Bernoulli **17**(1), 2011, 347–394
DOI: 10.3150/10-BEJ272

Simultaneous critical values for t-tests in very high dimensions

HONGYUAN CAO[1] and MICHAEL R. KOSOROK[2]

[1]*Department of Health Studies, 5841 South Maryland Avenue MC 2007, University of Chicago, Chicago, IL, 60637, USA. E-mail: hycao@uchicago.edu*
[2]*Department of Biostatistics and Department of Statistics and Operations Research, 3101 Mcgavran-Greenberg Hall, CB 7420, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA. E-mail: kosorok@unc.edu*

This article considers the problem of multiple hypothesis testing using t-tests. The observed data are assumed to be independently generated conditional on an underlying and unknown two-state hidden model. We propose an asymptotically valid data-driven procedure to find critical values for rejection regions controlling the k-familywise error rate (k-FWER), false discovery rate (FDR) and the tail probability of false discovery proportion (FDTP) by using one-sample and two-sample t-statistics. We only require a finite fourth moment plus some very general conditions on the mean and variance of the population by virtue of the moderate deviations properties of t-statistics. A new consistent estimator for the proportion of alternative hypotheses is developed. Simulation studies support our theoretical results and demonstrate that the power of a multiple testing procedure can be substantially improved by using critical values directly, as opposed to the conventional p-value approach. Our method is applied in an analysis of the microarray data from a leukemia cancer study that involves testing a large number of hypotheses simultaneously.

Keywords: empirical processes; FDR; high dimension; microarrays; multiple hypothesis testing; one-sample t-statistics; self-normalized moderate deviation; two-sample t-statistics

1. Introduction

Among the many challenges raised by the analysis of large data sets is the problem of multiple testing. Examples include functional magnetic resonance imaging, source detection in astronomy and microarray analysis in genetics and molecular biology. It is now common practice to simultaneously measure thousands of variables or features in a variety of biological studies. Many of these high-dimensional biological studies are aimed at identifying features showing a biological signal of interest, usually through the application of large-scale significance testing. The possible outcomes are summarized in Table 1.

Traditional methods that provide strong control of the familywise error rate (FWER $= P(V \geq 1)$) often have low power and can be unduly conservative in many applications. One way around this is to increase the number k of false rejections one is willing to tolerate. This results in a relaxed version of FWER, k-FWER $= P(V \geq k)$.

Benjamini and Hochberg [1] (hereafter referred to as "BH") pioneered an alternative. Define the false discovery proportion (FDP) to be the number of false rejections divided by the number of rejections (FDP $= V/(R \cup 1)$). The only effect of the $R \cup 1$ in the denominator is that the

Table 1. Outcomes when testing m hypotheses

Hypothesis	Accept	Reject	Total
Null true	U	V	m_0
Alternative true	F	S	m_1
Total	W	R	m

ratio V/R is set to zero when $R=0$. Without loss of generality, we treat FDP $= V/R$ and define the false discovery tail probability $FDTP = P(V \geq \alpha R)$, where α is pre-specified, based on the application. Several papers have developed procedures for FDTP control. We shall not attempt a complete review here, but mention the following: van der Laan, Dudoit and Pollard [26] proposed an augmentation-based procedure, Lehmann and Romano [18] derived a step-down procedure and Genoves and Wasserman [13] suggested an inversion-based procedure, which is equivalent to the procedure of [26] under mild conditions [13].

The false discovery rate (FDR) is the expected FDP. BH provided a distribution-free, finite-sample method for choosing a p-value threshold that guarantees that the FDR is less than a target level γ. Since this publication, there has been a considerable amount of research on both the theory and application of FDR control. Benjamini and Hochberg [2] and Benjamini and Yekutieli [3] extended the BH method to a class of dependent tests. A Bayesian mixture model approach to obtain multiple testing procedures controlling the FDR is considered in [11,21–24]. Wu [29] considered the conditional dependence model under the assumption of Donsker properties of the indicator function of the true state for each hypothesis and derived asymptotic properties of false discovery proportions and numbers of rejected hypotheses. A systematic study of multiple testing procedures is given in the book [9]. Other related work can be found in [6,7].

One challenge in multiple hypothesis testing is that many procedures depend on the proportion of null hypotheses, which is not known in reality. Estimating this proportion has long been known as a difficult problem. There have been some interesting developments recently, for example, the approach of [20] (see also [11,13,17,19]). Roughly speaking, these approaches are only successful under a condition which [13] calls the "purity" condition. Unfortunately, the purity condition depends on p-values and is hard to check in practice.

The general framework for k-FWER, FDTP, FDR control and the estimation of the proportion of alternative hypotheses is based on p-values which are assumed to be known in advance or can be accurately approximated. However, the assumption that p-values are always available is not realistic. In some special settings, approximate p-values have been shown to be asymptotically equivalent to exact p-values for controlling FDR [12,16]. However, these approximations are only helpful in certain simultaneous error control settings and are not universally applicable. Moreover, if the p-values are not reliable, any procedures derived later are problematic.

This motivates us to propose a method to find critical values directly for rejection regions to control k-FWER, FDTP and FDR by using one-sample and two-sample t-statistics. The advantage of using t-tests is that they require minimum conditions on the population, only existence of the fourth moment, which is relatively easily satisfied by most statistical distributions, rather than other stringent conditions such as the existence of the moment generating function. In addition, we approximate tail probabilities of both null and alternative hypotheses accurately, rather than

p-value approaches that only consider the case under null hypotheses. Thus, a better ranking of hypotheses is obtained. Furthermore, we propose a consistent estimate of the proportion of alternative hypotheses which only depends on test statistics. As long as the asymptotic distribution of the test statistic is known under the null hypothesis, we can apply our method to estimate this proportion, resulting in more precise cut-offs.

The BH procedure controls the FDR conservatively at $\pi_0\gamma$, where π_0 is the proportion of null hypotheses and γ is the targeted significance level. If π_0 is much smaller than 1, then the statistical power is greatly compromised. The power we use in this paper is $\mathrm{NDR} = E[S]/m_1$, as defined in [8]. In the situation that t-statistics can be used, our procedure gives a better approximation and more accurate critical values can be obtained by plugging in the estimate of π_0. The validity of our approach is guaranteed by empirical process methods and recent theoretical advances on self-normalized moderate deviations, in combination with Berry–Esseen-type bounds for central and non-central t-statistics.

To illustrate, we simulate a Markov chain, as in [25], of Bernoulli variables $(H_i), i = 1, \dots, 5000$, to indicate the true state of each hypothesis test ($H_i = 1$ if the alternative is true; $H_i = 0$ if the null is true). Conditional on the indicator, observations $x_{ij}, i = 1, \dots, 5000, j = 1, \dots, 80$, are generated according to the model $x_{ij} = \mu_i + \epsilon_{ij}$. The one-sample t-statistic is used to perform simultaneous hypothesis testing. Figure 1 shows the plot of 10 000 MCMC results of the realized and nominal FDR control based on the BH method for different control levels. From this plot, we can see that as the control level increases, the BH procedure becomes more and more conservative. For instance, the FDR actually obtained is 0.167 when the nominal level is set at 0.2, reflecting a significant loss in power.

The three methods of multiple testing control we utilize are k-FWER, FDTP and FDR. The criterion for using k-FWER is, asymptotically,

$$P(V \geq k) \leq \gamma. \tag{1.1}$$

Since we only apply our method when there are discoveries ($R > 0$), we need the FDTP, with a given proportion $0 < \alpha < 1$ and significance level $0 < \gamma < 1$, to satisfy, asymptotically,

$$P(V \geq \alpha R) \leq \gamma. \tag{1.2}$$

Similarly, the criterion for using FDR is, asymptotically,

$$FDR \leq \gamma \quad \text{or} \quad \int_0^1 P(V \geq \alpha R)\,\mathrm{d}\alpha \leq \gamma. \tag{1.3}$$

The main contributions of this paper are as follows: (1) Moderate deviation results which only require the finiteness of fourth moment, from which the statistic is computed in probability theory, are applied in multiple testing. Thus, the applicability of this procedure is dramatically expanded: it can deal with non-normal populations and even highly skewed populations. (2) The critical values for rejection regions are computed directly, which circumvents the intermediate p-value step. (3) An asymptotically consistent estimation of the proportion of alternative hypotheses is developed for multiple testing procedures under very general conditions.

The remainder of the paper is organized as follows. In Section 2, we present the basic data structure, our goals, the procedures and theoretical results for the one-sample t-test. Two-sample

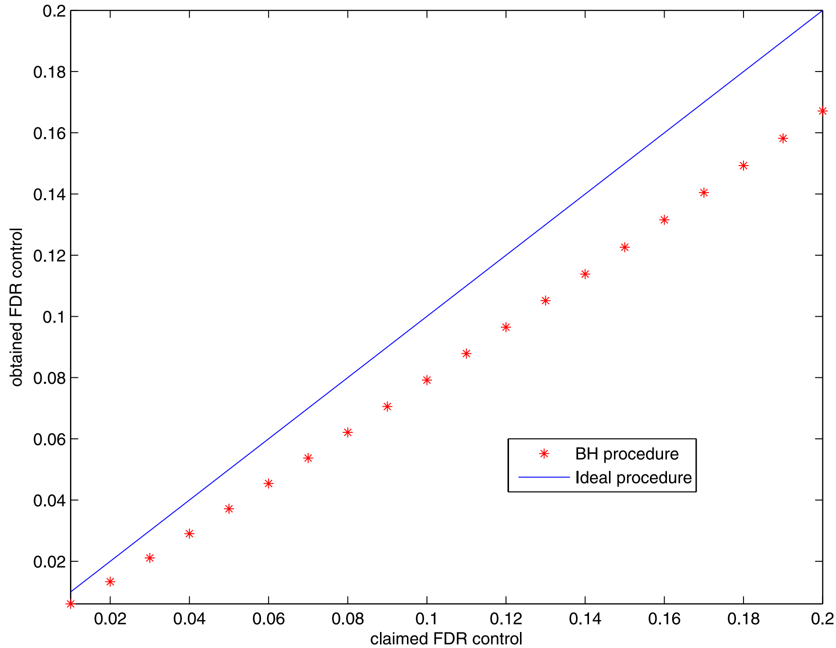

Figure 1. Claimed and obtained FDR control using the BH procedure.

t-test results are discussed in Section 3. Section 4 is devoted to numerical investigations using simulation and Section 5 applies our procedure to detect significantly expressed genes in a microarray study of leukemia cancer. Some concluding remarks and a discussion are given in Section 6. Proofs of results from Sections 2 and 3 are given in the Appendix.

2. One-sample t-test

In this section, we first introduce the basic framework for simultaneous hypothesis testing, followed by our main results. Estimation of the unknown proportion of alternative hypotheses π_1 is presented next. We conclude the section by presenting theoretical results for the special case of completely independent observations. This special setting is the basis for the more general main results and is also of independent interest since fairly precise rates of convergence can be obtained.

2.1. Basic framework

As a specific application of multiple hypothesis testing in very high dimensions, we use gene expression microarray data. At the level of single genes, researchers seek to establish whether each gene in isolation behaves differently in a control versus a treatment situation. If the transcripts are pairwise under two conditions, then we can use a one-sample t-statistic to test for differential expression.

The mathematical model is

$$X_{ij} = \mu_i + \epsilon_{ij}, \qquad 1 \le j \le n, 1 \le i \le m. \tag{2.1}$$

It should be noted that the following discussion is under this model and does not hold in general. Here, X_{ij} represents the expression level in the ith gene and jth array. Since the subjects are independent, for each i, $\epsilon_{i1}, \epsilon_{i2}, \ldots, \epsilon_{in}$ are independent random variables with mean zero and variance σ_i^2. The null hypothesis is $\mu_i = 0$ and the alternative hypothesis is $\mu_i \neq 0$. For the relationship between different genes, we propose the conditional independence model, as follows. Let (H_i) be a $\{0, 1\}$-valued stationary process and, given $(H_i)_{i=1}^m$, $X_{ij}, i = 1, \ldots, m$, are independently generated. The dependence is imposed on the hypothesis (H_i), where $H_i = 0$ if the null hypothesis is true and $H_i = 1$ if the alternative is true. From Table 1, we can see that $\sum_{i=1}^m H_i = m_1$ and $\sum_{i=1}^m (1 - H_i) = m_0$. It is assumed that $(H_i)_{i=1}^m$ satisfy a strong law of large numbers:

$$\frac{1}{m}\sum_{i=1}^m H_i \to \pi_1 \in (0, 1) \qquad \text{a.s.} \tag{2.2}$$

This condition is satisfied in a variety of scenarios, for example, the independent case, Markov models and stationary models. Consider the one-sample t-statistic

$$T_i = \sqrt{n}\bar{X}_i / S_i,$$

where

$$\bar{X}_i = \frac{1}{n}\sum_{j=1}^n X_{ij}, \qquad S_i^2 = \frac{1}{n-1}\sum_{j=1}^n (X_{ij} - \bar{X}_i)^2.$$

If we use t as a cut-off, then the number of rejected hypotheses and the number of false discoveries are, respectively,

$$R = \sum_{i=1}^m 1_{\{|T_i| \ge t\}}, \qquad V = \sum_{i=1}^m (1 - H_i) 1_{\{|T_i| \ge t\}}. \tag{2.3}$$

Under the null hypothesis, it is well known that T_i follows a Student t-distribution with $n - 1$ degrees of freedom if the sample is from a normal distribution. Asymptotic convergence to a standard normal distribution holds when the population is completely unknown, provided that it has a finite fourth moment under the null hypothesis. Moreover, under the alternative hypothesis, T_i can also be approximated by a normal distribution, but with a shift in location. We will show that

$$F_0(t) := P(|T_i| \ge t | H_i = 0) = P(|Z| \ge t)\bigl(1 + \mathrm{o}(1)\bigr) = 2\bar{\Phi}(t)\bigl(1 + \mathrm{o}(1)\bigr), \tag{2.4}$$

$$F_1(t) := P(|T_i| \ge t | H_i = 1) = \mathrm{E}\bigl[P\bigl(\bigl|Z + \sqrt{n}\mu_i/\sigma_i\bigr| \ge t | \mu_i, \sigma_i\bigr)\bigr]\bigl(1 + \mathrm{o}(1)\bigr), \tag{2.5}$$

uniformly for $t = o(n^{1/6})$ under some regularity conditions, where Z denotes the standard normal random variable, $\bar{\Phi}$ is the tail probability of the standard normal distribution and the critical values $t_{n,m}$ that control the FDTP and FDR asymptotically at prescribed level γ are bounded. These assumptions are fairly realistic in practice. We do not require the critical value for k-FWER to be bounded. Although we do not typically know m_1, $F_0(t)$ or $F_1(t)$ in practice, we need the following theorem – the proof of which is given in the Appendix – as the first step. We will shortly extend this result, in Theorem 2.2 below, to permit estimation of the unknown quantities.

Theorem 2.1. *Assume that* $E(\epsilon_{ij}|\mu_i, \sigma_i^2) = 0$, $\operatorname{Var}(\epsilon_{ij}|\mu_i, \sigma_i^2) = \sigma_i^2$, $\limsup E\epsilon_{ij}^4 < \infty$, $0 < \pi_1 < 1 - \alpha$ *and* (2.2) *is satisfied. Also, assume that there exist* $\epsilon_0 > 0$ *and* $c_0 > 0$ *such that*

$$P\bigl(\bigl|\sqrt{n}\mu_i/\sigma_i\bigr| \ge \epsilon_0 | H_i = 1\bigr) \ge c_0 \qquad \forall n \ge 1. \tag{2.6}$$

Let

$$\mu_m(t) = \alpha m_1 F_1(t) - (1-\alpha) m_0 F_0(t) \tag{2.7}$$

and

$$\sigma_m^2(t) = \alpha^2 m_1 F_1(t)\bigl(1 - F_1(t)\bigr) + (1-\alpha)^2 m_0 F_0(t)\bigl(1 - F_0(t)\bigr). \tag{2.8}$$

(i) *If* $t_{n,m}^{fdtp}$ *is chosen such that*

$$t_{n,m}^{fdtp} = \inf\{t : \mu_m(t)/\sigma_m(t) \ge z_\gamma\}, \tag{2.9}$$

where z_γ *is the* γ*th quintile of the standard normal distribution, then*

$$\lim_{m\to\infty} P(FDP \ge \alpha) = \lim_{m\to\infty} P(V \ge \alpha R) \le \gamma \tag{2.10}$$

holds.

(ii) *If* $t_{n,m}^{fdr}$ *is chosen such that*

$$t_{n,m}^{fdr} = \inf\left\{t : \frac{m_0 F_0(t)}{m_0 F_0(t) + m_1 F_1(t)} \le \gamma\right\}, \tag{2.11}$$

then

$$\lim_{m\to\infty} FDR = \lim_{m\to\infty} E(V/R) \le \gamma \tag{2.12}$$

holds.

(iii) *If* $t_{n,m}^{k\text{-}FWER}$ *is chosen such that*

$$t_{n,m}^{k\text{-}FWER} = \inf\bigl\{t : P\bigl(\eta(t) \ge k\bigr) \le \gamma\bigr\}, \tag{2.13}$$

where $\eta(t) \sim \operatorname{Poisson}(\theta(t))$ *and*

$$\theta(t) = m_o F_0(t),$$

then

$$\lim_{m\to\infty} k\text{-}FWER = \lim_{m\to\infty} P(V \geq k) \leq \gamma \tag{2.14}$$

holds.

Remark 2.1. In the next section, we use a Gaussian approximation for $F_0(t)$ and $F_1(t)$ for both FDTP and FDR, for which the critical values are shown to be bounded. In this case, m can be arbitrarily large, while the critical value remains bounded. Due to sparsity, we use a Poisson approximation for k-FWER, for which the critical value is no longer bounded as $m \to \infty$, and we require $\log m = \mathrm{o}(n^{1/3})$.

