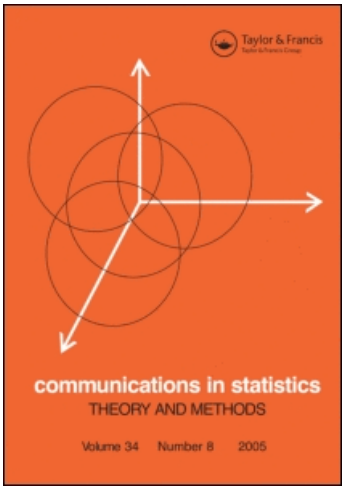

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597238

Bayesian Analysis of Multiple Hypothesis Testing with Applications to Microarray Experiments

M. C. Ausín[a]; M. A. Gómez-Villegas[b]; B. González-Pérez[b]; M. T. Rodríguez-Bernal[b]; I. Salazar[c]; L. Sanz[b]
[a] Departamento de Estadística, Universidad Carlos III de Madrid, Madrid, Spain [b] Departamento de Estadística e Investigación Operativa, Universidad Complutense de Madrid, Madrid, Spain [c] Departamento de Producción Animal, Universidad Complutense de Madrid, Madrid, Spain

Online publication date: 13 April 2011

To cite this Article Ausín, M. C. , Gómez-Villegas, M. A. , González-Pérez, B. , Rodríguez-Bernal, M. T. , Salazar, I. and Sanz, L.(2011) 'Bayesian Analysis of Multiple Hypothesis Testing with Applications to Microarray Experiments', Communications in Statistics - Theory and Methods, 40: 13, 2276 — 2291
To link to this Article: DOI: 10.1080/03610921003778183
URL: http://dx.doi.org/10.1080/03610921003778183

Communications in Statistics—Theory and Methods, 40: 2276–2291, 2011

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610921003778183

Bayesian Analysis of Multiple Hypothesis Testing with Applications to Microarray Experiments

M. C. AUSÍN[1], M. A. GÓMEZ-VILLEGAS[2], B. GONZÁLEZ-PÉREZ[2], M. T. RODRÍGUEZ-BERNAL[2], I. SALAZAR[3], AND L. SANZ[2]

[1]Departamento de Estadística, Universidad Carlos III de Madrid, Madrid, Spain
[2]Departamento de Estadística e Investigación Operativa, Universidad Complutense de Madrid, Madrid, Spain
[3]Departamento de Producción Animal, Universidad Complutense de Madrid, Madrid, Spain

Recently, the field of multiple hypothesis testing has experienced a great expansion, basically because of the new methods developed in the field of genomics. These new methods allow scientists to simultaneously process thousands of hypothesis tests. The frequentist approach to this problem is made by using different testing error measures that allow to control the Type I error rate at a certain desired level. Alternatively, in this article, a Bayesian hierarchical model based on mixture distributions and an empirical Bayes approach are proposed in order to produce a list of rejected hypotheses that will be declared significant and interesting for a more detailed posterior analysis. In particular, we develop a straightforward implementation of a Gibbs sampling scheme where all the conditional posterior distributions are explicit. The results are compared with the frequentist False Discovery Rate (FDR) methodology. Simulation examples show that our model improves the FDR procedure in the sense that it diminishes the percentage of false negatives keeping an acceptable percentage of false positives.

Keywords Empirical Bayes methods; False discovery rate; Gibbs sampler; Mixture models; Multiple hypothesis testing.

Mathematics Subject Classification Primary 62F15; Secondary 62F03.

1. Introduction

In this article, we are interested in the problem of multiple hypothesis testing where it is required to test more than one hypothesis at the same time. It is well known

Received October 21, 2009; Accepted March 15, 2010
Address correspondence to M. A. Gómez-Villegas, Departamento de Estadística e Investigación Operativa I, Universidad Complutense de Madrid. Plaza de Ciencias, 3, Ciudad Universitaria, 28040 Madrid, Spain; E-mail: ma.gv@mat.ucm.es

that if the number of hypotheses is very large and each single hypothesis is tested individually, there will be always a number of rejected hypotheses, even in the case that all of them are true. Thus, the probability of a false positive increases with the number of tests. For example, if the significance level is fixed at 0.05, for each single test, and a set of a hundred hypotheses are tested separately, the expected number of false positives will be five. Consequently, five hypothesis are expected to be rejected simply by chance.

Therefore, the multiple testing problem has been widely studied in very different fields in the literature, as Shaffer (1995) pointed out. Some recent references are Efron and Tibshirani (2007), Gordon et al. (2007), and Sun and Cai (2007). Recently, the field of genomics and, in particular, the DNA microarray experiments have influenced in the revitalization of multiple hypothesis testing procedures due to the requirement of testing thousands of hypotheses simultaneously. With DNA microarray experiments it is possible to obtain large data bases concerning the measurements of expression levels for thousands of genes simultaneously. One of the main objectives of these experiments is the identification of those genes that are differentially expressed, that is, those genes that vary their expression level according to the type of analyzed tissue. This situation can be formulated as a multiple hypothesis testing problem in which each individual hypothesis is associated with one single gene and the interest is to test simultaneously which of the thousands of genes are statistically significant. For a review on multiple hypothesis testing in microarrays experiments; see, e.g., Dudoit et al. (2003).

From a frequentist point of view, the procedure to carry out a multiple hypothesis test is based on controlling a particular measure related with the Type I error rate in order to produce a list of rejected hypotheses. This extends the case of a single hypothesis test where the rejection region is based on the control of the Type I error rate. The possible outcomes that may occur when testing multiple hypotheses simultaneously are resumed in Table 1, as proposed by Benjamini and Hochberg (1995), where N is the known number of hypotheses to test, N_T and N_F are the unknown number of true and false hypotheses, respectively, U, V, T, and S are non-observable random variables and R is an observable random variable representing the number of rejected hypotheses according to some significant rule.

