightarrow{a.s.} Q_{\underline{a}}^{-1}$ and because $Q_{\underline{a}}^{-1}(\alpha) \leq Q_a^{-1}(\alpha)$, the result follows from the previous theorem. $\square$

Recall that the oracle procedure is defined by $T_O(P^m) = Q^{-1}(\alpha)$. This procedure has the smallest FNR for all procedures that attain FDR $\leq \alpha$ up to sets of exponentially small probability. To see this, note that Q is increasing and $\widetilde{Q}$ is decreasing.

In the non-identifiable case, no data-based method can distinguish a and $\underline{a}$, so the performance of this oracle cannot be attained. We thus define the achievable oracle procedure T_{A0} to be analogous to T_{O} with $(1-\underline{a})t/G(t)$ replacing Q. In the identifiable case, $T_{\text{AO}} = T_{\text{O}}$.

The plug-in procedure that uses the estimator $\widehat{a}$ described in Theorem 3.2 asymptotically attains the performance of T_{AO} in the sense that $\mathsf{E}\,\Gamma(T_{\text{PI}}) = \alpha + o_P(1)$ and $\mathsf{E}\,\Xi(T_{\text{PI}}) = \mathsf{E}\,\Xi(T_{\text{AO}}) + o_P(1)$.

6 Confidence Thresholds

The distribution of the FDP need not be concentrated around its expected value. In practice, controlling the FDR does not offer high confidence that the FDP will be small. As an alternative, we develop methods in this section for controlling the probability of large values for FDP. Given c and α, a $(1-\alpha, c)$ *confidence threshold* is a procedure T such that $\mathsf{P}_G\{\Gamma(T) \leq c\} \geq 1-\alpha$. The remainder of this section shows three approaches to finding confidence thresholds.

6.1 Bootstrap Asymptotic Confidence Thresholds

In this section, we show how to construct a confidence threshold using a bias-corrected bootstrap. As a first guess, one might think of choosing T such that

$$P_{\widehat{G}}\{\Gamma^*(T) \leq c\} = 1-\alpha$$

where Γ^* is constructed from a bootstrap sample (described below). Unfortunately this fails. To see why, note that, for any t, standard bootstrap asymptotics yield

$$P_{\widehat{G}}\left\{\sqrt{m}(\Gamma^*(t) - \widehat{Q}(t)) \leq u\right\} \approx P_G\left\{\sqrt{m}(\Gamma(t) - Q(t)) \leq u\right\}$$

for all u. So, if $P_{\widehat{G}}\{\Gamma^*(t) \leq c\} = 1-\alpha$, then, with $u = \sqrt{m}(c - \widehat{Q}(T))$, we have

$$\begin{aligned}
1-\alpha &= \mathsf{P}_{\widehat{G}}\{\Gamma^*(t) \leq c\} \\
&= \mathsf{P}_{\widehat{G}}\left\{\sqrt{m}(\Gamma^*(t) - \widehat{Q}(t)) \leq u\right\} \\
&\approx \mathsf{P}_G\left\{\sqrt{m}(\Gamma(t) - Q(t)) \leq u\right\} \\
&= \mathsf{P}_G\left\{\Gamma(t) \leq c + (Q(t) - \widehat{Q}(t))\right\} \\
&\neq \mathsf{P}_G\{\Gamma(t) \leq c\}
\end{aligned}$$

because of the bias term $Q(t) - \widehat{Q}(t)$. Thus we will need to correct this bias. Moreover, we must be careful to account for the fact that the actual selected threshold T is a random variable.

6.1.1 Case I: a is Known.

It will be convenient to first solve the problem when a is known as the estimation of a introduces extra complexities. Let $\widehat{F}_m$ be the projection estimator of F defined in Section 3.3.

Bootstrap Confidence Threshold Algorithm Version I.

