Power Prediction in Large-scale Multiple Testing: A Fourier Approach

Ph.D Thesis Proposal

Avranil Sarkar
Department of Statistics, Carnegie Mellon University

January 13, 2010

Abstract

A problem that is frequently found in large-scale multiple testing is that, in the present stage of experiment (e.g. gene microarray, functional MRI), the signals are so faint that it is impossible to attain a desired level of testing power, and one has to enroll more samples in the follow-up experiment. Suppose we *are going to* enlarge the sample size by n times in the follow-up experiment, where $n > 1$ is not necessary an integer. A problem of great interest is, given data based on the *current* stage of experiment, how to predict the testing power after the sample size is enlarged by n times.

We consider test z-scores and model the test statistics in the current experiment as $X_j \sim N(\mu_j, 1)$, $1 \leq j \leq p$. We propose a Fourier approach to predicting the testing power after n replicates. The approach produces a very accurate prediction for moderately large values of n ($n \leq 7$). Finally, we discuss potential applications of this method on real data with emphasis on gene microarray data.

Acknowledgement : I would like to thank Peter Huggins for his valuable discussions.

1 Introduction

Until recently, "simultaneous inference" meant considering two or five or perhaps even 10 hypothesis tests at the same time. Rapid progress in technology, particularly in genomics and imaging, has vastly upped the ante for simultaneous inference problems giving rise to large scale multiple testing. Now 500 or 5,000 or even 50,000 tests may need to be evaluated simultaneously, raising new problems for the statistician, but also opening new analytic opportunities.

Simultaneous hypothesis testing begins with a collection of null hypotheses,

$$H_{01}, H_{02}, \ldots H_{0p}$$

corresponding test statistics, possibly not independent,

$$X_1, X_2, \ldots, X_p$$

and their corresponding p-values, with i^{th} p-value measuring how strongly x_i, the observed value of X_i, contradicts H_{0i}; "Large-scale" means that p is a big number.

Consider for example gene microarray data. In this type of data, for a large number of genes for two different groups of people, one is interested in finding which genes are differentially expressed between these two groups. This is equivalent to testing a large number of null hypotheses, one for each gene. The null hypothesis, for a particular gene, corresponds to that gene being not differentially expressed. Given a set of hypotheses to be tested and a set of test statistics, one for each hypothesis, a particular test statistic is said to contain a signal if the corresponding null hypothesis is false.

One of the problems frequently faced in large scale multiple testing is that the signals contained in the test statistics are faint. This can result in accepting a null hypothesis which is false. An obvious way of addressing this problem is to strengthen the signals by increasing the number of samples and perform the experiment again. Going back to the gene expression example, this just means to increase by n times the number of people in the two groups. Then the signal strength increases and so it becomes easier for any test procedure to correctly reject the false null hypotheses. n can be called as *replication multiplicity*. The larger the value of n, the easier it is to distinguish the null from the alternative.

However, collecting more samples is a costly procedure. So it is important to know beforehand from the available sample, what kind of results one can expect by enlarging the sample size. Several quantities are of interest. A few examples are given below.

- *Average power*. When a null hypothesis is rejected, it is a positive. A positive may be a true positive (TP) or a false positive (FP), depending on whether the hypothesis is correctly or incorrectly rejected. The average power of a procedure is the fraction of true positives that it yields. Larger sample size leads to increase in average power and so it is important to consider the prediction of the expected average power obtained from an enlarged sample.

- *False Discovery Rate (FDR)*. FDR is the fraction of false positive out of all positives. Control of the FDR has been widely accepted as a criterion in multiple testing (Benjamini and Hochberg, 1995). The FDR level serves as an important guideline for practitioner. Increase in sample size means decrease in FDR and how to predict it is of great interest.

- *Required replication multiplicity.* A larger sample usually means a larger power and a better control of the FDR. Given a desired average power, or a desired level of FDR, or both, it is of interest to know the minimum sample size or replication multiplicity that is required.

All the quantities of interest described above can be computed using the two key quantities ϵ, the proportion of non-null hypotheses, and PR, the (future) expected positive rate. Recall that given a test procedure, positive rate (PR) is the fraction of null hypotheses rejected. Future expected positive rate refers to the expected fraction of null hypotheses rejected by the test statistic computed on an enlarged sample, had it been available.

The problem of estimating ϵ has been studied in great detail before e.g see Genovese and Wasserman (2004), Cai and Jin (2009), Jin (2008), Jin and Cai (2006, 2007), Meishaussen and Rice (2006). The main focus of this thesis proposal is to study an efficient estimation procedure for the future expected positive rate, PR.

The next section gives a short review of the literature on multiple testing and false discovery rate. In §3, we describe the specific model used for the p-values as well as the Fourier approach for estimating a general functional. §4 defines the positive rate (PR) which is the main quantity of interest. We also propose an estimator for estimating PR and obtain the theoretical rates of convergence for its MSE. §5 describes simulation study and §6 gives a summary of the proposed research.

2 Review

In large scale multiple testing, hypothesis tests that incorrectly reject the null hypothesis are more likely to occur when one considers the family of tests as a whole. When dealing with a large number of hypotheses to be tested simultaneously, especially with a small proportion of hypotheses being false, as is the case in practice, attempts to control the probability of at least one false positive results in a very conservative test procedure. The sensitivity of the tests can be too low (Dudoit *et al.* (2002)).

The most significant development to overcome this kind of problem, is control of the false discovery rate (FDR) proposed by Benjamini & Hochberg (1995). In a list of rejected hypotheses, FDR controls the expected proportion of incorrectly rejected null hypotheses (type I errors). In practical terms, the FDR is the expected false positive rate. It is a less conservative procedure for comparison, with greater power than familywise error rate (FWER) control, at a cost of increasing the likelihood of obtaining type I errors.

Because of its useful interpretation, FDR is a very convenient scale to work on. For example, if we declare a collection of 100 tests with a maximum FDR of 0.05 to be actually true, then we expect a maximum of 5 hypotheses to be false positives. When controlling the FDR, an experimenter also needs to be aware of the sensitivity or false negative rate (FNR), as he/she does not want to lose too many of the true non-null hypotheses by setting the FDR too low. Thus, the increasing use of FDR needs to be accompanied by the sensitivity or FNR assessment. At least four factors determine the FDR characteristics of simulatneous multiple testing:

- the proportion of no-null hypotheses.
- the distribution of the signals.
- measurement variability
- sample size.

Only the latter is under the experimenter's control. Moreover, if the signals are too weak then an increase in sample size will also lead to stronger signals. Among other things, the analysis of FDR allows an assessment of sample size needed for testing. Knowing how many samples are needed has been a problem for many researchers, but no clear recommendation based on the FDR seems to be on offer. The standard sample size calculation from the traditional hypothesis testing framework, based on controlling the false positive rate has been studied before e.g. Dobbin *et al.*(2003), Wang and Chen, (2004) Yang *et al.* (2003), Gadbury *et al.* (2004). However these are not appropiate for FDR control.

Hence it is worthwhile to study the relationship between power of a test based on FDR control and the sample size required to achive a certain power. A method which can estimate the sample size needed in an experiment to achieve a pre-specified power is important.

