Multiple hypothesis testing on composite nulls using constrained p-values

Zhiyi Chi[*]

Department of Statistics
University of Connecticut
215 Glenbrook Road, U-4120
Storrs, CT 06269, USA
e-mail: zchi@stat.uconn.edu

Abstract: Multiple hypothesis testing often encounters composite nulls and intractable alternative distributions. In this case, using p-values that are defined as maximum significance levels over all null distributions ("p_{max}") often leads to very conservative testing. We propose constructing p-values via maximization under linear constraints imposed by data's empirical distribution, and show that these p-values allow the false discovery rate (FDR) to be controlled with substantially more power than p_{max}.

AMS 2000 subject classifications: Primary 62G10, 62H15; secondary 62G20.
Keywords and phrases: Multiple testing, composite null, p-value, empirical process, FDR, stopping time, DKW inequality.

1. Introduction

In multiple hypothesis testing, much research has been carried out on the case of simple nulls, where the data associated with true nulls follow a common distribution [8, 12]. On the other hand, the case of composite nulls, where the data associated with true nulls may follow different distributions, is often encountered and challenging.

Typically, for a multiple testing procedure, p-values are the only accessible information [14, 19, 20, 23, 31]. Thus, how p-values are defined is important to the performance of the procedure. For composite nulls, p-values are usually defined as maximum significance levels over all null distributions, resulting in conservative tests [19]. On the other hand, a Bayesian approach is possible [12], which assumes that there is a known mixture probability distribution on true nulls. Under the assumption, the overall distribution associated with true nulls can be obtained as a weighted integral of the null distributions, which essentially reduces testing on composite nulls to testing on simple nulls.

The article aims to develop an approach in between the above two. Its premise is that there exists a mixture probability distribution on true nulls, however, it

is unknown. In general, in cases like this, it is an important issue how to get extra information from data [7, 8]. One way is to assume that there is specific knowledge on both the null and alternative distributions, so that a few parameters suffice their characterization. Under the assumption, the parameters can be estimated from data [17]. However, in exploratory studies, oftentimes the goal is to identify novel signals with no prior knowledge on how they might look like; that is to say, no parametric forms of the alternative distribution can be presumed. In this case, there can be a wide choice of mixture probability distributions and alternative distributions. Roughly speaking, given a candidate mixture probability distribution, all the misfits to the data can be attributed to false nulls. This makes consistent estimation of the mixture probability distribution infeasible. At the same time, although there can be many different distributions associated with false nulls, since they and their composition are intractable, it is suitable to treat the data associated with false nulls as being sampled from a single, intractable, overall distribution.

From the above position, we shall study the case where there are only a finite number of null distributions. This case may serve as an approximation to the case where there are infinitely many null distributions; see the discussion in Section 6. Meanwhile, in pattern classification, it is common to deal with a finite number of distributions [4, 15]. On the one hand, for patterns that are known, their respective distributions can be learned beforehand as null distributions. On the other, in a novel environment, the composition of these patterns may be intractable due to the presence of false nulls, i.e., unknown novel patterns. Regardless of the specifics, the basic point is that the mixture of the null distributions is dominated by the empirical distribution of the data, except for a small error. As a result, the p-values can be defined via maximization over a range of linear combinations of the null distributions, with the coefficients constrained by the empirical distribution.

The foremost goal of the article is the construction of constrained p-values. These p-values can be used like any other p-values for multiple testing. To evaluate their efficacy, one has to choose an error criterion. We select the FDR due to the interest it has received in literature. There have been several FDR controlling procedures proposed with the aim to improve power [1, 10, 30]. We shall study the FDR control and power of some of these procedures based on the constrained p-values.

In the next section, we set up basic assumptions and review some FDR controlling procedures. In Section 3, we describe the initial step of the construction of constrained p-values, i.e., to get p-values under individual nulls. We also formulate some common types of p-values in ways that motivate the construction. In Section 4, we introduce two types of constrained p-values and state some results on the FDR control based on the p-values. Unfortunately, the proofs of the results are a bit involved, so we put them in the Appendix. In Section 5, we conduct numerical study on the FDR control and power of several procedures based on the constrained p-values and compare the results to other types of p-values. Section 6 concludes with some discussion.

2. Preliminaries

2.1. Distributional assumptions

Let $\{F_\theta, \theta \in \Theta\}$ be a family of distributions on $\mathbb{R}^d$. We make the following assumptions. First, given observations $X_1, \ldots, X_n \in \mathbb{R}^d$, the nulls to be tested are

$$H_i : X_i \sim F_\theta \text{ for some } \theta \in \Theta.$$

Second, the distribution under false nulls is a single distribution $G \notin \{F_\theta\}$, the population fraction of false nulls among the nulls is $a \in (0, 1)$, and Θ is endowed with a probability measure ν. Third, $X_1, \ldots, X_n$ are sampled as follows. First, draw $\eta_1, \ldots, \eta_n$ i.i.d. $\sim \mu = a\delta_* + (1 - a)\nu$, where δ_* is the point mass at $*$, an element not in Θ. Then conditional on the η_i's, draw X_i independently, such that if $\eta_i = *$, then $X_i \sim G$ and otherwise $X_i \sim F_{\eta_i}$. A null H_i is true if and only if $\eta_i \in \Theta$. The model is usually referred to as a random mixture model. We will refer to F_θ as a null distribution, and ν the mixture probability distribution of true nulls.

We will make two further assumptions. First, since $X_i \sim F = \int F_\theta \nu(d\theta)$ under true H_i, if ν is known, composite nulls can be reduced to simple nulls, which is not a case of our focus. Therefore, we shall assume that ν is unknown. Second, we assume that no specific knowledge on G is known, such as its parametric form. This is especially intended for the case where Θ is finite. Indeed, in the case where G is known, for $n \gg 1$, both a and ν as well as the parameters characterizing G can be estimated, e.g., by MLE, which effectively reduces the testing problem into one only involving simple nulls.

As an example, in many cases, the nulls are of the form $H_i : \theta_i > 0$ or $H_i : \theta_i \in [-\delta, \delta]$, where θ_i is a shift parameter, and there is little knowledge about the null distributions. Typically, one will then sample multiple observations and use their z-statistic to test each H_i. This way of testing underlies the belief that the z-statistic closely follows some $N(\mu, 1)$, hence placing itself under the mixture model. At the same time, under the normal approximation, except for special cases, the z-statistics associated with false nulls follow an unknown mixture of normal distributions, so the assumption that G is unknown still applies.

2.2. Step-up procedure for FDR control

Step-up (SU) procedures are widely used for multiple testing [16, 20, 28]. If $p_{(1)} \leq \cdots \leq p_{(n)}$ are sorted p-values, then an SU procedure rejects and only rejects nulls whose p-values have ranks no greater than $R = \max\{i : p_{(i)} \leq c_{i:n}\}$, with $\max \emptyset := 0$, where $c_{1:n}, \ldots, c_{n:n}$ are certain critical values [11].

To conduct multiple tests, one first has to specify an error criterion. There are quite a few such criteria available and a major line of research is to improve power while controlling the error [24, 25]. We shall use FDR as the error criterion. Several SU procedures have been proposed to control the FDR, all of

which set $c_{i:n} = \rho(i/n)$ for some function ρ. Perhaps the most well-known such procedure is the Benjamini-Hochberg procedure that uses $\rho(t) = \alpha t$, where α is the target FDR control level [1]. The procedure will be coded as BH. We shall also consider a few of its variants that aim to improve power.

The first variant is proposed by Storey, Taylor and Siegmund in [30]. Let

$$\widehat{\mathrm{FDR}}_\lambda^*(t) = \frac{\mathbf{1}\left\{t \leq \lambda\right\} \hat{\pi}_0^*(\lambda)t}{[R(t) \vee 1]/n} + \mathbf{1}\left\{t > \lambda\right\}$$

where $\lambda \in (0, 1)$ is a parameter, $R(t) = \#\{i : p_i \leq t\}$, and

$$\hat{\pi}_0^*(\lambda) = \frac{\#\{i : p_i > \lambda\} + 1}{(1 - \lambda)n}.$$

The procedure rejects and only rejects nulls whose p-values have ranks no greater than $R = \max\{i : \widehat{\mathrm{FDR}}_\lambda^*(p_{(i)}) \leq \alpha\}$. The procedure is an SU procedure, with

$$\rho(t) = \frac{\alpha t}{\hat{\pi}_0^*(\lambda)} \wedge \lambda. \tag{2.1}$$

It will be coded as STS.

We shall also consider the SU procedures proposed by Finner, Dickhaus, and Roters in [10]. Fix parameter $\kappa \in (0, 1]$ and let $f_\alpha(t) = t/\{t(1 - \alpha) + \alpha\}$. The procedures take $\rho_i(t) = f_\alpha^{-1}(t)\mathbf{1}\left\{t \leq f_\alpha(\kappa)\right\} + g_i(t)\mathbf{1}\left\{t > f_\alpha(\kappa)\right\}$, $i = 1, 2, 3$, respectively, with

$$g_1(t) = \frac{t - f_\alpha(\kappa)}{f_\alpha'(\kappa)} + \kappa, \quad g_2(t) = \frac{\kappa t}{f_\alpha(\kappa)}, \quad g_3(t) \equiv \kappa. \tag{2.2}$$

The procedures will be coded as F1, F2, and F3, respectively.

Recall that for multiple testing, if R is the number of rejected nulls and V that of rejected true nulls, then FDR $= \mathsf{E}\left[V/(R \vee 1)\right]$. By definition, if there are n nulls in total and N of them are true, then power $= \mathsf{E}[(R-V)/\{(n-N)\vee 1\}]$. Clearly, $R - V \leq n - N$, so power ≤ 1.

3. Significance levels defined via nested regions

3.1. P-*values under individual null distributions*

To construct suitable p-values for composite nulls, we need to first consider how to define p-values under individual F_θ. These p-values are the basic "building blocks". Let $\{D_t \subset \mathbb{R}^d : t \in \mathbb{R}\}$ be a family of Borel sets such that

D1. the family is increasing and right-continuous, i.e. $D_t = \bigcap_{s>t} D_s$ for any t;
D2. $\bigcup_t D_t = \mathbb{R}^d$; and
D3. $G(\bigcap_t D_t) = F_\theta(\bigcap_t D_t) = 0$, $\theta \in \Theta$.

For each t, denote the probabilities of D_t under G and F_θ by

$$\psi(t) = G(D_t), \quad \phi_\theta(t) = F_\theta(D_t). \tag{3.1}$$

By D1–D3, $\psi(t)$ and $\phi_\theta(t)$ are nondecreasing and right-continuous, with $\psi(-\infty) = \phi_\theta(-\infty) = 0$ and $\psi(\infty) = \phi_\theta(\infty) = 1$. Define

$$s(x) = \inf\{t : x \in D_t\}, \quad s_i = s(X_i), \quad i = 1, \ldots, n. \tag{3.2}$$

By D2, $s(x) < \infty$ is well-defined, and by D3, $s_i > -\infty$ almost surely. Denote

$$R_n(t) = \sum_{i=1}^{n} \mathbf{1}\{X_i \in D_t\}. \tag{3.3}$$

The next result is basic. Its part (1) implies that $\phi_\theta(s_i)$ can be used as p-values of X_i under F_θ and part (3) shows that $R_n(t)/n$ is the empirical distribution of s_i. Apparently, the p-values depend on D_t. While the selection of D_t can have a strong effect on multiple testing [5], the issue is beyond the scope of the article.

Proposition 3.1. *Under D1–3, (1) given θ, if $X_i \sim F_\theta$, then $s_i \sim \phi_\theta$, (2) $s_1, \ldots, s_n$ are i.i.d. $\sim Q$, with $Q(t) = (1-a) \int \phi_\theta(t)\, \nu(d\theta) + a\psi(t)$, and (3) for any t, $s_i \leq t \iff X_i \in D_t$ and hence $R_n(t) = \sum \mathbf{1}\{s_i \leq t\}$.*

Remark. (1) If the index set of a nested family $\{D_t,\, t \in I\}$ is an interval $I \neq \mathbb{R}$, then, by defining $D_s = \bigcap_{t \in I} D_t$ for $s \leq \inf I$ and $D_s = \mathbb{R}^d$ for $s \geq \sup I$, we can extend the nested family to one with index set $\mathbb{R}$, so that the discussion in the subsequent sections still applies.

(2) In single hypothesis tests, nested rejection regions are usually indexed by significance level. As already seen, other indices can be used as well, which is sometimes more natural and avoids potential problems caused by different regions having the same significance level. As an example, let X_i be real-valued. If we set $D_t = (-\infty, t]$, then $s_i = X_i$ and $\phi_\theta(s_i) = F_\theta(s_i)$, the lower-tail probability of X_i. If we set $D_t = [-t, \infty)$, then $s_i = -X_i$ and $\phi_\theta(s_i) = F_\theta([-s_i, \infty)) = F_\theta([X_i, \infty))$, the upper-tail probability of X_i. For F_θ continuous at 0, if we set $D_t = [-t, t]$ for $t \geq 0$ and $D_s = \{0\}$ for $s < 0$, then almost surely, $s_i = |X_i|$ and $\phi_\theta(s_i) = F_\theta([-s_i, s_i])$.

3.2. Maximum significance levels as p-*values*

The function

$$M_{\max}(t) = \sup_\theta \phi_\theta(t)$$

is nondecreasing with $M_{\max}(-\infty) = 0$ and $M_{\max}(\infty) = 1$. By Proposition 3.1, $M_{\max}(s_i)$ is the maximum significance level of X_i over all possible null distributions, which is a conventional p-value under composite nulls [19]. We henceforth denote $p_{i,\max} = M_{\max}(s_i)$.

It is known that using $p_{i,\max}$ as p-values, BH and its variants in (2.1) and (2.2) can control the FDR [3, 10, 30]. The issue here is the power of the procedures using $p_{i,\max}$. This will be studied in the simulations in Section 5.

