Markov chain Monte Carlo (MCMC) for polygenic prediction

Jian Zeng

j.zeng@uq.edu.au

Bayesian multiple regression

$$\mathbf{y} = \mathbf{1}_{\mathrm{n}}\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

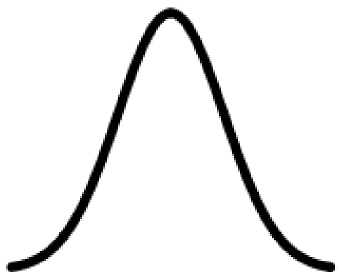
where

- **y** is a vector of *n* phenotypes,
- μ is the mean (can incorporate other covariates),
- **X** is the genotype matrix of *n* individuals for all SNPs,
- $\boldsymbol{\beta}$ is the vector of SNP effects,
- **e** is a vector of residuals, $\mathbf{e}$ ~ N(0, σ_e^2)

Bayesian methods mainly differ in the prior specification of SNP effects $\boldsymbol{\beta}$.

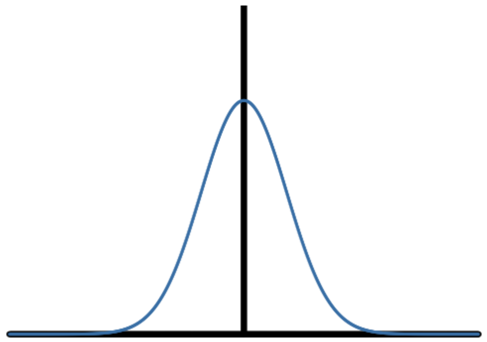
Prior distribution of SNP effects

Normal

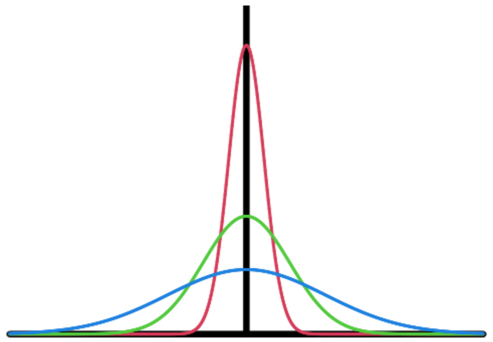


Zero-normal mixture

“Spike-and-slab” prior



Multi-component mixture



- We incorporate the prior knowledge into the estimation of SNP effects ($\boldsymbol{\beta}$) using **Bayes Theorem**:

$$P(\boldsymbol{\beta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\beta})P(\boldsymbol{\beta})$$

- Our target is to obtain the **joint posterior distribution** of the SNP effects and other model parameters ($\boldsymbol{\theta}$):

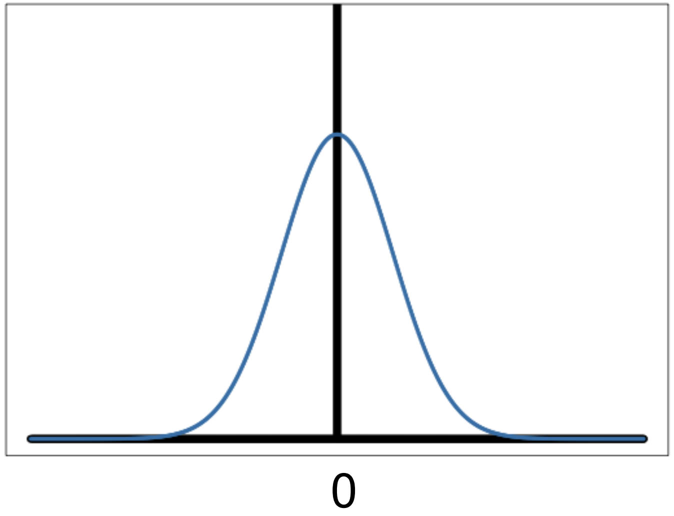
$$P(\boldsymbol{\theta}, \boldsymbol{\beta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\beta})P(\boldsymbol{\theta}, \boldsymbol{\beta})$$

- We can then use the posterior mean as a point estimate and the posterior variance to quantify uncertainty in estimation.

Spike-and-slab model (BayesC)

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

$$\beta_j \begin{cases} \sim N(0, \sigma_\beta^2) & \text{with probability } \pi \\ = 0 & \text{with probability } 1 - \pi \end{cases}$$



- The ‘slab’ component contains π, the probability that the SNP has nonzero effect (polygenicity).
- The ‘spike’ component contains $1 - \pi$, the probability that the SNP effect is zero.
- The variance σ_β^2 reflects the magnitude of effect sizes.