3 Methodology

In this section, first the hypotheses along with the distribution of the test statistic is described in detail. Next, the quantity of interest, the positive rate (PR) is explained and the method for estimating it is discussed.

3.1 Gaussian Model

Let there be p independent hypotheses to be tested, the null hypothesis $H_{0j} : \mu_j = 0$ vs. the alternative $H_{1j} : \mu_j \neq 0$ for $j = 1, \ldots, p$, where X_j, the test statistic for the j^{th} hypothesis computed from the available data, is modelled as an observation coming from $N(\mu_j, 1)$ independent of $X_{j'}$ for $j \neq j'$. Let ϵ_p be the proportion of false null hypotheses. Let $\mathcal{F}$ be the subset of $\{1, \ldots, p\}$ with $|\mathcal{F}| = p\epsilon_p$ such that if $j \in \mathcal{F}$, then H_{0j} is false for $j = 1, \ldots, p$. Also assume that if H_{0j} is false, then $\mu_j \sim g$ for some denstiy g. Here, ϵ_p, $\mathcal{F}$ and g are unknown. Then,

$$\begin{aligned} X_j &\overset{iid}{\sim} N(0,1) \, \forall \, j \in \mathcal{F}^c \\ X_{j'} &\overset{iid}{\sim} \int \phi(x-u)\, g(u)\, du \, \forall \, j' \in \mathcal{F}^c \\ X_j \text{ is independent of } & X_{j'} \, \forall j \in \mathcal{F}^c, \, j' \in \mathcal{F} \end{aligned} \tag{3.1}$$

where $\phi(\cdot)$ is the density of $N(0, 1)$.

Next, consider the future sample, which is assumed to be n times the size of the current sample with $n > 1$. Then the future sample can be considere as n replications of the current sample. For the j^{th} hypothesis, now there are n *i.i.d.* copies of the current test statistic X_j

each of which are distributed as $N(\mu_j, 1)$ with μ_j described as before for $j = 1, \ldots, p$. The test statistic for the future data is constructed by summing across the n replications and scaling by $\sqrt{n}$. Letting Y_j be the test statistic for the future data for the j^{th} hypothesis, it follows that $Y_j \sim N(\sqrt{n}\mu_j, 1)$ for $j = 1, \ldots, p$. In effect, the marginal density of the future test statistics is

$$Y_j \overset{iid}{\sim} N(0,1)\, \forall\, j \in \mathcal{F}^c \tag{3.2}$$
$$Y_{j'} \overset{iid}{\sim} \int \phi(x - \sqrt{n}u)\, g(u)\, du\, \forall\, j^{'} \in \mathcal{F}^c$$
$$Y_j \text{ is independent of } Y_{j'}\, \forall j\ \in \mathcal{F}^c,\ j^{'}\ \in \mathcal{F}$$

Now, assume that the enlarged sample is available, and consider the testing procedure that rejects the j^{th} null hypothsis H_{0j} if $|Y_j|$ exceeds some fixed threshold $t > 0$ for $j = 1, \ldots, p$. Then the future positive rate is nothing but the fraction of total number of rejected null hypotheses i.e. $\frac{1}{p}\sum_{j=1}^{p} I(|Y_j| > t)$. Then the future expected positive rate of §1 is

$$\begin{aligned} PR = PR(t;n) &= \frac{1}{p}\sum_{j=1}^{p} P(|Y_j| > t) \\ &= 1 - \frac{1}{p}E[\Psi(\mu_j; t, n)] \\ &= 1 - (1-\epsilon_p)\Psi(0; t, n) - \epsilon_p \int \Psi(u; t, n) g(u)\, du. \end{aligned} \tag{3.3}$$

where
$$\Psi(u; t, n) = 1 - [\bar{\Phi}(t - \sqrt{n}\cdot u) + \bar{\Phi}(t + \sqrt{n}\cdot u)], \tag{3.4}$$
with $\bar{\Phi} = 1 - \Phi$ being the survival function of $N(0,1)$ The central problem is then how to estimate $PR(t;n)$ from the sample availabe at the current stage i.e. X_j for $j = 1, \ldots, p$. In the next subsection a Fourier approach is proposed for estimating $PR(t;n)$.

3.2 A Fourier approach for estimation

In this set up, a Fourier approach was proposed for estimating ϵ_p in Jin (2008). Note that, $\epsilon_p = 1 - \frac{1}{p}\sum_{j=1}^{p} E[I(\mu_j = 0)]$. This approach can be generalized to estimate a much broader class of functionals of the form

$$T(h) = \frac{1}{p}\sum_{j=1}^{p} E[h(\mu_j)] \tag{3.5}$$

for some function h. It should be noted, that from (3.3) and (3.4), it follows that PR is of the form of (3.5) with h replaced by Ψ. Now for any general functional T of the form (3.5), the idea is to construct an appropriate function $f(x)$, and estimate $T(g)$ with

$$\hat{T}(X_1, X_2, \ldots, X_p) = \frac{1}{p}\sum_{j=1}^{p} f(X_j).$$

In fact, direct calculations show that

$$E[\hat{T}(X_1, X_2, \ldots, X_p)] = \frac{1}{p}\sum_{j=1}^{p} E[f(X_j)] = \frac{1}{p}\sum_{j=1}^{p} E[(f * \phi)(\mu_j)]$$

where $*$ is the usual convolution. So ideally, the estimator would be unbiased if it were possible to construct an f such that

$$f * \phi \equiv h. \tag{3.6}$$

However, for such an f to exist, in the frequency domain f should satisfy

$$\hat{f} \cdot \hat{\phi} = \hat{h}, \qquad \text{or} \qquad \hat{f}(\xi) = e^{\xi^2/2} \cdot \hat{h}(\xi). \tag{3.7}$$

where $\hat{r}$ denotes the Fourier transform of r and ϕ is the standard normal density, $\hat{\phi}(\xi) = e^{\xi^2/2}$. Generally, the function $(\hat{f}(\xi) = e^{\xi^2/2} \cdot \hat{h}(\xi))$ is not integrable and hence such an f does not exist. This is the case of PR with $h = \Psi$ and also of ϵ_p with $h(u) = 1_{\{u=0\}}$.

To overcome this difficulty i.e. to construct an f such that $\hat{f}$ is integrable and f approximately satistfies (3.7), a symmetric continuous function $\omega(\xi)$ is chosen, which will be referred to as a *kernel*, so that the function $\omega(\xi) \cdot e^{\xi^2/2} \cdot \hat{h}(\xi)$ is integrable. Then $\hat{f}(\xi)$ in (3.7) is replaced by

$$\hat{f}(\xi; \omega) = \omega(\xi) \cdot e^{\xi^2/2} \cdot \hat{h}(\xi). \tag{3.8}$$

By symmetry and inverse Fourier transformation, the unique f that satisfies (3.8) is

$$f(x; \omega) = \int \omega(\xi) \cdot e^{\xi^2/2} \cdot \hat{h}(\xi) \cos(\xi x)\, d\xi. \tag{3.9}$$

Note that a desirable kernel ω should be such that ω has sufficiently thin tail i.e. $\omega(\xi) \approx 0$ for large values of ξ in order to make the existence of f possible, but at the same time for small values of ξ, $\omega(\xi) \approx 1$ so that,

$$f(\cdot; \omega) * \phi \approx h. \tag{3.10}$$

In the literature, it is frequently seen that tampering a function significantly in the frequency domain may only result in a change that is uniformly small in the spatial domain. In this case, the ideal $f(\cdot)$ as described in (3.6) cannot be constructed. The function $f(\cdot, \omega)$ in (3.9) is an approximate version of $f(\cdot)$ where the approximation is done in the Fourier domain. The difference of $f(\cdot)$ and $f(\cdot, \omega)$ is uniformly small although $\hat{f}(\cdot)$ and $\hat{f}(\cdot, \omega)$ are significantly different.

