MARKOV CHAIN MONTE CARLO ALGORITHMS BEHIND THE REGRESSION WITH SUMMARY STATISTICS

By Xiang Zhu

University of Chicago

1. Overview. We describe the Markov chain Monte Carlo (MCMC) algorithms in terms of $\{S, R\}$, and then replace the unknown $\{S, R\}$ with their estimates $\{\widehat{S}, \widehat{R}\}$ in practice. This is similar to the likelihood derivation and prior specification in Zhu and Stephens (2016).

With minor modifications, we implement three prior distribution of β based on previous work.

- BVSR prior
  This prior modifies the prior distribution in Guan and Stephens (2011):

  $$\beta_j \sim \pi\mathcal{N}(0, \sigma_B^2) + (1-\pi)\delta_0, \tag{1.1}$$

  where δ_0 denotes the point mass at zero and $\sigma_B^2 := h(\pi\sum_{j=1}^p n^{-1}s_j^{-2})^{-1}$. The prior distributions of π and h are given by (1.4).
- BSLMM prior
  This prior modifies the prior distribution in Zhou, Carbonetto and Stephens (2013):

  $$\beta_j \sim \pi\mathcal{N}(0, \sigma_B^2 + \sigma_P^2) + (1-\pi)\mathcal{N}(0, \sigma_P^2), \tag{1.2}$$

  where the variances are given by

  $$\sigma_B^2 := h\rho(\pi\textstyle\sum_{j=1}^p n^{-1}s_j^{-2})^{-1},\ \ \sigma_P^2 := h(1-\rho)(\sum_{j=1}^p n^{-1}s_j^{-2})^{-1}, \tag{1.3}$$

  and the hyper-parameters $\{\pi, h, \rho\}$ are placed on independent priors

  $$\log\pi \sim \mathcal{U}(\log(1/p), \log 1),\ \ h \sim \mathcal{U}(0,1),\ \ \rho \sim \mathcal{U}(0,1). \tag{1.4}$$
- ASH prior
  This is the same as the prior distribution in Stephens (2016):

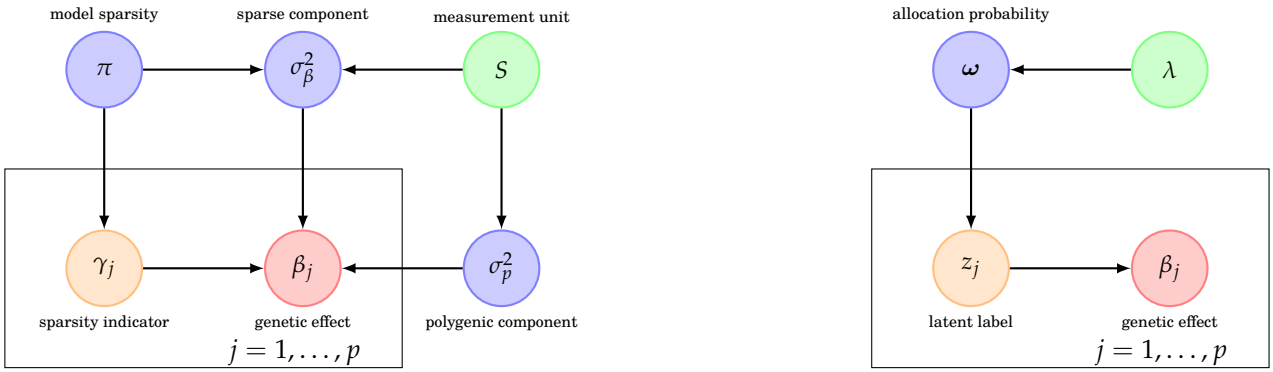
  $$\beta_j \sim \sum_{k=1}^K \omega_k\mathcal{N}(0, \sigma_k^2),\ \ \boldsymbol{\omega} \sim \mathcal{D}(\lambda, \ldots, \lambda),\ \ \lambda \sim \mathcal{U}(0, 10), \tag{1.5}$$

  where $\boldsymbol{\omega} := (\omega_1, \ldots, \omega_K)^\intercal$, K and $\{\sigma_k^2\}$ are pre-specified, and $\mathcal{D}$ denotes Dirichlet distribution.

