

# Laplace approximated EM Microarray Analysis: an empirical Bayes approach for comparative microarray experiments

Haim Bar\*, James Booth<sup>†</sup>, Elizabeth Schifano\*, Martin T. Wells<sup>‡</sup>

April 2009

## Abstract

A two groups mixed-effects model for the comparison of (normalized) microarray data from two treatment groups is considered. Most competing parametric methods that have appeared in the literature are obtained as special cases or by minor modification of the proposed model. Approximate maximum likelihood fitting is accomplished via a fast and scalable algorithm, which we call LEMMA (Laplace approximated EM Microarray Analysis). The posterior odds of treatment  $\times$  gene interactions, derived from the model, involve shrinkage estimates of both the interactions and of the gene specific error variances. Genes are classified as being associated with treatment based on the posterior odds and the local false discovery rate (fdr) with a fixed cutoff. Our model-based approach also allows one to declare the nonnull status of a gene by controlling the false discovery rate (FDR). It is shown by a detailed simulation study that the approach performs very well and is more numerically stable, when compared with well known competitors. We apply our proposed methodology

---

\*Cornell University, Department of Statistical Science, 301 Malott Hall, Ithaca, NY 14853, USA.

<sup>†</sup>Cornell University, Department of Biological Statistics and Computational Biology. The support of NSF Grant DMS 0805865 is gratefully acknowledged.

<sup>‡</sup>Cornell University, Department of Biological Statistics and Computational Biology and Department of Statistical Science. The support of NIH Grant R01-GM083606-01 is gratefully acknowledged.

to two microarray examples. In principle the parametric model and fitting procedure can be generalized to more complex designs and data settings.

*Keywords:* EM algorithm; empirical Bayes; Laplace approximation; *LEMMA*; *LIMMA*; linear mixed models; local false discovery rate; microarray analysis; mixture model, two-groups model.

## 1 Introduction

Microarray technologies have become a major data generator in the postgenomics era. Instead of working on a gene-by-gene basis, microarray technologies allow scientists to view the expression of thousands of genes from an experimental sample simultaneously. Due to the cost, it is common that thousands of genes are measured with a small number of replications, as a consequence, one faces a *large G, small n* problem, where  $G$  is the total number of genes and  $n$  is the number of replications. After preprocessing of the raw image data, the expression levels are often assumed to follow a two-groups model, that is, the expressions are each either *null* or *nonnull* with prior probability  $p_0$  or  $p_1 = 1 - p_0$ , respectively. The two-groups model plays an important role in the Bayesian microarray literature and is broadly applicable (Efron, 2008).

A general review of issues pertaining to microarray data analysis is provided in Allison et al. (2006). Here, we focus on statistical inference, and in particular, on what Allison et al. (2006) refer to as “consensus points 2 and 3”: the advantages of shrinkage methods, and controlling the false discovery rate. We review several inferential methods, and develop a unifying linear model approach.

Classical parametric statistics does not give a reliable methodology for determining differentially expressed genes. The large number of genes with

relatively few replications in microarray experiments yield variance estimates of the expression levels that are often unreliable. The classical t-test and F-test are generated under a heterogeneous error variance model assumption and do not enjoy the advantage gained by shrinkage estimation. The assumption that the variances are equal across all genes is unlikely to be true. Hypothesis tests based on a pooled common variance estimator for all genes have low power and can lead to misleading differential expression results (Wright and Simon, 2003; Cui et al., 2005).

An important observation is that although there are only a few replications for each gene, there are a very large number of total measurements. If one combined information across the genes (i.e. genome-wide shrinkage), then one can construct test procedures that have improved power. The SAM test (Tusher et al., 2001) and a moderated t-test in Efron et al. (2001) first used information across the genome-wide expression values by the addition of a data-based constant to the gene-specific standard errors.

The Bayesian approach seems to be particularly well-suited for combining information in expression data. Instead of directly modeling the variation of the expression data, Bayesian models are characterized by mixing measurements over a latent gene-specific component. Hierarchical Bayesian models have also been used for variance regularization by estimating moderate variances of individual genes. The estimated variances are calculated as weighted averages of the gene-specific sample variances and pooled variances across all genes. In particular, the regularized t-test proposed by Baldi and Long (2001) uses a hierarchical model and substitutes a Bayesian variance estimator based on a prior distribution in place of the usual variance estimate. A hierarchical gamma-gamma model was developed in Newton et al. (2001) for detecting changes of gene expression in a two-channel cDNA microarray experiment

and was extended to replicate chips with multiple conditions using a hierarchical lognormal-normal model in Kendzierski et al. (2003). Lonnstedt and Speed (2002) also used an empirical Bayes approach to derive the so-called B-statistic as the logarithm of the posterior odds of differential expression. Smyth (2004) extended the B-statistic to the linear models setting and has written the widely used *LIMMA R* package (R Development Core Team, 2007). The Bayes approach with a random gene error variance essentially yields test statistics equivalent to the moderated-t and F-statistics statistics of Wright and Simon (2003) and Cui et al. (2005). Compared to a number of existing tests, Cui et al. (2005) showed that their proposed F-type test, using a James-Stein type variance estimator, had the best or nearly the best power for detecting differentially expressed genes over a wide range of situations.

If the power of the F-type tests with a random gene error variance (i.e. shrinking the variances) is higher than in the case were the variance is fixed, why only model the random gene error variances but not the means? By putting a prior distribution on the means, Hwang and Liu (2007) proposed a new empirical Bayes likelihood ratio which shrinks both the means and variances. Their simulation studies indicate that the tests that shrink both the means and variances are more powerful than all the other tests existing in the literature. However, the Hwang and Liu (2007) test is computationally expensive and their estimation of the model hyperparameters is somewhat ad hoc and complicated.

The development of the empirical Bayes methodologies that improve the power of detecting differentially expressed genes essentially reduces to the issue of modeling whether gene-specific effects should be modeled as fixed or random. This question applies to effects on both the mean and the error variance. Thus, there are four combinations of fixed and random factors

leading to four models which we denote by FF, RF, FR and RR, where the first letter identifies whether the mean effects are fixed or random and the second letter does the same for the error variances. Two additional models denoted FH and RH, are obtained if the error variances are assumed to be homogeneous across genes. The FF category corresponds to the naive approach of applying t- or F-tests to each gene separately. The FR category includes the models in Wright and Simon (2003) and Cui et al. (2005). The gamma-gamma and log-normal-normal models of Newton et al. (2001) and Kendzierski et al. (2003) are of the RH type. The approach of Hwang and Liu (2007) falls in the RR category. The RR category also includes the B-statistic of Lonnstedt and Speed (2002) and Smyth (2004), with the caveat that the shrinkage of mean effects in numerator of the test statistic cancels with a key part of the shrinkage of the posterior variance in the denominator of the B-statistic resulting in a statistic of type FR. It also turns out that the mixture proportion parameter,  $p_1$ , is difficult to estimate in the models leading to the B-statistic, which may explain why it is set by default to 0.01, and not estimated, in the *LIMMA R* package. A summary of how previously proposed statistics fall into the six model categories is given in Table 1.

We propose a new and unified formulation of an empirical Bayes likelihood ratio procedure which shrinks both means and variances. In this paper we focus on a simple two-condition experimental setup, but the ANOVA formulation we posit in the next section allows for easy generalization to more than two conditions and comparisons based on a single sample of two channel arrays such as the more general designs in Kerr et al. (2000) and Smyth (2004). The methods of this article can, in principle, also be extended to a multivariate empirical Bayes model, for example, to analyze time-course data as in the extension of the B-statistic by Tai and Speed (2006); or multiple

Mean effect	Error variance	Methods
<b>Fixed</b>	<b>Fixed (Heterogenous)</b>	t-test/F-test
<b>Fixed</b>	<b>Fixed (<b>H</b>omogenous)</b>	$F_3$ in Cui and Churchill (2003)
<b>Fixed</b>	<b>R</b> andom	Wright and Simon (2003), Cui et al. (2005), Lonnstedt and Speed (2002), Smyth (2004)
<b>R</b> andom	<b>Fixed (Heterogenous)</b>	
<b>R</b> andom	<b>Fixed (<b>H</b>omogenous)</b>	Newton et al. (2001), Kendziorski et al. (2003)
<b>R</b> andom	<b>R</b> andom	$F_{SS}$ in Hwang and Liu (2007), Lonnstedt and Speed (2002), Smyth (2004)

Table 1: Models corresponding to combinations of fixed and random factors.

array platforms as is used in epigenomic data analysis (Figuroa et al., 2008).

We apply an approximate EM algorithm for fitting our proposed model. The integral needed to evaluate the complete data likelihood makes direct application of the EM algorithm intractable. However, a simple and accurate approximation is obtained via the Laplace approximation. The Laplace approximation makes the EM algorithm scalable, tractable, and extremely fast. Implementation of Bayes microarray models typically involves drawing MCMC samples from the posterior distribution of effects from all genes. MCMC sampling provides a mechanism to study the full Bayesian posterior distribution. However, there is a heavy computational burden that often makes the MCMC implementation infeasible. The Laplace approximation circumvents the generation of the thousands of gene effect parameters and gives a highly accurate approximation to the integral in the expression of the complete data likelihood. The Laplace approximated EM algorithm based analysis is the inspiration of the acronym *LEMMA* (**L**aplace approxi-

mated **EM Microarray Analysis**). R code and examples are available online at <http://www.stat.cornell.edu/lemma/>.

The paper is organized as follows. In Section 2, we introduce the necessary notions and the two-groups model along with the prior distribution specifications. In Section 3, we propose an approximate EM algorithm for fitting the RR model. We also explain in this section the cause of identifiability issue in the *LIMMA* model, and propose a solution. In Section 4 we approximate the posterior probability that a gene is nonnull and give a posterior t-statistic that is used as the test statistic to identify the differentially expressed genes. A detailed simulation study is given in Section 5. In Section 6 we apply our proposed methodology to two well known microarray examples: the ApoA1 data (Callow et al., 2000) and Colon Cancer data (Alon et al., 1999). We conclude the article in Section 7 with a brief discussion. In the appendix, we present an extension to multiple treatments and provide details on some of our previous derivations of the likelihood and empirical Bayes estimates.

## 2 Model and Notation

Let  $y_{ijg}$  denote the response (e.g. log expression ratio) of gene  $g$ , for subject (replicate)  $j$ , in treatment group  $i = 1, 2$ . We begin with the linear model,

$$y_{ijg} = \mu + \tau_i + \gamma_g + \psi_{ig} + \epsilon_{ijg}, \quad (1)$$

with a typical assumption concerning the errors being

$$\epsilon_{ijg} \sim \text{iid } N(0, \sigma_{\epsilon, g}^2) \quad (2)$$

for  $j = 1, \dots, n_{ig}$ , independently across genes and treatment groups. We impose the identifiability constraints,  $\tau_1 + \tau_2 = 0$  and  $\psi_{1g} + \psi_{2g} = 0$  for all  $g = 1, \dots, G$ . Then  $\tau = \tau_1 - \tau_2$  is the main effect of treatment, averaged

across genes, which we expect to be essentially zero in many applications, and  $\psi_g = \psi_{1g} - \psi_{2g}$ ,  $g = 1, \dots, G$ , are the gene specific treatment effects.

We further suppose that the genes fall into two groups, a *null* group in which  $\psi_g \equiv 0$  and a *nonnull* group in which  $\psi_g \neq 0$ . The primary goal is to classify genes as null or nonnull based on the observed responses. A probabilistic approach is to suppose that each gene has prior probability  $p_1$  of being nonnull (and  $p_0 = 1 - p_1$  of being null) and to use Bayes rule to determine the posterior probability given the data; specifically,

$$p_{1,g}(y_g) = \frac{p_1 f_{1,g}(y_g)}{p_0 f_{0,g}(y_g) + p_1 f_{1,g}(y_g)}, \quad (3)$$

where  $f_{1,g}(y_g)$  is the probability density of the responses for gene  $g$  implied by the nonnull model, and  $f_{0,g}(y_g)$  is the corresponding quantity if the gene is in the null group.

