View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com

ScienceDirect

Journal of Multivariate Analysis 99 (2008) 437–450

Journal of Multivariate Analysis

www.elsevier.com/locate/jmva

Simultaneous control of false positives and false negatives in multiple hypotheses testing

Nader Ebrahimi*

Division of Statistics, Northern Illinois University, DeKalb, IL 60115, USA

Received 17 July 2006

Available online 9 January 2007

Abstract

Multiple hypotheses testing is concerned with appropriately controlling the rate of false positives, false negatives or both when testing several hypotheses simultaneously. Nowadays, the common approach to testing multiple hypotheses calls for controlling the expected proportion of falsely rejected null hypotheses referred to as the false discovery rate (FDR) or suitable measures based on the positive false discovery rate (pFDR). In this paper, we consider the problem of determining levels that both false positives and false negatives can be controlled simultaneously. As our risk function, we use the expected value of the maximum between the proportions of false positives and false negatives, with the expectation being taken conditional on the event that at least one hypothesis is rejected and one is accepted, referred to as hybrid error rate (HER). We then develop, based on HER, an analog of p-value termed as h-value to test the individual hypotheses. The use of the new procedure is illustrated using the well-known public data set by Golub et al. [Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 386 (1999) 531–537] with Affymetrix arrays of patients with acute lymphoic leukemia and acute myeloid leukemia.

© 2007 Elsevier Inc. All rights reserved.

AMS 2000 subject classification: 62H15; 62C15

Keywords: False positives; False negatives; Step-up procedure

1. Introduction

In a single-hypothesis testing, suppose we wish to test a null hypothesis H_0 against an alternative H_a based on a test statistic Y. For a given rejection region Γ (the numerical values of the test statistic for which H_0 is rejected), if $Y \in \Gamma$ we reject H_0 and if $Y \notin \Gamma$ we accept H_0. Two types of decision errors occur. These two types of errors traditionally have been given the names, Types

* Fax: +1 815 753 6776.

E-mail address: nader@math.niu.edu.

0047-259X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.

doi:10.1016/j.jmva.2006.12.006

I and II errors. A Type I error occurs when $Y \in \Gamma$ but the null hypothesis is true and Type II error occurs when $Y \notin \Gamma$ but the null hypothesis is false. To choose a rejection region Γ, the most common approach is to consider all rejection regions that have a Type I error less than or equal to α. Then, among all these rejection regions select the one which has the lowest Type II error. Of course, this can be done if the alternative hypothesis is simple or uniformly most powerful (UMP) test exists.

Multiple testing has emerged as a statistical area of both practical and theoretical importance. The problem arises in many areas including neuroimaging, genomics, and astronomy. When testing multiple hypotheses, the situation becomes more complicated. Because for each test we have Types I and II errors and it becomes unclear how one should measure and control the overall error rate. In this paper, our goal is to present a measure of overall error rate. For simplicity of exposition, we use genomics terminology. We emphasize that our proposed measure can be applied to other areas as well.

Suppose we have microarray experiments which produce expression data on $m, m \geqslant 2$ genes (variables) for n samples (corresponding to individual microarray experiments). Let the data be arrayed as $m \times n$ matrix $X = (x_{ij}), i = 1, 2, \ldots, m, j = 1, \ldots, n$, with rows corresponding to genes and columns to individual microarray experiments. A typical microarray experiment measures several thousand genes simultaneously across different conditions. When testing for potential differential expression across those conditions, each gene is considered independently from one another. In other words, a t-test, ANOVA test or any other test is performed on each gene separately. The incidence of false positives (genes falsely called differentially expressed when they are not) or false negatives (genes falsely called not differentially expressed when they are) are dependent on the number of tests performed and the critical significance level (p-value cutoff).

Specifically, consider the problem of simultaneously testing m null hypotheses H_i, $i=1, \ldots, m$. Let $H_i = 0$ when the null hypothesis H_i is true, and $H_i = 1$ when the null hypothesis H_i is false. Here H_i is a statement made regarding gene $i, i = 1, \ldots, m$, and all m hypotheses are assumed to be known in advance.

For each gene, a null hypothesis is tested against the alternative hypothesis. Let S_0 and S_1 be the set of indices $1, \ldots, m$ corresponding to true and false null hypotheses, respectively, i.e., $S_0 = \{i : H_i = 0\}$ and $S_1 = \{i : H_i = 1\}$. Also, let $m_0 = |S_0|$ and $m_1 = |S_1| = m - m_0$ be the number of true and false null hypotheses, respectively. Here both S_0 and S_1 are unknown parameters. We can categorize the m tests in Table 1.

Putting $\frac{V}{R} = 0$ when $R = 0$ and $\frac{T}{m-R} = 0$ when $R = m$, we define the false discovery rate (FDR) and the false nondiscovery rate (FNR) by

$$FDR = E\left[\frac{V}{R} I(0 < R \leqslant m)\right] = E\left[\frac{V}{R}\middle| 0 < R \leqslant m\right] P(0 < R \leqslant m), \tag{1.1}$$

Table 1
Summary table for the multiple testing problem

	H_0 not rejected	H_0 rejected	Total
H_0 true	U	V	m_0
H_0 false	T	S	m_1
Total	$m - R$	R	m

and

$$FNR = E\left[\frac{T}{m-R}\, I(0 \leqslant R < m)\right] = E\left[\frac{T}{m-R}\,\middle|\, 0 \leqslant R < m\right] P(0 \leqslant R < m), \tag{1.2}$$

respectively. The FDR is the expected proportion of falsely rejected null hypotheses, and the FNR is the expected proportion of non-rejected null hypotheses that are incorrect. It should be mentioned that in Table 1, the number R of the rejected null hypotheses is an observable random variable, while U, V, T and S are unobservable random variables. When $m_0 = m$, then FDR $= P(0 < V \leqslant m_0) = P(V > 0)$. Note that in general $P(V > 0)$ is commonly referred to as familywise error rate (FWER), see Hochberg and Tamhane [12]. See also Benjamim and Hochberg [1] and Benjamin and Yekutieli [2] for more details about FNR and FDR. If the interest is to control Types I and II errors simultaneously, one can use the risk function of the form $bFNR + FDR$, $b \geqslant 0$, see Genovese and Wasserman [8,9].

Suppose we are interested in obtaining an error rate when positive (negative) findings have occurred, then the positive FDR, pFDR and positive FNR, pFNR are appropriate. Here,

$$pFDR = E\left[\frac{V}{R}\,\middle|\, 0 < R \leqslant m\right] \tag{1.3}$$

and

$$pFNR = E\left[\frac{T}{m-R}\,\middle|\, 0 \leqslant R < m\right], \tag{1.4}$$

see Storey [17,18].

