Journal of Multivariate Analysis 142 (2015) 106–116

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Capturing the severity of type II errors in high-dimensional multiple testing

Li He [a], Sanat K. Sarkar [b,*], Zhigen Zhao [b]

[a] Merck Research Laboratories, West Point, PA 19486, United States
[b] Department of Statistics, Temple University, Philadelphia, PA, 19122, United States

ARTICLE INFO

Article history:
Received 2 February 2014
Available online 24 August 2015

AMS 2000 subject classifications:
62J15
62C10
62F03

Keywords:
Bayes rule
Compound decision theory
Oracle procedure
Multiple testing
Weighted marginal false discovery rate
Weighted marginal false non-discovery rate

ABSTRACT

The severity of type II errors is frequently ignored when deriving a multiple testing procedure, even though utilizing it properly can greatly help in making correct decisions. This paper puts forward a theory behind developing a multiple testing procedure that can incorporate the type II error severity and is optimal in the sense of minimizing a measure of false non-discoveries among all procedures controlling a measure of false discoveries. The theory is developed under a general model allowing arbitrary dependence by taking a compound decision theoretic approach to multiple testing with a loss function incorporating the type II error severity. We present this optimal procedure in its oracle form and offer numerical evidence of its superior performance over relevant competitors.

1. Introduction

Simultaneous testing of multiple hypotheses is an integral part of analyzing high-dimensional data from modern scientific investigations like those in genomics, brain imaging, astronomy, and many others, making multiple testing an area of current importance and intense statistical research. A variety of multiple testing methods have been put forward in the literature from both frequentist and Bayesian perspectives. However, the theories behind the development of these methods are mostly driven by the overreaching goal of controlling an overall measure of type I errors or false discoveries, with other fundamentally important statistical issues often being ignored. For instance, in many of the aforementioned experiments there is a cost associated with the error of making a false discovery or missing a true discovery, and this cost increases with increasing severity of that error. This is an important issue not often taken into account when developing multiple testing procedures.

A Bayesian decision theoretic approach can yield a powerful multiple testing method not only incorporating costs of false and missed discoveries but also simultaneously addressing dependency, optimality, and multiplicity [12,13]. This motivates us to take a similar approach, but in a more general framework that conforms more to the present problem, that is, to address the aforementioned issue related to severity of errors. Before explaining this generalization, let us first briefly outline the approach taken in [12,13].

[*] Corresponding author.
E-mail addresses: li.he@merck.com (L. He), sanat@temple.edu (S.K. Sarkar), zhaozhg@temple.edu (Z. Zhao).

http://dx.doi.org/10.1016/j.jmva.2015.08.005
0047-259X/

Given a set of observations $\mathbf{X} = (X_1, \ldots, X_m) \sim f(\mathbf{x}, \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m) \in \{0, 1\}^m$, consider the problem of deciding between $H_i : \theta_i = 0$ and $\bar{H}_i : \theta_i = 1$ simultaneously for $i = 1, \ldots, m$, assuming that $X_i \mid \theta_i \overset{ind}{\sim} (1-\theta_i)f_0(x_i)+\theta_i f_1(x_i)$, for some given densities f_0 and f_1, and $\theta_i \sim \mathit{Bernoulli}(1 - \pi_0)$. Sun and Cai [12,13] started with the following uniformly weighted 0–1 loss function:

$$L_\lambda(\boldsymbol{\delta}(\mathbf{X}), \boldsymbol{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\{\lambda(1-\theta_i)\delta_i(\mathbf{X}) + \theta_i(1-\delta_i(\mathbf{X}))\}, \tag{1.1}$$

for a decision rule $\boldsymbol{\delta}(\mathbf{X}) = (\delta_1(\mathbf{X}), \ldots, \delta_m(\mathbf{X})) \in \{0, 1\}^m$, where λ is the relative cost of making a false discovery (type I error) to that of missing a true discovery (type II error) and assumed to be constant over all the hypotheses. They considered the Bayes rule associated with this loss function and showed that it is also optimal from a multiple testing point of view. Specifically, given any $\alpha \in (0, 1)$, there exists a $\lambda \equiv \lambda(\alpha)$ for which it controls the marginal false discovery rate,

$$\mathrm{mFDR} = \frac{E\left[\sum_{i=1}^{m}\delta_i(\mathbf{X})(1-\theta_i)\right]}{E\left[\sum_{i=1}^{m}\delta_i(\mathbf{X})\right]},$$

at α, and minimizes the marginal false non-discovery rate,

$$\mathrm{mFNR} = \frac{E\left[\sum_{i=1}^{m}\{1-\delta_i(\mathbf{X})\}\theta_i\right]}{E\left[\sum_{i=1}^{m}\{1-\delta_i(\mathbf{X})\}\right]},$$

among all decision rules defined in terms of statistics satisfying a monotone likelihood ratio condition (MLR) and controlling the mFDR at α. They expressed this optimal procedure in an alternative form using hypothesis specific test statistics defined in terms of the local FDR measure (Lfdr, Efron [5]), and called it the oracle procedure. They provided numerical evidence showing that their oracle procedure can outperform its competitors, such as those in [1,7].

Clearly, the loss function used in the above formulation is somewhat simplistic. It gives equal importance to all type I errors as well as to all type II errors. While it might be reasonable to treat the type I errors equally in terms of severity and attach a fixed cost to all of them, it is often unrealistic to do so for type II errors. For instance, in a microarray experiment, there might be a fixed cost of doing a targeted experiment to verify that each gene is active and the loss due to making a false discovery might be that cost (which is being wasted in case the gene is found to be inactive). However, it would be unrealistic to assume that the loss in identifying a truly active gene as inactive does not depend on how strong is the expected signal that has been missed. In fact, it might reasonably be proportional to the difference [3,14,11] or even to the squared difference between the expected values of the missed and no signals.

In other words, the above formulation needs to be generalized conforming it more to the reality in modern high-dimensional multiple testing. With that in mind, we consider testing $H_i : \mu_i = \mu_{i0}$ against its one or two-sided alternative, for some specified values μ_{i0}, simultaneously for $i = 1, \ldots, m$, under the following model:

$$\begin{aligned}
&\mathbf{X} \mid \boldsymbol{\mu}, \boldsymbol{\theta} \sim f(\mathbf{x} \mid \boldsymbol{\mu}), \quad \text{with } \boldsymbol{\mu} = (\mu_1, \ldots, \mu_m),\ \boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)\\
&\mu_i \mid \theta_i \sim (1-\theta_i)I(\mu_i = \mu_{i0}) + \theta_i h(\mu_i - \mu_{i0})\\
&\theta_i \sim \mathit{Bernoulli}(1-\pi_0),
\end{aligned} \tag{1.2}$$

given a density h, and under the following more general loss function:

$$L_{\lambda,s}(\boldsymbol{\delta}(\mathbf{X}), \boldsymbol{\mu}, \boldsymbol{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\{\lambda(1-\theta_i)\delta_i(\mathbf{X}) + s(\mu_i-\mu_{i0})\theta_i(1-\delta_i(\mathbf{X}))\}. \tag{1.3}$$

We do not impose any dependence restriction on $\mathbf{X}$, $\boldsymbol{\mu}$ or $\boldsymbol{\theta}$. It is assumed that there is only a baseline cost λ_1 for each type I error (which, as argued above, is reasonable for a point null hypothesis). For each type II error, however, we assume that the cost is λ_2, the baseline cost, times $s(\mu_i - \mu_{i0})$, a function s of $\mu_i - \mu_{i0}$ such that $s(0) = 0$ and is non-decreasing as μ_i moves away from μ_{i0}. We call $s(\cdot)$ the *severity function* for type II errors. Through this function, a penalty is being imposed on making a type II error for each H_i; the larger the value of $|\mu_i - \mu_{i0}|$ is, the more severe this penalty is. The λ equals λ_1/λ_2, the relative baseline cost of a type I error to a type II error. In other words, $\lambda/s(\mu_i - \mu_{i0})$ is the relative cost of a type I error to a type II error. The specific choice of $s(\cdot)$ will depend on how fast we want the cost of the type II error to increase as μ_i moves away from μ_{i0}.