In practice, of course, the mixture probability and the parameters that determine the null and nonnull densities have to be estimated. This estimation step depends upon additional assumptions, if any, that are made about the distribution of the responses. A basic question is whether gene-specific effects should be modeled as a fixed or random. This question applies to effects on both the mean and the error variance. Thus, there are four combinations of fixed and random factors leading to four models which we denote by FF, RF, FR and RR, where the first letter identifies whether the mean effects are fixed or random and the second letter does the same for the error variances. Two additional models, denoted FH and RH, are obtained if the error variances are assumed to be homogeneous across genes; i.e.  $\sigma_{\epsilon,g}^2 \equiv \sigma_{\epsilon}^2$ .

The ANOVA model (1) together with the distributional assumption (2) allows us to restrict attention to the sum and difference of gene-specific treatment means, respectively  $s_g = \bar{y}_{1,g} + \bar{y}_{2,g}$  and  $d_g = \bar{y}_{1,g} - \bar{y}_{2,g}$ , and the gene-

specific mean squared errors,

$$m_g = \sum_{i=1}^2 \sum_{j=1}^{n_{ig}} (y_{ijg} - \bar{y}_{i\cdot g})^2 / f_g,$$

where  $f_g = n_{1g} + n_{2g} - 2$ . Notice that  $s_g|g \sim N[2\mu + 2\gamma_g, \sigma_g^2]$ , where  $\sigma_g^2 \equiv \sigma_{\epsilon, g}^2(1/n_{1g} + 1/n_{2g})$ , and  $|g$  denotes conditioning on any gene-specific random effects. It follows that  $s_g$  carries no information about the gene-specific treatment effect  $\psi_g$ . For this reason, our estimation procedures use only the marginal likelihood based on the data  $(\{d_g\}, \{m_g\})$ . The model (1) together with assumption (2) also imply that  $d_g$  and  $m_g$  are conditionally independent, with  $d_g|g \sim (1-b_g)N_0 + b_gN_1$  independently of  $m_g|g \sim \sigma_{\epsilon, g}^2 \chi_{f_g}^2 / f_g$ , where  $b_g, g = 1, \dots, G$ , denote independent Bernoulli( $p_1$ ) latent indicators of nonnull status for the  $G$  genes,  $N_0$  and  $N_1$  denote normal variates with unequal means  $\tau$  and  $\tau + \psi_g \neq \tau$  respectively, but equal variances  $\sigma_g^2$ , and  $\chi_{f_g}^2$  denotes a chisquared variate with  $f_g$  degrees of freedom.

The family of parametric models considered in this paper is completed by specifying distributions for the gene-specific effects,  $\{\psi_g\}$  and  $\{\sigma_{\epsilon, g}^2\}$ . In what follows we suppose that, if the (nonnull) gene-specific interaction effects are modeled as random variates, they follow a normal distribution,

$$\psi_g \sim \text{i.i.d. } N(\psi, \sigma_\psi^2). \quad (4)$$

On the other hand, if the gene-specific variances are modeled as random variates, they are drawn from an inverse gamma distribution,

$$\sigma_{\epsilon, g}^{-2} \sim \text{i.i.d. } \text{Gamma}(\alpha, \beta), \quad (5)$$

where  $\alpha$  and  $\beta$  are shape and scale parameters. We note here that allowing the mean of the random effects distribution (4) to be non-zero is quite natural

in the two groups setting. We see no reason why the average gene-specific effect should be zero in the nonnull group. This assumption does not introduce any identifiability issues as it does in classical (one group) mixed models. In fact, it is worth contrasting (4) with the corresponding assumption in the models leading to the B-statistic given in Lonnstedt and Speed (2002) and Smyth (2004), where the mean of the random effects distribution is assumed to be zero. We argue later (in subsection 3.4) that this assumption may be the cause of the difficulty in estimating the mixture proportion parameter,  $p_1$ .

### 3 Estimation

In this section we describe in detail an approximate EM algorithm for fitting the RR model. Estimation for the other five models can be carried out by making appropriate modifications to this algorithm. The RR model has six parameters, two being the shape and scale of the distribution for the error variances given in (5). The remaining vector of parameters is  $(p_1, \tau, \psi, \sigma_\psi^2)$  which we denote by  $\phi$ .

Estimates of the hyperparameters,  $\alpha$  and  $\beta$ , are obtained by maximizing the marginal likelihood based on  $\{m_g\}$ , given by

$$\begin{aligned}
 L(\{m_g\}) &= \prod_{g=1}^G \int_0^\infty f(m_g | \sigma_{\epsilon,g}^2) f(\sigma_{\epsilon,g}^{-2}) d\sigma_{\epsilon,g}^{-2} \\
 &= \prod_{g=1}^G \frac{m_g^{\frac{f_g}{2}-1} \left(\frac{f_g}{2}\right)^{\frac{f_g}{2}} \Gamma\left(\frac{f_g}{2} + \alpha\right)}{\Gamma\left(\frac{f_g}{2}\right) \Gamma(\alpha) \beta^\alpha \left(\frac{m_g f_g}{2} + \frac{1}{\beta}\right)^{\frac{f_g}{2} + \alpha}}. \tag{6}
 \end{aligned}$$

### 3.1 EM Algorithm

We apply the EM algorithm to estimate  $\phi$ , with the latent indicators,  $\{b_g\}$ , playing the role of the missing data. Since  $d_g$  and  $m_g$  are conditionally independent given  $(b_g, \sigma_{\epsilon,g}^2)$ , the complete data likelihood for  $\phi$  based on  $(\{b_g\}, \{d_g\}, \{m_g\})$  is

$$L_C(\phi) = \prod_{g=1}^G \int L(b_g, d_g; \sigma_{\epsilon,g}^2) L(m_g; \sigma_{\epsilon,g}^2) f(\sigma_{\epsilon,g}^{-2}) d\sigma_{\epsilon,g}^{-2}, \quad (7)$$

where  $f(\sigma_{\epsilon,g}^{-2})$  represents the gamma density with shape  $\alpha$  and scale  $\beta$ .

The integral in (7) makes direct application of the EM algorithm intractable. However, a simple and accurate approximation is obtained via the Laplace approximation

$$L_C(\phi) \approx \tilde{L}_C(\phi) \equiv \prod_{g=1}^G L(b_g, d_g; \tilde{\sigma}_{\epsilon,g}^2) L(m_g; \tilde{\sigma}_{\epsilon,g}^2) f(\tilde{\sigma}_{\epsilon,g}^{-2}) \sqrt{-2\pi/\ell''(m_g; \tilde{\sigma}_{\epsilon,g}^2)}, \quad (8)$$

where  $\tilde{\sigma}_{\epsilon,g}^2$  is posterior mode of  $\sigma_{\epsilon,g}^2$  given  $m_g$  given by

$$\tilde{\sigma}_{\epsilon,g}^2 = \frac{f_g/2}{f_g/2 + \alpha + 1} m_g + \frac{\alpha + 1}{f_g/2 + \alpha + 1} \cdot \frac{1}{(\alpha + 1)\beta}. \quad (9)$$

Notice that the last three factors on the right side of (8) do not involve the parameter  $\phi$  and can therefore be ignored in the implementation of EM. In practice, we numerically maximize (6) with respect to  $\alpha$  and  $\beta$  to find their estimates, and use these in computing (9).

Denote the estimate after  $m$  iterations of EM by  $\phi^{(m)}$ . The  $(m + 1)$ st E-step consists of taking the conditional expectation of the logarithm of (7) given the observed data, using the current estimate,  $\phi^{(m)}$ . Using the Laplace

approximation (8), this is given by

$$\begin{aligned}
Q(\phi, \phi^{(m)}) &= E_{\phi^{(m)}} [\log L_C(\phi) | \{d_g\}, \{m_g\}] \\
&\approx E_{\phi^{(m)}} \left[ \log \tilde{L}_C(\phi) | \{d_g\}, \{m_g\} \right] \\
&\propto \sum_{g=1}^G E_{\phi^{(m)}} \{ \log L(b_g, d_g; \tilde{\sigma}_{\epsilon, g}^2) | d_g \} \\
&= \sum_{g=1}^G \left\{ p_{0,g}^{(m)} \log[p_{0,g} f_{0,g}(d_g)] + p_{1,g}^{(m)} \log[p_{1,g} f_{1,g}(d_g)] \right\} \\
&\equiv \tilde{Q}(\phi, \phi^{(m)})
\end{aligned} \tag{10}$$

where  $f_{0,g}$  and  $f_{1,g}$  denote  $N(\tau, \tilde{\sigma}_g^2)$  and  $N(\tau + \psi, \sigma_\psi^2 + \tilde{\sigma}_g^2)$  densities with  $\tilde{\sigma}_g^2 = \tilde{\sigma}_{\epsilon, g}^2 (1/n_{1g} + 1/n_{2g})$ ,  $p_{1,g} = E(b_g | d_g)$  and  $p_{0,g} + p_{1,g} = 1$ .

The M-step at the  $(m+1)$  iteration requires maximization of  $\tilde{Q}(\phi, \phi^{(m)})$  with respect to  $\phi$  to yield the updated estimate  $\phi^{(m+1)}$ . That is,

$$\phi^{(m+1)} = \arg \max_{\phi} \tilde{Q}(\phi, \phi^{(m)}) .$$

This leads to the following MLE update equations for  $p_1$ ,  $\tau$  and  $\psi$ ,

$$p_1^{(m+1)} = \frac{1}{G} \sum_{g=1}^G p_{1,g}^{(m)} , \tag{11}$$

$$\tau^{(m+1)} = \frac{\sum_{g=1}^G p_{0,g}^{(m)} d_g / \tilde{\sigma}_g^2}{\sum_{g=1}^G p_{0,g}^{(m)} / \tilde{\sigma}_g^2} , \tag{12}$$

and

$$\psi^{(m+1)} = \frac{\sum_{g=1}^G p_{1,g}^{(m)} (d_g - \tau^{(m+1)}) / (\sigma_\psi^{2(m)} + \tilde{\sigma}_g^2)}{\sum_{g=1}^G p_{1,g}^{(m)} / (\sigma_\psi^{2(m)} + \tilde{\sigma}_g^2)} , \tag{13}$$

while the update for  $\sigma_\psi^2$  is the solution of the equation,

$$\sum_{g=1}^G p_{1,g}^{(m)} \frac{1}{\sigma_\psi^2 + \tilde{\sigma}_g^2} = \sum_{g=1}^G p_{1,g}^{(m)} \frac{(d_g - \tau^{(m+1)} - \psi^{(m+1)})^2}{(\sigma_\psi^2 + \tilde{\sigma}_g^2)^2} . \tag{14}$$

Strictly speaking the update for  $\psi$  in (13) is conditional on the current value of  $\sigma_\psi^2$ . However, we have found this variant of EM to have almost identical convergence properties to the full EM in which  $Q$  is maximized jointly with respect to all four components of  $\phi$ .

### 3.2 Modifications for RF, RH, FR, FF, FH

The complete data likelihood for the RF model is

$$L_C = \prod_{g=1}^G L(b_g, d_g; \sigma_{\epsilon,g}^2) L(m_g; \sigma_{\epsilon,g}^2). \quad (15)$$

Since no integration is required to evaluate this likelihood the Laplace approximation is not needed in this case. As for the RR model we first estimate the error variances,  $\{\sigma_{\epsilon,g}^2\}$ , separately using the marginal likelihood for  $\{m_g\}$ . This results in the simple estimate,  $\hat{\sigma}_{\epsilon,g}^2 = m_g$ . The EM algorithm for estimating  $\phi$  then proceeds in an identical manner except that  $\tilde{\sigma}_g^2$  is replaced by  $\hat{\sigma}_g^2 = \hat{\sigma}_{\epsilon,g}^2(1/n_{1g} + 1/n_{2g})$ . The algorithm for the RH model is also similar with the marginal likelihood estimator of the homogeneous error variance given by  $\hat{\sigma}_\epsilon^2 = \sum_g m_g f_g / \sum_g f_g$ .