Having constructed $f(\cdot, \omega)$ in (3.9) , the functional $T(h)$ can be estimated with

$$\hat{T}(X_1, X_2, \ldots, X_p; \omega) = \frac{1}{p}\sum_{j=1}^{p} f(X_j;\ \omega). \tag{3.11}$$

In the next section, an estimator for $PR(t; n)$ is constructed using the approach described above and its asymptotic properties are studied in detail. In the process, the optimal kernel for estimating $PR(t; n)$ is also derived along with the rate of mean squared error for the estimator of $PR(t; n)$ using the optimal kernel.

4 Estimation of PR

In this section we construct an estimator for PR and obtain its rate of convergence. For constructing an estimator of PR, we apply the general framework of Fourier approach as introduced in Section 3.2 . From (3.3), it follows that

$$1 - PR = \frac{1}{p} E[\Psi(\mu_j)]$$

Now, PR is of the form $T(h)$ as in (3.5) with $h(\cdot) = \Psi(\cdot)$. From (3.9) and (3.11) it follows that for any kernel ω, an estimator $\widehat{PR}(\omega)$ of PR can be constructed of the form

$$\widehat{PR}(\omega) = 1 - \frac{1}{p} \sum_{j=1}^{p} f(X_j; \omega) \tag{4.1}$$

where

$$f(x; \omega) = \frac{1}{2\pi} \cdot \int \omega(\xi) \cdot e^{\xi^2/2} \cdot \hat{\Psi}(\xi) \cos(x\xi)\, d\xi, \tag{4.2}$$

where $\hat{\Psi}$ denotes the Fourier transform of Ψ with

$$\hat{\Psi}(\xi) = \frac{2t}{\sqrt{n}} \cdot e^{-\xi^2/(2n)} \cdot \frac{\sin(t\xi/\sqrt{n})}{t\xi/\sqrt{n}}. \tag{4.3}$$

Next, we calculate an upper bound for the bias and the variance of $\widehat{PR}(\omega)$. From §§3.1, $X_j \overset{iid}{\sim} N(\mu_j, 1)$, $\frac{1}{p}\{\#j : \mu_j \neq 0) = \epsilon_p$ and if $\mu_j \neq 0$ then $\mu_j \sim g$ for some density g for $j = 1, \ldots, p$. So this class of models can be parametrized by ϵ_p and g. We consider a very broad class $\mathcal{G}_1$ with

$$\mathcal{G}_1 = \{(\epsilon_p, g) : 0 \leq \epsilon_p \leq 1 \,\&\, g \,\text{is any density}\,\}$$

However, in a lot of real datasets, for example in the gene microarray data for Prostate, Leukemia and Colon data, it is believed that the proportions of signals is very small i.e. ϵ_p is very small. In these cases, the estimator of PR in (4.1) can be modified, to get a better rate of convergence. For this we consider also another class of models $\mathcal{G}_2$. Also, for $\mathcal{G}_2$, we only consider smooth densities g characterized by the tail behaviour of $\hat{g}$, the Fourier transform of g. Hence, we take

$$\mathcal{G}_2 = \{(\epsilon_p, g) : 0 \leq \epsilon_p \leq \epsilon_0 p^{-\beta},\, 0 \leq \beta < \frac{1}{2} \,\text{and}\, |\xi|^{\alpha} |\hat{g}(\xi)| \leq A \,\text{for large}\; |\xi|\} \tag{4.4}$$

which is smaller than $\mathcal{G}_1$. The class $\mathcal{G}_2$ has been considered before in Cai and Jin (2008).

In the following subsection, our goal is to find a uniform upper bound for the MSE for $\widehat{PR}$ over $\mathcal{G}_1$ for any kernel ω and then find an optimal kernel by minimizing the upper bound with respect to the kernel ω. Subsection (4.2) deals with the same problem for $\mathcal{G}_2$.

4.1 Estimating PR for $\mathcal{G}_1$

Lemma (4.1) gives a uniform upper bound of the MSE of $\widehat{PR}(\omega)$ over $\mathcal{G}_1$ for any given kernel ω.

Lemma 4.1 *Fix $n \geq 1$ and $t > 0$, and a kernel ω. Over the class $\mathcal{G}_1$,*

$$\big(E[\widehat{PR}(\omega)] - PR\big)^2 \leq \frac{t}{\pi\sqrt{n}} \int e^{-\xi^2/n} \cdot (\omega(\xi) - 1)^2 \, d\xi. \tag{4.5}$$

and

$$\mathrm{Var}(\widehat{PR}(\omega) \leq \frac{t}{\pi\sqrt{n}} \int \frac{1}{p} e^{(1-\frac{1}{n})\xi^2} \cdot \omega^2(\xi) \, d\xi. \tag{4.6}$$

Proof of lemma 4.1 is in the appendix. Lemma 4.1 gives an upper bound to the MSE of $\widehat{PR}(\omega)$,

$$MSE(\widehat{PR}(\omega)) \leq \frac{t}{\pi\sqrt{n}} \int e^{-\xi^2/n}[(\omega(\xi) - 1)^2 + \frac{1}{p} e^{\xi^2} \cdot \omega^2(\xi)] \, d\xi, \tag{4.7}$$

Now, the optimal kernel ω is derived in Lemma (4.2) by minimizing the right hand side of (4.7) using standard variation principle.

Lemma 4.2 *Fix $n \geq 1$ and $t > 0$. A continuous compactly-supported kernel that minimizes the right hand side of (4.7) is given by*

$$\tilde{\omega}(\xi) = \big(1 + \frac{1}{p} e^{\xi^2}\big)^{-1}, \qquad -\infty < \xi < \infty$$

Proof: From (4.7) it follows that the optimization problem amounts to minimizing

$$F(\omega) = \frac{1}{\pi} \frac{t}{\sqrt{n}} \int_{-\infty}^{\infty} \Big[(\omega(\xi) - 1)^2 e^{-\xi^2/n} + \frac{1}{p} \omega^2(\xi) e^{\xi^2(1-\frac{1}{n})}\Big] \, d\xi$$

with respect to ω. In order to minimize F, we use calculus of variation principle. Let ω_1 be any smooth and symmetric function. If F has a minimum at $\tilde{\omega}$, then $F(\tilde{\omega} + \epsilon\omega_1)$ should have a derivative equal to 0 with respect to ϵ at $\epsilon = 0$.

$$\frac{\partial F(\tilde{\omega} + \epsilon\omega_1)}{\partial \epsilon}\bigg|_{\epsilon=0} = \frac{2}{\pi} \frac{t}{\sqrt{n}} \int_{-\infty}^{\infty} \Big[(\tilde{\omega}(\xi) - 1) e^{-\xi^2/n} + \frac{1}{p} \tilde{\omega}(\xi) e^{\xi^2(1-\frac{1}{n})}\Big] \omega_1(\xi) \, d\xi$$

Using the fact that, $\frac{\partial F(\tilde{\omega}+\epsilon\omega_1)}{\partial \epsilon}\big|_{\epsilon=0} = 0$ and ω_1 smooth, we get

$$(\tilde{\omega}(\xi) - 1) e^{-\xi^2/n} + \frac{1}{p} \tilde{\omega}(\xi) e^{\xi^2(1-\frac{1}{n})} = 0 \, \forall \; \xi$$

Hence it follows $\tilde{\omega}(\xi) = \frac{1}{1+\frac{1}{p}e^{\xi^2}}$. □

Having obtained the optimal kernel $\tilde{\omega}$, we can construct the estimator $\widehat{PR}(\tilde{\omega})$ from (4.1). The following theorem characterizes the MSE of $\hat{P}R(\tilde{\omega})$.