Observe that $M_{\max}(t)$ can be written as

$$M_{\max}(t) = \sup_{\mu} \int \phi_\theta(t)\, d\mu(\theta), \qquad (3.4)$$

where the supremum is taken over all possible measures μ on Θ with $\mu(\Theta) \le 1$. If no information is available on ν or a, then the above unconstrained supremum is justified. If, on the other hand, it is known that ν and a satisfy certain conditions, then, by constraining the supremum with the conditions, it is possible to get significance levels closer to $p_{i,\mathrm{mix}}$, hence improving the performance of multiple testing. Here $p_{i,\mathrm{mix}} = M_{\mathrm{mix}}(s_i)$, with

$$M_{\mathrm{mix}}(t) = (1-a) \int \phi_\theta(t)\, d\nu(\theta). \qquad (3.5)$$

Indeed, since $\int \phi_\theta(t)\, d\nu(\theta)$ is the probability of D_t under the mixture of the null distributions, it is known that by using $p_{i,\mathrm{mix}}$ as the p-values, the power of BH is maximized when the FDR is controlled at a target level [2, 13, 29].

Remark. Since for $a > 0$, the largest possible value of $p_{i,\mathrm{mix}}$ is $1 - a < 1$, strictly speaking, $p_{i,\mathrm{mix}}$ are not significance levels. Nevertheless, for convenience, we shall still refer to them as significance levels or p-values.

4. Constructing *p*-values via constrained maximizations

4.1. Outlines

Henceforth, we consider the case where Θ is a finite set $\{\theta_k,\ k = 1,\dots,L\}$. Then the probability measure ν can be specified by

$$\boldsymbol{\nu} = (\nu_1,\dots,\nu_L)^\top, \quad \text{with} \ \ \nu_k = \nu\{\theta_k\}.$$

(Henceforth, boldfaced letters denote L-dimensional vectors.) Denote

$$\phi_k(t) = \phi_{\theta_k}(t), \quad \Delta = \{\boldsymbol{c} \in [0,1]^L :\ c_1 + \dots + c_L \le 1\}.$$

Let $s_i = s(X_i)$ as in (3.2). From Proposition 3.1, $s_1,\dots,s_n$ are i.i.d. $\sim Q$, with

$$Q(t) = (1-a)\boldsymbol{\nu}^\top \boldsymbol{\phi}(t) + a\psi(t),$$

and their empirical distribution function is $\mathbb{F}_n(t) = R_n(t)/n$, where R_n is as in (3.3). Now (3.4) and (3.5) can be rewritten as

$$M_{\max}(t) = \max_k \phi_k(t) = \sup\{\boldsymbol{c}^\top \boldsymbol{\phi}(t) : \boldsymbol{c} \in \Delta\},$$

$$M_{\mathrm{mix}}(t) = (1-a)\boldsymbol{\nu}^\top \boldsymbol{\phi}(t) = \sup\{\boldsymbol{c}^\top \boldsymbol{\phi}(t) : \boldsymbol{c} = (1-a)\boldsymbol{\nu}\}.$$

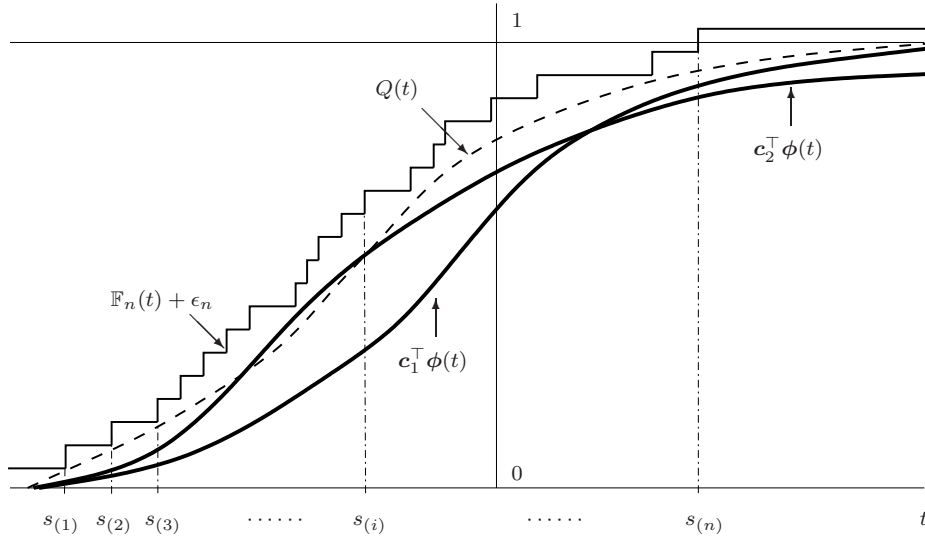At the same time, we still have $p_{i,\max} = M_{\max}(s_i)$ and $p_{i,\mathrm{mix}} = M_{\mathrm{mix}}(s_i)$.

FIG 1. *Constraints based on empirical distribution. Since a and $\boldsymbol{\nu}$ are unknown, the graph of $(1-a)\boldsymbol{\nu}^{\top}\boldsymbol{\phi}(t)$ is intentionally missing.*

The expressions for M_{max} and M_{mix} suggest a general approach to constructing p-values, that is, by constrained maximization of $\boldsymbol{c}^{\top}\boldsymbol{\phi}(s_i)$. Certainly, we hope to get some kind of p-values not so conservative as $p_{i,\mathrm{max}}$ but still conservative. Meanwhile, the closer we can make the p-values to $p_{i,\mathrm{mix}}$, the better. What constraints can be used then?

A straightforward idea is to find a suitable $C \subset \Delta$ and define p-values as

$$\sup\{\boldsymbol{c}^{\top}\boldsymbol{\phi}(s_i) : \boldsymbol{c} \in C\}.$$

Under the above considerations, we need $(1 - a)\boldsymbol{\nu} \in C \subsetneq \Delta$, which sometimes can be satisfied. For example, if it is known for sure that $a \leq a_0$, then, as $(1-a)\boldsymbol{\nu} \in \Delta' = \{\boldsymbol{c} \in \Delta : c_1 + \cdots + c_L \geq 1 - a_0\}$, one can set $C = \Delta'$. In general, however, are there constraints that are available under more general conditions?

It turns out that by exploiting properties of empirical distributions, a lot of constraints can be obtained. To get the constraints, we assume that $\phi_1, \ldots, \phi_L$ and ψ are continuous.

The constraints can be roughly divided into two types, as described next; see Fig. 1 for reference.

4.1.1. First type

Since $Q(t) - (1 - a)\boldsymbol{\nu}^{\top}\boldsymbol{\phi}(t) = a\psi(t)$ is increasing in t, the following inequalities

$$\boldsymbol{c}^{\top}[\boldsymbol{\phi}(t) - \boldsymbol{\phi}(s)] \leq Q(t) - Q(s), \quad t > s \geq -\infty \tag{A}$$

hold for $\boldsymbol{c} = (1-a)\boldsymbol{\nu}$, with $s = -\infty$ corresponding to $\boldsymbol{c}^\top \boldsymbol{\phi}(t) \leq Q(t)$. If Q were observable, the inequalities could serve as constraints on the maximization of $\boldsymbol{c}^\top \boldsymbol{\phi}(t)$. Since Q is in fact unobservable, we need to find an observable function $\tilde{Q}$ as a substitute. In general, $\tilde{Q}$ has to be made from the data so it will have some fluctuations. With this in mind, we should expect a new set of inequalities

$$\boldsymbol{c}^\top [\boldsymbol{\phi}(t) - \boldsymbol{\phi}(s)] \leq \tilde{Q}(t) - \tilde{Q}(s) + \text{``margin of error''}, \quad t > s \geq -\infty, \qquad \text{(B)}$$

which have to meet two criteria with high probability. First, the inequalities should hold for $(1-a)\boldsymbol{\nu}$, so that the resulting p-values are conservative. Second, the constraints imposed by (B) should not be too lax comparing with (A). Since

$$\sup_t |\mathbb{F}_n(t) - Q(t)| \xrightarrow{\text{P}} 0, \quad \text{as } n \to \infty,$$

a good choice for $\tilde{Q}$ seems to be $\mathbb{F}_n$, where $0 < \epsilon_n \ll 1$ such that, with high probability, $\mathbb{F}_n(t) - \mathbb{F}_n(s) + \epsilon_n \geq Q(t) - Q(s)$ for $t > s \geq -\infty$. Using probability inequalities, such ϵ_n can be found. Then the inequalities in (B) become

$$\boldsymbol{c}^\top [\boldsymbol{\phi}(t) - \boldsymbol{\phi}(s)] \leq \mathbb{F}_n(t) - \mathbb{F}_n(s) + \epsilon_n, \quad t > s \geq -\infty. \qquad (4.1)$$

In practice, for the infinitely many inequalities in (4.1), one has the freedom of choice. If, for example, only $\boldsymbol{c}^\top \boldsymbol{\phi}(t) \leq \mathbb{F}_n(t) + \epsilon_n$ are used as constraints due to concerns about computational cost, then $\boldsymbol{c}_1^\top \boldsymbol{\phi}(t)$ and $\boldsymbol{c}_2^\top \boldsymbol{\phi}(t)$ in Fig. 1 are functions satisfying the constraints.

Remark. (1) Under the constraints in (4.1), it is possible that $\boldsymbol{c}^\top \boldsymbol{\phi}(t) > Q(t)$, as seen from the graphs of $\boldsymbol{c}_2^\top \boldsymbol{\phi}$ and Q in Fig. 1.

(2) For different t, the value of $\boldsymbol{c}$ that maximizes $\boldsymbol{c}^\top \boldsymbol{\phi}(t)$ may be different. For example, in Fig. 1, for t around $s_{(i)}$, $\boldsymbol{c}_2$ is a more plausible maximizer than $\boldsymbol{c}_1$, whereas for t around $s_{(n)}$, the opposite is true.

4.1.2. Second type

Since the constraints in (4.1) are meant to be satisfied with high probability by all pairs $s < t$ simultaneously, they are not necessarily very tight for particular values of t. It is possible to replace some of the constraints with tighter ones based on local properties of the empirical distribution. Since the tails of a sample are often handled more carefully than its other part in hypothesis tests, we shall focus on $s_{(i)}$ with $i \ll n$. However, the discussion applies equally well to any small set of s_i.

The idea is as follows. Since $\boldsymbol{c}^\top \boldsymbol{\phi}(s_{(i)}) \leq Q(s_{(i)})$ is satisfied by $\boldsymbol{c} = (1-a)\boldsymbol{\nu}$, if we can find z_i, such that with high probability, $Q(s_{(i)}) \leq z_i$ for all $i \leq m_n$, where $m_n \ll n$, then, instead of $\boldsymbol{c}^\top \boldsymbol{\phi}(s_{(i)}) \leq \mathbb{F}_n(s_{(i)}) + \epsilon_n$, we can impose

$$\boldsymbol{c}^\top \boldsymbol{\phi}(s_{(i)}) \leq z_i, \quad i \leq m_n. \qquad (4.2)$$

To find such z_i, note that $Q(s_i)$ are i.i.d. $\sim \text{Unif}(0,1)$ as $Q = (1-a)\boldsymbol{\nu}^\top\boldsymbol{\phi} + a\psi$ is continuous by our assumption. Therefore, $Q(s_{(1)}), \ldots, Q(s_{(n)})$ have the same joint distribution as $S_1/S_{n+1}, \ldots, S_n/S_{n+1}$, where $S_i = \xi_1 + \cdots + \xi_i \sim \text{Gamma}(i,1)$, with ξ_k i.i.d. $\sim \text{Exp}(1)$ [27, Ch. 8]. Since $S_{n+1}/n \to 1$ almost surely as $n \to \infty$, given $0 < \beta < 1$, for $n \gg 1$, with high probability, $Q(s_{(i)}) \leq \gamma_i$, where γ_i is some random variable following $\text{Gamma}(i, 1/(n\beta))$. Thus z_i can be set equal to a high percentile of the Gamma distribution. In order for the constraints in (4.2) to be stronger than their counterparts in (4.1), one needs to have $z_i < \mathbb{F}_n(s_{(i)}) + \epsilon_n$, which has to be checked on a case-by-case basis for specific choice for ϵ_n.

4.2. Construction of p-values

To design concrete ways to construct p-values using the foregoing constraints, a principle we will follow is that a viable construction should yield p-values that are an increasing function of s_i. The increasing monotonicity naturally holds for $\phi_k(s_i)$, as they are p-values under individual null distributions. However, as different sets of constraints may be applied to different observations in calculating their p-values for composite nulls, the increasing monotonicity does not automatically hold and should be checked for specific constructions.

It is impossible to use all of the foregoing constraints as there are infinitely many of them. Besides, we need to take into account computational cost. The constraints we shall use can be divided into 3 categories. First, some "hard" constraints on $\boldsymbol{c} = (c_1, \ldots, c_L)$, the most basic ones being $c_k \geq 0$ and $\sum c_k \leq 1$. Second, upper bounds on $\boldsymbol{c}^\top\boldsymbol{\phi}(s_i)$ derived from (4.1) or (4.2), depending on the rank of s_i in $s_1, \ldots, s_n$. Third, upper bounds on $\boldsymbol{c}^\top[\boldsymbol{\phi}(t) - \boldsymbol{\phi}(s)]$, where $s < t$ belong to some pre-selected finite set of "check points".

We need to set some parameters first. To impose hard constraints on $\boldsymbol{c}$, fix a closed set $\Delta' \subset \Delta$, which must be *known for sure* to contain $(1-a)\boldsymbol{\nu}$. In most cases, $\Delta' = \Delta$ for lack of direct information about $\boldsymbol{\nu}$. To set the margin of error in the constraints of the form (4.1), fix $\epsilon_n > 0$ such that $\epsilon_n \to 0$ as $n \to \infty$. To impose constraints involving $\boldsymbol{c}^\top[\boldsymbol{\phi}(t) - \boldsymbol{\phi}(s)]$, let T_n be a finite set of check points. To set z_i in (4.2), fix $\beta \in (0,1)$ and m_n such that $m_n \ll n$ for large n.

Denote by $\bar{\Gamma}^*(z; a)$ the z-th *upper*-tail quantile of $\text{Gamma}(a, 1)$ and define

$$u_i = \begin{cases} \bar{\Gamma}^*\left(1/n; i\right)/(\beta n), & \text{if } i \leq m_n, \\ \mathbb{F}_n(s_{(i)}) + \epsilon_n & \text{otherwise, and} \end{cases} \tag{4.3}$$

Since by assumption Q is continuous, with probability 1, s_i are distinct from each other and $\mathbb{F}_n(s_{(i)}) = i/n$.