Our proposed loss function (1.3) is a non-uniformly weighted 0–1 loss function giving less and less weight to the type I error relative to the type II error as the type II error gets more and more severe as measured by the severity function. In this paper, we focus on deriving the theoretical form of an optimal multiple testing procedure from the Bayes rule under this general loss function. Given a severity function s, this Bayes rule provides an optimal multiple testing procedure in the sense of minimizing a measure of non-discoveries subject to controlling a measure of false discoveries at a specified level for a suitably chosen λ. These measures of false discoveries and false non-discoveries are of course different from the mFDR and mFNR, respectively, since we now need to account for the weights or penalties attached to the type II errors through the severity function that is not necessarily equal to one. We define these newer error rates as weighted mFDR and weighted mFNR and establish the aforementioned optimality result through these rates. We study the performance of this oracle optimal procedure with its relevant competitors through three numerical studies.

The remainder of the paper is organized as follows. The development of the Bayes rule under the loss function (1.3), its characterization as an optimal multiple testing procedure in the framework of weighted false discovery and false non-discovery rates, and our oracle multiple testing procedure are given in the next section. In Section 3, we present the results of three numerical studies providing evidence of this oracle procedure's superior performance over its relevant competitors. We end the paper with some concluding remarks in Section 4.

2. Optimal rules

Assuming that our problem is that of testing $H_i : \mu_i = 0$ simultaneously for $i = 1, \dots, m$ under the model (1.2) and the loss function $L_{\lambda,s}$ in (1.3), we do the following in this section: (i) determine the Bayes rule; (ii) show that the Bayes rule with an appropriately chosen λ provides an optimal multiple testing procedure in the sense of minimizing a measure of false non-discoveries among all rules that control a measure of false discoveries at a specified level; and (iii) express this optimal multiple testing procedure in terms of some test statistics to define the oracle procedure in this paper.

2.1. The Bayes rule

Let us first define

$$w_i(\mathbf{X}) = E\left[s(\mu_i) \mid \theta_i = 1, \mathbf{X}\right], \tag{2.1}$$

the average severity of type II errors conditional on the data **X** and $\theta_i = 1$. Then, we have the following:

Theorem 2.1. *Consider testing $H_i : \mu_i = 0$ simultaneously for $i = 1, \dots, m$ under the model (1.2) and the loss function (1.3). Then, the decision rule $\boldsymbol{\delta}^*(\mathbf{X}) = (\delta_1^*(\mathbf{X}), \dots, \delta_m^*(\mathbf{X}))$, where*

$$\delta_i^*(\mathbf{X}) = \begin{cases} 1 & \text{if } P(\theta_i = 0 \mid \mathbf{X}) \le \dfrac{w_i(\mathbf{X})}{\lambda} P(\theta_i = 1 \mid \mathbf{X}) \\ 0 & \text{if } P(\theta_i = 0 \mid \mathbf{X}) > \dfrac{w_i(\mathbf{X})}{\lambda} P(\theta_i = 1 \mid \mathbf{X}), \end{cases} \tag{2.2}$$

is the Bayes rule.

Proof. For any rule $\boldsymbol{\delta}(\mathbf{X}) = (\delta_1(\mathbf{X}), \dots, \delta_m(\mathbf{X}))$, we have

$$\begin{aligned} E\left[L_{\lambda,s}(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\delta}(\mathbf{X})) \mid \mathbf{X}\right] &= \frac{1}{m}\sum_{i=1}^m \{\lambda\delta_i(\mathbf{X})P(\theta_i = 0 \mid \mathbf{X}) + [1 - \delta_i(\mathbf{X})]E\left[s(\mu_i)I(\theta_i = 1) \mid \mathbf{X}\right]\} \\ &= \frac{1}{m}\sum_{i=1}^m \{\lambda\delta_i(\mathbf{X})P(\theta_i = 0 \mid \mathbf{X}) + [1 - \delta_i(\mathbf{X})]E\left[s(\mu_i) \mid \theta_i = 1, \mathbf{X}\right]P(\theta_i = 1 \mid \mathbf{X})\} \\ &= \frac{1}{m}\sum_{i=1}^m \{w_i(\mathbf{X})P(\theta_i = 1 \mid \mathbf{X}) + \delta_i(\mathbf{X})\left[\lambda P(\theta_i = 0 \mid \mathbf{X}) - w_i(\mathbf{X})P(\theta_i = 1 \mid \mathbf{X})\right]\}. \end{aligned}$$

Since the first term is constant with respect to **δ**, given **X**, it is clear that $\boldsymbol{\delta}^*(\mathbf{X})$ in (2.2) is the rule for which this conditional expectation is the minimum among all **δ**, and hence is Bayes. □

2.2. Optimal multiple testing procedure

Here we show that the aforementioned Bayes rule with an appropriately chosen λ provides an optimal multiple testing procedure in the sense of minimizing a measure of false non-discoveries among all rules that control a measure of false discoveries at a specified level. These measures of false discoveries and false non-discoveries are defined for any multiple testing rule **δ** as

$$\text{mFDR}^*(\boldsymbol{\delta}) = \frac{E\left[\sum_{i=1}^m \delta_i(\mathbf{X})(1 - \theta_i)w^*(\theta_i, \mu_i)\right]}{E\left[\sum_{i=1}^m \delta_i(\mathbf{X})w^*(\theta_i, \mu_i)\right]}, \tag{2.3}$$

and

$$\mathrm{mFNR}^*(\boldsymbol{\delta}) = \frac{E\left[\sum_{i=1}^{m}\{1-\delta_i(\mathbf{X})\}\theta_i w^*(\theta_i,\mu_i)\right]}{E\left[\sum_{i=1}^{m}\{1-\delta_i(\mathbf{X})\}\, w^*(\theta_i,\mu_i)\right]}, \tag{2.4}$$

respectively, where

$$w^*(\theta,\mu) = \begin{cases} 1 & \text{if } \theta = 0 \\ s(\mu) & \text{if } \theta = 1. \end{cases}$$

With $w^*(\theta_i, \mu_i)$ representing a weight associated with the ith hypothesis, these measures of false discoveries and false non-discoveries can be referred to as weighted mFDR and weighted mFNR, respectively. The severity function $s(\cdot)$ has an effect on these weighted measures through the weight function w^*. Suppose there are two severity functions s_1 and s_2, satisfying $s_1(\mu) > s_2(\mu), \forall\mu \neq 0$. Then for any decision rule $\boldsymbol{\delta}$, the mFDR* based on s_1 is smaller than that based on s_2, and the mFNR* based on s_1 is greater than that based on s_2. Particularly, if $s(\mu) > 1, \forall\mu \neq 0$, the corresponding mFDR* is a less conservative error rate than the mFDR; and if $s(\mu) = 1, \forall\mu \neq 0$, the corresponding mFDR* and mFNR* reduce to the mFDR and mFNR respectively.

