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

Jiashun Jin[1] and T. Tony Cai[2]

Abstract

An important issue raised by Efron [7] 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 paper, we consider the problem of estimating a null normal distribution, and a closely related problem, estimation of the proportion of non-null 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. Numerical performance of the estimators is investigated using both simulated and real data. In particular, we apply our

[1] Department of Statistics, Purdue University, West Lafayette, IN 47907. The research of Jiashun Jin is partially supported by NSF Grant DMS-0505423.

[2] Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104. The research of Tony Cai was supported in part by NSF Grants DMS-0306576 and DMS-0604954.

procedure to the analysis of breast cancer and HIV microarray data sets. The estimators perform favorably in comparison to existing methods.

Keywords: Empirical characteristic function, Fourier coefficients, multiple testing, null distribution, proportion of non-null effects, characteristic functions.

AMS 1991 subject classifications: Primary 62G10, 62G05; secondary 62G20.

1 Introduction

The analysis of massive data sets now commonly arising 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 (d.f.) F_0. That is,

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

Since the pioneering work of Benjamini and Hochberg [2], which introduced the False Discovery Rate (FDR)-controlling procedures, research on large-scale simultaneous testing has been very active. See, for example, [1, 4, 6, 7, 8, 10, 17, 19].

FDR procedures are based on the p-values, which measure the tail probability of the null distribution. Conventionally the null distribution is always assumed to be known. However, somewhat surprisingly, Efron pointed out in [7] that in many applications such an assumption would be incorrect. Efron [7] studied a data set on breast cancer, in which a gene microarray was generated for each patient in two groups, 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, though theoretically the null distribution should be the standard normal, empirically another null distribution (which Efron referred to as the *empirical null*) is found to be more appropriate. In fact, he found that $N(-0.02, 1.58^2)$ is a more appropriate null than $N(0,1)$; see Figure 1. A similar phenomenon is also found in the analysis of a microarray data set on HIV [7].

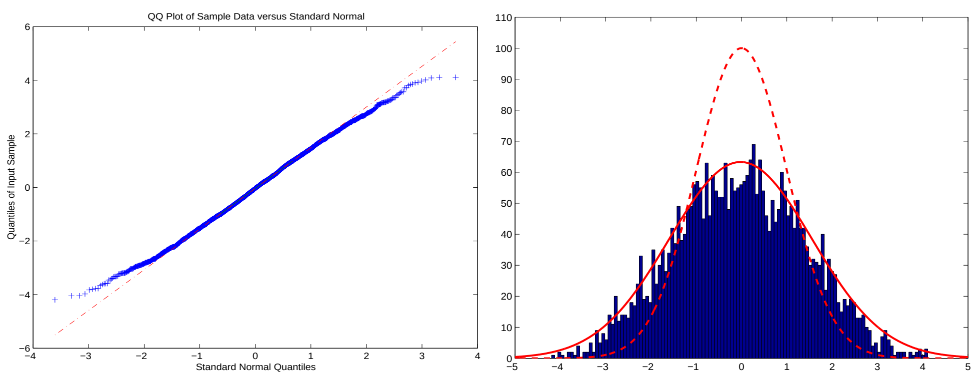

Figure 1: z-values of microarray data on breast cancer. Left panel: QQ-plot. Right panel: histogram and density curves of $N(0,1)$ (dashed) and $N(-0.02, 1.58^2)$. The plot suggests that the null is $N(-0.02, 1.58^2)$ rather than $N(0,1)$. See Efron [7] for further details.

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 which contains 100 independent cycles of simulations. In each cycle, 9000 samples are drawn from $N(0, 0.95^2)$ to represent the null effects, and 1000 samples are drawn from $N(2, 0.95^2)$ to represent the non-null effects. For each sample element X_j, p-values are calculated as $\bar{\Phi}^{-1}(X_j/0.95)$ and $\bar{\Phi}^{-1}(X_j)$, which represent the p-values under the *true* null and the *misspecified* null, respectively. The FDR procedure is then applied to both sets of p-values, where the FDR control parameter is set at 0.05. 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 $N(0, 0.95^2)$ as the null are considerably different. This, together with Efron's arguments, suggests that a careful study on estimating the null is indispensable.

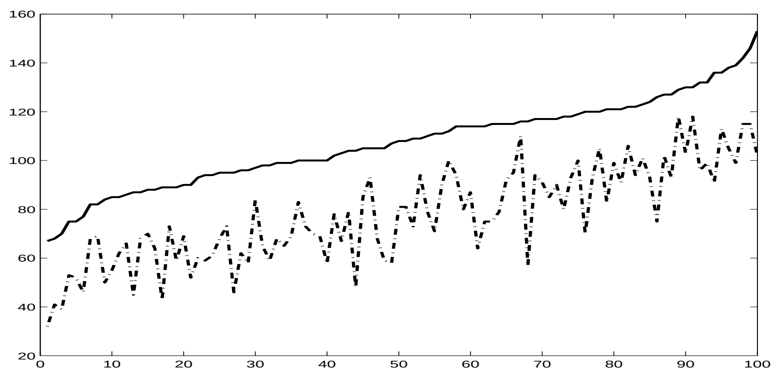

Figure 2: The solid and dashed curves represent the number of true positives for each cycle, using the true null and the misspecified null, respectively. For visualization, the numbers are sorted ascendingly with respect to those in the true null case.

Efron [7] introduced a method for estimating the null distribution based on the notion of "sparsity." There are several different ways to define sparsity [1]. The most intuitive one is that the proportion of non-null effects is small. In some applications, the case of "asymptotically vanishing sparsity" is of particular interest [1, 6]. This case refers to the situation where the proportion of non-null effects tends to zero as the number of hypotheses grows to infinity. In such a setting, heuristically, the influence of the non-null effects becomes more and more negligible and so the null can be reliably estimated asymptotically. In fact, Efron [7] suggested an approach which uses the center and half width of the central peak of the histogram for estimating the parameters of the null distribution.

In many applications it is more appropriate to model the setting as *non-sparse*, i.e., the proportion of non-null effects does *not* tend to zero when the number of hypotheses grows to infinity. In such settings, Efron's approach [7] 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, as 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 [7, 17, 20]. However, in the non-sparse case, neither is very informative as the relevant information about the null is highly distorted by the non-null effects in both of them. In this paper, we propose a new approach for estimating the null parameters by 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 non-null 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 outperform existing methods in simulations.

Beside the null distribution, the proportion of non-null effects is an important quantity. For example, the implementation of many recent procedures requires the knowledge of both the null and the proportion of non-null effects; see [8, 15, 19].

Developing good estimators for the proportion is a challenging task. Recent work includes that of Meinshausen and Rice [17], Swanepoel [20], Cai et al. [4], and Jin [13]. In this paper we extend the method of Jin [13] 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.

In addition to the theoretical properties, numerical performance of the estimators is investigated using both simulated and real data. In particular, we use our procedure to analyze the breast cancer [11] and HIV [21] microarray data that were analyzed in Efron [7]. The results indicate that our estimated null parameters lead to a more reliable identification of differentially expressed genes than that in [7].

The paper is organized as follows. In Section 2, after basic notations and definitions are reviewed, the estimators of the null parameters are defined in Section 2.1. The theoretical properties of the estimators are investigated in Sections 2.2 and 2.3. Section 2.4 discusses the extension to dependent data structures. Section 3 treats the estimation of the proportion of non-null effects. A simulation study is carried out in Section 4 to investigate numerical performance. In Section 5, we apply our procedure to the analysis of the breast cancer [11] and HIV [21] microarray data. Section 6 gives proofs of the main theorems.

2 Estimating the null distribution

As in Efron [7], we shall work on the z-scale and consider n test statistics

$$X_j \sim N(\mu_j, \sigma_j^2), \qquad 1 \le j \le n, \tag{2.1}$$

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

$$(\mu_j, \sigma_j) = (\mu_0, \sigma_0) \quad \text{if } H_j \text{ is true} \quad \text{and} \quad (\mu_j, \sigma_j) \ne (\mu_0, \sigma_0) \quad \text{if } H_j \text{ is untrue}, \tag{2.2}$$

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

Set $\mu = \{\mu_1, \ldots, \mu_n\}$ and $\sigma = \{\sigma_1, \ldots, \sigma_n\}$. Denote the proportion of non-null effects by

$$\epsilon_n = \epsilon_n(\mu, \sigma) = \frac{\#\{j : (\mu_j, \sigma_j) \neq (\mu_0, \sigma_0)\}}{n}. \tag{2.3}$$

We assume $\sigma_j \geq \sigma_0$ for all $1 \leq j \leq n$. That is, the standard deviation of a non-null effect is no less than that of a null effect. This is the case in a wide range of applications [7, 15]. To make the null parameters identifiable, we shall assume

$$\epsilon_n(\mu, \sigma) \leq \epsilon_0, \qquad \text{for some constant } 0 < \epsilon_0 < \tfrac{1}{2}. \tag{2.4}$$

Definition 2.1 *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 (2.4) is satisfied and $\sigma_j \geq \sigma_0$ for all $1 \leq j \leq n$.*

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

2.1 Estimating the null parameters

As mentioned in the Introduction, an informative approach for estimating the null distribution is to use the Fourier coefficients at suitable frequencies. In the literature, Fourier coefficients have been frequently used for statistical inference; see for example [9, 22]. We now use them to construct estimators for the null parameters.

Introduce the *empirical characteristic function*

$$\varphi_n(t) = \varphi_n(t; X_1, \ldots, X_n, n) = \frac{1}{n} \sum_{j=1}^{n} e^{itX_j}, \tag{2.5}$$

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

$\varphi(t) = \varphi_0(t) + \tilde{\varphi}(t)$, where $\varphi_0(t) = \varphi_0(t; \mu, \sigma, n) = (1 - \epsilon_n) \cdot e^{i\mu_0 t - \sigma_0^2 t^2/2}$ and

$$\tilde{\varphi}(t) = \tilde{\varphi}(t; \mu, \sigma, n) = \epsilon_n \cdot \text{Ave}_{\{j:\, (\mu_j, \sigma_j) \neq (\mu_0, \sigma_0)\}} \{e^{i\mu_j t - \sigma_j^2 t^2/2}\}, \tag{2.6}$$

which correspond to the null effects and non-null effects, respectively. Note that the identifiability condition $\epsilon_n \leq \epsilon_0 < 1/2$ ensures that $\varphi(t) \neq 0$ for all t.

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

$$\sigma_0^2(f; t) = -\frac{\frac{d}{dt}|f(t)|}{t \cdot |f(t)|}, \ \mu_0(f; t) = \frac{\text{Re}(f(t)) \cdot \text{Im}(f'(t)) - \text{Re}(f'(t)) \cdot \text{Im}(f(t))}{|f(t)|^2}, \tag{2.7}$$

where $\text{Re}(z)$ and $\text{Im}(z)$ denote respectively 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 $t \neq 0$.

