Web-based Supplementary Materials for "Bayesian Genome- and Epigenome-wide Association Studies with Gene Level Dependence"

Eric F. Lock and David B Dunson

September 23, 2016

Web Appendix A: Posterior Computation

This is a supplement to the article "Bayesian genome- and epigenome-wide association studies with gene level dependence," providing details on posterior computation for the methylation screening application described therein.

First, we estimate and fix a dictionary of normal kernels truncated between 0 and 1, which will be used as mixture components for the density at each CpG site. These are estimated as described in Section 5 of Lock and Dunson (2015). In particular, the number of kernels is determined by out-of-sample cross validation of the log posterior density. For the present application this yields $K = 8$ kernels that appropriately span the data range from 0 to 1.

Let $\Pi_{gm}^{(0)} = (\pi_{gm1}^{(0)}, \ldots, \pi_{gmK}^{(0)})$ be the kernel probability weights that define the generative distribution for gene g and site m for group 0, and let $\Pi_{gm}^{(1)}$ be the kernel probability weights for group 1. Under the null model H_{0gm}, the mixing weights are the same for both groups: $\Pi_{gm}^{(0)} = \Pi_{gm}^{(1)}$. The kernel weights are assumed to be generated from a Dirichlet(λ) distribution, where λ is a hyper-parameter that is inferred during the kernel estimation stage and fixed. Under H_{1gm}, $\Pi_{gm}^{(0)}$ and $\Pi_{gm}^{(1)}$ are considered independent realizations from Dirichlet(λ).

Under this framework, posterior draws from the gene-level prior model described in Section 3 of the main article are incorporated into Gibbs sampling for the kernel testing parameters. The full sampling algorithm is described below.

1. Draw the kernel that generated each observation $T^{(i)}_{gmn} \in 1, \ldots, K$ for genes $g = 1, \ldots, G$, markers $m = 1, \ldots, M_g$, classes $i = 0, 1$ and samples $n = 1, \ldots, N_i$. The conditional probability that the given value is realized from component k is

$$P(T^{(i)}_{mn} = k \mid X^{(i)}_m, \Pi^{(i)}_m) \propto \pi^{(i)}_{gmk} f(X^{(i)}_{gmn} | \mu_k, \sigma_k, [0, 1]),$$

where $f(\cdot)$ defines the density of a truncated normal distribution.

2. Designate null markers $H_{0,gm}$ for $g = 1, \ldots, G$, $m = 1, \ldots, M_g$. The conditional posterior probability is

$$P(H_{0,gm} \mid X, Y, p_g) = \frac{p_g \beta(\lambda)\beta(\vec{n}_m + \lambda)}{p_g \beta(\lambda)\beta(\vec{n}_m + \lambda) + (1 - p_g)\beta(\vec{n}^{(0)}_m + \lambda)\beta(\vec{n}^{(1)}_m + \lambda)},$$

where $\vec{n}^{(0)}_{gm} = (n^{(0)}_{gm1}, \ldots, n^{(0)}_{gmK})$ is the number of subjects in group 0 that belong to each kernel k in marker g, m, $\vec{n}^{(1)}_{gm}$ is defined similarly for group 1, and $\vec{n}_{gm} = \vec{n}^{(0)}_{gm} + \vec{n}^{(1)}_{gm}$.

3. Draw weights $\{\Pi^{(0)}_{gm}, \Pi^{(1)}_{gm}\}^M_{m=1}$. Under $H_{0,gm}$, $\Pi^{(0)}_{gm} = \Pi^{(1)}_{gm} \sim \text{Dirichlet}(\lambda + \vec{n}_{gm})$. Otherwise, $\Pi^{(0)}_{gm} \sim \text{Dirichlet}(\lambda + \vec{n}^{(0)}_{mg})$ and $\Pi^{(1)}_{gm} \sim \text{Dirichlet}(\lambda + \vec{n}^{(1)}_{mg})$.

4. Allocate gene-level Dirichlet indices C_g for $g = 1, \ldots, G$:

$$P(C_g = h \mid \theta_\cdot, H_{0,g\cdot}) \propto \pi_h \theta_h^{S_g} (1 - \theta_h)^{M_g - S_g}$$

where S_g is the number of null markers in gene g, $S_g = \sum_{m=1}^{M_g} \mathbb{1}(H_{0,gm})$.

5. Update the weights π_h for $h = 1, \ldots, H$. First, draw the stick-breaking weights $V_1, \ldots, V_{H-1}$ by

$$(V_h \mid C_\cdot) \sim \text{Beta}\left(1 + \sum_{g=1}^{G} \mathbb{1}(C_g = h), \alpha + \sum_{g=1}^{G} \mathbb{1}(C_g > h)\right),$$

with $V_H = 1$. Then set $\pi_h = V_h \prod_{l<h}(1 - V_l)$ for $h = 1, \ldots, H$.

