Genet. Sel. Evol. 33 (2001) 3–16

Original article

Hypothesis testing for the genetic background of quantitative traits

Luis Alberto GARCÍA-CORTÉS[a,*], Carlos CABRILLO[b], Carlos MORENO[a], Luis VARONA[c]

[a] Departamento de Genética, Facultad de Veterinaria, c/ Miguel Servet, 177, 50013 Zaragoza, Spain
[b] Instituto de Óptica "Daza de Valdés", CSIC, Serrano 121, 28006 Madrid, Spain
[c] Área de Producción Animal, Centro UdL-IRTA, Lleida, Spain

(Received 8 November 1999; accepted 29 September 2000)

Abstract – The testing of Bayesian point null hypotheses on variance component models have resulted in a tough assignment for which no clear and generally accepted method exists. In this work we present what we believe is a succeeding approach to such a task. It is based on a simple reparameterization of the model in terms of the total variance and the proportion of the additive genetic variance with respect to it, as well as on the explicit inclusion on the prior probability of a discrete component at origin. The reparameterization was used to bypass an arbitrariness related to the impropriety of uninformative priors onto unbounded variables while the discrete component was necessary to overcome the zero probability assigned to sets of null measure by the usual continuous variable models. The method was tested against computer simulations with appealing results.

animal breeding / prior distribution / Bayes factor / hypothesis testing / heritability

1. INTRODUCTION

Probably one of the most natural questions arising when dealing with quantitative traits is just whether the trait considered has a genetic background or not. In addition, in some circumstances such as the initial stages of genetic analysis, where the interest is aimed at predicting if the future artificial selection will produce or not a genetic response, this can be a very relevant question from both a genetic and economical point of view. Even in fields like animal breeding, where the effects of genes on economically interesting traits is in principle evident, there are some cases where hypothesis testing of the genetic background could be useful. For instance, it could be the case with chickens where the strong artificial selection in the past has reduced the genetic variability, or in small populations, where most of the alleles of the interesting genes

* Correspondence and reprints
E-mail: agarcor@posta.unizar.es

have disappeared by chance. Apparently quite a harmless question, hypothesis testing of the genetic background has resulted in a tough assignment.

Just to be specific, let us concentrate on hierarchical unbalanced mixed linear models, those commonly used in animal breeding, to analyse the genetic background of quantitative traits [12]. In these models, the genetic and environmental components of variance are traditionally estimated *via* either restricted maximum likelihood or through Bayesian inference [14]. Testing of the null hypothesis of the additive variance component of the genetic part of the model will answer whether genes influence the trait under analysis or not. At least in the case of hierarchical unbalanced mixed linear models, this task does not have a clear and generally accepted method. The null point hypothesis on variance components is exactly at the lower bound of the parametric space, where the asymptotic properties of the likelihood ratio tests fail [23]. Under the Bayesian framework, hypothesis testing is usually analysed by calculating the Bayes factor (BF) which retains good behavior even when the hypothesis to be tested is close or even at the boundary of the parametric space. Several methods to calculate BF, most of them approximated BF, have been described, such as the intrinsic BF [3], the fractional BF [21] or the posterior BF [1]. Moreover, numerical alternatives have been proposed, such as the reversible jump [10], incorporation of model indicators [4,7] or the harmonic mean estimator [20]. However, in spite of all of these efforts (see Kass and Raftery [17] for a nice review), the problem of testing the point null hypothesis on variance component models remains unsolved.

Gelman *et al.* [6] suggested that testing point hypothesis for continuous variables like variance components is not reasonable, simply because the probability of any set of null measure is zero for a continuous random variable. In this present article it is shown constructively that to test the null point hypothesis is indeed reasonable. In other words, we present here a method to implement the null-hypothesis test free of any inconsistency. We have identified two main causes as the origin for these possible inconsistencies: firstly, the zero probability assigned to isolated points in the usual probability models for continuous variables; secondly, the impropriety of the usual flat prior for unbounded variables. In the following we show by mixing continuous and discrete probability models how it is possible to overcome the former while for the latter a simple reparameterization of the standard model will do the task. Numerical simulations reveal the excellent behavior of our proposal.

2. MODELS

The mixed linear model considers both environmental and genetic effects. The model equation for the observed phenotypes ($\mathbf{y}$) is defined as

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

where $\mathbf{b}$ is a vector of environmental effects, $\mathbf{X}$ is the incidence matrix relating $\mathbf{b}$ and $\mathbf{y}$, $\mathbf{u}$ is a vector of sire effects, $\mathbf{Z}$ is the incidence matrix relating $\mathbf{u}$ and $\mathbf{y}$, and $\mathbf{e}$ is the vector of residuals.

Two models will be considered below: the standard between-sires variance model and the heritability model. This latter one will be revealed as superior as far as testing the null hypothesis is concerned.