Different measures concerning the Type I error rate have been proposed for the multiple hypothesis problem, including the family wise error rate, the per-comparison error rate and the per-family error rate (see, e.g., Dudoit et al., 2003; Shaffer, 1995). Most of these procedures are usually very conservative or ignore the multiplicity of the problem. Benjamini and Hochberg (1995) stated that, in some situations, it may be reasonable to assume a number of Type I errors, that is, rejected hypotheses that are actually true, provided they are only a few when

Table 1
Possible outcomes in a multiple hypothesis testing problem

	N^o accepted	N^o rejected	Total
True	U	V	N_T
False	T	S	N_F
	W	R	N

compared with the number of rejected hypotheses. For this reason, these authors introduced a less conservative measure, the False Discovery Rate (FDR), which is defined as the expected proportion of Type I errors among the rejected hypotheses,

$$FDR = E\left[\frac{V}{R}\,\middle|\, R > 0\right]\Pr(R > 0),$$

where V and R are defined in Table 1. Benjamini and Hochberg (1995) also derived a procedure to control the FDR at a certain level, α, for independent test statistics, which is based on the rejection of those null hypotheses corresponding to the ordered p-values, $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(k)}$, where

$$k = \max\{i : p_{(i)} \leq i\alpha/N\}.$$

In the context of microarray experiments, multiple hypothesis tests are usually performed in an initial exploratory step. In this initial step, it is mainly interesting to identify a number of gene subgroups that may be associated with important biological processes and might be potential candidates for a deeper posterior analysis. Therefore, the presence of a small number of false positives would not be important in this initial step of the analysis provided as many significant genes as possible can be obtained. Considering this fact, it seems reasonable to be more interested in reducing the Type II error, that is, no rejected hypotheses that are actually false, instead of minimizing the Type I error. These arguments make FDR one of the most popular error types considered in microarray experiments. However, although the FDR is less conservative than other approaches, the number of significant hypothesis obtained with the FDR procedure is still too small.

Bayesian procedures constitute an important class of methods for the identification of differentially expressed genes and whose applications in microarray data analysis can be found in the earlier work of Waller and Duncan (1969), Storey (2003) who discussed Bayesian interpretations of the p-value, Newton et al. (2001), Barbieri and Berger (2004), Lönnstedt and Britton (2005), Scott and Berger (2006), and Bansal (2007), among others. In Bayesian methods, the criterion for identifying differentially expressed genes is based on the posterior probability that a particular gene is differentially expressed given the data for all genes. Do et al. (2005) use a nonparametric Bayesian probability model for the distribution of gene intensities under different conditions.

The empirical Bayes approach, introduced by Robbins (1956, 1964), is used to make an analysis of microarray data as can be seen in Efron (2003). Empirical Bayes models in the context of microarray data have been developed by Efron et al. (2001), Kendziorski et al. (2003), and Datta and Datta (2005), among others. Casella (2001) estimates the hyperparameters by Gibbs sampling, using an empirical Bayes approach.

In this article, we propose a Bayesian approach in order to address the multiple hypothesis problem. We assume that a test statistic has been observed for each hypothesis. Thus, the test statistics follow a mixture of two distributions, one for each of the two hypotheses in the test. Assuming prior distributions for the model parameters, we are interested in estimating of the posterior probability that each hypothesis is true. This can be addressed using data augmentation techniques where indicator variables are introduced to simplify the likelihood and the derivation of

the posterior distributions. In particular, we implement a straightforward Gibbs sampling for testing multiple means under Gaussianity. Note that the Gaussian assumption is frequently imposed in the context of microarray experiments when data are usually preprocessed and normalized in order to compare gene expression information obtained from different sources. We also develop a detailed sensitivity analysis for the choice of the prior, propose an empirical Bayes model and compare our results with the frequentist FDR procedure.

The rest of this article is organized as follows. Section 2 proposes a full parametric Bayesian approach for multiple hypothesis testing. Section 3 applies the proposed approach for a multiple hypothesis test where the interest is in the means of Gaussian populations, which is the usual framework in microarray experiments. We define a Gibbs sampling algorithm and illustrate it using simulated data. A sensitivity analysis for the choice of prior distribution is also carried out. Section 4 deals with an empirical Bayes model that simplifies and improves the final results, with a couple of examples included to show how the methodology works. Section 5 includes some discussion and presents the final conclusions.

2. A Bayesian Approach

Consider the problem of multiple hypothesis testing which is given by,

$$H_{0i} : \theta_i \in \Theta_0 \quad \text{vs.} \quad H_{1i} : \theta_i \in \Theta_1,$$

for $i = 1, 2, \ldots, N$, where a test statistic $T_i = T_i(X_{i1}, \ldots, X_{in})$ is observed for each test $i = 1, \ldots, N$. Suppose that under the null H_{0i}, the test statistic follows a density $T_i \sim f(t_i \mid \theta_i \in \Theta_0, \phi)$, while under H_{1i}, the test statistic follows $T_i \sim f(t_i \mid \theta_i \in \Theta_1, \phi)$, where θ_i is the parameter of interest and ϕ is a nuisance parameter. Observe that for simplicity, we assume that ϕ is the same for all the hypotheses.

Furthermore, a natural Bayesian approach is to assume that there is a common prior probability p that $\theta_i \in \Theta_0$, for all i. Then, it can be observed that p is also the unknown proportion of the true null hypotheses.

From now on we will note, $T_i \mid H_{0i} \sim f_0(t_i)$ and $T_i \mid H_{1i} \sim f_1(t_i)$. Thus, the test statistics, T_i, comes from a mixture of both densities:

$$f(t_i \mid p, \boldsymbol{\theta}_i, \phi) = p f_0(t_i) + (1 - p) f_1(t_i), \tag{1}$$

for $i = 1, \ldots, N$. Then, assuming that the T_i are i.i.d. random variables, the likelihood can be written as,

$$l(\boldsymbol{\theta} \mid \boldsymbol{t}) = \prod_{i=1}^{N} f(t_i \mid p, \theta_i, \phi) = \prod_{i=1}^{N} [p f_0(t_i) + (1 - p) f_1(t_i)], \tag{2}$$

where $\boldsymbol{\theta} = (p, \phi, \theta_1, \ldots, \theta_N)$, $\boldsymbol{t} = (t_1, \ldots, t_N)$ and $t_i = T_i(x_{i1}, \ldots, x_{in})$.

In order to carry out Bayesian inference, we also need to define a prior distribution, $\pi(\boldsymbol{\theta})$, for the set of model parameters, $\boldsymbol{\theta}$. Given this prior and the likelihood (2), it is in general not easy to obtain an analytical expression for the posterior distribution, $\pi(\boldsymbol{\theta} \mid \boldsymbol{t}) \propto l(\boldsymbol{\theta} \mid \boldsymbol{t}) \pi(\boldsymbol{\theta})$. However, Bayesian inference may be performed using Markov Chain Monte Carlo (MCMC) methods, (for an overview see Robert and Casella, 2004). Under mild conditions, given an initial value, $\boldsymbol{\theta}^{(0)}$, the

MCMC approach can produce a Markov chain $\{\boldsymbol{\theta}^{(j)} : j = 1, \ldots, M\}$, where $\boldsymbol{\theta}^{(j)} = (p^{(j)}, \phi^{(j)}, \theta_1^{(j)}, \ldots, \theta_N^{(j)})$, which has equilibrium distribution, $\pi(\boldsymbol{\theta} \mid \boldsymbol{t})$. The MCMC algorithm is carried out by cycling repeatedly through draws of each parameter conditional on the remaining parameters.