Sequential construction. Denote by $p_{i,\text{seq}}$ the p-value of s_i constructed in this way. Then $p_{i,\text{seq}} = M_{n,\text{seq}}(s_i)$, where for any t,

$$M_{n,\text{seq}}(t) = \begin{cases} \sup\left\{\boldsymbol{c}^\top\boldsymbol{\phi}(t) : \boldsymbol{c} \in \Delta' \cap C_{n,\text{seq}}(t)\right\}, & \text{if } \Delta' \cap C_{n,\text{seq}}(t) \neq \emptyset, \\ 1, & \text{otherwise,} \end{cases} \tag{4.4}$$

with $C_{n,\text{seq}}(t)$ being the set of $\boldsymbol{c} \in \Delta$ satisfying the following conditions,

1) $\boldsymbol{c}^\top \boldsymbol{\phi}(s_{(j)}) \leq u_j$ for all $s_{(j)} \geq t$; and
2) $\mathbb{F}_n(t_2) - \mathbb{F}_n(t_1) + \epsilon_n \geq \boldsymbol{c}^\top [\boldsymbol{\phi}(t_2) - \boldsymbol{\phi}(t_1)]$ for $t_1, t_2 \in T_n$ with $t \leq t_1 < t_2$.

Global construction. Denote by $p_{i,\text{glb}}$ the p-value of s_i constructed in this way. Then $p_{i,\text{glb}} = M_{n,\text{glb}}(s_i)$, where for any t,

$$M_{n,\text{glb}}(t) = \begin{cases} \sup \left\{ \boldsymbol{c}^\top \boldsymbol{\phi}(t) : \boldsymbol{c} \in \Delta' \cap C_{n,\text{glb}} \right\}, & \text{if } \Delta' \cap C_{n,\text{glb}} \neq \emptyset, \\ 1, & \text{otherwise,} \end{cases} \quad (4.5)$$

with $C_{n,\text{glb}}$ being the set of $\boldsymbol{c} \in \Delta$ satisfying the following conditions,

1) $\boldsymbol{c}^\top \boldsymbol{\phi}(s_{(j)}) \leq u_j$ for all $j = 1, \ldots, n$; and
2) $\mathbb{F}_n(t_2) - \mathbb{F}_n(t_1) + \epsilon_n \geq \boldsymbol{c}^\top [\boldsymbol{\phi}(t_2) - \boldsymbol{\phi}(t_1)]$ for all $t_1, t_2 \in T_n$ with $t_1 < t_2$.

Remark. (1) Since $C_{n,\text{seq}}(t)$ and $C_{n,\text{glb}}$ are convex closed subsets of Δ and contain $\boldsymbol{0}$, the suprema in the definitions of $M_{n,\text{seq}}$ and $M_{n,\text{glb}}$ are attainable. Each of $C_{n,\text{seq}}(t)$ and $C_{n,\text{glb}}$ has the property that if $\boldsymbol{c}$ belongs to it, then so does any $\boldsymbol{d}$ with $0 \leq d_i \leq c_i$.

(2) For every given t, unlike $M_{\max}(t)$ and $M_{\text{mix}}(t)$, both $M_{n,\text{seq}}(t)$ and $M_{n,\text{glb}}(t)$ depends on $s_1, \ldots, s_n$ and hence are random.

(3) The first construction is dubbed "sequential" because each $p_{i,\text{seq}}$ is computed based on $s_j \geq s_i$: if we imagine that s_i are input one by one, starting with the largest one, then $p_{i,\text{seq}}$ can be computed only after all $s_j > s_i$ have been input. The second construction is dubbed "global" because each $p_{i,\text{glb}}$ is computed based on all s_j. While presumably imposing stronger constraints, the global construction has a higher computational cost.

(4) In both constructions, the constraints are linear, allowing each p-value to be computed by linear programming. It is easy to see that both allow parallelized computation of the p-values.

4.3. Application to FDR control

Once $p_{i,\text{seq}}$ and $p_{i,\text{glb}}$ are computed, they can be applied to testing procedures just as $p_{i,\text{mix}}$ and $p_{i,\text{max}}$ are. In this section, we state some theoretical results on the FDR control by BH based on $p_{i,\text{seq}}$ and $p_{i,\text{glb}}$.

The first result deals with $p_{i,\text{seq}}$. The tool for its proof is the optional sampling theorem (cf. [30]) and the Dvoretzky-Kiefer-Wolfowitz (DKW) inequality [21].

Theorem 4.1. *Suppose $\phi_1, \ldots, \phi_L$ and ψ are continuous and $\alpha < 1 - a$. Then for $n \geq 1$, provided $\exp(-2n\epsilon_n^2) \leq 1/2$,* BH *based on $p_{i,\text{seq}}$ attains*

$$\text{FDR} \leq \alpha + r_n + \mathsf{E}[\mathbf{1}\{R > 0\}/(R \vee 1)],$$

where $r_n = 2(1 + |T_n|) \exp(-2n\epsilon_n^2) + m_n[1/n + (\beta e^{1-\beta})^{n+1}]$.

The bound contains terms in addition to α. For appropriate ϵ_n and T_n, the term $2(1 + |T_n|)\exp(-2n\epsilon_n^2)$ is $o(1)$ as $n \to \infty$. Since $\beta e^{1-\beta} < 1$, if $m_n = o(n)$, then $r_n = o(1)$. Under certain conditions, R is of the same order as n. Thus, the bound shows the FDR can be asymptotically controlled at α. On the other hand, the simulations in Section 5 indicate that usually the realized FDR is substantially lower than α, which is perhaps not surprising because $p_{i,\mathrm{seq}} \geq p_{i,\mathrm{mix}}$.

Unlike the above result, the optional sampling theorem has not been able to apply to $p_{i,\mathrm{glb}}$. We will settle for an asymptotic result. For S, $T \subset \mathbb{R}$, denote $\delta(S, T) = \sup\{|s - t|, s \in S, t \in T\}$. A sequence of finite sets S_n is said to be increasingly dense in T if for any $r > 0$, $\delta(S_n, T \cap [-r, r]) \to 0$ as $n \to \infty$.

Theorem 4.2. *Suppose 1) $\phi_1, \ldots, \phi_L$ and ψ are continuous and 2) as $n \to \infty$, $\epsilon_n \to 0$, $n\epsilon_n^2 \to \infty$, $m_n = o(n)$, and T_n is increasingly dense in $\mathbb{R}$.*

Let $\Gamma_0 = \{\boldsymbol{c} \in \Delta : Q(t) - Q(s) \geq \boldsymbol{c}^\top[\boldsymbol{\phi}(t) - \boldsymbol{\phi}(s)], -\infty \leq s < t\}$. Define $m(t) = \sup\{\boldsymbol{c}^\top\boldsymbol{\phi}(t) : \boldsymbol{c} \in \Delta' \cap \Gamma_0\}$ and $t_* := \sup\{t \in \mathbb{R}: m(t) \leq \alpha Q(t)\}$. If there is $t_0 < t_*$, such that $m(t) < \alpha Q(t)$ on (t_0, t_*), then BH based on $p_{i,\mathrm{glb}}$ attains $\overline{\lim}_{n\to\infty} \mathrm{FDR} \leq \alpha$ and is asymptotically equivalent to the procedure that rejects H_i if and only if $s_i \leq t_*$.

The result is based on the observation that as $n \to \infty$, $M_{n,\mathrm{glb}}(t) \to m(t)$ under certain metric as well as a fixed point argument [12].

5. Numerical study

5.1. Setup

We next use simulations to compare the performances of multiple testing when different types of p-values are used. In the study, we only consider multiple tests for univariate observations. Given a data distribution $Q = (1-a)\sum_{k=1}^{L}\nu_k F_k + aG$ and $X_1, \ldots, X_n$ i.i.d. $\sim Q$, the nulls to be tested are H_i: "$X_i \sim F_k$ for some k", $i = 1, \ldots, n$. We use $F_k(X_i)$ as the p-values under individual null distributions; cf. the remark at the end of Section 3.1. The parameters of the data distributions used in the study are listed in Table 1.

Throughout, the target FDR control level is $\alpha = 0.25$, and, unless otherwise specified, the fraction of false nulls is $a = 0.05$. To calculate $p_{i,\mathrm{seq}}$ and $p_{i,\mathrm{glb}}$, in (4.3), we set $\epsilon_n = \sqrt{\ln n/n}$, $\beta = 0.95$, $m_n = n^{1/5}$, and in (4.4) and (4.5), we let

TABLE 1
Parameters of the data distributions for the simulation study. $t_{n,c}$ denotes the noncentral t distribution with n df and noncentrality c

	$F_1, \ldots, F_L$	$\nu_1, \ldots, \nu_L$	G
1	N(0,1), N(−1,1), N(−2,1)	.75, .15, .1	N(−4,1)
2	t_{20}, $t_{20,-1}$, $t_{20,-2}$	.75, .15, .1	$t_{20,-4}$
3	N(0,1), N(−1,1), N(−2,1)	.6, .25, .15	N(−4,1)
4	N(0,1), N(−1,1.5), N(−2,1.5)	.75, .15, .1	N(−4,1)
5	N(−i,1), i = 0,...,4	.65, .15, .1, .05, .05	N(−5,1)
6	N(5−i/5,1), i = 0,...,25	1/26	N(−1,1)
7	N(5−i/20,1), i = 0,...,100	$\propto \{i \vee (100-i) + 0.05\}$,	N(−1,1)

T_n be a set of $\lfloor (\ln n)^2 \rfloor$ equally spaced points that begins with $X_{(1)}$ and ends with $X_{(n)}$, and set $\Delta' = \{c \in \Delta : \sum c_k \geq 1 - a_0\}$, with $a_0 = 1$ or 0.1.

For each data distribution, the simulation proceeds as follows. First, 1000 samples are drawn, each one consisting of n i.i.d. X_i. Then, for each sample, different types of p-values are evaluated and different procedures are applied to the p-values. Finally, measures of performance are estimated by averaging over the samples. All the simulations are conducted in R language [22]; $p_{i,\mathrm{seq}}$ and $p_{i,\mathrm{glb}}$ are computed by the R linear programming package glpk.

5.2. Performances of BH based on different p-values

We apply BH to $p_{i,\mathrm{seq}}$, $p_{i,\mathrm{glb}}$, $p_{i,\mathrm{max}}$ and $p_{i,\mathrm{mix}}$, respectively, with $p_{i,\mathrm{seq}}$ and $p_{i,\mathrm{glb}}$ being computed by setting $\Delta' = \Delta$ in (4.4) and (4.5). The power and FDR are estimated by the sample averages of $(R-V)/(n-N)$ and $V/(R \vee 1)$, where N, R and V are the total numbers of true nulls, rejected nulls, and rejected true nulls, respectively. For the first six data distributions (cf. Table 1), we use $n = 5000$; for the last one, due to high computational intensity, we use $n = 2000$.

As seen from Table 2, each type of the p-values allows BH to control the FDR. On the other hand, $p_{i,\mathrm{mix}}$ yields substantially higher power than the others, while $p_{i,\mathrm{seq}}$ and $p_{i,\mathrm{glb}}$ yield substantially higher power than $p_{i,\mathrm{max}}$. To get an idea why this is the case, we compare the plots of the p-values. Since BH tests nulls by comparing $np_{(i)}/i$ and α, where $p_{(i)}$ is the ith smallest p-value of a given type, it is informative to graph $n\bar{p}_{(i)}/i$ vs i/n, where $\bar{p}_{(i)}$ is the sample average of $p_{(i)}$. As Fig. 2 shows, for small i/n, $np_{(i),\mathrm{seq}}/i$ and $np_{(i),\mathrm{glb}}/i$ are similar, explaining why the performances of BH are similar when it is applied to the two types of constrained p-values. At the same time, both $np_{(i),\mathrm{seq}}/i$ and $np_{(i),\mathrm{glb}}/i$ are substantially lower than $np_{(i),\mathrm{max}}/i$ and increase more rapidly than $np_{(i),\mathrm{mix}}/i$, which explains the differences in power of BH when it is applied to these different types of p-values. Thus, when ν_k are intractable, by utilizing properties of empirical processes to reduce over-evaluation of p-values, the power of BH can be significantly increased.

