DE GRUYTER DOI 10.1515/sagmb-2012-0076 — Statistical Applications in Genetics and Molecular Biology 2014; 13(1): 49–65

Miguel A. Gómez-Villegas*, Isabel Salazar and Luis Sanz

A Bayesian decision procedure for testing multiple hypotheses in DNA microarray experiments

Abstract: DNA microarray experiments require the use of multiple hypothesis testing procedures because thousands of hypotheses are simultaneously tested. We deal with this problem from a Bayesian decision theory perspective. We propose a decision criterion based on an estimation of the number of false null hypotheses (FNH), taking as an error measure the proportion of the posterior expected number of false positives with respect to the estimated number of true null hypotheses. The methodology is applied to a Gaussian model when testing bilateral hypotheses. The procedure is illustrated with both simulated and real data examples and the results are compared to those obtained by the Bayes rule when an additive loss function is considered for each joint action and the generalized loss 0–1 function for each individual action. Our procedure significantly reduced the percentage of false negatives whereas the percentage of false positives remains at an acceptable level.

Keywords: Bayes rule; Bayesian decision; multiple hypothesis; posterior expected loss; false positives; false negatives.

***Corresponding author: Miguel A. Gómez-Villegas,** Dpto. de Estadística e I.O., Facultad de Ciencias Matemáticas, Plaza de las Ciencias, 3, Universidad Complutense de Madrid, 28040–Madrid, Spain, e-mail: villegas@ucm.es
Isabel Salazar: Dpto. de Producción Animal, Facultad de Veterinaria, Universidad Complutense de Madrid, 28040–Madrid, Spain
Luis Sanz: Dpto. de Estadística e I.O., Facultad de Ciencias Matemáticas, Plaza de las Ciencias, 3, Universidad Complutense de Madrid, 28040–Madrid, Spain

1 Introduction

All statistical inference methods involve taking a decision. In multiple hypothesis testing problems and specifically in the context of microarray experiments, in which hundreds or thousands of hypotheses are simultaneously tested, the key point is to decide which hypotheses will be rejected and which will be accepted. Decision Theory is a theoretical framework that allows researches to globally study statistical inference problems as a unique type of problem: taking a decision. It also greatly reinforces the logic of the Bayesian approach (Berger, 1985).

From a Bayesian point of view, deciding between the null and the alternative hypothesis, when a single hypothesis is considered consists in computing both of the posterior probabilities and then deciding accordingly. The idea is similar when testing multiple hypotheses.

We consider the problem of simultaneously testing N hypotheses as follows

$$H_{0i}:\theta_i \in \Theta_{0i} \quad versus \quad H_{1i}:\theta_i \in \Theta_{1i}, \quad i=1, 2, \ldots, N \tag{1}$$

where $\Theta_{0i} \cup \Theta_{1i}=\Theta$ and $\Theta_{0i} \cap \Theta_{1i}=\emptyset$ for all $i=1, 2, \ldots, N$.

If the model is $f(x|\theta)$ with $x=(x_1, x_2, \ldots, x_N)$, where $x_i=(x_{i1}, x_{i2}, \ldots, x_{in})$ is observed from $f(x|\theta)$ for each $i=1, 2, \ldots, N$, and our beliefs about $\theta=(\theta_1, \ldots, \theta_N)$ can be expressed by means of $\pi(\theta)$, then the posterior distribution of θ is given by

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \tag{2}$$

The posterior probability for each of the null hypotheses, $Pr(\theta_i \in \Theta_{0i}|x)$, is obtained from the corresponding marginal distribution of θ_i for $i=1, \ldots, N$. Once the posterior probability of each null hypothesis is computed,

we must decide, on the basis of these probabilities, which of the null hypotheses will be rejected and which will be accepted. Therefore, it is necessary to choose a cutoff in order to reject all hypotheses whose posterior probability is less than or equal to this cutoff value.

To the best of our knowledge, Lehmann (1957a,b) was the first to consider the problem of multiple hypotheses as a frequentist decision problem. He developed an optimal procedure considering an additive loss function. Therefore, the problem of minimizing the risk function or the Bayes risk can be solved by minimizing separately each of the components of the problem and the procedure is then optimal for each comparison without taking into account the other comparisons.

Table 1 describes the possible outcomes from N hypothesis tests, where N_0 and N_1 are the unknown number of true and false null hypotheses, respectively, V is the number of true null hypotheses that are rejected (type I errors or false positives), T is the number of false null hypotheses that are accepted (type II errors or false negatives), and W and R are observable random variables representing the number of accepted and rejected null hypotheses, respectively.

Using a Neyman-Pearson frequentist approach, Spjøtvoll (1972) developed a powerful procedure for testing multiple hypotheses by maximizing the power for each individual hypothesis considering the *per-family error rate* (PFER). This is defined as $PFER=E[V]$, the expected number of false positives. The main result of the works of Lehmann and Spjøtvoll is that if an optimal procedure is used for each component of the problem, then the procedure is optimal for the joint multiple hypothesis problem.

From a Bayesian decision theory point of view, Duncan (1961, 1965) developed an optimal procedure to compare all couples of means for a one-way balanced design. Using normal and independent prior distributions, he derived a Bayes rule for when the loss function is additive and depends on the real mean differences. This rule does not depend on the number of comparisons and in this sense the Bayesian procedure for this problem has the same nature as procedures that control the *per-comparison error rate*, $PCER=E[V]/N$, because it ignores the multiplicity of the problem. However, the procedure proposed by Duncan depends on heterogeneity between the treatment means. Hochberg and Tamhane (1987) provide a detailed description of work by Lehmann, Spjøtvoll and Duncan.

Lewis and Thayer (2004) followed the approach of Shaffer (1999) and Duncan (1965) and applied Bayesian decision theory to the problem of testing multiple hypotheses in a design with random effects. They considered the 0–1 loss function and showed that a Bayes rule controls the *false discovery rate* (FDR), which provides theoretical support for the conclusions of Shaffer (1999). The methodologies of Lewis and Thayer (2004) and Shaffer (1999) can be considered semi-Bayesian because they control some frequentist error measures.

Sun and Cai (2007) developed a compound decision theory framework and derived an oracle rule based on z-values that minimizes the *false nondiscovery rate* (FNR) subject to a constraint on the FDR. Sun and McLain (2012) studied the problem in a compound decision theoretic framework and developed asymptotically valid and optimal procedures for testing composite null hypotheses in heteroscedastic models. They proposed the concept of a composite null distribution for heteroscedastic models and developed an optimal testing procedure that minimizes FNR subject to a constraint on FDR.

Scott and Berger (2006), in the context of DNA microarray experiments, dealt with the multiple hypothesis testing problem from a Bayesian decision theory perspective. In this type of experiment, one of the main objectives is to find active genes and therefore there are two possible actions for each gene: the gene is classified as active or as inactive. If an active gene is classified as inactive, the loss is proportional to the absolute value of the mean level expression for all genes; if an inactive gene is classified as active, the loss is one unit.

Table 1 Possible outcomes in a multiple hypothesis testing problem.

	Accept	Reject	Total
H_0 true	U	V	N_0
H_0 false	T	S	N_1
	W	R	N

Thus, for an active gene classified as inactive, the greater the mean level of the gene expression, the greater will be the loss. To specify this loss function it is only necessary to fix the constant of proportionality. The decision rule will be to choose the action with the smallest posterior expected loss for each gene.

In this paper we propose a Bayesian decision criterion based on estimation of the number of false null hypotheses, taking as an error measure the proportion of the posterior expected number of false positives with respect to the estimated number of true null hypotheses, denoted by $\widehat{FPr}$ and defined in Section 2. We show that this quantity is less than the posterior expected realized FDR, proposed by Genovese and Wasserman (2002, 2003), when the number of rejected null hypotheses is significantly less than the number of true null hypotheses, as occurs in the case of microarray experiments. The methodology is applied to a Gaussian model and is illustrated with simulated data and real data for DNA microarray experiments. The results are compared to those obtained using a Bayes rule methodology when an additive loss function is considered for each joint action and the 0–1 generalized loss function taking equal costs is considered for each individual action.

The results show that more false null hypotheses are detected with our criterion than with the Bayes rule. Therefore, the percentage of false negatives is reduced while the percentage of false positives remains at an acceptable level. This is especially suitable in the context of DNA microarray experiments, in which multiple hypothesis tests are used in most cases as a first exploratory step to identify genes that are differentially expressed for a posterior detailed analysis. In this way, a higher number of false positives can be allowed at an admissible proportion to obtain the greatest possible number of genes of interest (Dudoit et al., 2003).

