A Multivariate Empirical Bayes Statistic for Replicated Microarray Time Course Data

Yu Chuan Tai and Terence P. Speed

University of California, Berkeley, USA.
Walter and Eliza Hall Institute of Medical Research, Australia.

Abstract: In this paper we derive a one-sample multivariate empirical Bayes statistic (the MB-statistic) to rank genes in the order of differential expression from replicated microarray time course experiments . We do this by testing the null hypothesis that the expectation of a k-vector of a gene's expression levels is a multiple of 1_k, the vector of k 1s. The importance of moderation in this context is explained. Together with the MB-statistic we have the one-sample $\widetilde{T}^2$ statistic, a variant of the one-sample Hotelling T^2. Both the MB-statistic and $\widetilde{T}^2$ statistic can be used to rank genes in the order of evidence of nonconstancy, incorporating any correlation structure among time point samples and the replication. In a simulation study we show that the one-sample MB-statistic, $\widetilde{T}^2$ statistic, and moderated Hotelling T^2 statistic achieve the smallest number of false positives and false negatives, and all perform equally well. Several special and limiting cases of the MB-statistic are derived, and two-sample versions are described.

AMS 2000 subject classifications. Primary 62M10; secondary 62C12, 92D10.
Key words and phrases: microarray time course, multivariate empirical Bayes, moderation, one-sample, gene ranking, replication.
Supported by NIH through grants HG00047 (YCT) and R01 Lmo7609-01 (TPS).

Date: August 04, 2004

1. Introduction

Microarray time course experiments differ from other microarray experiments in that gene expression values at different time points can be correlated. This may happen when the mRNA samples at successive time points are taken from the same organism or cell culture. Such longitudinal experiments make it possible to monitor and study the temporal aspects of biological processes of interest for thousands of genes simultaneously. Two major categories of time course experiments are those involving periodic and developmental phenomena, respectively. Periodic time courses typically concern natural biological processes such as the cell cycle or circadian rhythms, where the time profiles follow regular patterns (Cho et al., 1998; Spellman et al., 1998; Storch et al., 2002). On the other hand, in developmental time course experiments, we measure gene expression levels at a series of times in a developmental process, or after applying a treatment such as a drug to the organism, tissue or cells (Chu et al. 1998; Wen et al. 1998; Tamayo et al. 1999). In this case we typically have few prior expectations concerning the temporal patterns of gene expression. The gene ranking method we develop in this paper is mainly for developmental time courses, although it is straightforward to apply it to periodic time courses.

A typical microarray time course dataset consists of expression measurements of G genes across k time points, under one or more biological conditions (*e.g.* wildtype versus mutant). The number of genes G ($10,000$-$20,000$) is very much larger than the number of time points k, which can be 5-10 for shorter, and 11-20 for longer time courses. Many such experiments are unreplicated due to cost or other limitations, and when replicates are done, the number n is typically quite small, say 2-5.

One of the statistical challenges here is to select nonconstant genes, that is, genes with *any* change in expression level over time. Such genes are of interest to biologists because they are often involved in the biological processes motivating the experiment. This challenge arises from the fact that there are very few time points, and very many genes. The series are usually so short that we cannot consider using standard time series methods such as the Fourier transform or wavelets. The gene ranking/selection problem in replicated microarray time course experiments is relatively new. The most widely used methods have focused on identifying differentially expressed genes for replicated microarray experiments across two or more independent sample groups (e.g. Baldi and Long 2001, Efron et al. 2001, Tusher et al. 2001, Dudoit et al. 2002, Lönnstedt and Speed 2002, Broberg 2003, Ge et al. 2003, Kendziorski et al. 2003, Reiner et al. 2003,

Smyth 2004). These methods may not be entirely appropriate as they do not address the fact that microarray time course samples are often correlated. Only a few methods have been proposed for ranking and selecting genes in the replicated microarray time course context. Hero and Fleury (2002) and Fleury et al. (2002) described a valuable method called Pareto front analysis to rank and select genes using multi-criterion optimization. Their Pareto fronts and the variants seek to select genes with large values for all the criteria (contrasts) of interest. Park et al. (2003) proposed a two-way anova model with permutations to compare temporal profiles from different experimental groups. Their approach is essentially in the same spirit as the standard F-test with adjusted p-values, and ignores the correlation structure among times. Guo et al. (2003) constructed a variant of the robust Wald-statistic, taking into account the possible within-subject correlations, to select genes with temporal changes. Bar-Joseph et al. (2003) used a maximization technique to compare temporal profiles reconstructed using B-splines from two nonhomogeneous time courses. Yuan et al. (2003) suggested hidden Markov models, incorporating the dependency among times, to select differentially expressed genes from multiple biological conditions. Luan and Li (2004) proposed a B-spline model-based method to identify differentially expressed genes from periodic time courses. Microarray time course experiments can involve correlations among the observations for the same gene. The ordinary F-statistic comparing times, as we would when comparing a number of groups, ignores such correlations, and is likely to be inefficient in the presence of temporal correlations. Even if it was effective, the question of estimating 10,000 different variances needs to be considered.

In this paper we develop a multivariate empirical Bayes LOD score we call the MB-statistic, and a $\widetilde{T}^2$ statistic to rank genes in order of evidence of nonconstancy in expression level over time, taking into account any correlations among observations at different times, and the replication. In essence, we want to measure change across time in relation to replicate variation within and correlation between times, keeping the analysis as gene-specific as we can.

Suppose that for each gene g, $g = 1, ..., G$, we have n_g independent time series, and that we model these as $i.i.d.$ random vectors from a multivariate normal distribution, with gene-specific means $\boldsymbol{\mu}_g$ and gene-specific covariance matrices $\boldsymbol{\Sigma}_g$. We use the natural conjugate priors for $\boldsymbol{\mu}_g$ and $\boldsymbol{\Sigma}_g$. i.e., an inverse Wishart prior for $\boldsymbol{\Sigma}_g$ and a dependent multivariate normal prior for $\boldsymbol{\mu}_g$. The details in this paper differ in two ways from the standard conjugate priors. First, we also have an indicator I_g such that

$I_g = 1$ for nonconstant genes and $I_g = 0$ for constant genes, and the priors for $\boldsymbol{\mu}_g$ differ in these two cases. Second, in order to get a simple closed form expression for the MB-statistic, we assume that the gene-specific covariance matrix $\boldsymbol{\Sigma}_g$ commutes with the $k \times k$ projection matrix $\mathbf{P} = k^{-1}\mathbf{1}_k\mathbf{1}_k'$, i.e., for all g, $\mathbf{P}\boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}_g\mathbf{P}$. Thus a $k \times k$ inverse Wishart prior for $\boldsymbol{\Sigma}_g$ is replaced by a $(k-1) \times (k-1)$ inverse Wishart prior for a part of $\boldsymbol{\Sigma}_g$ and an inverse gamma prior for the remainder. These two part priors are independent, see section 5.2 for details.

The multivariate empirical Bayes model proposed in this paper is motivated by the analogous univariate hierarchical model proposed in Lönnstedt and Speed (2002) for identifying differentially expressed genes in two-color comparative microarray experiments, and the more recent extensions by Smyth (2004). It is shown there that the univariate log posterior odds is equivalent to the square of the univariate moderated t-statistic $\tilde{t}_g^2$, when all the genes have the same degrees of freedom. Furthermore, Smyth (2004) derives improved hyperparameter estimates using the marginal sampling distributions of univariate moderated t-statistic $\tilde{t}_g$ and the sample variance s_g^2. Both the univariate log posterior odds (B-statistic) in Lönnstedt and Speed (2002) and Smyth (2004), and the univariate moderated t-statistic $\tilde{t}_g$ in Smyth (2004), consider just one parameter or contrast at a time in the null hypotheses. They are not for null hypotheses with two or more parameters or contrasts of interest simultaneously. Smyth (2004) introduces a partly-moderated F-statistic extending the univariate moderated t-statistic, which is the ordinary F-statistic from the linear model, with a moderated variance in the denominator. This partly-moderated F-statistic is useful for the simultaneous comparison of multiple uncorrelated coefficients or contrasts. However, as mentioned above, this is not the case in the microarray time-course experiment context. Both the MB-statistic and the $\widetilde{T}^2$ statistic derived in this paper allow multiple parameters simultaneously, while retaining the correlation structure among these parameter estimates, together with the moderation property, and can be applied to both single-channel and two-color microarray experiments.

This paper is organized as follows. After briefly explaining the rationale of moderation in the microarray time course context in section 2, we formally state the null and alternative hypotheses of the one-sample gene ranking problem. Section 4 shows the moderated versions of the standard likelihood-ratio statistic and the one-sample Hotelling T^2 statistic. We formally build up our multivariate empirical Bayes model and derive the MB statistic and the $\widetilde{T}^2$ statistic in section 5. Section 6 describes

how to estimate the hyperparameters of our multivariate empirical Bayes model, using the information across genes. In sections 7 and 8 we derive the expressions for the posterior odds under special and limiting cases of interest. In particular, we show in section 7 that the posterior odds under the constraint that time point samples are independent is equivalent to the product of univariate odds. Section 9 reports results from a simulation study we perform to compare the MB-statistic and the $\widetilde{T}^2$ statistic with other statistics. We discuss the extensions for future work of our multivariate empirical Bayes model in section 10.

2. Why moderation?

The idea of moderation has entered into the analysis of microarray data implicitly or explicitly in several forms (Tusher et al. 2001, Efron et al. 2001, Lönnstedt and Speed 2002, Broberg 2003, Smyth 2004). Smyth (2004) explicitly introduced the moderated t-statistic $\tilde{t}$ in the univariate general linear model setting, using parametric empirical Bayes. Here the rationale is briefly explained in the microarray time course context. Typically, genes with large overall amounts of change across time and with small replicate variances are the best candidates to follow up. However, given thousands of genes in a microarray time course experiment, the replicate variance-covariance matrices are very poorly estimated. Genes with small amounts of change over time and small replicate variances can have large between to within time F-statistics because these statistics are inflated by small denominators. We consider such genes as likely false-positives, i.e. incorrectly inferred to be changing over time. On the other hand, genes with large amounts of change over time, but with large replicate variance, may have small F-statistics, and such genes may be false-negatives. By moving (shrinking) the gene specific variance (or more generally, covariance matrix) toward a common value, estimated from the whole gene set, the total number of false positives and false negatives can probably be reduced.

3. Hypothesis testing

The gene ranking/selection problem described above can be formally stated as a hypothesis testing problem. However, in this paper we only seek a statistic for ranking genes in the order of evidence against the null hypothesis; we do not hope to obtain raw or adjusted p-values as in Ge et al. (2003).

For any single gene g, suppose that $\mathbf{X}_{g1}, ..., \mathbf{X}_{gn}$ are n $i.i.d.$ $k\times 1$ time course random

vectors from multivariate normal distribution with gene-specific mean vector $\boldsymbol{\mu}_g$ and covariance matrix $\boldsymbol{\Sigma}_g$, denoted by $N_k(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$, $g = 1, ..., G$. We make the multivariate normality assumption because relative or absolute gene expression measurements are approximately normal on the *log* scale, and this is the most convenient extension to the multivariate case. Our results are to be judged on their practical usefulness, not on the precise fit of our data to a multivariate normal distribution. However, as will be seen shortly, our final formulae involve the multivariate t distribution. Thus a measure of robustness is built in, and so our approach will probably be about as effective for elliptically distributed random vectors. To simpify the notation, the subscript g will be dropped for the rest of the paper. The statistical models presented in the remaining sections are for an arbitrary single gene g.

Following the notation in Bickel and Doksum (2001), the null hypothesis is denoted by H, while the alternative hypothesis is denoted by K. The null hypothesis corresponding to a gene's expression levels being constant is $H : \boldsymbol{\mu} = \mu_0\mathbf{1}$, $\boldsymbol{\Sigma} > 0$, where μ_0 is a scalar representing the expected value of the gene's expression level at any time point under H, and $\mathbf{1}$ is the $k\times 1$ constant vector of $1s$. The alternative hypothesis is $K : \boldsymbol{\mu} \neq \mu_0\mathbf{1}$, $\boldsymbol{\Sigma} > 0$. Later we consider the special case $\mu_0 = 0$.

4. The moderated LR-statistic

A standard likelihood-ratio statistic can be used directly to test the null hypothesis H against the alternative hypothesis K. According to standard multivariate results (e.g. Mardia et al. 2000), under the alternative hypothesis that there are no constraints on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, the maximum likelihood estimates are:

$$\hat{\boldsymbol{\mu}}_K = \overline{\mathbf{X}}, \qquad \widehat{\boldsymbol{\Sigma}}_K = \tfrac{n-1}{n}\mathbf{S},$$

where $\mathbf{S} = (n-1)^{-1}\sum_{i=1}^{n}(\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})'$ is the sample covariance matrix. As in Mardia et al. (2000), under the null hypothesis H, the maximum likelihood estimates for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are

$$\hat{\boldsymbol{\mu}}_H = \left(\frac{\mathbf{1}'\mathbf{S}^{-1}\overline{\mathbf{X}}}{\mathbf{1}'\mathbf{S}^{-1}\mathbf{1}}\right)\mathbf{1}, \qquad \widehat{\boldsymbol{\Sigma}}_H = \tfrac{n-1}{n}\mathbf{S} + \mathbf{d}\mathbf{d}',$$

where $\mathbf{d} = \hat{\boldsymbol{\mu}}_K - \hat{\boldsymbol{\mu}}_H$. The likelihood ratio statistic for testing H against K becomes

$$\mathbf{LR} = 2(\boldsymbol{l}_K^{max} - \boldsymbol{l}_H^{max}) = nlog\left(1 + \frac{n}{n-1}\mathbf{d}'\mathbf{S}^{-1}\mathbf{d}\right).$$

The statistic $n\mathbf{d}'\mathbf{S}^{-1}\mathbf{d}$ is the one-sample Hotelling T^2 statistic, and by section 5.3.1b in Mardia et al. (2000) it follows the Hotelling T^2 distribution $T^2(k-1, n-1)$ under H. Thus under the null hypothesis $((n-k+1)n\mathbf{d}'\mathbf{S}^{-1}\mathbf{d})/((n-1)(k-1))$ has an F-distribution with degrees of freedom $(k-1, n-k+1)$.

