Web-based Supplementary Materials for "Bayesian Modeling for Genetic Anticipation"

Philip S. Boonstra, Bhramar Mukherjee, Jeremy M. G. Taylor,
Mef Nilbert, Victor M. Moreno, and Stephen B. Gruber

Web Appendix A: Gibbs Sampler

Common elements of M1-M4

We take a Gibbs sampler data augmentation approach to the censored observations (Tanner and Wong, 1987). Lindley and Smith (1972) present posterior distributions for a general linear model and Carlin and Louis (2000) give the posterior distributions for a normal random-effects linear model. We introduced one auxiliary variable to simplify calculations. Let $V_{ij}|\sigma^2 \overset{iid}{\sim} \mathcal{INV}-\chi^2(5,\sigma^2)$. It can be shown (p. 303-304, Gelman et al., 2004) that

$$T^*_{ij}|\beta, b_i, V_{ij}, \sigma^2 \overset{ind}{\sim} \mathcal{N}\left(X_{ij}^\top\beta + Z_{ij}^\top b_i, V_{ij}\right)$$

Let t^* denote the vector of event times for all individuals, t^*_i the vector of event times for family i, V the diagonal matrix of $\{V_{ij}\}_{i,j}$ and V_i the diagonal matrix of $\{V_{ij}\}_j$. Also, let $\text{trunc}-\mathcal{N}(m,s^2,l,u)$ be the $\mathcal{N}(m,s^2)$ distribution truncated to the interval $[l,u)$ and $\phi_2(x,m,S)$ be the density of a bivariate normal vector with mean m and variance matrix S evaluted at x. Suppose we have current estimates of all parameters from time r. Then the Markov chain Monte Carlo (MCMC) algorithm proceeds as follows, letting "$\leftarrow$" denote stochastic assignment and "$=$" denote deterministic assignment, the latter included only for notational convenience:

1. $t_{ij}^{*(r+1)} \leftarrow \text{trunc}-\mathcal{N}\left(X_{ij}^\top\beta^{(r)} + Z_{ij}^\top b_i^{(r)}, V_{ij}^{(r)}, t_{ij}^{L*}, t_{ij}^{U*}\right)$
2. $V_{ij}^{(r+1)} \leftarrow \mathcal{INV}-\chi^2\left(6, \left[(t_{ij}^{*(r+1)} - X_{ij}\beta^{(r)} - Z_{ij}b_i^{(r)})^2 + 5\sigma^{2(r)}\right]/6\right)$
3. $\beta^{(r+1)} \leftarrow \mathcal{N}\left((X^\top {V^{(r+1)}}^{-1} X)^{-1} X^\top {V^{(r+1)}}^{-1}(t^{*(r+1)} - Zb^{(r)}), (X^\top {V^{(r+1)}}^{-1} X)^{-1}\right)$

(4a) $\nu_i = \left(Z_i^\top {V^{(r+1)}}^{-1}(t_i^{*(r+1)} - X_i\beta^{(r+1)}) + {\Sigma^{(r)}_{d_i^{(r)}}}^{-1}\mu^{(r)}_{d_i^{(r)}}\right)$

(4b) $\Lambda_i = \left(Z_i^\top {V^{(r+1)}}^{-1} Z_i + {\Sigma^{(r)}_{d_i^{(r)}}}^{-1} \right)^{-1}$

4. $b_i^{(r+1)} \leftarrow \mathcal{MVN}\left(\Lambda_i \nu_i, \Lambda_i\right)$

5. ${\sigma^2}^{(r+1)} \leftarrow \mathcal{G}\left(\frac{5n}{2} + 1, \frac{5}{2}\text{Trace}\left({V^{(r+1)}}^{-1}\right) + .01\right)$

It remains to sample from the conditional distributions of the parameters governing the distribution of the b_i's. M1, M2, and M3 can be grouped together, as the former two are special cases of the latter. In the case of M4, the MCMC algorithm branches slightly, and we present it separately.

Unique elements of M1-M3

Because the dimension of the parameter space in M3 potentially changes at each iteration, we use methodology developed by Stephens (2000) to achieve this.

