A Generalized Fourier Approach to Estimating the Null Parameters and Proportion of nonnull Effects in Large-Scale Multiple Testing

Jiashun Jin[1] and Jie Peng[2] and Pei Wang[3]
[1]Statistics Department, Carnegie Mellon University
[2]Statistics Department, University of California at Davis
[3]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center

November 19, 2009

Abstract

In a recent paper [4], Efron pointed out that an important issue in large-scale multiple hypothesis testing is that the null distribution may be unknown and need to be estimated. Consider a Gaussian mixture model, where the null distribution is known to be normal but both null parameters—the mean and the variance—are unknown. We address the problem with a method based on Fourier transformation. The Fourier approach was first studied by Jin and Cai [9], which focuses on the scenario where any non-null effect has either the same or a larger variance than that of the null effects. In this paper, we review the main ideas in [9], and propose a generalized Fourier approach to tackle the problem under another scenario: any non-null effect has a larger mean than that of the null effects, but no constraint is imposed on the variance. This approach and that in [9] complement with each other: each approach is successful in a wide class of situations where the other fails. Also, we extend the Fourier approach to estimate the proportion of non-null effects. The proposed procedures perform well both in theory and on simulated data.

Keywords: empirical null, Fourier transformation, generalized Fourier transformation, proportion of non-null effects, sample size calculation,

AMS 1991 subject classifications: Primary 62G10, 62G05; secondary 62H15, 62H20.

1 Introduction

Large-scale multiple testing is a recent area of active research in statistics, where one tests thousands or even millions of null hypotheses *simultaneously*:

$$H_j, \qquad j = 1, \ldots, n.$$

Associated with each null hypothesis is a test statistics X_j, which, depending on the situation, can be a summary statistic, a p-value, a regression coefficient, or a transform coefficient, etc.. We say that X_j contains a *null effect* if H_j is true, and contains a *non-null effect* if otherwise.

A convenient model is the Bayesian hierarchical model [4, 6] which we now describe. Fix $0 < \epsilon < 1$. For each $1 \leq j \leq n$, we flip a coin with probability ϵ of landing tail. If the coin lands head, we draw X_j from a common density function $f_0(x)$ which we call the *null density*. If the coin lands tail, we draw X_j from an individual density function $\xi_j(x)$, where ξ_j itself is randomly generated according to a fixed probability measure Ξ. In effect, X_j can be viewed as

samples from the density $f_1(x) \equiv \int \xi(x)d\Xi(\xi)$, which we call the *alternative density*; see [6, 8]. Marginally, X_j can be deemed as samples from the following two-component mixing density:

$$X_j \stackrel{iid}{\sim} (1 - \epsilon)f_0(x) + \epsilon f_1(x) \equiv f(x). \tag{1.1}$$

The parameter ϵ is closely related to the proportion of non-null effects (i.e., the fraction of null hypotheses that are untrue). In fact, under the Gaussian mixture model, the number of untrue hypothesis is distributed as Binomial with parameters n and ϵ. So when n is large, the difference between ϵ and the actual fraction $\leq O_p(\sqrt{\epsilon/n})$ and is usually negligible. For this reason, we call ϵ the proportion of the non-null effects in this paper.

The null density is the starting point for any testing procedures. In many scenarios, the null density is assumed as known. However, somewhat surprisingly, this assumption may be incorrect in some multiple testing situations as pointed out by Efron [4]. Efron illustrated his point with a breast cancer microarray data, which is based on 15 patients with 7 having BRCA1 mutation and 8 having BRCA2 mutation. For each patient, the same set of 3226 genes were measured and it is of interest to find which genes are differentially expressed. For each gene, a studentized-t score was calculated and then transformed to a z-score (see [4]) for the details). Efron argued that, although the theoretical null should be the standard normal $N(0, 1)$, another null density, $N(0.02, 2.50)$ seems to be more appropriate. Efron called the later the *empirical null* and demonstrated convincingly that it is better to use the empirical null instead of the theoretical null in many situations.

There are many possible reasons why the empirical null may be different from the theoretical null. Take the breast cancer microarray data for example, the studentized-t statistics may not be truly t-distributed due to failed distributional assumptions. There may be covariates (such as age of the patients) that has not been observed in the data. The correlation across different genes (also that across different arrays) has been neglected. All these factors may drive the empirical null far from the theoretical null.

Unfortunately, unlike the theoretical null, the empirical null is usually unknown. Thus how to estimate the empirical null is a problem of major interest.

1.1 Identifiability issue and constrained Gaussian mixture models

Note that in Model (1.1), some may call f_0 the null density, and some may call f_1 the null density. To resolve this issue, we fix a constant $\epsilon_0 \in (0, 1/2)$ and assume

$$0 < \epsilon \leq \epsilon_0,$$

so that the null density is tied to the majority of the hypotheses.

We adopt the Gaussian model as suggested in Efron [4]. In detail, let $\phi(\cdot)$ be the density of $N(0, 1)$. We assume that the null density f_0 is Gaussian with an unknown mean u_0 and an unknown variance σ_0^2:

$$f_0(x) = \frac{1}{\sigma_0}\phi(\frac{x - u_0}{\sigma_0}).$$

At the same time, we assume that the alternative density f_1 is a Gaussian mixture (both a location mixture and a scale mixture) with a bivariate mixing distribution $H(u, \sigma)$:

$$f_1(x) = \int \frac{1}{\sigma}\phi(\frac{x - u}{\sigma})dH(u, \sigma).$$

The marginal density of X_j is then

$$f(x) = f(x; u_0, \sigma_0, \epsilon, H) = (1 - \epsilon)\frac{1}{\sigma_0}\phi(\frac{x - u_0}{\sigma_0}) + \epsilon \int \frac{1}{\sigma}\phi(\frac{x - u}{\sigma})dH(u, \sigma). \tag{1.2}$$

With the Gaussian model, the problem of estimating the null density reduces into the problem of estimating the null parameters (u_0, σ_0^2).

However, the null density in the above Gaussian model is not always identifiable. This is because, without constraint on $H(\cdot, \cdot)$, f_1 can be very close or even identical to f_0. Fortunately,

there are many natural constraints that we can put on $H(\cdot, \cdot)$ to resolve this problem. Below are some examples.

Definition 1.1 *Fix $\epsilon_0 \in (0, 1/2)$, u_0, and $\sigma_0 > 0$. We say that $f(x) = f(x; u_0, \sigma_0, \epsilon, H)$ is a Gaussian mixture density constrained with Elevated Variances with respect to parameters $(u_0, \sigma_0, \epsilon_0)$ if it has the form as in (1.2), and that the proportion ϵ and the mixing distribution H satisfy*

$$0 < \epsilon \leq \epsilon_0, \qquad P_H(\sigma \geq \sigma_0) = 1, \qquad P_H\big((u, \sigma) \neq (u_0, \sigma_0)\big) = 1. \qquad (1.3)$$

We refer to the Gaussian model (1.2) with constraints in (1.3) as $GEV(u_0, \sigma_0, \epsilon_0)$.

For short, we write $GEV(u_0, \sigma_0, \epsilon_0)$ as GEV whenever there is no confusion. In the definition above, u and σ denote the location and scale parameters from the mixing distribution, and P_H denotes the probability under the mixing distribution $H(\cdot, \cdot)$. This models a situation where the variance associated with an individual non-null effect is no less than that of a null effect. The following lemma shows that, given (1.3), the triplets $(u_0, \sigma_0, \epsilon)$ are uniquely determined by $f(x)$ and the identifiability issue is therefore resolved.

Lemma 1.1 *Given a density $f(x) = f(x; u_0, \sigma_0, \epsilon, H)$ satisfying (1.2) and (1.3), the parameters u_0, σ_0, and ϵ are uniquely determined by $f(x)$.*

This lemma is proved in Section 5. Note that if we replace the constraint $P_H(\sigma \geq \sigma_0) = 1$ by $P(\sigma \leq \sigma_0) = 1$, then the identifiability issue persists (the construction of counter examples is elementary and we skip it).

Alternatively, we define GEM as follows.

Definition 1.2 *Fix $\epsilon_0 \in (0, 1/2)$, u_0, and $\sigma_0 > 0$. We say that $f(x) = f(x; \epsilon, u_0, \sigma_0, H)$ is a Gaussian mixture density constrained with Elevated Means with respect to parameters $(u_0, \sigma_0, \epsilon_0)$ if it has the form as in (1.2), and that the proportion ϵ and the mixing distribution H satisfy*

$$0 < \epsilon \leq \epsilon_0, \qquad P_H(u > u_0) = 1. \qquad (1.4)$$

We refer to the Gaussian model (1.2) with constraints in (1.4) as $GEM(u_0, \sigma_0, \epsilon_0)$.

GEM models a situation where the mean associated with an individual non-null effect is larger than that of a null effect. The following lemma is proved in Section 5.

Lemma 1.2 *Given a density $f(x) = f(x; u_0, \sigma_0, \epsilon, H)$ satisfying (1.2) and (1.4), the parameters u_0, σ_0, and ϵ are uniquely determined by $f(x)$.*

For the case where we replace the constraint $P_H(u > u_0) = 1$ in (1.4) with $P_H(u < u_0) = 1$, the discussion is similar. Also, we can relax the constraint to $P_H(u \geq u_0) = 1$. But by doing so we need some conditions on σ. For reasons of space, we skip the discussion along these two lines.

GEV and GEM are the two main models we study in this paper. Despite the additional constraints, both models are broad enough to accommodate many interesting cases that arise in real applications. In sections below, we discuss possible approaches to consistently estimating the null parameters in GEV and GEM.

1.2 A Fourier approach to estimating the null parameters in GEV

Conventionally, one estimates the null parameters with either empirical moments or extreme observations. However, in these quantities, the information containing the null parameters is highly distorted by the non-null effects. A non-orthodox approach is therefore necessary. In a recent work [9], Jin and Cai proposed a Fourier approach to estimating the null parameters in GEV. We now briefly explain the idea.