The remainder of the paper is organized as follows. The problem of testing multiple hypotheses, some frequentist error measures and the estimated proportion of false positives and false negatives that we propose are established in Section 2. In Section 3 we present two decision criteria: a Bayes rule and our proposed criterion based on estimation of the number of false null hypotheses. Section 4 addresses some properties of this new criterion. In Section 5 a Gaussian model is analyzed and examples with simulated and real data are developed. Finally, in Section 6 we present some comments and conclusions.

2 Multiple hypotheses testing and error measures

A traditional error measure in the context of multiple hypotheses as in (1) is the *family-wise error rate* (FWER), defined as $FWER=Pr(V\geq 1)$, the probability of obtaining at least one type I error or false positive. Shaffer (1999) modified Duncan's procedure to control FWER at the 0.05 significance level in a weak sense. She compared her methodology with different frequentist procedures and obtained similar results to those obtained by Benjamini and Hochberg (1995) for FDR control. They defined FDR as the expected proportion of false positives (erroneously rejected hypotheses) with respect to the rejected hypotheses, more precisely

$$FDR=\begin{cases} E[V/R] & \text{if } R>0 \\ 0 & \text{if } R=0 \end{cases} \tag{3}$$

Benjamini and Hochberg (1995) argued that in some situations it may be acceptable to tolerate some false positives provided that there are few in relation to the number of rejected null hypotheses. In the context of DNA microarray experiments, multiple hypotheses are often used as a first exploratory step to identify groups of genes that are differentially expressed in specific biological processes for subsequent further research. Thus, a slightly higher number of false positives may be acceptable at this stage of the analysis and it may also be of interest to reduce the number of false negatives. Therefore, FDR is the most commonly used error type in the frequentist approach in this context.

Genovese and Wasserman (2002, 2003) introduced the *realized FDR* and the *realized FNR*, which can be written as

$$rFDR=\frac{\sum_{i=1}^{N}(1-z_i)\delta_i}{\sum_{i=1}^{N}\delta_i}, \qquad rFNR=\frac{\sum_{i=1}^{N}z_i(1-\delta_i)}{\sum_{i=1}^{N}(1-\delta_i)}, \tag{4}$$

where $\delta_i=1$ if the null H_{0i} is rejected and $\delta_i=0$ otherwise, and $z_i=0$ if the null H_{0i} is true and $z_i=1$ if it is false (note the difference between the *rFDR* and the FDR in (3)). They also consider the posterior expected *rFDR* and *rFNR* (Müller et al., 2004; Do et al., 2005) defined as follows:

$$\begin{aligned}\overline{rFDR}&=E[rFDR|t]=\frac{\sum_{i=1}^{N}Pr(H_{0i}=0|\mathbf{t})\delta_i}{\sum_{i=1}^{N}\delta_i},\\ \overline{rFNR}&=E[rFNR|t]=\frac{\sum_{i=1}^{N}Pr(H_{0i}=1|\mathbf{t})(1-\delta_i)}{\sum_{i=1}^{N}(1-\delta_i)}.\end{aligned} \tag{5}$$

As we see later, it may be that $\overline{rFDR}$ and $\overline{rFNR}$ are acceptably small quantities while the proportion of false negatives, that is, the proportion of misclassified false null hypotheses, is significantly larger. This is why we consider the proportion of the posterior expected false positives and false negatives among the estimated number of true and false null hypotheses, respectively. In fact, consider

$$FP=\sum_{i=1}^{N}(1-z_i)\delta_i \quad \text{and} \quad FN=\sum_{i=1}^{N}z_i(1-\delta_i), \tag{6}$$

the number of false discoveries and false negatives realized, and

$$\overline{FP}=\sum_{i=1}^{N}Pr(H_{0i}=0|t)\delta_i \quad \text{and} \quad \overline{FN}=\sum_{i=1}^{N}Pr(H_{0i}=1|t)(1-\delta_i) \tag{7}$$

the corresponding posterior expected values of *FP* and *FN*. We consider the proportion of $\overline{FP}$ and $\overline{FN}$ with respect to the number of true and false null hypotheses, respectively, that is

$$\overline{FPr}=\frac{\overline{FP}}{N_0} \quad \text{and} \quad \overline{FNr}=\frac{\overline{FN}}{N_1}, \tag{8}$$

where N_0 and N_1 can be estimated by $\hat{N}_0=N\hat{p}$ and $\hat{N}_1=N(1-\hat{p})$ respectively, or through $\hat{N}_0=\sum_{i=1}^{N}Pr(H_{0i}=0|\mathbf{t})$ and $\hat{N}_1=\sum_{i=1}^{N}Pr(H_{0i}=1|\mathbf{t})$.

The quantities $\overline{FPr}$ and $\overline{FNr}$ represent the proportion of true and false misclassified null hypotheses, respectively, and are natural extensions of the probability of obtaining type I and type II errors in the case of a single hypothesis.

3 Bayesian decision criterion

In this section we consider two decision criteria from the Bayesian decision theory point of view. The first criterion is a traditional one (Duncan, 1965; Lewis and Thayer, 2004) and uses the Bayes rule when an additive loss function is considered for multiple hypothesis testing and the generalized 0–1 loss function is considered for each single test. The second criterion proposed is based on estimation of the number of false null hypotheses.

For a testing problem as in (1), we now present these two decision rules.

3.1 The Bayes rule

From a Bayesian decision theory perspective, to determine the null hypotheses that will be accepted and those that will be rejected, we must choose the action with the lowest posterior expected loss.

We first define the parametric and action spaces, which can be represented as

$$\Theta=\bigcup_{j=1}^{2^N}\Theta_j \quad \text{where} \quad \Theta_j=\bigcap_{i=1}^{N}\{H_{0i}=\varepsilon_{ij}\}, \quad j=1,\ldots,2^N, \tag{9}$$

where $\varepsilon_{ij}=0$ when H_{0i} is true and $\varepsilon_{ij}=1$ when H_{0i} is false, and

$$A=\bigcup_{j=1}^{2^N}A_j \quad \text{where} \quad A_j=\bigcap_{i=1}^{N}a_{e_{ij}i}, \quad j=1,\ldots,2^N, \tag{10}$$

where $e_{ij}=0$ if H_{0i} is accepted and $e_{ij}=1$ if H_{0i} is rejected. Therefore, $a_{e_{ij}i}$ is the individual action taken with respect to the hypothesis H_{0i} within the joint action A_j.

We can consider an additive loss function so that the problem of minimizing the posterior expected loss can be solved by minimizing each of its components. Then, when the joint action A_j is taken and the true value of the parameter is Θ_l, the loss function can be written as

$$L(\Theta_l,A_j)=\sum_{i=1}^{N}L_i(H_{0i}=\varepsilon_{il},a_{e_{ij}i}),$$

where

$$L_i(H_{0i}=\varepsilon_{il},a_{e_{ij}i})=\begin{cases}0 & \text{si} \quad e_{ij}=\varepsilon_{il}\\ C_{e_{ij}i} & \text{si} \quad e_{ij}\neq\varepsilon_{il}\end{cases}$$

is the 0–1 generalized loss function and represents the individual cost when the action $a_{e_{ij}i}$ is taken for H_{0i}, with $H_{0i}=\varepsilon_i$. Thus, C_{0i} represents the cost of a false negative and C_{1i} is the cost of a false positive for hypothesis i.

The individual expected posterior losses for the actions a_{0i} and a_{1i} are $C_{0i}Pr(H_{0i}=1|\mathbf{t})$ and $C_{1i}Pr(H_{0i}=0|t)$, respectively, where t is a statistic $t=t(x)$. Then, for each hypothesis H_{0i}, action a_{1i} is preferable to action a_{0i} if

$$C_{1i}Pr(H_{0i}=0|\mathbf{t})\leq C_{0i}Pr(H_{0i}=1|\mathbf{t}),$$

from which we can deduce the Bayes rule: for each $\mathbf{t}$, the null hypothesis H_{0i} will be rejected if

$$Pr(H_{0i}=0|\mathbf{t})\leq\frac{C_{0i}}{C_{0i}+C_{1i}} \tag{11}$$

given costs C_{0i} and C_{1i}, whose specification is not straightforward.

3.2 Criterion based on the estimated false null hypothesis (FNH) number

We propose a decision criterion based on an estimate of the FNH number given all data and prior knowledge. Let N_1 denote the number of false null hypotheses, as in Table 1. This procedure consists of rejecting $\hat{N}_1$ null hypotheses with lower posterior probability of being true, where $\hat{N}_1$ denotes an estimate of N_1. The objective is to obtain a Bayesian estimate of N_1. Suppose that a random vector $T_i=(X_{i1},\ldots,X_{in})$ is observed for each hypothesis and that p is the prior probability of $\theta_i\in\Theta_{0i}$ for all i.

