Estimating the Null and the Proportion of Nonnull Effects in Large-Scale Multiple Comparisons

Jiashun JIN and T. Tony CAI

An important issue raised by Efron in the context of large-scale multiple comparisons is that in many applications, the usual assumption that the null distribution is known is incorrect, and seemingly negligible differences in the null may result in large differences in subsequent studies. This suggests that a careful study of estimation of the null is indispensable. In this article we consider the problem of estimating a null normal distribution, and a closely related problem, estimation of the proportion of nonnull effects. We develop an approach based on the empirical characteristic function and Fourier analysis. The estimators are shown to be uniformly consistent over a wide class of parameters. We investigate the numerical performance of the estimators using both simulated and real data. In particular, we apply our procedure to the analysis of breast cancer and human immunodeficiency virus microarray datasets. The estimators perform favorably compared with existing methods.

KEY WORDS: Characteristic functions; Empirical characteristic function; Fourier coefficients; Multiple testing; Null distribution; Proportion of nonnull effects.

1. INTRODUCTION

The analysis of massive datasets now commonly used in scientific investigations poses many statistical challenges not present in smaller-scale studies. One such challenge is the need for large-scale *simultaneous testing* or *multiple comparisons*, in which thousands or even millions of hypotheses are tested simultaneously. In this setting, one considers a large number of null hypotheses $H_1, H_2, \ldots, H_n$, and is interested in determining which hypotheses are true and which are not. Associated with each hypothesis is a test statistic. When H_j is true, the test statistic X_j has a null distribution function F_0, that is,

$$(X_j | H_j \text{ is true}) \sim F_0.$$

Since the pioneering work of Benjamini and Hochberg (1995), which introduced the false discovery rate (FDR)-controlling procedures, research on large-scale simultaneous testing has been very active (see, e.g., Abravomich, Benjamini, Donoho, and Johnstone 2006; Cai, Jin, and Low 2007; Donoho and Jin 2004; Efron 2004; Efron, Tibshirani, Storey, and Tusher 2001; Genovese and Wasserman 2004; Meinhausen and Rice 2006; Storey, Dai, and Leek 2007).

FDR procedures are based on p values, which measure the tail probability of the null distribution. Conventionally the null distribution is always assumed known. However, somewhat surprisingly, Efron (2004) pointed out that in many applications, such an assumption would be incorrect. Efron studied a dataset on breast cancer in which a gene microarray was generated for each patient in two groups, the BRCA1 group and BRCA2 group. The goal was to determine which genes were differentially expressed between the two groups. For each gene, a p value was calculated using the classical t test. For convenience, Efron chose to work on the z scale through the transformation $X_j = \bar{\Phi}^{-1}(p_j)$, where $\bar{\Phi} = 1 - \Phi$ is the survival function of the standard normal distribution. Efron argued that although theoretically the null distribution should be the standard normal, empirically another null distribution (which Efron called the *empirical null*) is more appropriate. In fact, Efron found that $N(-.02, 1.58^2)$ is a more appropriate null than $N(0, 1)$ (Fig. 1). Efron also found a similar phenomenon in the analysis of a microarray dataset on human immunodeficiency virus (HIV).

Different choices of the null distribution can give substantially different outcomes in simultaneous multiple testing. Even a seemingly negligible estimation error of the null may result in large differences in subsequent studies. For illustration, we carried out an experiment containing 100 independent cycles of simulations. In each cycle, 9,000 samples are drawn from $N(0, .95^2)$ to represent the null effects, and 1,000 samples are drawn from $N(2, .95^2)$ to represent the nonnull effects. For each sample element X_j, p values are calculated as $\bar{\Phi}^{-1}(X_j/.95)$ and $\bar{\Phi}^{-1}(X_j)$, which represent the p values under the *true* null and the *misspecified* null. The FDR procedure is then applied to both sets of p values, with the FDR control parameter set at .05. The results, reported in Figure 2, show that the true positives obtained by using $N(0, 1)$ as the null and those obtained by using $N(0, .95^2)$ as the null are considerably different. This, together with Efron's arguments, suggests that a careful study on estimating the null is indispensable.

Efron (2004) introduced a method for estimating the null distribution based on the notion of "sparsity." Sparsity can be defined in several different ways (Abramovich et al. 2006). The most intuitive one is that the proportion of nonnull effects is small. In some applications, the case of "asymptotically vanishing sparsity" is of particular interest (Abramovich et al. 2006; Donoho and Jin 2004). This case refers to the situation in which the proportion of nonnull effects tends to zero as the number of hypotheses grows to infinity. In such a setting, heuristically, the influence of the nonnull effects becomes more and more negligible, and so the null can be reliably estimated asymptotically. In fact, Efron (2004) suggested an approach that uses the center and half-width of the central peak of the histogram to estimate the parameters of the null distribution.

In many applications it is more appropriate to model the setting as *nonsparse*, that is, the proportion of nonnull effects does

Jiashun Jin is Assistant Professor, Department of Statistics, Purdue University, West Lafayette, IN 47907 (E-mail: *jinj@stat.purdue.edu*). T. Tony Cai is Associate Professor, Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104 (E-mail: *tcai@wharton.upenn.edu*). The authors thank Bradley Efron for references and kindly sharing the datasets and Paul Shaman for a careful reading of our manuscript and suggestions that led to significant improvement of the presentation of the article. They also thank Herman Rubin, an associate editor, and referees for helpful comments and references. The research of Jin was supported in part by National Science Foundation (NSF) grant DMS-05-05423, and the research of Cai was supported in part by NSF grants DMS-03-06576 and DMS-06-04954.

Journal of the American Statistical Association
June 2007, Vol. 102, No. 478, Theory and Methods
DOI 10.1198/016214507000000167

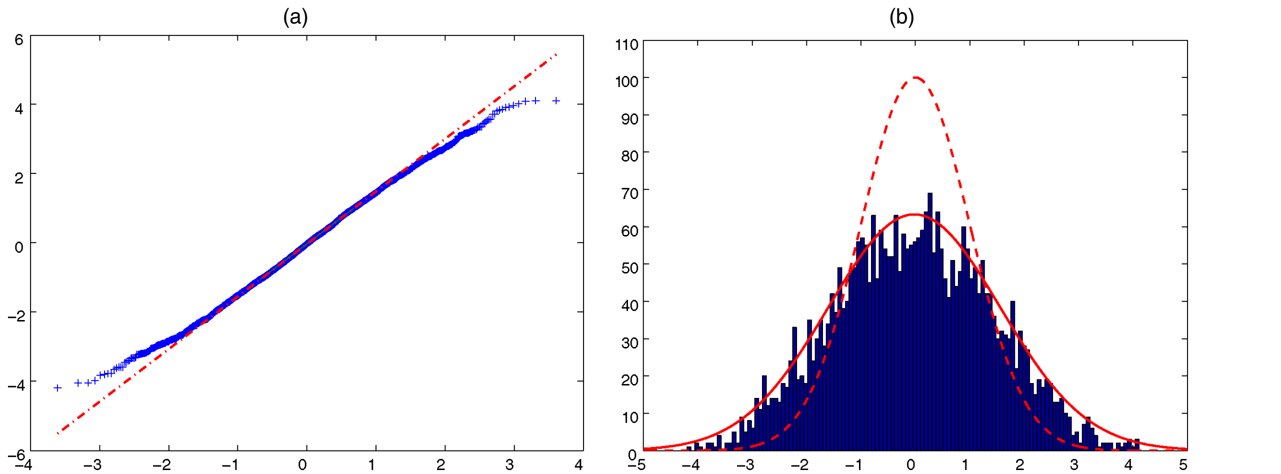

Figure 1. z-Values of Microarray Data on Breast Cancer. (a) Q–Q plot. (b) Histogram and density curves of N(0, 1) (— — —) and N(−.02, 1.58^2). The plot suggests that the null is N(−.02, 1.58^2) rather than N(0, 1). (See Efron 2004 for further details.)

not tend to zero when the number of hypotheses grows to infinity. In such settings, Efron's (2004) approach does not perform well, and it is not hard to show that the estimators of the null are generally inconsistent. Moreover, even when the setting is asymptotically vanishingly sparse and the estimators are consistent, it is still of interest to quantify the influence of sparsity on the estimators, because a small error in the null may propagate to large errors in subsequent studies.

Conventional methods for estimating the null parameters are based on either moments or extreme observations (Efron 2004; Meinshausen and Rice 2006; Swanepoel 1999). However, in the nonsparse case, neither of these is very informative, because the relevant information about the null is highly distorted by the nonnull effects in both of them. In this article we propose a new approach for estimating the null parameters using the empirical characteristic function and Fourier analysis as the main tools. The approach demonstrates that the information about the null is well preserved in the high-frequency Fourier coefficients, where the distortion of the nonnull effects is asymptotically negligible. The approach integrates the strength of several factors, including sparsity and heteroscedasticity, and provides good estimates of the null in a much broader range of situations than existing approaches do. The resulting estimators are shown to be uniformly consistent over a wide class of parameters and to outperform existing methods in simulations.

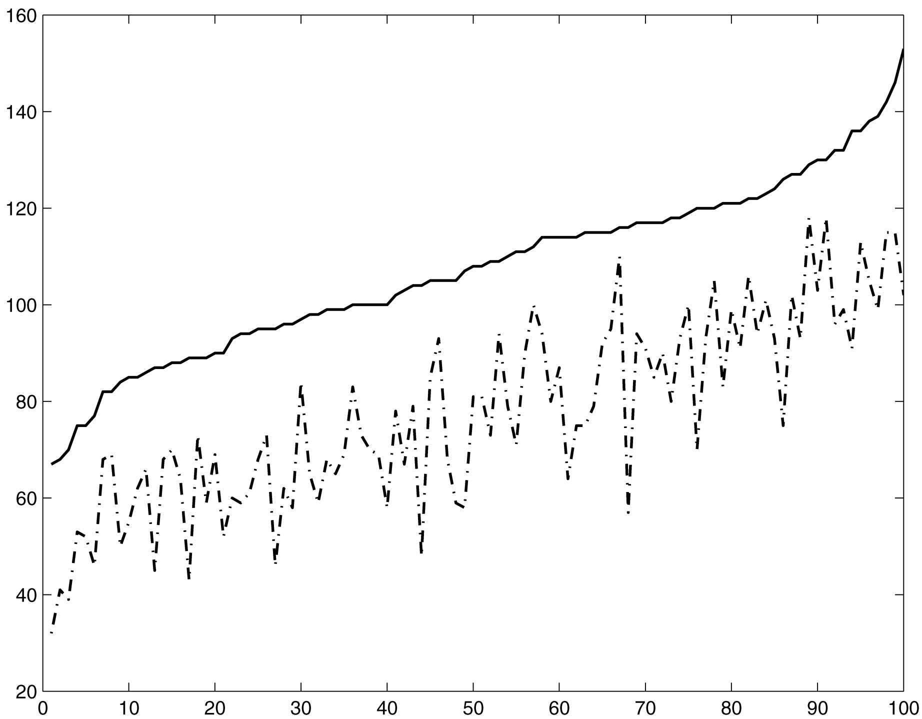

Figure 2. The Number of True Positives for Each Cycle, Using the True Null (—) and the Misspecified Null (— — —). For visualization, the numbers are sorted ascendingly with respect to those in the true null case.

