

Note

Extending Xu's Bayesian Model for Estimating Polygenic Effects Using Markers of the Entire Genome

Cajo J. F. ter Braak,¹ Martin P. Boer and Marco C. A. M. Bink

Biometris, Wageningen University and Research Centre, Wageningen, The Netherlands, 6700 AC

Manuscript received January 4, 2005
Accepted for publication April 4, 2005

ABSTRACT

In 2003, Xu obtained remarkably precise estimates of QTL positions despite the many markers simultaneously in his Bayesian model. We extend his model and Gibbs algorithm to ensure a valid posterior distribution and convergence to it, without changing the attractiveness of the method.

STIMULATED by MEUWISSEN *et al.* (2001), XU (2003) presented a Bayesian model for marker analysis in which each marker effect has its own specific variance with, on the log scale, a flat prior for each variance,

$$y_i = b_0 + \sum_{j=1}^p b_j x_{ij} + e_i,$$

where y_i denotes the phenotypic value for the i th individual ($i = 1, \dots, n$), b_0 is the population mean, b_j is the QTL effect associated with the j th marker ($j = 1, \dots, p$), x_{ij} is a dummy variable indicating genotype, and $e_i \sim N(0, \sigma_0^2)$ is the residual error. In a backcross (BC) population, the variable x_{ij} is either 1 or -1 , depending on whether the i th individual is homozygous or heterozygous. The priors of XU (2003) are $p(b_0) \propto 1$, $p(\sigma_0^2) \sim 1/\sigma_0^2$, $b_j \sim N(0, \sigma_j^2)$, and $p(\sigma_j^2) \propto 1/\sigma_j^2$; thus $p(\log(\sigma_j^2)) \propto 1$. This is a daring model because O'HAGAN (1994, Sect. 9.61) and GELMAN *et al.* (2004) warn that such priors yield improper (*i.e.*, invalid) posterior distributions in variance components models. Nevertheless, in simulations XU's (2003) Gibbs algorithm did very well in yielding precise estimates of the QTL effects and positions. The Bayesian estimates of the marker effects were all near zero for markers that did not coincide with a QTL. We were concerned that this might be an artifact of the Gibbs algorithm, which is prone to converge slowly for correlated parameters.

For illustration of the power of the method, Figure 1a shows results from our C++ implementation for the simulated BC population of Figure 6 of XU (2003), where $p = 301$, except that we used $n = 200$ instead of

300. The positions of the first six QTL stand out clearly with the lines with solid circles representing the effect means—the fifth QTL being spread over the interval [590–615]—whereas the remaining five QTL that each explain 0.6% of the phenotypic variance are barely visible, except perhaps for the tenth QTL. The 95%-credible intervals cover the true effects for the first six QTL, although only the first three exclude zero. The mean, 2.5%, and 97.5% points for markers not coinciding with QTL are typically near zero (<0.005 in absolute value). The estimates near 450 cM illustrate the danger of non-convergence. The peak is at 455 cM, close to the true QTL position. However, the LOD profile for these data (Figure 1b), calculated using a one-QTL model (DUPUIS and SIEGMUND 1999) and phenotypic data from which the true effects of the other QTL are subtracted, peaks at 450 cM, the scores at 445, 450, 455, and 460 cM being 7.1, 10.4, 9.5, and 7.2. The 1.5-LOD interval (DUPUIS and SIEGMUND 1999) includes both 450 and 455 cM. As Figure 1a contains little probability mass at 450 cM, it thereby suggests a more precise QTL position than warranted.