In this work we propose an alternative way to measure error rate on the problem of multiple hypotheses testing. Our proposed loss function simply combines false discovery and false non-discovery rates committed in such problems and use a suitable measure of this combination termed as HER. The HER risk function is a non-linear function of pFNR and pFDR and is more conservative in the sense that our procedure accepts more than pFDR. However, it is less conservative in the sense that it rejects more than the risk function $FDR + bFNR$ for many choices of b. Note that HER is always greater than or equal to both pFNR and pFDR.

Step-up and Step-down procedures represent extremely popular approaches to multiple testing. These procedures are based on different risk functions including FWER, FDR, FNR, pFDR and pFNR. We refer you to Dudoit et al. [6], Ge et al. [7], Sarkar [15], Delongchamp et al. [5], Cohen and Sackrowitz [3,4] and many references cited there for more details. In this paper we propose a procedure based on HER, an analog of p-value to multiple testing.

After some preliminaries, in Section 3 we present a formal definition of our risk function, HER and focus on the theoretical development under the independence assumption and under that all test statistics of the non-null are identically distributed. The emphasis in this measure is on both error rates. Thus, an immediate but important consequence of controlling HER is that it controls both expected false discovery and expected false non-discovery rates. Imitating the definition of q-value which is derived from pFDR, the new quantity called the h-value derived from HER is introduced and investigated in Section 3. An algorithm is also provided for computing h-value. In Section 4, we illustrate h-value in an example and compare it with q-value via simulation. We make concluding remarks in Section 5. Throughout the paper we take $\frac{0}{0} = 0$. Here "decreasing" means non-increasing and "increasing" means non-decreasing.

2. Preliminary result

In this section we give the result that is used in Section 3.

Lemma 1.

(a) *Suppose X is Binomial (n, p) and Y is Binomial (n, p^*). If $p \leqslant p^*$, then $[X|a \leqslant X \leqslant b] \underset{\text{st}}{\leqslant} [Y|a \leqslant Y \leqslant b]$ wherever $a \leqslant b$. (In general, for any two random variables X and Y, $[X|a \leqslant X \leqslant b]$ is stochastically smaller than $[Y|a \leqslant Y \leqslant b]$ (denoted by $[X|a \leqslant X \leqslant b] \underset{\text{st}}{\leqslant} [Y|a \leqslant Y \leqslant b]$) if and only if $\frac{P(X>x)}{P(a \leqslant X \leqslant b)} \leqslant \frac{P(Y>x)}{P(a \leqslant Y \leqslant b)}$ for all x.)*

(b) *Suppose X is Binomial (n, p). Then $E(X|1 \leqslant X \leqslant n-1) = \frac{np(1-p^{n-1})}{1-p^n-(1-p)^n}$.*

Proof. (a) From Lehmann [14], we know that $X \underset{\text{lr}}{\leqslant} Y$. (In general, X is said to be smaller than Y in the likelihood ratio order (denoted by $X \underset{\text{lr}}{\leqslant} Y$) if $\frac{P(X=x)}{P(Y=x)}$ decreases over the union of the supports of X and Y.) Now, the part (a) easily follows from Theorem 1.C.2. of Shaked and Shanthikumar [16].

(b) It is clear that

$$E(X|1 \leqslant X \leqslant n-1) = \frac{\sum_{x=1}^{n-1} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}}{1-p^n-(1-p)^n} = \frac{np(1-p^{n-1})}{1-p^n-(1-p)^n}. \qquad \square$$

3. New way of defining error rate

In this section we propose an alternative quantity to the pFDR and pFNR, which we call hybrid error rate or simply HER.

When both positive and negative findings have occurred, then one way to control both false discovery and false non-discovery rates is to consider the maximum of these two error rates and define HER to be

$$HER = E\left[\max\left(\frac{V}{R}, \frac{T}{m-R}\right)\middle| 0 < R < m\right]. \tag{3.1}$$

Although our proposed measure is more conservative than pFDR and pFNR, an advantage is that by controlling HER we automatically control both pFNR and pFDR. Two properties of the risk function HER are easily shown, yet are very important.

(a) Using the facts that $\max\left(\frac{V}{R}, \frac{T}{m-R}\right) \underset{\text{st}}{\geqslant} \frac{V}{R}$ and $\max\left(\frac{V}{R}, \frac{T}{m-R}\right) \underset{\text{st}}{\geqslant} \frac{T}{m-R}$ we get that

$$\begin{aligned} HER &\geqslant \max\left[E\left(\frac{V}{R}\middle| 0 < R < m\right), E\left(\frac{T}{m-R}\middle| 0 < R < m\right)\right] \\ &= \max\left[pFDR, pFNR\right]. \end{aligned}$$

Therefore, control of HER implies control of both pFDR and pFNR. As a result, any procedure that controls this measure also controls the pFDR and pFNR.

(b) In Table 1, if $m_0 = m$, then $HER = E\left[\frac{V}{R}\middle| 0 < R < m\right] = 1$ and if $m = m_1$, then $HER = E\left[\frac{T}{m-R}\middle| 0 < R < m\right] = 1$.

Remark 1. If $P(R = m) = 1$, then we will simply take HER = pFDR. Also, if $P(R = 0) = 1$, then we will take HER = pFNR.

3.1. Computing HER

Suppose we wish to perform m identical hypothesis tests $H_1, \ldots, H_m$ based on the independent and identically distributed test statistics $Y_1, \ldots, Y_m$ and significance region Γ. More specifically, assume that $Y_i|H_i \overset{\text{i.i.d.}}{\sim} (1 - H_i)F_0 + H_i F_1$ and $H_i \overset{\text{i.i.d.}}{\sim}$ Bernoulli$(1 - \pi_0)$ for $i = 1, \ldots, m$ and some null distribution F_0 and alternative distribution F_1. Here, π_0 is a prior probability that the null hypothesis H_i is true and $\pi_1(\pi_1 = 1 - \pi_0)$ is the probability that H_i is false. That is, H_i is Bernoulli random variable with $P(H_i = 0) = \pi_0$ and $P(H_i = 1) = \pi_1, i = 1, \ldots, m$.