When it comes to a density function, one usually pictures it as a smooth curve that spreads over the real line. Joseph Fourier taught us a different view point: a normal density $N(u, \sigma^2)$ is not only a bell shaped curve centered at u, but also a wave oscillate at the frequency u. In fact, the Fourier transform of the density $N(u, \sigma^2)$ can be decomposed into two components: the amplitude function determined by σ^2, and the phase function determined by u:

$$e^{-\sigma^2 t^2/2} \cdot e^{itu} \equiv \text{Amplitude} \cdot \text{Phase function}, \qquad i = \sqrt{-1}. \qquad (1.5)$$

3

Consequently, we can view a Gaussian mixture as a superposition of waves with different frequencies and different amplitudes.

We now invoke GEV. The above investigation gives rise to an interesting approach to estimating the null parameters. Denote the *empirical characteristic function* by

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

For an appropriately large frequency t, the stochastic fluctuation is negligible and ψ_n reduces to its non-stochastic counterpart—the *underlying characteristic function* $\psi(t) = E[\psi_n(t)]$. By direct calculations,

$$\psi(t) = \psi(t; u_0, \sigma_0, \epsilon, H) \equiv \psi_0(t)[1 + s(t)],$$

where

$$\psi_0(t) = \psi_0(t; u_0, \sigma_0, \epsilon) = (1 - \epsilon)e^{iu_0 t - \sigma_0^2 t^2/2},$$

and

$$s(t) = s(t; u_0, \sigma_0, \epsilon, H) = \frac{\epsilon}{1 - \epsilon} \int e^{i(u-u_0)t - (\sigma^2 - \sigma_0^2)t^2/2} dH(u, \sigma). \tag{1.6}$$

Now, with GEV and a little bit extra condition, $s(t) \approx 0$. For example, if we assume that $P_H(\sigma > \sigma_0) = 1$, then at a high frequency t,

$$|s(t)| \leq \frac{\epsilon}{1 - \epsilon} \int e^{-(\sigma^2 - \sigma_0^2)t^2/2} dH(u, \sigma) \approx 0. \tag{1.7}$$

This says that in GEV, as the frequency t tends to ∞, the waves corresponding to the alternative density damps faster than that associated with the null density. Therefore, the information containing the null parameters is asymptotically preserved in high frequency Fourier transform, where the distortion of non-null effects is negligible. In other words, for an appropriately large frequency t,

$$\psi_n(t) \approx \psi(t) \approx \psi_0(t).$$

Now, since $\psi_0(t)$ has a very simple form, we can solve (u_0, σ_0) (and also ϵ) from it.

The elaboration of the idea gives rise to the estimators in [9], which are proved to be uniformly consistent to the null parameters across a wide range of mixing distributions $H(\cdot, \cdot)$. It was also shown in [9] that these estimators attain the optimal rate of convergence. See the details therein. These works reveal that, somewhat surprisingly, the right place to estimate the null parameters is in the frequency domain, rather than in the spatial domain as one may have expected.

1.3 A generalized-Fourier approach to estimating the null parameters in GEM

Despite its encouraging performance in GEV, the above approach does not yield a satisfactory estimation in GEM. To see the point, we note that the key for the success of the above approach is (1.7), which critically depends on the assumption of $P_H(\sigma > \sigma_0) = 1$. Note that such an assumption does not hold in GEM. As a result, the above approach ceases to perform well.

Fortunately, there is an easy fix. The key is to replace the Fourier transformation by the generalized Fourier transformation (to be introduced below), so that in the frequency domain, the roles of the mean and the variance are sort of "swapped". In detail, let

$$\omega = -(1 + i)/\sqrt{2}, \qquad (\text{ note } \omega^2 = i).$$

For any density function $h(x)$, the generalized Fourier transformation is

$$\int h(x) \exp(\omega x) dx,$$

4

provided that the function $h(x)\exp(\omega x)$ is absolutely integrable. In particular, the generalized Fourier transform of the Gaussian density $N(u, \sigma^2)$ is

$$\exp\left(-\frac{ut}{\sqrt{2}}\right) \cdot \exp\left(i[-\frac{ut}{\sqrt{2}} + \frac{\sigma^2 t^2}{2}]\right) \equiv \text{Amplitude function} \cdot \text{Phase function.} \qquad (1.8)$$

Now, the amplitude is uniquely determined by the mean (compare with (1.5)).

The remaining part of the idea is similar to that in the preceding section. Denote the *generalized-empirical characteristic function* by

$$\varphi_n(t) = \varphi_n(t; X_1, \dots, X_n) = \frac{1}{n} \sum_{j=1}^{n} \exp(\omega t X_j).$$

For large n and an appropriately chosen t, one expects that the stochastic fluctuation is negligible, and that $\varphi_n(t)$ reduces approximately to the *generalized characteristic function*,

$$\varphi(t) = \varphi(t; u_0, \sigma_0, \epsilon, H) \equiv E[\varphi_n(t)].$$

Direct calculations show that

$$\varphi(t) = \varphi_0(t)[1 + r(t)],$$

where

$$\varphi_0(t) = \varphi_0(t; u_0, \sigma_0, \epsilon) = (1 - \epsilon) \exp(\omega u_0 t + i\sigma_0^2 t^2/2),$$

and

$$r(t) = r(t; u_0, \sigma_0, \epsilon, H) = \frac{\epsilon}{1 - \epsilon} \int \exp(\omega(u - u_0)t + i(\sigma^2 - \sigma_0^2)t^2/2)dH(u, \sigma). \qquad (1.9)$$

Recalling that $\omega = -(1 + i)/\sqrt{2}$, it is seen that

$$|r(t)| \leq \frac{\epsilon}{1 - \epsilon} \int \exp(-(u - u_0)t/\sqrt{2})dH(u, \sigma).$$

We now invoke GEM. Similarly, since that $P_H(u > u_0) = 1$, $r(t) \approx 0$ for large t. We expect that

$$\varphi_n(t) \approx \varphi(t) \approx \varphi_0(t).$$

Again, $\varphi_0(t)$ has a very simple form and we can solve (u_0, σ_0) (and also ϵ) from it. In fact, introduce two functionals $u_0(\cdot; t)$ and $\sigma_0^2(\cdot; t)$ by

$$u_0(g; t) = -\frac{\sqrt{2}}{|g(t)|} \frac{d}{dt}|g(t)|, \qquad \sigma_0^2(g; t) = \frac{\sqrt{2}\text{Re}(\omega \bar{g} g')}{t|g(t)|^2}, \qquad (1.10)$$

where g is any complex-valued differentiable function, and $|z|$, $Re(z)$ and $\bar{z}$ denote the module, the real part, and the complex conjugate of a complex number z, correspondingly. The following lemma says that plugging $g = \varphi_0$ into two functionals gives the desired parameters u_0 and σ_0^2, respectively.

Lemma 1.3 *For all $t \neq 0$, $u_0(\varphi_0; t) = u_0$ and $\sigma_0^2(\varphi; t) = \sigma_0^2$.*

Lemma 1.3 can be proved using elementary algebra, so we skip it. Taking $g = \varphi_n$ in (1.10), we expect to have

$$u_0(\varphi_n, t) \approx u_0(\varphi, t) \approx u_0, \qquad \sigma_0^2(\varphi_n, t) \approx \sigma_0^2(\varphi, t) \approx \sigma_0^2.$$

In this paper, we shall carefully study the bias and variance of $u_0(\varphi_n; t)$ and $\sigma_0^2(\varphi_n; t)$, and investigate which choices of t give a good tradeoff between the bias and the variance. We find out that as n tends to ∞, if we set t in an appropriate range, then both estimators are consistent with their estimands, uniformly so across a wide class of situations.

1.4 Estimating the proportion of non-null effects

Seemingly, the approach can be readily generalized to estimate the proportion of non-null effects ϵ. How to estimate the proportion has been the topic of many recent works in the area of large-scale multiple hypothesis testing. See for example [3, 5, 6, 8, 9, 10, 11, 12, 14]. There are two reasons for the enthusiasm. In some applications, the proportion is the quantity that is of direct interest [11]; while more often, knowing the proportion helps to improve many multiple testing procedures, such as the FDR procedure by Benjamini and Hochberg's [2], the local FDR procedure by Efron et al. [5] and the optimal discovery function by Storey [13]. See [8] for more discussions.

In Section 3, we extend the generalized Fourier approach to estimating the proportion in GEM. We discuss two different cases: (1) the null parameters are known; and (2) the null parameters are unknown. In both cases, we find that the estimators are uniformly consistent with the proportion across a wide class of situations.

We remark that the success of the Fourier approach for estimating the null parameters and the proportion is not coincidental. It roots from the key fact that the null density can be isolated from the alternative density in the high frequency Fourier coefficients. Naturally, we shall continue to find the Fourier approach to be successful in estimating many other quantities.

The remaining part of the paper is organized as follows. Section 2 studies the problem of estimating the null parameters in GEM. We show that by choosing an appropriate t, the estimators $u_0(\varphi_n;t)$ and $\sigma_0^2(\varphi_n;t)$ are consistent to the true parameters, uniformly across a wide class of situations. Section 3 studies the problem of estimating the proportion. While the studies in Sections 2–3 are asymptotic, we carry out a few simulation studies in Section 4, and investigate the performance of the proposed estimators for moderately large n. Section 5 contains the proofs for the theorems and lemmas, in the order they appear.

2 Main results

In this section, we limit our attention to GEM and study the estimation errors of $u_0(\varphi_n;t)$ and $\sigma_0^2(\varphi_n;t)$. Since the discussions are similar, we focus on that of $u_0(\varphi_n;t)$. For the asymptotic analysis, we adopt a framework where both ϵ and H may depend on n as n ranges from 1 to ∞ (denoted by ϵ_n and H_n). This covers a much broader situations than that when (ϵ, H) are fixed as n ranges from 1 to ∞.