(Step 1) Draw

$$H_1^* \ldots, H_m^* \sim \text{Bernoulli}(a)$$

and

$$P_i^* \mid H_i^* \sim (1 - H_i^*)U + H_i^* \widehat{F}_m.$$

(Step 2) Define $\Omega_c^*(t) = m^{-1} \sum_i 1\{P_i^* \leq t\} (1 - H_i^* - c)$ and let

$$T = \max\left\{t : \mathsf{P}_{\widehat{G}}\{\Omega_c^*(t) \leq -c\, \epsilon_m(\beta)\} \geq 1 - \beta\right\}$$

where $\beta = \alpha/2$.

Theorem 6.1. *Let T be defined as above. Then*

$$\mathsf{P}_G\{\Gamma(T) \leq c\} \geq 1 - \alpha + O\left(\frac{1}{\sqrt{m}}\right).$$

Proof. Define the following events:

$$\begin{aligned}
A_m^* &= \{\sqrt{m}\,(\Omega_c^*(T) - \widehat{\mu}_c(T)) \leq -\sqrt{m}\,\mu_c(T)\} \\
A_m &= \{\sqrt{m}\,(\Omega_c(T) - \mu_c(T)) \leq -\sqrt{m}\,\mu_c(T)\} \\
B_m &= \{\sqrt{m}\sup_t |\mu_c(t) - \widehat{\mu}_c(t)| \leq \sqrt{m}\,c\epsilon_m(\beta)\} = \{\sqrt{m}c\,\sup_t |G(t) - \widehat{G}(t)| \leq \sqrt{m}\,c\epsilon_m(\beta)\} \\
E_m^* &= \{\sqrt{m}(\Omega_c^*(T) - \widehat{\mu}_c(T)) \leq -\sqrt{m}\mu(T) + \sqrt{m}(\mu(T) - \widehat{\mu}_c(T)) - \sqrt{m}\,c\epsilon_m(\beta)\} \\
&= \{\sqrt{m}(\Omega_c^*(T) - \widehat{\mu}_c(T)) \leq -\sqrt{m}\,c\epsilon_m(\beta) - \sqrt{m}\widehat{\mu}_c(T))\} \\
&= \{\Omega_c^*(T) \leq -c\epsilon_m(\beta)\}.
\end{aligned}$$

First, note that $E_m^* \cap B_m \subset A_m^* \cap B_m$. Hence,

$$\begin{aligned}
\mathsf{P}_G(A_m^*) &\geq \mathsf{P}_G(A_m^* \cap B_m) \geq \mathsf{P}_G(E_m^* \cap B_m) = \mathsf{P}_G(E_m^*) + \mathsf{P}_G(B_m) - \mathsf{P}_G(E_m^* \cup B_m) \\
&\geq \mathsf{P}_G(E_m^*) + \mathsf{P}_G(B_m) - 1.
\end{aligned}$$

The threshold T was chosen so that $\mathsf{P}_{\widehat{G}}(E_m^*) \geq 1-\beta$. By Theorem 4.1, $\mathsf{P}_G(E_m) \geq 1-\beta+O(m^{-1/2})$, $\mathsf{P}_G(A_m^*) = \mathsf{P}_{\widehat{G}}(A_m^*)+O(m^{-1/2})$, and $\mathsf{P}_{\widehat{G}}(A_m^*) = \mathsf{P}_G(A_m)+O(m^{-1/2})$ almost surely. By, the DKW inequality,

$$P\{\sup_t |\widehat{G}(t)-G(t)| > \epsilon_m(\beta)\} \leq \beta$$

which implies that $\mathsf{P}_G(B_m) \geq 1-\beta$. Hence, $\mathsf{P}_G(A_m^*) \geq 1-2\beta = 1-\alpha$.

left off here

It follows that $\mathsf{P}(A_m) \geq 1-2\beta+O(m^{-1/2}) = 1-\alpha+O(m^{-1/2})$. □

6.1.2 Case II: a Unknown but Identifiable

We now modify the bootstrap procedure to incorporate an estimate of a.

Bootstrap Confidence Threshold Algorithm Version II.