2.2. Main results

Note that in Theorem 2.1, there are an unknown parameter m_1 and unknown functions $F_0(t)$ and $F_1(t)$ involved in $\mu_m(t)$ and $\sigma_m(t)$. For practical settings, we need to estimate these quantities. We will begin by assuming that we have a strongly consistent estimate of π_1 and will then provide one such estimate in the next section. Given $\mathcal{H}$, note that $p(t) = P(|T_i| \geq t) = (1 - H_i)P(|T_i| \geq t|H_i = 0) + H_i P(|T_i| \geq t|H_i = 1)$ can be estimated from the empirical distribution $\hat{p}_m(t)$ of $\{|T_i|\}$, where

$$\hat{p}_m(t) = \frac{1}{m}\sum_{i=1}^{m} I_{\{|T_i|\geq t\}}, \tag{2.15}$$

and that $P(|T_i| \geq t|H_i = 0)$ is close to $P(|Z| \geq t)$ when n is large, by (2.4). The next theorem, proved in the Appendix, provides a consistent estimate of the critical value $t_{n,m}$.

Theorem 2.2. *Let*

$$\nu_m(t) = \alpha\hat{p}_m(t) - 2(1 - \hat{\pi}_1)\bar{\Phi}(t) \tag{2.16}$$

and

$$\begin{aligned}\tau_m^2(t) &= \alpha^2\big(\hat{p}_m(t) - 2(1 - \hat{\pi}_1)\bar{\Phi}(t)\big)\left(1 - \frac{1}{\hat{\pi}_1}\big(\hat{p}_m(t) - 2(1 - \hat{\pi}_1)\bar{\Phi}(t)\big)\right)\\ &\quad + 2(1 - \alpha)^2(1 - \hat{\pi}_1)\bar{\Phi}(t)\big(1 - 2\bar{\Phi}(t)\big),\end{aligned} \tag{2.17}$$

where $\hat{\pi}_1$ is a strongly consistent estimate of π_1. Assume that the conditions of Theorem 2.1 are satisfied.

(i) *If $\hat{t}_{n,m}^{fdtp}$ is chosen such that*

$$\hat{t}_{n,m}^{fdtp} = \inf\left\{t : \frac{\sqrt{m}\,\nu_m(t)}{\tau_m(t)} \geq z_\gamma\right\}, \tag{2.18}$$

then

$$|\hat{t}_{n,m}^{fdtp} - t_{n,m}^{fdtp}| = \mathrm{o}(1) \qquad a.s. \tag{2.19}$$

(ii) *If* $\hat{t}^{fdr}_{n,m}$ *is chosen such that*

$$\hat{t}^{fdr}_{n,m} = \inf\left\{t : \frac{2(1-\hat{\pi}_1)\bar{\Phi}(t)}{\hat{p}_m(t)} \le \gamma\right\}, \tag{2.20}$$

then

$$|\hat{t}^{fdr}_{n,m} - t^{fdr}_{n,m}| = \mathrm{o}(1) \qquad a.s. \tag{2.21}$$

(iii) *If* $\hat{t}^{k\text{-}FWER}_{n,m}$ *is chosen such that*

$$\hat{t}^{k\text{-}FWER}_{n,m} = \inf\{t : P(\zeta(t) \ge k)\} \le \gamma, \tag{2.22}$$

where $\zeta(t) \sim \mathrm{Poisson}(\bar{\theta}(t))$ *and*

$$\bar{\theta}(t) = 2m(1-\hat{\pi}_1)\bar{\Phi}(t),$$

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

$$|\hat{t}^{k\text{-}FWER}_{n,m} - t^{k\text{-}FWER}_{n,m}| = \mathrm{o}(1) \qquad a.s. \tag{2.23}$$

Remark 2.2. This theorem deals with the general dependence case, where $(H_i)_1^m$ is assumed to follow a two-state hidden model and the data are generated independently conditional on $(H_i)_1^m$. The proof is mainly based on the independence case, which we present in Section 2.4 below, plus a conditioning argument.

2.3. Estimating π_1

In the previous section, we assumed that $\hat{\pi}_1$ was a consistent estimator of π_1. We now develop one such estimator. By the two-group nature of multiple testing, the test statistic is essentially a mixture of null and alternative hypotheses with proportion as a parameter. By virtue of moderate deviations, the distribution of t-statistics can be accurately approximated under both null and alternative hypotheses. However, for the alternative approximation, an unknown mean and variance are involved. So, we think of a functional transformation of the t-statistics which has a ceiling at 1 to first get a conservative estimate of π which is consistent under certain conditions. Let $c > 0$ and define $g_c(x) = \min(|x|, c)/c$. It is easy to see that g_c is a decreasing function of c, bounded by 1, and that the derivative $\frac{\mathrm{d}g_c}{\mathrm{d}c}$ is bounded by $1/c$. Hence, the function class $\{g_c\}$ indexed by c is a Donsker class and thus also Glivenko–Cantelli. Let

$$\hat{g}_c = \frac{1}{m}\sum_{i=1}^{m} g_c(T_i). \tag{2.24}$$

Theorem 2.3. *We have*

$$\pi_1 \ge \lim_{m\to\infty, n\to\infty} \sup_{c>0} \frac{\hat{g}_c - E(g_c(Z))}{1 - E(g_c(Z))} \qquad a.s.$$

If, in addition, we assume that

$$\sqrt{n}\mu_i/\sigma_i \to \infty \qquad \text{for all } i \text{ with } H_i = 1, i = 1, \dots, m, \text{ a.s. as } n \to \infty, \tag{2.25}$$

then

$$\pi_1 = \lim_{m\to\infty, n\to\infty} \sup_{c>0} \frac{\hat{g}_c - E(g_c(Z))}{1 - E(g_c(Z))} \qquad \text{a.s.},$$

where

$$E(g_c(Z)) = \frac{2}{c\sqrt{2\pi}}(1 - \mathrm{e}^{-c^2/2}) + 2\bar{\Phi}(c).$$

Proof. We can write

$$\begin{aligned}\hat{g}_c &= \frac{\sum_{i=1}^m 1_{\{H_i=0\}}}{m} \frac{\sum_{i=1}^m g_c(T_i) 1_{\{H_i=0\}}}{\sum_{i=1}^m 1_{\{H_i=0\}}} + \frac{\sum_{i=1}^m 1_{\{H_i=1\}}}{m} \frac{\sum_{i=1}^m g_c(T_i) 1_{\{H_i=1\}}}{\sum_{i=1}^m 1_{\{H_i=1\}}} \\ &:= \frac{m_0}{m} I + \frac{m_1}{m} II.\end{aligned}$$

Let $\mathcal{H} = \{H_i, 1 \le i \le m\}$. Conditional on $\mathcal{H}$, T_i, $1 \le i \le m$, are independent random variables. We consider I first. Let

$$A_m(c) = \frac{\sum_{i=1}^m g_c(T_i|\mathcal{H}) 1_{\{H_i=0\}}}{\sum_{i=1}^m 1_{\{H_i=0\}}} - \frac{\sum_{i=1}^m E(g_c(T_i|\mathcal{H}) 1_{\{H_i=0\}}}{\sum_{i=1}^m 1_{\{H_i=0\}}},$$

let E be the infinite sequence $1_{\{H_1=0\}}, 1_{\{H_2=0\}}, \dots$ and let F be the event that $\sum_{i=1}^m 1_{\{H_i=0\}} \to \infty$ as $m \to \infty$. By the assumption (2.2), we know that $P(F) = 1$. Thus,

$$P\Big(\lim_{m\to\infty} \sup_{c>0} |A_m(c)| = 0\Big) = E\Big[P\Big(\lim_{m\to\infty} \sup_{c>0} |A_m(c)| = 0 \big| E\Big)\Big] = 1,$$

where the second equality follows from the fact that, conditional on E, the terms in the sum are i.i.d. and thus the standard Glivenko–Cantelli theorem applies. Arguing similarly, based on conditioning on the sequence $1_{\{H_1=1\}}, 1_{\{H_2=1\}}, \dots,$ we can also establish that

$$\sup_{c>0} \left| \frac{\sum_{i=1}^m g_c(T_i|\mathcal{H}) 1_{\{H_i=1\}}}{\sum_{i=1}^m 1_{\{H_i=1\}}} - \frac{\sum_{i=1}^m E(g_c(T_i|\mathcal{H}) 1_{\{H_i=1\}}}{\sum_{i=1}^m 1_{\{H_i=1\}}} \right| \to 0 \qquad \text{a.s.}$$

Now, note that $II \le 1$. Thus, since $m_0/m \to (1 - \pi_1)$ a.s. and $m_1/m \to \pi_1$ a.s., we have that when $m \to \infty, n \to \infty$,

$$\begin{aligned}\hat{g}_c &\le (1 - \pi_1)E(g_c(Z)) + \pi_1 \qquad \text{a.s.} \\ &= E(g_c(Z)) + \big(1 - E(g_c(Z))\big)\pi_1.\end{aligned}$$

We now have the following lower bound for π_1:

$$\pi_1 \ge \lim_{m\to\infty, n\to\infty} \sup_{c>0} \frac{\hat{g}_c - E(g_c(Z))}{1 - E(g_c(Z))} \qquad \text{a.s.} \tag{2.26}$$

Define

$$\Delta_1 := (1-\pi_1)E(g_c(Z)) + \pi_1 \frac{1}{m_1} \sum_{i=1}^{m} E(g_c(T_i)|\mathcal{H}) 1_{\{H_i=1\}},$$

$$\Delta_2 := (1-\pi_1)E(g_c(Z)) + \pi_1 \frac{\sum_{i=1}^m E(g_c(Z+\sqrt{n}\mu_i/\sigma_i)) 1_{\{H_i=1\}}}{\sum_{i=1}^m 1_{\{H_i=1\}}}.$$

Letting $n \to \infty$, we have $\sup_{c>0} |\Delta_1 - \Delta_2| \to 0$ a.s. Also,

$$\begin{aligned}
\Delta_2 &= (1-\pi_1)E(g_c(Z)) \\
&\quad + \pi_1 \frac{1}{\sum_{i=1}^m 1_{\{H_i=1\}}} \sum_{i=1}^{m} E\biggl(g_c\biggl(Z + \frac{\sqrt{n}\mu_i}{\sigma_i}\biggr)\bigl(I_{\{|Z+\sqrt{n}\mu_i/\sigma_i| \ge c\}} + I_{\{|Z+\sqrt{n}\mu_i/\sigma_i| < c\}}\bigr)\biggr) H_i \\
&\ge (1-\pi_1)E(g_c(Z)) + \pi_1 \frac{\sum_{i=1}^m P(|Z+\sqrt{n}\mu_i/\sigma_i| \ge c) H_i}{\sum_{i=1}^m 1_{\{H_i=1\}}} \\
&\ge (1-\pi_1)E(g_c(Z)) + \pi_1 \\
&= E(g_c(Z)) + \pi_1\bigl(1 - E(g_c(Z))\bigr).
\end{aligned}$$

Note that

$$\sup_c |\hat{g}_c - \Delta_1| \to 0 \qquad \text{a.s. as } m \to \infty, n \to \infty.$$

Therefore,

$$\hat{g}_c \ge E(g_c(Z)) + \pi_1\bigl(1 - E(g_c(Z))\bigr) \qquad \text{a.s. as } m \to \infty, n \to \infty.$$

Thus, we obtain

$$\pi_1 \le \lim_{m\to\infty, n\to\infty} \sup_{c>0} \frac{\hat{g}_c - E(g_c(Z))}{1 - E(g_c(Z))} \qquad \text{a.s.} \tag{2.27}$$

□

As a consequence of this theorem, we propose the following estimate of π_1:

$$\hat{\pi}_1 := \sup_{c>0} \frac{\hat{g}_c - E(g_c(Z))}{1 - E(g_c(Z))}, \tag{2.28}$$

where

$$E(g_c(Z)) = \frac{2}{c\sqrt{2\pi}}(1 - \mathrm{e}^{-c^2/2}) + 2\bar{\Phi}(c).$$

Remark 2.3. If we use $\hat{\pi}_1$, as given in (2.28), then Theorem 2.2 yields a fully automated procedure to carry out multiple hypothesis testing in very high dimensions in practical data settings.

2.4. Consistency and rate of convergence under independence

In order to prove the main results in the general, possibly dependent, t-test setting, we need results under the assumption of independence between t-tests. Specifically, we assume in this section that $(T_i, H_i), i = 1, \dots, m$ are independent, identically distributed random variables with $\pi_1 = P(T_i = 1)$. This independence assumption can also yield stronger results than the more general setting and is of independent interest.

The next theorem, proved in the Appendix, provides a strong consistent estimate of the critical value $t_{n,m}$, as well as its rate of convergence.

Theorem 2.4. *Let*

$$\nu_m(t) = \alpha \hat{p}_m(t) - 2(1 - \pi_1)\bar{\Phi}(t) \tag{2.29}$$

and

$$\begin{aligned}\tau_m^2(t) = \alpha^2 \hat{p}_m(t)\big(1 - \hat{p}_m(t)\big) + 4\alpha(1 - \pi_1)\hat{p}_m(t)\bar{\Phi}(t) \\ + 2(1 - \pi_1)\bar{\Phi}(t)\big(1 - 2\alpha - 2(1 - \pi_1)\bar{\Phi}(t)\big).\end{aligned}$$

Assume the conditions of Theorem 2.1 with (2.2) replaced by the assumption that $(T_i, H_i), i = 1, \dots, m$, are i.i.d. and $\pi_1 = P(T_i = 1)$. Let $\mathcal{J} = \{i : H_i = 1\}$ be the set that contains the indices of alternative hypotheses. Also, assume that μ_i, σ_i are i.i.d. for $i \in \mathcal{J}$.

(i) *If $\hat{t}_{n,m}^{fdtp}$ is chosen such that*

$$\hat{t}_{n,m}^{fdtp} = \inf\left\{t : \frac{\sqrt{m}\nu_m(t)}{\tau_m(t)} \ge z_\gamma\right\}, \tag{2.30}$$

then

$$|\hat{t}_{n,m}^{fdtp} - t_{n,m}^{fdtp}| = \mathrm{O}\big(n^{-1/2} + m^{-1/2}(\log\log m)^{1/2}\big) \qquad a.s. \tag{2.31}$$

and

$$|\hat{t}_{n,m}^{fdtp} - t_{n,m}^{fdtp}| = \mathrm{O}(n^{-1/2} + m^{-1/2}) \qquad \text{in probability.} \tag{2.32}$$

Here, $t_{n,m}^{fdtp}$ is the critical value defined in (A.26).

(ii) *If $\hat{t}_{n,m}^{fdr}$ is chosen such that*

$$\hat{t}_{n,m}^{fdr} = \inf\left\{t : \frac{2(1 - \pi_1)\bar{\Phi}(t)}{\hat{p}_m(t)} \le \gamma\right\}, \tag{2.33}$$

then

$$|\hat{t}_{n,m}^{fdr} - t_{n,m}^{fdr}| = \mathrm{O}\big(n^{-1/2} + m^{-1/2}(\log\log m)^{1/2}\big) \qquad a.s. \tag{2.34}$$

and

$$|\hat{t}_{n,m}^{fdr} - t_{n,m}^{fdr}| = \mathrm{O}(n^{-1/2} + m^{-1/2}) \qquad \text{in probability.} \tag{2.35}$$

Here, $t_{n,m}^{fdr}$ *is the critical value defined in* (A.28).

(iii) *If* $\hat{t}_{n,m}^{k\text{-}FWER}$ *is chosen such that*

$$\hat{t}_{n,m}^{k\text{-}FWER} = \inf\{t : P(\zeta(t) \geq k)\} \leq \gamma, \tag{2.36}$$

where $\zeta(t) \sim \mathrm{Poisson}(\bar{\theta}(t))$ *and*

$$\bar{\theta}(t) = 2m(1 - \hat{\pi}_1)\bar{\Phi}(t),$$

then

$$|\hat{t}_{n,m}^{k\text{-}FWER} - t_{n,m}^{k\text{-}FWER}| = \mathrm{O}((\log m)^{-1/2}) \qquad a.s. \tag{2.37}$$

Here $t_{n,m}^{k\text{-}FWER}$ *is the critical value defined in* (A.30).

Remark 2.4. If $\alpha = \gamma$ in Theorem 2.4, then it is not difficult to see that $\hat{t}_{n,m}^{fdtp} - \hat{t}_{n,m}^{fdr} = \mathrm{O}(m^{-1/2})$ $a.s$. Therefore, (2.31) and (2.32) remain valid with $\hat{t}_{n,m}^{fdtp}$ replaced by $\hat{t}_{n,m}^{fdr}$. This shows that controlling FDTP is asymptotically equivalent to controlling FDR. This is also true in the more general dependence case. Thus, we will focus primarily on FDR in our numerical studies.

Remark 2.5. Note that π_1 is assumed to be known in order to get a precise rate of convergence for FDTP and FDR. If $\hat{\pi}_1$ is estimated with rate of convergence r_n, then the correct convergence rate for the "in probability" result for FDR and FDTP would involve an additional term $\mathrm{O}(r_n)$ added in (2.32) and (2.35). It is unclear what the correction would be for the almost sure rate in (2.31) and (2.34). These corrections are beyond the scope of this paper and will not be pursued further here. Note that the rate of $\hat{\pi}_1$ is not needed in the main results presented in Sections 2.1–2.3.

3. Two-sample t-test

In this section, the results of the previous section are extended to the two-sample t-test setting. The estimator of the unknown parameter π_1 remains the same as in the one-sample case, but with T_i in (2.24) being the two-sample, rather than one-sample, t-statistic. Theoretical results for the rates of convergence under independence are also presented, as in the previous section.