Then, in order to specify a model and a prior distribution $\pi(\boldsymbol{\theta})$ for making possible inferences via MCMC we introduce, as it is usually done in mixtures, iid latent variables z_i such that $P(z_i = 0 \mid p) = p$ and $P(z_i = 1 \mid p) = 1 - p$. Furthermore, we suppose that θ_i are iid variables with prior density $\pi(\theta_i \mid \phi)$ being z_i and θ_i independents whatever i, j. The model specification is completed by defining the conditional distribution of t_i given $z_i = 0$, θ_i and ϕ as a member of $f(t \mid \theta, \phi)$ for some $\theta \in \Theta_0$, and the conditional distribution of t_i given $z_i = 1$, θ_i and ϕ as a member of $f(t \mid \theta, \phi)$ for some $\theta \in \Theta_1$.

For simplicity, the conditional distributions $f(t_i \mid z_i = 0, \theta_i, \phi)$ and $f(t_i \mid z_i = 1, \theta_1, \phi)$ will be denoted by $f_0(t_i)$ and $f_1(t_i)$, respectively. The Bayesian specification will be completed by choosing a prior distribution for p, $\pi(p)$, and ϕ, $\pi(\phi)$. Using this latent variables, the likelihood (2) can be written as

$$l(\boldsymbol{\theta} \mid \boldsymbol{z}, \boldsymbol{t}) = f(\boldsymbol{t} \mid \boldsymbol{z}, \boldsymbol{\theta}_{-p}) f(\boldsymbol{z} \mid p) = p^{N_0}(1-p)^{N_1} \prod_{i:\, z_i = 0} f_0(t_i) \prod_{i:\, z_i = 1} f_1(t_i), \tag{3}$$

where $\boldsymbol{\theta}_{-p} = (\phi, \theta_1, \ldots, \theta_N)$, $N_0 = \sum_i I(z_i = 0) \sim Bin(N, p)$ is the amount of observations from the first component of the mixture and $N_1 = \sum_i I(z_i = 1) \sim Bin(N, 1-p)$ those from the second component. In fact, $N = N_0 + N_1$.

Then, the posterior distribution for $(\boldsymbol{\theta}, \boldsymbol{z})$ is given by

$$\pi(\boldsymbol{\theta}, \boldsymbol{z} \mid \boldsymbol{t}) \propto l(\boldsymbol{\theta} \mid \boldsymbol{z}, \boldsymbol{t}) \pi(\boldsymbol{\theta}) \tag{4}$$

and from (4) we can obtain the marginal posterior means for the parameters of interest.

3. The Gaussian Model: An Application to Microarrays

In this section, we address an application of the previous problem to the analysis of microarrays. In this context, as it was pointed out in the introduction, data are usually normalized and the interest is in the means of Gaussian distributions. In particular, each Gaussian variable might represent the measured expression of a certain gene from a DNA microarray. Then, we are interested in classifying the genes as active, when the mean is different from zero, and inactive, when the mean is equal to zero.

Thus, we consider the multiple hypothesis test problem given by,

$$H_{0i} : \mu_i = 0 \quad \text{vs.} \quad H_{1i} : \mu_i \neq 0, \quad i = 1, \ldots, N,$$

where the sample mean statistic $T_i = (X_{i1} + \cdots + X_{in})/n$ is observed for each test $i = 1, \ldots, N$. Under the null, the sample mean statistic follows a Gaussian distribution, $T_i \mid H_{0i} \sim N(0, 1/\phi)$, with zero mean and precision, ϕ, which is the inverse of the variance. Under the alternative, the sample mean statistic is $T_i \mid H_{1i} \sim$

$N(\mu_i, 1/\phi)$, where $\mu_i \neq 0$. Then, under this Gaussian model, the likelihood (3) for the parameter $\boldsymbol{\theta} = (p, \phi, \mu_1, \ldots, \mu_N)$ is given by:

$$\begin{aligned} l(\boldsymbol{\theta} \mid \boldsymbol{z}, \boldsymbol{t}) &\propto \left[\prod_{i:z_i=0} \left(p\phi^{1/2} \exp\left(-\frac{\phi}{2} t_i^2 \right) \right) \right] \\ &\quad \times \left[\prod_{i:z_i=1} \left((1-p)\, \phi^{1/2} \exp\left[-\left(\frac{\phi}{2} (t_i - \mu_i)^2 \right) \right] \right) \right] \\ &\propto p^{N_0} (1-p)^{N_1} \phi^{\frac{N}{2}} \exp\left(-\frac{\phi}{2} \sum_{i:z_i=0} t_i^2 \right) \exp\left[-\left(\frac{\phi}{2} \sum_{i:z_i=1} (t_i - \mu_i)^2 \right) \right]. \end{aligned} \tag{5}$$

Let us assume the following natural conjugate prior distributions (see Gelman et al., 2004),

$$p \sim Beta(\alpha, \beta), \tag{6}$$

$$\phi \sim Gamma(a/2, b/2), \tag{7}$$

$$\mu_i \mid \phi \sim N(0, 1/(c_i\phi)), \quad i = 1, \ldots, N. \tag{8}$$

Then, given these prior densities and the likelihood (5), we derive the conditional posterior distribution of each model parameter in order to construct the MCMC chain. Firstly, the conditional posterior probability that the statistic t_i has been generated by the first mixture component is,

$$\Pr(z_i = 0 \mid t_i, p, \phi, \mu_i) = \frac{p \exp\left(-\frac{\phi}{2} t_i^2\right)}{p \exp\left(-\frac{\phi}{2} t_i^2\right) + (1-p) \exp\left(-\frac{\phi}{2} (t_i - \mu_i)^2\right)}, \tag{9}$$

and by the second mixture component, $\Pr(z_i = 1 \mid t_i, p, \phi, \mu_i) = 1 - \Pr(z_i = 0 \mid t_i, p, \phi, \mu_i)$.