We also look at how linear programming works in the evaluation of the constrained p-values. For each $p_{(i),\mathrm{seq}}$ or $p_{(i),\mathrm{glb}}$, denote by $c_{1,(i)}, \ldots, c_{L,(i)}$ the coefficients obtained by the corresponding optimization in (4.4) or (4.5). Unlike the p-values, the coefficients exhibit very different patterns (Fig. 3). For each k,

TABLE 2
Power (top) and FDR (bottom) of BH based on different types of p-values

	1	2	3	4	5	6	7
$p_{i,\mathrm{seq}}$	4.93E-1	2.34E-1	4.52E-1	5.29E-4	4.64E-2	1.97E-2	3.39E-2
$p_{i,\mathrm{glb}}$	4.93E-1	2.33E-1	4.52E-1	5.29E-4	4.63E-2	1.96E-2	3.40E-2
$p_{i,\mathrm{max}}$	2.21E-1	3.59E-2	2.30E-1	4.71E-5	3.66E-3	3.78E-3	6.23E-3
$p_{i,\mathrm{mix}}$	7.70E-1	6.35E-1	6.86E-1	1.41E-1	4.47E-1	2.16E-1	5.66E-1
$p_{i,\mathrm{seq}}$	8.56E-2	8.14E-2	1.02E-1	6.65E-2	6.18E-2	5.46E-2	1.42E-2
$p_{i,\mathrm{glb}}$	8.55E-2	8.14E-2	1.02E-1	6.65E-2	6.18E-2	5.48E-2	1.42E-2
$p_{i,\mathrm{max}}$	2.66E-2	2.46E-2	3.79E-2	2.05E-2	1.02E-2	1.39E-2	2.15E-3
$p_{i,\mathrm{mix}}$	2.38E-1	2.40E-1	2.36E-1	2.42E-1	2.36E-1	2.37E-1	2.35E-1

Distribution 1 | Distribution 5
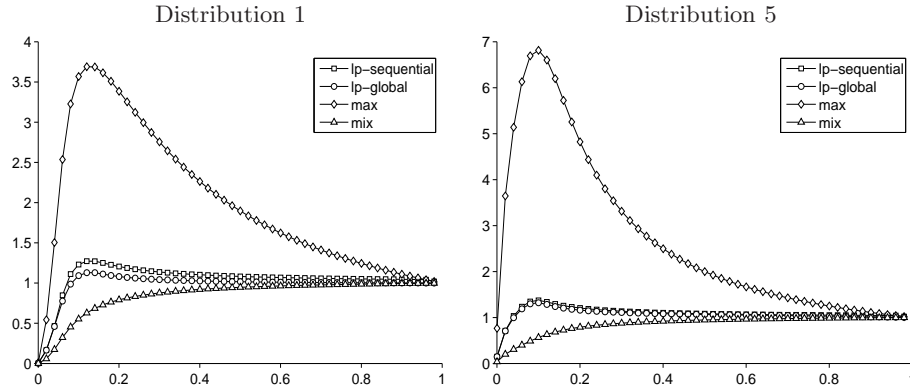
FIG 2. $n\bar{p}_{(i)}/i$ vs i/n for $p_{i,\text{seq}}$ ("lp-sequential"), $p_{i,\text{glb}}$ ("lp-global"), $p_{i,\text{max}}$ ("max"), and $p_{i,\text{mix}}$ ("mix").
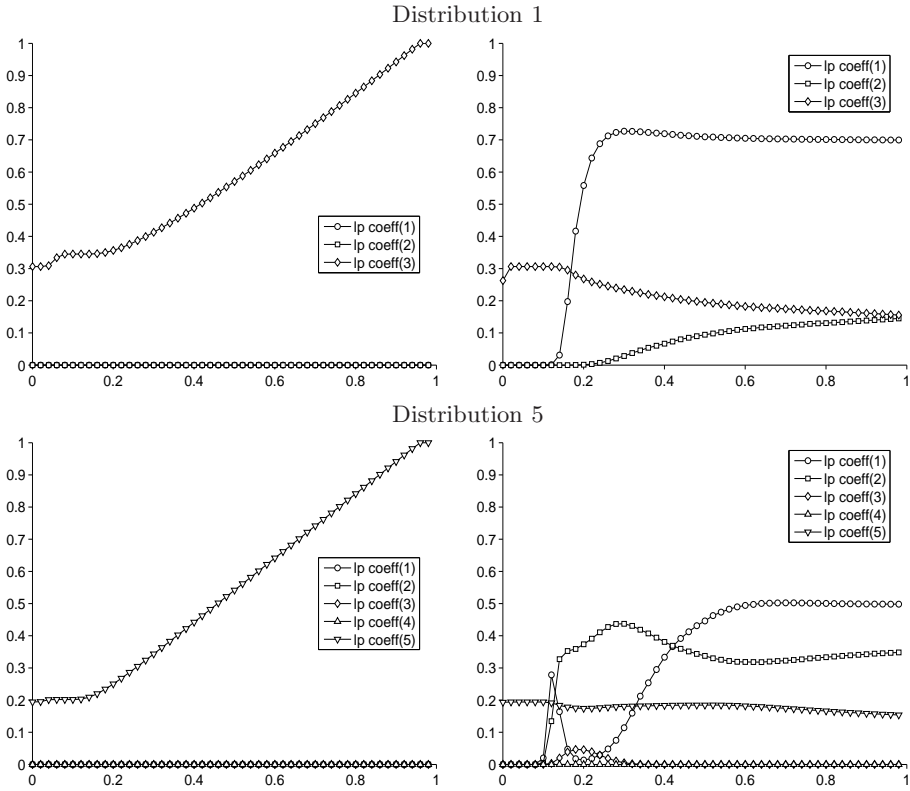
Distribution 1

Distribution 5

FIG 3. $c_{k,(i)}$ vs i/n, where $c_{1,(i)}, \ldots, c_{L,(i)}$ are the coefficients corresponding to $p_{(i),\text{seq}}$ (left) or $p_{(i),\text{glb}}$ (right).

TABLE 3

Powers of BH based on constrained p-values; cf. section 5.3 for detail. For each type of p-value, the 1st and 2nd rows are the sample mean and SD of $(R-V)/(n-N)$, respectively

	1	2	3	4	5	6	7
$p_{i,\text{seq}}$	4.93E-1	2.34E-1	4.52E-1	5.29E-4	4.64E-2	1.97E-2	3.39E-2
	5.40E-2	6.57E-2	5.37E-2	2.20E-3	3.32E-2	1.88E-2	3.56E-2
$p_{i,\text{glb}}$	4.93E-1	2.33E-1	4.52E-1	5.29E-4	4.63E-2	1.96E-2	3.40E-2
	5.40E-2	6.57E-2	5.37E-2	2.20E-3	3.32E-2	1.88E-2	3.56E-2
$p'_{i,\text{seq}}$	5.39E-1	2.96E-1	4.78E-1	8.50E-4	4.67E-2	1.98E-2	3.42E-2
	5.16E-2	6.88E-2	5.29E-2	2.87E-3	3.33E-2	1.90E-2	3.58E-2
$p'_{i,\text{glb}}$	5.39E-1	2.96E-1	4.78E-1	8.50E-4	4.67E-2	1.98E-2	3.42E-2
	5.16E-2	6.88E-2	5.29E-2	2.87E-3	3.33E-2	1.90E-2	3.58E-2

when i/n is small, the two types of $c_{k,(i)}$ are similar. However, as i/n increases, for $p_{(i),\text{seq}}$, all but one $c_{k,(i)}$ become 0, while for $p_{(i),\text{glb}}$, a more complicated combination of $c_{k,(i)}$ emerges. This difference may be partially due to how linear programming is implemented by the software used. However, it also suggests that c_k cannot be used as suitable estimates of ν_k.

5.3. Effects of stronger hard constraint

Observe that in Fig. 3, $\sum_k c_{k,(i)} \leq 0.4$ for small i/n. Since $a = 1 - \sum c_k$, this means that in the evaluation of the constrained p-values, 0.6 is counted as a feasible value of a, which is too high comparing to the true value $a = 0.05$. This suggests that, by imposing more constraints on $\sum c_k$, the power may be improved. We therefore simulate the scenario where it is known that $a \leq 0.1$. In this scenario, the hard constraint becomes $\boldsymbol{c} \in \Delta' = \{\boldsymbol{c} \in \Delta : \sum c_k \geq 0.9\}$.

Denote by $p'_{i,\text{seq}}$ and $p'_{i,\text{glb}}$ the p-values evaluated under the extra constraint. Table 3 compares the powers of BH when it is applied to $p_{i,\text{seq}}$, $p_{i,\text{glb}}$, $p'_{i,\text{seq}}$ and $p'_{i,\text{glb}}$, respectively. For each type of the p-values, the sample SDs of $(R-V)/(n-N)$ are also shown. For some of the data distributions (1–3), there is a small but significant increase in power when BH is applied to $p'_{i,\text{seq}}$ and $p'_{i,\text{glb}}$, while for the other data distributions, there is no significant difference. Fig. 4 shows the plots of $np_{(i)}/i$ for the 1st and 5th data distributions. Since all the rejected nulls are associated with $i \ll n$, we only compare the plots for $i \leq 0.05n$. Like $p_{(i),\text{seq}}$ and $p_{(i),\text{glb}}$, the plots for $p'_{(i),\text{seq}}$ and $p'_{(i),\text{glb}}$ are very close to each other. On the other hand, for the 1st data distribution, the latter two are significantly lower than the former two, which explains the improved power of BH when it is applied to $p'_{i,\text{seq}}$ and $p'_{i,\text{glb}}$. Finally, as Figs. 3 and 5 show, the new hard constraint on $\boldsymbol{c}$ substantially changes $c_{k(i)}$.

5.4. Comparisons between BH and its variants

For multiple testing on simple nulls with independent p-values, STS (2.1) and F1–3 (2.2) are more powerful than BH [10, 30]. They are also easy to implement. Since constrained p-values are computationally costly, one may ask if they can
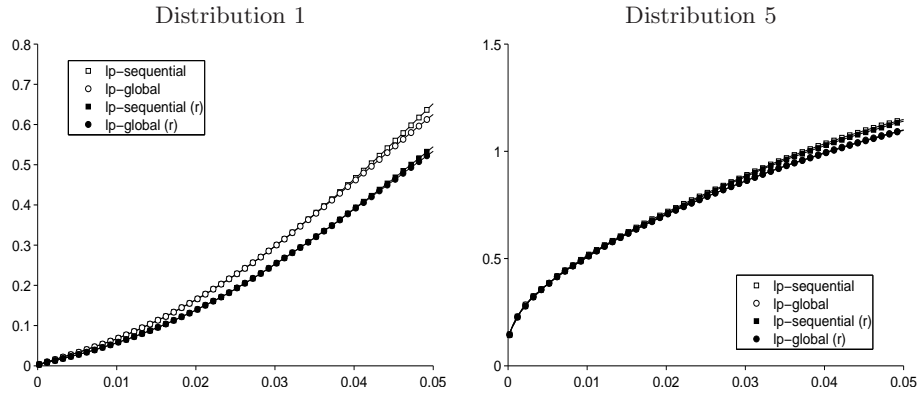
Distribution 1   Distribution 5

FIG 4. $n\bar{p}_{(i)}/i$ vs i/n, with $i/n \leq 0.05$. Plots with open symbols are for $p_{i,\mathrm{seq}}$ and $p_{i,\mathrm{glb}}$; those with closed symbols are for $p'_{i,\mathrm{seq}}$ and $p'_{i,\mathrm{glb}}$. See section 5.3 for detail.

Distribution 1

Distribution 5

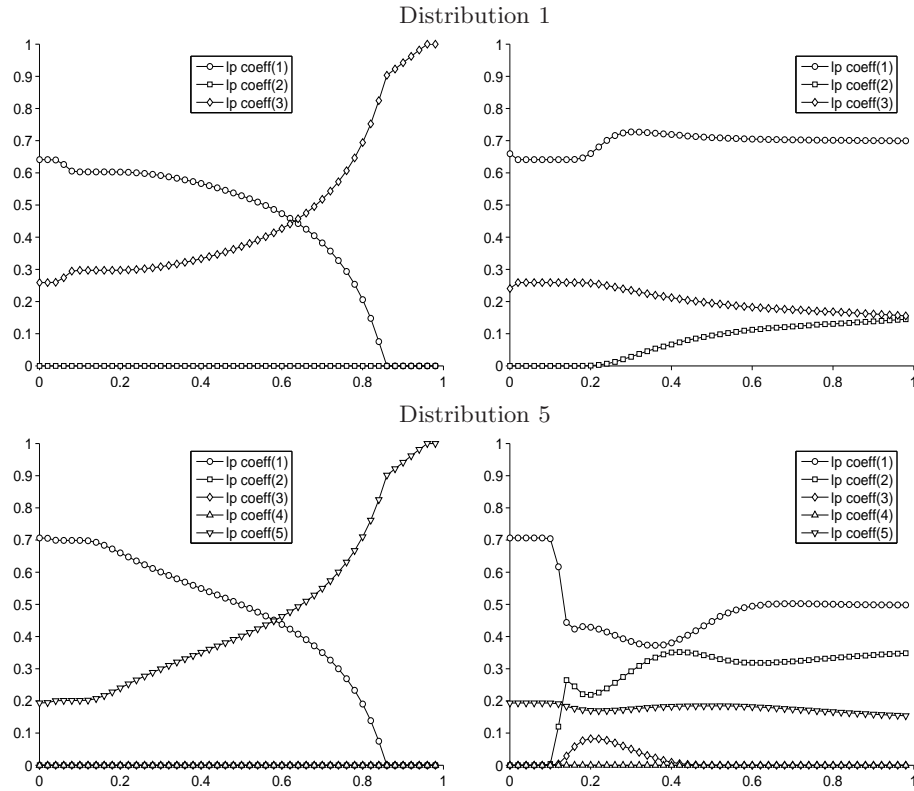FIG 5. $c_{k,(i)}$ vs i/n, where $c_{1,(i)}$, ..., $c_{L,(i)}$ are the coefficients to attain $p'_{(i),\mathrm{seq}}$ (left) or $p'_{(i),\mathrm{glb}}$ (right). See section 5.3 for detail.

TABLE 4

Powers of STS (top), F1–3 (bottom) based on different types of p-values; cf. section 5.4 for detail

	1	2	3	4	5	6	7
$p_{i,\text{seq}}$	4.72E-1	2.10E-1	4.30E-1	3.72E-4	4.03E-2	1.71E-2	2.84E-2
$p_{i,\text{glb}}$	4.89E-1	2.25E-1	4.46E-1	4.03E-4	4.15E-2	1.74E-2	2.90E-2
$p_{i,\text{max}}$	1.39E-1	1.40E-2	1.51E-1	2.36E-5	1.81E-3	2.06E-3	3.67E-3
$p_{i,\text{mix}}$	7.83E-1	6.56E-1	7.01E-1	1.92E-1	4.77E-1	2.35E-1	5.85E-1
$p_{i,\text{seq}}$	4.99E-1	2.37E-1	4.57E-1	5.46E-4	4.68E-2	1.98E-2	3.41E-2
$p_{i,\text{glb}}$	4.99E-1	2.37E-1	4.57E-1	5.46E-4	4.68E-2	1.98E-2	3.41E-2
$p_{i,\text{max}}$	2.23E-1	3.60E-2	2.32E-1	4.71E-5	3.66E-3	3.78E-3	6.22E-3
$p_{i,\text{mix}}$	7.80E-1	6.48E-1	6.97E-1	1.52E-1	4.61E-1	2.21E-1	5.77E-1

be dispensed with by STS and F1–3. In the context of the study, the question becomes, whether the power of these procedures when they are applied to $p_{i,\text{max}}$ can be as high as the power of BH when it is applied to $p_{i,\text{seq}}$ or $p_{i,\text{glb}}$.