Theorem 4.1 *Fix $n \geq 1$ and $t > 0$. For sufficiently large p, there is a constant $C = C(n,t) > 0$ such that*

$$MSE(\hat{PR}(\tilde{\omega})) \leq \frac{C \cdot n^2}{\log^2(p)} \cdot p^{-1/n}.$$

Proof of theorem 4.1 is in the appendix. Recall from (3.10) and (4.2), that our main goal behind the Fourier approach was to construct an f using a kernel ω such that

$$f(\cdot;\omega) * \phi(u) = \Psi(u)$$

With the optimal kernel $\tilde{\omega}$, the difference between

$$f(\cdot;\tilde{\omega}) * \phi(u) \qquad \text{and} \qquad \Psi(u)$$

is surprisingly small. See for example Figure 1, where we compare two functions for $t = 2$ and $n = 2, 4, 6, 8$. For $n \leq 4$, the difference between two functions is very small. As n increases, the approximation becomes less accurate.

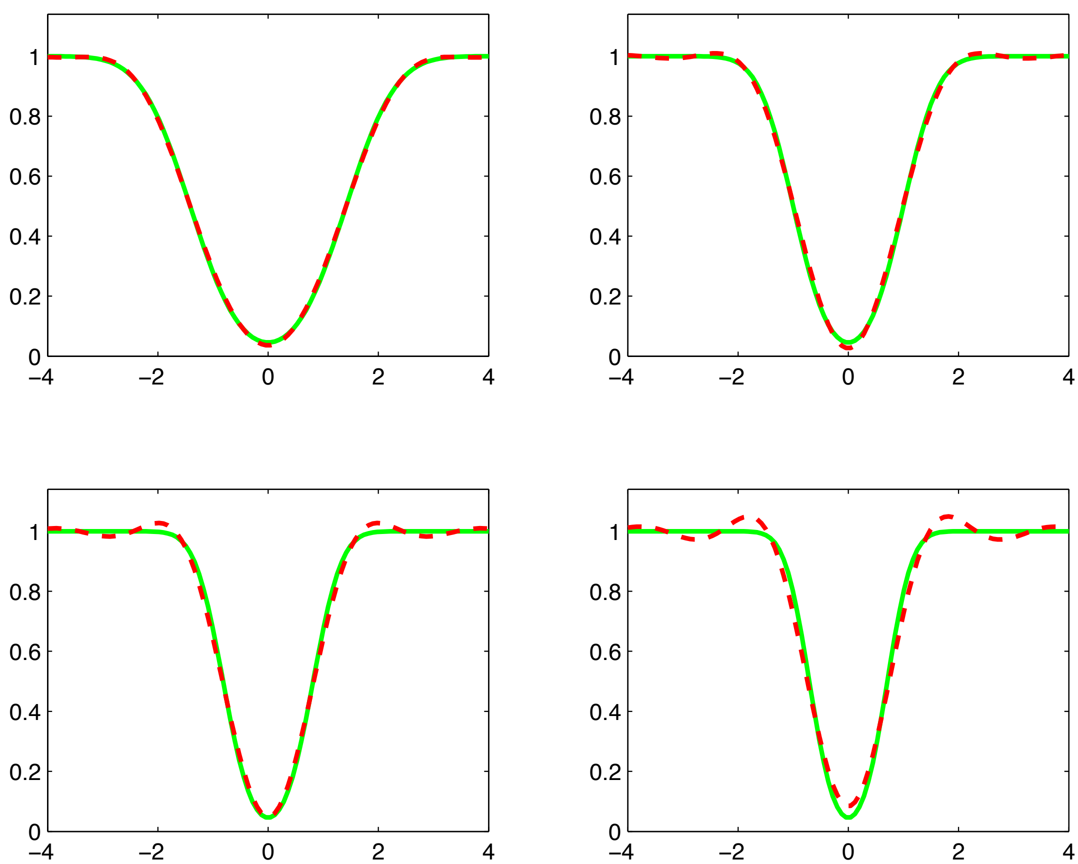

Figure 1: Display of $1 - f(\cdot;\tilde{\omega}) * \phi(u)$ (dashed) and $1 - \Psi(u)$ (solid) with $t = 2$, and $n = 2, 4, 6, 8$ from left to right then from top to bottom.

4.2 Estimating PR for $\mathcal{G}_2$

We now study the special case where ϵ_p is small and g is a smooth density, as in the class $\mathcal{G}_2$ in (4.4), a case that arises in many practical situations. In this case, it is possible to reduce considerably the rate of mean squared error of $\widehat{PR}$. In (4.1) for any kernel ω , it is possible to reduce the bias of $\widehat{PR}(\omega)$ without increasing its variance, by modifying the estimation procedure as explained below. First, fix any kernel ω. Recall that

$$PR(t;n) = (1-\epsilon_p)2\bar{\Phi}(t) + \epsilon_p \cdot \int (1-\Psi(u;t,n))\, g(u).$$
$$= 1 - \frac{(1-\epsilon_p)}{2\pi} \int \hat{\Psi}(\xi;t)\, d\xi - \frac{\epsilon_p}{2\pi} \cdot \int \Big[\int \hat{\Psi}(\xi;t,n) \cdot \cos(\xi u)\, d\xi\Big] g(u)\, du \qquad (4.8)$$

At the same time, direct calculations show that

$$E[\widehat{PR}(\omega)] = 1 - \frac{(1-\epsilon_p)}{2\pi} \int \omega(\xi) \cdot \hat{\Psi}(\xi;t)\, d\xi - \frac{\epsilon_p}{2\pi} \cdot \int \Big[\int \omega(\xi) \cdot \hat{\Psi}(\xi;t,n) \cdot \cos(\xi u)\, d\xi\Big] g(u)\, du. \qquad (4.9)$$

From (4.8) with (4.9) it follows that

$$\begin{array}{ll} Bias[\widehat{PR}(\omega)] = (1-\epsilon_p)b_0(\omega) + \epsilon_p b_1(\omega) & \text{where} \\ b_0(\omega) = \frac{1}{2\pi}\int (1-\omega(\xi)) \cdot \hat{\Psi}(\xi;t)\, d\xi \text{ and} & \\ b_1(\omega) = \frac{1}{2\pi} \cdot \int \big[\int (1-\omega(\xi)) \cdot \hat{\Psi}(\xi;t,n) \cdot \cos(\xi u)\, d\xi\big] g(u)\, du. & \end{array} \qquad (4.10)$$