For a given common abstract rejection region Γ, define HER(Γ) as we defined in (3.1):

$$HER(\Gamma) = E\left[\max\left[\frac{V(\Gamma)}{R(\Gamma)}, \frac{T(\Gamma)}{m - R(\Gamma)}\right] \middle| 0 < R(\Gamma) < m\right], \tag{3.2}$$

where $V(\Gamma)$ = the number {null $Y_i : Y_i \in \Gamma$}, $R(\Gamma)$ = the number $\{Y_i : Y_i \in \Gamma\}$, $T(\Gamma)$ = the number {alternative $Y_i : Y_i \notin \Gamma$}. In expression (3.2), when $m = m_0$ or $m = m_1$, $HER(\Gamma) = 1$. Let $\eta_1(\Gamma) = pFDR(\Gamma)$ and $\eta_2(\Gamma) = pFNR(\Gamma)$. Then, $\eta_1(\Gamma) = E\left(\frac{V(\Gamma)}{R(\Gamma)}|0 < R(\Gamma) \leqslant m\right) = \sum_{k=1}^{m} E\left(\frac{V(\Gamma)}{R(\Gamma)}|R(\Gamma) = k\right) \times P\left(R(\Gamma) = k|0 < R(\Gamma) \leqslant m\right) = \sum_{k=1}^{m} \frac{1}{k} E\left(V(\Gamma)|R(\Gamma) = k\right) P\left(R(\Gamma) = k|0 < R(\Gamma) \leqslant m\right)$. Since the statistics are independent and identically distributed, it follows that $V(\Gamma)|R(\Gamma) = k$ is a binomial random variable with probability of success $P(H = 0|Y \in \Gamma)$. Thus, $\eta_1(\Gamma) = \sum_{k=1}^{m} k\left(P(H = 0|Y \in \Gamma)\right) \times \frac{1}{k} P(R(\Gamma) = k|0 < R(\Gamma) \leqslant m) = P(H = 0|Y \in \Gamma)$. Similar arguments can be used to show that $\eta_2(\Gamma) = E\left(\frac{T(\Gamma)}{m - R(\Gamma)}|0 \leqslant R(\Gamma) < m\right) = P(H = 1|Y \in \Gamma)$.

Define,

$p_1(\Gamma) = P$ (null hypothesis is true and it is rejected),
$p_2(\Gamma) = P$ (null hypothesis is true and it is not rejected),
$p_3(\Gamma) = P$ (null hypothesis is false and it is rejected), and
$p_4(\Gamma) = P$ (null hypothesis is false and it is not rejected).

Then, it is clear that $\frac{p_1(\Gamma)}{p_1(\Gamma) + p_3(\Gamma)} = P(H = 0|Y \in \Gamma) = \eta_1(\Gamma)$ and $\frac{p_2(\Gamma)}{p_2(\Gamma) + p_4(\Gamma)} = P(H = 1|Y \notin \Gamma) = \eta_2(\Gamma)$.

Now, one can prove the following lemma:

Lemma 2. *Conditional on $R(\Gamma) = k$, for any k and $m - k$ bigger than or equal to one*, (a) *$V(\Gamma)$ and $T(\Gamma)$ are independent*, (b) *$V(\Gamma)$ is Binomial $(k, \eta_1(\Gamma))$ and* (c) *$T(\Gamma)$ is Binomial $(m - k, \eta_2(\Gamma))$.*

Proof. From Table 1,

$$P(V(\Gamma) = v, T(\Gamma) = t|R(\Gamma) = k) = \frac{P(V(\Gamma) = v, T(\Gamma) = t, R(\Gamma) = k)}{P(R(\Gamma) = k)}$$

$$= \frac{P(V(\Gamma)=v, T(\Gamma)=t, S(\Gamma)=k-v, U(\Gamma)=m-k-t)}{P(R(\Gamma)=k)}$$

$$= \frac{\frac{m!}{v!t!(k-v)!(m-k-t)!}p_1^v(\Gamma)\,p_2^t(\Gamma)p_3^{k-v}(\Gamma)p_4^{m-k-t}(\Gamma)}{\frac{m!}{k!(m-k)!}(p_1(\Gamma)+p_3(\Gamma))^k(p_2(\Gamma)+p_4(\Gamma))^{m-k}}$$

$$= \binom{k}{v}\left(\frac{p_1(\Gamma)}{p_1(\Gamma)+p_3(\Gamma)}\right)^v\left(\frac{p_3(\Gamma)}{p_1(\Gamma)+p_3(\Gamma)}\right)^{k-v}\binom{m-k}{t}\left(\frac{p_2(\Gamma)}{p_2(\Gamma)+p_4(\Gamma)}\right)^t$$
$$\times\left(\frac{p_4(\Gamma)}{p_2(\Gamma)+p_4(\Gamma)}\right)^{m-k-t}$$

$$= P(V(\Gamma)=v|R(\Gamma)=k)P(T(\Gamma)=t|R(\Gamma)=k).$$

This completes the proof of part (a). Parts (b) and (c) follow easily from the fourth equation in the proof of part (a). □

From Eq. (3.2), one can write

$$HER(\Gamma)=\left[\sum_{k=1}^{m-1}E\left\{\max\left(\frac{V(\Gamma)}{k},\frac{T(\Gamma)}{m-k}\right)\middle|R(\Gamma)=k\right\}P(R(\Gamma)=k)\right]$$
$$\times(P(1\leqslant R(\Gamma)\leqslant m-1))^{-1}.$$

Using Lemma 2 and the fact that $\max(a,b)=\frac{a+b+|a-b|}{2}$ we get

$$HER(\Gamma)=\frac{1}{2}(\eta_1(\Gamma)+\eta_2(\Gamma))+\frac{1}{2}\left\{\sum_{k=1}^{m-1}\left\{\sum_{v=0}^{k}\sum_{t=0}^{m-k}\left|\frac{v}{k}-\frac{t}{m-k}\right|b_1(v,\eta_1(\Gamma),k)\right.\right.$$
$$\left.\left.\times\, b_2(t,\eta_2(\Gamma),m-k)\right\}P(R(\Gamma)=k)\right\}(P(1\leqslant R(\Gamma)\leqslant m-1))^{-1}, \qquad (3.3)$$

where $b_1(v,\eta_1(\Gamma),k)=\frac{k!}{v!(k-v)!}(\eta_1(\Gamma))^v(1-\eta_1(\Gamma))^{k-v}$, $b_2(t,\eta_2(\Gamma),m-k)=\frac{(m-k)!}{t!(m-k-t)!}(\eta_2(\Gamma))^t(1-\eta_2(\Gamma))^{m-k-t}$. Note that $R(\Gamma)$ has Binomial distribution with parameters m and $p_1(\Gamma)+p_3(\Gamma)$.