To ensure a valid posterior we extend the prior to $p(\sigma_j^2) \propto (\sigma_j^2)^{-1+\delta}$, which yields a proper posterior for the QTL effect for $0 < \delta \leq 1/2$ (APPENDIX). Xu's prior ($\delta = 0$) is just excluded, because it yields a posterior of b_j with infinite mass near zero. The new prior requires a change in step 5 in Xu's Gibbs algorithm. In the notation of GELMAN *et al.* (2004), the new σ_j^2 must be sampled from $\text{inv } \chi^2(1 - 2\delta, b_j^2/(1 - 2\delta))$; that is, new $\sigma_j^2 = b_j^2 / \chi_{1-2\delta}^2$ with $\chi_{1-2\delta}^2$ a random number sampled from a chi-square distribution with $(1 - 2\delta)$ d.f. Another solution is to use an $\text{inv } \chi^2(\nu, s^2)$ prior (MEUWISSEN *et al.* 2001), yielding a valid posterior for positive ν and s^2 . However, with this approach we found that the 2.5

¹Corresponding author: Biometris, Wageningen University and Research Centre, Bornsesteeg 47, 6708 PD Wageningen, The Netherlands. E-mail: cajo.terbraak@wur.nl

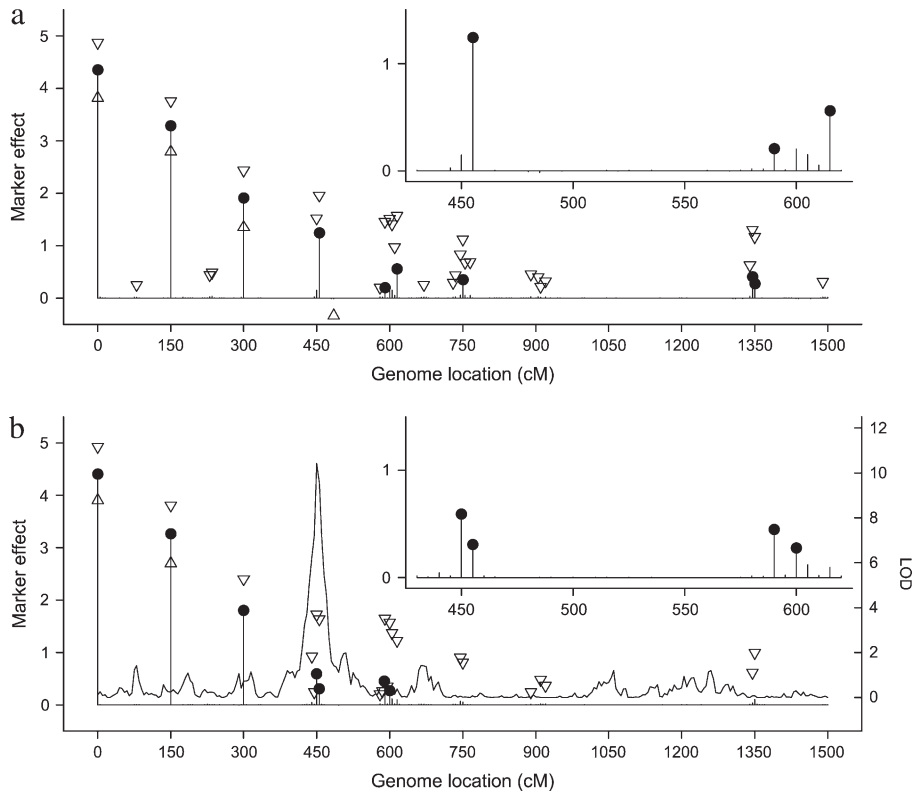


FIGURE 1.—Bayesian estimates of marker effects in a simulated BC family with 11 QTL every 150 cM (QTL effects 4.472, 3.162, 2.236, 1.581, 1.118, 0.7905, and, for QTL 7–11, 0.559 and $\sigma_j^2 = 10$). (a) Original analysis ($\delta = 0$); (b) reanalysis with $\delta = 0.001$ and swap step 5b to improve mixing. The lines with solid circles indicate posterior means, and triangles and inverted triangles indicate 2.5 and 97.5 percentiles, respectively. Symbols for values < 0.2 are not plotted to avoid cluttering. The profile (b) is LOD for the QTL at 450 cM. The insets zoom in at the interval 425–625 cM.

and 97.5% points for markers not coinciding with QTL were typically > 100 times larger than those with Xu's method, even for small values such as $\nu = s^2 = 10^{-4}$.