For all the fixed interaction effects models (FR, FF and FH) it is easily verified that  $d_g - \tau^{(m)} - \psi_g^{(m)} = 0$ . This implies that the EM update for the mixing parameter satisfies

$$p_1^{(m+1)} = \frac{1}{G} \sum_{g=1}^G \frac{p_1^{(m)}}{p_0^{(m)} \exp\left\{-\frac{(d_g - \tau^{(m)})^2}{2\hat{\sigma}_{\epsilon,g}^2}\right\} + p_1^{(m)}} > p_1^{(m)},$$

where  $\hat{\sigma}_{\epsilon,g}^2$  represents the appropriate  $\sigma_g^2$  estimator for the desired model. As a result the EM sequence for  $p_1$  always converges to 1, regardless of the starting value. An explanation for this behavior is that the mixture probability is not identifiable if the interaction effects are fixed.

### 3.3 Generalized *LIMMA*

The *LIMMA* model proposed by Smyth (2004) is a similar model to the one described in Section 2. A key difference is the assumption concerning the random interactions given in (4). The corresponding assumption in *LIMMA* is  $\psi_g | \sigma_{\epsilon,g}^2 \sim N(0, \eta \sigma_{\epsilon,g}^2)$ . This assumption, combined with (5), results in a closed form expression for the complete data likelihood (7), rendering the use of the Laplace approximation unnecessary. Another difference is that the parameter  $\tau$  is assumed to be zero in the *LIMMA* model. However, this difference has little bearing on the arguments that follow.

As noted in Section 2, it is somewhat unnatural and unnecessary to assume that the mean of the nonnull gene-specific effects is zero. Hence we consider a generalized *LIMMA* (*gLIMMA*) model with

$$\psi_g | \sigma_{\epsilon,g}^2 \sim N(\psi, \eta \sigma_{\epsilon,g}^2) \quad (16)$$

for the nonnull gene-specific interactions. The EM algorithm discussed earlier in this section can be implemented to fit this generalized model with minor modifications. Specifically, after using the Laplace approximation, the Q-function has the same form as (10) with  $\eta_g \tilde{\sigma}_{\epsilon,g}^2$  replacing  $\sigma_\psi^2 + \tilde{\sigma}_g^2$  as the variance in the nonnull density  $f_{1,g}$ , where  $\eta_g = \eta + 1/n_{1g} + 1/n_{2g}$ . This leads to update equations for  $p_1$  and  $\tau$  identical to (11) and (12) respectively. The update for  $\psi$  is

$$\psi^{(m+1)} = \frac{\sum_{g=1}^G p_{1,g}^{(m)} (d_g - \tau^{(m+1)}) / (\eta_g \tilde{\sigma}_{\epsilon,g}^2)}{\sum_{g=1}^G p_{1,g}^{(m)} / (\eta_g \tilde{\sigma}_{\epsilon,g}^2)},$$

and the update of  $\eta$  satisfies

$$\sum_{g=1}^G p_{1,g}^{(m)} \frac{1}{\eta_g} = \sum_{g=1}^G p_{1,g}^{(m)} \frac{(d_g - \tau^{(m+1)} - \psi^{(m+1)})^2}{\eta_g^2 \tilde{\sigma}_{\epsilon,g}^2}.$$

These updates simplify further if the sample sizes are the same for all genes.

### 3.4 Identifiability issues

Consider the *gLIMMA* model in the equal sample size case,  $n_{1g} = n_{2g} = n$ . Conditional on the error variance,  $\sigma_{\epsilon,g}^2$ ,  $d_g \sim N[\tau, (2/n)\sigma_{\epsilon,g}^2]$  if gene  $g$  is in the null group, and  $d_g \sim N[\tau + \psi, (\eta + 2/n)\sigma_{\epsilon,g}^2]$  if gene  $g$  is in the nonnull group. Hence, the posterior probability that a gene is nonnull simplifies to  $p_{1,g} = p_1/(p_0x_g + p_1)$ , where

$$x_g = \sqrt{\frac{\eta + 2/n}{2/n}} \exp \left\{ \frac{1}{2} \left[ \frac{(d_g - \tau - \psi)^2}{(\eta + 2/n)\sigma_{\epsilon,g}^2} - \frac{(d_g - \tau)^2}{(2/n)\sigma_{\epsilon,g}^2} \right] \right\}.$$

It follows that, if gene  $g$  is in the null group,

$$x_g \sim \sqrt{n} \exp \left\{ -\frac{1}{2} \left( X_g - \frac{\psi^2}{\eta\sigma_{\epsilon,g}^2} \right) \right\} \quad (17)$$

as  $n \rightarrow \infty$ , where  $X_g$  denotes a chisquared variate with one degree of freedom. (Applying analogous arguments under the *LEMMA* model assumptions leads to expression (17) with  $\eta\sigma_{\epsilon,g}^2$  replaced by  $\sigma_{\psi}^2$ .) In contrast, if gene  $g$  is in the nonnull group,  $x_g = O_p(\sqrt{n}e^{-n})$ . Consequently  $p_{1,g}$  converges to 0 or 1 and genes are guaranteed to be correctly classified as the sample size increases. However, if the number of replicates is small or moderate, which is often the case in practice, there is a greater degree of randomness in the classification of null genes, particularly if  $\psi^2/\eta$  is close to zero. This behavior may explain why the mixture parameter is set by default equal to 0.01, and not estimated, in the *LIMMA* R package (Smyth, 2005), where the underlying model assumes  $\psi = 0$ .

## 4 Inference

The posterior probability that gene  $g$  is nonnull is given by the expression (3). Its estimated value based on the RR model can be reexpressed in terms

of the likelihood ratio

$$\begin{aligned}
\frac{L_{0,g}}{L_{1,g}} &\equiv \frac{\hat{f}_{0,g}}{\hat{f}_{1,g}} = \frac{(2\pi\tilde{\sigma}_g^2)^{-1/2} \exp\{-(d_g - \hat{\tau})^2/2\tilde{\sigma}_g^2\}}{[2\pi(\hat{\sigma}_\psi^2 + \tilde{\sigma}_g^2)]^{-1/2} \exp\left\{-\frac{(d_g - \hat{\tau} - \hat{\psi})^2}{2(\hat{\sigma}_\psi^2 + \tilde{\sigma}_g^2)}\right\}} \\
&= \left(\frac{\tilde{\sigma}_g^2}{\hat{\sigma}_\psi^2 + \tilde{\sigma}_g^2}\right)^{-1/2} \exp\left\{-\frac{1}{2} \frac{[\hat{\lambda}_g(d_g - \hat{\tau}) + (1 - \hat{\lambda}_g)\hat{\psi}]^2}{\hat{\lambda}_g\tilde{\sigma}_g^2} + \frac{\hat{\psi}^2}{2\hat{\sigma}_\psi^2}\right\} \\
&\propto \left(\frac{\tilde{\sigma}_g^2}{\hat{\sigma}_\psi^2 + \tilde{\sigma}_g^2}\right)^{-1/2} \exp\left\{-\frac{1}{2}T_g^2\right\}, \tag{18}
\end{aligned}$$

with the constant of proportionality being  $\exp(\hat{\psi}^2/2\hat{\sigma}_\psi^2)$ , where

$$\lambda_g = \frac{1}{\sigma_g^2} \left( \frac{1}{\sigma_g^2} + \frac{1}{\sigma_\psi^2} \right)^{-1} = \frac{\sigma_\psi^2}{\sigma_\psi^2 + \sigma_g^2}.$$

The statistic  $T_g$  is a posterior t-statistic, being the ratio of the estimated posterior expectation of  $\psi_g$  to its estimated posterior standard deviation. The likelihood ratio has the same form for the RF and RH models with  $\tilde{\sigma}_{\epsilon,g}^2$  replaced by  $\hat{\sigma}_{\epsilon,g}^2$  and  $\hat{\sigma}_\epsilon^2$  respectively in  $\tilde{\sigma}_g^2$ . (Recall that  $\sigma_g^2 = \sigma_{\epsilon,g}^2(1/n_{1g} + 1/n_{2g})$ .) Test statistics for the fixed mean effects models, FR, FF and FH, are obtained as limits of  $T_g$  as  $\hat{\lambda}_g \rightarrow 1$ .

It is interesting to compare the likelihood-ratio (18) with the corresponding statistic under the *LIMMA* and *gLIMMA* assumptions discussed in the previous section. For these models  $\sigma_\psi^2$  is replaced by  $\eta\tilde{\sigma}_{\epsilon,g}^2$ , and so the shrinkage coefficient becomes

$$\lambda_g = \frac{\eta}{\eta + 1/n_{1g} + 1/n_{2g}}.$$

In particular, if the sample sizes are the same for all genes, then the amount of shrinkage is the same for all genes. Furthermore, if  $\psi$  is set equal to zero, as it is in *LIMMA*, then  $\lambda_g$  cancels out in the numerator and denominator and  $T_g$  reduces to the test statistic for the FR model,

$$T_g = \frac{d_g - \hat{\tau}}{\tilde{\sigma}_g}.$$

This has the same form as the moderated t-statistic of Smyth (2004) and Wright and Simon (2003) except for the subtraction of the average (null) gene effect,  $\tau$ , in the numerator and the use of the mode rather than the expected value of the posterior distribution of  $\sigma_{\epsilon,g}^2$  given  $m_g$  in the denominator.

For inference, we compare the posterior null probability,  $1 - p_{1,g}$  in (3), with a local fdr threshold to decide whether a gene is in the nonnull group. Alternatively, our model-based approach also allows one to declare the non-null status of a gene by controlling the false discovery rate (FDR), using the (Benjamini and Hochberg (1995)) (BH) procedure for any given level,  $q^*$ . Specifically, using the theoretical null-gene distributions of  $\{d_g\}$ , which are assumed to be  $N(\hat{\tau}, \tilde{\sigma}_g^2)$ , we obtain the p-values for the observed  $\{d_g\}$ . We denote the p-values by  $\{P_g\}$ , and find the largest index  $g'$  for which  $P_{g'}^F \leq q^* \times g'/G$ , where  $\{P_g^F\}$  is the sorted list of p-values. We declare all the genes with index smaller than or equal to  $g'$  (in the sorted list) as nonnull, and the FDR theorem guarantees that the expected false discovery rate is bounded by  $q^*$ .

## 5 Simulation Study

In the following section, we summarize several simulation studies that show the relative performance of the RR model and associated statistics. We used two different data generation set-ups, one according to the RR model and the other according to the *gLIMMA* model, to assess performance of various statistics in terms of average power and average number of false detections. We considered two metrics to determine null/nonnull status, one based on local false discovery rate (fdr) criteria (Efron, 2008) and the other based on test-specific genome-wide critical values. We also assessed

parameter estimation under the various models in terms of bias and variance.

## 5.1 Data Generation

The data generation process was identical for both scenarios, except for the specification of the variance of the gene-specific treatment effect. In both scenarios, we simulated  $S = 20$  datasets according to a mixture model with two groups, null and nonnull, each with  $G = 2000$  genes. In order to accomplish this, we created a random  $G \times S$  0 – 1 table,  $B = \{B_{gs}\}$ , with fixed marginals to establish null ( $B_{gs} = 0$ ) and nonnull ( $B_{gs} = 1$ ) status, such that  $\sum_{s=1}^{20} B_{gs} = 1$  (one nonnull gene in each row) and  $\sum_{g=1}^{2000} B_{gs} = 100$  (100 nonnull genes in each column/dataset). For each column, this set-up corresponded to the latent Bernoulli variables  $\{b_g\}$  with probability  $p_1 = 0.05$ . This further enabled us to assess average power and average number of false detections as described in Section 5.2 as a function of the mean nonnull gene-specific treatment effects,  $\psi$ .

We drew  $G \times S$  inverse gamma error variates with shape  $\alpha$  and scale  $\beta$ , one for each cell in the table  $B$ . By varying  $\alpha$  and  $\beta$ , we adjusted the amount of error variance variability present in the data. The values of  $\alpha$ ,  $\beta$ ,  $n_{1g} \equiv n_1$ , and  $n_{2g} \equiv n_2$  were chosen so that  $\text{mean}(\sigma_g^2) = 1$ . With  $n_1 = n_2 = 6$ , we set  $\alpha = 5$  and  $\beta = 1/12$  for the ‘low’ error variance variability; we set  $\alpha = 2.1$  and  $\beta = 10/33$  for the ‘high’ error variance variability. Hence the standard deviation (and also the coefficient of variation, CV) of  $\sigma_g^2$  for the former was  $3^{-1/2}$ , and for the latter was  $\sqrt{10}$ .