To simplify Eq. (3.3), we can write

$$\sum_{v=0}^{k}\sum_{t=0}^{m-k}\left|\frac{v}{k}-\frac{t}{m-k}\right|b_1(v,\eta_1(\Gamma),k)b_2(t,\eta_2(\Gamma),m-k)$$

$$=\sum_{v=0}^{k}\sum_{t=0}^{[\frac{m-k}{k}v]}\left(\frac{v}{k}-\frac{t}{m-k}\right)b_1(v,\eta_1(\Gamma),k)b_2(t,\eta_2(\Gamma),m-k)$$

$$-\sum_{v=0}^{k}\sum_{t=[\frac{m-k}{k}v]+1}^{m-k}\left(\frac{v}{k}-\frac{t}{m-k}\right)b_1(v,\eta_1(\Gamma),k)b_2(t,\eta_2(\Gamma),m-k)$$

$$=\eta_1(\Gamma)\sum_{v=0}^{k-1}b_1(v,\eta_1(\Gamma),k-1)B_2\left(\left[\frac{m-k}{k}(v+1)\right],\eta_2(\Gamma),m-k\right)$$

$$
\begin{aligned}
&-\eta_2(\Gamma)\sum_{v=0}^{k} b_1(v,\eta_1(\Gamma),k)B_2\left(\left[\frac{m-k}{k}v\right]-1,\eta_2(\Gamma),m-k-1\right)\\
&-\eta_1(\Gamma)\sum_{v=0}^{k-1} b_1(v,\eta_1(\Gamma),k-1)\left(1-B_2\left(\left[\frac{m-k}{k}(v+1)\right],\eta_2(\Gamma),m-k\right)\right)\\
&+\eta_2(\Gamma)\sum_{v=0}^{k} b_1(v,\eta_1(\Gamma),k)\left(1-B_2\left(\left[\frac{m-k}{k}v\right]-1,\eta_2(\Gamma),m-k-1\right)\right)\\
&=-2\eta_2(\Gamma)\sum_{v=0}^{k} b_1(v,\eta_1(\Gamma),k)B_2\left(\left[\frac{m-k}{k}v\right]-1,\eta_2(\Gamma),m-k-1\right)\\
&+2\eta_1(\Gamma)\sum_{v=0}^{k-1} b_1(v,\eta_1(\Gamma),k-1)B_2\left(\left[\frac{m-k}{k}(v+1)\right],\eta_2(\Gamma),m-k\right)\\
&-\eta_1(\Gamma)+\eta_2(\Gamma), \qquad (3.4)
\end{aligned}
$$

where $B_2(x,\eta_2(\Gamma),m-k)=\sum_{y=0}^{x} b_2(y,\eta_2(\Gamma),m-k)$ and $[x]$ is the largest integer less than or equal to x. Note that for $x<0$, $B_2(x,\eta_2(\Gamma),m-k)=0$.

Now, using Eq. (3.4), (3.3) reduces to

$$
\begin{aligned}
HER(\Gamma)=\eta_2(\Gamma)+\eta_1(\Gamma)&\left\{\sum_{k=1}^{m-1}\left\{\sum_{v=0}^{k-1} b_1(v,\eta_1(\Gamma),k-1)\right.\right.\\
&\left.\times B_2\left(\left[\frac{m-k}{k}(v+1)\right],\eta_2(\Gamma),m-k\right)\right\}P(R(\Gamma)=k)\Big\}\left(P(1\leqslant R(\Gamma)\leqslant m-1)\right)^{-1}\\
&-\eta_2(\Gamma)\left\{\sum_{k=1}^{m-1}\left\{\sum_{v=0}^{k} b_1(v,\eta_1(\Gamma),k)B_2\left(\left[\frac{m-k}{k}v\right]-1,\eta_2(\Gamma),m-k-1\right)\right\}\right.\\
&\left.\times\ P(R(\Gamma)=k)\right\}\left(P(1\leqslant R(\Gamma)\leqslant m-1)\right)^{-1}. \qquad (3.5)
\end{aligned}
$$

Note that $HER(\Gamma)$ in (3.5) is an expected risk or a Bayes risk with respect to the prior π_0. As an application of Eq. (3.5), for $m=2$, $HER(\Gamma)=\eta_1(\Gamma)+\eta_2(\Gamma)-\eta_1(\Gamma)\eta_2(\Gamma)$.

Remark 2. One can also write $HER(\Gamma)$ in Eq. (3.5) as

$$
\begin{aligned}
HER(\Gamma)=\eta_1(\Gamma)+\eta_2(\Gamma)&\left\{\sum_{k=1}^{m-1}\left\{\sum_{t=0}^{m-k-1} b_2(t,\eta_2(\Gamma),m-k-1)\right.\right.\\
&\left.\times\ B_1\left(\left[\frac{k}{m-k}(t+1)\right],\eta_1(\Gamma),k\right)\right\}P(R(\Gamma)=k)\Big\}\left(P(1\leqslant R(\Gamma)\leqslant m-1)\right)^{-1}\\
&-\eta_1(\Gamma)\left\{\sum_{k=1}^{m-1}\left\{\sum_{t=0}^{m-k} b_2(t,\eta_2(\Gamma),m-k)B_1\left(\left[\frac{k}{m-k}t\right]-2,\eta_1(\Gamma),k-1\right)\right\}\right.\\
&\left.\times\ P(R(\Gamma)=k)\right\}\left(P(0\leqslant R(\Gamma)\leqslant m-1)\right)^{-1}.
\end{aligned}
$$

Here $[x]$ means the smallest integer greater than or equal to x and $B_1(x, \eta_1(\Gamma), k) = \sum_{y=0}^{x} b_1(y, \eta_1(\Gamma), k)$. Note that $B_1(x, \eta_1(\Gamma), k) = 0$ for $x < 0$.

It is clear that $HER(\Gamma)$ in (3.5) is a non-linear function of $pFDR((\Gamma))$ and $pFNR((\Gamma))$.

Now, assume that F_0 and F_1 are continuous distributions with common support, with respective densities f_0 and f_1. Define the set of rejection regions $\{\beta_\lambda\}$ by

$$\beta_\lambda = \left\{ x : \frac{\pi_0 f_0(x)}{\pi_0 f_0(x) + \pi_1 f_1(x)} \leqslant \lambda \right\}, \tag{3.6}$$

where $0 \leqslant \lambda \leqslant 1$ and $\pi_1 = (1 - \pi_0)$. Let,

$$\lambda^* = \arg\min_{\lambda} [HER(\beta_\lambda)]. \tag{3.7}$$

We have the following result which is very useful for obtaining minimum of $HER(\Gamma)$ over all rejection regions Γ.