Also, it is easy to see that the conditional posterior distribution of p given the data and the rest of parameters is

$$\pi(p \mid \boldsymbol{t}, \boldsymbol{z}) \sim Beta(N_0 + \alpha, N_1 + \beta), \tag{10}$$

And, the conditional posterior distribution of ϕ given the data and the rest of parameters is

$$\pi(\phi \mid \boldsymbol{t}, \boldsymbol{z}, \mu_1, \ldots, \mu_N) \sim Gamma\left(\frac{a + 2N}{2}, \frac{1}{2} K \right), \tag{11}$$

where $K = b + \sum_{i=1}^{N} c_i \mu_i^2 + \sum_{i:z_i=0} t_i^2 + \sum_{i:z_i=1} (t_i - \mu_i)^2$.

Finally, for $i = 1, \ldots, N$, the conditional posterior distribution of μ_i given the data and the rest of parameters depends on the value of z_i and is given by

$$\pi(\mu_i \mid t_i, z_i = 0, \phi) \sim N\left(0, \frac{1}{c_i \phi}\right), \tag{12}$$

whereas

$$\pi(\mu_i \mid t_i, z_i = 1, \phi) \sim N\left(\frac{t_i}{1+c_i}, \frac{1}{(1+c_i)\phi}\right). \tag{13}$$

It can be observed that the conditional posterior distributions (9), (10), (11), (12), and (13) are explicit, and this allows us to implement a straightforward Gibbs sampling algorithm as follows.

1. Specify initial values $\boldsymbol{\theta}^{(0)} = \left(p^{(0)}, \phi^{(0)}, \mu_1^{(0)}, \ldots, \mu_N^{(0)}\right)$.
2. Update $z_i^{(j)}$, for $i = 1, \ldots, N$, by sampling from (9).
3. Update $p^{(j)}$ by sampling from (10).
4. Update $\phi^{(j)}$ by sampling from (11).
5. Update $\mu_i^{(j)}$ for $i = 1, \ldots, N$ by sampling from (12) if $z_i^{(j)} = 0$ and from (13) if $z_i^{(j)} = 1$.
6. Go to step 2.

Given an MCMC sample, $\left\{p^{(j)}, \phi^{(j)}, \mu_1^{(j)}, \ldots, \mu_N^{(j)}\right\}_{j=1}^{M}$, obtained from the Gibbs sampling algorithm, we can obtain estimates of the posterior marginal means by

$$\hat{p} = E[p \mid \boldsymbol{t}] \approx \frac{1}{M}\sum_{j=1}^{M} p^{(j)}, \tag{14}$$

$$\hat{\phi} = E[\phi \mid \boldsymbol{t}] \approx \frac{1}{M}\sum_{j=1}^{M} \phi^{(j)}, \tag{15}$$

and, for each $i = 1, \ldots, N$,

$$\hat{\mu}_i = E[\mu_i \mid \boldsymbol{t}] \approx \frac{1}{M}\sum_{j=1}^{M} \mu_i^{(j)}. \tag{16}$$

Using the posterior sample of the model parameters, we can also approximate the posterior probability of the alternative hypothesis by

$$\Pr(H_{1i} \mid \boldsymbol{t}) = \Pr(\mu_i \neq 0 \mid \boldsymbol{t}) = \Pr(z_i = 1 \mid \boldsymbol{t}) \approx \frac{1}{M}\sum_{j=1}^{M} I(z_i^{(j)} = 1), \tag{17}$$

for $i = 1, \ldots, N$.

Observe that we can use these posterior probabilities to solve the multiple hypothesis test problem. For example, we might consider rejecting the null hypothesis H_{0i} if $\Pr(z_i = 1 \mid \boldsymbol{t}) > 0.5$, for $i = 1, \ldots, N$. Using this criterion, the percentage of rejected hypotheses will be

$$R\% = \frac{\sum_{i=1}^{N} I\left(\Pr(z_i = 1 \mid \boldsymbol{t}) > 0.5\right)}{N} \times 100. \tag{18}$$

When the parameters μ_i are known, as they are with simulated data, we can compute the percentage of false positives (FP%) or Type I errors and the percentage of false

negatives (FN%) or Type II errors in order to show the behavior of our approach. Both percentages are given by the following expressions

$$FP\% = \frac{\sum_{i=1}^{N} I\left(\Pr(z_i = 1 \mid \boldsymbol{t}) > 0.5\right) \times I\left(\mu_i = 0\right)}{\sum_{i=1}^{N} I\left(\mu_i = 0\right)} \times 100, \tag{19}$$

$$FN\% = \frac{\sum_{i=1}^{N} I\left(\Pr(z_i = 1 \mid \boldsymbol{t}) \leq 0.5\right) \times I\left(\mu_i \neq 0\right)}{\sum_{i=1}^{N} I\left(\mu_i \neq 0\right)} \times 100. \tag{20}$$

3.1. *Simulation Results and Sensitivity Analysis*

In this section, we develop a simulation experiment to examine the performance of our approach. We simulate a microarray experiment with data on the expression levels of $N = 5{,}000$ genes, with $n = 5$ observations per gene, simultaneously tested in order to analyze which genes are differentially expressed.

In this context, the data are generated from a mixture of two Gaussian distributions, such that $x_{ij} \sim N(0, 1)$ with probability $p = 0.9$ and $x_{ij} \sim N(\mu_i, 1)$ with probability $1 - p = 0.1$, for $i = 1, \ldots, 5{,}000$ and $j = 1, \ldots, 5$, with μ_i taking different values in the interval $[-4, 4]$. Thus, the data on sample mean statistics, $t_i = (x_{i1} + \cdots + x_{i5})/5$, have been simulated from a mixture of two normal distributions, $0.9 \times N(0, 1/5) + 0.1 \times N(\mu_i, 1/5)$, for $i = 1, \ldots, 5{,}000$.

Given these simulated data, we apply our Gibbs sampling procedure using different values for the parameters of the prior distributions in order to develop a sensitivity analysis as follows.

- For the parameters (α, β) given in (6), we take the values $(1, 25)$, $(1, 1)$, and $(25, 1)$.
- For the parameters (a, b) given in (7), we take the values $(0, 0)$, $(1, 10)$, and $(10, 1)$.
- For the parameter c_i given in (8), we take the values 0.00001, 0.0001, 0.001, 0.01, 0.1, 0.2, and 0.4 for $i = 1, \ldots, 5{,}000$.