Note that the BVSR prior is a special case of the BSLMM prior where $\rho = 1$. The plate notations of BSLMM and ASH priors are given in Figure 1.

2. Rank-based strategy. When locally updating the SNP-specific parameters (e.g. genetic effect β_j and sparsity indicator γ_j for each SNP j) in the MCMC algorithms, we allocate more computational resources to SNPs with larger marginal association signals, using the rank-based strategy (Guan and Stephens, 2011). In particular, we first rank all the variants based on the single-SNP p-values and draw one SNP to update according to some probability distributions with decreasing probability. In our current implementation, we use a mixture distribution $q_p = 0.3u_p + 0.7g_p$, where u_p is a discrete uniform distribution and g_p is a geometric distribution truncated to $1, \ldots, p$ with its parameter chosen to give a mean of 2000.

Based on q_p, we introduce $Q(\cdot|\gamma)$, a proposal for the indicator γ. To propose a new value γ^* given the current value γ, we start by setting $\gamma^* = \gamma$ and then randomly choose one of the following:



(a) BSLMM prior (b) ASH prior

Fig 1: *Plate notations for two types of prior on β. Figure (1a) and (1b) correspond to Bayesian sparse linear mixed model (BSLMM) prior and Adaptive shrinkage (ASH) prior respectively.*

1. With probability P_a, draw SNP r according to q_p until $\gamma_r = 0$ and set $\gamma_r^* = 1$.
2. With probability P_r, draw SNP r uniformly from $\{j : \gamma_j = 1\}$ and set $\gamma_r^* = 0$.
3. With probability P_e, sample two SNPs by the above two steps and switch their indicators.

The default setting in our software is $P_a = P_r = 0.4, P_e = 0.2$.

3. BVSR prior. For RSS with BVSR prior, we use Metropolis-Hastings (MH) algorithm to obtain posterior samples of (γ, π, h) on the product space of $\{0,1\}^p \times (0,1) \times (0,1)$,

$$p(\gamma, \pi, h|\widehat{\beta}, S, R) \propto p(\widehat{\beta}|S, R, \gamma, \pi, h)p(\gamma|\pi)p(\pi)p(h). \tag{3.1}$$

Here we are exploiting the fact that β can be integrated out analytically to compute $p(\widehat{\beta}|S, R, \gamma, \pi, h)$:

$$\widehat{\beta}|S, R, \gamma, \pi, h \sim \mathcal{N}(\mathbf{0}, SRS + \sigma_B^2 M_\gamma M_\gamma^\intercal), \tag{3.2}$$

where $M := SRS^{-1}$ and M_γ denotes the sub-matrix of M restricted to those columns j for which $\gamma_j = 1$. We update γ using the rank-based proposal $Q(\cdot|\gamma)$. We update $\log \pi$ by adding a random number from $\mathcal{U}(-0.05, 0.05)$ to the current value, and update h by adding a random number from $\mathcal{U}(-0.1, 0.1)$ to the current value. New values of $\log \pi$ and h outside boundaries are reflected back.

For each simulated posterior draw of (γ, π, h), we sample β according to its conditional distributions given (γ, π, h) and $(\widehat{\beta}, S, R)$:

$$\beta_\gamma|\widehat{\beta}, S, R, \gamma, \pi, h \sim \mathcal{N}(\mu, \Omega^{-1}), \tag{3.3}$$

$$\beta_{-\gamma}|\widehat{\beta}, S, R, \gamma, \pi, h \sim \delta_0, \tag{3.4}$$

where β_γ and $\beta_{-\gamma}$ denote the subsets of β corresponding to the entries that $\gamma_j = 1$ and 0 respectively, δ_0 denotes the point mass at zero and,

$$\Omega := M_\gamma^\intercal (SRS)^{-1} M_\gamma + \sigma_B^{-2}(\gamma, \pi, h) I_{|\gamma|}, \tag{3.5}$$

$$\mu := \Omega^{-1} M_\gamma^\intercal (SRS)^{-1} \widehat{\beta}. \tag{3.6}$$

The marginal likelihood (3.2), up to some constant, can be written in terms of (Ω, μ),