We examined the danger of slow convergence (bad mixing) by adding a Metropolis step as step 5b in XU (2003) in which σ_j^2 and σ_{j+1}^2 are swapped ($j = 1, \dots, p - 1$) with an acceptance probability (A5). This step is not conditional on b_j and b_{j+1} as these are integrated out (BROWN *et al.* 2002) and ensures appropriate mixing of close marker variances and, thereby, of close marker effects.

Figure 1b shows the results obtained for $\delta = 0.001$ with the swap step. As in Figure 1a the major QTL stand out and the markers not coinciding with QTL have near zero means and percentage points. The improved mixing is visible near 450 cM where now the largest effect is where the LOD peaks instead of being at 455 cM. Also the peak at 615 cM in Figure 1a has decreased in importance in favor of the peak at 590 cM in Figure 1b, in agreement with the LOD profile for this QTL (not shown). In a one-QTL Bayesian model with $p = 1$ and fixed $\sigma_1^2 = 10$, (A6) in the APPENDIX is proportional to the posterior density of a QTL at the location of the marker. The posterior distribution and credible interval for a single QTL can thus be obtained by calculating (A6) at all markers and normalizing the obtained values to a sum of 1. We use this one-QTL model for its analytical tractability. We expect the multiple-QTL model to give credible intervals that are somewhat larger than those of the one-QTL model for phenotypic data from

which the true effects of the other QTL are subtracted. As 95% credible intervals we thus obtained for QTL 4–6 and 10 [450, 455], [590, 610], [150, 1330], and [75, 1425]. It is thus no surprise that QTL 6 and 10 do not stand out in Figure 1b. In further simulations we found that the swap step often has little effect on the results, showing that the original Gibbs algorithm mixes reasonably well. We view the swap step as an extra safeguard against bad mixing of consecutive σ_j^2 's. In addition, parameter expansion (GELMAN *et al.* 2004) could be beneficial for the mixing for each (b_j, σ_j^2) pair.

In this all-marker model, mixing has the drawback that the size of the QTL effect is shrunken compared to that of the unmixed case. The sum of b_j 's at 450 and 455 cM has a posterior mean of 1.39 in Figure 1a and a mean of 0.89 in Figure 1b, whereas the true QTL effect is 1.58. The reason is that the model shrinks each effect more the smaller its t -value is. For a better size estimate, nearby markers can be suppressed, as suggested by XU (2003) in the case of closely linked markers. We see virtue in the extended model in QTL selection models with epistatic and $G \times E$ effects, where the number of predictors at each location may be very large.

We thank Shizhong Xu and Jean-Luc Jannink for constructive comments that improved this note.

LITERATURE CITED

- ABRAMOWITZ, M., and I. A. STEGUN, 1972 *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*. Dover, New York.

- BOX, G. E. P., and G. C. TIAO, 1973 *Bayesian Inference in Statistical Analysis*. Wiley Classics, New York.
- BROWN, P. J., M. VANNUCCI and T. FEARN, 2002 Bayes model averaging with selection of regressors. *J. R. Stat. Soc. Ser. B* **64**: 519–536.
- DUPUIS, J., and D. SIEGMUND, 1999 Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics* **151**: 373–386.
- GELMAN, A., J. B. CARLIN, H. S. STERN and D. B. RUBIN, 2004 *Bayesian Data Analysis*, Ed. 2. Chapman & Hall, London.
- MEUWISSEN, T. H. E., B. J. HAYES and M. E. GODDARD, 2001 Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**: 1819–1829.
- O’HAGAN, A., 1994 *Bayesian Inference*. Arnold, London.
- XU, S., 2003 Estimating polygenic effects using markers of the entire genome. *Genetics* **163**: 789–801.