2.1. Between sires variance component model

This mixed linear model can be regarded as a hierarchical Bayesian model, where $\mathbf{u}|\sigma_u^2 \sim N\left(\mathbf{0}, \sigma_u^2\mathbf{I}\right)$ and $\mathbf{e}|\sigma_e^2 \sim N\left(\mathbf{0}, \sigma_e^2\mathbf{I}\right)$. In this case, we will test the null hypothesis on the between-sires variance component (σ_u^2), a quite usual simplifying assumption. For instance, Wang *et al.* [25], described a Bayesian implementation of the Gibbs sampler using this model. On the contrary, in animal breeding, dams are in general artificially inseminated and hence there is no environmental contact between the sire and their progeny. Therefore, differences between progeny groups are obviously explained just by the genes of the sire, making this model widely used in animal breeding for inferences about the heritability of the quantitative traits. Anyhow, the conclusions reached here do not depend on this simplification.

2.2. Heritability model

Here we define a model involving two alternative variables: the phenotypic variance, which is defined as, $\sigma^2 = \sigma_u^2 + \sigma_e^2$, and the coefficient of heritability, obtained from the classical expression $h^2 = 4\sigma_u^2/\left(\sigma_u^2 + \sigma_e^2\right)$. A similar model, including heritability and residual variance as unknowns, was analysed in Theobald *et al.* [24]. We will test for the presence of genes by testing the null hypothesis on the heritability. The heritability coefficient represents the proportion of variance explained by the genes involved in the performance trait. Usually, in the animal breeding context, $4\sigma_u^2$ is called the additive genetic variance.

The hierarchical model now considers $u|\sigma^2, h^2 \sim N\left(\mathbf{0}, 0.25\sigma^2h^2\mathbf{I}\right)$ and $e|\sigma^2, h^2 \sim N\left[\mathbf{0}, \sigma^2\left(1 - 0.25h^2\right)\mathbf{I}\right]$.

3. METHODS

Standard probability distributions for the description of continuous random variables rely on the Riemann integral to assign probabilities to intervals. Consequently, the probability of any set of zero measure, such as an isolated point, is also zero. However, this is just a property of a class of probability models. There is nothing fundamental impeding us to assign finite probabilities to isolated points belonging to the real line. But their random behavior should be described using discrete distribution functions rather than its continuous counterpart. Nor is there any inconsistency in mixing both types of distributions. More specifically, we would need a probability distribution, $F(x) = P(\theta \leqslant x)$, with discontinuities at each isolated point with finite probability. In view of all this, including the possibility of a finite probability of null genetic variance in a Bayesian inference scheme is just a matter of including a discontinuity at

zero in the prior. However, we would like to work with probability densities rather than with distributions as is usual in the field. At a first glance, this is a desperate task, since we need to derive $F(x)$ at a discontinuity. In other words, we would need the derivative function of the Heaviside step function $H(x)$ ($H(x) = 0$, $\forall x < 0$; $H(x) = 1$, $\forall x > 0$). Surprisingly enough, it is possible to give a rigorous mathematical meaning to such an object within what is called the theory of generalized functions. This is known as the Dirac delta and is usually denoted by $\delta(x)$. The interested reader can find a readable but quite rigorous introduction to the Dirac delta and the theory of generalized functions in Griffel [11]. For our purposes here, it suffices to know the following properties applied blindly, namely

$$\delta(x) = 0 \text{ for } x \neq 0, \int_{-\infty}^{\infty} \delta(x)\mathrm{d}x = 1 \qquad (1)$$
$$\int_0^a \delta(x) f(x)\mathrm{d}x = \frac{1}{2} f(0) \text{ if } a > 0.$$

Needless to say that the integrals in (1) are not of the Riemann kind but they must be understood in the generalized function sense. The reader can convince himself that given (1), the density probability distribution of a continuous and positive random variable with a finite probability of being exactly zero have the form

$$g(\theta) = P(H_0)\, 2\delta(\theta) + P(H_1)\, f(\theta) \qquad (2)$$

where θ is the variable to be tested, $f(\theta)$ is a sensible conventional probability density integrating to one, and $P(H_0)$ and $P(H_1)$ are the probabilities assigned to the null and alternative hypotheses respectively. We stress that although this goes beyond the classical function theory, there is nothing poorly defined within this framework. We will take equation (2) as the general form for our prior density probabilities. Except for the factor of two necessary for mathematical consistency when a discrete component is at the lower boundary of the parametric space, (2) coincides with the prior density used by Berger and Sellke [2].

After a prior density is defined, the posterior probability of both the null and the alternative hypotheses can be obtained from the marginal posterior provided by the Bayesian analysis

$$P(H_0|\mathbf{y}) = \lim_{\varepsilon \to 0} \int_0^{\varepsilon} g(\theta|\mathbf{y})\, \mathrm{d}\theta \qquad (3)$$

$$P(H_1|\mathbf{y}) = \lim_{\varepsilon \to 0} \int_{\varepsilon}^{a} g(\theta|\mathbf{y})\, \mathrm{d}\theta \qquad (4)$$

where a is the upper limit of the parametric space for θ (assumed positive). We stress that these integrals are mathematically rigorous under the generalized function theory. Usually, the Bayesian analysis is implemented *via* an MCMC algorithm, such as the Gibbs sampler. Although, the Dirac delta can be implemented numerically, it is far easier to use just $f(\theta)$ as an operational prior from

which $g(\theta)$ can be obtained as

$$g(\theta|\mathbf{y}) \propto \frac{g(\theta)}{f(\theta)} f(\theta|\mathbf{y}) \tag{5}$$

being $f(\theta|\mathbf{y})$ the conventional posterior obtained by using $f(\theta)$. Formula (5) can be regarded as an important sampling procedure, the usual tool for replacing *a priori* distributions for a sensitivity analysis [22].