In the microarray time course context, the number of replicates is typically smaller than the number of time points, and so $\mathbf{S}$ has less than full rank. Furthermore, as discussed above, we wish to moderate the sample covariance matrix. Our moderated $\mathbf{S}$ will take the form

$$\widetilde{\mathbf{S}} = \frac{\nu\mathbf{\Lambda} + (n-1)\mathbf{S}}{\nu + n - 1}$$

where $\nu > 0$ controls the degree of moderation, and $\mathbf{\Lambda}$ is the common $k \times k$ matrix toward which $\mathbf{S}$ is smoothed. In section 5.6 we give the theoretical reason for choosing this moderated variance-covariance matrix $\widetilde{\mathbf{S}}$. We explain in section 7.3 how we estimate ν and $\mathbf{\Lambda}$. Replacing $\mathbf{S}$ with $\widetilde{\mathbf{S}}$ in the LR-statistic, the moderated LR-statistic is

$$\widetilde{\mathbf{LR}} = 2(\boldsymbol{l}_K^{max} - \boldsymbol{l}_H^{max}) = nlog\left(1 + \frac{n}{n-1}\tilde{\mathbf{d}}'\widetilde{\mathbf{S}}^{-1}\tilde{\mathbf{d}}\right) \tag{1}$$

When all the genes have an equal number of replicates n, equation (1) is a monotonic increasing function of $n\tilde{\mathbf{d}}'\widetilde{\mathbf{S}}^{-1}\tilde{\mathbf{d}}$ and hence they are equivalent. We define the quadratic form $n\tilde{\mathbf{d}}'\widetilde{\mathbf{S}}^{-1}\tilde{\mathbf{d}} = ||n^{1/2}\widetilde{\mathbf{S}}^{-1/2}\tilde{\mathbf{d}}||^2$ to be the moderated one-sample Hotelling T^2 statistic. This is quite similar to the $\widetilde{T}^2$ statistic we derive in section 5. The moderated LR-statistic and the moderated one-sample Hotelling T^2 statistics are hybrids of likelihood and Bayesian statistics since $\widetilde{\mathbf{S}}$ is estimated using the multivariate empirical Bayes procedure we describe in section 7.3.

5. Multivariate hierarchical Bayesian model

5.1 Transformation

For each gene, let I be an indicator random variable such that

$$I = \begin{cases} 1 & \text{if } K \text{ is true} \\ 0 & \text{if } H \text{ is true.} \end{cases}$$

We suppose that I has a Bernoulli distribution with success probability p, $0 < p < 1$ across genes. Let $\mathbf{P} = k^{-1}\mathbf{1}_k\mathbf{1}_k'$ be the $k \times k$ projection matrix onto the rank 1 space of constant vectors, where $\mathbf{1}_k' = (1, ..., 1)$ a $k \times 1$ vector of 1s. Let $\mathbf{P}^c = \mathbf{I}_k - \mathbf{P}$ be the

projection onto the orthogonal complement of $R(\mathbf{P})$. We can write any vector $\boldsymbol{\mu} \in R^k$ as $\boldsymbol{\mu} = \mathbf{P}\boldsymbol{\mu} + \mathbf{P}^c\boldsymbol{\mu}$, and in the case $I = 0$, the second term $\mathbf{P}^c\boldsymbol{\mu}$ vanishes. We build up our multivariate hierarchical Bayesian model by assigning an inverse Wishart prior for the gene-specific covariance matrix $\boldsymbol{\Sigma}$ first:

$$\boldsymbol{\Sigma} \sim Inv\text{-}Wishart_{\nu}((\nu\boldsymbol{\Lambda})^{-1}).$$

Given $\boldsymbol{\Sigma}$, we assign multivariate normal priors for the gene-specific mean $\boldsymbol{\mu}$ for the case of nonconstant genes ($I = 1$) and constant genes ($I = 0$), respectively.

$$\begin{cases} \boldsymbol{\mu}|\boldsymbol{\Sigma}, I = 1 \sim N\left(0, \tau^{-1}\mathbf{P}\boldsymbol{\Sigma}\mathbf{P} + \kappa^{-1}\mathbf{P}^c\boldsymbol{\Sigma}\mathbf{P}^c\right) \\ \boldsymbol{\mu}|\boldsymbol{\Sigma}, I = 0 \sim N\left(0, \tau^{-1}\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}\right). \end{cases}$$

Given $\boldsymbol{\Sigma}$ and $I = 0$, the covariance matrix $\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}$ guarantees that $\boldsymbol{\mu}$ is a constant vector, while when $I = 1$, the extra component $\mathbf{P}^c\boldsymbol{\Sigma}\mathbf{P}^c$ adds further variance to $\boldsymbol{\mu}$ so that it becomes a nonconstant vector. The posterior odds are

$$\frac{P(I = 1|\overline{\mathbf{X}}, \mathbf{S})}{P(I = 0|\overline{\mathbf{X}}, \mathbf{S})} = \frac{p}{1-p}\frac{P(\overline{\mathbf{X}}, \mathbf{S}|I = 1)}{P(\overline{\mathbf{X}}, \mathbf{S}|I = 0)}.$$

The above equality holds by the Bayes rule. In order to obtain the full expression, we need to derive $P(\overline{\mathbf{X}}, \mathbf{S}|I = 1)$ and $P(\overline{\mathbf{X}}, \mathbf{S}|I = 0)$. We have

$$\begin{aligned} P(\overline{\mathbf{X}}, \mathbf{S}|I) &= \int P(\overline{\mathbf{X}}, \mathbf{S}|\boldsymbol{\Sigma}, I)P(\boldsymbol{\Sigma}|I)d\boldsymbol{\Sigma} \\ &= \int \left(\int P(\overline{\mathbf{X}}, \mathbf{S}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, I)P(\boldsymbol{\mu}|\boldsymbol{\Sigma}, I)d\boldsymbol{\mu} \right) P(\boldsymbol{\Sigma}|I)d\boldsymbol{\Sigma} \\ &= \int \left(\int P(\overline{\mathbf{X}}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, I)P(\mathbf{S}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, I)P(\boldsymbol{\mu}|\boldsymbol{\Sigma}, I)d\boldsymbol{\mu} \right) P(\boldsymbol{\Sigma}|I)d\boldsymbol{\Sigma} \\ &= \int \left(\int P(\overline{\mathbf{X}}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, I)P(\boldsymbol{\mu}|\boldsymbol{\Sigma}, I)d\boldsymbol{\mu} \right) P(\mathbf{S}|\boldsymbol{\Sigma}, I)P(\boldsymbol{\Sigma}|I)d\boldsymbol{\Sigma} \\ &= \int P(\overline{\mathbf{X}}|\boldsymbol{\Sigma}, I)P(\mathbf{S}|\boldsymbol{\Sigma}, I)P(\boldsymbol{\Sigma}|I)d\boldsymbol{\Sigma}. \end{aligned}$$

Thus, to derive the joint sampling distribution of $\overline{\mathbf{X}}$ and $\mathbf{S}$ given I, we need to derive $P(\overline{\mathbf{X}}|\boldsymbol{\Sigma}, I)$ first. When $I = 0$, we have

$$P(\overline{\mathbf{X}}|\boldsymbol{\Sigma}, I = 0) = \int P(\overline{\mathbf{X}}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, I = 0)P(\boldsymbol{\mu}|\boldsymbol{\Sigma}, I = 0)d\boldsymbol{\mu},$$

and the same holds for $P(\overline{\mathbf{X}}|\mathbf{\Sigma}, I = 1)$. It turns out that given $\mathbf{\Sigma}$ and $I = 0$, $\overline{\mathbf{X}}$ has a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $(n^{-1}\mathbf{\Sigma}+\tau^{-1}\mathbf{P\Sigma P})$. To get a closed-form expression for the joint sampling distribution $P(\overline{\mathbf{X}}, \mathbf{S}|I)$ and hence the posterior odds, we find it necessary to make an additional assumption, namely that $\mathbf{P\Sigma} = \mathbf{\Sigma P}$. It is shown in section 5.5 that this assumption gives us simple distributional results. With this assumption, given $\mathbf{\Sigma}$ and $I = 0$, $\overline{\mathbf{X}}$ is a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $(n^{-1}\mathbf{\Sigma} + \tau^{-1}\mathbf{\Sigma P})$. Similarly, given $\mathbf{\Sigma}$ and $I = 1$, $\overline{\mathbf{X}}$ is a multivariate normal distribution with mean $\mathbf{0}$ and the covariance matrix $(n^{-1}\mathbf{\Sigma} + \tau^{-1}\mathbf{\Sigma P} + \kappa^{-1}\mathbf{\Sigma P}^c)$.

For the rest of the paper, unless stated otherwise, we assume $\mathbf{P\Sigma} = \mathbf{\Sigma P}$, and we make use of the following lemma:

Lemma 1 *Suppose* $\mathbf{T}$ *is any* $k \times k$ *orthogonal matrix whose first row is constant. Write* $\mathbf{T} = (\mathbf{T}_0', \mathbf{T}_1')'$, *where* $\mathbf{T}_0$ *is the first row of* $\mathbf{T}$, *and* $\mathbf{T}_1$ *is the remainder. Then, for any* $\mathbf{\Sigma} > 0$ *satisfying* $\mathbf{P\Sigma} = \mathbf{\Sigma P}$, $\mathbf{T\Sigma T}' = \widetilde{\mathbf{\Sigma}}$ *is a* $k \times k$ *block diagonal matrix with the scalar* $\tilde{\sigma}^2 > 0$ *as the first block and* $(k-1) \times (k-1)$ *matrix* $\widetilde{\mathbf{\Sigma}}_1 > 0$ *as the second block: i.e.*

$$\mathbf{T\Sigma T}' = \widetilde{\mathbf{\Sigma}} = \begin{pmatrix} \tilde{\sigma}^2 & \mathbf{0} \\ \mathbf{0} & \widetilde{\mathbf{\Sigma}}_1 \end{pmatrix}$$

Proof:

From the condition $\mathbf{P\Sigma} = \mathbf{\Sigma P}$, we conclude that the row and column sums are all equal, i.e., for all i and j, $\sum_{i=1}^k \sigma_{ij} = \sum_{j=1}^k \sigma_{ij} = \alpha$, say. We have $\alpha > 0$ because $\mathbf{1}'\mathbf{\Sigma 1} = \alpha \mathbf{1}'\mathbf{1} = \alpha k > 0$ by positive definiteness. Now we can write

$$\mathbf{T\Sigma T}' = \begin{pmatrix} \mathbf{T}_0\mathbf{\Sigma T}_0' & \mathbf{T}_0\mathbf{\Sigma T}_1' \\ \mathbf{T}_1\mathbf{\Sigma T}_0' & \mathbf{T}_1\mathbf{\Sigma T}_1' \end{pmatrix}.$$

We now show that $\mathbf{T}_0\mathbf{\Sigma T}_1' = \mathbf{0}$. We have $\mathbf{T}_0\mathbf{\Sigma} = \alpha k^{-\frac{1}{2}}\mathbf{1}'$, and thus

$$\mathbf{T}_0\mathbf{\Sigma T}_1' = \alpha k^{-\frac{1}{2}}\mathbf{1}'\mathbf{T}_1' = \alpha k^{-\frac{1}{2}}(\mathbf{T}_1\mathbf{1})' = \mathbf{0},$$

since $\mathbf{T}_1\mathbf{1} = \mathbf{0}$.

We next show $\mathbf{T}_0\mathbf{\Sigma T}_0' > 0$ and $\mathbf{T}_1\mathbf{\Sigma T}_1' > 0$. First,

$$\mathbf{T}_0\mathbf{\Sigma T}_0' = \alpha k^{-\frac{1}{2}}(\mathbf{T}_0\mathbf{1})' = \alpha > 0.$$

Next, take a $(k-1)$-vector $\mathbf{a} \neq 0$. Then

$$\mathbf{a}'\mathbf{T}_1\mathbf{\Sigma T}_1'\mathbf{a} = (\mathbf{T}_1'\mathbf{a})'\mathbf{\Sigma}(\mathbf{T}_1'\mathbf{a}) > 0.$$

This proves that $\mathbf{T}_1\mathbf{\Sigma}\mathbf{T}_1^{'} > 0$, and the proof is complete.

As an example, let $\mathbf{T}$ be the Helmert matrix, where the ji–th element of $\mathbf{T}$ is defined as

$$\begin{cases} t_{ji} = 1/\sqrt{k} \quad \text{for} \quad j = 1, i = 1, ..., k \\ t_{ji} = 1/\sqrt{j(j-1)} \quad \text{for} \quad 2 \leq j \leq k, 1 \leq i \leq j-1 \\ t_{ji} = -(j-1)/\sqrt{j(j-1)} \quad \text{for} \quad 2 \leq j \leq k, i = j \\ t_{ji} = 0 \quad \text{for} \quad 2 \leq j \leq k-1, j+1 \leq i \leq k. \end{cases}$$

Here $\mathbf{T}$ is partitioned into its first row $\mathbf{T}_0$ $(1 \times k)$ and its last k–1 rows $\mathbf{T}_1$ $((k-1) \times k)$. Since the $\mathbf{X}_1, ..., \mathbf{X}_n$ are $i.i.d.$ $N(\boldsymbol{\mu}, \mathbf{\Sigma})$, the transformed random vectors $\mathbf{T}\mathbf{X}_i$ are also multivariate normally distributed with mean $\mathbf{T}\boldsymbol{\mu}$ and covariance matrix $\widetilde{\mathbf{\Sigma}}$, i.e.

$$\mathbf{TX_1}, \cdots, \mathbf{TX_n} | \mathbf{T}\boldsymbol{\mu}, \widetilde{\mathbf{\Sigma}} \quad \sim \quad N(\mathbf{T}\boldsymbol{\mu}, \widetilde{\mathbf{\Sigma}}).$$

By lemma 1, the matrix $\widetilde{\mathbf{\Sigma}}$ is a block diagonal matrix with $\tilde{\sigma}^2$ as the first block, and $\widetilde{\mathbf{\Sigma}}_1$ as the second block. Defining $\bar{x}_i = k^{-1} \sum_{j=1}^{k} X_{ij}$, then $\sqrt{k}\bar{x}_i$ and the random vector $\mathbf{T}_1\mathbf{X}_i$ are independent and normally distributed, with distributions

$$\begin{cases} \sqrt{k}\bar{x}_i | \mathbf{T}_0\mu, \tilde{\sigma}^2 \quad \sim \quad N(\mathbf{T}_0\boldsymbol{\mu}, \tilde{\sigma}^2), \\ \mathbf{T_1}\mathbf{X}_i | \mathbf{T}_1\boldsymbol{\mu}, \widetilde{\mathbf{\Sigma}}_1 \quad \sim \quad N(\mathbf{T}_1\boldsymbol{\mu}, \widetilde{\mathbf{\Sigma}}_1). \end{cases}$$

This transformation allows us to separate the gene expression changes into constant and non-constant changes.