Conditional on $\{b_i^{(r+1)}\}_i$, new components $\{\widetilde{\pi}, \widetilde{\mu}, \widetilde{\Sigma}\}$ are "born" to Equation (2) in the text in continuous time at some constant, user-supplied rate (the mixture proportions $\{\pi_1^{(r)}, \ldots, \pi_k^{(r)}, \widetilde{\pi}\}$ are normalized to sum to unity); the birth distributions are the priors. Simultaneously, each existing component "dies" in continuous time at a rate inversely related to how well it explains $\{b_i^{(r+1)}\}_i$. This is collectively referred to as the birth-death process. Components which fit the random effects well will die slowly, while those which have poor fit will die quickly. Run this process for a fixed time to propose a mixture distribution of (potentially) different dimensionality (step $s + \delta$), and then, conditional on the new dimensionality, $k^{(r+1)} \equiv k^{(s+\delta)}$, sample new values of the mixture components (step $s + 1$), to improve mixing of the chain. Readers are referred to the Stephens paper, particularly Algorithm 3.1 for further details.

6. (M3 only) $\{k^{(r+1)}, \{\pi_\ell^{(r+\delta)}, \mu_\ell^{(r+\delta)}, \Sigma_\ell^{(r+\delta)}\}_\ell\}$
   $\leftarrow$ birth-death process$(\{b_i^{(r+1)}\}, \xi^{(r)}, \kappa^{(r)}, \Psi^{(r)})$

(7a) (M3 only) $\lambda_{i\ell} = \pi_\ell^{(r+\delta)} \phi_2(b_i^{(r+1)}, \mu_\ell^{(r+\delta)}, \Sigma_\ell^{(r+\delta)})$

7. (M3 only) $d_i^{(r+1)} \leftarrow \mathcal{MULTINOM}(1; \lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{ik^{(r+1)}})$

(8a) $\theta_\ell = \sum_i^N 1[d_i^{(r+1)} = \ell]$

8. (M3 only) $(\pi_1^{(r+1)}, \pi_2^{(r+1)}, \ldots, \pi_{k^{(r+1)}}^{(r+1)})$
   $\leftarrow \mathcal{DIR}\left(1 + \theta_1, 1 + \theta_2, \ldots, 1 + \theta_{k^{(r+1)}}\right)$

(9a) $\Xi_\ell = \left(\theta_\ell {\Sigma_\ell^{(r+\delta)}}^{-1} + \kappa^{(r)} \right)^{-1}$

(9b) $\bar{b}_\ell = \sum_{i:d_i^{(r+1)}=\ell} b_i^{(r+1)}/\theta_\ell$

(9c) $\omega_\ell = \left(\theta_\ell {\Sigma_\ell^{(r+\delta)}}^{-1} \bar{b}_\ell + \kappa^{(r)}\xi^{(r)}\right)$

9. $\mu_\ell^{(r+1)} \leftarrow \mathcal{MVN}\left(\Xi_\ell \omega_\ell, \Xi_\ell\right)$

10. $\Psi^{(r+1)} \leftarrow \mathcal{WISH}\left(2g + 8k^{(r+1)}, \left(2h + 2\sum_{\ell=1}^{k^{(r+1)}} {\Sigma_\ell^{(r+\delta)}}^{-1}\right)\right)$

11. ${\Sigma_\ell^{(r+1)}}^{-1}$
$\leftarrow \mathcal{WISH}\left(8 + \theta_\ell, \left(2\Psi^{(r+1)} + \sum_{i:d_i^{(r+1)}=\ell} (b_i^{(r+1)} - \mu_\ell^{(r+1)})(b_i^{(r+1)} - \mu_\ell^{(r+1)})^\top\right)^{-1}\right)$

12. $\xi^{(r+1)} \leftarrow \mathcal{MVN}\left(\sum_\ell \mu_\ell^{(r+1)}/k^{(r+1)}, \left(k^{(r+1)}\kappa^{(r)}\right)^{-1}\right)$

13. $\kappa^{(r+1)} \leftarrow \mathcal{WISH}\left(2 + k^{(r+1)}, \left(2I_{2\times 2} + \sum_\ell (\mu_\ell^{(r+1)} - \xi^{(r+1)})(\mu_\ell^{(r+1)} - \xi^{(r+1)})^\top\right)^{-1}\right)$

It is easy to make a draw of b_{N+1}, the random effects of a newly observed pedigree, conditional upon current parameter estimates, or to calculate the posterior predictive density (marginalizing over the mixture in the case of M3).

