Bayesian Estimation of Genetic Parameters

Kenziro SAIO and Takeshi HAYASHI

*Faculty of Agriculture, The University of Tokyo, Yayoi 1-1-1
Bunkyo-ku Tokyo 113, Japan*

A bayesian approach for estimating genetic parameters of quantitative traits, such as heritability and genetic correlation coefficient, was evaluated. In this estimation procedure, prior information on genetic parameters to be estimated (such as the proper ranges e.g. [0, 1] for heritability and [-1, 1] for genetic correlation coefficient) was used in the construction of prior distributions of the bayesian procedure, where the joint posterior distribution of the parameters was obtained by replacing the other parameters as the phenotypic mean, variance and covariance with their classical estimators. Processing the expectation of each of the genetic parameters in relation to the joint posterior distribution by numerical integration gave the bayesian estimators. We call this bayesian method 'hybrid bayesian estimation' and its estimator 'hybrid bayesian estimator' as a combination of 'bayesian' and 'classical' method was used. Statistical properties of hybrid bayesian estimators with classical and modified ones, which are obtained by setting the classical estimates beyond the proper ranges to the nearest bound values, were investigated in the case of one trait and two traits using Monte-Carlo simulations. In the case of one trait, where only the heritability is of interest, simulations were carried out for two values of the heritability 0.2 and 0.8, while the phenotypic mean and variance were fixed at 0 and 1, respectively. It was shown that hybrid bayesian estimators have smaller m.s.e.'s (mean square errors) than classical and modified ones. In the case of two traits, where two heritabilities and a genetic correlation coefficient of two traits were jointly estimated, simulations were carried for two cases, first two heritabilities and a genetic correlation coefficient of 0.8, 0.2 and 0.0 and secondly 0.8, 0.2 and 0.5, respectively. The phenotypic mean vector and variance-covariance matrix were set at 0 and I_2 (2×2 identity matrix), respectively. Simulation results showed that for heritabilities, as in the case of one trait, hybrid bayesian estimators were superior to classical and modified ones while for the genetic correlation coefficient, the superiority of hybrid bayesian estimators was even more conspicuous. Therefore it is concluded that hybrid bayesian estimators of genetic parameters are superior to classical and modified ones and can be very useful in practice.

KEY WORDS: Bayesian estimation, genetic parameters, heritability, genetic correlation coefficient, half-sib data, Monte-Carlo simulation.

Introduction

The use of estimators of genetic parameters through a conventional or classical method is generally associated with large sampling errors and each of them is distributed beyond a proper range which each parameter should have (e.g. from 0 to 1 for heritability and from -1 to 1 for genetic correlation coefficient) (GILL and JENSEN 1968). It may be natural to consider that prior or empirical information (including proper ranges of distribution) on given genetic parameters should be employed (TORO and PRUÑONOSA 1984). The authors attempted to adopt a bayesian approach for estimating genetic parameters. In most cases, however, it may not be relevant to estimate by a bayesian method all the parameters on which a so-called posterior density of given genetic parameters depends. For example, parameters such as phenotypic variances

Received June 15, 1989

and/or covariances are well estimated by a conventional method such as maximum-likelihood or least-squares method without major difficulty. We considered that it may be preferable to estimate such parameters with a conventional method as a first step and then to estimate the other genetic parameters, using these estimated parameters as though known parameters, by a bayesian method. We tentatively call this method hybrid bayesian method. In this paper, we deal with the estimation of heritabilities and genetic correlation coefficients through the hybrid bayesian method and referred to sampling properties of their hybrid bayesian estimators, which were compared with those of conventional or classical ones, through Monte-Carlo simulations.

Bayesian Estimation

We will describe shortly the methods used for the bayesian estimation for continuous distributions.

Suppose that $p(\theta)$ denotes a so-called prior distribution density of a parameter vector θ which is also the marginal distribution density of θ , and $f(y|\theta)$ denotes a (conditional) distribution density of an observation. Then, the joint density of y and θ is $f(y|\theta)p(\theta)$; therefore the marginal distribution of y with respect to θ is $\int_R f(y|\theta)p(\theta)d\theta$ (where R is the distribution range of θ). Consequently, the conditional distribution of θ given y , which is the so-called posterior distribution of θ , is

$$\frac{f(y|\theta)p(\theta)}{\int_R f(y|\theta)p(\theta)d\theta}$$

and the conditional expectation of θ_i (i th element of the vector θ) given y is

$$\frac{\int_R \theta_i f(y|\theta)p(\theta)d\theta}{\int_R f(y|\theta)p(\theta)d\theta} \quad (1)$$

which is the usual bayesian estimator of θ_i .

Bayesian Estimation of Heritability from Half-Sib Data

We first carried out a bayesian estimation of heritability from half-sib data. We assume to have obtained the data of observations of a certain quantitative trait from n half-sib daughters in each of m sires or an equivalent mating system. We denote the measurement of the j th daughter of the i th sire by y_{ij} ($i = 1, \dots, m$; $j = 1, \dots, n$).