Along with the null distribution, the proportion of nonnull effects also is an important quantity. For example, implementation of many recent procedures requires the knowledge of both the null and the proportion of nonnull effects (see Efron et al. 2001; Lönnstedt and Speed 2002; Storey et al. 2007). Developing good estimators for the proportion is a challenging task. Recent work includes that of Meinshausen and Rice (2006), Swanepoel (1999), Cai et al. (2007), and Jin (2006). In this article we extend the method of Jin (2006) to the current setting of heteroscedasticity with an unknown null distribution. The estimator is shown to be uniformly consistent over a wide class of parameters.

Along with the theoretical properties, we investigated numerical performance of the estimators using both simulated and real data. In particular, we used our procedure to analyze the breast cancer (Hedenfalk, Duggen, Chen, et al. 2001) and HIV (Van't Wout et al.) microarray data analyzed by Efron (2004). The results indicate that our estimated null parameters lead to a more reliable identification of differentially expressed genes than that in Efron (2004).

The article is organized as follows. After reviewing basic notations and definitions, we define the estimators of the null parameters in Section 2.1. We investigate the theoretical properties of the estimators in Sections 2.2 and 2.3. In Section 2.4 we discuss the extension to dependent data structures. In Section 3 we cover estimation of the proportion of nonnull effects. In Section 4 we report on a simulation study carried out to investigate numerical performance. In Section 5 we apply our procedure to the analysis of the breast cancer and HIV microarray data. We give proofs of the main theorems in the Appendix.

2. ESTIMATING THE NULL DISTRIBUTION

Following Efron (2004), we work on the z-scale and consider n test statistics

$$X_j \sim \mathrm{N}(\mu_j, \sigma_j^2), \qquad 1 \le j \le n, \tag{1}$$

where μ_j and σ_j are unknown parameters. For a pair of *null parameters*, μ_0 and σ_0,

$$\begin{aligned} &(\mu_j, \sigma_j) = (\mu_0, \sigma_0) \quad \text{if } H_j \text{ is true} \qquad \text{and} \\ &(\mu_j, \sigma_j) \ne (\mu_0, \sigma_0) \quad \text{if } H_j \text{ is untrue,} \end{aligned} \tag{2}$$

and we are interested in estimating μ_0 and σ_0. We first consider the case in which $X_1, \dots, X_n$ are independent. We consider the dependent case in Section 2.4.

Set $\mu = \{\mu_1, \dots, \mu_n\}$ and $\sigma = \{\sigma_1, \dots, \sigma_n\}$. Denote the proportion of nonnull effects by

$$\epsilon_n = \epsilon_n(\mu, \sigma) = \frac{\#\{j : (\mu_j, \sigma_j) \ne (\mu_0, \sigma_0)\}}{n}. \tag{3}$$

We assume that $\sigma_j \ge \sigma_0$ for all $1 \le j \le n$; that is, the standard deviation of a nonnull effect is no less than that of a null effect. This is the case in a wide range of applications (Efron 2004; Lönnstedt and Speed 2002). To make the null parameters identifiable, we assume that

$$\epsilon_n(\mu, \sigma) \le \epsilon_0 \quad \text{for some constant } 0 < \epsilon_0 < \frac{1}{2}. \tag{4}$$

Definition 1. Fix $\epsilon_0 \in (0, 1/2)$, μ_0, and $\sigma_0 > 0$. We say that (μ, σ) is $(\mu_0, \sigma_0, \epsilon_0)$-eligible if (4) is satisfied and $\sigma_j \ge \sigma_0$ for all $1 \le j \le n$.

Throughout this article, we assume that (μ, σ) is $(\mu_0, \sigma_0, \epsilon_0)$-eligible.

2.1 Estimating the Null Parameters

As mentioned in Section 1, an informative approach to estimating the null distribution is to use the Fourier coefficients at suitable frequencies. In the literature, Fourier coefficients frequently have been used for statistical inference (see, e.g., Fan 1991; Zhang 1990). Here we use them to construct estimators for the null parameters.

Introduce the *empirical characteristic function*,

$$\varphi_n(t) = \varphi_n(t; X_1, \dots, X_n, n) = \frac{1}{n} \sum_{j=1}^{n} e^{itX_j}, \tag{5}$$

and its expectation, the *characteristic function*, $\varphi(t) = \varphi(t; \mu, \sigma, n) = \frac{1}{n} \sum_{j=1}^{n} e^{it\mu_j - \sigma_j^2 t^2/2}$, where $i = \sqrt{-1}$. The characteristic function φ naturally splits into two components, $\varphi(t) = \varphi_0(t) + \tilde{\varphi}(t)$, where $\varphi_0(t) = \varphi_0(t; \mu, \sigma, n) = (1 - \epsilon_n) \times e^{i\mu_0 t - \sigma_0^2 t^2/2}$ and

$$\begin{aligned} \tilde{\varphi}(t) &= \tilde{\varphi}(t; \mu, \sigma, n) \\ &= \epsilon_n \cdot \mathrm{Ave}_{\{j : (\mu_j, \sigma_j) \ne (\mu_0, \sigma_0)\}} \{ e^{i\mu_j t - \sigma_j^2 t^2/2} \}, \end{aligned} \tag{6}$$

which correspond to the null and nonnull effects. Note that the identifiability condition $\epsilon_n \le \epsilon_0 < 1/2$ ensures that $\varphi(t) \ne 0$ for all t.

We now use the foregoing functions to construct estimators for σ_0^2 and μ_0. For any $t \ne 0$ and any differentiable complex-valued function f such that $|f(t)| \ne 0$, we define the two functionals

$$\begin{aligned} \sigma_0^2(f; t) &= -\frac{\frac{d}{dt}|f(t)|}{t \cdot |f(t)|} \qquad \text{and} \\ \mu_0(f; t) &= \frac{\mathrm{Re}(f(t)) \cdot \mathrm{Im}(f'(t)) - \mathrm{Re}(f'(t)) \cdot \mathrm{Im}(f(t))}{|f(t)|^2}, \end{aligned} \tag{7}$$

where $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$ denote the real and imaginary parts of the complex number z. Simple calculus shows that evaluating the functionals at φ_0 gives the exact values of σ_0^2 and μ_0: $\sigma_0^2(\varphi_0; t) = \sigma_0^2$ and $\mu_0(\varphi_0; t) = \mu_0$ for all $t \ne 0$.

Inspired by this, we hope that for an appropriately chosen large t, $\varphi_n(t) \approx \varphi(t) \approx \varphi_0(t)$, so that the contribution of nonnull effects to the empirical characteristic function is negligible, which would then give rise to good estimates for σ_0^2 and μ_0. More specifically, we use $\sigma_0^2(\varphi_n; t)$ and $\mu_0(\varphi_n; t)$ as estimators for σ_0^2 and μ_0, and hope that by choosing an appropriate t, we have

$$\sigma_0^2(\varphi_n; t) \approx \sigma_0^2(\varphi; t) \approx \sigma_0^2(\varphi_0; t) \equiv \sigma_0^2 \tag{8}$$

and

$$\mu_0(\varphi_n; t) \approx \mu_0(\varphi; t) \approx \mu_0(\varphi_0; t) \equiv \mu_0. \tag{9}$$

There is clearly a trade-off in the choice of t. As t increases from 0 to ∞, the second approximations in (8) and (9) become increasingly accurate, but the first approximations become more unstable, because the variances of $\sigma_0^2(\varphi_n; t)$ and $\mu_0(\varphi_n; t)$ increase with t. Intuitively, we should choose a t

such that $\varphi_n(t)/\varphi(t) \approx 1$, so that φ can be estimated with first-order accuracy. Note that by the central limit theorem, $|\varphi_n(t) - \varphi(t)| = O_p(\frac{1}{\sqrt{n}})$, so t should be chosen such that $\varphi(t) \gg \frac{1}{\sqrt{n}}$.

We introduce the following method for choosing t, which is adaptive to the magnitude of the empirical characteristic function. For a given $\gamma \in (0, 1/2)$, set

$$\hat{t}_n(\gamma) = \hat{t}_n(\gamma; \varphi_n) = \inf\{t : |\varphi_n(t)| = n^{-\gamma}, 0 \le t \le \log n\}. \tag{10}$$

Once we determine the frequency $t = \hat{t}_n(\gamma)$, we have the following family of "plug in" estimators indexed by $\gamma \in (0, 1/2)$:

$$\hat{\sigma}_0^2 = \sigma_0^2(\varphi_n; \hat{t}_n(\gamma)) \qquad \text{and} \qquad \hat{\mu}_0 = \mu_0(\varphi_n; \hat{t}_n(\gamma)). \tag{11}$$

We show later (see Lemma A.3 in the App.) that $\hat{t}_n(\gamma)$ is asymptotically equivalent to the nonstochastic quantity

$$t_n(\gamma) = t_n(\gamma; \varphi) = \inf\{t : |\varphi(t)| = n^{-\gamma}, 0 \le t \le \log n\}, \tag{12}$$

and that the stochastic fluctuation of $\hat{t}_n(\gamma)$ is algebraically small and its effect is generally negligible. We note here that, by elementary calculus,

$$t_n(\gamma, \varphi) = [\sqrt{2\gamma \log n}/\sigma_0] \cdot (1 + o(1)), \qquad n \to \infty, \tag{13}$$

where $o(1)$ tends to 0 uniformly for all φ under consideration.

2.2 Uniform Consistency of the Estimators

We now show that the estimators $\hat{\sigma}_0^2$ and $\hat{\mu}_0$ given in (11) are consistent uniformly over a wide class of parameters. We introduce two nonstochastic bridging quantities, $\sigma_0^2(\varphi; t_n(\gamma))$ and $\mu_0(\varphi; t_n(\gamma))$, which correspond to σ_0^2 and μ_0. For each estimator, the estimation error can be decomposed into two components: the stochastic fluctuation and the difference between the true parameter and its corresponding bridging quantity,

$$\left|\sigma_0^2(\varphi_n; \hat{t}_n(\gamma)) - \sigma_0^2\right| \le \left|\sigma_0^2(\varphi_n; \hat{t}_n(\gamma)) - \sigma_0^2(\varphi; t_n(\gamma))\right| + \left|\sigma_0^2(\varphi; t_n(\gamma)) - \sigma_0^2\right| \tag{14}$$

and

$$\left|\mu_0(\varphi_n; \hat{t}_n(\gamma)) - \mu_0\right| \le \left|\mu_0(\varphi_n; \hat{t}_n(\gamma)) - \mu_0(\varphi; t_n(\gamma))\right| + \left|\mu_0(\varphi; t_n(\gamma)) - \mu_0\right|. \tag{15}$$

We consider the behavior of the two components separately. Fix constants $q > 0$ and $A > 0$, and introduce the set of parameters

$$\Lambda_n(q, A; \mu_0, \sigma_0, \epsilon_0) = \{(\mu, \sigma) \text{ is } (\mu_0, \sigma_0, \epsilon_0)\text{-eligible}, M_n^{(q)}(\mu, \sigma) \le A^q\}, \tag{16}$$

where $M_n^{(q)}(\mu, \sigma) = \text{Ave}_{\{j : (\mu_j, \sigma_j) \ne (\mu_0, \sigma_0)\}}\{(|\mu_j - \mu_0| + |\sigma_j^2 - \sigma_0^2|^{1/2})^q\}$. For a constant r, we say that a sequence $\{a_n\}_{n=1}^{\infty}$ is $\bar{o}(n^{-r})$ if for any $\delta > 0$, $n^{r-\delta}|a_n| \to 0$ as $n \to \infty$. The following theorem elaborates the magnitude of the stochastic component.

