The Annals of Statistics
2014, Vol. 42, No. 5, 2003–2025
DOI: 10.1214/14-AOS1249

PHASE TRANSITION AND REGULARIZED BOOTSTRAP IN LARGE-SCALE t-TESTS WITH FALSE DISCOVERY RATE CONTROL

BY WEIDONG LIU[1] AND QI-MAN SHAO[2]

Shanghai Jiao Tong University and The Chinese University of Hong Kong

Applying the Benjamini and Hochberg (B–H) method to multiple Student's t tests is a popular technique for gene selection in microarray data analysis. Given the nonnormality of the population, the true p-values of the hypothesis tests are typically unknown. Hence it is common to use the standard normal distribution $N(0,1)$, Student's t distribution t_{n-1} or the bootstrap method to estimate the p-values. In this paper, we prove that when the population has the finite 4th moment and the dimension m and the sample size n satisfy $\log m = o(n^{1/3})$, the B–H method controls the false discovery rate (FDR) and the false discovery proportion (FDP) at a given level α asymptotically with p-values estimated from $N(0,1)$ or t_{n-1} distribution. However, a phase transition phenomenon occurs when $\log m \geq c_0 n^{1/3}$. In this case, the FDR and the FDP of the B–H method may be larger than α or even converge to one. In contrast, the bootstrap calibration is accurate for $\log m = o(n^{1/2})$ as long as the underlying distribution has the sub-Gaussian tails. However, such a light-tailed condition cannot generally be weakened. The simulation study shows that the bootstrap calibration is very conservative for the heavy tailed distributions. To solve this problem, a regularized bootstrap correction is proposed and is shown to be robust to the tails of the distributions. The simulation study shows that the regularized bootstrap method performs better than its usual counterpart.

1. Introduction. Multiple Student's t tests often arise in many real applications, such as gene selection. Consider m tests on the mean values

$$H_{0i} : \mu_i = 0 \quad \text{versus} \quad H_{1i} : \mu_i \neq 0, \qquad 1 \leq i \leq m.$$

A popular procedure is to use the Benjamini and Hochberg (B–H) method to search for significant findings, with the false discovery rate (FDR) controlled at a given

Received October 2013; revised May 2014.

[1]Supported by NSFC, Grant No. 11201298 and No. 11322107, the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning, Shanghai Pujiang Program, Foundation for the Author of National Excellent Doctoral Dissertation of PR China and a grant from Australian Research Council.

[2]Supported in part by Hong Kong RGC GRF 603710 and 403513.

MSC2010 subject classifications. 62H15.

Key words and phrases. Bootstrap correction, false discovery rate, multiple t-tests, phase transition.

level $0 < \alpha < 1$; that is,

$$\mathsf{E}\left[\frac{V}{R \vee 1}\right] \le \alpha,$$

where V is the number of wrongly rejected hypotheses and R is the total number of rejected hypotheses. The seminal work of Benjamini and Hochberg (1995) is to reject the null hypotheses for which $p_i \le p_{(\hat{k})}$, where p_i is the p-value for H_{0i},

$$\hat{k} = \max\{0 \le i \le m : p_{(i)} \le \alpha i/m\} \tag{1}$$

and $p_{(1)} \le \cdots \le p_{(m)}$ are the ordered p-values. Let $T_1, \dots, T_m$ be Student's t test statistics

$$T_i = \frac{\bar{X}_i}{\hat{s}_{ni}/\sqrt{n}},$$

where

$$\bar{X}_i = \frac{1}{n}\sum_{k=1}^{n} X_{ki}, \qquad \hat{s}_{ni}^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_{ki} - \bar{X}_i)^2,$$

and $(X_{k1}, \dots, X_{km})'$, $1 \le k \le n$, are i.i.d. random samples from $(X_1, \dots, X_m)'$. When $T_1, \dots, T_m$ are independent and the true p-values p_i are known, Benjamini and Hochberg (1995) showed that the B–H method controls the FDR at level α.

In many applications, the distributions of X_i, $1 \le i \le m$, are non-Gaussian. Hence it is difficult to know the exact null distributions of T_i and the true p-values. When applying the B–H method, the p-values are actually estimators. According to the central limit theorem, it is common to use the standard normal distribution $N(0, 1)$ or Student's t distribution t_{n-1} to estimate the p-values, where t_{n-1} denotes the Student's t random variable with $n-1$ degrees of freedom. In a microarray analysis, Efron (2004) observed that the null distribution choices substantially affect the simultaneous inference procedure. However, a systematic theoretical study on the influence of the estimated p-values is still lacking. It is important to know how accurate $N(0, 1)$ and t_{n-1} calibrations can be. In this paper, we show that $N(0, 1)$ and t_{n-1} calibrations are accurate when $\log m = o(n^{1/3})$. Moreover, if the underlying distributions are symmetric, then the dimension can be as large as $\log m = o(n^{1/2})$. Under the finite 4th moment of X_i, the FDR and the false discovery proportion (FDP) of the B–H method with the estimated p-values $\hat{p}_{i,\Phi} = 2 - 2\Phi(|T_i|)$ or $\hat{p}_{i,\Psi_{n-1}} = 2 - 2\Psi_{n-1}(|T_i|)$ will converge to $\alpha m_0/m$, where m_0 is the number of true null hypotheses, $\Phi(t)$ is the standard normal distribution and $\Psi_{n-1}(t) = \mathsf{P}(t_{n-1} \le t)$. However, when $\log m \ge c_0 n^{1/3}$ for some $c_0 > 0$ and the distributions are asymmetric, $N(0, 1)$ and t_{n-1} calibrations may not work well, and a phase transition phenomenon occurs. Under $\log m \ge c_0 n^{1/3}$, the number of true alternative hypotheses $m_1 = \exp(o(n^{1/3}))$ and the average of skewnesses $\tau = \underline{\lim}_{m\to\infty} m_0^{-1} \sum_{i \in \mathcal{H}_0} |\mathsf{E}X_i^3/\sigma_i^3| > 0$, we show that the FDR of the B–H method satisfies $\lim_{(m,n)\to\infty} \mathrm{FDR} \geq \kappa$ for some constant $\kappa > \alpha$, where $\mathcal{H}_0 = \{i : \mu_i = 0\}$. Furthermore, if $\log m/n^{1/3} \to \infty$, then $\lim_{(m,n)\to\infty} \mathrm{FDR} = 1$. Similar results are proven for the false discovery proportion. This indicates that $N(0, 1)$ and t_{n-1} calibrations are inaccurate when the average of skewnesses $\tau \neq 0$ in the ultra high dimensional setting.

It is well known that the bootstrap is an effective way to improve the accuracy of an exact null distribution approximation. Fan, Hall and Yao (2007) showed that for the bounded noise, the bootstrap can improve the accuracy and allow a higher dimension $\log m = o(n^{1/2})$ on controlling the family-wise error rate. Delaigle, Hall and Jin (2011) showed that the bootstrap method has significant advantages in higher criticism. In this paper, we show that when the bootstrap calibration is used and $\log m = o(n^{1/2})$, the B–H method can asymptotically control FDR and FDP at level α. In our results, we assume the sub-Gaussian tails instead of the bounded noise in Fan, Hall and Yao (2007).

Although the bootstrap method allows for a higher dimension, the light-tailed condition cannot generally be weakened. The simulation study shows that the bootstrap method is very conservative for the heavy-tailed distributions. To solve this problem, we propose a regularized bootstrap method that is robust to the tails of the distributions. The proposed regularized bootstrap only requires a finite 6th moment, and the dimension can be as large as $\log m = o(n^{1/2})$.

It is also not uncommon in real applications for $X_1, \dots, X_m$ to be dependent. This results in a dependency between $T_1, \dots, T_m$. In this paper, we obtain some similar results for the B–H method under a general weak dependence condition. It should be noted that much work has been done on the robustness of the FDR/FDP controlling method against dependence. Benjamini and Yekutieli (2001) proved that the B–H procedure controlled FDR under positive regression dependency. Storey (2003), Storey, Taylor and Siegmund (2004) and Ferreira and Zwinderman (2006) imposed a dependence condition that required the law of large numbers for the empirical distributions under the null and alternative hypothesis. Wu (2008) developed FDR controlling procedures for the data coming from special models, such as the time series model. However, to satisfy the conditions in most of the existing methods, it is often necessary to assume that the number of true alternative hypotheses m_1 is asymptotically $\pi_1 m$ with some $\pi_1 > 0$. They exclude the sparse setting $m_1 = o(m)$, which is important in applications such as gene selection. For example, if $m_1 = o(m)$, then the conditions of Theorem 4 in Storey, Taylor and Siegmund (2004), and the conditions of the main results in Wu (2008) are not satisfied. In contrast, our results on FDR and FDP control under dependence allows $m_1 \leq \gamma m$ for some $\gamma < 1$.

The remainder of this paper is organized as follows. In Section 2.1, we show the robustness of and the phase transition phenomenon for the $N(0, 1)$ and t_{n-1} calibrations. In Section 2.2, we show that the bootstrap calibration can improve the FDR and FDP control. The regularized bootstrap method is proposed in Section 3. The results are extended to the dependence case in Section 4. The simulation study is presented in Section 5 and the proofs are postponed to Section 6. Throughout the paper, all constants such as γ, b_0, c_0 in the upper bounds and lower bounds do not depend on n and m.

2. Main results.

2.1. *Robustness and phase transition.* In this section, we assume that the Student's t test statistics $T_1, \dots, T_m$ are independent, and the results are extended to the dependent case in Section 4. Before stating the main theorems, we introduce some notation. Let $\hat{p}_{i,\Phi} = 2 - 2\Phi(|T_i|)$ and $\hat{p}_{i,\Psi_{n-1}} = 2 - 2\Psi_{n-1}(|T_i|)$ be the p-values calculated from the standard normal distribution and the t-distribution, respectively. Let FDR_Φ and $\mathrm{FDR}_{\Psi_{n-1}}$ be the FDR of the B–H method with $\hat{p}_{i,\Phi}$ and $\hat{p}_{i,\Psi_{n-1}}$ in (1), respectively. Similarly, we denote the false discovery proportions of the B–H method by FDP_Φ $(= \frac{V}{R\vee 1})$ and $\mathrm{FDP}_{\Psi_{n-1}}$. Recall that R is the total number of rejections. The critical values of the tests are then $\hat{t}_\Phi = \Phi^{-1}(1-\alpha R/(2m))$ and $\hat{t}_{\Psi_{n-1}} = \Psi_{n-1}^{-1}(1-\alpha R/(2m))$. Set $Y_i = (X_i - \mu_i)/\sigma_i$ with $\sigma_i^2 = \mathsf{Var}(X_i)$, $1 \le i \le m$.