To answer the question, we apply STS, F1, F2 and F3 to the p-values used in Section 5.2. We set $\lambda = 0.5$ in (2.1) for STS and $\kappa = 0.5$ in (2.2) for F1–3. Table 4 shows the powers of the procedures. The results from F1, F2, and F3 are identical, so they are grouped together. Comparing with Table 2, it is seen that the powers of BH, STS and F1–3 when they are applied to $p_{i,\text{max}}$ are similar, and all are substantially lower than the power of BH when it is applied to $p_{i,\text{seq}}$ and $p_{i,\text{glb}}$. The following observations can also be made. First, when these procedures are applied to $p_{i,\text{mix}}$, STS consistently has more power than the others. In contrast, when these procedures are applied to other types of the p-values, STS has the lowest power. Second, F1–3 in general has a little more power than BH when applied to each type of p-values. The results suggest that, when a is small, how p-values are defined has a stronger influence on multiple testing than how BH is modified, and variants of BH can have more or less power depending on the p-values being used.

To see how the procedures perform when the fraction of false nulls becomes larger, we repeat the simulations for moderate and large values of a, while keeping the other parameters unchanged. Tables 5 and 6 display the powers of the procedures in the simulations using the 6th set of parameters in Table 1. Since all the procedures control the FDR at or below $\alpha = 0.25$, the values of FDR are omitted. As seen from the tables, for all the values of a, the power of BH when it is applied to the constrained p-values is substantially higher then the powers of STS and F1–3 when they are applied to $p_{i,\text{max}}$. This again shows that the constrained p-values cannot be dispensed with by the variants of BH. The following observations can also be made. First, for each procedure, across the values of a, the constrained p-values yield more power than $p_{i,\text{max}}$, but less power than $p_{i,\text{mix}}$. Second, the powers of BH when it is applied to the two types of constrained p-values are similar, even when a is large. This is also the case for F2 and F3. In contrast, as a increases, STS gains more power from $p_{i,\text{glb}}$ than from $p_{i,\text{seq}}$, and even for moderate a, the gain is quite a lot. For F1, such gain in power from $p_{i,\text{glb}}$ is obvious only for large a. Third, as a increases, both STS and F1–3 become more powerful than BH when all of them are applied to

TABLE 5

Powers of BH (top), STS (middle) and F1–3 (bottom) based on different types of p-values, when $a = 0.05k$, $k = 1, \ldots, 6$. Except for a, the parameters are identical to the 6th group in Table 1. See section 5.4 for detail

a	.05	.1	.15	.2	.25	.3
$p_{i,\text{seq}}$	1.97E-2	4.56E-2	7.73E-2	1.18E-1	1.63E-1	2.05E-1
$p_{i,\text{glb}}$	1.96E-2	4.56E-2	7.73E-2	1.18E-1	1.63E-1	2.05E-1
$p_{i,\text{max}}$	3.78E-3	6.88E-3	1.13E-2	1.90E-2	2.95E-2	4.41E-2
$p_{i,\text{mix}}$	2.16E-1	4.33E-1	5.82E-1	6.83E-1	7.54E-1	8.07E-1
$p_{i,\text{seq}}$	1.71E-2	3.86E-2	6.53E-2	1.01E-1	1.41E-1	1.79E-1
$p_{i,\text{glb}}$	1.74E-2	4.16E-2	7.58E-2	1.26E-1	1.91E-1	2.61E-1
$p_{i,\text{max}}$	2.06E-3	3.14E-3	4.67E-3	7.66E-3	1.29E-2	2.14E-2
$p_{i,\text{mix}}$	2.35E-1	4.85E-1	6.60E-1	7.79E-1	8.59E-1	9.15E-1
$p_{i,\text{seq}}$	1.98E-2	4.60E-2	7.89E-2	1.23E-1	1.74E-1	2.25E-1
$p_{i,\text{glb}}$	1.98E-2	4.60E-2	7.89E-2	1.23E-1	1.74E-1	2.25E-1
$p_{i,\text{max}}$	3.78E-3	6.88E-3	1.13E-2	1.92E-2	3.01E-2	4.54E-2
$p_{i,\text{mix}}$	2.21E-1	4.56E-1	6.28E-1	7.52E-1	8.41E-1	9.04E-1

TABLE 6

Powers of the same procedures as in Table 5 when a is large. Differing values associated with F1 and F2–3 are marked by a and b respectively

a	.4	.5	.6
$p_{i,\text{seq}}$	2.96E-1	3.86E-1	4.76E-1
$p_{i,\text{glb}}$	2.96E-1	3.86E-1	4.76E-1
$p_{i,\text{max}}$	8.24E-2	1.27E-1	1.75E-1
$p_{i,\text{mix}}$	8.75E-1	9.15E-1	9.41E-1
$p_{i,\text{seq}}$	2.83E-1	4.41E-1	6.60E-1
$p_{i,\text{glb}}$	4.38E-1	6.56E-1	8.84E-1
$p_{i,\text{max}}$	5.36E-2	1.11E-1	2.02E-1
$p_{i,\text{mix}}$	9.74E-1	9.95E-1	1.00
$p_{i,\text{seq}}$	3.51E-1	5.08E-1	7.19E-1
$p_{i,\text{glb}}$	3.51E-1	5.83E-1[a], 5.08E-1[b]	1.00[a], 7.19E-1[b]
$p_{i,\text{max}}$	8.73E-2	1.41E-1	2.08E-1
$p_{i,\text{mix}}$	9.72E-1	9.95E-1	1.00

the constrained p-values. The difference is large when a is large. On the other hand, between STS and F1–3, the picture is more complicated. Depending on the value of a, each one can be more powerful than the other when applied to $p_{i,\text{glb}}$. However, when applied to $p_{i,\text{seq}}$ or $p_{i,\text{max}}$, F1–3 have more power, whereas when applied to $p_{i,\text{mix}}$, STS has more power, especially for moderate a ($\leq .3$). In summary, for moderate or large a, $p_{i,\text{glb}}$ yields more power than $p_{i,\text{seq}}$ and $p_{i,\text{max}}$, and STS and F1 tend to have more power than the other procedures.

The above results indicate that for large a, the two types of constrained p-value are no longer similar to each other. This is confirmed by Fig. 6. When $a = .5$, the plots for $p_{i,\text{seq}}$ and $p_{i,\text{glb}}$ are significantly different. It is also worth noting that in this case, for large i, $p_{(i),\text{glb}}$ are significantly smaller than $p_{(i),\text{mix}}$. The difference is also noticeable for $a = 0.4$ and even greater for $a = 0.6$ (results not shown). This suggests that, even though a cannot be estimated consistently, by exploiting the empirical distribution of observations, the linear programming for $p_{i,\text{glb}}$ can impose strong constraints on feasible values of a. Of course, it should be noted that this only occurs when a is large.
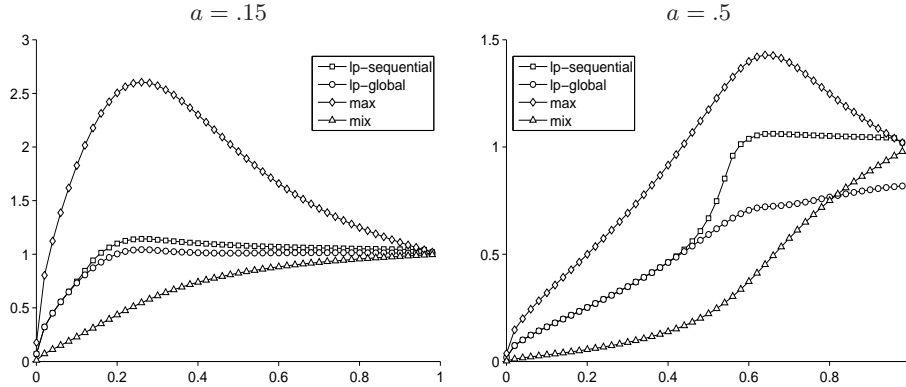
FIG 6. $n\bar{p}_{(i)}/i$ vs i/n *for different types of p-values in the simulations reported in Tables 5 and 6. The legend is identical to that of Figure 2.*

TABLE 7
Power (top) and the FDR (bottom) of BH for different sample sizes; cf. section 5.5 for detail

	Distribution 6				Distribution 7		
n	500	1000	2000	5000	500	1000	2000
$p_{i,\text{seq}}$	4.12E-2	3.10E-2	2.53E-2	1.97E-2	4.66E-2	3.77E-2	3.39E-2
$p_{i,\text{glb}}$	4.12E-2	3.10E-2	2.53E-2	1.96E-2	4.66E-2	3.77E-2	3.40E-2
$p_{i,\text{max}}$	1.66E-2	1.03E-2	6.20E-3	3.78E-3	1.46E-2	9.79E-3	6.23E-3
$p_{i,\text{mix}}$	2.61E-1	2.37E-1	2.29E-1	2.16E-1	5.68E-1	5.69E-1	5.66E-1
$p_{i,\text{seq}}$	5.71E-2	5.73E-2	5.73E-2	5.46E-2	1.11E-2	1.54E-2	1.42E-2
$p_{i,\text{glb}}$	5.71E-2	5.73E-2	5.72E-2	5.48E-2	1.11E-2	1.54E-2	1.42E-2
$p_{i,\text{max}}$	1.96E-2	1.99E-2	1.58E-3	1.39E-2	2.17E-3	5.53E-3	2.15E-3
$p_{i,\text{mix}}$	2.40E-1	2.39E-1	2.38E-1	2.37E-1	2.38E-1	2.37E-1	2.35E-1

5.5. Effects of sample size

Finally, we run simulations on BH with $n = 500, 1000, 2000,$ and 5000 for the first six distributions in Table 1, and $n = 500, 100,$ and 2000 for the last one. Table 7 shows the results for the 6th and 7th distributions. The results for the others show similar patterns. Across all values of n, the FDR is controlled. The power shows an decreasing trend when n increases. The trend is more pronounced for $p_{i,\text{seq}}$, $p_{i,\text{glb}}$ and $p_{i,\text{max}}$ than for $p_{i,\text{mix}}$.

6. Discussion

We have only considered the case where the number of null distributions is finite. Formally, it is straightforward to generalize the constrained maximization to the case where there are infinitely many null distributions. Generally speaking, however, the maximization will involve infinitely many degrees of freedom and it is unclear how to accommodate this with a finite number of observations. An alternative approach might be to partition the set of null distributions into a finite number of subsets and use the envelopes of the subsets

to compute p-values. More specifically, given a partition $\Theta_1, \ldots, \Theta_L$ of Θ, let $u_k(t) = \sup\{\phi_\theta(t) : \theta \in \Theta_k\}$ and $l_k(t) = \inf\{\phi_\theta(t) : \theta \in \Theta_k\}$. Then define, for example, $M_n(t) = \sup \boldsymbol{c}^\top \boldsymbol{u}(t)$, where the supremum is taken over $\boldsymbol{c} \in \Delta$ such that $\boldsymbol{c}^\top \boldsymbol{l}(t)$ is dominated by the empirical distribution function. One issue is how to select the partition. Too coarse partition will only yield loose constraints on c_k and too fine partition will result in many degrees of freedom. Either way, the obtained $M_n(t)$ may not be much different from unconstrained maximum significance levels.

As is known, the local FDR can be used for multiple testing [9]. For simple nulls, the local FDR is $(1 - a)f_0(x)/h(x)$, where f_0 is the density under true nulls and h the overall density of the data. If the number of null distributions is finite with densities $f_1, \ldots, f_L$, one could define a conservative local FDR as $\rho(x)/h(x)$, where $\rho(x) = \max\{\boldsymbol{c}^\top \boldsymbol{f}(x) \colon \boldsymbol{c} \in \Delta$ and $\boldsymbol{c}^\top \boldsymbol{f} \leq h\}$. Furthermore, if the dimension of the data is high, one could instead use the empirical distribution of its low-dimensional transformations to get constraints.

The article only considers the case where the data collected for different nulls are independent. Some progress has been made on multiple testing for time series [17]. Since the constraints used here are derived from empirical marginal distributions, they can also be applied to time series. At issue is the strength of the constraints, which is determined by how well the empirical marginal distribution approximates the underlying marginal distribution. For the i.i.d. case, this can be resolved by the DKW inequality. For other cases, if similar inequalities are available, then one may also get useful constraints.

Appendix

In this Appendix, the sup-norm of $f \in C(\mathbb{R})$ will be denoted by $\|f\|$.

A.1. Proof of Proposition 3.1

By Assumption D1, for any t,

$$s_i \leq t \iff X_i \in D_s \text{ for all } s > t \iff X_i \in D_t,$$

which not only gives (3) but also $\mathsf{P}\{s_i \leq t\} = \mathsf{P}\{X_i \in D_t\} = \phi_\theta(t)$ and hence (1). From (1) and the random mixture model, (2) follows.

A.2. Proof of Theorem 4.1

To prove the result, we will employ a stopping time technique and also rely on a few probability inequalities, esp. the DKW inequality [21]. Let

$$\tau = \sup\left\{t : \ M_{n,\mathrm{seq}}(t) \leq \alpha[R_n(t) \vee 1]/n\right\},$$

where $R_n(t)$ is as in (3.3). We will show that BH is equivalent to a thresholding procedure with τ as the cut-off, making it possible to apply the optional sampling theorem. Together with a few probability inequalities, this will give Theorem 4.1.

Recall that by assumption, ϕ_k and ψ are continuous. Denote $F = \boldsymbol{\nu}^\top \boldsymbol{\phi}$. Then $F \in C(\mathbb{R})$ and by Proposition 3.1, under true H_i, s_i are i.i.d. $\sim F$.

Lemma A.1. $M_{n,\text{seq}}$ *is always nondecreasing with* $M_{n,\text{seq}}(-\infty) = 0$. *Almost surely, 1)* $M_{n,\text{seq}}$ *is continuous at every t other than* $s_1, \ldots, s_n$ *and 2)* $M_{n,\text{seq}}$ *is left-continuous and has a right limit at each* s_i.

Proof. If $s < t$, then $C_{n,\text{seq}}(s) \subset C_{n,\text{seq}}(t)$ and $\phi_k(s) \le \phi_k(t)$ for every k. Therefore, $M_{n,\text{seq}}(s) \le M_{n,\text{seq}}(t)$. By $\phi_k(-\infty) = 0$, $M_{n,\text{seq}}(-\infty) = 0$.