For each  $\psi \in \{0, 1, 2, \dots, 6\} \equiv \Psi$ , we used the same table  $B$  and the same  $G \times S$  inverse gamma error variates. In the RR data generation, we set  $\sigma_\psi^2 = 1$ ; in the *gLIMMA* data generation, we set  $\eta = 1/3$  so that  $\sigma_\psi^2 = \eta \times \text{mean}(\sigma_{\epsilon,g}^2)$ . For both generation schemes, we set  $\tau = 0$ , as only the random gene effect

models (RR, RF, RH) are capable of estimating  $\tau$ .

We generated  $y_{ijg}$  according to equations (1) and (2) with the above parameter specifications, which allowed us to compute  $\{d_g\}$  and  $\{m_g\}$ . While we only present results for these specific parameter value settings, numerous simulations were performed with a variety of sample sizes  $n_i$ ,  $i = 1, 2$ , non-null probabilities  $p_1$ , and gene-specific treatment variances  $\sigma_{\psi}^2$ . In addition, we also considered using the log-normal distribution to generate the error variance  $\sigma_{\epsilon,g}^2$  rather than the inverse gamma distribution. We found that our results are insensitive to these different settings, and the results presented below portray an accurate summary of the performance of the methods.

## 5.2 Data Analysis and Results

We consider two metrics for determining null and nonnull status of genes. The first method was based on computing empirical quantile critical values, which required the computation of an empirical null distribution. The second method did not require this computation, and was based on local fdr criteria (Efron et al., 2001; Efron, 2005). After estimating the necessary parameters using the EM algorithm described in Section 3.1, we computed the desired test statistics and compared their performances in terms of average power and average number of false detections.

### 5.2.1 Empirical Quantile Critical Values

Since the distribution of many of our test statistics is unknown, we defined a test-specific critical value,  $T_c$ , as the 0.95 quantile among the  $1900 \times 20$  null genes. By design, this resulted in an average level of 0.05 for each test. For a generic test statistic  $T_{gs}$  corresponding to the  $B_{gs}$  cell in table  $B$ , the average power for the corresponding test was determined by the proportion

of nonnull genes correctly declared nonnull based on the critical value  $T_c$ ; i.e.,

$$\text{average power} = \frac{\sum_{g,s} B_{gs} I(T_{gs} \geq T_c)}{\sum_{g,s} B_{gs}}. \quad (19)$$

First, we compared the tests corresponding to the six versions of the ANOVA model (1) with error distribution (2), and the associated test for the *gLIMMA* model. Figure 1 shows the average power for each test plotted as a function of  $\psi$  for the two generation schemes (columns) under both low and high error variance variability (rows). Note that the observed power at  $\psi = 0$  is greater than 0.05 since this case does not correspond to the null distribution, which additionally requires  $\sigma_\psi^2 = 0$ .

We see in each case that the RR statistics display the highest average power for all  $\psi \in \Psi$ . In general, the random gene models (RR, RF, RH) demonstrate higher average power than their corresponding fixed gene counterparts. As expected, Figure 1 also shows that the average power in the homogeneous error variance models (RH, FH) decreases as the error variance variability increases. Additionally, it is evident from the plots that the *gLIMMA* statistics do not perform as well as the RR statistics, even when the data was generated under *gLIMMA* assumptions.

### 5.2.2 Local fdr

Efron et al. (2001) and Efron (2005) defined local fdr as

$$fdr(y_g) = \Pr(\text{null} | Y = y_g) \quad (20)$$

for the posterior probability of a gene  $g$  being in the null group. Note that this is precisely  $1 - p_{1,g}(y_g)$ , where  $p_{1,g}(y_g)$  is given by (3). Following the suggestions of Efron (2005), we used  $fdr(y_g) \leq 0.20$  as the cutoff point for declaring nonnull status. Thus, any gene with nonnull posterior probability greater than 0.80 was deemed nonnull. We used the knowledge of the

true null/nonnull classification to determine the average power and average number of false detections as a function of  $\psi$ .

Because we could not estimate  $p_1$  for all six models, we only considered the statistics associated with RR, RF and RH. For comparison, we also considered the *gLIMMA* statistics, two types of  $B$  statistics computed under the *LIMMA* assumptions, and local fdr statistics computed from the *locfdr* (Efron et al., 2008) R package (referred to as ‘Efron’ for simplicity). Two types of  $B$  statistics were used to differentiate between those statistics computed with the *LIMMA* (Smyth, 2005) R package default value of  $p_1 = 0.01$  (referred to as ‘B (p=default)’) and those computed with the true value of  $p_1$  specified in the simulation (referred to as ‘B (p=true)’).

Let  $fdr_{gs}$  be the local fdr statistic corresponding to the  $B_{gs}$  cell in table  $B$ . Analogous to equation (19), we defined average power in this scenario as

$$\text{average power} = \frac{\sum_{g,s} B_{gs} I(fdr_{gs} \leq 0.20)}{\sum_{g,s} B_{gs}}. \quad (21)$$

Furthermore, we defined the average number of false detections as

$$\text{average number of false detections} = \sum_{g,s} (1 - B_{gs}) I(fdr_{gs} \leq 0.20). \quad (22)$$

Figure 2 shows the average power and average number of false detections for data generated under the RR model with both low and high error variance variability. The RR statistics display higher average power than others, except for RF. However, the average number of false detections for the RF statistics, as well as for the RH statistics, appear quite poor in comparison to RR. For RR, there are less than ten false detections on average for each  $\psi$ , whereas for RF the number of false detections in the hardest case ( $\psi = 0$ ) is greater than 35 on average. The results under the *gLIMMA* data generation model shown in Figure 3 are similar. Notably, *gLIMMA* does not perform as well as RR even when the data is generated under the *gLIMMA* model.

### 5.3 Estimation Performance

In all of the simulations described above, we imposed an inverse gamma distribution on the error variance,  $\sigma_{\epsilon,g}^2$ , and estimated this quantity using the posterior mode,  $\tilde{\sigma}_{\epsilon,g}^2$ . We computed the posterior mode using marginal maximum likelihood estimates  $\hat{\alpha}$  and  $\hat{\beta}$ , as noted in Section 3.1. Figure 4 shows the distribution of estimates over  $S = 20$  replications for each  $\psi \in \Psi$  for the low error variance variability case in both the RR (top panel) and *gLIMMA* (bottom panel) data generation schemes. For both schemes, it is apparent that the estimates are quite close to their true values.

The RR, RH, RF, and *gLIMMA* models all used the EM algorithm (appropriately modified for each model) to estimate the model parameters. Under the RR data generation scheme, Figure 5 shows the distribution of estimates over  $S = 20$  replications for each  $\psi \in \Psi$ . Figure 5(a) indicates that the estimation of  $p_1$  was apparently quite challenging for the RF model, while the other three models tended to do a much better job. As expected, the accuracy of estimation of  $p_1$  improves as  $\psi$  increases. Indeed, the RR and RH models have very high precision and accuracy in the estimation of  $p_1$  when  $\psi = \{4, 5, 6\}$ . The estimation of  $\tau$  is quite good overall, as the scale of the y-axis in Figure 5(b) is quite small. Evidently, the RF model performed the worst in this regard, as it tended to under-estimate  $\tau$ . In Figure 5(c), we see that  $\psi$  is predominantly under-estimated by all models, especially by RF and *gLIMMA*. The RR and RH models tend to be least biased in the estimation of  $\psi$ . Finally, Figure 5(d) shows the distribution of  $\sigma_\psi^2$  estimates for RR, RH, and RF. Here we find that  $\sigma_\psi^2$  is predominantly over-estimated by the three models, with RR and RF being the least and most biased, respectively. Recall that the *gLIMMA* assumptions specify the gene-specific treatment effect variance as  $\eta\sigma_{\epsilon,g}^2$  rather than  $\sigma_\psi^2$ . Thus, while the data was

generated under the RR model assumptions with  $\sigma_\psi^2 = 1$ , *gLIMMA* can only estimate  $\eta$ . By our simulation design, estimates of  $\eta$  should be approximately 1/3. Figure 7(a) displays the corresponding boxplot for these estimates of  $\eta$ . For *gLIMMA*,  $\eta$  tends to be over-estimated for larger values of  $\psi$ , as well. It appears as though both RF and *gLIMMA* compensate for their severe under-estimation of  $\psi$  by over-estimating  $\sigma_\psi^2$  or  $\eta$ , respectively.

Figure 6 and Figure 7(b) show the analogous plots for the *gLIMMA* data generation scheme. Qualitatively, the results are quite similar. It is interesting to note, however, that *gLIMMA* does not perform better when the data is actually generated under the *gLIMMA* model. In general, the RR model is the most successful in the estimation of the model parameters under both data generation scenarios.

## 6 Examples

We applied the RR model to several microarray datasets. For illustration purposes, we provide our analysis of two publicly available, two-channel gene expression microarray datasets that were previously analyzed: the ApoA1 data (Callow et al., 2000) and Colon Cancer data (Alon et al., 1999).

### 6.1 ApoA1 Data

The ApoA1 experiment (Callow et al., 2000) used gene targeting in embryonic stem cells to produce mice lacking apolipoprotein A-1, a gene known to play a critical role in high density lipoprotein (HDL) cholesterol levels. Originally, 5,600 expressed sequence tags (EST) were selected. In our analysis, we used the data and normalization method provided with the *LIMMA* R package (Smyth, 2005), which consists of 5,548 ESTs, from 8 control (wild type “black six”) mice and 8 “knockout” (lacking ApoA1) mice. Common reference RNA

was obtained by pooling RNA from the control mice, and was used to perform expression profiling for all 16 mice.

The response of interest,  $y_{ijg}$ , is the  $\log_2$  fluorescence ratio (with respect to the common reference) where  $g$  is one of 5,548 genes,  $j = 1, \dots, 8$  (mouse number), and  $i$  is the population index (control and knockout). Using the EM algorithm, we obtained estimates for the parameters in our RR model. Figure 8(a) depicts the histogram of the 2000  $d_g$  statistics. The smooth black curve shows the fitted mixture distribution, drawn using the average estimated error variance. The smooth blue and red curves correspond to the average fitted distributions of the null and nonnull groups, respectively. Per-gene fitted distributions are plotted in light colors (note that the nonnull probability is very small, so only gene-specific distributions of the null group, in light blue, can be observed in this case). The mean-effect parameter estimates we obtained are  $\hat{\tau} = 0.007$  and  $\hat{\psi} = 0.682$ ,  $\hat{\sigma}_\psi^2 = 0.874$ .

Figure 8(c) depicts the histogram of the  $m_g$  statistics and the fitted distribution. The estimates for the shape and scale parameters of the error variance distribution are 1.87 and 11.11, respectively. The empirical mean and variance of  $\{\tilde{\sigma}_{\epsilon,g}^2\}$  are 0.078 and 0.004.

We used the parameter estimates to establish the RR test statistics, and the p-values for the hypotheses that genes are in the null group. Figure 8(b) depicts the Benjamini-Hochberg adjusted p-values. The red, solid points represent the genes that were declared nonnull, using a (liberal) FDR threshold of 0.2. Using the FDR criteria, we detected 25 nonnull genes.

Using the RR statistics and Efron's 0.2 threshold for local fdr, we detected 9 nonnull genes, including the ApoA1 gene and others that are closely related to it. The top eight genes had local fdr values of nearly zero, while the ninth had a much higher value of 0.08. Figure 8(d) depicts the RR local

fdr statistics, and the red, solid points represent the genes that were declared nonnull using a local fdr threshold of 0.2. The top eight genes (using either the FDR or the local fdr criteria) are also identified (among others) when using the *LIMMA* and *locfdr* R packages, and were confirmed to be differentially expressed in the knockout versus the control line by an independent assay.