Recall that m_1 is the number of true alternative hypotheses. Throughout this paper, we assume $m_1 \le \gamma m$ for some $\gamma < 1$, which includes the important sparse setting $m_1 = o(m)$.

THEOREM 2.1. *Suppose $X_1, \dots, X_m$ are independent and $\log m = o(n^{1/2})$. Assume that $\max_{1\le i\le m} \mathsf{E}Y_i^4 \le b_0$ for some constant $b_0 > 0$ and*

$$\mathrm{Card}\{i : |\mu_i/\sigma_i| \ge 4\sqrt{\log m/n}\} \to \infty. \tag{2}$$

Then

$$\lim_{(n,m)\to\infty} \frac{\mathrm{FDR}_\Phi}{(m_0/m)\alpha\kappa_\Phi} = 1 \quad \textit{and} \quad \lim_{(n,m)\to\infty} \frac{\mathrm{FDR}_{\Psi_{n-1}}}{(m_0/m)\alpha\kappa_{\Psi_{n-1}}} = 1,$$

where

$$\kappa_\Phi = \mathsf{E}\big[\hat{\kappa}_\Phi I\{\hat{\kappa}_\Phi \le 2(\alpha - \alpha\gamma)^{-1}\}\big],$$
$$\hat{\kappa}_\Phi = \frac{\sum_{i\in\mathcal{H}_0}\{\exp(\hat{t}_\Phi^3 \mathsf{E}X_i^3/(\sqrt{n}\sigma_i^3)) + \exp(-\hat{t}_\Phi^3 \mathsf{E}X_i^3/(\sqrt{n}\sigma_i^3))\}}{2m_0}$$

and $\kappa_{\Psi_{n-1}}$ is defined in the same way. For the false discovery proportion, we have

$$\frac{\mathrm{FDP}_\Phi}{(m_0/m)\alpha\hat{\kappa}_\Phi} \to 1 \quad \textit{and} \quad \frac{\mathrm{FDP}_{\Psi_{n-1}}}{(m_0/m)\alpha\hat{\kappa}_{\Psi_{n-1}}} \to 1$$

in probability as $(n,m)\to\infty$.

Let $\tau = \underline{\lim}_{m\to\infty} m_0^{-1}\sum_{i\in\mathcal{H}_0}|\mathsf{E}Y_i^3|$. We have the following corollary.

COROLLARY 2.1. *Assume that the conditions in Theorem 2.1 are satisfied.*

(i) *If* $\log m = o(n^{1/3})$, *then we have*

$$\lim_{(n,m)\to\infty} \mathrm{FDR}_{\Phi}/(\alpha m_0/m) = 1 \quad \textit{and} \quad \mathrm{FDP}_{\Phi}/(\alpha m_0/m) \to 1 \qquad \textit{in probability}.$$

(ii) *Suppose* $\log m \geq c_0 n^{1/3}$ *for some* $c_0 > 0$ *and* $m_1 = \exp(o(n^{1/3}))$. *Also assume that* $\tau > 0$. *Then* $\underline{\lim}_{(n,m)\to\infty} \mathrm{FDR}_{\Phi} \geq \beta$ *and* $\lim_{(n,m)\to\infty} \mathsf{P}(\mathrm{FDP}_{\Phi} \geq \beta) = 1$ *for some constant* $\beta > \alpha$.

(iii) *Suppose* $\log m/n^{1/3} \to \infty$ *and* $m_1 = \exp(o(n^{1/3}))$. *Assume that* $\tau > 0$. *Then we have* $\lim_{(n,m)\to\infty} \mathrm{FDR}_{\Phi} = 1$ *and* $\mathrm{FDP}_{\Phi} \to 1$ *in probability.*

The same conclusions hold for $\mathrm{FDR}_{\Psi_{n-1}}$ *and* $\mathrm{FDP}_{\Psi_{n-1}}$.

Theorem 2.1 and Corollary 2.1 show that when $\log m = o(n^{1/3})$, $N(0,1)$ and t_{n-1} calibrations are accurate. Note that only a finite 4th moment of Y_i is required. Furthermore, if the skewnesses $\mathsf{E}Y_i^3 = 0$ for $i \in \mathcal{H}_0$, then the dimension can be as large as $\log m = o(n^{1/2})$. However, a phase transition occurs if the average of skewnesses $\tau > 0$, for example, for the exponential distribution. The FDR and FDP of the B–H method are greater than α as long as $\log m \geq c_0 n^{1/3}$ and converge to one when $\log m/n^{1/3} \to \infty$.

Under a finite 4th moment of X_i, Cao and Kosorok (2011) prove the robustness of Student's t test statistics and $N(0,1)$ calibration in the control of FDR and FDP. They require $m_1/m \to c$ for some $0 < c < 1$, which does not cover the sparse case.

Corollary 2.1 also indicates that the choice of asymptotic null distributions is important in the study of large-scale testing problems. When the dimension is much larger than the sample size, simply using the null limiting distribution to estimate the true p-values may result in larger FDR and FDP. This is further verified by our simulation study in Section 5.

In Theorem 2.1 and Corollary 2.1, we require technical condition (2). Actually, this condition is nearly optimal for the FDP results. If the number of true alternative hypotheses m_1 is fixed as $m \to \infty$, then Proposition 2.1 below shows that even for the true p-values, the B–H method is unable to control FDP at any level $0 < \xi < 1$ with overwhelming probability. Note that (2) is only slightly stronger than $m_1 \to \infty$.

Let $\mathrm{FDP}_{\mathrm{true}}$ be the false discovery proportion of the B–H method, with the true p-values p_i, $1 \leq i \leq m$. Let $U(0,1)$ be the uniform random variable on $(0,1)$.

PROPOSITION 2.1. *Assume that* m_1 *is fixed as* $m \to \infty$ *and* $X_1, \ldots, X_m$ *are independent. Suppose that* $p_i \sim U(0,1)$ *for* $i \in \mathcal{H}_0$. *For any* $0 < \xi < 1$, *we have*

$$\underline{\lim}_{(n,m)\to\infty} \mathsf{P}(\mathrm{FDP}_{\mathrm{true}} \geq \xi) \geq \eta$$

for some $\eta > 0$, *where* η *may depend on* m_1 *and* ξ.

Proposition 2.1 indicates that $m_1 \to \infty$ is a necessary condition for FDP control. In contrast, the control of FDR does not need $m_1 \to \infty$ when $\log m = o(n^{1/3})$. However, FDR_Φ and $\mathrm{FDR}_{\Psi_{n-1}}$ may still converge to one if $\log m/n^{1/3} \to \infty$ and $\tau > 0$.

PROPOSITION 2.2. *Suppose m_1 is fixed as $m \to \infty$, $X_1, \ldots, X_m$ are independent and $\log m = o(n^{1/2})$. Assume that $\max_{1\le i\le m} \mathsf{E}Y_i^4 \le b_0$ for some constant $b_0 > 0$.*

(i) *If $\log m = o(n^{1/3})$ and $p_i \sim U(0,1)$ for $i \in \mathcal{H}_0$, then* $\overline{\lim}_{(n,m)\to\infty} \mathrm{FDR}_\Phi \le \alpha$.

(ii) *Suppose $\log m/n^{1/3} \to \infty$. Assume that $\tau > 0$. We have* $\lim_{(n,m)\to\infty} \mathrm{FDR}_\Phi = 1$.

The same conclusions remain valid for $\mathrm{FDR}_{\Psi_{n-1}}$.

2.2. *Bootstrap calibration.* In this section, we show that the bootstrap procedure can improve the accuracy of FDR and FDP control. Write $\mathcal{X}_i = \{X_{1i}, \ldots, X_{ni}\}$. Let $\mathcal{X}_{ki}^* = \{X_{1ki}^*, \ldots, X_{nki}^*\}$, $1 \le k \le N$, be resamples drawn randomly with replacement from $\mathcal{X}_i$. Let T_{ki}^* be Student's t test statistics constructed from $\{X_{1ki}^* - \bar{X}_i, \ldots, X_{nki}^* - \bar{X}_i\}$. We use $G_{N,m}^*(t) = \frac{1}{Nm}\sum_{k=1}^N \sum_{i=1}^m I\{|T_{ki}^*| \ge t\}$ to approximate the null distribution and define the p-values by $\hat{p}_{i,B} = G_{N,m}^*(|T_i|)$. Let FDR_B and FDP_B denote the FDR and FDP of the B–H method with $\hat{p}_{i,B}$ in (1), respectively.

THEOREM 2.2. *Suppose that $\max_{1\le i\le m} \mathsf{E}e^{tY_i^2} \le K$ for some constants $t > 0$ and $K > 0$, and the conditions in Theorem 2.1 are satisfied.*

(i) *If $\log m = o(n^{1/3})$, then we have*

$$\lim_{(n,m)\to\infty} \mathrm{FDR}_B/(\alpha m_0/m) = 1 \quad \text{and} \quad \mathrm{FDP}_B/(\alpha m_0/m) \to 1 \quad \text{in probability.} \tag{3}$$

(ii) *If $\log m = o(n^{1/2})$ and $m_1 \le m^\eta$ for some $\eta < 1$, then (3) holds.*

Another common bootstrap method is to estimate the p-values individually by $\breve{p}_{i,B} = G_i^*(T_i)$, where $G_i^*(t) = \frac{1}{N}\sum_{k=1}^N I\{T_{ki}^* \ge t\}$; see Fan, Hall and Yao (2007) and Delaigle, Hall and Jin (2011). Similar results to those achieved in Theorem 2.2 can be obtained if N is large enough. Let $\mathrm{FDR}_{\breve{B}}$ and $\mathrm{FDP}_{\breve{B}}$ be the FDR and FDP of the B–H method with $\breve{p}_{i,B}$, respectively. The following result holds.

