

Bayesian QTL Mapping

Fei Zou

Department of Biostatistics

Gillings School of Global Public Health

University of North Carolina-Chapel Hill

fzou@bios.unc.edu

Bayesian QTL Mapping with Variable Selection

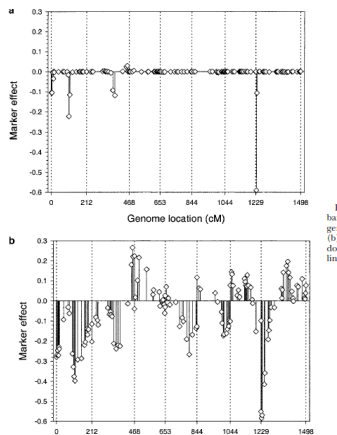
- When $p \gg n$, it might be more desirable to shrink most of the parameters towards 0.
- Let's consider the following model again
$$y_i = \mu + \sum_{j=1}^p \beta_j x_{ij} + e_i$$
 where $e_i \sim N(0, \sigma^2)$
with the following priors
 - $p(\mu) \propto 1$
 - $\sigma^2 \propto \frac{1}{\sigma^2}$
 - $\beta_j \propto N(0, \sigma_j^2)$
 - $\sigma_j^2 \propto \frac{1}{\sigma_j^2}$
- Note: improper priors used for μ, σ^2 and σ_j^2 s.

Bayesian QTL Mapping with Variable Selection

- conditional probabilities of the parameters are now
 - $\mu \mid (\{y_i\}, \sigma^2, \{\sigma_j^2\}, \{\beta_j\}) \sim N(\frac{1}{n} \sum_i (y_i - \sum_j \beta_j x_{ij}), \frac{1}{n} \sigma^2)$
 - $\sigma^2 \mid (\{y_i\}, \mu, \{\sigma_j^2\}, \{\beta_j\}) \sim$ scaled inverted- χ_n^2 and similarly the conditional probability of σ_j^2 follows a scaled inverted- χ_1^2 .
 - $\beta_j \mid (\{y_i\}, \mu, \sigma^2, \{\sigma_k^2\}_{k \neq j}, \{\beta_j\}) \sim N(\bar{\beta}_j, s_j^2)$ [**exercise**: find $\bar{\beta}_j$ and s_j^2].
- All parameters can be sampled via Gibbs sampler: Great!

Bayesian QTL Mapping

- Shrinkage phenomena is clear and seems to work reasonably well for QTL mapping



Bayesian QTL Mapping

- Note: improper priors used for μ, σ^2 and σ_j^2 s.
- Hobert and Casella (1996 JASA) showed that even when all conditional distributions are proper, the posterior may not be proper!

Bayesian QTL Mapping

- Example: consider the following simple mixed effect model $y_{ij} = \mu + u_i + \epsilon_{ij}$, $i = 1, \dots, k$ $j = 1, \dots, J$ where u_i s (random effects) are iid $N(0, \sigma_u^2)$ and ϵ_{ij} s are iid $N(0, \sigma^2)$. Let $\theta = (\mu, \{u_j\}, \sigma^2, \sigma_u^2)$
 - $\mu \sim \pi(\mu) \propto 1$; $\sigma^2 \sim \pi(\sigma^2) \propto \frac{1}{\sigma^2}$
 - $u_j \sim N(0, \sigma_u^2)$ $\sigma_u^2 \sim \pi(\sigma_u^2) \propto \frac{1}{\sigma_u^2}$
- All conditional distributions are proper:
 - $p(\mu | \{u_j\}, \sigma^2, \sigma_u^2, \{y_{ij}\}) \sim N(\frac{1}{n} \sum_{i,j} (y_{ij} - u_j), \frac{1}{n} \sigma^2)$
 - $p(u_j | \{u_k\}_{k \neq j}, \sigma^2, \sigma_u^2, \{y_{ij}\}) \sim N(\frac{\sum_j (y_{ij} - \mu)}{1/\sigma_u^2 + n/\sigma^2}, \frac{1}{1/\sigma_u^2 + n/\sigma^2})$
 - similarly $p(\sigma^2 | \mu, \{u_j\}, \sigma_u^2, \{y_{ij}\})$ and $p(\sigma_u^2 | \mu, \{u_j\}, \sigma^2, \{y_{ij}\})$ can be shown to be proper [**exercise**: find the conditional densities of σ^2 and σ_u^2].
- However, $\int p(\mu, \{u_j\}, \sigma^2, \sigma_u^2 | \{y_{ij}\}) d\theta = \infty$ which is not proper!!!! [**exercise**: prove this claim]
- Be careful when using improper priors