5.2 Priors

The prior for $\widetilde{\mathbf{\Sigma}}$ is first set through the independent priors for $\tilde{\sigma}^2$ and $\widetilde{\mathbf{\Sigma}}_1$. We suppose that $\tilde{\sigma}^2$ and $\widetilde{\mathbf{\Sigma}}_1$ are independently distributed, with an inverse gamma distribution with shape parameter $\xi/2$ and scale parameter $\xi\lambda^2/2$, and an inverse Wishart distribution with degrees of freedom ν and scale matrix $\nu\mathbf{\Lambda}$, respectively (Gelman et al. 2000), i.e.

$$\begin{cases} \tilde{\sigma}^2 \sim inv\text{–}gamma\left(\frac{1}{2}\xi, \frac{1}{2}\xi\lambda^2\right), \\ \\ \widetilde{\mathbf{\Sigma}}_1 \sim Inv\text{–}Wishart_{\nu}((\nu\mathbf{\Lambda})^{-1}). \end{cases} \tag{2}$$

The prior for $\mathbf{T}\boldsymbol{\mu}$ has four parts. We assign independent priors to $\mathbf{T}_0\boldsymbol{\mu}$ and $\mathbf{T}_1\boldsymbol{\mu}$, separately for the cases $I = 1$ and $I = 0$. For the case $I = 1$, priors are

$$\begin{cases} \mathbf{T}_0\boldsymbol{\mu} | \tilde{\sigma}^2, I = 1 \sim N\left(\theta, \kappa^{-1}\tilde{\sigma}^2\right), \\ \\ \mathbf{T}_1\boldsymbol{\mu} | \widetilde{\mathbf{\Sigma}}_1, I = 1 \sim N\left(\mathbf{0}, \eta^{-1}\widetilde{\mathbf{\Sigma}}_1\right), \end{cases} \tag{3}$$

where $\theta \geq 0$ is the mean, and $\kappa > 0$ and $\eta > 0$ are scale parameters. When $I = 0$, $\mathbf{T}_1\boldsymbol{\mu} = \mathbf{0}$ with probability 1. Thus, the priors in this case are

$$\begin{cases} \mathbf{T}_0\boldsymbol{\mu}|\tilde{\sigma}^2, I = 0 \sim N\left(\theta, \kappa^{-1}\tilde{\sigma}^2\right), \\ \\ \mathbf{T}_1\boldsymbol{\mu}|\widetilde{\boldsymbol{\Sigma}}_1, I = 0 \quad \equiv \mathbf{0}. \end{cases} \tag{4}$$

It is reasonable to assume $P(\mathbf{T}_0\boldsymbol{\mu}|\tilde{\sigma}^2, I = 0) = P(\mathbf{T}_0\boldsymbol{\mu}|\tilde{\sigma}^2, I = 1)$ for large genome-wide arrays since there is no obvious reason why the expected grand mean of the expression levels for nonconstant genes should differ from that of constant genes. For two-color comparative microarray experiments, it is also reasonable to assume $\theta = 0$.

5.3 Likelihoods

Define $\bar{\bar{x}} = n^{-1}\sum_{i=1}^{n}\bar{x}_i$ and $\mathbf{T}_1\overline{\mathbf{X}} = n^{-1}\sum_{i=1}^{n}\mathbf{T}_1\mathbf{X}_i$. By a standard property of the multivariate normal distribution, $\bar{\bar{x}}$ and $\mathbf{T}_1\overline{\mathbf{X}}$ are independent. The independent sums of squares and products associated with $\bar{\bar{x}}$ and $\mathbf{T}_1\overline{\mathbf{X}}$ are $s^2 = (n-1)^{-1}\sum_{i=1}^{n}(\bar{x}_i - \bar{\bar{x}})^2$, and $\mathbf{S}_1 = (n-1)^{-1}\sum_{i=1}^{n}(\mathbf{T}_1\mathbf{X}_i - \mathbf{T}_1\overline{\mathbf{X}})(\mathbf{T}_1\mathbf{X}_i - \mathbf{T}_1\overline{\mathbf{X}})'$, respectively. It is well known that $\bar{\bar{x}}$ and s^2 are sufficient statistics for $\mathbf{T}_0\boldsymbol{\mu}$ and $\tilde{\sigma}^2$, while $\mathbf{T}_1\overline{\mathbf{X}}$ and $\mathbf{S}_1$ are sufficient statistics for $\mathbf{T}_1\boldsymbol{\mu}$ and $\widetilde{\boldsymbol{\Sigma}}_1$, respectively. It is sufficient to set up the likelihoods of these four random variables for our hierarchical model, and these follow from

$$\begin{cases} \sqrt{k}\bar{\bar{x}} \mid \mathbf{T}_0\boldsymbol{\mu}, \tilde{\sigma}^2 \sim N\left(\mathbf{T}_0\boldsymbol{\mu}, n^{-1}\tilde{\sigma}^2\right), \\ ks^2 \mid \tilde{\sigma}^2 \sim (n-1)^{-1}\tilde{\sigma}^2\chi^2_{n-1}, \\ \\ \mathbf{T}_1\overline{\mathbf{X}} \mid \mathbf{T}_1\boldsymbol{\mu}, \widetilde{\boldsymbol{\Sigma}}_1 \sim N\left(\mathbf{T}_1\boldsymbol{\mu}, n^{-1}\widetilde{\boldsymbol{\Sigma}}_1\right), \\ \\ \mathbf{S}_1 \mid \widetilde{\boldsymbol{\Sigma}}_1 \sim Wishart_{n-1}\left((n-1)^{-1}\widetilde{\boldsymbol{\Sigma}}_1\right). \end{cases} \tag{5}$$

5.4 Univariate joint sampling distribution

Once the priors and likelihoods are set, the joint sampling distributions can be determined for the cases $I = 1$ and $I = 0$. For the case $I = 1$, the joint sampling distribution is written as

$$P(\sqrt{k}\bar{\bar{x}}, ks^2, \mathbf{T}_1\overline{\mathbf{X}}, \mathbf{S}_1|I = 1)$$

$$= \int\int\int\int P(\sqrt{k}\bar{\bar{x}}, ks^2, \mathbf{T}_1\overline{\mathbf{X}}, \mathbf{S}_1|\mathbf{T}_0\boldsymbol{\mu}, \tilde{\sigma}^2, \mathbf{T}_1\boldsymbol{\mu}, \widetilde{\boldsymbol{\Sigma}}_1, I = 1)$$

$$P(\mathbf{T}_0\boldsymbol{\mu}, \tilde{\sigma}^2, \mathbf{T}_1\boldsymbol{\mu}, \widetilde{\boldsymbol{\Sigma}}_1|I = 1)\,\mathrm{d}\mathbf{T}_0\boldsymbol{\mu}\,\mathrm{d}\tilde{\sigma}^2\,\mathrm{d}\mathbf{T}_1\boldsymbol{\mu}\,\mathrm{d}\widetilde{\boldsymbol{\Sigma}}_1.$$

Since we assume that $(\mathbf{T}_0\boldsymbol{\mu}, \tilde{\sigma}^2)$ and $(\mathbf{T}_1\boldsymbol{\mu}, \widetilde{\boldsymbol{\Sigma}}_1)$ are independent, and $(\sqrt{k}\bar{\tilde{x}}, ks^2)$ and $(\mathbf{T}_1\overline{\mathbf{X}}, \mathbf{S}_1)$ are also independent, given $(\mathbf{T}_0\boldsymbol{\mu}, \tilde{\sigma}^2)$ and $(\mathbf{T}_1\boldsymbol{\mu}, \widetilde{\boldsymbol{\Sigma}}_1)$, respectively, the above expression becomes

$$\begin{aligned}
&\int\int P(\sqrt{k}\bar{\tilde{x}}, ks^2|\mathbf{T}_0\boldsymbol{\mu}, \tilde{\sigma}^2, I=1)P(\mathbf{T}_0\boldsymbol{\mu}, \tilde{\sigma}^2|I=1)\,\mathrm{d}\mathbf{T}_0\boldsymbol{\mu}\,\mathrm{d}\tilde{\sigma}^2 \times \\
&\int\int P(\mathbf{T}_1\overline{\mathbf{X}}, \mathbf{S}_1|\mathbf{T}_1\boldsymbol{\mu}, \widetilde{\boldsymbol{\Sigma}}_1, I=1)P(\mathbf{T}_1\boldsymbol{\mu}, \widetilde{\boldsymbol{\Sigma}}_1|I=1)\,\mathrm{d}\mathbf{T}_1\boldsymbol{\mu}\,\mathrm{d}\widetilde{\boldsymbol{\Sigma}}_1 \qquad (6) \\
&= P(\sqrt{k}\bar{\tilde{x}}, ks^2|I=1)P(\mathbf{T}_1\overline{\mathbf{X}}, \mathbf{S}_1|I=1),
\end{aligned}$$

where

$$P(\sqrt{k}\bar{\tilde{x}}, ks^2|I=1) = \int P(\sqrt{k}\bar{\tilde{x}}|\tilde{\sigma}^2, I=1)P(ks^2|\tilde{\sigma}^2, I=1)P(\tilde{\sigma}^2|I=1)\mathrm{d}\tilde{\sigma}^2.$$

Now $\sqrt{k}\bar{\tilde{x}}|\tilde{\sigma}^2, I=1$ is normally distributed with mean θ and variance $(n^{-1}+\kappa^{-1})\,\tilde{\sigma}^2$. The joint sampling distribution of $\sqrt{k}\bar{\tilde{x}}$ and ks^2 is therefore

$$P(\sqrt{k}\bar{\tilde{x}}, ks^2|I=1) = \frac{\Gamma\left(\frac{n+\xi}{2}\right) ks^{2\left(\frac{n-1}{2}-1\right)} \left(\frac{n-1}{2}\right)^{\frac{1}{2}(n-1)} \left(\frac{\xi\lambda^2}{2}\right)^{\frac{1}{2}\xi}}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{\xi}{2}\right)\sqrt{2\pi(n^{-1}+\kappa^{-1})}} \times$$

$$\left(\frac{\frac{(\sqrt{k}\bar{\tilde{x}}-\theta)^2}{n^{-1}+\kappa^{-1}} + ks^2(n-1) + \lambda^2\xi}{2}\right)^{-\frac{1}{2}(n+\xi)}.$$

Following Smyth (2004), the univariate moderated t-statistic is defined as

$$\tilde{t} = n^{\frac{1}{2}}(k^{\frac{1}{2}}\bar{\tilde{x}} - \theta)\tilde{s}^{-1},$$

where $\tilde{s}^2 = (n-1+\xi)^{-1}((n-1)ks^2 + \xi\lambda^2)$ is the reciprocal of the posterior mean of $\tilde{\sigma}^{-2}$ given ks^2. By Smyth (2004), the statistic $\tilde{t}$ is independent of s^2 and distributed as a scaled t with $n-1+\xi$ degrees of freedom and scale parameter $\sqrt{\kappa^{-1}(n+\kappa)}$ (Gelman et al. 2000). Thus

$$P(\tilde{t}|I=1) = \frac{\Gamma\left(\frac{n+\xi}{2}\right)}{\Gamma\left(\frac{n+\xi-1}{2}\right)}\pi^{-\frac{1}{2}}(n-1+\xi)^{-\frac{1}{2}} \times$$

$$\left(\frac{\kappa}{n+\kappa}\right)^{\frac{1}{2}} \left(1 + \left(\frac{1}{n-1+\xi}\right)\left(\frac{\kappa}{n+\kappa}\right)\tilde{t}^2\right)^{-\frac{1}{2}(n+\xi)}.$$

The distribution of $\tilde{t}$ given $I = 0$ is the same as that given $I = 1$.

