Indian Journal of Animal Sciences **76** (10): 860–861, October 2006

Estimation of heritability by Bayesian approach

V K BHATIA[1] and AMRIT KUMAR PAUL[2]

Indian Agricultural Statistics Research Institute, New Delhi 110 012 India

Received: 14 May 2005; Accepted: 26 July 2006

Key words: Bayesian approach, Heritability estimates

For the half-sib model, we know that the unbiased estimator for 'between sire' variance component (σ_s^2) may take negative values. If, one attempts to restrict the value of σ_s^2 to be non-negative then this will destroy its unbiasedness property and more importantly, further complicates the distribution theory of σ_s^2. A second difficulty is the sensitivity of inferences to departure from underlying assumptions. To overcome these difficulties, following Box and Tiao (1975) Bayesian approach has been adopted to estimate the variance components σ_s^2 and σ_e^2. Kumar *et al.* (2004) obtained Bayesian estimates of heritability in animal breeding experiment.

Bayesian analysis of data involves treating the parameters as random variables and finding their joint posterior distribution given the data. The joint distribution of the unknown random variables and the data (y) can be written as $p(y, \mu, \sigma_s^2, \sigma_e^2) = p(y/\mu, \sigma_s^2, \sigma_e^2)\ p(\mu)\ p(\sigma_s^2, \sigma_e^2)$

By taking different prior distributions for σ_s^2 and σ_e^2, the posterior distribution given the data can be obtained using Markov Chain Monte Carlo (MCMC) method and Gibbs sampling (Gelfand and Smith 1990). It may be pointed out here that prior distribution represents a population of possible parametric values, from which θ of current interest has been drawn. This prior distribution should be taken in such a way that the posterior distribution follows a known parametric form. In MCMC, we construct a stochastic process that has the desired posterior distribution as its stationary distribution and then simulate the process. We begin with a set of starting values for μ, σ_s^2, σ_e^2 and then successively generate values from the conditional posterior distribution of each parameter, conditioning on the most recently generated values of the other parameters at each step.

For the set of parameters values $\theta_1, \theta_2, \ldots \theta_k$, Gibbs sampling proceeds as follows:

Given an arbitrary set of starting values $\theta_1^{(0)}, \theta_2^{(0)}, \ldots \theta_k^{(0)}$ draw

$\theta_1^{(1)} \sim [\theta_1/\theta_2^{(0)}, \theta_3^{(0)}, \ldots \theta_k^{(0)}]$

$\theta_2^{(1)} \sim [\theta_2/\theta_1^{(0)}, \theta_3^{(0)}, \ldots \theta_k^{(0)}]$

$\theta_3^{(1)} \sim [\theta_3/\theta_1^{(0)}, \theta_2^{(0)}, \theta_4^{(0)}, \ldots \theta_k^{(0)}]$

$\vdots \quad \vdots$

$\theta_k^{(1)} \sim [\theta_k/\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}, \ldots \theta_{k-1}^{(0)}]$

Thus, each variable is visited in the natural order and a cycle generates k random variables. A sample of such draws can then be used to make inferences about the population.

For the present study, priors can be specified as,

$$\mu \sim N(\mu_0, \sigma_0^2),\ \sigma_s^2 \sim IG\ (a_1, b_1) \text{ and } \sigma_e^2 \sim IG\ (a_2, b_2)$$

where μ_0, σ_0^2, a_1, b_1, a_2 and b_2 are assumed to be known. Here IG refers to inverse gamma. With this knowledge of prior distribution and data at our disposal we can easily obtain the posterior distribution of the unknown parameters. The estimation of moments of the posterior distribution will ultimately result in estimates of the unknown variance component which in turn yield an estimate of the heritability.

