

MARGINAL ASYMPTOTICS FOR THE “LARGE p , SMALL n ” PARADIGM: WITH APPLICATIONS TO MICROARRAY DATA¹

BY MICHAEL R. KOSOROK AND SHUANGGE MA

University of North Carolina and Yale University

The “large p , small n ” paradigm arises in microarray studies, image analysis, high throughput molecular screening, astronomy, and in many other high dimensional applications. False discovery rate (FDR) methods are useful for resolving the accompanying multiple testing problems. In cDNA microarray studies, for example, p -values may be computed for each of p genes using data from n arrays, where typically p is in the thousands and n is less than 30. For FDR methods to be valid in identifying differentially expressed genes, the p -values for the nondifferentially expressed genes must simultaneously have uniform distributions marginally. While feasible for permutation p -values, this uniformity is problematic for asymptotic based p -values since the number of p -values involved goes to infinity and intuition suggests that at least some of the p -values should behave erratically. We examine this neglected issue when n is moderately large but p is almost exponentially large relative to n . We show the somewhat surprising result that, under very general dependence structures and for both mean and median tests, the p -values are simultaneously valid. A small simulation study and data analysis are used for illustration.

1. Introduction. The “large p , small n ” paradigm [27] arises in microarray studies, image analysis, high throughput molecular screening, astronomy, and in many other high-dimensional applications. Microarrays, in particular, are capable of monitoring the gene expression of thousands of genes and have become routine in biomedical research. Microarray studies of phenotypic variation can lead to a better treatment assignment and so there has been an increasing demand for novel statistical tools analyzing such data. Representative recent developments utilize both semiparametric methods [8, 13, 14, 29] and penalized methods [11, 12].

Although statistical analysis with microarray data has been one of the most investigated areas, theoretical studies of the relevant asymptotic properties remain rare (for important exceptions to this, see [8, 14, 25]). The paucity of such research is partly caused by the abnormality of the microarray data structure. For example, with cDNA microarrays, the dimension of the covariate p (number of genes) is

Received May 2005; revised August 2006.

¹Supported in part by Grant CA075142 from the National Cancer Institute.

AMS 2000 subject classifications. Primary 62A01, 62H15; secondary 62G20, 62G30.

Key words and phrases. Brownian bridge, Brownian motion, empirical process, false discovery rate, Hungarian construction, marginal asymptotics, maximal inequalities, median tests, microarrays, t -tests.

usually much larger than the sample size n (number of arrays). Typically, data is aggregated across the n arrays to form test statistics for each of p genes, resulting in large scale multiple testing. False discovery rate (FDR) methods (see [1]) are used to account for this multiplicity in order to successfully identify which among thousands of monitored genes are significantly differentially expressed.

For FDR methods to be valid for identifying differentially expressed genes, the p -values for the nondifferentially expressed genes must simultaneously have uniform distributions marginally. While this is feasible for permutation based p -values, it is unclear whether such uniformity holds for p -values based on asymptotic approximations. For instance, suppose that we wish to use t -tests for each of the p genes and to compute approximate p -values based on the normal approximation for simplicity. If the data is not normally distributed, we would have to rely on the central limit theorem. Unfortunately, it is unclear whether this will work for all of the tests simultaneously. The issue is that the number of p -values involved goes to infinity and intuition suggests that at least some of the p -values should behave erratically. In this paper, we examine this neglected issue when n is allowed to be moderately large but p is almost exponentially larger than n . We show the somewhat surprising result that, under arbitrary dependence structures for both mean and median tests, the p -values are simultaneously valid.

To further clarify ideas, consider a simple one-sample cDNA microarray study. Note that this data setting and the following discussions can be easily extended to incorporate loop designs as in [29]. Studies using Affymetrix genechip data can be included in the same framework after some modification. Denote Y_{ij} and Z_{ij} as the background-corrected log-ratios and log-intensities (as in [14]), for array $i = 1, \dots, n$ and gene $j = 1, \dots, p$. Consider the following simplified partial linear model:

$$(1) \quad Y_{ij} = \mu_j + h_i(Z_{ij}) + \varepsilon_{ij},$$

where μ_j are the fixed gene effects, $h_i(Z_{ij})$ are the smooth array-specific normalization effects (constrained to be mean zero within array) and ε_{ij} are mean zero random errors. The constraints are for model identifiability. For simplicity of exposition, we have omitted other potentially important terms in our model, such as possible print-tip effects. We note, however, that the theory we present in this paper can extend readily to these richer models.

Models similar to (1) have been investigated by [8, 14]. In [14], asymptotic properties based on least squares estimation are established assuming fixed p and $n \rightarrow \infty$. It is shown that μ_j and h_i can both be consistently estimated with optimal convergence rates. In [8], partial consistency type asymptotics are established. It is proved that when n is fixed and $p \rightarrow \infty$, h_i can be consistently estimated by an estimator \hat{h}_i , although μ_j cannot be consistently estimated. If we let $X_{ij} = \mu_j + \varepsilon_{ij}$ and $\tilde{X}_{ij} = Y_{ij} - \hat{h}_i(Z_{ij})$, the results of [8] can be restated as $\max_{1 \leq i \leq n} \max_{1 \leq j \leq p} |\tilde{X}_{ij} - X_{ij}| = o_P(1)$. This process of removing the h_i effects

is referred to as “normalization,” and the fact that the uniform difference between \tilde{X}_{ij} and X_{ij} goes to zero means that the normalization process is consistent. This permits the use of the normalized array-specific gene effects \tilde{X}_{ij} for inference in place of the true array-specific gene effects X_{ij} . However, because n is fixed, the permissible inference tools at the gene level are restricted to exact methods, such as permutation tests.

The goal of our paper is to study normalization and inference when the number of arrays $n \rightarrow \infty$ slowly while the number of genes $p \gg n$. This is essentially the asymptotic framework considered in van der Laan and Bryan [25], who show that provided the range of expression levels is bounded, the sample means consistently estimate the mean gene effects uniformly across genes whenever $\log p = o(n)$. We extend the results of [8, 14, 25] in three important ways. First, uniform consistency results are extended to general empirical distribution functions and sample medians. Second, a precise Brownian bridge approximation to the empirical distribution function is developed and utilized to establish uniform validity of marginal p -values based on the normal approximation. In other words, we develop a central limit theorem for the large p small n setting. Third, these results are further extended to allow for incorporating the normalization process. We find that the rate requirement ranges from $\log p = o(n^{1/2})$ to $\log p = o(n^{1/5})$, depending on the choice of test statistic and the data assumptions.

An important consequence of these results is that approximate p -values based on normalized gene expression data can be validly applied to FDR methods for identifying differentially expressed genes. We refer to this kind of asymptotic regime as “marginal asymptotics” (see also [18]) because the focus of the inference is at the marginal (gene) level, even though the results are uniformly valid over all genes. The main conclusion of our paper is that the marginal asymptotic regime is valid even if the number of genes increases almost exponentially relative to the number of arrays, that is, $\log p = o(n^\gamma)$ for some $\gamma > 0$. Qualitatively, this seems to be the correct order of asymptotics for microarray experiments with a moderate number, say approximately 30, of replications. The main technical tools we use include maximal inequalities, a specialized Hungarian construction for the empirical distribution function, and a precise bound on the modulus of continuity of Brownian motion.

In this article, we mainly focus on the simplified “one-sample” design. Most microarray studies have in fact more complicated designs. However, they can usually be decomposed and analyzed in the one-sample and/or two-sample design framework. Consider for example the Apo A1 data in [5]. For each chip, the red and green channels are the same genes from different samples/conditions. Thus loosely speaking, we have a one-sample study in the normalization stage, where the outcome is the log-ratio of red over green intensities. After normalization, genes from two subpopulations are compared. We then have a two-sample comparison. For more complicated cDNA study designs, for example the loop design,

similar decompositions can be carried out. For the Affymetrix studies, normalization and statistical models may differ significantly from those used for cDNA studies. However, if the gene effects are of greatest interest, then after normalization, the Affymetrix studies are reasonably close to cDNA studies. Extending the theoretical results to the two-sample setting is straightforward (see [19]), and we include a two-sample study of estrogen data in Section 5.2 below.

The article is organized as follows. In Section 2 we present two discrepancy measures for assessing p -value accuracy. In Section 3 we present the main results on asymptotic consistency of estimates and p -values for both mean and median based tests. The impact of normalization is studied in Section 4. In Section 5 a small simulation study and data analysis are presented. A discussion is given in Section 6, while proofs are presented in Section 7.

2. Assessing p -value consistency. Suppose we have p hypothesis tests with p -values $q_{(p)} \equiv \{q_1, \dots, q_p\}$ but only know the estimated p -values $\hat{q}_{(p)} \equiv \{\hat{q}_1, \dots, \hat{q}_p\}$. An important question is how accurate must $\hat{q}_{(p)}$ be in order for inference based on $\hat{q}_{(p)}$ to be asymptotically equivalent to inference based on $q_{(p)}$? For this paper, the chief hypothesis testing issue is controlling the FDR asymptotically in p . To fix ideas, suppose the indices $J_n = \{1, \dots, p\}$ for the hypothesis tests are divided into two groups, J_{0p} and J_{1p} , where some null hypotheses hold for all $j \in J_{0p}$ and some alternatives hold for all $j \in J_{1p}$. We will assume that q_j is uniformly distributed for all $j \in J_{0p}$ and that q_j has distribution F_1 for all $j \in J_{1p}$, where $F_1(t) \geq t$ for all $t \in [0, 1]$ and F_1 is strictly concave with $\lim_{t \downarrow 0} F_1(t)/t = \infty$. Let $\lambda_p \equiv \#J_{0p}/p$ be the proportion of true null hypotheses, and assume $\lambda_p \rightarrow \lambda_0 \in (0, 1]$ as $p \rightarrow \infty$. Also let $\tilde{F}_p(t) \equiv p^{-1} \sum_{j=1}^p \mathbf{1}\{q_j \leq t\}$, where $\mathbf{1}\{A\}$ is the indicator of A , and assume $\tilde{F}_p(t)$ converges uniformly in t to $F_0(t) \equiv \lambda_0 t + (1 - \lambda_0)F_1(t)$.

The estimate of FDR proposed by [23] (see also [10]) for a p -value threshold of $t \in [0, 1]$ is $\widehat{FDR}_l(t) \equiv \tilde{\lambda}(l)t/(\tilde{F}_p(t) \vee (1/p))$, where $\tilde{\lambda}(l) \equiv (1 - \tilde{F}_p(l))/(1 - l)$ is a conservative estimate of λ_0 , in that $\tilde{\lambda}(l) \rightarrow \lambda_*$ in probability, where $\lambda_0 \leq \lambda_* \leq 1$, and where $a \vee b$ denotes the maximum of a and b . The quantity l is the tuning parameter and is constrained to be in $(0, 1)$ with decreasing bias as l gets closer to zero. Because of the upward bias in $\tilde{\lambda}(l)$, if $\tilde{\lambda}(l)$ is distinctly < 1 , then one can be fairly confident that $\lambda_0 < 1$.

We first consider $\lambda_0 < 1$. The asymptotic FDR for the procedure rejecting all hypotheses corresponding to indices with $p_j \leq t$ is $r_0(t) \equiv \lambda_0 t/(\lambda_0 t + (1 - \lambda_0)F_1(t))$. [24] demonstrates that under fairly general dependencies among the p -values $q_{(p)}$, $\tilde{F}_p(t)$ converges to $F_0(t)$, and thus $\widehat{FDR}_l(t)$ converges in probability to $r_*(t) \equiv (\lambda_*/\lambda_0)r_0(t)$. Our assumptions on F_1 ensure that $r_0(t)$ is monotone increasing with derivative $\dot{r}_0(t)$ bounded over $t \in (\delta, 1]$ by $(4\delta)^{-1}$. Thus, for each $\rho \in [0, \lambda_*]$, there exists a $t \in [0, 1]$ with $r_*(t) = \rho$ and $r_0(t) \leq \rho$. Thus using $\widehat{FDR}_l(t)$ to control FDR is asymptotically valid, albeit conservative.

Suppose all we have available is $\hat{q}_{(p)}$. Now we estimate F_0 with $\hat{F}_p(t) = p^{-1} \sum_{j=1}^p \mathbf{1}\{\hat{q}_j \leq t\}$ and λ_* with $\hat{\lambda}_l(t) \equiv (1 - \hat{F}_p(l))/(1 - l)$. The previous results will all hold for $\widehat{FDR}_l(t) \equiv \hat{\lambda}(l)t/(\hat{F}_p(t) \vee (1/p))$, provided \hat{F}_p is uniformly consistent for F_0 . We now show that a sufficient condition for this is $E_{1p}(\hat{q}_{(p)}, q_{(p)}) \equiv \max_{1 \leq j \leq p} |\hat{q}_j - q_j| \rightarrow 0$ in probability. Under this condition, there exists a positive sequence $\varepsilon_p \downarrow 0$ such that $P(E_{1p}(\hat{q}_{(p)}, q_{(p)}) > \varepsilon_p) \rightarrow 0$ in probability. Accordingly, we have with probability tending to 1 that for any $t \in [\varepsilon_p, 1 - \varepsilon_p]$, $\tilde{F}_p(t - \varepsilon_p) \leq \hat{F}_p(t) \leq \tilde{F}_p(t + \varepsilon_p)$. Thus, by continuity of F_0 , uniform consistency of \hat{F}_p follows from uniform consistency of \tilde{F}_p . In summary, the above procedure for controlling FDR is asymptotically valid when $\lambda_0 < 1$, provided $E_{1p}(\hat{q}_{(p)}, q_{(p)})$ goes to zero in probability.

The above result does not hold when $\lambda_0 = 1$. Since this forces $\lambda_* = 1$, we will have warning whenever $\hat{\lambda}(l)$ or $\tilde{\lambda}(l)$ is close to 1. The main difficulty is that $r_0(t) = 1$ for all $t \in (0, 1]$, and thus the previous asymptotic arguments will not hold. This issue is confronted in Fan, Hall and Yao [7], who utilize large deviation results to build on an earlier version of the present paper [19] for controlling FDR using t -tests and bootstrap calibration. They show that controlling the error in the p -value ratios, $\max_{1 \leq j \leq p} |\hat{q}_j/q_j - 1|$, is needed. An additional advantage of controlling the p -value ratios, is that the relative order of significantly small p -values can be better determined. This is useful, for example, for prioritizing genes for follow-up studies.

The ratio-error measure we propose is

$$E_{2p}^{(\alpha)}(\hat{q}_{(p)}, q_{(p)}) \equiv \max_{1 \leq j \leq p} \left(\mathbf{1}\{q_j > \alpha/(2p)\} \left| \frac{\hat{q}_j}{q_j} - 1 \right| + \mathbf{1}\{q_j \leq \alpha/(2p), \hat{q}_j > \alpha/p\} \right).$$

We require $E_{2p}^{(\alpha)}(\hat{q}_{(p)}, q_{(p)}) \rightarrow 0$, in probability, for all $\alpha \in (0, 1]$. This error is not quite as stringent as the uniform ratio error given in the previous paragraph, but it is nearly so, as seen in the following simple lemma.