We assume the following model:

$$y_{ij} = \mu + s_i + d_{ij} \quad (i = 1, \dots, m; \quad j = 1, \dots, n \text{ for each } i), \quad (2)$$

where y_{ij} : the measurement of the j th daughter of sire i ,

μ : the general mean,

s_i : the variate on the effect of sire i , which is distributed as $N(0, \sigma_s^2)$,

d_{ij} : the variate on the effect of the j th daughter of sire i , which is distributed as $N(0, \sigma_d^2)$;

and all of the variates are distributed independently of each other. This model can be mentioned in other words; y_{ij} ($i = 1, \dots, m$; $j = 1, \dots, n$) are independently and normally

distributed as $N(\mu, \sigma_P^2)$, where σ_P^2 is the so-called phenotypic variance and $\sigma_P^2 = \sigma_s^2 + \sigma_d^2$. The joint distribution of y_{ij} ($i = 1, \dots, m$; $j = 1, \dots, n$) is the joint distribution of independently and identically distributed $y_1, \dots, y_m$, where $y_i = (y_{i1}, \dots, y_{in})$ ($i = 1, \dots, m$) is the vector variate with the n elements $y_{i1}, \dots, y_{in}$, and is distributed as the multivariate normal distribution $N(\mu, \Sigma)$, where the n element mean vector μ and the $n \times n$ covariance matrix Σ are as follows;

$$\mu = \begin{pmatrix} \mu \\ \cdot \\ \cdot \\ \mu \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_P^2 & \sigma_s^2 & \dots & \sigma_s^2 \\ \sigma_s^2 & \sigma_P^2 & \dots & \sigma_s^2 \\ \cdot & \cdot & \dots & \cdot \\ \sigma_s^2 & \sigma_s^2 & \dots & \sigma_P^2 \end{pmatrix} \quad (3)$$

And also $\sigma_s^2 = \sigma_G^2/4$ (KEMPTHORNE 1956) or $\sigma_s^2 = h^2 \sigma_P^2/4$, where σ_G^2 is the additive genetic variance and h^2 is the heritability which is σ_G^2/σ_P^2 .

Consequently, the density of $y_1, \dots, y_m$ can be expressed using the three parameters h^2 , σ_P^2 , and μ . Let be the conditional density of $y_1, \dots, y_m$ given h^2 , σ_P^2 , and μ , and the prior density of h^2 , σ_P^2 , and μ denoted by $f(y_i|h^2, \sigma_P^2, \mu)$ and $g(h^2, \sigma_P^2, \mu)$ respectively. Hence the bayesian estimator of h^2 is

$$\frac{\int \int \int_R h^2 [\prod_i f(y_i|h^2, \sigma_P^2, \mu)] g(h^2, \sigma_P^2, \mu) dh^2 d\sigma_P^2 d\mu}{\int \int \int_R [\prod_i f(y_i|h^2, \sigma_P^2, \mu)] g(h^2, \sigma_P^2, \mu) dh^2 d\sigma_P^2 d\mu} \quad (4)$$

However, since a joint prior distribution of these three parameters cannot be easily conceived, we must devise an alternative or modified method. Thus σ_P^2 and μ can be estimated, as a first step, by the conventional or classical method which is expected to produce considerably well-behaved estimators of these two parameters, and then by replacing them with their estimates, we proceed to one parameter bayesian estimation of h^2 as if σ_P^2 and μ were known. We should like, hereafter, to call this method 'hybrid bayesian method' as a combination of 'bayesian' and 'classical' method was used.

Hybrid Bayesian Estimation of Heritability from Half-Sib Data

To carry out the hybrid bayesian estimation of h^2 from the half-sib data mentioned above, we must use the approximate conditional density of y_i ($i = 1, \dots, m$) given h^2 in which μ and Σ are substituted for $\hat{\mu}$ and $\tilde{\Sigma}$ in the exact conditional distribution of y_i given h^2 . This approximate distribution is $N(\hat{\mu}, \tilde{\Sigma})$, where

$$\hat{\mu} = \begin{pmatrix} \mu \\ \cdot \\ \cdot \\ \mu \end{pmatrix}, \quad \tilde{\Sigma} = \begin{pmatrix} \hat{\sigma}_P^2 & \hat{\sigma}_s^2 & \dots & \hat{\sigma}_s^2 \\ \hat{\sigma}_s^2 & \hat{\sigma}_P^2 & \dots & \hat{\sigma}_s^2 \\ \cdot & \cdot & \dots & \cdot \\ \hat{\sigma}_s^2 & \hat{\sigma}_s^2 & \dots & \hat{\sigma}_P^2 \end{pmatrix} \quad (5)$$

and $\hat{\mu} = \bar{y}_{..} (= \sum_i \sum_j y_{ij}/mn)$ is the grand mean, $\hat{\sigma}_P^2 = (B + (n-1)W)/n$ is the unbiased estimate of the phenotypic variance, and $\hat{\sigma}_s^2 = h^2 \hat{\sigma}_P^2/4$, where $B = n \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2/(m-1)$

is the between sire mean square, $W = \sum_i \sum_j (y_{ij} - y_i.)^2 / m(n-1)$ is the within sire mean square. The density of this multivariate normal distribution can be denoted by $\tilde{f}(y_i|h^2)$, and $g*(h^2)$ denotes the marginal prior density of h^2 . Then the hybrid bayesian estimator of h^2 is

$$\frac{\int_a^b h^2 [\prod_i \tilde{f}(y_i|h^2)] g*(h^2) dh^2}{\int_a^b [\prod_i \tilde{f}(y_i|h^2)] g*(h^2) dh^2} \quad (6)$$

where $[a, b]$ is the distribution range of $g*(h^2)$.