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 non-null 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, respectively, and hope that by choosing an appropriate t,

$$\sigma_0^2(\varphi_n; t) \ \approx \ \sigma_0^2(\varphi; t) \ \approx \sigma_0^2(\varphi_0; t) \equiv \sigma_0^2, \tag{2.8}$$

$$\mu_0(\varphi_n; t) \ \approx \ \mu_0(\varphi; t) \ \approx \mu_0(\varphi_0; t) \equiv \mu_0. \tag{2.9}$$

There is clearly a tradeoff in the choice of t. As t increases from 0 to ∞, the second approximations in (2.8) and (2.9) 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

$$\hat{t}_n(\gamma) = \hat{t}_n(\gamma; \varphi_n) = \inf\{t : \ |\varphi_n(t)| = n^{-\gamma}, \ 0 \le t \le \log n\}. \tag{2.10}$$

Once we decide on the frequency $t = \hat{t}_n(\gamma)$, we have the following family of 'plug in' estimators which are indexed by $\gamma \in (0, 1/2)$:

$$\hat{\sigma}_0^2 = \sigma_0^2(\varphi_n; \hat{t}_n(\gamma)) \quad \text{and} \quad \hat{\mu}_0 = \mu_0(\varphi_n; \hat{t}_n(\gamma)). \tag{2.11}$$

We mention here that it will be shown later in Lemma 6.3 that $\hat{t}_n(\gamma)$ is asymptotically equivalent to the non-stochastic quantity

$$t_n(\gamma) = t_n(\gamma; \varphi) = \inf\{t : \ |\varphi(t)| = n^{-\gamma}, \ 0 \le t \le \log n\}, \tag{2.12}$$

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

$$t_n(\gamma, \varphi) = [\sqrt{2\gamma \log n}/\sigma_0] \cdot (1 + o(1)), \qquad n \to \infty, \tag{2.13}$$

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 (2.11) are consistent uniformly over a wide class of parameters. Introduce two non-stochastic 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, respectively. For each estimator, the estimation error can be decomposed into two components: one is the stochastic fluctuation and the other is the difference between the true parameter and its corresponding bridging quantity,

$$|\sigma_0^2(\varphi_n; \hat{t}_n(\gamma)) - \sigma_0^2| \le |\sigma_0^2(\varphi_n; \hat{t}_n(\gamma)) - \sigma_0^2(\varphi; t_n(\gamma))| + |\sigma_0^2(\varphi; t_n(\gamma)) - \sigma_0^2|, \tag{2.14}$$

$$|\mu_0(\varphi_n; \hat{t}_n(\gamma)) - \mu_0| \le |\mu_0(\varphi_n; \hat{t}_n(\gamma)) - \mu_0(\varphi; t_n(\gamma))| + |\mu_0(\varphi; t_n(\gamma)) - \mu_0|. \tag{2.15}$$

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

$$\Lambda_n(q, A; \mu_0, \sigma_0, \epsilon_0) = \{(\mu, \sigma) \text{ is } (\mu_0, \sigma_0, \epsilon_0)\text{-eligible}, \; M_n^{(q)}(\mu, \sigma) \leq A^q\}, \tag{2.16}$$

where $M_n^{(q)}(\mu, \sigma) = \text{Ave}_{\{j:\,(\mu_j,\sigma_j)\neq(\mu_0,\sigma_0)\}}\{(|\mu_j - \mu_0| + |\sigma_j^2 - \sigma_0^2|^{1/2})^q\}$. 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 2.1 *Fix constants $\gamma, \epsilon_0 \in (0, 1/2)$, $q \geq 3$, and $A > 0$. As $n \to \infty$, except for an event with probability $\bar{o}(n^{-c_1})$,*

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

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

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

$$c_1 = c_1(q, \gamma) = \begin{cases} (q/2 - 1 - \gamma)/2, & q < 4, \\ (q/2 - 1 - \gamma), & 4 \leq q \leq 4 + 2\gamma, \\ (q - 1 - 2\gamma)/3, & q > 4 + 2\gamma. \end{cases} \tag{2.17}$$

Theorem 2.1 says that the stochastic components in (2.14) and (2.15) are both algebraically small, uniformly over Λ_n.

We now consider the non-stochastic components in (2.14) and (2.15). As defined in (2.6), $\tilde{\varphi}(t)$ naturally factors into $\tilde{\varphi}(t) = e^{i\mu_0 t - \sigma_0^2 t^2/2} \cdot \psi(t)$, where

$$\psi(t) = \psi(t; \mu, \sigma, n) = \epsilon_n \cdot \text{Ave}_{\{j:\,(\mu_j,\sigma_j)\neq(\mu_0,\sigma_0)\}} e^{i(\mu_j - \mu_0)t - (\sigma_j^2 - \sigma_0^2)t^2/2}. \tag{2.18}$$

Lemma 6.5 in Section 6 tells us that there is a constant $C > 0$ such that uniformly for all $(\mu_0, \sigma_0, \epsilon_0)$-eligible parameters (μ, σ), $|\sigma_0^2(\varphi; t_n(\gamma)) - \sigma_0^2| \leq C \cdot |\psi'(t_n(\gamma))|/t_n(\gamma)$ and $|\mu_0(\varphi; t_n(\gamma)) - \mu_0| \leq C \cdot |\psi'(t_n(\gamma))|$; see details therein. Combining these with Theorem 2.1 gives the following theorem, which is proved in Section 6.

Theorem 2.2 *Fix constants* $\gamma, \epsilon_0 \in (0, 1/2)$*,* $q \geq 3$*, and* $A > 0$*. For all* t*,* $\sup_{\{\Lambda_n(q,A;\mu_0,\sigma_0,\epsilon_0)\}} |\psi'(t)| \leq A \cdot \epsilon_0$*. 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*,*

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

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

Consequently, $\sigma_0^2(\varphi_n; \hat{t}_n(\gamma))$ *is uniformly consistent for* σ_0^2 *over* $\Lambda_n(q, A; \mu_0, \sigma_0, \epsilon_0)$*. Additionally, 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, they are in fact 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 I. *Asymptotically vanishing sparsity.* Sparsity is a natural phenomenon found in many scientific fields such as genomics, astronomy, and image processing. As mentioned before, 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, and among them are *moderately sparse* and *very sparse*, where $\epsilon_n = n^{-\beta}$ for some parameter β satisfying $\beta \in (0, 1/2)$ and $\beta \in (1/2, 1)$, respectively [1, 6]. Lemma 6.5 shows that uniformly over Λ_n, $|\psi'(t_n(\gamma))| \leq O(\epsilon_n(\mu, \sigma))$. Theorem 2.2 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 case and the very sparse case.

Example II. *Heteroscedasticity.* It is natural in many applications to find that a non-null 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 non-null effects by

$$\tau_n = \tau_n(\mu, \sigma) = \min_{\{j:\, (\mu_j, \sigma_j) \neq (\mu_0, \sigma_0)\}} \{\sigma_j^2 - \sigma_0^2\}. \tag{2.19}$$

Lemma 6.5 shows that $|\psi'(t_n(\gamma))| \leq O(\epsilon_n e^{-\gamma \log(n) \tau_n(\mu,\sigma)})$. So $\psi'(t_n(\gamma)) = o(1)$ if, say, $\tau_n \geq \frac{\log \log n}{\log n}$, and $\psi'(t_n(\gamma))$ is algebraically small if $\tau_n \geq c_0$ for some constant $c_0 > 0$.

Example III. *Gaussian hierarchical model.* The Gaussian hierarchical model is widely used in statistical inference, as well as in microarray analysis; see Efron [7], for example. A simple version of the model is where $\sigma_j \equiv \sigma_0$ and the means μ_j associated with non-null effects are modeled as samples from a density function h, $(\mu_j | H_j \text{ is untrue}) \stackrel{iid}{\sim} h$. It is not hard to show that $|\psi'(t_n(\gamma))| \leq \epsilon_n \cdot | \int e^{i t_n(\gamma) u} [(u - \mu_0) h(u)] du|$, 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 [16], $|\psi'(t_n(\gamma))| = o(t_n^{-k}(\gamma))$ if the k-th derivative of $h(u)$ is absolutely integrable. In particular, if h is Gaussian, say $N(a, b^2)$, then $|\psi'(t_n(\gamma))| \leq O(\epsilon_n \cdot |t_n(\gamma)| \cdot n^{-\gamma b^2})$ and is algebraically small.

We note here that sparsity, heteroscedasticity, and the smoothness of h can occur at the same time, which makes the convergence even faster. In a sense, our approach combines the strengths of sparsity, heteroscedasticity, and the smoothness of the density h. The approach can thus be viewed as an extension of Efron's approach, as it is consistent not only in the asymptotically vanishingly sparse case, but also in many interesting non-sparse cases. Additionally, in the asymptotically vanishingly sparse case, the convergence rates of our estimators can be substantially faster than those of Efron. For example, this may occur when the data set is both sparse and heteroscedastic.

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 (2.1)-(2.2) and has been widely used in the literature; see for example [7, 10]. The model treats the test statistics X_j as samples from a two-component Gaussian mixture:

$$X_j \sim (1-\epsilon)N(\mu_0, \sigma_0^2) + \epsilon N(\mu_j, \sigma_j^2), \qquad 1 \leq j \leq n, \tag{2.20}$$

where (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. As 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 non-stochastic component, $|\sigma_0^2(\varphi_n; \hat{t}_n(\gamma)) - \sigma_0^2| \leq |\sigma_0^2(\varphi_n; \hat{t}_n(\gamma)) - \sigma_0^2(\varphi; t_n(\gamma))| + |\sigma_0^2(\varphi; t_n(\gamma)) - \sigma_0^2|$. Note that the non-stochastic component only contains marginal effects and is unrelated to dependence structures. We thus need only to study the stochastic component, or to extend Theorem 2.1. In fact, once Theorem 2.1 is extended to the dependent case, the extension of Theorem 2.2 follows directly by arguments similar to those given in the proof of Theorem 2.2. For reasons of space, we shall focus on two dependent structures: the strongly (α)-mixing case and the short-range dependent case. Denote the strongly mixing coefficients by $\alpha(k) = \sup_{\{1 \leq t \leq n\}} \alpha(\sigma(X_s, s \leq t), \sigma(X_s, s \geq t+k))$, 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 $\alpha(k) \leq Bk^{-d}$ for some positive constants B and d. In the short-range dependent case, we suppose $\alpha(k) = 0$ when $k \geq n^{\tau}$ for some constant $\tau \in (0, 1)$.

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

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

Note that this technical condition is not essential and can be relaxed. The following theorem treats the strongly mixing case and is proved in [14, Section 7].

Theorem 2.3 *Fix $d > 1.5$, $q \geq 3$, $\gamma \in (0, \frac{d-1.5}{2d+2.5})$, $A > 0$, $a > 0$, and $B > 0$. Suppose $\alpha(k) \leq Bk^{-d}$ for all $1 \leq k \leq n$. 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,*

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

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, whose proof is similar to that of Theorem 2.3 and is thus omitted.

Theorem 2.4 *Fix $q \geq 3$, $\tau \in (0, 1)$, $\gamma \in (0, \frac{1-\tau}{2})$, $A > 0$, $a > 0$, and $B > 0$. Suppose $\alpha(k) = 0$ for all $k \geq n^\tau$. 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,*

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

We mention that consistency for more general dependent settings is possible provided the following two key requirements are satisfied. First, there is 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 inequality in the proof for the independent case, and use [3, Theorem 1.3] in the proof of Theorem 2.3. Second, the standard deviation of $\varphi_n(t_n(\gamma))$ has 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 non-null effects

The development of useful estimator for the proportion of non-null effects, together with the corresponding statistical analysis, poses many challenges. Recent work includes those of Meinshausen and Rice [17], Swanepoel [20], Cai, et al. [4], and Jin [13]. See also [8, 10]. The first two approaches only provide consistent estimators under a condition which Genovese and Wasserman call "purity" [10]. These approaches do not perform well in the current setting as the purity condition is not satisfied; see Lemma 3.1 for details. Cai et al. [4] largely focuses on a very sparse setting, and so a more specific model is needed. Jin [13] considers estimating the proportion of nonzero normal means but concentrates on the homoscedastic case with known null parameters. This motivates a 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 [13] can be extended to the heteroscedastic setting here. Fix $\gamma \in (0, \frac{1}{2})$. The following estimator is proposed in [13] for the homoscedastic case:

$$\hat{\epsilon}_n(\gamma) = \hat{\epsilon}_n(\gamma; X_1, \ldots, X_n, n) = \sup_{\{0 \le t \le \sqrt{2\gamma \log n}\}} \{1 - \Omega_n(t; X_1, \ldots, X_n, n)\}, \tag{3.1}$$

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

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

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

Theorem 3.1 *For any $\gamma \in (0, 1/2)$, $q \ge 1$, and $A > 0$, except for an event with algebraically small probability,* $\lim_{n \to \infty} \big(\sup_{\{\Theta_n(\gamma; q, A, \mu_0, \sigma_0, \epsilon_0)\}} \{|\frac{\hat{\epsilon}_n(\gamma)}{\epsilon_n(\mu, \sigma)} - 1|\}\big) = 0$.

Roughly speaking, the estimator is consistent if the proportion is asymptotically larger than $1/\sqrt{n}$. The case where the proportion is asymptotically smaller than

$1/\sqrt{n}$ is very challenging, and usually it is very hard to construct consistent estimates without a more specific model; see [4, 6] for more discussion.

We now turn to the case where 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 (3.1) for estimation of 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 3.2 below describes how $(\hat{\sigma}_0, \hat{\mu}_0)$ affects the estimation accuracy of $\hat{\epsilon}_n^*$.