Theorem 2.2. *Consider the model in (1.2). Suppose there exists a testing procedure $\boldsymbol{\delta}_0(\mathbf{X}) = (\delta_{10}(\mathbf{X}), \ldots, \delta_{m0}(\mathbf{X}))$ such that $\delta_{i0}(\mathbf{X})$ is defined as in (2.2) and $\mathrm{mFDR}^*(\boldsymbol{\delta}_0) = \alpha$. Let $\boldsymbol{\delta}(\mathbf{X})$ be any other rule such that $\mathrm{mFDR}^*(\boldsymbol{\delta}) \leq \alpha$. Then $\mathrm{mFNR}^*(\boldsymbol{\delta}_0) \leq m\mathrm{FNR}^*(\boldsymbol{\delta})$.*

Proof. First note that

$$\sum_{i=1}^{m} E\left[\{\delta_{i0}(\mathbf{X}) - \delta_i(\mathbf{X})\}\left\{P(\theta_i = 0 \mid \mathbf{X}) - \frac{w_i(\mathbf{X})}{\lambda}P(\theta_i = 1 \mid \mathbf{X})\right\}\right] \leq 0, \tag{2.5}$$

according to (2.2), and

$$\sum_{i=1}^{m} E\left[\{\delta_{i0}(\mathbf{X}) - \delta_i(\mathbf{X})\}\left\{P(\theta_i = 0 \mid \mathbf{X}) - \frac{\alpha}{1-\alpha}w_i(\mathbf{X})P(\theta_i = 1 \mid \mathbf{X})\right\}\right] \geq 0, \tag{2.6}$$

from the assumption, $m\mathrm{FDR}^*(\boldsymbol{\delta}) \leq \alpha = m\mathrm{FDR}^*(\boldsymbol{\delta}_0)$. From (2.5) and (2.6), we get

$$\sum_{i=1}^{m} E\left[\{\delta_{i0}(\mathbf{X}) - \delta_i(\mathbf{X})\}\, w_i(\mathbf{X})P(\theta_i = 1 \mid \mathbf{X})\left(\frac{1}{\lambda} - \frac{\alpha}{1-\alpha}\right)\right] \geq 0,$$

which implies that

$$\sum_{i=1}^{m} E\left[\delta_{i0}(\mathbf{X})w_i(\mathbf{X})P(\theta_i = 1 \mid \mathbf{X})\right] \geq \sum_{i=1}^{m} E\left[\delta_i(\mathbf{X})w_i(\mathbf{X})P(\theta_i = 1 \mid \mathbf{X})\right], \tag{2.7}$$

since

$$\frac{\alpha}{1-\alpha} = \frac{\sum_{i=1}^{m} E\left[\delta_{i0}(\mathbf{X})P(\theta_i = 0 \mid \mathbf{X})\right]}{\sum_{i=1}^{m} E\left[\delta_{i0}(\mathbf{X})w_i(\mathbf{X})P(\theta_i = 1 \mid \mathbf{X})\right]} \leq \frac{1}{\lambda}.$$

Thus, we have from (2.7)

$$E\left[\sum_{i=1}^{m}\left\{\frac{1-\delta_{i0}(\mathbf{X})}{\sum_{i=1}^{m} E\left[\{1-\delta_{i0}(\mathbf{X})\}\, w_i(\mathbf{X})P(\theta_i = 1 \mid \mathbf{X})\right]} - \frac{1-\delta_i(\mathbf{X})}{\sum_{i=1}^{m} E\left[\{1-\delta_i(\mathbf{X})\}\, w_i(\mathbf{X})P(\theta_i = 1 \mid \mathbf{X})\right]}\right\}\right.$$
$$\left.\times\left\{P(\theta_i = 0 \mid \mathbf{X}) - \frac{w_i(\mathbf{X})}{\lambda}P(\theta_i = 1 \mid \mathbf{X})\right\}\right] \geq 0.$$

This implies that

$$\frac{1 - m\mathrm{FNR}^*(\boldsymbol{\delta}_0)}{m\mathrm{FNR}^*(\boldsymbol{\delta}_0)} \geq \frac{1 - m\mathrm{FNR}^*(\boldsymbol{\delta})}{m\mathrm{FNR}^*(\boldsymbol{\delta})},$$

that is, $m\mathrm{FNR}^*(\boldsymbol{\delta}_0) \leq m\mathrm{FNR}^*(\boldsymbol{\delta})$, as desired. □

Remark 2.1. Theorem 2.2 improves the work of Sun and Cai [12] in the following senses: (1) it accommodates situations where penalties or weights associated with type II errors can be assessed through a severity function and incorporated into the development of a multiple testing procedure; and (2) it provides a rule that is optimal among all procedures controlling the $m\mathrm{FDR}^*$ at level α without any distributional restriction on the corresponding test statistics. Next, we will prove the existence of such a procedure $\boldsymbol{\delta}_0(\mathbf{X})$.

We can express the optimal procedure $\boldsymbol{\delta}_0(\mathbf{X})$ in Theorem 2.2 in terms of the following test statistics:

$$T_i(\mathbf{X}) = \frac{P(\theta_i = 0 \mid \mathbf{X})}{P(\theta_i = 0 \mid \mathbf{X}) + w_i(\mathbf{X})P(\theta_i = 1 \mid \mathbf{X})}, \quad i = 1, \dots, m. \tag{2.8}$$

The statistic T_i will be referred to as generalized local fdr (Glfdr). When $s(\mu) = 1$, it reduces to the usual definition of the local fdr (Lfdr) of Efron [4] under independence and to the test statistic defined in [13] under arbitrary dependence. We consider decision rules of the form $\boldsymbol{\delta}(\mathbf{T}, c) = (\delta(T_1, c), \dots, \delta(T_m, c))$, where

$$\delta(T_i, c) = \begin{cases} 1 & \text{if } T_i \leq c \\ 0 & \text{if } T_i > c, \end{cases} \tag{2.9}$$

with c being such that $m\mathrm{FDR}^*(\boldsymbol{\delta}(\mathbf{T}, c)) \leq \alpha$. In this paper we assume that $\mathbf{X}$ is continuous and hence $m\mathrm{FDR}^*(\boldsymbol{\delta}(\mathbf{T}, c))$ is continuous in c. Before we state our oracle procedure more explicitly in terms of the distributions of T_i's, we give the following proposition asserting the existence of such a c.