Replacing (2) in (5) and assuming $P(H_0) = P(H_1) = 0.5$, making our prior uninformative (following the "canonical" Bayesian approach we will include all our prior information at the model level) results in

$$g(\theta|\mathbf{y}) \propto \frac{\delta(\theta) + 0.5 f(\theta)}{f(\theta)} f(\theta|\mathbf{y}) = \left(\frac{\delta(\theta)}{f(\theta)} + 0.5\right) f(\theta|\mathbf{y}).$$

Then, integrating this function for the hypothesis intervals, following the formulas in (3) and (4)

$$P(H_0|\mathbf{y}) \propto \lim_{\varepsilon \to 0} \int_0^{\varepsilon} \left(\frac{\delta(\theta)}{f(\theta)} + 0.5\right) f(\theta|\mathbf{y})\, \mathrm{d}\theta$$
$$P(H_1|\mathbf{y}) \propto \lim_{\varepsilon \to 0} \int_{\varepsilon}^{a} \left(\frac{\delta(\theta)}{f(\theta)} + 0.5\right) f(\theta|\mathbf{y})\, \mathrm{d}\theta$$

and using the properties (1) of the Dirac delta

$$P(H_0|\mathbf{y}) \propto \frac{f(\theta = 0|\mathbf{y})}{2 f(\theta = 0)}$$
$$P(H_1|\mathbf{y}) \propto \int_0^{a} \frac{1}{2} f(\theta|\mathbf{y})\, \mathrm{d}\theta = \frac{1}{2}.$$

Since $P(H_0|\mathbf{y}) + P(H_1|\mathbf{y}) = 1$ we have

$$P(H_0|\mathbf{y}) = \frac{f(\theta = 0|\mathbf{y})}{f(\theta = 0|\mathbf{y}) + f(\theta = 0)} \tag{6}$$

$$P(H_1|\mathbf{y}) = \frac{f(\theta = 0)}{f(\theta = 0|\mathbf{y}) + f(\theta = 0)} \tag{7}$$

so that the ratio, which represents the Bayes factor against the null hypothesis, is given by

$$\frac{P(H_1|\mathbf{y})}{P(H_0|\mathbf{y})} = \frac{f(\theta = 0)}{f(\theta = 0|\mathbf{y})}. \tag{8}$$

Testing the null hypothesis only requires the ordinate at zero of the conventional marginal posterior density. This result is equivalent to that presented in Berger and Sellke [2].

3.1. Testing the null hypothesis on the between sires variance component model

Once the model has been established and under the assumption that it conveys all our prior information, we will take flat priors on the levels of **b** and the residual variance component while for σ_u^2 we use (2) with a flat $f(\theta)$ with density k. All together this amounts to

$$g\left(\sigma_u^2\right) = \delta\left(\sigma_u^2\right) + 0.5k \quad \text{if} \quad \sigma_u^2 \in \left[0, \frac{1}{3}\sigma_e^2\right].$$

We know *a priori* that the between sires variance component should not exceed the third part of the residual variance [24]. From the causal point of view, the between sires variance component involves the fourth part of the variance explained by the genes, and the residual variance involves three fourths of the variance explained by the genes plus the variance due to the environment.

The posterior distribution is

$$g\left(\mathbf{b}, \sigma_u^2, \sigma_e^2|\mathbf{y}\right) \propto f\left(\mathbf{y}|\mathbf{b}, \sigma_u^2, \sigma_e^2\right) f\left(\mathbf{b}\right) g\left(\sigma_u^2\right) f\left(\sigma_e^2\right).$$

An operational flat prior on σ_u^2 is used to implement the Bayesian analysis. Now, the posterior distribution is proportional to

$$f\left(\mathbf{b}, \sigma_u^2, \sigma_e^2|\mathbf{y}\right) \propto |\mathbf{V}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\mathbf{y} - \mathbf{Xb}\right)' \mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{Xb}\right)\right\},$$

where **V** is presented as a function of the variance components

$$\mathbf{V} = \sigma_u^2 \mathbf{ZZ}' + \sigma_e^2 \mathbf{I}.$$

Replacing this formula in the joint posterior distribution

$$f\left(\mathbf{b}, \sigma_u^2, \sigma_e^2|\mathbf{y}\right) \propto \left|\sigma_u^2 \mathbf{ZZ}' + \sigma_e^2 \mathbf{I}\right|^{-\frac{1}{2}} \times \exp\left\{-\frac{1}{2}\left(\mathbf{y} - \mathbf{Xb}\right)' \left(\sigma_u^2 \mathbf{ZZ}' + \sigma_e^2 \mathbf{I}\right)^{-1}\left(\mathbf{y} - \mathbf{Xb}\right)\right\}.$$