Theorem 3.2 *Fix $\epsilon_0 \in (0, 1/2)$, $\gamma \in (0, 1/2)$, $q \geq 1$, 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 $B_n^c \cap D_n^c$*

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

Results in previous sections show that, under mild conditions, the estimation errors of $(\hat{\mu}_0, \hat{\sigma}_0)$ are algebraically small, and so is $\hat{\epsilon}_n^*(\gamma) - \hat{\epsilon}_n(\gamma)$. In the non-sparse case, such differences are negligible and both $\hat{\epsilon}_n(\gamma)$ and $\hat{\epsilon}_n^*(\gamma)$ are consistent. The sparse case, especially when the proportion is algebraically small, is more subtle. In this case a more specific model is often needed. See Cai et al. [4].

We now compare our procedure with those in Meinshausen and Rice [17] and in Cai et al. [4]. We begin by introducing the aforementioned purity condition. If we model the p-values of the test statistics as samples from a mixing density, $(1-\epsilon)U(0,1) + \epsilon h$, where $U(0,1)$ and h are the marginal densities of the p-values for the null effects and non-null effects respectively. The purity condition is defines as $\text{essinf}_{\{0<p<1\}} h(p) = 0$. Meinshausen and Rice [17] propose a confidence lower bound for ϵ that is valid for all h. Despite this advantage, the lower bound is generally conservative and inconsistent. In fact, the purity condition is necessary for the lower bound to be consistent. Similar results can be found in Genovese and Wasserman [10]. Unfortunately, the purity condition generally does not hold in our settings.

Lemma 3.1 *Let the test statistics X_j be given as in (2.20). If the marginal distribution $F(\mu, \sigma)$ satisfies either $P_F\{\sigma > 1\} \neq 0$ or $P_F\{\sigma = 1\} = 1$, but $P_F\{\mu > 0\} \neq 0$ and $P_F\{\mu < 0\} \neq 0$, then the purity condition does not hold.*

Cai et al. [4] consider a very sparse setting for a two-point Gaussian mixture model where 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 where 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 for the simulation study is three-fold: to investigate how different choices of γ affect the estimation errors, to compare the performance of our approach with that in Efron [7], 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. Set $\sigma_0 = 1/\sqrt{2}$ and $\mu_0 = -1/2$ throughout this section. We take $n = 10000$, $\epsilon = 0.1$, and $a = 0.75$, 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 $U(a, a + 0.5)$, 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 non-null 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 $\gamma = 0.01, 0.02, \ldots, 0.5$.

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

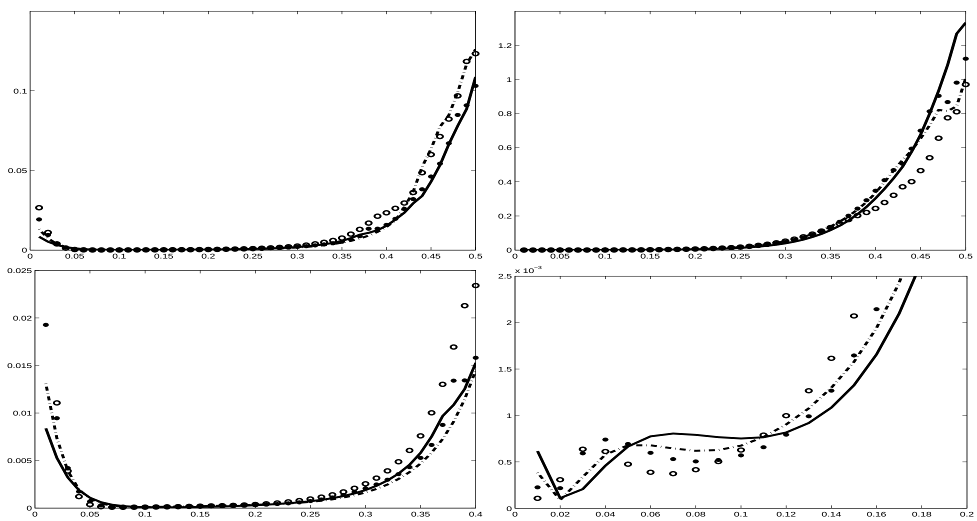

Figure 3: x-axis: γ. y-axis: mean squared error (MSE). Top row: MSE for $\hat{\sigma}_0(\gamma)$ (left) and $\hat{\mu}_0(\gamma)$ (right). The four different curves (solid, dashed, dot, and circle) correspond to $a = 0.75, 1.00, 1.25$, and 1.50. Bottom row: zoom in.

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 $(0.1, 0.15)$. 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 $\gamma = 0.1$, for example, the mean squared errors for $\hat{\sigma}_0(\gamma)$ and $\hat{\mu}_0(\gamma)$ are of magnitude 10^{-4} and 10^{-3}, respectively. These suggest the use of the following estimators for simplicity, where we take $\gamma = 0.1$:

$$\hat{\sigma}_0^* = \sigma_0(\varphi_n; \hat{t}_n(0.1)), \qquad \hat{\mu}_0^* = \mu_0(\varphi_n; \hat{t}_n(0.1)). \tag{4.1}$$

We now compare $(\hat{\sigma}_0^*, \hat{\mu}_0^*)$ with the estimators in Efron [7]. Recall that one major difference between the two approaches is that Efron's estimators are not consistent for the non-sparse case, while 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, $0.05, 0.10, 0.15$, and 0.20, 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 [7], and finally repeat the experiment for 100 independent cycles. The results are reported in Figures 4 - 5.

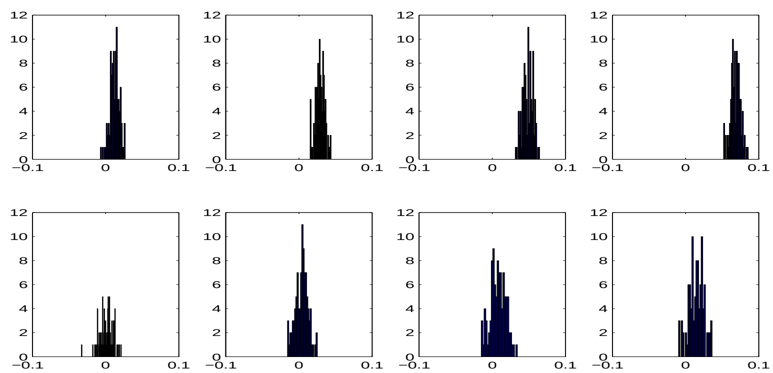

Figure 4: Histograms for the estimation errors of Efron's estimator for σ_0 (top row) and $\hat{\sigma}_0^*$ (bottom row). From left to right: $\epsilon = 0.05$, 0.10, 0.15, and 0.20.

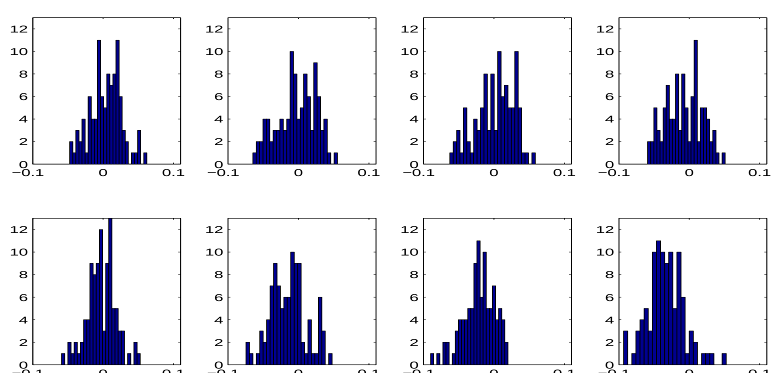

Figure 5: Histograms for the estimation errors of Efron's estimator for μ_0 (top row) and $\hat{\mu}_0^*$ (bottom row). From left to right: $\epsilon = 0.05$, 0.10, 0.15, and 0.20.

The results show that our estimator of σ_0^2 is more accurate than that of Efron [7], and the difference becomes more prominent as ϵ increases. In fact, when ϵ ranges between 0.05 and 0.2, the estimation errors of $\hat{\sigma}_0^*$ are of the order 10^{-2}, while those of Efron's estimator could get 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 0.02 across different choices of ϵ.

However, the above comparison is only for moderately large n. With a much larger n, the previous theory (Theorem 2.2) predicts that the estimation errors of $(\hat{\sigma}_0^*, \hat{\mu}_0^*)$ will become substantially smaller as $(\hat{\sigma}_0^*, \hat{\mu}_0^*)$ is consistent for (σ_0, μ_0). In comparison, the errors of Efron's estimators will not become substantially smaller as the estimators are not consistent. To illustrate this point, we carry out a small scale simulation experiment. We take $\epsilon = 0.1$ and $a = 1$ as before, while we 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 mean squared errors (MSE), and repeat the process for 30 independent cycles. The results are reported in Table 1, and they support the asymptotic analysis.

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	0.816	0.276	0.047	0.031
MSE for μ_0	Efron's approach	8.916	5.905	3.957	3.617
	Our approach	5.807	3.019	1.106	0.538

Table 1: Mean squared errors (MSE) for various values of n. The corresponding MSE equals the value in each cell times 10^{-4}.

Finally, we investigate the performance of the proposed procedures for dependent data. Fix $n = 10^4$, $\epsilon = 0.1$, and $a = 1$, and let L range from 0 to 250 with an increment of 5. For each L, 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 block-wise dependent (block size equal to $L+1$) and the marginal distribution of each z_j is $N(0,1)$. At the same time, 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 are reported in Figure 6, which suggests that the estimation errors increase as the range of dependency increases. However, 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.

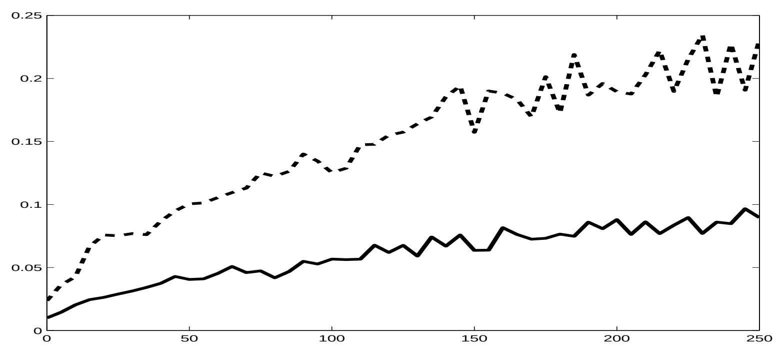

Figure 6: x-axis: L. y-axis: root mean squared error for $\hat{\mu}_0^*$ (dashed) and $\hat{\sigma}_0^*$ (solid).

5 Applications to microarray analysis

We now apply the proposed procedures to the analysis of the breast cancer and HIV microarray data sets that were analyzed in Efron [7]. The R code for our procedures is available on the web at *http://www.stat.purdue.edu/~jinj/Research/software*. The z-scores for both data sets 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 only comment briefly on the HIV data.

The breast cancer data was 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 reported on the same set of $N = 3226$ genes. For the j-th gene, the two-sample t-test comparing the seven BRCA1 responses with the eight BRCA2 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 [7], $X_j = \bar{\Phi}^{-1}(p_j) = \bar{\Phi}^{-1}(\bar{F}_{13}(y_j))$, where $\bar{\Phi}$ and $\bar{F}_{13}$ are the survival functions of $N(0,1)$ and t-distribution with 13 degrees of freedom, respectively.

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 j-th gene is not differentially expressed. Since 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 for example, [5, 7]. The normality assumption is also reasonable here, as the marginal density of non-null effects can generally be well approximated by Gaussian mixtures; see [7, Page 99]. Particularly, 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 includes three parts: estimating the null parameters (σ_0, μ_0), estimating the proportion of non-null effects, and implementing the local FDR approach proposed by Efron et al. [8].

The first part is estimating (σ_0, μ_0). We apply $(\hat{\sigma}_0^*, \hat{\mu}_0^*)$ (defined in (4.1)) as well as the estimators used by Efron [7] to the z-scores. For the breast cancer data, our procedure yields $(\hat{\sigma}_0^*, \hat{\mu}_0^*) = (1.5277, -0.0525)$, while Efron's estimators give $(\hat{\sigma}_0, \hat{\mu}_0) = (1.616, -0.082)$.