While the classical estimator of h^2 using the variance component estimators is

$$\hat{h}^2 = [4(B - W)/n] / \hat{\sigma}_P^2 \quad (\text{BULMER 1980}).$$

Prior Distributions of Heritability or Any Other Genetic Parameters with a Finite Proper Range

Prior or empirical information about the heritability of a certain trait in a certain crop or livestock is a concept expressed with some fuzzy set, the elements of which are values of the heritability. Usually the membership function is continuous in a rough sense, and a prior distribution density can be considered as some normalized membership function of the set. By the normalization, the value of the finite integration on the function over some range becomes the unity. Then, in most cases, it may be natural that a prior distribution of h^2 is considered as some unimodal or uniform distribution within a certain range included in $[0, 1]$ interval. Consequently $g*(h^2)$ in the previous section can be considered as a modified beta distribution density with an arbitrary distribution range $[a, b]$ (in this case the range must be within $[0, 1]$ interval, but generally it may be only finite), and we will call tentatively this distribution a generalized beta distribution. It is very convenient to assume that this includes uniform distributions. In conventional consideration, the generalized beta distribution can be used as a prior distribution of any other genetic parameter defined in a finite range, say genetic correlation coefficient as mentioned later, as well as heritability. The generalized beta density can be written as follows: for $a < b$,

$$h(\theta) = \begin{cases} \frac{(b-a)}{B(\alpha, \beta)} (\theta-a)^{\alpha-1} (b-\theta)^{\beta-1}, & a \leq \theta \leq b \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

where θ is any genetic parameter with a finite proper range, $B(\alpha, \beta)$ is a beta function, and the interval $[a, b]$ is included in the proper range of the parameter (e.g. $[0, 1]$ in heritability, and $[-1, 1]$ in genetic correlation coefficient). This distribution will be denoted as Beta $(\alpha, \beta; a, b)$, and a mean and a variance are

$$\text{mean : } \frac{a\beta + b\alpha}{\alpha + \beta}, \quad \text{variance : } \frac{(b-a)^2 \alpha \beta}{(\alpha + \beta + 1)(\alpha + \beta)^2} \quad (8)$$

Beta $(\alpha, \beta; 0, 1)$ is a usual beta distribution, and furthermore Beta $(1, 1; a, b)$ is a uniform distribution in a range $[a, b]$. Since the former two parameters α and β are related to

a shape of this distribution, we will call them shape parameters, and the latter two parameters a and b will be called range parameters for the same reason.

If one wishes to take a multimodal distribution as some prior distribution, a mixed distribution of two or more generalized beta distributions must be one of the best candidates of such distribution types, and this could be used as an approximation of a distribution of the discrete type by mixing several generalized beta distributions which have small variances if a discrete prior distribution is needed. In the case of a discrete prior distribution, however, the integration on $p(\theta)$ in formula (1) can be changed to an appropriate summation, and hence probably a continuous distribution need not be used as such an approximation of a discrete prior distribution.

Monte-Carlo Simulation Results for Hybrid Bayesian Heritability Estimators

Some statistical properties of hybrid bayesian estimators were examined by means of Monte-Carlo simulation.

Since a data-set $y_i (i = 1, \dots, m)$ must be obtained from an independent sample of size m from an n -dimensional multivariate normal population $N(\mu, \Sigma)$ in (3), m pseudo-random n -dimensional vectors in $N(\mu, \Sigma)$ can be considered to be one data-set in the simulation. The first simulation was executed, where the original three parameters h^2, σ_p^2 and μ relating to μ and Σ were fixed at 0.2, 1 and 0 respectively, and sample sizes m and n were fixed at $m = 10$ and $n = 20$. In the second simulation h^2 was set at 0.8, while the values of other parameters were unchanged.

In any following simulations, we can set $\mu = 0$ and $\sigma_p^2 = 1$ without loss of generality to examine the statistical properties of estimators of the heritability, since the values themselves of these parameters are independent of heritability estimators.

As mentioned in the previous section, the approximate conditional density of y_i given h^2 is the density of $N(\hat{\mu}, \tilde{\Sigma})$, and written precisely as follows;

$$\tilde{f}(y_i|h^2) = (2\pi)^{-m/2} |\tilde{\Sigma}|^{-1/2} \exp[-1/2(y_i - \hat{\mu})' \tilde{\Sigma}^{-1} (y_i - \hat{\mu})]. \quad (9)$$

where the element $\hat{\mu}$ of the mean vector $\hat{\mu}$ and the element $\hat{\sigma}_p^2$ (which is also the factor of the element of $\tilde{\sigma}_s^2$) of the covariance matrix $\tilde{\Sigma}$ are calculated as (5) from an observed data-set. In (6), by substituting this density for $\tilde{f}(y_i|h^2)$ and any density (7) of generalized beta distributions restricted in their range parameters as $a(\geq 0)$ and $b(\leq 1)$ for $g^*(h^2)$, a hybrid bayesian estimate can be estimated by calculation of (6) through some appropriate method of numerical integration. The results of the 300 runs of simulation in uniform prior distributions are shown in table 1, in terms of frequency distributions and statistical properties of three kinds of estimators, the hybrid bayesian estimator h^{*2} , the classical estimator $\hat{h}^2$ and modified classical estimator $\tilde{h}^2$ in which 0 and 1 are used for under-zero and over-one $\hat{h}^2$ estimates respectively. The method of numerical integration adopted here is the iteration of the Gauss 6 points method by partitioning (MCCALLA 1967). Comparison of effects of prior distributions in the hybrid bayesian estimation is given in table 2 by estimation from some identical data-sets with varying parameters of the generalized beta