$$p(\widehat{\beta}|S, R, \gamma, \pi, h) \propto \sigma_B^{-|\gamma|} |\Omega|^{-1/2} \exp\{\mu^\intercal \mathbf{q}_\gamma / 2\}, \tag{3.7}$$

where $\mathbf{q}_\gamma$ denotes the subset of $\mathbf{q} := S^{-1}\beta$ corresponding to the entries that $\gamma_j = 1$. The matrix computation in a single step of the MCMC algorithm above involves one Cholesky decomposition of Ω and three triangular linear systems. Hence, the computational cost for each iteration of MCMC is $\mathcal{O}(|\gamma|^3 + 3|\gamma|^2)$, where $|\gamma|$ denotes the number of non-zero entries in γ.

To improve precision, we can use Rao-Blackwellized estimates. For SPIP, we have

$$\Pr(\gamma_j = 1|\widehat{\beta}, S, R) = \mathrm{E}(\Pr(\gamma_j = 1|\widehat{\beta}, S, R, \xi_{-j})) \approx M^{-1}\textstyle\sum_{i=1}^{M}\Pr(\gamma_j = 1|\widehat{\beta}, S, R, \xi_{-j}^{(i)})$$

where ξ_{-j} stands for $\{\beta_{-j}, \gamma_{-j}, \pi, h\}$, γ_{-j} and β_{-j} denote the vectors γ and β excluding the jth coordinate and $\xi_{-j}^{(i)}$ denotes the ith MCMC sample from the posterior distribution of ξ_{-j}. For the posterior mean of the multiple-SNP effect at SNP j, we have

$$\mathrm{E}(\beta_j|\widehat{\beta}, S, R) = \mathrm{E}(\mathrm{E}(\beta_j|\widehat{\beta}, S, R, \xi_{-j})) \approx M^{-1}\textstyle\sum_{i=1}^{M}\mathrm{E}(\beta_j|\widehat{\beta}, S, R, \gamma_j = 1, \xi_{-j}^{(i)})\Pr(\gamma_j = 1|\widehat{\beta}, S, R, \xi_{-j}^{(i)}).$$

To obtain the Rao-Blackwellized estimates, we only need $p(\gamma_j|\widehat{\beta}, S, R, \xi_{-j})$ and $p(\beta_j|\widehat{\beta}, S, R, \gamma_j, \xi_{-j})$:

$$\begin{aligned}
\frac{\Pr(\gamma_j = 1|\widehat{\beta}, S, R, \xi_{-j})}{\Pr(\gamma_j = 0|\widehat{\beta}, S, R, \xi_{-j})} &= \frac{\pi}{1-\pi}\sqrt{\frac{s_j^2}{s_j^2+\sigma_B^2}}\exp\left\{\frac{1}{2(\sigma_B^{-2}+s_j^{-2})}\left(\frac{\hat{\beta}_j}{s_j^2} - \sum_{i\neq j}\frac{r_{ij}\beta_i}{s_i s_j}\right)^2\right\} \\
\beta_j|\widehat{\beta}, S, R, \gamma_j = 1, \xi_{-j} &\sim \mathcal{N}\left(\frac{1}{\sigma_B^{-2}+s_j^{-2}}\left(\frac{\hat{\beta}_j}{s_j^2} - \sum_{i\neq j}\frac{r_{ij}\beta_i}{s_i s_j}\right), \frac{1}{\sigma_B^{-2}+s_j^{-2}}\right) \\
\beta_j|\widehat{\beta}, S, R, \gamma_j = 0, \xi_{-j} &\sim \delta_0
\end{aligned}$$

where r_{ij} is the (i,j)-th entry of R.