2.1 Asymptotic framework

Recall that the test statistics X_j are iid samples from

$$f(x) = f(x; u_0, \sigma_0, \epsilon_n, H_n, n) = (1 - \epsilon_n)\frac{1}{\sigma_0}\phi(\frac{x - u_0}{\sigma_0}) + \epsilon_n \int \frac{1}{\sigma}\phi(\frac{x - u}{\sigma})dH_n(u, \sigma). \quad (2.11)$$

As before, fix $\epsilon_0 \in (0, 1/2)$. We suppose that for any $n \geq 1$,

$$0 < \epsilon_n \leq \epsilon_0. \quad (2.12)$$

Of course, the condition can be relaxed so that it only holds for sufficiently large n.

Also, fixing $A > 0$, assume that

$$u_0 \geq -A, \qquad \sigma_0^2 \leq A. \quad (2.13)$$

In addition, we assume that for any $n \geq 1$,

$$P_{H_n}(u > u_0) = 1, \qquad P_{H_n}(\sigma^2 \leq A) = 1. \quad (2.14)$$

These conditions are relatively relaxed, except for the second one in (2.14). We need this condition to control the variance of the estimators (whether this condition can be significantly

relaxed is an open question, which we leave to the future study). In short, we focus the study on the class of marginal densities as follows,

$$\Lambda_n(\epsilon_0, A) = \{f(x) = f(x; u_0, \sigma_0, \epsilon_n, H_n, n) \text{ has the form as in (2.11) that satisfies (2.12)-(2.14)}\}.$$

For any $t > 0$, it follows from the triangle inequality that

$$|u_0(\varphi_n, t) - u_0| \leq |u_0(\varphi_n, t) - u_0(\varphi, t)| + |u_0(\varphi, t) - u_0|.$$

On the right hand side, the first term is the stochastic term, and the second term is the bias term. Seemingly, the performance of the estimator depends on the choice of t. Larger t tends to give a larger stochastic fluctuation but a smaller bias. It turns out that the interesting range of t is $O(\sqrt{\log n})$. In light of this, we calibrate t through a parameter γ by

$$t = t_n(\gamma) = \sqrt{\gamma \log n}, \qquad \gamma > 0.$$

We now study the stochastic term and the bias term separately.

2.2   The stochastic term

We need the following definition.

Definition 2.1 *Fixing a constant r, we say that a sequence $\{b_n\}_{n=1}^{\infty}$ is $\bar{o}(n^{-r})$ if $n^{r-\delta}|b_n| \to 0$ as $n \to \infty$, for all $\delta > 0$. Especially, when $r = 0$, we write $\bar{o}(1)$.*

First, we study the stochastic fluctuation of $\varphi_n(t)$ and $\varphi'_n(t)$. The following lemmas are proved in Section 5.

Lemma 2.1 *Fix $\epsilon_0 \in (0, 1/2)$, $A > 0$, and $\gamma \in (0, 1/A)$. As n tends to ∞,*

$$\sup_{\{f \in \Lambda_n(\epsilon_0, A)\}} \{\mathrm{Var}(\varphi_n(t_n(\gamma)))\} \leq n^{A\gamma - 1}, \tag{2.15}$$

and

$$\sup_{\{f \in \Lambda_n(\epsilon_0, A)\}} \{\mathrm{Var}(\varphi'_n(t_n(\gamma)))\} \lesssim 4A^2 \gamma \log(n) \cdot n^{A\gamma - 1}. \tag{2.16}$$

The upper bounds in (2.15)-(2.16) may be conservative, especially when ϵ_n is small. See the proof for the details (we say two positive sequences $a_n \lesssim b_n$ if $a_n/b_n \leq 1 + o(1)$ for sufficiently large n).

We now relate the stochastic fluctuations of $u_0(\varphi_n; t)$ and $\sigma_0^2(\varphi_n; t)$ to that of $\varphi_n(t)$ and $\varphi'_n(t)$. This is achieved by the following lemma.

Lemma 2.2 *Let $u_0(\cdot; \cdot)$ and $\sigma_0^2(\cdot; \cdot)$ be defined as in (1.10). Fix $t > 0$. For any differentiable complex-valued functions f and g satisfying $|f(t)| \neq 0$ and $|g(t)| \neq 0$,*

$$|u_0(g,t) - u_0(f,t)| \leq \frac{1}{|f(t)|^2} \big[(|u_0(g,t)| \cdot (|f(t)| + |g(t)|) + \sqrt{2}|g'(t)|) |f(t) - g(t)| + \sqrt{2}|f(t)| \cdot |(f(t) - g(t))'|,$$

and

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

Apply Lemma 2.2 with $f = \varphi_n$, $g = \varphi$. Intuitively,

$$\varphi_n(t) \approx \varphi(t), \qquad \varphi'_n(t) \approx \varphi'(t).$$

Also, when $t = t_n(\gamma)$,

$$\varphi_n(t_n(\gamma)) = \bar{o}(1).$$

We therefore expect to have

$$|u_0(\varphi_n, t_n(\gamma)) - u_0(\varphi; t_n(\gamma))| \leq C \frac{1}{|\varphi(t(\gamma))|} |\varphi_n(t(\gamma)) - \varphi(t(\gamma))| + \frac{|\varphi'(t(\gamma))|}{|\varphi(t(\gamma))|^2} |\varphi'_n(t(\gamma)) - \varphi'(t(\gamma))| \leq \bar{o}(n^{A\gamma - 1}).$$

As a result, we have the following lemma.

7

Lemma 2.3 *Fix $\epsilon_0 \in (0, 1/2)$, $A > 0$, and $\gamma \in (0, 1/A)$. As n tends to ∞, except for a probability that tends to 0,*

$$\sup_{\{f \in \Lambda_n(\epsilon_0, A)\}} \{|u_0(\varphi_n; t_n(\gamma)) - u_0(\varphi; t_n(\gamma))|\} \leq \bar{o}(n^{(A\gamma-1)/2}),$$

and

$$\sup_{\{f \in \Lambda_n(\epsilon_0, A)\}} \{|\sigma_0^2(\varphi_n; t_n(\gamma)) - \sigma_0^2(\varphi; t_n(\gamma))|\} \leq \bar{o}(n^{(A\gamma-1)/2}).$$

In conclusion, except for a probability that tends to 0, the stochastic fluctuation of either estimator is of the order of $n^{(A\gamma-1)/2}$. Note that the exponent $(A\gamma - 1)/2 < 0$.

2.3 The bias term

We now discuss the bias term. The following lemma is proved in Section 5.

Lemma 2.4 *Fix $\epsilon_0 \in (1/2)$ and $A > 0$. Let $r(t) = r(t; u_0, \sigma_0, \epsilon_n, H_n)$ be as in (1.9). For any $t > 0$ and $f \in \Lambda_n(\epsilon_0, A)$, there exists a universal constant $C > 0$ such that*

$$|u_0(\varphi; t) - u_0| \leq C|r'(t)|,$$
$$|\sigma_0^2(\varphi; t) - \sigma_0^2| \leq C|r'(t)|/t.$$

Write for short $t_n = t_n(\gamma)$. Under mild conditions, $r'(t_n) \to 0$. We now show some examples where this is the case.

Example 1. The non-null effects are sparse. In this case, we suppose that the parameter ϵ_n tends to 0 as n tends to ∞ at a rate faster than that of $1/t_n$. By the proof of Lemma 2.3,

$$|r'(t_n)| \leq \frac{\epsilon_n}{1 - \epsilon_n} \left[\frac{\sqrt{2}}{et_n} + At_n \right].$$

So as long as $\epsilon_n t_n \to 0$, $|r'(t_n)| \to 0$, regardless of the distribution of $H_n(\cdot, \cdot)$ (of course, the condition of $P_{H_n}(u > u_0) = 1$ is still needed).

Example 2. Elevated means. In this case, we suppose that the mean corresponding to the null density is elevated by at least a small amount $\delta_n > 0$:

$$P_{H_n}(u \geq u_0 + \delta_n) = 1.$$

Recall that $u_0 \geq -A$ and that for any $H_n \in \Lambda_n(\epsilon_0, A)$, $P_{H_n}(|\sigma^2 - \sigma_0^2| \leq A) = 1$. Similar to the proof of Lemma 2.3,

$$|r'(t_n)| \leq \frac{\epsilon_n}{1 - \epsilon_n} \left(\delta_n + \frac{\sqrt{2}}{et_n} + At_n \right) e^{-\delta_n t_n/\sqrt{2}}.$$

As a result, we have the following lemma, whose proof is elementary so we omit it.

Lemma 2.5 *If there is some constant $c_0 > 0$ such that*

$$\liminf_{n \to \infty} \left\{ \frac{\delta_n t_n}{\sqrt{2} \log(t_n)} \right\} \geq (c_0 + 1), \tag{2.17}$$

then $|r'(t_n)| \lesssim A\epsilon_n t_n^{-c_0}$.

As a result, as $n \to \infty$, the bias $\to 0$ if (2.17) holds for some constant $c_0 > 0$, whether ϵ_n tends to 0 or not.

Example 3. When the bivariate random variables (u, σ^2) have a smooth joint density. We re-center u and σ^2 by letting $\delta = u - u_0$ and $\kappa = (\sigma^2 - \sigma_0^2)/2$. Denote the joint density of (δ, κ) by $h_n(\cdot, \cdot)$. We show that the $r'(t_n) = o(1)$ under mild smoothness conditions on $h_n(\cdot, \cdot)$. In detail, for each fixed $\delta > 0$, let $h_n^{FT}(\cdot|\delta)$ be the Fourier transform of the conditional density $h_n(\kappa|\delta)$. Fix $\alpha > 0$. Suppose that there is a generic constant $C > 0$ such that for all δ in the range,

$$|h_n^{FT}(t_n|\delta)| \leq C(1 + |t_n|)^{-\alpha}, \qquad |\frac{d}{dt} h_n^{FT}(t_n|\delta)| \leq C(1 + |t_n|)^{-(\alpha+1)}. \tag{2.18}$$

We have the following lemma, whose proof is elementary so we omit it.