Proposition 2.1. *For the decision rule in (2.9) with T_i defined in (2.8), $m\mathrm{FDR}^*(\boldsymbol{\delta}(\mathbf{T}, c))$ is non-decreasing in c.*

We will prove this proposition by making use of the following two lemmas.

Lemma 2.1. *Consider the ratio of expectations $E_{H_1}[\delta(T, c)]/E_{H_0}[\delta(T, c)]$, for any random variable T having distribution H_1 in the numerator and distribution H_0 in the denominator. It is non-decreasing (non-increasing) in $c > 0$, if $dH_1(t)/dH_0(t)$ is non-decreasing (non-increasing) in t.*

Proof. The ratio can be expressed as the expectation, $E_{H_c^*}\varphi(T)$, of the non-decreasing function $\varphi(T) = dH_1(T)/dH_0(T)$, where H_c^* is such that

$$dH_c^*(t) = \delta(t, c)dH_0(t)/E_{H_0}[\delta(T, c)].$$

Since $\delta(t, c)$ is totally positive of order two (TP_2) in (t, c), that is, it satisfies the inequality

$$\delta(t, c)\ \delta(t', c') \geq \delta(t, c')\ \delta(t', c), \quad \forall t < t',\ c < c',$$

the lemma follows from the following result [8]: The expectation of a non-decreasing (non-increasing) function of a random variable $Y \sim g(y, \theta)$, with $g(y, \theta)$ being TP_2 in (y, θ), is non-decreasing (non-increasing) in θ. □

Remark 2.2. Sun and Cai [12] derived the above result for the collection of decisions based on the test statistics satisfying the MLR condition. Note that our proof, which is different, does not rely on any such condition.

Lemma 2.2. *Given two distributions $f_0(\mathbf{x})$ and $f_1(\mathbf{x})$ of a random vector $\mathbf{X}$, define $T(\mathbf{X}) = af_0(\mathbf{X})/\{af_0(\mathbf{X}) + bf_1(\mathbf{X})\}$, for any constants $a, b > 0$. Let $H_i(t) = P_{f_i}(T(\mathbf{X}) \leq t),\ 0 < t < 1$, for $i = 0, 1$. Then, $dH_1(t)/dH_0(t) = a(1 - t)/bt$.*

Proof. Since

$$[T(\mathbf{X}) - t][I(T(\mathbf{X}) \leq t) - I(T(\mathbf{X}) \leq t \pm \epsilon)] \leq 0, \quad \forall 0 < t < 1,\ \epsilon > 0,$$

by taking expectations of both sides in this inequality with respect to

$$\mathbf{X} \sim \frac{a}{a+b}f_0(\mathbf{x}) + \frac{b}{a+b}f_1(\mathbf{x}),$$

we have

$$a(1-t)[H_0(t) - H_0(t \pm \epsilon)] \leq bt[H_1(t) - H_1(t \pm \epsilon)], \quad \forall 0 < t < 1,\ \epsilon > 0.$$

The desired result then follows by letting $\epsilon \to 0$. □

Proof of Proposition 2.1. Let $G_{i,\boldsymbol{\mu}}^{(j)}$ denote the conditional distribution of $T_i(\mathbf{X})$ given $\theta_i = j$ and $\boldsymbol{\mu}$, for $j = 0, 1$. Then, from (2.3), we note that

$$m\text{FDR}^*(\boldsymbol{\delta}(\mathbf{T}, c)) = \frac{\pi_0 \sum_{i=1}^{m} G_{i,0}(c)}{\pi_0 \sum_{i=1}^{m} G_{i,0}(c) + (1-\pi_0) \sum_{i=1}^{m} G_{i,1}(c)},$$

where

$$G_{i,0}(c) = \int G_{i,\boldsymbol{\mu}}^{(0)}(c) h(\boldsymbol{\mu}|\theta_i = 0) d\boldsymbol{\mu}$$

$$\text{and} \quad G_{i,1}(c) = \int s(\mu_i) G_{i,\boldsymbol{\mu}}^{(1)}(c) h(\boldsymbol{\mu}|\theta_i = 1) d\boldsymbol{\mu},$$

with $h(\boldsymbol{\mu}|\theta_i = 0)$ and $h(\boldsymbol{\mu}|\theta_i = 1)$ representing the joint distribution of $\boldsymbol{\mu}$ conditionally given $\theta_i = 0$ and $\theta_i = 1$, respectively.

$$\begin{aligned}
\frac{1 - m\text{FDR}^*(\boldsymbol{\delta}(\mathbf{T}, c))}{m\text{FDR}^*(\boldsymbol{\delta}(\mathbf{T}, c))} &= \frac{E\left[\sum_{i=1}^{m} \delta(T_i, c)\theta_i \omega^*(\theta_i, \mu_i)\right]}{E\left[\sum_{i=1}^{m} \delta(T_i, c)(1-\theta_i)\omega^*(\theta_i, \mu_i)\right]} \\
&= \frac{E\left[\sum_{i=1}^{m} \delta(T_i, c) s(\mu_i) I(\theta_i = 1)\right]}{E\left[\sum_{i=1}^{m} \delta(T_i, c) I(\theta_i = 0)\right]} \\
&= \frac{1-\pi_0}{\pi_0}\left(\frac{1}{m}\sum_{i=1}^{m} \beta_i\right) \frac{E_{G_1}\left[\delta(T, c)\right]}{E_{G_0}\left[\delta(T, c)\right]},
\end{aligned} \tag{2.10}$$

where $G_1(t) = \sum_{i=1}^{m} w_i \tilde{G}_{i,1}(t)$, $G_0(t) = \frac{1}{m}\sum_{i=1}^{m} G_{i,0}(t)$, $\tilde{G}_{i,1}(t) = G_{i,1}(t)/\beta_i$, and $w_i = \beta_i / \sum_{j=1}^{m} \beta_j$, with $\beta_i = \int s(\mu_i) h(\boldsymbol{\mu}|\theta_i = 1) d\boldsymbol{\mu}$. The proposition will be proved from Lemma 2.1 if we can show that $dG_1(t)/dG_0(t)$ is a non-increasing function of t, since the left hand side of proposition (2.10) is a decreasing function of $m\text{FDR}^*(\boldsymbol{\delta}(\mathbf{T}, c))$.