Interestingly, assuming no other genes are in the nonnull group, the true value of  $p$  is 0.00144, and our EM estimate is 0.0039, while Efron's estimates using the MLE and CME methods are -0.036, -0.083, respectively. As we mentioned earlier, the *LIMMA* R package does not provide an estimate for  $p_1$ , and uses a default value of 0.01.

## 6.2 Colon Cancer Data

The data analyzed by Alon et al. (1999) consists of 2000 ESTs in 40 tumor and 22 normal colon tissue samples. Of the 40 patients involved in the study, 22 supplied both tumor and normal tissue samples. In their analysis, Alon et al. (1999) used an Affymetrix oligonucleotide array complementary to more than 6,500 human genes and expressed sequence tags (ESTs), and a two-way clustering method to identify families of genes and tissues based on expression patterns in the data set. Do et al. (2005) used a Bayesian mixture model to analyze the same data set and estimated the probability of differential expression. Using empirical Bayes methods they obtained a point estimate  $\hat{p}_0 = 0.39$  and contrasted it with the posterior marginal probability distribution of  $p_0$  from the non-parametric Bayesian model, which they fit using MCMC simulations. The empirical Bayes estimate for  $p_0$  was far out in the right tail of the posterior distribution, which they argued, might lead to underestimating the posterior probability of being in the nonnull group

(differentially expressed genes). They propose using posterior expected FDR (Genovese and Wasserman, 2002) thresholds to calibrate between a desired false discovery rate and the number of significant genes. For example, with FDR=0.2, they find 1,938 nonnull genes.

Using our estimation method, we obtain  $\hat{p}_1 = 0.36$  (or  $\hat{p}_0 = 0.64$ ). In this data set, the empirical mean and variance of  $\tilde{\sigma}_{\epsilon,g}^2$  are 0.47 and 0.021, respectively, with estimates  $\hat{\alpha} = 10.42$  and  $\hat{\beta} = 0.22$ .

According to the RR model, the (nonnull) mean effect of the interaction term is estimated by  $\hat{\psi} = -0.03$  and the variance by  $\hat{\sigma}_{\psi}^2 = 0.12$ . This suggests that there may be two nonnull groups (over- and under-expressed groups of genes). Figure 9(a) shows the histogram of the 2000  $d_g$  statistics. The light blue and purple curves represent the (per gene) fitted distributions for the null and nonnull groups, respectively. The smooth black curve shows the fitted mixture distribution, drawn using the average estimated error variance.

Figure 9(c) shows the histogram of the  $m_g$  statistics and the fitted distributions. Figure 9(b) shows the BH-adjusted p-values of all the genes based on the estimated null distributions. Using an FDR level of 0.2 we detect 107 nonnull genes. Figure 9(d) depicts the RR local fdr test statistics, and with null posterior probability threshold of 0.2 we detect 170 nonnull genes.

## 7 Discussion

In the previous sections, we demonstrated that our approach can lead to six different models depending on the assumptions imposed on the gene-specific effects. Interestingly, the test statistics associated with these models have been considered independently in the literature in various forms, but to our knowledge, this is the first time they have been categorized as special cases of the same model. The RR model, in which both the nonnull gene-specific

interaction effects and gene-specific variances are modeled as random variates, leads to James Stein-type (shrinkage) estimation of the parameters. Specifically, the RR statistics enjoy shrinkage in both the numerator and denominator of a posterior t-statistic, resulting in powerful test statistics while maintaining few false positives in our simulation studies. Using a Laplace approximation to make the EM algorithm tractable, our approach yields stable parameter estimates, even for the notoriously difficult parameter  $p_1$ .

Since our approach is model-based, it can be easily generalized to other situations. For example, the methods described in this paper can be extended to deal with multiple treatments, multiple nonnull groups, and multivariate responses. The extension to multiple treatments is described in the appendix. The other two extensions are currently being investigated.

**Acknowledgement:**

We would like to thank Yoav Benjamini, Daniel Yekutieli, Peng Liu, and Gene Hwang for their useful comments.

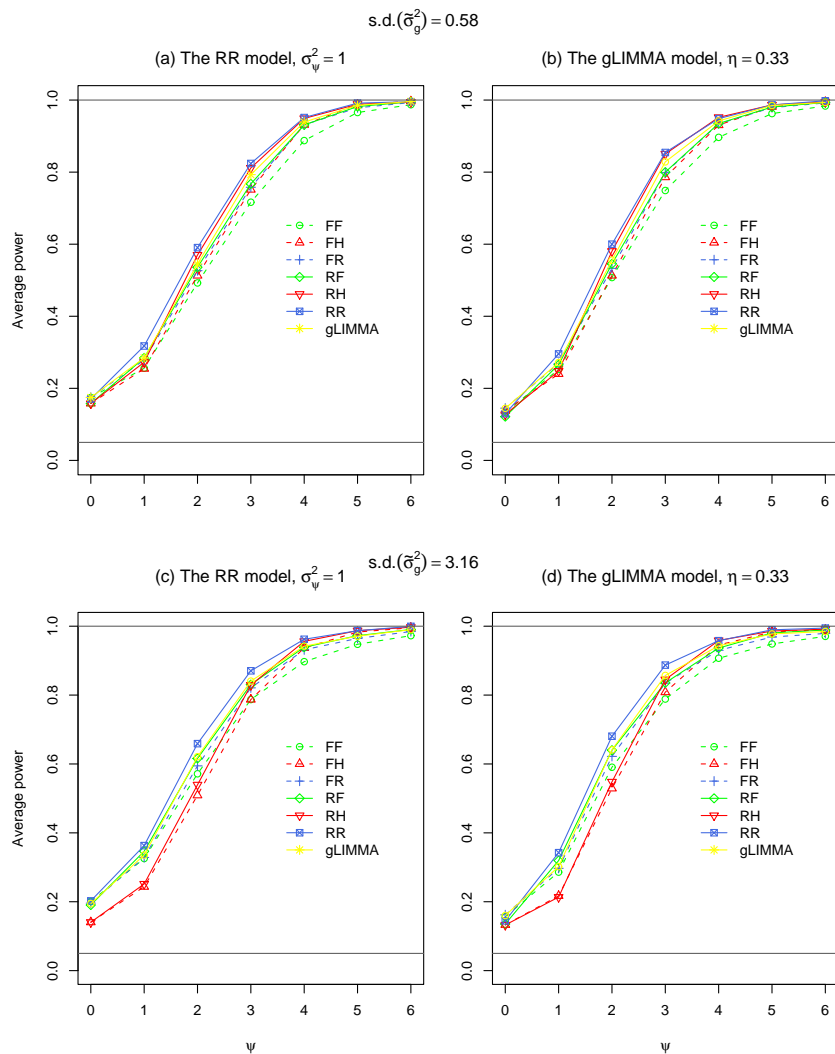


Figure 1: Average Power for Empirical Quantile Analysis under RR ((a) and (c)) and *gLIMMA* ((b) and (d)) data generation models, with  $n_1 = n_2 = 6$  samples,  $G = 2000$  genes, and  $p_1 = 0.05$  probability of nonnull status. Plots (a) and (b) show performance with low error variance variability ( $CV=0.58$ ), whereas plots (c) and (d) show performance with high error variance variability ( $CV=3.16$ ).

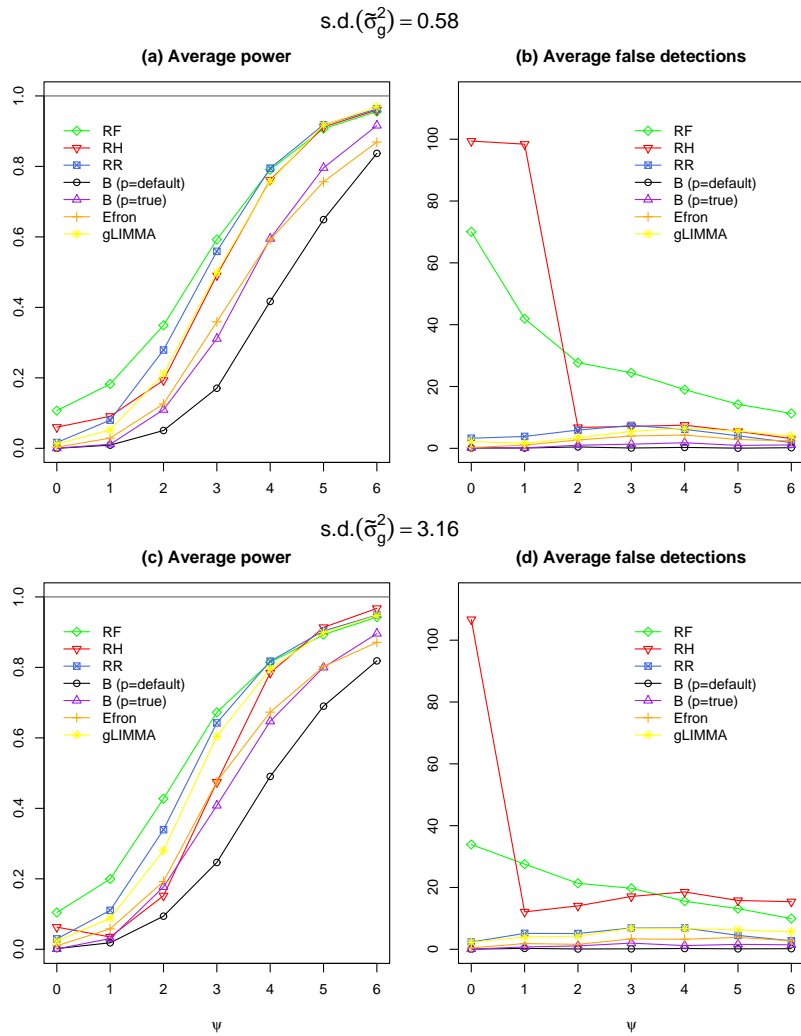


Figure 2: Average Power ((a) and (c)) and Number of False Detections ((b) and (d)) for RR model using local fdr cutoff= 0.20, with  $n_1 = n_2 = 6$  samples,  $G = 2000$  genes, and  $p_1 = 0.05$  probability of nonnull status. Plots (a) and (b) show performance with low error variance variability ( $CV=0.58$ ), whereas plots (c) and (d) show performance with high error variance variability ( $CV=3.16$ ).

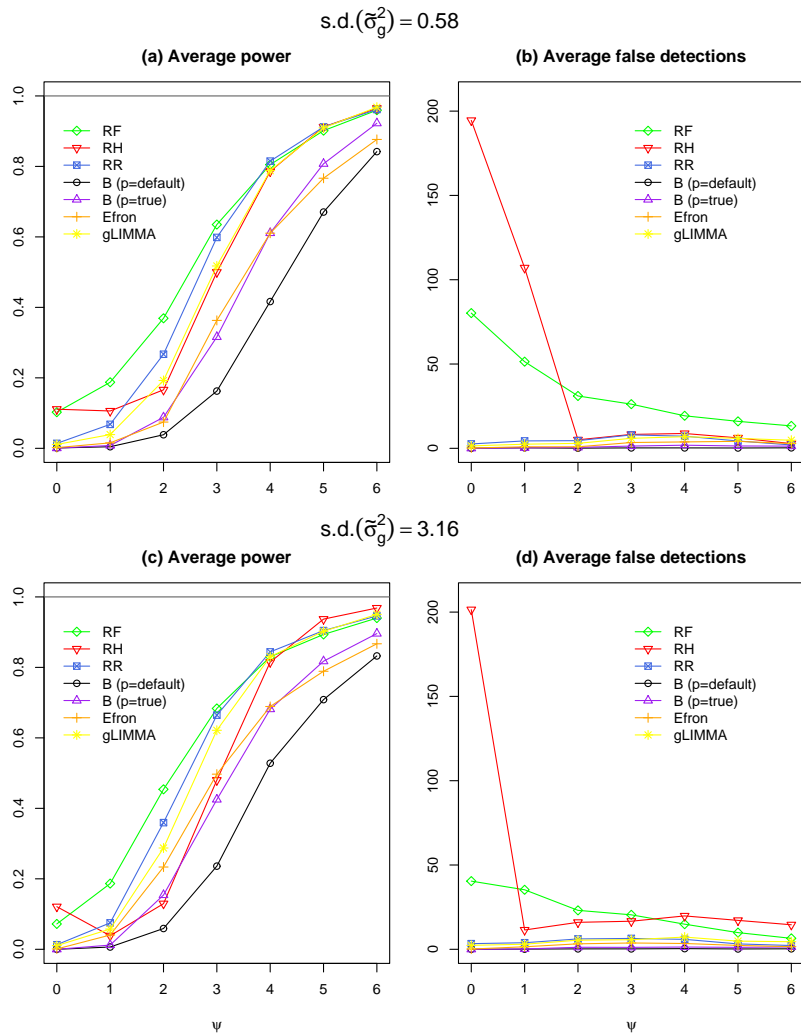


Figure 3: Average Power ((a) and (c)) and Number of False Detections ((b) and (d)) for *gLIMMA* model using local fdr cutoff= 0.20, with  $n_1 = n_2 = 6$  samples,  $G = 2000$  genes, and  $p_1 = 0.05$  probability of nonnull status. Plots (a) and (b) show performance with low error variance variability ( $CV=0.58$ ), whereas plots (c) and (d) show performance with high error variance variability ( $CV=3.16$ ).