(Step 0) Compute $\widehat{a}$, a consistent estimate of a.

(Step 1) Draw $H_1^* \ldots, H_m^* \sim \text{Bernoulli}(\widehat{a})$ and $P_i^* \mid H_i^* \sim (1-H_i^*)U+H_i^*\widehat{F}_m$.

(Step 2) Define $\Omega_c^*(t) = \sum_i 1\{P_i^* \leq t\}\,(1-H_i^*-c)$ and let

$$T = \max\left\{t : \mathsf{P}_{\widehat{G}}\{\Omega_c^*(t) \leq -\delta_m\} \geq 1-\beta\right\}$$

where $\delta_m = c\epsilon_m(\beta)+\rho_m$, $\beta = \alpha/3$, and ρ_m is defined by

$$\mathsf{P}\{|a-\widehat{a}| > \rho_m\} \leq \beta.$$

The quantity ρ_m can, for instance, be obtained from Theorem 3.3.

Theorem 6.2. *Let T be defined as above. Then*

$$\mathsf{P}_G\{\Gamma(T) \leq c\} \geq 1-\alpha+O\left(\frac{1}{\sqrt{m}}\right).$$

Proof. The proof is similar to that for version I except that the bias term is

$$|\widehat{\mu}_c(T)-\mu_c(T)| \leq |a-\widehat{a}|+c|\widehat{G}(T)-G(T)|$$

instead of $c|\widehat{G}(T)-G(T)|$. Also, the bootstrap process has asymptotic covariance kernel

$$\widehat{K}(s,t) = -(1-\widehat{a})(1-c)[\widehat{a}ct\widehat{F}_m(s)+(1-c)((s\wedge t)-(1-\widehat{a})st)].$$

Note that $\widehat{K}(s,t) \xrightarrow{P} K_\Omega(s,t)$. The rest of the proof is similar. □

6.1.3 Case III: a Unknown and Not Identifiable

Let $\widehat{a}$ be a consistent estimator of $\underline{a}$. Version III of the bootstrap is identical to version II except we use this conservative estimator.

THEOREM 6.3. *Let T be defined as in version 3. Then,*

$$\mathsf{P}_G\{\Gamma(T) \leq c\} \geq 1 - \alpha + O\left(\frac{1}{\sqrt{m}}\right).$$

LEMMA 6.1. *For all t and any $0 < c < 1$,*

$$\mathsf{P}_{a,F}\{\Omega_c(t) \leq 0\} \geq \mathsf{P}_{\underline{a},\underline{F}}\{\Omega_c(t) \leq 0\}.$$

For any random T and any $0 < c < 1$,

$$\mathsf{P}_{a,F}\{\Omega_c(T) \leq 0\} \geq \mathsf{P}_{\underline{a},\underline{F}}\{\Omega_c(T) \leq 0\} + O\left(\frac{1}{\sqrt{m}}\right).$$

PROOF. For fixed threshold t, the D_is which are the summands in Ω_c can take only three values: $-c$, 0, and $1-c$. The distribution of D_i under (a, F) is stochastically smaller than the distribution of D_i under $(\underline{a}, \underline{F})$, using the result of Proposition 3.1. The first statement follows.

For the second statement, note that $G = (1-a)U + aF = (1-\underline{a})U + \underline{aF}$ with $\underline{a} \leq a$ and $F \leq \underline{F}$.

The mean $\Omega_c(t)$ under (a, F) is $\mu_c(t) = (1-c)(1-a)t - caF(t)$. The mean $\Omega_c(t)$ under $(\underline{a}, \underline{F})$ is $\widetilde{\mu}_c(t) = (1-c)(1-\underline{a})t - c\underline{aF}(t)$. Thus, $\widetilde{\mu}_c(t) - \mu_c(t) = (a - \underline{a})t > 0$, so $\widetilde{\mu}_c(t) > \mu_c(t)$ for all t.