Now given any kernel ω, $b_0(\omega)$ is known. So, if we use an estimator $\hat{\epsilon}_p$ for ϵ_p, and then estimate PR by

$$\widehat{PR}(\omega, \hat{\epsilon}_p) = \widehat{PR}(\omega) + (1-\hat{\epsilon}_p)b_0$$

then,

$$Bias[\widehat{PR}(\omega, \hat{\epsilon}_p)] = Bias(\hat{\epsilon}_p) \cdot b_0(\omega) + \epsilon_p b_1(\omega) \qquad (4.11)$$

In case of $\mathcal{G}_2$ where ϵ_p is very small, $Bias(\hat{\epsilon}_p)$ is much smaller than $(1-\epsilon_p)$ and so comparing (4.10) with (4.11), it is easy to see that the bias for $\widehat{PR}(\omega, \hat{\epsilon}_p)$ is smaller than $\widehat{PR}(\omega)$. From the proof of Theorem (4.2) below, it follows that the rate of variance of $\widehat{PR}(\omega, \hat{\epsilon}_p)$ is the same as that of $\widehat{PR}(\omega)$, given any kernel ω. The problem of estimating ϵ_p, the proportion of non-null hypotheses, has been extensively studied in Jin(2008). Estimation for ϵ_p is discussed below.

4.2.1 Estimation of proportion (ϵ_p)

Estimation of the proportion of non-null hypotheses in multiple testing is a very well-studied problem, see for example Jin(2008) and Cai and Jin(2009). As mentioned in subsection (3.2), the Fourier approach can be applied for estimating ϵ_p. The estimator proposed in Jin(2008), based on this approach, is

$$\tilde{\epsilon}_p = 1 - \frac{1}{p}\sum_{j=1}^{p} \int \tilde{r}(\xi) e^{t^2\xi^2/2} \cos(t\xi X_j)\, d\xi \qquad t = \sqrt{2\gamma_0 \log p} \qquad (4.12)$$

for an appropiately chosen γ_0 between 0 and $\frac{1}{2}$ and the kernel $\tilde{r}$ is any symmetric density on $[-1, 1]$. It follows from Cai and Jin(2009) that $\tilde{\epsilon}_p$ cannot achieve theoretically the optimal rate in the class $\mathcal{G}_2$ although simulations show it performs very well numerically. Here we propose another estimator $\hat{\epsilon}_p$, a slightly modified version of $\tilde{\epsilon}_p$ which performs numerically at least as good as $\tilde{\epsilon}_p$ and also achieves theoretically the optimal rate. The estimator we propose here is,

$$\hat{\epsilon}_p = 1 - \frac{1}{p} \sum_{j=1}^{p} \int r(\xi) e^{t^2\xi^2/2} \cos(t\xi X_j)\, d\xi \qquad t = \sqrt{2\gamma_0 \log p} \tag{4.13}$$

where $\gamma_0 \in (0, \frac{1}{2})$ and the kernel r is a smooth density on $[-1, 1]$ with no mass on $[-\delta, \delta]$ for some small δ. The choice of an appropiate γ_0 and δ as well as r will be discussed in §5.1 .

4.2.2 Rate of MSE for PR

It follows from (4.11), the final estimator proposed for estimating PR in $\mathcal{G}_2$ is, for any given kernel ω,

$$\widehat{PR}(\omega, \hat{\epsilon}_p) = \widehat{PR}(\omega) + \frac{(1-\hat{\epsilon}_p)}{2\pi} \int (1 - \omega(\xi)) \cdot \hat{\Psi}(\xi; t)\, d\xi \tag{4.14}$$

Following the same procedure as in lemma 4.1, the optimal kernel ω^* in this case can be obtained by minimizing the MSE of $\widehat{PR}(\omega, \epsilon_p)$ as a function of ω. It can be shown that the optimal kernel for estimating PR is,

$$\omega^*(\xi, \epsilon_p, \alpha) = \frac{1}{1 + \frac{(\log p)^\alpha}{p\epsilon_p^2} e^{\xi^2}} \qquad -\infty < \xi < \infty \tag{4.15}$$

Proof of (4.15) can be obtained by minor modifications of lemma 4.1, and hence is omitted. Using the optimal kernel ω^* in (4.15), the rate of MSE of $\widehat{PR}(\omega^*, \hat{\epsilon}_p)$ in the class $\mathcal{G}_2$ is given in the following theorem.

Theorem 4.2 *Fix $n \geq 1$ and $t > 0$. For sufficiently large p, there is a constant $C > 0$ such that*

$$MSE[\widehat{PR}(\omega^*, \hat{\epsilon}_p)] \leq \frac{C \cdot n^2}{\log^{2+\alpha(1-1/n)}(p)} \cdot \epsilon_p^{2(1-1/n)} p^{-1/n}.$$

The proof of the above theorem can be obtained by minor modifications of the Theorem 4.1 and hence is omitted. As we can see above, the kernel giving the optimal rate of convergence for the class $\mathcal{G}_2$ depends on ϵ_p as well as α which are unknown in practice. We shall discuss more about this problem in §6.

5 Simulation study

In this section, we discuss the choice of the kernel r and δ for estimating the proportion of non-null effects (ϵ_p), as described in (4.13). Then we also test the performance of the resulting estimator by simulation. We also test the performance of $\widehat{PR}(\omega^*, \hat{\epsilon}_p)$ in (4.14) by simulations.

5.1 Simulation study for estimating proportion

In this section, we discuss the choice of the tuning parameters γ_0, δ, and r for estimating ϵ_p. First consider the problem of estimating ϵ_p which involves the choice of the tuning parameter γ_0. Our proposed estimator from (4.13) is

$$\hat{\epsilon}_p = 1 - \frac{1}{p}\sum_{j=1}^{p}\int r(\xi)e^{t^2\xi^2/2}\cos(t\xi X_j)\,d\xi, \qquad t = \sqrt{2\gamma_0 \log p} \tag{5.1}$$

Recall from (4.13), r is a symmetric density on $[-1,1]$ with no mass on $[-\delta,\delta]$ for some small δ. We choose $\delta = 0.01$ and

$$r(\xi) = C \cdot e^{\frac{1}{1-\xi^2}}, \qquad \delta < |\xi| < 1$$

and for $\tilde{\epsilon}_p$ in (4.12) we again choose

$$\tilde{r}(\xi) = C \cdot e^{\frac{1}{1-\xi^2}}, \qquad |\xi| < 1$$

Simulation results show that choosing $\gamma_0 \in [0.2, 0.25]$ gives good numerical result for both $\tilde{\epsilon}_p$ and $\hat{\epsilon}_p$. Numerically, their performance depends both on the signal strength and on p. Here, we do simulations for different signal strengths for $p = 5000$ in the following way : for the signal strength s, we assume under H_1, $\mu \sim U(s, s+1)$ with s fixed in $[1,4]$. The value for ϵ_p is taken to be 10%. For each value of s simulate in the following way.