How to estimate the joint posterior distribution?

$$\hat{\boldsymbol{\beta}} = E(\boldsymbol{\beta}|\boldsymbol{y}) = \int_{\beta_1} \dots \int_{\beta_m} \beta_m (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{2\sigma_e^2}\right\} \prod_{j=1}^{m}\left[(\sigma_\beta^2)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\}\pi + \varphi_0(1-\pi)\right] \mathrm{d}\beta_m \dots \mathrm{d}\beta_1$$

Cannot solve the posterior distribution of $\boldsymbol{\beta}$ directly

Same issue for the other parameters $(\mu, \beta, \sigma_\beta^2, \pi, \sigma_e^2)$

Use Markov chain Monte Carlo (MCMC) algorithm!

Markov chain

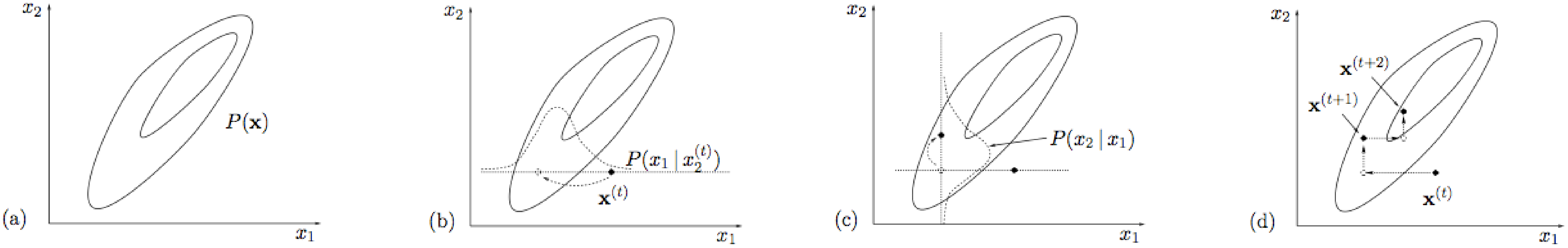
A sequence of samples where each sample depends only on the previous one (memoryless). This property allows the algorithm to gradually explore the distribution.

Monte Carlo

Using random sampling to perform numerical estimation, e.g., integrating over a probability distribution by averaging over samples.

Gibbs Sampling

A special case of MCMC to sample from posterior distribution of each parameter **conditional** on all other parameters.



The key is to derive $P(x_1|x_2)$ and $P(x_2|x_1)$

Figure source

To run Gibbs sampling, we need to derive the full conditional distribution for each parameter

- $P\left(\mu \middle| \boldsymbol{y}, \boldsymbol{\beta}, \sigma_\beta^2, \pi, \sigma_e^2\right)$
- $P\left(\beta_j \middle| \boldsymbol{y}, \boldsymbol{\beta}_{-j}, \sigma_\beta^2, \pi, \sigma_e^2\right)$
- $P\left(\sigma_\beta^2 \middle| \boldsymbol{y}, \boldsymbol{\beta}, \pi, \sigma_e^2\right)$
- $P\left(\pi \middle| \boldsymbol{y}, \boldsymbol{\beta}, \sigma_\beta^2, \sigma_e^2\right)$
- $P\left(\sigma_e^2 \middle| \boldsymbol{y}, \boldsymbol{\beta}, \sigma_\beta^2, \pi\right)$

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

$$\beta_j \begin{cases} \sim N(0, \sigma_\beta^2) & \text{with probability } \pi \\ = 0 & \text{with probability } 1 - \pi \end{cases}$$

Posterior joint distribution

Scaled inverse chi-square distribution

$$P\left(\mu,\boldsymbol{\beta},\sigma_\beta^2,\pi,\sigma_e^2\middle|\boldsymbol{y}\right)$$

$$\propto P\left(\boldsymbol{y}\middle|\mu,\boldsymbol{\beta},\sigma_\beta^2,\pi,\sigma_e^2\right)P(\mu)P\left(\boldsymbol{\beta}\middle|\sigma_\beta^2,\pi\right)P\left(\sigma_\beta^2\right)P(\pi)P\left(\sigma_e^2\right)$$

Likelihood

Flat prior

Beta distribution

Point-normal mixture

These priors are chosen for mathematical convenience, such that the full conditional distribution belongs to the same family as the prior (known as **conjugate priors**).