Communicating editor: C. HALEY

APPENDIX

Posterior of b_j : XU’s (2003) Bayesian model and our extension of it can be expressed without reference to variance components σ_j^2 . Let the prior of b_j given σ_j^2 be $N(0, \sigma_j^2)$ and the prior for σ_j^2 be $p(\sigma_j^2) \propto (\sigma_j^2)^{\delta-1}$. Then, by integrating out σ_j^2 , the (unconditional) prior for b_j becomes

$$p(b_j) \propto \int_0^\infty (\sigma_j^2)^{\delta-1} \sigma_j^{-1} \exp(-\frac{1}{2} b_j^2 \sigma_j^{-2}) d\sigma_j^2 \propto (b_j^2)^{\delta-1/2}. \quad (\text{A1})$$

The last proportionality uses BOX and TIAO’s (1973) Equation A2.1.2. The prior is improper, for $\delta = \frac{1}{2}$ flat and for $\delta = 0$ a hyperbola. For the posterior distribution of b_j , given b_0 and the other QTL effects b_k ($k \neq j$), we obtain, with $y_i^* = y_i - b_0 - \sum_{k \neq j} b_k x_{ik}$,

$$p(b_j | \mathbf{y}) \propto (b_j^2)^{\delta-1/2} \exp[-\frac{1}{2} \sum_{i=1}^n (y_i^* - b_j x_{ij})^2 / \sigma_0^2] \propto (b_j^2)^{\delta-1/2} \exp(-\frac{1}{2} f_j (b_j - \hat{b}_j)^2), \quad (\text{A2})$$

where $f_j = \sigma_0^{-2} \mathbf{x}_j^T \mathbf{x}_j$ and $\hat{b}_j = (\mathbf{x}_j^T \mathbf{x}_j)^{-1} \mathbf{x}_j^T \mathbf{y}^*$, the least-squares estimate of b_j given the other parameters. For $\delta = 0$, the integral $C = \int_{-\infty}^\infty p(b_j | \mathbf{y}) db_j$ does not exist, as can be shown as follows. For any ε between 0 and 1,

$$\int_\varepsilon^1 |b_j^{-1}| \exp(-\frac{1}{2} f_j (b_j - \hat{b}_j)^2) db_j > a \int_\varepsilon^1 b_j^{-1} db_j = a [\log(b_j)]_\varepsilon^1 = -a \log(\varepsilon) \rightarrow \infty \quad \text{for } \varepsilon \downarrow 0$$

with a the minimum of $\exp(-\frac{1}{2} f_j (b_j - \hat{b}_j)^2)$ over b_j in the interval $[0, 1]$; a is a positive number. As the integrand is positive over the whole range of b_j , $C = \infty$; the posterior of b_j is thus improper and no valid Bayesian analysis is available. For $0 < \delta \leq \frac{1}{2}$,

$$C = 2^\delta \hat{b}_j^{-\delta} \exp(-\frac{1}{2} \hat{b}_j^2 f_j) \Gamma(\delta) M(\delta, \frac{1}{2}, \frac{1}{2} \hat{b}_j^2 f_j) < \infty, \quad (\text{A3})$$

with $\Gamma(\cdot)$ the gamma function and $M(\cdot, \cdot, \cdot)$ the first Kummer function (ABRAMOWITZ and STEGUN 1972). For $\delta = \frac{1}{2}$, $C = (2\pi\sigma^2/\mathbf{x}^T \mathbf{x})^{1/2}$ and (A2) is the normal density. Equation A2 has an infinite mode at 0 and, for $\hat{b}_j/se(\hat{b}_j) > 2(1 - 2\delta)^{1/2}$, a finite local mode between $\frac{1}{2}\hat{b}_j$ and \hat{b}_j . The second mode thus starts to appear if, for small δ , the z -ratio > 2 . This may also explain why $\delta = 0$ works well in practice to pick up QTL.