Table 2 presents the estimated posterior means of p and ϕ, obtained using (14) and (15), respectively. Table 2 also shows the percentage of rejected hypothesis (R%) that is, the percentage of differentially expressed genes, obtained using (18), and the percentage of false positives (FP%) and false negatives (FN%), obtained using (19) and (20), respectively, for different values of α, β and $c = c_i$, for $i = 1, \ldots, N$, and a noninformative gamma prior for ϕ with parameters $(a, b) = (0, 0)$. It can be observed that for a fixed value of c, the estimations $\hat{p}$ and $\hat{\phi}$, and the values of $R\%$, $FP\%$, and $FN\%$ are robust with respect to the prior distribution of p, because they take similar values for different choices of the prior parameters α and β.

However, it seems that the parameter $c_i = c$, defined in (8), has a great influence on the results since we can observe significant differences in the estimations obtained for different values of c. Moreover, for very small or very large values of c, the percentage of rejected hypotheses is much smaller than 10% and the percentage of false negatives is very large. Therefore, it seems that a value of c around 0.1 would be appropriate, since it produces a small value for $FN\%$, keeping a reasonable value for $FP\%$. Then, we conclude that there is sensitivity to the choice of c, which implies a sensitivity to the choice of the prior variance for the means μ_i. In order to solve this problem, we will develop in the next section an empirical Bayesian model to

Table 2
Estimated posterior means of p and ϕ, percentage of rejected hypothesis ($R\%$) and percentage of false positives ($FP\%$) and false negatives ($FN\%$) using a non informative gamma prior distribution for the nuisance parameter, ϕ, with $a = 0$ and $b = 0$

		c						
(α, β)		0.00001	0.0001	0.001	0.01	0.1	0.2	0.4
(1, 25)	$\hat{p}$	0.94	0.93	0.92	0.89	0.82	0.80	0.77
	$\hat{\phi}$	3.95	4.50	5.01	5.60	4.88	3.90	2.96
	$R\%$	5.28	6.22	6.88	7.92	9.02	8.22	7.70
	$FP\%$	0	0.02	0.07	0.36	1.08	0.58	0.22
	$FN\%$	51.91	43.53	37.89	30.78	26.59	29.87	31.69
(1, 1)	$\hat{p}$	0.95	0.94	0.92	0.90	0.83	0.81	0.79
	$\hat{\phi}$	3.89	4.46	4.99	5.54	4.81	3.86	2.93
	$R\%$	5.20	6.16	6.88	7.78	8.76	8.04	7.56
	$FP\%$	0	0.02	0.07	0.27	0.90	0.47	0.16
	$FN\%$	52.64	44.08	37.89	31.33	27.50	30.60	32.42
(25, 1)	$\hat{p}$	0.95	0.94	0.93	0.90	0.83	0.82	0.79
	$\hat{\phi}$	3.91	4.47	4.99	5.54	4.80	3.85	2.93
	$R\%$	5.24	6.22	6.86	7.80	8.70	8.00	7.54
	$FP\%$	0	0.02	0.07	0.29	0.88	0.45	0.16
	$FN\%$	52.28	43.53	38.07	31.33	27.87	30.78	32.60

select an appropriate value for c, which will lead to adequate values for $R\%$, $FN\%$, and $FP\%$.

Firstly, in order to explore the influence of the choice of the gamma prior distribution for ϕ, given in (7), Tables 3 and 4 show the estimated posterior means of p and ϕ, the percentage of rejected hypothesis and percentage of false positives and false negatives for the same values of α, β, and c considered in Table 2 and two gamma prior distributions for ϕ with parameters (1, 10) and (10, 1), respectively. It can be observed that there is no much sensitivity to the choice of the gamma prior parameters, (a, b), because the estimations obtained in Tables 3 and 4 are very similar to those obtained in Table 2.

4. An Empirical Bayes Model

In this section, we consider the same Gaussian model defined in the previous section where $\boldsymbol{\theta} = (p, \phi, \mu_1, \ldots, \mu_N)$ is the set of parameters of interest, whose prior distributions are given by (6), (7), and (8), respectively. We illustrated in the previous section that there is no sensitivity to the choice of the prior parameters, except for the choice of c_i.

In order to solve the lack of robustness to the choice of c_i we could consider a hierarchical Bayesian model where for example $c_i \sim Gamma(e, f)$, for $i = 1, \ldots, N$. However, we have observed in practice that the results are highly dependent on the selection of the hyperparameters (e, f).

Table 3
Estimated posterior means of p and ϕ, percentage of rejected hypothesis ($R\%$) and percentage of false positives ($FP\%$) and false negatives ($FN\%$) using a gamma prior distribution for the nuisance parameter, ϕ, with $a = 1$ and $b = 10$

		c						
(α, β)		0.00001	0.0001	0.001	0.01	0.1	0.2	0.4
(1, 25)	$\hat{p}$	0.94	0.93	0.92	0.89	0.82	0.80	0.77
	$\hat{\phi}$	3.86	4.44	4.94	5.50	4.81	3.86	2.94
	$R\%$	5.18	6.2	6.84	7.82	8.86	8.14	7.66
	$FP\%$	0	0.02	0.07	0.29	0.99	0.49	0.18
	$FN\%$	52.82	43.72	38.25	31.15	27.32	29.87	31.69
(1, 1)	$\hat{p}$	0.95	0.94	0.93	0.90	0.83	0.81	0.79
	$\hat{\phi}$	3.86	4.41	4.91	5.46	4.74	3.82	2.91
	$R\%$	5.18	6.16	6.80	7.76	8.68	8.02	7.60
	$FP\%$	0	0.02	0.07	0.27	0.88	0.47	0.18
	$FN\%$	52.82	44.08	38.62	31.51	28.05	30.78	32.24
(25, 1)	$\hat{p}$	0.95	0.94	0.93	0.90	0.84	0.82	0.79
	$\hat{\phi}$	3.82	4.41	4.91	5.45	4.73	3.81	2.90
	$R\%$	5.14	6.18	6.76	7.74	8.58	7.98	7.48
	$FP\%$	0	0.02	0.04	0.27	0.79	0.43	0.16
	$FN\%$	53.19	43.90	38.80	31.69	28.23	30.78	33.15

