Biometrics 000, 000–000 DOI: 000
000 0000

Web-based Supplementary Materials for Bayesian hierarchical modeling for subject-level response classification in peptide microarray immunoassays by Gregory Imholte and Raphael Gottardo

This paper has been submitted for consideration for publication in *Biometrics*

Web Appendix A

[Figure 1 about here.]

In the absence of significant plate replication within subjects, and because we cannot truly know which vaccine subjects are respondents against which peptides, it is difficult to empirically address whether slide and peptide effects are additive. Our RV144 placebo data can partially address the additivity of baseline peptide effects and plate effects. As discussed in the paper, twenty RV144 subjects received placebo and were measured on two separate slides. Under the paired pepBayes model, differences between peptide averages before and after placebo treatment (i.e. quantities $z_{ip} = \bar{y}_{ip1} - \bar{y}_{ip0}$ as in the paper) should have a symmetric distribution centered around the difference between the slides' respective plate effects, $(\mu_{i1} - \mu_{i0})$. Moreover, if we plot z_{ip} as a function of average intensity across the paired slides $\frac{1}{2}(\bar{y}_{ip1} + \bar{y}_{ip0})$, observations should approximately center around a constant difference $(\mu_{i1} - \mu_{i0})$. Figure 1 shows this plot for each of the twenty placebo subjects, along with a scatterplot smoother. In many subjects, the scatterplot smoother is reasonably constant across most of the range of average intensities, and also in most cases the scatterplot smoother lies either entirely above or below the horizontal y = 0 line. Although the smoother line is rarely perfectly horizontal, the additivity assumption of baseline peptide effect and plate effects at least roughly holds is a useful approximation to explain variability in peptide intensities.

Web Appendix B

Sum to zero constraint for slide effects

Let $S = 2N$, the number of slides, and let J_n denote an $n \times n$ matrix filled with ones. We define a positive semi-definite matrix

$$\Sigma_\mu = I_S - \frac{J_S}{S}$$

and let slide effects $\boldsymbol{\mu} \sim \text{Normal}(\mathbf{0}, \Sigma_\mu)$. To work with this prior, we represent $\boldsymbol{\mu}$ with an $(S-1)$-dimensional normal random variable. We define an $S \times (S-1)$ matrix Q via the eigen-decomposition of Σ_μ, taking the columns of Q to be the eigenvectors of Σ_μ corresponding to the positive eigenvalues, i.e.

$$\Sigma_\mu = Q I_{S-1} Q^T.$$

Next, we define a design matrix X between the slide effects $\boldsymbol{\mu}$ and the observations $\boldsymbol{y}$. We denote the total number of probers per slide as $N_{probe} = \sum_p^P n_p$, and we define an $S * N_{probe} \times S$ design matrix

$$X = \begin{pmatrix} \mathbf{1}_{N_{probe}} & \mathbf{0} & & \\ \mathbf{0} & \ddots & \ddots & \\ & \ddots & \ddots & \mathbf{0} \\ & & \mathbf{0} & \mathbf{1}_{N_{probe}} \end{pmatrix}.$$

Let $\boldsymbol{\mu}^*$ be an $S-1$ random vector with multivariate normal distribution $\text{Normal}(\mathbf{0}, I_{S-1})$, define $X^* \equiv XQ$, and let $\boldsymbol{\mu} = Q\boldsymbol{\mu}^*$. When defined as such, $\boldsymbol{\mu}$ has a degenerate multivariate normal distribution with mean $\mathbf{0}$ and covariance $QI_{S-1}Q^T = \Sigma_\mu$, as desired. Also note that $X^*\boldsymbol{\mu}^* = XQ\boldsymbol{\mu}^* = X\boldsymbol{\mu}$, which is useful for deriving the full conditional distribution of $\boldsymbol{\mu}^*$.