4. BSLMM prior. We propose a component-wise MCMC algorithm for RSS with BSLMM prior. First, we re-parameterize the multiple-SNP effect sizes β_j as follows

$$\beta_j|\gamma_j = 1, \pi, h, \rho, S = \sqrt{\sigma_B^2 + \sigma_P^2}\cdot\tilde{\beta}_j \tag{4.1}$$

$$\beta_j|\gamma_j = 0, \pi, h, \rho, S = \sigma_P\cdot\tilde{\beta}_j \tag{4.2}$$

where the standardized effect sizes $\tilde{\beta}_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, for $j \in \{1, \ldots, p\}$. Equivalently,

$$\beta = B\widetilde{\beta},\ \ \widetilde{\beta} \sim \mathcal{N}(\mathbf{0}, I_p) \tag{4.3}$$

where the scaling matrix B is diagonal with the jth diagonal b_j defined as

$$b_j = \sigma_P\mathbf{1}\{\gamma_j = 0\} + \sqrt{\sigma_B^2 + \sigma_P^2}\mathbf{1}\{\gamma_j = 1\}. \tag{4.4}$$

The new parameterization could help speed up the convergence of MCMC, since $\tilde{\beta}$ are independent with (γ, π, h, ρ) *a priori*. We then draw posterior samples of $(\tilde{\beta}, \gamma, \pi, h, \rho)$ iteratively.

- Given $(\widetilde{\beta}, \pi, h, \rho)$, we update γ by a standard MH algorithm, where the proposal is $Q(\cdot|\gamma)$.
- Given (γ, π, h, ρ), we update $\widetilde{\beta}$ by a mixture of global and local moves. With probability P_{g}, we draw a new value of $\widetilde{\beta}$ from its full conditional,

  $$\widetilde{\beta}|\widehat{\beta}, S, R, \gamma, \pi, h, \rho \sim \mathcal{N}((BS^{-1}RS^{-1}B + I)^{-1}BS^{-2}\widehat{\beta}, (BS^{-1}RS^{-1}B + I)^{-1}). \tag{4.5}$$

  With probability $1 - P_{\mathrm{g}}$, we randomly pick a SNP j according to the distribution q_p and draw $\tilde{\beta}_j$ from its full conditional

  $$\tilde{\beta}_j|\widehat{\beta}, S, R, \tilde{\beta}_{-j}, \gamma, \pi, h, \rho \sim \mathcal{N}\left(\frac{b_j s_j \ell_j}{s_j^2 + b_j^2}, \frac{s_j^2}{s_j^2 + b_j^2}\right),\ \ell_j := \frac{\hat{\beta}_j}{s_j} - \sum_{i\neq j}\frac{r_{ij}b_i\tilde{\beta}_i}{s_i}. \tag{4.6}$$

- Given $(\widetilde{\boldsymbol{\beta}}, \gamma, h, \rho)$, we update π by a Metropolis algorithm, where the proposal is symmetric Gaussian random walks on $\log((\pi - p^{-1})/(1-\pi))$.
- Given $(\widetilde{\boldsymbol{\beta}}, \gamma, \pi, \rho)$, we update h by a Metropolis algorithm, where the proposal is symmetric Gaussian random walks on $\log(h/(1-h))$.
- Given $(\widetilde{\boldsymbol{\beta}}, \gamma, \pi, h)$, we update ρ by a Metropolis algorithm, where the proposal is symmetric Gaussian random walks on $\log(\rho/(1-\rho))$.

The most computationally intensive step is drawing $\widetilde{\boldsymbol{\beta}}$ from a p-dimensional multivariate normal distribution (4.5). For each draw, one Cholesky decomposition of $BS^{-1}RS^{-1}B + I$ and two triangular linear systems are required. Since matrix R is banded with some bandwidth w (Wen and Stephens, 2010), the matrix $BS^{-1}RS^{-1}B + I$ also has the same bandwidth and therefore, the per-iteration cost of the algorithm above is at most $\mathcal{O}(pw^2 + 2p^2)$. For all the simulations, we set $P_{\mathrm{g}} = 0.05$. For the analysis of adult height data, we set $P_{\mathrm{g}} = 0.001$ (the default value in our software).

5. ASH prior. For each SNP j, we introduce a latent label $z_j \in [K]$ to denote the mixture component index for β_j in the ASH prior, that is, $\beta_j | z_j = k \sim \mathcal{N}(0, \sigma_k^2)$, $k \in [K]$, $j \in [p]$. Let $\boldsymbol{z} = (z_1, \ldots, z_p)^\intercal$. To fit the RSS model with ASH prior, we simulate the posterior draws of $(\boldsymbol{\beta}, \boldsymbol{z}, \boldsymbol{\omega}, \lambda)$ by the following component-wise MCMC procedure.