Let $(P^m, H^m) \sim (a, F)$ and $(\widetilde{P}^m, \widetilde{H}^m) \sim (\underline{a}, \underline{F})$. Define $\Omega_c(\cdot)$ from (P^m, H^m) and $\widetilde{\Omega}_c(\cdot)$ from $(\widetilde{P}^m, \widetilde{H}^m)$, put both processes on a common probability space, and make them independent. Let

$$\Delta(t) = \sqrt{m}\left[(\widetilde{\Omega}_c(t) - \widetilde{\mu}_c(t)) - (\Omega_c(t) - \mu_c(t))\right].$$

Then,

$$
\begin{aligned}
\mathsf{P}\{\Omega_c(T) \le 0\} &= \mathsf{P}\Big\{\sqrt{m}(\Omega_c(T) - \mu_c(T)) \le -\sqrt{m}\,\mu_c(T)\Big\} \\
&\ge \mathsf{P}\Big\{\sqrt{m}(\Omega_c(T) - \mu_c(T)) \le -\sqrt{m}\,\widetilde{\mu}_c(T)\Big\} \\
&= \mathsf{P}\Big\{\sqrt{m}(\widetilde{\Omega}_c(T) - \widetilde{\mu}_c(T)) - \Delta(T) \le -\sqrt{m}\,\widetilde{\mu}_c(T)\Big\} \\
&= \mathsf{P}\Big\{\sqrt{m}(\widetilde{\Omega}_c(T) - \widetilde{\mu}_c(T)) \le -\sqrt{m}\,\mu_c(T) + \Delta(T)\Big\} \\
&= \mathsf{P}\Big\{\sqrt{m}\,\widetilde{\Omega}_c(T) \le \Delta(T)\Big\} \\
&= \mathsf{P}\Big\{\widetilde{\Omega}_c(T) \le m^{-1/2}\sup_t |\Delta(T)|\Big\} \\
&= \mathsf{P}\Big\{\widetilde{\Omega}(T) \le 0\Big\} + O\left(\frac{1}{\sqrt{m}}\right),
\end{aligned}
$$

since $\sup_t |\Delta(t)| = O_P(1)$. $\square$

Proof of Theorem. Following the same proof as before, we conclude that

$$\mathsf{P}_{\underline{a},\underline{F}}\{\Gamma(T) \le c\} \ge 1 - \alpha + O(m^{-1/2}).$$

From the previous Lemma, $\mathsf{P}_{a,F}\{\Omega(T) \le 0\} \ge \mathsf{P}_{\underline{a},\underline{F}}\{\Omega(T) \le 0\} + O(m^{-1/2})$. It follows that

$$\mathsf{P}_{a,F}\{\Gamma(T) \le c\} \ge \mathsf{P}_{\underline{a},\underline{F}}\{\Gamma(T) \le c\} \ge 1 - \alpha + O(m^{-1/2}),$$

as claimed. $\square$

Remark 6.1. It is also possible to formulate a conditional bootstrap in which the P_i's are kept fixed at their observed values. In this one draw $H_i^* \mid P_i \sim \text{Bernoulli}(1 - \widehat{q}(P_i))$ where $\widehat{q}(P_i) = \widehat{\mathsf{P}}\{H_i = 0 \mid P_i\} = (1 - \widehat{a})/\widehat{g}(P_i)$. Again, a bias correction is needed that now involves $\widehat{g}$.

6.2 Closed-Form Expression for Asymptotic Confidence Threshold

As an alternative to the bootstrap, we derive in this section a formula for a confidence threshold that is asymptotically valid. As usual, we begin by assuming a is known.