Theorem 1. Fix constants $\gamma, \epsilon_0 \in (0, 1/2)$, $q \ge 3$, and $A > 0$. As $n \to \infty$, except for an event with probability $\bar{o}(n^{-c_1})$,

$$\sup_{\{\Lambda_n(q, A; \mu_0, \sigma_0, \epsilon_0)\}} \left|\sigma_0^2(\varphi_n; \hat{t}_n(\gamma)) - \sigma_0^2(\varphi; t_n(\gamma))\right| \le 3c_2 \cdot \log^{1/2}(n) \cdot n^{\gamma - 1/2}$$

and

$$\sup_{\{\Lambda_n(q, A; \mu_0, \sigma_0, \epsilon_0)\}} \left|\mu_0(\varphi_n; \hat{t}_n(\gamma)) - \mu_0(\varphi; t_n(\gamma))\right| \le \sqrt{2\gamma} c_2 \cdot \log(n) \cdot n^{\gamma - 1/2},$$

where $c_2 = c_2(\sigma_0, q, \gamma) = 2\sigma_0^2 \cdot \sqrt{\max\{3, q - 1 - 2\gamma\}}$ and

$$c_1 = c_1(q, \gamma) = \begin{cases} (q/2 - 1 - \gamma)/2, & q < 4 \\ (q/2 - 1 - \gamma), & 4 \le q \le 4 + 2\gamma \\ (q - 1 - 2\gamma)/3, & q > 4 + 2\gamma. \end{cases} \tag{17}$$

Theorem 1 says that the stochastic components in (14) and (15) are both algebraically small uniformly over Λ_n.

We now consider the nonstochastic components in (14) and (15). As defined in (6), $\tilde{\varphi}(t)$ naturally factors into $\tilde{\varphi}(t) = e^{i\mu_0 t - \sigma_0^2 t^2/2} \cdot \psi(t)$, where

$$\psi(t) = \psi(t; \mu, \sigma, n) = \epsilon_n \cdot \text{Ave}_{\{j : (\mu_j, \sigma_j) \ne (\mu_0, \sigma_0)\}} e^{i(\mu_j - \mu_0)t - (\sigma_j^2 - \sigma_0^2)t^2/2}. \tag{18}$$

Lemma A.5 tells us that there is a constant $C > 0$ such that uniformly for all $(\mu_0, \sigma_0, \epsilon_0)$-eligible parameters (μ, σ), $|\sigma_0^2(\varphi; t_n(\gamma)) - \sigma_0^2| \le C \cdot |\psi'(t_n(\gamma))|/t_n(\gamma)$ and $|\mu_0(\varphi; t_n(\gamma)) - \mu_0| \le C \cdot |\psi'(t_n(\gamma))|$; see the details therein. Combining these with Theorem 1 gives the following theorem, which is proved in the Appendix.

Theorem 2. Fix constants $\gamma, \epsilon_0 \in (0, 1/2)$, $q \ge 3$, and $A > 0$. For all t, $\sup_{\{\Lambda_n(q, A; \mu_0, \sigma_0, \epsilon_0)\}} |\psi'(t)| \le A \cdot \epsilon_0$. Moreover, there is a constant $C = C(\gamma, q, A, \epsilon_0, \mu_0, \sigma_0)$ such that, except for an event with algebraically small probability, for any $(\mu, \sigma) \in \Lambda_n(q, A; \mu_0, \sigma_0, \epsilon_0)$ and all sufficiently large n,

$$\left|\sigma_0^2(\varphi_n; \hat{t}_n(\gamma)) - \sigma_0^2\right| \le C\left(\frac{|\psi'(t_n(\gamma))|}{\sqrt{\log n}} + \log^{1/2}(n) \cdot n^{\gamma - 1/2}\right)$$

and

$$\left|\mu_0(\varphi_n; \hat{t}_n(\gamma)) - \mu_0\right| \le C\left(\left|\psi'(t_n(\gamma))\right| + \log(n) \cdot n^{\gamma - 1/2}\right).$$

Consequently, $\sigma_0^2(\varphi_n; \hat{t}_n(\gamma))$ is uniformly consistent for σ_0^2 over $\Lambda_n(q, A; \mu_0, \sigma_0, \epsilon_0)$. In addition, if $\psi'(t_n(\gamma)) = o(1)$, then $\mu_0(\varphi_n; \hat{t}_n(\gamma))$ is consistent for μ_0 as well.

We remark here that $\mu_0(\varphi_n; \hat{t}_n(\gamma))$ is uniformly consistent for μ_0 over any subset $\Lambda_n^* \subset \Lambda_n$ with $\sup_{\{\Lambda_n^*\}}\{|\psi'(t_n(\gamma))|\} = o(1)$. Although at first glance the convergence rates are relatively slow, in fact they are much faster in many situations.

2.3 Convergence Rate: Examples and Discussions

We now show that under mild conditions, the convergence rates of $\sigma_0^2(\varphi_n; \hat{t}_n(\gamma))$ and $\mu_0(\varphi_n; \hat{t}_n(\gamma))$ can be significantly improved and sometimes are algebraically fast.

Example 1: Asymptotically Vanishing Sparsity. Sparsity is a natural phenomenon found in many scientific fields, including genomics, astronomy, and image processing. As mentioned earlier, asymptotically vanishing sparsity refers to the case where $\epsilon_n(\mu, \sigma) \to 0$ (as $n \to \infty$). Several models for sparsity have been considered in the literature; among these are *moderately sparse* and *very sparse* models, where $\epsilon_n = n^{-\beta}$ for some parameter β satisfying $\beta \in (0, 1/2)$ and $\beta \in (1/2, 1)$ (Abramovich et al. 2006; Donoho and Jin 2004). Lemma A.5 shows that uniformly over Λ_n, $|\psi'(t_n(\gamma))| \le O(\epsilon_n(\mu,\sigma))$. Theorem 2 then yields the fact that the estimation errors of $\sigma_0^2(\varphi_n; \hat{t}_n(\gamma))$ and $\mu_0(\varphi_n; \hat{t}_n(\gamma))$ are algebraically small for both the moderately sparse and very sparse cases.

Example 2: Heteroscedasticity. It is natural in many applications to find that a nonnull effect has an elevated variance. A test statistic consists of two components: signal and noise. An elevation of variance occurs when the signal component contributes extra variance. Denote the minimum elevation of the variance for the nonnull effects by

$$\tau_n = \tau_n(\mu,\sigma) = \min_{\{j:\,(\mu_j,\sigma_j)\neq(\mu_0,\sigma_0)\}} \{\sigma_j^2 - \sigma_0^2\}. \qquad (19)$$

Lemma A.5 shows that $|\psi'(t_n(\gamma))| \le O(\epsilon_n e^{-\gamma \log(n)\tau_n(\mu,\sigma)})$. Thus $\psi'(t_n(\gamma)) = o(1)$ if, say, $\tau_n \ge \frac{\log\log n}{\log n}$, and $\psi'(t_n(\gamma))$ is algebraically small if $\tau_n \ge c_0$ for some constant $c_0 > 0$.

Example 3: Gaussian Hierarchical Model. The Gaussian hierarchical model is widely used in statistical inference as well as in microarray analysis (see, e.g., Efron 2004). A simple version of the model is one in which $\sigma_j \equiv \sigma_0$ and the means μ_j associated with nonnull effects are modeled as samples from a density function h, $(\mu_j|H_j \text{ is untrue}) \overset{\text{iid}}{\sim} h$. It is not hard to show that $|\psi'(t_n(\gamma))| \le \epsilon_n \cdot |\int e^{it_n(\gamma)u}[(u-\mu_0)h(u)]\,du|$, where the integral is the Fourier transform of the function $(u-\mu_0)h(u)$ at frequency $t_n(\gamma)$. By the Riemann–Lebesgue lemma (Mallat 1998), $|\psi'(t_n(\gamma))| = o(t_n^{-k}(\gamma))$ if the kth derivative of $h(u)$ is absolutely integrable. In particular, if h is Gaussian, [say $\mathrm{N}(a, b^2)$], then $|\psi'(t_n(\gamma))| \le O(\epsilon_n \cdot |t_n(\gamma)| \cdot n^{-\gamma b^2})$ and is algebraically small.

We note here that sparsity, heteroscedasticity, and the smoothness of h can occur simultaneously, making convergence even faster. In a sense, our approach combines the strengths of sparsity, heteroscedasticity, and smoothness of the density h. Thus the approach can be viewed as an extension of Efron's approach, because it is consistent not only in the asymptotically vanishingly sparse case, but also in many interesting nonsparse cases. In addition, in the asymptotically vanishingly sparse case, the convergence rates of our estimators can be substantially faster than those of Efron. This may occur when the dataset is both sparse and heteroscedastic, for example.

Remark. The theory developed in Sections 2.1–2.3 can be naturally extended to the Gaussian hierarchical model, which is the Bayesian counterpart of model (1)–(2) and has been widely used in the literature (see, e.g., Efron 2004; Genovese and Wasserman 2004). The model treats the test statistics X_j as samples from a two-component Gaussian mixture,

$$X_j \sim (1-\epsilon)\mathrm{N}(\mu_0, \sigma_0^2) + \epsilon \mathrm{N}(\mu_j, \sigma_j^2), \qquad 1 \le j \le n, \quad (20)$$

where the (μ_j, σ_j) are samples from a bivariate distribution $F(\mu,\sigma)$. The previous results can be naturally extended to this model.

2.4 Extension to Dependent Data Structures

We now consider the proposed approach for dependent data. Because the discussions are similar, we focus on $\sigma_0^2(\varphi; \hat{t}_n(\gamma))$. Recall that the estimation error splits into a stochastic component and a nonstochastic component, $|\sigma_0^2(\varphi_n; \hat{t}_n(\gamma)) - \sigma_0^2| \le |\sigma_0^2(\varphi_n; \hat{t}_n(\gamma)) - \sigma_0^2(\varphi; t_n(\gamma))| + |\sigma_0^2(\varphi; t_n(\gamma)) - \sigma_0^2|$. Note that the nonstochastic component contains only marginal effects and is unrelated to dependence structures. Thus we need only study the stochastic component, or to extend Theorem 1. In fact, once Theorem 1 is extended to the dependent case, the extension of Theorem 2 follows directly by arguments similar to those given in the proof of Theorem 2. For reasons of space, we focus on two dependent structures: the strongly (α)-mixing case and the short-range dependent case. Denote the strongly mixing coefficients by $\alpha(k) = \sup_{\{1\le t\le n\}} \alpha(\sigma(X_s, s \le t), \sigma(X_s, s \ge t+k))$, where $\sigma(\cdot)$ is the σ-algebra generated by the random variables specified in the brackets and $\alpha(\Sigma_1, \Sigma_2) = \sup_{\{E_1\in\Sigma_1, E_2\in\Sigma_2\}} |P\{E_1 \cap E_2\} - P\{E_1\}P\{E_2\}|$ for any two σ-algebras Σ_1 and Σ_2. In the strongly mixing case, we suppose that $\alpha(k) \le Bk^{-d}$ for some positive constants B and d. In the short-range dependent case, we suppose that $\alpha(k) = 0$ when $k \ge n^\tau$ for some constant $\tau \in (0,1)$.