5.5 Multivariate joint sampling distribution

In this section we derive the joint sampling distribution of $\mathbf{T}_1\overline{\mathbf{X}}$ and $\mathbf{S}_1$. When $I = 1$, it can be written as

$$P(\mathbf{T}_1\overline{\mathbf{X}}, \mathbf{S}_1|I=1) = \int P(\mathbf{T}_1\overline{\mathbf{X}}|\widetilde{\boldsymbol{\Sigma}}_1, I=1)P(\mathbf{S}_1|\widetilde{\boldsymbol{\Sigma}}_1, I=1)P(\widetilde{\boldsymbol{\Sigma}}_1|I=1)\mathrm{d}\widetilde{\boldsymbol{\Sigma}}_1.$$

Since $P(\mathbf{T}_1\overline{\mathbf{X}}|\widetilde{\boldsymbol{\Sigma}}_1, I=1) = \int P(\mathbf{T}_1\overline{\mathbf{X}}|\mathbf{T}_1\boldsymbol{\mu}, \widetilde{\boldsymbol{\Sigma}}_1, I=1)P(\mathbf{T}_1\boldsymbol{\mu}|\widetilde{\boldsymbol{\Sigma}}_1, I=1)\mathrm{d}\mathbf{T}_1\boldsymbol{\mu}$, the first term inside the integral, the distribution of $\mathbf{T}_1\overline{\mathbf{X}}$, given $\widetilde{\boldsymbol{\Sigma}}_1$ and $I = 1$, is $(k-1)$-variate normal with mean $\mathbf{0}$ and variance $(n^{-1}+\eta^{-1})\,\widetilde{\boldsymbol{\Sigma}}_1$. Therefore, the joint sampling distribution of $\mathbf{T}_1\overline{\mathbf{X}}$ and $\mathbf{S}_1$, given $I = 1$ is

$$P(\mathbf{T}_1\overline{\mathbf{X}}, \mathbf{S}_1|I=1) =$$

$$\int P(\mathbf{T}_1\overline{\mathbf{X}}|\widetilde{\boldsymbol{\Sigma}}_1, I=1)P(\mathbf{S}_1|\widetilde{\boldsymbol{\Sigma}}_1, I=1)P(\widetilde{\boldsymbol{\Sigma}}_1|I=1)\mathrm{d}\widetilde{\boldsymbol{\Sigma}}_1 \tag{7}$$

$$= \int (2\pi)^{-\frac{1}{2}(k-1)} \left|(n^{-1}+\eta^{-1})\widetilde{\boldsymbol{\Sigma}}_1\right|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(\mathbf{T}_1\overline{\mathbf{X}})'\left((n^{-1}+\eta^{-1})\widetilde{\boldsymbol{\Sigma}}_1\right)^{-1}\mathbf{T}_1\overline{\mathbf{X}}\right\}\times$$

$$\frac{(n-1)^{\frac{1}{2}(k-1)(n-1)}|\widetilde{\boldsymbol{\Sigma}}_1|^{-\frac{1}{2}(n-1)}|\mathbf{S}_1|^{\frac{1}{2}(n-k-1)}exp\left\{-\frac{1}{2}tr(\widetilde{\boldsymbol{\Sigma}}_1^{-1}(n-1)\mathbf{S}_1)\right\}}{2^{\frac{1}{2}(n-1)(k-1)}\boldsymbol{\Gamma}_{k-1}(\frac{n-1}{2})}\times$$

$$\frac{\nu^{\frac{1}{2}(k-1)\nu}|\boldsymbol{\Lambda}|^{\frac{\nu}{2}}|\widetilde{\boldsymbol{\Sigma}}_1|^{-\frac{1}{2}(\nu+k)}exp\left\{-\frac{1}{2}tr(\nu\boldsymbol{\Lambda}\widetilde{\boldsymbol{\Sigma}}_1^{-1})\right\}}{2^{\frac{1}{2}\nu(k-1)}\boldsymbol{\Gamma}_{k-1}(\frac{\nu}{2})}\mathrm{d}\widetilde{\boldsymbol{\Sigma}}_1$$

$$= \frac{(2\pi)^{-\frac{1}{2}(k-1)}(n-1)^{\frac{1}{2}(k-1)(n-1)}|\mathbf{S}_1|^{\frac{1}{2}(n-k-1)}|\boldsymbol{\Lambda}|^{\frac{\nu}{2}}\nu^{\frac{1}{2}(k-1)\nu}\,(n^{-1}+\eta^{-1})^{-\frac{1}{2}(k-1)}}{2^{\frac{1}{2}(n-1)(k-1)}2^{\frac{1}{2}\nu(k-1)}\boldsymbol{\Gamma}_{k-1}(\frac{n-1}{2})\boldsymbol{\Gamma}_{k-1}(\frac{\nu}{2})}\times$$

$$\int |\widetilde{\boldsymbol{\Sigma}}_1|^{-\frac{1}{2}(n+\nu+k)}\times$$

$$exp\left\{-\frac{1}{2}\left((\mathbf{T}_1\overline{\mathbf{X}})'\left((n^{-1}+\eta^{-1})\,\widetilde{\boldsymbol{\Sigma}}_1\right)^{-1}\mathbf{T}_1\overline{\mathbf{X}} + tr(\widetilde{\boldsymbol{\Sigma}}_1^{-1}(n-1)\mathbf{S}_1) + tr(\nu\boldsymbol{\Lambda}\widetilde{\boldsymbol{\Sigma}}_1^{-1})\right)\right\}\mathrm{d}\widetilde{\boldsymbol{\Sigma}}_1.$$

The last integrand in equation (7) is

$$\int |\widetilde{\mathbf{\Sigma}}_1|^{-\frac{1}{2}(n+\nu+k)} \times \\ exp\left\{-\tfrac{1}{2} tr\left(\widetilde{\mathbf{\Sigma}}_1^{-1}\left(\left(n^{-1}+\eta^{-1}\right)^{-1}\mathbf{T}_1\overline{\mathbf{X}\mathbf{X}}'\mathbf{T}_1' + (n-1)\mathbf{S}_1 + \nu\mathbf{\Lambda}\right)\right)\right\} \mathrm{d}\widetilde{\mathbf{\Sigma}}_1. \tag{8}$$

Using the probability density function of the inverse Wishart distribution (Gelman et al. 2000), equation (8) integrates to

$$\frac{2^{\frac{1}{2}(k-1)(n+\nu)}\mathbf{\Gamma}_{k-1}(\frac{n+\nu}{2})}{\left|\left(n^{-1}+\eta^{-1}\right)^{-1}\mathbf{T}_1\overline{\mathbf{X}\mathbf{X}}'\mathbf{T}_1' + (n-1)\mathbf{S}_1 + \nu\mathbf{\Lambda}\right|^{\frac{1}{2}(n+\nu)}}. \tag{9}$$

Equation (7) becomes

$$P(\mathbf{T}_1\overline{\mathbf{X}}, \mathbf{S}_1 | I=1) =$$

$$\frac{\mathbf{\Gamma}_{k-1}\left(\frac{n+\nu}{2}\right)}{\mathbf{\Gamma}_{k-1}\left(\frac{n-1}{2}\right)\mathbf{\Gamma}_{k-1}\left(\frac{\nu}{2}\right)} \times (n-1)^{\frac{1}{2}(k-1)(n-1)}\nu^{-\frac{1}{2}(k-1)n}\pi^{-\frac{1}{2}(k-1)}\left(n^{-1}+\eta^{-1}\right)^{-\frac{1}{2}(k-1)}$$

$$\times \frac{|\mathbf{\Lambda}|^{-\frac{1}{2}n}|\mathbf{S}_1|^{\frac{1}{2}(n-k-1)}}{|\mathbf{I}_{k-1} + ((n^{-1}+\eta^{-1})\nu\mathbf{\Lambda})^{-1}(\mathbf{T}_1\overline{\mathbf{X}})(\mathbf{T}_1\overline{\mathbf{X}})' + \left(\frac{\nu\mathbf{\Lambda}}{n-1}\right)^{-1}\mathbf{S}_1|^{\frac{1}{2}(n+\nu)}}$$

Thus, given $I=1$, $\mathbf{T}_1\overline{\mathbf{X}}$ and $\mathbf{S}_1$ follow a Student-Siegel distribution (Aitchison and Dunsmore, page 257). Following Aitchison and Dunsmore's notation, this distribution is denoted by

$$StSi_{k-1}(\nu; \mathbf{0}, (n^{-1}+\eta^{-1})\mathbf{\Lambda}; n-1, (n-1)^{-1}\nu\mathbf{\Lambda}).$$

5.6 Multivariate moderated t-statistic

A multivariate moderated t-statistic $\tilde{\mathbf{t}}$ is defined, in the hope that it will serve as a useful complement to the multivariate log posterior odds. We write

$$\tilde{\mathbf{t}} = n^{\frac{1}{2}}\widetilde{\mathbf{S}}_1^{-\frac{1}{2}}\mathbf{T}_1\overline{\mathbf{X}}, \tag{10}$$

where $\widetilde{\mathbf{S}}_1$ is the inverse of the posterior mean of $\widetilde{\mathbf{\Sigma}}_1^{-1}$ given $\mathbf{S}_1$

$$\widetilde{\mathbf{S}}_1 = [E(\widetilde{\mathbf{\Sigma}}_1^{-1}|\mathbf{S}_1)]^{-1} = \frac{(n-1)\mathbf{S}_1 + \nu\mathbf{\Lambda}}{n-1+\nu}.$$

By Gupta and Nagar (2000), the Jacobian transformation from $\mathbf{T}_1\overline{\mathbf{X}}$ to $\tilde{\mathbf{t}}$ is $J(\mathbf{T}_1\overline{\mathbf{X}} \rightarrow \tilde{\mathbf{t}}) = |n^{-1/2}\widetilde{\mathbf{S}}_1^{1/2}|$. Substituting for $\mathbf{T}_1\overline{\mathbf{X}}$ in terms of $\tilde{\mathbf{t}}$ in the joint sampling distribution of $\mathbf{T}_1\overline{\mathbf{X}}$ and $\mathbf{S}_1$, and multiplying the resulting expression by $J(\mathbf{T}_1\overline{\mathbf{X}} \rightarrow \tilde{\mathbf{t}})$, the joint sampling distribution of $\tilde{\mathbf{t}}$ and $\mathbf{S}_1$ is derived:

$$P(\tilde{\mathbf{t}}, \mathbf{S}_1|I=1) =$$

$$\frac{\mathbf{\Gamma}_{k-1}\left(\frac{n+\nu-1}{2}\right)}{\mathbf{\Gamma}_{k-1}\left(\frac{n-1}{2}\right)\mathbf{\Gamma}_{k-1}\left(\frac{\nu}{2}\right)} \frac{|\mathbf{S}_1|^{\frac{1}{2}(n-k-1)}}{|\frac{\nu\mathbf{\Lambda}}{n-1}|^{\frac{1}{2}(n-1)}|\mathbf{I}_{k-1}+(\frac{\nu\mathbf{\Lambda}}{n-1})^{-1}\mathbf{S}_1|^{\frac{1}{2}(n+\nu-1)}} \times$$

$$\pi^{-\frac{1}{2}(k-1)}\frac{\Gamma\left(\frac{n+\nu}{2}\right)}{\Gamma\left(\frac{n+\nu-k+1}{2}\right)}\left(\frac{n+\eta}{\eta}\right)^{-\frac{1}{2}(k-1)}(n-1+\nu)^{-\frac{1}{2}(k-1)}\times$$

$$\left(1+\frac{1}{n-1+\nu}\left(\frac{\eta}{n+\eta}\right)\tilde{\mathbf{t}}'\tilde{\mathbf{t}}\right)^{-\frac{1}{2}(n+\nu)}.$$

The above expression factorizes into parts involving $\mathbf{S}_1$ only and $\tilde{\mathbf{t}}$ only, proving that $\tilde{\mathbf{t}}$ and $\mathbf{S}_1$ are independent. It is apparent that $\tilde{\mathbf{t}}$ has a multivariate t-distribution with $n+\nu-k+1$ degrees of freedom, scale parameter $n+\nu-1$, covariance matrix $\eta^{-1}(n+\eta)\mathbf{I}_{k-1}$, and mean vector $\mathbf{0}$. This distribution is denoted by $\tilde{\mathbf{t}}|I=1 \sim \mathbf{t}_{k-1}(n+\nu-k+1, n+\nu-1, \mathbf{0}, \eta^{-1}(n+\eta)\mathbf{I}_{k-1})$ (Gupta and Nagar 2000) with probability density function

$$P(\tilde{\mathbf{t}}|I=1) = \pi^{-\frac{1}{2}(k-1)}\frac{\Gamma\left(\frac{n+\nu}{2}\right)}{\Gamma\left(\frac{n+\nu-k+1}{2}\right)}\left(\frac{n+\eta}{\eta}\right)^{-\frac{1}{2}(k-1)}(n-1+\nu)^{-\frac{1}{2}(k-1)}\times \left(1+\frac{1}{n-1+\nu}\left(\frac{\eta}{n+\eta}\right)\tilde{\mathbf{t}}'\tilde{\mathbf{t}}\right)^{-\frac{1}{2}(n+\nu)}. \tag{11}$$

It is straightforward to see that $\tilde{\mathbf{t}}|I=0 \sim \mathbf{t}_{k-1}(n+\nu-k+1, n+\nu-1, \mathbf{0}, \mathbf{I}_{k-1})$. Given $I=1$, $\mathbf{S}_1$ is distributed as generalized type-II beta distribution with parameters $(n-1)/2$, $\nu/2$, scale matrix $\nu\mathbf{\Lambda}/(n-1)$, and location matrix $\mathbf{0}$. The distribution is denoted by $GB_{k-1}^{II}((n-1)/2, \nu/2, \nu\mathbf{\Lambda}/(n-1), \mathbf{0})$ (Gupta and Nagar 2000) with probability density function

$$P(\mathbf{S}_1|I=1) = \frac{1}{\boldsymbol{\beta}_{k-1}(\frac{n-1}{2}, \frac{\nu}{2})}\frac{|\mathbf{S}_1|^{\frac{1}{2}(n-k-1)}}{|\frac{\nu\mathbf{\Lambda}}{n-1}|^{\frac{1}{2}(n-1)}|\mathbf{I}_{k-1}+(\frac{\nu\mathbf{\Lambda}}{n-1})^{-1}\mathbf{S}_1|^{\frac{1}{2}(n+\nu-1)}}.$$

The marginal sampling distribution of $\mathbf{S}_1$ does not depend on I so that $P(\mathbf{S}_1|I=0) = P(\mathbf{S}_1|I=1)$.