Table 1. Frequency distributions and statistical properties of estimators of heritability in Monte-Carlo simulation

class	$h^2=0.2$			h^2		
	h^{2*}	$\hat{h}^2$	$\tilde{h}^2$	h^{2*}	$\hat{h}^2$	$\tilde{h}^2$
under 0	0	42	0	0	2	0
[0.0, 0.1)	9	68	110	0	4	6
[0.1, 0.2)	107	72	72	6	10	10
[0.2, 0.3)	90	62	62	14	11	11
[0.3, 0.4)	57	27	27	12	17	17
[0.4, 0.5)	20	14	14	40	30	30
[0.5, 0.6)	12	8	9	44	27	27
[0.6, 0.7)	5	2	2	70	46	46
[0.7, 0.8)	0	4	4	65	39	39
[0.8, 0.9)	0	0	0	46	26	26
[0.9, 1.0)	0	0	0	3	27	88
over 1	0	0	0	0	61	0
total	300	300	300	300	300	300
mean	0.257	0.176	0.181	0.624	0.733	0.688
variance	0.0154	0.0277	0.0252	0.0314	0.1117	0.0677
m.s.e.	0.0186	0.0282	0.0255	0.0621	0.1158	0.0799
maximum	0.689	0.778	0.778	0.914	0.726	1.000
minimum	0.074	-0.136	0.000	0.108	-0.044	0.000

Note: h^{2*} , $\hat{h}^2$, $\tilde{h}^2$ =hybrid bayesian estimator, classical estimator and modified estimator, respectively.

m =number of sires

n =number of dams within sires.

m.s.e.=mean square error.

The number of simulations was 300.

distribution. Although 300 as the number of simulations may not be sufficient to analyse statistical properties, the increase in the number of simulation runs may not be practical due to the considerably large CPU time required for numerical integration even by using a very high speed computer HITACHI M-680H in the Computer Centre of the University of Tokyo. This fact never tells that the hybrid bayesian estimation of heritability is not practical, since in an actual estimation we need only a few sets of calculations.

Concluding Remarks about Hybrid Bayesian Estimation of Heritability

The hybrid bayesian estimation gives well-behaved estimators even in the case of [0,1] uniform prior distribution which means that no prior information is available. However, highly incorrect prior distributions give bad results. Then [0,1] uniform prior distribution should be adopted when we have little and/or doubtful prior information. Though in general hybrid bayesian estimators seem to have a slightly larger bias than classical estimators as the expectation of estimators may be close to the expectation of the prior distribution, hybrid bayesian estimators show smaller variances and hence smaller mean square errors (m.s.e.'s) except in the cases of highly incorrect prior information. It can be considered

Table 2. Statistical properties of hybrid bayesian estimators with respect to various prior generalized beta distributions
a) $m=10$, $n=20$, $h^2=0.2$

shape parameter α	parameter β	range a	parameter b	exp	mean	variance	m.s.e.	minimum	maximum
1.0	1.0	0.0	1.0	0.50	0.257	0.0154	0.0186	0.074	0.689
		0.0	0.5	0.25	0.208	0.0055	0.0055	0.073	0.399
		0.5	1.0	0.75	0.621	0.0008	0.1779	0.581	0.749
4.0	6.0	0.0	1.0	0.40	0.316	0.0031	0.0165	0.209	0.504
		0.0	0.5	0.20	0.292	0.0016	0.0010	0.206	0.402
		0.5	1.0	0.70	0.565	0.0001	0.1333	0.552	0.603
5.0	5.0	0.0	1.0	0.50	0.388	0.0031	0.0384	0.278	0.571
		0.0	0.5	0.25	0.388	0.0010	0.0020	0.265	0.421
		0.5	1.0	0.75	0.581	0.0001	0.1453	0.563	0.636

b) $m=10$, $n=20$, $h^2=0.8$

shape parameter α	parameter β	range a	parameter b	exp	mean	variance	m.s.e.	minimum	maximum
1.0	1.0	0.0	1.0	0.50	0.624	0.0314	0.0621	0.108	0.914
		0.0	0.5	0.25	0.372	0.0043	0.1877	0.104	0.458
		0.5	1.0	0.75	0.737	0.0056	0.0095	0.590	0.915
4.0	6.0	0.0	1.0	0.40	0.484	0.0080	0.1081	0.240	0.697
		0.0	0.5	0.20	0.387	0.0017	0.1724	0.234	0.456
		0.5	1.0	0.70	0.603	0.0008	0.0398	0.555	0.705
5.0	5.0	0.0	1.0	0.50	0.551	0.0072	0.0693	0.312	0.747
		0.0	0.5	0.25	0.410	0.0009	0.1531	0.290	0.461
		0.5	1.0	0.75	0.631	0.0013	0.0298	0.567	0.750

Note: exp=expectation of prior distribution.

m.s.e.=mean square error.

The number of simulations was 300.

that mean square errors are the best criterion for the evaluation of usefulness of estimators. Above all hybrid bayesian estimation cannot give wrong estimates beyond the $[0,1]$ range.