In this case, it is unnecessary to carry out the whole Bayesian analysis to realise the logically absurd results that the procedure offers. We can analyse just the ordinate at the origin of the Gibbs conditional of σ_u^2

$$f\left(\sigma_u^2 = 0|\mathbf{b}, \sigma_e^2, \mathbf{y}\right) \propto \left(\sigma_e^2\right)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_e^2}\left(\mathbf{y} - \mathbf{Xb}\right)' \left(\mathbf{y} - \mathbf{Xb}\right)\right\}$$

which is a positive value for any value of σ_e^2 greater than 0. Using the Rao-Blackwell argument, the average of these conditional densities, that is, the marginal density of σ_u^2, will also be greater than zero at $\sigma_u^2 = 0$, *i.e.*,

$$f\left(\sigma_u^2 = 0|\mathbf{y}\right) = a > 0.$$

Posterior odds of the null and the alternative hypothesis can be obtained from formula (6) and (7)

$$P(H_0|\mathbf{y}) = \frac{f\left(\sigma_u^2 = 0|\mathbf{y}\right)}{f\left(\sigma_u^2 = 0|\mathbf{y}\right) + f\left(\sigma_u^2 = 0\right)} = \frac{a}{a+k} \tag{9}$$

$$P(H_1|\mathbf{y}) = \frac{f\left(\sigma_u^2 = 0\right)}{f\left(\sigma_u^2 = 0|\mathbf{y}\right) + f\left(\sigma_u^2 = 0\right)} = \frac{k}{a+k} \tag{10}$$

that is, this model leads to a result that depends on k, the ordinate of the prior density. Considering the arbitrariness of k, the result turns out useless when the prior is improper.

3.2. Testing the null hypothesis on the heritability-phenotypic variance model

Again, we consider flat unbounded prior distributions for $\mathbf{b}$ and σ^2. The prior distribution for h^2 is now given by

$$g\left(h^2\right) = \delta\left(h^2\right) + 0.5 \quad \text{if} \quad h^2 \in [0,1].$$

Hobert and Casella [13] studied the effect of improper priors on the propriety of the posterior distributions in cases of hierarchical linear models. They show that in general, propriety of the posterior is mathematically difficult to probe or impossible, and they propose to always use proper priors to painlessly circumvent this problem. This is precisely the advantage of the heritability model since now, for the relevant variable h^2, the prior probability density, although it is flat is perfectly proper.

As in the previous case, we consider an operational proper prior distribution of the heritability coefficient similar to $g(h^2)$ but ignoring the Dirac delta at zero

$$f\left(h^2\right) = 1 \quad \text{if} \quad h^2 \in [0,1].$$

The posterior distribution of the parameters given data is

$$f\left(\mathbf{b}, \sigma^2, h^2|\mathbf{y}\right) \propto f\left(\mathbf{y}|\mathbf{b}, \sigma^2, h^2\right) f\left(\mathbf{b}, \sigma^2, h^2\right) \propto f\left(\mathbf{y}|\mathbf{b}, \sigma^2, h^2\right)$$

that is,

$$f\left(\mathbf{b}, \sigma^2, h^2|\mathbf{y}\right) \propto |\mathbf{V}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\mathbf{y} - \mathbf{X}\mathbf{b}\right)' \mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\mathbf{b}\right)\right\} \tag{11}$$

with,

$$\mathbf{V} = \sigma^2\left[0.25h^2\left(\mathbf{Z}\mathbf{Z}' - \mathbf{I}\right) + \mathbf{I}\right]. \tag{12}$$

Before replacing (11) in (12), we reduce $\mathbf{V}$ to its canonical form in order to avoid its repeated inversion. Lin [18] described this procedure in an EM-REML context.

$$\mathbf{V}^{-1} = \sigma^{-2}\mathbf{U}\left[0.25h^2\mathbf{D} + \mathbf{I}\right]^{-1}\mathbf{U}' \tag{13}$$

where the columns of $\mathbf{U}$ contain the eigenvectors of $(\mathbf{ZZ}' - \mathbf{I})$ and $\mathbf{D}$ is a diagonal matrix containing their corresponding eigenvalues. Cases including animals related by coancestry will include an extra non-diagonal matrix. For these cases, Lin [19] proposed the simultaneous diagonalization of two matrices, a procedure that could be used here if necessary. Replacing (13) in (11) :

$$f\left(\mathbf{b}, \sigma^2, h^2|\mathbf{y}\right) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \left|0.25h^2\mathbf{D} + \mathbf{I}\right|^{-\frac{1}{2}} \times \exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{y} - \mathbf{Xb}\right)' \mathbf{U}\left[0.25h^2\mathbf{D} + \mathbf{I}\right]^{-1} \mathbf{U}'\left(\mathbf{y} - \mathbf{Xb}\right)\right\}.$$

Some remarks about the Gibbs sampler implementation to obtain the marginal posterior distribution of h^2 are presented in the Appendix.

Probabilities of the null and alternative hypotheses were obtained using formulas (6) and (7) and the conventional marginal posterior $f\left(h^2|\mathbf{y}\right)$.

$$P\left(H_0|\mathbf{y}\right) = \frac{f\left(h^2 = 0|\mathbf{y}\right)}{f\left(h^2 = 0|\mathbf{y}\right) + 1} \qquad (14)$$

$$P\left(H_1|\mathbf{y}\right) = \frac{1}{f\left(h^2 = 0|\mathbf{y}\right) + 1}.$$

This time the prior distribution is a well-defined proper distribution yielding well-defined unambiguous posterior probabilities.