The second part of the analysis is estimating the proportion of non-null effects.

	Our estimator	Local FDR	MR	CJL
Our Estimated Null	0.0040	0.0128	0.0033	0
Efron's Estimated Null	0	0	0.0098	0

Table 2: Estimated proportion of non-null effects for the breast cancer data.

We implement our procedure as well as Meinshausen and Rice's [17] approach and the approach of Cai et al. [4] (which we denote by MR and CJL respectively for short), to the z-scores of the breast cancer data. The bounding function a_n^* for MR estimator is set as $1.25 \times \sqrt{2\log\log n}/\sqrt{n}$, and the a_n for CJL estimator is set as $\sqrt{2\log\log n}/\sqrt{n}$; see [4] for details. Using the estimated null parameters either obtained by Efron's approach or obtained by our approach, we apply each of these procedures to the z-scores. In addition, the local FDR approach also provides an estimate for the proportion automatically. The results are reported in Table 2.

In the last part of the analysis we implement the local FDR thresholding procedure proposed in [8] with the z-scores of the breast cancer data. For any given FDR-control parameter $q \in (0,1)$, the procedure calculates a score for each data point and determines a threshold t_q at the same time. 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 false discovery rate at level q, $E[\frac{\#\text{False Discoveries}}{\#\text{Total Discoveries}}] \leq q$. See [8] 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 data set. Also, three different estimators for the proportion report 0. These suggest that either the proportion of signals (differentially expressed genes) is small and/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

$q \geq 0.91$. For example, the number of total discoveries equal to 167 when $q = 0.92$, and equal to 496 when $q = 0.94$. Take $q = 0.94$, for example, since for any $q \in (0, 1)$, the number of true discoveries approximately equal to $(1-q)$ times the number of total discoveries [7], this suggests a total of 30 true discoveries. The result is consistent with biological discoveries. Among the 496 genes which are identified to be differentially expressed by the local FDR procedures, 17 of them have been discovered in the study by Hedenfalk et al. [11]. 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, Hs.73798. We also identified several genes whose functions are associated with the cell cycle, including PCNA, CCNA2, and CKS2. These genes are found to be significant by Storey et al. [19]. The results indicate that our estimated null parameters lead to reliable identification of differentially expressed genes.

Similarly, for the HIV data, our estimators give $(\hat{\sigma}_0^*, \hat{\mu}_0^*) = (0.7709, -0.0806)$, while Efron's method gives $(\hat{\sigma}_0, \hat{\mu}_0) = (0.738, -0.082)$. With $q = 0.05$, 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.

6 Proofs of the main results

We now prove Theorems 2.1, 2.2, and 3.1. The proof of Theorem 3.2 is similar to those of Theorems 2.2 and 3.1 and so is omitted. As 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 are given in [14].

Lemma 6.1 *Let $\sigma_0^2(\cdot;\cdot)$ and $\mu_0(\cdot;\cdot)$ be defined as in (2.7). Fix $t > 0$. For any* *differentiable complex-valued functions* f *and* g *satisfying* $|f(t)| \neq 0$ *and* $|g(t)| \neq 0$*,*

$$|\sigma_0^2(f,t) - \sigma_0^2(g,t)| \leq \frac{|g(t)|}{t|f(t)|^2}\big[\big(2t\cdot|\sigma_0^2(g,t)| + |\frac{g'(t)}{g(t)}|\big)|f(t)-g(t)| + |f'(t)-g'(t)| + r_n^{(1)}(t)\big],$$

$$|\mu_0(f,t) - \mu_0(g,t)| \leq \frac{|g(t)|}{|f(t)|^2}\cdot\big[\big(2|\mu_0(g,t)| + |\frac{g'(t)}{g(t)}|\big)\cdot|f(t)-g(t)| + |f'(t)-g'(t)| + r_n^{(2)}(t)\big],$$

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

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

$$|\varphi_n(\hat{t}_n) - \varphi(\hat{t}_n)| \leq O_p(\sqrt{\log n}/\sqrt{n}), \qquad |\varphi_n'(\hat{t}_n) - \varphi'(\hat{t}_n)| \leq O_p(\sqrt{\log n}/\sqrt{n}). \quad (6.1)$$

Applying Lemma 6.1 with $f = \varphi_n$, $g = \varphi$, and $t = \hat{t}_n(\gamma)$, we have

$$|\sigma_0^2(\varphi_n, \hat{t}_n(\gamma)) - \sigma_0^2(\varphi, \hat{t}_n(\gamma))|$$
$$\sim n^{\gamma}\big(3\sigma_0^2|\varphi_n(\hat{t}_n(\gamma)) - \varphi(\hat{t}_n(\gamma))| + \frac{1}{\hat{t}_n(\gamma)}|\varphi_n'(\hat{t}_n(\gamma)) - \varphi'(\hat{t}_n(\gamma))|\big) \sim O(n^{\gamma-\frac{1}{2}}\sqrt{\log n}),$$

and Theorem 2.1 follows. We now study (6.1) in detail.

Lemma 6.2 *Set* $W_0(\varphi_n;n) = W_0(\varphi_n;n,X_1,\ldots,X_n) = \sup_{0\leq t\leq \log n}|\varphi_n(t) - \varphi(t)|$*. Fix* $q_1 > 3$*. Let* $\Lambda_n(q,A;\mu_0,\sigma_0,\epsilon_0)$ *be given as in Theorem 2.1. When* $n \to \infty$*,*

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

Lemma 6.2 implies that except for an event with algebraically small probability, $|\varphi(\hat{t}_n) - \varphi(t_n)| \leq W_0(\varphi_n;n) \leq \sqrt{2q_1\log n}/\sqrt{n}$. 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 6.3 *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 2.1. When* $n \to \infty$*,*

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

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

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

By Lemma 6.3, except for an event with algebraically small probability, $|\hat{t}_n(\gamma) - t_n(\gamma)| \leq \pi_0 \cdot n^{\gamma-1/2}$, and consequently $|\varphi_n'(\hat{t}_n(\gamma)) - \varphi'(\hat{t}_n(\gamma))| \leq W_1(\varphi_n, \gamma, \pi_0; n)$. The following lemma describes the tail behavior of W_1.

Lemma 6.4 *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 $(\mu, \sigma) \in \Lambda_n(q, A; \mu_0, \sigma_0, \epsilon_0)$, $\bar{s}_n \leq C_1$,*

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

where $c_1(q, \gamma)$ is as in Theorem 2.1. As a result, except for an event with algebraically small probability, $|\varphi_n'(\hat{t}_n(\gamma)) - \varphi'(\hat{t}_n(\gamma))| \leq W_1(\varphi_n, \gamma, \pi; n) \leq O(\sqrt{\log n}/\sqrt{n})$.

We have now elaborated the inequalities in (6.1). 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 6.5 *Fix $q \geq 3$ and $A > 0$, with $\psi(t)$ and τ_n as defined in (2.18) and (2.19) respectively, 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*

$$|\sigma_0^2(\varphi, t) - \sigma_0^2| \leq \frac{|r'(t)|}{t} \cdot \frac{1 + |r(t)|}{|1 + r(t)|^2} \leq C|\psi'(t)|/t, \tag{6.2}$$

$$|\mu_0(\varphi, t) - \mu_0| \leq |r'(t)| \cdot \frac{1 + |r(t)|}{|1 + r(t)|^2} \leq C|\psi'(t)|. \tag{6.3}$$

Additionally, uniformly for all $(\mu, \sigma) \in \Lambda_n(q, A; \mu_0, \sigma_0, \epsilon_0)$ and all $t > 0$,

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

(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 Theorem 2.1, 2.2, and 3.1.

Proof of Theorem 2.1: Since 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

$$B_0 = \{W_0(\varphi_n; n) \leq \sqrt{2q_1 \log n}\}, \quad B_1 = \{W_1(\varphi_n; n) \leq \frac{s_n\sqrt{(q-2)\log n} + 2s_n^2}{\sqrt{n}}\}.$$

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 defined as in Theorem 2.1. Use Lemma 6.2 and Lemma 6.4, $P\{B_0^c\} \leq \bar{o}(n^{-q_1/3})$ and $P\{B_1^c\} \leq \bar{o}(n^{-c_1(q,\gamma)})$; since $c_1(q, \gamma) < q_1/3$, $P\{B_0^c \cup B_1^c\} \leq \bar{o}(n^{-c_1(q,\gamma)})$. We now focus on $B_0 \cap B_1$. By triangle inequality, $|\sigma_0^2(\varphi_n; \hat{t}_n) - \sigma_0^2(\varphi; t_n)| \leq |\sigma_0^2(\varphi_n; \hat{t}_n) - \sigma_0^2(\varphi; \hat{t}_n)| + |\sigma_0^2(\varphi; \hat{t}_n) - \sigma_0^2(\varphi; t_n)|$. Note that by the choice of π_0 and Lemma 6.3, $|\hat{t}_n - t_n| \leq \pi_0 \cdot n^{\gamma - 1/2}$ for sufficiently large n, it thus follows from Lemma 6.5 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 $c_2(\sigma_0, q, \gamma) > \sigma_0^2\sqrt{2q_1}$, so to show the claim, it suffices to show that as $n \to \infty$,

$$|\sigma_0^2(\varphi_n; \hat{t}_n) - \sigma_0^2(\varphi; \hat{t}_n)| \leq 3\sigma_0^2 \cdot \sqrt{2q_1 \log n} \cdot n^{\gamma-1/2} \cdot (1 + o(1)), \qquad \text{over } B_0 \cap B_1. \quad (6.4)$$

We now show (6.4). Over the event $B_0 \cap B_1$, recall $|\hat{t}_n - t_n| \leq \pi_0 n^{\gamma-1/2}$, so by (2.13), $\hat{t}_n \sim t_n \sim \sqrt{2\gamma \log n}/\sigma_0$; by Lemma 6.5, this implies $\sigma_0^2(\varphi, \hat{t}_n) \sim \sigma_0^2$ and $|\varphi'(\hat{t}_n)|/|\varphi(\hat{t}_n)| \sim \sigma_0^2 \hat{t}_n \sim \sigma_0^2 t_n$. Moreover, since $|\varphi_n(\hat{t}_n) - \varphi(\hat{t}_n)| \leq \sqrt{2q_1 \log n}/\sqrt{n}$, it follows that $|\varphi(\hat{t}_n)|/(\hat{t}_n|\varphi_n(\hat{t}_n)|^2) \sim (1/t_n)n^{\gamma}$. Lastly, by Lemma 6.4, $|\varphi_n'(\hat{t}_n) - \varphi'(\hat{t}_n)| \leq O(\sqrt{\log n}/\sqrt{n})$. Combining these, (6.4) follows directly by applying Lemma 6.1 with $f = \varphi_n$, $g = \varphi$, and $t = \hat{t}_n$. □

Proof of Theorem 2.2: Note that, by triangle inequality, $|\sigma_0^2(\varphi_n; \hat{t}_n(\gamma)) - \sigma_0^2| \leq |\sigma_0^2(\varphi_n; \hat{t}_n(\gamma) - \sigma_0^2(\varphi; t_n(\gamma))| + |\sigma_0^2(\varphi; t_n(\gamma)) - \sigma_0^2|$. Theorem 2.2 now follows directly from Theorem 2.1 and Lemma 6.5. $\square$

Proof of Theorem 3.1: 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 $\Omega(t) = E[\Omega_n(t)]$, $\Psi_n^*(t) = \sup_{\{0 \leq s \leq t\}}\{1 - \Omega_n(s)\}$, and $\Psi^*(t) = \sup_{\{0 \leq s \leq t\}}\{1 - \Omega(s)\}$. Note that it is sufficient to show that when $n \to \infty$, (a) except for an event with algebraically small probability, $\sup_{\{(\mu,\sigma)\in\Theta_n\}} |\Psi_n^*(t_n) - \Psi^*(t_n)| \leq O(\log^{-3/2}(n) \cdot n^{\gamma-1/2})$, 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

$$\sup_{0 \leq t \leq t_n} |\Omega_n(t) - \Omega(t)| \leq 2\int_0^1 (1-\xi)e^{\frac{t_n^2\xi^2}{2}} \sup_{0 \leq t \leq t_n} |\mathrm{Re}(\varphi_n(t)) - \mathrm{Re}(\varphi(t))| d\xi. \tag{6.5}$$