PROPOSITION 2.3. *Suppose that $N \ge m^{2+\delta}$ for some $\delta > 0$, $\max_{1\le i\le m} \mathsf{E}e^{tY_i^2} \le K$ for some constants $t > 0$ and $K > 0$, and $\log m = o(n^{1/2})$. Assume that $X_1, \ldots, X_m$ are independent.*

(i) *If* (2) *holds, then the results of Theorem* 2.2(i) *and* (ii) *hold for* $\mathrm{FDR}_{\breve{B}}$ *and* $\mathrm{FDP}_{\breve{B}}$.

(ii) *Suppose that* m_1 *is fixed and* $p_i \sim U(0,1)$ *for* $i \in \mathcal{H}_0$. *If* $\log m = o(n^{1/2})$, *then we have* $\overline{\lim}_{(n,m)\to\infty} \mathrm{FDR}_{\breve{B}} \le \alpha$.

Fan, Hall and Yao (2007) proved that the bootstrap calibration accurately controls the family-wise error rate if $\log m = o(n^{1/2})$ and $\mathsf{P}(|Y_i| \le C) = 1$ for $1 \le i \le m$. Our result on FDR control only requires the sub-Gaussian tails, which is a weaker requirement than the bounded noise.

The bootstrap method has often been used in multiple Student's t tests in real applications. Fan, Hall and Yao (2007) and Delaigle, Hall and Jin (2011) have proven that the bootstrap method provides more accurate p-values than the normal or t_{n-1} approximation for the light-tailed distributions. Theorem 2.2 and Proposition 2.3 show that the bootstrap method allows a higher dimension $\log m = o(n^{1/2})$ for FDR control as long as $\max_{1\le i\le m} \mathsf{E}e^{tY_i^2} \le K$. However, some real data may not satisfy such a light-tailed condition. The simulation study in Section 5 also indicates that the bootstrap calibration does not always outperform the $N(0,1)$ or t_{n-1} calibrations.

3. Regularized bootstrap in large-scale tests. In this section, we introduce a regularized bootstrap method that is robust for heavy-tailed distributions, and the dimension m can be as large as $e^{o(n^{1/2})}$. For the regularized bootstrap method, the finite 6th moment condition is enough. Let $\lambda_{ni} \to \infty$ be a regularization parameter. Define

$$\hat{X}_{ki} = X_{ki} I\{|X_{ki}| \le \lambda_{ni}\}, \qquad 1 \le k \le n, 1 \le i \le m.$$

Write $\hat{\mathcal{X}}_i = \{\hat{X}_{1i}, \dots, \hat{X}_{ni}\}$. Let $\hat{\mathcal{X}}_{ki}^* = \{\hat{X}_{1ki}^*, \dots, \hat{X}_{nki}^*\}$, $1 \le k \le N$, be resamples drawn independently and uniformly with replacement from $\hat{\mathcal{X}}_i$. Let $\hat{T}_{ki}^*$ be Student's t test statistics constructed from $\{\hat{X}_{1ki}^* - \hat{X}_i, \dots, \hat{X}_{nki}^* - \hat{X}_i\}$, where $\hat{X}_i = \frac{1}{n}\sum_{k=1}^n \hat{X}_{ki}$. We use $\hat{G}_{N,m}^*(t) = \frac{1}{Nm}\sum_{k=1}^N \sum_{i=1}^m I\{|\hat{T}_{ki}^*| \ge t\}$ to approximate the null distribution and define the p-values by $\hat{p}_{i,RB} = \hat{G}_{N,m}^*(|T_i|)$. Let FDR_{RB} and FDP_{RB} be the FDR and FDP of the B–H method with $\hat{p}_{i,RB}$ in (1), respectively.

THEOREM 3.1. *Assume that* $\max_{1\le i\le m} \mathsf{E}X_i^6 \le K$ *for some constant* $K > 0$. *Suppose* $X_1, \dots, X_m$ *are independent*, (2) *holds and* $\min_{1\le i\le m} \sigma_{ii} \ge c_1$ *for some* $c_1 > 0$. *Let* $c_2(n/\log m)^{1/6} \le \lambda_{ni} \le c_3(n/\log m)^{1/6}$ *for some* $c_2, c_3 > 0$.

(i) *If* $\log m = o(n^{1/3})$, *then*

$$\lim_{(n,m)\to\infty} \mathrm{FDR}_{RB}/(\alpha m_0/m) = 1 \quad \textit{and} \quad \mathrm{FDP}_{RB}/(\alpha m_0/m) \to 1 \tag{4}$$

in probability.

(ii) *If* $\log m = o(n^{1/2})$ *and* $m_1 \le m^\eta$ *for some* $\eta < 1$, *then* (4) *remains valid.*

In Theorem 3.1, we only require $\max_{1\le i\le m} \mathsf{E}X_i^6 \le K$, which is much weaker than the moment condition in Theorem 2.2.

As in Section 2.2, we can also estimate the p-values individually by $\breve{p}_{i,RB} = \hat{G}_i^*(T_i)$, where $\hat{G}_i^*(t) = \frac{1}{N}\sum_{k=1}^N I\{\hat{T}_{ki}^* \ge t\}$. Let $\mathrm{FDR}_{\breve{RB}}$ and $\mathrm{FDP}_{\breve{RB}}$ be the FDR and FDP of the B–H method with $\breve{p}_{i,RB}$, respectively. We have the following result.

PROPOSITION 3.1. *Suppose that* $N \ge m^{2+\delta}$ *for some* $\delta > 0$, $\max_{1\le i\le m} \mathsf{E}X_i^6 \le K$ *for some constant* $K > 0$, $\min_{1\le i\le m}\sigma_{ii} \ge c_1$ *for some* $c_1 > 0$ *and* $c_2(n/\log m)^{1/6} \le \lambda_{ni} \le c_3(n/\log m)^{1/6}$ *for some* $c_2, c_3 > 0$. *Assume that* $X_1, \dots, X_m$ *are independent.*

(i) *Suppose that* (2) *holds. Then Theorem* 3.1(i) *and* (ii) *hold for* $\mathrm{FDR}_{\breve{RB}}$ *and* $\mathrm{FDP}_{\breve{RB}}$.

(ii) *Suppose that* m_1 *is fixed and* $p_i \sim U(0,1)$ *for* $i \in \mathcal{H}_0$. *If* $\log m = o(n^{1/2})$, *then we have* $\overline{\lim}_{(n,m)\to\infty} \mathrm{FDR}_{\breve{RB}} \le \alpha$.

Theorem 3.1 does not cover the case when m_1 is fixed. However, if $\breve{p}_{i,RB}$, $1 \le i \le m$ are used, then Proposition 3.1 shows that the FDR can be controlled when m_1 is fixed and $\log m = o(n^{1/2})$. Actually, when m_1 is fixed and $\log m = o(n^{1/3})$, by the proof of Propositions 2.2 and 3.1, we can show that $\overline{\lim}_{(n,m)\to\infty} \mathrm{FDR}_{RB} \le \alpha$. It is unclear whether the similar result holds for FDR_{RB} when the dimension becomes larger, that is, $\log m = o(n^{1/2})$. However, under (2), Theorem 3.1 only requires $N \ge 1$ because we use the average of all m variables. Hence $\hat{p}_{i,RB}$ have the significant advantage on the computational cost over $\breve{p}_{i,RB}$. Moreover, Proposition 2.1 indicates that (2) is nearly necessary for FDP control. Note that when one has FDP control, one can also have FDR control, but the reverse is not true, as Proposition 2.1 shows. Because FDR control is about the FDP average, studying FDP is more appealing in applications than FDR control.

In the regularized bootstrap method, we must choose the regularization parameter λ_{ni}. By Theorem 1.2 in Wang (2005), equation (2.2) in Shao (1999) and the proof of Theorem 3.1, we have

$$\mathsf{P}(|\hat{T}_{ki}^*| \ge t|\hat{\mathcal{X}}) = \frac{1}{2}G(t)\left[\exp\left(\frac{t^3}{\sqrt{n}}\hat{\kappa}_i(\lambda_{ni})\right) + \exp\left(-\frac{t^3}{\sqrt{n}}\hat{\kappa}_i(\lambda_{ni})\right)\right](1 + o_{\mathsf{P}}(1)),$$

uniformly for $0 \le t \le o(n^{1/4})$, where $G(t) = 2 - 2\Phi(t)$, $\hat{\mathcal{X}} = \{\hat{\mathcal{X}}_1, \dots, \hat{\mathcal{X}}_m\}$,

$$\hat{\kappa}_i(\lambda_{ni}) = \frac{1}{n\hat{\sigma}_i^3}\sum_{k=1}^n (\hat{X}_{ki} - \hat{X}_i)^3 \quad \text{and} \quad \hat{\sigma}_i^2 = \frac{1}{n}\sum_{k=1}^n (\hat{X}_{ki} - \hat{X}_i)^2. \tag{5}$$

Also,

$$\mathsf{P}(|T_i| \ge t) = \frac{1}{2}G(t)\left[\exp\left(\frac{t^3}{\sqrt{n}}\kappa_i\right) + \exp\left(-\frac{t^3}{\sqrt{n}}\kappa_i\right)\right](1 + o(1)),$$

uniformly for $0 \le t \le o(n^{1/4})$, where $\kappa_i = \mathsf{E}Y_i^3$. A good choice of λ_{ni} is to make $\hat{\kappa}_i(\lambda_{ni})$ approach κ_i. As κ_i is unknown, we propose the following cross-validation method.