Posterior joint distribution

$$P\left(\mu, \boldsymbol{\beta}, \sigma_\beta^2, \pi, \sigma_e^2 \middle| \boldsymbol{y}\right)$$

$$\propto P\left(\boldsymbol{y} \middle| \mu, \boldsymbol{\beta}, \sigma_\beta^2, \pi, \sigma_e^2\right) P(\mu) P\left(\boldsymbol{\beta} \middle| \sigma_\beta^2, \pi\right) P\left(\sigma_\beta^2\right) P(\pi) P\left(\sigma_e^2\right)$$

$$\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{\left(\mathbf{y}-\mathbf{1}\mu-\sum_j \mathbf{X}_j\beta_j\right)'\left(\mathbf{y}-\mathbf{1}\mu-\sum_j \mathbf{X}_j\beta_j\right)}{2\sigma_e^2}\right\}$$

← Likelihood

$$\times \prod_{j=1}^{m} \left[\left(\sigma_\beta^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\} \pi + \varphi_0 (1-\pi) \right]$$

← Prior for $\boldsymbol{\beta}$: point-normal mixture

$$\times \left(\sigma_\beta^2\right)^{-\frac{\upsilon_\beta+2}{2}} exp\left\{-\frac{\upsilon_\beta \tau_\beta^2}{2\sigma_\beta^2}\right\}$$

← Prior for σ_β^2 : scaled inverse chi-square distribution

$$\times \left(\sigma_e^2\right)^{-\frac{\upsilon_e+2}{2}} exp\left\{-\frac{\upsilon_e \tau_e^2}{2\sigma_e^2}\right\}$$

← Prior for σ_e^2 : scaled inverse chi-square distribution

$$\times \pi^{a-1} (1-\pi)^{b-1}$$

← Prior for π : beta distribution

Full conditional distribution for μ

$$P\left(\mu, \boldsymbol{\beta}, \sigma_\beta^2, \pi, \sigma_e^2 \middle| \boldsymbol{y}\right)$$

$$\propto P\left(\boldsymbol{y} \middle| \mu, \boldsymbol{\beta}, \sigma_\beta^2, \pi, \sigma_e^2\right) P(\mu) P\left(\boldsymbol{\beta} \middle| \sigma_\beta^2, \pi\right) P\left(\sigma_\beta^2\right) P(\pi) P(\sigma_e^2)$$

$$\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j \beta_j\right)'\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j \beta_j\right)}{2\sigma_e^2}\right\}$$

$$\times \prod_{j=1}^{m} \left[\left(\sigma_\beta^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\} \pi + \varphi_0 (1 - \pi) \right]$$

$$\times \left(\sigma_\beta^2\right)^{-\frac{\upsilon_\beta + 2}{2}} exp\left\{-\frac{\upsilon_\beta \tau_\beta^2}{2\sigma_\beta^2}\right\}$$

$$\times (\sigma_e^2)^{-\frac{\upsilon_e + 2}{2}} exp\left\{-\frac{\upsilon_e \tau_e^2}{2\sigma_e^2}\right\}$$

$$\times \pi^{a-1} (1 - \pi)^{b-1}$$

Full conditional distribution for μ

$$P\left(\mu \middle| \boldsymbol{y}, \boldsymbol{\beta}, \sigma_\beta^2, \pi, \sigma_e^2\right)$$

$$\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j \beta_j\right)'\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j \beta_j\right)}{2\sigma_e^2}\right\}$$

$$\sim N\left(\frac{\mathbf{1}'\left(\mathbf{y} - \sum_j \mathbf{X}_j \beta_j\right)}{n}, \frac{\sigma_e^2}{n}\right)$$

Full conditional distribution for β_j

$$P\left(\mu, \boldsymbol{\beta}, \sigma_\beta^2, \pi, \sigma_e^2 \middle| \boldsymbol{y}\right)$$

$$\propto P\left(\boldsymbol{y} \middle| \mu, \boldsymbol{\beta}, \sigma_\beta^2, \pi, \sigma_e^2\right) P(\mu) P\left(\boldsymbol{\beta} \middle| \sigma_\beta^2, \pi\right) P\left(\sigma_\beta^2\right) P(\pi) P(\sigma_e^2)$$

$$\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j \beta_j\right)'\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j \beta_j\right)}{2\sigma_e^2}\right\}$$

$$\times \prod_{j=1}^{m} \left[\left(\sigma_\beta^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\} \pi + \varphi_0(1 - \pi)\right]$$

$$\times \left(\sigma_\beta^2\right)^{-\frac{\upsilon_\beta + 2}{2}} exp\left\{-\frac{\upsilon_\beta \tau_\beta^2}{2\sigma_\beta^2}\right\}$$

$$\times (\sigma_e^2)^{-\frac{\upsilon_e + 2}{2}} exp\left\{-\frac{\upsilon_e \tau_e^2}{2\sigma_e^2}\right\}$$

$$\times \pi^{a-1} (1 - \pi)^{b-1}$$

Full conditional distribution for β_j

$$P\left(\beta_j \middle| \boldsymbol{y}, \boldsymbol{\beta}_{-j}, \sigma_\beta^2, \pi, \sigma_e^2\right)$$

$$\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{\left(\mathbf{y}-\mathbf{1}\mu-\sum_j \mathbf{X}_j\beta_j\right)'\left(\mathbf{y}-\mathbf{1}\mu-\sum_j \mathbf{X}_j\beta_j\right)}{2\sigma_e^2}\right\}$$

$$\times \left(\sigma_\beta^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\}\pi + \varphi_0(1-\pi)$$