Unique Elements of M4

Recall the Dirichlet Process mixture (DPM) prior on b_i can be written as

$$\begin{aligned} b_i | d_i, \mu_{d_i}, \Sigma_{d_i} &\sim \mathcal{MVN}(\mu_{d_i}, \Sigma_{d_i}) \\ \mu_{d_i}, \Sigma_{d_i}^{-1} | G &\sim G \\ G | \alpha, G_0 &\sim \mathcal{DP}(\alpha, G_0(\mu, \Sigma)) \end{aligned}$$

where α is the precision parameter and $G_0(\mu, \Sigma)$ the base distribution of the Dirichlet Process prior. It is required then to sample $\{\mu_\ell^{(r+1)}, {\Sigma_\ell^{(r+1)}}^{-1}\}_\ell$, $\{d_i^{(r+1)}\}_i$, $\alpha^{(r+1)}$, $G_0^{(r+1)}$, and the necessary hyperparameters, all conditional on $\{b_i^{(r+1)}\}_i$. We use the `DPdensity` function found in the `R` package, `DPpackage` (Jara, 2007), which is an implementation of Algorithm 8 in Neal (2000). We refer readers to the paper for details, but the process is actually quite similar in spirit to the birth-death process proposed by Stephens. m draws of $\{\widetilde{\mu}, \widetilde{\Sigma}\}$ are made from $G_0^{(r)}(\mu, \Sigma)$, where m is an integer pre-set by the user. $b_i^{(r+1)}$ is assigned to an existing cluster proportional to the product of the cluster's current size and the cluster density evaluated at $b_i^{(r+1)}$ or to a new cluster proportional to the product of $\alpha^{(r+1)}/m$ and the density of the normal distribution with parameters $\{\widetilde{\mu}, \widetilde{\Sigma}\}$ evaluated

at $b_i^{(r+1)}$ (because of exchangeability, it does not matter which of the m draws of $\{\widetilde{\mu}, \widetilde{\Sigma}\}$ is used). Once the assignments have been shifted in this way, Gibbs steps can be used (assuming conjugacy) to draw $\alpha^{(r+1)}$, $G_0^{(r+1)}(\mu, \Sigma)$, etc., similar to above.

The one significant change between M1-M3 and M4 is in the posterior predictive density of b_{N+1}: because M4 allows for a countably infinite number of clusters, a newly observed pedigree should, with positive probability, be within a new cluster of families. Specifically, we wish to draw from

$$\begin{aligned} & p(b_{N+1}^{(r+1)} | \{b_i^{(r+1)}\}_i) \\ & = \int \int p(b_{N+1} | \{\mu_{d_{N+1}}^{(r+1)}, \Sigma_{d_{N+1}}^{(r+1)^{-1}}\} \\ & \qquad dp(\{\mu_{d_{N+1}}^{(r+1)}, \Sigma_{d_{N+1}}^{(r+1)^{-1}}\} | \{\mu_\ell^{(r+1)}, \Sigma_\ell^{(r+1)^{-1}}\}_\ell, \{b_i^{(r+1)}\}_i) \\ & \qquad dp(\{\mu_\ell^{(r+1)}, \Sigma_\ell^{(r+1)^{-1}}\}_\ell, \{b_i^{(r+1)}\}_i). \end{aligned}$$

See for example MacEachern and Müller (1998). To approximate this integral, at each iteration, let $\theta_\ell = \sum_i^N 1[d_i^{(r+1)} = \ell]$ and draw d_{N+1} such that

$$\Pr(d_{N+1} = \ell) = \begin{cases} \alpha^{(r+1)}(N + \alpha^{(r+1)})^{-1} & \ell = k^{(r+1)} + 1 \\ \theta_\ell (N + \alpha^{(r+1)})^{-1} & \ell \leq k^{(r+1)} \end{cases}$$

If $d_{N+1} \leq k^{(r+1)}$, then set $\{\mu_{d_{N+1}}^{(r+1)}, \Sigma_{d_{N+1}}^{(r+1)^{-1}}\}$ to correspond to its existing cluster. Otherwise, draw $\{\mu_{d_{N+1}}^{(r+1)}, \Sigma_{d_{N+1}}^{(r+1)^{-1}}\}$ from $G_0^{(r+1)}(\mu, \Sigma)$. Finally, make a draw from $b_{N+1} | \{\mu_{d_{N+1}}^{(r+1)}, \Sigma_{d_{N+1}}^{(r+1)^{-1}}\}$, which is just a bivariate normal distribution.