To show 1) and 2), it suffices to show a) $M_{n,\text{seq}}$ is left-continuous and b) $M_{n,\text{seq}}$ is right-continuous at every $t \notin S = \{s_1, \ldots, s_n\}$ and has a right limit at every s_i.

a) Fix t. If $0 < t - u \ll 1$, then $[u, t)$ has no point in $T_n \cup S$. Thus $C_{n,\text{seq}}(u) = C_{n,\text{seq}}(t)$. Denote $K = C_{n,\text{seq}}(t)$. It is not hard to see that K is compact and $\boldsymbol{c}^\top \boldsymbol{\phi}(s)$ is uniformly continuous in $(\boldsymbol{c}, s) \in K \times \mathbb{R}$. Then $\sup_{\boldsymbol{c} \in K} \boldsymbol{c}^\top \boldsymbol{\phi}(s)$ is continuous in s, implying $M_{n,\text{seq}}(u) \to M_{n,\text{seq}}(t)$ as $u \uparrow t$. Thus $M_{n,\text{seq}}$ is left-continuous.

b) Since $M_{n,\text{seq}}$ is nondecreasing, it has a right limit at every t. It remains to show that at $t \notin S$, $M_{n,\text{seq}}$ is right-continuous. Given $t \notin S$, if $0 < u - t \ll 1$, then $[t, u)$ has no point in $T_n \cup S$, thus $C_{n,\text{seq}}(u) = C_{n,\text{seq}}(t)$. Then the right-continuity follows from the same argument for the left-continuity. $\square$

Lemma A.2. *(1) There is a nonrandom $b_0 > -\infty$, such that $\tau \ge b_0$ absolutely. (2) Almost surely, $\tau < \infty$. (3) Almost surely, $M_{n,\text{seq}}(\tau) \vee (\alpha/n) \le \alpha[R_n(\tau) \vee 1]/n$.*

Proof. (1) Since

$$M_{n,\text{seq}}(t) \le u(t) := \sum_k \phi_k(t)$$

and $\boldsymbol{\phi}$ is continuous with $\boldsymbol{\phi}(-\infty) = \boldsymbol{0}$, $\tau \ge b_0 := \sup\{t : u(t) \le \alpha/n\} > -\infty$. (2) Almost surely, $u := s_{(n)} \vee \max T_n < \infty$. For $t > u$, $M_{n,\text{seq}}(t) = \sup\{\boldsymbol{c}^\top \boldsymbol{\phi}(t) : \boldsymbol{c} \in \Delta'\}$. Since $(1 - a)\boldsymbol{\nu} \in \Delta'$ and $1 - a > \alpha$, for all $t \gg u$, $M_{n,\text{seq}}(t) > \alpha$, giving $\tau < \infty$. (3) follows from (1), (2) and that $M_{n,\text{seq}}$ is left-continuous and R_n is nondecreasing. $\square$

Lemma A.3. BH *based on $p_{i,\text{seq}}$ rejects H_i if and only if $s_i \le \tau$.*

Proof. By definition, BH based on $p_{i,\text{seq}}$ rejects H_i if and only if $s_i \le s_{(R)}$, where $R = \max\{i : M_{n,\text{seq}}(s_{(i)}) \le \alpha i/n\}$, with $\max \emptyset = 0$. Clearly $s_{(R)} \le \tau$, so any H_i rejected by BH has $s_i \le \tau$. On the other hand, suppose $s_i \le \tau$. Let j be the largest integer with $s_{(j)} \le \tau$. Then $M_{n,\text{seq}}(s_{(j)}) \le M_{n,\text{seq}}(\tau) \le \alpha[R_n(\tau) \vee 1]/n = \alpha j/n$, with the second inequality due to Lemma A.2. Thus $j \le R$. As $s_i \le s_{(j)} \le s_{(R)}$, H_i is rejected by BH. $\square$

Let N be the total number of true nulls. For each t, let $\mathcal{F}_t$ be the σ-field generated by s_i and η_i that are "observable" in $[t, \infty)$, i.e.,

$$\mathcal{F}_t = \mathcal{F}(N, \mathbf{1}\{s_i \ge s\}, \tilde{\eta}_i(s), \ s \ge t, \ i = 1, \ldots, n), \text{ where } \tilde{\eta}_i(s) = \begin{cases} \eta_i & \text{if } s_i \ge s, \\ \emptyset & \text{if } s_i < s, \end{cases}$$

with $\emptyset \notin \{*\} \cup \Theta$ denoting a value being missing. Then $R_n(t-) = n - \#\{i : s_i \geq t\}$ and $V_n(t-) = N - \#\{i : s_i \geq t, \eta_i \in \Theta\}$ are $\mathcal{F}_t$-measurable, where

$$V_n(t) = \sum_{i=1}^{n} \mathbf{1}\{X_i \in D_t,\, \eta_i \in \Theta\}. \tag{A.1}$$

Likewise, for $s \geq t$, $R_n(s)$ and $V_n(s)$ are $\mathcal{F}_t$-measurable. It is seen that $\{\mathcal{F}_t, t \in \mathbb{R}\}$ is a backward filtration, i.e., $\mathcal{F}_t \subset \mathcal{F}_s$ for $t > s$.

Lemma A.4. *For $t \in \mathbb{R}$, $M_{n,\mathrm{seq}}(t)$ is $\mathcal{F}_t$-measurable.*

Proof. It suffices to show that given t and $0 \leq z < 1$, $\{M_{n,\mathrm{seq}}(t) > z\} \in \mathcal{F}_t$. Let S be a dense countable subset of Δ'. Note that $S \cap C_{n,\mathrm{seq}}(t)$ may not be dense in $\Delta' \cap C_{n,\mathrm{seq}}(t)$. By the properties of $C_{n,\mathrm{seq}}(t)$ noted at the end of Section 4.2, it can be seen that $M_{n,\mathrm{seq}}(t) > z$ if and only if for some fixed rational number $r > 0$ and every k, there is $\boldsymbol{c} \in T_{r,k} = \{\boldsymbol{c} \in S : (\boldsymbol{c} - \mathbf{1}/k)^\top \boldsymbol{\phi}(t) \geq z + r - 2L/k\}$ such that $(\boldsymbol{c} - \mathbf{1}/k)_+ \in C_{n,\mathrm{seq}}(t)$, where $\mathbf{1} = (1,\ldots,1)^\top$ and $\boldsymbol{d}_+$ stands for a point with coordinates $d_i \vee 0$. That is,

$$\{M_{n,\mathrm{seq}}(t) > z\} = \bigcup_{r \in \mathbb{Q}_+} \bigcap_{k=1}^{\infty} \bigcup_{\boldsymbol{c} \in T_{r,k}} \{(\boldsymbol{c} - \mathbf{1}/k)_+ \in C_{n,\mathrm{seq}}(t)\},$$

where $\mathbb{Q}_+$ denotes the set of positive rational numbers. Since $T_{r,k}$ is countable, by the above expression, it suffices to show that for *any* $\boldsymbol{c}$, the event $\{\boldsymbol{c} \in C_{n,\mathrm{seq}}(t)\}$ belongs to $\mathcal{F}_t$, which follows easily from the definition of $C_{n,\mathrm{seq}}(t)$. □

Lemma A.5. *Fix $t \in \mathbb{R}$. If $F(t) > 0$, then for any $s \leq t$,*

$$\mathsf{E}[V_n(s-) \mid \mathcal{F}_t] = \frac{F(s)V_n(t-)}{F(t)}, \quad a.s.$$

Proof. We already know that $V_n(t-)$ is $\mathcal{F}_t$-measurable. It suffices to show

$$\mathsf{E}[\mathbf{1}_E V_n(s-)] = F(s)\mathsf{E}[\mathbf{1}_E V_n(t-)]/F(t), \quad \text{any } E \in \mathcal{F}_t. \tag{A.2}$$

As t is fixed, it is more convenient to write $\mathcal{F}_t$ as

$$\mathcal{F}_t = \mathcal{F}(N, \hat{s}_i, \hat{\eta}_i, \, i = 1,\ldots,n), \quad \text{where } \hat{s}_i = \begin{cases} s_i & \text{if } s_i \geq t, \\ -\infty & \text{if } s_i < t, \end{cases} \quad \hat{\eta}_i = \tilde{\eta}_i(t).$$

Given $E \in \mathcal{F}_t$, there is a Borel function $f(x, a_i, b_i, i = 1,\ldots,n)$ in $x \in \{1,\ldots,n\}$, $a_i \in [-\infty,\infty)$, and $b_i \in \Theta \cup \{*\} \cup \{\emptyset\}$, such that $\mathbf{1}_E = f(N, \hat{s}_i, \hat{\eta}_i, i = 1,\ldots,n)$ ([26], p. 172). Given $I, J \subset \{1,\ldots,n\}$, let f_{IJ} be a function only in a_i and b_i with $i \notin I$ as follows,

$$f_{IJ}(a_i, b_i, i \notin I) = f(|I| + |J|, a_i, b_i, i \notin I, a_k = -\infty, b_k = \emptyset, k \in I).$$

Since $\hat{s}_i, \hat{\eta}_i$ are $\mathcal{F}_t$-measurable, $f_{IJ}(\hat{s}_i, \hat{\eta}_i, i \notin I)$ is $\mathcal{F}_t$-measurable. We claim

$$\mathbf{1}_E V_n(s-) = \sum_{I \cap J = \emptyset} A_{IJ} B_I, \tag{A.3}$$

where for I, $J \in \{1, \ldots, n\}$,

$$A_{IJ} = f_{IJ}(\hat{s}_i, \hat{\eta}_i, i \notin I) \prod_{i \in J} \mathbf{1}\{s_i \geq t, \eta_i \in \Theta\} \prod_{i \notin I \cup J} \mathbf{1}\{\eta_i = *\},$$

$$B_I = \prod_{i \in I} \mathbf{1}\{s_i < t, \eta_i \in \Theta\} \times \sum_{k \in I} \mathbf{1}\{s_k < s, \eta_k \in \Theta\}$$

$$= \sum_{k \in I} \mathbf{1}\{s_k < s, \eta_k \in \Theta\} \prod_{i \in I \setminus \{k\}} \mathbf{1}\{s_i < t, \eta_i \in \Theta\}.$$

Assuming (A.3) is true for now, observe that for each fixed disjoint pair I and J, A_{IJ} and B_I are independent, because A_{IJ} is determined by $(s_i, \eta_i, i \notin I)$, while B_I is determined by $(s_i, \eta_i, i \in I)$. Furthermore, $\mathsf{E}(A_{IJ})$ is independent of the value of s, while by the continuity of F,

$$\mathsf{E}(B_I) = \sum_{k \in I} (1-a)^{|I|} F(s) F(t)^{|I|-1} = |I| \left[(1-a)F(t)\right]^{|I|} F(s)/F(t).$$

It follows that

$$\mathsf{E}[\mathbf{1}_E V_n(s-)] = \frac{LF(s)}{F(t)}, \quad \text{with} \quad L = \sum_{I \cap J = \emptyset} \mathsf{E}(A_{IJ}) |I| \left[(1-a)F(t)\right]^{|I|}.$$

In particular, letting $s = t$, it is seen that $L = \mathsf{E}[\mathbf{1}_E V_n(t-)]$ and (A.2) follows.

Finally, to prove (A.3), let $\mathcal{I} = \{i : s_i < t, \eta_i \in \Theta\}$, $\mathcal{J} = \{i : s_i \geq t, \eta_i \in \Theta\}$. Since $\mathcal{I} \cap \mathcal{J} = \emptyset$,

$$\mathbf{1}_E V_n(s-) = \sum_{I \cap J = \emptyset} \mathbf{1}_E V_n(s-) \mathbf{1}\{\mathcal{I} = I, \, \mathcal{J} = J\}.$$

Comparing to (A.3), it suffices to show that for every disjoint pair I and J,

$$\mathbf{1}_E V_n(s-) \mathbf{1}\{\mathcal{I} = I, \, \mathcal{J} = J\} = A_{IJ} B_I. \tag{A.4}$$

Since $\mathcal{I} \cup \mathcal{J} = \{i : \eta_i \in \Theta\}$, given disjoint I and J, if $\mathcal{I} = I$ and $\mathcal{J} = J$, then

$$\mathbf{1}_E = f(N, \hat{s}_i, \hat{\eta}_i, \, i = 1, \ldots, n)$$
$$= f(|I| + |J|, \hat{s}_i, \hat{\eta}_i, \, i \notin I, \hat{s}_k = -\infty, \hat{\eta}_k = \emptyset, \, k \in I) = f_{IJ}(\hat{s}_i, \hat{\eta}_i, \, i \notin I)$$

and $V_n(s-) = |\{k : s_k < s, \eta_k \in \Theta\}| = \sum_{k \in I} \mathbf{1}\{s_k < s, \eta_k \in \Theta\}$. As a result,

$$\mathbf{1}_E V_n(s-) \mathbf{1}\{\mathcal{I} = I, \, \mathcal{J} = J\}$$
$$= f_{IJ}(\hat{s}_i, \hat{\eta}_i, i \notin I) \left[\sum_{k \in I} \mathbf{1}\{s_k < s, \eta_k \in \Theta\}\right] \mathbf{1}\{\mathcal{I} = I, \, \mathcal{J} = J\}.$$

On the other hand, by the definition of A_{IJ} and B_I,

$$A_{IJ} B_I = f_{IJ}(\hat{s}_i, \hat{\eta}_i, i \notin I) \left[\sum_{k \in I} \mathbf{1}\{s_k < s, \eta_k \in \Theta\}\right]$$
$$\times \prod_{i \in I} \mathbf{1}\{s_i < t, \eta_i \in \Theta\} \prod_{i \in J} \mathbf{1}\{s_i \geq t, \eta_i \in \Theta\} \prod_{i \notin I \cup J} \mathbf{1}\{\eta_i = *\}.$$