Let’s introduce an indicator variable δ_j

If $\delta_j = 0,$ then $\beta_j = 0$
If $\delta_j = 1,$ then β_j is in non-zero component

Full conditional distribution for β_j

If $\delta_j = 1$

$$P\left(\beta_j \middle| \boldsymbol{y}, \delta_j = 1, \boldsymbol{\beta}_{-j}, \sigma_\beta^2, \pi, \sigma_e^2\right)$$

$$\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j\beta_j\right)'\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j\beta_j\right)}{2\sigma_e^2}\right\} \times \left(\sigma_\beta^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\}$$

$$\sim N\left(\frac{\mathbf{X}_j'(\mathbf{y} - \mathbf{1}\mu - \sum_{k\neq j} \mathbf{X}_k' \beta_k)}{\mathbf{X}_j'\mathbf{X}_j + \sigma_e^2/\sigma_\beta^2}, \frac{\sigma_e^2}{\mathbf{X}_j'\mathbf{X}_j + \sigma_e^2/\sigma_\beta^2}\right)$$

Full conditional distribution for δ_j

To improve mixing, we sample δ_j unconditional on β_j

$$P(\delta_j|\boldsymbol{y}, else) = \int_{\beta_j} P(\delta_j, \beta_j|\boldsymbol{y}, else)\mathrm{d}\beta_j$$

In other words, we sample δ_j and β_j jointly:

$$P(\beta_j, \delta_j|\boldsymbol{y}, else) \propto P(\beta_j|\delta_j, \boldsymbol{y}, else)P(\delta_j|\boldsymbol{y}, else)$$

Full conditional distribution for δ_j

$$P(\delta_j|\boldsymbol{y}, else) = \int_{\beta_j} P(\delta_j, \beta_j|\boldsymbol{y}, else)\mathrm{d}\beta_j = \int_{\beta_j} P(\boldsymbol{y}|\delta_j, \beta_j, else)P(\beta_j)P(\delta_j)\mathrm{d}\beta_j$$

$$= P(\boldsymbol{y}|\delta_j, else)P(\delta_j)$$

Therefore,

$$\Pr(\delta_j = 1|\boldsymbol{y}, else) = \frac{P(\boldsymbol{y}|\delta_j = 1, else)P(\delta_j = 1)}{P(\boldsymbol{y}|\delta_j = 0, else)P(\delta_j = 0) + P(\boldsymbol{y}|\delta_j = 1, else)P(\delta_j = 1)}$$

Full conditional distribution for σ_β^2

$$P\left(\mu,\boldsymbol{\beta},\sigma_\beta^2,\pi,\sigma_e^2\middle|\boldsymbol{y}\right)$$

$$\propto P\left(\boldsymbol{y}\middle|\mu,\boldsymbol{\beta},\sigma_\beta^2,\pi,\sigma_e^2\right)P(\mu)P\left(\boldsymbol{\beta}\middle|\sigma_\beta^2,\pi\right)P\left(\sigma_\beta^2\right)P(\pi)P\left(\sigma_e^2\right)$$

$$\propto (\sigma_e^2)^{-\frac{n}{2}}\exp\left\{-\frac{\left(\mathbf{y}-\mathbf{1}\mu-\sum_j\mathbf{X}_j\beta_j\right)'\left(\mathbf{y}-\mathbf{1}\mu-\sum_j\mathbf{X}_j\beta_j\right)}{2\sigma_e^2}\right\}$$

$$\times\prod_{j=1}^{m}\left[\left(\sigma_\beta^2\right)^{-\frac{1}{2}}\exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\}\pi+\varphi_0(1-\pi)\right]$$

$$\times\left(\sigma_\beta^2\right)^{-\frac{\upsilon_\beta+2}{2}}exp\left\{-\frac{\upsilon_\beta\tau_\beta^2}{2\sigma_\beta^2}\right\}$$

$$\times(\sigma_e^2)^{-\frac{\upsilon_e+2}{2}}exp\left\{-\frac{\upsilon_e\tau_e^2}{2\sigma_e^2}\right\}$$