Simultaneous Estimation of Heritabilities and Genetic Correlation Coefficients from Half-Sib Multivariate Data

The above-mentioned model can be directly applied to a multivariate case where two or more traits are observed. Then our p -variate model in p traits observations will be as follows:

$$\underset{(p \times 1)}{y_{ij}} = \mu + s_i + d_{ij} (i = 1, \dots, m; j = 1, \dots, n \text{ for each } i), \quad (10)$$

where y_{ij} : the observation vector of p traits of the j th daughter of sire i ,
 μ : the general mean p -vector,
 s_i : the p -variate vector on the effect of sire i , which is distributed as $N(0, \Sigma_s)$,
 $(p \times p)$
 d_{ij} : the p -variate vector on the effect of the j th daughter of sire i , which is distributed as $N(0, \Sigma_d)$;
 $(p \times p)$

and $s_i (i = 1, \dots, m)$ and $d_{ij} (i = 1, \dots, m; j = 1, \dots, n)$ are distributed independently of each other. This model also implies that $y_{ij} (i = 1, \dots, m; j = 1, \dots, n)$ are independently and normally distributed as $N(\mu, \Sigma_P)$, where Σ_P is the so-called phenotypic covariance matrix and $\Sigma_P = \Sigma_s + \Sigma_d$. The joint distribution of $y_{ij} (i = 1, \dots, m; j = 1, \dots, n)$ is the joint distribution of independently and identically distributed $y_1, \dots, y_m$, where $y_i (pn \times 1) = (y_{i1}, \dots, y_{in})$ ($i = 1, \dots, m$) is the vector variate with pn elements of n p -component vectors $y_{i1}, \dots, y_{in}$, and is distributed as the multivariate normal distribution $N(\mu_0, \Sigma_0)$, where the pn element mean vector μ_0 and the $pn \times pn$ covariance matrix Σ_0 are as follows;

$$\mu_0 = \begin{pmatrix} \mu \\ \cdot \\ \cdot \\ \cdot \\ \mu \end{pmatrix}, \quad \Sigma_0 = \begin{pmatrix} \Sigma_P & \Sigma_s & \dots & \Sigma_s \\ \Sigma_s & \Sigma_P & \dots & \Sigma_s \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \Sigma_s & \Sigma_s & \dots & \Sigma_P \end{pmatrix} \quad (11)$$

Also

$$\Sigma_s = \Sigma_G/4 = PHR_GHP/4, \quad (12)$$

where Σ_G is the additive genetic matrix, and P and H are diagonal matrices as follows:

$$\begin{matrix} P \\ (p \times p) \end{matrix} = \begin{pmatrix} \sqrt{\sigma_{P1}^2} & & 0 \\ & \cdot & \\ 0 & & \cdot \\ & & & \sqrt{\sigma_{Pp}^2} \end{pmatrix}, \quad \begin{matrix} H \\ (p \times p) \end{matrix} = \begin{pmatrix} \sqrt{h_1^2} & & 0 \\ & \cdot & \\ 0 & & \cdot \\ & & & \sqrt{h_p^2} \end{pmatrix}, \quad (13)$$

and R_G is the genetic correlation matrix as follows:

$$\begin{matrix} R_G \\ (p \times p) \end{matrix} = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \cdot & \cdot & \dots & \cdot \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix}, \quad (14)$$

where σ_{Pk}^2 and $h_k^2 (k = 1, \dots, p)$ are the phenotypic variance and the heritability, respectively, of the k th trait, and $r_{kl} (k, l = 1, \dots, p; k \neq l)$ is the genetic correlation coefficient between the k th trait and the l th trait. The conditional distribution density of observations $y_1, \dots, y_m$ given H, R_G, Σ_P and μ , that is, given heritabilities, genetic means about all p traits (also P needs to be given, but P is automatically determined according to the given Σ_P since the diagonal elements of P are the square roots of the diagonal elements of Σ_P), is

$$\prod_i f(y_i | H, R_G, \Sigma_P, \mu), \quad (15)$$

where $f(\cdot | \dots)$ indicates a density of the multivariate normal distribution $N(\mu_0, \Sigma_0)$, with μ_0

and Σ_0 being determined by the given parameters written in the right side of ']' of $f(.|....)$.

As in a univariate case, the hybrid bayesian estimation can be considered as the most relevant and practical method. In the estimation, Σ_P and μ are estimated by the classical method using the 'between sire' mean squares and mean products matrix, the 'within sire' matrix and the grand mean vector of the observation vectors in extending the univariate form to the multivariate form. The gained estimates $\hat{\Sigma}_P$ and $\hat{\mu}$ are used instead of Σ_P and μ in (15), respectively. The approximate conditional density is written as follows:

$$\prod_i \tilde{f}(y_i|H, R_G) \quad (16)$$

where $\tilde{f}(.|..)$ is the substituted $f(.|....)$ with $\hat{\Sigma}_P$ and $\hat{\mu}$ instead of Σ_P and μ .

The joint prior distribution of H and R_G , that is $h_1^2, \dots, h_p^2$, and $r_{12}, \dots, r_{(p-1)p}$ ($p(p+1)/2$ parameters at total), could be considered as various types of distributions, even though it may be natural that each marginal distribution is a generalized beta distribution. Though the independency of $h_1^2, \dots, h_p^2$ could be assumed, all the parameters of $h_1^2, \dots, h_p^2, r_{12}, \dots, r_{(p-1)p}$ cannot be distributed independently as a whole due to the necessary restriction that the additive genetic covariance matrix Σ_G must be positive definite (HILL and THOMPSON 1978). Fortunately the positive definiteness of Σ_G is equivalent to that of R_G for (12) since $\sigma_{P_i}^2$ and h_i^2 ($i = 1, \dots, p$) are positive. Hence, except when some special information is available, H and R_G may be distributed independently. However the elements of $R_G, r_{12}, \dots, r_{(p-1)p}$, must be distributed with the positive definiteness of R_G . Consequently, in the case of $p = 2$, the joint prior distribution of the three parameters, h_1^2, h_2^2 and r_{12} , can be considered as an independent joint of three generalized beta distributions, the distribution ranges of which are within interval $[0,1]$ in h_1^2 and h_2^2 and $[-1,1]$ in r_{12} .