LEMMA 1. *Suppose $E_{2p}^{(\alpha)}(\hat{q}_{(p)}, q_{(p)}) \rightarrow 0$ in probability, as $p \rightarrow \infty$, for every $\alpha \in (0, 1]$. Then $\max_{j \in J_{0p}} |\hat{q}_j/q_j - 1| \rightarrow 0$ in probability, as $p \rightarrow \infty$.*

The proof easily follows from the marginal uniformity of q_j when $j \in J_{0n}$, since this implies $P(\min_{j \in J_{0p}} q_j \leq \alpha/(2p)) \leq \sum_{j \in J_{0p}} P(q_j \leq \alpha/(2p)) \leq \alpha/2$ and since α can be chosen arbitrarily close to 0.

The $E_{2p}^{(\alpha)}$ error measure also requires that when $q_j \leq \alpha/(2p)$, \hat{q}_j must satisfy the most stringent Bonferroni correction at level α . We now argue that $E_{2p}^{(\alpha)}(\hat{q}_{(p)}, q_{(p)}) \rightarrow 0$ in probability for every $\alpha \in (0, 1]$ ensures that controlling FDR using $\widehat{FDR}_l(t)$ is asymptotically valid.

We need to modify our previous conditions to assume that $\lambda_p \rightarrow 1$; that $t^{-\delta} F_1(t) \rightarrow \infty$, as $t \rightarrow 0$, for some $\delta \in (0, 1)$; and that $(1 - \lambda_p)^{-1} = o(p^{1-\delta})$. Define $FDR_{(p)}(t)$ to be the sequence of true FDRs at significance level t , $\tilde{G}_p(t)$ to be the observed empirical distribution of the $p\lambda_p$ p -values from null hypotheses, and let $r_{(p)}(t) \equiv \lambda_p t / (\lambda_p t + (1 - \lambda_p) F_1(t))$. The second assumption strengthens the degree of concavity of F_1 in the left tail, while the third assumption prevents $r_{(p)}$ from converging to 1 too rapidly. As we did previously, we will control FDR at the level ρ by finding a $t \in (0, 1)$ such that $\widehat{FDR}_l(t) \leq \rho$ and rejecting hypotheses corresponding to estimated p -values \hat{q}_j for which $\hat{q}_j \leq t$. The following lemma tells us when this procedure will be asymptotically valid for controlling FDR.

LEMMA 2. Fix $\rho, l \in (0, 1)$. In addition to the given conditions, assume that $E_{2p}^{(\alpha)}(\hat{q}_{(p)}, q_{(p)}) \rightarrow 0$ in probability for every $\alpha \in (0, 1]$, and that

$$(2) \quad \sup_{t \in [t_p, 1]} |\widehat{FDR}_l(t) - (\lambda_{*}/\lambda_0)r_{(p)}(t)| \rightarrow 0 \quad \text{and} \quad \sup_{t \in [t_p, 1]} |t^{-1}\tilde{G}_p(t) - 1| \rightarrow 0,$$

in probability, where $t_p = (1 - \lambda_p)^{1/(1-\delta)}$. Let $\hat{t}_p = \sup\{t : \widehat{FDR}_l(\hat{t}_p) \leq \rho\}$ for all $p \geq 1$. Then $FDR_{(p)}(\hat{t}_p) \vee \rho \rightarrow \rho$, in probability.

The following lemma provides a sufficient condition for (2).

LEMMA 3. Condition (2) holds when q_1, \dots, q_p are independent.

REMARK 1. We now demonstrate that when $\lambda_0 = 1$, $E_{1p}(\hat{q}_{(p)}, q_{(p)}) \rightarrow 0$ in probability does not in general guarantee asymptotic control of FDR using \widehat{FDR}_l . First note that since $\widehat{FDR}_l(t)$, $r_{(p)}(t)$ and $\tilde{\lambda}(l)$ are uniformly bounded almost surely, we have that (2) implies $\sup_{t \in [t_p, 1]} |FDR_{(p)}(t) - r_{(p)}(t)| \rightarrow 0$ by bounded convergence. Assume for illustration that $F_1(t) = t^{1/2}$ and $\lambda_p = 1 - p^{-1/3}$, and note these satisfy the required assumptions for Lemma 2. Fix $\rho \in (0, 1)$, and note that $t_p = \sqrt{\rho / ((1 - \rho)(p^{1/2} - 1))}$ solves $r_{(p)}(t_p) = \rho$ for all p large enough. Now $pt_p \rightarrow \infty$ but $t_p \rightarrow 0$. Thus uniform estimation of t_p will not be enough to control FDR at the level ρ since $r_{(p)}(t) \rightarrow 1$, as $p \rightarrow \infty$, for all $t \in (0, 1]$. In particular, we need $E_{1p}(\hat{q}_{(p)}, q_{(p)}) = o_P(p^{-1/4})$.

In summary, using $\widehat{FDR}_l(t)$ to control FDR is asymptotically valid under reasonable regularity conditions, provided $E_{1p}(\hat{q}_{(p)}, q_{(p)}) \rightarrow 0$ in probability when $\lambda_0 < 1$, or $E_{2p}^{(\alpha)}(\hat{q}_{(p)}, q_{(p)}) \rightarrow 0$ in probability, for every $\alpha \in (0, 1]$, when $\lambda_0 = 1$. For the remainder of the paper, we will drop the assumptions on the distributions of the p -values $q_{(p)}$, except that p -values corresponding to null hypotheses will still be uniform. Our focus hereafter is on conditions under which estimated p -values are uniformly consistent via E_{1p} and $E_{2p}^{(\alpha)}$.

3. Marginal asymptotics. The results of this section are based on the true data (without normalization error). For each $n \geq 1$, let $X_{1(n)}, \dots, X_{n(n)}$ be a sample of i.i.d. vectors (e.g., microarrays) of length p_n , where the dependence within vectors is allowed to be arbitrary. Denote the j th component (e.g., gene) of the i th vector $X_{ij(n)}$, that is, $X_{i(n)} = (X_{i1(n)}, \dots, X_{ip_n(n)})'$. Also let the marginal distribution of $X_{1j(n)}$ be denoted $F_{j(n)}$, and let $\hat{F}_{j(n)}(t) = n^{-1} \sum_{i=1}^n \mathbf{1}\{X_{ij(n)} \leq t\}$, for all $t \in \mathbb{R}$ and each $j = 1, \dots, p_n$.

We first establish, in Section 3.1, uniform consistency of the marginal empirical distribution function estimator and also uniformity of a Brownian bridge approximation to the standardized empirical distribution. These results are used in Sections 3.2 and 3.3 to study inferential properties of the marginal means and medians. Note that both the mean and median are functionals of the empirical distribution function.

3.1. *Consistency of marginal empirical distribution functions.* This section, consisting of Theorems 1 and 2 below, is the basis for the results of Sections 3.2 and 3.3. The two theorems are somewhat surprising, high-dimensional extensions of two classical univariate results for empirical distribution functions: the celebrated Dvoretzky, Kiefer and Wolfowitz inequality [6] as refined by Massart [20], and the celebrated Komlós, Major and Tusnády Hungarian construction [15] as refined by Bretagnolle and Massart [3]. The extensions utilize maximal inequalities based on Orlicz norms (see Chapter 2.2 of [26]). For any real random variable Y and any $d \geq 1$, let $\|Y\|_{\psi_d}$ denote the Orlicz norm for $\psi_d(x) = e^{x^d} - 1$, that is, $\|Y\|_{\psi_d} = \inf\{C > 0: E[e^{|Y|^d/C^d} - 1] \leq 1\}$. Note that $\|\cdot\|_{\psi_d}$ increases with d (up to a constant depending on d) and that $\|\cdot\|_{\psi_1}$ dominates all L_p norms (up to a constant depending on p). We denote the uniform norm $\|\cdot\|_\infty$.

The first theorem yields simultaneous consistency of all $\hat{F}_{j(n)}$'s.

THEOREM 1. *For a universal constant $0 < c_0 < \infty$ and all $n, p_n \geq 2$,*

$$(3) \quad \left\| \max_{1 \leq j \leq p_n} \|\hat{F}_{j(n)} - F_{j(n)}\|_\infty \right\|_{\psi_2} \leq c_0 \sqrt{\frac{\log p_n}{n}}.$$

In particular, if $n \rightarrow \infty$ and $\log p_n = o(n)$, then the left-hand side of (3) $\rightarrow 0$.

REMARK 2. The rate on the right-hand side of (3) is sharp, in the sense that there exist sequences of datasets, where $(\log p_n/n)^{-1/2} \max_{1 \leq j \leq p_n} \|\hat{F}_{j(n)} - F_{j(n)}\|_\infty \rightarrow c > 0$, in probability, as $n \rightarrow \infty$. In particular, this is true if the genes are all independent, $n, p_n \rightarrow \infty$ with $\log p_n = o(n)$, and $c = 1/2$.

The second theorem shows that the standardized empirical processes $\sqrt{n}(\hat{F}_{j(n)} - F_{j(n)})$ can be simultaneously approximated by Brownian bridges in

a manner which preserves the original dependence structure in the data. For example, if the original data has *weak dependence*, as defined in [24], then so will the approximating Brownian bridges. To this end, let $\mathcal{F}_{j(n)}$ be the smallest σ -field making all of $X_{1j(n)}, \dots, X_{nj(n)}$ measurable, $1 \leq j \leq p_n$, and let \mathcal{F}_n be the smallest σ -field making all of $\mathcal{F}_{1(n)}, \dots, \mathcal{F}_{p_n(n)}$ measurable.

THEOREM 2. For universal constants $0 < c_1, c_2 < \infty$ and all $n, p_n \geq 2$,

$$(4) \quad \left\| \max_{1 \leq j \leq p_n} \left\| \sqrt{n}(\hat{F}_{j(n)} - F_{j(n)}) - B_{j(n)}(F_{j(n)}) \right\|_{\infty} \right\|_{\psi_1} \leq \frac{c_1 \log n + c_2 \log p_n}{\sqrt{n}},$$

for some stochastic processes $B_{1(n)}, \dots, B_{p_n(n)}$ which are conditionally independent given \mathcal{F}_n and for which each $B_{j(n)}$ is a standard Brownian bridge with conditional distribution given \mathcal{F}_n depending only on $\mathcal{F}_{j(n)}$, $1 \leq j \leq p_n$.

3.2. Inference for marginal sample means. For each $1 \leq j \leq p_n$, assume for this section that $F_{j(n)}$ has finite mean $\mu_{j(n)}$ and standard deviation $\sigma_{j(n)} > 0$. Let $\bar{X}_{j(n)}$ be the sample mean of $X_{1j(n)}, \dots, X_{nj(n)}$. The following corollary yields simultaneous consistency of the marginal sample means.

COROLLARY 1. Assume the closure of the support of $F_{j(n)}$ is a compact interval $[a_{j(n)}, b_{j(n)}]$ with $a_{j(n)} \neq b_{j(n)}$. Under the conditions of Theorem 1 and with the same constant c_0 , we have for all $n, p_n \geq 2$,

$$(5) \quad \left\| \max_{1 \leq j \leq p_n} |\bar{X}_{j(n)} - \mu_{j(n)}| \right\|_{\psi_2} \leq c_0 \sqrt{\frac{\log p_n}{n}} \max_{1 \leq j \leq p_n} |b_{j(n)} - a_{j(n)}|.$$

REMARK 3. Note that Corollary 1 slightly extends the large p small n consistency results of [25] by allowing the range of the support to increase with n provided it does not increase too rapidly.

The following corollary relaxes the somewhat restrictive bounded support assumption at the expense of decreasing the rate of convergence.

COROLLARY 2. Assume that there exist constants $0 < k_1, k_2 < \infty$ and $r \geq 1$ such that $P(|X_{1j(n)} - \mu_{j(n)}| > u) \leq k_1 e^{-k_2 u^r}$, for all $u > 0$ and all $1 \leq j \leq p_n$. Then, provided $n \rightarrow \infty$ and $p_n \geq 2$, we have

$$(6) \quad \max_{1 \leq j \leq p_n} |\bar{X}_{j(n)} - \mu_{j(n)}| = O_P \left(\sqrt{\frac{\log p_n}{n}} (\log p_n + \log n)^{1/r} + (np_n)^{-k_2} \right).$$

In particular, the right-hand-side goes to zero provided $\log p_n = o(n^{r/(2+r)})$.

Now suppose we wish to test the marginal null hypothesis $H_0^{j(n)} : \mu_{j(n)} = \mu_{0,j(n)}$ with $T_{j(n)} = \sqrt{n}(\bar{X}_{j(n)} - \mu_{0,j(n)})/\hat{\sigma}_{j(n)}$, where $\hat{\sigma}_{j(n)}$ is a location-invariant and consistent estimator of $\sigma_{j(n)}$. To use FDR, we need uniformly consistent estimates of the p -values of these tests. Permutation methods can be used. An easier way is to use $\hat{\pi}_{j(n)} = 2\Phi(-|T_{j(n)}|)$, where Φ is the standard normal distribution function, but we need to show this is valid. For the estimator $\hat{\sigma}_{j(n)}$, we require $\hat{\sigma}_{j(n)}/\sigma_{j(n)}$ to be uniformly consistent for 1, that is, $E_{0n} \equiv \max_{1 \leq j \leq p_n} |\hat{\sigma}_{j(n)}^2/\sigma_{j(n)}^2 - 1| = o_P(1)$. One choice for $\hat{\sigma}_{j(n)}^2$ that satisfies this requirement is the sample variance $S_{j(n)}^2$ for $X_{1j(n)}, \dots, X_{nj(n)}$:

COROLLARY 3. *Assume $n \rightarrow \infty$, with $p_n \geq 2$ and $\log p_n = o(n^\gamma)$ for some $\gamma \in (0, 1]$. The following are true under the given assumptions.*

(i) *Assume the closure of the support of $F_{j(n)}$ is compact as in Corollary 1, and let $d_n \equiv \max_{1 \leq j \leq p_n} \sigma_{j(n)}^{-1} |b_{j(n)} - a_{j(n)}|$. Then $E_{0n} = O(n^{-1}) + o_P(d_n^2 n^{\gamma/2-1/2})$. In particular, if $d_n = O(1)$, then $E_{0n} = o_P(1)$.*

(ii) *Assume there exist constants $0 < k_1, k_2 < \infty$ such that $P(\sigma_{j(n)}^{-1} |X_{1j(n)} - \mu_{j(n)}| > x) \leq k_1 e^{-k_2 x^2}$ for all $x > 0$, $1 \leq j \leq p_n$, and all $n \geq 1$. Then $E_{0n} = O(n^{-1}) + O_P((np_n)^{-k_2}) + o_P(n^{3\gamma/2-1/2})$. In particular, if $\gamma \in (0, 1/3]$, then $E_{0n} = o_P(1)$.*

This approach leads to uniformly consistent p -values:

COROLLARY 4. *Assume as $n \rightarrow \infty$ that $p_n \geq 2$, $\log p_n = o(n^\gamma)$, for some $\gamma \in (0, 1/2]$, and $E_{0n} = o_P(1)$. Then there exist standard normal random variables $Z_{1(n)}, \dots, Z_{p_n(n)}$ which are conditionally independent given \mathcal{F}_n , with each $Z_{j(n)}$ having conditional distribution given \mathcal{F}_n depending only on $\mathcal{F}_{j(n)}$, $1 \leq j \leq p_n$, such that the following hold for $\pi_{(n)} \equiv \pi_{1(n)}, \dots, \pi_{p_n(n)}$ and $\hat{\pi}_{(n)} \equiv \hat{\pi}_{1(n)}, \dots, \hat{\pi}_{p_n(n)}$, where*

$$(7) \quad \pi_{j(n)} \equiv 2\Phi\left(-\left|Z_{j(n)} + \frac{\sqrt{n}(\mu_{j(n)} - \mu_{0,j(n)})}{\sigma_{j(n)}}\right|\right), \quad 1 \leq j \leq p_n :$$

(a) *Provided the support of $F_{j(n)}$ is bounded as in part (i) of Corollary 3:*

(i) $E_{1p_n}(\hat{\pi}_{(n)}, \pi_{(n)}) = O_P(E_{0n}^{1/2}) + o_P(n^{\gamma-1/2}d_n)$. In particular, the error goes to zero if $\hat{\sigma}_{j(n)}^2 = S_{j(n)}^2$ and $d_n = O(1)$.