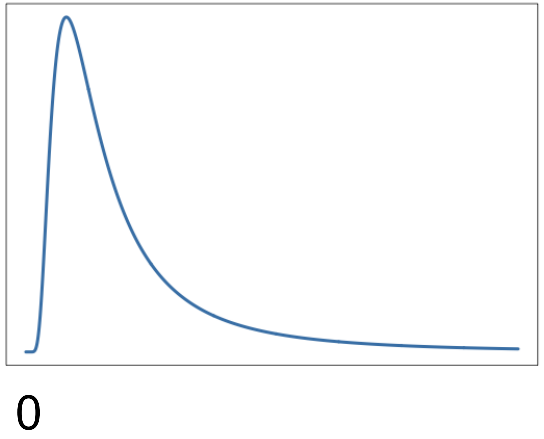
$$\times\pi^{a-1}(1-\pi)^{b-1}$$

Full conditional distribution for σ_β^2

$$P\left(\sigma_\beta^2 \middle| \boldsymbol{y}, \boldsymbol{\beta}, \pi, \sigma_e^2\right)$$

$$\propto \prod_{j=1}^{m}\left[\left(\sigma_\beta^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\}\right]^{\delta_j} \times \left(\sigma_\beta^2\right)^{-\frac{\upsilon_\beta+2}{2}} exp\left\{-\frac{\upsilon_\beta \tau_\beta^2}{2\sigma_\beta^2}\right\}$$

$$\sim \chi^{-2}\left(\tilde{\upsilon}_\beta = \upsilon_\beta + \sum_j \delta_j \,, \tilde{\tau}_\beta^2 = \frac{\sum_j \beta_j^2 + \upsilon_\beta \tau_\beta^2}{\tilde{\upsilon}_\beta}\right)$$



Full conditional distribution for π

$$P\left(\mu, \boldsymbol{\beta}, \sigma_\beta^2, \pi, \sigma_e^2 \middle| \boldsymbol{y}\right)$$

$$\propto P\left(\boldsymbol{y} \middle| \mu, \boldsymbol{\beta}, \sigma_\beta^2, \pi, \sigma_e^2\right) P(\mu) P\left(\boldsymbol{\beta} \middle| \sigma_\beta^2, \pi\right) P\left(\sigma_\beta^2\right) P(\pi) P\left(\sigma_e^2\right)$$

$$\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j\beta_j\right)'\left(\mathbf{y} - \mathbf{1}\mu - \sum_j \mathbf{X}_j\beta_j\right)}{2\sigma_e^2}\right\}$$

$$\times \prod_{j=1}^{m} \left[\left(\sigma_\beta^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{\beta_j^2}{2\sigma_\beta^2}\right\} \pi + \varphi_0(1-\pi)\right] \longrightarrow$$

Only depends on the indicator variable δ_j

$$\prod_{j=1}^{m} \left[\pi^{\delta_j} + (1-\pi)^{(1-\delta_j)}\right]$$

$$\times \left(\sigma_\beta^2\right)^{-\frac{v_\beta+2}{2}} exp\left\{-\frac{v_\beta \tau_\beta^2}{2\sigma_\beta^2}\right\}$$

$$\times (\sigma_e^2)^{-\frac{v_e+2}{2}} exp\left\{-\frac{v_e \tau_e^2}{2\sigma_e^2}\right\}$$

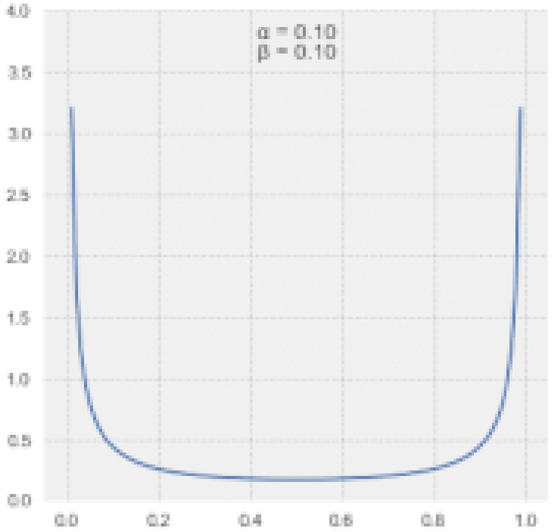
$$\times \pi^{a-1}(1-\pi)^{b-1}$$

Full conditional distribution for π

$$P\left(\pi \middle| \boldsymbol{y}, \boldsymbol{\beta}, \sigma_\beta^2, \sigma_e^2\right)$$

$$\propto \prod_{j=1}^{m} \left[\pi^{\delta_j} + \ (1-\pi)^{(1-\delta_j)}\right] \times \pi^{a-1}(1-\pi)^{b-1}$$

$$\sim Beta\left(a + \sum\nolimits_j \delta_j \ , b + m - \sum\nolimits_j \delta_j\right)$$