3.1. Basic set-up and results

When two groups, such as a control and an experimental group, are independent, which we assume here, a natural statistic to use is the two-sample t-statistic. As far as possible, we adopt the same notation as used in the one-sample case, and we assume that (2.2) holds. We observe the random variables

$$X_{ij} = \mu_i + \epsilon_{ij}, \qquad 1 \leq j \leq n_1, 1 \leq i \leq m, \qquad Y_{ij} = \nu_i + \omega_{ij}, \qquad 1 \leq j \leq n_2, 1 \leq i \leq m,$$

with the index i denoting the ith gene, j indicating the jth array, μ_i representing the mean effect for the ith gene from the first group and ν_i representing the mean effect for the ith gene from the second group. The sampling processes for the two groups are assumed to be independent of each other. The sample sizes n_1 and n_2 are assumed to be of the same order, that is, $0 < b_1 \le n_1/n_2 \le b_2 < \infty$. We will also assume that for each i, $\epsilon_{i1}, \epsilon_{i2}, \dots, \epsilon_{in_1}$ are independent random variables with mean zero and variance σ_i^2; $\omega_{i1}, \omega_{i2}, \dots, \omega_{in_2}$ are independent random variables with mean zero and variance τ_i^2. The null hypothesis is $\mu_i = \nu_i$, the alternative hypothesis is $\mu_i \ne \nu_i$ and the dependence is assumed to be generated in the same manner as the dependence in the one-sample setting. Consider the two-sample t-statistic

$$T_i^* = \frac{\bar{X}_i - \bar{Y}_i}{\sqrt{S_{1i}^2/n_1 + S_{2i}^2/n_2}},$$

where

$$\bar{X}_i = \frac{1}{n_1}\sum_{j=1}^{n_1} X_{ij}, \qquad \bar{Y}_i = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_{ij},$$

$$S_{1i}^2 = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1}(X_{ij} - \bar{X}_i)^2, \qquad S_{2i}^2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(Y_{ij} - \bar{Y}_i)^2.$$

Then

$$R = \sum_{i=1}^{m} 1_{\{|T_i^*| \ge t\}}, \qquad V = \sum_{i=1}^{m}(1 - H_i)1_{\{|T_i^*| \ge t\}}. \tag{3.1}$$

The two-sample t-statistic is one of the most commonly used statistics to construct confidence intervals and carry out hypothesis testing for the difference between two means. There are several premises underlying the use of two-sample t-tests. It is assumed that the data have been derived from populations with normal distributions. Based on the fact that $S_{1i} \to \sigma_i$, $S_{2i} \to \tau_i$ a.s., with moderate violation of the assumption, statisticians quite often recommend using the two-sample t-test, provided the samples are not too small and the samples are of equal or nearly equal size. When the populations are not normally distributed, it is a consequence of the central limit theorem that two-sample t-tests remain valid. A more refined confirmation of this validity under non-normality based on moderate deviations is shown in [4]. Furthermore, under the alternative hypothesis, the asymptotic results still hold, but with a shift in location similar to the one-sample case under certain conditions, that is,

$$P(|T_i^*| \ge t | H_i = 0) = P(|Z| \ge t)\big(1 + \mathrm{o}(1)\big),$$

$$P(|T_i^*| \ge t | H_i = 1) = P\left(\left|Z + \frac{\mu_i - \nu_i}{B_{n_1,n_2}}\right| \ge t\right)\big(1 + \mathrm{o}(1)\big),$$

uniformly in $t = \mathrm{o}(n^{1/6})$, where $B_{n_1,n_2}^2 = \sigma_i^2/n_1 + \tau_i^2/n_2$. Under the assumption of (2.2), asymptotic critical values to control FDTP, FDR and k-FWER are very similar to the one-sample t-test case with the one-sample t-statistic T_i replaced by the two-sample t-statistic T_i^*. The following theorem, proved in the Appendix, is analogous to Theorem 2.1 and is a necessary first step.

Theorem 3.1. *Assume that* $E(\epsilon_{ij}|\mu_i,\ \sigma_i^2) = 0$, $E(\omega_{ij}|\nu_i,\ \tau_i^2) = 0$, $\mathrm{Var}(\epsilon_{ij}|\mu_i, \sigma_i^2) = \sigma_i^2$, $\mathrm{Var}(\omega_{ij}|\nu_i, \tau_i^2) = \tau_i^2$, $\limsup E\epsilon_{ij}^4 < \infty$, $\limsup E\tau_{i,j}^4 < \infty$, $0 < \pi_1 < 1 - \alpha$ *and that* (2.2) *is satisfied. Assume that there exist* ϵ_0 *and* c_0 *such that*

$$P\left(\left|\frac{\mu_i - \nu_i}{B_{n_1,n_2}}\right| \geq \epsilon_0 \middle| H_i = 1\right) \geq c_0 \qquad \text{for all } n_1, n_2. \tag{3.2}$$

The conclusions of Theorem 2.1 then hold with the one-sample t-statistic T_i replaced by the two-sample t-statistic T_i^*.

3.2. Main results

The unknown parameter m_1 and functions $F_0(t)$ and $F_1(t)$ in Theorem 3.1 are estimated similarly as in the one-sample case with the one-sample t-statistic replaced by its two-sample counterpart. The following theorem, the proof of which is given in the Appendix, gives our main results for two-sample t-tests.

Theorem 3.2. *Assume that the conditions in Theorem 3.1 are satisfied. Replace the one-sample t-statistic T_i by the two-sample t-statistic T_i^* in Theorem 2.2. Let $\hat{\pi}_1$ be a strong consistent estimate of π_1, as in (2.28), using the two-sample t-statistic T_i^*.*

(i) *If $\hat{t}_{n,m}^{fdtp}$ is chosen such that*

$$\hat{t}_{n,m}^{fdtp} = \inf\left\{t : \frac{\sqrt{m}\nu_m(t)}{\tau_m(t)} \geq z_\gamma\right\}, \tag{3.3}$$

then

$$|\hat{t}_{n,m}^{fdtp} - t_{n,m}^{fdtp}| = \mathrm{o}(1) \qquad a.s. \tag{3.4}$$

(ii) *If $\hat{t}_{n,m}^{fdr}$ is chosen such that*

$$\hat{t}_{n,m}^{fdr} = \inf\left\{t : \frac{2(1-\hat{\pi}_1)\bar{\Phi}(t)}{\hat{p}_m(t)} \leq \gamma\right\} \tag{3.5}$$

then

$$|\hat{t}_{n,m}^{fdr} - t_{n,m}^{fdr}| = \mathrm{o}(1) \qquad a.s. \tag{3.6}$$

(iii) *If $\hat{t}_{n,m}^{k\text{-}FWER}$ is chosen such that*

$$\hat{t}_{n,m}^{k\text{-}FWER} = \inf\{t : P(\zeta(t) \geq k)\} \leq \gamma, \tag{3.7}$$

where $\zeta(t) \sim \text{Poisson}(\bar{\theta}(t))$ *and*

$$\bar{\theta}(t) = 2m(1-\hat{\pi}_1)\bar{\Phi}(t),$$

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

$$|\hat{t}_{n,m}^{k\text{-}FWER} - t_{n,m}^{k\text{-}FWER}| = \mathrm{o}(1) \qquad a.s. \tag{3.8}$$

Remark 3.1. $\hat{\pi}_1$ can be estimated via (2.28) by using two-sample t-statistics. Theorem 2.3 is applicable in the two-sample setting, as well as in the one-sample case, and consistency follows. Thus, Theorem 3.2 gives a fully automated procedure to conduct multiple hypothesis testing using two-sample t-statistics after we plug in the $\hat{\pi}_1$ given in (2.28).

3.3. Consistency and rate of convergence under independence

Results for the independence setting are needed for the proofs of the main results, as was the case for one-sample t-tests. We can, once again, obtain more precise estimation compared with the general dependence case. The following theorem, proved in the Appendix, gives us conditions and conclusions using two-sample t-statistics for controlling FDTP and FDR asymptotically, as well as rates of convergence under the assumption that (T_i, H_i) are independent of each other for $1 \le i \le m$. Assume that π_1 is the proportion of the alternative hypotheses among m hypothesis tests, that is, $\pi_1 = P(H_i = 1)$. Let $\mathcal{J} = \{i : H_i = 1\}$.

Theorem 3.3. *Assume the conditions of Theorem 3.1 are satisfied. Rather than (2.2), we assume that (T_i, H_i) are independent and identically distributed. In addition, $\pi_1 = P(T_1 = 1)$ and μ_i, σ_i are i.i.d. for $i \in \mathcal{J}$. Let*

$$p(t) = P(|T_1^*| \ge t), \tag{3.9}$$

$$a_1(t) = \alpha p(t) - (1-\pi_1)P(|T_1^*| \ge t | H_1 = 0), \tag{3.10}$$

$$\begin{aligned} b_1^2(t) = {} & \alpha^2 p(t)\bigl(1-p(t)\bigr) + 2\alpha(1-\pi_1)p(t)P(|T_1^*| \ge t|H_1 = 0) \\ & + (1-\pi_1)P(|T_1^*| \ge t|H_1 = 0)\bigl(1-2\alpha-(1-\pi_1)P(|T_1^*| \ge t|H_1 = 0)\bigr), \end{aligned}$$

$$\hat{p}_m(t) = \frac{1}{m}\sum_{i=1}^{m} I_{\{|T_i^*| \ge t\}}, \tag{3.11}$$

$$\nu_m(t) = \alpha\hat{p}_m(t) - 2(1-\pi_1)\bar{\Phi}(t), \tag{3.12}$$

and

$$\begin{aligned} \tau_m^2(t) = {} & \alpha^2\hat{p}_m(t)\bigl(1-\hat{p}_m(t)\bigr) + 4\alpha(1-\pi_1)\hat{p}_m(t)\bar{\Phi}(t) \\ & + 2(1-\pi_1)\bar{\Phi}(t)\bigl(1-2\alpha-2(1-\pi_1)\bar{\Phi}(t)\bigr). \end{aligned}$$

The conclusions of Theorem 2.4 then hold with the one-sample t-statistics T_i replaced by the two-sample t-statistics T_i^*.

Remark 3.2. In the above sections, we developed our theorems based on two-sided tests. The results for the case of one-sided tests are very similar, but with the rejection region $\{T_i \geq t\}$ for each test. We omit the details.

4. Numerical studies

In this section, we present numerical studies based on simulated data and compare the power of our approach with [1] (BH) and [23] (ST) approaches using one-sample t-statistics. The results for using two-sample t-statistics are very similar and so we omit the details here.

4.1. Simulation study 1

We investigate the results for the i.i.d. case first. Recall the model

$$X_{ij} = \mu_i + \epsilon_{ij}, \qquad 1 \leq i \leq m, 1 \leq j \leq n.$$

We set the signal using $\mu_i \sim \mathit{Unif}(0.5, 1)$ or $\mu_i \sim \mathit{Unif}(-1, -0.5)$, which is of the correct order for the standardized error term. Here, the number of hypothesis tests is $m = 10\,000$, which is the same for all following simulation studies, unless otherwise noted. The proportion of alternatives $\pi_1 = 0.2$ and the error term $t(4)$ are used just to illustrate the asymptotic results. We vary the number of arrays n from 20 to 50 to 300 to evaluate our asymptotic approximation. Empirical distributions of FDTP, FDR and k-FWER based on 100 000 repetitions are treated as the gold standard since they have almost negligible Monte Carlo error. The samples are generated to evaluate our proposed method based on asymptotic theory. Specifically, for each sample, we calculate the sample paths of the following quantities indexed by t: $\sqrt{m}\nu_m(t)/\tau_m(t)$ for studying FDTP, $2(1 - \hat{\pi}_1)\bar{\Phi}(t)/\hat{p}_m(t)$ for studying FDR and $P(\text{Poisson}(2m(1 - \hat{\pi}_1)\bar{\Phi}(t)) \geq 10)$ for studying 10-FWER (here, we choose $k = 10$ just for the purposes of illustration). $\hat{\pi}_1$ is defined as in (2.28).

Figure 2 shows the overlay of the true path and 100 random estimated paths for FDTP, FDR and k-FWER, respectively. As n increases, we see that the true path and estimated paths are fairly close to each other, which, in turn, validates our asymptotic theory. We can see that the slopes of FDTP and 10-FWER are very steep, which means a small change in the critical value results in a large change in the level of control, while the FDR has a flatter trend.

4.2. Simulation study 2

Under the same set-up as in the previous section, we simulate data with different error terms: standard normal ($N(0, 1)$), Student t with one degree of freedom (Cauchy), Student t with four degrees of freedom ($t(4)$), Student t with ten degrees of freedom ($t(10)$), Laplace and exponential. Note that, except for the Cauchy error term, all of the error terms satisfy the condition

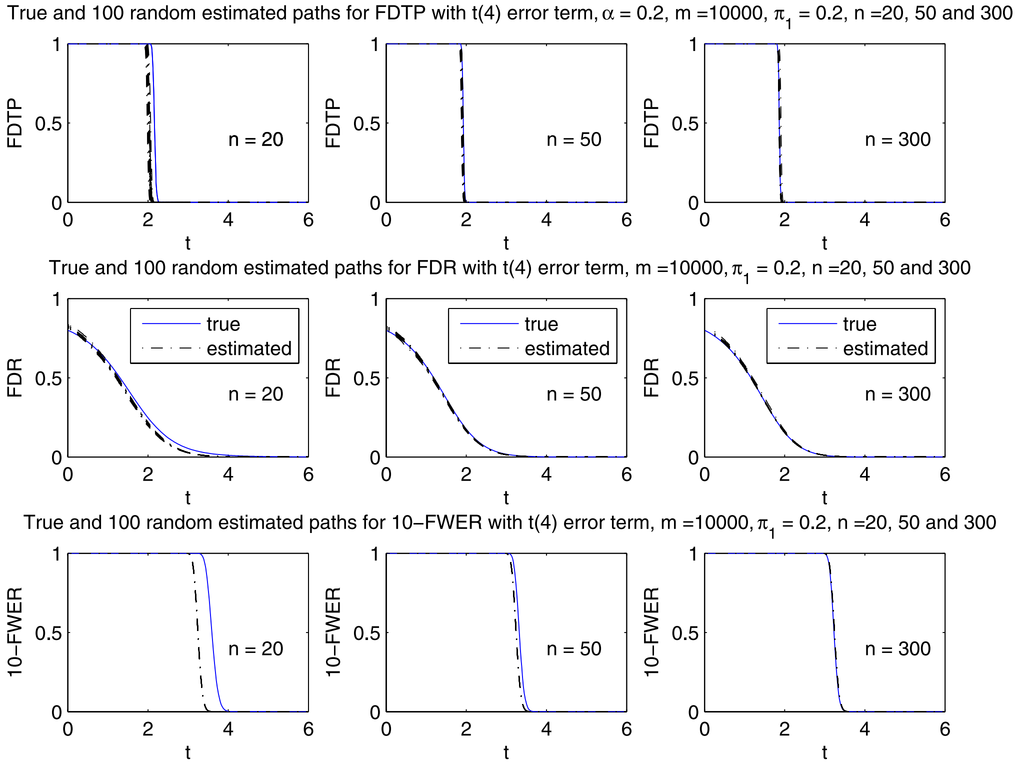

Figure 2. Overlay of true and 100 random estimated sample paths with respect to cut-off t for the three procedures under differing sample sizes.

of finite fourth moment. Empirical distributions of FDTP, FDR and k-FWER based on 100 000 repetitions are treated as the gold standard for obtaining true critical values. Each scenario is repeated 1000 times to evaluate our proposed method for estimating the critical value based on asymptotic theory. We control FDR at different levels (from 0.01 to 0.2) to get true and estimated critical values. Asymptotically, the estimated critical value $\hat{t}$ based on our theory should be very close to the true critical value t and lie on a diagonal line of the square. From Figure 3, the estimated critical values $\hat{t}$ do not match the true critical value t under the Cauchy error since the Cauchy distribution does not have finite fourth moment. For the Cauchy distribution, even the central limit theorem does not hold since it does not have finite mean. As the number of arrays n increases, the estimated critical values $\hat{t}$ match the true critical values t better under symmetric error terms ($N(0, 1), t(4), t(10)$ and Laplace), but not quite so well under asymmetric errors (e.g., exponential errors). The difficulty with the exponential error terms suggests the value of conducting research to derive higher order approximations. We plan to undertake this in the near future.

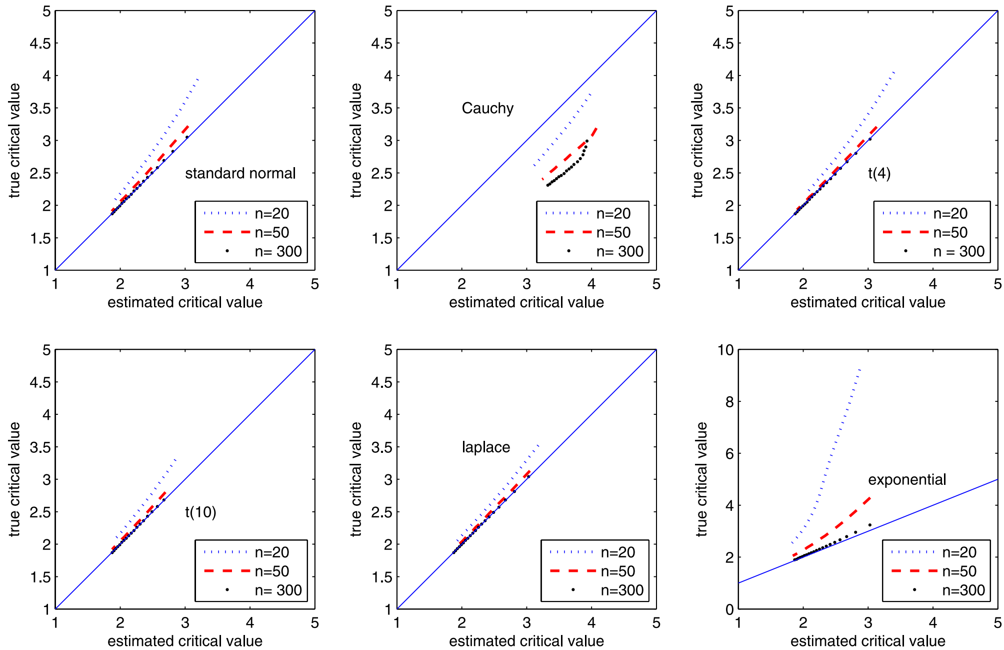

Figure 3. Comparison of true and estimated critical values using FDR for different error terms and numbers of arrays n.