Moreover, similar to the proof of Lemma 7.2 in [13], we have that for fixed $q > 3/2$, $\sup_{\{(\mu,\sigma)\in\Theta_n\}} \sup_{\{0 \leq t \leq t_n\}} |\mathrm{Re}(\varphi_n(t)) - \mathrm{Re}(\varphi(t))| \leq O(\sqrt{\log n}/\sqrt{n})$ except for an event with probability $\sim 2\log^2(n) \cdot n^{-2q/3}$. Elementary calculus yields $|\Psi_n^*(t_n) - \Psi^*(t_n)| \leq O(\sqrt{\log n}/\sqrt{n}) \cdot \int_0^1 (1-\xi)e^{(\gamma \log n)\cdot\xi^2} d\xi = O(\log^{-3/2}(n) \cdot n^{\gamma-1/2})$, 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 $\hat{\phi}_{\delta_j(t)}(\xi) = \exp(\frac{(1-\sigma_j^2)t^2\xi^2}{2})$ and $\hat{\rho}(\xi) = \max\{1 - |\xi|, 0\}$. So by the Fourier Inversion Theorem [16, Page 22],

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

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

$$1 - \Omega(t) = \epsilon_n \cdot \mathrm{Ave}_{\{j:\,(\mu_j,\sigma_j)\neq(0,1)\}}\{1 - \phi_{\delta_j(t)} * \rho(t\mu_j)\}. \tag{6.6}$$

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 $\max\{a_n, b_n\} \to \infty$, so by (6.6) and the definition of Θ_n, $\sup_{\{(\mu,\sigma)\in\Theta_n\}} \left|\frac{1-\Omega(t_n)}{\epsilon_n} - 1\right| = o(1)$. Note that $0 \leq \phi_{\delta_j(t)} * \psi(t) \leq 1$ for all t, so by (6.6) and the definition of Ψ^*, $\Omega(t_n) \leq \Psi^*(t_n) \leq \epsilon_n$; as a result, $\left|\frac{1-\Psi^*(t_n)}{\epsilon_n} - 1\right| \leq \left|\frac{1-\Omega(t_n)}{\epsilon_n} - 1\right|$, and (b) follows directly. □

Acknowledgments

We thank Bradley Efron for references and kindly sharing the data sets. We thank Paul Shaman for a careful reading of our manuscript and for suggestions which lead to significant improvement of the presentation of the paper. We also thank Herman Rubin, an Associate editor, and referees for helpful comments and references.

References

[1] Abramovich, F. and Benjamini, Y. and Donoho, D. and Johnstone, I. (2006). Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.* **34**, 584-653.

[2] 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.

[3] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes*, 2nd edition, Springer Verlag, New York.

[4] Cai, T. and Jin, J. and Low, M. (2005). Estimation and confidence sets for sparse normal mixtures, *manuscript*.

[5] Cui, X. and Churchill, G. (2003). Statistical tests for different expression in cDNA microarray experiments. *Genome Biology* **4**:210.

[6] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32**, 962-994.

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

[8] EFRON, B., TIBSHIRANI, R., STOREY, J., and TUSHER, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* **96**, 1151-1160.

[9] FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257-1272.

[10] GENOVESE, C. and WASSERMAN, L. (2004). A stochastic process approach to false discovery control, *Ann. Statist.*, **32**, 1035-1061.

[11] HEDENFALK, I. and DUGGEN, D. and CHEN, Y. et al. (2001). Gene expression profiles in hereditary breast cancer, *New Engl. Jour. Medicine* **344**, 539-548.

[12] HOEFFDING, W. (1963). Probability inequality for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13-30.

[13] JIN, J. (2006). Proportion of nonzero normal means: universal oracle equivalences and uniformly consistent estimations, *manuscript*.

[14] JIN, J. and CAI, T. (2006). Estimating the null and the proportion of non-null effects in large-scale multiple comparisons, *technical report*.

[15] LÖNNSTEDT, I. and SPEED, T. (2002). Replicated microarray data. *Statistica Sinica* **12**, 31-46.

[16] MALLAT, S. (1998). *A Wavelet Tour of Signal Processing*, Academic Press.

[17] MEINSHAUSEN, M. and RICE, J. (2006). Estimating the proportion of false null hypotheses among a large number of independent tested hypotheses. *Ann. Statist.* **34**, 373-393.

[18] RUBIN, H. and SETHURAMAN, J. (1965). Probabilities of moderate deviations. *Sankhyā* **27**, 325-346.

[19] STOREY, J. D. and DAI, J. Y. and LEEK, J. T. (2005). The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments. *Working Paper, University of Washington.*

[20] SWANEPOEL, J. W. H. (1999). The limiting behavior of a modified maximal symmetric $2s$-spacing with applications. *Ann. Statist.* **27**, 24-35.

[21] VAN'T WOUT, A. et al. (2003). Cellular gene expression upon human immunodeficiency virus type 1 infection of $CD4^+$-T-Cell lines. *J. Virology* **77**, 1392-1402.

[22] ZHANG, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18**, 806-831.

7 Appendix

7.1 Proof of Theorem 2.3

For short, write $\hat{t}_n = \hat{t}_n(\gamma)$ and $t_n = t_n(\gamma)$. The following two lemmas are proved in Section 7.1.1 and Section 7.1.2 respectively.

Lemma 7.1 *With $\alpha(\cdot)$ and $\tilde{\Lambda}_n(a, B, q, A)$ as in Theorem 2.3. Fix $r \in (1.5, d-(2d+2.5)\gamma)$. As $n \to \infty$, uniformly for all $(\mu, \sigma) \in \tilde{\Lambda}_n(a, B, q, A)$, except for an event with a probability of $\bar{o}(n^{-2r/3})$, $\sup_{\{0 \leq t \leq \log n\}} |\varphi_n(t) - \varphi(t)| = \bar{o}(n^{-(d-r)/(2d+2.5)})$.*

Lemma 7.2 *With $\alpha(\cdot)$ and $\tilde{\Lambda}_n(a, B, q, A)$ as in Theorem 2.3. Fix $\gamma \in (0, \frac{d-1.5}{2d+2.5})$ and an integer $k \geq 0$. As $n \to \infty$, for all $(\mu, \sigma) \in \tilde{\Lambda}_n(a, B, q, A)$, $\sup_{\{0 \leq t \leq \log n\}}\{|\varphi_n^{(k)}(t) - \varphi^{(k)}(t)|\} \leq O_p(\log^{ak}(n))$, and $[\varphi_n^{(k)}(t_n) - \varphi^{(k)}(t_n)] = O_p(\log^{(a+1/2)k}(n)/\sqrt{n})$.*

To show the theorem, it is sufficient to show that

$$|\varphi_n(\hat{t}_n) - \varphi(\hat{t}_n)| = O_p(1/\sqrt{n}), \qquad |\varphi_n'(\hat{t}_n) - \varphi'(\hat{t}_n)| = O_p(\log^{a+1/2}(n)/\sqrt{n}). \quad (7.1)$$

In fact, by triangle inequality,

$$|\sigma_0^2(\varphi_n; \hat{t}_n) - \sigma_0^2(\varphi; t_n)| \leq |\sigma_0^2(\varphi_n; \hat{t}_n) - \sigma_0^2(\varphi; \hat{t}_n)| + |\sigma_0^2(\varphi; \hat{t}_n) - \sigma_0^2(\varphi; t_n)|. \quad (7.2)$$

Once (7.1) is proved, by similar arguments as in the proof of Theorem 6.3,

$$|\hat{t}_n - t_n| = O_p(n^{\gamma - 1/2}), \quad (7.3)$$

it thus follows from Lemma 6.5 that

$$|\sigma_0^2(\varphi; \hat{t}_n) - \sigma_0^2(\varphi; t_n)| = o_p(|\hat{t}_n - t_n|) = o_p(n^{\gamma - 1/2}). \quad (7.4)$$

At the same time, by (7.3) and Lemma 6.5, except for an event with asymptotically vanishing probability, $|\varphi_n(\hat{t}_n)|/|\varphi_n(\hat{t}_n)|^2 \sim n^{\gamma}$, $\sigma_0^2(\varphi; \hat{t}_n) \sim \sigma_0^2$, and $|\varphi'(\hat{t}_n)|/|\varphi(\hat{t}_n)| \sim \sigma_0^2 \hat{t}_n$; applying Lemma 6.1 with $f = \varphi_n$, $g = \varphi$, and $t = \hat{t}_n$, it follows that

$$|\sigma_0^2(\varphi_n; \hat{t}_n) - \sigma_0^2(\varphi; \hat{t}_n)| = O_p(n^{\gamma - 1/2}). \quad (7.5)$$

The theorem follows directly by inserting (7.4) and (7.5) into (7.2).

We now show (7.1). Since the proofs are similar, we only show the first equality. Applying Lemma 7.1 with $r = (1.5 + d - (2d + 2.5)\gamma)/2$, it follows that there is an event A_n such that $P\{A_n^c\}$ is algebraically small and

$$\sup_{\{0 \leq t \leq \log n\}} |\varphi_n(t) - \varphi(t)| \leq \bar{o}(n^{-\frac{1}{2} \cdot (\gamma + \frac{d-1.5}{2d+2.5})}), \qquad \text{over } A_n. \quad (7.6)$$

By similar arguments as in the proof of Lemma 6.3, it follows that

$$|\hat{t}_n(\gamma) - t_n(\gamma)| \leq \bar{o}(n^{\frac{1}{2}\cdot(\gamma - \frac{d-1.5}{2d+2.5})}), \qquad \text{over } A_n, \tag{7.7}$$

notice the exponent is negative. Now, let ℓ be the smallest integer satisfying $(\ell + 1) \cdot |\gamma - \frac{d-1.5}{2d+2.5}| > 1$. By Taylor expansion, for some ξ falling between $\hat{t}_n$ and t_n,

$$\varphi_n(\hat{t}_n) - \varphi(t_n) = \sum_{k=0}^{\ell} \frac{\varphi_n^{(k)}(t_n) - \varphi^{(k)}(t_n)}{k!}(\hat{t}_n - t_n)^k + \frac{\varphi_n^{(\ell+1)}(\xi) - \varphi^{(\ell+1)}(\xi)}{(\ell+1)!}(\hat{t}_n - t_n)^{\ell+1}.$$

Notice that by the choice of ℓ and (7.7), $(\hat{t}_n - t_n)^{\ell+1} = \bar{o}(1/\sqrt{n})$ over A_n, the claim follows directly from Lemma 7.2. □

7.1.1 Proof of Lemma 7.1

Applying [3, Theorem 1.3] with $b = 2$, $q = n^{(d-r)/(d+1.25)}$, and $\epsilon = \sqrt{32r\log n}/\sqrt{q}$ gives $P\{|\mathrm{Re}(\varphi_n(t) - \varphi(t))| \geq \epsilon\} \leq \bar{o}(n^{-r})$ and $P\{|\mathrm{Im}(\varphi_n(t) - \varphi(t))| \geq \epsilon\} \leq \bar{o}(n^{-r})$, it thus follows

$$P\{|\varphi_n(t) - \varphi(t)| \geq \sqrt{2}\epsilon\} \leq \bar{o}(n^{-r}). \tag{7.8}$$

The remaining part of the proof is similar to that of Lemma 6.2 so we keep it brief. Fix $\delta \in (1/2, \infty)$, with the same grid and similar arguments as in Lemma 6.2, it follows that

$$P\{\sup_{\{0 \leq t \leq \log n\}} |\varphi_n(t_k) - \varphi(t)| \geq (\sqrt{2}\epsilon + \frac{1}{\sqrt{n}})\} \leq I + II, \tag{7.9}$$

where $I = P\{\sup_{\{1 \leq k \leq n^\delta \log n\}} |\varphi_n(t_k) - \varphi(t_k)| \geq \sqrt{2}\epsilon\}$ and $II \leq P\{n^{-\delta}\sup_t\{|\varphi_n'(t) - \varphi'(t)|\} \geq \frac{1}{\sqrt{n}}\}$. The key for the proof is to show that

$$\mathrm{Var}(\frac{1}{n}\sum_{j=1}^{n}|X_j|) \leq C\log^{2a}(n)/n. \tag{7.10}$$

In fact, once (7.10) is proved, then on one hand, by (7.8), $I \leq n^\delta \log(n) \cdot \bar{o}(n^{-r}) = \bar{o}(n^{-(r-\delta)})$. On the other hand, by similar arguments as in the proof of Lemma 6.2,

$$II \leq P\{\frac{1}{n}\sum_{j=1}^{n}(|X_j| - E|X_j|) \geq n^{\delta-1/2} - s_n\} \lesssim \frac{1}{n^{2\delta-1}} \cdot \mathrm{Var}(\frac{1}{n}\sum_{j=1}^{n}|X_j|) = \bar{o}(n^{-2\delta}),$$

where $s_n \equiv \frac{1}{n}\sum_{j=1}^n E|X_j| \leq C\log^a(n)$ as $\max_j\{|\mu_j| + |\sigma_j|\} \leq \log^a(n)$. The claims follows by taking $\delta = r/3$.