8

Lemma 2.6 *Suppose (2.18) holds for some constant $C > 0$ and $\alpha > 0$. Then there is a generic constant $C > 0$ such that*
$$|r'(t_n)| \leq C\epsilon_n |t_n|^{-2(\alpha+1)},$$

Note that $r'(t_n) \to 0$ in a much broader setting than that in this example.

Combining Lemma 2.3-2.4 and the above examples, we have the following theorem.

Theorem 2.1 *Fix $\epsilon_0 \in (0, 1/2)$, $A > 0$, and $\gamma \in (0, 1/A)$. Suppose that when n tends to ∞, at least one of the three conditions below holds:*

a. $\lim_{n\to\infty}(\epsilon_n \cdot t_n(\gamma)) = 0$,

b. $P_{H_n}(u > u_0 + \delta_n) = 1$, *where δ_n satisfies (2.17) for some constant $c_0 > 0$.*

c. *(2.18) holds for some parameter $\alpha > 0$.*

Then the estimators $u_0(\varphi_n; t_n(\gamma))$ and $\sigma_0^2(\varphi_n; t_n(\gamma))$ are consistent with respect to the null parameters u_0 and σ_0^2, respectively, uniformly across all densities in $\Lambda_n(\epsilon_0, A)$ that satisfy one or more of the conditions (a), (b), and (c).

We remark that while choosing $\gamma \in (0, 1/A)$ ensures consistency, different choices of γ affect the convergence rate of the estimators. The optimal choice of γ depends on unknown parameters and is hard to set. In Section 4, we investigate how to choose γ with simulated data. In our experience, when A is not very large, it is usually appropriate to choose $\gamma \approx 0.2$.

We also remark that in Theorem 2.1 (as well as Theorems 3.1, 3.2 below), we have assumed independence of the test statistics X_j. When the test statistics X_j are correlated, the bias of the estimators remain the same, but the variance of the estimators may inflate by a factor. On the other hand, if the correlation is relatively weak, the estimators continue to perform well. In Section 4, we investigate an simulation example with block-wise dependence among X_j. The simulation results suggest that the estimators continue to perform well when the block size is small (e.g. ≤ 100). See Section 4 and Figure 3 for the details.

3   Estimating the proportion of non-null effects

The proportion has an identifiability issue that is very similar to that of the null parameters. The issue can also be resolved similarly in GEV and GEM. In [9, 8, 3], we have carefully investigated the problem of estimating the proportion in GEV. Similarly to estimating the null parameters, a Fourier approach was introduced (e.g. [8, 9]). Compared to existing approaches in the literature (e.g. [5, 10, 11, 12, 14, 6]), the Fourier approach was proven to be successful in a much broader setting. Especially, it was shown to be successful without the so-called *purity* condition, a notion introduced in [6]. Later in [3], the approach was shown to also attain the optimal rate of convergence over a wide class of situations.

We now shift our attention to GEM. Despite the encouraging development, the Fourier approach in [8, 9] ceases to perform well in this case. In fact, in this case, it can be shown that none of the aforementioned approaches is uniformly consistent with the proportion. Therefore, it is necessary to develop a new approach.

In this section, we propose a new approach to estimating the proportion by using the generalized Fourier transformation, as a natural extension of the ideas in preceding sections. We discuss two cases separately: the case where the null parameters are known, and the case where the null parameters are unknown. In both cases, we show that under mild conditions, the proposed approach is uniformly consistent with the proportion.

3.1   Known null parameters

Recall that
$$\varphi_0(t) = (1 - \epsilon_n)e^{\omega u_0 t + i\sigma_0^2 t^2/2}.$$

The key observation is that, when the null parameters (u_0, σ_0^2) are known, ϵ_n can be easily solved from $\varphi_0(t)$ by

$$\epsilon_n \equiv 1 - e^{-\omega u_0 t - i\sigma_0^2 t^2/2} \varphi_0(t).$$

Inspired by this, we introduce the functional

$$\epsilon_n(g; t, u_0, \sigma_0^2) = 1 - e^{-\omega u_0 t - i\sigma_0^2 t^2/2} g(t). \tag{3.19}$$

where $g(t)$ is any complex-valued function. Recall that

$$\varphi_n(t) \approx \varphi(t) \approx \varphi_0(t).$$

By the continuity of the functional, we hope that for an appropriately chosen t,

$$\epsilon_n(\varphi_n; t, u_0, \sigma_0^2) \approx \epsilon_n(\varphi_0; t, u_0, \sigma_0^2) \equiv \epsilon_n.$$

We now analyze the variance and the bias of this estimator. As before, let $t_n(\gamma) = \sqrt{\gamma \log n}$. For any $H_n(\cdot, \cdot)$ satisfying $P_{H_n}(\sigma^2 \leq A) = 1$, direct calculations show that

$$\mathrm{Var}(\epsilon_n(\varphi_n; t_n(\gamma), u_0, \sigma_0^2) \leq \frac{1}{n} E[e^{-\sqrt{2} t_n(\gamma)(X_1 - u_0)}] \leq \frac{1}{n}[(1 - \epsilon_n)n^{\sigma_0^2 \gamma} + \epsilon_n n^{A\gamma}],$$

so the standard deviation of the estimator is of the order of $o(\epsilon_n)$ when

$$n^{A\gamma - 1} = o(\epsilon_n^2).$$

At the same time, by elementary calculus, the bias of the estimator equals to

$$\left| E[\epsilon_n(\varphi_n; t_n, u_0, \sigma_0^2)] - \epsilon_n \right| = \epsilon_n \cdot \left| \int e^{\omega(u - u_0)t_n + i(\sigma^2 - \sigma_0^2)t_n^2/2} dH_n(u, \sigma) \right|,$$

which is of the order of $o(\epsilon_n)$ if either of the aforementioned conditions (b) or (c) holds. Combining these gives the following theorem.

Theorem 3.1 *Fix u_0, $A > 0$, $\sigma_0^2 \in (0, A)$, $\gamma \in (0, 1/A)$, and a sequence of positive numbers b_n satisfying $\lim_{n \to \infty} b_n = 0$. Consider a sequence of parameters $\epsilon_n \in (0, 1)$ and a sequence of bivariate distribution $H_n(u, \sigma)$ such that for sufficiently large n, $P_{H_n}(u > u_0, \sigma^2 \leq A) = 1$ and $\epsilon_n^{-2} n^{A\gamma - 1} \leq b_n$. Also, suppose that when n tends to ∞, at least one of the two conditions below holds:*

b. $P_{H_n}(u > u_0 + \delta_n) = 1$, *where δ_n satisfies (2.17) for some constant $c_0 > 1$.*

c. *(2.18) holds for some parameter $\alpha > 0$.*

Then as $n \to \infty$, except for a probability that tends to zero,

$$\left| \frac{\epsilon_n(\varphi_n; t_n(\gamma), u_0, \sigma_0^2)}{\epsilon_n} - 1 \right| \to 0,$$

uniformly for all ϵ_n and $H_n(\cdot, \cdot)$ satisfying the conditions above.

In other words, $\epsilon_n(\varphi_n; t_n(\gamma))$ is uniformly consistent with ϵ_n provided that either (b) or (c) holds, and that the variance of the estimator is of a smaller order than that of ϵ_n^2. The latter is satisfied when ϵ_n tends to 0 slowly enough.

3.2 Unknown null parameters

When the null parameters are unknown, a natural approach is to estimate the null parameters using the approach in Section 2 first, then plug in the estimated values to estimate the proportion. In other words, we first estimate the null parameters by

$$\hat{u}_0(\gamma) = u_0(\varphi_n; t_n(\gamma)), \qquad \hat{\sigma}_0^2(\gamma) = \sigma_0^2(\varphi_n; t_n(\gamma)). \tag{3.20}$$

We then estimate the proportion by the plugging estimator,

$$\epsilon_n(\varphi_n; t_n(\gamma), \hat{u}_0(\gamma), \hat{\sigma}_0^2(\gamma)) = 1 - e^{-\omega \hat{u}_0(\gamma) t_n(\gamma) - i \hat{\sigma}_0^2(\gamma) t_n^2(\gamma)/2} \varphi_n(t_n(\gamma)).$$

Note that the bias of both $\hat{u}_0(\gamma)$ and $\hat{\sigma}_0^2(\gamma)$ are typically of the order of $o(\epsilon_n)$, and their variance are of the same order as that of $\epsilon_n(\varphi_n; \gamma)$. Therefore, replacing (u_0, σ_0^2) by $(\hat{u}_0(\gamma), \hat{\sigma}_0^2(\gamma))$ does not increase either the bias or variability of the estimator. The following theorem is proved in Section 5.

Theorem 3.2 *Fix u_0, $A > 0$, $\sigma_0^2 \in (0, A)$, $\gamma \in (0, 1/A)$, and a sequence of positive numbers b_n satisfying $\lim_{n\to\infty} b_n = 0$. Consider a sequence of parameters $\epsilon_n \in (0, 1)$ and a sequence of bivariate distribution $H_n(u, \sigma)$ such that for sufficiently large n, $P_{H_n}(u > u_0, \sigma^2 \le A) = 1$ and $\epsilon_n^{-2} n^{A\gamma - 1} \le b_n$. Also, suppose that when n tends to ∞, at least one of the two conditions below holds:*

b. $P_{H_n}(u > u_0 + \delta_n) = 1$, *where δ_n satisfies (2.17) for some constant $c_0 > 1$.*

c. *(2.18) holds for some parameter $\alpha > 0$.*

Then as $n \to \infty$, except for a probability that tends to zero,

$$\left| \frac{\epsilon_n(\varphi_n; t_n(\gamma), \hat{u}_0(\gamma), \hat{\sigma}_0^2(\gamma))}{\epsilon_n} - 1 \right| \to 0,$$

uniformly for all ϵ_n and $H_n(\cdot, \cdot)$ satisfying the conditions above.