5.7 Posterior Odds

The posterior odds are the probability that the expected time course $\boldsymbol{\mu}$ is nonconstant ($i.e., I=1$) over the probability that $\boldsymbol{\mu}$ is constant ($i.e.$, $I=0$), given the sufficient statistics $\tilde{t}, ks^2, \tilde{\mathbf{t}}$, and $\mathbf{S}_1$. Following Smyth's (2004) notation, we denote the multivariate posterior odds by

$$\mathbf{O} = \frac{P(I=1|\tilde{t}, ks^2, \tilde{\mathbf{t}}, \mathbf{S}_1)}{P(I=0|\tilde{t}, ks^2, \tilde{\mathbf{t}}, \mathbf{S}_1)} = \frac{p}{1-p}\frac{P(\tilde{t}, ks^2, \tilde{\mathbf{t}}, \mathbf{S}_1|I=1)}{P(\tilde{t}, ks^2, \tilde{\mathbf{t}}, \mathbf{S}_1|I=0)} = \frac{p}{1-p}\frac{P(\tilde{\mathbf{t}}|I=1)}{P(\tilde{\mathbf{t}}|I=0)}.$$

The last equality holds because $\tilde{t}, s^2, \tilde{\mathbf{t}}$, and $\mathbf{S}_1$ are mutually independent and the distributions of $\tilde{t}$, s^2 and $\mathbf{S}_1$ do not depend on I. Plugging in the density functions of $\tilde{\mathbf{t}}$ in section 5.6, and defining $\widetilde{T}^2 = \tilde{\mathbf{t}}'\tilde{\mathbf{t}}$, $\mathbf{O}$ become

$$\begin{aligned}\mathbf{O} &= \frac{p}{1-p}\left(\frac{\eta}{n+\eta}\right)^{\frac{1}{2}(k-1)}\left(\frac{n-1+\nu+\tilde{\mathbf{t}}'\tilde{\mathbf{t}}}{n-1+\nu+\left(\frac{\eta}{n+\eta}\right)\tilde{\mathbf{t}}'\tilde{\mathbf{t}}}\right)^{\frac{1}{2}(n+\nu)} \\ &= \frac{p}{1-p}\left(\frac{\eta}{n+\eta}\right)^{\frac{1}{2}(k-1)}\left(\frac{n-1+\nu+\widetilde{T}^2}{n-1+\nu+(\frac{\eta}{n+\eta})\widetilde{T}^2}\right)^{\frac{1}{2}(n+\nu)}.\end{aligned} \tag{12}$$

Following the tradition in genetics, the *log* base 10 of the above expression is called the LOD score. To distinguish it from the LOD score (also called the B-statistic) in the univariate model of Lönnstedt and Speed (2002) and Smyth (2004), the multivariate LOD score in this paper is called the MB-statistic,

$$MB = log_{10}\mathbf{O}. \tag{13}$$

When all genes have the same number of replicates n, equation (13) is a monotonic increasing function of $\widetilde{T}^2 = \tilde{\mathbf{t}}'\tilde{\mathbf{t}}$. This shows that the MB statistic is equivalent to the $\widetilde{T}^2$ statistic when n is the same across genes, and therefore, one is encouraged to use the $\widetilde{T}^2$ statistic in this case since it does not require the estimation of η and leads to the same rankings as equation (13). Under H, $(k-1)^{-1}\widetilde{T}^2$ has an F distribution with degrees of freedom $(k-1, n+\nu-k+1)$, or equivalently, $(n+\nu-k+1)^{-1}(n+\nu-1)\widetilde{T}^2$ has a Hotelling T^2 distribution $T^2(k-1, n+\nu-1)$. It turns out that $\mathbf{O}$ (or equivalently, the MB-statistic) simplifies in several special and limiting cases of interest. We discuss

these in sections 7 and 8.

6. Hyperparameter estimation

Section 5 shows that the MB-statistic for assessing whether or not a time course is constant depends on $(k^2 - k + 6)/2$ hyperparameters: $\nu, \mathbf{\Lambda}, \eta$, and p. In practice, we need to estimate these hyperparameters, and plug in our estimates into the formula for $\widetilde{\mathbf{S}}_1$, $\tilde{\mathbf{t}}$, $\mathbf{O}$,...etc. Slightly abusing our notation, we will use the same symbols for these estimates, relying on context to make it clear whether we are assuming the hyperparameters known or not. In our multivariate model, many more hyperparameters need to be estimated, compared to the univariate models in Lönnstedt and Speed (2002) and Smyth (2004), both of which have 4 hyperparameters. Smyth (2004) derives closed form estimators for the hyperparameters in the univariate linear model setting, using the marginal sampling distributions of the statistic $\tilde{t}$ and the sample variance s^2, and shows that the estimators are better than the simple estimators in Lönnstedt and Speed (2002). Following Smyth (2004), the aim of this section is to derive estimators for the hyperparameters in our multivariate model. In general, the hyperparameter η associated with the nonconstant case $I = 1$ is estimated based on only a small subset of genes, while ν and $\mathbf{\Lambda}$ are estimated using the whole gene set. Instead of estimating the proportion of differentially expressed genes p, we only plug in an user-defined value for p since the choice of p does not affect the rankings of genes based on the MB-statistic.

6.1 EB estimation of ν and $\mathbf{\Lambda}$

The hyperparameter ν determines the degree of smoothness between $\mathbf{S}_1$ and $\mathbf{\Lambda}$. The method we use to estimate ν builds on that used to estimate d_0 in section 6.2 in Smyth (2004). However, unlike d_0 in Smyth (2004), ν is associated with the (k-1) $\times$ (k-1) dimensional matrix $\widetilde{\mathbf{\Sigma}}_1$. Therefore, a method appropriate to this multivariate framework is needed. Let $\mathbf{T}_{1j}$ be the *j-th* row of $\mathbf{T}_1$. Then $\mathbf{T}_{1j}\mathbf{X}_1$,...,$\mathbf{T}_{1j}\mathbf{X}_n$ are n independent scalar random variables. Since $\mathbf{T}_1$ has k-1 rows, it follows that we have k-1 sets of such data. Let $\hat{\nu}_j$ be the estimated prior degrees of freedom for the *j-th* set of data based on the method proposed in section 6.2 in Smyth (2004). Our estimation of ν is based on the following two-step strategy. For the simulation study, the constraint that $\nu \geq k+5$ is enforced because this guarantees that the third moment of $\widetilde{\mathbf{\Sigma}}_1$ exists (Gupta and Nagar 2000) and the corresponding $\widetilde{\mathbf{\Sigma}}$s simulated are more stable. Furthermore, such constraint guarantees the positive definiteness of the estimated $\mathbf{\Lambda}$. As the first step, set ν as $\hat{\nu} = max(mean(\hat{\nu}_j), k + 5)$. This estimated $\hat{\nu}$ is used to estimate $\mathbf{\Lambda}$. Once $\mathbf{\Lambda}$

is estimated, $\hat{\nu}$ is reset to be $\hat{\nu} = mean(\hat{\nu}_j)$. In practice, one can even just plug in a user-defined value ν_0 which gives the desired amount of smoothing. In such a case, the first step sets $\hat{\nu} = max(\nu_0, k+5)$. This $\hat{\nu}$ is used to estimate $\mathbf{\Lambda}$. After $\mathbf{\Lambda}$ is estimated, $\hat{\nu}$ can be reset to the user-defined value ν_0.

Our estimate of $\mathbf{\Lambda}$ comes after the first step in the estimation of ν. Section 5.6 shows that under our model $\mathbf{S}_1$ follows the generalized type-II beta distribution, with the expectation $(\nu-k)^{-1}\nu\mathbf{\Lambda}$. By the weak law of large numbers, $\overline{\mathbf{S}}_1$ converges in probability to $(\nu-k)^{-1}\nu\mathbf{\Lambda}$. We thus estimate $\mathbf{\Lambda}$ by $\hat{\nu}^{-1}(\hat{\nu}-k)\overline{\mathbf{S}}_1$. If $\hat{\nu}\to\infty$, then $\mathbf{\Lambda}$ is estimated by $\overline{\mathbf{S}}_1$. It is shown in our simulation study that this estimate of $\mathbf{\Lambda}$ is close to its true value. The above estimates work well on real data. A theoretical analysis of the estimation of our hyperparameters will be given in a later paper. For the moment we content ourselves with obtaining reasonable estimates.

6.2 EB estimation of η

The hyperparameter η is related to the moderated $\tilde{\mathbf{t}}$ of nonconstant genes. The method we use to estimate η builds on that of estimating v_0 in Smyth (2004), except that we now need to deal with the multivariate case. Let $\tilde{t}_j$ be the jth element of $\tilde{\mathbf{t}}$, $j = 1,..,k-1$. i.e., $\tilde{t}_j = n^{1/2}\tilde{s}_j^{-1}\mathbf{T}_{1j}\overline{\mathbf{X}}$ where $s_j^2 = (n-1)^{-1}\sum_{i=1}^n(\mathbf{T}_{1j}\mathbf{X}_i - \mathbf{T}_{1j}\overline{\mathbf{X}})^2$, $\tilde{s}_j^2 = (n-1+\nu)^{-1}((n-1)s_j^2 + \nu\lambda_j^2)$ and λ_j^2 is the jth diagonal element of $\mathbf{\Lambda}$. As in section 6.3 in Smyth (2004), each $\tilde{t}_j$ gives an estimate of η, call it $\hat{\eta}_j$, based on the top $p/2$ portion of genes with the largest $|\tilde{t}_j|$. We set $\hat{\eta}$ to be the mean of $\hat{\eta}_j$, $j = 1,...,k-1$.

7. Special cases

This section derives expressions for the posterior odds under additional constraints on the distributions of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$.

7.1 $\mathbf{\Sigma} = \sigma^2\mathbf{I}_k$.

The case $\mathbf{\Sigma} = \sigma^2\mathbf{I}_k$ corresponds to the fact that the time samples are independent across different time points, and the variances at different times are equal. This special case is of interest because it corresponds to the standard microarray analysis problem of comparing k different independent groups (*e.g.* k treatments). It is clear that if $\mathbf{\Sigma} = \sigma^2\mathbf{I}_k$, then $\mathbf{P\Sigma} = \mathbf{\Sigma P}$ and $\widetilde{\mathbf{\Sigma}} = \sigma^2\mathbf{I}_k$. Suppose that the prior for σ^2 is

$$\sigma^2 \sim inv\text{-}gamma\left(\frac{1}{2}\nu, \frac{1}{2}\nu\lambda^2\right).$$

Define $s_j^2 = (n-1)^{-1}\sum_{i=1}^n(\mathbf{T}_{1j}\mathbf{X}_i - \mathbf{T}_{1j}\overline{\mathbf{X}})^2$, $\tilde{s}_j^2 = (n-1+\nu)^{-1}((n-1)s_j^2 + \nu\lambda^2)$, and $\tilde{t}_j = n^{1/2}\mathbf{T}_{1j}\overline{\mathbf{X}}\tilde{s}_j^{-1}$, $j = 1,..,k-1$. In this case, the posterior odds are equivalent to a

product of $k-1$ independent univariate odds:

$$\mathbf{O} = \frac{p}{1-p}\left(\frac{\eta}{n+\eta}\right)^{\frac{1}{2}(k-1)} \prod_{j=1}^{k-1}\left(\frac{n-1+\nu+\tilde{t_j}^2}{n-1+\nu+\left(\frac{\eta}{n+\eta}\right)\tilde{t_j}^2}\right)^{\frac{1}{2}(n+\nu)}, \tag{14}$$

and the MB-statistic is equivalent to the sum of $k-1$ univariate B-statistics.

7.2 $\boldsymbol{\Sigma} = \sigma_1^2\mathbf{P} + \sigma_2^2(\mathbf{I}_k - \mathbf{P})$.

This slightly more general case corresponds to an assumption that the distribution of the $\mathbf{X}_i$ is exchangeable across times. It should and does lead to the same expression as we obtained in section 7.1. Set the priors for σ_1^2 and σ_2^2 as

$$\sigma_1^2 \sim \textit{inv-gamma}\left(\frac{1}{2}\xi, \frac{1}{2}\xi\lambda^2\right),$$
$$\sigma_2^2 \sim \textit{inv-gamma}\left(\frac{l}{2}l, \frac{1}{2}lm^2\right),$$

$\xi > 0$ and $l > 0$. Define $\tilde{s}_j^2 = (n-1+l)^{-1}((n-1)s_j^2 + lm^2)$, where s_j^2 is the same as that defined in section 7.1. Denote the individual univariate moderated t-statistics by $\tilde{t}_j = n^{1/2}\mathbf{T}_{1j}\overline{\mathbf{X}}\tilde{s}_j^{-1}$, $j = 1, .., k-1$. Using the fact that $\mathbf{TPT}'$ is a $k \times k$ $\mathbf{0}$ matrix with the $(1,1)$-element replaced with a 1, $\widetilde{\boldsymbol{\Sigma}}$ is found to be a diagonal matrix with σ_1^2 as the first diagonal element, and σ_2^2 in the remaining $k-1$ diagonal entries. The posterior odds become

$$\mathbf{O} = \frac{p}{1-p}\left(\frac{\eta}{n+\eta}\right)^{\frac{1}{2}(k-1)} \prod_{j=1}^{k-1}\left(\frac{n-1+l+\tilde{t_j}^2}{n-1+l+\left(\frac{\eta}{n+\eta}\right)\tilde{t_j}^2}\right)^{\frac{1}{2}(n+l)}, \tag{15}$$

essentially the same as equation (14).