4. RESULTS ON SIMULATIONS

4.1. Simulation

In order to ascertain the value of the method for testing the null hypothesis, nine cases with three different true heritabilities (0, 0.1 and 0.2) and three different data sizes (500, 1 000 and 2 000) were analyzed. Twenty replicates were simulated for each case. All replicates included 25 fixed levels of $\mathbf{b}$, whose true values were arbitrarily sampled from a uniform distribution within the interval (100–200). The true phenotypic variance was arbitrarily set to 100 in all cases. The number of sires was 50, 100 or 200, depending on the data set size. True values of the sires were sampled from a normal distribution with null mean and the sire variance was equal to 0, 2.5 or 5, depending on the heritability. Records were obtained from a fixed level and a sire was chosen at random. The residual variance equaled 100, 97.5 or 95, depending on the heritability. These 180 data samples were analysed using a Gibbs sampler algorithm. The chain length was 1 100 and the first 100 iterations were discarded as burn-in. Marginal densities at the origin were obtained *via* the Rao-Blackwell argument.

4.2. Results

The average results of the twenty replicates are presented in Table I. The fourth column presents the ordinates at zero of the operational posterior distribution of the heritability coefficient given the data. The fifth column presents

Table I. Ordinate at zero of the conventional marginal posterior distribution, posterior probability of the null hypothesis, and Bayes Factor against both the null and the alternative hypothesis. Average results of 20 replicates.

Case	h^2	Sires	$f\left(h^2=0\|\mathbf{y}\right)$	$p\left(H_0\|\mathbf{y}\right)$	BF(1/0)*	BF(0/1)*	LLRT**
1	0	50	8.2753	0.8401	20–0–0–0	4–10–6–0	0.9505(2)
2	0.1	50	4.4263	0.7030	20–0–0–0	10–8–2–0	1.3560(1)
3	0.2	50	2.1099	0.4394	13–3–3–1	16–3–1–0	6.2034(11)
4	0	100	11.4483	0.8931	20–0–0–0	2–6–12–0	0.6105(0)
5	0.1	100	1.8600	0.4544	14–1–4–1	17–3–0–0	2.8713(7)
6	0.2	100	0.2785	0.1561	3–10–2–5	20–0–0–0	7.6374(14)
7	0	200	20.8802	0.9430	20–0–0–0	0–0–0–20	0.7093(1)
8	0.1	200	4.0054	0.4294	12–6–1–1	15–3–2–0	7.2953(13)
9	0.2	200	0.2203	0.0996	2–4–5–9	20–0–0–0	21.2338(20)

* The 20 replicates classified as four types: no evidence, substantial evidence, strong evidence and decisive evidence.

** Average of the log likelihood ratio test of the 20 replicates. Between parentheses, the number of replicates where the null hypothesis was rejected with $p = 0.05$.

the probability of the null hypothesis calculated from formula (14) for each replicate. Both the sixth and the seventh column correspond to the BF against both the null and alternative hypotheses respectively. As Kass and Raftery [17] suggest, we considered no evidence against the hypothesis when the BF was smaller than 3.2, substantial evidence when the BF was within 3.2 and 10, strong evidence when the BF was within 10 and 100 and decisive evidence when the BF was greater than 100. Thus, from the 20 replicates analysed for case 6, there was no evidence against the null hypothesis in 3 replicates, the evidence was substantial for 10 replicates, strong for 2 replicates and decisive for 5 replicates. The last column of this table shows the average of the log likelihood ratio test (LLRT) for the twenty replicates of each case. It also shows between parentheses, the number of replicates where the null hypothesis was rejected with a significance probability of 0.05.

In Figure 1, the operational marginal posterior density $f\left(h^2|\mathbf{y}\right)$ is illustrated for the first replicate of cases 1 to 3. Taking case 3 as an example, its value at the origin, *i.e.*, $f\left(h^2=0|\mathbf{y}\right)$, corresponded to 0.3810. The posterior probabilities of the null and alternative hypotheses are then given by

$$P\left(H_0|\mathbf{y}\right) = \frac{0.3810}{0.3810+1} = 0.2959$$

$$P\left(H_1|\mathbf{y}\right) = \frac{1}{0.3810+1} = 0.6041.$$

Cases 1, 4 and 7 of Table I show that the posterior probability of the null hypothesis increased when the number of data increased and the true heritability was null. On the contrary, the probability of the null hypothesis was decreased by increasing the number of data as illustrated in the remaining cases where the true heritability was set to 0.1 or 0.2. Table I reveals an excellent behavior of our approach as far as testing the null hypothesis is concerned.