Denoting $H_{0i}=0$ if H_{0i} is true, $H_{0i}=1$ if H_{0i} is false and $p=Pr(H_{0i}=0|p)$ with $1-p=Pr(H_{0i}=1|p)$ for $i=1, \ldots, N$ and assuming that the N hypotheses are independent, then $H_{0i}|p \sim Bernoulli(1-p)$ and since $N_1=\sum_{i=1}^{N} H_{0i}$ we conclude that $N_1|p \sim Binomial(N, 1-p)$ and we can estimate N_1 using, for example, the mean of this distribution.

The unknown parameter p can be estimated using its posterior distribution, $\pi(p|\mathbf{t})$, where $\mathbf{t}=(t_1, \ldots, t_N)$ and $t_i=(x_{i1}, \ldots, x_{in})$ for $i=1, \ldots, N$. For example, $\hat{p}=E_{\pi(p|\mathbf{t})}[p]$, the posterior mean of $\pi(p|\mathbf{t})$. Finally, $N_1|\hat{p} \sim Binomial(N, 1-\hat{p})$ and we can estimate N_1 by $\hat{N}_1=E[N_1|\hat{p}]=N(1-\hat{p})$.

In this way, we consider as a decision criterion rejecting of the $\hat{N}_1$ null hypotheses with the lowest posterior probability and accepting the others, avoiding the problem of choosing the constants C_{0i} and C_{1i} required to apply the Bayes rule defined in (11).

Although it is known that genes are linked in a complex form and therefore are not independent, the independence assumption is useful to model the data in a pragmatic way and to check whether the observed differences in the expression level are significant based solely upon a comparison gene-by-gene. Independence in microarray data is considered largely in literature, as it is the case of Benjamini and Hochberg (1995), Baldi and Long (2001), Genovese and Wasserman (2003), Storey (2003), Chen and Sarkar (2004), Müller et al. (2004), Do et al. (2005), Lönnstedt and Britton (2005), Scott and Berger (2006) and Sun and McLain (2012) between others.

4 Properties of the FNH criterion

First, we note that the FNH criterion is equivalent to the Bayes rule defined in (11) for concrete costs C_{0i} and C_{1i}, depending on the multiple hypothesis testing problem.

In fact, to apply the Bayes rule given in (11), the costs C_{0i} and C_{1i} must first be set for all $i=1, \ldots, N$. Then the cutoff $C_{0i}/(C_{0i}+C_{1i})$ is obtained for each hypothesis. However, with the FNH criterion we reject the estimated number of null hypotheses, $\hat{N}_1$, with the lowest posterior probability of being true and the cutoff is directly obtained from the highest posterior probability of the rejected null hypotheses. That is, $p_{\hat{N}_1}=Pr(H_{(0\hat{N}_1)}=0|\mathbf{t})$ will be the cutoff, where $Pr(H_{(0i)}=0|\mathbf{t})$ represents the ordered posterior probabilities.

Then, setting $C_{0i}/(C_{0i}+C_{1i})$ equal to $p_{\hat{N}_1}$, we obtain the relation between the costs C_{0i} and C_{1i} corresponding to the FNH criterion, that is,

$$\frac{C_{1i}}{C_{0i}}=\frac{1-p_{\hat{N}_1}}{p_{\hat{N}_1}}. \tag{12}$$

Therefore, if the costs C_{0i} and C_{1i}, $i=1, \ldots, N$, satisfy relation (12), then the Bayes rule will provide equivalent results to the FNH criterion. In general, following this reasoning, we could write the costs C_{0i} and C_{1i} in this way, where $p_k=C_{0i}/(C_{0i}+C_{1i})$ is the cutoff to rejection of each null hypothesis and k is the number of null hypotheses rejected.

Second, we show that if the costs for false negatives are equal and positive, that is, if $C_{0i}=C>0$ for $i=1, \ldots, N$, then N_1 is the smallest number of rejected null hypotheses with which a zero posterior expected loss could be obtained.

To prove this result, we need the following proposition, which shows the behavior of the posterior expected loss as a function of k.

Proposition 1. *For fixed values of the costs for false negatives, C_{0i}, $i=1, \ldots, N$, the posterior expected loss for the Bayes rule that reject k hypotheses, $\rho(\pi(\boldsymbol{\theta}|\mathbf{x}), a_k^*)$, is a decreasing function of k, the number of null hypotheses rejected.*

That is, if $k_1<k_2$, then $\rho(\pi(\boldsymbol{\theta}|\mathbf{x}), a_{k_1}^*) \geq \rho(\pi(\boldsymbol{\theta}|\mathbf{x}), a_{k_2}^*)$, where

$$\rho(\pi(\boldsymbol{\theta}|\mathbf{x}),a_k^*)=\sum_{i=1}^{k}C_{1i}Pr(H_{(0i)}=0|\mathbf{t})+\sum_{i=k+1}^{N}C_{0i}Pr(H_{(0i)}=1|\mathbf{t})$$

and a_k^* is the posterior Bayes action by which k null hypotheses will be rejected.

Proof. See Appendix A.

On the other hand, if the costs for false negatives are equal, that is, if $C_{0i}=C$ for $i=1, \ldots, N$, the posterior expected loss for the Bayes rule that leads us to reject k null hypotheses can be expressed as follows:

$$\rho(\pi(\boldsymbol{\theta}|\mathbf{x}),a_k^*)=C\left(\frac{1}{p_k}\sum_{i=1}^{k}Pr(H_{(0i)}=0|\mathbf{t})+N_1-k\right). \tag{13}$$

Proof. See Appendix B.

Therefore, if $C_{0i}=C>0$ for all $i=1, \ldots, N$, then $\rho(\pi(\boldsymbol{\theta}|\mathbf{x}),a_k^*)\geq 0$, because the expected posterior loss is a decreasing function of k,

$$\begin{aligned}\rho(\pi(\boldsymbol{\theta}|\mathbf{x}),a_k^*)\geq\rho(\pi(\boldsymbol{\theta}|\mathbf{x}),a_N^*)&=C\left(\frac{1}{p_N}\sum_{i=1}^{N}Pr(H_{(0i)}=0|\mathbf{t})+N_1-N\right)\\&=C\left(\frac{Np}{p_N}-Np\right)\geq 0\end{aligned}$$

and then $\rho(\pi(\boldsymbol{\theta}|\mathbf{x}),a_k^*)=0$ only if

$$\frac{1}{p_k}\sum_{i=1}^{k}Pr(H_{(0i)}=0|\mathbf{t})=k-N_1,$$

and since $\frac{1}{p_k}\sum_{i=1}^{k}Pr(H_{(0i)}=0|\mathbf{t})\geq 0$ it follows that k should be greater than or equal to N_1. Therefore, N_1 is the smallest number of rejected null hypotheses with which a zero posterior expected loss could be obtained.

Finally, we note that if we reject the N_1 null hypotheses with the lowest posterior probability of being true, we obtain the same expected posterior number of false positives and false negatives.

In fact, if we reject the N_1 null hypotheses with the lowest posterior probability of being true, from (7) we can write

$$\overline{FP}=\sum_{i=1}^{N_1}Pr(H_{(0i)}=0|\mathbf{t}) \quad \text{and} \quad \overline{FN}=\sum_{i=N_1+1}^{N}Pr(H_{(0i)}=1|\mathbf{t}), \tag{14}$$

and using (20) we have that $\overline{FP}=\overline{FN}$, that is, we obtain the same expected posterior number of false positives and false negatives.

Thus, if the FNH criterion is used to reject the $\hat{N}_1$ null hypotheses with the lowest posterior probability and accept the rest, the estimated posterior number of false positives will be very similar to the estimated posterior number of false negatives as we will show in the next section.

Then, as the posterior expected loss is a decreasing function of the number of null hypotheses rejected, *i*) the more hypotheses that are rejected the lower will be the expected posterior loss obtained; however, it is possible that the number of false positives will significantly increase; and *ii*) the fewer hypotheses that are rejected the fewer will be the false positives, but a higher posterior expected loss will be obtained, along with an increase in the number of false negatives.

Therefore, we can conclude that our criterion properly addresses these issues. On the one hand, if the costs for false negatives are equal and positive ($C_{0i}=C>0$ for all $i=1, \ldots, N$), then it follows that the average number of false null hypotheses is the smallest number of the null hypotheses to be rejected to obtain zero loss. On the other hand, on rejecting the average number of false null hypotheses, we obtain that the expected posterior number of false positives and false negatives are equal.

5 Gaussian model

Consider the multiple testing problem $H_{0i}:\mu_i=0$ versus $H_{1i}:\mu_i\neq 0$, $i=1, \ldots, N$ (Dudoit et al., 2003; Lönnstedt and Britton, 2005; Scott and Berger, 2006; Storey et al., 2007; Cabras, 2010; Ausín et al., 2011). We assume that the sample mean $T_i=(X_{i1}+ \ldots +X_{in})/n$ is obtained for each hypothesis.