7.3 $\boldsymbol{\mu} = \mathbf{0}$

A case of special interest is $H : \boldsymbol{\mu} = \mathbf{0}$, i.e., the gene stays at 0 over time. This is a sub hypothesis of $H : \boldsymbol{\mu} = \mu_0\mathbf{1}$. In this case, the assumption $\mathbf{P}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\mathbf{P}$ can be dropped. The priors are

$$\boldsymbol{\Sigma} \sim \textit{Inv-Wishart}_\nu((\nu\boldsymbol{\Lambda})^{-1}),$$
$$\boldsymbol{\mu}|\boldsymbol{\Sigma}, I = 1 \sim N_k(\mathbf{0}, \eta^{-1}\boldsymbol{\Sigma}),$$
$$\boldsymbol{\mu}|\boldsymbol{\Sigma}, I = 0 \sim N_k(\mathbf{0}, \mathbf{0}),$$

and it is straightforward to show that the posterior odds $\mathbf{O}$ become

$$\mathbf{O} = \frac{p}{1-p}\left(\frac{\eta}{n+\eta}\right)^{\frac{1}{2}k}\left(\frac{n-1+\nu+\widetilde{T}^2}{n-1+\nu+\left(\frac{\eta}{n+\eta}\right)\widetilde{T}^2}\right)^{\frac{1}{2}(n+\nu)}. \tag{16}$$

In this case, the multivariate moderated t-statistic is $\tilde{\mathbf{t}} = n^{\frac{1}{2}}\widetilde{\mathbf{S}}^{-\frac{1}{2}}\overline{\mathbf{X}}$, where $\widetilde{\mathbf{S}} = (n-1+\nu)^{-1}((n-1)\mathbf{S}+\nu\mathbf{\Lambda})$, and $\widetilde{T}^2 = \tilde{\mathbf{t}}'\tilde{\mathbf{t}}$. Under H, $k^{-1}\widetilde{T}^2$ has an F distribution with degrees of freedom $(k, n+\nu-k)$. The $\widetilde{T}^2$ statistic is identical to the one-sample moderated Hotelling T^2 statistic in section 4. It is easy to show that, here, as in section 5.6, $\mathbf{S}$ is independent of I and is distributed as a generalized type II beta distribution, with dimension parameter k, degrees of freedom $(n-1)/2$ and $\nu/2$, and location and scale matrices $(n-1)^{-1}\nu\mathbf{\Lambda}$ and $\mathbf{0}$, respectively, namely

$$P(\mathbf{S}|I=1) = P(\mathbf{S}|I=0) = \frac{1}{\boldsymbol{\beta}_k(\frac{n-1}{2}, \frac{\nu}{2})} \frac{|\mathbf{S}|^{\frac{1}{2}(n-k-2)}}{|\frac{\nu\mathbf{\Lambda}}{n-1}|^{\frac{1}{2}(n-1)}|\mathbf{I}_k + \left(\frac{\nu\mathbf{\Lambda}}{n-1}\right)^{-1}\mathbf{S}|^{\frac{1}{2}(n+\nu-1)}}.$$

To calculate $\widetilde{\mathbf{S}}$, we need to estimate ν and $\mathbf{\Lambda}$. This is similar to section 6.1 except that we are dealing with a $k \times k$ matrix. The parameter ν is first estimated based on the transformed data $\mathbf{TX}_1$,...,$\mathbf{TX}_n$, and then we apply the same procedure as in section 6.1; putting ν as $\hat{\nu} = max(mean(\hat{\nu}_j), k+6), j = 1, ..., k$. This estimated $\hat{\nu}$ is used to estimate $\mathbf{\Lambda}$. Once $\mathbf{\Lambda}$ is estimated, $\hat{\nu}$ is reset to be $\hat{\nu} = mean(\hat{\nu}_j)$. As in section 6.1, we can even just plug in a user-defined value ν_0 which gives the desired amount of smoothing. In such case, the first step sets $\hat{\nu} = max(\nu_0, k+6)$. This $\hat{\nu}$ is used to estimate $\mathbf{\Lambda}$. After $\mathbf{\Lambda}$ is estimated, $\hat{\nu}$ is reset to the user-defined value ν_0. The matrix $\mathbf{\Lambda}$ is estimated by $\hat{\nu}^{-1}(\hat{\nu}-k-1)\bar{\mathbf{S}}$. If $\hat{\nu} = \infty$, $\mathbf{\Lambda}$ is estimated by $\bar{\mathbf{S}}$. We also use this procedure described above to estimate the $\widetilde{\mathbf{S}}$ for the one-sample moderated Hotelling T^2 statistic in section 4.

When $k = 1$, the above expression reduces to the univariate posterior odds in Lönnstedt and Speed (2002) and Smyth (2004).

7.4 $n = 1$.

When $n = 1$, that is, when there is no replication at all, each gene has its own unkown variability. The multivariate moderated t-statistic becomes $\tilde{\mathbf{t}} = \mathbf{\Lambda}^{-1/2}\mathbf{T}_1\overline{\mathbf{X}}$. The posterior odds are obtained by plugging in $n = 1$ in the equation (12), and are found to be a function of $\mathbf{T}_1\overline{\mathbf{X}}$ only. Since there is no replication, our hyperparameters must be assigned values, for example from previous experiments.

7.5 $k = 2$.

When $k = 2$, i.e. when there are only two time points, the alternative hypothesis states that there is change between these two time points. Our multivariate model should and does reduce to the univariate model in Lönnstedt and Speed (2002) and Smyth

(2004). The posterior odds become

$$\mathbf{O} = \frac{p}{1-p}\left(\frac{\eta}{n+\eta}\right)^{\frac{1}{2}}\left(\frac{n-1+\nu+\tilde{t}^2}{n-1+\nu+\left(\frac{\eta}{n+\eta}\right)\tilde{t}^2}\right)^{\frac{1}{2}(n+\nu)}. \tag{17}$$

8. Limiting cases

We now list a number of limiting cases, for most of which the answer is either known or expected.

8.1 $\nu \to \infty$.

In this case, the smoothed multivariate variance-covariance matrix $\widetilde{\mathbf{S}}_1$ reduces to $\mathbf{\Lambda}$. The multivariate t-statistic is thus $\tilde{\mathbf{t}}_\infty = n^{1/2}\mathbf{\Lambda}^{-1/2}\mathbf{T}_1\overline{\mathbf{X}}$, and $\widetilde{T}^2_\infty = \tilde{\mathbf{t}}'_\infty\tilde{\mathbf{t}}_\infty$, and the posterior odds become

$$\mathbf{O} = \frac{p}{1-p}\left(\frac{\eta}{n+\eta}\right)^{\frac{1}{2}(k-1)} exp\left(\frac{1}{2}\left(\frac{n}{n+\eta}\right)\widetilde{T}^2_\infty\right).$$

8.2 $\nu \to \infty$ and $\mathbf{\Sigma} = \sigma^2\mathbf{I}_k$.

As in section 7.1, define $\tilde{t}_{\infty j} = n^{1/2}\lambda^{-1}\mathbf{T}_{1j}\overline{\mathbf{X}}$, $j = 1, ..., k-1$. The posterior odds become

$$\mathbf{O} = \frac{p}{1-p}\left(\frac{\eta}{n+\eta}\right)^{\frac{1}{2}(k-1)} exp\left(\frac{1}{2}\left(\frac{n}{n+\eta}\right)\sum_{j=1}^{k-1}\tilde{t}^2_{\infty j}\right).$$

8.3 $\nu \to 0$.

In this case, the multivariate moderated t-statistic reduces to the ordinary unmoderated multivariate t-statistic, $\mathbf{t} = n^{\frac{1}{2}}\mathbf{S}_1^{-\frac{1}{2}}\mathbf{T}_1\overline{\mathbf{X}}$. Let $T^2 = \mathbf{t}'\mathbf{t}$. The posterior odds become

$$\mathbf{O} = \frac{p}{1-p}\left(\frac{\eta}{n+\eta}\right)^{\frac{1}{2}(k-1)} \times \left(\frac{n-1+T^2}{n-1+\left(\frac{\eta}{n+\eta}\right)T^2}\right)^{\frac{n}{2}}.$$

Here $\mathbf{S}_1^{-\frac{1}{2}}$ can be obtained by using a g-inverse.

8.4 $\nu \to 0$ **and** $\mathbf{\Sigma} = \sigma^2\mathbf{I}_k$**.**

In this case, the univariate moderated t-statistic in section 7.1 reduces to the unmoderated t-statistic, $\tilde{t}_j = t_j = n^{1/2}\mathbf{T}_{1j}\overline{\mathbf{X}}s_j^{-1}$, $j = 1, ..., k-1$. The posterior odds become

$$\mathbf{O} = \frac{p}{1-p}\left(\frac{\eta}{n+\eta}\right)^{\frac{1}{2}(k-1)}\prod_{j=1}^{k-1}\left(\frac{n-1+t_j^2}{n-1+\left(\frac{\eta}{n+\eta}\right)t_j^2}\right)^{\frac{1}{2}n}.$$

8.5 $\nu \to \infty$ and $\boldsymbol{\mu} = \mathbf{0}$.

In the case of $\nu \to \infty$ and $\boldsymbol{\mu} = \mathbf{0}$, the multivariate moderated t-statistic in section 7.3 becomes $\tilde{\mathbf{t}}_\infty = n^{\frac{1}{2}} \boldsymbol{\Lambda}^{-\frac{1}{2}} \overline{\mathbf{X}}$, and $\widetilde{T}^2_\infty = \tilde{\mathbf{t}}'_\infty \tilde{\mathbf{t}}_\infty$. The posterior odds become

$$\mathbf{O} = \frac{p}{1-p} \left(\frac{\eta}{n+\eta} \right)^{\frac{1}{2}k} exp \left(\frac{1}{2} \left(\frac{n}{n+\eta} \right) \widetilde{T}^2_\infty \right).$$

8.6 $\nu \to 0$ and $\boldsymbol{\mu} = \mathbf{0}$.

In this case, the multivariate unmoderated t-statistic is $\mathbf{t} = n^{\frac{1}{2}} \mathbf{S}^{-\frac{1}{2}} \overline{\mathbf{X}}$, and $T^2 = \mathbf{t}'\mathbf{t}$. The posterior odds become

$$\mathbf{O} = \frac{p}{1-p} \left(\frac{\eta}{n+\eta} \right)^{\frac{1}{2}k} \left(\frac{n-1+T^2}{n-1+\left(\frac{\eta}{n+\eta} \right) T^2} \right)^{\frac{1}{2}n}.$$

9. Simulation Study

9.1 Method

In this section we report on a small simulation study based on an actual example we have met. We simulate 100 data sets, each with $20,000$ genes. The genes are simulated independently, which we regard as an assumption that makes sense to compare methods, but it should be kept in mind that gene expression measures in real data can be quite dependent. In each simulated data set, 400 out of the $20,000$ genes are assigned to be nonconstant. i.e., $p = 0.02$. Each gene is simulated with three independent replicates ($n = 3$) and eight time points ($k = 8$). The other hyperparameters are: $\nu = 13$, $\xi = 3$, $\lambda^2 = 0.3$, $\theta = 0$ (two-color experiments), $\kappa = 0.02$, $\eta = 0.08$, and

$$\boldsymbol{\Lambda} = \begin{pmatrix} 14.69 & 0.57 & 0.99 & 0.40 & 0.55 & 0.51 & -0.23 \\ 0.57 & 15.36 & 1.22 & 0.84 & 1.19 & 0.91 & 0.86 \\ 0.99 & 1.22 & 14.41 & 2.47 & 1.81 & 1.51 & 1.07 \\ 0.40 & 0.84 & 2.47 & 17.05 & 2.40 & 2.32 & 1.33 \\ 0.55 & 1.19 & 1.81 & 2.40 & 15.63 & 3.31 & 2.75 \\ 0.51 & 0.91 & 1.51 & 2.32 & 3.31 & 13.38 & 3.15 \\ -0.23 & 0.86 & 1.07 & 1.33 & 2.75 & 3.15 & 12.90 \end{pmatrix} \times 10^{-3}.$$

The correlation matrix of $\boldsymbol{\Lambda}$ is

$$\begin{pmatrix} 1 & 0.04 & 0.07 & 0.03 & 0.04 & 0.04 & -0.02 \\ 0.04 & 1 & 0.08 & 0.05 & 0.08 & 0.06 & 0.06 \\ 0.07 & 0.08 & 1 & 0.16 & 0.12 & 0.11 & 0.08 \\ 0.03 & 0.05 & 0.16 & 1 & 0.15 & 0.15 & 0.09 \\ 0.04 & 0.08 & 0.12 & 0.15 & 1 & 0.23 & 0.20 \\ 0.04 & 0.06 & 0.11 & 0.15 & 0.23 & 1 & 0.24 \\ -0.02 & 0.06 & 0.08 & 0.09 & 0.19 & 0.24 & 1 \end{pmatrix},$$

and we see clear evidence of serial correlation. The statistics compared are the (1) MB-statistic, or equivalently, the $\widetilde{T}^2$ statistic; (2) MB-statistic in the special case $\mathbf{\Sigma} = \sigma^2\mathbf{I}_k$ (section 7.1); (3) MB-statistic in the limiting case $\nu \to \infty$ (section 8.1); (4) MB-statistic in the limiting case $\nu \to 0$ (section 8.3); (5) ordinary F-statistic from a two-way anova with time and replicate effects; (6) partly-moderated F-statistic proposed in Smyth (2004) from a two-way anova model with time and replicate effects; (7) one-sample moderated Hotelling T^2 statistic $||n^{1/2}\widetilde{\mathbf{S}}^{-1/2}\tilde{\mathbf{d}}||^2$ derived in section 4, or equivalently, the moderated LR-statistic, where the degree of moderation and the common matrix toward which each sample covariance matrix moves is estimated by using the same method estimating ν and $\mathbf{\Lambda}$ in section 6.1, see section 7.3 for a discussion on estimating $\widetilde{\mathbf{S}}$. (8) The variance across time course replicates $(nk-1)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{k}(X_{ij}-\bar{\bar{x}})^2$. A comphrehensive comparison among all the published methods using both simulations and real data will be given in a later study. Here each of the eight statistics incorporates either none (e.g. variance) or one (ordinary F-statistic) or more of the followings: moderation, correlation structure, and replicate variance and thus can be used to show the importance of the above properties. It is not appropriate to set the prior degrees of freedom ν to be a very small number, since we have the constraint that $\nu \geq k-1$. We choose ν to be $k+5=13$ because it simulates more stable $\mathbf{\Sigma}$s across genes.

9.2 Results

Figure 1 displays examples of simulated nonconstant genes as in a-c, and constant genes as in d-f. The expected time course $\boldsymbol{\mu}$ of simulated constant genes may center around $\mathbf{0}$ as in e, or away from but parallel with $\mathbf{0}$, as in d and f.