By the definition of $\mathcal{I}$ and $\mathcal{J}$,

$$\prod_{i\in I}\mathbf{1}\left\{s_i < t,\ \eta_i \in \Theta\right\}\prod_{i\in J}\mathbf{1}\left\{s_i \geq t,\ \eta_i \in \Theta\right\}\prod_{i\notin I\cup J}\mathbf{1}\left\{\eta_i = *\right\} = 1$$

$$\Longleftrightarrow\ I \subset \mathcal{I},\ J \subset \mathcal{I},\ \text{and}\ \{1,\dots,n\}\setminus(I\cup J) \subset \{1,\dots,n\}\setminus(\mathcal{I}\cup\mathcal{J})$$

$$\Longleftrightarrow\ \mathbf{1}\left\{\mathcal{I} = I,\ \mathcal{J} = J\right\} = 1.$$

Therefore, (A.4) is proved. □

Lemma A.6. *Given $\delta > 0$, let $Z(t) = V_n(t-)/[F(t)\vee\delta]$. Then $\{Z(t),\ \mathcal{F}_t\}$ is a left-continuous backward* super*martingale.*

Proof. Since $V_n(t-)$ is left-continuous and $F(t)\vee\delta \geq \delta$ is continuous, $Z(t)$ is left-continuous. To show that $\{Z(t),\mathcal{F}_t\}$ is a backward supermartingale, it suffices to show that for any $s < t$, $\mathsf{E}[Z(s)\,|\,\mathcal{F}_t] \leq Z(t)$. If $V_n(t-) = 0$, then $V_n(s-) = 0$ for any $s \leq t$ and $\mathsf{E}[Z(s)\,|\,\mathcal{F}_t] = 0 = Z(t)$. If $V_n(t-) > 0$, then there has to be $F(t) > 0$. By Lemma A.5,

$$\mathsf{E}[Z(s)\,|\,\mathcal{F}_t] = KV_n(t-),\ \text{with}\ K = \frac{F(s)}{[F(s)\vee\delta]F(t)}.$$

By $0 \leq F(s) \leq F(t)$, $K \leq 1/[F(t)\vee\delta]$, hence showing the claim. □

In light of the Lemmas, our intention is to apply the optional sampling theorem to τ and $\{Z(t),\ \mathcal{F}_t\}$. There is one minor problem with this, i.e., the index t starts at ∞ instead of a finite value. We can get around the problem using truncation. Let b_0 be as in part (1) of Lemma A.2. For $c > b_0$, define

$$\tau_c = \sup\left\{t \leq c:\ M_{n,\mathrm{seq}}(t) \leq \alpha[R_n(t)\vee 1]/n\right\}.$$

By Lemma A.2, the set on the right hand side is nonempty, so τ_c is well-defined. Furthermore, almost surely, for all $c \gg 1$, $\tau_c = \tau$.

Lemma A.7. *For each $c > b_0$, τ_c is a stopping time of the backward filtration $\{\mathcal{F}_t,\ t \leq c\}$. Furthermore, for the $Z(t)$ in Lemma A.6, $\mathsf{E}[Z(\tau_c)] \leq (1-a)n$.*

Proof. To show that τ_c is a stopping time of the backward filtration $\mathcal{F}_t$, it suffices to show $\{\tau_c \geq t\} \in \mathcal{F}_t$ for every $t \leq c$. By the same argument for (3) of Lemma A.2, $M_{n,\mathrm{seq}}(\tau_c) \leq \alpha[R_n(\tau_c)\vee 1]/n$. Therefore, $\{\tau_c \geq t\} = \{g(s) \leq 0$ for some $s \in [t,c]\}$, where $g(s) = M_{n,\mathrm{seq}}(s) - \alpha[R_n(s)\vee 1]/n$. Since the latter event can be reduced to $\{g(t) \leq 0\}\cup\{g(s_i) \leq 0$ for some $s_i \in (t,c]\}$, it is in $\mathcal{F}_t$.

Let $X_t = -Z(c-t)$ and $\tilde{\mathcal{F}}_t = \mathcal{F}_{c-t}$ for $0 \leq t \leq c - b_0$ and $T = c - \tau_c$. By Lemma A.6, $\{X_t, \tilde{\mathcal{F}}_t, 0 \leq t \leq c - b_0\}$ is a right-continuous *sub*martingale with a last element X_{c-b_0} and T is a stopping time of $\{\tilde{\mathcal{F}}_t, 0 \leq t \leq c - b_0\}$. By the optional sampling theorem (cf. [18], Ch. 1, Thm 3.22), $\mathsf{E}(X_T) \geq \mathsf{E}(X_0)$. Consequently, $\mathsf{E}[Z(\tau_c)] \leq \mathsf{E}[Z(c)] = \mathsf{E}\{V_n(c-)/[F(c)\vee\delta]\} \leq (1-a)n$. □

We shall need a few probability inequalities. To get the first one, for each t, let

$$\mathcal{E}(t) = \{\boldsymbol{c} \in \Delta:\ \boldsymbol{c}\ \text{satisfies conditions 1) and 2)}\}$$

where the conditions are

1) $\boldsymbol{c}^\top \boldsymbol{\phi}(s_{(j)}) \le \mathbb{F}_n(s_{(j)}) + \epsilon_n$ for all $s_{(j)} \ge t$; and
2) $\mathbb{F}_n(t_2) - \mathbb{F}_n(t_1) + \epsilon_n \ge \boldsymbol{c}^\top [\boldsymbol{\phi}(t_2) - \boldsymbol{\phi}(t_1)]$ for $t_1, t_2 \in T_n$ with $t \le t_1 < t_2$.

Lemma A.8. *If* $\exp(-2n\epsilon_n^2) \le \frac{1}{2}$, *then*

$$\mathsf{P}\left\{(1-a)\boldsymbol{\nu} \notin \mathcal{E}(t) \text{ for some } t\right\} \le 2(1 + |T_n|) \exp(-2n\epsilon_n^2).$$

Proof. Recall $Q \in C(\mathbb{R})$. By the DKW inequality in [21], for any $\lambda > 0$ with $\exp(-2n\lambda^2) \le \frac{1}{2}$, $\mathsf{P}\{\sup_t [Q(t) - \mathbb{F}_n(t)] \ge \lambda\} \le \exp(-2n\lambda^2)$. By $(1-a)\boldsymbol{\nu}^\top \boldsymbol{\phi} \le Q$,

$$\mathsf{P}\left\{(1-a)\boldsymbol{\nu}^\top \boldsymbol{\phi}(t) \ge \mathbb{F}_n(t) + \lambda \text{ for some } t\right\} \le 2\exp(-2n\lambda^2).$$

On the other hand, we will show that given $x \in \mathbb{R}$,

$$\mathsf{P}\left\{\sup_{t \ge x} \{[Q(t) - Q(x)] - [\mathbb{F}_n(t) - \mathbb{F}_n(x)]\} \ge \lambda\right\} \le 2\exp(-2n\lambda^2). \tag{A.5}$$

Assuming (A.5) is true for now, it gives

$$\mathsf{P}\left\{\begin{array}{c} Q(t) - Q(t_i) \ge \mathbb{F}_n(t) - \mathbb{F}_n(t_i) + \lambda \\ \text{for some } t_i \in T_n \text{ and } t > t_i \end{array}\right\} \le 2|T_n| \exp(-2n\lambda^2).$$

Let $\lambda = \epsilon_n$. By $Q(t) - Q(t_i) \ge (1-a)\boldsymbol{\nu}^\top [\boldsymbol{\phi}(t) - \boldsymbol{\phi}(t_i)]$ for $t > t_i$, the Lemma follows.

It remains to get (A.5). Let $y = Q(x)$. By quantile transformation,

$$\sup_{t \ge x}\{[Q(t) - Q(x)] - [\mathbb{F}_n(t) - \mathbb{F}_n(x)]\}$$
$$\sim \xi = \sup_{s \ge y}\{s - y - [\mathbb{G}_n(s) - \mathbb{G}_n(y)]\},$$

where $\mathbb{G}_n$ is the empirical distribution of $U_i = Q(X_i)$ i.i.d. $\sim \text{Unif}(0,1)$. Let $V_i = U_i - y + \mathbf{1}\{U_i \le y\}$. Then V_i are i.i.d. $\sim \text{Unif}(0,1)$ and $\xi = \sup_{0 \le s \le 1-y}[s - \mathbb{G}'_n(s)]$, where $\mathbb{G}'_n$ is the empirical distribution of V_i. Applying DKW inequality to ξ, it is seen that (A.5) follows. $\square$

To get the rest of the probability inequalities we need, let

$$\mathcal{E}' = \left\{\boldsymbol{c} \in \Delta : \boldsymbol{c}^\top \boldsymbol{\phi}(s_{(i)}) \le \Gamma^*(1/n; i)/(\beta n) \text{ for all } i \le m_n\right\} \tag{A.6}$$

Lemma A.9. *Let* u_i *be defined as in* (4.3). *Then*

$$\mathsf{P}\{(1-a)\boldsymbol{\nu} \notin \mathcal{E}'\} \le \mathsf{P}\left\{Q(s_{(i)}) > u_i \text{ for some } i = 1, \dots, m_n\right\}$$
$$\le m_n \left[1/n + (\beta e^{1-\beta})^{n+1}\right].$$

Proof. We follow the comment just below (4.2) and use the notation therein. Given $\beta \in (0,1)$, let $t = 1/\beta - 1 > 0$. Then by exponential inequality (cf. [27],

p. 852, Eq. (a)) for $k \geq 1$, $\mathsf{P}\{\xi_1 + \cdots + \xi_k \leq \beta k\} \leq \{\mathsf{E}e^{t(\beta - \xi_1)}\}^k = (\beta e^{1 - \beta})^k$. For $i = 1, \ldots, m_n$, since $u_i = \bar{\Gamma}^*(1/n; i)/(\beta n)$, then

$$\mathsf{P}\left\{Q(s_{(i)}) > u_i\right\} = \mathsf{P}\left\{\frac{\sum_{k=1}^{i} \xi_k}{\sum_{k=1}^{n+1} \xi_k} > \frac{\bar{\Gamma}^*(1/n; i)}{\beta n}\right\}$$

$$\leq \mathsf{P}\left\{\sum_{k=1}^{i} \xi_k > \bar{\Gamma}^*(1/n; i)\right\} + \mathsf{P}\left\{\sum_{k=1}^{n+1} \xi_k \leq \beta n\right\} \leq 1/n + (\beta e^{1 - \beta})^{n+1},$$

where the last inequality is due to $\sum_{k=1}^{i} \xi_k \sim \mathrm{Gamma}(i, 1)$. It is then seen the second inequality claimed in the Lemma holds. Since $(1 - a)\boldsymbol{\nu} \notin \mathcal{E}'$ implies $Q(s_{(i)}) > \bar{\Gamma}^*(1/n; i)/(\beta n)$ for some $i \leq m_n$, the first inequality also holds. $\quad\square$

Proof of Theorem 4.1. First, by Lemma A.3, $R = R_n(\tau)$ and $V = V_n(\tau)$. Let $\delta = \alpha/[n(1 - a)]$. For $c > b_0$, by Lemma A.7, $\mathsf{E}\left[V_n(\tau_c-)/\{F(\tau_c) \vee \delta\}\right] \leq (1 - a)n$. Let $c \uparrow \infty$. By dominated convergence,

$$\mathsf{E}\left[\frac{V_n(\tau-)}{F(\tau) \vee \delta}\right] \leq (1 - a)n. \tag{A.7}$$

On the other hand, let $\Gamma = \{(1 - a)\boldsymbol{\nu} \in C_{n,\mathrm{seq}}(t) \text{ for all } t\}$. Then

$$\mathrm{FDR} = \mathsf{E}\left[\frac{V}{R \vee 1}\right] = \mathsf{E}\left[\frac{V_n(\tau-)}{R \vee 1}\right] + \mathsf{E}\left[\frac{V_n(\tau) - V_n(\tau-)}{R \vee 1}\right]$$

$$\leq \mathsf{E}\left[\frac{V_n(\tau-)}{R \vee 1}\,\bigg|\,\Gamma\right]\mathsf{P}(\Gamma) + \mathsf{P}(\Gamma^c) + \mathsf{E}\left[\frac{V_n(\tau) - V_n(\tau-)}{R \vee 1}\right].$$

By Lemma A.2, $[nM_{n,\mathrm{seq}}(\tau)/\alpha] \vee 1 \leq R \vee 1$. On the other hand, by the definition of $M_{n,\mathrm{seq}}$, conditional on Γ, $M_{n,\mathrm{seq}}(\tau) \geq (1 - a)F(\tau)$. Thus, by (A.7),

$$\mathsf{E}\left[\frac{V_n(\tau-)}{R \vee 1}\,\bigg|\,\Gamma\right]\mathsf{P}(\Gamma) \leq \mathsf{E}\left[\frac{V_n(\tau-)}{\{n(1 - a)F(\tau)/\alpha\} \vee 1}\,\bigg|\,\Gamma\right]\mathsf{P}(\Gamma)$$

$$\leq \frac{\alpha}{n(1 - a)}\mathsf{E}\left[\frac{V_n(\tau-)}{F(\tau) \vee \delta}\right] \leq \alpha.$$

It is easy to see that $C_{n,\mathrm{seq}}(t) \supset \mathcal{E}(t) \cap \mathcal{E}'$. Therefore, by Lemmas A.8 and A.9, $\mathsf{P}(\Gamma^c) \leq 2(1 + |T_n|)\exp(-2n\epsilon_n^2) + m_n[1/n + (\beta e^{1 - \beta})^{n+1}]$. Finally, note that $R = 0$ implies $V_n(\tau) - V_n(\tau-) = 0$ while $V_n(\tau) - V_n(\tau-) \geq 2$ implies at least two true nulls have the same value of s_i, which is a null event. Therefore, $V_n(\tau) - V_n(\tau-) \leq \mathbf{1}\{R > 0\}$ a.s. This finishes the proof. $\quad\square$

A.3. Proofs of Theorem 4.2

For each $r \geq 0$, let

$$\Gamma_r = \left\{\boldsymbol{c} \in \Delta : Q(t) - Q(s) + r \geq \boldsymbol{c}^\top[\boldsymbol{\phi}(t) - \boldsymbol{\phi}(s)], -\infty \leq s < t\right\}.$$

It is easy to see that for $0 \leq s < r$, $(1 - a)\boldsymbol{\nu} \in \Gamma_s \subset \Gamma_r$ and $\Gamma_s = \bigcap_{r > s} \Gamma_r$.