To illustrate use of the FNH criterion, together with the proportion $\overline{FPr}$ defined in (8), we compare the results with those obtained when using $\overline{rFDR}$ in (5) and the Bayes rule at equal costs $C_{0i}=C_{1i}$ for all $i=1, \ldots, N$. We consider the Gaussian model of Ausín et al. (2011) and use an empirical Bayes method to estimate the model parameters and compute the posterior probability of each null hypothesis. That is, we assume that the statistics T_i are, under the null hypothesis, normally distributed random variables with zero mean and unknown accuracy ϕ, $T_i|H_{0i}\sim N(0,1/\sqrt{\phi})$, whereas these statistics have mean $\mu_i\neq 0$ under the alternative hypothesis $T_i|H_{0i}\sim N(\mu_i,1/\sqrt{\phi})$, where μ_i, $i=1, \ldots, N$, are the parameters of interest.

The variance of the two components of the mixture model are assumed to be equal as it is usually considered in the literature (Lönnstedt and Speed, 2002; Kendziorski et al., 2003; Storey, 2003; Scott and Berger, 2006; De la Horra, 2007). If the variance of both components are not the same, we can proceed in an analogous way taking a prior distribution for this new parameter or estimating it by empirical Bayes methods.

Therefore, we should consider for T_i a mixture of both distributions as follows:

$$f(t_i|p,\mu_i,\phi)=pf_0(t_i|\phi)+(1-p)f_1(t_i|\mu_i,\phi) \tag{15}$$

for all $i=1, \ldots, N$, where p is the prior probability of the null hypothesis and $f_0(t_i|\phi)$ and $f_1(t_i|\mu_i,\phi)$ are the densities under the null and the alternative hypotheses, respectively. Then, assuming that the statistics T_i are independent and identically distributed random variables, the likelihood can be written as:

$$l(\boldsymbol{\theta}|\mathbf{t})=\prod_{i=1}^{N} f(t_i|p,\mu_i,\phi)=\prod_{i=1}^{N}[pf_0(t_i|\phi)+(1-p)f_1(t_i|\mu_i,\phi)],$$

where $\boldsymbol{\theta}=(p,\phi,\mu_1, \ldots, \mu_N)$, $\mathbf{t}=(t_1, \ldots, t_N)$ and $t_i=T_i(x_{i1}, \ldots, x_{in})$.

To apply Bayesian inference, we need to define a prior distribution for the model parameters, θ, for which we consider the following conjugate prior distributions (Ausín et al., 2011):

$$\begin{aligned} &p\sim Beta(\alpha,\beta),\\ &\phi\sim Gamma(a/2,\ b/2),\\ &\mu_i|\phi\sim N\left(0,\frac{1}{\sqrt{c_i\phi}}\right),\quad i=1,\ldots,N, \end{aligned} \tag{16}$$

where the parametrization of the gamma distribution is chosen in this way to simplify the calculations. Other parametrizations lead to the same results.

In general, it is not easy to obtain an analytical expression for the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{t})\propto l(\boldsymbol{\theta}|\mathbf{t})\pi(\boldsymbol{\theta})$. However, Bayesian inference may be applied using Monte Carlo methods based on Markov chains (MCMC). We apply a Gibbs sampling to estimate the model parameters and the posterior probability of each null hypothesis using the algorithm of Ausín et al. (2011). Then we use an empirical Bayes approach to estimate the parameter $c=c_i$, for all $i=1, \ldots, N$, associated with the variance of the prior distribution assumed for μ_i in (16).

5.1 Simulation results

To illustrate how the error measures $\overline{FPr}$ and $\overline{FNr}$ in (8), and the FNH criterion work and to compare the results with those obtained using $\overline{rFDR}$ and $\overline{rFNR}$ in (5) and the Bayes rule with equal costs $C_{0i}=C_{1i}$ for all $i=1, \ldots, N$, we performed a simulation with $N=5000$ hypotheses and $n=5$ observations per hypothesis for a mixture of two normal distributions as in (15).

We generated three data sets such that $x_{ij} \sim N(0, 1)$ with probability p and $x_{ij} \sim N(\mu_i, 1)$ with probability $1-p$ for $i=1, \ldots, 5000$ and $j=1, \ldots, 5$ and for values of $p=0.7$, 0.8 and 0.9, as is usually done for microarray experiments. For μ_i, $i=1, \ldots, 5000$, different values were chosen in the interval $[-10, 10]$. Next, for each data set, we calculated the sample means $t_i = \sum_{j=1}^{5} x_{ij}/5$, $i=1, \ldots, 5000$, so that the data t_i are equivalent to simulations from a mixture $pN(0,1/\sqrt{5})+(1-p)N(\mu_i,1/\sqrt{5})$, with $p=0.7$, 0.8 and 0.9.

Given these simulated data, we applied the Gibbs sampling procedure using $(\alpha, \beta)=(1, 1)$ and (a, b)=(0, 0) for the prior distributions in (16), since the procedure is robust with respect to choice of these parameters (Ausín et al., 2011). For the parameters c_i, the same value c was chosen for $i=1, \ldots, 5000$ and was estimated using an empirical Bayes approach. Appendix C contains the Matlab code to implement the methodology. Table 2 shows the posterior estimates of the parameters c, p and ϕ for the different values of p considered and different values of the hyper-parameters. It is evident that our procedure yields good estimates of p and ϕ and moreover it can be seen that the results are robust against the choice of the hyper-parameters of the prior distributions.

The proportions of false positives and false negatives for the Bayes rule with equal costs derived from (5) and (8) are estimated by the following expressions:

$$\widehat{FPr}_B = \frac{\widehat{FP}_B}{\hat{N}_0}, \quad \widehat{FNr}_B = \frac{\widehat{FN}_B}{\hat{N}_1}, \quad \widehat{rFDR}_B = \frac{\widehat{FP}_B}{R_B}, \quad \widehat{rFNR}_B = \frac{\widehat{FN}_B}{N-R_B} \tag{17}$$

where $\widehat{FP}_B = \sum_{i=1}^{N} I(\hat{P}_{0i} \le 0.5)\hat{P}_{0i}$, $\hat{N}_0 = \sum_{i=1}^{N} \hat{P}_{0i}$, $\widehat{FN}_B = \sum_{i=1}^{N} I(\hat{P}_{1i} > 0.5)\hat{P}_{1i}$, $\hat{N}_1 = \sum_{i=1}^{N} \hat{P}_{1i}$; with $\hat{P}_{0i} = \hat{Pr}(H_{0i}=0|\mathbf{t},\alpha,\beta,a,b,\hat{c})$, $\hat{P}_{1i} = \hat{Pr}(H_{0i}=1|\mathbf{t},\alpha,\beta,a,b,\hat{c})$ and R_B is the number of hypotheses rejected according to the Bayes rule with equal costs $C_{0i}=C_{1i}$, for all $i=1, \ldots, N$.

Table 3 shows the posterior estimate of c and results for the Bayes rule with equal costs $C_{0i}=C_{1i}$ for all $i=1, \ldots, N$, including the percentage of rejected null hypotheses (R_B%), $\widehat{FPr}_B$, $\widehat{FNr}_B$, $\widehat{rFDR}_B$ and $\widehat{rFNR}_B$, all computed according to (17).

The estimated percentages of false positives and false negatives for the FNH criterion derived from (5) and (8) are given by

$$\widehat{FPr}_{FNH} = \frac{\widehat{FP}_{FNH}}{\hat{N}_0}, \quad \widehat{FNr}_{FNH} = \frac{\widehat{FN}_{FNH}}{\hat{N}_1}, \tag{18}$$

$$\widehat{rFDR}_{FNH} = \frac{\widehat{FP}_{FNH}}{R_{FNH}}, \quad \widehat{rFNR}_{FNH} = \frac{\widehat{FN}_{FNH}}{N-R_{FNH}},$$

where $\widehat{FP}_{FNH} = \sum_{i=1}^{N} I(\hat{P}_{0i} \le p_{\hat{N}_1})\hat{P}_{0i}$, $\hat{N}_0 = \sum_{i=1}^{N} \hat{P}_{0i}$, $\widehat{FN}_{FNH} = \sum_{i=1}^{N} I(\hat{P}_{1i} > p_{\hat{N}_1})\hat{P}_{1i}$, $\hat{N}_1 = \sum_{i=1}^{N} \hat{P}_{1i}$, with $\hat{P}_{0i} = \hat{Pr}(H_{0i}=0|\mathbf{t},\alpha,\beta,a,b,\hat{c})$ $\hat{P}_{1i} = \hat{Pr}(H_{0i}=0|\mathbf{t},\alpha,\beta,a,b,\hat{c})$ and R_{FNH} is the number of hypotheses obtained rejected on application of the FNH criterion.

