

August 9, 2007

## 1 Web Appendix A

### Special and Limiting Cases for Longitudinal Posterior Odds

This section derives the posterior odds under some constraints on the prior distributions of the  $\boldsymbol{\mu}_d$ ,  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$ .

1.  $\nu \rightarrow \infty$ . The posterior odds do not involve  $\mathbf{W}$  anymore, and become

$$\mathbf{O} = \frac{p}{1-p} \left( \frac{n+\beta}{\beta} \right)^{\frac{k}{2}} \prod_{d=1}^D \left( \frac{\beta_d}{n_d+\beta_d} \right)^{\frac{k}{2}} \exp \{tr(\mathbf{B} + \mathbf{M}_H - \mathbf{M}_K)\}.$$

In contrast, when  $\nu \rightarrow 0$ , the posterior odds are just equation (5) with  $\nu$  replaced by 0. This means that gene-specific  $\mathbf{W}$ s are so different such that moderation is not able to move these gene-specific  $\mathbf{W}$  at all. Thus, the degree of moderation reflects how similar these gene-specific within group sums of squares are. The more similar they are, the greater degree of moderation our proposed statistic can perform.

2.  $n_d$  is the same across genes,  $\beta_d \rightarrow 0$ ,  $\beta \rightarrow 0$ .

In this case, when the condition specific sample sizes are the same across genes, the posterior odds are equivalent to the likelihood-based moderated Wilks' lambda statistic  $\widehat{\mathbf{LR}}$  (equation 1) in section 2.1. In addition, if  $\nu \rightarrow \infty$ , the posterior odds become  $\exp(tr\mathbf{B})$ . If  $\nu \rightarrow 0$ , the posterior odds are equivalent to classical likelihood-based Wilks' lambda.

3.  $D = 1$ ,  $\beta \rightarrow \infty$ ,  $\boldsymbol{\alpha} = \boldsymbol{\alpha}_1 = \mathbf{0}$ .

In this one-sample case, the between sums of squares  $\mathbf{B}$  vanishes and under the null, the expected time course stays at 0 and the posterior odds should and indeed do reduce to those in Tai and Speed (2006)

$$\mathbf{O} = \frac{p}{1-p} \left( \frac{\beta_1}{n_1+\beta_1} \right)^{\frac{k}{2}} \left( \frac{n_1-1+\nu+\tilde{T}^2}{n_1-1+\nu+\left(\frac{\beta_1}{n_1+\beta_1}\right)\tilde{T}^2} \right)^{\frac{1}{2}(n_1+\nu)},$$

where  $\tilde{T}^2 = \tilde{\mathbf{t}}' \tilde{\mathbf{t}}$ ,  $\tilde{\mathbf{t}} = n_1^{\frac{1}{2}} \tilde{\mathbf{S}}^{-\frac{1}{2}} \bar{\mathbf{X}}_1$  and  $\tilde{\mathbf{S}}$  is the moderated sample variance-covariance matrix

$$\tilde{\mathbf{S}} = \frac{(n_1 - 1)\mathbf{S} + \nu\mathbf{\Lambda}}{n_1 - 1 + \nu}$$

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

4.  $D = 2$ ,  $\beta \rightarrow 0$ ,  $\beta_d \rightarrow 0$ .

In this two-sample case, we define  $\tilde{\mathbf{S}}$  to be the smoothed pooled sample variance-covariance matrix, and  $\tilde{\mathbf{t}}$  being the two-sample moderated multivariate  $t$ -statistic

$$\tilde{\mathbf{S}} = (n_1 + n_2 - 2 + \nu)^{-1} \left\{ \sum_{d=1}^2 \sum_{i=1}^{n_d} (\mathbf{X}_{di} - \bar{\mathbf{X}}_d)(\mathbf{X}_{di} - \bar{\mathbf{X}}_d)' + \nu\mathbf{\Lambda} \right\},$$

$$\tilde{\mathbf{t}} = (n_1^{-1} + n_2^{-1})^{-\frac{1}{2}} \tilde{\mathbf{S}}^{-\frac{1}{2}} (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2).$$

It turns out that when  $n_1$  and  $n_2$  are identical across genes, the posterior odds are equivalent to  $\tilde{T}^2 = \tilde{\mathbf{t}}' \tilde{\mathbf{t}}$ ,

$$\mathbf{O} \propto \left(1 + \tilde{T}^2\right)^{\frac{1}{2}(n_1 + n_2 + \nu)},$$

which is the same as the independent two-sample moderated Hotelling  $T^2$  statistic  $\tilde{T}^2$  proposed in Tai and Speed (2006). Furthermore, if  $k = 1$ ,  $\tilde{\mathbf{t}}' \tilde{\mathbf{t}} = \tilde{t}^2$  is the square of the moderated univariate  $t$ -statistic defined in Smyth (2004).