4   Simulations

In this section, we conduct simulation studies for investigating the performance of the proposed estimators of $(u_0, \sigma_0^2, \epsilon)$ with a finite n. We write for short

$$\hat{u}_0(\gamma) = u_0(\varphi_n; t_n(\gamma)), \qquad \hat{\sigma}_0^2(\gamma) = \sigma_0^2(\varphi_n; t_n(\gamma)), \qquad \hat{\epsilon}_n(\gamma) = \epsilon_n(\varphi_n; t_n(\gamma), \hat{u}_0(\gamma), \hat{\sigma}_0^2(\gamma)).$$

Specifically, we are interested in four aspects: (1) how different choices of γ affect the estimation errors of $\hat{u}_0(\gamma)$, $\hat{\sigma}_0^2(\gamma)$ and $\hat{\epsilon}_n(\gamma)$; and what γ values we should recommend in practice; (2) the effect of different choices of the proportion ϵ and the mixing distribution $H_n(\cdot, \cdot)$; (3) the effect of larger n; and (4) the effect of dependent structures.

Example 1. *Different choices of the tuning parameter γ.* In this example, we let $n = 50,000$, $(u_0, \sigma_0^2) = (-1, 1)$, and $\epsilon = 0.025 \times (1, 2, 3, 4, 8)$. We choose 20 different γ ranging from 0.01 to 0.5 with equal inter-distances. For each combination of (ϵ, γ), we conduct an experiment with the following four steps.

- *Step 1.* For each $1 \le j \le n(1 - \epsilon)$, draw $X_j \sim N(u_0, \sigma_0^2)$ to represent a null effect.

- *Step 2.* For each $n(1 - \epsilon) + 1 \le j \le n$, draw independently a sample $u \sim \text{Uniform}(1, 2)$ and a sample $\sigma \sim \text{Uniform}(0.5, 1.5)$. Then, draw $X_j \sim N(u, \sigma^2)$ to represent a non-null effect.

- *Step 3.* Calculate $\hat{u}_0(\gamma), \hat{\sigma}_0^2(\gamma)$, and $\hat{\epsilon}_n(\gamma)$.

- *Step 4.* Repeat Steps 1–3 for 100 times.

The results are reported in Figure 1, from which we can see that the MSE are the smallest when $\gamma \in (0.15, 0.25)$. Also, the MSE are not sensitive to different choices of γ: they remain about the same for different $\gamma \in (0.15, 0.25)$. All of the three estimators $\hat{u}_0(\gamma), \hat{\sigma}_0^2(\gamma)$, and $\hat{\epsilon}_n(\gamma)$ have satisfactory performances: when $\gamma = 0.2$, the MSE are as small as the order of 10^{-4}. Somewhat surprisingly, in this example, different ϵ do not have a prominent effect on the MSE.

Example 2. *The effect of different mixing distribution $H_n(\cdot, \cdot)$.* In this example, we set $n = 50,000$, $(u_0, \sigma_0^2, \epsilon) = (-1, 1, 0.05)$, and choose 20 different γ ranging from 0.01 to 0.5 with equal inter-distances. Compared to Example 1, we conduct experiments with different choices of the
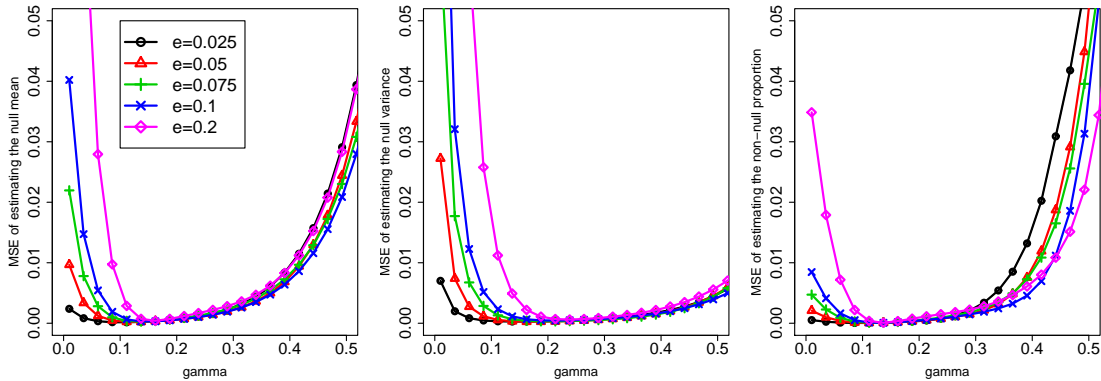
11

Figure 1: MSE for $\hat{u}_0(\gamma)$ (**left**), $\hat{\sigma}_0^2(\gamma)$ (**middle**) and $\hat{\epsilon}_n(\gamma)$ (**right**) for 100 repetitions. The x-axis displays γ, and the y-axis displays the MSE. The null parameters $(u_0, \sigma_0^2) = (-1, 1)$. Different colors of the curves represent different values of ϵ.

Table 1: MSE for $\hat{u}_0(\gamma)$, $\hat{\sigma}_0^2(\gamma)$, and $\hat{\epsilon}_n(\gamma)$ for different n, where we take $\gamma = 0.2$. The parameters $(u_0, \sigma_0^2, \epsilon) = (-1, 1, 0.05)$. For the mixing distribution $H_n(u, \sigma)$, $u \sim \text{Uniform}(1, 2)$ and $\sigma \sim \text{Uniform}(0.5, 1.5)$ independently. In each cell, the MSE equals the cell value times 10^{-4}.

n	10^4	3×10^4	5×10^4	8×10^4	10^5
MSE for $\hat{u}_0(\gamma)$	41.28	10.46	5.66	4.01	2.73
MSE for $\hat{\sigma}_0^2(\gamma)$	16.47	6.93	2.36	1.81	1.48
MSE for $\hat{\epsilon}_n(\gamma)$	20.13	5.28	4.17	2.87	2.01

mixing distribution $H_n(\cdot, \cdot)$. We consider two scenarios. In the first scenario, independently, $(u - u_0) \sim \text{Gamma}(10, 0.25)$ ($\text{Gamma}(k, \theta)$ is the Gamma distribution with shape parameter k and scale parameter θ), and $\sigma \sim \text{Uniform}(0.5, 1.5)$. The parameters $(10, 0.25)$ are chosen such that the mean value of the random variable u is 1.5, the same as that in the preceding example. In the second scenario, independently, $u \sim \text{Uniform}(1, 2)$ and $\sigma \sim \text{Gamma}(10, 0.1)$.

For each scenario and each γ, we run experiments following Steps 1–4 as in Example 1, but with the current choice of $H_n(\cdot, \cdot)$. The MSE for $\hat{u}_0(\gamma)$, $\hat{\sigma}_0^2(\gamma)$ and $\hat{\epsilon}_n(\gamma)$ are reported in Figure 2. From this figure, a similar conclusion can be drawn: the estimators perform well in both scenarios, with the MSE as small as 10^{-4}–10^{-3}. The best range of γ is $(0.15, 0.2)$. In this range, the MSE is relatively insensitive to different choices of γ.

It is noteworthy that in the first scenario, the support of random variable u is not bounded away from the null parameter u_0. It is also noteworthy that in the second scenario, σ is unbounded such that the assumption (2.14) is violated. Despite the seeming challenges in these two scenarios, the proposed approach continues to perform well. This suggests that the proposed approaches are successful in a broader situations than that considered in Sections 2 and 3.

Example 3. _The effect of larger n._ In this example, we fix $(u_0, \sigma_0^2, \epsilon) = (-1, 1, 0.05)$. Since the MSE is relatively insensitive to different choices of γ, we fix $\gamma = 0.2$. For the mixing distribution $H_n(\cdot, \cdot)$, we let $u \sim \text{Uniform}(1, 2)$ and $\sigma \sim \text{Uniform}(0.5, 1.5)$, independently of each other. According to the asymptotic analysis in preceding sections, we understand that the performance of proposed estimators improves when n increases. In this example, we validate this point by choosing $n = 10^4 \times (1, 3, 5, 8, 10)$. For each n, we run experiments following Steps 1–4 as in Example 1. The results are summarized in Table 1. The MSE of all $\hat{u}_0(\gamma)$, $\hat{\sigma}_0^2(\gamma)$ and $\hat{\epsilon}_n(\gamma)$ decreases as n increases. This fits well with the asymptotic analysis in Sections 2 and 3.
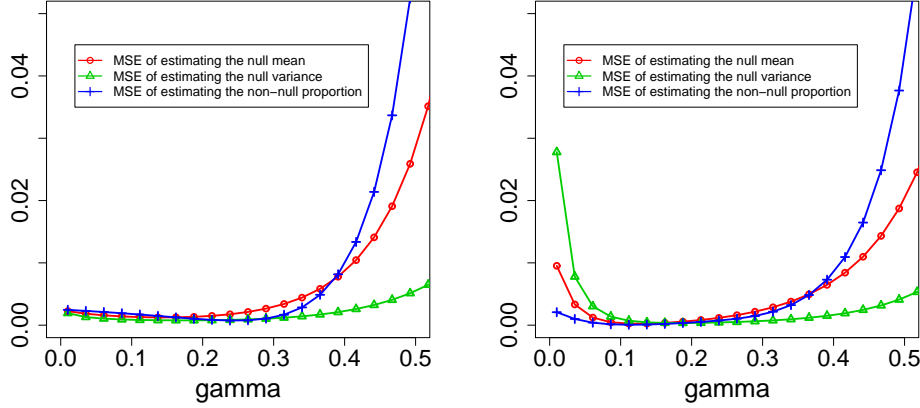
Figure 2: MSE for $\hat{u}_0(\gamma)$ (*red*), $\hat{\sigma}_0^2(\gamma)$ (*green*) and $\hat{\epsilon}_n(\gamma)$ (*blue*) for Scenario 1 (**left**) and Scenario 2 (**right**) considered in Example 2. The x-axis displays γ and the y-axis displays the MSE. $n = 50,000$ and $(u_0, \sigma_0^2, \epsilon) = (-1, 1, 0.05)$.