Since $T_i(\mathbf{X}) = \pi_0 f_{i,0}(\mathbf{X}) / \{\pi_0 f_{i,0}(\mathbf{X}) + (1-\pi_0)\beta_i f_{i,1}^*(\mathbf{X})\}$, and $G_{i,0}$ and $\tilde{G}_{i,1}$ are the cdf's of $T_i(\mathbf{X})$ under the distributions $f_{i,0}(\mathbf{x}) = f(\mathbf{x} \mid \theta_i = 0)$ and

$$f_{i,1}^*(\mathbf{x}) = \frac{1}{\beta_i} \int s(\mu_i) f(\mathbf{x} \mid \theta_i = 1, \boldsymbol{\mu}) h(\boldsymbol{\mu}|\theta_i = 1) d\boldsymbol{\mu},$$

respectively, we see from Lemma 2.2 that $d\tilde{G}_{i,1}(t) = \frac{\pi_0}{(1-\pi_0)\beta_i}\left(\frac{1}{t} - 1\right) dG_{i,0}(t)$, for any $0 < t < 1$. Thus,

$$\begin{aligned}
\left(\sum_{i=1}^{m} \beta_i\right) dG_1(t) &= \sum_{i=1}^{m} \beta_i d\tilde{G}_{i,1}(t) \\
&= \sum_{i=1}^{m} \frac{\beta_i \pi_0 (1-t)}{\beta_i (1-\pi_0) t} dG_{i,0}(t) = \frac{m\pi_0}{1-\pi_0}\left(\frac{1}{t} - 1\right) dG_0(t),
\end{aligned}$$

implying that $dG_1(t)/dG_0(t)$ is non-increasing in $t \in (0, 1)$, as desired. Thus, the proposition is proved. □

Given Proposition 2.1, we are now ready to define our oracle procedure in the following:

Definition 2.1 (*The Oracle Procedure*)**.** Consider the multiple testing procedure $\boldsymbol{\delta}(\mathbf{T}, c^*)$, where

$$c^* = \sup\left\{t : m\text{FDR}^*(\boldsymbol{\delta}(\mathbf{T}, t)) \le \alpha\right\}. \tag{2.11}$$

This is a generalized version of the oracle procedure of Sun and Cai [12]. It is developed not only under any dependence structure among $(\mathbf{X}, \boldsymbol{\mu})$ but also it allows the alternatives to vary across tests and each type II error to be weighted by a measure of severity. Moreover, for its optimality, any specific property, like the monotone likelihood ratio property that Sun and Cai [12] assumed for the underlying test statistics, is not required.

Remark 2.3. Let $fdr_i(\mathbf{X}) = P(\theta_i = 0|\mathbf{X})$ and $d_i(\mathbf{X}) = fdr_i(\mathbf{X})/T_i(\mathbf{X})$. Then, it is to be noted that the $m\mathrm{FDR}^*(\boldsymbol{\delta}(\mathbf{T}, t))$ can be expressed as follows:

$$m\mathrm{FDR}^*(\boldsymbol{\delta}(\mathbf{T}, t)) = \frac{\sum_{i=1}^{m} E\left[I(T_i(\mathbf{X}) < t) fdr_i(\mathbf{X})\right]}{\sum_{i=1}^{m} E\left[I(T_i(\mathbf{X}) < t) fdr_i(\mathbf{X}) + I(T_i(\mathbf{X}) < t)(1 - fdr_i(\mathbf{X}))w_i(\mathbf{X})\right]}$$

$$= \frac{\sum_{i=1}^{m} E\left[I(T_i(\mathbf{X}) < t) T_i(\mathbf{X}) d_i(\mathbf{X})\right]}{\sum_{i=1}^{m} E\left[I(T_i(\mathbf{X}) < t) d_i(\mathbf{X})\right]}. \tag{2.12}$$

3. Numerical studies related to the oracle procedure

We carried out three numerical studies to see how our procedure in its oracle form compares with its relevant competitors for the problem of testing $\mu_i = 0$ against $\mu_i \neq 0$, $i = 1, \ldots, m$, with $s(\mu) = \mu^2$, under the following model:

$$\begin{aligned} &\mathbf{X} \mid \boldsymbol{\mu}, \boldsymbol{\theta} \sim N_m(\boldsymbol{\mu}, \Sigma), \quad \text{with } \Sigma = (1-\rho)I_m + \rho 1_m 1_m' \\ &\mu_i \mid \theta_i \overset{\text{ind}}{\sim} (1-\theta_i) I(\mu_i = 0) + \theta_i h(\mu_i) \\ &\theta_i \overset{\text{i.i.d.}}{\sim} Bernoulli(1 - \pi_0). \end{aligned} \tag{3.1}$$

In the first two numerical studies, we assume $\rho = 0$ so that (X_i, μ_i, θ_i), $i = 1, \ldots, m$ are independently distributed, while in the third numerical study, we assume the dependence case with $\rho > 0$.

Often a multiple testing procedure can be seen as first ranking the hypotheses according to a measure of significance, based on some test statistic, p-value, or local fdr, before choosing a cut-off point for the significance measure to determine which hypotheses are to be declared significant subject to control over a certain error rate, such as FDR or mFDR, at a specified level. Such ranking plays an important role in a procedure's performance, and can itself be used as a basis to compare with another procedure controlling a different error rate. More specifically, between two procedures providing the same number of discoveries, the one with better ranking should provide more true discoveries. The first numerical study was designed to make such ranking comparison between the Sun and Cai [12] and our oracle procedures that control two different measures of false discoveries, even though one is a generalized version of the other.

The second numerical study was conducted in light of Theorem 2.2, that is, to investigate if the mFNR* is indeed the smallest for our procedure in comparison with others that control the mFDR*, the procedure based on Lfdr scores [12] and the p-value based procedure.

In the third simulation study, we provide some insight into the performance of our oracle procedure under dependence. Although we have theoretically proved the existence and the optimality property of our proposed procedure under arbitrary dependence, the calculations of $w_i(\mathbf{X})$'s and the test statistics $T_i(\mathbf{X})$'s require the knowledge of the dependence structure. The construction of data-driven version of our procedure, particular under dependence, presents additional complexities and newer theoretical challenges that require a careful and special attention in a separate communication. In this numerical study, we investigate how well our oracle procedure derived without accounting for dependence information performs when dependence does indeed exist.

Towards understanding what significance measure is being used to rank the hypotheses in our procedure, we note that under the independence case of model (3.1) ($\rho = 0$), the $m\mathrm{FDR}^*(\boldsymbol{\delta}(\mathbf{T}, t))$ given in Remark 2.3 reduces to the following:

$$mFDR^*(\delta(\mathbf{T}, t)) = \frac{E\left(I(T(\mathbf{X}) \leq t) T(\mathbf{X}) d(\mathbf{X})\right)}{E\left(I(T(\mathbf{X}) \leq t) d(\mathbf{X})\right)},$$

with $T(\mathbf{X}) \equiv T_1(\mathbf{X})$ and $d(\mathbf{X}) \equiv d_1(\mathbf{X})$. The numerator and denominator expectations in the above ratio can be approximated (for large m) by $\frac{1}{m}\sum_{i=1}^{m}(I(T_i(\mathbf{X}) \leq t) T_i(\mathbf{X}) d_i(\mathbf{X}))$ and $\frac{1}{m}\sum_{i=1}^{m}(I(T_i(\mathbf{X}) \leq t) d_i(\mathbf{X}))$, respectively, resulting in an estimate of $mFDR^*(\delta(\mathbf{T}, t))$ at t as follows:

$$\widehat{mFDR^*}(\delta(\mathbf{T}, t)) = \frac{\sum_{i=1}^{m} I(T_i(\mathbf{X}) \leq t) T_i(\mathbf{X}) d_i(\mathbf{X})}{\sum_{i=1}^{m} I(T_i(\mathbf{X}) \leq t) d_i(\mathbf{X})}.$$

Let $T_{(1)}, \ldots, T_{(m)}$ be the ordered versions of $T_1(\mathbf{X}), \ldots, T_m(\mathbf{X})$, and $H_{(i)}$ and $d_{(i)}(\mathbf{X})$ be respectively the null hypothesis and the d-value corresponding to $T_{(i)}(\mathbf{X})$. Then, our oracle procedure can be described approximately as follows:

Find

$$k = \max\left\{ j : \frac{\sum_{i=1}^{j} T_{(i)}(\mathbf{X}) d_{(i)}(\mathbf{X})}{\sum_{i=1}^{j} d_{(i)}(\mathbf{X})} \le \alpha \right\}, \tag{3.2}$$

and reject $H_{(i)}$ for all $i = 1, \ldots, k$.