- Given $(\boldsymbol{\omega}, \lambda)$, we update $(\boldsymbol{\beta}, \boldsymbol{z})$ by a random-scan Gibbs step. For each iteration, we randomly pick a SNP j according to the distribution q_p and then draw (β_j, z_j) from its full conditional. Let $\boldsymbol{\theta}_{-j} := (\boldsymbol{\beta}_{-j}, \boldsymbol{z}_{-j}, \boldsymbol{\omega}, \lambda)$. We first sample z_j from $[K]$ and then sample β_j given z_j,

$$\frac{\mathsf{Pr}(z_j = k | \widehat{\boldsymbol{\beta}}, S, R, \boldsymbol{\theta}_{-j})}{\mathsf{Pr}(z_j = 1 | \widehat{\boldsymbol{\beta}}, S, R, \boldsymbol{\theta}_{-j})} = \sqrt{\frac{s_j^2 + \sigma_1^2}{s_j^2 + \sigma_k^2}} \cdot \frac{\exp\{b_j^2 s_j^4 (s_j^2 + \sigma_1^2)^{-1}/2\}}{\exp\{b_j^2 s_j^4 (s_j^2 + \sigma_k^2)^{-1}/2\}} \tag{5.1}$$

$$\beta_j | \widehat{\boldsymbol{\beta}}, S, R, z_j = k, \boldsymbol{\theta}_{-j} \sim \mathcal{N}\left(\frac{\sigma_k^2 s_j^2 b_j}{\sigma_k^2 + s_j^2}, \frac{\sigma_k^2 s_j^2}{\sigma_k^2 + s_j^2}\right), \; b_j := \frac{\hat{\beta}_j^2}{s_j^2} - \sum_{i \neq j} \frac{R_{ij}\beta_i}{s_i s_j} \tag{5.2}$$

  We leave the remaining components $(\boldsymbol{\beta}_{-j}, \boldsymbol{z}_{-j})$ unchanged in the current iteration.
- Given $(\boldsymbol{\beta}, \boldsymbol{z}, \lambda)$, we draw $\boldsymbol{\omega}$ from its full conditional

$$\boldsymbol{\omega} | \widehat{\boldsymbol{\beta}}, S, R, \boldsymbol{\beta}, \boldsymbol{z}, \lambda \sim \mathcal{D}(n_1(\boldsymbol{z}) + \lambda, \ldots, n_K(\boldsymbol{z}) + \lambda), \tag{5.3}$$

  where $n_k(\boldsymbol{z})$ is the number of entries in $\{j : z_j = k\}$ for each $k \in [K]$.
- Given $(\boldsymbol{\beta}, \boldsymbol{z}, \boldsymbol{\omega})$, we update λ by a Metropolis algorithm, where the proposal distribution is a symmetric Gaussian random walk on $\log(\lambda/(10-\lambda))$.

6. Small world proposal. To improve the convergence rate of the MCMC schemes, we use the "small-world" proposal (Guan and Krone, 2007) as an add-on for every Metropolis step in our main algorithms above. Specifically, with probability 0.3 in each iteration, a long-range move is made by compounding randomly many (from 2 to 20) local proposals.

References.

Guan, Y. and Krone, S. (2007). Small-world MCMC and convergence to multi-modal distributions: From slow mixing to fast mixing. *The Annals of Applied Probability* **17** 284–304.

Guan, Y. and Stephens, M. (2011). Bayesian variable selection regression for genome-wide association studies, and other large-scale problems. *The Annals of Applied Statistics* **5** 1780–1815.

Stephens, M. (2016). False discovery rates: A new deal. *bioRxiv*.

Wen, X. and Stephens, M. (2010). Using linear predictors to impute allele frequencies from summary or pooled genotype data. *The Annals of Applied Statistics* **4** 1158–1182.

Zhou, X., Carbonetto, P. and Stephens, M. (2013). Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genetics* **9** e1003264.

Zhu, X. and Stephens, M. (2016). Bayesian large-scale multiple regression with summary statistics from genome-wide association studies. *bioRxiv*.