Example 4. *The effect of dependence.* In this example, we fix $n = 50,000$, $(u_0, \sigma_0^2, \epsilon) = (-1, 1, 0.05)$. We investigate how the dependent structures may affect the performance of the proposed procedures. For each L ranging from 1 to 250 with an increments of 10, we generate samples as follows.

1. For each $1 \leq j \leq n(1 - \epsilon)$, set $(\mu_j, \sigma_j) = (u_0, \sigma_0)$.

2. For each $n(1 - \epsilon) + 1 \leq j \leq n$, draw $\mu_j \sim \text{Uniform}(1, 2)$ and $\sigma_j \sim \text{Uniform}(0.5, 1.5)$.

3. Draw $w_1, \cdots, w_{n+L}$ independently from $N(0, 1)$. For $1 \leq j \leq n$, let $z_j = \sum_{k=j}^{k=j+L} \frac{w_k}{\sqrt{L+1}}$. Note that marginally $z_j \sim N(0, 1)$.

4. For $1 \leq j \leq n$, let $X_j = \mu_j + \sigma_j \cdot z_j$.

The data generated in this way is block-wise dependent, with the block size being controlled by L. Fix $\gamma = 0.2$. We calculate $\hat{u}_0(\gamma)$, $\hat{\sigma}_0^2(\gamma)$, and $\hat{\epsilon}_n(\gamma)$, and repeat the experiment for 100 times. We then calculate the MSE. The results are summarize in Figure 3. While the MSE increase with the block size L, we also note that the MSE remain small when, say, $L \leq 50$ (all three curves fall below 0.02). This suggests that the proposed methods are relatively robust for short-range dependence.

5 Proofs

5.1 Proof of Lemma 1.1 and 1.2

We prove Lemma 1.1 first. Consider two density functions $f_k(x) = f_k(x; u_0^{(k)}, (\sigma_0^2)^{(k)}, \epsilon^{(k)}, H^{(k)})$ that satisfy (1.2)-(1.3), $k = 1, 2$. For short, denote $(u_k, \sigma_k^2, \epsilon_k, H_k) = (u_0^{(k)}, (\sigma_0^2)^{(k)}, \epsilon^{(k)}, H^{(k)})$. Suppose $f_1 = f_2$. We want to show that $(u_1, \sigma_1, \epsilon_1) = (u_2, \sigma_2, \epsilon_2)$. Note that the Fourier transformation of f_1 and f_2 must be identical. By direct calculations, with $s_i(t)$ as defined in (1.6),

$$(1 - \epsilon_1)e^{itu_1 - \frac{\sigma_1^2 t^2}{2}}(1 + s_1(t)) = (1 - \epsilon_2)e^{itu_2 - \frac{\sigma_2^2 t^2}{2}}(1 + s_2(t)). \tag{5.21}$$

We first show $\sigma_1 = \sigma_2$. By (5.21),

$$e^{-\frac{(\sigma_2^2 - \sigma_1^2)t^2}{2}} = \left| e^{it(u_1 - u_2)} \frac{(1 - \epsilon_1)(1 + s_1(t))}{(1 - \epsilon_2)(1 + s_2(t))} \right|. \tag{5.22}$$
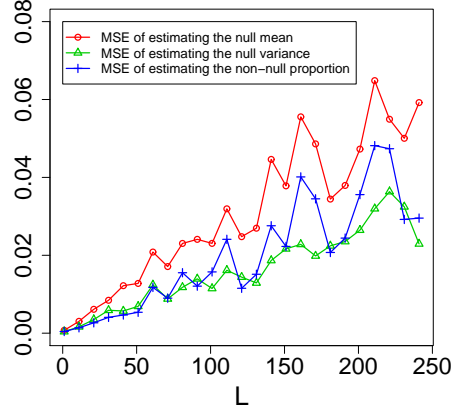
13

Figure 3: MSE for $\hat{u}_0(\gamma)$ (*red*), $\hat{\sigma}_0^2(\gamma)$ (*green*) and $\hat{\epsilon}_n(\gamma)$ (*blue*) when the data X_j are block-wise dependent with a block size L (displayed in the x-axis; see Example 4 for the details). The parameters $(u_0, \sigma_0^2, \epsilon) = (-1, 1, 0.05)$. For the mixing distribution $H_n(u, \sigma)$, $u \sim \mathrm{Uniform}(1, 2)$ and $\sigma \sim \mathrm{Uniform}(0.5, 1.5)$ independently.

Note that $|s_k(t)| \leq \epsilon_k/(1 - \epsilon_k) < 1$, where $|\epsilon_k| \leq \epsilon_0 < 1/2$. Therefore, the right hand side of (5.22) is bounded away from both 0 and ∞ by a constant. Letting t tend to ∞ implies that $\sigma_1 = \sigma_2$.

Next, we show $(u_1, \epsilon_1) = (u_2, \epsilon_2)$. By (5.21) and $\sigma_1 = \sigma_2$,

$$(1 - \epsilon_1)(1 + s_1(t)) = (1 - \epsilon_2)e^{it(u_2 - u_1)}(1 + s_2(t)). \tag{5.23}$$

Fix a small positive number $a > 0$, let $\phi_a(t)$ be the density of $N(0, 1/a)$. Times $\phi_a(t)$ to both sides of (5.23) and integrate in terms of t. By direct calculations and Fubini's theorem, the left hand side of (5.23) is

$$(1 - \epsilon_1) + \epsilon_1 \int \frac{a}{\sqrt{\sigma^2 - \sigma_1^2 + a^2}} \exp\left(-\frac{(u - u_1)^2}{2(\sigma^2 - \sigma_1^2 + a^2)}\right) dH_1(u, \sigma), \tag{5.24}$$

and the right hand side of (5.23) is

$$(1 - \epsilon_2)e^{-\frac{(u_2 - u_1)^2}{2a^2}} + \epsilon_2 \int \frac{a}{\sqrt{\sigma^2 - \sigma_2^2 + a^2}} \exp\left(-\frac{(u - u_1)^2}{2(\sigma^2 - \sigma_2^2 + a^2)}\right) dH_2(u, \sigma). \tag{5.25}$$

Note that by Dominant Convergence Theorem (DCT), any fixed $H(\cdot, \cdot)$ satisfying $P_H(\sigma \geq \sigma_0) = 1$ and $P_H\big((u, \sigma) = (u_0, \sigma_0)\big) = 1$,

$$\lim_{a \to 0} \int \frac{a}{\sqrt{\sigma^2 - \sigma_0^2 + a^2}} \exp\left(-\frac{(u - u_0)^2}{2(\sigma^2 - \sigma_0^2 + a^2)}\right) dH(u, \sigma) = 0. \tag{5.26}$$

Combining (5.24)–(5.26) gives $(1 - \epsilon_1) = \lim_{a \to \infty}[(1 - \epsilon_2)\exp(-\frac{(u_2 - u_1)^2}{2a^2})]$, which immediately implies $(u_1, \epsilon_1) = (u_2, \epsilon_2)$. This proves Lemma 1.1.

Consider Lemma 1.2. The difference is now that both f_i satisfy (1.2)-(1.4). Similarly, suppose $f_1 \equiv f_2$. We want to show that $(u_1, \sigma_1, \epsilon_1) = (u_2, \sigma_2, \epsilon_2)$. By direct calculations, the generalized Fourier transform of f_k are $(1 - \epsilon_k)e^{\omega u_k t + i\sigma_k^2 t^2}[1 + r_k(t)]$, $k = 1, 2$. It follows that

$$e^{\omega t(u_1 - u_2)} \cdot e^{i(\sigma_1^2 - \sigma_2^2)t^2} = \left(\frac{1 - \epsilon_2}{1 - \epsilon_1}\right)\left(\frac{1 + r_2(t)}{1 + r_1(t)}\right). \tag{5.27}$$

Let $t \to \infty$ on both sides. By the condition of $P_{H_k}(u > u_k) = 1$, $r_k(t) \to 0$. Comparing the modules of both sides gives $u_1 = u_2$ and $\epsilon_1 = \epsilon_2$. Combining this with (5.27), $e^{i(\sigma_1^2 - \sigma_2^2)t^2} = \frac{1 + r_2(t)}{1 + r_1(t)}$. Letting $t \to \infty$, the right hand side tends to 1. Therefore, $\sigma_1 = \sigma_2$. $\square$

14

5.2 Proof of Lemma 2.1

It is sufficient to show that for any $t > 0$ and $f \in \Lambda_n(\epsilon_0, A)$,

$$\text{Var}(\varphi_n(t)) \leq \frac{1}{n} e^{-\sqrt{2}u_0 t + \sigma_0^2 t^2} [(1 - \epsilon_n) + \epsilon_n e^{(A - \sigma_0^2)t^2}], \tag{5.28}$$

and

$$\text{Var}(\varphi_n'(t)) \lesssim \frac{1}{n} e^{-\sqrt{2}u_0 t + \sigma_0^2 t^2} [(1 - \epsilon_n)(\sigma_0^2 + 2u_0^2 + 4\sigma_0^4 t^2) + \epsilon_n (A + 2u_0^2 + 4A^2 t^2) e^{(A - \sigma_0^2)t^2}]. \tag{5.29}$$

In fact, once these are proved, the claim follows from (2.12)-(2.14) by taking $t = t_n(\gamma)$.