Theorem 1. *If for any λ, $E\left(\max\left(\frac{V(\beta_\lambda)}{k}, \frac{T(\beta_\lambda)}{m-k}\right)\Big| R(\beta_\lambda) = k\right)$ is decreasing in k, then given π_0, $HER(\beta_{\lambda^*})$ minimizes $HER(\Gamma)$ among all measurable Γ.*

Proof. Recall that by the Neyman–Pearson Lemma, the $\{\beta_\lambda\}$ form a set of MP (most powerful) rejection regions. For each $0 \leqslant \alpha \leqslant 1$, there exists a β_λ such that $P(Y \in \beta_\lambda \Big| H = 0) = \alpha$ (see [14]). Now consider any measurable Γ. Then, there exists a β_λ such that $P(Y \in \Gamma \Big| H = 0) = P(Y \in \beta_\lambda \Big| H = 0)$. Since $\{\beta_\lambda\}$ are MP, it follows that $P(Y \in \Gamma | H = 1) \leqslant P(Y \in \beta_\lambda | H = 1)$ and $P(Y \notin \beta_\lambda \Big| H = 1) + P(Y \in \beta_\lambda \Big| H = 0) \leqslant 1$. Consequently,

$$\begin{aligned}
\eta_1(\Gamma) &= \frac{\pi_0 P(Y \in \Gamma \Big| H = 0)}{\pi_0 P(Y \in \Gamma \Big| H = 0) + \pi_1 P(Y \in \Gamma \Big| H = 1)} \\
&\geqslant \frac{\pi_0 P(Y \in \beta_\lambda \Big| H = 0)}{\pi_0 P(Y \in \beta_\lambda \Big| H = 0) + \pi_1 P(Y \in \beta_\lambda \Big| H = 1)} = \eta_1(\beta_\lambda)
\end{aligned}$$

and

$$\begin{aligned}
\eta_2(\Gamma) &= \frac{\pi_1 P(Y \notin \Gamma \Big| H = 1)}{\pi_1 P(Y \notin \Gamma \Big| H = 1) + \pi_0 P(Y \notin \Gamma \Big| H = 0)} \\
&\geqslant \frac{\pi_1 P(Y \notin \beta_\lambda \Big| H = 1)}{\pi_1 P(Y \notin \beta_\lambda \Big| H = 1) + \pi_0 P(Y \notin \beta_\lambda \Big| H = 0)} = \eta_2(\beta_\lambda)
\end{aligned}$$

and $\eta_1(\beta_\lambda) + \eta_2(\beta_\lambda) \leqslant 1$.

Now, since $[V(\Gamma)\Big|R(\Gamma)=k] \sim$ Binomial $(k, \eta_1(\Gamma))$, from Lemma 1, part (a) it follows that $[V(\Gamma)\Big|R(\beta_\lambda)=k]$ is stochastically larger than $[V(\beta_\lambda)\Big|R(\beta_\lambda)=k]$. Similarly, since $[T(\Gamma)\Big|R(\Gamma)=k] \sim$ Binomial $(m-k, \eta_2(\Gamma))$, it follows that $[T(\Gamma)\Big|R(\Gamma)=k]$ is stochastically larger than $[T(\beta_\lambda)\Big|R(\beta_\lambda)=k]$. Also, $R(\Gamma) \sim$ Binomial $(m, \pi_0\alpha+\pi_1 P(Y\in\Gamma\Big|H=1))$ and $\pi_0\alpha+\pi_1 P(Y\in\Gamma\Big|H=1)\leqslant \pi_0\alpha+\pi_1 P(Y\in\beta_\lambda\Big|H=1)$. Thus, from Lemma 1, part (a), we get that $[R(\Gamma)\Big|1\leqslant R(\Gamma)\leqslant m-1]$ is stochastically less than $[R(\beta_\lambda)\Big|1\leqslant R(\beta_\lambda)\leqslant m-1]$.

Under our assumptions, using these properties and the fact that $\eta_1(\beta_\lambda)+\eta_2(\beta_\lambda)\leqslant 1$,

$$\begin{aligned}
HER(\Gamma) &= \sum_{k=1}^{m-1}\left[E\left(\max\left(\frac{V(\Gamma)}{k},\frac{T(\Gamma)}{m-k}\right)\Big|R(\Gamma)=k\right)\right]\frac{P(R(\Gamma)=k)}{P(1\leqslant R(\Gamma)\leqslant m-1)}\\
&\geqslant \sum_{k=1}^{m-1}\left[E\left(\max\left(\frac{V(\beta_\lambda)}{k},\frac{T(\beta_\lambda)}{m-k}\right)\Big|R(\beta_\lambda)=k\right)\right]\frac{P(R(\Gamma)=k)}{P(1\leqslant R(\Gamma)\leqslant m-1)}\\
&\geqslant \sum_{k=1}^{m-1}\left[E\left(\max\left(\frac{V(\beta_\lambda)}{k},\frac{T(\beta_\lambda)}{m-k}\right)\Big|R(\beta_\lambda)=k\right)\right]\frac{P\left(R(\beta_\lambda)=k\right)}{P(1\leqslant R(\beta_\lambda)\leqslant m-1)}\\
&= HER(\beta_\lambda).
\end{aligned}$$

This completes the proof. □

The following example gives an application of Theorem 1.

Example 1. Suppose $Y_i\Big|H_i \overset{\text{i.i.d.}}{\sim}(1-H_i)\cdot N(0,1)+H_i\cdot N(2,1)$ and $H_i \overset{\text{i.i.d.}}{\sim}$ Bernoulli(0.2), $i=1,2,3$. Also suppose we want to minimize $HER(\Gamma)$ over all measures Γ. By Theorem 1 since $E\left(\max\left(\frac{V(\beta_\lambda)}{k},\frac{T(\beta_\lambda)}{m-k}\right)\Big|R(\beta_\lambda)=k\right)=\eta_1(\beta_\lambda)+\eta_2(\beta_\lambda)-2\eta_1(\beta_\lambda)\eta_2(\beta_\lambda)$ for $k=1,2$ we only have to consider rejection regions of the form

$$\beta_\lambda=\left\{x:\ \frac{0.8\phi(x)}{0.8\phi(x)+0.2\phi(x-2)}\leqslant\lambda\right\}.$$

By calculating $HER(\beta_\lambda)$ using Eq. (3.5) and minimizing it with respect to λ we get $\lambda^*\approx 0.38$ which implies $\beta_{0.38}=\{Y\geqslant 2.96\}$. Therefore, $\min_\Gamma(HER(\Gamma))=HER(\beta_{0.38})=0.101$ and this occurs at $\Gamma=\beta_{0.38}=\{Y\geqslant 2.96\}$.