Web Appendix C

Monte Carlo Expectation Conditional Maximization

We find a posterior maximum using an Monte Carlo Expectation Conditional Maximization (MCECM) algorithm. At iteration $k+1$, we estimate $Q(\theta|\theta^{(k)}) = \text{E}\left[L(X, Z, \theta)|X, \theta^{(k)}\right]$ (the E-step). We then maximize $Q(\theta|\theta^{(k)})$ in a cyclic manner, generating $\theta^{(k+1)}$ (the M-step). Incorporating missing data typically eases the E and M steps such that they are fast, closed form operations that are simpler than directly optimizing the likelihood of the observed data. Jointly optimizing θ in $Q(\theta|\theta^{(k)})$ may remain intractable and parameters may instead be

updated in blocks, holding the remaining parameters fixed. Described in Meng and Rubin (1993), this is the Expectation Conditional Maximization algorithm which resembles the stable cyclical coordinate ascent optimization method. Because we are unable to generate a closed form expectation $\mathrm{E}\left[L(X, Z, \theta)|X, \theta^{(k)}\right]$, we approximate it with Monte Carlo methods (Wei and Tanner, 1990).

The MVT residual error distributions and the univariate-t distributions on random effects $\alpha_{ip0}, \alpha_{ip1}$ are implemented via a Normal-Gamma representation of a Student-t random variable. For each parameter $\alpha_{ip0}, \alpha_{ip1}$, we augment the prior distribution with parameters u_{ip0}, u_{ip1} having Gamma $\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ distributions. Similarly, we augment the prior distribution for residual errors with parameters w_{ip0}, w_{ip1} having Gamma $\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ distributions. Given these new latent variables, the full conditional distributions of numerous parameters become familiar, tractable distributions.

The probe responses $\boldsymbol{y}$ are observed data, while parameters $\boldsymbol{\gamma}, \boldsymbol{u}, \boldsymbol{w}$ are treated as missing data. We represent the remaining model parameters, except for random effects $\boldsymbol{\alpha}$, as θ. The random effects $\boldsymbol{\alpha}$ must be integrated from the likelihood to stabilize MCECM inference, but doing so causes the E step to lose a closed form solution. We cannot simultaneously integrate random effects $\boldsymbol{\alpha}$ and the weights $(\boldsymbol{u}, \boldsymbol{w})$ in a closed form, but we can separately integrate $(\boldsymbol{u}, \boldsymbol{w})$ or $\boldsymbol{\alpha}$ from the complete data likelihood to closed forms. We exploit these two separate forms to complete the E step using Monte Carlo and numerical integration techniques. Because of conditional independence relations in our model, we need only compute this expectation on a single observation $\boldsymbol{y}_{ip}$.

E-Step First we apply the Law of Total Expectation to separate the E-step calculation

into cases $\gamma_{ip} = 0$ or 1:

$$\mathrm{E}\left[l(\boldsymbol{y}_{ip}, \boldsymbol{u}_{ip}, \boldsymbol{w}_{ip}, \gamma_{ip}, \theta)|\boldsymbol{y}_{ip}, \theta^{(k)}\right] =$$

$$\sum_{j=0}^{1} \mathrm{E}\left[l(\boldsymbol{y}_{ip}, \boldsymbol{u}_{ip}, \boldsymbol{w}_{ip}, \gamma_{ip} = j, \theta)|\boldsymbol{y}_{ip}, \theta^{(k)}, \gamma_{ip} = j\right] \pi(\gamma_{ip} = j|\boldsymbol{y}_{ip}, \theta^{(k)}).$$

The probabilities $\pi(\gamma_{ip} = j|\boldsymbol{y}_{ip}, \theta^{(k)})$ are not available in closed form. We apply scaled, shifted univariate Gauss-Hermite quadratures to $\pi(\boldsymbol{y}_{ip0}, \boldsymbol{y}_{ip1}, \alpha_{ip0}|\gamma_{ip} = 0, \theta)$ and $\pi(\boldsymbol{y}_{ip0}, \boldsymbol{y}_{ip1}, \boldsymbol{\alpha}_{ip}|\gamma_{ip} = 1, \theta)$ to compute $\pi(\boldsymbol{y}_{ip}|\gamma_{ip} = j, \theta^{(k)})$. The prior for γ_{ip} is Bern(ω_p), and an application of Baye's Theorem produces