In other words, our procedure can be seen as ranking the hypotheses according to the increasing values of $T_i(\mathbf{X})$, the Glfdr scores corresponding to the H_i's, before determining the cut-off point $t \in \{T_{(1)}(\mathbf{X}), \ldots, T_{(m)}(\mathbf{X})\}$ to control the mFDR*; whereas, the Sun–Cai oracle procedure does the same in terms of the Lfdr scores to control the mFDR.

3.1. Numerical study 1

We considered using a measure of non-discoveries to compare the rankings provided by the Sun–Cai and our oracle procedures. More specifically, we wanted to see how these procedures compare in terms of not discovering the *most important* signals (i.e., the signals that are truly and highly significant), given the same number of discoveries made by each of them. The measure of non-discoveries is defined with weights assigned to the signals according to their magnitudes using our chosen severity function $s(\mu) = \mu^2$ to capture these *most important* signals with greater certainty.

With that in mind, we generated $m = 1000$ observations according to the model (3.1). Here we chose $\rho = 0, \pi_0 = 0.95$ and

$$h(\mu_i) = \pi_{11} N(\mu_-, \tau^2) + \pi_{12} N(\mu_+, \tau^2),$$

with $\pi_{11} = 0.2$, $\mu_- = -1.5$, $\mu_+ = 1$, and $\tau = 0.5$. We then calculated the values of Glfdr given in (2.8), which can be written for this model as $Glfdr_i = \frac{\pi_0 \phi(x_i)}{\pi_0 \phi(x_i) + \pi_1 H(x_i)}$ with

$$\begin{aligned} H(x_i) = {} & \pi_{11}\left[\frac{1}{\sqrt{1+\tau^2}}\phi\left(\frac{x_i - \mu_-}{\sqrt{1+\tau^2}}\right)\frac{\tau^2}{1+\tau^2} + \frac{(\tau^2 x_i + \mu_-)^2}{(1+\tau^2)}\right] \\ & + \pi_{12}\left[\frac{1}{\sqrt{1+\tau^2}}\phi\left(\frac{x_i - \mu_+}{\sqrt{1+\tau^2}}\right)\frac{\tau^2}{1+\tau^2} + \frac{(\tau^2 x_i + \mu_-)^2}{(1+\tau^2)}\right]. \end{aligned} \tag{3.3}$$

We ordered these values of Glfdr increasingly as $Glfdr_{(1)} \le \cdots \le Glfdr_{(m)}$. Let $H_{(i)}$ be the null hypothesis corresponding to $Glfdr_{(i)}$, for $i = 1, \ldots, m$. For each given $R = 1, 2, \ldots, m$, we marked the first R null hypothesis to be rejected and the rest to be accepted. With $\theta_{(i)} = 0$ or 1 indicating whether the null hypothesis $H_{(i)}$ is true or false (with $\mu_{(i)}$ being the true signal), respectively, we then calculated the weighted type II errors $\sum_{j=R+1}^{m} \theta_{(j)} \mu_{(j)}^2$. We replicated these steps 2000 times and averaged the 2000 values of the weighted type II errors before obtaining the simulated value of $\beta^*(R)$, the expected weighted type II errors (or non-discoveries) given R rejections (or discoveries). The red curve in Fig. 1 shows the plot of $\beta^*(R)$ against R. The similar plot was obtained for the Lfdr score and is shown using the green curve in this figure. As seen from this figure, between the Sun–Cai and our oracle procedures, ours can potentially be more powerful in the sense of producing a smaller amount of weighted type II errors associated with missing the *most important* signals.

3.2. Numerical study 2

In this numerical study, we again consider model (3.1) with $\rho = 0$ and $h(\mu_i) = \pi_{11} N(\mu_+, \tau^2) + \pi_{12} N(\mu_- \tau^2)$ but now we chose $\mu_+ = 4$, $\mu_- = -1$, $\tau = 0.5$ and $\pi_0 = 0.95$. The rejection region for our oracle procedure is $\{X_i : X_i \le c_l \text{ or } X_i \ge c_u\}$ for each H_i, with the cut-offs c_l and c_u being determined as in the following steps:

(i) For a given $0 < t < 1$, solve the following equation for z to obtain $c_l^{(t)}$ and $c_u^{(t)}$:

$$t\pi_1 H(z) - \pi_0(1-t)\phi(z) = 0$$

where $H(\cdot)$ is defined in (3.3).

(ii) Calculate

$$mFDR^* = \frac{\pi_0 \Psi(c_l^{(t)}, c_u^{(t)})}{\pi_0 \Psi(c_l^{(t)}, c_u^{(t)}) + \pi_1 \left\{ \pi_{11} E_{\mu_1}\left[\mu_1^2 \Psi(c_l^{(t)} - \mu_1, c_u^{(t)} - \mu_1)\right] + \pi_{12} E_{\mu_2}\left[\mu_2^2 \Psi(c_l^{(t)} - \mu_2, c_u^{(t)} - \mu_2)\right]\right\}},$$

where $\mu_1 \sim N(\mu_-, \tau^2)$, $\mu_2 \sim N(\mu_+, \tau^2)$, and $\Psi(c_l^{(t)}, c_u^{(t)}) = 1 - \Phi(c_u^{(t)}) + \Phi(c_l^{(t)})$, with Φ being the cdf of $N(0, 1)$.

(iii) Repeat the above two steps until we find t^* such that the mFDR* is α.

(iv) c_l and c_u are then determined as $c_l^{(t^*)}$ and $c_u^{(t^*)}$.

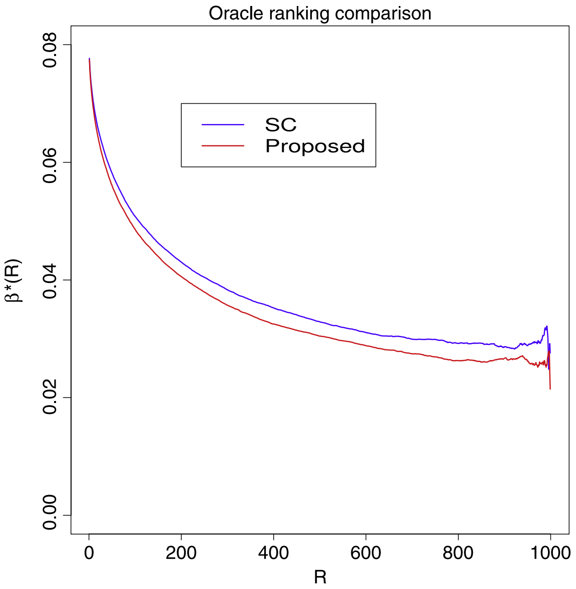


Fig. 1. Simulated average weighted type II errors.