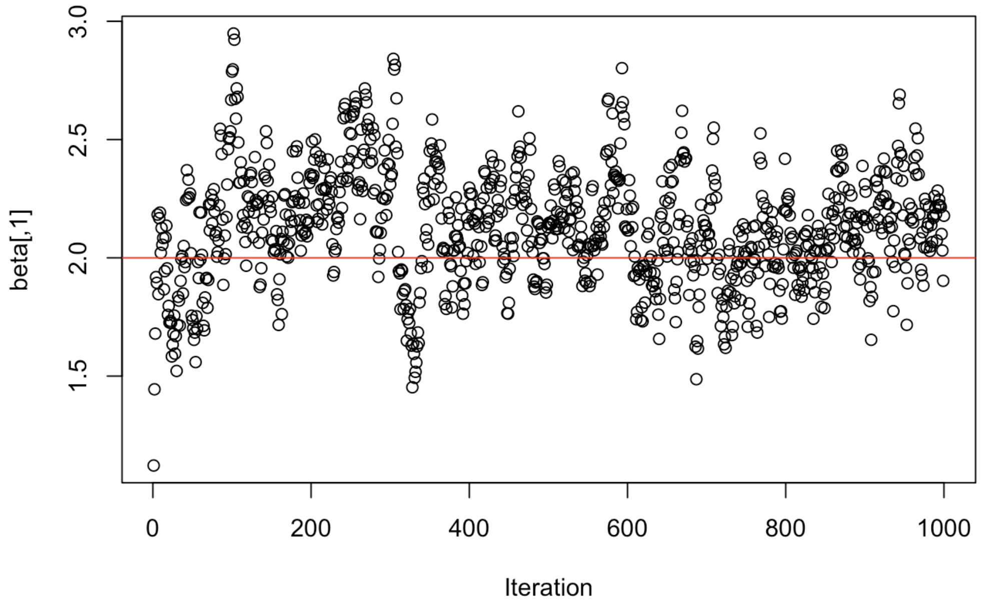
Gibbs sampling

- Set starting values for (μ, $\boldsymbol{\delta}$, $\boldsymbol{\beta}$, σ_β^2, π, σ_e^2)
- Then (for many iterations)
  - For each SNP, sample δ_j, β_j conditional on other parameters
  - Sample μ, σ_β^2, π, σ_e^2 with updated $\boldsymbol{\delta}$, $\boldsymbol{\beta}$

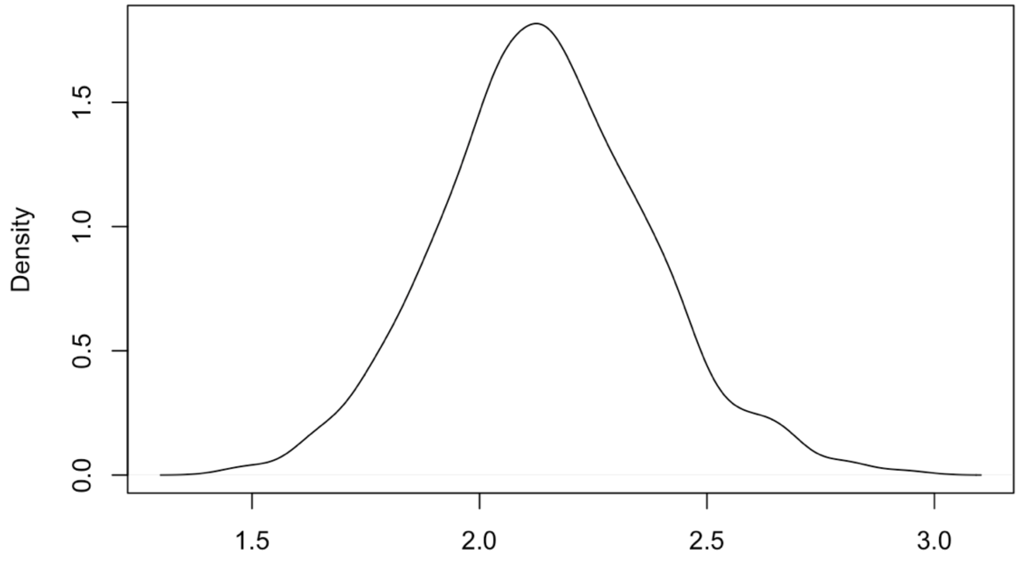
Samples reconstruct posterior distributions of parameters