Web Appendix B: Figures

- Figure 1 displays the external data used to create the pseudo-AOO data (Engholm et al., 2010). Each birth cohort has a piecewise incidence rate for colorectal cancer.
- Figure 2 gives kernel density estimates of b_{1i} for all families from each model. Qualitatively, neither M3 nor M4 was able to cluster the mutation subtypes.

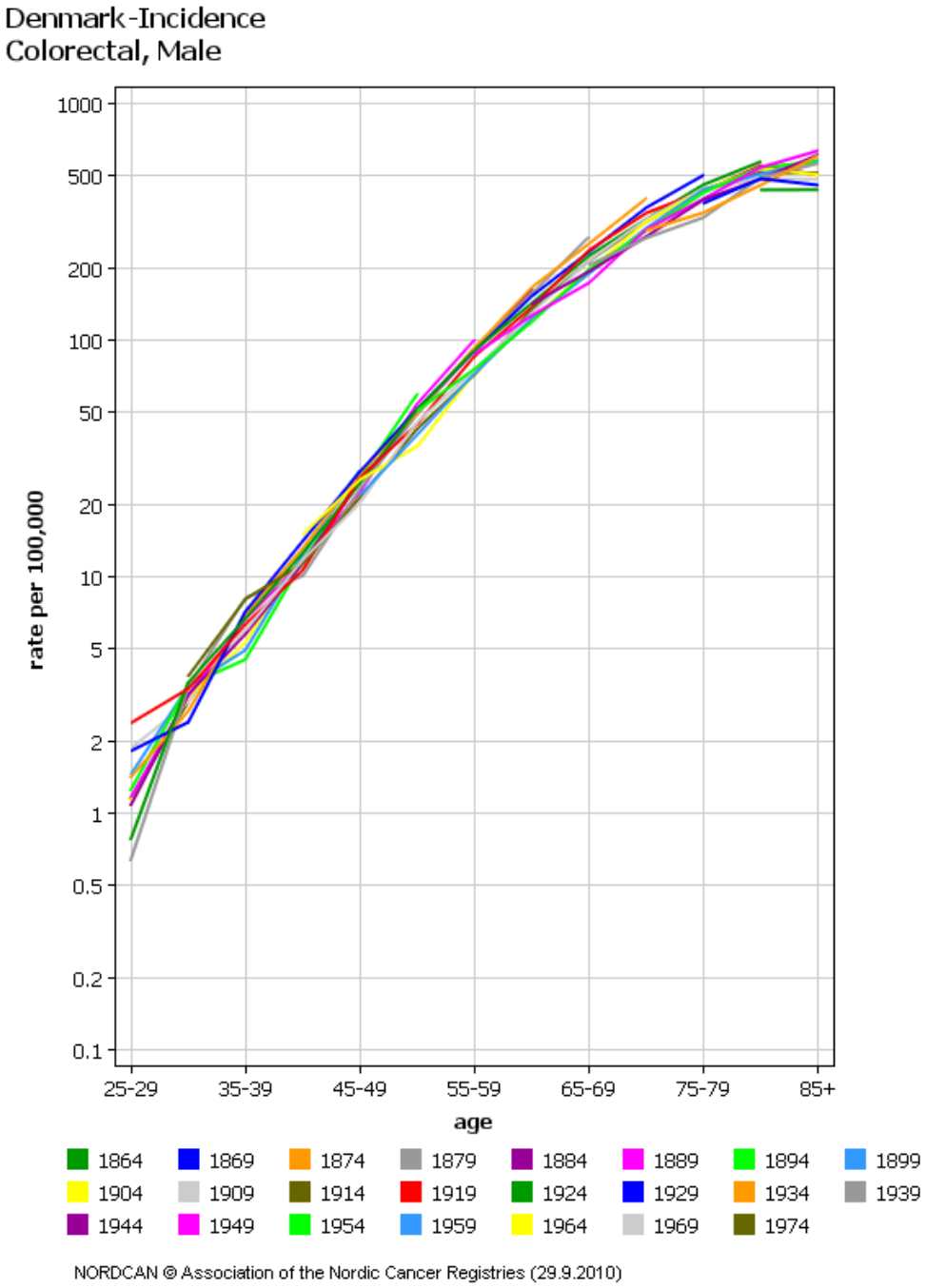

Figure 1: Pictorial representation of colorectal incidence data for males by birth cohort taken from NORDCAN (Engholm et al., 2010).