4.3. Simulation study 3

The above results are from the independent test setting. We carried out similar simulation studies for the dependent setting and found that the corresponding plots are quite similar to the above results and the same conclusions can be drawn. To see whether our proposed method obtains the claimed level of control, we use a hidden Markov chain to generate dependent indicators $H_i, i = 1, \ldots, m$. Conditional on $H_i, i = 1, \ldots, m$, the data is generated independently. The transition probability of the hidden Markov chain is set to

$$\begin{pmatrix} 1 - p_1 p_1 \\ p_0 1 - p_0 \end{pmatrix},$$

where p_1 is the transition probability from 0 to 1 and p_0 is the transition probability from 1 to 0. In the simulation, $p_0 = 0.8$ and $p_1 = 0.2$. Based on the limiting stationary distribution, the alternative proportion should be $\pi_1 = p_1/(p_0 + p_1)$. Under the null hypothesis, we simulate data from four error terms ($N(0, 1)$, $t(4)$, Laplace and exponential) and, under the alternative hypothesis, we simulate data with mean effects half from $Unif(0.1, 0, 5)$ and half from $Unif(-0.5, -0.1)$, plus the same four error terms. Figure 4 uses FDR as the control criterion. For different control levels γ, we compare the claimed level of control and the actually obtained level of control

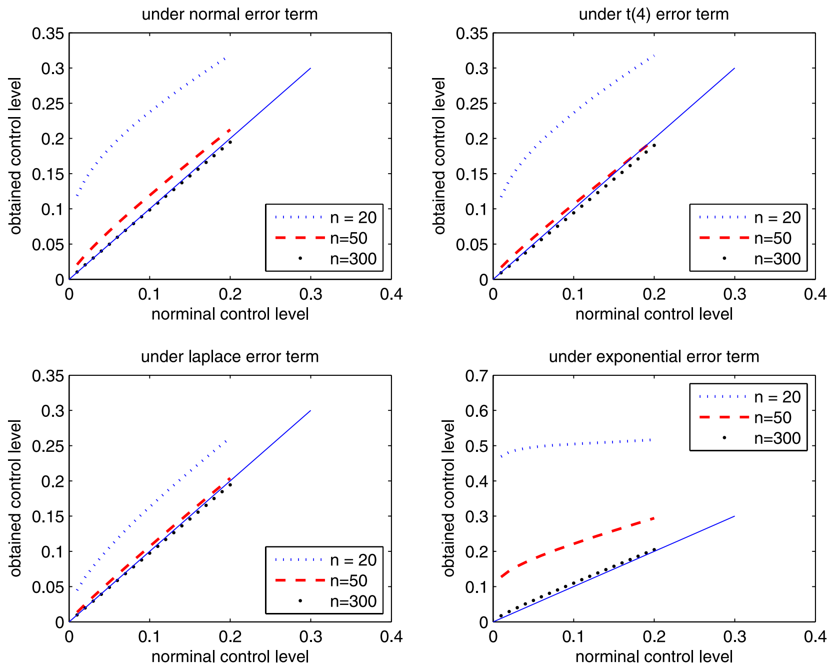

Figure 4. Comparison of nominal and obtained control level for different error terms and numbers of arrays n.

based on our method for different numbers of arrays: small ($n = 20$), medium ($n = 50$) and large ($n = 300$).

From Figure 4, we can see that when the number of arrays n is small ($n = 20$), we do not, in general, achieve the claimed level of control. If we have a medium sample size ($n = 50$), the obtained level of control is very close to the nominal level of control and the results are almost perfect if we have a large number of arrays ($n = 300$), even for the asymmetric exponential error term. This strongly supports our theoretical predictions but suggests that higher order approximations would be useful in some settings.

To see the performance of our method using 10-FWER, Table 2 summarizes the control level actually obtained for different error terms and numbers of arrays n when the nominal control

Table 2. Obtained control level using 10-FWER with nominal control level 0.05

n	$N(0, 1)$	$t(4)$	Laplace	Exponential
20	0.998 (9.0e−05)	0.90 (7.0e−03)	0.81 (1.1e−02)	1 (0)
50	0.52 (1.2e−02)	0.14 (9.1e−03)	0.17 (1.2e−02)	1 (0)
300	0.076 (3.8e−03)	0.031 (2.8e−03)	0.05 (2.7e−03)	0.82 (4.6e−03)

level is 0.05. The obtained control level is incorrect when the number of arrays n is small, which can be deduced from the samples paths of 10-FWER given in Figure 1. It has a very steep slope, so when n is small, the approximation is crude and there is a noticeable difference between the estimated critical value and the true critical value, yielding a big difference in the control level. For large sample sizes, the obtained control level is reasonably good because our asymptotic theory begins to take effect. The exponential error setting appears not to perform as well as the other error settings.

4.4. Simulation study 4

All previous numerical studies involve the alternative proportion estimate $\hat{\pi}_1$ defined in (2.28). In this section, we investigate numerically how this estimate is affected by number of arrays n and compare with the alternative estimate proposed by [23]. The first simulation set-up is similar to the one in the previous section. We drew $N = 1000$ sets of data as follows. Dependent indicators $H_i, i = 1, \ldots, m$, are generated from a hidden Markov chain with the limiting alternative proportion $\pi_1 = 0.2$. Conditional on these, a vector of expected values, $\mu = (\mu_1, \ldots, \mu_m)$, was constructed. The expected values for the true null hypotheses were set to 0 with standard normal noise, whereas the expected values for the alternative hypotheses were drawn from $\mathit{Unif}(0.1, 0.5)$ plus standard normal noise. Correspondingly, 1000 replications of the proportion estimate $\hat{\pi}_1$ were calculated using (2.28). The root means square error (RMSE) is given as

$$\mathit{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\bigl(\hat{\pi}_1^{(n)} - \pi_1^{(n)}\bigr)^2},$$

where $\hat{\pi}_1^{(n)}$ is the estimate of π_1 for the nth simulated data set and $\pi_1^{(n)}$ is the truth. Table 3 summarizes the effect of n. As the number of arrays n increases, the RMSE gets smaller, which validates our asymptotic prediction.

In the second simulation, we compare our proportion estimate with the one using spline smoothing proposed by [23]. Recall the proportion estimate $\pi_0(\lambda) = \#\{p_i > \lambda; i = 1, \ldots, m\}/(m(1-\lambda))$. The smoothing approach proceeds as follows: first, $\pi_0(\lambda)$ are calculated over a (fine) grid of λ; then, a natural cubic spline y with three degrees of freedom is fitted to $(\lambda, \hat{\pi}_0(\lambda))$; finally, π_0 is estimated by $\hat{\pi}_0 = y(1)$. The simulation set-up is similar to the previous one, except that we have two groups here with $n_1 = 70$ and $n_2 = 80$. We change the alternative proportion to compare the performances of our approach (π_1^{ck}) with the spline smoothing approach (π_1^{st}) in Table 4. They produce very similar results; both are conservative, with less bias using our approach and less variance using the spline smoothing approach. The advantage of our approach is that it

Table 3. RMSE for $N = 1000$ estimated values of π_1

n	20	50	300
RMSE	0.0156	0.0136	0.0104

Table 4. Proportion estimate comparison

π_1	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
$\hat{\pi}_1^{ck}$	0.044	0.091	0.141	0.182	0.217	0.255	0.289	0.335	0.365
$\hat{\pi}_1^{st}$	0.041	0.081	0.125	0.161	0.195	0.236	0.276	0.323	0.355
$sd(\hat{\pi}_1^{ck})$	0.042	0.043	0.041	0.040	0.046	0.041	0.047	0.042	0.038
$sd(\hat{\pi}_1^{st})$	0.039	0.041	0.036	0.040	0.041	0.038	0.034	0.036	0.031

is computationally very fast, while the spline smoothing approach requires that p-values are first obtained using permutation, which is computationally much more intensive than our approach (which can be computed directly from the t-statistics).

4.5. Comparison with BH and ST procedures

In this section, we compare our approach with the BH and ST procedures under the dependence structure described in [29]. We also use a hidden Markov model to simulate the indicator function $H_i, i = 1, \dots, m$. Conditional on $H_i, i = 1, \dots, m$, the data is generated independently. The number of hypotheses tested $m = 5000$ and the number of arrays $n = 80$. The data generating mechanism is otherwise the same as in the independence case. First, we construct a one-sample t-statistic and apply our procedure to obtain the critical value for the rejection region. We then obtain p-values and q-values, and apply the BH and ST procedures to decide which genes are significantly expressed. We now briefly describe the BH procedure. Let p_i be the marginal p-value of the ith test, $1 \le i \le m$, and let $p_{(1)} \le \cdots \le p_{(m)}$ be the order statistics of $p_1, \dots, p_m$. Given a control level $\gamma \in (0, 1)$, let

$$r = \max\{i \in \{0, 1, \dots, m+1\} : p_{(i)} \le \gamma i/m\},$$

where $p_0 = 0$ and $p_{(m+1)} = 1$. The BH procedure rejects all hypotheses for which $p_{(i)} \le p_{(r)}$. If $r = 0$, then all hypotheses are accepted. The q-value in [23] is similar to the well-known p-value, except that it is a measure of significance in terms of FDR, rather than type I error, and an estimate of alternative proportion is plugged in, based on available p-values, as described in the previous section. We revisit the motivating example and give a plot of the claimed FDR and actually obtained FDR by using the proposed critical value method. From Figure 5, we can see that our procedure controls the FDR at the claimed level asymptotically, although somewhat liberally for finite samples, and has better power at the same target FDR level compared with the BH and ST procedures.

5. Applications to microarray analysis

We now apply the proposed procedure to the analysis of a leukemia cancer data set [14] in order to identify differentially expressed genes between AML and ALL. For the original data, see

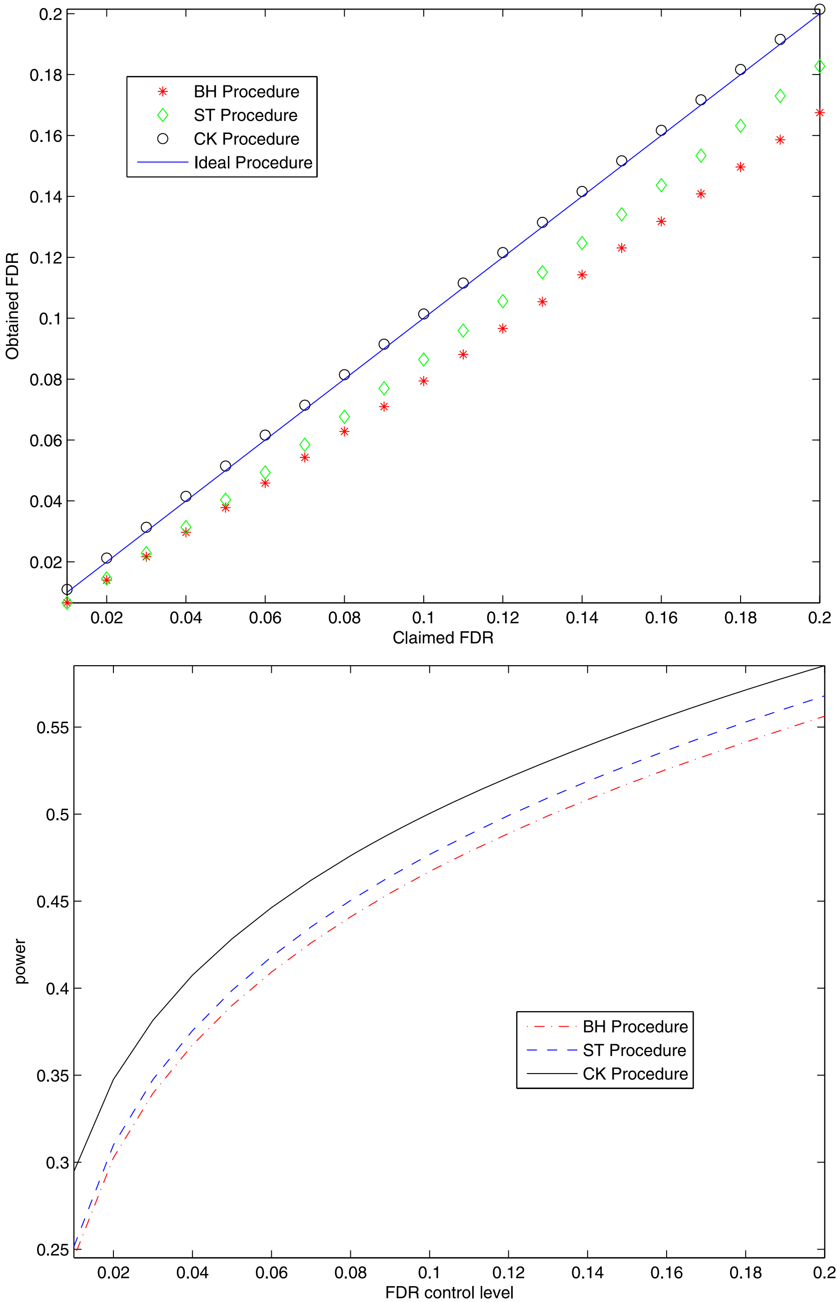

Figure 5. FDR control and power comparison.

http://www.broad.mit.edu/cgi-bin/cancer/datasets.cgi. In this analysis, we use the methodology developed for the dependence case. The raw data consist of $m = 7129$ genes and 72 samples coming from two classes: 47 in class ALL (acute lymphoblastic leukemia) and 25 in class AML (acute myeloid leukemia). Our simulation results showed reasonable performance of the procedure for a moderate sample size in this range. For each gene location, the two-sample t-statistic comparing the 47 ALL responses with the 25 AML responses was computed. Using our proposed approach for the dependent case, we find the critical value for controlling FDR at level γ,

$$\hat{t}_{n,m}^{fdr} = \inf\left\{t : \frac{2(1-\hat{\pi}_1)\bar{\Phi}(t)}{\hat{p}_m(t)} \le \gamma\right\},$$

where $\hat{p}_m = \sum_{i=1}^{m} 1_{\{|T_i|\ge t\}}/m$ and $\hat{\pi}_1$ is estimated by (2.28).

In Figure 6, we plot the FDR level and the number of significantly expressed genes by our (CK) procedure, BH procedure and the q-value based Storey–Tibshirani (ST) procedure. From the plot, we can see that our procedure detects the largest number of significant genes, followed by the ST procedure and then the BH procedure, which is the most conservative one. At FDR level 0.01, we detected 870 genes, the ST procedure detected 778 genes and the BH procedure detected 614 genes. Using the two-sample t-test, similarly to the higher power of our approach in simulation studies, we detected all of the genes that the other two approaches detected. The

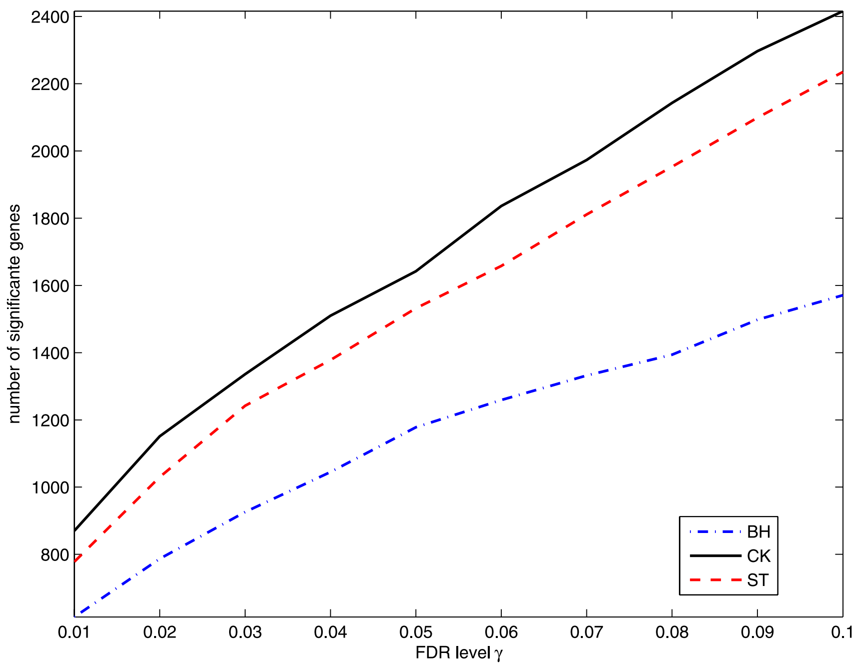

Figure 6. Comparison between our (CK) procedure, the ST procedure and the BH procedure using real data.

BH procedure is very conservative at the expense of power loss. The ST procedure requires permutation to obtain p-values, while our procedure gets the critical value directly and is thus faster in terms of computation. The estimation of π_1 is 0.467 by our procedure and 0.477 by the ST procedure. These results can serve as a first exploratory step for more refined analyses concerning these significant genes. Another issue may be that the critical value approach based on asymptotic FDR control may not be conservative enough in some settings.

6. Concluding remarks and discussion

We have presented a new approach for the significance analysis of thousands of features in high-dimensional biological studies. The approach is based on estimating the critical values of the rejection regions for high-dimensional multiple hypothesis testing, rather than the conventional p-value approaches in the literature. We developed a detailed method that can be used to identify differentially expressed genes in microarray experiments. The proposed procedure performs well for large samples, reasonably well for intermediate samples and not quite as well for small samples, and appears to perform better than existing alternatives under realistic sample sizes. Our method is also computationally faster than the competing approaches. The potential for improvement in small-sample performance motivates the need for a second-order expansion of our theoretical work. In addition, we have proposed a new consistent estimate of the proportion of alternative hypotheses under certain conditions. Numerical studies demonstrate that our methodology fits the truth well and improves the statistical power in multiple testing. Extensions of the current work can be pursued in several directions.