Gibbs sampling

Trace plot



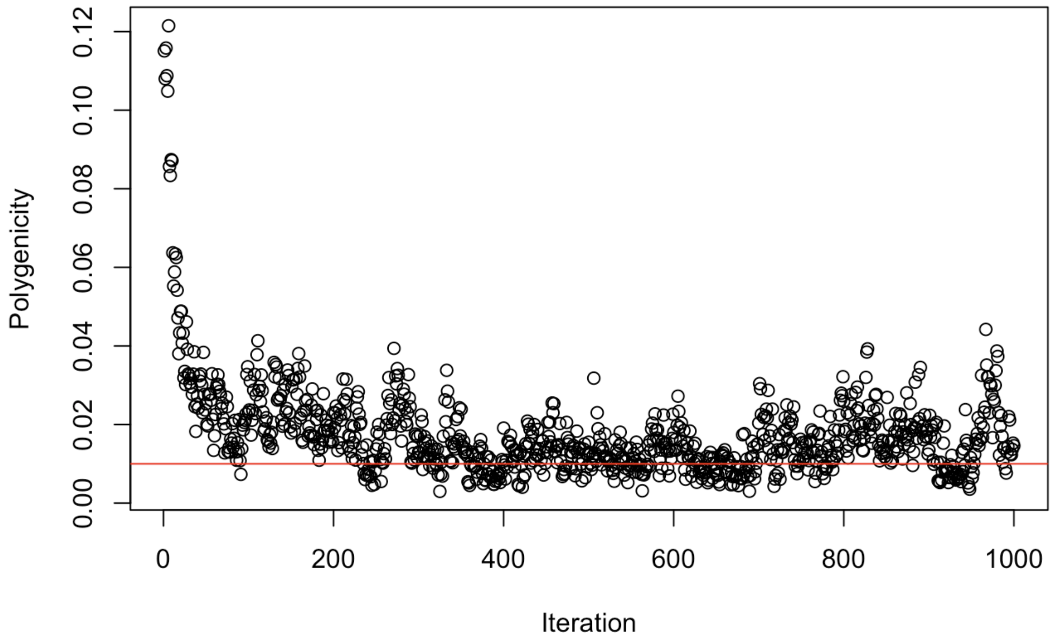
Posterior distribution

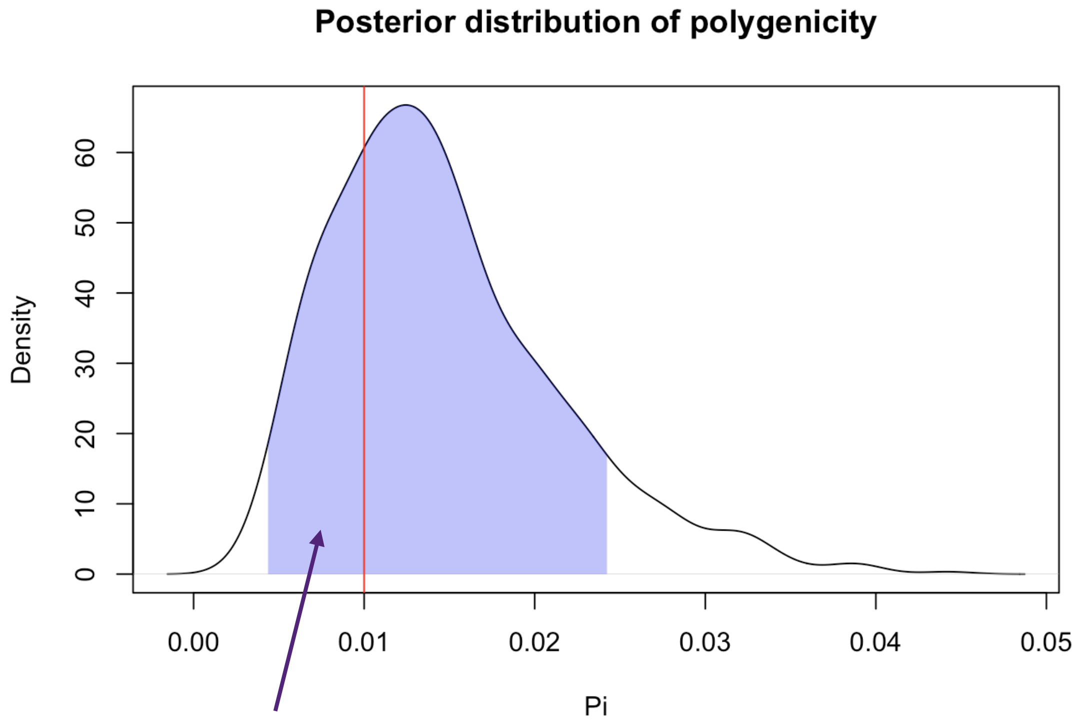


Posterior mean is used as the point estimate of the SNP effect

Gibbs sampling

Trace plot





90% highest probability density (HPD) as credible interval

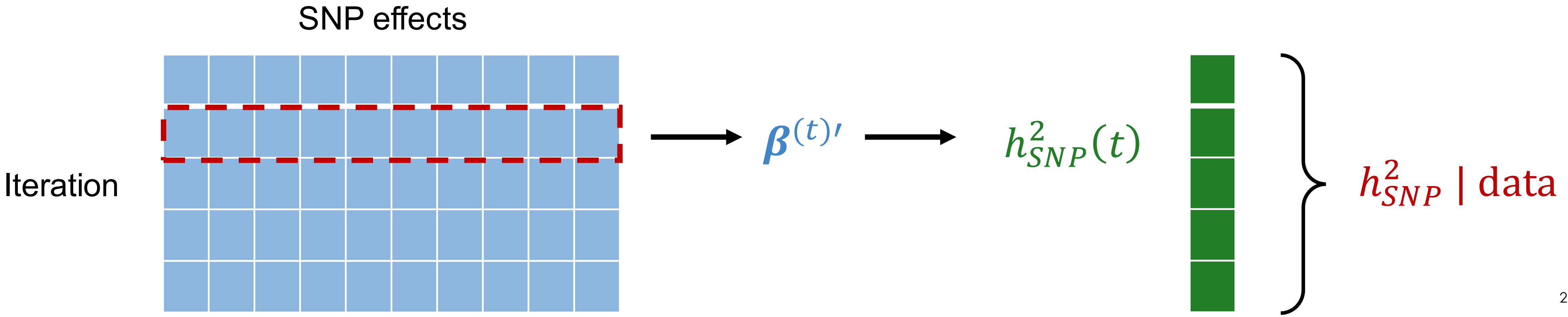
Estimation of SNP-based heritability

By definition,

$$h^2_{SNP} = \frac{Var(\mathbf{X}\boldsymbol{\beta})}{Var(\mathbf{X}\boldsymbol{\beta}) + \sigma^{\mathbf{2}}_{\boldsymbol{e}}}$$

Of course, we don't know the true $\boldsymbol{\beta}$. However, from MCMC sampling, we "observed" the sampled values of $\boldsymbol{\beta}$ from its joint posterior distribution. In each iteration (*t*),

$$h^2_{SNP}(t) = \frac{Var(\boldsymbol{X\beta}^{(t)})}{Var(\boldsymbol{X\beta}^{(t)}) + \sigma_{\boldsymbol{e}}^{\mathbf{2},(t)}}$$



Estimation of SNP-based heritability

By definition,

$$h^2_{SNP} = \frac{Var(\mathbf{X}\boldsymbol{\beta})}{Var(\mathbf{X}\boldsymbol{\beta}) + \sigma^{\mathbf{2}}_{\boldsymbol{e}}}$$

Of course, we don't know the true $\boldsymbol{\beta}$. However, from MCMC sampling, we "observed" the sampled values of $\boldsymbol{\beta}$ from its joint posterior distribution. In each iteration (*t*),

$$h^2_{SNP}(t) = \frac{Var(\boldsymbol{X}\boldsymbol{\beta}^{(t)})}{Var(\boldsymbol{X}\boldsymbol{\beta}^{(t)}) + \sigma_{\boldsymbol{e}}^{\mathbf{2},(t)}}$$

Over a total of *T* MCMC samples, the posterior mean provides the point estimate:

$$E[h^2_{SNP}] = \frac{1}{T}\sum_{t=1}^{T} h^2_{SNP}(t)$$

Summary

- Markov chain Monte Carlo (MCMC) is a technique to draw samples from a posterior distribution for Bayesian inference of model parameters.
- MCMC sampling allows us to observe the parameter values at each iteration – we can derive any function of them and compute quantities that are not explicitly modeled, e.g., h_{SNP}^2.
- It's important to assess the convergence of MCMC by running a longer chain or multiple chains and examine the changes in prediction accuracy.

Recommended reading

1. Habier D, *et al*. Extension of the bayesian alphabet for genomic selection. BMC *Bioinformatics*. 2011 May 23;12:186. (**BayesC**)
2. Lloyd-Jones LR, *et al*. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat Commun* 10, 5086 (2019). (**SBayesR; MCMC details in Supplementary**)