Table 4
Estimated posterior means of p and ϕ, percentage of rejected hypothesis ($R\%$) and percentage of false positives ($FP\%$) and false negatives ($FN\%$) using a gamma prior distribution of the nuisance parameter, ϕ, with $a = 10$ and $b = 1$

		c						
(α, β)		0.00001	0.0001	0.001	0.01	0.1	0.2	0.4
(1, 25)	$\hat{p}$	0.94	0.93	0.92	0.89	0.82	0.80	0.77
	$\hat{\phi}$	3.93	4.52	5.03	5.61	4.88	3.91	2.97
	$R\%$	5.24	6.22	6.88	7.90	9.04	8.30	7.70
	$FP\%$	0	0.02	0.07	0.34	1.10	0.65	0.22
	$FN\%$	52.28	43.53	37.89	30.78	26.59	29.69	31.69
(1, 1)	$\hat{p}$	0.95	0.94	0.92	0.90	0.83	0.81	0.78
	$\hat{\phi}$	3.90	4.48	4.99	5.55	4.82	3.87	2.94
	$R\%$	5.2	6.22	6.82	7.80	8.78	8.08	7.56
	$FP\%$	0	0.02	0.07	0.29	0.92	0.47	0.16
	$FN\%$	52.64	43.53	38.43	31.33	27.50	30.24	32.42
(25, 1)	$\hat{p}$	0.95	0.94	0.92	0.90	0.83	0.81	0.79
	$\hat{\phi}$	3.91	4.48	4.99	5.55	4.81	3.86	2.93
	$R\%$	5.24	6.22	6.88	7.84	8.74	8.04	7.56
	$FP\%$	0	0.02	0.07	0.31	0.90	0.47	0.16
	$FR\%$	52.28	43.53	37.89	31.15	27.69	30.60	32.42

Alternatively, in this section, we use an empirical Bayes approach to estimate the parameter $c = c_i$, for $i = 1, \ldots, N$. Once we obtain an estimated value, $\hat{c}$, we can generate samples of the posterior distribution, $\pi(\boldsymbol{\theta} \mid \boldsymbol{t}, \alpha, \beta, a, b, \hat{c})$, and obtain the corresponding estimators of the posterior marginal means of p, ϕ and μ_i, for $i = 1, \ldots, N$, through (14), (15), and (16), respectively, by using the Gibbs sampling algorithm defined in the previous section, as proposed by Casella (2001).

Our proposal, to estimate the parameter c, is to obtain by maximum likelihood the value $\hat{c}$ such that

$$\hat{c} = \arg\max_{c} m(\boldsymbol{t} \mid \alpha, \beta, a, b, c)$$

where

$$\begin{aligned} m(\boldsymbol{t} \mid \alpha, \beta, a, b, c) &= \int l(\boldsymbol{\theta} \mid \boldsymbol{t}) \pi(\boldsymbol{\theta} \mid \boldsymbol{t}, \alpha, \beta, a, b, c) d\boldsymbol{\theta} \\ &= \int \prod_{i=1}^{N} f(t_i \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta} \mid \boldsymbol{t}, \alpha, \beta, a, b, c) d\boldsymbol{\theta}, \end{aligned} \tag{21}$$

where $f(t_i \mid \boldsymbol{\theta})$ is given in (1).

It is not easy to evaluate analytically the integral given in (21). However, it can be approximated using posterior samples obtained from the Gibbs sampling algorithm defined in the previous section. For each value of c, we can obtain a posterior sample,

$$\left\{\boldsymbol{\theta}^{(j)}(c)\right\}_{j=1}^{M} = \left\{p^{(j)}, \phi^{(j)}, \mu_1^{(j)}, \ldots, \mu_N^{(j)}\right\}_{j=1}^{M}$$

and then use the following Monte Carlo approximation of (21),

$$\hat{m}(\boldsymbol{t} \mid \alpha, \beta, a, b, c) = \frac{1}{M} \sum_{j=1}^{M} \prod_{i=1}^{N} f\left(t_i \mid \boldsymbol{\theta}^{(j)}(c)\right).$$

Therefore, our estimation of c is given by

$$\hat{c} = \arg\max_{c} \frac{1}{M} \sum_{j=1}^{M} \prod_{i=1}^{N} f\left(t_i \mid \boldsymbol{\theta}^{(j)}(c)\right).$$

Observe that in the maximization process it is required to run a Gibbs sampling for every evaluation of the function to maximize. However, this is not an important problem in terms of computational cost because the proposed Gibbs sampling, being completely explicit, is very fast. For example, less than one minute is required to run 20,000 iterations (discarding the first 10,000 as burning iterations).

Table 5 shows the maximum likelihood estimator of c computed for each case, together with the estimated values of p and ϕ and the percentage of rejected hypotheses computed from (18), the estimated FDR and the estimated FNR. The results show, in coherence with those obtained in Tables 2, 3, and 4, that our approach seems to select appropriate optimal values for c, leading to adequate numbers of rejected hypotheses and small percentages of false negatives and positives as it is shown with the estimated values of FDR and FNR.

Table 5

Maximum likelihood estimator of c, estimated posterior means of p and ϕ, percentage of rejected hypothesis ($R\%$) and estimated FDR ($\widehat{FDR}$) and estimated FNR ($\widehat{FNR}$), using different priors for $p \sim Beta(\alpha, \beta)$ and $\phi \sim Gamma(a/2, b/2)$

(a, b)	(α, β)	$\hat{c}_{opt}$	$\hat{p}$	$\hat{\phi}$	$R\%$	$\widehat{FDR}$	$\widehat{FNR}$
(0, 0)	(1, 25)	0.0860	0.83	5.06	9.14	0.130	0.291
	(1, 1)	0.0739	0.84	5.15	8.90	0.111	0.296
	(25, 1)	0.0911	0.84	4.91	8.74	0.103	0.303
(1, 10)	(1, 25)	0.0865	0.83	4.97	8.90	0.113	0.298
	(1, 1)	0.0673	0.85	5.15	8.74	0.103	0.303
	(25, 1)	0.0792	0.84	4.99	8.66	0.098	0.307
(10, 1)	(1, 25)	0.0865	0.83	5.06	9.16	0.130	0.289
	(1, 1)	0.0735	0.84	5.17	8.90	0.1019	0.296
	(25, 1)	0.0705	0.84	5.19	8.88	0.109	0.296