Table 2 Posterior estimates for c, p and ϕ for different values of p and different priors for $p \sim Beta(\alpha, \beta)$ and $\phi \sim Gamma(a/2, b/2)$.

(a,b)		$p=0.7$			$p=0.8$			$p=0.9$		
	(α, β)	$\hat{c}$	$\hat{p}$	$\hat{\phi}$	$\hat{c}$	$\hat{p}$	$\hat{\phi}$	$\hat{c}$	$\hat{p}$	$\hat{\phi}$
(0, 0)	(1, 25)	0.0053	0.66	5.63	0.0041	0.78	5.74	0.0074	0.87	5.19
	(1, 1)	0.0053	0.67	5.61	0.0034	0.78	5.88	0.0066	0.88	5.25
	(25, 1)	0.0073	0.67	4.90	0.0060	0.78	5.28	0.0074	0.88	5.17
(5, 1)	(1, 25)	0.0064	0.66	5.22	0.0031	0.78	5.96	0.0075	0.87	5.19
	(1, 1)	0.0054	0.67	5.57	0.0046	0.78	5.60	0.0097	0.88	4.97
	(25, 1)	0.0050	0.67	5.73	0.0034	0.78	5.87	0.0065	0.88	5.25
(1,10)	(1, 25)	0.0051	0.66	5.62	0.0046	0.78	5.53	0.0065	0.88	5.21
	(1, 1)	0.0055	0.67	5.44	0.0034	0.78	5.78	0.0074	0.88	5.11
	(25, 1)	0.0052	0.67	5.55	0.0046	0.78	5.50	0.0068	0.88	5.16

Table 3 Results for application of the Bayes rule with prior distributions $p\sim Beta(1, 1)$, $\phi\sim Gamma(0, 0)$ for different values of p and the corresponding posterior estimate of c.

	$\hat{c}$	$R_B\%$	$\widehat{FPr}_B$	$\widehat{FNr}_B$	$\widehat{rFDR}_B$	$\widehat{rFNR}_B$
p=0.7	0.0053	28.80	0.0092	0.1501	0.0214	0.0699
p=0.8	0.0034	18.96	0.0047	0.1438	0.0195	0.0385
p=0.9	0.0066	9.76	0.0018	0.1906	0.0163	0.0251

Table 4 shows the posterior estimate of c and results obtained by applying the FNH criterion, including the estimated percentage of false null hypotheses $(\hat{N}_1\%)$ and the highest estimated posterior probability for which each null hypothesis is rejected, $p_{\hat{N}_1}=\hat{Pr}(H_{(0\hat{N}_1)}=0|\mathbf{t},\alpha,\beta,a,b)$, where $\hat{Pr}(H_{(0i)}=0|\mathbf{t},\alpha,\beta,a,b)$ is the ordered posterior probabilities. Table 4 also shows results for $\overline{FPr}_{FNH}$, $\overline{FNr}_{FNH}$, $\overline{rFDR}_{FNH}$ and $\overline{rFNR}_{FNH}$, all computed according to (18).

As observed in Tables 3 and 4, the FNH criterion is less conservative than the Bayes rule for equal costs, in the sense that more null hypotheses are rejected. Moreover, our criterion detects a higher percentage of false null hypotheses because we obtain a lower proportion of errors for false negatives, while the proportion of errors for false positives remains at an acceptable level. In addition, $\overline{FPr}$ and $\overline{FNr}$ allow us to reject more null hypotheses than $rFDR$ and $rFNR$. It can be observed in Tables 3 and 4 that rejection more hypotheses according to the FNH criterion reduces the $\overline{FNr}$ error with $\overline{FPr}$ remaining at low levels, while $\overline{FDR}$ increases significantly.

Therefore, we can conclude that the FNH criterion and the error measures $\overline{FPr}$ and $\overline{FNr}$ are more appropriate in our examples than the Bayes rule with equal costs and the error measures $rFDR$ and $rFNR$ Furthermore, they are particularly suitable for DNA microarray experiments, in which tests of multiple hypotheses are often used as a first exploratory step to identify genes that are potentially differentially expressed for subsequent more detailed analysis. Thus, the test may be able to support a greater number of false positives, as long as their proportion is admissible, to identify the largest possible number of genes of interest.

Moreover, as the posterior expected loss is a decreasing function of k, the number of hypotheses rejected, we can also conclude that the posterior expected loss is lower using the FNH criterion than using the Bayes rule, because our criterion indicates rejection of a greater number of null hypotheses, as observed in Figure 1.

Figure 2 shows the error measures computed according to (17) as a function of k for the simulated data with p=0.9. It is evident that $\overline{FPr}$ allows us to reject more null hypotheses than $rFDR$. We might even reject more hypotheses than those rejected by the FNH criterion. Furthermore, we can see that $\widehat{rFDR}\approx\widehat{rFNR}$ using the Bayes rule with equal costs and $\widehat{FP}\approx\widehat{FN}$ using the FNH criterion and that the number of rejected null hypotheses with the FNH criterion is an intermediate point between the point at which equilibrium is reached using $rFDR$ and $rFNR$ and the point at which equilibrium is reached using $\overline{FPr}$ and $\overline{FNr}$.

5.2 Real data results

In this section, we apply the Bayes rule and the FNH criterion to colon cancer data from Alon et al. (1999) to identify differentially expressed genes. Alon et al. (1999) used Affymetrix oligonucleotide arrays to monitor

Table 4 Results according to the FNH criterion with prior distributions $p\sim Beta(1, 1)$, and $\phi\sim Gamma(0, 0)$ for different values of p and the corresponding posterior estimate of c.

	$\hat{c}$	$\hat{N}_1\%$	$p_{\hat{N}_1}$	$\widehat{FPr}_{FNH}$	$\widehat{FNr}_{FNH}$	$\widehat{rFDR}_{FNH}$	$\widehat{rFNR}_{FNH}$
p=0.7	0.0053	33.18	0.8176	0.0559	0.1121	0.1127	0.0556
p=0.8	0.0034	21.72	0.8343	0.0302	0.1085	0.1088	0.0301
p=0.9	0.0066	11.88	0.8706	0.0201	0.1480	0.1493	0.0199

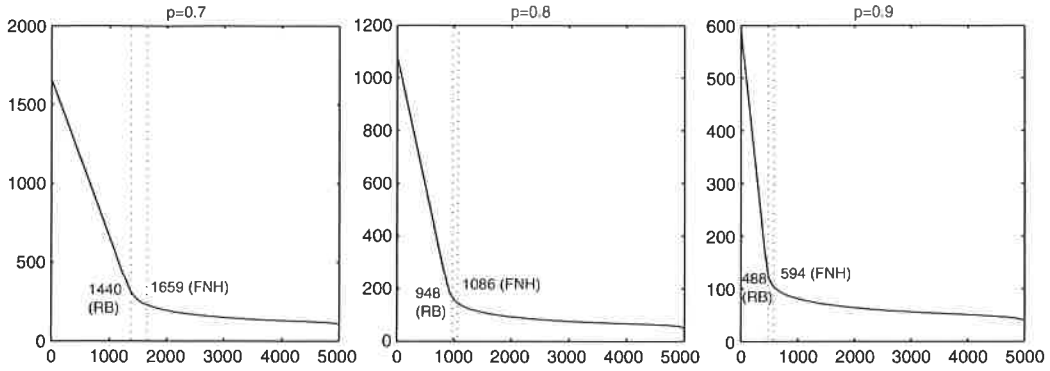

Figure 1 The posterior expected loss according to (13) with C=1 as a function of the number of null hypotheses rejected using simulated data for different values of p.

the expression of over 6500 human genes in 40 tumor and 22 normal colon tissue samples. The samples were taken from 40 different patients, with 22 patients supplying both a tumor and a normal tissue sample. They focused on the 2000 genes with highest minimal intensity across the samples. Further details are available at http://www.stat.ucla.edu/wxl/research/microarray/DBC/index.htm and http://microarray.princeton.edu/oncology/.

Thus, the microarray data matrix for this set has 2000 rows and 62 columns. Alon et al. (1999) did not list the tissues consecutively, so we rearranged the data so that normal tissues are labeled from 1 to 22 and tumor tissues from 23 to 62.

To determine if there are significant differences between the level of expression in normal tissue and that in tumor tissue for each gene, we used the statistic T_i. We consider the Gaussian model with prior distributions given by (16) and with c_i=c for i=1, …, 2000. For the parameters of the priors we choose the same values as in the simulation case: p~*Beta*(1, 1), and ϕ~*Gamma*(0, 0).