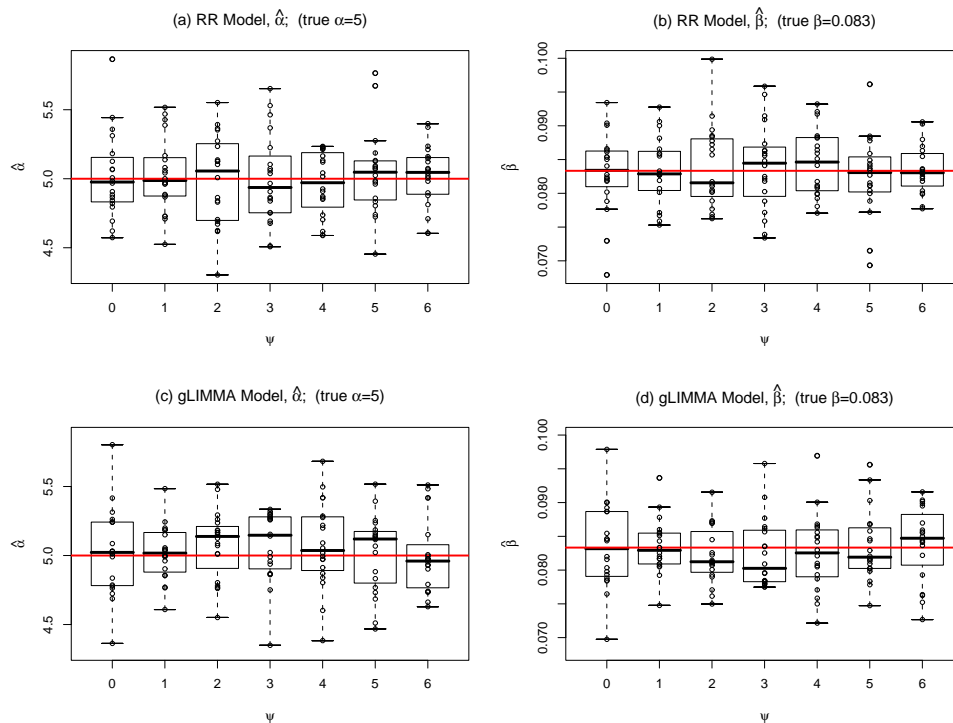


Figure 4: Estimates of shape  $\alpha$  and scale  $\beta$  for  $\sigma_{\epsilon,g}^2$  for the RR data generation scheme ((a) and (b)) and the *gLIMMA* data generation scheme ((c) and (d)).

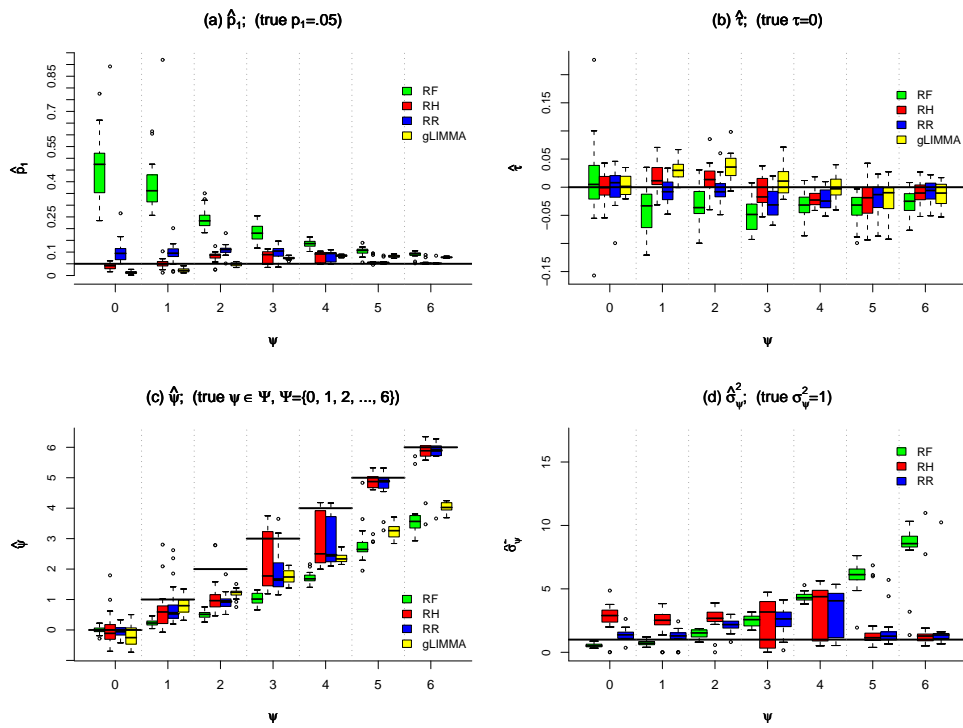


Figure 5: Estimates of  $\phi$  using EM with simulation parameters  $n_1 = n_2 = 6$  and  $G = 2000$  using low error variance variability ( $\alpha = 5$  and  $\beta = 1/12$ ) under the RR data generation scheme.

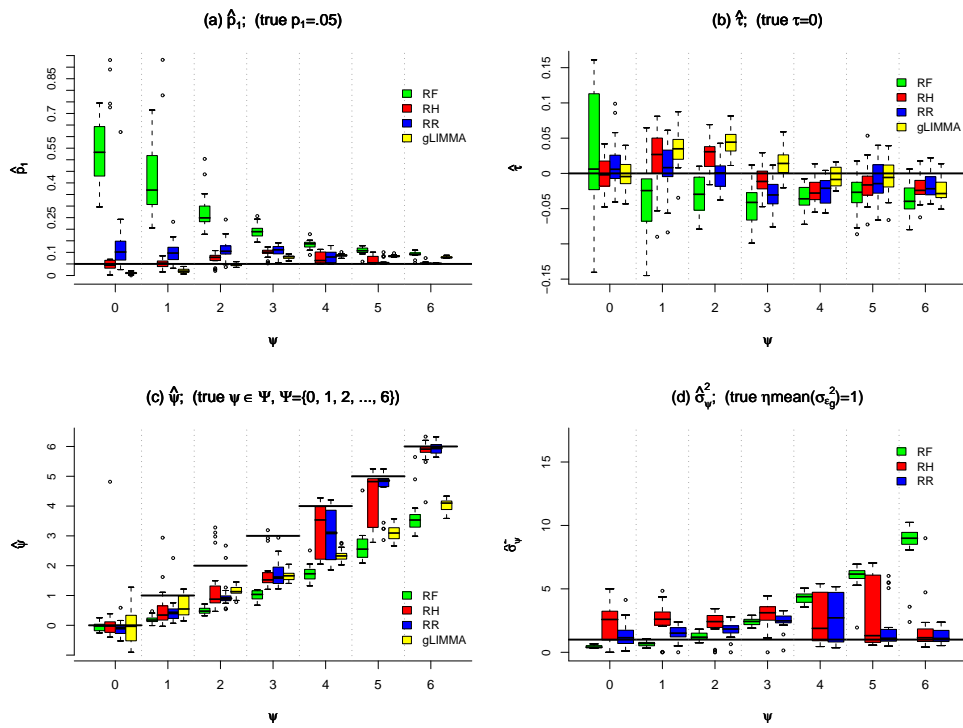


Figure 6: Estimates of  $\phi$  using EM with simulation parameters  $n_1 = n_2 = 6$  and  $G = 2000$  using low error variance variability ( $\alpha = 5$  and  $\beta = 1/12$ ) under the *gLIMMA* data generation scheme.

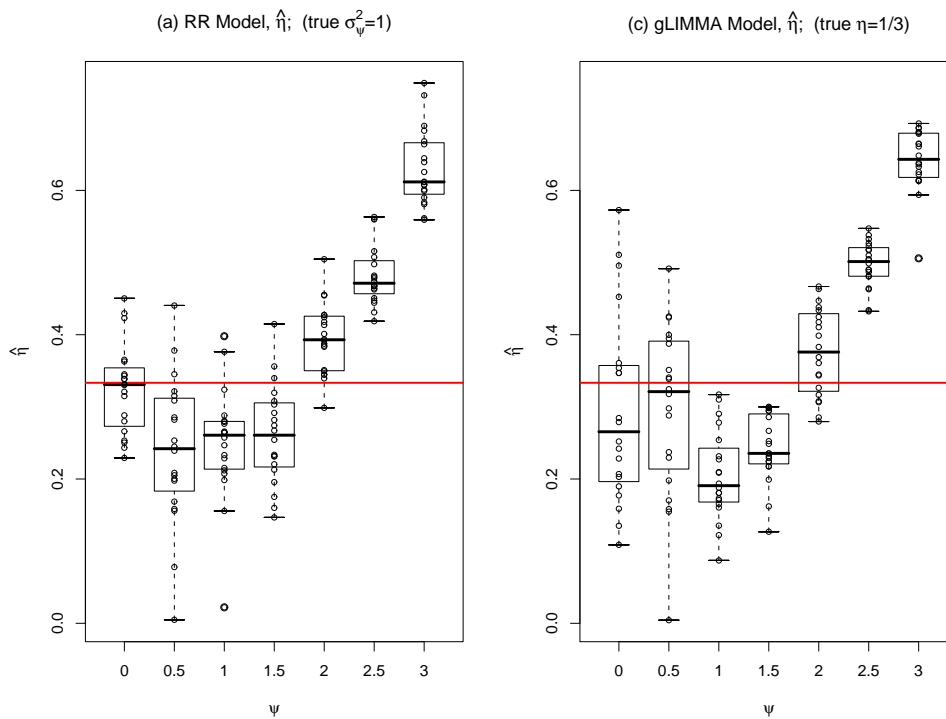


Figure 7: Estimates of  $\eta$  using EM with simulation parameters  $n_1 = n_2 = 6$  and  $G = 2000$  using low error variance variability ( $\alpha = 5$  and  $\beta = 1/12$ ) under the RR (a) and *gLIMMA* (b) data generation schemes.