(ii) *If, in addition, $E_{0n} = O_P(n^{-\gamma})$ and $d_n = O(n^{-3\gamma/2+1/2})$, then $E_{2p_n}^{(\alpha)}(\hat{\pi}_{(n)}, \pi_{(n)}) = o_P(1)$, for all $\alpha \in (0, 1]$. In particular, the conditions hold for $\hat{\sigma}_{j(n)}^2 = S_{j(n)}^2$, $d_n = O(1)$ and $\gamma \in (0, 1/3]$.*

(b) *Provided the $F_{j(n)}$ have sub-Gaussian tails as in part (ii) of Corollary 3:*

(i) $E_{1p_n}(\hat{\pi}_{(n)}, \pi_{(n)}) = O_P(E_{0n}^{1/2}) + o_P(n^{3\gamma/2-1/2})$. In particular, the error goes to zero if $\hat{\sigma}_{j(n)}^2 = S_{j(n)}^2$ and $\gamma \in (0, 1/3]$.

(ii) If, in addition, $E_{0n} = O_P(n^{-\gamma})$ and $\gamma \in (0, 1/4]$, then $E_{2p_n}^{(\alpha)}(\hat{\pi}_{(n)}, \pi_{(n)}) = o_P(1)$, for all $\alpha \in (0, 1]$. In particular, the conditions hold for $\hat{\sigma}_{j(n)}^2 = S_{j(n)}^2$ and $\gamma \in (0, 1/5]$.

REMARK 4. Corollary 4 tells us that if we assume bounded ranges of the distributions and if uniform convergence of p -values is sufficient, then the approximate p -values are asymptotically valid, provided we use the sample standard deviation for t -tests and $\log p_n = o(n^{1/2})$. If we also need the ratios of the p -values to be valid, then we need $\log p_n = o(n^{1/3})$. If we weaken our bounded range assumption to only requiring uniformly sub-Gaussian tails, then the rate requirements for uniform convergence and ratio convergence of the p -values become $\log p_n = o(n^{1/3})$ and $\log p_n = o(n^{1/5})$, respectively.

It has been shown that performance of gene-level p -values can be improved through assuming homogeneity of variance across genes [9]. If we are willing to make this assumption, that is, that $\sigma_{j(n)}^2 = \sigma^2$ for all $1 \leq j \leq p_n$, then Corollary 4 can be strengthened if we estimate σ^2 with $\hat{\sigma}^2 \equiv p_n^{-1} \sum_{i=1}^{p_n} S_{j(n)}^2$:

COROLLARY 5. Assume the conditions of part (b)(ii) of Corollary 4 are strengthened to require $0 < \sigma_{j(n)}^2 = \sigma^2 < \infty$ and $\hat{\sigma}_{j(n)}^2 = \hat{\sigma}^2$, $1 \leq j \leq p_n$. Then $E_{0n} = O_P(n^{-1/2})$ and $E_{2p_n}^{(\alpha)}(\hat{\pi}_{(n)}, \pi_{(n)}) = o_P(1)$ for all $\alpha \in (0, 1]$, provided $\gamma \in (0, 1/4]$.

REMARK 5. This yields a relaxation from requiring $\log p_n = o(n^{1/5})$ to only requiring $\log p_n = o(n^{1/4})$. The other error rates in Corollary 4 do not appear to benefit from faster convergence of E_{0n} .

3.3. *Inference for marginal sample medians.* Assume each $F_{j(n)}$ has median $\xi_{j(n)}$ and is continuous in a neighborhood of $\xi_{j(n)}$ with density $f_{j(n)}$. In this section, we do not require the support of $F_{j(n)}$ to be compact, but we do assume there exist $\eta, \tau > 0$ such that

$$(8) \quad \min_{1 \leq j \leq p_n} \inf_{x: |x - \xi_{j(n)}| \leq \eta} f_{j(n)}(x) \geq \tau.$$

Denote the sample median for $X_{1j(n)}, \dots, X_{nj(n)}$ as $\hat{\xi}_{j(n)}$, that is, let $\hat{\xi}_{j(n)} = \inf\{x : \hat{F}_{j(n)}(x) \geq 1/2\}$. The following gives uniform consistency of $\hat{\xi}_{j(n)}$.

COROLLARY 6. Under condition (8) (for some $\eta, \tau > 0$) and the conditions of Theorem 1, we have that

$$(9) \quad \max_{1 \leq j \leq p_n} |\hat{\xi}_{j(n)} - \xi_{j(n)}| = O_P\left(\frac{\log(n \vee p_n)}{n} + \sqrt{\frac{\log p_n}{n}}\right).$$

Now assume we wish to test the marginal null hypotheses $H_0^{j(n)}: \xi_{j(n)} = \xi_{0,j(n)}$ with $U_{j(n)} = 2\sqrt{n} \hat{f}_{j(n)}(\hat{\xi}_{j(n)} - \xi_{0,j(n)})$, where $\hat{f}_{j(n)}$ is a consistent estimator of $f_{j(n)}(\xi_{j(n)})$. As discussed in [16], this is a good choice of median test because it converges rapidly to its limiting Gaussian distribution and appears to have better moderate sample size performance compared to other median tests. As with the marginal mean test, we need consistent estimates of the p -values of these tests. We now study the consistency of the p -value estimates $\hat{\pi}'_{j(n)} = 2\Phi(-|U_{j(n)}|)$. We need additional conditions. Assume there exist $\eta, \tau > 0$ and $M < \infty$ such that (8) holds and, moreover, that

$$(10) \quad \max_{1 \leq j \leq p_n} \sup_{x: |x - \xi_{j(n)}| \leq \eta} f_{j(n)}(x) \leq M$$

and

$$(11) \quad \max_{1 \leq j \leq p_n} \sup_{\varepsilon \leq \eta} \sup_{u: |u| \leq \varepsilon} \frac{|f_{j(n)}(\xi_{j(n)} + u) - f_{j(n)}(\xi_{j(n)})|}{\varepsilon^{1/2}} \leq M.$$

Let $E'_{0n} \equiv \max_{1 \leq j \leq p_n} |\hat{f}_{j(n)} - f_{j(n)}(\xi_{j(n)})|$. One simple choice for $\hat{f}_{j(n)}$ is

$$(12) \quad \tilde{f}_{j(n)} = \frac{\hat{F}_{j(n)}(\hat{\xi}_{j(n)} + h_n) - \hat{F}_{j(n)}(\hat{\xi}_{j(n)} - h_n)}{2h_n},$$

where the window width h_n goes to zero in probability, as $n \rightarrow \infty$. The following corollary shows that this estimator is uniformly consistent.

COROLLARY 7. *Assume condition (10) holds and that $\log p_n = o(n^\gamma)$ for some $\gamma \in (0, 1]$, as $n \rightarrow \infty$. Suppose also that $\hat{f}_{j(n)} = \tilde{f}_{j(n)}$ for all $1 \leq j \leq p_n$, $h_n = o_P(1)$ and $h_n^{-1} = O_P(n^{(1-\gamma)/2})$. Then $E'_{0n} = o_P(h_n^{-1} n^{-(1-\gamma)/2}) + O_P(h_n^{1/2}) = o_P(1)$. In particular, the selection $h_n = O_P(n^{-(1-\gamma)/4})$ and $h_n^{-1} = O_P(n^{5(1-\gamma)/24})$ yields $E'_{0n} = O_P(n^{-(1-\gamma)/8})$, while the choice $h_n = O_P(n^{-(1-\gamma)/3})$ and $h_n^{-1} = O_P(n^{(1-\gamma)/3})$ yields a further improvement to $E'_{0n} = O_P(n^{-(1-\gamma)/6})$.*

There are many other possible kernel estimators with other choices of window widths that will also work, but we will not pursue them here.

We are now ready for the following corollary about p -value consistency.

COROLLARY 8. *Assume as $n \rightarrow \infty$ that $p_n \geq 2$, $\log p_n = o(n^\gamma)$, for some $\gamma \in (0, 1/3]$, and $E'_{0n} = o_P(1)$. Assume also that conditions (8), (10) and (11), for some $\eta, \tau > 0$ and $M < \infty$, hold. Then*

$$(13) \quad E_{1p_n}(\hat{\pi}'_{j(n)}, \pi'_{j(n)}) = o_P(1),$$

where

$$(14) \quad \pi'_{j(n)} = 2\Phi(-|Z_{j(n)} + 2\sqrt{n} f_{j(n)}(\xi_{j(n)})(\xi_{j(n)} - \xi_{0,j(n)})|),$$

and, for each $n \geq 1$, $Z_{1(n)}, \dots, Z_{p_n(n)}$ are standard normals conditionally independent given \mathcal{F}_n and for which each $Z_{j(n)}$ has conditional distribution given \mathcal{F}_n depending only on $\mathcal{F}_{j(n)}$, $1 \leq j \leq p_n$. In particular, (13) holds if $\hat{f}_{j(n)} = \tilde{f}_{j(n)}$ for all $1 \leq j \leq p_n$, $h_n = o_P(1)$ and $h_n^{-1} = O_P(n^{(1-\gamma)/2})$. If, moreover, we require $\gamma \in (0, 1/5]$, $h_n = O_P(n^{-(1-\gamma)/4})$ and $h_n^{-1} = O_P(n^{5(1-\gamma)/24})$, then $E_{2p_n}^{(\alpha)}(\hat{\pi}'_{(n)}, \pi'_{(n)}) = o_P(1)$, for every $\alpha \in (0, 1]$.

REMARK 6. It is unclear if any improvements in the choice of γ are possible through combining information across genes ($1 \leq j \leq p_n$), as was done in Corollary 5 for the mean inference setting.

4. Impact of microarray normalization. In this section, we consider the affect of normalization on the theory presented in Sections 2–3. For the simple normalization model (1), this may require the \hat{h}_i 's to be uniformly consistent at the rate $O_P(n^\beta \log n)$, for some $\beta \in [1/2, 1)$. This requirement seems reasonable for certain estimation methods, including the method described in [8], which benefits from the so-called “blessing of dimensionality.” In these methods, data across all genes within each array are utilized for estimating the h_i 's. Since the number of genes p_n usually increases nearly exponentially relative to the number of microarrays, the number of observations available for estimating the h_i 's is many orders of magnitude higher than n , even after taking into account dependencies within arrays and the fact that the number of arrays is increasing in n . For this particular facet of our problem, the large number of genes actually works in our favor. A variant of this argument can also be found in [18].

Let $\tilde{X}_{i(n)} = (\tilde{X}_{i1(n)}, \dots, \tilde{X}_{ip_n(n)})'$ be an approximation of the “true data” $X_{i(n)}$, $1 \leq i \leq n$, and define $\hat{\epsilon}_n \equiv \max_{1 \leq j \leq p_n; 1 \leq i \leq n} |\tilde{X}_{ij(n)} - X_{ij(n)}|$. With proper, partially consistent normalization, the true gene effects $\{X_{ij(n)}, 1 \leq j \leq p_n, 1 \leq i \leq n\}$ should be uniformly consistently estimated by the residuals from the normalization $\{\tilde{X}_{ij(n)}, 1 \leq j \leq p_n, 1 \leq i \leq n\}$. We now examine conditions under which the results of Sections 3.2 and 3.3 carry through after normalization, that is, after $\hat{\epsilon}_n = o_P(1)$.

4.1. *Inference for marginal sample means.* Let $\check{X}_{j(n)}$ and $\check{S}_{j(n)}^2$ be the sample mean and variance for the normalized sample $\tilde{X}_{1j(n)}, \dots, \tilde{X}_{nj(n)}$. Define $\check{E}_{0n} \equiv \max_{1 \leq j \leq p_n} |\check{S}_{j(n)}^2/S_{j(n)}^2 - 1|$ and assume throughout this section that $\min_{1 \leq j \leq p_n} \sigma_{j(n)} \geq \tau > 0$ for all $n \geq 1$. We have the following.

LEMMA 4. (a) $\max_{1 \leq j \leq p_n} |\check{X}_{j(n)} - \bar{X}_{j(n)}| = O_P(\hat{\epsilon}_n)$.

(b) Assume $n \rightarrow \infty$, with $p_n \geq 2$ and $\log p_n = o(n^\gamma)$ for some $\gamma \in (0, 1]$.

(i) If the support of $F_{j(n)}$ is bounded and $d_n = O(1)$, $\check{E}_{0n} = O_P(\hat{\epsilon}_n)$.

(ii) If the sub-Gaussian tail hypothesis of Corollary 3 holds and $\gamma \in (0, 1/3]$, $\check{E}_{0n} = O_P(\hat{\epsilon}_n)$.

The above leads to the following result. Define first $\check{\pi}_{(n)} \equiv \check{\pi}_{1(n)}, \dots, \check{\pi}_{p_n(n)}$, where $\check{\pi}_{j(n)} \equiv 2\Phi(-|\check{T}'_{j(n)}|)$, $\check{T}'_{j(n)} \equiv \sqrt{n}(\check{X}_{j(n)} - \mu_{0,j(n)})/\check{S}_{j(n)}$, $1 \leq j \leq p_n$.

COROLLARY 9. Assume the conditions of Corollary 4. Then:

- (a) Provided the support of $F_{j(n)}$ is bounded and $d_n = O(1)$, we have
 - (i) $E_{1p_n}(\check{\pi}_{(n)}, \pi_{(n)}) = o_P(1)$ when $\hat{\epsilon}_n = o_P(n^{-1/2})$; and (ii) $E_{2p_n}^{(\alpha)}(\check{\pi}_{(n)}, \pi_{(n)}) = o_P(1)$, for all $\alpha \in (0, 1]$, when $\gamma \in (0, 1/3]$ and $\hat{\epsilon}_n = O_P(n^{-1/2-\gamma/2})$.
 - (b) Provided $F_{j(n)}$ has sub-Gaussian tails as in part (b) of Corollary 4, we have (i) $E_{1p_n}(\check{\pi}_{(n)}, \pi_{(n)}) = o_P(1)$ when $\gamma \in (0, 1/3]$ and $\hat{\epsilon}_n = o_P(n^{-1/2})$; and (ii) $E_{2p_n}^{(\alpha)}(\check{\pi}_{(n)}, \pi_{(n)}) = o_P(1)$, for all $\alpha \in (0, 1]$, when $\gamma \in (0, 1/5]$ and $\hat{\epsilon}_n = O_P(n^{-1/2-\gamma/2})$.