3.2. The h-value

Imitating the definition of q-value derived from pFDR, see Storey [17], in this section, we propose h-value which is obtained from HER. Recall that $Y_i|H_i \overset{\text{i.i.d}}{\sim}(1-H_i)F_0+H_iF_1, i=1,\ldots,m, H_i \sim$ Bernoulli$(1-\pi_0), i=1,\ldots,m$, and F_0 and F_1 have a common support.

First, consider the test of a single hypothesis, say H_i, with nested level α rejection regions Γ_α such that $\Gamma_{\alpha'}\subset\Gamma_\alpha$ for $0\leqslant\alpha'\leqslant\alpha\leqslant 1$ and $P(Y_i\in\Gamma_\alpha|H_i=0)\leqslant\alpha$, for $0\leqslant\alpha\leqslant 1$. Suppose we are interested in using the test statistic Y_i to carry out a two-tailed test. Then p-value and h-value can

also be defined more generally. Nested rejection regions $\Gamma_\alpha = [-\infty, -c_\alpha] \cup [c_\alpha, \infty]$ are such that $P(|Y_i| > c_\alpha \Big| H = 0) = \alpha$. The p-value for the observed value $Y_i = y_i$ is (see [14]),

$$p\text{-value}(y_i) = \min_{\{\Gamma_\alpha : y_i \in \Gamma_\alpha\}} P(|Y_i| > c_\alpha \Big| H = 0) = P(|Y_i| > y_i \Big| H = 0). \tag{3.8}$$

The p-value (y_i) in Eq. (3.8) gives a measure of strength of the observed statistic with respect to making Type I error. More specifically, the p-value (y_i) is the minimum Type I error over all possible rejection regions Γ_α containing the observed value $Y_i = y_i$.

We define an analogous quantity in terms of HER.

Definition 1. For m hypothesis, $m \geqslant 2$ given that we observe the random variables $Y_1, \ldots, Y_m$ to be $Y_1 = y_1, \ldots, Y_m = y_m$, we define the h-value of y_i to be

$$h\text{-value}(y_i) = \min_{\{\Gamma_\alpha \,:\, y_i \in \Gamma_\alpha\}} HER(\Gamma_\alpha), \tag{3.9}$$

where $HER(\Gamma_\alpha)$ is obtained from Eq. (3.5) using the facts that $\eta_1(\Gamma_\alpha) = P(H = 0 \Big| |Y_i| > c_\alpha)$ and $\eta_2(\Gamma_\alpha) = P(H = 1 \Big| |Y_i| \leqslant c_\alpha)$. The following remarks are in order with regard to the definition of h-value:

(1) It is known that pFDR is stochastically an increasing function of α and the definition of q-value will force this monotonicity. However, with pFDR being an increasing function of α and pFNR being a decreasing function of α (both in stochastic sense), one can expect that HER is a function like a parabola curve. This is similar to the sum of Types I and II error rates as a function of α for a single hypothesis testing problem.
(2) h-value is more conservative than q-value because h-value controls both errors. To make the h-value less conservative, in practical applications, one can take h-value to be large compared to the q-value. (See Section 4.)

In this paper we limit ourselves to the case where we reject the null hypothesis on the basis of m independent observed p-values, $p_1, \ldots, p_m$. As shown by Storey [18], for rejections based on p-values all the rejection regions are of the form $[0, \gamma]$ for some $0 \leqslant \gamma \leqslant 1$. The nested rejection region $\Gamma_\gamma = [0, \gamma]$, abbreviated by γ leads to h-value $(p_i) = \min_{\{\gamma \geqslant p_i\}} HER(\gamma)$.

3.3. Estimating HER and h-value

Suppose P is the random p-value resulting from any test. Then, under independence the p-values are exchangeable and $P(P \leqslant c|H_0) = c$ and $P(P \leqslant c|H_1) = G(c)$. That is, G is the cumulative distribution function of P under the alternative hypothesis. Now, in terms of p-value we write Eq. (3.5) as

$$\begin{aligned} HER(\gamma) = \eta_2(\gamma) + \eta_1(\gamma) \Bigg\{ \sum_{k=1}^{m-1} \Bigg\{ \sum_{v=0}^{k-1} b_1(v, \eta_1(\gamma), k-1) \\ \times B_2\left(\left[\frac{m-k}{k}(v+1) \right], \eta_2(\gamma), m-k \right) \Bigg\} b_3(m,k) \Bigg\} \left(1 - b_3(m,0) - b_3(m,m)\right)^{-1} \\ - \eta_2(\Gamma) \Bigg\{ \sum_{k=1}^{m-1} \Bigg\{ \sum_{v=0}^{k} b_1(v, \eta_1(\gamma), k) B_2\left(\left[\frac{m-k}{k} v \right] - 1, \eta_2(\gamma), m-k-1 \right) \Bigg\} b_3(m,k) \Bigg\} \end{aligned}$$

$$\times (1 - b_3(m, 0), b_3(m, m))^{-1}, \tag{3.10}$$

where $b_3(m, k) = \frac{m!}{k!(m-k)!}(\pi_0\gamma + \pi_1 G(\gamma))^k (1 - \pi_0\gamma - \pi_1 G(\gamma))^{m-k}$, $\pi_1 = 1 - \pi_0$,

$$\eta_1(\gamma) = \frac{\pi_0\gamma}{\pi_0\gamma + \pi_1 G(\gamma)}, \tag{3.11}$$

and

$$\eta_2(\gamma) = \frac{\pi_1(1 - G(\gamma))}{1 - \pi_0\gamma - \pi_1 G(\gamma)}. \tag{3.12}$$