Swap step: Let \mathbf{G} be a 2×2 diagonal matrix with elements σ_j^2 and σ_{j+1}^2 , $\mathbf{X} = [\mathbf{x}_j | \mathbf{x}_{j+1}]$, $\mathbf{V} = \mathbf{X}^T \mathbf{X} + \sigma_0^2 \mathbf{G}^{-1}$, and $\tilde{\mathbf{y}}$ the phenotypic data residualized with respect to b_0 and all current marker effects, except b_j and b_{j+1} . Because $p(b_j, b_{j+1} | \tilde{\mathbf{y}}, \sigma_0^2, \mathbf{G})$ is bivariate normal $N(\tilde{\mathbf{b}}, \sigma_0^2 \mathbf{V}^{-1})$ with $\tilde{\mathbf{b}} = \mathbf{V}^{-1} \mathbf{X}^T \tilde{\mathbf{y}}$, the posterior distribution of the variance components σ_j^2 and σ_{j+1}^2 has density

$$\begin{aligned} p(\sigma_j^2, \sigma_{j+1}^2 | \tilde{\mathbf{y}}, \sigma_0^2) &= \frac{p(b_j, b_{j+1}, \sigma_j^2, \sigma_{j+1}^2 | \tilde{\mathbf{y}}, \sigma_0^2)}{p(b_j, b_{j+1} | \tilde{\mathbf{y}}, \sigma_0^2, \mathbf{G})} = 2\pi\sigma_0 |\mathbf{V}|^{-1/2} p(b_j = \tilde{b}_1, b_{j+1} = \tilde{b}_2, \sigma_j^2, \sigma_{j+1}^2 | \tilde{\mathbf{y}}, \sigma_0^2) \\ &\propto (\sigma_j^2 \sigma_{j+1}^2)^{\delta-1} |\mathbf{I} + \sigma_0^{-2} \mathbf{X}^T \mathbf{X} \mathbf{G}|^{-1/2} \exp\left(\frac{\tilde{\mathbf{y}}^T \mathbf{X} \mathbf{V}^{-1} \mathbf{X}^T \tilde{\mathbf{y}}}{2\sigma_0^2}\right) \end{aligned} \quad (\text{A4})$$

(BOX and TIAO 1973; O’HAGAN 1994; GELMAN *et al.* 2004). If $\sigma_j^2 = a$ and $\sigma_{j+1}^2 = b$ at the beginning of the swap step (the additional Metropolis step 5b), the acceptance probability of the swap to $\sigma_j^2 = b$ and $\sigma_{j+1}^2 = a$ is

$$\alpha = \min(1, p(\sigma_j^2 = b, \sigma_{j+1}^2 = a | \tilde{\mathbf{y}}, \sigma_0^2) / p(\sigma_j^2 = a, \sigma_{j+1}^2 = b | \tilde{\mathbf{y}}, \sigma_0^2)). \quad (\text{A5})$$

The two conditional probabilities in (A5) are calculated by inserting (b, a) and (a, b) for $(\sigma_j^2, \sigma_{j+1}^2)$ in (A4),

respectively. Note that \mathbf{G} and \mathbf{V} in (A4) depend on σ_j^2 and σ_{j+1}^2 . The calculation is easy and quick as (A4) contains the determinant and inverse of 2×2 matrices only.

Posterior of σ_j^2 : The univariate analogue of (A4) is

$$p(\sigma_j^2 | \mathbf{y}^*, \sigma_0^2) \propto (\sigma_j^2)^{\delta-1} (1 + f_j \sigma_j^2)^{-1/2} \exp\left(\frac{\hat{b}_j^2 f_j^2 \sigma_j^2}{1 + f_j \sigma_j^2}\right). \quad (\text{A6})$$

The posterior of σ_j^2 can be shown to be proper for $0 < \delta < 1/2$.