We now show (7.10). Applying [3, Corollary 1.1] with $p = 1.5$, $q = r = 6$,

$$\mathrm{Var}(\frac{1}{n}\sum_{j=1}^n |X_j|) = \frac{1}{n^2}\sum_{j,k}\mathrm{Cov}(|X_j|, |X_k|) \leq \frac{C}{n^2}\sum_{j,k}\alpha^{2/3}(|j-k|)\|X_j\|_6\|X_k\|_6.$$

By $\max_j\{|\mu_j| + |\sigma_j|\} \leq \log^a(n)$, $\|X_j\|_6 \leq C\log^a(n)$ for all $1 \leq j \leq n$; since $\alpha(k) \leq Bk^{-d}$ with $d > 1.5$, (7.10) follows by observing $\sum_{j,k}\alpha^{2/3}(|j-k|) \leq Cn\sum_{k=1}^\infty \alpha^{2/3}(k) \leq Cn\sum_{k=1}^\infty k^{-2d/3} = O(n)$. □

7.1.2 Proof of Lemma 7.2

Consider the first claim. By direct calculations,

$$|\varphi_n^{(k)}(t) - \varphi^{(k)}(t)| = |\frac{1}{n}\sum_{j=1}^n (iX_j)^k e^{itX_j} - E[\frac{1}{n}\sum_{j=1}^n (iX_j)^k e^{itX_j}]| \leq \frac{1}{n}\sum_{j=1}^n [|X_j|^k + E|X_j|^k],$$

where the right hand side does not depend on t. Since $\max_{\{j\}}\{|\mu_j| + |\sigma_j|\} \leq B\log^a(n)$, the claim follows directly from $E|X_j|^k \leq C\cdot(|\mu_j|^k + |\sigma_j|^k) \leq C\cdot\log^{ak}(n)$, $\forall\ 1 \leq j \leq n$, where $C = C(k)$ is a generic constant.

Consider the second claim. Introduce an event $D_n = \{\max_j\{|X_j|\} \leq 3B\log^{a+1/2}(n)\}$. By $\max_{\{j\}}\{|\mu_j| + |\sigma_j|\} \leq B\log^a(n)$ and direct calculations,

$$P\{D_n^c\} \leq \sum_j P\{|X_j| \geq 3B\log^{a+1/2}(n)\} \leq 2n\bar{\Phi}(3\sqrt{\log n} - 1) = \bar{o}(n^{-1}), \quad (7.11)$$

where $\bar{\Phi}$ is the survival function of $N(0,1)$. To show the claim, it suffices to show

$$E[(\varphi_n^{(k)}(t_n) - \varphi^{(k)}(t_n))\cdot 1_{\{D_n\}}]^2 = O(\log^{(2a+1)k}(n)/n). \quad (7.12)$$

Now, first, observe that $|x|^k\exp(-\frac{(x-\mu_j)^2}{2\sigma_j^2}) = o(1)$, where $o(1) \to 0$ as $n \to \infty$, uniformly for all $|x| \geq 3B\log^{a+1/2}(n)$ and (μ_j, σ_j) satisfying $|\mu_j| + |\sigma_j| \leq B\log^a(n)$; combining this with (7.11) gives $|E(\varphi_n(t_n)\cdot 1_{\{D_n^c\}})| \leq \frac{1}{n}\sum_{j=1}^n E(|X_j|^k\cdot 1_{\{D_n^c\}}) = \bar{o}(n^{-1})$. Notice that $E\varphi_n^{(k)}(t_n) = \varphi^{(k)}(t_n)$, we thus have

$$E[(\varphi_n^{(k)}(t_n) - \varphi^{(k)}(t_n))\cdot 1_{\{D_n\}}] = -E[\varphi_n^{(k)}(t_n)\cdot 1_{\{D_n^c\}}] = \bar{o}(1/n). \quad (7.13)$$

Second, as $\max_{\{j\}}\{|X_j|\} \leq 3B\log^{a+1/2}(n)$ over D_n, by Billingsley's inequality [3, Page 22],

$$\begin{aligned}\mathrm{Var}(\varphi_n^{(k)}(t_n)\cdot 1_{\{D_n\}}) &= \frac{1}{n^2}\sum_{j_1,j_2}\mathrm{Cov}((iX_{j_1})^k\cdot e^{itX_{j_1}}\cdot 1_{\{D_n\}},(iX_{j_2})^k\cdot e^{itX_{j_2}}\cdot 1_{\{D_n\}})\\ &\leq \frac{C}{n^2}B^2\log^{(2a+1)k}(n)\sum_{j_1,j_2}\alpha(|j_1-j_2|)\leq O(\log^{(2a+1)k}(n)/n).\end{aligned}$$

Combining this with (7.13) gives (7.12). ☐

7.2 Proof of Lemma 6.1

For short, we drop t from the functions whenever there is no confusion. For the first claim, by direct calculations, we have:

$$\sigma_0^2(g,t)-\sigma_0^2(f,t)=\frac{\frac{d}{dt}|f|}{t|f|}-\frac{\frac{d}{dt}|g|}{t|g|}=I+II+III,$$

where $I=(1-\frac{|g|^2}{|f|^2})\cdot\sigma_0^2(g,t)$, $II=\frac{1}{t|f|^2}\cdot[\mathrm{Re}(g')\cdot\mathrm{Re}(f-g)+\mathrm{Im}(g')\cdot\mathrm{Im}(f-g)+\mathrm{Re}(g)\cdot\mathrm{Re}((f-g)')+\mathrm{Im}(g)\cdot\mathrm{Im}((f-g)')]$, and $III=\frac{1}{t|f|^2}\cdot[\mathrm{Re}(f-g)\cdot\mathrm{Re}((f-g)')+\mathrm{Im}(f-g)\cdot\mathrm{Im}((f-g)')]$. Now, firstly, using triangle inequality,

$$|I|\leq\frac{|\sigma_0^2(g,t)|}{|f|^2}\cdot\big||f|^2-|g|^2\big|\leq\frac{|\sigma_0^2(g,t)|}{|f|^2}\big(2|g|\cdot|f-g|+|f-g|^2\big);$$

secondly, using Cauchy-Schwartz inequality, $|\mathrm{Re}(z)\mathrm{Re}(w)+\mathrm{Im}(z)\mathrm{Im}(w)|\leq|z|\cdot|w|$ for any complex numbers z and w, so it follows that

$$|II|\leq\frac{1}{t|f|^2}\cdot[|g'|\cdot|f-g|+|g|\cdot|(f-g)'|],\qquad |III|\leq\frac{1}{t|f|^2}\cdot|f-g|\cdot|(f-g)'|;$$

combining these gives

$$|\sigma_0^2(g,t)-\sigma_0^2(f,t)|\leq\frac{1}{t|f|^2}\big[(2t\cdot|\sigma_0^2(g,t)|\cdot|g|+|g'|)|f-g|+|g|\cdot|(f-g)'|+\tilde{r}_n^{(1)}\big],$$

where $\tilde{r}_n^{(1)}=t\cdot|\sigma_0^2(g,t)|\cdot|f-g|^2+|f-g|\cdot|(f-g)'|$, and the claim follows directly.

For the second claim, by direct calculations:

$$\mu_0(g,t) - \mu_0(f,t) = \frac{\mathrm{Re}(f')\mathrm{Im}(f) - \mathrm{Re}(f)\mathrm{Im}(f')}{|f|^2} - \frac{\mathrm{Re}(g')\mathrm{Im}(g) - \mathrm{Re}(g)\mathrm{Im}(g')}{|g|^2}$$
$$= I + II + III,$$

where $I = (1 - \frac{|g|^2}{|f|^2}) \cdot \mu_0(g,t)$, $II = \frac{1}{|f|^2} \cdot [(\mathrm{Re}(g') \cdot \mathrm{Im}(f-g) - \mathrm{Im}(g') \cdot \mathrm{Re}(f-g)) + (\mathrm{Im}(g) \cdot \mathrm{Re}((f-g)') - \mathrm{Re}(g) \cdot \mathrm{Im}((f-g)'))]$, and $III = \frac{1}{|f|^2}[\mathrm{Re}((f-g)') \cdot \mathrm{Im}(f-g) - \mathrm{Re}(f-g) \cdot \mathrm{Im}((f-g)')]$. As in the first part,

$$|I| \leq \frac{|\mu_0(g,t)|}{|f|^2}[2|g| \cdot |f-g| + |f-g|^2],$$

$$|II| \leq \frac{1}{|f|^2} \cdot [|g'| \cdot |f-g| + |g| \cdot |(f-g)'|], \qquad |III| \leq \frac{1}{|f|^2} \cdot |(f-g)'| \cdot |f-g|;$$

combining these gives

$$|\mu_0(g,t) - \mu_0(f,t)| \leq \frac{1}{|f|^2} \cdot \big[(2|\mu_0(g,t)| \cdot |g| + |g'|) \cdot |f-g| + |g| \cdot |(f-g)'| + \tilde{r}_n^{(2)}\big],$$

where $\tilde{r}_n^{(2)} = |\mu_0(g,t)| \cdot |f-g|^2 + |f-g| \cdot |(f-g)'|$, and the claim follows. □

7.3 Proof of Lemma 6.2

Lay out a grid $t_k = k/n^\delta$, for $k = 1, \ldots, n^\delta \log n$ and $\delta \in (1/2, q_1/2)$. For any $0 \leq t \leq \log n$, pick the closest grid point t_k, so that $|t_k - t| \leq n^{-\delta}$ and

$$|\varphi_n(t) - \varphi(t)| \leq |\varphi_n(t_k) - \varphi(t_k)| + |(\varphi_n(t) - \varphi(t)) - (\varphi_n(t_k) - \varphi(t_k))|,$$

where the second term is $\leq n^{-\delta} \cdot \sup_t |\varphi_n'(t) - \varphi'(t)|$. Write:

$$\frac{\sqrt{2q_1 \log n}}{\sqrt{n}} = \lambda_1(q_1, n) + \lambda_2(q_1, n),$$

where $\lambda_1(q_1,n) = \frac{\sqrt{2q_1 \log n} - 2\log\log n/\sqrt{2q_1 \log n}}{\sqrt{n}}$ and $\lambda_2(q_1,n) = \frac{2\log\log n/\sqrt{2q_1\log n}}{\sqrt{n}}$. It thus follows that

$$P\{\sup_{0 \leq t \leq \log n} |\varphi_n(t) - \varphi(t)| \geq \frac{\sqrt{2q_1 \log n}}{\sqrt{n}}\} \leq I + II, \tag{7.14}$$

where $I = P\{\sup_{1\le k\le n^\delta \log n} |\varphi_n(t_k)-\varphi(t_k)| \ge \lambda_1(q_1,n)\}$, and $II = P\{n^{-\delta}\cdot \sup_t |\varphi_n'(t)-\varphi'(t)| \ge \lambda_2(q_1,n)\}$.