Now, we need to estimate the right-hand side of Eq. (3.10). Following Storey and Tibshirani [19], a conservative estimate of π_0 and consequently π_1 is

$$\hat{\pi}_0(\lambda) = \frac{\#\{p_i > \lambda\}}{(1 - \lambda)m}$$

for some well chosen λ, where $p_1, \ldots, p_m$ are the observed p-value. See also the algorithm below for more details. A natural estimate of $P(P \leqslant \gamma) = \pi_0\gamma + \pi_1 G(\gamma)$ is

$$\hat{P}(P \leqslant \gamma) = \frac{\#\{p_i \leqslant \gamma\}}{m}.$$

By plugging these quantities into the right side of Eqs. (3.11) and (3.12), $\eta_1(\gamma)$ and $\eta_2(\gamma)$ are estimated by

$$\hat{\eta}_1(\gamma) = \frac{\hat{\pi}_0\gamma m}{\#\{p_i \leqslant \gamma\}},$$
$$\hat{\eta}_2(\gamma) = \frac{\hat{\pi}_1(1 - G(\gamma))}{\#\{p_i > \gamma\}}.$$

Replacing $\eta_1(\gamma)$ and $\eta_2(\gamma)$, with $\hat{\eta}_1(\gamma)$ and $\hat{\eta}_2(\gamma)$, respectively, and $\pi_0\gamma + \pi_1 G(\gamma)$ by $\frac{1}{m}\sum_{i=1}^{m} I(p_i \leqslant \gamma)$ in Eq. (3.10) we obtain $\hat{HER}(\gamma)$ which is an estimate of $HER(\gamma)$. Generally speaking, since p-values of truly alternative hypotheses will tend to be close to zero, a natural choice for $G(u)$ is a non-decreasing concave function with $G(0) = 0$ and $G(1) = 1$.

Now we can estimate the h-value of feature i by simply plugging $\hat{HER}(\gamma)$ into Eq. (3.9),

$$\hat{h}(p_i) = \min_{\gamma \geqslant p_i} \hat{HER}(\gamma). \tag{3.13}$$

The following is general algorithm for estimating h-values from the list of p-values.

1. Let $p(1) \leqslant p(2) \leqslant \cdots \leqslant p(m)$ be the ordered p-values for the m hypothesis tests. This also denotes the ordering of the features in terms of their evidence against the null hypothesis.
2. For a range of λ, say $\lambda = 0, .01, \ldots, .95$ calculate

$$\hat{\pi}_0(\lambda) = \frac{\#\{p_i > \lambda\}}{m(1 - \lambda)}.$$

3. There are many approaches for estimating a density function at a boundary, see Hall and Park [11]. In this paper we follow the method proposed by Storey and Tibshirani [19]. Let $\hat{f}$ be the natural cubic spline with 3 df of $\hat{\pi}_0(\lambda)$ on λ. Set the estimate of π_0 to be

$$\hat{\pi}_0 = \hat{f}(1).$$

4. Noting that for $P(R = m) = 1, HER(\gamma) = pFDR(\gamma)$ a natural estimate of h at $p(m)$ is $\hat{h}(p(m)) = \hat{q}(p(m))$, where $\hat{q}(p(m))$ is the estimate of q-value at $p(m)$.
5. Calculate, for $i = 1, \ldots, m-1$, $\hat{h}(p(i)) = \min(\hat{HER}_{p(i) \leqslant \gamma < p(i+1)}(\gamma), \hat{h}(p(i+1)))$.

4. Example and simulation study

In this section we first apply our method to an example. We then compare our procedure with Storey's procedure.

4.1. Example

A common goal in DNA microarray experiments is to detect genes that show differential expression across two or more biological conditions. In this scenario, the "features" are the genes, and they are tested against the null hypothesis that there is no differential gene expression. One of the goals of Golub et al. [10] was to identify genes that are differentially expressed in patients with two types of leukemia, acute lymphoblastic leukemia (ALL, Class 1) and acute myeloid leukemia (AML, Class 2). Gene expression levels were measured using Affymetrix high density oligonucleotide arrays containing $m = 6817$ human genes. The learning set comprises $n = 38$ samples, 27 ALL cases and 11 AML patients. (Data available at http://www.genome.wi.mit.edu/MPR.) We eliminated some of the genes by truncating very high and very low expression levels and removing genes whose truncated expression showed no variation. This left $m = 3051$ genes. There were no missing values. The data were summarized by a 3051×38 matrix $X = (x_{ij})$, where x_{ij} denotes the expression level for gene i in mRNA sample j.

This data set was used here to illustrate an application of our proposed procedure. We tested each gene for differential expression between ALL and AML patients by using Welch two-sample t-statistic. The two-sample t-statistic for the gene i, allowing for the possibility that two classes have different variances, is then computed for $i = 1, \ldots, 3051$. We next calculated null version of $t_1, t_2, \ldots, t_{3051}$ when there is no differential gene expression. Because it is not clearly valid to assume that t_i follow t-distribution, we calculated them by a permutation method. More specifically, we consider all possible ways to assign $n = 38$ arrays to $n_1 = 27$ arrays from ALL and $n_2 = 11$ arrays from AML. Under the null hypothesis there is no differential gene expression and therefore, t-statistic should have the same distribution regardless of how we make these assignments. The labels on the arrays are randomly scrambled, and the t-statistic are recomputed. Therefore, for $B = 1000$ permutations of the array labels we get a set of null statistics $t_1^{0b}, \ldots, t_{3051}^{0b}$, $b = 1, \ldots, 1000$. The p-value for gene i, $i = 1, \ldots, 3051$ was calculated by

$$p_i = \frac{\sum_{b=1}^{B} \#\{\left|t_j^{0b}\right| > |t_i|,\ j = 1, \ldots, 3051\}}{(3051)B}.$$

We estimated the $h_1, \ldots, h_{3051}$ for differential gene expression between ALL and AML by using the algorithm presented in Section 3. Take, for instance, $G(u) = 1 - \Phi\left(\Phi^{-1}(u) - \sqrt{n}\, w\right)$, where Φ is the cumulative distribution of standard normal distribution and $n = \frac{n_1 n_2}{n_1 + n_2} = \frac{(27)(11)}{38} = 7.8$, see Hung et al. [13]. We have decided to declare a gene to be differentially expressed if its raw p-value is less than 5% and its h-value is less than 15%. As we mentioned in the previous section, to make our procedure less conservative we work with slightly larger h-value. We found 143 genes to have their p-values being less than 5%, and out of these 143 genes, 82 genes had their h-values being less then 15% when $w = .4, .2$ and $.01$. That is, our declared significant genes did

Table 2
Estimated *pFDR*, *pFNR* and *pHER*

	Using q	Using h
$n_1 = n_2 = 10$	$pF\hat{D}R = .123$	$pF\hat{D}R = .108$
	$pF\hat{N}R = .128$	$pF\hat{N}R = .109$
	$pH\hat{E}R = .132$	$pH\hat{E}R = .115$
$n_1 = n_2 = 20$	$pF\hat{D}R = .081$	$pF\hat{D}R = .083$
	$pF\hat{N}R = .095$	$pF\hat{N}R = .092$
	$pH\hat{E}R = .111$	$pH\hat{E}R = .105$
$n_1 = n_2 = 50$	$pF\hat{D}R = .073$	$pF\hat{D}R = .076$
	$pF\hat{N}R = .078$	$pF\hat{N}R = .074$
	$pH\hat{E}R = .091$	$pH\hat{E}R = .090$

not change for different values of w. Also, out of these 143 genes, 98 had their q-values less than .10. If the cutoff point for q is .10, then the declared significant genes using q is more than our procedure.