REMARK 7. Provided $\hat{\epsilon}_n = O_P(n^{-1/2})$ and utilizing arguments similar to those used in Corollary 5, Lemma 4 yields that $\max_{1 \leq j \leq p_n} |\check{S}_{j(n)}^2/\sigma_{j(n)}^2 - 1| = O_P(n^{-1/2})$ under the assumption of common variances. This means that the requirement that $\log p_n = o(n^{1/5})$ can be relaxed to $\log p_n = o(n^{1/4})$ for achieving uniform consistency of p -value ratios under sub-Gaussian tails, as discussed in Remark 5, even in the presence of normalization.

4.2. *Inference for marginal sample medians.* The effect of normalization on medians is trickier than its effect on means. The essence of our argument involves an assessment of how well $\tilde{F}_{j(n)}(t) \equiv n^{-1} \sum_{i=1}^n \mathbf{1}\{\tilde{X}_{ij(n)} \leq t\}$ approximates $\hat{F}_{j(n)}(t)$. We need the following strengthening of (10):

$$(15) \quad \limsup_{n \rightarrow \infty} \max_{1 \leq j \leq p_n} \|f_{j(n)}\|_\infty \leq \tilde{M},$$

for some $\tilde{M} < \infty$. The result below requires a precise bound on the modulus of continuity of Brownian motion (see Lemma 10 in Section 7 below).

THEOREM 3. Assume condition (15) holds for some $\tilde{M} < \infty$ and that $\log p_n = o(n^\gamma)$ for some $\gamma \in (0, 1]$. Then the following are true.

- (i) If $\hat{\epsilon}_n = o_P(1)$, then $\max_{1 \leq j \leq p_n} \|\tilde{F}_{j(n)} - \hat{F}_{j(n)}\|_\infty = o_P(1)$.
- (ii) If, in addition, $\gamma \in (0, 1/2]$ and $n^{1-\gamma}(\log n)\hat{\epsilon}_n = O_P(1)$, then also

$$\max_{1 \leq j \leq p_n} \|\tilde{F}_{j(n)} - \hat{F}_{j(n)}\|_\infty = o_P(n^{-(1-\gamma)}).$$

REMARK 8. Note that the signed rank test $R_{j(n)}$ studied in [18] can be written as a normalization of $\sqrt{n} \int_{\mathbb{R}} [\hat{F}_{j(n)}(u) - \hat{F}_{j(n)}(-u)] d\hat{F}_{j(n)}(u)$, while the sign test can be written as a normalization of $\sqrt{n} \int_{\mathbb{R}} \text{sign}(u) d\hat{F}_{j(n)}(u)$. Thus part (ii) of Theorem 3 allows us to replace $\hat{F}_{j(n)}$ with $\tilde{F}_{j(n)}$ in both of these statistics, $1 \leq j \leq p_n$, without destroying simultaneous consistency and validity of the normal p -value approximation assured by Theorems 1 and 2.

We now show that Theorem 3 can also be used to verify that the asymptotic results for the median tests of Section 3.3 hold for normalized data. For $j = 1, \dots, p_n$, define the approximate sample median $\check{\xi}_{j(n)} = \inf\{r : \tilde{F}_{j(n)}(r) \geq 1/2\}$ and $\check{f}_{j(n)} \equiv (\tilde{F}_{j(n)}(\check{\xi}_{j(n)} + h_n) - \tilde{F}_{j(n)}(\check{\xi}_{j(n)} - h_n))/(2h_n)$. Corollaries 10 and 11 validate the median approach after normalization.

COROLLARY 10. Assume there are no ties in the normalized data $\{\tilde{X}_{ij(n)}\}$.

(i) Under the conditions of Theorem 3, part (i), $\max_{1 \leq j \leq p_n} |\check{\xi}_{j(n)} - \hat{\xi}_{j(n)}| = o_P(1)$.

(ii) Assume the conditions of Theorem 3, part (ii), but with $\gamma \in (0, 1/3]$. Then $\max_{1 \leq j \leq p_n} |\check{\xi}_{j(n)} - \hat{\xi}_{j(n)}| = o_P(n^{-3(1-\gamma)/4})$.

(iii) Assume in addition the conditions of Corollary 7 with $\check{f}_{j(n)}$ replacing $\tilde{f}_{j(n)}$. Then $\max_{1 \leq j \leq p_n} |\check{f}_{j(n)} - \hat{f}_{j(n)}| = o_P(h_n^{-1}n^{-3(1-\gamma)/4})$.

COROLLARY 11. Assume the conditions of Corollary 10, part (iii), and the conditions of Corollary 8, except $\check{\pi}'_{j(n)} \equiv 2\Phi(-2\sqrt{n}\check{f}_{j(n)}|\check{\xi}_{j(n)} - \xi_{0,j(n)}|)$ replaces $\pi'_{j(n)}$, for all $1 \leq j \leq p_n$. Assume also that $n^{1-\gamma}(\log n)\hat{\epsilon}_n = O_P(1)$ in all cases. Then all of the conclusions of Corollary 8 follow with $\check{\pi}'_{(n)} \equiv \check{\pi}'_{1(n)}, \dots, \check{\pi}'_{p_n(n)}$ replacing $\hat{\pi}'_{(n)}$.

REMARK 9. Note that we require in some instances that $\hat{\epsilon}_n$ converge to zero faster than $O_P(n^{-1/2})$. This is possible since we can borrow strength across the $p_n \gg n$ genes to increase the normalization precision. We have thus shown that the normal approximation for p -values, even after normalization, can yield asymptotically valid FDR control as discussed in Section 2.

5. Numerical studies.

5.1. *One-sample simulation study.* We used a small simulation study to assess the finite sample performance of the following one-sample methodologies: (1) the mean-based comparison with unpooled variance estimate; (2) the mean-based comparison with pooled variance estimate; (3) the median-based comparison and (4) the signed rank test approach. We set the number of genes to $p = 2000$

and the number of arrays to $n = 10, 50$ and 100 . For simplicity, we assume the genes are independently distributed and the first 200 genes are differentially expressed. We consider the following marginal gene expression distributions: Models 1 and 3: $\text{Uniform}[-\sqrt{3}v + 1.5, \sqrt{3}v + 1.5]$ for differentially expressed genes and $\text{Uniform}[-\sqrt{3}v, \sqrt{3}v]$ otherwise. Models 2 and 4: $N(1.5, v)$ for differentially expressed genes and $N(0.0, v)$ otherwise. In Models 1 and 2, $v = 1.0$; in Models 3 and 4, $v \sim \text{Uniform}[0.25, 1.75]$ and marginal variances differ as much as seven-fold. Genes with marginal p -values < 0.001 are classified as differentially expressed. We compute \widehat{FDR} using the method in Section 2. For simplicity, we fix $l = 0.2$. We show in Table 1 the medians of the true FDR and \widehat{FDR} based on 200 replicates.

As sample size increases, the estimated FDRs become closer to the true FDRs for Models 1 and 2. For the pooled mean, median and signed rank approaches, the approach discussed in Section 2 provides a conservative control of the true FDR. For extremely small sample size ($n = 10$), the pooled mean approach is still valid, whereas the signed rank approach cannot identify any differentially expressed genes (denoted “NA” in Table 1). For the mean test without pooling, controlling the FDR with the approach given in Section 2 does not work well for small samples (especially under Models 2 and 3). As expected, pooling does not work when the variances are homogeneous (Models 3 and 4). With the ongoing

TABLE 1
One-sample simulation study results for the mean, pooled mean, median and signed rank statistics under Models 1, 2, 3 and 4. n : sample size

n	Mean		Pooled mean		Median		Signed rank	
	FDR	\widehat{FDR}	FDR	\widehat{FDR}	FDR	\widehat{FDR}	FDR	\widehat{FDR}
	Model 1							
10	0.113	0.008	0.005	0.010	0.190	0.035	NA	NA
50	0.019	0.009	0.009	0.009	0.008	0.009	0.000	0.010
100	0.012	0.009	0.007	0.009	0.009	0.012	0.000	0.010
	Model 2							
10	0.086	0.009	0.010	0.010	0.048	0.027	NA	NA
50	0.016	0.009	0.009	0.009	0.001	0.010	0.000	0.010
100	0.012	0.009	0.009	0.009	0.001	0.010	0.000	0.010
	Model 3							
10	0.123	0.009	0.016	0.010	0.143	0.027	NA	NA
50	0.020	0.009	0.024	0.009	0.005	0.009	0.000	0.010
100	0.015	0.009	0.029	0.009	0.012	0.009	0.000	0.010
	Model 4							
10	0.090	0.009	0.026	0.010	0.040	0.022	NA	NA
50	0.015	0.009	0.024	0.009	0.000	0.010	0.000	0.010
100	0.010	0.009	0.024	0.009	0.000	0.010	0.000	0.010

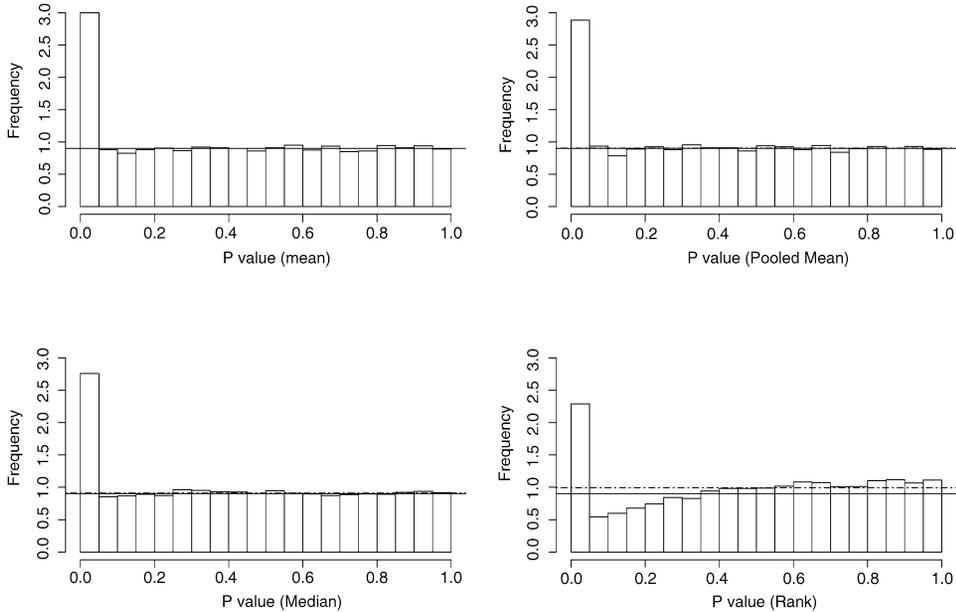


FIG. 1. *Simulation study: plots of marginal p -values. The solid horizontal line is the true proportion of nondifferentially expressed genes; the dotted line is the estimate.*

development of microarray technology, a typical study may have more than 50 arrays. Thus FDR control based on the mean, pooled mean (when pooling is justified), median and rank approaches will often be valid. Simulations with correlated expression levels and other marginal distributions yield similar conclusions.

We show in Figure 1 the histogram of the marginal p -values for a single replication from Model 1 with $n = 50$. For the mean, pooled mean and median approaches, the distributions seem to be a mixture of a uniform and a point mass at zero; the marginal p -value distribution for nondifferentially expressed genes for the rank approach is also close to uniform. With the rank approach, the proportion of nondifferentially expressed genes is overestimated. Other diagnostics show that the FDR approach is well behaved.

5.2. Estrogen data. These datasets were first presented in [28] and [22]. Their common expression matrix monitors 7129 genes in 49 breast tumor samples. The data were obtained by applying the Affymetrix gene chip technology. The response describes the lymph nodal (LN) status, an indicator of disease severity. 25 samples are positive and 24 samples are negative. The goal is to identify genes differentially expressed between positive and negative samples from the 3332 genes passing the first step of processing described in [5]. A base 2 log transform of the gene expressions is first applied. We compute marginal p -values based on the mean (without pooling), pooled mean and median approaches, and the Wilcoxon

test, which is a two-sample version of the signed rank test. See [19] for additional details on the two-sample methods. We use $l = 0.2$ for tuning and genes with p -values < 0.001 are considered differentially expressed. With the four approaches, 196 (0.012), 237 (0.011), 68 (0.040) and 156 (0.016) genes are identified, respectively, with the numbers in parentheses being the estimated FDRs. There are reasonable overlaps between the genes identified with the different approaches.

6. Discussion. The main results of this paper are that marginal (gene specific) estimates and asymptotic-based p -values are uniformly consistent in normalized microarray experiments with n replications—regardless of the dependencies between genes—provided the number of genes p_n satisfies $\log p_n = o(n^\gamma)$, for a suitable $\gamma \in (0, 1]$, depending on the setting. In other words, the number of genes can increase almost exponentially fast relative to the number of arrays. This seems to be realistic for microarray studies. Note that p_n can be even larger for certain distribution free statistics (see [18]). These results also hold for two-sample comparisons (see [19]).

We note that the numerical studies seem to support the theoretical results of the paper, although some procedures work better than others. For certain statistics, the required sample sizes are too large to be practical without pooling or some other method of borrowing strength across genes (see [9]). We acknowledge that a number of other issues, such as the effects of marginal distributions and normalization, were not evaluated in Section 5. A more thorough simulation study addressing these points would be useful. A theoretical limitation is that the asymptotics developed are not yet accurate enough to provide precise guidelines on sample size. This pursuit will likely require some assumptions on the dependencies between genes. Such assumptions are out of place in the present paper since a strength of the paper is the absence of assumptions on gene interdependence. Because of this generality, the results of this paper should be a useful point of departure for future, more refined asymptotic analyses of microarray experiments.

7. Proofs.

PROOF OF LEMMA 2. Since $E_{2p}^{(\alpha)}(\hat{q}_{(p)}, q_{(p)}) \rightarrow 0$ in probability for every $\alpha \in (0, 1]$, there exist decreasing sequences $0 < \alpha_p, \varepsilon_p \downarrow 0$ with $\limsup_{p \rightarrow \infty} P(E_{2p}^{(\alpha_p)}(\hat{q}_{(p)}, q_{(p)}) > \varepsilon_p) = 0$ and $\limsup_{p \rightarrow \infty} P(|\hat{\lambda}(l)/\tilde{\lambda}(l) - 1| > \varepsilon_p) = 0$. Thus $\mathbf{1}\{q_j \leq t(1 - \varepsilon_p)\} \leq \mathbf{1}\{\hat{q}_j \leq t\} \leq \mathbf{1}\{q_j \leq t(1 + \varepsilon_p)\}$ for all $t \in [2t_p, 1]$ and all $1 \leq j \leq p$, with large probability. Hence also

$$\begin{aligned} \widetilde{FDR}_l(t(1 + \tilde{\varepsilon}_p(t))) \frac{1 - \varepsilon_p}{1 + \tilde{\varepsilon}_p(t)} &\leq \widehat{FDR}_l(t) \\ &\leq \widetilde{FDR}_l(t(1 - \varepsilon_p)) \frac{1 + \varepsilon_p}{1 - \varepsilon_p}, \end{aligned}$$

for all n large enough, where $\tilde{\varepsilon}_p(t) \equiv \varepsilon_p \wedge (t^{-1} - 1)$. Hence by condition (2),

$$(16) \quad \begin{aligned} & \sup_{t \in [2t_p, 1]} |\widehat{FDR}_l(t) - r_{(p)}(t)| \\ & \leq \sup_{t \in [2t_p, 1]} |r_{(p)}(t(1 + \tilde{\varepsilon}_p(t))) - r_{(p)}(t(1 - \varepsilon_p))| + o_P(1). \end{aligned}$$

Note that for any $g : [0, 1] \mapsto [0, 1]$ with derivative \dot{g} , we have for all $t \in [0, 1]$ and $\varepsilon \in (-1, 1 \wedge (t^{-1} - 1))$, that $|g(t(1 + \varepsilon)) - g(t)| \leq \varepsilon(1 - \varepsilon)^{-1} \sup_{t \in [0, 1]} |t \dot{g}(t)|$. Applying this to $t \mapsto r_{(p)}(t)$, with derivative $\dot{r}_{(p)}$, we obtain that the right-hand side of (16) = $o_P(1)$, provided $|t \dot{r}_{(p)}(t)|$ is uniformly bounded over all $t \in [0, 1]$ and $p \geq 1$. The provision holds since $0 \leq F_1(t) - t f_1(t) \leq F_1(t)$ by concavity of F_1 , where f_1 is the derivative of F_1 , and thus $t \dot{r}_{(p)}(t) = \lambda_p t (1 - \lambda_p) (F_1(t) - t f_1(t)) (\lambda_p t + (1 - \lambda_p) F_1(t))^{-2} \leq 1/4$.