First, as stated above, the precision of the asymptotic approximations has room for improvement in small-to-moderately-small sample sizes, suggesting that a second-order expansion would be valuable. Second, in the dependence case, it would be of interest to see how the rate of convergence could be derived under various assumptions on the form of the dependence. Thirdly, the plug-in estimator π_1 is consistent, but somewhat ad hoc. Complete, theoretical properties of this estimator remain to be explored. Last, but not least, we only considered a fixed proportion π_1 of alternative hypotheses. It is of great interest also to consider the sparsity setting, in which $\pi_1 \to 0$ as $m \to \infty$, and to see what patterns emerge.

Appendix: Proofs of main results

Our main tools are limit theorems of empirical processes, Berry–Esseen bounds and self-normalized moderate deviations for one- and two-sample t-statistics.

A.1. Preliminary lemmas

We first state a non-uniform Berry–Esseen inequality for nonlinear statistics.

Lemma A.1 **([5]).** *Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent random variables with $E\xi_i = 0$, $\sum_{i=1}^n E\xi_i^2 = 1$ and $E|\xi_i|^3 < \infty$. Let $W_n = \sum_{i=1}^n \xi_i$ and $\Delta = \Delta(\xi_1, \ldots, \xi_n)$ be a measurable* *function of* $\{\xi_i\}$. *Then*

$$
\begin{aligned}
&|P(W_n + \Delta \le z) - \Phi(z)| \\
&\quad \le P\bigl(|\Delta| > (|z|+1)/3\bigr) \qquad\qquad (\text{A.1})\\
&\qquad + C(|z|+1)^{-3}\Biggl(\|\Delta\|_2 + \sum_{i=1}^{n}\bigl(E\xi_i^2\bigr)^{1/2}\bigl(E(\Delta-\Delta_i)^2\bigr)^{1/2} + \sum_{i=1}^{n} E|\xi_i|^3\Biggr).
\end{aligned}
$$

This is [5], Theorem 2.2, and the proof can be found there. The next lemma provides a Berry–Esseen bound for non-central t-statistics.

Lemma A.2. *Let* $X, X_1, \dots, X_n$ *be i.i.d. random variables with* $E(X) = 0$, $\sigma^2 = EX^2$ *and* $EX^4 < \infty$. *Let*

$$
\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.
$$

Then

$$
\left| P\left(\frac{\sqrt{n}(\bar{X}+c)}{s_n} \le x\right) - \Phi\bigl(x - \sqrt{n}c/\sigma\bigr)\right| \le K\frac{(1+|x|)}{(1+|x-\sqrt{n}c/\sigma|)\sqrt{n}} \qquad (\text{A.2})
$$

for any c *and* x, *where* K *is a finite constant that may depend on* σ *and* EX^4.

Proof. Without loss of generality, assume that $x \ge 0$ and $\sigma = 1$. Using the fact that

$$
1 - |t| \le (1+t)^{1/2} \le 1 + |t| \qquad \text{for } t \ge -1, \qquad (\text{A.3})
$$

we have

$$
xs_n = x(1 + s_n^2 - 1)^{1/2} \le x(1 + |s_n^2 - 1|) \qquad (\text{A.4})
$$

and

$$
xs_n \ge x(1 - |s_n^2 - 1|). \qquad (\text{A.5})
$$

Therefore,

$$
\begin{aligned}
P\left(\frac{\sqrt{n}(\bar{X}+c)}{s_n} \le x\right) &= P\bigl(\sqrt{n}(\bar{X}+c) \le xs_n\bigr) \\
&\le P\bigl(\sqrt{n}\bar{X} \le x - \sqrt{n}c + x|s_n^2 - 1|\bigr).
\end{aligned} \qquad (\text{A.6})
$$

We now apply (A.1) with $\xi_i = X_i/\sqrt{n}$, $W_n = \sqrt{n}\bar{X}$ and

$$
z = x - \sqrt{n}c, \qquad \Delta = -x|s_n^2 - 1|, \qquad \Delta_i = -x|s_{n,i}^2 - 1|,
$$

where $s_{n,i}^2$ is defined as s_n^2 with 0 replacing X_i.

Noting that

$$s_n^2 - 1 = \frac{1}{n-1}\left(\sum_{j=1}^n (X_j^2 - 1) - n\bar{X}^2\right) + \frac{1}{n-1},$$

$$s_{n,i}^2 - 1 = \frac{1}{n-1}\left(\sum_{j\neq i}(X_j^2 - 1) - n(\bar{X} - X_i/n)^2\right),$$

we have

$$E|s_n^2 - 1|^2 \leq KEX^4/n \tag{A.7}$$

and

$$\begin{aligned} E(s_n^2 - s_{n,i}^2)^2 &= \frac{1}{(n-1)^2} E\big((X_i^2 - 1) - n\bar{X}^2 + n(\bar{X} - X_i/n)^2 + 1\big)^2 \\ &= \frac{1}{(n-1)^2} E\big((X_i^2 - 1) - X_i\big(2(\bar{X} - X_i/n) + X_i/n\big) + 1\big)^2 \\ &\leq \frac{2}{(n-1)^2} E\big(2(X_i^2 - 1)^2 + 2 + X_i^2\big(2(\bar{X} - X_i/n) + X_i/n\big)^2\big) \qquad (A.8) \\ &\leq \frac{2}{(n-2)^2}\big(4EX^4 + 6 + EX_i^2\big(8(\bar{X} - X_i/n)^2 + 2EX_i^2/n\big)\big) \\ &\leq KEX^4/n^2. \end{aligned}$$

It follows from (A.7) and (A.8) that

$$\|\Delta\|_2 \leq K\frac{|x|\sqrt{EX^4}}{\sqrt{n}},$$

$$P\left(|\Delta| > \frac{|z|+1}{3}\right) \leq K\frac{|x|\sqrt{EX^4}}{\sqrt{n}(1+|z|)},$$

$$\sum_{i=1}^n (E\xi_i^2)^{1/2}\big(E(\Delta - \Delta_i)^2\big)^{1/2} \leq K\frac{|x|\sqrt{EX^4}}{\sqrt{n}}$$

and

$$\sum_{i=1}^n E|\xi_i|^3 \leq \frac{EX^3}{\sqrt{n}}.$$

Therefore, by (A.1),

$$\big|P\big(\sqrt{n}\bar{X} \leq x - \sqrt{n}c + x|s_n^2 - 1|\big) - \Phi\big(x - \sqrt{n}c\big)\big| \leq \frac{K(1+|x|)}{(1+|x-\sqrt{n}c|)\sqrt{n}}. \tag{A.9}$$

Similarly,

$$P\left(\frac{\sqrt{n}(\bar{X}+c)}{s_n} \le x\right) \ge P\big(\sqrt{n}\bar{X} \le x - \sqrt{n}c - x|s_n^2 - 1|\big)$$

and

$$\big|P\big(\sqrt{n}\bar{X} \le x - \sqrt{n}c - x|s_n^2 - 1|\big) - \Phi\big(x - \sqrt{n}c\big)\big| \le \frac{K(1+|x|)}{(1+|x-\sqrt{n}c|)\sqrt{n}}. \tag{A.10}$$

This proves (A.2). □

We also need a moderate deviation for the non-central t-statistics, as given in the following lemma.

Lemma A.3. *Suppose that $X, X_i, i = 1, \dots, n$, are independent identically distributed random variables. Let*

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}, \qquad s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2.$$

If X satisfies $E|X|^4 < \infty$, $E(X^2) = \sigma^2 > 0$ and $E(X) = 0$, then

$$P\left(\left|\frac{\sqrt{n}(\bar{X}+c)}{s_n}\right| \ge t\right) = P\big(|Z + c\sqrt{n}/\sigma| \ge t\big)\big(1 + \mathrm{o}(1)\big) \tag{A.11}$$

uniformly in c and $t = \mathrm{o}(n^{1/6})$. Here, and in the sequel, Z denotes a standard normal random variable.

Proof. When t is bounded, (A.11) follows from Lemma A.2. Consider large t with $t = \mathrm{o}(n^{1/6})$. We need the following result of [27,28]:

$$P\left(\frac{\sqrt{n}(\bar{X}+c)}{s_n} \ge t\right) = \big(1 - \Phi\big(t - c\sqrt{n}/\sigma\big)\big)\big(1 + \mathrm{o}(1)\big) \tag{A.12}$$

uniformly in $|c\sqrt{n}/\sigma| \le t/5$ and $t = \mathrm{o}(n^{1/6})$. We note that following the same lines as their proof, we can see that (A.12) remains valid for $-t/5 \le c\sqrt{n}/\sigma \le t$. We write

$$P\left(\left|\frac{\sqrt{n}(\bar{X}+c)}{s_n}\right| \ge t\right) = P\left(\frac{\sqrt{n}(\bar{X}+c)}{s_n} \ge t\right) + P\left(\frac{\sqrt{n}(-\bar{X}-c)}{s_n} \ge t\right).$$

By (A.12), the remark above and the fact that

$$1 - \Phi(t+x) = \mathrm{o}\big(1 - \Phi(t-x)\big)$$

for $x \ge 1$ (recall here that we assume t is large), (A.11) holds for $-t \le c\sqrt{n}/\sigma \le t$. Now, assume $|c|\sqrt{n}/\sigma > t$. Then, by (A.2),

$$\left| P\left(\left| \frac{\sqrt{n}(\bar{X}+c)}{s_n} \right| \ge t \right) - P\left(\left| Z + c\sqrt{n}/\sigma \right| \ge t \right) \right| = \mathrm{o}(1).$$

Since $|c|\sqrt{n}/\sigma > t$, we have $P(|Z + c\sqrt{n}/\sigma| \ge t) \ge 1/2$ and hence

$$P\left(\left| \frac{\sqrt{n}(\bar{X}+c)}{s_n} \right| \ge t \right) = P\left(\left| Z + c\sqrt{n}/\sigma \right| \ge t \right)\bigl(1 + \mathrm{o}(1)\bigr).$$

This completes the proof of (A.11). □

The lemma below shows that $t_{n,m}$ defined in (A.26) under independence is bounded.

Lemma A.4. *Assume that there exist $\varepsilon_0 > 0$ and $c_0 > 0$ such that*

$$P\bigl(\left| \sqrt{n}\mu_1/\sigma_1 \right| \ge \varepsilon_0 \bigr) \ge c_0. \tag{A.13}$$

Let $t_{n,m}$ satisfy (A.37). Then

$$t_{n,m} \le t_0, \tag{A.14}$$

where t_0 is the solution of

$$\alpha\pi_1 c_0 \exp\bigl((t_0 - \varepsilon_0)\varepsilon_0\bigr) = 12(1 + t_0 - \varepsilon_0). \tag{A.15}$$

Proof. It suffices to show that

$$\sqrt{m}E\xi_1(t_0) \ge (\mathrm{var}(\xi_1(t_0)))^{1/2} z_\gamma. \tag{A.16}$$

It is easy to see that $P(|Z + a| \ge t_0)$ is a monotone increasing function of $a > 0$. Hence,

$$\begin{aligned} P\bigl(\left| Z + \sqrt{n}\mu_1/\sigma_1 \right| \ge t_0 \bigr) &\ge P\bigl(\left| Z + \sqrt{n}\mu_1/\sigma_1 \right| \ge t_0, \left| \sqrt{n}\mu_1/\sigma_1 \right| \ge \varepsilon_0 \bigr) \\ &\ge P(|Z + \varepsilon_0| \ge t_0) P\bigl(\left| \sqrt{n}\mu_1/\sigma_1 \right| \ge \varepsilon_0 \bigr) \\ &\ge c_0 P(|Z + \varepsilon_0| \ge t_0) \ge c_0\bigl(1 - \Phi(t_0 - \varepsilon_0)\bigr) \\ &\ge \frac{c_0}{3(1 + t_0 - \varepsilon_0)} \exp\bigl(-(t_0 - \varepsilon_0)^2/2\bigr) \\ &\ge \frac{c_0}{3(1 + t_0 - \varepsilon_0)} \exp\bigl(-t_0^2/2 + (t_0 - \varepsilon_0)\varepsilon_0\bigr). \end{aligned} \tag{A.17}$$

Here, we use the fact that

$$\frac{1}{2}\mathrm{e}^{-x^2/2} \ge 1 - \Phi(x) \ge \frac{1}{\sqrt{2\pi}(1+x)}\mathrm{e}^{-x^2/2} \qquad \text{for } x \ge 0.$$

Under the null hypothesis $H_1 = 0$, which corresponds to $\mu_i = 0$, we apply Lemma A.3 and obtain

$$P(|T_1| \geq t | H_1 = 0) = P(|Z| \geq t)\bigl(1 + \mathrm{o}(1)\bigr) \tag{A.18}$$

uniformly in $t = \mathrm{o}(n^{1/6})$.

Under the alternative hypothesis $H_1 = 1$, we apply Lemma A.3 to $X_{ij} - \mu_i$ and obtain

$$\begin{aligned} P(|T_1| \geq t | H_1 = 1) &= P\bigl(\bigl|\sqrt{n}(\bar{X}_1 - \mu_1 + \mu_1)/s_1\bigr| \geq t | H_1 = 1\bigr) \\ &= E[P(|Z + \sqrt{n}\mu_1/\sigma_1)| \geq t | \mu_1, \sigma_1)]\bigl(1 + \mathrm{o}(1)\bigr) \\ &= P\bigl(\bigl|Z + \sqrt{n}\mu_1/\sigma_1\bigr| \geq t\bigr)\bigl(1 + \mathrm{o}(1)\bigr) \end{aligned} \tag{A.19}$$

uniformly in $t = \mathrm{o}(n^{1/6})$.

Also, note that

$$\begin{aligned} P(|T_1| \geq t) &= P(|T_1| \geq t, H_1 = 0) + P(|T_1| \geq t, H_1 = 1) \\ &= (1 - \pi_1)P(|T_1| \geq t | H_1 = 0) + \pi_1 P(|T_1| \geq t | H_1 = 1) \\ &= (1 - \pi_1)P(|Z| \geq t)\bigl(1 + \mathrm{o}(1)\bigr) \\ &\quad + \pi_1 P\bigl(\bigl|Z + \sqrt{n}\mu_1/\sigma_1\bigr| \geq t\bigr)\bigl(1 + \mathrm{o}(1)\bigr). \end{aligned} \tag{A.20}$$

By (A.34), (A.18), (A.20) and (A.17),

$$\begin{aligned} E\xi_1(t_0) &= \alpha(1 - \pi_1)P(|Z| \geq t_0)\bigl(1 + \mathrm{o}(1)\bigr) + \alpha\pi_1 P\bigl(\bigl|Z + \sqrt{n}\mu_1/\sigma_1\bigr| \geq t_0\bigr)\bigl(1 + \mathrm{o}(1)\bigr) \\ &\quad - (1 - \pi_1)P(|Z| \geq t_0)\bigl(1 + \mathrm{o}(1)\bigr) \\ &\geq \alpha\pi_1 \frac{c_0}{6(1 + t_0 - \varepsilon_0)} \exp\bigl(-t_0^2/2 + (t_0 - \varepsilon_0)\varepsilon_0\bigr) - 2P(Z \geq t_0) \\ &\geq \frac{\alpha\pi_1 c_0}{6(1 + t_0 - \varepsilon_0)} \exp\bigl(-t_0^2/2 + (t_0 - \varepsilon_0)\varepsilon_0\bigr) - \mathrm{e}^{-t_0^2/2} \\ &= \mathrm{e}^{-t_0^2/2}\left(\frac{\alpha\pi_1 c_0}{6(1 + t_0 - \varepsilon_0)} \exp\bigl((t_0 - \varepsilon_0)\varepsilon_0\bigr) - 1\right) \\ &= \mathrm{e}^{-t_0^2/2}, \end{aligned} \tag{A.21}$$

by (A.15) and the definition of t_0. It is easy to see that $E\xi_1^2 \leq 1$ and $\operatorname{var}(\xi_1(t_0)) \leq 1$ in particular. Thus, by (A.21),

$$\frac{\sqrt{m}E\xi_1(t_0)}{(\operatorname{var}(\xi_1(t)))^{1/2}} \geq \sqrt{m}\mathrm{e}^{-t_0^2/2} \geq z_\gamma, \tag{A.22}$$

provided that m is large enough. This proves (A.16). □

The following i.i.d. results are essential for the general results.

Lemma A.5. *Assume the conditions of Theorem 2.1 with (2.2) replaced by the assumption that $(T_i, H_i), i = 1, \ldots, m$ are i.i.d. and $\pi_1 = P(T_i = 1)$. Let $\mathcal{J} = \{i : H_i = 1\}$ be the set that contains the indices of alternative hypotheses. Also, assume that μ_i, σ_i are i.i.d. for $i \in \mathcal{J}$. Let*

$$p(t) = P(|T_1| \geq t), \tag{A.23}$$

$$a_1(t) = \alpha p(t) - (1 - \pi_1) F_0(t) \tag{A.24}$$

and

$$\begin{aligned} b_1^2(t) = {} & \alpha^2 p(t)\big(1 - p(t)\big) + 2\alpha(1 - \pi_1) p(t) F_0(t) \\ & + (1 - \pi_1) F_0(t)\big(1 - 2\alpha - (1 - \pi_1) F_0(t)\big). \end{aligned} \tag{A.25}$$

(i) *If $t_{n,m}^{fdtp}$ is chosen such that*

$$t_{n,m}^{fdtp} = \inf\big\{t : \sqrt{m} a_1(t)/b_1(t) \geq z_\gamma\big\}, \tag{A.26}$$

then

$$\lim_{m\to\infty} P(FDP \geq \alpha) = \lim_{m\to\infty} P(V \geq \alpha R) \leq \gamma \tag{A.27}$$

holds.