Data-driven choice of λ_{ni}. We propose to choose $\hat{\lambda}_{ni} = |\bar{X}_i| + \hat{s}_{ni}\lambda$, where λ will be selected as follows. Split the samples into two parts $\mathcal{I}_0 = \{1, \dots, n_1\}$ and $\mathcal{I}_1 = \{n_1 + 1, \dots, n\}$ with sizes $n_0 = [n/2]$ and $n_1 = n - n_0$, respectively. For $\mathcal{I} = \mathcal{I}_0$ or $\mathcal{I}_1$, let

$$\hat{\kappa}_{i,\mathcal{I}} = \frac{1}{|\mathcal{I}|\hat{s}^3_{ni,\mathcal{I}}} \sum_{k \in \mathcal{I}} (X_{ki} - \bar{X}_{i,\mathcal{I}})^3, \qquad \hat{s}^2_{ni,\mathcal{I}} = \frac{1}{|\mathcal{I}|} \sum_{k \in \mathcal{I}} (X_{ki} - \bar{X}_{i,\mathcal{I}})^2,$$

$$\bar{X}_{i,\mathcal{I}} = \frac{1}{|\mathcal{I}|} \sum_{k \in \mathcal{I}} X_{ki}.$$

Let $\hat{\kappa}_{i,\mathcal{I}}(\lambda_{ni})$, with $\lambda_{ni} = |\bar{X}_{i,\mathcal{I}}| + \hat{s}_{ni,\mathcal{I}}\lambda/2$, be defined as in (5) based on $\{\hat{X}_{ki}, k \in \mathcal{I}\}$. Define the risk

$$R_j(\lambda) = \sum_{i=1}^{m} \bigl(\hat{\kappa}_{i,\mathcal{I}_j}(\lambda_{ni}) - \hat{\kappa}_{i,\mathcal{I}_{1-j}}\bigr)^2.$$

We choose λ by

$$\hat{\lambda} = \mathop{\arg\min}_{0<\lambda<\infty} \bigl\{R_0(\lambda) + R_1(\lambda)\bigr\}. \tag{6}$$

The final regularization parameter is $\hat{\lambda}_{ni} = |\bar{X}_i| + \hat{s}_{ni}\hat{\lambda}$.

The numerical performance comparison between the data-driven choice $\hat{\lambda}_{ni}$ and the theoretical choice [e.g., $(n/\log m)^{1/6}$] is given in Section 5. In addition, it is important to investigate the theoretical property of $\hat{\lambda}_{ni}$ and to see whether Theorem 3.1 still holds when $\hat{\lambda}_{ni}$ is used. We leave this for future work.

4. FDR control under dependence. To generalize the results to the dependent case, we introduce a class of correlation matrices. Let $\mathbf{A} = (a_{ij})$ be a symmetric matrix. Let k_m and s_m be positive numbers. Assume that for every $1 \le j \le m$,

$$\operatorname{Card}\bigl\{1 \le i \le m : |a_{ij}| \ge k_m\bigr\} \le s_m. \tag{7}$$

Let $\mathcal{A}(k_m, s_m)$ be the class of symmetric matrices satisfying (7). Let $\mathbf{R} = (r_{ij})$ be the correlation matrix of $\mathbf{X}$. We introduce the following two conditions:

(C1) Suppose that $\max_{1\le i<j\le m} |r_{ij}| \le r$ for some $0 < r < 1$ and $\mathbf{R} \in \mathcal{A}(k_m, s_m)$ with $k_m = (\log m)^{-2-\theta}$ and $s_m = O(m^\rho)$ for some $\theta > 0$ and $0 < \rho < (1-r)/(1+r)$.

(C1*) Suppose that $\max_{1\le i<j\le m} |r_{ij}| \le r$ for some $0 < r < 1$. For each X_i, assume that the number of variables X_j that are dependent with X_i is no more than s_m.

(C1) and (C1*) impose the weak dependence between $X_1, \dots, X_m$. In (C1), each variable can be highly correlated with other s_m variables and weakly correlated with the remaining variables. (C1*) is stronger than (C1). For each X_i, (C1*) requires the independence between X_i and other $m - s_m$ variables.

Recall that $m_1 \le \gamma m$ for some $\gamma < 1$.

THEOREM 4.1. *Assume that* $\max_{1\le i\le m} \mathsf{E}Y_i^4 \le b_0$ *for some constant* $b_0 > 0$, *and* (2) *holds*.

(i) *If* $\log m = O(n^{\zeta})$ *for some* $0 < \zeta < 3/23$ *and* (C1) *is satisfied*, *then we have*

$$\lim_{(n,m)\to\infty} \frac{\mathrm{FDR}_\Phi}{(m_0/m)\alpha} = 1 \quad \textit{and} \quad \frac{\mathrm{FDP}_\Phi}{(m_0/m)\alpha} \to 1 \qquad \textit{in probability.} \tag{8}$$

(ii) *Under* $\log m = o(n^{1/3})$ *and* (C1*), (8) *also holds*.

The same conclusions hold for $\mathrm{FDR}_{\Psi_{n-1}}$ *and* $\mathrm{FDP}_{\Psi_{n-1}}$.

For the bootstrap and regularized procedures, we have similar results.

THEOREM 4.2. *Suppose that* $\max_{1\le i\le m} \mathsf{E}e^{tY_i^2} \le K$ *and* (2) *is satisfied*.

(1) *Under the conditions of* (i) *or* (ii) *in Theorem* 4.1, *we have*

$$\lim_{(n,m)\to\infty} \frac{\mathrm{FDR}_B}{(m_0/m)\alpha} = 1 \quad \textit{and} \quad \frac{\mathrm{FDP}_B}{(m_0/m)\alpha} \to 1 \qquad \textit{in probability.} \tag{9}$$

(2) *Under* (C1*), $\log m = o(n^{1/2})$ *and* $m_1 \le m^{\eta}$ *for some* $\eta < 1$, (9) *holds*.

THEOREM 4.3. *Suppose that* $\max_{1\le i\le m} \mathsf{E}X_i^6 \le K$ *for some constant* $K > 0$, $\min_{1\le i\le m} \sigma_{ii} \ge c_1$ *for some* $c_1 > 0$ *and* (2) *is satisfied*. *Let* $c_2(n/\log m)^{1/6} \le \lambda_{ni} \le c_3(n/\log m)^{1/6}$ *for some* $c_2, c_3 > 0$.

(1) *Under the conditions of* (i) *or* (ii) *in Theorem* 4.1, *we have*

$$\lim_{(n,m)\to\infty} \frac{\mathrm{FDR}_{RB}}{(m_0/m)\alpha} = 1 \quad \textit{and} \quad \frac{\mathrm{FDP}_{RB}}{(m_0/m)\alpha} \to 1 \qquad \textit{in probability.} \tag{10}$$

(2) *Under* (C1*), $\log m = o(n^{1/2})$ *and* $m_1 \le m^{\eta}$ *for some* $\eta < 1$, (10) *holds*.

Theorems 4.1–4.3 imply that the B–H method remains valid asymptotically for weak dependence. As the phase transition phenomenon caused by the growth of the dimension, it would be interesting to investigate when the B–H method will fail to control the FDR as the correlation becomes stronger.

5. Numerical study. In this section, we conduct a small simulation to verify the phase transition phenomenon. Let

$$X_i = \mu_i + (\varepsilon_i - \mathsf{E}\varepsilon_i), \qquad 1 \le i \le m, \tag{11}$$

where $(\varepsilon_1, \dots, \varepsilon_m)'$ are i.i.d. random variables. We consider three models for ε_i and μ_i.

Model 1. ε_i is the exponential random variable with parameter 1. Let $\mu_i = 2\sigma\sqrt{\log m/n}$ for $1 \le i \le m_1$ with $m_1 = 0.05m$ and $\mu_i = 0$ for $m_1 < i \le m$, where $\sigma^2 = \mathsf{Var}(\varepsilon_i)$.

Model 2-1. ε_i is the Gamma random variable with parameter $(0.5, 1)$. Let $\mu_i = 4\sigma\sqrt{\log m/n}$ for $1 \le i \le m_1$ with $m_1 = 0.05m$ and $\mu_i = 0$ for $m_1 < i \le m$.

Model 2-2. ε_i is the Gamma random variable with parameter $(0.5, 1)$. Let $m_1 = 0$.

In all three models, the average of skewness is $\tau > 0$. We generate $n = 30, 50$ independent random samples from (11). In our simulation, α is taken to be $0.1, 0.2, 0.3$ and m is taken to be 500, 1000, 3000. For computational reasons, we only consider the estimated p-values $\hat{p}_{i,B}$ and $\hat{p}_{i,RB}$ in the bootstrap and regularized bootstrap procedures, respectively. The number of bootstrap resamples is taken to be $N = 200$. We use FDR_B, FDR_{RB} and FDR^*_{RB} to denote the FDR of the B–H method with bootstrap, regularized bootstrap with data-driven $\hat{\lambda}_{ni}$ and regularized bootstrap with theoretical $\lambda_{ni} = (n/\log m)^{1/6}$, respectively. The simulation is replicated 1000 times and the empirical FDR and power for $m = 3000$ are summarized in Tables 1 and 2. To save space, we leave the simulation results for $m = 500$ and 1000 in the supplementary material of Liu and Shao (2014). The empirical power is defined by the average ratio between the number of correct rejections and m_1. Due to the nonzero skewness and $m \gg \exp(n^{1/3})$, the empirical FDR_Φ and $\mathrm{FDR}_{\Psi_{n-1}}$ are much larger than the target FDR. The bootstrap method and the regularized bootstrap method with data-driven $\hat{\lambda}_{ni}$ provide more accurate approximations for the true p-values. Thus the empirical FDR_B and FDR_{RB} are much closer to α than FDR_Φ and $\mathrm{FDR}_{\Psi_{n-1}}$. For Models 1, 2-1 and 2-2, the bootstrap method and the proposed regularized bootstrap method with data-driven $\hat{\lambda}_{ni}$ perform quite similarly. In addition, the data-driven $\hat{\lambda}_{ni}$ performs much better than the theoretical λ_{ni}. All of four methods perform better as the sample size n grows from 30 to 50, although the empirical FDR_Φ and $\mathrm{FDR}_{\Psi_{n-1}}$ still exhibit a serious departure from α.

Next, we consider the following two models to compare the performance between the four methods when the distributions are symmetric and heavy tailed.

Model 3. ε_i is Student's t distribution with 4 degrees of freedom. Let $\mu_i = 2\sqrt{\log m/n}$ for $1 \le i \le m_1$ with $m_1 = 0.1m$ and $\mu_i = 0$ for $m_1 < i \le m$.