LEMMA 6.2. *Let $t_0 = Q^{-1}(c)$, and assume $0 < t_0 < 1$. If $t_m - t_0 = O(m^{-1/2})$, $\Omega(t_m) - \mu(t_m) = \Omega(t_0) + o_P(m^{-1/2})$. Thus, if $u_m = vm^{-1/2} + o(m^{-1/2})$ for some v,*

$$\mathsf{P}\{\Omega(t_m) \leq \mu(t_m) + u_m\} - \mathsf{P}\{\Omega(t_0) \leq u_m\} = o(1).$$

PROOF. Note that $\mu(t_0) = 0$. Note also that $|\Omega(t_m) - \Omega(t_0)| \leq \max\{c, 1 - c\} m^{-1} \sum_i |1\{P_i \leq t_m\} - 1\{P_i \leq t_0\}| \leq |\widehat{G}(t_m) - \widehat{G}(t_0)|$ which is Binomial$(m, |G(t_m) - G(t_0)|)/m$ and has variance of order $m^{-3/2}$. The first claim follows by subtracting the means and multiplying by $\sqrt{m}$. The second claim is immediate. □

LEMMA 6.3. *Let $t_0 = Q^{-1}(c)$ and let $K_\Omega(s,t)$ be the covariance kernel defined in (18). Assume that $F \neq U$. Define*

$$t_m \equiv t_m(\alpha) = t_0 - \frac{z_\alpha}{\sqrt{m}} \frac{\sqrt{K_\Omega(t_0,t_0)}}{1 - a - cg(t_0)}.$$

Then

$$\mathsf{P}\{\Gamma(t_m) \leq c\} = 1 - \alpha + O(m^{-1/2}).$$

PROOF. We have

$$\begin{aligned} \mathsf{P}\{\Gamma(t_m) \leq c\} &= \mathsf{P}\{\Omega(t_m) - \mu(t_m) \leq -\mu(t_m)\} \\ &= \mathsf{P}\left\{\sqrt{m}\frac{\Omega(t_0)}{\sqrt{K_\Omega(t_0,t_0)}} \leq -\frac{\sqrt{m}\mu(t_m)}{\sqrt{K_\Omega(t_0,t_0)}}\right\} + o(1), \end{aligned}$$

from Lemma 6.2. It suffices, in light of Theorem 4.1 and Lemma 6.2, to show that

$$-\sqrt{m}\frac{\mu(t_m)}{\sqrt{K_\Omega(t_0,t_0)}} \to z_\alpha.$$

Now, $\mu(t_0) = 0$, hence,

$$\begin{aligned} \mu(t) &= (t - t_0)\mu'(t_0) + o(|t - t_0|) \\ &= (t - t_0)(1 - a - cg(t_0)) + o(|t - t_0|). \end{aligned}$$

Hence,

$$\mu(t_m) = (t_m - t_0)(1 - a - cg(t_0)) + o(m^{-1/2}).$$

The result follows from the definition of t_m. $\square$

It would be tempting to substitute $\widehat{t}_0 = \widehat{Q}^{-1}(c)$ for t_0 in the definition of t_m and then use the estimate t_m. Unfortunately, much like the bootstrap, this adds an asymptotic bias which prevents the resulting threshold from having correct asymptotic coverage. Again, a bias correction is needed.

THEOREM 6.4. *Let $\widehat{a}$ be a consistent estimator of $\underline{a}$. Let $t_0 = Q^{-1}(c)$, and $\widehat{t}_0 = \widehat{Q}^{-1}(c)$. Assume that $F \neq U$. Let*

$$T = \widehat{t}_0 - \frac{\widehat{J}}{\sqrt{m}}$$

where

$$\widehat{J} = \frac{z_{\alpha/2}\left(\sqrt{\widehat{K}_{Q^{-1}}(\widehat{t}_0, \widehat{t}_0)} + \widehat{g}(\widehat{t}_0)\right) + 2\sqrt{\log m}}{1 - \widehat{a} - c\widehat{g}(\widehat{t}_0)}$$

and

$$\widehat{K}_{Q^{-1}}(s,t) = \frac{\widehat{K}_Q(\widehat{Q}^{-1}(s), \widehat{Q}^{-1}(t))}{\widehat{Q'}(\widehat{Q}^{-1}(s))\widehat{Q'}(\widehat{Q}^{-1}(t))},$$

$$\widehat{K}_Q(s,t) = \frac{(1-\widehat{a})^2 st}{\widehat{G}^2(s)\widehat{G}^2(t)}\left[\widehat{G}(s \wedge t) - \widehat{G}(s)\widehat{G}(t)\right].$$