5.  $k = 1$ .

In the case when there is only 1 time point and  $D$  biological conditions, the posterior odds become

$$\mathbf{O} = \frac{p}{1-p} \left( \frac{n + \beta}{\beta} \right)^{\frac{1}{2}} \prod_{d=1}^D \left( \frac{\beta_d}{n_d + \beta_d} \right)^{\frac{1}{2}} \left( \frac{TSS + m_H + \nu\lambda}{WSS + m_K + \nu\lambda} \right)^{\frac{1}{2}(n + \nu)},$$

where

$$TSS = \sum_{d=1}^D \sum_{i=1}^{n_d} (x_{di} - \bar{x})^2$$

$$WSS = \sum_{d=1}^D \sum_{i=1}^{n_d} (x_{di} - \bar{x}_d)^2$$

$$m_H = (n^{-1} + \beta^{-1})^{-1} (\bar{x} - \alpha)^2$$

$$m_K = \sum_{d=1}^D (n_d^{-1} + \beta_d^{-1})^{-1} (\bar{x}_d - \alpha_d)^2$$

are total and within sums of squares, and quantities involving (condition-specific) prior means, respectively. Thus, as Lönnstedt et al. (2005), we obtain the same form of fully moderated F-statistic in the univariate case.

## 2 Web Appendix B

### 2.1 Hyperparameter Estimation for the Longitudinal Model

This section describes the estimation of hyperparameters:  $\nu$ ,  $\mathbf{\Lambda}$ ,  $\alpha_d$ ,  $\alpha$ ,  $\beta_d$ ,  $\beta$ . We only seek simple estimates which produce reasonable results here. Better estimates, e.g. ones maximizing a marginal likelihood, will be considered in the future.

#### 2.1.1 EB estimation of $\nu$ , $\mathbf{\Lambda}$

The hyperparameters  $\nu$  and  $\mathbf{\Lambda}$  associated with the priors for  $\mathbf{\Sigma}$  are estimated based on within sums of squares and products  $\mathbf{W}$ , which is equivalent to the pooled sample variance-covariance matrix  $\mathbf{S}$ . We estimate  $\nu$  using the procedure described in Tai and Speed (2006). In summary, we have  $k$  sets of  $n$  scalar random variables from the  $j$ -th element of  $\mathbf{X}_{di}$ ,  $d = 1, \dots, D, i = 1, \dots, n_d$ . Our two-step strategy in Tai and Speed (2006) applies directly: (1) setting  $\hat{\nu}$  to be  $\max(\text{mean}(\hat{\nu}), k + 6)$ , and after  $\mathbf{\Lambda}$  is estimated, (2) setting  $\hat{\nu}$  back to be  $\text{mean}(\hat{\nu})$ . Here,  $\mathbf{\Lambda}$  is estimated using the prior for  $\mathbf{\Sigma}$  by the first moment of the inverse Wishart distribution, *i.e.*  $\hat{\mathbf{\Lambda}} = \hat{\nu}^{-1}(\hat{\nu} - k - 1)\bar{\mathbf{S}}$ , where  $\bar{\mathbf{S}}$  is the average pooled sample variance-covariance matrix across all  $G$  genes. If  $\nu \rightarrow \infty$ , then  $\hat{\mathbf{\Lambda}} = \bar{\mathbf{S}}$ .