Now note that for any $p > 2$, $r_{(p)}$ is strictly increasing, $r_{(p)}(1) = 1$ and

$$(17) \quad r_{(p)}(2t_p) = \left(1 + \frac{(1 - \lambda_p) F_1(2t_p)}{\lambda_p 2t_p} \right)^{-1} \rightarrow 0,$$

since $F_1(2p^{-1}) / (2p^{-1})^\delta \rightarrow \infty$ by assumption and $t_p^{-(1-\delta)} = (1 - \lambda_p)^{-1}$. Next note that the sequence of solutions t_p^* of $r_{(p)}(t_p^*) = \lambda_0 \rho / \lambda_*$ is unique and satisfies $2t_p \leq t_p^* \leq 1$ for all p large enough. These facts, combined with the fact that (16) = $o_P(1)$, now force both $\widehat{FDR}_l(2t_p) = o_P(1)$ and $\widehat{FDR}_l(t_p^*) = \rho + o_P(1)$. This now implies, with probability tending to 1, that $\hat{t}_p \geq 2t_p$ for all p large enough. Hence $r_{(p)}(\hat{t}_p) \vee (\lambda_0 / \lambda_*) \rho = (\lambda_0 / \lambda_*) \rho + o_P(1)$, and we are done since $\sup_{t \in [t_p, 1]} |FDR_{(p)}(t) - r_{(p)}(t)| \rightarrow 0$ via Remark 1. \square

PROOF OF LEMMA 3. Let \tilde{F}_{1p} be the empirical distribution function of the p -values associated with alternative hypotheses. Then (2) will follow provided $\sup_{t \in [t_p, 1]} |t^{-1} \tilde{G}_p(t) - 1| = o_P(1)$ and $\sup_{t \in [t_p, 1]} |\tilde{F}_{1p}(t) / F_1(t) - 1| = o_P(1)$. This holds by Lemma 5 below since $pt_p \rightarrow \infty$ and $pF_1(t_p) \rightarrow \infty$, as $p \rightarrow \infty$. \square

LEMMA 5. Let X_1, \dots, X_n be i.i.d. with continuous distribution F with support on $[0, 1]$. Let \hat{F}_n be the empirical distribution, and let $\{u_n\} \in [0, 1]$ satisfy $nF(u_n) \rightarrow \infty$. Then $\sup_{u \in [u_n, 1]} |\hat{F}_n(u) / F(u) - 1| = o_P(1)$.

PROOF. By the continuity of F , we can assume that $F(u) = u$ without loss of generality since $F(X)$ is uniform. Let $s_n = nu_n/2$, and note that

$$P(n\hat{F}_n(u_n) \geq s_n) = P\left(\frac{n\hat{F}_n(u_n) - nu_n}{\sqrt{nu_n}} \geq \frac{s_n - nu_n}{\sqrt{nu_n}} \right) \rightarrow 1,$$

as $n \rightarrow \infty$, since $(s_n - nu_n) / \sqrt{nu_n} \rightarrow \infty$ and $(nu_n)^{-1/2} (n\hat{F}_n(u_n) - nu_n) = O_P(1)$. Thus $\sup_{u \in [u_n, 1]} |\hat{F}_n(u) / u - 1| \leq \sup_{u \in [0, 1] : n\hat{F}_n(u) \geq s_n} |\hat{F}_n(u) / u - 1| +$

$o_P(1)$, and the desired result now follows from Theorem 2 of [17] since \hat{F}_n is an uncensored Kaplan–Meier estimator under a reversed time scale. \square

PROOF OF THEOREM 1. Define $V_{j(n)} \equiv \sqrt{n} \|\hat{F}_{j(n)} - F_{j(n)}\|_\infty$, and note that by Corollary 1 of [20], $P(V_{j(n)} > x) \leq 2e^{-2x^2}$, for all $x \geq 0$ and any distribution $F_{j(n)}$. This inequality is a refinement of the celebrated result of Dvoretzky, Kiefer and Wolfowitz [6], given in their Lemma 2, and the extension to distributions with discontinuities is standard. Using Lemma 6 below, we obtain $\|V_{j(n)}\|_{\psi_2} \leq \sqrt{3/2}$ for all $1 \leq j \leq p_n$. Now, by Lemma 7 below combined with the fact that $\limsup_{x,y \rightarrow \infty} \psi_2(x)\psi_2(y)/\psi_2(xy) = 0$, we have that there exists a universal constant $c_* < \infty$ with $\|\max_{1 \leq j \leq p_n} V_{j(n)}\|_{\psi_2} \leq c_* \sqrt{\log(1 + p_n)} \sqrt{3/2}$ for all $n \geq 1$. The desired result now follows for the constant $c_0 = \sqrt{6}c_*$, since $\log(k + 1) \leq 2 \log k$ for any $k \geq 2$. \square

LEMMA 6. *Let X be a random variable with $P(|X| > x) \leq Ke^{-Cx^p}$ for every $x > 0$ and constants K, C and $p \geq 1$. Then $\|X\|_{\psi_p} \leq ((1 + K)/C)^{1/p}$.*

LEMMA 7. *Let ψ be a convex, nondecreasing, nonzero function with $\psi(0) = 0$ and $\limsup_{x,y \rightarrow \infty} \psi(x)\psi(y)/\psi(cxy) < \infty$ for a finite constant c . Then, for any random variables X_1, \dots, X_m , $\|\max_{1 \leq i \leq m} X_i\|_\psi \leq K\psi^{-1}(m) \times \max_{1 \leq i \leq m} \|X_i\|_\psi$, for a constant K depending only on ψ .*

PROOF OF LEMMAS 6 AND 7. These are Lemmas 2.2.1 and 2.2.2, respectively, of [26], and the proofs can be found therein. \square

PROOF OF THEOREM 2. Let $U_j, j = 1, \dots, p_n$, be i.i.d. uniform random variables independent of the data. Then, by Theorem 4 below, we have for each $1 \leq j \leq p_n$ that there exists a measurable map $g_{j(n)} : \mathbb{R}^n \times [0, 1] \mapsto C[0, 1]$ where $B_{j(n)} = g_{j(n)}(X_{1j(n)}, \dots, X_{nj(n)}, U_j)$ is a Brownian bridge with

$$(18) \quad P(\sqrt{n} \|\sqrt{n}(\hat{F}_{j(n)} - F_{j(n)}) - B_{j(n)}(F_{j(n)})\|_\infty > x + 12 \log n) \leq 2e^{-x/6},$$

for all $x \geq 0$. Note that this construction generates an ensemble of Brownian bridges $B_{1(n)}, \dots, B_{p_n(n)}$ that may be dependent when the components in $X_{1(n)} = (X_{11(n)}, \dots, X_{1p_n(n)})'$ are dependent. However, each $B_{j(n)}$ only depends on the information contained in $\mathcal{F}_{j(n)}$ and the independent uniform random variable U_j . Thus $B_{j(n)}$ depends on \mathcal{F}_n only through the information contained in $\mathcal{F}_{j(n)}$, and the ensemble of Brownian bridges is conditionally independent given \mathcal{F}_n . Note also the validity of (18) for all $n \geq 2$.

Define $V_{j(n)} = ((\sqrt{n}/(\log n)) \|\sqrt{n}(\hat{F}_{j(n)} - F_{j(n)}) - B_{j(n)}(F_{j(n)})\|_\infty - 12)^+$, where u^+ is the positive part of u . By Lemma 6, expression (18) implies that $\|V_{j(n)}\|_{\psi_1} \leq 18/\log n$. Reapplying the result that $\log(k + 1) \leq 2 \log k$ for any

$k \geq 2$, we now have, by the fact that $\limsup_{x,y \rightarrow \infty} \psi_1(x)\psi_1(y)/\psi_1(xy) = 0$ combined with Lemma 7, that there exists a universal constant $0 < c_2 < \infty$ for which $\|\max_{1 \leq j \leq p_n} V_{j(n)}\|_{\psi_1} \leq c_2 \log p_n / (\log n)$. Now (4) follows, for $c_1 = 12$, from the definition of $V_{j(n)}$. \square

THEOREM 4. *For $n \geq 2$, let Y_1, \dots, Y_n be i.i.d. real random variables with distribution G (not necessarily continuous), and let U_0 be a uniform random variable independent of Y_1, \dots, Y_n . Then there exists a measurable map $g_n: \mathbb{R}^n \times [0, 1] \mapsto C[0, 1]$ such that $B = g_n(Y_1, \dots, Y_n, U_0)$ is a standard Brownian bridge satisfying, for all $x \geq 0$,*

$$(19) \quad P(\sqrt{n} \|\sqrt{n}(\hat{G}_n - G) - B(G)\|_\infty > x + 12 \log n) \leq 2e^{-x/6},$$

where \hat{G}_n is the empirical distribution of Y_1, \dots, Y_n .

PROOF. By Theorem 20.4 of [2], there exists a measurable $h_0: [0, 1] \mapsto [0, 1]^2$ such that $(U_1, U_2) \equiv h_0(U_0)$ is a pair of independent uniforms. Moreover, standard arguments yield the existence of a function $h_n: \mathbb{R}^n \times [0, 1] \mapsto [0, 1]^n$ such that $(V_1, \dots, V_n) \equiv h_n(Y_1, \dots, Y_n, U_1)$ is a sample of i.i.d. uniforms and $(Y_1, \dots, Y_n) = (\psi(V_1), \dots, \psi(V_n))$, where $\psi(u) \equiv \inf\{x: G(x) \geq u\}$. U_1 is needed to handle possible discontinuities in G .

Let \hat{H}_n be the empirical distribution for V_1, \dots, V_n , and note that

$$(20) \quad \sqrt{n}(\hat{H}_n(G(x)) - G(x)) = \sqrt{n}(\hat{G}_n(x) - G(x)) \quad \forall x \in \mathbb{R}$$

by design. Now by the Hungarian construction (Theorem 1) of Bretagnolle and Massart [3], there exists a Brownian bridge B depending only on V_1, \dots, V_n and U_2 such that $P(\sqrt{n} \sup_{u \in [0,1]} |\sqrt{n}(\hat{H}_n(u) - u) - B(u)| > x + 12 \log n) \leq 2e^{-x/6}$, for all $x \geq 0$, and thus by (20),

$$(21) \quad P(\sqrt{n} \|\sqrt{n}(\hat{G}_n - G) - B(G)\|_\infty > x + 12 \log n) \leq 2e^{-x/6} \quad \forall x \geq 0.$$

By Lemma 8 below, we can take B to be $f_n(V_1, \dots, V_n, U_2)$, where $f_n: [0, 1]^{n+1} \mapsto D[0, 1]$ is measurable and $D[0, 1]$ has the Skorohod rather than uniform metric, since both $t \mapsto \sqrt{n}(\hat{H}_n(t) - t)$ and $t \mapsto B(t)$ are Borel measurable on the Skorohod space $D[0, 1]$. Since $P(B \in C[0, 1]) = 1$, and since the uniform and Skorohod metrics are equivalent on $C[0, 1]$, we now have that f_n is also measurable with respect to the uniform topology. Thus the map $g_n: \mathbb{R}^n \times [0, 1] \mapsto C[0, 1]$ defined by the composition $(Y_1, \dots, Y_n, U_0) \mapsto (V_1, \dots, V_n, U_2) \mapsto B$ is Borel measurable, and (19) follows. \square

LEMMA 8. *Given two random elements X and Y in a separable metric space \mathbb{X} , there exist a Borel measurable $f: \mathbb{X} \times [0, 1] \mapsto \mathbb{X}$ and a uniform random variable Z independent of X such that $Y = f(X, Z)$ almost surely.*

PROOF. The result and proof are given in [21]. While Skorohod’s paper does not specify uniformity of Z , this readily follows without loss of generality. \square

PROOF OF COROLLARY 1. Apply Theorem 1 and the identity

$$(22) \quad \int_{[a_{j(n)}, b_{j(n)}]} x [d\hat{F}_{j(n)}(x) - dF_{j(n)}(x)] \\ = - \int_{[a_{j(n)}, b_{j(n)}]} [\hat{F}_{j(n)}(x) - F_{j(n)}(x)] dx. \quad \square$$

PROOF OF COROLLARY 2. Let $U_{j(n)} = (\log p_n + \log n)^{1/r} \vee \max_{1 \leq i \leq n} |X_{ij(n)} - \mu_{j(n)}|$, for $1 \leq j \leq p_n$, and note that the integration by parts formula yields

$$|\bar{X}_{j(n)} - \mu_{j(n)}| \leq 2U_{j(n)} \|\hat{F}_{j(n)} - F_{j(n)}\|_\infty \\ + \int_{U_{j(n)}}^\infty \mathbb{P}(|X_{1j(n)} - \mu_{j(n)}| > u) du \\ = O_P(\log n + \log p_n)^{1/p} \times O_P\left(\sqrt{\frac{\log p_n}{n}}\right) + O_P(np_n)^{-k_2},$$

by Lemma 7, Theorem 1, and by the definition of $U_{j(n)}$. \square

PROOF OF COROLLARY 3. Let $S_{j(n)}^2$ be the sample variance version with n in the denominator, and let $\tilde{S}_{j(n)}^2$ be the version with denominator $n - 1$. Then $\tilde{S}_{j(n)}^2/\sigma_{j(n)}^2 - 1 = O(n^{-1}) + (1 + o(1))(S_{j(n)}^2/\sigma_{j(n)}^2 - 1)$ and thus we can assume the denominator is n after adding the term $O(n^{-1})$. Note that

$$\left| \frac{S_{j(n)}^2}{\sigma_{j(n)}^2} - 1 \right| \leq \left| n^{-1} \sum_{i=1}^n \frac{(X_{ij(n)} - \mu_{j(n)})^2}{\sigma_{j(n)}^2} - 1 \right| + \left(\frac{\bar{X}_{j(n)} - \mu_{j(n)}}{\sigma_{j(n)}} \right)^2.$$

For part (i), we apply Corollary 1 twice, once for the data $(X_{ij(n)} - \mu_{j(n)})^2/\sigma_{j(n)}^2$ and once for the data $(X_{ij(n)} - \mu_{j(n)})/\sigma_{j(n)}$. This gives us

$$\max_{1 \leq j \leq p_n} \left| \frac{S_{j(n)}^2}{\sigma_{j(n)}^2} - 1 \right| \leq O_P\left(\sqrt{\frac{\log p_n}{n}} d_n^2 + \frac{\log p_n}{n} d_n^2\right) \\ = o_P(d_n^2 n^{\gamma/2-1/2}),$$

since $n^{\gamma/2-1/2} = o(1)$ by assumption. This yields part (i).