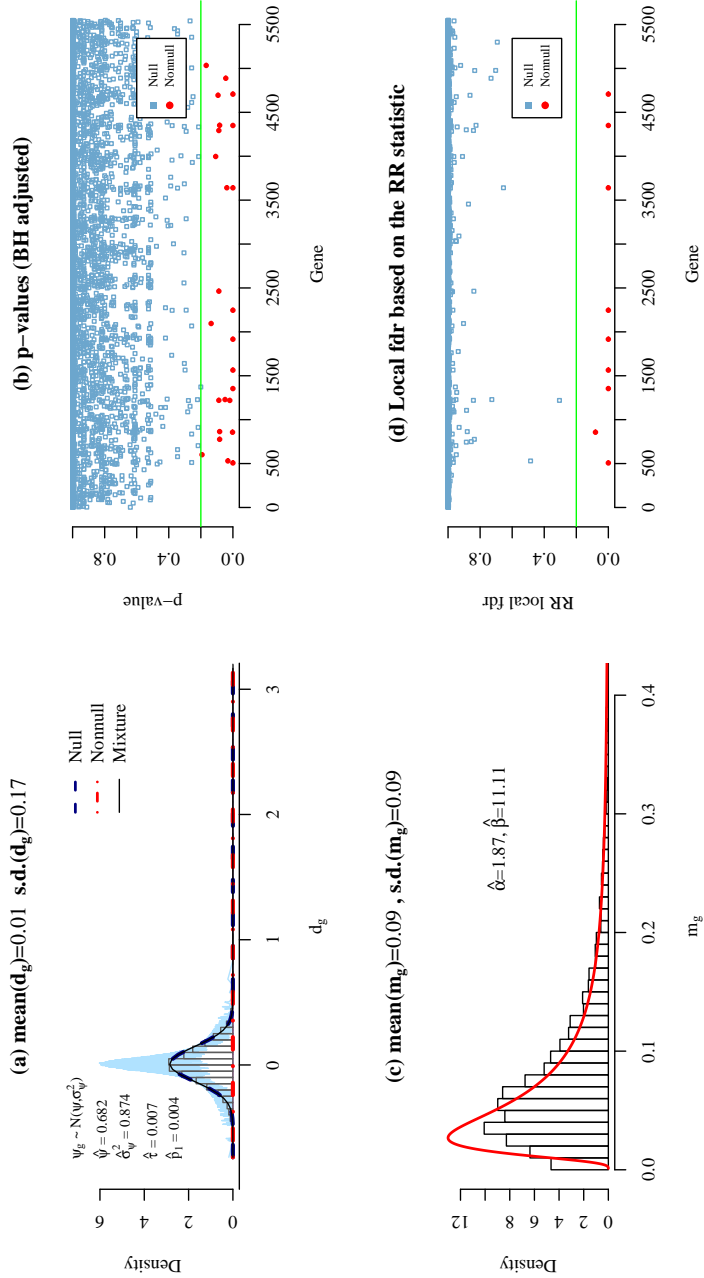


Figure 8: (a) The histogram of the 5548  $d_g$  statistics from the ApoA1 data set and the fitted distributions. The light blue curve represent the (per gene) fitted distributions for the null group. The smooth black curve shows the fitted mixture distribution, drawn using the average estimated error variance across all genes. (b) The Benjamini-Hochberg adjusted p-values for all genes. Using an FDR level of 0.2, we detect 25 nonnull genes. (c) The histogram of the  $m_g$  test statistics and the fitted distribution. The empirical mean and s.d. of  $\{\hat{\sigma}_{\epsilon,g}^2\}$  are 0.078 and 0.062. (d) The RR test statistics of all the genes. Using a 0.2 threshold for the posterior probability, we declare 9 genes to be nonnull.

Colon cancer data (Alon, 1999)

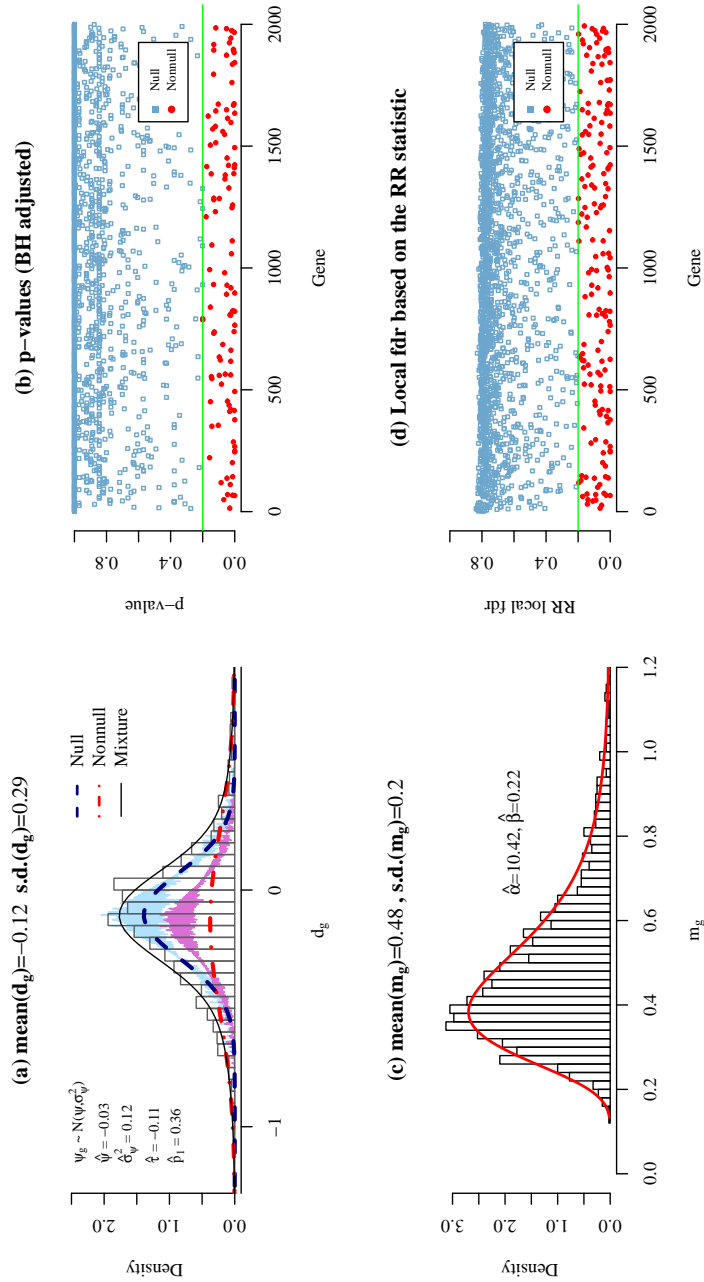


Figure 9: (a) The histogram of the 2000  $d_g$  statistics from the colon cancer data set (Alon et al., 1999) and the fitted distributions. The light blue and purple curves represent the (per gene) fitted distributions for the null and nonnull groups, respectively. The smooth black curve shows the fitted mixture distribution, drawn using the average estimated error variance across all genes. (b) The Benjamini-Hochberg adjusted p-values for all genes. Using an FDR level of 0.2, we detect 107 nonnull genes. (c) The histogram of the  $m_g$  test statistics and the fitted distribution. The empirical mean and s.d. of  $\{\hat{\sigma}_{\epsilon,g}^2\}$  are 0.47 and 0.15. (d) The RR test statistics of all the genes. Using a 0.2 threshold for the posterior probability, we declare 170 genes to be nonnull.

## Appendix

In this section we provide details on some of our previous derivations, and elaborate on the case of multiple treatments.

### Empirical Bayes Estimates for $\alpha$ and $\beta$

To obtain an estimate of the error variance in the random error case, recall that

$$m_g | \sigma_{\epsilon,g}^2 \sim \frac{\sigma_{\epsilon,g}^2}{f_g} \chi_{f_g}^2 \equiv \text{Gamma} \left( \frac{f_g}{2}, \frac{2\sigma_{\epsilon,g}^2}{f_g} \right). \quad (23)$$

We maximize the marginal density of  $m_g$  numerically to obtain maximum likelihood estimates of  $\alpha$  and  $\beta$ . Given the conditional distribution in (23), we find the marginal density of  $m_g$  by integrating out  $\sigma_{\epsilon,g}^2$ . Specifically,

$$\begin{aligned} f(m_g) &= \int_0^\infty f(m_g | \sigma_{\epsilon,g}^2) f(\sigma_{\epsilon,g}^{-2}) d\sigma_{\epsilon,g}^{-2} \\ &= \int_0^\infty \left[ \frac{m_g^{\frac{f_g}{2}-1} \exp(-\frac{m_g f_g}{2\sigma_{\epsilon,g}^2})}{\Gamma\left(\frac{f_g}{2}\right) \left(\frac{2\sigma_{\epsilon,g}^2}{f_g}\right)^{\frac{f_g}{2}}} \right] \left[ \frac{\exp(\sigma_{\epsilon,g}^{-2} \beta^{-1})}{\Gamma(\alpha) \beta^\alpha} (\sigma_{\epsilon,g}^{-2})^{\alpha-1} \right] d\sigma_{\epsilon,g}^{-2} \\ &= \frac{m_g^{\frac{f_g}{2}-1} \left(\frac{f_g}{2}\right)^{\frac{f_g}{2}}}{\Gamma\left(\frac{f_g}{2}\right) \Gamma(\alpha) \beta^\alpha} \int_0^\infty (\sigma_{\epsilon,g}^{-2})^{\frac{f_g}{2}+\alpha-1} \exp\left[-\sigma_{\epsilon,g}^{-2} \left(\frac{m_g f_g}{2} + \frac{1}{\beta}\right)\right] d\sigma_{\epsilon,g}^{-2} \\ &= \frac{m_g^{\frac{f_g}{2}-1} \left(\frac{f_g}{2}\right)^{\frac{f_g}{2}}}{\Gamma\left(\frac{f_g}{2}\right) \Gamma(\alpha) \beta^\alpha} \cdot \frac{\Gamma\left(\frac{f_g}{2} + \alpha\right)}{\left(\frac{m_g f_g}{2} + \frac{1}{\beta}\right)^{\frac{f_g}{2}+\alpha}}. \end{aligned} \quad (24)$$

The final equality in (24) results from noting that the integral in the third equality is proportional to a  $\text{Gamma}(f_g/2 + \alpha, [\beta^{-1} + m_g f_g/2]^{-1})$  density. We maximize  $\sum_g \log(f(m_g))$  with respect to  $\alpha$  and  $\beta$  to obtain the empirical Bayes estimates  $\hat{\alpha}$  and  $\hat{\beta}$ .

The joint distribution of  $m_g$  and  $\sigma_{\epsilon,g}$  is given by

$$f(m_g, \sigma_{\epsilon,g}^{-2}) = \frac{m_g^{f_g/2-1} f_g^{f_g/2} (\sigma_{\epsilon,g}^{-2})^{\alpha-1+f_g/2} \exp\left\{-\sigma_{\epsilon,g}^{-2} \left[\frac{m_g}{2} f_g + \frac{1}{\beta}\right]\right\}}{\Gamma\left(\frac{f_g}{2}\right) 2^{f_g/2} \Gamma(\alpha) \beta^\alpha}.$$

So, conditional on  $m_g$ ,

$$\sigma_{\epsilon,g}^{-2} \sim \text{Gamma}(\alpha + f_g/2, (m_g f_g/2 + 1/\beta)^{-1}).$$

Hence, the conditional expectation is

$$\begin{aligned} E(\sigma_{\epsilon,g}^2 | m_g) &= \frac{f_g/2}{f_g/2 + \alpha - 1} m_g + \frac{\alpha + 1}{f_g/2 + \alpha - 1} \cdot \frac{1}{(\alpha + 1)\beta} \\ &\approx \frac{f_g/2}{f_g/2 + \alpha - 1} m_g + \frac{\alpha + 1}{f_g/2 + \alpha - 1} \bar{m}, \end{aligned}$$

and the conditional mode is

$$\begin{aligned} \text{Mode}(\sigma_{\epsilon,g}^2 | m_g) &= \frac{f_g/2}{f_g/2 + \alpha + 1} m_g + \frac{\alpha + 1}{f_g/2 + \alpha + 1} \cdot \frac{1}{(\alpha + 1)\beta} \\ &\approx \frac{f_g/2}{f_g/2 + \alpha + 1} m_g + \frac{\alpha + 1}{f_g/2 + \alpha + 1} \bar{m}. \end{aligned}$$

Note that using the approximation of the mode,  $\bar{m} \approx [(\alpha + 1)\beta]^{-1}$ , in both the posterior mean and posterior mode yields a shrinkage-estimator form. Equivalently, we could replace  $(\alpha + 1)$  with  $(\alpha - 1)$  in the conditional expectation and the conditional mode, and obtain shrinkage toward the sample mean of  $\{m_g\}$ .