Now, fix constants $a > 0$, $B > 0$, $q \ge 3$, and $A > 0$, introduce the following set of parameters, which we denote by $\tilde{\Lambda}_n(a, B, q, A) = \tilde{\Lambda}_n(a, B, q, A; \epsilon_0, \mu_0, \sigma_0)$:

$$\Big\{(\mu,\sigma) \in \Lambda_n(q, A; \mu_0, \sigma_0, \epsilon_0), \\ \max_{\{1\le j\le n\}} \{|\mu_j| + |\sigma_j|\} \le B\log^a(n)\Big\}.$$

Note that this technical condition is not essential and can be relaxed. The following theorem treats the strongly mixing case and was proven by Jin and Cai (2006, sec. 7).

Theorem 3. Fix $d > 1.5$, $q \ge 3$, $\gamma \in (0, \frac{d-1.5}{2d+2.5})$, $A > 0$, $a > 0$, and $B > 0$. Suppose that $\alpha(k) \le Bk^{-d}$ for all $1 \le k \le n$. As $n \to \infty$, uniformly for all $(\mu,\sigma) \in \tilde{\Lambda}_n(a, B, q, A)$, except for an event with asymptotically vanishing probability,

$$\big|\sigma_0^2(\varphi_n; \hat{t}_n(\gamma)) - \sigma_0^2(\varphi; t_n(\gamma))\big| \le \bar{o}\big(n^{\gamma-1/2}\big),$$
$$\big|\mu_0(\varphi_n; \hat{t}_n(\gamma)) - \mu_0(\varphi; t_n(\gamma))\big| \le \bar{o}\big(n^{\gamma-1/2}\big).$$

An interesting question is whether this result holds for all $\gamma \in (0, 1/2)$; we leave this for future study. The following theorem concerns the short-range dependent case, the proof of which is similar to that of Theorem 3 and thus is omitted.

Theorem 4. Fix $q \ge 3$, $\tau \in (0,1)$, $\gamma \in (0, \frac{1-\tau}{2})$, $A > 0$, $a > 0$, and $B > 0$. Suppose that $\alpha(k) = 0$ for all $k \ge n^\tau$. As $n \to \infty$, uniformly for all $(\mu,\sigma) \in \tilde{\Lambda}_n(a, B, q, A)$, except for an event with asymptotically vanishing probability,

$$\big|\sigma_0^2(\varphi_n; \hat{t}_n(\gamma)) - \sigma_0^2(\varphi; t_n(\gamma))\big| \le \bar{o}\big(n^{\gamma-(1-\tau)/2}\big),$$
$$\big|\mu_0(\varphi_n; \hat{t}_n(\gamma)) - \mu_0(\varphi; t_n(\gamma))\big| \le \bar{o}\big(n^{\gamma-(1-\tau)/2}\big).$$

We mention that consistency for more general dependent settings is possible provided that the following two key requirements are satisfied. First, there must be an exponential-type inequality for the tail probability of $|\varphi_n(t) - \varphi(t)|$ for all $t \in (0, \log n)$; we use Hoeffding's (1963) inequality in the proof for the independent case, and use (Bosq 1998, thm. 1.3) in the proof of Theorem 3. Second, the standard deviation of $\varphi_n(t_n(\gamma))$ must have a smaller order than that of $\varphi(t_n(\gamma))$, so that the approximation $\varphi_n(t_n(\gamma))/\varphi(t_n(\gamma)) \approx 1$ is accurate to the first order.

3. ESTIMATING THE PROPORTION OF NONNULL EFFECTS

The development of useful estimator for the proportion of nonnull effects, together with the corresponding statistical analysis, poses many challenges. Recent work includes that of Meinshausen and Rice (2006), Swanepoel (1999), Cai et al. (2007), and Jin (2006) (see also Efron et al. 2001; Genovese and Wasserman 2004). The first two approaches provide consistent estimators only under a condition that Genovese and Wasserman (2004) called "purity." These approaches do not perform well in the current setting, because the purity condition is not satisfied; see Lemma 1 for details. Cai et al. (2007) focused largely on a very sparse setting, and so a more specific model is needed. Jin (2006) considered estimating the proportion of nonzero normal means but concentrated on the homoscedastic case with known null parameters. This motivates careful study of estimation of the proportion in the current setting.

We begin by first assuming that the null parameters are known. In this case the approach of Jin (2006) can be extended to the heteroscedastic setting here. Fix $\gamma \in (0, \frac{1}{2})$. The following estimator was proposed by Jin (2006) for the homoscedastic case:

$$\hat{\epsilon}_n(\gamma) = \hat{\epsilon}_n(\gamma; X_1, \ldots, X_n, n) = \sup_{\{0 \le t \le \sqrt{2\gamma \log n}\}} \{1 - \Omega_n(t; X_1, \ldots, X_n, n)\}, \qquad (21)$$

where $\Omega_n(t; X_1, \ldots, X_n, n) = \int_{-1}^{1} (1 - |\xi|)(\mathrm{Re}(\varphi_n(t\xi; X_1, \ldots, X_n, n)e^{-i\mu_0 t\xi + \sigma_0^2 t^2 \xi^2/2}))\, d\xi$. This estimator continues to be consistent for the current heteroscedastic case. Set

$$\Theta_n(\gamma; q, A, \mu_0, \sigma_0, \epsilon_0) = \left\{ (\mu, \sigma) \in \Lambda_n(q, A; \mu_0, \sigma_0, \epsilon_0), \; \Delta_n \ge \frac{\log\log n}{\log n}, \epsilon_n(\mu, \sigma) \ge n^{\gamma - 1/2} \right\},$$

where $\Delta_n = \Delta_n(\mu, \sigma) = \min_{\{j:\,(\mu_j, \sigma_j) \ne (\mu_0, \sigma_0)\}} \{\max\{|\mu_j - \mu_0|^2, |\sigma_j^2 - \sigma_0^2|\}\}$.

Theorem 5. For any $\gamma \in (0, 1/2)$, $q \ge 1$, and $A > 0$, except for an event with algebraically small probability, $\lim_{n\to\infty}(\sup_{\{\Theta_n(\gamma; q, A, \mu_0, \sigma_0, \epsilon_0)\}} \{|\frac{\hat{\epsilon}_n(\gamma)}{\epsilon_n(\mu, \sigma)} - 1|\}) = 0$.

Roughly speaking, the estimator is consistent if the proportion is asymptotically larger than $1/\sqrt{n}$. The case in which the proportion is asymptotically smaller than $1/\sqrt{n}$ is very challenging, and in most cases constructing consistent estimates is very difficult without a more specific model (see Cai et al. 2007; Donoho and Jin 2004 for more discussion).

We now turn to the case in which the null parameters (μ_0, σ_0) are unknown. A natural approach is to first use the proposed procedures in Section 2.1 to obtain estimates for μ_0 and σ_0 (say $\hat{\mu}_0$ and $\hat{\sigma}_0$) and then plug them into (21) to estimate the proportion. This yields the estimate $\hat{\epsilon}_n^*(\gamma; \hat{\mu}_0, \hat{\sigma}_0) = \hat{\epsilon}_n^*(\gamma; \hat{\mu}_0, \hat{\sigma}_0, X_1, \ldots, X_n, n)$. Theorem 6 describes how $(\hat{\sigma}_0, \hat{\mu}_0)$ affects the estimation accuracy of $\hat{\epsilon}_n^*$.

Theorem 6. Fix $\epsilon_0 \in (0, 1/2)$, $\gamma \in (0, 1/2)$, $q \ge 1$, and $A > 0$. As $n \to \infty$, suppose that except for an event B_n with algebraically small probability, $\max\{|\hat{\mu}_0 - \mu_0|^2, |\hat{\sigma}_0^2 - \sigma_0^2|\} = o(\frac{1}{\log n})$. Then there are a constant $C = C(\gamma, q, A, \mu_0, \sigma_0, \epsilon_0) > 0$ and an event D_n with algebraically small probability, such that over $B_n^c \cap D_n^c$,

$$|\hat{\epsilon}_n^*(\gamma; \hat{\mu}_0, \hat{\sigma}_0) - \hat{\epsilon}_n(\gamma)| \le C \cdot \left[\log^{-3/2}(n) \cdot n^{\gamma - 1/2} + \log n \cdot |\hat{\sigma}_0^2 - \sigma_0^2| + \sqrt{\log n} \cdot |\hat{\mu}_0 - \mu_0|\right].$$

Results in previous sections show that under mild conditions, the estimation errors of $(\hat{\mu}_0, \hat{\sigma}_0)$ are algebraically small, as is $\hat{\epsilon}_n^*(\gamma) - \hat{\epsilon}_n(\gamma)$. In the nonsparse case, such differences are negligible, and both $\hat{\epsilon}_n(\gamma)$ and $\hat{\epsilon}_n^*(\gamma)$ are consistent. The sparse case is more subtle, especially when the proportion is algebraically small. In this case, a more specific model is often needed (see Cai et al. 2007).

We now compare our procedure with those of Meinshausen and Rice (2006) and Cai et al. (2007). We begin by introducing the aforementioned purity condition. If we model the p values of the test statistics as samples from a mixing density $(1 - \epsilon)\mathrm{U}(0, 1) + \epsilon h$, where $\mathrm{U}(0, 1)$ and h are the marginal densities of the p values for the null effects and nonnull effects. The purity condition is defined as $\operatorname{essinf}_{\{0<p<1\}} h(p) = 0$. Meinshausen and Rice (2006) proposed a confidence lower bound for ϵ that is valid for all h. Despite this advantage, however, the lower bound is generally conservative and inconsistent. In fact, the purity condition is necessary for the lower bound to be consistent. Similar results were given by Genovese and Wasserman (2004). Unfortunately, the purity condition generally does not hold in our settings.

Lemma 1. Let the test statistics X_j be given as in (20). If the marginal distribution $F(\mu, \sigma)$ satisfies either $P_F\{\sigma > 1\} \ne 0$ or $P_F\{\sigma = 1\} = 1$, but $P_F\{\mu > 0\} \ne 0$ and $P_F\{\mu < 0\} \ne 0$, then the purity condition does not hold.

Cai et al. (2007) considered a very sparse setting for a two-point Gaussian mixture model in which the proportion is modeled as $n^{-\beta}$ with $\beta \in (\frac{1}{2}, 1)$. Their estimator is consistent whenever consistent estimation is possible, and it attains the optimal rate of convergence. In a sense, their approach complements our method; the former deals with a very sparse but more specific model, and the latter deals with a more general model in which the level of sparsity is much lower.

4. SIMULATION EXPERIMENTS

We now turn to the numerical performance of our estimators of the null parameters. The goal of the simulation study is threefold: to investigate how different choices of γ affect the estimation errors, to compare the performance of our approach with that of Efron (2004), and to investigate the performance of the proposed approach for dependent data. We leave the study for real data to Section 5.

We first investigate the effect of γ on the estimation errors. We set $\sigma_0 = 1/\sqrt{2}$ and $\mu_0 = -1/2$ throughout this section. We take $n = 10{,}000$, $\epsilon = .1$, and $a = .75, 1.00, 1.25$, and 1.50 for the following simulation experiment:

Step 1. (Main step). For each a, first generate $n\epsilon$ pairs of (μ_j, σ_j) with μ_j from N(0, 1) and σ_j from the uniform distribution $U(a, a + .5)$, and then generate a sample from $N(\mu_j, \sigma_j^2)$ for each pair of (μ_j, σ_j). These $n\epsilon$ samples represent the nonnull effects. In addition, generate $n \cdot (1-\epsilon)$ samples from $N(\mu_0, \sigma_0^2)$ to represent the null effects.