For part (ii), we apply Corollary 2 twice instead of Corollary 1, once for the data $(X_{ij(n)} - \mu_{j(n)})^2/\sigma_{j(n)}^2$ with $r = 1$, and once for the data $(X_{ij(n)} - \mu_{j(n)})/\sigma_{j(n)}$

with $r = 2$. This immediately yields

$$\begin{aligned} \max_{1 \leq j \leq p_n} \left| \frac{S_{j(n)}^2}{\sigma_{j(n)}^2} - 1 \right| &\leq O_P \left(\sqrt{\frac{\log p_n}{n}} (\log p_n + \log n) + (np_n)^{-k_2} \right) \\ &\quad + O_P \left(\sqrt{\frac{\log p_n}{n}} (\log p_n + \log n)^{1/2} + (np_n)^{-k_2} \right)^2 \\ &= o_P(n^{3\gamma/2-1/2} + n^{2\gamma-1}) + O_P((np_n)^{-k_2} + (np_n)^{-2k_2}), \end{aligned}$$

which yields the desired result since $\gamma \in (0, 1]$. \square

PROOF OF COROLLARY 4. We begin by establishing the results for $E_{1p_n}(\pi_{(n)}, \hat{\pi}_{(n)})$ [result (i)] for both parts (a) and (b). Note that for any $x \in \mathbb{R}$ and any $y > 0$, $|\Phi(xy) - \Phi(x)| \leq 0.25 \times (|1 - y| \vee |1 - 1/y|)$. Thus

$$\begin{aligned} (23) \quad &\max_{1 \leq j \leq p_n} |\hat{\pi}_{j(n)} - \hat{\pi}_{j(n)}^*| \\ &\leq \frac{1}{2} \left(\max_{1 \leq j \leq n} (\hat{\sigma}_{j(n)} \vee \sigma_{j(n)}) \left| \frac{1}{\hat{\sigma}_{j(n)}} - \frac{1}{\sigma_{j(n)}} \right| \right) \\ &= O_P(E_{0n}^{1/2}), \end{aligned}$$

where $\hat{\pi}_{j(n)}^* \equiv 2\Phi(-|T_{j(n)}^*|)$ and $T_{j(n)}^* \equiv \sqrt{n}(\bar{X}_{j(n)} - \mu_{0,j(n)})/\sigma_{j(n)}$. Now, for part (a) Theorem 2 yields

$$\begin{aligned} &\max_{1 \leq j \leq p_n} |\hat{\pi}_{j(n)}^* - \pi_{j(n)}| \\ &= O_P \left(\frac{c_1 \log n + c_2 \log p_n}{\sqrt{n}} \times \max_{1 \leq j \leq p_n} \frac{|b_{j(n)} - a_{j(n)}|}{\sigma_{j(n)}} \right) \\ &= o_P(n^{\gamma-1/2}d_n), \end{aligned}$$

and thus $E_{1p_n}(\pi_{j(n)}, \hat{\pi}_{j(n)}) = O_P(E_{0n}^{1/2}) + o_P(n^{\gamma-1/2}d_n)$.

For part (b), let $U_n \equiv n^{\gamma/2} \vee (\max_{1 \leq j \leq p_n, 1 \leq i \leq n} |X_{ij(n)} - \mu_{j(n)}|/\sigma_{j(n)})$ and note that $U_n = n^{\gamma/n}$ for large n with high probability, since the maximum over all $|X_{ij(n)} - \mu_{j(n)}|/\sigma_{j(n)}$ is $O_P(\sqrt{\log p_n + \log n}) = o_P(n^{\gamma/2})$ by the sub-Gaussian tail assumption. Now let the empirical distribution of the data $(X_{ij(n)} - \mu_{j(n)})/\sigma_{j(n)}$, for $i = 1, \dots, n$, be denoted $\hat{F}_{j(n)}^0$, and let $F_{j(n)}^0$ denote the distribution of $(X_{1j(n)} - \mu_{j(n)})/\sigma_{j(n)}$. We now have

$$\begin{aligned} (24) \quad &\frac{\sqrt{n}(\bar{X}_{j(n)} - \mu_{j(n)})}{\sigma_{j(n)}} \\ &= -\sqrt{n} \int_{\mathbb{R}} (\hat{F}_{j(n)}^0(x) - F_{j(n)}^0(x)) dx \end{aligned}$$

$$\begin{aligned}
 &= -\sqrt{n} \int_{-r_n}^{r_n} (\hat{F}_j^0(x) - F_{j(n)}^0(x)) dx \\
 &\quad + \sqrt{n} \int_{r_n}^{\infty} (F_{j(n)}^0(-x) - (1 - F_{j(n)}^0(x))) dx + R_n,
 \end{aligned}$$

where R_n is only nonzero when $U_n \neq n^{\gamma/2}$, but this latter condition occurs with probability $\rightarrow 0$ as $n \rightarrow \infty$. Hence, $R_n = o_P(n^{-1})$ at least. Thus

$$(24) = - \int_{-r_n}^{r_n} B_{j(n)}^0(F_{j(n)}^0(x)) dx + o_P(\log p_n n^{\gamma/2-1/2}) + O(\sqrt{n} k_1 e^{-k_2 n^\gamma})$$

by Theorem 2 applied to $\sqrt{n}(\hat{F}_{j(n)}^0 - F_{j(n)}^0)$, where $B_{j(n)}^0$ is the associated Brownian bridge approximation and the errors do not depend on j . Hence

$$\begin{aligned}
 (24) &= - \int_{\mathbb{R}} B_{j(n)}^0(F_{j(n)}^0(x)) dx \\
 &\quad + \int_{|x| > n^{\gamma/2}} B_{j(n)}^0(F_{j(n)}^0(x)) dx + o_P(\log p_n n^{\gamma/2-1/2}).
 \end{aligned}$$

However, $\int_{|x| > n^{\gamma/2}} B_{j(n)}^0(F_{j(n)}^0(x)) dx$ is mean zero normal with variance

$$\begin{aligned}
 &\int_{|x| > n^{\gamma/2}} \int_{|y| > n^{\gamma/2}} [F_{j(n)}^0(x \wedge y) - F_{j(n)}^0(x)F_{j(n)}^0(y)] dx dy \\
 &= 2 \int_{-\infty}^{-n^{\gamma/2}} \int_{-\infty}^y F_{j(n)}^0(x) dx \bar{F}_{j(n)}^0(x) dy + 2 \int_{n^{\gamma/2}}^{\infty} \int_y^{\infty} \bar{F}_{j(n)}^0(x) dx F_{j(n)}^0(y) dy \\
 &\quad + 2 \int_{-\infty}^{-n^{\gamma/2}} F_{j(n)}^0(x) dx \int_{n^{\gamma/2}}^{\infty} \bar{F}_{j(n)}^0(x) dx \\
 &\leq 2 \int_{n^{\gamma/2}}^{\infty} \int_y^{\infty} k_1 e^{-k_2 x^2} dx dy + 2 \left(\int_{n^{\gamma/2}}^{\infty} k_1 e^{-k_2 x^2} dx \right)^2 \\
 &= o(e^{-k_2 n^\gamma}),
 \end{aligned}$$

where $\bar{F}_{j(n)}^0 \equiv 1 - F_{j(n)}^0$. We now have

$$(25) \quad \frac{\sqrt{n}(\bar{X}_{j(n)} - \mu_{j(n)})}{\sigma_{j(n)}} = Z_{j(n)} + o_P(\log p_n n^{\gamma/2-1/2}),$$

where $Z_{j(n)} \equiv - \int_{\mathbb{R}} B_{j(n)}^0(F_{j(n)}^0(x)) dx$ can be easily shown to be standard normal, and the error term again does not depend on j . Thus $E_{1p_n}(\pi_{(n)}, \hat{\pi}_{(n)}) = O_P(E_{0n}^{1/2}) + o_P(\log p_n n^{\gamma/2-1/2}) = O_P(E_{0n}^{1/2}) + o_P(n^{3\gamma/2-1/2})$.

We now establish the results for $E_{2p_n}^{(\omega)}(\pi_{(n)}, \hat{\pi}_{(n)})$ [result (ii)] for both parts (a) and (b). Let $\bar{\Phi}(x)$ be the standard normal tail probability at x , let $\tilde{T}_{j(n)} = Z_{j(n)} + \sqrt{n}(\mu_{j(n)} - \mu_{0,j(n)})/\sigma_{j(n)}$, and define $D_{0n} \equiv \max_{1 \leq j \leq p_n} |T_{j(n)}^* - \tilde{T}_{j(n)}|$.

Then, by Lemma 9 below, we have $\bar{\Phi}(|T_{j(n)}|)/\bar{\Phi}(|\tilde{T}_{j(n)}|) \rightarrow 1$, provided $T_{j(n)}^2 - \tilde{T}_{j(n)}^2 \rightarrow 0$. Note that we only need the ratios of the p -values to converge when $2\bar{\Phi}(|\tilde{T}_{j(n)}|) > \alpha/(2p_n)$, implying $|\tilde{T}_{j(n)}| \leq \sqrt{2\log(4p_n/(\sqrt{2\pi}\alpha))}$, for all n large enough, since $\bar{\Phi}(z) \leq \phi(z)/z$ for $z > 0$. Since, for $1 \leq j \leq p_n$,

$$(26) \quad |T_{j(n)}^2 - \tilde{T}_{j(n)}^2| \leq E_{0n} \tilde{T}_{j(n)}^2 + \frac{\sigma_{j(n)}^2}{\hat{\sigma}_{j(n)}^2} (2D_{0n} |\tilde{T}_{j(n)}| + D_{0n}^2),$$

$$\max_{1 \leq j \leq p_n} |T_{j(n)}^2 - \tilde{T}_{j(n)}^2| \leq o_P(n^\gamma E_{0n} + n^{\gamma/2} D_{0n}) + O_P(D_{0n}^2).$$

For part (a), Theorem 2 yields that $D_{0n} = O_P(\log p_n n^{-1/2} d_n)$, and the right-hand side of (26) becomes $o_P(n^\gamma E_{0n} + n^{3\gamma/2-1/2} d_n + n^{2\gamma-1} d_n^2)$. For part (b), we use (25) to obtain that the right-hand side of (26) becomes $o_P(n^\gamma E_{0n} + n^{2\gamma-1/2})$. We are done if $\max_{1 \leq j \leq p_n} \mathbf{1}\{\pi_{j(n)} \leq \alpha/(2p_n), \hat{\pi}_{j(n)} > \alpha/p_n\} = 0$ with high probability as $n \rightarrow \infty$. Assume without loss of generality that $p_n \rightarrow \infty$, since otherwise the proof is trivial. Now, since $\bar{\Phi}(x) \geq x^{-1}\phi(x)(1 - x^{-2})$ for $x > 0$, we have for large n that $2\bar{\Phi}(|\tilde{T}_{j(n)}|) \leq \alpha/(2p_n)$ implies $\tilde{T}_{j(n)}^2/2 + \log |\tilde{T}_{j(n)}| \geq \log(4p_n/(\sqrt{2\pi}\alpha)) - o(1)$, where $o(1)$ does not depend on j . Thus by (26),

$$(27) \quad (\log p_n) \max_{1 \leq j \leq p_n} \left| \frac{T_{j(n)}^2}{\tilde{T}_{j(n)}^2} - 1 \right| \leq o_P(1 + n^\gamma E_{0n} + n^{\gamma/2} D_{0n}).$$

This means that

$$(28) \quad \begin{aligned} \frac{2\bar{\Phi}(|T_{j(n)}|)}{\alpha/p_n} &\leq \exp\left(-\frac{T_{j(n)}^2}{2} - \log |T_{j(n)}| + \log\left(\frac{2p_n}{\sqrt{2\pi}\alpha}\right)\right) \\ &\leq \exp\left(\log\left(\frac{2p_n}{\sqrt{2\pi}\alpha}\right)\right) \\ &\quad - \log\left(\frac{4p_n}{\sqrt{2\pi}\alpha}\right) + o_P(1 + n^\gamma E_{0n} + n^{\gamma/2} D_{0n}) \\ &= \exp(-\log 2 + o_P(1 + n^\gamma E_{0n} + n^{\gamma/2} D_{0n})), \end{aligned}$$

where again the error terms do not depend on j . Now previous results indicate that for the conditions under either part (a) or (b), the left-hand side of (28) is bounded uniformly over $1 \leq j \leq p_n$ by $1/2 + o_P(1)$, and the desired conclusions follow. \square

LEMMA 9. For $x_n, y_n \geq 0$, $\bar{\Phi}(x_n)/\bar{\Phi}(y_n) \rightarrow 1$ if and only if $x_n^2 - y_n^2 \rightarrow 0$.

PROOF. Assume $x_n^2 - y_n^2 \rightarrow 0$. Without loss of generality, $x_n \rightarrow c \in [0, \infty]$. If $c < \infty$, obviously $\bar{\Phi}(x_n)/\bar{\Phi}(y_n) \rightarrow 1$. If $c = \infty$, $\bar{\Phi}(x_n)/\bar{\Phi}(y_n) - \phi(x_n)/\phi(y_n) \rightarrow 0$ by a standard approximation for $\bar{\Phi}$, and $\bar{\Phi}(x_n)/\bar{\Phi}(y_n) \rightarrow 1$ again.