(ii) *If $t_{n,m}^{fdr}$ is chosen such that*

$$t_{n,m}^{fdr} = \inf\left\{t : \frac{(1 - \pi_1) F_0(t)}{p(t)} \leq \gamma\right\}, \tag{A.28}$$

then

$$\lim_{m\to\infty} FDR = \lim_{m\to\infty} E(V/R) \leq \gamma \tag{A.29}$$

holds.

(iii) *If $t_{n,m}^{k\text{-}FWER}$ is chosen such that*

$$t_{n,m}^{k\text{-}FWER} = \inf\big\{t : P\big(\eta(t) \geq k\big) \leq \gamma\big\}, \tag{A.30}$$

where $\eta(t) \sim \text{Poisson}(\theta(t))$ and

$$\theta(t) = m(1 - \pi_1) F_0(t),$$

then

$$\lim_{m\to\infty} k\text{-}FWER = \lim_{m\to\infty} P(V \geq k) \leq \gamma \tag{A.31}$$

holds.

Proof. We first prove the i.i.d. case for one-sample t-statistics. By (2.3),

$$
\begin{aligned}
\alpha R - V &= \alpha \sum_{i=1}^{m} I_{\{|T_i| \geq t\}} - \sum_{i=1}^{m} (1 - H_i) I_{\{|T_i| \geq t\}} \\
&= \sum_{i=1}^{m} (H_i + \alpha - 1) I_{\{|T_i| \geq t\}} \\
&= \sum_{i=1}^{m} \alpha I_{\{|T_i| \geq t\}} I_{\{H_i = 1\}} + \sum_{i=1}^{m} (\alpha - 1) I_{\{|T_i| \geq t\}} I_{\{H_i = 0\}} \\
&= \sum_{i=1}^{m} \alpha I_{\{|T_i| \geq t\}} \big(1 - I_{\{H_i = 0\}}\big) + \sum_{i=1}^{m} (\alpha - 1) I_{\{|T_i| \geq t\}} I_{\{H_i = 0\}} \\
&= \sum_{i=1}^{m} \big(\alpha I_{\{|T_i| \geq t\}} - I_{\{|T_i| \geq t\}} I_{\{H_i = 0\}}\big) \\
&= \sum_{i=1}^{m} \xi_i,
\end{aligned}
$$

where

$$
\xi_i := \xi_i(t) = \alpha I_{\{|T_i| \geq t\}} - I_{\{|T_i| \geq t\}} I_{\{H_i = 0\}}
$$

is obviously a Donsker class indexed by t [15]. Hence,

$$
P(V \geq \alpha R) = P\left(\sum_{i=1}^{m} \xi_i(t) \leq 0\right). \tag{A.32}
$$

Note that since ξ_i are independent random variables, we can apply the uniform central limit theorem to choose t so that

$$
P\left(\sum_{i=1}^{m} \xi_i(t) \leq 0\right) \leq \gamma. \tag{A.33}
$$

To this end, we need the mean and variance of ξ_i. Without loss of generality, we use ξ_1 as an example, since ξ_i are i.i.d. random variables. Thus,

$$
\begin{aligned}
E\xi_1 &= \alpha P(|T_1| \geq t) - P(|T_1| \geq t, H_1 = 0) \\
&= \alpha P(|T_1| \geq t) - P(H_1 = 0) P(|T_1| \geq t | H_1 = 0) \\
&= \alpha P(|T_1| \geq t) - (1 - \pi_1) P(|T_1| \geq t | H_1 = 0).
\end{aligned} \tag{A.34}
$$

Similarly,

$$
\begin{aligned}
E\xi_1^2 &= E\big(\alpha^2 I_{\{|T_1|\ge t\}} + (1-2\alpha) I_{\{|T_1|\ge t\}} I_{\{H_1=0\}}\big) \\
&= \alpha^2 P(|T_1|\ge t) + (1-2\alpha)(1-\pi_1) P(|T_1|\ge t|H_1=0)
\end{aligned} \tag{A.35}
$$

and

$$
\begin{aligned}
\operatorname{var}(\xi_1) &= E\xi_1^2 - (E\xi_1)^2 \\
&= \alpha^2 P(|T_1|\ge t) + (1-2\alpha)(1-\pi_1) P(|T_1|\ge t|H_1=0) \\
&\quad - \{\alpha P(|T_1|\ge t) - (1-\pi_1) P(|T_1|\ge t|H_1=0)\}^2 \\
&= \alpha^2 P(|T_1|\ge t)\big(1 - P(|T_1|\ge t)\big) \\
&\quad + (1-\pi_1) P(|T_1|\ge t|H_1=0)\big(1 - 2\alpha - (1-\pi_1) P(|T_1|\ge t|H_1=0)\big) \\
&\quad + 2\alpha(1-\pi_1) P(|T_1|\ge t) P(|T_1|\ge t|H_1=0).
\end{aligned} \tag{A.36}
$$

Now, define

$$
t_{n,m} = \inf\left\{t : \frac{\sqrt{m} E\xi_1(t)}{(\operatorname{var}(\xi_1(t)))^{1/2}} \ge z_\gamma\right\}. \tag{A.37}
$$

By Lemma A.4, $t_{n,m}$ is bounded and hence the uniform central limit theorem yields

$$
\begin{aligned}
P\left(\sum_{i=1}^m \xi_i(t_{n,m}) \le 0\right) &= P\bigg(\frac{\sum_{i=1}^m (\xi_i(t_{n,m}) - E\xi_i(t_{n,m}))}{(\sum_{i=1}^m \operatorname{var}(\xi_i(t_{n,m})))^{1/2}} \\
&\qquad \le -\frac{\sum_{i=1}^m E\xi_i(t_{n,m})}{(\sum_{i=1}^m \operatorname{var}(\xi_i(t_{n,m})))^{1/2}}\bigg) \\
&\le P\bigg(\frac{\sum_{i=1}^m (\xi_i(t_{n,m}) - E\xi_i(t_{n,m}))}{(\sum_{i=1}^m \operatorname{var}(\xi_i(t_{n,m})))^{1/2}} \le -z_\gamma\bigg) \\
&\to \Phi(-z_\gamma) = \gamma.
\end{aligned} \tag{A.38}
$$

This proves (A.27).

Note that

$$
\begin{aligned}
FDR &= \int_0^1 P(FDTP \ge x)\,\mathrm{d}x \\
&= \int_0^1 P(V \ge xR)\,\mathrm{d}x \\
&= \int_0^1 P\left(\sum_1^m \xi_i \le 0\right)\mathrm{d}x \\
&= \int_0^1 P\left(N(0,1) \le \frac{-\sqrt{m} E\xi_1}{\sqrt{\operatorname{Var}\xi_1}}\right)\mathrm{d}x.
\end{aligned}
$$

Letting $m \to +\infty$, $P(N(0,1) \le -\sqrt{m}E\xi_1/\sqrt{\operatorname{Var}\xi_1})$ is either 0 or 1, depending on the sign of $E\xi_1$. Thus, the range of x that makes this probability 1 satisfies

$$E\xi_1 = xP(|T_1| \ge t) - (1-\pi_1)P(|T_1| \ge t | H_1 = 0) < 0$$

and the corresponding $x < (1-\pi_1)P(|T_1| \ge t|H_1 = 0)/P(|T_1| \ge t)$. In order to control FDR at level γ, we require

$$\frac{(1-\pi_1)P(|T_1| \ge t|H_1 = 0)}{P(|T_1| \ge t)} \le \gamma.$$

This proves (A.28).

For the k-FWER, we use the characteristic function method. Let $\eta_i = (1 - H_i)I_{\{|T_i| \ge t\}}$,

$$\begin{aligned} E\mathrm{e}^{\mathrm{i}s\sum_{i=1}^m \eta_i} &= \prod_{i=1}^m E\mathrm{e}^{\mathrm{i}s\eta_i} \\ &= \prod_{i=1}^m [\mathrm{e}^{\mathrm{i}s}(1-\pi_1)F_0 + 1 - (1-\pi_1)F_0] \\ &= \left[1 + \frac{1}{m}m(1-\pi_1)F_0(\mathrm{e}^{\mathrm{i}s} - 1)\right]^m \\ &\to \mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}s}-1)}, \end{aligned}$$

where $m_0F_0 \to \lambda$ as $m \to \infty$ and λ is the parameter for the Poisson distribution such that

$$P\big(\operatorname{Poiss}(\lambda) \ge k\big) \le \gamma. \qquad \square$$

The following functional central limit theorem is needed in the proof of Theorem 2.1:

Lemma A.6. *Suppose the triangular array* $\{f_{ni}(\omega, t), i = 1, \ldots, m_n, t \in T\}$ *consists of independent processes within rows and is almost measurable Suslin analytic set* (*AMS*)(*see page* 25 *in* [15]). *Let*

$$X_n(\omega, t) \equiv \sum_{i=1}^{m_n}[f_{ni}(\omega, t) - Ef_{ni}(\cdot, t)]. \tag{A.39}$$

Assume:

(A) *the* $\{f_{ni}\}$ *are manageable, with envelopes* $\{F_{ni}\}$ *which are also independent within rows*;
(B) $H(s,t) = \lim_{n\to\infty} EX_n(s)X_n(t)$ *exists for every* $s, t \in T$;
(C) $\limsup_{n\to\infty} \sum_{i=1}^{m_n} E^*F_{ni}^2 < \infty$;
(D) $\lim_{n\to\infty} \sum_{i=1}^{m_n} E^*F_{ni}^2 1\{F_{ni} > \epsilon\} = 0$ *for each* $\epsilon > 0$;
(E) $\rho(s,t) = \lim_{n\to\infty} \rho_n(s,t)$, *where*

$$\rho_n(s,t) \equiv \left(\sum_{i=1}^{m_n} E|f_{ni}(\cdot, s) - f_{ni}(\cdot, t)|^2\right)^{1/2}$$

exists for every $s, t \in T$ and, for all deterministic sequences $\{s_n\}$ and $\{t_n\}$ in T, if $\rho(s_n, t_n) \to 0$, then $\rho_n(s_n, t_n) \to 0$.

Then X_n converges weakly on $l^\infty(T)$ to a tight mean-zero Gaussian process X concentrated on $UC(T, \rho)$, with covariance $H(s, t)$.

Proof. The definitions involved in this lemma and the proof can be found in [15], Theorem 11.16. Below, we verify that, conditional on $\mathcal{H}$, $f_{ni}(\omega, t) = \xi_i(\omega, t)/\sqrt{m}$ satisfy the conditions in Lemma A.6. Since $\xi_i(\omega, t)$ is the difference between two monotone bounded functions, it is clear that, conditional on $\mathcal{H}$, $\xi_i(\omega, t)/\sqrt{m}$ is *AMS*, manageable and has envelopes $\alpha/\sqrt{m}$. Also,

$$\begin{aligned}
EX_n(s)X_n(t) &= EE[X_n(s)X_n(t)|\mathcal{H}] \\
&= EE\left[\frac{\sum_{i=1}^m (\xi_i(s)|\mathcal{H} - E\xi_i(s)|\mathcal{H})}{\sqrt{m}} \frac{\sum_{j=1}^m (\xi_j(t)|\mathcal{H} - E\xi_j(t)|\mathcal{H})}{\sqrt{m}}\right] \\
&= EE\frac{\sum_{i=1}^m (\xi_i(s)|\mathcal{H} - E\xi_i(s)\mathcal{H})(\xi_i(t)|\mathcal{H} - E\xi_i(t)\mathcal{H})}{m} \\
&= \frac{1}{m}E\sum_{i=1}^m E(\xi_i(s)|\mathcal{H})(\xi_i(t)|\mathcal{H}) - \sum_{i=1}^m E(\xi_i(s)|\mathcal{H})E(\xi_i(t)|\mathcal{H}) \\
&= \frac{1}{m}E\sum_{i=1}^m \bigl(\alpha^2 H_i + (1-\alpha)^2(1-H_i)\bigr)EI_{\{|T_i| \ge t \cup s|\mathcal{H}\}} \\
&\quad - \sum_{i=1}^m [\alpha H_i + (1-\alpha)(1-H_i)]^2 EI_{\{|T_i| \ge s\mathcal{H}\}} EI_{\{|T_i| \ge t|\mathcal{H}\}} \\
&= \frac{1}{m}E\sum_{i=1}^m \bigl(\alpha^2 H_i F_1(t \cup s) + (1-\alpha)^2(1-H_i)F_0(t \cup s)\bigr) \\
&\quad - \sum_{i=1}^m [\alpha^2 H_i + (1-\alpha)^2(1-H_i)][H_i F_1(s) + (1-H_i)F_0(s)] \\
&\qquad \times [H_i F_1(t) + (1-H_i)F_0(t)] \\
&= \frac{1}{m}E\sum_{i=1}^m \bigl[\alpha^2 H_i\bigl(F_1(t \cup s) - F_1(t)F_1(s)\bigr) \\
&\qquad + (1-\alpha)^2(1-H_i)\bigl(F_0(t \cup s) - F_0(t)F_0(s)\bigr)\bigr] \\
&\to \pi_1\alpha^2\bigl(F_1(t \cup s) - F_1(t)F_1(s)\bigr) + (1-\pi_1)(1-\alpha)^2\bigl(F_0(t \cup s) - F_0(t)F_0(s)\bigr) \\
&\equiv H(s, t),
\end{aligned}$$

which is the same as $q^2(t)$ when $s = t$. (C) is easily satisfied. For all $\epsilon > 0$, there exists an N_0 such that $\alpha/N_0 < \epsilon$, so $\lim_{m\to\infty}\sum_{i=1}^{m} E\alpha^2/m 1\{\alpha/\sqrt{m} > \epsilon\} = \lim_{m\to\infty}\sum_{i=1}^{N_0-1}\alpha^2/m = 0$, which verifies (D). Similarly, we can show that (E) is satisfied and thus the functional central limit theorem holds. □

Let

$$\begin{aligned}G(t) &= \alpha\pi_1 E P\bigl(\bigl|Z + \sqrt{n}\mu_1/\sigma_1\bigr| \ge t\bigr) - (1-\alpha)(1-\pi_1)P(|Z| \ge t)\\ &= \alpha\pi_1 E P\bigl(\bigl|Z + \sqrt{n}|\mu_1|/\sigma_1\bigr| \ge t\bigr) - (1-\alpha)(1-\pi_1)P(|Z| \ge t)\end{aligned}$$

and

$$t_1 = \inf\{t : G(t) = 0\}. \tag{A.40}$$

The following lemma is needed in the proof of consistency.

Lemma A.7. *Assume that* $0 < \pi_1 < 1-\alpha$ *and* (A.13) *is satisfied. Then*

$$G(t)\begin{cases} < 0 & \text{for } t < t_1,\\ = 0 & \text{for } t = t_1,\\ > 0 & \text{for } t > t_1.\end{cases} \tag{A.41}$$

Moreover, $G'(t_1) \ge \mathrm{e}^{-t_0^2/2}/\sqrt{2\pi}$.

Proof. We first observe that $0 < t_1 \le t_0$ by the fact that $G(0) < 0$, $G(t_0) > \mathrm{e}^{-t_0^2/2} > 0$ in (A.21) and $G(t)$ is a continuous function.

To prove (A.41), it suffices to show that there exists a $t_2 > t_1$ such that $G(t)$ is increasing in $[0, t_2]$ and decreasing in $[t_2, \infty)$. To this end, consider the derivative of G:

$$\begin{aligned}G'(t) &= -\alpha\pi_1 E\bigl(\phi\bigl(t - \sqrt{n}|\mu_1|/\sigma_1\bigr) + \phi\bigl(t + \sqrt{n}|\mu_1|/\sigma_1\bigr)\bigr) + 2(1-\alpha)(1-\pi_1)\phi(t)\\ &= \frac{\mathrm{e}^{-t^2/2}}{\sqrt{2\pi}}\biggl\{-\alpha\pi_1 E\biggl(\exp\biggl(-\frac{n\mu_1^2}{2\sigma_1^2} + \frac{\sqrt{n}|\mu_1|t}{\sigma_1}\biggr) + \exp\biggl(-\frac{n\mu_1^2}{2\sigma_1^2} - \frac{\sqrt{n}|\mu_1|t}{\sigma_1}\biggr)\biggr)\\ &\qquad + 2(1-\alpha)(1-\pi_1)\biggr\}.\end{aligned} \tag{A.42}$$

Let

$$\begin{aligned}H(t) = -\alpha\pi_1 E\biggl(&\exp\biggl(-\frac{n\mu_1^2}{2\sigma_1^2} + \frac{\sqrt{n}|\mu_1|t}{\sigma_1}\biggr)\\ &+ \exp\biggl(-\frac{n\mu_1^2}{2\sigma_1^2} - \frac{\sqrt{n}|\mu_1|t}{\sigma_1}\biggr)\biggr) + 2(1-\alpha)(1-\pi_1).\end{aligned}$$