Consider (5.28). Direct calculations show that

$$\text{Var}(\varphi_n(t)) \leq \frac{1}{n} E[|e^{\omega t X_j}|^2] \leq \frac{1}{n} E[e^{-\sqrt{2}t X_j}].$$

Direct calculations show that

$$E[e^{-\sqrt{2}t X_j}] = e^{-\sqrt{2}t u_0 + \sigma_0^2 t^2} [(1 - \epsilon_n) + \epsilon_n \int e^{-\sqrt{2}(u - u_0)t + (\sigma^2 - \sigma_0^2)t^2} dH_n(u, \sigma)].$$

The claim follows by $u_0 > -A$, $\sigma_0^2 \leq A$, and $P_{H_n}(u > u_0, \sigma_0^2 \leq A) = 1$.

Consider (5.29). Similarly,

$$\text{Var}(\varphi_n'(t)) \leq \frac{1}{n} E[|\omega X_j e^{\omega t X_j}|^2] \leq \frac{1}{n} E[X_j^2 e^{-\sqrt{2}t X_j}].$$

By direct calculations,

$$E[X_j^2 e^{-\sqrt{2}t X_j}] = I + II, \tag{5.30}$$

where

$$I = (1 - \epsilon_n)[\sigma_0^2 + (-u_0 + \sqrt{2}\sigma_0^2 t)^2] e^{-\sqrt{2}u_0 t + \sigma_0^2 t^2},$$

and

$$II = \epsilon_n \int [\sigma^2 + (-u + \sqrt{2}\sigma^2 t)^2] e^{-\sqrt{2}ut + \sigma^2 t^2} dH_n(u, \sigma).$$

By Schwartz inequality,

$$(-u_0 + \sqrt{2}\sigma_0^2 t)^2 \leq 2(u_0^2 + 2\sigma_0^4 t^2),$$

$$(-u + \sqrt{2}\sigma^2 t^2)^2 = (-u_0 - (u - u_0) + \sqrt{2}\sigma^2 t^2)^2 \leq 2(u_0^2 + 2\sigma^2 t^2 + (u - u_0)^2).$$

So

$$I \leq (1 - \epsilon_n)[\sigma_0^2 + 2u_0^2 + 4\sigma_0^4 t^2] e^{-\sqrt{2}u_0 t + \sigma_0^2 t^2}, \tag{5.31}$$

and

$$II \leq \epsilon_n [\int (\sigma^2 + 2u_0^2 + 4\sigma^4 t^2) e^{-\sqrt{2}ut + \sigma^2 t^2} dH_n(u, \sigma) + 2e^{-\sqrt{2}u_0 t} \int (u - u_0)^2 e^{-\sqrt{2}(u - u_0)t + \sigma^2 t^2} dH_n(u, \sigma)]. \tag{5.32}$$

Note that $\sup_{x>0} 2x^2 e^{-\sqrt{2}tx} = 2/(et^2)$. It follows

$$\int (u - u_0)^2 e^{-\sqrt{2}(u - u_0)t + \sigma^2 t^2} dH_n(u, \sigma)] \leq [2/(et^2)] \int e^{\sigma^2 t^2} dH_n(u, \sigma). \tag{5.33}$$

Inserting (5.33) into (5.32) and recalling that $P_{H_n}(u > u_0, \sigma^2 \leq A) = 1$,

$$II \leq \epsilon_n [A + 2u_0^2 + 4A^2 t^2 + \frac{2}{et^2}] e^{-\sqrt{2}u_0 t + At^2}. \tag{5.34}$$

Inserting (5.31) and (5.34) into (5.30) gives the claim. □

15

5.3 Proof of Lemma 2.2

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

$$u_0(g,t) - u_0(f,t) = \frac{\frac{d}{dt}|f|}{|f|} - \frac{\frac{d}{dt}|g|}{|g|} = I + II,$$

where $I = (1 - \frac{|g|^2}{|f|^2}) \cdot u_0(g,t)$, $II = \frac{1}{|f|^2} \cdot [\text{Re}(g') \cdot \text{Re}(f-g) + \text{Im}(g') \cdot \text{Im}(f-g) + \text{Re}(f) \cdot \text{Re}((f-g)') + \text{Im}(f) \cdot \text{Im}((f-g)')]$. Now, firstly, using triangle inequality,

$$|I| \leq \frac{|u_0(g,t)|}{|f|^2} \cdot \left| |f|^2 - |g|^2 \right| \leq \frac{|u_0(g,t)|}{|f|^2} ((|f| + |g|)|f-g|);$$

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

$$|II| \leq \frac{\sqrt{2}}{|f|^2} \cdot [|g'| \cdot |f-g| + |f| \cdot |(f-g)'|].$$

Combining these gives

$$|u_0(g,t) - u_0(f,t)| \leq \frac{1}{|f|^2} \left[(|u_0(g,t)|(|f| + |g|) + \sqrt{2}|g'|) \cdot |f-g| + \sqrt{2}|g| \cdot |(f-g)'| \right].$$

Consider the second claim. By direct calculations,

$$\sigma_0^2(g,t) - \sigma_0^2(f,t) = I + II,$$

where $I = (1 - \frac{|g|^2}{|f|^2}) \cdot u_0(g,t)$, $II = \frac{\sqrt{2}}{t|f|^2} \cdot [(\text{Re}(\omega \bar{g} g') - \text{Re}(\omega \bar{f} f')]$. Similarly,

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

$$|II| \leq \frac{\sqrt{2}}{t|f|^2} \cdot [|g'| \cdot |f-g| + |f| \cdot |(f-g)'|].$$

Combining these gives

$$|\sigma_0^2(g,t) - \sigma_0^2(f,t)| \leq \frac{1}{t|f|^2} \cdot \left[(|\sigma_0^2(g,t)| \cdot t \cdot (|f| + |g|) + \sqrt{2}|g'|) \cdot |f-g| + \sqrt{2}|f| \cdot |(f-g)'| \right].$$

$\square$

5.4 Proof of Lemma 2.3

Write $t_n = t_n(\gamma)$ for short. Introduce the event

$$A_n = \{\max\{|\varphi_n(t_n) - \varphi(t_n)|, \ |\varphi_n'(t_n) - \varphi'(t_n)|\} \leq \log^{3/2}(n)\}.$$

Applying Lemma 2.1, $P(A_n^c) \to 0$, uniformly for all $f \in \Lambda_n(\epsilon_0, A)$. To show the claim, it is sufficient to show that the inequalities hold over the event A_n. Since the proofs are similar, we only prove the first one.

We claim that (a). $1/\varphi(t_n)| \leq \bar{o}(1)$ over event A_n, (b). $|\varphi_n(t_n)| \sim |\varphi(t_n)|$ over event A_n, and (c). $|u_0(\varphi; t_n)| \leq \bar{o}(1)$. Consider (a) and (b). By $\epsilon_n \leq \epsilon_0 < 1/2$ and elementary calculus, $|r(t)| \leq \epsilon_n/(1-\epsilon_n) \leq \epsilon_0/(1-\epsilon_0)$. The claim follows from

$$|\varphi(t)| \geq (1 - |r(t)|)|\varphi_0(t)| \geq \frac{1 - 2\epsilon_0}{1 - \epsilon_0} e^{-u_0 t_n/\sqrt{2}}.$$

By the definition of A_n,

$$|\varphi_n(t_n) - \varphi(t_n)| \leq \bar{o}(n^{(A\gamma - 1)/2}),$$

and the claims follow. Consider (c). By Lemma 2.4, $|u_0(\varphi_n; t)| \leq |u_0| + |r'(t_n)|$. Write

$$r'(t) = \frac{\epsilon_n}{1 - \epsilon_n} \int [\omega(u - u_0) + i(\sigma^2 - \sigma_0^2)t] e^{\omega(u - u_0)t + i(\sigma^2 - \sigma_0^2)t^2/2} dH_n(u, \sigma).$$

Since that $\sup_{x>0}\{xe^{-x}\} = 1/e$ and $P_{H_n}(u > u_0, \sigma^2 \leq A) = 1$, the claim follows from

$$|r'(t_n)| \leq \int [|u - u_0| e^{-(u - u_0)t_n/\sqrt{2}} + |\sigma^2 - \sigma_0^2| t_n] dH_n(u, \sigma) \leq (\sqrt{2}/(et_n)) + At_n.$$

Finally, combine (a)-(c) with Lemma 2.2,

$$|u_0(\varphi_n; t_n) - u_0(\varphi; t_n)| \leq \bar{o}(1) \cdot [|\varphi_n(t_n) - \varphi(t_n)| + |\varphi_n'(t_n) - \varphi'(t_n)|],$$

and the claim follows. $\qquad\square$

5.5 Proof of Lemma 2.4

For simplicity, drop t from $\varphi(t)$, $\varphi_0(t)$, and $r(t)$ whenever there is no confusion. Consider the first claim. Recalling that $|\varphi| = |\varphi_0| \times |1 + r|$,

$$\frac{d}{dt}|\varphi(t)| = (\frac{d}{dt}|\varphi_0|) \cdot |1 + r| + |\varphi_0| \cdot \frac{d}{dt}|1 + r|.$$

Using the definition of $u_0(\varphi; t)$ and Lemma 1.3, it follows from direct calculations that

$$|u_0(\varphi; t) - u_0| = \frac{\sqrt{2}}{|1 + r(t)|} \frac{d}{dt}|1 + r(t)|. \tag{5.35}$$

Moreover,

$$\frac{d}{dt}|1 + r(t)| = \frac{r'(t)(1 + \bar{r}(t)) + (1 + r(t))\bar{r}'(t)}{2|1 + r(t)|}. \tag{5.36}$$

By that $P_{H_n}(u > u_0) = 1$,

$$|r(t)| \leq \frac{\epsilon_n}{1 - \epsilon_n} |\int e^{-(u - u_0)t/\sqrt{2} + i(-(u - u_0)t/\sqrt{2} + (\sigma^2 - \sigma_0^2)t^2/2} dH_n(u, \sigma)| \leq \frac{\epsilon_n}{1 - \epsilon_n}. \tag{5.37}$$

Combining(5.35)–(5.37) gives the claim.

