

Three Different Gibbs Samplers for BayesB Genomic Prediction

A.S. Leaflet R2867

Hao Cheng, Research Assistant; Rohan Fernando, Professor;
Dorian Garrick, Professor, Department of Animal Science

Summary and Implications

Typical implementations of genomic prediction utilize Markov chain Monte Carlo (MCMC) sampling to estimate effects. Metropolis-Hastings (MH) is a commonly-used algorithm. We considered three different Gibbs samplers to speed up BayesB, a commonly-used model for genomic prediction. These differ in the manner they sample the marker effect, the locus-specific variance and the indicator variable. They are a single-site Gibbs Sampler, a blocking Gibbs Sampler and a Gibbs Sampler with pseudo prior. These three versions of BayesB are about twice as fast as the one using a MH algorithm.

Introduction

In whole-genome analyses, the number of marker covariates is often much larger than the number of observations. Bayesian multiple regression models are widely used in genomic selection to address this problem of the number of effects to estimate exceeding the number of observations. In most Bayesian analyses of whole-genome data, inferences are based on Markov chains constructed to have a stationary distribution equal to the posterior distribution of the unknown parameters of interest. This is often done by employing a Gibbs sampler where samples are drawn from the full-conditional distribution of the parameters. BayesB was introduced by Meuwissen et al, (2001) for which they used a MH algorithm to jointly sample the marker effect and the locus-specific variance. For each locus in each MCMC iteration, 100 MH cycles were used, which makes BayesB computationally intensive.

However, Gibbs sampler can be used for the BayesB method instead of the MH algorithm. We show that by introducing a Bernoulli indicator variable in BayesB, indicating whether the marker effect for a locus is zero or non-zero, the marker effect and locus-specific variance can be sampled using a Gibbs sampler. We have considered three different versions of the Gibbs sampler to sample the marker effect, the locus-specific variance and the indicator

variable. The objectives of this research are to introduce these samplers and to study their performance.

Materials and Methods

Here we present three Gibbs Samplers for BayesB. The first is a single-site Gibbs sampler, where all parameters are sampled from their full conditional distributions. The second is a blocking Gibbs sampler, where the indicator variable and marker effects are sampled from their joint full-conditional distribution, which might improve mixing because the indicator variable and the marker effect are highly dependent. The third is a Gibbs sampler where a pseudo prior is used for the marker effect when the indicator variable is zero. Godsill has shown that the marginal posterior for parameters in the model do not depend on the pseudo priors chosen for the parameters that are not in the model. It has been suggested to choose the full conditional distribution for the marker effect when it is in the model as the pseudo prior. However, in BayesB, use of the exact full conditional distribution as the pseudo prior will require MH. Thus, a distribution close to the full conditional is employed.

Results and Discussion

Prediction accuracies of different versions of BayesB are identical. These three versions using Gibbs are about twice as fast as the one using a MH algorithm.

Table 1. Computing time (in seconds) for 40,000 iterations of BayesB with Metropolis-Hastings (MH), a single-site Gibbs Sampler, a blocking Gibbs Sampler and a Gibbs Sampler with pseudo prior.

Replica- tions	MH	Single- site Gibbs	Blocking Gibbs	Gibbs with pseudo prior
1	90,009	52,452	44,726	47,043
2	89,688	59,407	52,737	47,051
3	89,811	52,451	44,816	47,156
Speedup	1.0	1.7	2.0	1.9

Acknowledgments

Hao Cheng is funded by Endowment of the Lush Chair.