Once c_l and c_u are determined, the mFNR* of the oracle procedure is calculated as follows:

$$mFNR^* = \frac{\pi_1\left\{\pi_{11}E_{\mu_1}\left\{\mu_1^2[1-\Psi(c_l-\mu_1, c_u-\mu_1)]\right\}+\pi_{12}E_{\mu_2}\left\{\mu_2^2[1-\Psi(c_l-\mu_2, c_u-\mu_2)]\right\}\right\}}{\pi_0[1-\Psi(c_l, c_u)]+\pi_1\left\{\pi_{11}E_{\mu_1}\left\{\mu_1^2[1-\Psi(c_l-\mu_1, c_u-\mu_1)]\right\}+\pi_{12}E_{\mu_2}\left\{\mu_2^2[1-\Psi(c_l-\mu_2, c_u-\mu_2)]\right\}\right\}}.$$

For the p-value based procedure, $c_l^{(t)}$ and $c_u^{(t)}$ in the step (i) are respectively the $t/2$th and $(1-t/2)$th quantiles of $N(0, 1)$ distribution. For the Lfdr based procedure, $c_l^{(t)}$ and $c_u^{(t)}$ are obtained as in our proposed procedure with $s(\cdot) = 1$. Then for both procedures, we follow the same steps (ii)–(iv) as above to determine the corresponding c_l and c_u, and mFNR*.

Fig. 2 compares the mFNR* values of these three procedures across different values of π_{11} and for some values of α. Our proposed procedure does indeed have the smallest mFNR* among these three procedures.

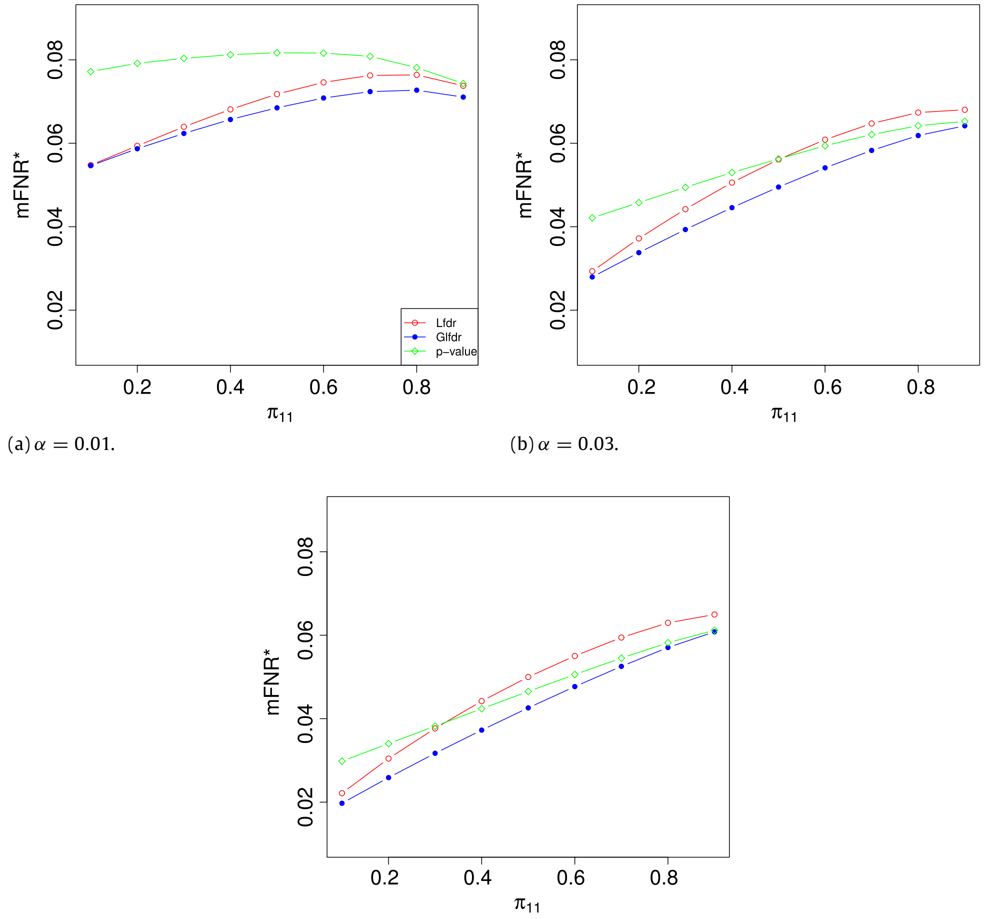
3.3. Numerical study 3

Here, we generated an observation $\mathbf{X} = (X_1, \ldots, X_m)$ according to Model (3.1) with $h(\mu_i) = 0.2N(\mu_+, \tau^2) + 0.8N(\mu_-, \tau^2)$, where $m = 1000$, $\mu_+ = 4$, $\mu_- = -1$, $\tau = 0.5$, $\pi_0 = 0.95$, and ρ is chosen from $\{0.1, 0.2, \ldots, 0.9\}$, and then applied the above three oracle procedures. We replicated the above step 1000 times to obtain the simulated mFDR* and mFNR* of the three procedures. From Fig. 3, we see that, for all ρ values, the mFDR* is still controlled by all three procedures and our proposed procedure still has the lowest mFNR*. These results suggest that even when we do not make use of the dependence information in deriving our oracle procedure, it can still be valid under positive dependence and be more powerful than its relevant competitors.

4. Concluding remarks

The decision theoretic approach to a multiple testing problem is not new. Other relevant work includes Sarkar et al. [10] and Peña et al. [9]. Nevertheless, the idea of incorporating the severity of type II errors has not been fully explored previously in the literature. We have developed the theory behind our idea from a compound decision theoretic point of view considering a loss function that incorporates the type II error severity. The consideration of type II error severity into the loss function allows us to re-formulate the work of Sun and Cai [12] in a more general framework involving newer, generalized forms of marginal false discovery and marginal false non-discovery rates. Newer theoretical results generalizing and often improving the existing ones are given in this process. We now have the theory for developing a much wider class of multiple testing procedures constructed from a decision theoretic point of view. Some of the newer methods in this class, those corresponding to non-constant type II error severity, are seen to have better performance in their oracle forms, as shown in our numerical studies, than those with constant type II error severity (i.e., those in [12] and some standard p-value based procedures).