For I, a direct generalization of Hoeffding's inequality [12] to complex-valued random variables gives:

$$I \le (n^\delta \log n) 4 e^{-\frac{1}{4} n \lambda_1^2(q_1,n)} = 4n^\delta \log n \cdot e^{-\frac{q_1 \log n}{2} + \log\log n (1 - \frac{\log\log n}{2q_1 \log n})} \tag{7.15}$$

$$\lesssim (4n^\delta \log n)(n^{-q_1/2}\log n) = 4n^{\delta - q_1/2}\log^2 n. \tag{7.16}$$

For II, direct calculations show that $\sup_t |\varphi_n'(t) - \varphi'(t)| \le \frac{1}{n}\cdot\sum_{j=1}^n (|X_j| + E|X_j|)$. Denote $s_n = \frac{1}{n}\sum_{j=1}^n E|X_j|$ for short, it follows from Chebyshev's inequality that:

$$II \le P\{\frac{1}{n}\sum_{j=1}^n (|X_j| + E|X_j|) \ge n^\delta \cdot \lambda_2(q,n)\} \tag{7.17}$$

$$= P\{\frac{1}{n}\sum_{j=1}^n (|X_j| - E|X_j|) \ge n^\delta \cdot \lambda_2(q,n) - 2s_n\} = O(n^{-2\delta}\frac{\log^2(\log(n))}{\log(n)}), \tag{7.18}$$

where we have used the fact that s_n is uniformly bounded from above by a constant $C(q, A, \mu_0, \sigma_0) < \infty$. Inserting (7.15) - (7.18) to (7.14) and taking $\delta = q_1/6$ give:

$$P\{\sup_{0\le t\le \log n} |\varphi_n(t) - \varphi(t)| \ge \frac{\sqrt{2q_1\cdot \log n}}{\sqrt{n}}\} = 4\log^2(n)\cdot n^{-q_1/3}\cdot(1+o(1)), \qquad q_1 > 3.$$

This concludes the proof of Lemma 6.2. □

7.4 Proof of Lemma 6.3.

For short, write $\hat{t}_n = \hat{t}_n(\gamma)$, $t_n = t_n(\gamma)$, $\varphi_n(t) = \varphi_n(t; X_1, \ldots, X_n, n)$, $\varphi(t) = \varphi(t;\mu,\sigma,n)$, and $\Lambda_n = \Lambda_n(q, A; \mu_0, \sigma_0, \epsilon_0)$. We claim that for sufficiently large n, $|\varphi(t)|$ is monotonely decreasing in t over $[\log\log n, \infty)$. In fact, using Lemma 6.5, when $n \to \infty$, $\inf_{\{t\ge \log\log n\}}\{\sigma_0^2(\varphi; t)\} = \sigma_0^2\cdot(1+o(1)) > 0$; recall that

$$\frac{d}{dt}|\varphi(t)| = -t\cdot|\varphi(t)|\cdot\sigma_0^2(\varphi,t), \tag{7.19}$$

the monotonicity follows directly.

We now focus on the event $D_n = \{W_0(\varphi_n; n) \leq \sqrt{2q_1 \log n}/\sqrt{n}\}$. Recall that $|\varphi(t_n)| = |\varphi_n(\hat{t}_n)| = n^{-\gamma}$, so

$$\big||\varphi(\hat{t}_n)| - |\varphi(t_n)|\big| = \big||\varphi(\hat{t}_n)| - |\varphi_n(\hat{t}_n)|\big| \leq |\varphi(\hat{t}_n) - \varphi_n(\hat{t}_n)| \leq \sqrt{2q_1 \log n}/\sqrt{n}; \quad (7.20)$$

combining (7.19) and (7.20) and using Taylor expansion, there is a ξ falling between t_n and $\hat{t}_n$ such that

$$|\hat{t}_n - t_n| = \left|\frac{|\varphi(\hat{t}_n)| - |\varphi(t_n)|}{|\varphi|'(\xi)}\right| \leq \frac{\sqrt{2q_1 \log n}/\sqrt{n}}{\xi \cdot |\varphi(\xi)| \cdot |\sigma_0^2(\varphi, \xi)|}. \quad (7.21)$$

At the same time, elementary calculus shows

$$(1 - 2\epsilon_0) e^{-\sigma_0^2 t^2/2} \leq |\varphi(t)| \leq e^{-\sigma_0^2 t^2/2}, \qquad \forall\, t > 0. \quad (7.22)$$

Combining (7.20) and (7.22), it follows that $\hat{t}_n \geq \log\log n$ for sufficiently large n. Since $|\varphi(t)|$ is monotone over $[\log\log n, \infty)$, so (7.20) and (7.22) further imply that $|\varphi(\xi)| \sim n^{-\gamma}$ and $\xi \sim \hat{t}_n \sim t_n \sim \sqrt{2\gamma \log n}/\sigma_0$; these, together with Lemma 6.5, imply that $\sigma_0^2(\varphi, \xi) \sim \sigma_0^2$. Inserting these into (7.21) gives $|\hat{t}_n - t_n| \lesssim \frac{\sqrt{2q_1 \log n}/\sqrt{n}}{\sigma_0^2 \cdot t_n \cdot n^{\gamma}} \sim \frac{1}{\sigma_0} \cdot \sqrt{q_1/\gamma} \cdot n^{\gamma - 1/2}$. □

7.5 Proof of Lemma 6.4

Lay out a grid $t_k = (t_n(\gamma) - \tau_0 n^{\gamma - 1/2}) + \frac{k}{n^{\delta}}$, for $1 \leq k \leq 2\tau_0 n^{\delta + \gamma - 1/2}$ and $\delta \in [1/2, \infty)$. For any $t \in [t_k, t_{k+1}]$,

$$|\varphi_n'(t) - \varphi'(t)| \leq |\varphi_n'(t_k) - \varphi'(t_k)| + n^{-\delta} \cdot \left(\sup_{|\xi - t_n(\gamma)| \leq \tau_0 \cdot n^{\gamma - 1/2}} |\varphi_n''(\xi) - \varphi''(\xi)| \right). \quad (7.23)$$

By direct calculations and the definition of $\bar{s}_n$,

$$|\varphi_n''(\xi) - \varphi''(\xi)| \leq \frac{1}{n} \sum_{j=1}^{n} (X_j^2 + E[X_j^2]) \equiv \frac{1}{n} \sum_{j=1}^{n} (X_j^2 - E[X_j^2]) + 2\bar{s}_n^2,$$

it thus follows that:

$$|\varphi_n'(t) - \varphi'(t)| \leq |\varphi_n'(t_k) - \varphi'(t_k)| + n^{-\delta} \cdot \left[\frac{1}{n} \sum_{j=1}^{n} (X_j^2 - E(X_j^2)) + 2\bar{s}_n^2\right] \quad (7.24)$$

$$\leq |\varphi_n'(t_k) - \varphi'(t_k)| + n^{-\delta} \cdot [\frac{1}{n} \sum_{j=1}^{n} (X_j^2 - E(X_j^2))] + \frac{2\bar{s}_n^2}{\sqrt{n}}. \quad (7.25)$$

Now, denote $q_1 = q/2 - 1$ for short, write:

$$\frac{\bar{s}_n(\sqrt{(q-2)\log n} + 2\bar{s}_n)}{\sqrt{n}} = \frac{\bar{s}_n(\sqrt{2q_1 \log n} + 2\bar{s}_n)}{\sqrt{n}} = \lambda_1(q,n) + \lambda_2(q,n) + \frac{2\bar{s}_n^2}{\sqrt{n}}, \quad (7.26)$$

where $\lambda_1(q_1, n) = \big(\bar{s}_n\sqrt{2q_1 \log n} - (\frac{\log\log n}{2\bar{s}_n\sqrt{2q_1 \log n}})\big)/\sqrt{n}$ and $\lambda_2(q_1, n) = \big(\frac{\log\log n}{2\bar{s}_n\sqrt{2q_1\log n}}\big)/\sqrt{n}$. Compare (7.26) with (7.24) - (7.25) gives:

$$P\{\sup_{|t-t_n(\gamma)|\leq \pi_0\cdot n^{\gamma-1/2}} |\varphi_n'(t) - \varphi'(t)| \geq \frac{\bar{s}_n \cdot (\sqrt{2q_1 \log n} + 2\bar{s}_n)}{\sqrt{n}}\} \leq I + II,$$

where $I = P\{\sup_{1\leq k\leq 2\pi_0 n^{\delta+\gamma-1/2}} |\varphi_n'(t_k) - \varphi'(t_k)| \geq \lambda_1(q_1, n)\}$, and $II = P\{n^{-\delta} \cdot [\frac{1}{n}\sum_{j=1}^n (X_j^2 - EX_j^2)] \geq \lambda_2(q_1, n)\}$.

For I, by [18, Theorem 1] and direct calculations,

$$I \leq (2\pi_0 n^{\delta+\gamma-1/2})\cdot \bar{o}(e^{-\frac{1}{2}n\lambda_1^2(q_1,n)}) \leq (2\pi_0 n^{\delta+\gamma-1/2})\cdot\bar{o}(n^{-q_1}) = \bar{o}(n^{\delta+\gamma-1/2-q_1}). \quad (7.27)$$

For II, we study for the case $q < 4$ and the case $q \geq 4$ separately.

For the case $q < 4$, set $\delta = (q_1 + 1 - \gamma)/2 > 1/2$, by Chebyshev's inequality,

$$II = P\{\frac{1}{n}\sum_{i=1}^n X_j^2 \geq \bar{s}_n^2 + n^\delta \cdot \lambda_2(q_1, n)\} \leq \frac{\bar{s}_n^2}{\bar{s}_n^2 + n^\delta \cdot \lambda_2(q_1, n)} \leq \bar{o}(n^{1/2-\delta}), \quad (7.28)$$

where we have used the fact that $\bar{s}_n^2$ is uniformly bounded from above by a constant $C_1 = C_1(q, A, \mu_0, \sigma_0) < \infty$. Notice that the choice of δ satisfies $\delta + \gamma - 1/2 - q_1 = 1/2 - \delta = (1+\gamma-q/2)/2$, combining (7.27) and (7.28) gives $I + II \leq \bar{o}(n^{(1+\gamma-q/2)/2})$.

For the case $q \geq 4$, notice that $\frac{1}{n}\sum_{j=1}^n E(X_j^2 - E[X_j^2])^2$ is uniformly bounded from above by a constant $C_2 = C_2(q, A, \mu_0, \sigma_0) < \infty$, it follows from Chebyshev's inequality that

$$II \leq \left(\frac{C_2}{\lambda_2^2(q_1, n)\cdot n \cdot n^{2\delta}}\right) = \bar{o}(n^{-2\delta}). \quad (7.29)$$

Set $\delta = \max\{1/2, (q-1-2\gamma)/6\}$, combining (7.27) and (7.29) gives:

$$I + II \leq \begin{cases} \bar{o}(n^{\gamma+1-q/2}), & 4 \leq q \leq 4 + 2\gamma, \\ \bar{o}(n^{(2\gamma+1-q)/3}), & q > 4 + 2\gamma. \end{cases}$$

This finishes the proof of Lemma 6.4. □

7.6 Proof of Lemma 6.5

First, we show (6.2). Write $|\varphi(t)| = |\varphi_0(t)| \cdot |1 + r(t)|$, recall that $\sigma_0^2(\varphi_0; t) \equiv \sigma_0^2 t$, so

$$\frac{\frac{d}{dt}|\varphi(t)|}{|\varphi(t)|} = \frac{\frac{d}{dt}|\varphi_0(t)| \cdot |1 + r(t)| + |\varphi_0(t)| \cdot \frac{d}{dt}|1 + r(t)|}{|\varphi_0(t)| \cdot |1 + r(t)|} = -\sigma_0^2 t + \frac{\frac{d}{dt}|1 + r(t)|}{|1 + r(t)|},$$

and it follows that $\sigma_0^2(\varphi, t) - \sigma_0^2 = -\frac{d}{dt}(|1 + r(t)|)/(t \cdot |1 + r(t)|)$, which yields (6.2) by direct calculations.