Then

$$
\begin{aligned}
H'(t) &= -\alpha\pi_1 E\left\{\frac{\sqrt{n}|\mu_1|}{\sigma_1}\exp\left(\frac{\sqrt{n}|\mu_1|t}{\sigma_1}-\frac{n\mu_1^2}{2\sigma_1^2}\right)-\frac{\sqrt{n}|\mu_1|}{\sigma_1}\exp\left(-\frac{\sqrt{n}|\mu_1|t}{\sigma_1}-\frac{n\mu_1^2}{2\sigma_1^2}\right)\right\}\\
&= -\alpha\pi_1 E\left\{\frac{\sqrt{n}|\mu_1|}{\sigma_1}\mathrm{e}^{-n\mu_1^2/(2\sigma_1^2)}\left(\exp\left(\frac{\sqrt{n}|\mu_1|t}{\sigma_1}\right)-\exp\left(-\frac{\sqrt{n}|\mu_1|t}{\sigma_1}\right)\right)\right\}<0
\end{aligned}
\tag{A.43}
$$

for all $t > 0$. Therefore, $H(t)$ is monotone decreasing. Taking into account the facts that $H(0) > 0$ by assumption, $\pi_1 < 1 - \alpha$ and $H(+\infty) < 0$, we conclude that $H(t)$ has only one zero point, say, t_2. Moreover, $H(t) > 0$ for $t < t_2$ and $H(t) < 0$ for $t > t_2$. This is also true for $G'(t)$, by (A.42). Hence, $G(t)$ is increasing for $t < t_2$ and decreasing for $t > t_2$. Note that since $G(0) < 0$, $G(t_0) > 0$ and $G(+\infty) = 0$, we can see that $G(t)$ has a unique zero point t_1 and $t_2 > t_1$. Since $G(t)$ is increasing for $0 < t < t_2$, we have $G'(t_1) > 0$. We now prove that $G'(t_1) \geq \mathrm{e}^{-t_0^2/2}/\sqrt{2\pi}$. It follows from the proof of (A.21) that

$$G(t_0) \geq \mathrm{e}^{-t_0^2/2}. \tag{A.44}$$

Recalling that $G'(t) = \frac{\mathrm{e}^{-t^2/2}}{\sqrt{2\pi}}H(t)$ and H is decreasing, we have

$$
\begin{aligned}
G(t_0) = G(t_0) - G(t_1) &= \int_{t_1}^{t_0} G'(s)\,\mathrm{d}s \leq \int_{t_1}^{t_0}\frac{\mathrm{e}^{-s^2/2}}{\sqrt{2\pi}}H(t_1)\,\mathrm{d}s\\
&\leq H(t_1)\bigl(1-\Phi(t_1)\bigr) \leq H(t_1)\mathrm{e}^{-t_1^2/2} = G'(t_1)\sqrt{2\pi}.
\end{aligned}
\tag{A.45}
$$

This proves $G'(t_1) \geq \mathrm{e}^{-t_0^2/2}/\sqrt{2\pi}$. □

A.2. Proof of Theorem 2.1

We now return to show our main theorem under dependence. Let $\mathcal{H} = \{H_i, 1 \leq i \leq m\}$. To prove (i), following along the same lines as the proof of Lemma A.5, we need to obtain the asymptotic distribution of

$$P(V \geq \alpha R) = P\left(\sum_{i=1}^m \xi_i(t) \leq 0\right), \tag{A.46}$$

where

$$\xi_i(t) = \alpha I_{\{|T_i|\geq t\}} - I_{\{|T_i|\geq t\}}I_{\{H_i=0\}} = (\alpha + H_i - 1)I_{\{|T_i|\geq t\}} = [\alpha H_i - (1-\alpha)(1-H_i)]I_{\{|T_i|\geq t\}}.$$

Note that

$$P(|T_i| \geq t|\mathcal{H}) = (1-H_i)P(|T_i| \geq t|H_i = 0) + H_i P(|T_i| \geq t|H_i = 1).$$

Given $\mathcal{H}$, $\xi_i(t)$, $1 \le i \le m$, are independent random variables. The conditional mean equals

$$
\begin{aligned}
E\left(\sum_{i=1}^m \xi_i | \mathcal{H}\right) &= \sum_{i=1}^m \bigl\{\alpha E\bigl(I_{\{H_i=0\}}|\mathcal{H}\bigr) P(|T_i| \ge t | H_i = 0) + \alpha E\bigl(I_{\{H_i=1\}}|\mathcal{H}\bigr) P(|T_i| \ge t | H_i = 1) \\
&\qquad - E\bigl(I_{\{H_i=0\}}|\mathcal{H}\bigr) P(|T_i| \ge t | H_i = 0)\bigr\} \\
&= \sum_{i=1}^m \{\alpha(1 - H_i) P(|T_i| \ge t | H_i = 0) + \alpha H_i P(|T_i| \ge t | H_i = 1) \\
&\qquad - (1 - H_i) P(|T_i| \ge t | H_i = 0)\} \\
&= \alpha \sum_{i=1}^m \{H_i P(|T_i| \ge t | H_i = 1)\} - (1 - \alpha) \sum_{i=1}^m \{(1 - H_i) P(|T_i| \ge t | H_1 = 0)\} \\
&= \alpha m_1 F_1(t) - (1 - \alpha) m_0 F_0(t).
\end{aligned}
$$

Next, we calculate the conditional variance of $\sum_{i=1}^m \xi_i(t)$, given $\mathcal{H}$:

$$
\begin{aligned}
\operatorname{var}\left(\sum_{i=1}^m \xi_i(t) | \mathcal{H}\right) &= \operatorname{var}\left(\sum_{i=1}^m [\alpha H_i - (1 - \alpha)(1 - H_i)] I_{\{|T_i| \ge t | \mathcal{H}\}}\right) \\
&= \sum_{i=1}^m \bigl(\alpha^2 H_i + (1 - \alpha)^2 (1 - H_i)\bigr) \operatorname{var}\bigl(I_{\{|T_i| \ge t | \mathcal{H}\}}\bigr) \\
&= \alpha^2 m_1 F_1(t)\bigl(1 - F_1(t)\bigr) + (1 - \alpha)^2 m_0 F_0(t)\bigl(1 - F_0(t)\bigr).
\end{aligned}
$$

From (2.7) and (2.8),

$$
\frac{\mu_m(t)}{\sigma_m(t)} = \sqrt{m} \frac{\mu_m(t)/m}{\sqrt{\sigma_m^2(t)/m}}.
$$

By the fact that $m_1/m \to \pi_1$ a.s., we have

$$
\mu_m(t)/m \to \alpha \pi_1 F_1(t) - (1 - \alpha)(1 - \pi_1) F_0(t) \qquad \text{a.s.} \tag{A.47}
$$

and

$$
\begin{aligned}
\sigma_m^2(t)/m &\to \alpha^2 \pi_1 F_1(t)\bigl(1 - F_1(t)\bigr) \\
&\qquad + (1 - \alpha)^2 (1 - \pi_1) F_0(t)\bigl(1 - F_0(t)\bigr) = q^2(t) \qquad \text{a.s.,}
\end{aligned} \tag{A.48}
$$

which is smaller than $\operatorname{var}(\xi_1(t))$, due to the fact that

$$
\operatorname{var} X = E(\operatorname{var}(X|Y)) + \operatorname{var}(E(X|Y))
$$

for any two random variables X and Y. By (A.16), we can see that the critical value defined at (2.9) is bounded. Thus, conditional on $\mathcal{H}$, we can use the functional central limit theorem on $\sum_{i=1}^m \xi_i(t)/\sqrt{m}$, by virtue of Lemma A.6. The limit is a Gaussian process with continuous sample paths. Hence,

$$\begin{aligned}
P\left(\sum_{i=1}^m \xi_i(t) \le 0\right) &= E\left(E 1_{\{\sum_{i=1}^m \xi_i(t)/\sqrt{m} \le 0\}} \middle| \mathcal{H}\right) \\
&= E\left\{P\left(\sum_{i=1}^m \xi_i/\sqrt{m} - \sum_{i=1}^m E(\xi_i|\mathcal{H})/\sqrt{m} \le \frac{-\sum_{i=1}^m E(\xi_i|\mathcal{H})\sigma_m(t)}{\sqrt{m}\sigma_m(t)} \middle| \mathcal{H}\right)\right\} \\
&\le E\left\{P\left(\sum_{i=1}^m \xi_i/\sqrt{m} - \sum_{i=1}^m E(\xi_i|\mathcal{H})/\sqrt{m} \le \frac{-\sum_{i=1}^m E(\xi_i|\mathcal{H})}{\sigma_m(t)} \frac{\sigma_m(t)}{\sqrt{m}} \middle| \mathcal{H}\right)\right\} \\
&\le E\left\{P\left(N(0,1)q(t) \le -z_\gamma q(t)\right)\right\} \\
&\to P\left(N(0,1) \le -z_\gamma\right) = \gamma \qquad \text{as } m \to \infty.
\end{aligned}$$

This proves (2.9).

(ii) can be proven similarly. The characteristic function method can be used to prove (iii).

A.3. Proof of Theorem 2.2

We first prove (i), and (ii) follows along the same lines as the independent case, plus a conditional argument. Without loss of generality, we use T_1 as a representative that comes from the alternative. We have to show that

$$|\hat{t}_{n,m} - t_{n,m}| = \mathrm{o}(1) \qquad \text{a.s.} \tag{A.49}$$

We first prove that

$$|\hat{t}_{n,m} - t_1| = \mathrm{o}(1) \qquad \text{a.s.}, \tag{A.50}$$

where t_1 is defined as in (A.40). It suffices to show that for any $\varepsilon > 0$,

$$\frac{\sqrt{m}\nu_m(t_1+\varepsilon)}{\tau_m(t_1+\varepsilon)} \ge z_\gamma \tag{A.51}$$

and

$$\frac{\sqrt{m}\nu_m(s)}{\tau_m(s)} < z_\gamma \qquad \text{for all } s \le t_1 - \varepsilon. \tag{A.52}$$

Recall that $\hat{p}_m(t) = \frac{1}{m}\sum_{i=}^m I_{\{|T_i| \ge t\}}$. Given $\mathcal{H}$, by the uniform law of the iterated logarithm (see, e.g., [10]),

$$\hat{p}_m(t) - \frac{1}{m}\sum_{i=1}^m \{(1-H_i)F_0(t) + H_i F_1(t)\} = \mathrm{o}(m^{-1/2}(\log\log m)^{1/2}) \qquad \text{a.s.}$$

By the strong law of large number,

$$\frac{1}{m}\sum_{i=1}^{m}\{(1-H_i)F_0(t)+H_iF_1(t)\}\to(1-\pi_1)F_0(t)+\pi_1F_1(t)\qquad\text{a.s.}\tag{A.53}$$

So

$$\hat{p}_m(t)\to(1-\pi_1)F_0(t)+\pi_1F_1(t)\qquad\text{a.s.}$$

Recall that

$$\nu_m(t)=\alpha\hat{p}_m(t)-2(1-\hat{\pi}_1)\bar{\Phi}(t).$$

By (A.2), our strong consistent estimate $\hat{\pi}_1$ described in Section 2.3 and the continuous mapping theorem, we have

$$\sup_t\Big|\nu_m(t)-\Big\{\alpha\big((1-\pi_1)F_0(t)+\alpha\pi_1F_1(t)\big)-(1-\pi_1)P(|Z|\ge t)\Big\}\Big|\to 0\qquad\text{a.s.,}\tag{A.54}$$

which, together with (A.20) and the definition of G, implies that

$$\sup_{0\le t\le 1+t_0}|\nu_m(t)-G(t)|\to 0\qquad\text{a.s.}\tag{A.55}$$

In particular, since $G(t_1+\varepsilon)>0$ for $0<\varepsilon<t_2-t_1$, we have

$$\nu_m(t_1+\varepsilon)\ge G(t_1+\varepsilon)/2\qquad\text{a.s.}\tag{A.56}$$

for sufficiently large m and, therefore, $\sqrt{m}\nu_m(t_1+\epsilon)\ge z_\gamma\tau_m(t_1+\epsilon)$. This proves (A.51).

Similarly, since $G(t)$ is increasing and $G(t_1-\varepsilon)<0$, we have

$$\max_{s\le t_1-\varepsilon}\nu_m(s)\le G(t_1-\varepsilon)/2\qquad\text{a.s.}\tag{A.57}$$

for sufficiently large m. Hence, (A.52) holds. This proves (A.50).

Following the same lines as the proof of (A.50), we have

$$|t_{n,m}-t_1|=\mathrm{o}(1).\tag{A.58}$$

This completes the proof of (A.49).

For k-FWER, let η_0 be the number that satisfies $P(\mathrm{Poiss}(\eta_0)\ge k)\le\gamma$. Let $t_{0,m}=t_{n,m}^{k\text{-}FWER}$ and $t_m=\hat{t}_{m,n}^{k\text{-}FWER}$. Thus, by definition, $t_{0,m}$ is the t that satisfies $(1-\pi_1)mF_{\mathrm{o}}(t)=\eta_0$ and t_m is the t that satisfies $2(1-\hat{\pi}_1)m\bar{\Phi}(t)=\eta_0$. We then have $\frac{(1-\pi_1)F_0(t_{0,m})}{(1-\hat{\pi}_1)2\bar{\Phi}(t_m)}=1$, which implies that

$$\begin{aligned}&\frac{F_0(t_{0,m})}{2\bar{\Phi}(t_m)}=\frac{1-\hat{\pi}_1}{1-\pi_1}=1+\mathrm{o}_P(1)\\ \Longrightarrow\quad&\frac{\bar{\Phi}(t_{0,m})}{\bar{\Phi}(t_m)}\big(1+\mathrm{O}(n^{-1/2})\big)=1+\mathrm{o}_P(1)\end{aligned}$$

$$\begin{aligned}
&\Longrightarrow \quad \frac{\bar{\Phi}(t_{0,m})}{\bar{\Phi}(t_m)} = 1 + \mathrm{o}_P(1) \\
&\Longrightarrow \quad \frac{t_m}{t_{0,m}} \mathrm{e}^{-t_{0,m}^2/2 + t_m^2/2} = 1 + \mathrm{o}_P(1) \\
&\Longrightarrow \quad R\mathrm{e}^{-t_{0,m}^2/2 + R^2 t_{0,m}^2/2} = R\mathrm{e}^{-(1-R^2)t_{0,m}^2/2} = 1 + \mathrm{o}_P(1).
\end{aligned}$$

Hence, $R = t_m/t_{0,m} \to 1$ in probability. Thus,

$$t_{0,m}^2 - t_m^2 = \mathrm{o}_P(1) \quad \Longrightarrow \quad |t_{0,m} - t_m| = \frac{\mathrm{o}_P(1)}{1 + |t_{0,m} + t_m|} = \mathrm{O}_p((\log m)^{-1/2})$$

since $t_m = \mathrm{o}_P(n^{1/6})$ and $\log m = \mathrm{o}(n^{1/3})$.

A.4. Proof of Theorem 2.4

In this section, we give the proof of the rate of convergence for the i.i.d. case by using the one-sample t-statistic. Let $p(t) = P(|T_1| \geq t)$ and let

$$\hat{p}_m(t) = \frac{1}{m} \sum_{i=1}^{m} I_{\{|T_i| \geq t\}}.$$

By the Glivenko–Cantelli theorem,

$$\sup_t |\hat{p}_m(t) - p(t)| \to 0 \qquad \text{a.s.} \tag{A.59}$$

and, by the Donsker theorem,

$$\sup_t |\hat{p}_m(t) - p(t)| = \mathrm{O}(m^{-1/2}) \qquad \text{in probability.} \tag{A.60}$$

By the uniform law of the iterated logarithm,

$$\sup_t |\hat{p}_m(t) - p(t)| = \mathrm{O}(m^{-1/2}(\log\log m)^{1/2}) \qquad \text{a.s.} \tag{A.61}$$

We define strong consistent estimators of $E\xi_1(t)$ and $\mathrm{var}(\xi_1(t))$ by $\nu_m(t)$ and $\tau_m^2(t)$, respectively, where

$$\nu_m(t) = \alpha \hat{p}_m(t) - (1 - \pi_1) P(|Z| \geq t) \tag{A.62}$$

and

$$\begin{aligned}
\tau_m^2(t) = {} & \alpha^2 \hat{p}_m(t)\bigl(1 - \hat{p}_m(t)\bigr) + 2\alpha(1 - \pi_1)\hat{p}_m(t) P(|Z| \geq t) \\
& + (1 - \pi_1) P(|Z| \geq t)\bigl(1 - 2\alpha - (1 - \pi_1) P(|Z| \geq t)\bigr).
\end{aligned} \tag{A.63}$$

We now define an estimator of $t_{n,m}$ by

$$\hat{t}_{n,m} = \inf\left\{t : \frac{\sqrt{m}\,\upsilon_m(t)}{\tau_m(t)} \geq z_\gamma\right\}. \tag{A.64}$$

For FDTP, we have to show that

$$|\hat{t}_{n,m} - t_{n,m}| = \mathrm{O}\left(\frac{1}{\sqrt{n}} + \left(\frac{\log\log m}{m}\right)^{1/2}\right) \qquad \text{a.s.} \tag{A.65}$$

and

$$|\hat{t}_{n,m} - t_{n,m}| = \mathrm{O}(n^{-1/2} + m^{-1/2}) \qquad \text{in probability.} \tag{A.66}$$

Below, we prove (A.65) and (A.66). We will show that

$$|\hat{t}_{n,m} - t_1| = \mathrm{O}\left(\left(\frac{1}{n}\right)^{1/2} + \left(\frac{\log\log m}{m}\right)^{1/2}\right) \qquad \text{a.s.,} \tag{A.67}$$

$$|t_{n,m} - t_1| = O\left(\left(\frac{1}{n}\right)^{1/2} + \left(\frac{\log\log m}{m}\right)^{1/2}\right) \qquad \text{a.s.} \tag{A.68}$$

By the uniform law of the iterated logarithm,

$$\sup_t |\hat{p}_m(t) - p(t)| = \mathrm{O}\left(\left(\frac{\log\log m}{m}\right)^{1/2}\right) \qquad \text{a.s.} \tag{A.69}$$

Therefore, we have

$$\sup_t \left|\upsilon_m(t) - [\alpha p(t) - (1-\pi_1)P(|Z| \geq t)]\right| = \mathrm{O}\left(\left(\frac{\log\log m}{m}\right)^{1/2}\right) \qquad \text{a.s.} \tag{A.70}$$

Note that

$$\begin{aligned}
&\alpha p(t) - (1-\pi_1)P(|Z| \geq t) - G(t) \\
&\quad = \alpha(1-\pi_1)\big(P(|T_1| \geq t | H_1 = 0) - P(|Z| \geq t)\big) \\
&\qquad + \alpha\pi_1\big(P(|T_1| \geq t | H_1 = 1) - EP\big(\big|Z + \sqrt{n}\mu_1/\sigma_1\big| \geq t\big)\big).
\end{aligned}$$