1. Generate μ from $U(s, s+1)$ and then generate an observation X from $(1-\epsilon_p)N(0,1) + \epsilon_p N(\mu, 1)$.

2. Repeat step 1, p times and estimate ϵ_p using $\hat{\epsilon}_p$ as well as $\tilde{\epsilon}_p$ with $\gamma_0 = 0.2$.

3. Repeat steps 1 & 2, 100 times and compute the mean squared error of $\hat{\epsilon}_p$.

Figure (2) illustrates the performance of $\tilde{\epsilon}_p$ and $\hat{\epsilon}_p$. Both seem to perform equally good. However, since $\hat{\epsilon}_p$ is theoretically optimal, we use $\hat{\epsilon}_p$ as an estimator of ϵ_p.

5.2 Simulation study for estimating PR

We now look at the numerical performance of our estimator. In this section, our objective is to predict the positive rate for n replications using the available data. Then we compare it with the actual positive rate for n replications.

Set $p = 10,000$. Take the range of the threshold value t to be $(1,3)$ which is usually the most interesting range for practical purposes. Set the proportion of non-null effects $\epsilon_p = 0.10$. For the number of replications n, take $n = 2$ and $n = 4$. Now, for each value of n, we do the following steps:

1. Generate $p\epsilon_p$ values of μ from $U(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}} + 1)$.

2. For each such value of μ generate an observation from $N(\mu, 1)$. Generate $p(1-\epsilon_p)$ observations from $N(0,1)$.

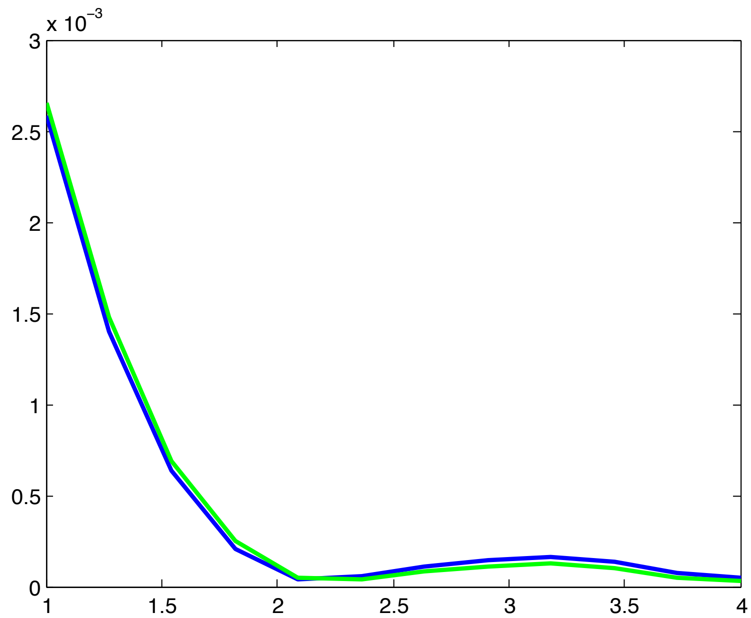

Figure 2: The plot displays the mean squared error of $\tilde{\epsilon}_p$ (green) and $\hat{\epsilon}_p$ (blue). The mean squared error is plotted on the $y-$axis versus the signal strength s along the $x-$axis. $\hat{\epsilon}_p$ does slightly better for weaker signals whereas $\tilde{\epsilon}_p$ does slightly better for strong signals, but overall there is not too much difference.

3. Using our estimator $PR^*(t, n)$ predict the positive rate for each value of the threshold t.

4. Repeat steps, 2 and 3 for 100 independent cycles.

As it can be seen in Figure (3), for the case $n = 2$, the estimated PR almost merges with the true PR with very low variance. For the case, $n = 4$, the bias and variance is slightly larger, but the estimated PR crve is significantly higher than the PR from the current data.

6 Proposed research

The next logical steps for this reseach falls into three categories: studying the minimax risk for estimating PR in the class $\mathcal{G}_2$, trying to find an estimator for PR which adapts to $\mathcal{G}_2$ assuming the parameters in $\mathcal{G}_2$ are unknown, and applying the estimation procedure for PR to real datasets. This section describes our proposed work in details.

6.1 Minimax risk

From theorem 4.2, it follows that the rate of MSE of $\widehat{PR}(\omega^*, \hat{\epsilon}_p)$ for estimating PR in $\mathcal{G}_2$ is

$$MSE[\widehat{PR}(\omega^*, \hat{\epsilon}_p)] \leq \frac{C \cdot n^2}{\log^{2+\alpha(1-1/n)}(p)} \cdot \epsilon_p^{2(1-1/n)} p^{-1/n}.$$

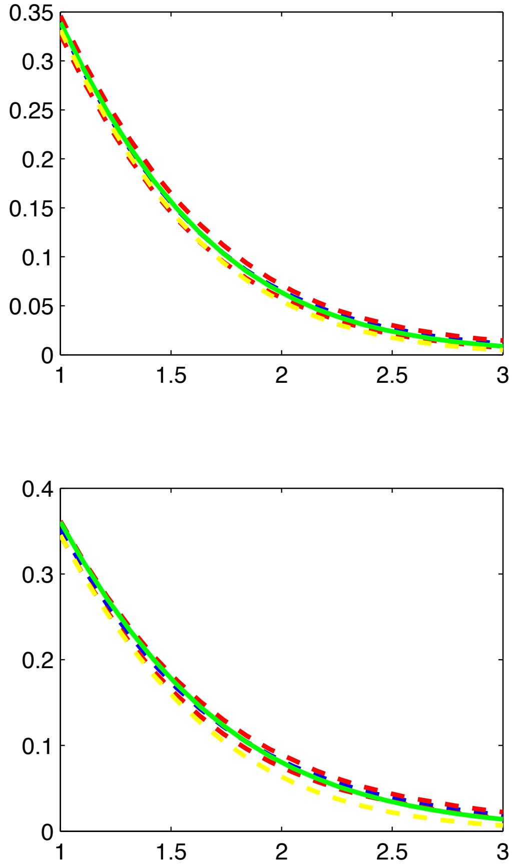

Figure 3: Display of the positive rate (PR) for threshold values $t \in [1, 3]$. The top row is for $\epsilon_p = 5\%$ and the bottom row is for $\epsilon_p = 10\%$ with $n = 2$ and $n = 4$ replications from left to right. The solid line(green) is the true PR and the blue dashed line is the estimated PR. The yellow dashed line is the PR from the current data. The red dashed line is the 95% confidence interval.

Now the minimax rate for estimating PR in $\mathcal{G}_2$ is defined to be

$$\mathcal{R}(\mathcal{G}_2) = \inf_{\hat{T}} \sup_{\mathcal{G}_2} E[\hat{T} - PR]^2$$

It will be interesting to calculate the minimax rate and compare with the MSE of $\widehat{PR}(\omega^*, \hat{\epsilon}_p)$. An estimator is said to be optimal if it achieves the minimax rate. So by computing the minimax rate we can compare our proposed estimator to see if it is optimal or not, or how far is it from the optimal rate.