Assume also that $\widehat{g}$ is continuous, is a consistent estimator of g and that $\widehat{g}(t_0) - g(t_0) = O_P(m^{-\delta})$ for some $\delta > 0$. Then,

$$\mathsf{P}\{\Gamma(T) \leq c\} \geq 1 - \alpha + o(1).$$

PROOF. First, recall that $\sqrt{m}(\widehat{t}_0 - t_0) \rightsquigarrow N(0, K_{Q^{-1}}(c,c))$. Let

$$t_m = t_0 - \frac{J}{\sqrt{m}}$$

where

$$J = \frac{z_{\alpha/2}\left(\sqrt{K_\Omega(t_0, t_0)} + g(t_0)\right) + 2\sqrt{\log m}}{1 - a - cg(t_0)}.$$

Now,

$$\mathsf{P}\{\Gamma(T) \le c\} = \mathsf{P}\{\Omega(T) \le 0\} = \mathsf{P}\{\Omega(t_m) + [\Omega(T) - \Omega(t_m)] \le 0\}\,.$$

Let $I_i = 1\{P_i \le T\}$ and $J_i = 1\{P_i \le t_m\}$. Then, with probability at lest $1 - 2/m^2$, for some $\widetilde{t}$ between T and t,

$$\begin{aligned}
|\Omega(T) - \Omega(t_m)| &\le \frac{1}{m}\sum_i |1 - H_i - c|\,|I_i - J_i| \\
&\le \max\{c, 1-c\}\frac{1}{m}\sum_i |I_i - J_i| \\
&\le \frac{1}{m}\sum_i |I_i - J_i| \\
&\le |\widehat{G}(T) - \widehat{G}(t_m)| \\
&\le 2\sqrt{\frac{\log m}{m}} + |G(T) - G(t_m)| \\
&= 2\sqrt{\frac{\log m}{m}} + |T - t_m| g(\widetilde{t}) \\
&\le 2\sqrt{\frac{\log m}{m}} + |t_0 - \widehat{t}_0| g(t_0) + o_P(m^{-1/2}).
\end{aligned}$$

Let

$$\Delta_m = \frac{g(t_0) z_{(\alpha/2)} \sqrt{K_{Q^{-1}}(c,c)} + 2\sqrt{\log m}}{\sqrt{m}}.$$

So,

$$\begin{aligned}
\mathsf{P}\{\Gamma(T) \le c\} &= \mathsf{P}\{\Omega_c(T) \le 0\} \\
&\ge \mathsf{P}\{\Omega_c(t_m) \le -\Delta_m,\ |\Omega(T) - \Omega(t_m)| < \Delta_m\} \\
&\ge \mathsf{P}\{\Omega_c(t_m) \le -\Delta_m\} + \mathsf{P}\{|\Omega(T) - \Omega(t_m)| < \Delta_m\} - 1.
\end{aligned}$$

Then,

$$\mathsf{P}\{|\Omega_c(T) - \Omega_c(t_m)| > \Delta_m\} \;\le\; \mathsf{P}\Big\{\sqrt{m}|\widehat{t}_0 - t_0| + o_P(1) > z_{\alpha/2}\Big\} \to \frac{\alpha}{2}.$$

Also,

$$\mathsf{P}\{\Omega_c(t_m) \le -\Delta_m\} \;=\; \mathsf{P}\left\{\sqrt{m}\frac{\Omega_c(t_m) - \mu(t_m)}{\sqrt{K(t_0,t_0)}} \le \frac{-\sqrt{m}\Delta_m - \sqrt{m}\mu(t_m)}{\sqrt{K(t_0,t_0)}}\right\}.$$

The right hand side tends to $z_{\alpha/2}$.