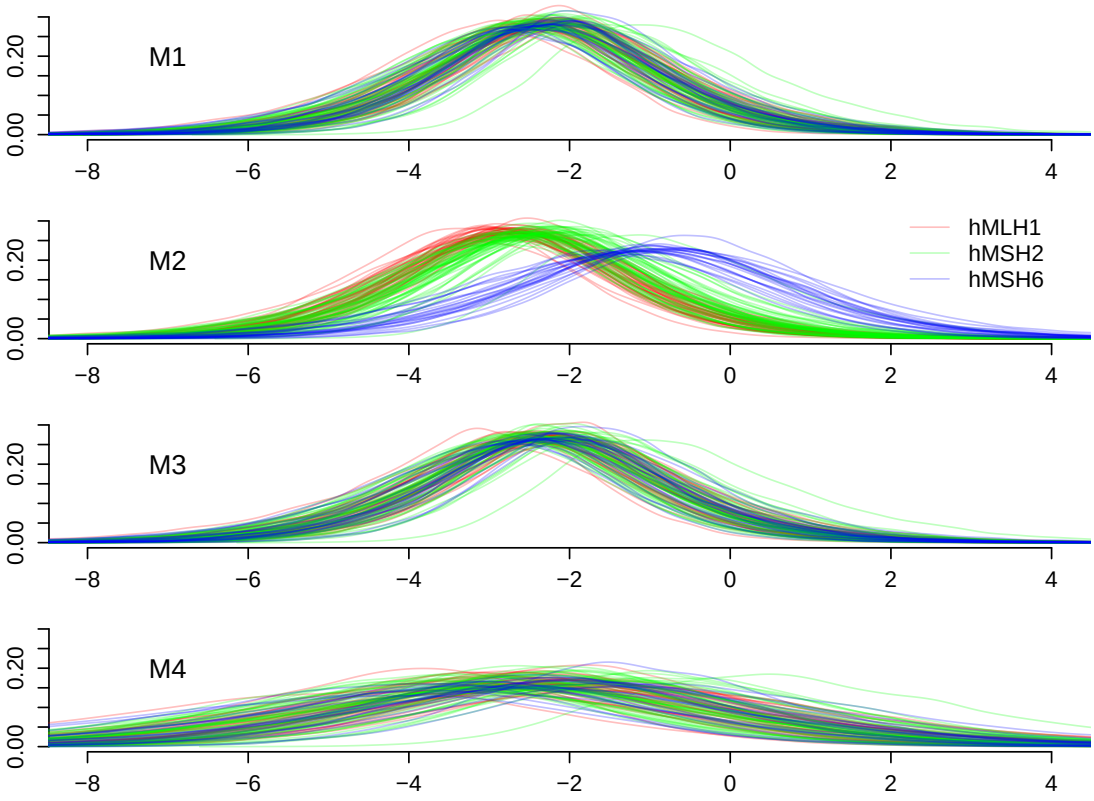

Figure 2: Kernel density estimates of b_{1i} for all families from each model, grouped by mutation subtype.

References

Carlin, B. and Louis, T. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall/CRC, 2nd edition.

Engholm, G., Ferlay, J., Christensen, N., Bray, F., Gjerstorff, M. L., Klint, A., Køtlum, J. E., Ólafsdóttir, E., Pukkala, E., and Storm, H. H. (2010). NORDCAN: Cancer incidence, mortality, prevalence and prediction in the Nordic countries, Version 3.6. Association of the Nordic Cancer Registries. Danish Cancer Society. Oct. 2010 <http://www.ancr.nu>.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian data analysis*, pages 303–304. Texts in Statistical Science. Chapman & Hall/CRC, 2nd edition.

Jara, A. (2007). Applied Bayesian non- and semi-parametric inference using DPpackage. *Rnews* **7,** 17–26.

Lindley, D. and Smith, A. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society: Series B* **34,** 1–41.

MacEachern, S. N. and Müller, P. (1998). Estimating mixture of Dirchlet Process models. *Journal of Computational and Graphical Statistics* **7,** 223–238.

Neal, R. M. (2000). Markov Chain sampling methods for Dirichlet Process mixture models. *Journal of Computational and Graphical Statistics* **9,** 249–265.

Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of components – an alternative to reversible jump methods. *The Annals of Statistics* **28,** 40–74.

Tanner, M. and Wong, W. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82,** 528–540.