$$\begin{aligned}
\pi(\gamma_{ip} = 1|\boldsymbol{y}_{ip}, \theta) &= \frac{\pi(\boldsymbol{y}_{ip}|\gamma_{ip} = 1, \theta)\pi(\gamma_{ip} = 1, \theta)}{\pi(\boldsymbol{y}_{ip}|\gamma_{ip} = 0, \theta)\pi(\gamma_{ip} = 0, \theta) + \pi(\boldsymbol{y}_{ip}|\gamma_{ip} = 1, \theta)\pi(\gamma_{ip} = 1, \theta)} \\
&= \frac{\pi(\boldsymbol{y}_{ip}|\gamma_{ip} = 1, \theta)\pi(\gamma_{ip} = 1)\pi(\theta)}{\pi(\boldsymbol{y}_{ip}|\gamma_{ip} = 0, \theta)\pi(\gamma_{ip} = 0)\pi(\theta) + \pi(\boldsymbol{y}_{ip}|\gamma_{ip} = 1, \theta)\pi(\gamma_{ip} = 1)\pi(\theta)} \\
&= \frac{\pi(\boldsymbol{y}_{ip}|\gamma_{ip} = 1, \theta)\omega_p}{\pi(\boldsymbol{y}_{ip}|\gamma_{ip} = 0, \theta)(1 - \omega_p) + \pi(\boldsymbol{y}_{ip}|\gamma_{ip} = 1, \theta)\omega_p}.
\end{aligned}$$

We denote $\pi(\gamma_{ip} = 1|\boldsymbol{y}_{ip}, \theta^{(k)}) = \tau_{ip}^{(k)}$. As mentioned before, $E\left[l(\boldsymbol{y}_{ip}, \boldsymbol{u}_{ip}, \boldsymbol{w}_{ip}, \gamma_{ip} = j, \theta)|\boldsymbol{y}_{ip}, \theta^{(k)}, \gamma_{ip} = j\right]$ does not easily admit a closed form expression. We use importance sampling to generate a Monte Carlo estimate of the expectation, as the distribution $(\boldsymbol{u}_{ip}, \boldsymbol{w}_{ip}|\gamma_{ip} = j, \boldsymbol{y}_{ip}, \theta^{(k)}, \boldsymbol{\alpha}_{ip})$ is easily sampled. We first sample a value $\boldsymbol{\alpha}_{ip}^{(s)}$ from an instrumental distribution h, and then we draw $(\boldsymbol{u}_{ip}^{(s)}, \boldsymbol{w}_{ip}^{(s)})$ from the distribution $(\boldsymbol{u}_{ip}, \boldsymbol{w}_{ip}|\boldsymbol{y}_{ip}, \theta^{(k)}, \gamma_{ip} = j, \boldsymbol{\alpha}_{ip}^{(s)})$. In the case $\gamma_{ip} = 0$ (dropping ip subscripts for clarity),

$$E\left[l(\boldsymbol{y}, u_0, \boldsymbol{w}, \gamma = 0, \theta)|\boldsymbol{y}, \theta^{(k)}, \gamma = 0\right] =$$

$$\int_{-\infty}^{\infty} E\left[l(\boldsymbol{y}, u_0, \boldsymbol{w}, \gamma = 0, \theta)|\boldsymbol{y}, \theta^{(k)}, \gamma = 0, \alpha_0\right] \pi(\alpha_0|\boldsymbol{y}, \theta^{(k)}, \gamma = 0) d\alpha_0 =$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{\mathbf{R}^2} l(\boldsymbol{y}, \theta, u_0, \gamma = 0, \boldsymbol{w})\pi(\boldsymbol{w}, u_0|\boldsymbol{y}, \theta^{(k)}, \gamma = 0, \alpha_0)\pi(\alpha_0|\boldsymbol{y}, \theta^{(k)}, \gamma = 0) d\boldsymbol{w} du_0 d\alpha_0 =$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{\mathbf{R}^2} l(\boldsymbol{y}, \theta, \gamma = 0, u_0, \boldsymbol{w},)\frac{\pi(\alpha_0|\boldsymbol{y}, \theta^{(k)}, \gamma = 0)}{h(\alpha_0)}\pi(\boldsymbol{w}, u_0|\boldsymbol{y}, \theta^{(k)}, \gamma = 0, \alpha_0)h(\alpha_0) d\boldsymbol{w} du_0 d\alpha_0 \approx$$

$$\sum_{m=1}^{M} l(\boldsymbol{y}, u_0^{(m)}, \boldsymbol{w}^{(m)}, \gamma = 0, \theta)\frac{\pi(\alpha_0^{(m)}|\gamma = 0, \boldsymbol{y}, \theta^{(k)})}{h(\alpha_0^{(m)})}.$$