Now, if a is unknown, we use the estimate $\widehat{a}$. Let $T_{\underline{a}}$ be the confidence threshold attained above, treating $\underline{a}$ as known. The method of proof given for Theorem 5.2 – substituting $Q^{-1}(c)$ and $\widehat{Q}^{-1}(c)$ for $Q^{-1}(\alpha)$ and $\widehat{Q}^{-1}(\alpha)$ respectively – yields that

$$\mathsf{P}_{a,F}\{\Gamma(T) \leq c\} \geq \mathsf{P}_{a,F}\{\Gamma(T_{\underline{a}}) \leq c\} + o(1).$$

The result follows. $\square$

Remark 6.2. The asymptotic approach does not require bootstrapping, but does require density estimation. This is analogous to the choices faced in estimating the standard error of a median.

6.3 Exact Confidence Thresholds

In this section, we will construct confidence thresholds that are valid for finite samples.

Let $0 < \alpha < 1$. Given $V_1, \ldots, V_k$, let $\varphi_k(v_1, \ldots, v_k)$ be a non-randomized level α test of the null hypothesis that $V_1, \ldots, V_k$ are drawn IID from a Uniform$(0,1)$ distribution. Define $p_0^m(h^m) = (p_i \;:\; h_i = 0, 1 \leq i \leq m)$ and $m_0(h^m) = \sum_{i=1}^m (1 - h_i)$. and

$$\mathcal{U}_\alpha(p^m) = \left\{h^m \in \{0,1\}^m \,:\, \varphi_{m_0(h^m)}\left(p_0^m(h^m)\right) = 0\right\},$$

Note that as defined, $\mathcal{U}_\alpha$ always contains the vector $(1, 1, \ldots, 1)$.

For example, if we define the empirical CDF of the null p-values by

$$S(t; h^m, p^m) = \frac{\sum_i I\{p_i \leq t\}(1 - h_i)}{\sum_i (1 - h_i)},$$

we can use the Kolmogorov-Smirnov one-sample test against the Uniform$(0,1)$.

Also, let

$$\begin{aligned} \mathcal{G}_\alpha(p^m) &= \{\Gamma(\cdot, h^m, p^m) \,:\, h^m \in \mathcal{U}_\alpha(p^m)\} && (19) \\ \mathcal{M}_\alpha(p^m) &= \{m_0(h^m) \,:\, h^m \in \mathcal{U}_\alpha(p^m)\}. && (20) \end{aligned}$$

Then, we have the following theorem which follows from standard results on inverting hypothesis tests to construct confidence sets.

THEOREM 6.5. *For all $0 < a < 1$, $F \in \mathcal{F}$, and positive integers m,*

$$
\begin{aligned}
\mathsf{P}_{a,F}\{H^m \in \mathcal{U}_\alpha(P^m)\} &\geq 1-\alpha && (21)\\
\mathsf{P}_{a,F}\{M_0 \in \mathcal{M}_\alpha(P^m)\} &\geq 1-\alpha && (22)\\
\mathsf{P}_{a,F}\{\Gamma(\cdot, H^m, P^m) \in \mathcal{G}_\alpha\} &\geq 1-\alpha && (23)\\
\mathsf{P}_{a,F}\{\Gamma(T_C) \leq c\} &\geq 1-\alpha, && (24)
\end{aligned}
$$

where

$$T_C = \sup\{t:\ \Gamma(t; h^m, P^m) \leq c \text{ and } h^m \in \mathcal{U}_\alpha(P^m)\}. \qquad (25)$$

In particular, T_C is a $(1-\alpha, c)$ confidence threshold procedure.

Because $\mathcal{U}_\alpha$ always contains $(1, 1, \ldots, 1)$, the pointwise infimum of functions in $\mathcal{G}_\alpha$ will be zero. However, there is a non-trivial upperbound

$$\mathcal{S}_\alpha(t) = \sup\{\Gamma(t):\ \Gamma \in \mathcal{G}_\alpha(P^m)\}, \qquad (26)$$

which satisfies $\inf_{a,F} \mathsf{P}_{a,F}\{\Gamma(t, H^m, P^m) \leq \mathcal{S}_\alpha(t), \text{ for all } t\} \geq 1-\alpha$.