Estimation of the model parameters and the posterior probability of the null hypothesis was performed using the procedure proposed by Ausín et al. (2011). We obtained $\hat{c}$=0.0041 $\hat{p}$=0.75 and $\hat{\phi}$=0.00059. Table 5 shows the results obtained by applying the Bayes rule with equal costs C_{0i}=C_{1i} for i=1, …, 2000 and the FNH criterion; that is, the percentage of genes with differential expression (R%) and the estimated proportion of the posterior expected count of false positives ($\widehat{FPr}$) and false negatives ($\widehat{FNr}$), together with the posterior expected realized FDR and FNR ($\widehat{rFDR}$ and $\widehat{rFNR}$) according to (17) and (18). We also found that the highest estimated posterior probability for which each null hypothesis is rejected when applying the FNH criterion, $p_{\hat{N}_1}$=0.7002.

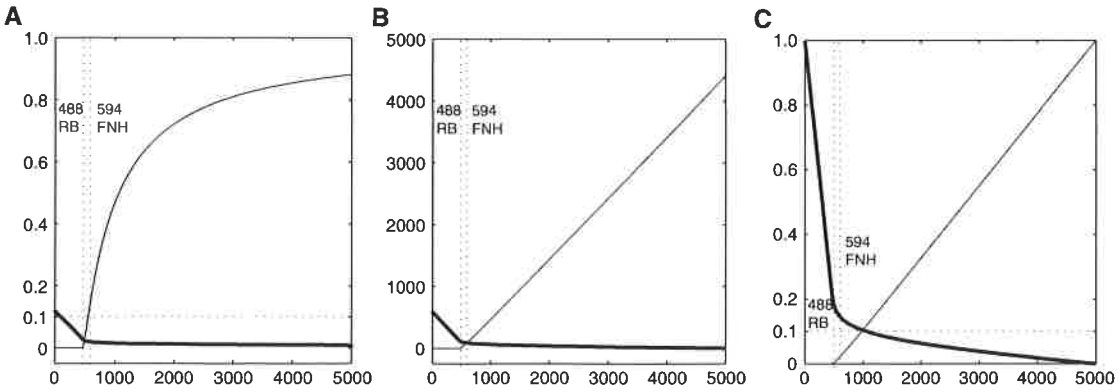

Figure 2 (A) $\overline{rFDR}$ (solid line) and $\overline{rFNR}$ (thick line), (B) $\overline{FP}$ (solid line) and $\overline{FN}$ (thick line) and (C) $\overline{FPr}$ (solid line) and $\overline{FNr}$ (thick line) as a function of the number of null hypotheses rejected using simulated data with p=0.9.

Table 5 Results according to the Bayes rule with equal costs, $C_{0i}=C_{1i}$ for all $i=1, \ldots, 2000$, and the FNH criterion with prior distributions $p\sim Beta(1, 1)$, and $\phi\sim Gamma(0, 0)$ and $\hat{c}=0.0041$ for colon cancer data from Alon et al. (1999).

	R%	$\widehat{FPr}$	$\widehat{FNr}$	$\widehat{rFDR}$	$\widehat{rFNR}$
Bayes rule	21.95	0.0195	0.1782	0.0669	0.0567
FNH criterion	24.90	0.0434	0.1297	0.1310	0.0429

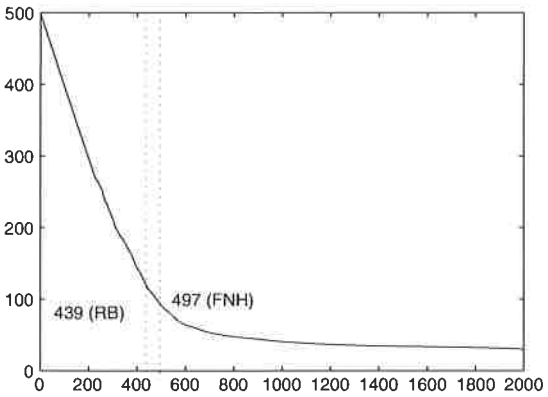

Figure 3 Posterior expected loss as a function of the number of null hypotheses rejected for colon cancer data from Alon et al. (1999) with $(\alpha, \beta)=(1, 1)$, (a, b)=(0, 0) and $\hat{c}=0.0041$.

It can be observed that, as in the case of simulated data, a higher percentage of genes with differential expression is detected and a lower proportion of false negatives is obtained with our criterion than with the Bayes rule, while the proportion of false positives remains at an acceptable level, which is desirable in this context. Figure 3 shows the expected posterior loss as a function of the number of hypotheses rejected (k). It is evident that our procedure provides a lower posterior expected loss than the Bayes rule, in agreement with the simulated results. Finally, Figure 4 shows the error measures computed according to (17) and (18) as a function of k for colon cancer data from Alon et al. (1999). It is evident that these error measures show the same behavior as for the simulated data.

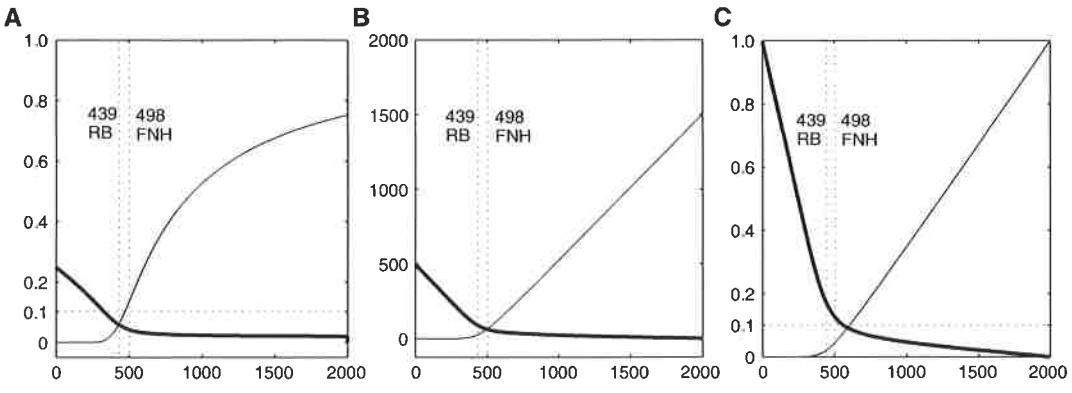

Figure 4 $\widehat{rFDR}$ (solid line) and $\widehat{rFNR}$ (thick line), b) $\widehat{FP}$ (solid line) and $\widehat{FN}$ (thick line) and c) $\widehat{FPr}$ (solid line) and $\widehat{FNr}$ (thick line) as a function of the number of rejected null hypotheses for colon cancer data from Alon et al. (1999) with $(\alpha, \beta)=(1, 1)$, (a, b)=(0, 0) and $\hat{c}=0.0041$.

6 Conclusions and comments

The procedure proposed here, which consists of estimating the number of false null hypotheses that will be rejected, takes into account not only information provided by the observations corresponding to each hypothesis but also the joint information provided by all of the hypotheses. Moreover, it is very straightforward to implement.

In comparison with the Bayes rule, the FNH criterion detects a higher percentage of false null hypotheses, resulting in a more powerful test since a lower percentage of false negatives is obtained. In addition, the rate of false positives is kept within acceptable levels, as observed in Tables 3 and 4. Furthermore, it is not necessary to fix the costs C_{0i} and C_{1i} for each hypothesis for our decision criterion, which is one of the problems that arises when applying the standard Bayes rule because the costs are not easy to specify in most cases.

Furthermore, when the costs for false negatives are fixed, the posterior expected loss is a decreasing function of k, the number of hypotheses that are rejected; from another point of view, the posterior expected loss is also a decreasing function of the highest probability at which every null hypothesis is rejected. Therefore, using our approach in this case, the posterior expected loss is lower than if the Bayes rule is used provided that such a probability is >0.5, which occurs in all the cases we have discussed. Also note that since the posterior expected loss is a decreasing function of k, as more hypotheses are rejected, a lower posterior expected loss is obtained along with a lower proportion of false negatives, albeit with an increase in the percentage of false positives. Conversely, if fewer null hypotheses are rejected, the percentage of false positives decreases and the posterior expected loss increases with the percentage of false negatives. In this sense, the FNH criterion combines very adequately both situations. On the one hand, if the costs for false negatives are equal, that is $C_{0i}=C>0$ for all $i=1, \ldots, N$, then it follows that the average number of false null hypotheses is the smallest number of null hypotheses to be rejected for zero loss. On the other hand, for rejection of the average number of false null hypotheses the expected posterior number of false positives and false negatives is equal and matches the number of null hypotheses rejected when using our FNH decision criterion.