#### 2.1.2 EB estimation of $\alpha_d$ , $\alpha$ , $\beta_d$ , $\beta$

The estimations of the hyperparameters associated with the means  $\boldsymbol{\mu}_d$  and  $\boldsymbol{\mu}$  are done using simple method of moments from the prior distributions. The hyperparameters associated with  $\boldsymbol{\mu}_d$  (*i.e.*  $\alpha_d$  and  $\beta_d$ ) are estimated using the top  $p \times 100\%$  genes according to some simple statistic roughly measuring the differential expression. On the other hand, those associated with  $\boldsymbol{\mu}$  (*i.e.*  $\alpha$  and  $\beta$ ) are estimated using the bottom  $(1 - p) \times 100\%$  genes according to the same simple statistic. Here, we propose the following statistic for ranking genes for this purpose only

$$\Delta = \sum_{i \neq j} (\bar{\mathbf{X}}_{d_i} - \bar{\mathbf{X}}_{d_j})' (\bar{\mathbf{X}}_{d_i} - \bar{\mathbf{X}}_{d_j}).$$

Thus, our estimates are

$$\hat{\boldsymbol{\alpha}}_d = \overline{\overline{\mathbf{X}}_d}, \quad \hat{\boldsymbol{\alpha}} = \overline{\overline{\mathbf{X}}},$$

$$\hat{\beta}_d = \frac{\text{tr}(\overline{\mathbf{S}})}{\text{tr}(\overline{\mathbf{U}}_d)}, \quad \hat{\beta} = \frac{\text{tr}(\overline{\mathbf{S}})}{\text{tr}(\overline{\mathbf{U}})},$$

where  $\text{tr}$  refers to the trace operator,  $\overline{\overline{\mathbf{X}}_d}$  and  $\overline{\overline{\mathbf{X}}}$  are average time course vectors for the  $d$ -th biological condition across the top  $p \times 100\%$  genes, and all the conditions across the bottom  $(1-p) \times 100\%$  genes based on  $\Delta$ , respectively. The matrices  $\mathbf{U}_d$  and  $\mathbf{U}$  are the estimated gene-specific variance-covariance matrices for the gene-specific means  $\boldsymbol{\mu}_d$  and  $\boldsymbol{\mu}$

$$\mathbf{U}_d = (\overline{\mathbf{X}}_d - \hat{\boldsymbol{\alpha}}_d)(\overline{\mathbf{X}}_d - \hat{\boldsymbol{\alpha}}_d)', \quad \mathbf{U} = (\overline{\mathbf{X}} - \hat{\boldsymbol{\alpha}})(\overline{\mathbf{X}} - \hat{\boldsymbol{\alpha}})',$$

and  $\overline{\overline{\mathbf{U}}_d}$  and  $\overline{\overline{\mathbf{U}}}$  are their averages again across the top  $p \times 100\%$  and the bottom  $(1-p) \times 100\%$  genes, respectively. The numerator of  $\hat{\beta}_d$  and  $\hat{\beta}$  ( $\overline{\mathbf{S}}$ ) are average pooled sample variance-covariance matrices across the top  $p \times 100\%$  and the bottom  $(1-p) \times 100\%$  genes, respectively. In the limiting case that  $\boldsymbol{\alpha}_d = \boldsymbol{\alpha} = 0$ ,  $\beta_d \rightarrow 0$ , and  $\beta \rightarrow 0$ , we set  $\hat{\beta}_d = \hat{\beta} = 0$ .

## 3 Web Appendix C

### 3.1 Hyperparameter Estimation for the Cross-sectional Model

This section describes the estimation of hyperparameters associated with the cross-sectional model.

#### 3.1.1 EB estimation of $\nu$ , $\lambda$

The hyperparameters  $\nu$  and  $\lambda$  are associated with the variance  $\sigma^2$ . As indicated in section 3.1.2, the prior for  $\sigma^2$  is inverse-gamma distribution with shape  $\nu/2$  and scale  $\nu\lambda^2/2$  parameters. Suppose  $a$  and  $b$  are the prior mean and variance of  $\sigma^2$ . Then

$$a = (\nu - 2)^{-1}\nu\lambda^2, \quad b = (\nu - 2)^{-2}(\nu - 4)2\nu^2\lambda^4.$$