Though it may not be necessary to treat cases of $p \geq 3$ if the objective is to analyse the statistical properties of estimators of a genetic correlation coefficient, it is never easy to show some natural joint prior distribution of $r_{12}, \dots, r_{(p-1)p}$. Therefore the joint prior distribution of h_i^2 and r_{ij} ($i, j = 1, \dots, p; i < j$) is denoted by the following equation:

$$g(h_1^2, \dots, h_p^2, r_{12}, \dots, r_{(p-1)p}) = g_1(h_1^2) \dots g_p(h_p^2) g_r(r_{12}, \dots, r_{(p-1)p}), \quad (17)$$

where each of $g_i(.)$ ($i = 1, \dots, p$) is a prior density of the heritability h_i^2 and may be considered as a generalized beta density (7) with a distribution range within $[0,1]$ and $g_r(.....)$ is a joint prior density of $r_{12}, \dots, r_{(p-1)p}$ and each univariate marginal density can also be expressed as a generalized beta distribution with a distribution range within $[-1,1]$.

The hybrid bayesian estimator of any of h_i^2 and r_{ij} ($i, j = 1, \dots, p; i < j$) is as follows:

$$\frac{\int_R \dots \int_{a_1}^{b_1} \dots \int_{a_p}^{b_p} x \{ \prod_i \tilde{f}_0(y_i, H, R_G) \} dh_1^2 \dots dh_p^2 dr_{12} \dots dr_{(p-1)p}}{\int_R \dots \int_{a_1}^{b_1} \dots \int_{a_p}^{b_p} \prod_i \tilde{f}_0(y_i, H, R_G) dh_1^2 \dots dh_p^2 dr_{12} \dots dr_{(p-1)p}} \quad (17)$$

where (a_i, b_i) ($i = 1, \dots, p$) are the distribution ranges of $g_i(.)$ ($i = 1, \dots, p$) respectively, R is a $p(p-1)/2$ dimensional distribution range of $g_r(., \dots, .)$, x is any of h_i^2 or r_{ij} ($i, j = 1, \dots, p; i < j$), and $\tilde{f}_0(y_i, H, R_G) = \tilde{f}(y_i|H, R_G) g_1(h_1^2) g_2(h_2^2) \dots g_p(h_p^2) g_r(r_{12}, \dots, r_{(p-1)p})$.

Table 3. Statistical properties of estimators of heritabilities and genetic correlation.

a) $m=10$, $n=20$, $h_1^2=0.8$, $h_2^2=0.2$, $r_{12}=0.0$

	h_1^2			h_2^2			r_{12}		
	h^{*1^2}	$\hat{h}_1^2$	$\tilde{h}_1^2$	h^{*2^2}	$\hat{h}_2^2$	$\tilde{h}_2^2$	r^{*12}	$\hat{r}_{12}$	$\tilde{r}_{12}$
mean	0.647	0.782	0.726	0.248	0.182	0.192	-0.021	-0.091	-0.054
variance	0.031	0.114	0.065	0.019	0.035	0.031	0.140	0.570	0.366
m.s.e.	0.054	0.113	0.070	0.021	0.035	0.031	0.139	0.571	0.365
under 0, -1	0	0	0	0	14	0	0	8	0
over 1	0	25	0	0	0	0	0	7	0
maximum	0.907	1.648	1.000	0.683	0.784	0.784	0.724	1.410	1.000
minimum	0.213	0.113	0.113	0.059	-0.163	0.000	-0.688	-2.513	-1.000

b) $m=10$, $n=20$, $h_1^2=0.8$, $h_2^2=0.2$, $r_{12}=0.50$

	h_1^2			h_2^2			r_{12}		
	h^{*1^2}	$\hat{h}_1^2$	$\tilde{h}_1^2$	h^{*2^2}	$\hat{h}_2^2$	$\tilde{h}_2^2$	r^{*12}	$\hat{r}_{12}$	$\tilde{r}_{12}$
mean	0.650	0.777	0.718	0.254	0.182	0.189	0.319	0.717	0.456
variance	0.034	0.125	0.071	0.018	0.032	0.029	0.111	4.471	0.255
m.s.e.	0.058	0.124	0.077	0.021	0.032	0.029	0.143	4.464	0.253
under 0, -1	0	0	0	0	17	0	0	1	0
over 1	0	27	0	0	0	0	0	15	0
maximum	0.915	1.716	1.000	0.696	0.783	0.783	0.823	1.000	18.799
minimum	0.162	0.054	0.054	0.072	-0.144	0.000	-0.381	-1.000	-2.310

Note: h_1^{*2} , h_2^{*2} , r^{*12} =hybrid bayesian estimators of heritabilities and genetic correlation. $\hat{h}_1^2$, $\hat{h}_2^2$, $\hat{r}_{12}$ =classical estimators of heritabilities and genetic correlation. $\tilde{h}_1^2$, $\tilde{h}_2^2$, $\tilde{r}_{12}$ =modified classical estimators of heritabilities and genetic correlation.

m.s.e.=mean square error

under 0, -1=number of runs under 0 in heritabilities and under -1 in genetic correlation.