Therefore, we can conclude that the FNH criterion and the error measures $\overline{FDr}$ and $\overline{FNr}$ are more appropriate in our examples than the Bayes rule with equal costs and the error measures $\overline{FDR}$ and $\overline{FNR}$. Furthermore, they are particularly desirable in the context of DNA microarray experiments, in which tests of multiple hypotheses are used in many cases as a first exploratory step to identify groups of genes that are potentially differentially expressed for subsequent more detailed analyses. Thus, the test may be able to support a greater number of false positives, as long as their proportion is admissible, to reach the largest possible number of genes of interest.

Acknowledgements: This work was supported by grants MTM 2008-03282/MTM and GR 58/08. The authors gratefully acknowledge the very constructive comments and suggestions of two anonymous referees who have contributed to improve the quality of the paper.

Appendix A

Proof of Proposition 1

First we show that for fixed values of the costs for false negatives, C_{0i}, $i=1, \ldots, N$, the posterior expected loss for the Bayes rule that rejects k hypotheses, $\rho(\pi(\boldsymbol{\theta}|\mathbf{x}), a_k^*)$ is a function of the cutoff p_k or equivalently a function of k, the number of rejected null hypotheses.

In fact, setting $p_k=C_{0i}/(C_{0i}+C_{1i})$, we can write $\rho(\pi(\boldsymbol{\theta}|\mathbf{x}), a_k^*)$ as follows,

$$\begin{aligned}\rho(\pi(\boldsymbol{\theta}|\mathbf{x}),a_k^*)&=\sum_{i=1}^{k}C_{1i}Pr(H_{(0i)}=0|\mathbf{t})+\sum_{i=k+1}^{N}C_{0i}Pr(H_{(0i)}=1|\mathbf{t})\\&=\frac{(1-p_k)}{p_k}\sum_{i=1}^{k}C_{0i}Pr(H_{(0i)}=0|\mathbf{t})+\sum_{i=k+1}^{N}C_{0i}Pr(H_{(0i)}=1|\mathbf{t}),\end{aligned}\tag{19}$$

where a_k^* is the posterior Bayes action by which the kth null hypothesis will be rejected. Then, for fixed values of C_{0i}, $i=1, \ldots, N$, the posterior expected loss for the Bayes rule, $\rho(\pi(\boldsymbol{\theta}|\mathbf{x}),a_k^*)$, is a function of the cutoff p_k or equivalently a function of k, the number of rejected null hypotheses.

Finally we show that $\rho(\pi(\boldsymbol{\theta}|\mathbf{x}),a_k^*)$ is a decreasing function on k.

Let k_1 and k_2 be such that $k_1<k_2$; then $p_{k_1}<p_{k_2}$ and $\frac{(1-p_{k_2})}{p_{k_2}}<\frac{(1-p_{k_1})}{p_{k_1}}$, and

$$\begin{aligned}\rho(\pi(\boldsymbol{\theta}|\mathbf{x}),a_{k_2}^*)&=\frac{(1-p_{k_2})}{p_{k_2}}\sum_{i=1}^{k_2}C_{0i}Pr(H_{(0i)}=0|\mathbf{t})+\sum_{i=k_2+1}^{N}C_{0i}Pr(H_{(0i)}=1|\mathbf{t})=\frac{(1-p_{k_2})}{p_{k_2}}\sum_{i=1}^{k_1}C_{0i}Pr(H_{(0i)}=0|\mathbf{t})\\&+\sum_{i=k_1+1}^{N}C_{0i}Pr(H_{(0i)}=1|\mathbf{t})-\sum_{i=k_1+1}^{k_2}C_{0i}Pr(H_{(0i)}=1|\mathbf{t})+\frac{(1-p_{k_2})}{p_{k_2}}\sum_{i=k_1+1}^{k_2}C_{0i}Pr(H_{(0i)}=0|\mathbf{t})\\&<\frac{(1-p_{k_1})}{p_{k_1}}\sum_{i=1}^{k_1}C_{0i}Pr(H_{(0i)}=0|\mathbf{t})+\sum_{i=k_1-1}^{N}C_{0i}Pr(H_{(0i)}=1|\mathbf{t})-\sum_{i=k_1+1}^{k_2}C_{0i}Pr(H_{(0i)}=1|\mathbf{t})\\&+\frac{(1-p_{k_2})}{p_{k_2}}\sum_{i=k_1+1}^{k_2}C_{0i}Pr(H_{(0i)}=0|\mathbf{t})=\rho(\pi(\boldsymbol{\theta}|\mathbf{x}),a_{k_1}^*)-\sum_{i=k_1+1}^{k_2}C_{0i}Pr(H_{(0i)}=1|\mathbf{t})\\&+\frac{(1-p_{k_2})}{p_{k_2}}\sum_{i=k_1+1}^{k_2}C_{0i}Pr(H_{(0i)}=1|\mathbf{t})=\rho(\pi(\boldsymbol{\theta}|\mathbf{x}),a_{k_1}^*)-\sum_{i=k_1+1}^{k_2}C_{0i}+\sum_{i=k_1+1}^{k_2}C_{0i}Pr(H_{(0i)}=0|\mathbf{t})\\&+\frac{(1-p_{k_2})}{p_{k_2}}\sum_{i=k_1+1}^{k_2}C_{0i}Pr(H_{(0i)}=0|\mathbf{t})=\rho(\pi(\boldsymbol{\theta}|\mathbf{x}),a_{k_1}^*)-\sum_{i=k_1+1}^{k_2}C_{0i}+\frac{1}{p_{k_2}}\sum_{i=k_1+1}^{k_2}C_{0i}Pr(H_{(0i)}=0|\mathbf{t})\\&\leq\rho(\pi(\boldsymbol{\theta}|\mathbf{x}),a_{k_1}^*)-\sum_{i=k_1+1}^{k_2}C_{0i}+\sum_{i=k_1+1}^{k_2}C_{0i}=\rho(\pi(\boldsymbol{\theta}|\mathbf{x}),a_{k_1}^*)\end{aligned}$$

Appendix B

In fact, if $C_{0i}=C$ for $i=1, \ldots, N$, then the second term of (19) can be written as

$$\begin{aligned}C\sum_{i=k+1}^{N}Pr(H_{(0i)}=1|\mathbf{t})&=C\sum_{i=k+1}^{N}(1-Pr(H_{(0i)}=0|\mathbf{t}))\\&=C\left(N-k-\sum_{i=k+1}^{N}Pr(H_{(0i)}=0|\mathbf{t})\right)\\&=C\left(N-k-Np+\sum_{i=1}^{N}Pr(H_{(0i)}=0|\mathbf{t})\right)\\&=C\left(N_1-k+\sum_{i=1}^{N}Pr(H_{(0i)}=0|\mathbf{t})\right),\end{aligned}\tag{20}$$

where $p=Pr(H_{0i}=0|p)$ is the prior probability that each null hypothesis is true. Then, substituting (20) in (19), we immediately obtain (13).

Appendix C

Matlab code for simulated data

1. Codes of Matlab for function M–file for function to maximize:

```
function EB = EB (c,t,al,bet,a,b,m,N)
p= ;phi= ;mu=zeros(1,N);EB=0;burnin= ; iters= ;
for iter=1:burnin+iters
    prob0=p*exp(-phi*(t.^2)/2);
    prob1=prob0+(1-p)*exp(-phi*((t-mu).^2)/2);
    uni=rand(1,N).*prob1;z=zeros(1,N);z=uni>prob0;
    N1=sum(z);N0=N-N1;I1=find(z);
    p=betarnd(al+N0,bet+N1);
    bast=b+sum(t(z==0).^2)+sum((t(I1)-mu(I1)).^2)+sum(c*(mu-m).^2);
    phi=gamrnd((a+2*N)/2,2/bast);
    mu(I1)=normrnd((c*m(I1)+t(I1))./(c+1),1./sqrt((c+1)*phi));
    mu(z==0)=normrnd(m(z==0),1./sqrt(c*phi));
    if iter>burnin;EB=EB+prod((p*(phi^0.5)*exp(-phi*(t.^2)/2))+((1-14  p)*(phi^0.5)*
    exp(-phi*((t-mu).^2)/2)));end
end
EB=-EB/iters
```