Consider the second claim. Write

$$\varphi' = \varphi_0'(1 + r) + \varphi_0 r'.$$

We have $\bar{\varphi}\varphi' = |1 + r|^2 \bar{\varphi}_0 \varphi_0' + |\varphi_0|^2(1 + \bar{r})r'$, and so

$$\mathrm{Re}(\omega\bar{\varphi}\varphi) = |1 + r|^2 \mathrm{Re}(\omega\bar{\varphi}_0\varphi_0) + |\varphi_0|^2 \mathrm{Re}(\omega(1 + \bar{r})r').$$

Therefore,

$$|\sigma_0^2(\phi; t) - \sigma_0^2| \leq \frac{|\mathrm{Re}(\omega(1 + \bar{r}(t))r')|}{t|1 + r(t)|^2} \leq C(\epsilon_0)|r'(t)|/t,$$

and the claim follows directly. $\qquad\square$

5.6 Proof of Theorem 3.1

Write for short $t_n = t_n(\gamma)$ and $\epsilon_n(\cdot; t_n) = \epsilon_n(\cdot; t_n, u_0, \sigma_0^2)$. By triangle inequality,

$$|\epsilon_n(\varphi_n; t_n) - \epsilon_n| \leq |\epsilon_n(\varphi_n; t_n) - \epsilon_n(\varphi; t_n)| + |\epsilon_n(\varphi; t_n) - \epsilon_n|.$$

Compare this with the desired claim. It is sufficient to show that $E[|\epsilon_n(\varphi_n; t_n) - \epsilon_n(\varphi; t_n)|^2] \leq n^{A\gamma - 1}$ and $|\epsilon_n(\varphi; t_n)/\epsilon_n - 1|$ tends to 0 in a speed that does not depends on ϵ_n and A_n.

Consider the first term first. By the definition of the functional $\epsilon_n(\cdot; t_n)$ (i.e. (3.19)),

$$|\epsilon_n(\varphi_n; t_n) - \epsilon_n(\varphi; t_n)| \leq |e^{-\omega u_0 t_n - i\sigma_0^2 t_n^2/2}(\varphi_n(t_n) - \varphi(t_n))| \leq e^{\frac{u_0 t_n}{\sqrt{2}}} |\varphi_n(t_n) - \varphi(t_n)|.$$

At the same time, by the definitions of $\varphi_n(\cdot)$ and $\varphi(\cdot)$ and elementary calculus,

$$E|\varphi_n(t_n) - \varphi(t_n)|^2 = \frac{1}{n} \mathrm{Var}(e^{\omega t_n X_1}) \leq \frac{1}{n} E[e^{-\sqrt{2} t_n X_1}].$$

Combining these gives,

$$E(|\epsilon_n(\varphi_n; t_n) - \epsilon_n(\varphi; t_n)|^2) \leq \frac{1}{n} e^{\sqrt{2} t_n u_0} E[e^{-\sqrt{2} t_n X_1}],$$

where by direct calculations and the assumptions of $P_{H_n}(u > u_0, \sigma^2 \leq A) = 1$ and $\sigma_0^2 \leq A$, the last term is no greater than

$$\frac{1}{n} \left\{ (1 - \epsilon_n) e^{t_n^2 \sigma_0^2} + \epsilon_n \int e^{-\sqrt{2} t_n(u - u_0) + t_n^2 \sigma^2} dH_n(u, \sigma) \right\} \leq n^{A\gamma - 1}.$$

Combining these gives the first claim.

Consider the second claim. Recall that $\epsilon_n = \epsilon_n(\varphi_0; t_n)$, that $\varphi(t) = \varphi_0(t)(1 + r(t))$ (see (1.9)), and that $\varphi_0(t) = (1 - \epsilon_n) e^{-\omega u_0 t_n - i\sigma_0^2 t_n^2/2}$. By the definition of the functional $\epsilon_n(\cdot, t_n)$,

$$\epsilon_n(\varphi; t_n) - \epsilon_n = e^{-\omega u_0 t_n - i\sigma_0^2 t_n^2/2}(\varphi(t_n) - \varphi_0(t_n)) = (1 - \epsilon_n) r(t_n).$$

It then follows from the definition of $r(\cdot)$ that

$$|\epsilon_n(\varphi; t_n) - \epsilon_n| \leq |(1 - \epsilon_n) r(t_n)| = \epsilon_n \left| \int e^{\omega(u - u_0) t_n + i(\sigma^2 - \sigma_0^2) t_n^2/2} dH_n(u, \sigma) \right|. \tag{5.38}$$

Suppose condition (b) holds. Then $P_{H_n}(u > u_0 + \delta_n) = 1$, where δ_n satisfies (2.17) with some constant $c_0 > 0$. It follows from (5.38) and elementary calculus that as $n \to \infty$,

$$|\epsilon_n(\varphi; t_n)/\epsilon_n - 1| \leq \int e^{-\frac{(u - u_0)t}{\sqrt{2}}} dH_n(u, \sigma) \to 0. \tag{5.39}$$

Suppose condition (c) holds. Let $\delta = u - u_0, \kappa = \sigma^2 - \sigma_0^2$. By the definitions of $g(\kappa|\delta)$ and $g(\delta)$ and elementary Fourier analysis,

$$\int e^{\omega(u - u_0) t_n + i(\sigma^2 - \sigma_0^2) t_n^2/2} dH_n(u, \sigma) = \int e^{\omega \delta t_n + i\kappa t_n^2/2} g(\kappa|\delta) h(\delta) d\kappa d\delta = \int e^{\omega \delta t_n} g^{FT}(t_n^2/2; \delta) h(\delta) d\delta.$$

By the assumptions, $P_{H_n}(\delta > 0) = 1$ and $g^{FT}(t) \leq C(1 + |t|)^{-\alpha}$, so

$$\left| \int e^{\omega \delta t_n} g^{FT}(t_n^2/2; \delta) h(\delta) d\delta \right| \leq \int e^{-\frac{\delta t_n}{\sqrt{2}}} |g^{FT}(t_n^2/2; \delta)| h(\delta) d\delta \leq C(1 + t_n^2/2)^{-k} \longrightarrow 0,$$

Combining these with (5.38) gives the claim. □

5.7 Proof of Theorem 3.2

Write for short $t_n = t_n(\gamma)$, $\hat{u}_0 = u_0(\varphi_n; t_n)$, and $\hat{\sigma}_0^2 = \sigma_0^2(\varphi_n; t_n)$. By the definitions of $\epsilon_n(\cdot; t, u, \sigma)$,

$$|\epsilon_n(\varphi_n; t_n, \hat{u}_0, \hat{\sigma}_0^2) - \epsilon_n(\varphi_n; t_n, u_0, \sigma_0^2)| \leq |e^{-\omega(\hat{u}_0 - u_0)t_n - i(\hat{\sigma}_0^2 - \sigma_0^2)t_n^2} - 1| \cdot |e^{-\omega u_0 t_n - i\sigma_0^2 t_n^2}\varphi_n(t_n)|,$$

where we note that by the definition of the functional $\epsilon_n(\cdot; t, u, \sigma)$, the last term $\leq 1 + |\epsilon_n(\varphi_n; t_n, u_0, \sigma_0^2)|$. By Lemmas 2.3–2.4, except for a small probability that tends to 0 as n tends to ∞,

$$|\hat{u}_0 - u_0|t_n \leq t_n|r'(t_n)| + \bar{o}(n^{(A\gamma-1)/2}), \qquad |\hat{\sigma}_0^2 - \sigma_0^2|t_n^2 \leq t_n|r'(t_n)| + \bar{o}(n^{A\gamma-1)/2}).$$

At the same time, by Theorem 3.1, except for a small probability that tends to 0 as n tends to ∞,

$$\epsilon_n(\varphi_n; t_n, u_0, \sigma_0^2) \sim \epsilon_n.$$

Combine these, as n tends to ∞, except for a small probability that tends to 0.

$$|\epsilon_n(\varphi_n; t_n, \hat{u}_0, \hat{\sigma}_0^2) - \epsilon_n(\varphi_n; t_n, u_0, \sigma_0^2)| \leq t_n|r'(t_n)|,$$

which, by Lemmas 2.5–2.6, tends to 0. This concludes the proof. $\qquad\square$

References

[1] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B.* **57**, 289-300.

[2] BENJAMINI, Y. and KRIEGER, A. and YEKUTIELI, D (2005). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, **93** (3), 491-507.

[3] CAI, T. and JIN, J. (2009). Optimal rate of convergence for estimating the null parameters and the proportion of non-null effects, *to appear in Ann. Statist.*

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

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

[6] GENOVESE, C. and WASSERMAN, L. (2002). A stochastic process approach to false discovery control. *Ann. Statist.* **32** (3), 1035-1061.

[7] HALL, P. and JIN, J. (2009). Innovated Higher Criticism for detecting sparse signals in correlated noise. *Ann. Statist., to appear.*

[8] JIN, J. (2008). Proportion of nonzero normal means: oracle equivalence and uniformly consistent estimators. *J. Roy. Statist. Soc. B.* **70**(3), 461-493.

[9] JIN, J. and CAI, T. (2006). Estimating the null and the proportion of non-null effects in large-scale multiple comparisons. *J. Amer. Statist. Assoc.*, **102**, 496-506.

[10] LANGAAS, M. and LINDQVIST, B. H. and FERKINGSTAD, E. (2005). Estimating the proportion of true null hypotheses, with applications to DNA microarray data. *J. R. Statist. Soc. B.*, **67**, 555-572.

[11] MEINSHAUSEN, M. and RICE, J. (2004). Estimating the proportion of false null hypothesis among a large number of independent tested hypotheses. *Ann. Statist.*, **34** (1), 373–393.

[12] STOREY, J. D. (2002). A direct approach to false discovery rate. *J. R. Stat. Soc. Ser. B.*, **64**, 479–498.

[13] STOREY, J. D. (2007). The optimal discovery procedure: A new approach to simultaneous significance testing. *J. R. Stat. Soc. Ser. B.*, **69**, 347–368.

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