4.2. A simulation study

In this section we carry out a simulation study to compare our method with Storey's method. We consider two groups and the problem of simultaneously testing $m = 1000$ independent tests $H_1, \ldots, H_{1000}$ for these two groups. In each case the null hypothesis is either true ($H_i = 0$) or not ($H_i = 1$). (One can think of $m = 1000$ as 1000 genes and $H_i = 0$ means that the ith gene is not differentially expressed between two groups.) For a given i, n_1 observations were generated from $N(\mu_i, 2)$ (Group 1) and n_2 observations were generated from $N(\mu_i^*, 2)$ (Group 2), $i = 1, \ldots, 1000$. We assume that $\mu_i = \mu_i^*$ for $i = 1, \ldots, 900$ and $\mu_i^* - \mu_i = .2$, $\mu_i = 2$ for $i = 901, \ldots, 950$. We also take $\mu_i^* - \mu_i = 1.4$ and $\mu_i = 2$ for $i = 951, \ldots, 1000$. Thus the total number of hypotheses tested is 1000 of which the null hypothesis is true $m_0 = 900$ times and the alternative is true $m_1 = 100$ times. For a given i, $i = 1, \ldots, 1000$, we test for equality of two means μ_i and μ_i^* by using two-sample t-test. It gives us p-value for each of the hypotheses. Call these $p_1, \ldots, p_{1000}$, evidence in these 1000 tests. Then, we estimated $h_1, \ldots, h_{1000}$ and $q_1, \ldots, q_{1000}$ based on $p(1), \ldots, p(m)$. To declare hypotheses to be null or alternative, we use cutoff points of .05, .1 and .07 for raw p-values, h-values and q-values, respectively. We repeat this 1000 times and estimate *pFDR*, *pFNR* and *HER* using both q and h. Table 2 gives the estimates for different values of n_1 and n_2. It seems that both q and h work very well for large and moderate sample sizes. However, for small sample size it seems that our procedure does a better job.

5. Conclusion

In this paper we have proposed HER as a measure of error rate in multiple hypothesis testing and studied its properties. Its advantage to commonly used measures is that it considers both FDR and false non-discovery rate simultaneously. We have also developed, based on the risk function HER, an analog of p-value termed as h-value for testing multiple hypothesis. Based on our limited simulation study, since for small sample size h-value does a better job than q-value, we recommend using this measure. For large to moderate sample sizes we recommend using

both measures. The h-value is more conservative than q-value in the sense that it accepts the null hypothesis more than q-value. In practice, however, to make this measure less conservative, one can reject the null hypothesis using larger h-values, that is, a larger cutoff point for h-value than for q-value. A question which remains unanswered yet is how to specify cutoff points for q-value and h-value.

Acknowledgment

I am so grateful for the helpful comments of the reviewers of my manuscript.

References

[1] Y. Benjamin, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. Roy. Statist. Soc. Ser. B 57 (1995) 289–300.
[2] Y. Benjamin, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency, Ann. Statist. 29 (2001) 1165–1188.
[3] A. Cohen, H.B. Sackrowitz, Decision theory results for one sided multiple comparison procedures, Ann. Statist. 33 (2005) 126–144.
[4] A. Cohen, H.B. Sackrowitz, More on the inadmissibility of step-up, J. Multivariate Anal. 98 (2007) 481–492.
[5] R.R. Delongchamp, J.F. Boyer, J.J. Chen, R.R. Kodell, Multiple testing strategy for analyzing cDNA array data on gene expression, Biometrics 60 (2004) 774–782.
[6] S. Dudoit, J.P. Shaffer, J.C. Boldrick, Multiple hypothesis testing in microarray experiments, Statist. Sci. 18 (2003) 71–103.
[7] Y. Ge, S. Dudoit, T. Speed, Resampling-based multiple testing for microarray data analysis, Tests 12 (2003) 1–77.
[8] C. Genovese, L. Wasserman, Operating characteristics and extensions of the false discovery rate procedure, J. Roy. Statist. Soc. B 64 (2002) 499–517.
[9] C. Genovese, L. Wasserman, A stochastic process approach to false discovery control, Ann. Statist. 32 (2004) 1035–1061.
[10] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeck, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, E.S. Lander, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 386 (1999) 531–537.
[11] P. Hall, B.U. Park, Bandwidth choice for local polynomial estimation of smooth boundaries, J. Multivariate Anal. 91 (2004) 240–261.
[12] Y. Hochberg, A. Tamhane, Multiple Comparison Procedures, Wiley, New York, 1987.
[13] H.M.J. Hung, R.T. O'Neil, P. Bauer, K. Kohne, The behavior of P-value when the alternative hypothesis is true, Biometrics 53 (1997) 11–22.
[14] E.L. Lehmann, Testing Statistical Hypotheses, second ed., Springer, New York, 1986.
[15] S.K. Sarkar, FDR-controlling stepwise procedures and their false negative rates comparisons. in: A.C. Tamhane, Peter Westfall (Eds.), The Third International Conference on Multiple Comparison, special issues, J. Statist. Plann. Inference 125 (2004) 119–137.
[16] M. Shaked, J.G. Shanthikumar, Stochastic Orders and their Applications, Academic Press, New York, 1994.
[17] J.D. Storey, A direct approach to false discovery rates, J. Roy. Statist. Soc. B 64 (2002) 479–498.
[18] J.D. Storey, The positive false discovery rate: a Bayesian interpretation and the q-value, Ann. Statist. 31 (2003) 2013–2035.
[19] J.D. Storey, R. Tibshirani, Statistical significance for genomic studies, Proc. Nat. Acad. Sci. 100 (2003) 9440–9445.