Agrociencia

ELICITATION OF PRIOR DISTRIBUTIONS TO ESTIMATE VARIANCE COMPONENTS

Ignacio **Luna-Espinoza**[1,2]*, Paulino **Pérez-Rodríguez**[1], José **Crossa**[1,3], Osval Antonio **Montesinos-López**[4], Juan Manuel **Romero-Padilla**[1]

[1]Colegio de Postgraduados Campus Montecillo. Carretera México-Texcoco km 36.5, Montecillo, Texcoco, State of Mexico, Mexico. C. P. 56264.
[2]Universidad del Istmo. Ciudad Universitaria S/N, Ciudad Ixtepec, Oaxaca, Mexico. C. P. 70110.
[3]Centro Internacional de Mejoramiento de Maíz y Trigo. Carretera México-Veracruz km 45, El Batán, Texcoco, State of Mexico, Mexico. C. P. 56237.
[4]Universidad de Colima. Facultad de Telemática. Avenida Universidad 333, Las Víboras, Colima, Mexico. C. P. 28040.
* Author for correspondence: ignaciolunaespinoza@gmail.com

ABSTRACT

In the Bayesian context, when variance components are modeled in normal hierarchical models, the inverted gamma distribution (IG) is typically used as the prior density for each component. However, the literature indicates that this prior density is highly informative, and thus the Half Cauchy distribution (HC) is recommended. The aim of this study was to evaluate, using simulation (in the context of high-dimensional data such as in the case of genomic selection applications), the suitability of the scaled inverse chi-squared $(\chi_{v,S}^{-2})$ distribution, which belongs to the family of scaled inverse gamma distributions, and HC as prior densities for the variance components in the Bayesian Ridge regression model. The evaluation was carried out when the number of observations in the response variable is greater than the number of predictor variables ($n > p$) as well as in high dimensions ($n << p$). The Bayesian learning of the posterior distribution was evaluated using the Hellinger distance (HD). The results of the Bayesian analysis were also compared with those obtained with the restricted maximum likelihood (REML). Results indicate that when $n > p$, the REML method underestimates the variance of the random effect, whereas in scenarios in which $n << p$, the method overestimates the same parameter when the variance of the error is large (greater than or equal to 6.0) and gives consistent estimations when the error variance is moderate (equal to 1.0). On the other hand, under prior distribution $(\chi_{v,S}^{-2})$ and in both scenarios ($n > p$) and $n << p$, it was observed that the parameters can be overestimated or underestimated, depending on the fixed values used to simulate the data. For the case of the HC prior distribution, the credibility intervals for both the variance of the effects of the predictor variables and the variance of the error contain the true values of the parameters, and their precisions increase with the sample size.

Keywords: Bayesian learning, scale inverse chi-squared distribution, Half Cauchy distribution, Hellinger distance.

Citation: Luna-Espinoza I, Pérez-Rodríguez P, Crossa J, Montesinos-López OA, Romero-Padilla JM. 2026. Elicitation of prior distributions to estimate variance components. **Agrociencia.** https://doi.org/ 10.47163/agrociencia.v60i1.3350

Editor in Chief:
Dr. Fernando C. Gómez Merino

Received: June 05, 2025.
Approved: January 12, 2026.
Published in Agrociencia:
January 30, 2026.

INTRODUCTION

For the past two decades, the development of data-acquisition technology has enabled diverse gadgets to take anywhere between thousands and millions of measurements of one single unit of interest. For example, DNA microarray technology measures the expression levels of thousands of genes. Using the information gathered, it is possible to understand the biological regulation mechanisms and develop new drugs. In these data, the number p of predictor variables amounts to thousands, while the number n of individuals involved in the experiment is much lower, that is, $n \ll p$ (Giraud, 2015). Another example of data in high dimensions is raised when taking images and videos, where large image databases are being continually gathered, such as medical or astrophysical images. Each image can be made up of millions of pixels or voxels. In the case of medical images, the number p of pixels is generally much greater than the number n of patients in the study.

From the statistics and computer science point of view, the challenge is determining the function $\Phi(\cdot)$, in such a way that $\Phi(x_1, ..., x_p)$ predicts the response variable y exactly and accurately. Working with massive data is extremely challenging because separating useful information from noise is almost impossible (Giraud, 2015). This problem is known as the curse of dimensionality. Although the proposals found in the literature show the existence of countless functions $\Phi(\cdot)$, linear models $y_i = \mu + \sum_{j=1}^{p} x_{ij}\beta_j + \epsilon_i$ are some of the best to carry out predictions in high dimensions. Under the Bayesian approach, different prior distributions have been assigned to regression coefficients β_j, giving rise to the "Bayesian Alphabet" (Gianola, 2013).