Although the normalizing constant for $\pi(\alpha_{ip0}^{(m)}|\gamma_{ip} = 0, \boldsymbol{y}_{ip}, \theta^{(k)})$ is available from our previous numerical integration, in practice we use the unnormalized version of importance sampling

and use importance weights

$$q_{ip0}^{(l)} \equiv \frac{\frac{\pi(\alpha_{ip0}^{(l)}|\gamma_{ip}=0,\boldsymbol{y}_{ip},\theta^{(k)})}{h(\alpha_{ip0}^{(l)})}}{\sum_{m=1}^{M}\frac{\pi(\alpha_{ip0}^{(m)}|\gamma_{ip}=0,\boldsymbol{y}_{ip},\theta^{(k)})}{h(\alpha_{ip0}^{(m)})}}.$$

For the case $\gamma_{ip} = 1$, we separately sample the weights and random effects for $\boldsymbol{y}_{ip0}$ and $\boldsymbol{y}_{ip1}$ in a similar fashion to the case $\gamma_{ip} = 0$. Our M-step function is then

$$\begin{aligned} Q(\theta|\theta^{(k)}) \approx \sum_{m=1}^{M} & l(\boldsymbol{y}_{ip}, u_{ip0}^{(m)}, \boldsymbol{w}_{ip}^{(m)}, \gamma_{ip}=0, \theta^{(k)}) q_{ip0}^{(m)}(1-\tau_{ip}^{(k)}) + \\ & l(\boldsymbol{y}_{ip0}, u_{ip0}^{(m)}, w_{ip0}^{(m)}, \gamma_{ip}=1, \theta^{(k)}) q_{ip1}^{(m,0)} \tau_{ip}^{(k)} + \\ & l(\boldsymbol{y}_{ip1}, u_{ip1}^{(m)}, w_{ip1}^{(m)}, \gamma_{ip}=1, \theta^{(k)}) q_{ip1}^{(m,1)} \tau_{ip}^{(k)}. \end{aligned}$$

The M-step Q function then yields manageable expressions for conditional maximizations.

Web Appendix D

[Figure 2 about here.]

References

Meng, X.-L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm : A general framework. *Biometrika* **80,** 267–278.

Wei, G. C. G. and Tanner, M. (1990). A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms. *Journal of the American Statistical Association* **85,** 699–704.