## Maximum Likelihood Estimation of $\phi$

Recall that in the RR model we use the Laplace approximation (8), hence the (approximate) complete likelihood is

$$\begin{aligned}
\tilde{L}_C(\phi) &= \prod_{g=1}^G L(b_g, d_g; \tilde{\sigma}_g^2) \\
&= \prod_{g=1}^G \int L(b_g; p_1) L(d_g | b_g; \psi_g, \tilde{\sigma}_g^2) f(\psi_g) d\psi_g \\
&= \prod_{g=1}^G \int \left[ p_1^{b_g} (1-p_1)^{1-b_g} (2\pi\tilde{\sigma}_g^2)^{-\frac{1}{2}} (2\pi\sigma_\psi^2)^{-\frac{b_g}{2}} \exp\left\{-\frac{1-b_g}{2\tilde{\sigma}_g^2} (d_g - \tau)^2\right\} \right. \\
&\quad \left. \times \int \exp\left\{-\frac{b_g}{2\tilde{\sigma}_g^2} (d_g - \tau - \psi_g)^2 - \frac{b_g}{2\sigma_\psi^2} (\psi_g - \psi)^2\right\} \right] d\psi_g.
\end{aligned} \tag{25}$$

with log-likelihood

$$\begin{aligned}
\ell(\phi) &= \sum_{g=1}^G [(1-b_g) \log(1-p_1) + b_g \log(p_1)] \\
&\quad - \sum_{g=1}^G \left[ \frac{b_g}{2} \log(2\pi\sigma_\psi^2) + \frac{1}{2} \log(2\pi\tilde{\sigma}_g^2) \right] \\
&\quad - \frac{1}{2} \sum_{g=1}^G b_g \log\left( (2\pi)^{-1} \left( \frac{1}{\tilde{\sigma}_g^2} + \frac{1}{\sigma_\psi^2} \right) \right) - \sum_{g=1}^G (1-b_g) \frac{(d_g - \tau)^2}{2\tilde{\sigma}_g^2} \\
&\quad - \sum_{g=1}^G \frac{b_g}{2} \left[ \frac{1}{\sigma_\psi^2 + \tilde{\sigma}_g^2} (d_g - \tau - \psi)^2 \right].
\end{aligned} \tag{26}$$

The estimates (11-14) are obtained by maximizing the log-likelihood with respect to the parameters,  $\phi$ .

Although the Laplace approximation is not necessary in the RF and RH models, note that the complete likelihoods and log-likelihoods for these models are identical to equations (25) and (26), with  $\tilde{\sigma}_g^2$  replaced by  $\hat{\sigma}_g^2$  and  $\hat{\sigma}_\epsilon^2$  (as defined in Section 3.2), respectively.

## Multiple Treatments

In the general case we assume  $t \geq 2$  treatments  $i = 1, 2, \dots, t$  assigned to  $t$  groups of  $n_{1,g}, n_{2,g}, \dots, n_{t,g}$  subjects indexed by  $j_1 = 1, \dots, n_{1,g}, \dots, j_t = 1, \dots, n_{t,g}$ , and we use the model defined by (1) and (2). Here, we impose a standard (fixed effect) constraint

$$\sum_{i=1}^t \psi_{ig} = 0.$$

The distributions for the gene-specific effects in the multiple-treatment case are assumed to follow a normal distribution,

$$\boldsymbol{\psi}_g \sim \text{iid } N_t(\boldsymbol{\psi}, \sigma_\psi^2 (\mathbf{I}_t - \bar{\mathbf{J}}_t)),$$

where  $\mathbf{I}_t - \bar{\mathbf{J}}_t$  is the  $t \times t$  centering matrix,  $\boldsymbol{\psi}$  is a  $t$ -dimensional vector, and  $\sigma_\psi^2$  is a scalar. The test statistic  $m_g$  is defined as

$$m_g = \sum_{i=1}^t \sum_{j=1}^{n_{ig}} (y_{ijg} - \bar{y}_{i \cdot g})^2 / f_g,$$

where  $f_g = n_{1g} + \dots + n_{tg} - t$ , and we use  $m_g$  as before, to estimate  $\alpha$  and  $\beta$ .

To estimate the rest of the parameters in the RR model, we use the  $(t - 1)$ -dimensional vector) test statistics

$$\mathbf{d}_g = \mathbf{H} \bar{\mathbf{Y}}_g,$$

where

$$\mathbf{H} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 & \dots & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1/\sqrt{t(t-1)} & 1/\sqrt{t(t-1)} & 1/\sqrt{t(t-1)} & \dots & -(t-1)/\sqrt{t(t-1)} \end{bmatrix},$$

$$\bar{\mathbf{Y}}_g = (\bar{Y}_{1g}, \bar{Y}_{2g}, \dots, \bar{Y}_{tg})'.$$

Derivations similar to the ones we used to obtain the estimates in Section 3 lead the same estimate for  $p_1$  and to the following estimates, analogous to (12) and (13):

$$\begin{aligned} (\mathbf{H}\boldsymbol{\tau}^{(m+1)})' &= \left[ \sum_{g=1}^G p_{0,g}^{(m)} \mathbf{d}'_g \boldsymbol{\Lambda}_0^{-1} \right] \left[ \sum_{g=1}^G p_{0,g}^{(m)} \boldsymbol{\Lambda}_0^{-1} \right]^{-1}, \\ (\mathbf{H}\boldsymbol{\psi}^{(m+1)})' &= \left[ \sum_{g=1}^G p_{1,g}^{(m)} (\mathbf{d}_g - \mathbf{H}\boldsymbol{\tau}^{(m+1)})' (\boldsymbol{\Lambda}_0 + \boldsymbol{\Lambda}_A)^{-1} \right] \left[ \sum_{g=1}^G p_{1,g}^{(m)} (\boldsymbol{\Lambda}_0 + \boldsymbol{\Lambda}_A)^{-1} \right]^{-1}, \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\Lambda}_A &= \sigma_\psi^{2(m)} \mathbf{I}_{t-1}, \\ \boldsymbol{\Lambda}_0 &= \tilde{\sigma}_{\epsilon,g}^2 \mathbf{H} [\text{diag}_i (1/n_{ig})] \mathbf{H}'. \end{aligned}$$

The update for  $\tilde{\sigma}_\psi^2$  is the solution to the equation,

$$\sum_{g=1}^G p_{1,g}^{(m)} \cdot \text{tr} \left( (\boldsymbol{\Lambda}_0 + \boldsymbol{\Lambda}_A)^{-1} \right) = \sum_{g=1}^G p_{1,g}^{(m)} \left( \mathbf{d}_g - \mathbf{H}\boldsymbol{\xi}^{(m+1)} \right)' (\boldsymbol{\Lambda}_0 + \boldsymbol{\Lambda}_A)^{-2} \left( \mathbf{d}_g - \mathbf{H}\boldsymbol{\xi}^{(m+1)} \right).$$

where  $\boldsymbol{\xi}^{(m+1)} = \boldsymbol{\tau}^{(m+1)} + \boldsymbol{\psi}^{(m+1)}$ .

The likelihood ratio test statistic has a similar form as (18),

$$\frac{L_{0,g}}{L_{1,g}} = |\mathbf{I} - \boldsymbol{\Lambda}_g|^{-1/2} \exp \left\{ -\frac{1}{2} \boldsymbol{\Gamma}' (\boldsymbol{\Lambda}_0^{-1} \boldsymbol{\Lambda}_g^{-1}) \boldsymbol{\Gamma} \right\} \exp \left\{ -\frac{1}{2} \tilde{\sigma}_\psi^{-2} (\mathbf{H}\hat{\boldsymbol{\psi}})' (\mathbf{H}\hat{\boldsymbol{\psi}}) \right\},$$

where

$$\begin{aligned} \boldsymbol{\Gamma} &= \left[ \boldsymbol{\Lambda}_g (\mathbf{d}_g - \mathbf{H}\hat{\boldsymbol{\tau}}) + (\mathbf{I} - \boldsymbol{\Lambda}_g) (\mathbf{H}\hat{\boldsymbol{\psi}}) \right], \\ \boldsymbol{\Lambda}_g &= (\boldsymbol{\Lambda}_A + \boldsymbol{\Lambda}_0)^{-1} \boldsymbol{\Lambda}_A, \\ \mathbf{I} - \boldsymbol{\Lambda}_g &= (\boldsymbol{\Lambda}_A + \boldsymbol{\Lambda}_0)^{-1} \boldsymbol{\Lambda}_0. \end{aligned}$$

## References

- ALLISON, D. B., CUI, X., PAGE, G. and SABRIPOUR, M. (2006). Microarray data analysis: from disarray to consolidation and consensus. *Nature Genetics*, **7** 55–65.
- ALON, BARKAI, NOTTERMAN, GISH, YBARRA, MACK and LEVINE (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences*, **96** 6745–6750.
- BALDI, P. and LONG, A. D. (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 HOCHBERG, Y. (1995). Controlling the false discovery rate—a practical and powerful approach to multiple testing. *Journal Of The Royal Statistical Society Series B*, **57** 499–517.
- CALLOW, M. J., DUDOIT, S., GONG, E. L., SPEED, T. P. and RUBIN, E. M. (2000). Microarray expression profiling identifies genes with altered expression in HDL-deficient mice. *Genome Res*, **10** 2022–9.
- CUI, X. and CHURCHILL, G. (2003). Statistical tests for differential expression in cDNA microarray experiments. *Genome Biology*, **4**.
- CUI, X., HWANG, J. T. G., QUI, J., BLADES, N. J. and CHURCHILL, G. A. (2005). Improved statistical tests for differential gene expression by shrinking variance components. *Biostatistics*, **6** 59–75.
- DO, K.-A., MULLER, P. and TANG, F. (2005). A bayesian mixture model for differential gene expression. *Appl. Stat*, **54** 627–644.

- EFRON, B. (2005). Local false discovery rates. URL <http://www-stat.stanford.edu/brad/papers/False.pdf>.
- EFRON, B. (2008). Microarrays, empirical bayes and the two groups model. *Statistical Science*, **23** 1–22.
- EFRON, B., TIBSHIRANI, R., STOREY, J. and TUSHER, V. (2001). Empirical bayes analysis of microarray experiment. *JASA*, **96** 1151–1160.
- EFRON, B., TURNBULL, B. B. and NARASIMHAN, B. (2008). *locfdr: Computes local false discovery rates*. R package version 1.1-6.
- FIGUEROA, M. E., REIMERS, M., THOMPSON, R. F., YE, K., LI, Y., SELZER, R. R., FRIDRIKSSON, J., PAIETTA, E., WIERNIK, P., GREEN, R. D., GREALLY, J. M. and MELNICK, A. (2008). An integrative genomic and epigenomic approach for the study of transcriptional regulation. *PLoS ONE*, **3**.
- GENOVESE, C. and WASSERMAN, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *Journal Of The Royal Statistical Society Series B*, **64** 499–517.
- HWANG, J. T. G. and LIU, P. (2007). Optimal tests shrinking both means and variances applicable to microarray data. Currently under review.
- KENDZIORSKI, C. M., NEWTON, M. A., LAN, H. and GOULD, M. N. (2003). On parametric empirical bayes methods for comparing multiple groups using replicated gene expression profiles. *Statistics in Medicine*, **22** 3899–3914.
- KERR, M., MARTIN, M. and CHURCHILL, G. (2000). Analysis of variance in microarray data. *Journal of Computational Biology*, **7** 819–837.

- LONNSTEDT, I. and SPEED, T. (2002). Replicated microarray data. *Statistica Sinica*, **12** 31–46.
- NEWTON, M. A., KENDZIORSKI, C. M., RICHMOND, C. S., BLATTNER, F. R. and TSUI, K. W. (2001). On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data. *Computational Biology*, **8** 37–52.
- R DEVELOPMENT CORE TEAM (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- SMYTH, G. K. (2004). Linear models for empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, **3**. Article 2.
- SMYTH, G. K. (2005). *Limma: linear models for microarray data*. Springer, New York, 397–420.
- TAI, Y. C. and SPEED, T. P. (2006). A multivariate empirical bayes statistic for replicated microarray time course data. *Annals of Statistics*, **34** 2387–2412.
- TUSHER, V. G., TIBSHIRANI, R. and CHU, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.*, **98** 5116–5121.
- WRIGHT, G. W. and SIMON, R. M. (2003). A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics*, **19** 2448–2455.