Lemma A.10. *Let $r > 0$. Then* $\mathsf{P}\{\Gamma_0 \subset C_{n,\mathrm{glb}} \subset \Gamma_r\} \to 1$ *as* $n \to \infty$.

Proof. Let $\mathcal{E}_n = \{\|\mathbb{F}_n - Q\| \le \epsilon_n/2\}$ and $\tilde{\mathcal{E}}_n = \{Q(s_{(i)}) \le u_i, \text{ all } i = 1, \ldots, m_n\}$, where u_i are defined as in (4.3). It is seen that $\mathcal{E}_n \cap \tilde{\mathcal{E}}_n$ implies $\Gamma_0 \subset C_{n,\mathrm{glb}}$. Since $Q \in C(\mathbb{R})$, the DKW inequality [21] gives $\mathsf{P}(\mathcal{E}_n^c) \le 2\exp(-n\epsilon_n^2/2)$. As $n\epsilon_n^2 \to \infty$, $\mathsf{P}(\mathcal{E}_n) \to 1$. On the other hand, by Lemma A.9 and $m_n = o(n)$, $\mathsf{P}(\tilde{\mathcal{E}}_n^c) \to 0$. It follows that $\mathsf{P}\{\Gamma_0 \subset C_{n,\mathrm{glb}}\} \ge \mathsf{P}(\mathcal{E}_n \cap \tilde{\mathcal{E}}_n) \to 1$.

Given $r > 0$, fix $C > 0$ and $\varepsilon > 0$, such that

$$\max_k[\phi_k(-C) + 1 - \phi_k(C)] + Q(-C) + 1 - Q(C) < r,$$

$$\max_k |\phi_k(s) - \phi_k(t)| + |Q(s) - Q(t)| < r, \text{ if } |s - t| < \varepsilon,$$

Denote by S the set of s_i and $\mathrm{sp}(Q)$ the support of Q [6]. Define event $\mathcal{E}_n' = \{\delta(S, [-C, C] \cap \mathrm{sp}(Q)) < \varepsilon\}$. We next show that, conditional on $\mathcal{E}_n \cap \mathcal{E}_n'$,

$$\boldsymbol{c}^\top \boldsymbol{\phi}(t) \le Q(t) + 2\epsilon_n + 2r, \quad \boldsymbol{c} \in C_{n,\mathrm{glb}}. \tag{A.8}$$

Indeed, if $t \in [-C, C] \cap \mathrm{sp}(Q)$, then there is s_i with $|t - s_i| < \varepsilon$. For $\boldsymbol{c} \in C_{n,\mathrm{glb}}$, $\boldsymbol{c}^\top \boldsymbol{\phi}(t) \le \boldsymbol{c}^\top \boldsymbol{\phi}(s_i) + \max_k |\phi_k(t) - \phi_k(s_i)|$. So by the selection of ε, $\boldsymbol{c}^\top \boldsymbol{\phi}(t) \le \mathbb{F}_n(s_i) + \epsilon_n + r \le Q(s_i) + 2\epsilon_n + r < Q(t) + 2\epsilon_n + 2r$. If $t \le -C$, then $\boldsymbol{c}^\top \boldsymbol{\phi}(t) \le \max \phi_k(-C) \le r \le Q(t) + r$. If $t \ge C$, then $\boldsymbol{c}^\top \boldsymbol{\phi}(t) \le 1 \le Q(t) + r$. Finally, if $t \in [-C, C] \setminus \mathrm{sp}(Q)$, let $z = \inf\{s \ge t : s \in \mathrm{sp}(Q) \text{ or } s = \infty\}$. Since z is either in $[-C, C] \cap \mathrm{sp}(Q)$ or greater than C, $\boldsymbol{c}^\top \boldsymbol{\phi}(t) \le \boldsymbol{c}^\top \boldsymbol{\phi}(z) \le Q(z) + 2\epsilon_n + 2r$. Since Q is continuous, $Q(z) = Q(t)$. Thus conditional on $\mathcal{E}_n \cap \mathcal{E}_n'$, (A.8) holds.

Similarly, it can be shown that if $\delta(T_n, [-C, C]) < \varepsilon$, then conditional on $\mathcal{E}_n$, $\boldsymbol{c}^\top[\boldsymbol{\phi}(t_2) - \boldsymbol{\phi}(t_1)] < Q(t_2) - Q(t_1) + 3\epsilon_n + 4r$ for $t_1 < t_2$.

As a result, $\mathcal{E}_n \cap \mathcal{E}_n'$ implies $C_{n,\mathrm{glb}} \subset \Gamma_\sigma$, where $\sigma = 3\epsilon_n + 4r$. As $n \to \infty$, $\mathsf{P}(\mathcal{E}_n \cap \mathcal{E}_n') \to 1$ and $\epsilon_n \to 0$. Since r is arbitrary, the proof is complete. $\square$

Lemma A.11. *(1) $M_{n,\mathrm{glb}}(t)$, $n \ge 1$, and $m(t)$ are continuous. (2) As $n \to \infty$,* $\|M_{n,\mathrm{glb}} - m\| \xrightarrow{\mathrm{P}} 0$.

Proof. Denote $K_n := \Delta' \cap C_{n,\mathrm{glb}}$.

(1) If $K_n = \emptyset$, then by definition, $M_{n,\mathrm{glb}}(t) \equiv 1$. On the other hand, if $K_n \ne \emptyset$, then, since K_n is compact and $\boldsymbol{\phi} \in C(\mathbb{R})$ is bounded, $\boldsymbol{c}^\top \boldsymbol{\phi}(t)$, $\boldsymbol{c} \in K_n$ as a family of functions in t are equicontinuous and uniformly bounded. It follows that $M_{n,\mathrm{glb}} \in C(\mathbb{R})$. Likewise, $m \in C(\mathbb{R})$.

(2) Given $\varepsilon > 0$, there is $r > 0$ such that $\delta(\Gamma_0, \Gamma_r) < \varepsilon$. Conditional on $\Gamma_0 \subset C_{n,\mathrm{glb}}$, $m(t) \le M_{n,\mathrm{glb}}(t)$ for all t. On the other hand, conditional on $C_{n,\mathrm{glb}} \subset \Gamma_r$, $K_n \ne \emptyset$ so for any t, there is $\boldsymbol{c}(t) \in K_n$ with $M_{n,\mathrm{glb}}(t) = \boldsymbol{c}(t)^\top \boldsymbol{\phi}(t)$. Meanwhile, there is $\boldsymbol{c}_0(t) \in \Gamma_0$ such that $|\boldsymbol{c}(t) - \boldsymbol{c}_0(t)| \le \varepsilon$. Then

$$|M_{n,\mathrm{glb}}(t) - \boldsymbol{c}_0(t)^\top \boldsymbol{\phi}(t)| \le |\boldsymbol{c}(t) - \boldsymbol{c}_0(t)||\boldsymbol{\phi}(t)| \le \sqrt{L}\varepsilon$$

$$\implies M_{n,\mathrm{glb}}(t) \le \boldsymbol{c}_0(t)^\top \boldsymbol{\phi}(t) + \sqrt{L}\varepsilon \le m(t) + \sqrt{L}\varepsilon.$$

Thus, $\{\Gamma_0 \subset C_{n,\mathrm{glb}} \subset \Gamma_r\} \subset \{0 \le M_{n,\mathrm{glb}}(t) - m(t) \le \sqrt{L}\varepsilon \text{ all } t\}$. Because ε is arbitrary, by Lemma A.10, $\|M_{n,\mathrm{glb}} - m\| \xrightarrow{\mathrm{P}} 0$. $\square$

Proof of Theorem 4.2. The proof follows closely the one in [12]. By the continuity of m and Q, for any $0 < \epsilon < t_* - t_0$,

$$\delta = \min\left\{\inf_{t\in(t_0+\epsilon,t_*-\epsilon)}[\alpha Q(t) - m(t)],\ \inf_{t>t_*+\epsilon}[m(t) - \alpha Q(t)]\right\} > 0.$$

Let $Q_n(t) = [R_n(t) \vee 1]/n$. As $n \to \infty$, by $\|Q_n - Q\| \xrightarrow{\mathrm{P}} 0$ and $\|M_n - m\| \xrightarrow{\mathrm{P}} 0$, the probability that

$$\min\left\{\inf_{t\in(t_0+\epsilon,t_*-\epsilon)}[\alpha Q_n(t) - M_n(t)],\ \inf_{t>t_*+\epsilon}[M_n(t) - \alpha Q_n(t)]\right\} \geq \delta/2$$

tends to 1, implying $\mathsf{P}\{|\tau - t_*| \leq \epsilon\} \to 1$. Therefore, $\tau \xrightarrow{\mathrm{P}} t_*$, which leads to the last claim of the theorem. Since $\alpha Q(t) > m(t) \geq 0$ for some $t < t_*$, we have $Q(t_*) \geq Q(t) > 0$. Thus, by the Weak Law of Large Numbers and dominated convergence,

$$\mathrm{FDR} = \mathsf{E}\left[\frac{V_n(\tau)/n}{Q_n(\tau)}\right] \to \frac{(1-a)\boldsymbol{\nu}^\top \boldsymbol{\phi}(t_*)}{Q(t_*)} \leq \frac{m(t_*)}{Q(t_*)} = \alpha,$$

where the last equality is due to the continuity of m and Q at t_*. □

References

[1] BENJAMINI, Y. AND HOCHBERG, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 1, 289–300. MR1325392

[2] BENJAMINI, Y. AND HOCHBERG, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Educ. Behav. Statist.* **25**, 1, 60–83.

[3] BENJAMINI, Y. AND YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29**, 4, 1165–1188. MR1869245

[4] BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A., AND STONE, C. J. (1984). *Classification and regression trees.* Wadsworth Statistics/Probability Series. Wadsworth Advanced Books and Software, Belmont, CA. MR0726392

[5] CHI, Z. (2008). False discovery rate control with multivariate *p*-values. *Electron. J. Statist.* **2**, 368–411. MR2411440

[6] CHOW, Y. S. AND TEICHER, H. (1997). *Probability theory: independence, interchangeability, martingales*, 3 ed. Springer Texts in Statistics. Springer-Verlag, New York. MR1476912

[7] EFRON, B. (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *J. Amer. Statist. Assoc.* **99**, 465 (Mar.), 96–104. MR2054289

[8] EFRON, B. (2008). Microarrays, empirical Bayes and the two-groups model. *Statist. Sci.* **23**, 1, 1351–1377. MR2431866

[9] EFRON, B., TIBSHIRANI, R., STOREY, J. D., AND TUSHER, V. G. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* **96**, 456, 1151–1160. MR1946571

[10] FINNER, H., DICKHAUS, T., AND ROTERS, M. (2009). On the false discovery rate and an asymptotically optimal rejection curve. *Ann. Statist.* **37**, 2, 596–618. MR2502644

[11] FINNER, H. AND ROTERS, M. (1998). Asymptotic comparison of step-down and step-up multiple test procedures based on exchangeable test statistics. *Ann. Statist.* **26**, 2, 505–524. MR1626043

[12] GENOVESE, C. AND WASSERMAN, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *J. R. Stat. Soc. Ser. B* **64**, 3, 499–517. MR1924303

[13] GENOVESE, C. AND WASSERMAN, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.* **32**, 3, 1035–1061. MR2065197

[14] GENOVESE, C. AND WASSERMAN, L. (2006). Exceedance control of the false discovery proportion. *J. Amer. Statist. Assoc.* **101**, 476, 1408–1417. MR2279468

[15] HASTIE, T., TIBSHIRANI, R., AND FRIEDMAN, J. (2001). *The elements of statistical learning.* Springer Series in Statistics. Springer-Verlag, New York. Data mining, inference, and prediction. MR1851606

[16] HOCHBERG, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75**, 4, 800–802. MR0995126

[17] JIN, J. AND CAI, T. T. (2007). Estimating the null and the proportional of nonnull effects in large-scale multiple comparisons. *J. Amer. Statist. Assoc.* **102**, 478, 495–506. MR2325113

[18] KARATZAS, I. AND SHREVE, S. E. (1991). *Brownian motion and stochastic calculus*, Second ed. Graduate Texts in Mathematics, Vol. **113**. Springer-Verlag, New York. MR1121940

[19] LEHMANN, E. L. AND ROMANO, J. P. (2005). Generalizations of the familywise error rate. *Ann. Statist.* **33**, 3, 1138–1154. MR2195631

[20] LEHMANN, E. L., ROMANO, J. P., AND SHAFFER, J. P. (2005). On optimality of stepdown and stepup multiple test procedures. *Ann. Statist.* **33**, 3, 1084–1108. MR2195629

[21] MASSART, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.* **18**, 3, 1269–1283. MR1062069

[22] R DEVELOPMENT CORE TEAM. (2005). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

[23] SARKAR, S. K. (2008). Two-stage stepup procedures controlling FDR. *J. Statist. Plann. & Inference* **138**, 4, 1072–1084. MR2384506

[24] SHAFFER, J. P. (2004). Optimality results in multiple hypothesis testing. In *The First Erich L. Lehmann Symposium – Optimality*. IMS Lecture NOTEs Monogr. Ser., Vol. **44**. Inst. Math. Statist., Beachwood, OH, 11–35. MR2118559

[25] SHAFFER, J. P. (2006). Recent developments towards optimality in multiple hypothesis testing. In *Optimality*. IMS Lecture NOTEs Monogr. Ser., Vol. **49**. Inst. Math. Statist., Beachwood, OH, 16–32. MR2337828

[26] SHIRYAYEV, A. N. (1984). *Probability*. Graduate Texts in Mathematics, Vol. **95**. Springer-Verlag, New York. MR0737192

[27] SHORACK, G. R. AND WELLNER, J. A. (1986). *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York. MR0838963

[28] SIMES, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73**, 3, 751–754. MR0897872

[29] STOREY, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B* **64**, 3, 479–498. MR1924302

[30] STOREY, J. D., TAYLOR, J. E., AND SIEGMUND, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Stat. Soc. Ser. B* **66**, 1, 187–205. MR2035766

[31] VAN DER LAAN, M. J., DUDOIT, S., AND POLLARD, K. S. (2004). Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. *Stat. Appl. Genet. Mol. Biol. 3*, Art. 15, 27 pp. (electronic). MR2101464