6.2 Adapting to unknown class

Recall that the class $\mathcal{G}_2$ where we are trying to estimate PR in §§ 4.2 is defined to be

$$\mathcal{G}_2 = \{(\epsilon_p, g) : 0 \leq \epsilon_p \leq \epsilon_0 p^{-\beta},\ 0 \leq \beta < \frac{1}{2} \text{ and } |\xi|^\alpha |\hat{g}(\xi)| \leq A \text{ for large } |\xi|\}$$

Note that β corresponds to the proportion of non-null hypotheses and α corresponds to the unknown density of the true signals. For practical purposes, both of these parameters are unknown. However, from (4.14) and (4.15), it follows that the estimator $\widehat{PR}(\omega^*, \hat{\epsilon}_p)$ depends on both α and β. Hence, it will be worthwhile to find out a way to circumvent these kind of problems, either by changing the estimation procedure to adapt to these unknown parameters or by using plug-in estimators for the unknown parameters. In that case we also need to find out how this adjustment affects the rate of MSE of the estimator.

6.3 Application on real data

For real data, we will focus mainly on DNA microarray data. This kind of data consists of expression measurements for a large number of genes measured over a number of subjects, consisting of controls and cases. Our objective is to test which genes are differentially expressed. We select a fraction, say $\frac{1}{3}$ of the controls and the cases, and predict the PR for twice the sample size. The replication multiplicity, n, is equal to 2 in this case. Then, we can compare the predicted PR with the PR computed from the remaining $\frac{2}{3}$ of the subjects. The specific datasets we will consider are the colon data and the leukemia data(Golub *et al* (1999)).

7 Appendix

7.1 Proof of Theorem 4.1

Denote the bias of $PR_p^*(t,n)$ by $b^*(\omega;t,n,p)$. Using (7.8) we get,

$$\big|b^*(\omega;t,n,p)\big| \leq \frac{t}{\pi\sqrt{n}} \int |\omega^*(\xi)-1| \cdot |\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}| \cdot e^{-\xi^2/2n}\, d\xi.$$

Substituting ω^* from Lemma 4.2, we get

$$\big|b^*(\omega;t,n,p)\big| \leq \frac{t}{\pi\sqrt{n}} \int \Big[\frac{\frac{1}{p}e^{\xi^2}}{1+\frac{1}{p}e^{\xi^2}} |\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}| \cdot e^{-\xi^2/2n}\Big]\, d\xi.$$

We introduce some more notations for simplicity. Let $a(\xi) = \frac{\frac{1}{p}e^{\xi^2(1-\frac{1}{2n})}}{1+\frac{1}{p}e^{\xi^2}}$, $b(\xi) = \big|\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}\big|$ and

$$I_1 = \int \frac{\frac{1}{p}e^{\xi^2(1-\frac{1}{2n})}}{1+\frac{1}{p}e^{\xi^2}} |\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}|\, d\xi. \tag{7.1}$$

Then, $I_1 = \int a(\xi)b(\xi)\, d\xi$. The change of variable, $\xi = \sqrt{\log p} + \frac{\eta}{2\sqrt{\log p}}$ in (7.1), gives

$$a(\xi) = \frac{\frac{1}{p}e^{\left(\log p + \eta + \frac{\eta^2}{4\log p}\right)(1-\frac{1}{2n})}}{1+\frac{1}{p}e^{\left(\log p+\eta+\frac{\eta^2}{4\log p}\right)}} = p^{-\frac{1}{2n}} \frac{e^{(\eta+\frac{\eta^2}{4\log p})(1-\frac{1}{2n})}}{1+e^{(\eta+\frac{\eta^2}{4\log p})}}$$

$$\text{and } b(\xi) = \Big|\frac{\sin\left(\frac{t\sqrt{\log p}}{\sqrt{n}}(1+\frac{\eta}{2\log p})\right)}{\frac{t\sqrt{\log p}}{\sqrt{n}}(1+\frac{\eta}{2\log p})}\Big| \leq \frac{\sqrt{n}}{t\sqrt{\log p}} \frac{1}{|1+\frac{\eta}{2\log p}|}\ , \text{ and hence}$$

$$I_1 = \int a(\xi)b(\xi)\, d\xi \leq \frac{\sqrt{n}p^{-\frac{1}{2n}}}{t\log p} \int \frac{1}{|1+\frac{\eta}{2\log p}|} \frac{e^{(\eta+\frac{\eta^2}{4\log p})(1-\frac{1}{2n})}}{1+e^{(\eta+\frac{\eta^2}{4\log p})}}\, d\eta$$

As $p \to \infty$, by Dominated Convergence Theorem,

$$\int \frac{1}{|1+\frac{\eta}{2\log p}|} \frac{e^{(\eta+\frac{\eta^2}{4\log p})(1-\frac{1}{2n})}}{1+e^{(\eta+\frac{\eta^2}{4\log p})}}\, d\eta \sim \int \frac{e^{\eta(1-\frac{1}{2n})}}{1+e^{\eta}} d\eta \sim 2n \cdot C_1 \Rightarrow I_1 \lesssim C_1\Big(\frac{n^{3/2}p^{-\frac{1}{2n}}}{t\log p}\Big) \tag{7.2}$$

Combining (7.1) and (7.2) gives,

$$\big|b^*(\omega;t,n,p)\big| \lesssim C_1 \cdot p^{-(\frac{1}{2n})} \frac{n}{\log p} \leq C_1 \cdot p^{-(\frac{1}{2n})} \frac{n}{\log p}\,, \tag{7.3}$$

uniformly for all ϵ_p between 0 and 1. Using (7.11),

$$\begin{aligned} \mathrm{Var}(PR_p^*(t,n)) &\leq \frac{1}{p} \cdot E\Big[\Big(\frac{t}{\pi\sqrt{n}} \int \omega^*(\xi) \frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}} e^{(1-\frac{1}{n})\xi^2/2} \cos(\xi X_1)\, d\xi\Big)^2\Big] \\ &\leq \frac{t^2}{p\pi^2 n}\Big(\int |\omega^*(\xi)| |\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}| e^{(1-\frac{1}{n})\xi^2/2}\, d\xi\Big)^2 \end{aligned}$$

Inserting ω^* from Lemma 4.2 and using triangle inequality and symmetry around 0 gives,

$$\mathrm{Var}(PR_p^*(t,n)) \le 4\frac{t^2}{p\pi^2 n}\Big(\int_0^\infty \frac{1}{1+\frac{1}{p}e^{\xi^2}}\big[\big|\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}\big|e^{\xi^2(1-\frac{1}{n})\frac{1}{2}}\big]\,d\xi\Big)^2 \tag{7.4}$$

We introduce some more notations.

$$\text{Let}\ , I_{21} = \int_0^\infty \big|\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}\big|\frac{e^{\xi^2(1-\frac{1}{n})\frac{1}{2}}}{1+\frac{1}{p}e^{\xi^2}}\,d\xi \tag{7.5}$$

Also let, $a_1(\xi) = \frac{e^{\xi^2(1-\frac{1}{n})\frac{1}{2}}}{1+\frac{1}{p}e^{\xi^2}}$ and, as before, $b(\xi) = \big|\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}\big|$. Then, $I_{21} = \int_0^\infty b(\xi)a_1(\xi)\,d\xi$. Using the change of variable, $\xi = \sqrt{\log p} + \frac{\eta}{2\sqrt{\log p}}$ in I_{21} we get,

$$a_1(\xi) = \frac{e^{(\log p+\eta+\frac{\eta^2}{4\log p})(1-\frac{1}{n})\frac{1}{2}}}{1+\frac{1}{p}e^{(\log p+\eta+\frac{\eta^2}{4\log p})}} = p^{(1-\frac{1}{n})\frac{1}{2}}\frac{e^{(\eta+\frac{\eta^2}{4\log p})(1-\frac{1}{n})\frac{1}{2}}}{1+e^{\eta+\frac{\eta^2}{4\log p}}}\ \text{and as before}\ b(\xi) \le \frac{\sqrt{n}}{t\sqrt{\log p}}\frac{1}{|1+\frac{\eta}{2\log p}|}$$