REMARK 6.3. If there is some substantive reason to bound M_0 from below, then $\mathcal{G}_\alpha$ will have a non-trivial lower bound as well.

REMARK 6.4. At first glance, computation of $\mathcal{U}_\alpha$ would appear to require an exponential-time algorithm. However, for broad classes of tests, including the Kolmogorov-Smirnov test, it is possible construct $\mathcal{U}_\alpha$ in polynomial time. We have fast algorithms and alternative tests which we will present and analyze in a forthcoming paper.

Appendix: Algorithm for Finding $\widehat{F}_m$

Here, we restrict our attention to the case in which we take $\widehat{F}$ as piecewise constant on the same grid as $\mathbb{G}$. When F is concave, the algorithm works in the same way with the sharper piecewise linear approximation.

Step 0. Begin by constructing an initial estimate of F that is a CDF. For example, we can define H to be the piecewise constant function that takes the following values on the P_is

$$H(P_{(i)}) = \max_{j \leq i} \frac{\widehat{G}(P_{(j)}) - (1-\widehat{a})P_{(j)}}{\widehat{a}}.$$

Step 1. Identify the segment with the biggest absolute difference between $\widehat{G}$ and $(1-\widehat{a})U + \widehat{a}H$.

Step 2. Determine how far and in what direction (up or down) this segment can be moved while keeping H a CDF and minimizing $||\widehat{G} - (1-\widehat{a})U + \widehat{a}H||_\infty$.

Step 3. If the segment can be moved, move it and go to Step 1. Else go to Step 4.

Step 4. If no segment can be moved to reduce $||\widehat{G} - (1-\widehat{a})U + \widehat{a}H||_\infty$, STOP.

If the current segment is part of a contiguous block of segments where one segment in the block can be moved to reduce $||\widehat{G} - (1-\widehat{a})U + \widehat{a}H||_\infty$, move the segment at the end of the contiguous block of segments that provides the greatest reduction in $||\widehat{G} - (1-\widehat{a})U + \widehat{a}H||_\infty$. Go to Step 1.

References

Abramovich, F., Benjamini, Y., Donoho, D. and Johnstone, I. (2000). Adapting to unknown sparsity by controlling the false discovery rate. Technical report 2000-19. Department of Statistics. Stanford University.

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.

Benjamini, Y. and Hochberg, Y. (2000). *Journal of Educational Behavior, Statistics.* ATTN

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

Efron, B., Tibshirani R., Storey, J., and Tusher, V. (2001). Empirical Bayes Analysis of a Microarray Experiment. *Journal of the American Statistical Association*, 96, 1151-1160.

Genovese, C. R. and Wasserman, L. (2001). Operating Characteristics and Extensions of the FDR Procedure. *Journal of the Royal Statistical Society*

B, to appear.

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

Hochberg, Y. and Benjamini, Y. (1989). *Statistics in Medicine*. ATTN (?Plug in)

Steck, G. P. (1971). Rectangle Probabilities for Uniform Order Statistics and the Probability That the Empirical Distribution Function Lies Between Two Distribution Functions. *The Annals of Mathematical Statistics*, 42, 1–11.

Storey, J. (2001a). A direct approach to false discovery rates. *Journal of the Royal Statistical Society B*, to appear.

Storey, J. (2001b). The positive False Discovery Rate: A Bayesian interpretation and the q-value. Technical Report, Stanford University Dept. of Statistics, `http://www-stat.stanford.edu/ jstorey`.

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

van der Vaart, A. (1998). *Asymptotic Statistics*, Cambridge University Press, Cambridge.

Yekutieli, D. and Benjamini, Y. (1999). *Journal of Statistical Planning and Inference*. ATTN