The number of simulations was taken 100 times.

Monte-Carlo Simulation Results of Joint Hybrid Bayesian Estimation of Heritabilities of two Traits and a Genetic Correlation Coefficient between them

As mentioned in the previous section, a joint prior density of heritabilities h_1^2 and h_2^2 of two traits and a genetic correlation coefficient r_{12} between them will be considered as an independent joint of three generalized beta densities $g_1(h_1^2)$, $g_2(h_2^2)$ and $g_r(r_{12})$. If we have no intrinsic prior information about these parameters, it will be appropriate that we consider that any of the three densities is uniform in a proper possible range in each parameter, that is, both $g_1(\cdot)$ and $g_2(\cdot)$ are densities of Beta(1,1;0,1) (uniform in (0,1)) and $g_r(\cdot)$ is a density of Beta(1,1;-1,1) (uniform in (-1,1)). In the following Monte-Carlo simulation, the joint prior distribution of h_1^2 , h_2^2 and r_{12} is considered to be an independent

Table 4. Frequency distributions of estimators

class	$r_{12}=0.0$		$r_{12}=0.5$	
	r^*_{12}	$\hat{r}_{12}$	r^*_{12}	$\hat{r}_{12}$
under -1.0	0	8	0	1
[-1.0, -0.9)	0	1	0	1
[-0.9, -0.8)	0	1	0	0
[-0.8, -0.7)	0	3	0	0
[-0.7, -0.6)	6	8	0	0
[-0.6, -0.5)	4	4	0	1
[-0.5, -0.4)	10	3	0	3
[-0.4, -0.3)	12	6	6	1
[-0.3, -0.2)	7	3	3	4
[-0.3, -0.2)	7	3	3	4
[-0.2, -0.1)	4	5	6	1
[-0.1, 0.0)	6	3	3	3
[0.0, 0.1)	10	4	12	7
[0.1, 0.2)	8	7	7	4
[0.2, 0.3)	10	3	3	4
[0.3, 0.4)	12	3	10	2
[0.4, 0.5)	2	6	14	4
[0.5, 0.6)	3	5	11	6
[0.6, 0.7)	4	2	14	9
[0.7, 0.8)	2	2	8	4
[0.8, 0.9)	0	2	3	6
[0.9, 1.0)	0	0	0	7
over 1.0	0	7	0	15
Rate of estimableness (%)	100	86	100	83

Note: r^*_{12} =hybrid bayesian estimator of r_{12} .

$\hat{r}_{12}$ =classical estimator of r_{12} .

joint of two Beta(1,1;0,1)'s and one Beta(1,1;-1,1). Simulations were carried out for two cases of $h_1^2 = 0.8$, $h_2^2 = 0.2$, $r_{12} = 0.0$ and $h_1^2 = 0.8$, $h_2^2 = 0.2$, $r_{12} = 0.5$. With respect to the other parameters $\Sigma_P = I_2$, and $\mu = 0$ are assumed, where I_2 is 2×2 identity matrix. The results of 100 runs are listed in table 3, where classical estimators and hybrid bayesian estimators are compared in two cases for true values of three parameters. Table 4 shows the frequency distributions of estimators of the genetic correlation coefficient. In the case of two or more traits, the number of genetic parameters to be estimated increase with the number of traits and the time spent in a simulation is much longer than that in one trait case. Therefore in this case we selected 100 as the number of simulations. Although 100 simulation runs may not be sufficient, the statistical properties of each estimator could be revealed roughly, if not precisely. It is evident that hybrid bayesian estimators, which give smaller m.s.e's than classical estimators, are well-behaved statistical tools.

Concluding Remarks about Joint Hybrid Bayesian Estimation of Heritabilities and Genetic Correlation Coefficient

As shown in table 3, hybrid bayesian estimators of heritabilities, $h_i^{*2} (i = 1, 2)$, have more desirable properties as in the case of one trait. In the estimation of genetic correlation coefficients, the superiority of hybrid bayesian estimator r_{12}^* to classical $\hat{r}_{12}$ and modified $\tilde{r}_{12}$ is more conspicuous than in the heritability. In the classical estimation $\hat{r}_{12}$ is given as $\hat{\sigma}_{G12} / \sqrt{\hat{h}_1^2 \hat{\sigma}_{P1}^2 \hat{h}_2^2 \hat{\sigma}_{P2}^2}$, where $\hat{\sigma}_{G12}$ is the classical estimator of genetic covariance between two traits, which is computed from the off-diagonal component of the between and within sire mean square and products matrix. Therefore non positive values of $\hat{h}_i^2$ make the classical and modified estimation of the genetic correlation coefficient impossible. As shown in table 4, such situations often occur. In the simulation, the number of runs in which $\hat{r}_{12}$ was not obtained was 17 for $r_{12} = 0.0$ and 14 for $r_{12} = 0.5$ out of 100. Even when $\hat{r}_{12}$ could be obtained, it was unstable, showed large sampling errors and m.s.e. and often took values beyond the proper range $[-1, 1]$, as shown in table 3. For these problems the classical and modified estimators of the genetic correlation coefficient are of little use in practice.