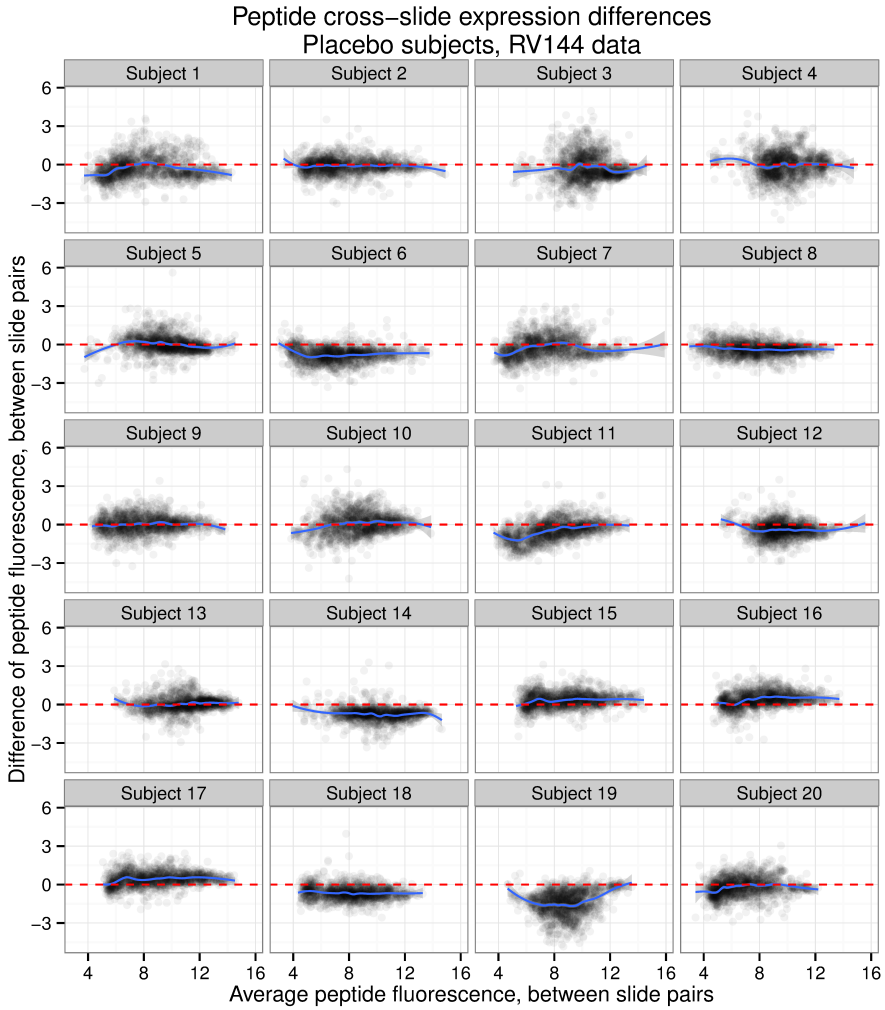

Figure 1. A loess scatterplot smoother traces the mean difference in fluorescence intensity between slide pairs versus the mean fluorescence across slide pairs, separately by each placebo subject. Under an additivity assumption, the scatterplot smoother should run approximately horizontally.

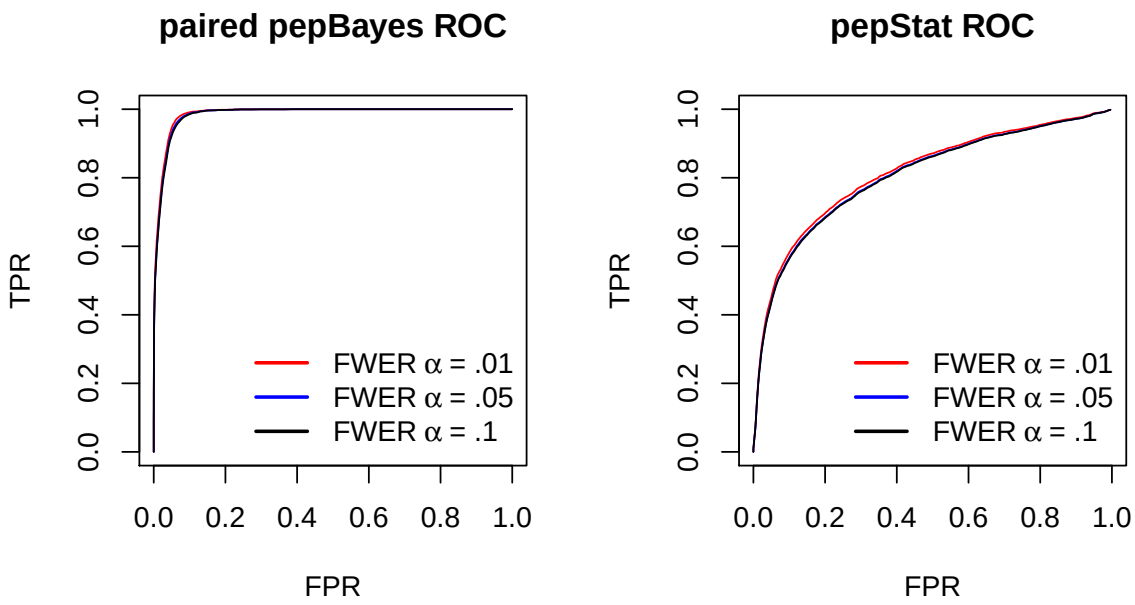

Figure 2. ROC curves for RV144 data with positive peptide positions called at different family-wise error rate α levels. At a variety of cutoffs, the results remain very similar