And  $\nu$  and  $\lambda$  can be estimated by

$$\hat{\nu} = \frac{2\hat{a}^2}{\hat{b}} + 4, \quad \hat{\lambda} = \sqrt{\frac{(\hat{\nu}-2)\hat{a}}{\hat{\nu}}},$$

where  $\hat{a} = \bar{\sigma}^2$ ,  $\hat{b} = \widehat{Var}(\hat{\sigma}^2)$ , and  $\hat{\sigma}^2$  is the gene-specific pooled replicate variances across biological condition and time

$$\hat{\sigma}^2 = \frac{\sum_{d=1}^D \sum_{j=1}^{K_d} (n_{d,j} - 1) s_{d,j}^2}{\sum_{d=1}^D \sum_{j=1}^{K_d} (n_{d,j} - 1)}, \quad s_{d,j}^2 = \frac{\sum_{i=1}^{n_{d,j}} (y_{d,j,i} - \bar{y}_{d,j})^2}{(n_{d,j} - 1)} I(n_{d,j} > 1).$$

### 3.1.2 EB estimation of $\alpha_d$ , $\alpha$ , $\Omega_d$ , $\Omega$

### 3.1.3 Multi-sample case

The estimation of the hyperparameters associated with the prior means  $\alpha_d$ ,  $\alpha$ ,  $\Omega_d$ , and  $\Omega$  is done using the method of moments from the prior distributions. As in the longitudinal case, we use a simple statistic roughly measuring differential expression to rank genes for estimation purpose only. The simple statistic we propose is

$$\Delta = \sum_{i \neq j} (\hat{\theta}_{d_i} - \hat{\theta}_{d_j})' (\hat{\theta}_{d_i} - \hat{\theta}_{d_j}).$$

Thus, the estimates are

$$\begin{aligned} \hat{\alpha}_d &= \bar{\hat{\theta}}_d & \hat{\alpha} &= \bar{\hat{\theta}} \\ \hat{\omega}_{d,j} &= \frac{\bar{\hat{\sigma}}_d^2}{\bar{\mathbf{U}}_{d,j}} & \hat{\omega}_j &= \frac{\bar{\hat{\sigma}}^2}{\bar{\mathbf{U}}_j}, \end{aligned}$$

where  $\bar{\hat{\theta}}_d$  and  $\bar{\hat{\theta}}$  are average estimated regression coefficients for the  $d$ -th biological condition across the top  $p \times 100\%$  genes, and all the conditions across the bottom  $(1 - p) \times 100\%$  genes, respectively. The matrices  $\mathbf{U}_d$  and  $\mathbf{U}$  are the estimated gene-specific variance-covariance matrices for  $\theta_d$  and  $\theta$ , and  $\mathbf{U}_{d,j}$ ,  $\mathbf{U}_j$  are their corresponding  $(j, j)$ -th diagonal elements.

$$\mathbf{U}_d = (\hat{\theta}_d - \hat{\alpha}_d)(\hat{\theta}_d - \hat{\alpha}_d)'$$

$$\mathbf{U} = (\hat{\theta} - \hat{\alpha})(\hat{\theta} - \hat{\alpha})'$$

The estimates  $\hat{\omega}_{d,j}$  and  $\hat{\omega}_j$  are based on the top  $p \times 100\%$  genes, and all the conditions across the bottom  $(1 - p) \times 100\%$  genes, respectively. In the case that flat priors are used for  $\theta_d$  and  $\theta$  (*i.e.*,  $\omega_{d,j} \rightarrow 0$  and  $\omega_j \rightarrow 0$ ), we just set  $\hat{\omega}_{d,j} = \hat{\omega}_j = 0$

### 3.1.4 One-sample case

In the one-sample case, the subscript  $d$  above is dropped and hyperparameters  $\alpha$  and  $\Omega$  are estimated using the same method as what we have just described. The

univariate hyperparameters  $\alpha_0$  is estimated by  $\hat{\alpha}_0 = \bar{\hat{\beta}}_0$ , the average least squares estimate for  $\beta_0$  across all genes. The hyperparameter  $\omega$  is estimated by  $\hat{\omega} = c\bar{\sigma}^2$ , where  $c = \text{mean}(\hat{\beta}_0 - \hat{\alpha}_0)^{-2}$ . Again,  $\hat{\omega} = 0$  if  $\omega \rightarrow 0$ .