Bayesian QTL Mapping

- Note: ter Braak et al (2005 Genetics) showed that the posterior showed about for QTL is not proper either
- The posterior has two modes for β_j s, one (infinity mass) at 0 and another at the true parameter value. The method works well likely due to the fact that all MCMC samples are trapped around the 2nd mode.
- ter Braak et al (2005) suggested the following prior modification to ensure a proper posterior:
 - $\sigma^2 \propto (\sigma^2)^{-1+\delta}$
 - $\sigma_j^2 \propto (\sigma_j^2)^{-1+\delta}$

which yields a proper posterior for the QTL effect when $0 < \delta \leq 1/2$.

Bayesian QTL Mapping

- Thresholding must be employed in the Bayesian method discussed in previous lecture for variable selection purpose
- Alternatively, may employ Composite Space representation of the QTL model (Yi 2004 Genetics)
 - Re-represent the linear regression model as
$$y_i = \mu + \sum_{j=1}^p \gamma_j \beta_j x_{ij} + e_i$$
where $\gamma_j = 1$ or 0 depending on whether marker j is a QTL or not. Doing this way, the number of total QTL equals $\sum_{j=1}^p \gamma_j$.
 - Priors:
 - 1) $p(\gamma_j) = \omega^{\gamma_j} (1 - \omega)^{1 - \gamma_j}$ where ω is pre-specified, such as 1/2
 - 2) $\mu \sim N(\mu_0, \kappa_0^2)$.
 - 3) $\beta_j \sim N(0, \kappa^2)$ or $\beta_j | \gamma_j \sim (1 - \gamma_j)N(0, \kappa^2) + \gamma_j N(0, c\kappa^2)$ where c is a predetermined large number. The later has been used by George and McCulloch (1993) and Dellaportas et al. (2002) for a linear regression model.
 - 4) $\sigma^2 \sim \text{Inverse-Gamma}(u, v)$

Extension to Gene-gene interaction

- in all models discussed thus far, genetics effects from multiple QTL are treated additive (i.e. no gene-gene interactions)
- Genes may interact with each other, i.e. epistasis effect

Example:

		QTL1	
		AA	AB
QTL2	AA	1	2
	AB	2	1

- Problem: both genes would be missed by discussed methods
- Solution: split data into groups based on genotypes of one QTL and do stratified QTL mapping within each group, which requires prior knowledge which gene interacts with other genes
 - challenging problem.
- Question: for p putative genes, how many possible two way interactions? how many three way interactions?

Multiple QTL Mapping

- Model with gene-gene interaction

$$y_i = \mu + \sum_j \beta_j x_{ij} + \sum_{j \neq k} \beta_{jk} x_{ij} x_{ik} + e_i$$

- similar Bayesian approaches can be developed but the number of two-way interaction terms increases dramatically

- Model with other non-genetic covariates

- Additive model:

$$y_i = \mu + \sum_j \beta_j x_{ij} + \sum_k \xi_k z_{ik} + e_i \text{ or}$$

- Interaction model:

$$y_i = \mu + \sum_j \beta_j x_{ij} + \sum_k \xi_k z_{ik} + \sum_{jk} \beta_{jk} x_{ij} z_{ik} + e_i$$