On the other hand, r_{12}^* can always be obtained. Moreover r_{12}^* exhibits much smaller variance and m.s.e. than $\hat{r}_{12}$ and $\tilde{r}_{12}$ and takes values within $[-1, 1]$ as shown in table 3. Therefore r_{12}^* is superior to and more useful in practice than $\hat{r}_{12}$ and $\tilde{r}_{12}$. We should note that the results presented here were obtained with a uniform prior distribution in $[-1, 1]$ in the absence of prior information except for the proper range about the genetic correlation.

Though the hybrid bayesian estimation of genetic parameters is suitable, there is room for improvement of this method. One of the problems is how to incorporate the information contained in data into the shape of prior distribution function. In this regard the use of classical estimator as prior information could be considered. Moreover the hybrid bayesian estimation of genetic parameters in other mating systems such as selfing should be formulated. These subjects will be discussed elsewhere.

Literature Cited

- BULMER, M.G. 1980. Biometrical analysis of outbred populations. In "The mathematical theory of quantitative genetics." Clarendon press, Oxford. 73 ~ 104.
- GILL, J.L. and E.L. JENSEN 1968. Probability of obtaining negative estimates of heritability. *Biometrics* 24:517 ~ 526.
- HILL, W.G. and R. THOMPSON 1978. Probabilities of non-positive definite between-group or genetic covariance matrices. *Biometrics* 34:429 ~ 439.
- KENPTHORNE, O. 1957. Inheritance of quantitative characters in a random mating population. In "An introduction to genetic statistics." John Wiley & Sons, Inc., London. 308 ~ 342.
- MCCALLA, T.R. 1967. Gauss-Legendre Quadrature. In "Introduction to numerical methods and Fortran programming." John Wiley & Sons, Inc. 280 ~ 287.
- TORO, M.A. and J.V. PRUÑONOSA 1984. The use of prior information in the estimation of heritability by parent-offspring regression. *Genet. Sel. Evol.* 16:177 ~ 184.

遺伝パラメータのベイズ推定

斎尾乾二郎・林 武司

(東京大学農学部, 東京都文京区, 〒113)

計量形質の遺伝率や遺伝相関係数などの遺伝パラメータの推定において, 従来の方法(ここでは古典的推定という)によって得られる推定量は標本誤差が大きく, しばしばその値がパラメータのもつ固有の範囲(遺伝率では $[0, 1]$, 遺伝相関係数では $[-1, 1]$) を越えてしまうなどの問題点がある. これらの問題点を解消するために, 事前の情報を取り入れるベイズ推定を導入した. 推定すべきパラメータに関する事前情報は, その固有の範囲の上で定義される事前分布にとりこまれ, 得られたデータと組み合わせられパラメータの事後分布が与えられる. しかし事後分布は他のパラメータ, 表現型平均や表現型分散, 共分散などにも依存している. これらのパラメータに関しては, 最尤法やモーメント法など従来の方法によって得られる推定量が, 良い統計的性質を備えていることが知られている. よって表現型平均や表現型分散, 共分散などのパラメータはそれらの古典的推定量で置き換えて事後分布を構成し, 推定すべき各遺伝パラメータの事後分布にもとずく期待値を数値積分によって求め, ベイズ推定量を得ることにした. このようにして古典的推定量と組み合わせるベイズ推定量を得る方法を, 我々は“ハイブリッドベイズ推定”と呼びその推定量を“ハイブリッドベイズ推定量”と名付けた. 本論文ではハイブリッドベイズ推定量の統計的性質をモンテカルロシミュレーションにより半兄弟モデルについて調べ, 古典的推定量, および古典的推定量でその値が固有の範囲を越えたものはその境界の値にした修正された古典的推定量との比較を行なった. 遺伝率だけを推定する一形質に着目した場合と遺伝率と遺伝相関係数の推定を同時に行なう二形質に着目した場合について調べた. 一形質の場合, 事前分布として遺伝率が固有の範囲内にその値をとる, という事以外には利用できる事前情報がないうち $[0, 1]$ 上の一様分布を, それ以外にも事前情報があるときにはそれを反映させたベータ分布を用いた. いずれの場合もハイブリッドベイズ推定量は, その値が固有の範囲内にあり, 不適当な事前分布を用いた場合を除いて, 他の二つの推定量に比べ平均二乗誤差が小さいなどの望ましい性質を備えていることが示された(Table 1, 2). また二形質の場合, 事前分布としての遺伝率については $[0, 1]$ 上の, 遺伝相関係数については $[-1, 1]$ 上の一様分布をそれぞれ用いた. 遺伝率については一形質の場合と同様の結果がえられ, 遺伝相関係数についてはハイブリッドベイズ推定量のもつ望ましい性質がより顕著に示された(Table 3, 4). つまり古典的推定量は推定不能の場合があり, かつ可能な場合も著しく固有の範囲を逸脱することがあるのに対しハイブリッドベイズ推定量はつねに推定可能であり, 平均二乗誤差が小さくその値は常に固有の範囲内にあった.

遺伝率, 遺伝相関係数などの遺伝パラメータは選抜のさいに重要な指標であるにもかかわらず, その古典的推定量は信頼のおける値が得られる場合が少なく従来育種の実場面での有用性はあまり大きくなかった. しかしシミュレーションの結果からみられるようにハイブリッドベイズ推定によって得られる推定量は望ましい統計的性質を備えておりその値には信頼がおける. よってハイブリッドベイズ推定の導入により遺伝率, 遺伝相関係数などの遺伝パラメータの育種における実用性が増大すると思われる.