TABLE 1
Comparison of FDR (FDR $= \alpha$, $m = 3000$)

	$n = 30$			$n = 50$		
α	**0.1**	**0.2**	**0.3**	**0.1**	**0.2**	**0.3**
			exp(1)			
FDR_Φ	0.3811	0.4791	0.5527	0.2931	0.3975	0.4809
FDR_Ψ	0.3127	0.4184	0.4987	0.2508	0.3569	0.4422
FDR_B	0.0712	0.1810	0.2866	0.0926	0.1939	0.2930
FDR_{RB}	0.0712	0.1810	0.2866	0.0926	0.1940	0.2931
FDR^*_{RB}	0.2520	0.3727	0.4642	0.2109	0.3234	0.4153
			Gamma(0.5, 1), $m_1 = 0.05m$			
FDR_Φ	0.5036	0.5826	0.6384	0.4009	0.4946	0.5634
FDR_Ψ	0.4492	0.5400	0.6034	0.3629	0.4623	0.5348
FDR_B	0.0735	0.1756	0.2847	0.0855	0.1889	0.2930
FDR_{RB}	0.0735	0.1756	0.2847	0.0854	0.1889	0.2930
FDR^*_{RB}	0.2614	0.4144	0.5180	0.2326	0.3650	0.4633
			Gamma(0.5, 1), $m_1 = 0$			
FDR_Φ	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
FDR_Ψ	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
FDR_B	0.0110	0.0350	0.0630	0.0270	0.0630	0.1010
FDR_{RB}	0.0110	0.0350	0.0630	0.0270	0.0630	0.1010
FDR^*_{RB}	0.4450	0.7780	0.9570	0.5170	0.8081	0.9490

Model 4. $\varepsilon_i = \varepsilon_{i1} - \varepsilon_{i2}$, where ε_{i1} and ε_{i1} are independent lognormal random variables with parameters (0, 1). Let $\mu_i = 4\sqrt{\log m/n}$ for $1 \le i \le m_1$ with $m_1 = 0.1m$ and $\mu_i = 0$ for $m_1 < i \le m$.

For these two models, the normal approximation performs the best on the control of FDR; see Tables 3 and 4. FDR_B is much smaller than α, so the bootstrap method is quite conservative. This is mainly due to the heavy tails of the $t(4)$ and lognormal distributions. The regularized bootstrap method works much better than the bootstrap method to control FDR. Table 4 shows that it also has a higher power (power_{RB}) than the bootstrap method (power_B). Hence the proposed regularized bootstrap is more robust than the commonly used bootstrap method.

Finally, we examine the FDP control of the B–H method when m is small and p-values are known. To this end, we consider Model 5 in which the exact null distributions are known.

Model 5. Let ε_i be i.i.d. $N(0,1)$ random variables. Let $\mu_i = 2\sqrt{\log m/n}$ for $1 \le i \le m_1$ and $\mu_i = 0$ for $m_1 < i \le m$, where $m_1 = 0$, 1 and 5.

In Figure 1, we plot the curve of the tailed probability of FDP based on 5000 replications, that is, $\sum_{i=1}^{5000} I\{\mathrm{FDP}_i \ge t\}/5000$, where FDP_i is the true FDP in the ith replication. From Figure 1, we can see that when m_1 is small, the B–H method works unfavorably on FDP control. For example, the empirical probability of FDP > 0.4 is 1 when $m_1 = 0$, 0.35 when $m_1 = 1$ and 0.12 when $m_1 = 5$.

TABLE 2
Comparison of power (FDR $= \alpha$)

m	α	$n = 30$ 0.1	0.2	0.3	$n = 50$ 0.1	0.2	0.3
				exp(1)			
3000	power_Φ	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	power_Ψ	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000
	power_B	0.9642	0.9984	0.9999	0.9987	1.0000	1.0000
	power_{RB}	0.9648	0.9983	0.9998	0.9989	1.0000	1.0000
	power^*_{RB}	0.9997	1.0000	1.0000	1.0000	1.0000	1.0000
				Gamma(0.5, 1), $m_1 = 0.05m$			
3000	power_Φ	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	power_Ψ	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	power_B	0.9996	1.0000	1.0000	1.0000	1.0000	1.0000
	power_{RB}	0.9996	1.0000	1.0000	1.0000	1.0000	1.0000
	power^*_{RB}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

This phenomenon is in accord with Proposition 2.1. In contrast, as indicated by Theorem 2.1, the performance of FDP control improves when m_1 increases.

6. Proof of main results. We begin the proof by showing a uniform law of large numbers (13), which plays a key role in the proof of main results. According to Theorem 1.2 in Wang (2005) and equation (2.2) in Shao (1999), we have for

TABLE 3
Comparison of FDR (FDR $= \alpha$)

m	α	$n = 30$ 0.1	0.2	0.3	$n = 50$ 0.1	0.2	0.3
				$t(4)$			
3000	FDR_Φ	0.1158	0.2137	0.3087	0.1028	0.1984	0.2920
	FDR_Ψ	0.0713	0.1569	0.2464	0.0773	0.1638	0.2551
	FDR_B	0.0381	0.1093	0.1946	0.0542	0.1348	0.2238
	FDR_{RB}	0.0609	0.1439	0.2341	0.0722	0.1591	0.2500
	FDR^*_{RB}	0.0636	0.1476	0.2380	0.0733	0.1603	0.2512
				Lognormal(0, 1)			
3000	FDR_Φ	0.0807	0.1706	0.2656	0.0745	0.1627	0.2574
	FDR_Ψ	0.0442	0.1146	0.1983	0.0523	0.1282	0.2175
	FDR_B	0.0008	0.0148	0.0509	0.0071	0.0441	0.1056
	FDR_{RB}	0.0323	0.0956	0.1761	0.0488	0.1239	0.2129
	FDR^*_{RB}	0.0006	0.0268	0.1124	0.0487	0.1235	0.2116

TABLE 4
Comparison of power (FDR $= \alpha$)

		$n=30$			$n=50$		
m	α	**0.1**	**0.2**	**0.3**	**0.1**	**0.2**	**0.3**
				$t(4)$			
3000	power_Φ	0.9075	0.9413	0.9589	0.9109	0.9449	0.9621
	power_Ψ	0.8765	0.9250	0.9483	0.8936	0.9357	0.9564
	power_B	0.8291	0.9036	0.9362	0.8712	0.9262	0.9509
	power_{RB}	0.8655	0.9200	0.9456	0.8903	0.9347	0.9557
	power^*_{RB}	0.8685	0.9215	0.9464	0.8912	0.9351	0.9560
				Lognormal(0, 1)			
3000	power_Φ	0.8639	0.9009	0.9229	0.8613	0.9021	0.9256
	power_Ψ	0.8322	0.8810	0.9085	0.8420	0.8898	0.9169
	power_B	0.5881	0.7688	0.8385	0.7193	0.8307	0.8783
	power_{RB}	0.8141	0.8711	0.9019	0.8374	0.8865	0.9149
	power^*_{RB}	0.5438	0.7986	0.8785	0.8368	0.8866	0.9149

$0 \le t \le o(n^{1/4})$,

$$\begin{aligned}\mathsf{P}\big(|T_i - \sqrt{n}\mu_i/\hat{s}_n| \ge t\big) &= \frac{1}{2}G(t)\biggl[\exp\biggl(-\frac{t^3}{3\sqrt{n}}\kappa_i\biggr) + \exp\biggl(\frac{t^3}{3\sqrt{n}}\kappa_i\biggr)\biggr] \\ &\quad \times \big(1 + o(1)\big),\end{aligned} \tag{12}$$

where $o(1)$ is uniformly in $1 \le i \le m$, $G(t) = 2 - 2\Phi(t)$ and $\kappa_i = \mathsf{E}Y_i^3$.

For any $b_m \to \infty$ and $b_m = o(m)$, we first prove that, under (C1*) and $\log m = o(n^{1/2})$ [or (C1) and $\log m = O(n^\zeta)$ for some $0 < \zeta < 3/23$],

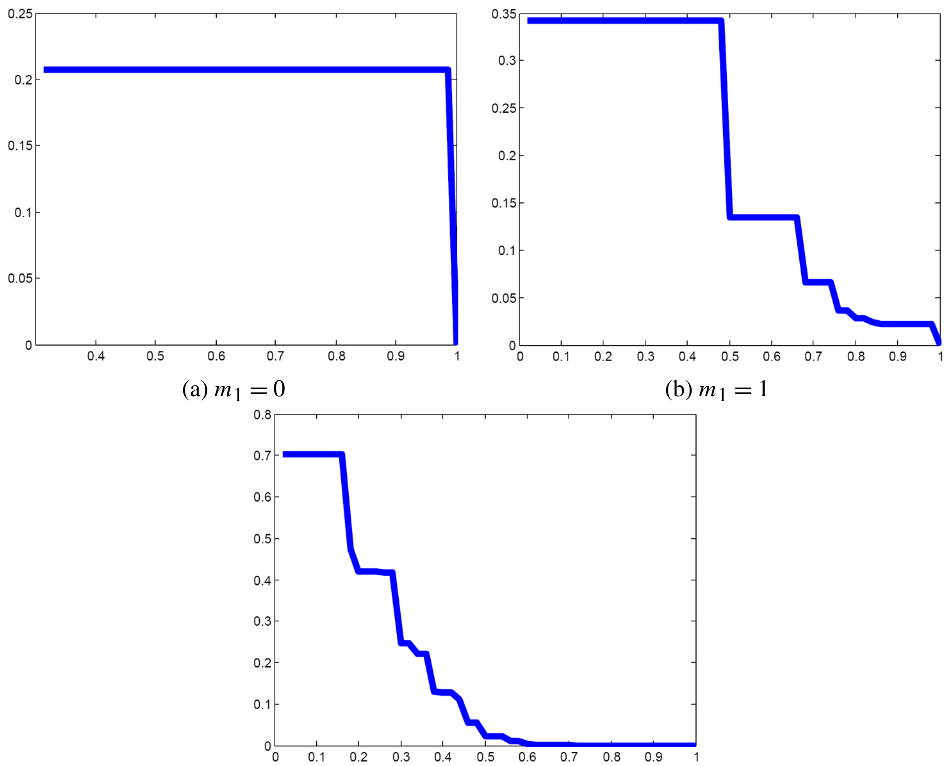
$$\sup_{0 \le t \le G_\kappa^{-1}(b_m/m)} \left| \frac{\sum_{i \in \mathcal{H}_0} I\{|T_i| \ge t\}}{m_0 G_\kappa(t)} - 1 \right| \to 0 \tag{13}$$

in probability, where

$$G_\kappa(t) = \frac{1}{2m_0} G(t) \sum_{i \in \mathcal{H}_0} \biggl[\exp\biggl(-\frac{t^3}{3\sqrt{n}}\kappa_i\biggr) + \exp\biggl(\frac{t^3}{3\sqrt{n}}\kappa_i\biggr)\biggr] =: G(t)\hat{\kappa}_\Phi(t)$$

and $G_\kappa^{-1}(t) = \inf\{y \ge 0 : G_\kappa(y) = t\}$ for $0 \le t \le 1$. Note that for $0 \le t \le o(\sqrt{n})$, $G_\kappa(t)$ is a strictly decreasing and continuous function. Let $z_0 < z_1 < \cdots < z_{d_m} \le 1$ and $t_i = G_\kappa^{-1}(z_i)$, where $z_0 = b_m/m$, $z_i = b_m/m + b_m^{2/3} e^{i^\delta}/m$, $d_m = [\{\log((m - b_m)/b_m^{2/3})\}^{1/\delta}]$ and $0 < \delta < 1$, which will be specified later. Note that $G_\kappa(t_i)/G_\kappa(t_{i+1}) = 1 + o(1)$ uniformly in i, and $t_0/\sqrt{2\log(m/b_m)} = 1 + o(1)$. Then to prove (13), it is enough to show that

$$\sup_{0 \le j \le d_m} \left| \frac{\sum_{i \in \mathcal{H}_0} I\{|T_i| \ge t_j\}}{m_0 G_\kappa(t_j)} - 1 \right| \to 0 \tag{14}$$