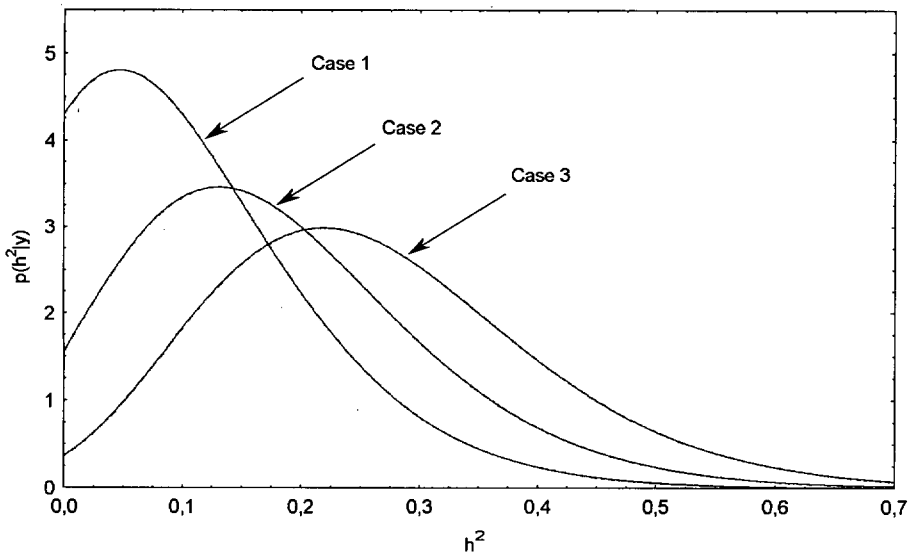

Figure 1. Conventional marginal posterior distribution of the heritability coefficient for the first replicate of cases 1 to 3.

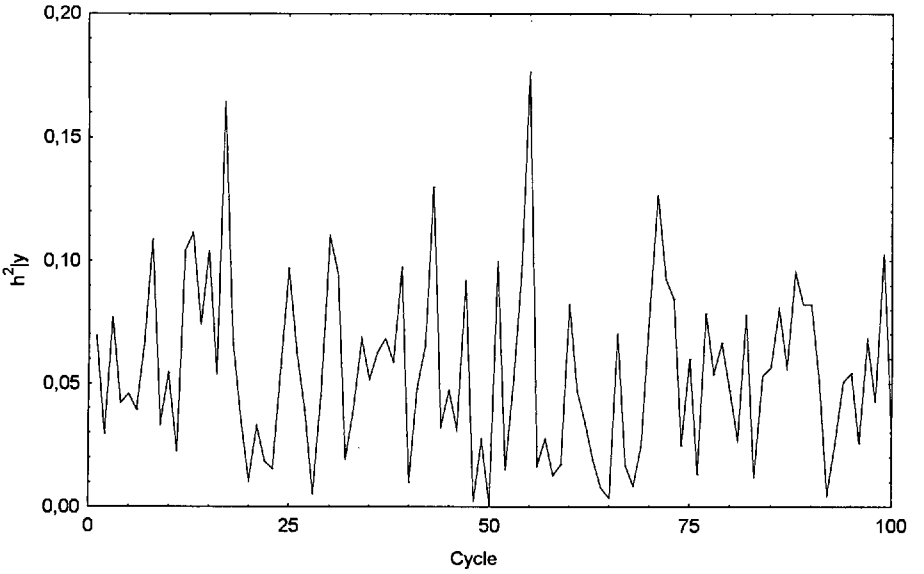

Figure 2. Path of the first 100 cycles after burn-in of the last replicate of case 1.

The average sample sizes of the Gibbs chains over the 20 replicates, obtained as in Geyer [8], are presented in Table II. These results show the fast mixing of the Markov chains in all cases. For instance, Figure 2 contains the path of the first 100 cycles after burn-in of the last replicate of case 1.

Table II. Average effective sample size for the 9 cases and phenotypic variance and heritability.

Case	σ^2	h^2
1	193.4	376.3
2	192.7	287.4
3	225.6	539.8
4	332.0	441.4
5	262.0	691.4
6	475.7	311.3
7	554.9	708.5
8	455.2	779.5
9	630.0	744.3

5. DISCUSSION

In the light of the previous results, some comments are necessary concerning the Bayesian hypothesis testing presented here. In the first place, the failure of the variance component model along with the success of the heritability one does not imply any inconsistency of the Bayesian inference approach. On the contrary it yields a consistent output. The arbitrariness of the posterior probabilities is a direct reflection of the arbitrariness of the prior density. We are purposely avoiding the word prior probability, since the prior density used does not lead to a probability. Bayes' theorem is a well-established mathematical result relating probabilities and as such, consistency is guaranteed when it is applied to probabilities. But we are pushing its application outside the realm in which the theorem is proven. It is not surprising if along the way we get some useless or apparently absurd result. Notice also that the arbitrariness of k is caused by the arbitrariness of the prior density present only because the restriction of a total area equaling one does not apply. It has nothing to do with the subjective character of Bayesian inference.

It can be argued, however, that in fact k is not so arbitrary and although usually not stated explicitly it is an infinitesimal quantity. Putting it mathematically, the prior uninformative density is not just a constant function but the limit of a succession of proper flat distributions, *i.e.*,

$$f\left(\sigma_u^2\right) = \lim_{k \to 0} k \quad \text{if} \quad k \in \left[0, k^{-1}\right]. \tag{15}$$

Operationally, (15) implies taking the limit $k \to 0$ after performing the calculations, thus making any arbitrariness disappear. Immediately, formulae (9) and (10) led us to an "absurd" result, namely the posterior probability of the null hypothesis equal to one and zero for the alternative, whatever the evidence added by the data. It then seems, that there is some inconsistency in the Bayesian inference itself that invalidates its application to certain questions or models. But again this was only apparent. The Bayesian approach yielded a reasonable output. Although, definition (15) led to a probability

between 0 and 1, the limit process spoiled the countable additivity, and again the associated "probability" was not well defined. More specifically, it assigned zero probability to any finite interval (although it gives the correct answer for the probability ratio between two finite intervals). The Bayesian inference is doing nothing else than reflecting this "*a priori*" absurd.