## 4 Web Table 1

Hyperparameter	True Value	Mean	SD
$\beta_1$	1	0.7	0.06
$\beta_2$	1	0.8	0.07
$\beta_3$	1	0.7	0.08
$\beta$	1	2.7	0.02
$\nu$	13	5.7	0.07
$\lambda_1$	0.292	0.34	0.004
$\lambda_2$	0.292	0.34	0.005
$\lambda_3$	0.172	0.20	0.003
$\lambda_4$	0.183	0.21	0.003
$\lambda_5$	0.273	0.32	0.005
$\lambda_6$	0.303	0.35	0.006
$\lambda_7$	0.202	0.23	0.003
$\lambda_8$	0.192	0.22	0.004

Table 1: Mean and SD for the estimated hyperparameters.  $\lambda$ s refer to the diagonal elements of  $\Lambda$

## 5 Web Figure 1

## 6 Web Figure 2

## References

- Lönnstedt, I., Rimini, R., and Nilsson, P. (2005). Empirical Bayes microarray ANOVA and grouping cell lines by equal expression levels. *Statistical Applications in Genetics and Molecular Biology*, 4(1):article 7.
- Lönnstedt, I. and Speed, T. P. (2002). Replicated microarray data. *Statistica Sinica*, 12:31–46.
- 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(1):article 3.

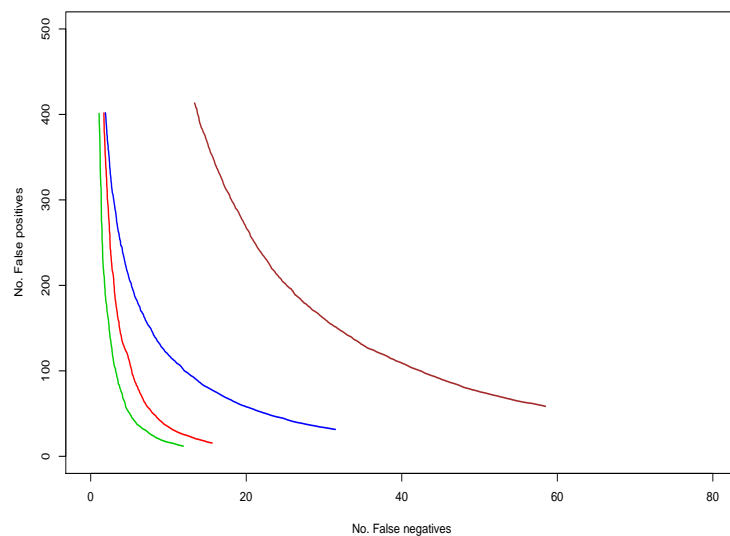


Figure 1: Average number of false positives versus number of false negatives of our fully moderated Wilks' lambda (red), and likelihood-ratio moderated Wilks' lambda (blue), our fully moderated Wilks' lambda with all true hyperparameters (green), and the moderated F-statistic (brown).

Tai, Y. C. and Speed, T. P. (2006). A multivariate empirical Bayes statistic for replicated microarray time course data . *Annals of Statistics*, 34(5):2387–2412.

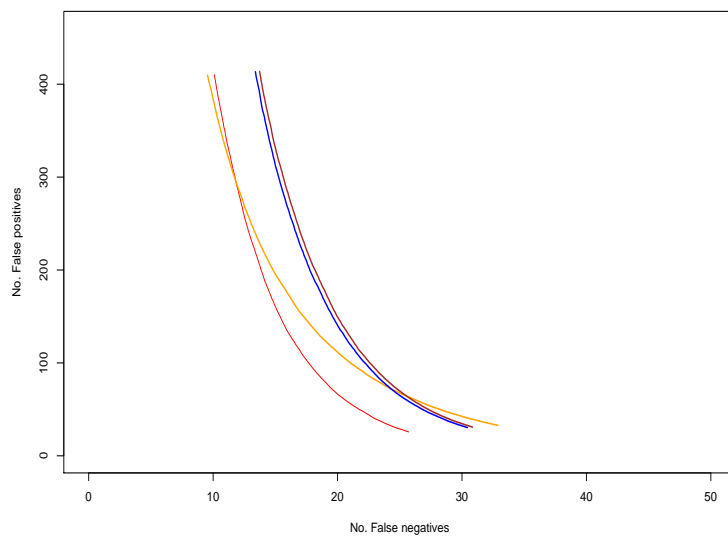


Figure 2: Average number of false positives versus number of false negatives of our  $MB$ -statistic with B-spline basis of 6 degrees of freedom (red),  $MB$ -statistic with B-spline basis of 3 degrees of freedom (orange), fully moderated likelihood-ratio statistic (blue), and the moderated F-statistic (brown).