Step 2. For the samples obtained in step 1, implement $\hat{\sigma}(\gamma) = \sigma_0(\varphi_n; \hat{t}_n(\gamma))$ and $\hat{\mu}_0(\gamma) = \mu_0(\varphi_n; \hat{t}_n(\gamma))$ for each $\gamma = .01, .02, \ldots, .5$.

Step 3. Repeat steps 1 and 2 for 100 independent cycles.

The results, reported in Figure 3, suggest that the best choice of γ for both $\hat{\sigma}_0(\gamma)$ and $\hat{\mu}_0(\gamma)$ are in the range (.1, .15). With γ in this range, the performance of the estimators is not very sensitive to different choices of γ, and both estimators are accurate. Taking $\gamma = .1$, for example, the mean squared errors (MSEs) for $\hat{\sigma}_0(\gamma)$ and $\hat{\mu}_0(\gamma)$ are of magnitude 10^{-4} and 10^{-3}. These suggest using the following estimators for simplicity, where we take $\gamma = .1$:

$$\hat{\sigma}_0^* = \sigma_0(\varphi_n; \hat{t}_n(.1)) \qquad \text{and} \qquad \hat{\mu}_0^* = \mu_0(\varphi_n; \hat{t}_n(.1)). \quad (22)$$

We now compare $(\hat{\sigma}_0^*, \hat{\mu}_0^*)$ with the estimators of Efron (2004). Recall that one major difference between the two approaches is that Efron's estimators are not consistent for the nonsparse case, whereas ours are. It is thus of interest to make comparisons at different levels of sparsity. To do so, we set a at 1, and let ϵ take four different values (.05, .10, .15, and .20) to represent different levels of sparsity. For each ϵ, we first generate samples according to the main step in the aforementioned experiment, then implement $(\hat{\sigma}_0^*, \hat{\mu}_0^*)$ and the estimators of Efron (2004), and finally repeat the experiment for 100 independent cycles. The results are reported in Figures 4 and 5.

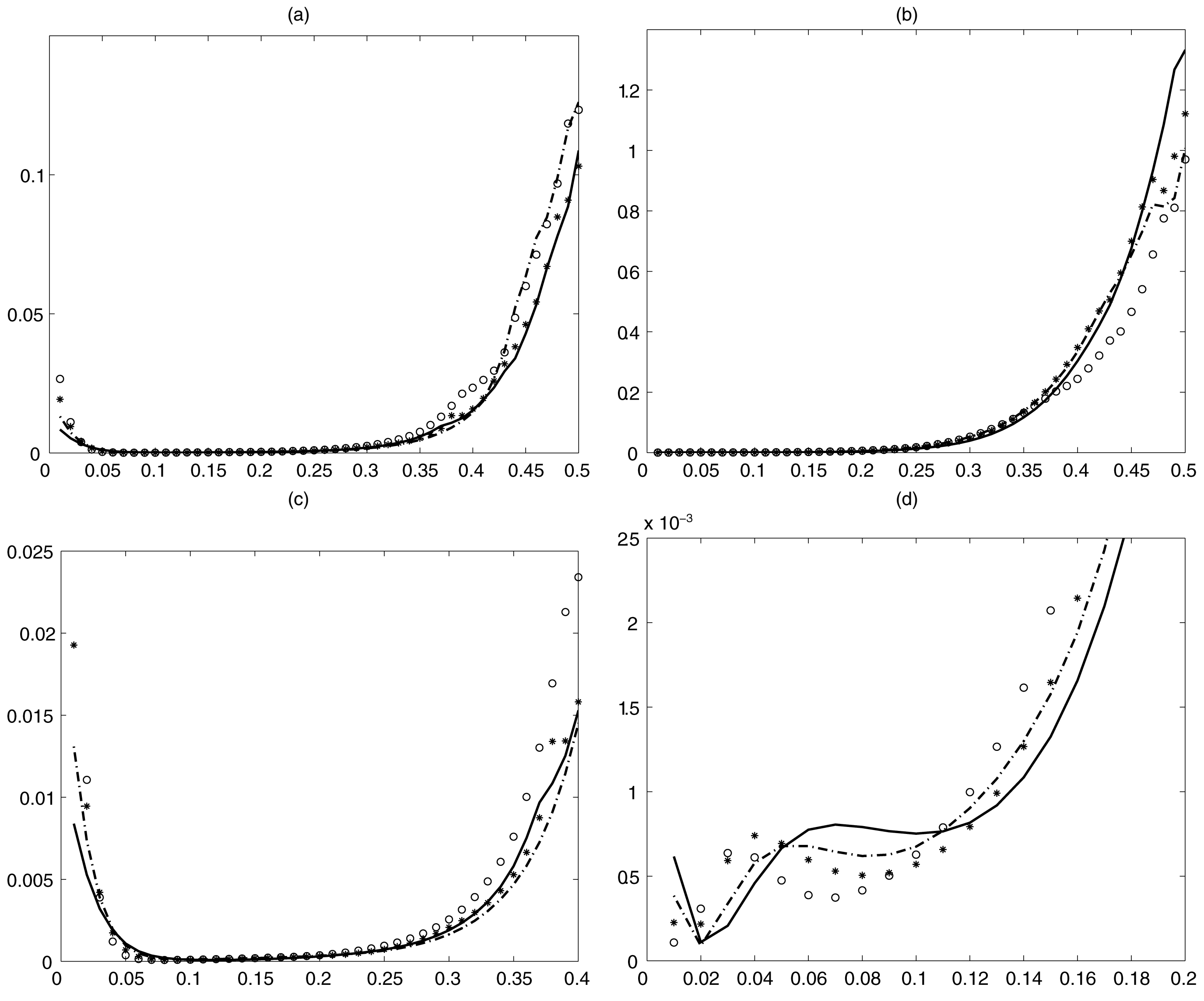

Figure 3. MSE for $\hat{\sigma}_0(\gamma)$ (a) and $\hat{\mu}_0(\gamma)$ (b) for $\gamma \in (0, 1/2]$. The four different curves (—, $-\cdot-\cdot$, $***$, and $\circ\circ\circ$) correspond to $a = .75$, 1.00, 1.25, and 1.50. [(c) and (d)] Zoom in.

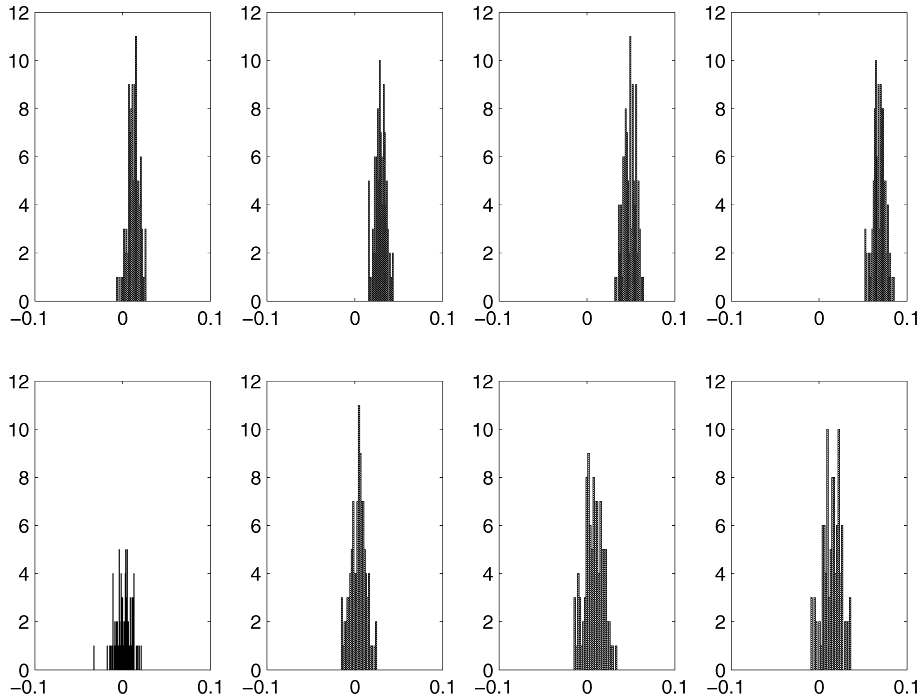

Figure 4. Histograms for the Estimation Errors of Efron's Estimator for $\hat{\sigma}_0$ (top row) and $\hat{\sigma}_0^*$ (bottom row). From left to right: $\epsilon = .05, .10, .15$, and .20.

The results show that our estimator of σ_0^2 is more accurate than that of Efron (2004), with the difference becoming more prominent as ϵ increases. In fact, when ϵ ranges between .05 and .2, the estimation errors of $\hat{\sigma}_0^*$ are of the order 10^{-2}, whereas those of Efron's estimator could become as large as the order 10^{-1}. On the other hand, the two estimators of μ_0 are almost equally accurate, and the estimation errors for both approaches fluctuate around .02 across different choices of ϵ.

But, the foregoing comparison is only for moderately large n. With a much larger n, Theorem 2 predicts that the estimation errors of $(\hat{\sigma}_0^*, \hat{\mu}_0^*)$ will become substantially smaller, because $(\hat{\sigma}_0^*, \hat{\mu}_0^*)$ is consistent for (σ_0, μ_0). In comparison, the errors of Efron's estimators will not become substantially smaller, be-

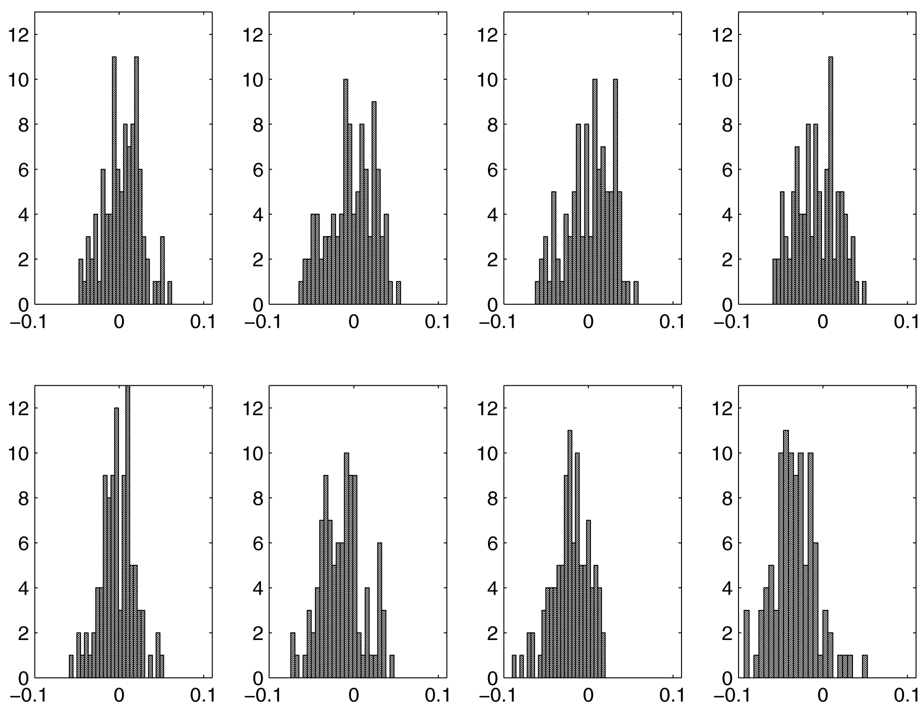

Figure 5. Histograms for the Estimation Errors of Efron's Estimator for $\hat{\mu}_0$ (top row) and $\hat{\mu}_0^*$ (bottom row). From left to right: $\epsilon = .05, .10, .15$, and .20.