The above argumentation does not mean that every Bayesian method based on an unbounded flat prior is wrong. It simply means that it is necessary to be aware of the improper character of the prior. Depending on the problem addressed, inconsistencies will show up or not. Choosing heritability and phenotypic variance as the basic variables (as a matter of fact, a time honored election in genetics) is just to re-encode the information contained in two unbounded variables in one representing the total and the proportion of the second with respect to the total (obviously bounded). The impropriety inherent to a flat prior on unbounded variables is still there but behind the scene on the total variance. The relevant variable, that is, heritability, is now armed with a perfectly proper prior density leading to a well-defined prior probability. The impropriety has been hidden rather than eliminated but this is enough for our purposes.

The conventional likelihood ratio test implemented on the 180 replicates also showed good results, but it only provides whether the observed likelihood ratio is within the rejecting interval or not. Frequentist approaches do not provide probabilities of both the null and the alternative hypotheses, as the BF does. Furthermore, the hypothesis test is implemented integrating out the variables of the model using the BF, while LLRT only the most probable values are considered. Future research should be aimed to the numerical comparison of the frequentist properties of LLRT and BF.

With regard to the possible outcomes of a Bayesian analysis, Jaynes [15,16] proposed a very interesting idea, namely, that a Bayesian statistician can never say that a result is absurd. Saying that a result is logically absurd implies that we are using information that is not included in the prior or the data, although the prior should include all information we have before the analysis. Hence, what we are saying really is that our prior knowledge about the problem was not well modeled. This was the rational which guided our search. In view of the results, it has resulted in being a very useful one.

REFERENCES

[1] Aitkin M., Posterior Bayes factors (with discussion), J. Roy. Stat. Soc. B. 53 (1991) 111–142.

[2] Berger J.O., Sellke T., Testing a null hypothesis: The irreconcilability of significance levels and evidence, J. Am. Stat. Assoc. 82 (1987) 112–122.

[3] Berger J.O., Perrichi L., The intrinsic Bayes factor for model selection and prediction, J. Am. Stat. Assoc. 91 (1996) 109–122.

[4] Carlin B.P., Chib S., Bayesian model choice *via* Markov chain Monte Carlo, J. Roy. Stat. Soc B. 57 (1995) 473–484.

[5] Gelfand A., Smith A.F.M., Sampling based approaches to calculating marginal densities, J. Am. Stat. Assoc. 85 (1990) 398–409.

[6] Gelman A., Carlin J.B., Stern H.S., Rubin D.B., Bayesian data analysis, Chapman and Hall, London, 1995.

[7] George E.I., McCullogh R.E., Variable Selection *via* Gibbs Sampling, J. Am. Stat. Assoc. 88 (1993) 881–889.
[8] Geyer C.J., Practical Markov chain Monte Carlo (with discussion), Stat. Sci. 7 (1992) 467–511.
[9] Gilks W.R., Wild P., Adaptative rejection sampling for Gibbs sampling, Appl. Stat. 41 (1992) 337–348.
[10] Green P.J., Reversible Jump Markov Chain Monte Carlo computation and Bayesian model determination, Biometrika 82 (1995) 711–732.
[11] Griffel D.H., Applied functional analysis, Ellis Horwood limited, Chichester, 1981.
[12] Henderson C.R., Applications of linear models in animal breeding, University of Guelph, Ontario, 1984.
[13] Hobert J.P., Casella G., The effect of improper priors on Gibbs sampling in hierarchical linear mixed models, J. Am. Stat. Assoc. 91 (1996) 1461–1473.
[14] Hofer A., Variance component estimation in animal breeding: A review, J. Anim. Breed. Gen. 115 (1998) 247–265.
[15] Jaynes E.T., Bayesian methods: General background, in: Justice J.H. (Ed.), Maximum entropy and Bayesian methods in applied statistics, Cambridge University Press, Cambridge, 1986, pp. 1–25.
[16] Jaynes E.T., Monkeys, Kangaroos and N, in: Justice J.H. (Ed.), Maximum entropy and Bayesian methods in applied statistics, Cambridge University Press, Cambridge, 1986, pp. 1–25.
[17] Kass R.E., Raftery A.E., Bayes factors, J. Am. Stat. Assoc. 90 (1995) 773–795.
[18] Lin C.Y., Application of singular value decomposition to restricted maximum likelihood estimation of variance components, J. Dairy Sci. 70 (1987) 2680–2684.
[19] Lin C.Y., Four equivalent sets of mixed model equations with relationship matrix for estimation of genetic parameters, J. Anim. Sci. 66 (1988) 1627–1635.
[20] Newton M.A., Raftery A.E., Approximate Bayesian inference with the weighted likelihood bootstrap, J. Roy. Statist. Soc. B. 56 (1994) 3–48.
[21] O' Hagan A., Fractional Bayes Factor for model comparison (with discussion), J. Roy. Stat. Soc. B 57 (1995) 99–138.
[22] Sorensen D., Gibbs sampling in quantitative genetics, Tech. report. No. 82, Nat. Inst. of Anim. Sci., Foulum, 1996.
[23] Stram D.O., Lee J.W., Variance components testing in the longitudinal mixed effect model, Biometrics 50 (1994) 1171–1177.
[24] Theobald C.M., Firat M.Z., Thompson R., Gibbs sampling, adaptative rejection sampling and robustness to prior specification for a mixed linear model, Genet. Sel. Evol. 29 (1997) 57–72.
[25] Wang C.S., Rutledge J.J., Gianola D., Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs, Genet. Sel. Evol. 26 (1994) 91–115.