2. Create an M–file to obtain the maximum of the function EB using the code "fminbnd"

3. Given the previously estimated value of c, next sentences estimate the rest of the parameters and measures used in the approach.

```
    clear;seed=;rand('state',seed);randn('state',seed);
    N=;n=;sig=;ptrue=;phitrue=n*(sig)^-2;mutrue=linspace( , ,N);
    al=;bet=;a=;b=;m=zeros(1,N);c= *ones(1,N);
    meds=mutrue;uni=rand(1,N);ztrue=uni>ptrue;meds(uni<ptrue)=0;
    for i=1:n;x(i,:)=normrnd(meds,sig);end
    t=mean(x);s=std(x);
    p=; phi=; mu=zeros(1,N);
    Ez=zeros(1,N);Emu1=zeros(1,N);Ep=0;Ephi=0;
    fid = fopen('pphi.txt','wt');fid2 = fopen('tzmu.txt','wt');
    fid3 = fopen('mpphi.txt','wt');burnin=;iters=;
    for iter=1:burnin+iters
        prob0=p*exp(-phi*(t.^2)/2);
        prob1=prob0+(1-p)*exp(-phi*((t-mu).^2)/2);
        uni=rand(1,N).*prob1;z=zeros(1,N);z=uni>prob0;
        if iter>burnin;Ez=Ez+z;end
        N1=sum(z);N0=N-N1;I1=find(z);p=betarnd(al+N0,bet+N1);
        if iter>burnin;Ep=Ep+p;mp=Ep/(iter-burnin);end
        bast=b+sum(t(z==0).^2)+sum((t(I1)-mu(I1)).^2)+sum(c.*(mu-m).^2);
        phi=gamrnd((a+2*N)/2,2/bast);
        if iter>burnin;Ephi=Ephi+phi;mphi=Ephi/(iter-burnin);end
        if iter>burnin;fprintf(fid3,'%6.4f %6.4f/n',[mp; mphi]);end
        mu(I1)=normrnd((c(I1).*m(I1)+t(I1))./(c(I1)+1),1./sqrt((c(I1)+1)*phi));
        mu(z==0)=normrnd(m(z==0),1./sqrt(c(z==0)*phi));
        if iter>burnin;Emu1(I1)=Emu1(I1)+mu(I1);end
        if iter>burnin;fprintf(fid,'%6.4f %6.4f/n',[p; phi]);end
    end
    Emu1=Emu1./Ez;Ez=Ez/iters;load pphi.txt;load mpphi.txt
```

```
pest=[ 1-sum(ztrue)/N ptrue mean(pphi(:,1))]
phiest=[phitrue mean(pphi(:,2))]
zest=[ztrue;Ez]';muest=[meds;Emu1]';tzmu=[t; ztrue;Ez;meds;Emu1]';
fprintf(fid2,'%6.4f %6.4f %6.4f %6.4f %6.4f/n',[t;ztrue;Ez;meds;Emu1]);
fclose('all');
subplot(2,2,1),plot(pphi(:,1)),title('p')
subplot(2,2,2),plot(pphi(:,2)),title('phi')
subplot(2,2,3),plot(mpphi(:,1)),title('mp')
subplot(2,2,4),plot(mpphi(:,2)),title('mphi')
RB=(sum(Ez>.5)*100)/N
R1=round(N*(1-mean(pphi(:,1))));
N1=(R1*100)/N
EzO=sort(Ez,'descend');pN1=1-EzO(R1)
V=cumsum(1-EzO);F=cumsum(EzO);T=sum(Ez)-F;FPr=V./V(N);
FNr=T./sum(Ez);R=(1:1:N);I=V./R;II=T./(N-R);
for i=1:N-1;EI(i)=I(i);EII(i)=II(i);end
rFDR=[EI,V(N)/N];rFNR=[EII,0];
FPrB=FPr((RB*N)/100)
FNrB=FNr((RB*N)/100)
rFDRB=rFDR((RB*N)/100)
rFNRB=rFNR((RB*N)/100)
FPrFNH=FPr(R1)
FNrFNH= FNr(R1)
rFDRFNH=rFDR(R1)
rFNRFNH=rFNR(R1)
```

For real data change the following lines
line 17 by: clear;x = xlsread('File name.xls'); %This file must contain a column with mean differences
line 18 by: N= ;
lines 20, 21 and 22 by: t=x';
line 44 by: pest=mean(pphi(:,1))
line 45 by: phiest=mean(pphi(:,2))
line 46 by: tzmu=[t;Ez;Emu1]';
line 47 by: fprintf(fid2,'%6.4f %6.4f %6.4f/n',[t;Ez;Emu1]);

References

Alon, U., N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack and A. J. Levine (1999): "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," Proc. Natl. Acad. Sci. USA, 96, 6745–6750.

Ausín, M. C., M. A. Gómez-Villegas, B. González-Pérez, M. T. Rodríquez-Bernal, I. Salazar and L. Sanz (2011): "Bayesian analysis of multiple hypothesis testing with applications to microarray experiments," Commun Stat-Theor M, 40, 2276–2291.

Baldi, P. and A. D. Long (2001): "A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes," Bioinformatics, 17, 509–519.

Benjamini, Y. and Y. Hochberg (1995): "Controlling the false discovery rate: a practical and powerful approach to multiple testing," J. Roy. Stat. Soc. Ser. B, 57, 289–300.

Berger, J. (1985): Statistical decision theory and Bayesian analysis, 2nd ed. New York: Springer-Verlag.

Cabras, S. (2010): "A note on multiple testing for composite null hypotheses," J. Stat. Plan. Infer., 140, 659–666.

Chen, J. and S. K. Sarkar (2004): "Multiple testing of response rates with a control: a Bayesian stepwise approach," J. Stat. Plan. Infer., 125, 3–16.

De la Horra, J. (2007): "Bayesian robustness of the positive false discovery rate," Commun Stat-Theor M, 36, 1905–1914.

Do, K.-A., P. Müller and F. Tang (2005): "A Bayesian mixture model for differential gene expression," J. R. Stat. Soc. Ser. C, 54, 627–644.

Dudoit, S., J. P. Shaffer, and J. C. Boldrick (2003): "Multiple hypothesis testing in microarray experiments," Stat. Sci., 18, 71–103.

Duncan, D. B. (1961): "Bayes rules for a common multiple comparisons problem and related Student-t problems," Ann. Math. Stat., 32, 1013–1033.

Duncan, D. B. (1965): "A Bayesian approach to multiple comparisons," Technometrics, 7, 171–222.

Genovese, C. and L. Wasserman (2002): "Operating characteristics and extensions of the false discovery rate procedure," J. Roy. Stat. Soc. Ser. B, 64, 499–517.

Genovese, C. and L. Wasserman (2003): Bayesian and frequentist multiple testing. In Bernardo, J. M., Bayarri, M., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., West, M. (Eds.), Bayesian Statistics 7. Oxford, UK: Oxford University Press, pp. 145–162.

Hochberg, Y. and A. C. Tamhane (1987): Multiple Comparison Procedures. New York: John Wiley.

Kendziorski, C., M. Newton, H. Lan, M. N. y Gould (2003): "On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles," Stat. Med., 22, 3899–3914.

Lehmann, E. L. (1957a): "A theory of some multiple decision problems, I," Ann. Math. Stat., 28, 1–25.

Lehmann, E. L. (1957b): "A theory of some multiple decision problems, II," Ann. Math. Stat., 28, 547–572.

Lewis, C. and D. T. Thayer (2004): "A loss function related to the FDR for random effects multiple comparisons," J. Stat. Plan. Infer., 125, 49–58.

Lönnstedt, I. and T. Britton (2005): "Hierarchical Bayes models for cDNA microarray gene expression," Biostatistics, 6, 279–291.

Lönnstedt, I. and T. Speed (2002): "Replicated microarray data," Stat. Sinica, 12, 31–46.

Müller, P., G. Parmigiani, C. Robert and J. Rousseau (2004): "Optimal sample size for multiple testing: the case of gene expression microarrays," J. Am. Stat. Assoc., 468:990–1001.

Scott, J. G. and J. O. Berger (2006): "An exploration of aspects of Bayesian multiple testing," J. Stat. Plan. Infer., 136, 2144–2162.

Shaffer, J. P. (1999): "A semi-Bayesian study of Duncan's Bayesian multiple comparison procedure," J. Stat. Plan. Infer., 82, 197–213.

Spjøtvoll E. (1972): "On the optimality of some multiple comparison procedures," Ann. Math. Stat., 43, 398–411.

Storey, J. D. (2003): "The positive false discovery rate: a Bayesian interpretation and the q-value," Ann. Stat., 31, 2013–2035.

Storey, J. D., J. Y. Day and J. T. Leek (2007): "The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments," Biostatistics, 8, 414–432.

Sun, W. and T. T. Cai (2007): "Oracle and adaptive compound decision rules for false discovery rate control," J. Am. Stat. Assoc., 102, 901–912.

Sun, W., and A. C. McLain (2012): "Multiple testing of composite null hypotheses in heteroscedastic models," J. Am. Stat. Assoc., 107, 673–687.