Table 1. MSE for Various Values of n

n		10^4	4×10^4	1.6×10^5	6.4×10^5
MSE for σ_0	Efron's approach	9.100	8.564	8.415	8.567
	Our approach	.816	.276	.047	.031
MSE for μ_0	Efron's approach	8.916	5.905	3.957	3.617
	Our approach	5.807	3.019	1.106	.538

NOTE: The corresponding MSE equals the value in each cell times 10^{-4}.

cause the estimators are not consistent. To illustrate this point, we carry out a small-scale simulation experiment. We take $\epsilon = .1$ and $a = 1$ as before and let $n = 10^4$, 4×10^4, 1.6×10^5, and 6.4×10^5. For each n, we generate samples according to the main step, calculate the MSEs, and repeat the process for 30 independent cycles. The results, reported in Table 1, support the asymptotic analysis.

Finally, we investigate the performance of the proposed procedures for dependent data. We fix $n = 10^4$, $\epsilon = .1$, and $a = 1$ and let L range from 0 to 250 in increments of 5. For each L, we generate $n + L$ samples $w_1, w_2, \ldots, w_{n+L}$ from N(0, 1) and let $z_j = (\sum_{k=j}^{k=j+L} w_k)/\sqrt{L+1}$, so that $\{z_j\}_{j=1}^n$ are blockwise dependent (blocksize equal to $L+1$) and the marginal distribution of each z_j is N(0, 1). At the same time, we generate the mean vector μ and the vector of standard deviations σ according to the main step, let $X_j = \mu_j + \sigma_j \cdot z_j$, and implement $(\hat{\mu}_0^*, \hat{\sigma}_0^*)$ to $\{X_j\}_{j=1}^n$. We then repeat the process for 100 independent cycles. The results, reported in Figure 6, suggest that the estimation errors increase as the range of dependency increases. But, when $L \leq 100$, for example, the estimation errors are still relatively small, especially those for σ_0^*. This suggests that the procedures are relatively robust to short-range dependency.

5. APPLICATIONS TO MICROARRAY ANALYSIS

We now apply the proposed procedures to the analysis of the breast cancer and HIV microarray datasets analyzed by Efron (2004). The R code for our procedures is available at *http://www.stat.purdue.edu/~jinj/Research/software*. The z scores for both datasets can be downloaded from this site as well; they were kindly provided by Bradley Efron. The R code for Efron's procedures and related software can be downloaded from *http://cran.rproject.org/src/contrib/Descriptions/locfdr.html*. For reasons of space, we focus on the breast cancer data and comment only briefly on the HIV data.

The breast cancer data were based on 15 patients diagnosed with breast cancer, 7 with the BRCA1 mutation and 8 with the BRCA2 mutation. Each patient's tumor was analyzed on a separate microarray, and the microarrays were reported on the same set of $N = 3{,}226$ genes. For the jth gene, the two-sample t test comparing the seven BRCA1 responses with the eight BRCA2 responses was computed. The t score, y_j, was first converted to the p value by $p_j = \bar{F}_{13}(y_j)$, and was then converted to the z scale (Efron 2004), $X_j = \bar{\Phi}^{-1}(p_j) = \bar{\Phi}^{-1}(\bar{F}_{13}(y_j))$, where $\bar{\Phi}$ and $\bar{F}_{13}$ are the survival functions of N(0, 1) and t distribution with 13 degrees of freedom.

We model X_j as $N(\mu_j, \sigma_j^2)$ variables with weakly dependent structure, and for a pair of unknown parameters (μ_0, σ_0), $(\mu_j, \sigma_j) = (\mu_0, \sigma_0)$ if and only if the jth gene is not differentially expressed. Because X_j is transformed from the t score, which has been standardized by the corresponding standard error, it is reasonable to assume that the null effects are homogeneous and that all effects are homoscedastic (see, e.g., Cui and Churchill 2003; Efron 2004). The normality assumption is also

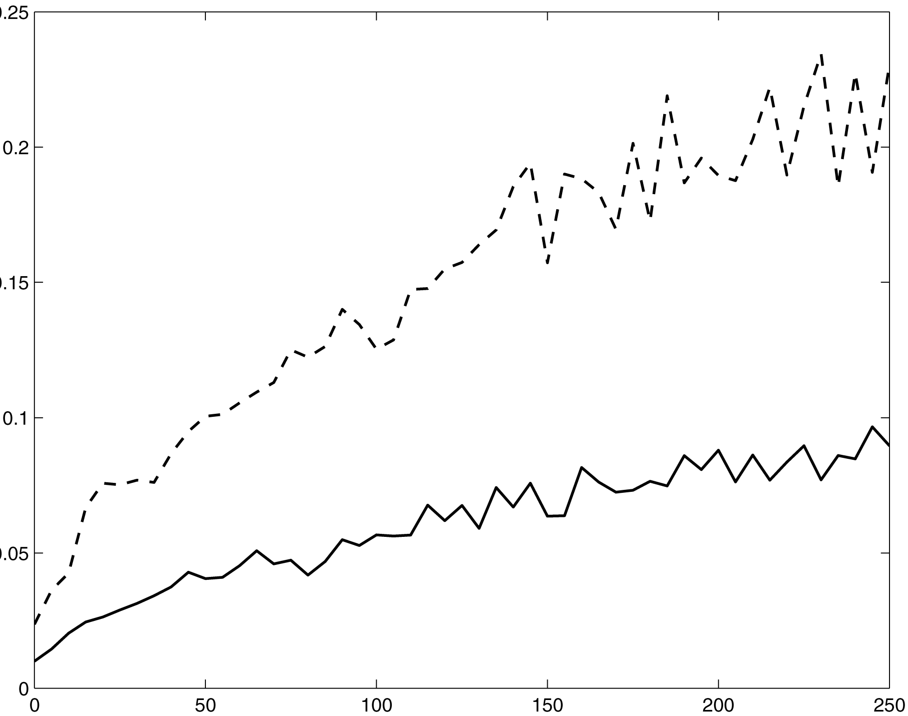

Figure 6. Root MSE for $\hat{\mu}_0^*$ (– – –) and $\hat{\sigma}_0^*$ (—).

reasonable here, because the marginal density of nonnull effects generally can be well approximated by Gaussian mixtures (see Efron 2004, p. 99). In particular, it is well known that the set of all Gaussian mixing densities is dense in the set of all density functions under the ℓ^1 metric.

We now proceed with the data analysis. The analysis comprises three parts: estimating the null parameters (σ_0, μ_0), estimating the proportion of nonnull effects, and implementing the local FDR approach proposed by Efron et al. (2001).

The first part is estimating (σ_0, μ_0). We apply $(\hat{\sigma}_0^*, \hat{\mu}_0^*)$ [defined in (22)] as well as the estimators used by Efron (2004) to the z scores. For the breast cancer data, our procedure yields $(\hat{\sigma}_0^*, \hat{\mu}_0^*) = (1.5277, -.0525)$, whereas Efron's estimators give $(\hat{\sigma}_0, \hat{\mu}_0) = (1.616, -.082)$.

The second part of the analysis involves estimating the proportion of nonnull effects. We implement our procedure as well as the approaches of Meinshausen and Rice (2006) and Cai et al. (2007) (which we denote by MR and CJL for short) to the z scores of the breast cancer data. The bounding function a_n^* for the MR estimator is set as $1.25 \times \sqrt{2\log\log n}/\sqrt{n}$, and the a_n for the CJL estimator is set as $\sqrt{2\log\log n}/\sqrt{n}$ (see Cai et al. 2007 for details). Using the estimated null parameters obtained using either Efron's approach or our approach, we apply each of these procedures to the z scores. In addition, the local FDR approach also automatically provides an estimate for the proportion. The results are reported in Table 2.

In the last part of the analysis, we implement the local FDR thresholding procedure proposed by Efron et al. (2001) with the z scores of the breast cancer data. For any given FDR-control parameter $q \in (0, 1)$, the procedure simultaneously calculates a score for each data point and determines a threshold t_q. A hypothesis is rejected if the score exceeds the threshold and is accepted otherwise. If we call a rejected hypothesis a "discovery," then the local FDR thresholding procedure controls the expected FDR at level q, $E[(\#\text{false discoveries})/(\#\text{total discoveries})] \le q$ (see Efron et al. 2001 for details).

With Efron's estimated null parameters, for any fixed $q \in (0, 1)$, the local FDR procedures report no rejections for the breast cancer dataset. Also, three different estimators for the proportion report 0. These suggest that either the proportion of signals (differentially expressed genes) is small or the signal is very weak.

In contrast, with our estimated null parameters, the estimated proportions are small but nonzero. Furthermore, the local FDR procedures report rejections when $q \ge .91$. For example, the number of total discoveries is 167 when $q = .92$ and 496 when $q = .94$. Take $q = .94$, for example, because for any $q \in (0, 1)$, the number of true discoveries is approximately equal to $(1 - q)$ times the number of total discoveries (Efron 2004), so we have approximately a total of 30 true discoveries. The result is consistent with biological discoveries. Among the 496 genes identified as being differentially expressed by the local FDR procedures, 17 of them have been discovered in the study by Hedenfalk et al. (2001). The corresponding Unigene cluster IDs are Hs.182278, Hs.82916, Hs.179661, Hs.119222, Hs.10247, Hs.469, Hs.78996, Hs.11951, Hs.79078, Hs.9908, Hs.5085, Hs.171271, Hs.79070, Hs.78934, Hs.469, Hs.197345, and Hs.73798. We also identified several genes whose functions are associated with the cell cycle, including PCNA, CCNA2, and CKS2. These genes were found to be significant by Storey et al. (2007). The results indicate that our estimated null parameters lead to reliable identification of differentially expressed genes.

Table 2. Estimated Proportion of Nonnull Effects for the Breast Cancer Data

	Our estimator	*Local FDR*	*MR*	*CJL*
Our estimated null	.0040	.0128	.0033	0
Efron's estimated null	0	0	.0098	0

Similarly, for the HIV data, our estimators give $(\hat{\sigma}_0^*, \hat{\mu}_0^*) = (.7709, -.0806)$, whereas Efron's method gives $(\hat{\sigma}_0, \hat{\mu}_0) = (.738, -.082)$. With $q = .05$, the local FDR procedures report 59 total discoveries with our estimated null parameters and 80 with Efron's estimated null parameters; the latter yields slightly more signals.

APPENDIX: PROOFS OF THE MAIN RESULTS

We now prove Theorems 1, 2, and 5. The proof of Theorem 6 is similar to those of Theorems 2 and 5 and thus is omitted. Because the proofs for the estimators of σ_0^2 and μ_0 are similar, we focus on σ_0^2. We first collect a few technical results and outline the basic ideas. The proofs of these preparatory lemmas were given by Jin and Cai (2006).