APPENDIX

The Gibbs sampler implementation of the conventional heritability model

The Gibbs sampling algorithm is implemented by successively sampling from the full conditional distribution of each variable in the model [5]. In this appendix we show the Gibbs conditionals used to implement the heritability model.

The Gibbs conditional distribution of $\mathbf{b}$ is a multivariate normal distribution with the following expectation and variance

$$E\left(\mathbf{b}|\sigma^2,h^2,\mathbf{y}\right)=\left(\mathbf{X}'\mathbf{U}\left[0.25h^2\mathbf{D}+\mathbf{I}\right]^{-1}\mathbf{U}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{U}\left[0.25h^2\mathbf{D}+\mathbf{I}\right]^{-1}\mathbf{U}'\mathbf{y}$$

$$\text{Var}\left(\mathbf{b}|\sigma^2,h^2,\mathbf{y}\right)=\sigma^2\left(\mathbf{X}'\mathbf{U}\left[0.25h^2\mathbf{D}+\mathbf{I}\right]^{-1}\mathbf{U}'\mathbf{X}\right)^{-1}.$$

To circumvent the calculations involved in these formulas in cases with a large number of elements in $\mathbf{b}$, each element of $\mathbf{b}$ can be sampled in a univariate way. The Gibbs conditional distribution of σ^2 has an inverted chi square form

$$f\left(\sigma^2|\mathbf{b},h^2,\mathbf{y}\right)\propto\left(\sigma^2\right)^{-\frac{n}{2}}\exp\left\{-\frac{\mathbf{Q}}{2\sigma^2}\right\}$$

where $\mathbf{Q}=\mathbf{s}'\left[0.25h^2\mathbf{D}+\mathbf{I}\right]^{-1}\mathbf{s}$ and $\mathbf{s}=\mathbf{U}'\left(\mathbf{y}-\mathbf{X}\mathbf{b}\right)$. Although there are methods available to sample from inverted chi square distributions, we used the adaptive rejection algorithm [9]. The algorithm requires knowing, by proportionality, the logarithm of the conditional and its derivative

$$\log\left[f\left(\sigma^2|\mathbf{b},h^2,\mathbf{y}\right)\right]\propto-0.5n\log\left(\sigma^2\right)-0.5\sigma^{-2}\mathbf{Q}$$

$$\frac{\partial\log\left[f\left(\sigma^2|\mathbf{b},h^2,\mathbf{y}\right)\right]}{\partial\sigma^2}\propto-0.5\sigma^{-2}n+0.5\left(\sigma^2\right)^{-2}\mathbf{Q}.$$

The Gibbs conditional distribution of h^2, taking into account that $0.25h^2\mathbf{D}+\mathbf{I}$ is diagonal

$$f\left(h^2|\mathbf{b},\sigma^2,\mathbf{y}\right)\propto\prod_{i=1}^{n}\left(0.25h^2d_i+1\right)^{-\frac{1}{2}}\exp\left\{-0.5\sigma^{-2}\sum_{i=1}^{n}s_i^2\left(0.25h^2d_i+1\right)^{-1}\right\}$$

where n is the rank of $\mathbf{V}$, *i.e.*, the number of observed data, d_i is the ith element of the diagonal of $\mathbf{D}$ and s_i is the ith element of $\mathbf{s}$. The adaptive rejection algorithm can also be used to sample from the Gibbs conditional of h^2. The logarithm of the conditional and its derivative are

$$\begin{aligned}\log\left[f\left(h^2|\mathbf{b},\sigma^2,\mathbf{y}\right)\right]\propto&-0.5\sum_{i=1}^{n}\log\left(0.25h^2d_i+1\right)\\&-0.5\sigma^{-2}\sum_{i=1}^{n}s_i^2\left(0.25h^2d_i+1\right)^{-1}\\\frac{\partial\log\left[f\left(h^2|\mathbf{b},\sigma^2,\mathbf{y}\right)\right]}{\partial h^2}\propto&-0.5\sum_{i=1}^{n}0.25d_i\left(0.25h^2d_i+1\right)^{-1}\\&+0.5\sigma^{-2}\sum_{i=1}^{n}0.25d_is_i^2\left(0.25h^2d_i+1\right)^{-2}\end{aligned}$$

The marginal posterior density of h^2, especially the density at zero, can be obtained following the Rao-Blackwell argument, that is, by averaging the conditional densities of each cycle of the algorithm.