(a) $m_1 = 0$ (b) $m_1 = 1$

(c) $m_1 = 5$

FIG. 1. *Tailed probability of FDP with* $\alpha = 0.2$ *and* $n = 50$. *The* y*-axis values are the empirical tailed probabilities* $\sum_{i=1}^{5000} I\{\mathrm{FDP}_i \geq t\}/5000$.

in probability. Under (C1), define

$$\mathcal{S}_j = \{i \in \mathcal{H}_0 : |r_{ij}| \geq (\log m)^{-2-\theta}\}, \qquad \mathcal{S}_j^c = \mathcal{H}_0 - \mathcal{S}_j,$$

and under (C1*), define

$$\mathcal{S}_j = \{i \in \mathcal{H}_0 : X_i \text{ is dependent with } X_j\}.$$

We claim that, under (C1*) and $\log m = o(n^{1/2})$ [or (C1) and $\log m = O(n^{\zeta})$ for some $0 < \zeta < 3/23$], for any $\varepsilon > 0$ and some $\gamma_1 > 0$,

$$\begin{aligned} I_2(t) &:= \mathsf{E}\biggl(\sum_{i \in \mathcal{H}_0} \{I\{T_i \geq t\} - \mathsf{P}(|T_i| \geq t)\}\biggr)^2 \\ &\leq C m_0^2 G_\kappa^2(t) \biggl(\frac{1}{m_0 G_\kappa(t)} + \frac{\exp((r+\varepsilon)t^2/(1+r))}{m^{1-\rho}} + (\log m)^{-1-\gamma_1}\biggr) \end{aligned} \tag{15}$$

uniformly in $t \in [0, K\sqrt{\log m}]$ for all $K > 0$. Take $(1+\gamma_1)^{-1} < \delta < 1$. Given (15) and $G_\kappa^{-1}(b_m/m) \sim \sqrt{2\log(m/b_m)}$, for any $\varepsilon > 0$, we have

$$
\begin{aligned}
&\sum_{j=0}^{d_m} \mathsf{P}\biggl(\biggl|\frac{\sum_{i\in\mathcal{H}_0} I\{T_i \geq t_j\}}{m_0 G_\kappa(t_j)} - 1\biggr| \geq \varepsilon\biggr) \\
&\quad \leq \sum_{j=0}^{d_m} \mathsf{P}\biggl(\biggl|\frac{\sum_{i\in\mathcal{H}_0}(I\{T_i \geq t_j\} - \mathsf{P}(|T_i| \geq t_j))}{m_0 G_\kappa(t_j)}\biggr| \geq \varepsilon/2\biggr) \\
&\quad \leq C\biggl(\frac{1}{m_0 G_\kappa(t_0)} + \sum_{j=1}^{d_m} \frac{1}{m_0 G_\kappa(t_j)} + d_m m^{-1+\rho+((2r+2\varepsilon)/(1+r))+o(1)} \\
&\qquad\qquad + d_m(\log m)^{-1-\gamma_1}\biggr) \\
&\quad \leq C\Biggl(b_m^{-1} + b_m^{-2/3}\sum_{j=1}^{d_m} e^{-j^\delta} + o(1)\Biggr) = o(1).
\end{aligned}
$$

This proves (14).

To prove (15), we need the following lemma, which is proven in the supplementary material Liu and Shao (2014).

LEMMA 6.1. (i) *Suppose that* $\log m = O(n^{1/2})$. *For any* $\varepsilon > 0$,

$$
\max_{j\in\mathcal{H}_0} \max_{i\in\mathcal{S}_j\setminus j} \mathsf{P}\bigl(|T_i| \geq t, |T_j| > t\bigr) \leq C\exp\bigl(-(1-\varepsilon)t^2/(1+r)\bigr) \tag{16}
$$

uniformly in $t \in [0, o(n^{1/4}))$.

(ii) *Suppose that* $\log m = O(n^\zeta)$ *for some* $0 < \zeta < 3/23$. *We have for any* $K > 0$

$$
\mathsf{P}\bigl(|T_i| > t, |T_j| > t\bigr) = (1 + A_n)\mathsf{P}\bigl(|T_i| > t\bigr)\mathsf{P}\bigl(|T_j| > t\bigr) \tag{17}
$$

uniformly in $0 \leq t \leq K\sqrt{\log m}$, $j \in \mathcal{H}_0$ *and* $i \in \mathcal{S}_j^c$, *where* $|A_n| \leq C(\log m)^{-1-\gamma_1}$ *for some* $\gamma_1 > 0$.

Set $f_{ij}(t) = \mathsf{P}(|T_i| \geq t, |T_j| \geq t) - \mathsf{P}(|T_i| \geq t)\mathsf{P}(|T_j| \geq t)$. Note that under (C1*) $f_{ij} = 0$ when $j \in \mathcal{H}_0 \setminus \mathcal{S}_i$. We have

$$
\begin{aligned}
I_2(t) &\leq \sum_{i\in\mathcal{H}_0}\sum_{j\in\mathcal{S}_i} \mathsf{P}\bigl(|T_i| \geq t, |T_j| \geq t\bigr) + \sum_{i\in\mathcal{H}_0}\sum_{j\in\mathcal{H}_0\setminus\mathcal{S}_i} f_{ij}(t) \\
&\leq C m_0 G_\kappa(t) + C\frac{\exp((r+2\varepsilon)t^2/(1+r))}{m^{1-\rho}} m_0^2 G_\kappa^2(t) + A_n m_0^2 G_\kappa^2(t),
\end{aligned}
$$

where the last inequality follows from Lemma 6.1 and $G_\kappa(t) = G(t)e^{o(1)t^2}$ for $t = o(\sqrt{n})$. This proves (15).

6.1. *Proof of Theorem* 2.1 *and Corollary* 2.1. We only prove the theorem for $\hat{p}_{i,\Phi}$. The proof for $\hat{p}_{i,\Psi_{n-1}}$ is exactly the same when $G(t)$ is replaced with $2 - 2\Psi_{n-1}(t)$. By Lemma 1 in Storey, Taylor and Siegmund (2004), we can see that the B–H method with $\hat{p}_{i,\Phi}$ is equivalent to the following procedure: reject H_{0i} if and only if $\hat{p}_{i,\Phi} \le \hat{t}_0$, where

$$\hat{t}_0 = \sup\left\{0 \le t \le 1 : t \le \frac{\alpha \max(\sum_{1\le i\le m} I\{\hat{p}_{i,\Phi} \le t\}, 1)}{m}\right\}.$$

It is equivalent to reject H_{0i} if and only if $|T_i| \ge \hat{t}$, where

$$\hat{t} = \inf\left\{t \ge 0 : 2 - 2\Phi(t) \le \frac{\alpha \max(\sum_{1\le i\le m} I\{|T_i| \ge t\}, 1)}{m}\right\}.$$

By the continuity of $\Phi(t)$ and the monotonicity of the indicator function, it is easy to see that

$$\frac{mG(\hat{t})}{\max(\sum_{1\le i\le m} I\{|T_i| \ge \hat{t}\}, 1)} = \alpha,$$

where $G(t) = 2 - 2\Phi(t)$. Let $\mathcal{M}$ be a subset of $\{1, 2, \dots, m\}$ satisfying $\mathcal{M} \subset \{i : |\mu_i/\sigma_i| \ge 4\sqrt{\log m/n}\}$ and $\mathrm{Card}(\mathcal{M}) \le \sqrt{n}$. By $\max_{1\le i\le m} \mathsf{E}Y_i^4 \le K$ and Markov's inequality, for any $\varepsilon > 0$,

$$\mathsf{P}\Big(\max_{i\in\mathcal{M}} |\hat{s}_{ni}^2/\sigma_i^2 - 1| \ge \varepsilon\Big) = O(1/\sqrt{n}).$$

This, together with (2) and (12), implies that there exist some $c > \sqrt{2}$ and some $b_m \to \infty$,

$$\mathsf{P}\left(\sum_{i=1}^{m} I\{|T_i| \ge c\sqrt{\log m}\} \ge b_m\right) \to 1. \tag{18}$$

This implies that $\mathsf{P}(\hat{t} \le G^{-1}(\alpha b_m/m)) \to 1$. Given (13) and $G_\kappa(t) \ge G(t)$, it follows that $\mathsf{P}(\hat{t} \le G_\kappa^{-1}(\alpha b_m/m)) \to 1$. Therefore, by (13)

$$\frac{\sum_{i\in\mathcal{H}_0} I\{|T_i| \ge \hat{t}\}}{m_0 G_\kappa(\hat{t})} \to 1$$

in probability. Note that

$$G(\hat{t}) = \frac{\alpha\hat{m}}{m} + \frac{\alpha m_0}{m} \frac{\sum_{i\in\mathcal{H}_0} I\{|T_i| \ge \hat{t}\}}{m_0},$$

where $\hat{m} = \sum_{i\in\mathcal{H}_1} I\{|T_i| \ge \hat{t}\}$. With probability tending to one,

$$G(\hat{t}) = \frac{\alpha\hat{m}}{m} + \frac{\alpha m_0}{m} G(\hat{t})\hat{\kappa}_\Phi\big(1 + o(1)\big) \ge \frac{\alpha m_0}{m} G(\hat{t})\hat{\kappa}_\Phi\big(1 + o(1)\big). \tag{19}$$