Lemma A.1. Let $\sigma_0^2(\cdot;\cdot)$ and $\mu_0(\cdot;\cdot)$ be defined as in (7). Fix $t > 0$. For any differentiable complex-valued functions f and g satisfying $|f(t)| \ne 0$ and $|g(t)| \ne 0$,

$$\begin{aligned} &|\sigma_0^2(f,t) - \sigma_0^2(g,t)| \\ &\quad \le \frac{|g(t)|}{t|f(t)|^2}\left[\left(2t \cdot |\sigma_0^2(g,t)| + \left|\frac{g'(t)}{g(t)}\right|\right)|f(t) - g(t)| \right. \\ &\qquad \left. + |f'(t) - g'(t)| + r_n^{(1)}(t)\right] \end{aligned}$$

and

$$\begin{aligned} &|\mu_0(f,t) - \mu_0(g,t)| \\ &\quad \le \frac{|g(t)|}{|f(t)|^2} \cdot \left[\left(2|\mu_0(g,t)| + \left|\frac{g'(t)}{g(t)}\right|\right) \cdot |f(t) - g(t)| \right. \\ &\qquad \left. + |f'(t) - g'(t)| + r_n^{(2)}(t)\right], \end{aligned}$$

where $r_n^{(1)}(t) = \frac{1}{|g(t)|} \cdot [t \cdot |\sigma_0^2(g,t)| \cdot |f(t) - g(t)|^2 + |f(t) - g(t)| \cdot |f'(t) - g'(t)|]$ and $r_n^{(2)}(t) = \frac{1}{|g(t)|} \cdot [|\mu_0(g,t)| \cdot |f(t) - g(t)|^2 + |f(t) - g(t)| \cdot |f'(t) - g'(t)|]$.

Heuristically, $|\varphi(\hat{t}_n)|/|\varphi_n(\hat{t}_n)|^2 \sim n^\gamma$, $\sigma_0^2(\varphi, \hat{t}_n) \sim \sigma_0^2$, $|\varphi'(\hat{t}_n)|/|\varphi(\hat{t}_n)| \sim \sigma_0^2 \hat{t}_n$, and

$$\begin{aligned} |\varphi_n(\hat{t}_n) - \varphi(\hat{t}_n)| &\le O_p(\sqrt{\log n}/\sqrt{n}), \\ |\varphi_n'(\hat{t}_n) - \varphi'(\hat{t}_n)| &\le O_p(\sqrt{\log n}/\sqrt{n}). \end{aligned} \tag{A.1}$$

Applying Lemma A.1 with $f = \varphi_n$, $g = \varphi$, and $t = \hat{t}_n(\gamma)$, we have

$$\begin{aligned} &\left|\sigma_0^2(\varphi_n, \hat{t}_n(\gamma)) - \sigma_0^2(\varphi, \hat{t}_n(\gamma))\right| \\ &\quad \sim n^\gamma \left(3\sigma_0^2 \left|\varphi_n(\hat{t}_n(\gamma)) - \varphi(\hat{t}_n(\gamma))\right| \right. \\ &\qquad \left. + \frac{1}{\hat{t}_n(\gamma)}\left|\varphi_n'(\hat{t}_n(\gamma)) - \varphi'(\hat{t}_n(\gamma))\right|\right) \\ &\quad \sim O\big(n^{\gamma - 1/2}\sqrt{\log n}\big), \end{aligned}$$

and Theorem 1 follows. We now study (A.1) in detail.

Lemma A.2. Set $W_0(\varphi_n; n) = W_0(\varphi_n; n, X_1, \ldots, X_n) = \sup_{0\le t\le \log n} |\varphi_n(t) - \varphi(t)|$. Fix $q_1 > 3$. Let $\Lambda_n(q, A; \mu_0, \sigma_0, \epsilon_0)$ be given as in Theorem 1. When $n \to \infty$,

$$\sup_{\{(\mu,\sigma)\in\Lambda_n(q,A;\mu_0,\sigma_0,\epsilon_0)\}} P\{W_0(\varphi_n; n) \ge \sqrt{2q_1 \log n}/\sqrt{n}\} \le 4\log^2(n) \cdot n^{-q_1/3} \cdot (1 + o(1)).$$

Lemma A.2 implies that except for an event with algebraically small probability, $|\varphi(\hat{t}_n) - \varphi(t_n)| \le W_0(\varphi_n; n) \le \sqrt{2q_1 \log n}/\sqrt{n}$. This naturally leads to a precise description of the stochastic behavior of $|\hat{t}_n(\gamma) - t_n(\gamma)|$ given in the following lemma.

Lemma A.3. Let $q_1 > 0$ and let $\Lambda_n(q, A; \mu_0, \sigma_0, \epsilon_0)$, $\hat{t}_n(\gamma)$, and $t_n(\gamma)$ be given as in Theorem 1. When $n \to \infty$,

$$\sup_{\{(\mu,\sigma)\in\Lambda_n(q,A;\mu_0,\sigma_0,\epsilon_0)\}} \left\{|\hat{t}_n(\gamma) - t_n(\gamma)| \cdot 1_{\{W_0(\varphi_n;n)\le\sqrt{2q_1\log n}/\sqrt{n}\}}\right\} \le \frac{1}{\sigma_0}\sqrt{\frac{q_1}{\gamma}} n^{\gamma-1/2}(1 + o(1)).$$

We now study $|\varphi_n'(\hat{t}_n) - \varphi'(\hat{t}_n)|$. Pick a constant $\pi_0 > \frac{1}{\sigma_0}\sqrt{q_1/\gamma}$ and set

$$\begin{aligned} W_1(\varphi_n, \gamma, \pi_0; n) &= W_1(\varphi_n, \gamma, \pi_0; n, X_1, \ldots, X_n) \\ &= \sup_{|t-t_n(\gamma)|\le\pi_0\cdot n^{\gamma-1/2}} |\varphi_n'(t) - \varphi'(t)|. \end{aligned}$$

By Lemma A.3, except for an event with algebraically small probability, $|\hat{t}_n(\gamma) - t_n(\gamma)| \le \pi_0 \cdot n^{\gamma-1/2}$, and, consequently, $|\varphi_n'(\hat{t}_n(\gamma)) - \varphi'(\hat{t}_n(\gamma))| \le W_1(\varphi_n, \gamma, \pi_0; n)$. The following lemma describes the tail behavior of W_1.

Lemma A.4. Fix $\gamma \in (0, 1/2)$, $\pi_0 > \frac{1}{\sigma_0}\sqrt{q_1/\gamma}$ and set $\bar{s}_n^2 = \frac{1}{n}\sum_{j=1}^n E[X_j^2]$. There exist constants C_1 and $C_2 > 0$ such that for any $(\mu, \sigma) \in \Lambda_n(q, A; \mu_0, \sigma_0, \epsilon_0)$, $\bar{s}_n \le C_1$,

$$P\left\{W_1(\varphi_n, \gamma, \pi_0; n) \ge \bar{s}_n \cdot \frac{\sqrt{(q-2)\log n} + 2\bar{s}_n}{\sqrt{n}}\right\} \le C_2 \cdot n^{-c_1(q,\gamma)},$$

where $c_1(q, \gamma)$ is as in Theorem 1. As a result, except for an event with algebraically small probability, $|\varphi_n'(\hat{t}_n(\gamma)) - \varphi'(\hat{t}_n(\gamma))| \le W_1(\varphi_n, \gamma, \pi; n) \le O(\sqrt{\log n}/\sqrt{n})$.

We have now elaborated the inequalities in (A.1). The only missing piece is the following lemma, which gives the basic properties of $\sigma_0^2(\varphi; t)$ and $\mu_0(\varphi; t)$.

Lemma A.5. Fix $q \ge 3$ and $A > 0$, with $\psi(t)$ and τ_n as defined in (18) and (19) write $\psi(t) = \epsilon_n g(t)$ and $r(t) = \frac{\epsilon_n}{1-\epsilon_n} r(t)$. For all $(\mu_0, \sigma_0, \epsilon_0)$-eligible (μ, σ) and all $t > 0$, there is a constant $C > 0$ such that

$$|\sigma_0^2(\varphi, t) - \sigma_0^2| \le \frac{|r'(t)|}{t} \cdot \frac{1 + |r(t)|}{|1 + r(t)|^2} \le C|\psi'(t)|/t \tag{A.2}$$

and

$$|\mu_0(\varphi, t) - \mu_0| \le |r'(t)| \cdot \frac{1 + |r(t)|}{|1 + r(t)|^2} \le C|\psi'(t)|. \tag{A.3}$$

In addition, uniformly for all $(\mu, \sigma) \in \Lambda_n(q, A; \mu_0, \sigma_0, \epsilon_0)$ and all $t > 0$, the following hold:

(a1) $|g(t)| \le e^{-(\tau_n^2 t^2)/2} \le 1$, $|g'(t)| \le A$, $|g''(t)| \le C(1 + A^2)$, $|g'''(t)| \le C(1 + A^3)$, and $|g'(t)| \le Ae^{-\tau_n t^2/2} + \min\{A^2 te^{-\tau_n t^2/2}, \frac{2}{et}\}$

(a2) Consequently, $|\varphi'(t)|/|\varphi(t)| = \sigma_0^2 \cdot t \cdot (1 + o(1))$

(a3) The second derivative of $\sigma_0^2(\varphi; t)$ is uniformly bounded, and $\sigma_0^2(\varphi; t) \to \sigma_0^2$, $\frac{d}{dt}\sigma_0^2(\varphi; t) \to 0$ as $t \to \infty$.

Similarly, both $\mu_0(\varphi; t)$ and its first two derivatives are uniformly bounded for all $t > 0$, and $\frac{d}{dt}\mu_0(\varphi; t) \to 0$ if $\mu_0(\varphi; t) \to \mu_0$.

We now prove Theorems 1, 2, and 5.

Proof of Theorem 1

Because the arguments are similar, we prove the first claim only. Write $\hat{t}_n = \hat{t}_n(\gamma)$, $t_n = t_n(\gamma)$, and $W_1(\varphi_n; n) = W_1(\varphi_n, \gamma, \pi_0; n)$. Pick constants q_1 and π_0 such that $1 < q_1/\max\{3, (q - 1 - 2\gamma)\} < 2$ and $\pi_0 > \frac{1}{\sigma_0}\sqrt{q_1/\gamma}$. Introduce events

$$B_0 = \{W_0(\varphi_n; n) \le \sqrt{2q_1 \log n}\},$$

$$B_1 = \left\{W_1(\varphi_n; n) \le \frac{s_n\sqrt{(q-2)\log n} + 2s_n^2}{\sqrt{n}}\right\}.$$

Note that the choice of q_1 satisfies $c_1(q, \gamma) < q_1/3$ and $c_2(\sigma_0, q, \gamma) > \sigma_0^2\sqrt{2q_1}$, where $c_1(q, \gamma)$ and $c_2(\sigma_0, q, \gamma)$ are as defined in Theorem 1. Use Lemmas A.2 and A.4, $P\{B_0^c\} \le \bar{o}(n^{-q_1/3})$ and $P\{B_1^c\} \le \bar{o}(n^{-c_1(q,\gamma)})$; because $c_1(q, \gamma) < q_1/3$, $P\{B_0^c \cup B_1^c\} \le \bar{o}(n^{-c_1(q,\gamma)})$. We now focus on $B_0 \cap B_1$. By the triangle inequality, $|\sigma_0^2(\varphi_n; \hat{t}_n) - \sigma_0^2(\varphi; t_n)| \le |\sigma_0^2(\varphi_n; \hat{t}_n) - \sigma_0^2(\varphi; \hat{t}_n)| + |\sigma_0^2(\varphi; \hat{t}_n) - \sigma_0^2(\varphi; t_n)|$. Note that by the choice of π_0 and Lemma A.3, $|\hat{t}_n - t_n| \le \pi_0 \cdot n^{\gamma-1/2}$ for sufficiently large n, it follows from Lemma A.5 that $|\sigma_0^2(\varphi; \hat{t}_n) - \sigma_0^2(\varphi; t_n)| \sim o(|\hat{t}_n - t_n|) = o(n^{\gamma-1/2})$. Recall that $c_2(\sigma_0, q, \gamma) > \sigma_0^2\sqrt{2q_1}$; thus, to show the claim, it suffices to show that as $n \to \infty$,

$$|\sigma_0^2(\varphi_n; \hat{t}_n) - \sigma_0^2(\varphi; \hat{t}_n)| \le 3\sigma_0^2 \cdot \sqrt{2q_1 \log n} \cdot n^{\gamma-1/2} \cdot (1 + o(1)) \quad \text{over } B_0 \cap B_1. \tag{A.4}$$