Convergence can be diagnosed using the approach of Gelman *et al.* (1995). From the multiple chains of Gibbs sampler from overdispersed starting values, Gelman *et al.* (1995) potential scale reduction factor, $(R)^{1/2}$ which assesses between-chain and within-chain variation can be computed. Values of Gelman *et al.* (1995) statistics near one for all of the model parameters is evidence that the distribution of the Gibbs iterations is reasonably close to stationary (posterior) distribution. A sample of draws can then be used to make inferences about the population to diminish the effect of the starting distribution, and as such some starting observations will be discarded as burn in.

The application of Bayesian statistical techniques has also been illustrated by subjecting the simulated data sets to Bayesian procedure of estimating variance components. This technique is principally based on the posterior distribution of the unknown parameters by utilizing data and some prior distribution of the parameters. In the present investigation inverse gamma distribution IG (a_1, b_1) has been taken as the prior distribution as it belongs to the conjugate family of normal distribution and the required posterior and conditional distributions can be obtained in the known form. Because prior distribution for population variation is not known, the

Present address: [1]Principal Scientist, [2]Scientist.

Table 1. Comparison of Bayes estimates of heritability of different parametric values in balanced, unbalanced and outlier data sets

Population parameters	Parametric values (h^2)				
	0.1	0.2	0.3	0.4	0.5
		Balanced 20, 20, 20, 20			
Sample estimate	**–0.0966**	**–0.0515**	**0.0652**	**0.1383**	**0.2275**
Bayes (balanced)	0.1752	0.2025	0.3274	0.4362	0.5268
Bias	0.0752	0.0025	0.0274	0.0362	0.0268
		Unbalanced 15, 20,18, 20			
Sample estimate	**–0.1296**	**–0.1234**	**0.0351**	**0.1786**	**0.2304**
Bayes (unbalanced)	0.2134	0.2008	0.3649	0.5085	0.5861
Bias	0.1134	0.0008	0.0649	0.1085	0.0861
		Outlier (+10) 20, 20, 20, 20			
Sample estimate	**–0.1237**	**0.0943**	**0.1453**	**0.1569**	**0.0324**
Bayes (Outlier)	0.1060	0.2506	0.3343	0.3431	0.2035
Bias	0.0060	0.0506	0.0343	–0.0567	–0.2965

parametric values in the prior distribution are set to the low values of $a_t=b_t=0.001$. In this situation deliberately bad samples (samples having estimate either inadmissible or having large bias) are considered so as to examine the usefulness of Bayesian technique. The simulated samples were further subjected to Gibbs Sampling by employing the technique of Monte Carlo Markov Chain (MCMC) method and results obtained are presented in Table 1. In the present investigation, 5 Gibbs sequences of length 100 000 were used to obtain draws from the posterior distributions of the model parameters given the data. The first 25 000 draws of each chain were discarded and then every 150th draw was saved. The five chain yielded a posterior sample of 2 500 approximately uncorrelated draws. An examination of plots and $(R)^{½}$ values based on the 2 500 draws indicated that Gibbs sampler reached approximate convergence. As the sample size (n) has very little effect on the estimation of variances, so it was considered to be small. From the results it is seen that even very bad samples resulted very good estimates of population parameter h^2 in all the situations of balanced, unbalanced and having outliers. This thus advocates that if one takes into account many samples including good and bad then it is highly probable that Bayesian method will yield very good results in the sense that the estimates will have least bias and MSE.

SUMMARY

The applicability of the Bayesian techniques was also studied thoroughly and the inference from the results are very encouraging in the sense that the estimates obtained are very close to the parametric values of heritability.

REFERENCES

Box G E P and Tiao G C. 1973. *Bayesian Inference in Statistical Analysis Reading*. Addision Wesley Publishing Co., MA.

Gelfand A E and Smith A F M. 1990. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**: 398–409.

Gelman A, Carlin J B, Stern H S and Rubin D B. 1995. *Bayesian Data Analysis*. Chapman and Hall, New York.

Kumar Sanjeev, Rao A R and Bhatia V K. 2004. Bayesian estimation of heritability in animal breeding experiments under 2-way nested classification. *Journal of Indian Society of Agricultural Statistics* **58**(3): 352–62.