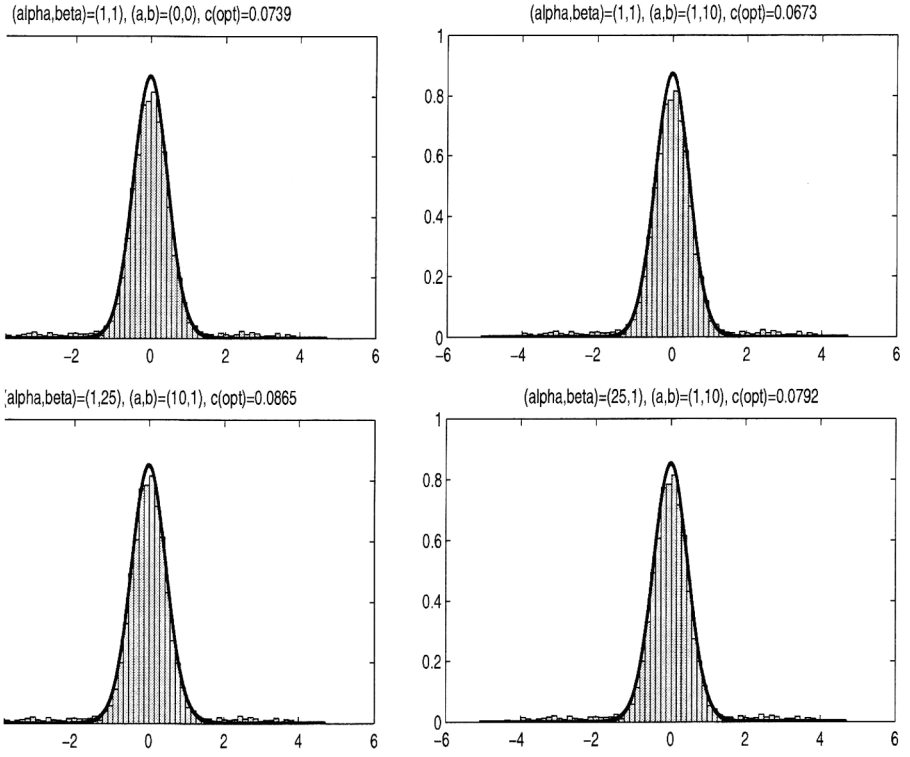

Figure 1. Diagnostic check: histograms are of normalized differences from simulated data from the Gaussian–Gamma model with the fitted predictive density, for different parameter values.

The data were simulated from the Gaussian–Gamma model. The histograms and the fitted densities are shown in Fig. 1.

In comparison with the FDR procedure, Table 6 shows the percentage of rejected hypotheses together with the estimated values of FDR and FNR obtained by the empirical Bayes approach, using a $Gamma(0, 0)$ and a $Beta(1, 1)$ as priors for ϕ and p, respectively, and the corresponding maximum likelihood estimator for c ($\hat{c}_{MLE}$). It can be observed that the estimated percentage of rejected hypotheses is closer to the percentage simulated in the mixture (10%) than the percentages estimated by the FDR method with the levels usually used in the literature for control the FDR. Moreover, the estimated FNR is significantly smaller than the FNR obtained with the FDR procedure, whereas the estimated FDR is acceptable. Moreover, it has been included another two estimated values of c in order to compare our results with those obtained by the FDR procedure.

4.1. *Application to Microarrays*

In this section, the proposed procedure is applied to the data set of colon cancer from Alon et al. (1999) to identify differentially expressed genes. Alon et al. (1999) used Affymetrix oligonucleotide arrays to monitor expressions of over 6,500 human gene expressions in 40 tumor and 22 normal colon tissue samples. The samples were taken from 40 different patients, with 22 patients supplying both a tumour and a normal tissue sample. They focused on the 2,000 genes with highest minimal intensity across the samples. Further details 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 2,000 rows and 62 columns. In Alon et al. (1999), the tissues are not listed consecutively, but here we have rearranged the data so that the normal tissues are labeled from 1 to 22 and the tumour tissues from 23–62.

Table 7 shows, for different values of the hyperparameters α, β, a, and b, the maximum likelihood estimator of c, together with the estimations of p, ϕ and the percentage of rejected hypotheses. It can be shown that, just like the simulated data, these results are robust with respect to the choice of prior distribution parameters.

Using $Gamma(0, 0)$ and $Beta(1, 1)$ distributions as priors for ϕ and p, respectively, and the corresponding maximum likelihood estimator for c, it can

Table 6

Percentage of rejected hypothesis and estimated FDR ($\widehat{FDR}$) and estimated FNR ($\widehat{FNR}$) using the empirical Bayes (EB) approach with $(\alpha, \beta) = (1, 1)$ and $(a, b) = (0, 0)$, compared with the frequentist FDR method for different values of c

	EB method			FDR method	
	$\hat{c}_{MLE} = 0.0739$	$\hat{c} = 0.014$	$\hat{c} = 0.0327$	$\alpha = 0.05$	$\alpha = 0.10$
$R\%$	8.90	7.94	8.56	2.74	4.72
$\widehat{FDR}$	0.111	0.048	0.094	0.047	0.085
$\widehat{FNR}$	0.296	0.334	0.313	0.783	0.633

Table 7
Maximum likelihood estimator of c, estimated posterior means of p and ϕ and percentage of rejected hypothesis ($R\%$), using different priors for $p \sim Beta(\alpha, \beta)$ and $\phi \sim Gamma(a/2, b/2)$. Colon cancer data

(a, b)	(α, β)	$\hat{c}_{opt}$	$\hat{p}$	$\hat{\phi}$	$R\%$
(0, 0)	(1, 25)	0.0040	0.72	0.00069	24.25
	(1, 1)	0.0041	0.75	0.00059	21.95
	(25, 1)	0.0041	0.76	0.00047	21.65
(1, 10)	(1, 25)	0.0060	0.71	0.00065	24.40
	(1, 1)	0.0056	0.74	0.00060	22.80
	(25, 1)	0.0048	0.75	0.00058	21.90
(10, 1)	(1, 25)	0.0040	0.71	0.00073	24.90
	(1, 1)	0.0054	0.73	0.00063	23.55
	(25, 1)	0.0047	0.75	0.00061	22.75