Thus $\mathsf{P}(\hat{\kappa}_\Phi \le m/(\alpha m_0) + \varepsilon) \to 1$ for any $\varepsilon > 0$. Let $\hat{\kappa}_\Phi^* = \hat{\kappa}_\Phi I\{\hat{\kappa}_\Phi \le 2(\alpha(1-\gamma))^{-1}\}$. Note that $m/(\alpha m_0) + \varepsilon \le 2(\alpha(1-\gamma))^{-1}$. We have

$$\frac{\mathrm{FDP}_\Phi}{(m_0/m)\alpha\hat{\kappa}_\Phi^*} = \frac{\sum_{i\in\mathcal{H}_0} I\{|T_i| \ge \hat{t}\}}{m_0 G_\kappa(\hat{t})} \frac{\hat{\kappa}_\Phi}{\hat{\kappa}_\Phi^*}(1+o(1)) \to 1$$

in probability. Then for any $\varepsilon > 0$,

$$\mathrm{FDR}_\Phi \le (1+\varepsilon)\frac{m_0}{m}\alpha \mathsf{E}\hat{\kappa}_\Phi^* + \mathsf{P}\left(\mathrm{FDP}_\Phi \ge (1+\varepsilon)\frac{m_0}{m}\alpha\hat{\kappa}_\Phi^*\right)$$

and

$$\mathrm{FDR}_\Phi \ge (1-\varepsilon)\frac{m_0}{m}\alpha \mathsf{E}\hat{\kappa}_\Phi^* - 2\big(\alpha(1-\gamma)\big)^{-1}\mathsf{P}\left(\mathrm{FDP}_\Phi \le (1-\varepsilon)\frac{m_0}{m}\alpha\hat{\kappa}_\Phi^*\right).$$

This proves Theorem 2.1. Corollary 2.1(1) follows directly from Theorem 2.1 and $\mathsf{P}(\hat{t} \le \sqrt{2\log m}) \to 1$.

To prove Corollary 2.1(2), we first assume that $\frac{\alpha m_0}{m}\hat{\kappa}_\Phi \le 1-\eta$ for some $(1-\eta)/\alpha > 1$. So, by (19) and the condition $m_1 = \exp(o(n^{1/3}))$, with probability tending to one, $G(\hat{t}) \le 2\alpha\eta^{-1}\hat{m}/m \le 2\alpha\eta^{-1}m^{-1+o(1)}$. Hence, $\hat{t} \ge c\sqrt{\log m}$ for any $c < \sqrt{2}$. Recall that $\tau = \underline{\lim}_{m\to\infty} m_0^{-1}\sum_{i\in\mathcal{H}_0}|\mathsf{E}Y_i^3| > 0$. Set

$$\mathcal{H}_{01} = \{i \in \mathcal{H}_0 : |\mathsf{E}Y_i^3| \ge \tau/8\}.$$

According to the definition of τ and $|\mathsf{E}Y_i^3| \le (\mathsf{E}(Y_i^4))^{3/4} \le b_0^{3/4}$, $m_0^{-1}|\mathcal{H}_{01}^c|\tau/8 + b_0^{3/4}m_0^{-1}|\mathcal{H}_{01}| \ge \tau/2$. This implies that $|\mathcal{H}_{01}| \ge \tau b_0^{-3/4}m_0/4$. Hence, we can get $m_0^{-1}\sum_{i\in\mathcal{H}_0}|\mathsf{E}Y_i^3|^2 \ge c_\tau$ for some $c_\tau > 0$. It follows from Taylor's expansion of the exponential function and $\hat{t} \ge c\sqrt{\log m}$ that $\hat{\kappa}_\Phi \ge 1+\epsilon$ for some $\epsilon > 0$. However, if $\frac{\alpha m_0}{m}\hat{\kappa}_\Phi > 1-\eta$, then $\hat{\kappa}_\Phi \ge 1+\epsilon$ for some $\epsilon > 0$. This yields that $\mathsf{P}(\hat{\kappa}_\Phi \ge 1+\epsilon) \to 1$ for some $\epsilon > 0$. So we have $\kappa_\Phi \ge 1+\epsilon$ for some $\epsilon > 0$. Note that $m_0/m \to 1$. We prove Corollary 2.1(2).

We next prove Corollary 2.1(3). By the inequality $e^x + e^{-x} \ge |x|$, $\mathsf{P}(\hat{\kappa}_\Phi \le m/(\alpha m_0) + \varepsilon) \to 1$, we obtain that

$$\frac{\sum_{i\in\mathcal{H}_0}(\hat{t}^3/\sqrt{n})|\mathsf{E}Y_i^3|}{2m_0} \le m/(\alpha m_0) + \varepsilon$$

with probability tending to one. By $\tau > 0$, we have $\mathsf{P}(\hat{t} \le cn^{1/6}) \to 1$ for some constant $c > 0$. Thus $\mathsf{P}(G(\hat{t}) \ge \exp(-2cn^{1/3}) \to 1$. Because $\hat{m}/m \le \exp(-Mn^{1/3})$ for any $M > 0$, and given (19), we have

$$\frac{\alpha m_0}{m}\hat{\kappa}_\Phi \to 1$$

in probability. Hence, $\kappa_\Phi \to 1/\alpha$ as $m_0/m \to 1$. The proof is finished.

6.2. *Proof of Theorems 2.2 and 4.2.* Let $\hat{\kappa}_i = \frac{1}{n\hat{s}_{ni}^3}\sum_{k=1}^n (X_{ki} - \bar{X}_i)^3$. Define the event

$$\mathbf{F} = \left\{ \max_{1\le i\le m} \frac{1}{n\hat{s}_{ni}^4} \sum_{k=1}^n (X_{ki} - \bar{X}_i)^4 \le K_1, \max_{1\le i\le m} |\hat{\kappa}_i - \kappa_i| \le K_2\sqrt{\log m/n} \right\}$$

for some large $K_1 > 0$ and $K_2 > 0$. We first suppose that $\mathsf{P}(\mathbf{F}) \to 1$. Let $G_i^\star(t) = \mathsf{P}^*(|T_{ki}^*| \ge t)$ be the conditional distribution of T_{ki}^* given $\mathcal{X} = \{\mathcal{X}_1, \dots, \mathcal{X}_m\}$. Note that, given $\mathcal{X}$ and on the event $\mathbf{F}$,

$$\begin{aligned} G_i^\star(t) &= \frac{1}{2}G(t)\left[\exp\left(-\frac{t^3}{3\sqrt{n}}\hat{\kappa}_i\right) + \exp\left(\frac{t^3}{3\sqrt{n}}\hat{\kappa}_i\right)\right](1+o(1)) \\ &= \frac{1}{2}G(t)\left[\exp\left(-\frac{t^3}{3\sqrt{n}}\kappa_i\right) + \exp\left(\frac{t^3}{3\sqrt{n}}\kappa_i\right)\right](1+o(1)) \end{aligned}$$

uniformly in $0 \le t \le o(n^{1/4})$. Hence, given $\mathcal{X}$ and on the event $\mathbf{F}$,

$$\frac{G_i^\star(t)}{\mathsf{P}(|T_i - \sqrt{n}\mu_i/\hat{s}_n| \ge t)} = 1 + o(1) \tag{20}$$

uniformly in $1 \le i \le m$ and $0 \le t \le o(n^{1/4})$. Put

$$\hat{G}_\kappa(t) = \frac{1}{2m}G(t)\sum_{1\le i\le m}\left[\exp\left(-\frac{t^3}{3\sqrt{n}}\kappa_i\right) + \exp\left(\frac{t^3}{3\sqrt{n}}\kappa_i\right)\right].$$

Set $\hat{c}_m = \hat{G}_\kappa^{-1}(b_m/m)$. Note that, given $\mathcal{X}$, T_{ki}^*, $1 \le k \le N$, $1 \le i \le m$, are independent. Hence, as (13), we can show that for any $b_m \to \infty$,

$$\sup_{0\le t\le \hat{c}_m}\left|\frac{G_{N,m}^*(t)}{\hat{G}_\kappa(t)} - 1\right| \to 0 \tag{21}$$

in probability. For $t = O(\sqrt{\log m})$, under the conditions of Theorem 3.2, we have $\hat{G}_\kappa(t)/G_\kappa(t) = 1 + o(1)$. So, it is easy to see that (13) still holds when $G_\kappa^{-1}(b_m/m)$ is replaced by $\hat{G}_\kappa^{-1}(b_m/m)$. This implies that for any $b_m \to \infty$,

$$\sup_{0\le t\le \hat{c}_m}\left|\frac{\sum_{i\in\mathcal{H}_0} I\{|T_i| \ge t\}}{m_0 G_{N,m}^*(t)} - 1\right| \to 0 \tag{22}$$

in probability.

Let

$$\hat{t}_0 = \sup\left\{0 \le t \le 1 : t \le \frac{\alpha\max(\sum_{1\le i\le m} I\{\hat{p}_{i,B} \le t\}, 1)}{m}\right\}.$$

Then we have

$$\hat{t}_0 = \frac{\alpha\max(\sum_{1\le i\le m} I\{\hat{p}_{i,B} \le \hat{t}_0\}, 1)}{m}.$$

According to (12) and (20) we have, given $\mathcal{X}$ and on the event $\mathbf{F}$, $G_i^*(c\sqrt{\log m}) = m^{-c^2/2+o(1)}$ for any $c > \sqrt{2}$ uniformly in i. So, by Markov's inequality, for any $\varepsilon > 0$, we have $\mathsf{P}(G_{N,m}^*(c\sqrt{\log m}) \le m^{-c^2/2+\varepsilon}) \to 1$. By (2) and (18), we have $\mathsf{P}(\hat{t}_0 \ge \alpha b_m/m) \to 1$ for some $b_m \to \infty$. It follows from (22) that

$$\frac{\sum_{i\in\mathcal{H}_0} I\{\hat{p}_{i,B} \le \hat{t}_0\}}{m_0\hat{t}_0} \to 1$$

in probability. This finishes the proof of Theorem 2.2(1), (2) and Theorem 4.2 if we can show that $\mathsf{P}(\mathbf{F}) \to 1$. Without loss of generality, we can assume that $\mu_i = 0$ and $\sigma_i = 1$. We first show that for some constant $K_1 > 0$,

$$\mathsf{P}\biggl(\max_{1\le i\le m}\biggl|\sum_{k=1}^{n}(X_{ki}^4 - \mathsf{E}X_{ki}^4)\biggr| \ge K_1 n\biggr) = o(1). \tag{23}$$