Now assume $\bar{\Phi}(x_n)/\bar{\Phi}(y_n) \rightarrow 1$. Again without loss of generality, $x_n \rightarrow c \in [0, \infty]$. If $c < \infty$, obviously $x_n^2 - y_n^2 \rightarrow 0$. But $c = \infty$ implies $y_n \rightarrow \infty$. Since $\bar{\Phi}(x_n)/\bar{\Phi}(y_n) - y_n\phi(x_n)/(x_n\phi(y_n)) = o(1)$, the result follows. \square

PROOF OF COROLLARY 5. From (26) and (28), we know that the desired result will follow provided $n^\gamma E_{0n} + n^{\gamma/2} D_{0n} = O_P(1)$. Since $D_{0n} = o_P(n^{3\gamma/2-1/2})$ by (25), it suffices to verify that $E_{0n} = O_P(n^{-\gamma})$ for any $\gamma \in (0, 1/4]$. This will certainly hold if $E_{0n} = O_P(n^{-1/2})$. Now,

$$\begin{aligned} & E|\hat{\sigma}^2/\sigma^2 - 1| \\ & \leq O\left[\frac{1}{n} + p_n^{-1} \sum_{j=1}^{p_n} E\left|n^{-1} \sum_{i=1}^n \frac{(X_{ij(n)} - \mu_{j(n)})^2}{\sigma^2} - 1 - \left(\frac{\bar{X}_{j(n)} - \mu_{j(n)}}{\sigma}\right)^2\right|\right] \\ & \leq O\left[\frac{1}{n} + \sqrt{\max_{1 \leq j \leq p_n} E\left(n^{-1} \sum_{i=1}^n \frac{(X_{ij(n)} - \mu_{j(n)})^2}{\sigma^2} - 1\right)^2}\right] \\ & = O(n^{-1/2}), \end{aligned}$$

since $E[(X_{ij(n)} - \mu_{j(n)})^2/\sigma^2 - 1]^2$ is uniformly bounded. \square

PROOF OF COROLLARY 6. That the left-hand side of (9) is $o_P(1)$ follows from condition (8) combined with Theorem 1. By the definition of the sample median, we have that $\hat{F}_{j(n)}(\hat{\xi}_{j(n)}) - F_{j(n)}(\xi_{j(n)}) \equiv H_{j(n)}$, where $|H_{j(n)}| \leq 1/n$. This now implies that $\hat{F}_{j(n)}(\hat{\xi}_{j(n)}) - F(\hat{\xi}_{j(n)}) + F(\hat{\xi}_{j(n)}) - F(\xi_{j(n)}) = H_{j(n)}$. The result now follows from the mean value theorem and condition (8). \square

PROOF OF COROLLARY 7. We have $2\sqrt{n}h_n E'_{0n} \leq \max_{1 \leq j \leq p_n} \sqrt{n} \times |F_{j(n)}(\hat{\xi}_{j(n)} + h_n) - F_{j(n)}(\xi_{j(n)} + h_n)| + \max_{1 \leq j \leq p_n} \sqrt{n} |F_{j(n)}(\hat{\xi}_{j(n)} - h_n) - F_{j(n)}(\xi_{j(n)} - h_n)| + \max_{1 \leq j \leq p_n} \sqrt{n} |F_{j(n)}(\xi_{j(n)} + h_n) - F_{j(n)}(\xi_{j(n)} - h_n) - 2f_{j(n)}(\xi_{j(n)})h_n| + \max_{1 \leq j \leq p_n} \sqrt{n} |\hat{F}_{j(n)}(\hat{\xi}_{j(n)} + h_n) - F_{j(n)}(\hat{\xi}_{j(n)} + h_n) - \hat{F}_{j(n)}(\hat{\xi}_{j(n)} - h_n) + F_{j(n)}(\hat{\xi}_{j(n)} - h_n)| = O_P(\sqrt{n} \max_{1 \leq j \leq p_n} |\hat{\xi}_{j(n)} - \xi_{j(n)}| + h_n^{3/2}) + o_P(n^{\gamma-1/2}) + O_P(\max_{1 \leq j \leq p_n} |B_{j(n)}(F_{j(n)}(\hat{\xi}_{j(n)} + h_n)) - B_{j(n)}(F_{j(n)}(\hat{\xi}_{j(n)} - h_n))|) = o_P(n^{\gamma/2}) + O_P(h_n^{3/2})$, via conditions (10) and (11) and Corollary 6. \square

PROOF OF COROLLARY 8. Now, for some $\xi_{j(n)}^*$ in between $\xi_{j(n)}$ and $\hat{\xi}_{j(n)}$, we have $f_{j(n)}(\xi_{j(n)}^*)(\hat{\xi}_{j(n)} - \xi_{j(n)}) = F_{j(n)}(\hat{\xi}_{j(n)}) - F_{j(n)}(\xi_{j(n)})$. Using the conditions of the corollary we are proving, we also have

$$(29) \quad |(f_{j(n)}(\xi_{j(n)}^*) - f_{j(n)}(\xi_{j(n)}))(\hat{\xi}_{j(n)} - \xi_{j(n)})| \leq M|\hat{\xi}_{j(n)} - \xi_{j(n)}|^{3/2},$$

with probability approaching 1, for all n large enough. Hence

$$\begin{aligned}
 (30) \quad U_{j(n)}^* &\equiv 2\sqrt{n}f_{j(n)}(\xi_{j(n)})(\hat{\xi}_{j(n)} - \xi_{0,j(n)}) \\
 &= \sqrt{n}(F_{j(n)}(\hat{\xi}_{j(n)}) - F_{j(n)}(\xi_{j(n)})) + \Delta_{j(n)},
 \end{aligned}$$

where $\max_{1 \leq j \leq p_n} |\Delta_{j(n)}| = o_P(n^{3\gamma/4-1/4})$, by Corollary 6.

Now note that

$$\begin{aligned}
 &\sqrt{n}(F_{j(n)}(\hat{\xi}_{j(n)}) - F_{j(n)}(\xi_{j(n)})) \\
 &= -\sqrt{n}(\hat{F}_{j(n)}(\hat{\xi}_{j(n)}) - F_{j(n)}(\hat{\xi}_{j(n)}) - \hat{F}_{j(n)}(\xi_{j(n)}) + F_{j(n)}(\xi_{j(n)})) \\
 &\quad - \sqrt{n}(\hat{F}_{j(n)}(\xi_{j(n)}) - F_{j(n)}(\xi_{j(n)})) + \sqrt{n}(\hat{F}_{j(n)}(\hat{\xi}_{j(n)}) - F_{j(n)}(\hat{\xi}_{j(n)})) \\
 &\equiv -A_{j(n)} - V_{j(n)} + C_{j(n)},
 \end{aligned}$$

where $C_{j(n)} = \sqrt{n}H_{j(n)}$ and $H_{j(n)}$ is defined in the proof of Corollary 6 with $|H_{j(n)}| \leq 1/n$. Hence $C_{j(n)} = O_P(n^{-1/2})$, uniformly over $1 \leq j \leq p_n$. Theorem 2 tells us that we can, uniformly over $1 \leq j \leq p_n$, replace $A_{j(n)}$ and $V_{j(n)}$ with $A'_{j(n)} = B_{j(n)}(F_{j(n)}(\hat{\xi}_{j(n)})) - B_{j(n)}(F_{j(n)}(\xi_{j(n)}))$ and $V'_{j(n)} = B_{j(n)}(1/2)$, with error $o_P(n^{\gamma-1/2})$. Note that $Z_{j(n)} \equiv 2B_{j(n)}(1/2)$ are standard normals and that $B_{j(n)}(t) = W_{j(n)}(t) - tW_{j(n)}(1)$ for all $t \in [0, 1]$, for some standard Brownian motions $W_{j(n)}$. Thus, by the symmetry properties of Brownian motion,

$$\begin{aligned}
 |A'_{j(n)}| &\leq \sqrt{\hat{\delta}_{j(n)}} \left[\sup_{0 \leq t \leq 1} |W'_{j(n)}(t)| + \sup_{0 \leq t \leq 1} |W''_{j(n)}(t)| \right] + \hat{\delta}_{j(n)} |W_{j(n)}(1)| \\
 &\equiv \tilde{A}_{j(n)}(\hat{\delta}_{j(n)}),
 \end{aligned}$$

where $\hat{\delta}_{j(n)} \equiv M|\hat{\xi}_{j(n)} - \xi_{j(n)}|$ and M is as defined in (9), and where $W_{j(n)}$, $W'_{j(n)}$ and $W''_{j(n)}$ are Brownian motions.

Thus $\max_{1 \leq j \leq p_n} \tilde{A}_{j(n)}(k_n r_n) \leq O_P(\sqrt{\log p_n \max_{1 \leq j \leq p_n} |\hat{\xi}_{j(n)} - \xi_{j(n)}|}) = o_P(n^{3\gamma/4-1/4})$ by Corollary 6. Combining this with (30), we obtain

$$\begin{aligned}
 (31) \quad U_{j(n)}^* &= Z_{j(n)} + 2\sqrt{n}f_{j(n)}(\xi_{j(n)})(\xi_{j(n)} - \xi_{0,j(n)}) + o_P(n^{3\gamma/4-1/4}) \\
 &= \tilde{U}_{j(n)} + o_P(n^{3\gamma/4-1/4}),
 \end{aligned}$$

where the errors are uniform in j and $\tilde{U}_{j(n)} \equiv Z_{j(n)} + 2\sqrt{n}f_{j(n)}(\xi_{j(n)})(\xi_{j(n)} - \xi_{0,j(n)})$, since $o_P(n^{\gamma-1/2}) = o_P(n^{3\gamma/4-1/4}) \times o_P(1)$.

Since $U_{j(n)} - U_{j(n)}^* = (\hat{f}_{j(n)}/f_{j(n)}(\xi_{j(n)}) - 1)U_{j(n)}^*$, the structure of the remainder of the proof parallels the latter part of the proof of Corollary 4, but with E_{0n} replaced by $(E'_{0n})^2$ and D_{0n} replaced by $D'_{0n} \equiv U_{j(n)}^* - \tilde{U}_{j(n)}$. Thus for $E_{1p_n}(\hat{\pi}'_{j(n)}, \pi'_{j(n)}) = o_P(1)$, we need $E'_{0n} = o_P(1)$ and $D'_{0n} = o_P(1)$. Since $D'_{0n} = o_P(n^{3\gamma/4-1/4})$ by (31), the result follows for all $\gamma \in (0, 1/3]$. In order for

$E_{2p_n}^{(\alpha)}(\hat{\pi}'_{j(n)}, \pi'_{j(n)}) = o_P(1)$, we need $n^\gamma (E'_{0n})^2 = O_P(1)$ and $n^{\gamma/2} D'_{0n} = O_P(1)$. Assuming $\gamma \in (0, 1/5]$, $h_n = O_P(n^{-(1-\gamma)/4})$, $h_n^{-1} = O_P(n^{-5(1-\gamma)/24})$, we have, by Corollary 7, $E'_{0n} = O_P(n^{-(1-\gamma)/6})$. Combining with (31) yields $n^\gamma (E'_{0n})^2 + n^{\gamma/2} D'_{0n} = O_P(n^{4\gamma/3-1/3} + n^{5\gamma/4-1/4})$. The conclusion follows. \square

PROOF OF LEMMA 4. Part (a) is obvious. The results for part (b) follow from Corollary 3, provided we show that $\check{E}_{0n}^* \equiv \max_{1 \leq j \leq p_n} \sigma_{j(n)}^{-2} |\check{S}_{j(n)}^2 - S_{j(n)}^2| = O_P(\hat{\varepsilon}_n)$. Without loss of generality, we will assume that the denominators for both $\check{S}_{j(n)}$ and $S_{j(n)}$ are n instead of $n - 1$. Thus

$$\begin{aligned} |\check{S}_{j(n)}^2 - S_{j(n)}^2| &\leq \sigma_{j(n)}^{-2} |(\check{X}_{j(n)} - \mu_{j(n)})^2 - (\bar{X}_{j(n)} - \mu_{j(n)})^2| \\ &\quad + \left| \frac{1}{n} \sum_{i=1}^n \left(\frac{\check{X}_{ij(n)} - \mu_{j(n)}}{\sigma_{j(n)}} \right)^2 - \left(\frac{X_{ij(n)} - \mu_{j(n)}}{\sigma_{j(n)}} \right)^2 \right| \\ &\leq 2\hat{\varepsilon}_n^2 + 2\hat{\varepsilon}_n \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{X_{ij(n)} - \mu_{j(n)}}{\sigma_{j(n)}} \right)^2}. \end{aligned}$$

Part (i) now follows easily. For part (ii), we utilize Corollary 2 and the sub-Gaussian tail assumption with $r = 1$ and with $\bar{X}_{j(n)}$ in the corollary replaced by $n^{-1} \sum_{i=1}^n (X_{ij(n)} - \mu_{j(n)})^2 / \sigma_{j(n)}^2$ and with $\mu_{j(n)}$ replaced by 1. With these replacements, $\max_{1 \leq j \leq p_n} |\bar{X}_{j(n)} - \mu_{j(n)}| = o_P(1)$, and we are done. \square

PROOF OF COROLLARY 9. This is a consequence of Lemma 4. Let $\check{D}_{0n} \equiv \sqrt{n}(\check{X}_{j(n)} - \bar{X}_{j(n)})/\sigma_{j(n)}$, and note that $\check{D}_{0n} = O_P(\sqrt{n}\hat{\varepsilon}_n)$. Thus, to ensure normalization does not affect the conclusions of Corollary 4, we need the respective rates of \check{E}_{0n} and \check{D}_{0n} to not exceed the rates of E_{0n} and D_{0n} at the appropriate points. For part (a)(i) when $d_n = O(1)$, we need $\check{E}_{0n} = o_P(1)$ and $\check{D}_{0n} = o_P(1)$, which holds if $\hat{\varepsilon}_n = o_P(n^{-1/2})$. For part (b)(i), we need $E_{0n} = O_P(n^{-\gamma})$ and $\check{D}_{0n} = O_P(n^{-\gamma/2})$, which holds if $\hat{\varepsilon}_n = O_P(n^{-1/2-\gamma/2})$, since $\gamma \in (0, 1/2]$. The remaining arguments are similar. \square

PROOF OF THEOREM 3. Define $\tilde{H}_n = \max_{1 \leq j \leq p_n} \|\tilde{F}_{j(n)} - \hat{F}_{j(n)}\|_\infty$ and, for each $\delta \geq 0$, $\hat{H}_n(\delta) = \max_{1 \leq j \leq p_n} \sup_{|s-t| \leq \delta} |\hat{F}_{j(n)}(s) - \hat{F}_{j(n)}(t)|$. Suppose now that for some positive, nonincreasing sequences $\{s_n, \delta_n\}$, with $\delta_n \rightarrow 0$, we have $\hat{H}_n(\delta_n) = o_P(s_n)$ and $P(\hat{\varepsilon}_n > \delta_n) = o(1)$. Then, by the definition of $\hat{\varepsilon}_n$,

$$(32) \quad \tilde{H}_n = \tilde{H}_n \mathbf{1}\{\hat{\varepsilon}_n \leq \delta_n\} + \tilde{H}_n \mathbf{1}\{\hat{\varepsilon}_n > \delta_n\} \leq \hat{H}_n(\delta_n) + o_P(s_n) = o_P(s_n).$$

Now, by Theorem 2 and condition (15), we have for any $\delta_n \downarrow 0$, $\sqrt{n}\hat{H}_n(\delta_n) \leq \max_{1 \leq j \leq p_n} \sup_{|s-t| \leq \delta_n} \sqrt{n}|\hat{F}_{j(n)}(s) - F_{j(n)}(s) - \hat{F}_{j(n)}(t) + F_{j(n)}(t)| + \sqrt{n}\tilde{M}\delta_n \leq$

$\max_{1 \leq j \leq p_n} \sup_{|s-t| \leq \delta_n} |B_{j(n)}(F_{j(n)}(s)) - B_{j(n)}(F_{j(n)}(t))| + O_P((\log n + \log p_n)/\sqrt{n} + \sqrt{n}\delta_n)$. Combining this with a reapplication of condition (15) plus Lemma 10 below (a modulus of continuity bound for Brownian motion),

$$(33) \quad \sqrt{n}\hat{H}_n(\delta_n) \leq O_P\left(\sqrt{(\log p_n)\delta_n \log(1/\delta_n)} + \frac{\log n + \log p_n}{\sqrt{n}} + \sqrt{n}\delta_n\right).$$

Using the fact $\hat{\varepsilon}_n = o_P(1)$, we can find a positive, decreasing sequence $\delta_n \rightarrow 0$ with $\hat{\varepsilon}_n = o_P(\delta_n)$. Now, by applying (32) with $s_n = 1$, we obtain result (i): $\hat{H}_n = o_P(1)$. For result (ii), we use the fact $\log p_n = o(n^\gamma)$ to construct a positive, nondecreasing sequence $r_n \rightarrow \infty$ with $r_n \log p_n = o(n^\gamma)$ and $r_n/\log n = o(1)$. Since $n^{1-\gamma}(\log n)\hat{\varepsilon}_n = O_P(1)$, we have

$$\begin{aligned} & n^{1-\gamma}\hat{\varepsilon}_n \log(1/\hat{\varepsilon}_n) \\ &= n^{1-\gamma}(\log n) \frac{\hat{\varepsilon}_n}{\log n} \left(\log\left(\frac{1}{n^{1-\gamma}(\log n)\hat{\varepsilon}_n}\right) + \log(n^{1-\gamma} \log n) \right) = O_P(1). \end{aligned}$$