Hence we get,

$$\begin{aligned} I_{21} &\le \frac{\sqrt{n}}{t}\frac{p^{(1-\frac{1}{n})\frac{1}{2}}}{\log p}\int_{-\infty}^\infty \frac{e^{(\eta+\frac{\eta^2}{4\log p})(1-\frac{1}{n})\frac{1}{2}}}{1+e^{\eta+\frac{\eta^2}{4\log p}}}\frac{\mathbf{1}(\eta > -2\log p)}{|1+\frac{\eta}{2\log p}|}\,d\eta \\ &\sim \frac{\sqrt{n}}{t}\frac{Cp^{(1-\frac{1}{n})\frac{1}{2}}}{\log p}\int_{-\infty}^\infty \frac{e^{\eta(1-\frac{1}{n})\frac{1}{2}}}{1+e^\eta}d\eta = C\cdot\frac{\sqrt{n}}{t}\frac{p^{(1-\frac{1}{n})\frac{1}{2}}}{\log p} \end{aligned} \tag{7.6}$$

The last approximation in (7.6) follows from Dominated Convergence Theorem. Inserting I_{21} in (7.4) gives,

$$\mathrm{Var}(PR_p^*(t,n)) \le C\cdot\frac{p^{-\frac{1}{n}}}{\log^2 p}\Big(1+\frac{n^{3/2}}{t\log^{3/2}p}\Big)^2 \sim C\cdot\frac{p^{-\frac{1}{n}}}{\log^2 p} \tag{7.7}$$

Together, (7.3) and (7.7) gives,

$$\mathrm{MSE}(PR_p^*(t,n)) \lesssim C\cdot\Big(p^{-\frac{1}{n}}\frac{n^2}{\log^2 p}+\frac{p^{-\frac{1}{n}}}{\log^2 p}\Big) \sim C\cdot p^{-\frac{1}{n}}\frac{n^2}{\log^2 p}$$

7.2 Proof of Lemma 4.1

Consider the first claim. For short, denote the bias by $b(\omega;t,n,p) = E[PR_p(\omega;t,n,X_1,\ldots,X_p)] - PR(t;n)$. By (4.3) and (4.9),

$$\begin{aligned} |b(\omega;t,n,p)| &= \frac{t}{\pi\sqrt{n}}\Big|\int(\omega(\xi)-1)\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}e^{-\xi^2/2n}\cdot\big[(1-\epsilon)+\epsilon\int\cos(\xi u)\,dF_p(u)\big]d\xi\Big| \\ &\le \frac{t}{\pi\sqrt{n}}\int|\omega(\xi)-1|\cdot\big|\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}\cdot|e^{-\xi^2/2n}\,d\xi. \end{aligned} \tag{7.8}$$

Use Hölder inequality,

$$\Big(\int |\omega(\xi)-1|\cdot|\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}|\cdot e^{-\xi^2/2n}\,d\xi\Big)^2 \leq \Big(\int (\omega(\xi)-1)^2\cdot e^{-\xi^2/n}\,d\xi\Big)\cdot\Big(\int \Big(\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}\Big)^2\,d\xi\Big), \tag{7.9}$$

where by elementary calculus,

$$\int \Big(\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}\Big)^2\,d\xi = \frac{\sqrt{n}}{t}\int \frac{\sin^2(\eta)}{\eta^2}\,d\eta = \sqrt{n}\pi/t. \tag{7.10}$$

Inserting (7.9) and (7.10) into (7.8) gives the first claim.

Consider the second claim. By definition and symmetry

$$\begin{aligned} \mathrm{Var}[PR_p(\omega;t,n,)] &= \frac{1}{p}\cdot \mathrm{Var}\Big(\frac{t}{\pi\sqrt{n}}\int \omega(\xi)\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}e^{(1-\frac{1}{n})\xi^2/2}\cos(\xi X_1)\,d\xi\Big) \\ &\leq \frac{1}{p}\cdot E\Big[\Big(\frac{t}{\pi\sqrt{n}}\int \omega(\xi)\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}e^{(1-\frac{1}{n})\xi^2/2}\cos(\xi X_1)\,d\xi\Big)^2\Big]. \end{aligned} \tag{7.11}$$

Since $|\cos(\xi X_1)| \leq 1$,

$$E\Big[\Big(\int \omega(\xi)\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}e^{(1-\frac{1}{n})\xi^2/2}\cos(\xi X_1)\,d\xi\Big)^2\Big] \leq \Big(\int |\omega(\xi)|\cdot|\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}|\cdot e^{(1-\frac{1}{n})\xi^2/2}\,d\xi\Big)^2. \tag{7.12}$$

Now, by similar argument,

$$\Big(\int |\omega(\xi)|\cdot|\frac{\sin(\xi t/\sqrt{n})}{\xi t/\sqrt{n}}|\cdot e^{(1-\frac{1}{n})\xi^2/2}\,d\xi\Big)^2 \leq \frac{t}{\pi\sqrt{n}}\int \omega^2(\xi)e^{(1-\frac{1}{n})\xi^2}\,d\xi. \tag{7.13}$$

Inserting (7.12) and (7.13) into (7.11) gives the second claim. □

References

[1] Abramovich, F., Benjamini, Y., 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] Cai, T., Jin, J. and Low, M. (2007). Estimation and confidence sets for sparse normal mixtures. *Ann. Statist.* **35** 2421-2449.

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

[5] Donoho, D. and Jin, J. (2006). Asymptotic minimaxity of false discovery rate thresholding for sparse exponential data. *Ann. Statist.* **34** 2980-3018.

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

[7] Efron, B. (2008). Microarrays, empirical Bayes, and the two-groups model. *Statist. Sci.* **23** 1-22.

[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] Jin, J. (2008). Proportion of nonzero normal means: universal oracle equivalences and uniformly consistent estimations. *J. Roy. Statist. Assoc. Ser. B* **70**(3) 461-493.

[12] Jin, J. and Cai, T. (2006). Estimating the null and the proportion of non-null effects in large-scale multiple comparisons. *ArXiv.math/0611108v1.*

[13] Jin, J. and Cai. T. (2007). Estimating the null and the proportion of non-null effects in large-scale multiple comparisons. *J. Amer. Statist. Assoc.* **102** 495-506.

[14] Mallat, S. (1998). *A Wavelet Tour of Signal Processing.* 2nd ed. Academic Press.

[15] Meinshausen, M. and Rice, J. (2006). Estimating the proportion of false null hypothesis among a large number of independent tested hypotheses. *Ann. Statist* **34** 373-39.

[16] Storey, J. D. (2002). A direct approach to false discovery rate. *J. Roy. Statist. Soc. Ser. B* **64** 479-498.

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

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