6. Update the atoms θ_h for $h = 1, \ldots, H$:

$$(\theta_h \mid C_\cdot, H_{0,\cdot\cdot}) \sim \text{Beta}\left(a + \tilde{S}_h, b + \tilde{M}_h - \tilde{S}_h\right),$$

where $\tilde{M}_h$ is the total number of markers in genes allocated to cluster h, and $\tilde{S}_h$ is the number of null markers:

$$\tilde{M}_h = \sum_{\{g:C_g=h\}} M_g \ , \ \tilde{S}_h = \sum_{\{g:C_g=h\}} S_g.$$

Set $p_g = \theta_{C_g}$ for $g = 1, \ldots, G$.

We use a simple uniform prior for the base distribution of p_g $(a = b = 1)$. For the high-throughput data considered, computing is not trivial, costing approximately 30 seconds per Gibbs cycle. However, less than 1% of computing time is spent on the draws for the gene-level prior parameters (steps 4-6). We find that draws mix well and converge very quickly to a stationary posterior. We run two parallel chains, with different initializations, for 1000 cycles, with the first 200 treated as burn-in. Figure 1 shows good agreement of estimated gene-level prior probabilities between the two chains.

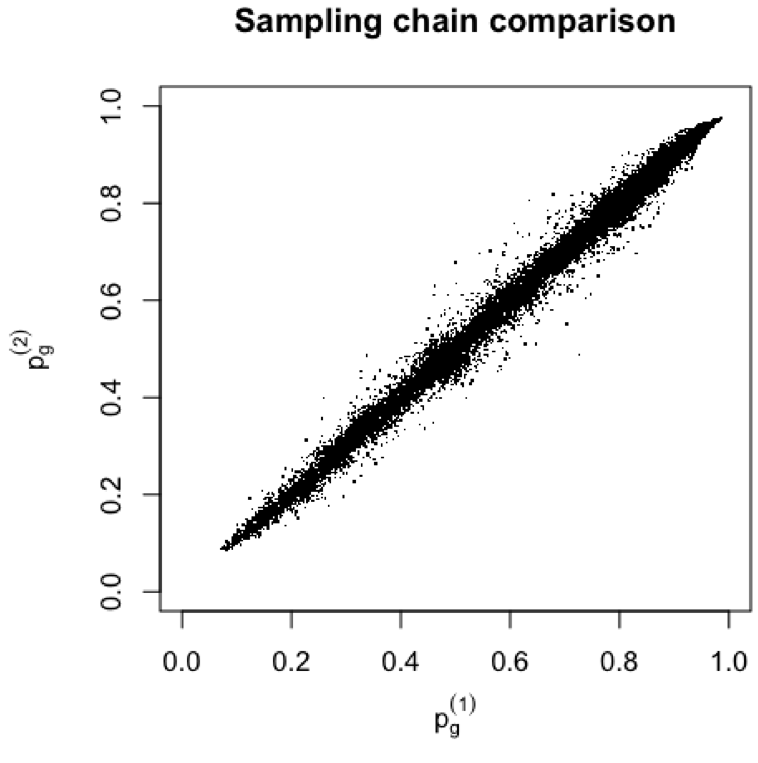

Figure 1: Scatterplot of estimated gene-level prior from two independent sampling chains.

Web Appendix B: Simulation ROC Curves

Here we give additional results for the simulation study described in Section 4 of the main manuscript. Figure 2 gives receiver operating characteristic (ROC) curves showing the proportion of markers with false positive or true positive classification (where 'positive' corresponds to the alternative) as the threshold on the posterior probability, p-value, or FDR is varied. Results are shown for the bimodal and Beta simulations; the ROC curve for the null simulation is trivial, as there are no true positives. In both cases the Bayesian hierarchical model has uniformly better classification performance than alternatives. The ROC curves for separate estimation are close, indicating that the rank ordering of probabilities are similar despite the improved accuracy of the hierarchical model.

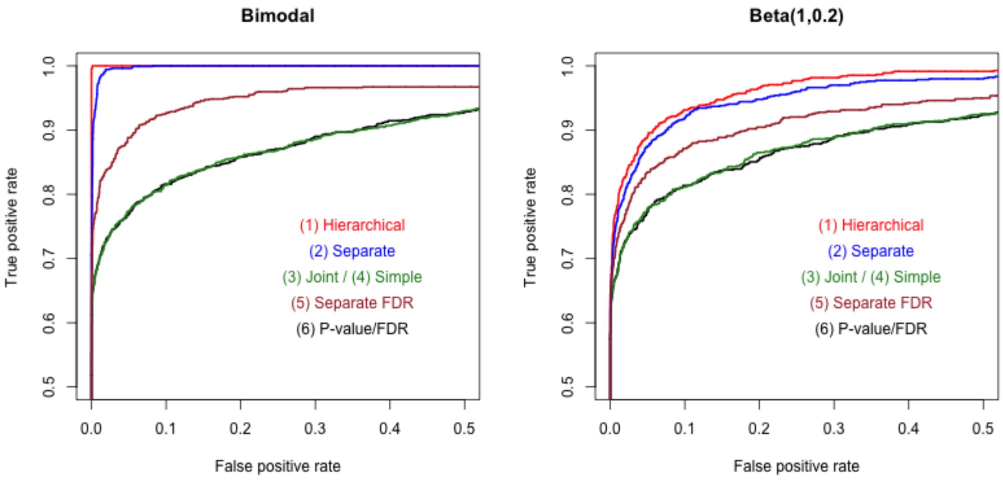

Figure 2: ROC curves obtained by varying the threshold on the posterior probability, p-value, or FDR for various methods. Curves are combined for the joint and simple Bayesian models, and for the uncorrected p-value and FDR, as the rank ordering of markers between these methods do not change.