For $1 \le i \le n$, put

$$\hat{X}_{ki} = X_{ki} I\{|X_{ki}| \le \sqrt{n/\log m}\}, \qquad \check{X}_{ki} = X_{ki} - \hat{X}_{ki}.$$

Then, for large n,

$$\begin{aligned}
&\mathsf{P}\biggl(\max_{1\le i\le m}\biggl|\sum_{k=1}^{n}(\check{X}_{ki}^4 - \mathsf{E}\check{X}_{ki}^4)\biggr| \ge K_1 n/2\biggr)\\
&\quad \le nm \max_{1\le i\le m} \mathsf{P}\bigl(|X_{1i}| \ge \sqrt{n/\log m}\bigr)\\
&\quad \le C\exp(\log m + \log n - tn/\log m)\\
&\quad = o(1).
\end{aligned}$$

Let $Z_{ki} = \hat{X}_{ki}^4 - \mathsf{E}\hat{X}_{ki}^4$. By the inequality $|e^s - 1 - s| \le s^2 e^{\max(s,0)}$ and $1 + s \le e^s$, we have for $\eta = 2^{-1}t(\log m)/n$ and some large K_1

$$\begin{aligned}
&\mathsf{P}\biggl(\max_{1\le i\le m}\biggl|\sum_{k=1}^{n} Z_{ki}\biggr| \ge K_1 n/2\biggr)\\
&\quad \le \sum_{i=1}^{m}\mathsf{P}\biggl(\sum_{k=1}^{n} Z_{ki} \ge K_1 n/2\biggr) + \sum_{i=1}^{m}\mathsf{P}\biggl(-\sum_{k=1}^{n} Z_{ki} \ge K_1 n/2\biggr)\\
&\quad \le \sum_{i=1}^{m}\exp(-\eta K_1 n/2)\biggl[\prod_{k=1}^{n}\exp(\eta Z_{ki}) + \prod_{k=1}^{n}\exp(-\eta Z_{ki})\biggr]\\
&\quad \le 2\sum_{i=1}^{m}\exp\bigl(-\eta K_1 n/2 + \eta^2 n\mathsf{E}Z_{1i}^2 e^{\eta|Z_{1i}|}\bigr)\\
&\quad \le C\exp\bigl(\log m - tK_1(\log m)/4\bigr)\\
&\quad = o(1).
\end{aligned}$$

This proves (23). By replacing X_{ki}^4, $\eta = 2^{-1}t(\log m)/n$ and $K_1 n/2$ with X_{ki}^3, $\eta = 2^{-1}t\sqrt{(\log m)/n}$ and $K_1\sqrt{n\log m}/2$, respectively, in the above proof, we can show that

$$\mathsf{P}\left(\max_{1\le i\le m}\left|\frac{1}{n}\sum_{k=1}^{n}(X_{ki}^3 - \mathsf{E}X_{ki}^3)\right| \ge K_1\sqrt{(\log m)/n}\right) = o(1). \tag{24}$$

Similarly, we have

$$\mathsf{P}\left(\max_{1\le i\le m}\left|\frac{1}{n}\sum_{k=1}^{n}(X_{ki}^2 - \mathsf{E}X_{ki}^2)\right| \ge K_1\sqrt{(\log m)/n}\right) = o(1) \tag{25}$$

and

$$\mathsf{P}\left(\max_{1\le i\le m}\left|\frac{1}{n}\sum_{k=1}^{n}(X_{ki} - \mathsf{E}X_{ki})\right| \ge K_1\sqrt{(\log m)/n}\right) = o(1). \tag{26}$$

Combining (23)–(26), we prove that $\mathsf{P}(\mathbf{F}) \to 1$.

6.3. *Proof of Theorems* 3.1 *and* 4.3. Let

$$\hat{\mathbf{F}} = \left\{\max_{1\le i\le m}\frac{1}{n\hat{\sigma}_i^4}\sum_{k=1}^{n}(\hat{X}_{ki} - \hat{X}_i)^4 \le K_1,\ \max_{1\le i\le m}\left|\hat{\kappa}_i(\lambda_{ni}) - \kappa_i\right| \le K_2\sqrt{\log m/n}\right\}.$$

By the proof of Theorems 2.2 and 4.2, it is enough to show that $\mathsf{P}(\hat{\mathbf{F}}) \to 1$. Recall that $\hat{X}_{ki} = X_{ki}I\{|X_{ki}| \le \lambda_{ni}\}$ and put $Z_{ki} = \hat{X}_{ki}^4 - \mathsf{E}\hat{X}_{ki}^4$. Take $\eta = (\log m)/n$. We have

$$\begin{aligned}
&\mathsf{P}\left(\max_{1\le i\le m}\left|\sum_{k=1}^{n} Z_{ki}\right| \ge K_1 n/2\right)\\
&\quad \le 2\sum_{i=1}^{m}\exp\left(-\eta K_1 n/2 + \eta^2 n\mathsf{E}Z_{1i}^2 e^{\eta|Z_{1i}|}\right)\\
&\quad \le C\exp\bigl(2\log m - K_1(\log m)/4\bigr)\\
&\quad = o(1).
\end{aligned}$$

Similarly, by replacing $\hat{X}_{ki}^4$, $\eta = (\log m)/n$ and $K_1 n/2$ with $\hat{X}_{ki}^3$, $\eta = \sqrt{(\log m)/n}$ and $K_1\sqrt{n\log m}/2$, respectively, in the above proof, we can show that

$$\mathsf{P}\left(\max_{1\le i\le m}\left|\frac{1}{n}\sum_{k=1}^{n}(\hat{X}_{ki}^3 - \mathsf{E}\hat{X}_{ki}^3)\right| \ge K_1\sqrt{(\log m)/n}\right) = o(1).$$

Also, using the above arguments, it is easy to show that

$$\mathsf{P}\left(\max_{1\le i\le m}\left|\frac{1}{n}\sum_{k=1}^{n}(\hat{X}_{ki}^2 - \mathsf{E}\hat{X}_{ki}^2)\right| \ge K_1\sqrt{(\log m)/n}\right) = o(1)$$

and

$$\mathsf{P}\left(\max_{1\le i\le m}\left|\frac{1}{n}\sum_{k=1}^{n}(\hat{X}_{ki}-\mathsf{E}\hat{X}_{ki})\right|\ge K_1\sqrt{(\log m)/n}\right)=o(1).$$

Note that

$$\max_{1\le i\le m}\mathsf{E}|X_{1i}|^3 I\{|X_{1i}|\ge\lambda_{ni}\}\le C\sqrt{\frac{\log m}{n}}\max_{1\le i\le m}\mathsf{E}X_{1i}^6$$

and

$$\max_{1\le i\le m}\mathsf{E}|X_{1i}|^2 I\{|X_{1i}|\ge\lambda_{ni}\}\le C\left(\frac{\log m}{n}\right)^{2/3}\max_{1\le i\le m}\mathsf{E}X_{1i}^6.$$

This proves that $\mathsf{P}(\hat{\mathbf{F}})\to 1$.

6.4. *Proof of Theorem* 4.1. Recall that

$$\frac{mG(\hat{t})}{\max(\sum_{1\le i\le m}I\{|T_i|\ge\hat{t}\},1)}=\alpha.$$

From (18), we have $\mathsf{P}(\hat{t}\ge G^{-1}(\alpha b_m/m))\to 1$. The theorem follows from (13) and the fact that $G_\kappa(t)/G(t)=1+o(1)$ uniformly in $t\in[0,o(n^{1/6}))$.

6.5. *Proof of Propositions* 2.1, 2.2, 2.3 *and* 3.1. To save space, the proof of these propositions is given in the supplementary material Liu and Shao (2014).

Acknowledgments. The authors would like to thank the Associate Editor and two referees for their valuable comments, which have helped to improve the quality and presentation of this paper.

SUPPLEMENTARY MATERIAL

Supplement to "Phase transition and regularized bootstrap in large-scale t-tests with false discovery rate control" (DOI: 10.1214/14-AOS1249SUPP; .pdf). The supplementary material includes part of numerical results and the proof of Lemma 6.1 and Propositions 2.1, 2.2, 2.3 and 3.1.

REFERENCES

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 Stat. Methodol.* **57** 289–300. MR1325392

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

CAO, H. and KOSOROK, M. R. (2011). Simultaneous critical values for t-tests in very high dimensions. *Bernoulli* **17** 347–394. MR2797995

DELAIGLE, A., HALL, P. and JIN, J. (2011). Robustness and accuracy of methods for high dimensional data analysis based on Student's t-statistic. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73** 283–301. MR2815777

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

FAN, J., HALL, P. and YAO, Q. (2007). To how many simultaneous hypothesis tests can normal, Student's t or bootstrap calibration be applied? *J. Amer. Statist. Assoc.* **102** 1282–1288. MR2372536

FERREIRA, J. A. and ZWINDERMAN, A. H. (2006). On the Benjamini–Hochberg method. *Ann. Statist.* **34** 1827–1849. MR2283719

LIU, W. and SHAO, Q. (2014). Supplement to "Phase transition and regularized bootstrap in large-scale t-tests with false discovery rate control." DOI:10.1214/14-AOS1249SUPP.

SHAO, Q.-M. (1999). A Cramér type large deviation result for Student's t-statistic. *J. Theoret. Probab.* **12** 385–398. MR1684750

STOREY, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann. Statist.* **31** 2013–2035. MR2036398

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 Stat. Methodol.* **66** 187–205. MR2035766

WANG, Q. (2005). Limit theorems for self-normalized large deviation. *Electron. J. Probab.* **10** 1260–1285 (electronic). MR2176384

WU, W. B. (2008). On false discovery control under dependence. *Ann. Statist.* **36** 364–380. MR2387975

DEPARTMENT OF MATHEMATICS
AND INSTITUTE OF NATURAL SCIENCES
SHANGHAI JIAO TONG UNIVERSITY
SHANGHAI
CHINA
E-MAIL: weidongl@sjtu.edu.cn

DEPARTMENT OF STATISTICS
THE CHINESE UNIVERSITY OF HONG KONG
SHATIN, N.T., HONG KONG
CHINA