The idea of weighting hypotheses or p-values while developing multiple testing methods in an FDR but non-decision theoretic framework has been proposed before. Benjamini and Hochberg [2] considered weighting the hypotheses in the original definition of the FDR to define the weighted FDR and proposed a weighted version of their 1995 FDR controlling

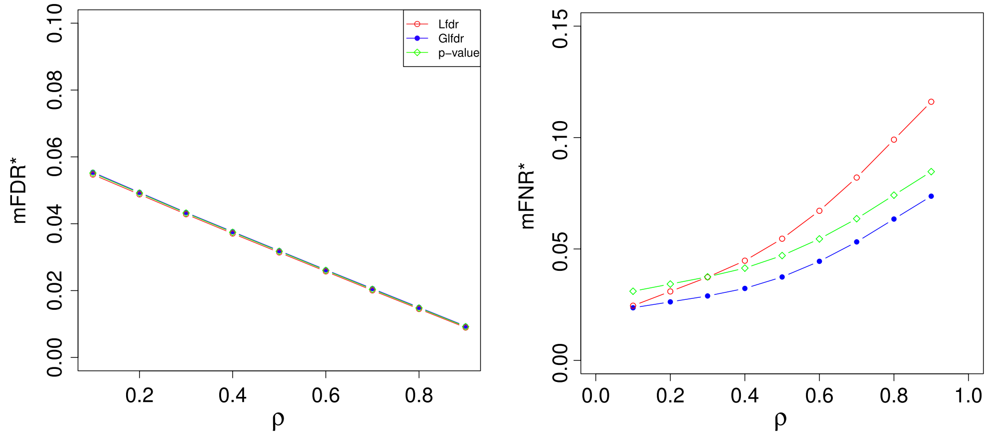


(a) $\alpha = 0.01$.

(b) $\alpha = 0.03$.

(c) $\alpha = 0.05$.

Fig. 2. Comparison of the three procedures with exact mFDR* control under independence: the oracle procedure based on the Glfdr (Blue), the oracle procedure based on the Lfdr statistic (red), and the p-value based procedure (Green) based on Model (3.1) with $\pi_0 = 0.95$, π_{11} varying from 0 to 1, $\mu_- = -1$, and $\mu_+ = 4$.



(a) mFDR*.

(b) mFNR*.

Fig. 3. Comparison of the three procedures with exact mFDR* control under independence for equi-correlated dependence structure: the oracle procedure based on the Glfdr (Blue), the oracle procedure based on the Lfdr statistic (red), and the p-value based procedure (Green) based on the model given in (3.1) with $\pi_0 = 0.95$, $h(\mu_i) = 0.2N(4, 0.25) + 0.8N(-1, 0.25)$ and $\rho \in \{0.1, 0.2, \dots, 0.9\}$. The significance level α is chosen to be 0.05.

method, the so-called BH method, that controls this weighted FDR. Genovese et al. [6], on the other hand, weighted each p-value and developed a BH type method controlling the usual FDR based on these weighted p-values. Our concern in this paper has been to define weighted versions of not only the marginal FDR but also the marginal FNR from their original definitions before providing a theoretical framework for the development of our procedure. Our approach to defining weighted mFDR and weighted mFNR is similar to Benjamini and Hochberg [2]. We attach weights to the hypotheses, although they are chosen to effectively act only on the false nulls. More specifically, we have

$$\text{mFDR}^*(\boldsymbol{\delta}(\mathbf{T}, c)) = \frac{E\left[\sum_{i=1}^{m} I(T_i < c, \theta_i = 0)\right]}{E\left[\sum_{i=1}^{m} I(T_i < c, \theta_i = 0) + \sum_{i=1}^{m} I(T_i < c, \theta_i = 1)s(\mu_i)\right]},$$

and

$$\text{mFNR}^*(\boldsymbol{\delta}(\mathbf{T}, c)) = \frac{E\left[\sum_{i=1}^{m} I(T_i > c, \theta_i = 1)s(\mu_i)\right]}{E\left[\sum_{i=1}^{m} I(T_i > c, \theta_i = 1)s(\mu_i) + \sum_{i=1}^{m} I(T_i > c, \theta_i = 0)\right]}.$$

The weight is assigned to a false null hypothesis according to its signal strength. It does not depend on whether acceptance or rejection of the false null contributes to a measure of false non-discoveries or false discoveries in the form of a penalty or boon. It is important to point out that our weights for all the hypotheses do not add up to m, contrary to what one might conclude from Benjamini and Hochberg [2]. In fact, a careful study of Benjamini and Hochberg [2] would reveal that such a restriction on the weights is not necessary in their paper, even though they have assumed it.

Derivation of an optimal multiple testing procedure incorporating type II error severity in its oracle form has been our primary focus in this paper. Now that we have this oracle procedure, a data-driven version of it with similar optimal property can potentially be constructed. However, construction of such an optimal data-driven procedure depends heavily on the underlying model and the chosen severity function, requiring newer efforts and techniques. We therefore leave this for a future communication. Also, a more comprehensive study of the procedure in terms of its sensitivity under varying choice of the severity function is also on our agenda for future research.

Acknowledgments

The authors thank the referee who made several useful suggestions. The research of Li He is supported by Merck Research Fellowship. Sanat K. Sarkar's research is supported by NSF Grants DMS-1006344 and DMS-1208735. Zhigen Zhao's research is supported by NSF[http://dx.doi.org/10.13039/100000001] Grant DMS-1208735.

References

[1] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B 57 (1) (1995) 289–300.
[2] Y. Benjamini, Y. Hochberg, Multiple hypotheses testing with weights, Scand. J. Statist. 24 (3) (1997) 407–418.
[3] D.B. Duncan, A Bayesian approach to multiple comparisons, Technometrics 7 (1965) 171–222.
[4] B. Efron, Large-scale simultaneous hypothesis testing, J. Amer. Statist. Assoc. 99 (465) (2004) 96–104.
[5] B. Efron, Large-scale Inference, Empirical Bayes Methods for Estimation, Testing, and Prediction, Cambridge University Press, 2010.
[6] C.R. Genovese, K. Roeder, L. Wasserman, False discovery control with p-value weighting, Biometrika 93 (3) (2006) 509–524.
[7] C. Genovese, L. Wasserman, Operating characteristics and extensions of the false discovery rate procedure, J. R. Stat. Soc. Ser. B 64 (3) (2002) 499–517.
[8] S. Karlin, Y. Rinott, Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions, J. Multivariate Anal. 10 (4) (1980) 467–498.
[9] E.A. Peña, J.D. Habiger, W. Wu, Power-enhanced multiple decision functions controlling family-wise error and false discovery rates, Ann. Statist. 39 (1) (2011) 556–583.
[10] S.K. Sarkar, T. Zhou, D. Ghosh, A general decision theoretic formulation of procedures controlling FDR and FNR from a Bayesian perspective, Statist. Sinica 18 (3) (2008) 925–945.
[11] J.G. Scott, J.O. Berger, An exploration of aspects of Bayesian multiple testing, J. Statist. Plann. Inference 136 (7) (2006) 2144–2162.
[12] W. Sun, T.T. Cai, Oracle and adaptive compound decision rules for false discovery rate control, J. Amer. Statist. Assoc. 102 (479) (2007) 901–912.
[13] W. Sun, T.T. Cai, Large-scale multiple testing under dependence, J. R. Stat. Soc. Ser. B 71 (2) (2009) 393–424.
[14] R.A. Waller, D.B. Duncan, A Bayes rule for the symmetric multiple comparisons problems, J. Amer. Statist. Assoc. 64 (1969) 1484–1503.