Thus, if we set $\delta_n = r_n/(n^{1-\gamma} \log n)$, $\hat{\varepsilon}_n = o_P(\delta_n)$. We also have, by (33),

$$\begin{aligned} \hat{H}_n(\delta_n) &= O_P\left(\sqrt{\frac{1}{n} \times \frac{r_n \log p_n}{n^{1-\gamma}} \times \frac{\log n^{1-\gamma} + \log \log n - \log r_n}{\log n}} + o\left(\frac{1}{n^{1-\gamma}}\right)\right) \\ &= o_P(n^{-(1-\gamma)}). \end{aligned}$$

The proof is done after reapplying (32) with $s_n = n^{-(1-\gamma)}$. \square

LEMMA 10. Let $W : [0, 1] \mapsto \mathbb{R}$ be a standard Brownian motion. Then there exists a universal constant $k_0 < \infty$ such that, for all $0 < \delta \leq 1/2$,

$$\left\| \sup_{|s-t| \leq \delta} |W(s) - W(t)| \right\|_{\psi_2} \leq k_0 \sqrt{\delta \log(1/\delta)}.$$

PROOF. Denote $Z(\delta) \equiv \sup_{|s-t| \leq \delta} |W(s) - W(t)|$. Lemma 1.1.1 of [4] yields that for some $C_0 < \infty$, $P(Z(\delta) > x\sqrt{\delta}) \leq C_0\delta^{-1}e^{-x^2/4}$ for all $x > 0$ and $\delta \in (0, 1/2]$. Letting $u = x\sqrt{\delta}$ we obtain $P(Z(\delta) > u) \leq C_0\delta^{-1}e^{-u^2/(4\delta)}$. Since $-u^2/(4\delta) + \log(1/\delta)$ is increasing in δ for all $u > 2$, we obtain for some $k_* < \infty$ that $P(Z(\delta) > u) \leq k_*e^{-u^2/2}$ for all $u > 0$, since $\delta \leq 1/2$. Now the desired result follows for some k_0 not depending on δ via Lemma 7. \square

PROOF OF COROLLARY 10. Result (i) follows directly from part (i) of Theorem 3 and Theorem 1 plus the fact that the absence of ties yields

$$(34) \quad \begin{aligned} O_P(n^{-1}) &= \tilde{F}_{j(n)}(\check{\xi}_{j(n)}) - F_{j(n)}(\xi_{j(n)}) \\ &= o_P(1) + F_{j(n)}(\check{\xi}_{j(n)}) - F_{j(n)}(\xi_{j(n)}), \end{aligned}$$

where the errors are uniform in j . To prove result (ii), we first utilize part (ii) of Theorem 3 and a reapplication of Theorem 1 to reduce the uniform error in (34) to $o_P(n^{-(1-\gamma)/2})$, obtaining $C_n \equiv \max_{1 \leq j \leq p_n} |\check{\xi}_{j(n)} - \hat{\xi}_{j(n)}| = o_P(n^{-(1-\gamma)/2})$. Using an alternative expansion to (34) followed by part (ii) of Theorem 3, we now obtain $O_P(n^{-1}) = \tilde{F}_{j(n)}(\check{\xi}_{j(n)}) - \hat{F}_{j(n)}(\hat{\xi}_{j(n)}) = \tilde{F}_{j(n)}(\check{\xi}_{j(n)}) - \hat{F}_{j(n)}(\check{\xi}_{j(n)}) + [\hat{F}_{j(n)}(\check{\xi}_{j(n)}) - F_{j(n)}(\check{\xi}_{j(n)}) - \hat{F}_{j(n)}(\hat{\xi}_{j(n)}) + F_{j(n)}(\hat{\xi}_{j(n)})] + F_{j(n)}(\check{\xi}_{j(n)}) - F_{j(n)}(\hat{\xi}_{j(n)}) \equiv o_P(n^{-(1-\gamma)}) + [H_{j(n)}] + (1 + o_P(1))f_{j(n)}(\xi_{j(n)})(\check{\xi}_{j(n)} - \hat{\xi}_{j(n)})$, where the errors are uniform in j . Define $\check{H}_n \equiv \max_{1 \leq j \leq p_n} |H_{j(n)}|$, $q_n \equiv n^{-(1-\gamma)/2}$ and $\delta_n \equiv q_n/4$. Note that $P(C_n > \delta_n) = o(1)$, $P(\max_{1 \leq j \leq p_n} |\hat{\xi}_{j(n)} - \xi_{j(n)}| > q_n) = o(1)$ and $P(\max_{1 \leq j \leq p_n} |\check{\xi}_{j(n)} - \xi_{j(n)}| > q_n) = o(1)$. Thus

$$\begin{aligned} \sqrt{n}\check{H}_n &= O_P\left(\max_{1 \leq j \leq p_n} \sup_{|s-t| \leq \delta_n: 0 \leq s, t \leq q_n} |B_{j(n)}(F_{j(n)}(\xi_{j(n)} + s)) \right. \\ &\quad \left. - B_{j(n)}(F_{j(n)}(\xi_{j(n)} + t))\right) \\ &\quad + o_P(n^{-(1/2-\gamma)}) \\ &= O_P(\sqrt{\log p_n \delta_n \log(q_n/\delta_n)}) + o_P(n^{-(1/2-\gamma)}), \end{aligned}$$

where the last equality follows from Lemma 10 and the rescaling properties of Brownian motion. Thus $\check{H}_n = o_P(n^{-3(1-\gamma)/4})$, and part (ii) follows.

For part (iii), recycling arguments yields $\check{H}_n \equiv h_n \sqrt{n} \max_{1 \leq j \leq p_n} |\check{f}_{j(n)} - \hat{f}_{j(n)}| \leq \max_{1 \leq j \leq p_n} \sqrt{n} |\tilde{F}_{j(n)}(\check{\xi}_{j(n)} + h_n) - \hat{F}_{j(n)}(\check{\xi}_{j(n)} + h_n) + \hat{F}_{j(n)}(\check{\xi}_{j(n)} + h_n) - \hat{F}_{j(n)}(\hat{\xi}_{j(n)} + h_n)| + \max_{1 \leq j \leq p_n} \sqrt{n} |\tilde{F}_{j(n)}(\check{\xi}_{j(n)} - h_n) - \hat{F}_{j(n)}(\check{\xi}_{j(n)} - h_n) + \hat{F}_{j(n)}(\check{\xi}_{j(n)} - h_n) - \hat{F}_{j(n)}(\hat{\xi}_{j(n)} - h_n)| = O_P(\max_{1 \leq j \leq p_n} \sqrt{n} \|\tilde{F}_{j(n)} - \hat{F}_{j(n)}\|_\infty) + o_P(n^{-(1/2-\gamma)}) + o_P(n^{3\gamma/4-1/4}) + O_P(\max_{1 \leq j \leq p} |B_{j(n)}(F_{j(n)}(\check{\xi}_{j(n)} + h_n)) - B_{j(n)}(F_{j(n)}(\hat{\xi}_{j(n)} + h_n))|) + O_P(\max_{1 \leq j \leq p} |B_{j(n)}(F_{j(n)}(\check{\xi}_{j(n)} - h_n)) - B_{j(n)}(F_{j(n)}(\hat{\xi}_{j(n)} - h_n))|) = o_P(n^{3\gamma/4-1/4}) + O_P(\max_{1 \leq j \leq p_n} \sup_{|s-t| \leq \delta'_n} |B_{j(n)}(s) - B_{j(n)}(t)|) = o_P(n^{3\gamma/4-1/4}) + O_P(\sqrt{\delta'_n \log(1/\delta'_n)})$, where $\delta'_n = n^{-3(1-\gamma)/4}$, ensuring that $P(C_n > \delta'_n) = o(1)$. Hence $\check{H}_n = O_P(n^{-3(1-\gamma)/8} \times \sqrt{\log n}) = o_P(n^{3\gamma/4-1/4})$, and the desired result follows. \square

PROOF OF COROLLARY 11. Note that since $|\check{f}_{j(n)} - f_{j(n)}(\xi_{j(n)})| \leq |\check{f}_{j(n)} - \hat{f}_{j(n)}| + |\hat{f}_{j(n)} - f_{j(n)}(\xi_{j(n)})|$, $\max_{1 \leq j \leq p_n} |\check{f}_{j(n)} - f_{j(n)}(\xi_{j(n)})| = o_P(h_n^{-1} n^{-3(1-\gamma)/4}) + o_P(h_n^{-1} n^{-(1-\gamma)/2}) + O_P(h_n^{1/2}) = o_P(h_n^{-1} n^{-(1-\gamma)/2}) + O_P(h_n^{1/2})$, where the second-to-last equality comes from Corollary 7. Now by part (ii) of Corollary 10, $2\sqrt{n}f_{j(n)}(\xi_{j(n)})(\check{\xi}_{j(n)} - \hat{\xi}_{j(n)}) = o_P(n^{3\gamma/4-1/4})$. Thus, considering (31) and the lines following, none of the error rates in the proof of Corollary 8 is altered after replacing $\hat{\xi}_{j(n)}$ and $\hat{f}_{j(n)}$ with $\check{\xi}_{j(n)}$ and $\check{f}_{j(n)}$, respectively. \square

Acknowledgments. The authors thank Editor Jianqing Fan, an Associate Editor and three referees for several very helpful suggestions.

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] BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. Wiley, New York. [MR1324786](#)
- [3] BRETAGNOLLE, J. and MASSART, P. (1989). Hungarian construction from the nonasymptotic viewpoint. *Ann. Probab.* **17** 239–256. [MR0972783](#)
- [4] CSÖRGŐ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, New York. [MR0666546](#)
- [5] DUDOIT, S., FRIDLAND, J. and SPEED, T. P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *J. Amer. Statist. Assoc.* **97** 77–87. [MR1963389](#)
- [6] DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669. [MR0083864](#)
- [7] FAN, J., HALL, P. and YAO, Q. (2005). To how many simultaneous hypothesis tests can normal, Student’s t or bootstrap calibration be applied? Unpublished manuscript.
- [8] FAN, J., PENG, H. and HUANG, T. (2005). Semilinear high-dimensional model for normalization of microarray data: A theoretical analysis and partial consistency (with discussion). *J. Amer. Statist. Assoc.* **100** 781–813. [MR2201010](#)
- [9] FAN, J., TAM, P., VANDE WOUDE, G. and REN, Y. (2004). Normalization and analysis of cDNA microarrays using within-array replications applied to neuroblastoma cell response to a cytokine. *Proc. Natl. Acad. Sci. USA* **101** 1135–1140.
- [10] GENOVESE, C. and WASSERMAN, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** 499–517. [MR1924303](#)
- [11] GHOSH, D. and CHINNAIYAN, A. M. (2005). Classification and selection of biomarkers in genomic data using LASSO. *J. Biomedicine and Biotechnology* **2005** 147–154.
- [12] GUI, J. and LI, H. (2005). Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. *Bioinformatics* **21** 3001–3008.
- [13] HUANG, J., KUO, H.-C., KOROLEVA, I., ZHANG, C.-H. and BENTO SOARES, M. (2003). A semilinear model for normalization and analysis of cDNA microarray data. Technical Report 321, Dept. Statistics and Actuarial Science, Univ. Iowa.
- [14] HUANG, J., WANG, D. and ZHANG, C.-H. (2005). A two-way semilinear model for normalization and analysis of cDNA microarray data. *J. Amer. Statist. Assoc.* **100** 814–829. [MR2201011](#)
- [15] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent τ_v ’s and the sample df. *I. Z. Wahrsch. Verw. Gebiete* **32** 111–131. [MR0375412](#)
- [16] KOSOROK, M. R. (1999). Two-sample quantile tests under general conditions. *Biometrika* **86** 909–921. [MR1741986](#)
- [17] KOSOROK, M. R. (2002). On global consistency of a bivariate survival estimator under univariate censoring. *Statist. Probab. Lett.* **56** 439–446. [MR1898723](#)
- [18] KOSOROK, M. R. and MA, S. (2005). Comment on “Semilinear high-dimensional model for normalization of microarray data: A theoretical analysis and partial consistency,” by J. Fan, H. Peng and T. Huang. *J. Amer. Statist. Assoc.* **100** 805–807.

- [19] KOSOROK, M. R. and MA, S. (2005). Marginal asymptotics for the “large p , small n ” paradigm: With applications to microarray data. Technical Report 188, Dept. Biostatistics and Medical Informatics, Univ. Wisconsin, Madison.
- [20] MASSART, P. (1990). The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Ann. Probab.* **18** 1269–1283. [MR1062069](#)
- [21] SKOROHOD, A. V. (1976). On a representation of random variables. *Theory Probab. Appl.* **21** 628–632. [MR0428369](#)
- [22] SPANG, R., BLANCHETTE, C., ZUZAN, H., MARKS, J., NEVINS, J. and WEST, M. (2001). Prediction and uncertainty in the analysis of gene expression profiles. In *Proc. German Conference on Bioinformatics GCB 2001* (E. Wingender, R. Hofestädt and I. Liebich, eds.) 102–111.
- [23] STOREY, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** 479–498. [MR1924302](#)
- [24] STOREY, J. D., TAYLOR, J. E. and SIEGMUND, E. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discover rates: A unified approach. *J. R. Stat. Soc. Ser. B Methodol.* **66** 187–205. [MR2035766](#)
- [25] VAN DER LAAN, M. J. and BRYAN, J. (2001). Gene expression analysis with the parametric bootstrap. *Biostatistics* **2** 445–461.
- [26] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York. [MR1385671](#)
- [27] WEST, M. (2003). Bayesian factor regression models in the “large p , small n ” paradigm. In *Bayesian Statistics 7* (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, eds.) 733–742. Oxford Univ. Press. [MR2003537](#)
- [28] WEST, M., BLANCHETTE, C., DRESSMAN, H., HUANG, E., ISHIDA, S., SPANG, R., ZUZAN, H., OLSON, J. A. JR., MARKS, J. R. and NEVINS, J. R. (2001). Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc. Natl. Acad. Sci. USA* **98** 11,462–11,467.
- [29] YANG, Y. H., DUDOIT, S., LUU, P. and SPEED, T. P. (2001). Normalization for cDNA microarray data. In *Microarrays: Optical Technologies and Informatics* (M. L. Bittner, Y. Chen, A. N. Dorsal and E. R. Dougherty, eds.) 141–152. *Proc. SPIE* **4266**.

DEPARTMENT OF BIostatISTICS
UNIVERSITY OF NORTH CAROLINA-CHAPEL HILL
3101 MCGAVRAN-GREENBERG, CB 7420
CHAPEL HILL, NORTH CAROLINA 27599
USA
E-MAIL: kosorok@unc.edu

DIVISION OF BIostatISTICS
YALE UNIVERSITY
60 COLLEGE ST., LEPH 209
NEW HAVEN, CONNECTICUT 06520
USA
E-MAIL: shuangge.ma@yale.edu