Table 1 compares the means and standard deviations of the hyperparameter estimates of the diagonal elements of $\mathbf{\Lambda}$ (λ_j^2), $j=1,...,k-1$ with their true values. The mean estimate of $\mathbf{\Lambda}$ is very close to the true $\mathbf{\Lambda}$, and the standard deviations are very

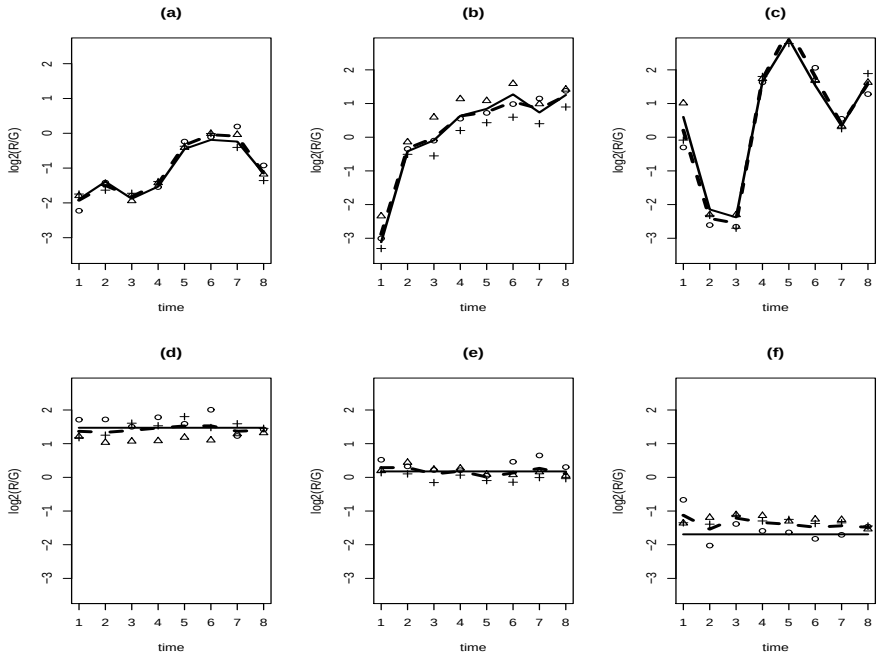

Figure 1: Plots of simulated nonconstant a-c and constant d-f genes. The solid line is the expected time course vector $\boldsymbol{\mu}$, and the dashed line is the sample average time course vector $\overline{\mathbf{X}}$. Points of the same shape are time point samples for the same replicate.

small. The hyperparameter η is always under-estimated (mean=0.026, SD=0.002), which agrees with section 8 in Smyth (2004), where v_0 was usually over-estimated. The hyperparameter ν is also always under-estimated (mean=7.024, SD=0.19), which is fine because we do not want to over smooth the gene specific sum of squares matrix $\mathbf{S}_1$.

To examine the relationship between the $\widetilde{T}^2$ statistic and the true deviation from constancy, the log_{10} transformed $\widetilde{T^2}$ statistic from one simulated dataset is plotted against the Mahalanobis distance between the expected time course vector $\boldsymbol{\mu}$ and its projection onto the rank 1 constant space $\bar{\boldsymbol{\mu}} = \mathbf{P}\boldsymbol{\mu}$ (Figure 2). The squared Mahalanobis distance is defined by $d(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})^2 = (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}})$. Figure 2 clearly shows that the $log_{10}\widetilde{T}^2$ are positively correlated with $d(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})$, and most of the 400 true nonconstant genes achieve higher $\widetilde{T}^2$ statistics than the constant genes.

Figures 3 and 4 plot the average numbers of false positives against average numbers of false negatives at different cutoffs. The lines in Figure 3 from left to right represent

Hyperparameters	True Value$\times 10^3$	Mean$\times 10^3$	SD$\times 10^3$
λ_1^2	14.69	14.71	0.16
λ_2^2	15.36	15.37	0.17
λ_3^2	14.41	14.43	0.15
λ_4^2	17.05	17.04	0.19
λ_5^2	15.63	15.63	0.15
λ_6^2	13.38	13.40	0.15
λ_7^2	12.90	12.92	0.17

Table 1: The means and standard deviations (SD) of diagonal elements of estimated $\mathbf{\Lambda}$.

the: MB-statistic ($\widetilde{T}^2$), one-sample moderated Hotelling T^2 statistic (indistinguishable from the MB-statistic), MB-statistic with $\mathbf{\Sigma} = \sigma^2\mathbf{I}_k$ (section 7.1), MB-statistic with $\nu \to \infty$ (section 8.1), partly-moderated F-statistic (Smyth 2004), ordinary F-statistic, MB-statistic with $\nu \to 0$ (section 8.3), and variance. The MB-statistic and the $\widetilde{T}^2$ statistic attain almost the same number of false positives and number of false negatives as the one-sample moderated Hotelling T^2 statistic (Figures 3 and 4). The importance of moderation is highlighted by comparing the lines of the MB-statistic, the MB-statistic in the limiting case $\nu \to \infty$ of section 8.1, and the MB-statistic in the limiting case that $\nu \to 0$ of section 8.3. Both of these limiting cases achieve higher aggregate false positives and false negatives (Figure 3). This result supports the view stated in section 2 that moderation is important. In particular, the case $\nu \to 0$ (no moderation at all) produces much higher numbers of false positives and false negatives. This is likely due to the poor estimation of sample variance-covariance matrices with a small number of replicates. Indeed, the ordinary unmoderated F-statistic which ignores the correlation structure achieves smaller numbers of false positives and false negatives than the unmoderated MB-statistic. A similar situation also arises in the microarray discrimination context, see section 7 of Dudoit et al. (2002). The partly-moderated F-statistic (Smyth 2004) which ignores the dependency among times behaves like the MB-statistic with the special case $\mathbf{\Sigma} = \sigma^2\mathbf{I}_k$ in section 7.1 (Figures 3 and 4). Moreover, it achieves fewer false positives and false negatives than the ordinary F-statistic (Figure 3). Figure 3 also demonstrates the importance of incorporating the correlation structure among time points. The MB-statistic, $\widetilde{T}^2$, and the one-sample moderated

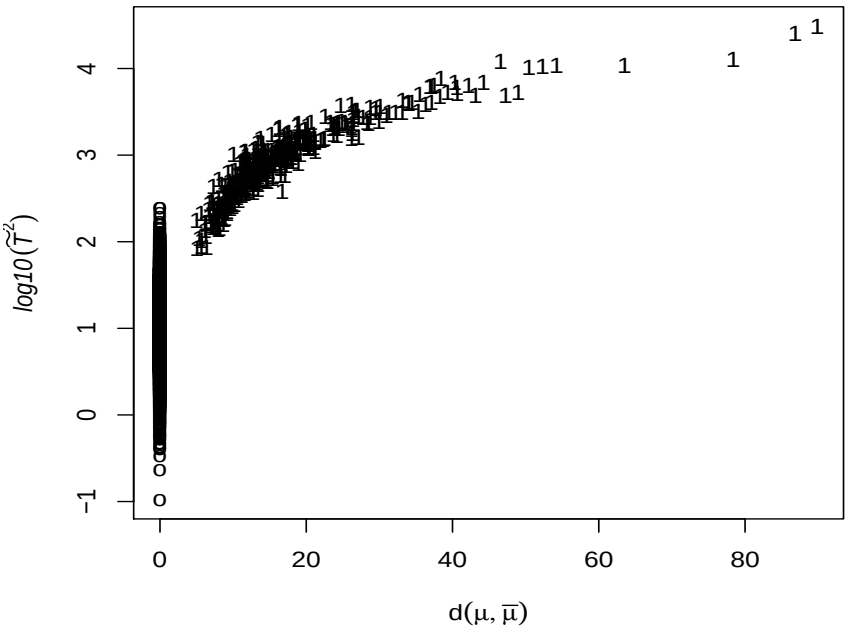

Figure 2: The $log_{10}\widetilde{T}^2$ statistic versus the true deviation from constancy $d(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})$ for one simulated data set. Here 1 denotes nonconstant, and o constant genes.

Hotelling T^2 statistic perform better than the partly-moderated F-statistic in Smyth (2004) and the ordinary F-statistic; the former incorporate the correlation structure among time points, whereas the latter do not. However, we observe that the amount of moderation given by the partly-moderated F-statistic in Smyth (2004) is usually much less than that given by the MB-statistic. When there are a large number of residual degrees of freedom from the linear model , the partly-moderated F-statistic (Smyth 2004) behaves very much like the ordinary F-statistic. This suggests that the lower number of false positives and number of false negatives from the MB-statistic than the partly-moderated F-statistic (Smyth 2004) involve both the incorporation of correlation structures and the amounts of moderation. The incorporation of correlation structure into the analysis probably has more impact on the results when there are true biological correlations, see section 10. As expected, the simple variance statistic across replicates which totally ignores the replicate variances perform the worst. This demonstrates the importance of incorporating the replicate variances into any statistic.

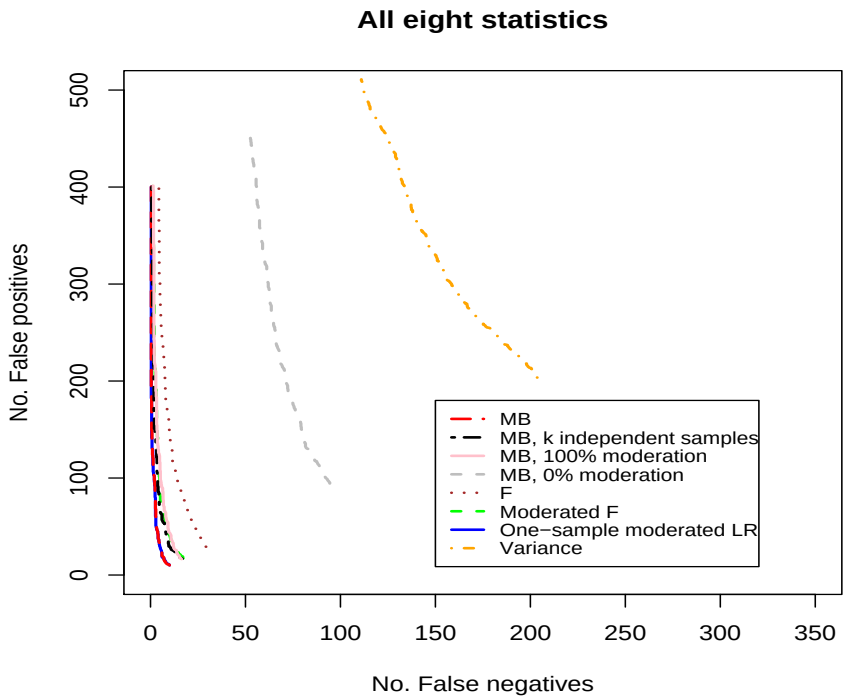

Figure 3: Number of false positives versus number of false negatives of all the eight statistics.

10. Discussion

In this paper we have proposed the multivariate empirical Bayes log posterior odds (MB-statistics) for replicated microarray time course experiments to rank genes for evidence of differential expression over time. We have shown in the simulation study that the MB-statistic, the $\widetilde{T}^2$ statistic, and the one-sample moderated Hotelling T^2 statistic perform best among all the eight statistics compared. This is not entirely surprising given that we simulated data under our model, but the comparisons are still informative. In practice, we consider the MB-statistic (or the $\widetilde{T}^2$ statistic) performs as well as the moderated LR-statistic (or the one-sample moderated Hotelling T^2 statistic), and one of the values of our multivariate empirical Bayes framework is that it provides a natural way to estimate the one-sample moderated Hotelling T^2 statistic (section 7.3), while the likelihood-based approach alone does not provide such an estimate.

A question which naturally arises is when we should suppose $\boldsymbol{\Sigma}$ general rather than

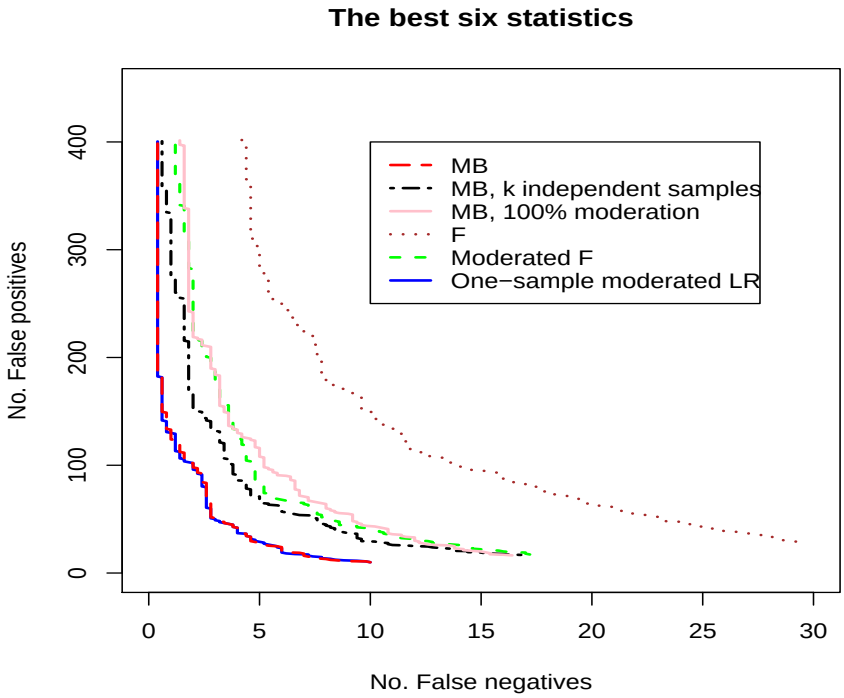

Figure 4: A closer look of Figure 3. Number of false positives versus number of false negatives of the best six statistics.