Web Appendix C: Hyperparameter Sensitivity

Here we present the results of a simulation study to assess the sensitivity and affect of hyperparameters for the hierarchical gene-level prior. Data

Table 1: Average expected misclassification probability over all markers, for different hyperparameters

	Null	Bimodal	Beta $(1, 0.2)$
$\text{Beta}(1,1)$, $\alpha = 0.1$	0.0001	0.005	0.079
$\text{Beta}(1,1)$, $\alpha = 1$	0.0003	0.005	0.079
$\text{Beta}(1,1)$, $\alpha = 10$	0.0004	0.009	0.079
$\text{Beta}(19,1)$, $\alpha = 0.1$	0.0001	0.014	0.081
$\text{Beta}(19,1)$, $\alpha = 1$	0.0002	0.014	0.081
$\text{Beta}(19,1)$, $\alpha = 10$	0.0012	0.015	0.081
$\text{Beta}(M,1)$, $\alpha = 0.1$	0.00004	0.071	0.086
$\text{Beta}(M,1)$, $\alpha = 1$	0.00006	0.071	0.087
$\text{Beta}(M,1)$, $\alpha = 10$	0.00006	0.070	0.085

are simulated exactly as in Section 4 of the main manuscript, using the likelihood framework of Example 2.1. As in Section 4, we simulate gene-level probabilities for a global null hypothesis, a bimodal scenario where markers in 20% of genes are alternative and the other 80% are null, and where gene-level probabilities are generated from a $\text{Beta}(1, 0.2)$ distribution. For each scenario, we compute the posterior for 9 different choices for the model hyperparamaters. We use a factorial design with 3 different values of the DP hyperparameter α ($\alpha = 0.1, 1$, and 10), and three different parameter values for the beta base distribution. For the beta base distribution we consider the default uniform ($a = b = 1$), a more conservative choice ($a = 19, b = 1$), and an even more conservative choice motivated by a prior adjustment for multiplicity ($a = M, b = 1$ where M is the total number of markers).

Table 1 gives the results for each set of hyperparameters, summarized by the average expected misclassification probability over all markers. As expected, the more conservative choices for the beta base distribution lead to better performance under a global null, but the effect is minor. The most conservative base distribution, $\text{Beta}(M, 1)$, performs substantially worse for the bimodal case, where it has less flexibility to detect alternative genes. For the global null, smaller values of α perform better, as this implies larger clusters and more similarity between the genes; the effect of α is negligible for the other scenarios.

Web Appendix D: Shared Kernel Simulation

Here we present a simulation study using continuous data and shared kernels, as in the methylation application described in Section 5 of the main manuscript. Continuous data from 0 to 1 are generated from a mixture of truncated normal distributions (kernels); for data generation we use the same 8 kernels that are estimated for the methylation application (see Web Appendix A). For null markers the mixing weights that characterize the continuous distribution are the same for two classes, generated from a uniform Dirichlet distribution. For alternative markers the mixing weights for each class are generated independently from uniform distributions. Other aspects of the simulation design are analogous to that in Section 4 of the main manuscript, including the sample size (80 in each group), number of genes (500), and number of markers per gene (selected uniformly from $2, \ldots, 20$). The simulation scenarios follow those in Section 4 of the main manuscript, with gene-level probabilities for a global null hypothesis, a bimodal scenario where markers in 20% of genes are alternative and the other 80% are null, and where gene-level probabilities are generated from a Beta$(1, 0.2)$ distribution.

For each scenario, we compute the posterior using the sampling algorithm given in Web Appendix A using the gene-level hierarchical model. As in Section 4, we also consider a model with a separate uniform prior for each gene, a joint model with a shared prior for all genes, and a simple model with fixed prior probability 0.5. The results are given in Table 2. The advantages of the hierarchical model are apparent, but the hypotheses are reasonably well discriminated even for the simple model. We also consider a smaller sample size of $N = 40$ (20 in each group); here, each individual marker provides less information to discriminate the hypotheses, and the advantages of the hierarchical approach are more substantial.

Table 2: Average expected misclassification probability over all markers, for different estimation schemes.

N=160	Null	Bimodal	Beta $(1, 0.2)$
Hierarchical	0.01%	0.50%	2.1%
Separate	0.62%	0.99%	2.3%
Joint	0.01%	3.67%	3.1%
Simple	4.1%	4.9%	4.6%
N=40	Null	Bimodal	Beta $(1, 0.2)$
Hierarchical	0.10%	0.96%	10.1%
Separate	6.56%	8.11%	13.0%
Joint	0.10%	16.9%	14.7%
Simple	24.0%	25.4%	24.7%

References

Lock, E. F. and Dunson, D. B. (2015). Shared kernel Bayesian screening. *Biometrika* **102,** 829–842.