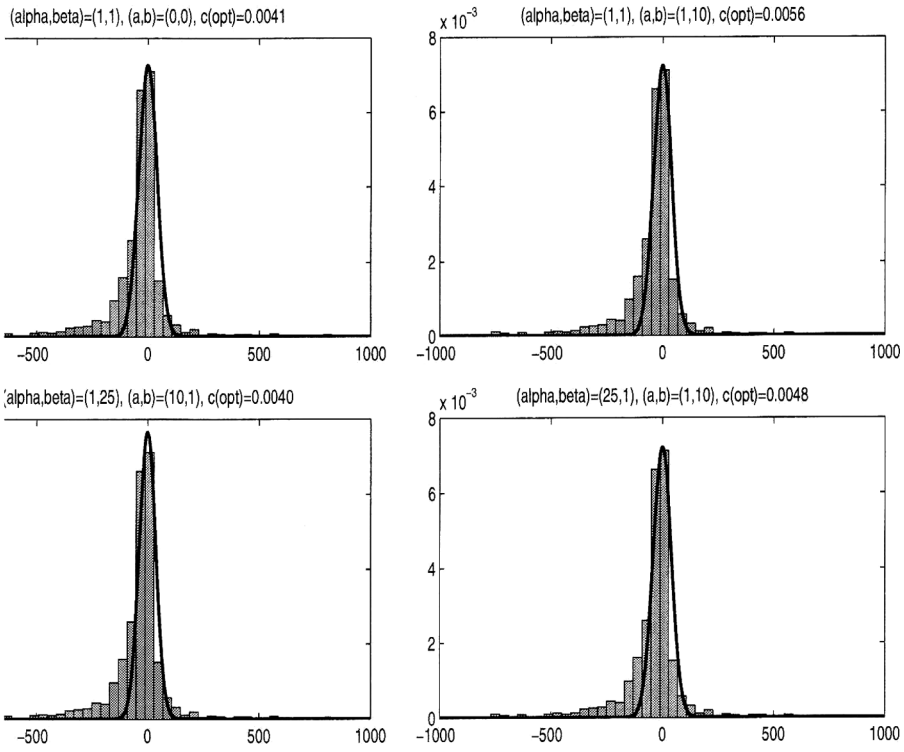

Figure 2. Diagnostic check: histograms are of differences from Colon cancer data. Curves are fitted densities, for different parameter values.

be observed that 21.95% of genes are differentially expressed with our empirical procedure, whereas with the procedure proposed by Benjamini and Hochberg (1995) for control the FDR, this percentage is reduced to a 6% for an $\alpha = 0.05$ and 10.90% if $\alpha = 0.1$.

It must be observed that Alon et al. (1999) analyzed more than 6,500 genes at first, but after they focused their analysis in 2,000 genes with highest minimal intensity across the samples. We have dealt with these 2,000 genes, therefore a high percentage of rejected hypothesis is obtained.

Our approach leads to a relatively straightforward procedure to identify genes differentially expressed. Moreover, the inferences are based on the posterior probabilities and then the influence of each gene can be evaluated. For example, in the data set of colon cancer from Alon et al. (1999), we obtained 223 genes that are differentially expressed with posterior probability equal to 1.

Figure 2 shows how our Gaussian–Gamma model fits data. Plotted on each histogram is the fitted marginal density from de Gaussian–Gamma model. It can be seen that the fit captures the basic attributes of the data, even though that there is margin for improvement.

5. Conclusions and Comments

Our approach significantly improves the percentage of the Type II errors, or false negatives, with respect to the FDR procedure, while the percentage of false positives or Type I errors is quite acceptable. That is to say, our model is less conservative than the FDR procedure, which is an objection usually marked in the literature.

In the simulated example that we dealt with, the performance of our Bayesian methods improves the results obtained with the FDR procedure in the sense that it reduces significantly the percentage of false negatives with respect to the FDR method. Moreover, our procedure identifies the genes that are really differentially expressed.

Furthermore, the procedure is robust with respect to the parameters of the prior distributions, as great changes in these values do not cause important changes either in the parameter estimations or in the estimation of false positives and false negatives.

The sensitivity, which was initially observed to the choice of the parameter associated with the prior variance of the means, has been solved by using an empirical Bayes approach that provides good results when all the observations are used to estimate the value of the parameter c in the μ_i's prior distribution. This methodology implies that we use a prior distribution which depends on the observed data. However, as Casella (2001) pointed out, data dependent priors are perfectly valid and are frequently used in the Bayesian literature.

Acknowledgments

The authors are grateful to an anonymous referee for his helpful comments, which helped to improve an earlier version of this article. This research was partially supported by the Ministerio de Ciencia e Innovación, Spain, grant MTM 2008–03282 and Universidad Complutense de Madrid–Banco Santander Central Hispano. The first author was also partially supported by the Ministerio de Ciencia e Innovación, Spain, grant MTM 2008–00166.

References

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

Bansal, N. K. (2007). Decision theoretic Bayesian hypothesis testing with the selection goal. *Statist. Dec.* 25:19–39.

Barbieri, M., Berger, J. (2004). Optimal predictive model selection. *Ann. Statist.* 32:870–897.

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

Casella, G. (2001). Empirical Bayes Gibbs sampling. *Biostatistics* 2(4):485–500.

Datta, S., Datta, S. (2005). Empirical Bayes screening of many *p*-values with applications to microarray studies. *Bioinformatics* 21(9):1987–1944.

Do, K.-A., Müller, P., Tang, F. (2005). A nonparametric Bayesian mixture model for gene expression. *J. Roy. Statist. Soc. C* 54(3):627–644.

Dudoit, S., Shaffer, J. P., Boldrick, J. C. (2003). Multiple hypothesis testing in microarray experiments. *Statist. Sci.* 18(1):71–103.

Efron, B. (2003). Robbins, empirical Bayes and microarrays. *Ann. Statist.* 31:366–378.

Efron, B., Tibshirani, R. (2007). On testing the significance of sets of genes. *Ann. Appl. Statist.* 1(1):107–129.

Efron, B., Tibshirani, R., Storey, J. D., Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* 96:1151–1160.

Gordon, A., Glazko, G., Qiu, X., Ykovlev, A. (2007). Control of the mean number of false discoveries, Bonferroni and stability of multiple testing. *Ann. Appl. Statist.* 1(1):179–190.

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

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

Newton, M. A., Kendziorski, C. M., Richmond, C. S., Blatter, F. R., Tsui, K. W. (2001). On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data. *J. Comp. Biol.* 8:37–52.

Robert, C. P., Casella, G. (2004). *Monte Carlo Statistical Methods*. 2nd ed. New York: Springer.

Robbins, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* Vol. 1. Berkeley: University California Press, pp. 157–163.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* 35:1–20.

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

Shaffer, J. P. (1995). Multiple hypothesis testing: a review. *Ann. Rev. Psychol.* 46:561–584.

Storey, J. D. (2003). The positive false discovery rate: a Bayesian interpretation and the *q*-value. *Ann. Statist.* 31(6):2013–2035.

Sun, W., Cai, T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *J. Amer. Statist. Assoc.* 102(479):901–912.

Waller, R. A., Duncan, D. B. (1969). A Bayes rule for the symmetric multiple comparison problem. *J. Amer. Statist. Assoc.* 64:1484–1503.