$\mathbf{\Sigma} = diag(\sigma_1^2, ..., \sigma_k^2)$ or $\mathbf{\Sigma} = \sigma^2 \mathbf{I}_k$ as in section 7.1. We have encountered a number of microarray time series experiments where there are no biological reasons associated with the mRNA samples why there should be any correlation between gene expression levels at different times, but nevertheless, the estimated $\mathbf{\Lambda}$ is clearly far from diagonal. It is not hard to think of non-biological reasons why temporal correlations may exist, all of the "common cause" type: the samples from different points in the time course may have been processed at the same or similar real times, in the same way, using similarly treated whole organisms, features which leave their imprint on the resulting gene expression data no matter how well we normalize before analysis. Perhaps the reader may see such causes of association as "fixed" rather than "random", and object to their being incorporated as correlations. We tend to agree, but feel that by allowing a general $\mathbf{\Lambda}$ and hence $\mathbf{\Sigma}$ (aparting from our commuting constraints), we permit real biological as well as technical temporal associations to be incorporated into the analysis. Whether there is a loss of power when a general $\mathbf{\Sigma}$ is used rather than, say, a

diagonal $\mathbf{\Sigma}$, is a question we plan to address in the near future. Another question we plan to investigate in the future is the effect of assuming the same $\mathbf{\Sigma}$ over all genes, rather than assigning different covariances for the cases $I = 1$ and $I = 0$. Or more generally, the effect of assuming the same $\mathbf{\Sigma}$ for the case $(I = 1)$, rather than different $\mathbf{\Sigma}$s corresponding to different temporal profile classes of genes.

Among the other methods which have been proposed in the literature for this gene ranking/selection problem from replicated microarray time course data, the Pareto front analysis described in Hero and Fleury (2002) and Fleury et al. (2002) make many uses of contrasts and can be applied to both short and long time courses. The ANOVA model Park et al. (2003) suggested is essentially the same as an unmoderated F-test, which does not incorporate the possible dependency among times, and does not do any moderation as in Smyth (2004). However, based on our simulation results and practical experience, moderation appears to be necessary. Genes selected by the standard F-test without any moderation at all tend to have small amounts of changes over time and very tight replicates (i.e., the replicate variances are very small). The B-spline based approach suggested by Bar-Joseph et al. (2003) and Luan and Li (2004) seem to perform better on data with longer time courses, but may not be appropriate for shorter time series, which are very common among microarray time course experiments. The likelihood vector-based method proposed in Guo et al. (2003) and the hidden Markov model proposed in Yuan et al. (2003), take into account the dependency among times, and can be applied to both short and long time courses.

The assumption of $\mathbf{P\Sigma} = \mathbf{\Sigma P}$ guarantees the possibility of our mathematical calculations, and hence our closed-form formula for the MB-statistic. One question which naturally arises is the impact of such constraint on the rankings of genes. From the practical point of view, the impact of this constraint on gene rankings is very slight. The correlations between the rankings of the one-sample MB-statistic *with* the commuting assumption and the moderated Hotelling T^2 statistic *without* the constraint from the actual examples we have met are typically very high (over 0.99). The correlations between rankings from our simulated data are also over 0.99. It is shown in section 7.3 that if the null hypothesis is $H : \boldsymbol{\mu} = \mathbf{0}$ (i.e. the gene stays at 0 over time) with the alternative $K : \boldsymbol{\mu} \neq \mathbf{0}$, instead of the null and alternative discussed here, then the assumption that $\mathbf{P\Sigma} = \mathbf{\Sigma P}$ can be dropped. This special case is useful for the one-sample gene ranking problem from two-channel microarray experiments (e.g. cDNA microarrays), where relative temporal profiles are measured, or two-sample gene

ranking problem from both single- and two-channel microarray experiments. For the two-sample problem, see later this section for details.

Our title refers to repicated microarray time course data, but we have not been very explicit about the nature of the replicates. We could have genes spotted in replicate on the microarray slide, we could have so-called technical replicates, where mRNA from the same biological specimen is split, and hybridized to multiple arrays, or we could have biological replicates, where mRNA samples are taken from different organisms. Naturally, the variability between gene expression measurements across these different types of replicates will be different, and probably gene-specific. Which level of variability is relevant to the investigator seeking genes whose absolute or relative temporal profiles are non-constant is a choice for the analyst in a given context, and will naturally depend on what is available. For example, we know of one investigator who had just one set of microarray time course measurements, but these were carried out on slides for which a subset of the genes were spotted down in quadruplicate. It is nature in this case we want to make use of this partial replication, and one may to do so would be to use the approach in this paper, estimating a common $\mathbf{\Lambda}$ for all genes, using data from the subset of genes which are replicated. Similarly, we know of a different investigator who carried out four separate time-course experiments, using only three biologically distinct samples of cells, one being the basis of a pair of technical replicates. Here we tried and compared two strategies, one treating the data as four replicate series, and the other as just three replicates, using the average of the two technical replicates. Neither approach is entirely satisfactory, but even for a single time point, we have been unable to derive a moderated t- or B-statistic which incorporates multiple strata of variability. Finding a closed-form statistic seems to us to be an open and probably hard problem, essentially equivalent to wanting closed form expressions for the Bayesian analysis of normal variance components with conjugate priors, something presently carried out by MCMC.

As originally defined, the empirical Bayes model proposed in this paper applies only to the one-sample problem of detecting nonconstant genes from a single biological condition. However, with minimal changes, it also applies to the two-sample problem of comparing time course profiles. Suppose that $(\mathbf{Z}_1, \mathbf{Y}_1)$..., $(\mathbf{Z}_n, \mathbf{Y}_n)$ are *i.i.d.* random pairs corresponding to two different biological conditions, *e.g.* wildtype and mutant. Then $\mathbf{X}_1 = \mathbf{Z}_1 - \mathbf{Y}_1$,...,$\mathbf{X}_n = \mathbf{Z}_n - \mathbf{Y}_n$ are the *i.i.d.* difference time course vectors between these two conditions, and $\boldsymbol{\mu}_X = \boldsymbol{\mu}_Z - \boldsymbol{\mu}_Y$. The null hypothesis

$H : \boldsymbol{\mu}_X = \mu_0\mathbf{1}$ corresponds to the null hypothesis that the expected time profiles $\boldsymbol{\mu}_Z$ and $\boldsymbol{\mu}_Y$ of these two biological conditions have the same shape, and equation (13) can be applied directly to this problem. Moreover, $H : \boldsymbol{\mu}_X = \mathbf{0}$ further corresponds to the null hypothesis that $\boldsymbol{\mu}_Z$ and $\boldsymbol{\mu}_Y$ are identical, and it is straightforward to use the special case in section 7.3 here. The unpaired case can be dealt with in a similar manner. Suppose $\mathbf{Z}_1, ..., \mathbf{Z}_m$ and $\mathbf{Y}_1, ..., \mathbf{Y}_n$ are *i.i.d.* random vectors from $N(\boldsymbol{\mu}_Z, \boldsymbol{\Sigma})$ and $N(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma})$, respectively. The sample mean $\overline{\mathbf{Z}} \sim N(\boldsymbol{\mu}_Z, m^{-1}\boldsymbol{\Sigma})$ and $\overline{\mathbf{Y}} \sim N(\boldsymbol{\mu}_Y, n^{-1}\boldsymbol{\Sigma})$ permit us to test $H : \boldsymbol{\mu}_Z = \boldsymbol{\mu}_Y$ using a $\tilde{\mathbf{t}}$-statistic with numerator $(m^{-1} + n^{-1})^{-1/2}(\overline{\mathbf{Z}} - \overline{\mathbf{Y}})$ and denominator the square root of the smoothed pooled covariance matrix $\widetilde{\mathbf{S}} = (n + m - 2 + \nu)^{-1}((m-1)\mathbf{S}_Z + (n-1)\mathbf{S}_Y + \nu\boldsymbol{\Lambda})$.

The multivariate empirical Bayes model here can also be extended to more general multivariate empirical Bayes regression models allowing the comparisons among time course profiles from multiple biological conditions simultaneously, while taking into account the correlation structure and moderating the estimate of the variance-covariance matrix (Tai and Speed 2004, in preparation). The multivariate empirical Bayes regression models in this more general context differ from the univariate empirical Bayes linear models in Smyth (2004) in that Smyth (2004) considers each coefficient or contrast individually and independently, while in Tai and Speed (2004), all the contrasts are considered at the same time and correlated with each other. This extension allows comparisons of time course profiles among multiple biological conditions (eg. wildtype, mutant1, mutant2,...etc.) with different sample sizes by setting up appropriate contrasts. Moreover, by using appropriate contrasts, we can also deal with the detection of genes having specific patterns in the one-sample case, and genes having different specific patterns in the multi-sample case.

Acknowledgments

We thank Gordon Smyth and Ingrid Lönnstedt for reading our earlier drafts of this paper and their valuable comments. Special thanks go to Ben Bolstad for his assistance with our simulations. We also appreciate many constructive comments on our final manuscript from Darlene Goldstein, Avner Bar-Hen, Alfred Hero, Cavan Reilly, and Christina Kendziorski. Finally, we would like to acknowledge John Ngai, Karen Vranizan, Mary Wildermuth, Moriah Szpara, Carina Howell, and Jason Dugas for their helpful discussions on biological background of time course experiments and access to their data to test the methods developed in this paper. This work was supported by the NIH through grants HG00047 (YCT) and R01 Lmo7609-01 (TPS).

References

Aitchison, J. and Dunsmore, I. R. (1975). Statistical prediction analysis. *Cambridge University Press*.

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.

Bar-Joseph, Z., Gerber, G., Simon, I., Gifford, D. and Jaakkola, T. (2003). Comparing the continuous representation of time-series expression profiles to identify differentially expressed genes. *Proc. Natl. Acad. Sci. USA* **100** 10146-10151.

Bickel, P. J. and Doksum, K. A. (2001). Mathematical Statistics vol 1. *Prentice Hall*.

Broberg, P. (2003). Statistical methods for ranking differentially expressed genes. *Genome Biology,* **4** R41.

Cho, R. J., Campbell, M. J., Winzeler, E. A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T. G., Gabrielian, A. E., Landsman, D., Lockhart, D. J., Davis, R. W. (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. *Molecular Cell,* **2** 65-73.

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P. O., Herskowitz, I. (1998). The transcriptional program of sporulation in budding yeast. *Science,* **282** 699-705.

Dudoit, S., Yang, Y. H., Speed, T. P., Callow, M. J. (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statist. Sinica* **12** 111-139.

Dudoit, S., Fridlyand, J., Speed, T. P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *J. Amer. Statist. Assoc.* **97** 77-87.

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

Fleury, G., Hero, A., Yoshida, S., Carter, T., Barlow, C., Swaroop, A. (2002). Pareto analysis for gene filtering in microarray experiments. *Proc. XI European Signal Processing Conference, Toulouse France.*

Ge, Y., Dudoit, S., Speed, T. P.(2003). Resampling-based multiple testing for microarray data analysis. *Test* **12** 1-77.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). Bayesian data analysis. *Chapman & Hall/CRC.*

Guo, X., Qi, H., Verfaillie, C., Pan, W. (2003). Statistical significance analysis of longitudinal gene expression data. *Bioinformatics* **19** 1628-1635.

Gupta, A. K. and Nagar, D. K. (2000). Matrix variate distributions. *Chapman & Hall/CRC*

Hero, A. and Fleury, F. (2002). Posterior Pareto front analysis for gene filtering. *Proc. of Workshop on Genomic Signal Processing and Statistics (GENSIPS) Raleigh NC.*

Lönnstedt, I. and Speed, T. P. (2002). Replicated microarray data. *Statist. Sinica* **12** 31-46.

Luan, Y., and Li, H. (2004). Model-based methods for identifying periodically expressed genes based on time course microarray gene expression data. *Bioinformatics* **20** 332-339.

Mardia, K. V., Kent, J. T., Bibby, J. M. (2000). Multivariate Analysis. *Academic Press.*

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.

Park, T., Yi, S. G., Lee, S., Lee, S. Y., Yoo, D. H., Ahn, J. I., Lee, Y. S. (2003). Statistical tests for identifying differentially expressed genes in time-course microarray experiments. *Bioinformatics* **19** 694-703.

Reiner, A., Yekutieli, D., Benjamini, Y. (2003). Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19** 368-375.

Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* **3** no.1 article 3.

Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D., Futcher, B. Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Molecular Biology of the Cell,* **9**:3273-3297.

Storch, K. F., Lipan, O., Leykin, I., Viswanathan, N., Davis, F. C., Wong, W. H., Weitz, C.J. (2002). Extensive and divergent circadian gene expression in liver and heart. *Nature.* **417** (6884):78-83. Epub 2002 Apr 21. Erratum in: Nature 2002 Aug 8;418(6898):665.

Tai, Y. C. and Speed, T. P. (2004). General multivariate empirical Bayes regression models for replicated microarray time course data. *Manuscript in preparation.*

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E. S., Golub, T. R. (1999). Interpreting patterns of gene ex-

pression with self-organizing maps: Methods and application to hematopoietic differentiation *Proc. Natl. Acad. Sci. USA* **96**: 2907-2912.

Tusher, V. G. Tibshirani, R., and Chu, G. (2001).Significant analysis of microarrays applied to the ionizing radition response. *Proc. Natl. Acad. Sci. USA* **98** 5116-5121.

Wen, X., Fuhrman, S., Michaels, G. S., Carr, D. B., Smith, S., Barker, J. L., Somogyi, R. (1998). Large-scale temporal gene expression mapping of central nervous system development *Proc. Natl. Acad. Sci. USA* **95**: 334-339.

Yuan, M., Kendziorski, C., Park, F., Porter, J. R., Hayes, K., and Bardfield, C.A. (2003). Hidden markov models for microarray time course data in multiple biological conditions. *Technical Report #178, Department of Biostatistics and Medical Informatics, University of Wisconsin.*

Yu Chuan Tai
Division of Biostatistics, University of California, Berkeley, 367 Evans Hall, Berkeley, CA 94720-3860, USA.
E-mail: yuchuan@stat.berkeley.edu

Terence P. Speed
Department of Statistics, University of California, Berkeley, 367 Evans Hall, Berkeley CA 94720-3860, USA.
E-mail: terry@stat.berkeley.edu

Division of Genetics and Bioinfomatics, the Walter and Eliza Hall Institute of Medical Research, Australia.
E-mail: terry@wehi.edu.au