We now show (A.4). Over the event $B_0 \cap B_1$, recall that $|\hat{t}_n - t_n| \le \pi_0 n^{\gamma-1/2}$, so by (13), $\hat{t}_n \sim t_n \sim \sqrt{2\gamma \log n}/\sigma_0$; by Lemma A.5, this implies that $\sigma_0^2(\varphi, \hat{t}_n) \sim \sigma_0^2$ and $|\varphi'(\hat{t}_n)|/|\varphi(\hat{t}_n)| \sim \sigma_0^2\hat{t}_n \sim \sigma_0^2 t_n$. Moreover, because $|\varphi_n(\hat{t}_n) - \varphi(\hat{t}_n)| \le \sqrt{2q_1 \log n}/\sqrt{n}$, it follows that $|\varphi(\hat{t}_n)|/(\hat{t}_n|\varphi_n(\hat{t}_n)|^2) \sim (1/t_n)n^\gamma$. Finally, by Lemma A.4, $|\varphi_n'(\hat{t}_n) - \varphi'(\hat{t}_n)| \le O(\sqrt{\log n}/\sqrt{n})$. Combining these, (A.4) follows directly by applying Lemma A.1 with $f = \varphi_n$, $g = \varphi$, and $t = \hat{t}_n$.

Proof of Theorem 2

Note that by the triangle inequality, $|\sigma_0^2(\varphi_n; \hat{t}_n(\gamma)) - \sigma_0^2| \le |\sigma_0^2(\varphi_n; \hat{t}_n(\gamma) - \sigma_0^2(\varphi; t_n(\gamma))| + |\sigma_0^2(\varphi; t_n(\gamma)) - \sigma_0^2|$. Theorem 2 now follows directly from Theorem 1 and Lemma A.5.

Proof of Theorem 5

Without loss of generality, set $\mu_0 = 0$ and $\sigma_0 = 1$. Write $t_n = \sqrt{2\gamma \log n}$, $\epsilon_n = \epsilon_n(\mu, \sigma)$, $\varphi_n(t) = \varphi_n(t; X_1, \ldots, X_n, n)$, $\varphi(t) = \varphi(t; \mu, \sigma, n)$, $\Omega_n(t) = \Omega_n(t; X_1, \ldots, X_n, n)$, and $\Theta_n = \Theta_n(\gamma; q, A, \mu_0, \sigma_0, \epsilon_0)$. Set $\Omega(t) = E[\Omega_n(t)]$, $\Psi_n^*(t) = \sup_{\{0\le s\le t\}}\{1 - \Omega_n(s)\}$, and $\Psi^*(t) = \sup_{\{0\le s\le t\}}\{1 - \Omega(s)\}$. Note that it is sufficient to show that when $n \to \infty$, (a) except for an event with algebraically small probability, $\sup_{\{(\mu,\sigma)\in\Theta_n\}} |\Psi_n^*(t_n) - \Psi^*(t_n)| \le O(\log^{-3/2}(n) \cdot n^{\gamma-1/2})$, and (b) $\sup_{\{(\mu,\sigma)\in\Theta_n\}} |\frac{\Psi^*(t_n)}{\epsilon_n} - 1| = o(1)$.

We first show (a). By symmetry, $|\Psi_n^*(t_n) - \Psi^*(t_n)|$ does not exceed

$$\sup_{0\le t\le t_n} |\Omega_n(t) - \Omega(t)| \le 2\int_0^1 (1 - \xi)e^{t_n^2\xi^2/2} \sup_{0\le t\le t_n} \left|\mathrm{Re}(\varphi_n(t)) - \mathrm{Re}(\varphi(t))\right| d\xi. \tag{A.5}$$

Moreover, similar to the proof of lemma 7.2 of Jin (2006), we have that for fixed $q > 3/2$, $\sup_{\{(\mu,\sigma)\in\Theta_n\}} \sup_{\{0\le t\le t_n\}} |\mathrm{Re}(\varphi_n(t)) - \mathrm{Re}(\varphi(t))| \le O(\sqrt{\log n}/\sqrt{n})$, except for an event with probability $\sim 2\log^2(n) \cdot n^{-2q/3}$. Elementary calculus yields $|\Psi_n^*(t_n) - \Psi^*(t_n)| \le O(\sqrt{\log n}/\sqrt{n}) \cdot \int_0^1 (1 - \xi)e^{(\gamma \log n)\cdot\xi^2} d\xi = O(\log^{-3/2}(n) \cdot n^{\gamma-1/2})$, and (a) follows.

We now show (b). Let $\hat{f}$ be the Fourier transform of f, and let $\phi_{\delta_j(t)}(x)$ be the density function of $N(0, \delta_j^2(t))$ with $\delta_j(t) = t(\sigma_j^2 - 1)^{1/2}$. Set $\rho(x) = 2(1 - \cos(x))/x^2$ for $x \neq 0$ and $\rho(0) = 1$. Elementary calculus shows that $\hat{\phi}_{\delta_j(t)}(\xi) = \exp((1 - \sigma_j^2)t^2\xi^2/2)$ and $\hat{\rho}(\xi) = \max\{1 - |\xi|, 0\}$. So, by the Fourier inversion theorem (Mallat 1998, p. 22),

$$\begin{aligned}\Omega(t) &= \frac{1}{n}\sum_{j=1}^{n}\int_{-1}^{1}(1 - |\xi|)\exp\left(\frac{(1 - \sigma_j^2)t^2\xi^2}{2}\right)\cos(t\mu_j\xi)\,d\xi \\ &= \frac{1}{n}\sum_{j=1}^{n}\int_{-1}^{1}\hat{\phi}_{\delta_j(t)}(\xi)\hat{\rho}(\xi)\cos(t\mu_j\xi)\,d\xi \\ &= \frac{1}{n}\sum_{j=1}^{n}\phi_{\delta_j(t)} * \rho(t\mu_j),\end{aligned}$$

where $*$ is the usual convolution. Because $\phi_{\delta_j(t)} * \rho(t\mu_j) = 1$ when $(\mu_j, \sigma_j) = (0, 1)$,

$$1 - \Omega(t) = \epsilon_n \cdot \mathrm{Ave}_{\{j:\,(\mu_j,\sigma_j)\neq(0,1)\}}\big\{1 - \phi_{\delta_j(t)} * \rho(t\mu_j)\big\}. \tag{A.6}$$

Note that $\phi_{a_n} * \rho(b_n) \to 0$ for any sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ satisfying $\max\{a_n, b_n\} \to \infty$, and thus by (A.6) and the definition of Θ_n, $\sup_{\{(\mu,\sigma)\in\Theta_n\}} |\frac{1-\Omega(t_n)}{\epsilon_n} - 1| = o(1)$. Note that $0 \leq \phi_{\delta_j(t)} * \psi(t) \leq 1$ for all t, so by (A.6) and the definition of Ψ^*, $\Omega(t_n) \leq \Psi^*(t_n) \leq \epsilon_n$; as a result, $|\frac{1-\Psi^*(t_n)}{\epsilon_n} - 1| \leq |\frac{1-\Omega(t_n)}{\epsilon_n} - 1|$, and (b) follows directly.

[Received May 2006. Revised November 2006.]

REFERENCES

Abramovich, F., Benjamini, Y., Donoho, D., and Johnstone, I. (2006), "Adapting to Unknown Sparsity by Controlling the False Discovery Rate," *The Annals of Statistics*, 34, 584–653.

Benjamini, Y., and Hochberg, Y. (1995), "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society*, Ser. B, 57, 289–300.

Bosq, D. (1998), *Nonparametric Statistics for Stochastic Processes* (2nd ed.), New York: Springer-Verlag.

Cai, T., Jin, J., and Low, M. (2007), "Estimation and Confidence Sets for Sparse Normal Mixtures," *The Annals of Statistics*, in press.

Cui, X., and Churchill, G. (2003), "Statistical Tests for Different Expression in cDNA Microarray Experiments," *Genome Biology*, 4, 210.

Donoho, D., and Jin, J. (2004), "Higher Criticism for Detecting Sparse Heterogeneous Mixtures," *The Annals of Statistics*, 32, 962–994.

Efron, B. (2004), "Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis," *Journal of the American Statistical Association*, 99, 96–104.

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

Fan, J. (1991), "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems," *The Annals of Statistics*, 19, 1257–1272.

Genovese, C., and Wasserman, L. (2004), "A Stochastic Process Approach to False Discovery Control," *The Annals of Statistics*, 32, 1035–1061.

Hedenfalk, I., Duggen, D., Chen, Y., et al. (2001), "Gene Expression Profiles in Hereditary Breast Cancer," *New England Journal of Medicine*, 344, 539–548.

Hoeffding, W. (1963), "Probability Inequality for Sums of Bounded Random Variables," *Journal of the American Statistical Association*, 58, 13–30.

Jin, J. (2006), "Proportion of Nonzero Normal Means: Universal Oracle Equivalences and Uniformly Consistent Estimations," technical report, Purdue University, Statistics Department.

Jin, J., and Cai, T. (2006), "Estimating the Null and the Proportion of Non-Null Effects in Large-Scale Multiple Comparisons," available at *arXiv:math.ST/0611108* and *http://www.amstat.org/publications/jasa/supplemental_materials/*.

Lönnstedt, I., and Speed, T. (2002), "Replicated Microarray Data," *Statistica Sinica*, 12, 31–46.

Mallat, S. (1998), *A Wavelet Tour of Signal Processing*, New York: Academic Press.

Meinshausen, M., and Rice, J. (2006), "Estimating the Proportion of False Null Hypotheses Among a Large Number of Independent Tested Hypotheses," *The Annals of Statistics*, 34, 373–393.

Storey, J. D., Dai, J. Y., and Leek, J. T. (2007), "The Optimal Discovery Procedure for Large-Scale Significance Testing, With Applications to Comparative Microarray Experiments," *Biostatistics*, in press.

Swanepoel, J. W. H. (1999), "The Limiting Behavior of a Modified Maximal Symmetric 2*s*-Spacing With Applications," *The Annals of Statistics*, 27, 24–35.

Van't Wout, A., Lehrma, G., Mikheeva, S., O'Keeff, G., Katze, M., Bumgarner, R., Geiss, G., and Mullins, J. (2003), "Cellular Gene Expression Upon Human Immunodeficiency Virus Type 1 Infection of CD4$^+$T-Cell Lines," *Journal of Virology*, 77, 1392–1402.

Zhang, C.-H. (1990), "Fourier Methods for Estimating Mixing Densities and Distributions," *The Annals of Statistics*, 18, 806–831.