From (A.2), we obtain

$$P(|T_1| \geq t | H_1 = 0) - P(|Z| \geq t) = \mathrm{O}\left(\frac{1}{\sqrt{n}}\right) \qquad \text{a.s.} \tag{A.71}$$

and

$$P(|T_1| \geq t | H_1 = 1) - EP\big(\big|Z + \sqrt{n}\mu_1/\sigma_1\big| \geq t\big) = \mathrm{O}\left(\frac{1}{\sqrt{n}}\right) \qquad \text{a.s.} \tag{A.72}$$

Thus, we have

$$\sup_t \left|\alpha p(t) - (1-\pi_1)P(|Z| \geq t) - G(t)\right| = \mathrm{O}\left(\frac{1}{\sqrt{n}}\right) \qquad \text{a.s.} \tag{A.73}$$

Taking into account (A.70), we have

$$\sup_t |v_m(t) - G(t)| \leq c_2\left(\frac{1}{\sqrt{n}} + \left(\frac{\log\log m}{m}\right)^{1/2}\right) \qquad \text{a.s.} \tag{A.74}$$

for some constant $0 < c_2 < \infty$. Below, we show that there exists a finite constant $c_3 > 0$ such that

$$t_1 - c_3\left(\frac{1}{\sqrt{n}} + \left(\frac{\log\log m}{m}\right)^{1/2}\right) < \hat{t}_{n,m} < t_1 + c_3\left(\frac{1}{\sqrt{n}} + \left(\frac{\log\log m}{m}\right)^{1/2}\right). \tag{A.75}$$

Recalling (A.74), we have, for $\epsilon = c_3(\frac{1}{\sqrt{n}} + (\frac{\log\log m}{m})^{1/2})$, that

$$\begin{aligned} v_m(t_1+\epsilon) &\geq G(t_1+\epsilon) - c_2\left(\frac{1}{\sqrt{n}} + \left(\frac{\log\log m}{m}\right)^{1/2}\right) \\ &= G(t_1) + \epsilon G'(t_1+\theta_1) - c_2\left(\frac{1}{\sqrt{n}} + \left(\frac{\log\log m}{m}\right)^{1/2}\right) \\ &\geq c_1\epsilon - c_2\left(\frac{1}{\sqrt{n}} + \left(\frac{\log\log m}{m}\right)^{1/2}\right) > 2\left(\frac{\log\log m}{m}\right)^{1/2}, \end{aligned}$$

provided that c_3 is chosen large enough: here, $0 \leq \theta_1 \leq \epsilon$ and we have used Lemma A.7. For sufficiently large m, we have

$$\sqrt{m}v_m(t_1+\epsilon) > \tau_m(t_1+\epsilon)z_\gamma.$$

This proves that

$$\hat{t}_{n,m} - t_1 \leq c_3\left(\left(\frac{1}{n}\right)^{1/2} + \left(\frac{\log\log m}{m}\right)^{1/2}\right) \qquad \text{a.s.}$$

Similarly, we have

$$\hat{t}_{n,m} - t_1 \geq -c_3\left(\left(\frac{1}{n}\right)^{1/2} + \left(\frac{\log\log m}{m}\right)^{1/2}\right) \qquad \text{a.s.}$$

This proves (A.67).

Following the same line of proof, we have

$$|t_{n,m} - t_1| = \mathrm{O}\left(\frac{1}{\sqrt{n}} + \left(\frac{\log\log m}{m}\right)^{1/2}\right) \qquad \text{a.s.}$$

If we use

$$\sup_t |\hat{p}_m(t) - p(t)| = \mathrm{O}(m^{-1/2}) \qquad \text{in probability,} \tag{A.76}$$

based on the Donsker theorem instead of (A.69), using the same line of the proof of the a.s. convergence rate, we can obtain the rate of convergence in probability, which is

$$|\hat{t}_{n,m} - t_{n,m}| = \mathrm{O}(n^{-1/2} + m^{-1/2}) \qquad \text{in probability.}$$

This completes the proof of (A.65).

Similarly, the critical value for FDR control is bounded, due to the fact that

$$EP\left(\left|Z + \frac{\sqrt{n}\mu_1}{\sigma_1}\right| \geq t\right) \leq 1.$$

By (A.60), (A.61), (A.71) and (A.72), we have

$$\sup_t \left|\frac{m_0 F_0(t)}{m_0 F_0(t) + m_1 F - 1(t)} - \frac{2(1-\pi_1)\bar{\Phi}(t)}{\hat{p}_m(t)}\right| = \mathrm{O}\left(n^{-1/2} + \left(\frac{\log\log m}{m}\right)^{1/2}\right) \qquad \text{a.s.,}$$

$$\sup_t \left|\frac{m_0 F_0(t)}{m_0 F_0(t) + m_1 F - 1(t)} - \frac{2(1-\pi_1)\bar{\Phi}(t)}{\hat{p}_m(t)}\right| = \mathrm{O}\big(n^{-1/2} + (m)^{-1/2}\big) \qquad \text{in probability.}$$

Noting that $2(1-\pi_1)\bar{\Phi}(t)/[2(1-\pi_1)\bar{\Phi}(t) + EP(|Z + \sqrt{n}\mu_1/\sigma_1| \geq t)]$ is a monotone decreasing continuous function with respect to t, combined with the definitions of $(t_{n,m}^{fdr})$ and $(\hat{t}_{n,m}^{fdr})$, (2.34) and (2.35) hold.

The proof of k-FWER is the same as that given in Theorem 2.2.

A.5. Proof of Theorem 3.1

For the two-sample t-statistic, the only part we need to show is the boundedness of $t_{n,m}$ under independence, which will imply the boundedness in the general dependence case, as happens with the one-sample t-statistic. The remaining results follows along the same lines as the proof in the one sample t-statistic setting. Based on Lemma A.8 below, plus (3.1), and using the same line of proof as in the one-sample t-statistic case, the boundedness of $t_{n,m}$ holds for two-sample t-statistics.

The proof of the boundedness of $t_{n,m}$ is based on the following asymptotic distribution of T_i^* under the alternative hypothesis.

Lemma A.8. *Suppose that $X, X_1, \ldots, X_{n_1}$ are independent and identically distributed random variables from a population with mean μ_1 and variance σ_1^2, and $Y, Y_1, \ldots, Y_{n_2}$ are independent and identically distributed random variables from another population with mean μ_2 and variance σ_2^2. Assume the sampling processes are independent of each other. Also, assume that there are*

$0 < c_1 \le c_2 < \infty$ such that $c_1 \le n_1/n_2 \le c_2$. Let

$$T^* = \frac{\bar{X} - \bar{Y}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}, \tag{A.77}$$

where

$$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i, \qquad \bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i, \tag{A.78}$$

$$s_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2 \quad \text{and} \quad s_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2. \tag{A.79}$$

If $EX^4 < \infty$ and $EY^4 < \infty$, then

$$P(|T^*| \ge t) = P\left(\left|Z + \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right| \ge t\right)\bigl(1 + \mathrm{o}(1)\bigr), \tag{A.80}$$

uniformly in $t = \mathrm{o}(n^{1/6})$, where $n = \max\{n_1, n_2\}$.

Proof. The proof of this lemma is very similar to the proof of Lemma A.3 and so we omit the details. □

A.6. Proof of Theorem 3.2

This follows the same arguments as in the one-sample t-statistic case, by virtue of Lemma A.8.

A.7. Proof of Theorem 3.3

When we plug in an estimator of $P(|T_i^*| \ge t)$,

$$\hat{p}_m(t) = \frac{1}{m}\sum_{i=1}^{m} I_{\{|T_i^*| \ge t\}},$$

the proof of the two-sample t-statistic case follows along the same lines as its one-sample counterpart, except that we have to show the rate of convergence under the alternative hypothesis for the two-sample t-statistic. This follows from the following lemma, which completes the proof of Theorem 3.3.

Lemma A.9. *Let $X, X_1, \ldots, X_{n_1}$ be i.i.d. random variables from a population with mean μ_1 and variance σ_1^2, and $Y, Y_1, \ldots, Y_{n_2}$ be i.i.d. random variables from another population with mean μ_2*

and variance σ_2^2. The sampling processes are assumed to be independent of each other. Assume that there are $0 < c_1 \le c_2 < \infty$ such that $c_1 \le n_1/n_2 \le c_2$. Let T^* be defined as in Lemma A.8. If $E|X|^4 < \infty$ and $E|Y|^4 < \infty$, then

$$\begin{aligned}\left|P(T^* \le x) - \Phi\left(x - \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right)\right| \\ \le \frac{K(1+|x|)}{(1+|x-(\mu_1-\mu_2)/\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}|)\sqrt{\min\{n_1,n_2\}}},\end{aligned} \tag{A.81}$$

where K is a finite constant that may depend on $\sigma_1^2, \sigma_2^2, E|X|^3, E|Y|^3, EX^4$ and EY^4.

Proof. Without loss of generality, we assume that $n_1 = b_1 n$, $n_2 = b_2 n$, $b_1 + b_2 = 1$ with $b_1 > 0$ and $b_2 > 0$. Note that

$$\begin{aligned}P(T^* \le x) &= P\left(\frac{\bar{X} - \mu_1 - (\bar{Y} - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} + \frac{\mu_1 - \mu_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \le x\right) \\ &= P\left(\frac{\bar{X} - \mu_1 - (\bar{Y} - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} + \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \le x\frac{\sqrt{s_1^2/n_1 + s_2^2/n_2}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right) \\ &\le P\left(\frac{\bar{X} - \mu_1 - (\bar{Y} - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \le x - \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} + x\left|\frac{s_1^2/n_1 + s_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} - 1\right|\right),\end{aligned}$$

where we make use of (A.3). We now apply (A.1) with $\xi_i = \frac{(X_i - \mu_1)/n_1}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$ for $1 \le i \le n_1$ and $\xi_i = -\frac{(Y_i - \mu_2)/n_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$ for $n_1 + 1 \le i \le n_1 + n_2$. Let

$$z = x - \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}},$$

$$\Delta = -x\left|\frac{s_1^2/n_1 + s_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} - 1\right|,$$

$$\Delta_i = -x\left|\frac{s_{1,i}^2/n_1 + s_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} - 1\right|$$

for $1 \le i \le n_1$, and

$$\Delta_i = -x\left|\frac{s_1^2/n_1 + s_{2,i}^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} - 1\right|$$

for $n_1+1 \le i \le n_1+n_2$, where $s_{1,i}^2$ is defined as s_1^2 with 0 replacing X_i and $s_{2,i}^2$ is defined as s_2^2 with 0 replacing Y_i. Noting that

$$\frac{s_1^2/n_1+s_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}-1=\frac{1}{\sigma_1^2/n_1+\sigma_2^2/n_2}[(s_1^2-\sigma_1^2)/n_1+(s_2^2-\sigma_2^2)/n_2],$$

we have, by (A.7), that

$$E\left|\frac{s_1^2/n_1+s_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}-1\right|^2 \le K\frac{EX^4+EY^4}{n}.$$

For $1\le i\le n_1$,

$$\begin{aligned}&E\left(\frac{s_1^2/n_1+s_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}-\frac{s_{1i}^2/n_1+s_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}\right)^2\\&\quad=\frac{1}{n_1^2(\sigma_1^2/n_1+\sigma_2^2/n_2)^2}E(s_1^2-s_{1i}^2)^2\le\frac{KEX^4}{n^2},\end{aligned}$$

by (A.8). Similarly, for $n_1+1\le i\le n_1+n_2$, we have

$$E\left(\frac{s_1^2/n_1+s_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}-\frac{s_1^2/n_1+s_{2i}^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}\right)^2=\frac{1}{n_2^2(\sigma_1^2/n_1+\sigma_2^2/n_2)^2}E(s_2^2-s_{2i}^2)\le\frac{KEY^4}{n^2}.$$

It follows that

$$\begin{aligned}\|\Delta\|_2&\le K\frac{|x|\sqrt{EX^4+EY^4}}{\sqrt{n}},\\P\left(|\Delta|>\frac{|z|+1}{3}\right)&\le K\frac{E|\Delta|}{|z|+1}\le K\frac{\|\Delta\|_2}{|z|+1}\le K\frac{|x|\sqrt{EX^4+EY^4}}{\sqrt{n}(|z|+1)},\\\sum_{i=1}^n(E\xi_i^2)^{1/2}\big(E(\Delta-\Delta_i)^2\big)^{1/2}&\le K\frac{\sqrt{(\sigma_1^2+\sigma_2)(EX^4+EY^4)}}{\sqrt{n}},\\\sum_{i=1}^n E|\xi_i|^3&\le K\frac{E|X|^3+E|Y|^3}{\sqrt{n}}.\end{aligned}$$

Therefore, by (A.1),

$$\begin{aligned}&\left|P\left(\frac{\bar{X}-\mu_1-(\bar{Y}-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\le x-\frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}+x\left|\frac{s_1^2/n_1+s_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}-1\right|\right)\right.\\&\quad\left.-\Phi\left(x-\frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\right)\right|\le K\frac{1+|x|}{(1+|x-(\mu_1-\mu_2)/\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}|)\sqrt{n}}.\end{aligned}$$

Similarly,

$$
\begin{aligned}
P(T^* \le x) &= P\left(\frac{\bar{X} - \mu_1 - (\bar{Y} - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} + \frac{\mu_1 - \mu_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \le x\right) \\
&\ge P\left(\frac{\bar{X} - \mu_1 - (\bar{Y} - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \le x - \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} - x\left|\frac{s_1^2/n_1 + s_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} - 1\right|\right)
\end{aligned}
$$

and

$$
\begin{aligned}
&\left|P\left(\frac{\bar{X} - \mu_1 - (\bar{Y} - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \le x - \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} - x\left|\frac{s_1^2/n_1 + s_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} - 1\right|\right)\right. \\
&\quad \left. - \Phi\left(x - \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right)\right| \le K \frac{1 + |x|}{(1 + |x - (\mu_1 - \mu_2)/\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}|)\sqrt{n}}.
\end{aligned}
$$

This proves (A.81). □

Acknowledgements

The authors are grateful to the Associate Editor and two anonymous reviewers for valuable comments and suggestions which improved the paper. The second author was partially supported by grants CA075142 and CA142538 from the U.S. National Institutes of Health and grant DMS-0904184 from the U.S. National Science Foundation.

References

[1] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 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. Stat.* **25** 60–83.

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

[4] Cao, H. (2007). Moderate deviations for two sample t-statistics. *ESAIM Probab. Stat.* **11** 264–271. MR2320820

[5] Chen, L.H.Y. and Shao, Q.M. (2007). Normal approximation for nonlinear statistics using a concentration inequality approach. *Bernoulli* **13** 581–599. MR2331265

[6] Chi, Z. (2007). On the performance of FDR control: Constraints and a partial solution. *Ann. Statist.* **35** 1409–1431. MR2351091

[7] Chi, Z. and Tan, Z. (2008). Positive false discovery proportions: Intrinsic bounds and adaptive control. *Statist. Sinica* **18** 837–860. MR2440397

[8] Craiu, R. and Sun, L. (2008). Choosing the lesser evil: Trade-off between false discovery rate and non-discovery rate. *Statist. Sinica* **18** 861–879. MR2440073

[9] Dudoit, S. and van der Laan, M.J. (2008). *Multiple Testing Procedures with Applications to Genomics*. New York: Springer. MR2373771
[10] Dudley, R.M. and Philipp, W. (1983). Invariance principles for sums of Banach space valued random elements and empirical processes. *Z. Wahrsch. Verw. Gebiete* **62** 509–552. MR0690575
[11] Efron, B., Tibshirani, R., Storey, J.D. and Tusher, V.G. (2001). Empirical bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* **96** 1151–1160. MR1946571
[12] Fan, J., Hall, P. and Yao, Q. (2007). To how many simultaneous hypothesis tests can normal, Student's t or bootstrap calibration be appplied? *J. Amer. Statist. Assoc.* **102** 1282–1288. MR2372536
[13] Genoves, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.* **32** 1035–1061. MR2065197
[14] Golub, T.R. *et al*. (1999). Molecular classifcation of cancer: Class discovery and class prediction by gene expression monitoring. *Science* **286** 531–537.
[15] Kosorok, M. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer.
[16] Kosorok, M. and Ma, S. (2007). Marginal asymptotics for the "large p, small n" paradigm: With application to microarray data. *Ann. Statist.* **35** 1456–1486. MR2351093
[17] Langaas, M. and Lindqvist, B. (2005). Estimating the proportion of true null hypotheses, with application to DNA microarray data. *J. Roy. Statist. Soc. Ser. B* **67** 555–572. MR2168204
[18] Lehmann, E.L. and Romano, J.P. (2005). Generalizations of the familywise error rate. *Ann. Statist.* **33** 1138–1154. MR2195631
[19] Meinshausen, N and Bühlmann, P. (2005). Lower bounds for the number of false null hypotheses for multiple testing of associations. *Biometrika* **92** 893–907. MR2234193
[20] Meinshausen, N and Rice, J. (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.* **34** 373–393. MR2275246
[21] Storey, J. (2002). A direct approach to false discovery rates. *J. Roy. Statist. Soc. Ser. B* **64** 479–498. MR1924302
[22] Storey, J. (2003). The positive false discoery rate: A bayesian interpretation and the q-value. *Ann. Statist.* **31** 2013–2035. MR2036398
[23] Storey, J. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100** 9440–9445. MR1994856
[24] Storey, J., Taylor, J. and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. *J. Roy. Statist. Soc. Ser. B* **66** 187–205. MR2035766
[25] Sun, W. and Cai, T. (2009). Large-scale multiple testing under dependencey. *J. Roy. Statist. Soc. Ser. B* **71** 393–424.
[26] 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**: Article 15 (electronic). MR2101464
[27] Wang, Q. and Hall, P. (2009). Relative errors in central limit theorem for Student's t statistics with applications. *Statist. Sinica* **19** 343–354. MR2487894
[28] Wang, Q. (2008). Absolute and relative errors in central limit theorem for self-normalized sums: Review and new results. Unpublished manuscript.
[29] Wu, W. (2008). On false discovery control under dependence. *Ann. Statist.* **36** 364–380. MR2387975

Received June 2009 and revised January 2010