Next, we show (6.3). For short, we drop t from all expressions whenever there is no confusion. Since $\varphi = \varphi_0(1 + r)$, $\mathrm{Re}(\varphi) = \mathrm{Re}(\varphi_0) + \mathrm{Re}(r)\mathrm{Re}(\varphi_0) - \mathrm{Im}(r)\mathrm{Im}(\varphi_0)$, and $\mathrm{Im}(\varphi) = \mathrm{Im}(\varphi_0) + \mathrm{Im}(r)\mathrm{Re}(\varphi_0) + \mathrm{Re}(r)\mathrm{Im}(\varphi_0)$; it can be showed that

$$\mathrm{Re}(\varphi') \cdot \mathrm{Im}(\varphi) - \mathrm{Im}(\varphi') \cdot \mathrm{Re}(\varphi) = I + II, \tag{7.30}$$

where $I = -|1 + r|^2 \mu_0 |\varphi_0|^2$, and $II = |\varphi_0|^2 \cdot [-\mathrm{Im}(r') + \mathrm{Re}(r')\mathrm{Im}(r) - \mathrm{Im}(r')\mathrm{Re}(r)]$. The proof of (7.30) is long, so we leave it to the end of this section. Now,

$$\mu_0(\varphi; t) = -\frac{I + II}{|\varphi|^2} = \mu_0 + \frac{\mathrm{Im}(r') - \mathrm{Re}(r')\mathrm{Im}(r) + \mathrm{Re}(r)\mathrm{Im}(r')}{|1 + r|^2},$$

so by Cauchy-Schwartz inequality,

$$|\mu_0(\varphi, t) - \mu_0| = \frac{|\mathrm{Im}(r') - \mathrm{Re}(r')\mathrm{Im}(r) + \mathrm{Im}(r')\mathrm{Re}(r)|}{|1 + r|^2} \leq |r'| \cdot \frac{1 + |r|}{|1 + r|^2},$$

and (6.3) follows directly.

Next, we show (a1) and (a3). (a2) follows directly from (a1) and direct calculations, so we omit it.

We now show (a1). For the 5 inequalities, the proofs for the first 4 are similar, so we only show the second one and the last one. First, consider the second inequality. Use Hölder's inequality,

$$\mathrm{Ave}_{\{j:\, (\mu_j, \sigma_j) \neq (\mu_0, \sigma_0)\}} \{|\mu_j - \mu_0| + (\sigma_j^2 - \sigma_0^2)^{1/2}\} \leq A, \tag{7.31}$$

$$\mathrm{Ave}_{\{j:\, (\mu_j, \sigma_j) \neq (\mu_0, \sigma_0)\}} \{(\sigma_j^2 - \sigma_0^2)\} \leq A^2. \tag{7.32}$$

Note that $\sup_{\{x\geq 0\}} xe^{-x^2/2} = 1/e \leq 1$, direct calculations show that

$$\begin{aligned} |g'(t)| &\leq \mathrm{Ave}_{\{j:\,(\mu_j,\sigma_j)\neq(\mu_0,\sigma_0)\}}\{e^{-\frac{(\sigma_j^2-\sigma_0^2)t^2}{2}}\cdot[|\mu_j-\mu_0|+(\sigma_j^2-\sigma_0^2)t]\} \\ &\leq \mathrm{Ave}_{\{j:\,(\mu_j,\sigma_j)\neq(\mu_0,\sigma_0)\}}\{|\mu_j-\mu_0|+(\sigma_j^2-\sigma_0^2)^{1/2}\cdot[(\sigma_j^2-\sigma_0^2)^{1/2}t\cdot e^{-\frac{(\sigma_j^2-\sigma_0^2)t^2}{2}}]\} \\ &\leq \mathrm{Ave}_{\{j:\,(\mu_j,\sigma_j)\neq(\mu_0,\sigma_0)\}}\{|\mu_j-\mu_0|+(\sigma_j^2-\sigma_0^2)^{1/2}\}, \end{aligned} \tag{7.33}$$

the second inequality follows directly by using (7.31). Second, consider the last inequality. By the definition of τ_n and (7.31), it is seen

$$\mathrm{Ave}_{\{j:\,(\mu_j,\sigma_j)\neq(\mu_0,\sigma_0)\}}\{e^{-\frac{(\sigma_j^2-\sigma_0^2)t^2}{2}}\cdot|\mu_j-\mu_0|\}\leq Ae^{-\frac{\tau_n t^2}{2}}. \tag{7.34}$$

At the same time, notice that $\sup_{\{x\geq 0\}} xe^{-x/2} = 2/e$, so $e^{-\frac{(\sigma_j^2-\sigma_0^2)t^2}{2}}\cdot(\sigma_j^2-\sigma_0^2)t \leq \min\{e^{-\tau_n t^2/2}\cdot(\sigma_j^2-\sigma_0^2)t, 2/(et)\}$, and it follows from (7.32) that

$$\mathrm{Ave}_{\{j:\,(\mu_j,\sigma_j)\neq(\mu_0,\sigma_0)\}}\{e^{-\frac{(\sigma_j^2-\sigma_0^2)t^2}{2}}\cdot(\sigma_j^2-\sigma_0^2)t\}\leq\min\{A^2e^{-\tau_n t^2/2}t, 2/(et)\}. \tag{7.35}$$

The claim follows by combining (7.33) - (7.35).

Next, we show (a3). As the proofs are similar, we only show that corresponds to σ_0^2. By (6.2), $|\sigma_0^2(\varphi;t)-\sigma_0^2|\to 0$ uniformly; by (a1), it is not hard to show that $\sigma_0^2(\varphi;t)$ and its first two derivatives are all uniformly bounded; so all remains to show is that $\frac{d}{dt}\sigma_0^2(\varphi;t)\to 0$ uniformly. Observe that for any twice differentiable function f and $\Delta>0$, $|\frac{f(t+\Delta)-f(t)}{\Delta}-f'(t)|\leq\sup_{\{s\}}\{|f''(s)|\}\Delta$, so it follows $|f'(t)|\leq\{\sup_{\{s\}}\{|f''(s)|\}\Delta+\frac{1}{\Delta}\sup_{\{s,s'\geq t\}}\{|f(s)-f(s')|\}$; the claim follows by taking $\Delta=\sqrt{\sup_{\{s,s'\geq t\}}\{|f(s)-f(s')|\}/\sup_{\{s\}}\{|f''(s)|\}}$ and $f(t)=\sigma_0^2(\varphi;t)$.

Lastly, we validate (7.30). Write $\mathrm{Re}(\varphi')=\mathrm{Re}(\varphi_0')+\mathrm{Re}(r')\mathrm{Re}(\varphi_0)+\mathrm{Re}(r)\mathrm{Re}(\varphi_0')-\mathrm{Im}(r')\mathrm{Im}(\varphi_0)-\mathrm{Im}(r)\mathrm{Im}(\varphi_0')$, and $\mathrm{Im}(\varphi')=\mathrm{Im}(\varphi_0')+\mathrm{Im}(r')\mathrm{Re}(\varphi_0)+\mathrm{Im}(r)\mathrm{Re}(\varphi_0')+$

$\mathrm{Re}(r')\mathrm{Im}(\varphi_0) + \mathrm{Re}(r)\mathrm{Im}(\varphi_0')$, we have

$$\mathrm{Re}(\varphi') \cdot \mathrm{Im}(\varphi) - \mathrm{Im}(\varphi') \cdot \mathrm{Re}(\varphi) = \mathrm{Re}(\varphi_0')[\mathrm{Im}(\varphi_0) + \mathrm{Im}(r)\mathrm{Re}(\varphi_0) + \mathrm{Re}(r)\mathrm{Im}(\varphi_0)]$$
$$+ \mathrm{Re}(r')\mathrm{Re}(\varphi_0)[\mathrm{Im}(\varphi_0) + \mathrm{Im}(r)\mathrm{Re}(\varphi_0) + \mathrm{Re}(r)\mathrm{Im}(\varphi_0)] + \mathrm{Re}(r)\mathrm{Re}(\varphi_0')[\mathrm{Im}(\varphi_0)+$$
$$\mathrm{Im}(r)\mathrm{Re}(\varphi_0) + \mathrm{Re}(r)\mathrm{Im}(\varphi_0)] - \mathrm{Im}(r')\mathrm{Im}(\varphi_0)[\mathrm{Im}(\varphi_0) + \mathrm{Im}(r)\mathrm{Re}(\varphi_0) + \mathrm{Re}(r)\mathrm{Im}(\varphi_0)]$$
$$- \mathrm{Im}(r)\mathrm{Im}(\varphi_0')[\mathrm{Im}(\varphi_0) + \mathrm{Im}(r)\mathrm{Re}(\varphi_0) + \mathrm{Re}(r)\mathrm{Im}(\varphi_0)] - \mathrm{Im}(\varphi_0')[\mathrm{Re}(\varphi_0) + \mathrm{Re}(r)$$
$$\mathrm{Re}(\varphi_0) - \mathrm{Im}(r)\mathrm{Im}(\varphi_0)] - \mathrm{Im}(r')\mathrm{Re}(\varphi_0)[\mathrm{Re}(\varphi_0) + \mathrm{Re}(r)\mathrm{Re}(\varphi_0) - \mathrm{Im}(r)\mathrm{Im}(\varphi_0)]-$$
$$\mathrm{Im}(r)\mathrm{Re}(\varphi_0')[\mathrm{Re}(\varphi_0) + \mathrm{Re}(r)\mathrm{Re}(\varphi_0) - \mathrm{Im}(r)\mathrm{Im}(\varphi_0)] - \mathrm{Re}(r')\mathrm{Im}(\varphi_0)[\mathrm{Re}(\varphi_0)+$$
$$\mathrm{Re}(r)\mathrm{Re}(\varphi_0) - \mathrm{Im}(r)\mathrm{Im}(\varphi_0)] - \mathrm{Re}(r)\mathrm{Im}(\varphi_0')[\mathrm{Re}(\varphi_0) + \mathrm{Re}(r)\mathrm{Re}(\varphi_0) - \mathrm{Im}(r)\mathrm{Im}(\varphi_0)];$$

by cancellations, this reduces to

$$\mathrm{Re}(\varphi_0') \cdot [\mathrm{Im}(\varphi_0) + \mathrm{Re}(r)\mathrm{Im}(\varphi_0)] + \mathrm{Re}(r')\mathrm{Re}(\varphi_0)[\mathrm{Im}(r)\mathrm{Re}(\varphi_0)] + \mathrm{Re}(r)\mathrm{Re}(\varphi_0')$$
$$[\mathrm{Im}(\varphi_0) + \mathrm{Re}(r)\mathrm{Im}(\varphi_0)] - \mathrm{Im}(r')\mathrm{Im}(\varphi_0)[\mathrm{Im}(\varphi_0) + \mathrm{Re}(r)\mathrm{Im}(\varphi_0)] - \mathrm{Im}(r)\mathrm{Im}(\varphi_0')$$
$$[\mathrm{Im}(r)\mathrm{Re}(\varphi_0)] - \mathrm{Im}(\varphi_0') \cdot [\mathrm{Re}(\varphi_0) + \mathrm{Re}(r)\mathrm{Re}(\varphi_0)] - \mathrm{Im}(r')\mathrm{Re}(\varphi_0)[\mathrm{Re}(\varphi_0) + \mathrm{Re}(r)$$
$$\mathrm{Re}(\varphi_0)] - \mathrm{Im}(r)\mathrm{Re}(\varphi_0') \cdot [-\mathrm{Im}(r)\mathrm{Im}(\varphi_0)] - \mathrm{Re}(r')\mathrm{Im}(\varphi_0)[-\mathrm{Im}(r)\mathrm{Im}(\varphi_0)]$$
$$- \mathrm{Re}(r)\mathrm{Im}(\varphi_0')[\mathrm{Re}(\varphi_0) + \mathrm{Re}(r)\mathrm{Re}(\varphi_0)];$$

by recombinations, this reduces to $|1+r|^2 \cdot [\mathrm{Re}(\varphi_0')\mathrm{Im}(\varphi_0) - \mathrm{Re}(\varphi_0)\mathrm{Im}(\varphi_0')] + |\varphi_0|^2 \cdot [-\mathrm{Im}(r') + \mathrm{Re}(r')\mathrm{Im}(r) - \mathrm{Im}(r')\mathrm{Re}(r)]$. Note that $[\mathrm{Re}(\varphi_0')\mathrm{Im}(\varphi_0) - \mathrm{Re}(\varphi_0)\mathrm{Im}(\varphi_0')] = -\mu_0|\varphi_0|^2$, (7.30) follows directly. $\square$