In this study, errors were assumed to be independent with a normal distribution, with a mean of zero and a variance of ϕ_ϵ, that is, $\epsilon_i \sim \text{iid}N(0, \phi_\epsilon)$; whereas β_j regression coefficients are assigned normal independent prior distributions with a mean of zero and a variance of ϕ_β, that is, $\beta_j \sim \text{iid}N(0, \phi_\beta)$. It is worth raising the question of which are the prior densities that must be assigned to the ϕ_β and ϕ_ϵ variance components, in such a way that they enable a full Bayesian learning.

Until a little more than two decades ago, when variance components were modeled in normal hierarchical models (in models structured from frameworks that allow the analysis of data organized into levels or nested groups) (Gelman and Hill, 2007), the IG prior distribution with the shape parameter equal to the scale parameter IG (ε, ε) was typically used. According to Spiegelhalter *et al.* (1996), the frequent use of this distribution was due to the fact that, in the BUGS software, it was the only prior distribution available for modeling variance components. Its wide use can be explained by an argument provided by Fink (1997), who points out that before the general availability of computers, joint prior distributions were used in order for the product of the prior density function and the likelihood to be analytically treatable.

In this context, the full conditional distribution of each variance component is also IG, which facilitates the computational implementation of the Gibbs sampler to generate samples from the posterior distribution. Otherwise, if the full conditional distribution of the variance component is complex (when its kernel does not correspond with that of a known univariate density function), the implementation of a Markov Chain Monte Carlo (MCMC) sampling technique must be implemented to obtain samples from the posterior distribution, which entails a high computational cost, such as the requirement for large amounts of computer time, particularly in high-dimensional settings. In contrast, assigning the IG prior to each variance component does not present the above issue.

Due to the widespread use of the IG prior distribution to model variance components, it is worth asking how appropriate the use of this prior distribution is for the modeling of variance components. A pioneering study evaluating this distribution as non-informative prior was conducted by Browne and Draper (2006). They point out that the use of this prior density must be taken with caution, since the IG distribution (ε, ε) has a peak near zero, and this can lead to problems when the random component variance is close to zero. Coincidentally, Gelman (2006) argues that the selection of small values for hyperparameter ε in the prior IG density (ε, ε) causes the posterior distribution of the variance components to produce small values; that is, a noticeable shrinkage occurs of the posterior distribution towards zero which causes an imperfect Bayesian learning.

On the other hand, Polson and Scott (2012) point out that it is not appropriate to assign the IG prior distribution to scale parameters in higher levels of normal hierarchical models; instead, they recommend the use of the HC density, particularly when proper prior distributions are required. The recommendation is supported by the fact that the HC prior distribution has a good performance close to zero and shows no drastic influence in other parts of the parametric space, which provides a solid justification for its routine use. The HC distribution is heavy-tailed, which implies the frequent occurrence of extreme values. This characteristic may facilitate the Bayesian learning of the posterior distribution of the corresponding variance component, as it leads to a broad exploration of the parametric space. Additionally, Polson and Scott (2012) point out that in contexts in which the regression coefficient vector is sparse, that is, where most coefficients are zero and only few have relevant effects, IG conjugated prior distribution can severely distort inferences.

In the framework of the Bayesian alphabet, Gianola *et al.* (2009) implicitly expose that prior distribution $(\chi^{-2}_{v,s})$ in the BayesA and BayesB models has a strong influence on the variance of the effects of the predictor variables. These results are ratified by Lehermeier *et al.* (2013), who point out the strong influence of hyperparameters in the predictive ability of the BayesA and BayesB models. The choice of a large value for the scale of the parameter of prior distribution $(\chi^{-2}_{v,s})$ assigned to the variance of the effects leads to an overfitting of the data. Meanwhile, a very small value leads to underfitting due to a noticeable shrinking of the effects. In both cases, the predictive ability of the models is reduced.

In the context of Bayesian learning, as the sample size increases, the influence of the prior distribution should disappear gradually, hence the prior distribution would have little importance in large samples (Bernardo and Smith, 1994). This result is valid for parameters with maximum likelihood estimators that are identifiable (Gianola, 2013). However, in schemes in which $n << p$, Bayesian learning may arise only for n parameters or functions of n parameters, since $p - n$ parameters are not identifiable; hence, when $n << p$ the prior distributions will always be important, and their influence will never disappear, even with large samples (Gianola, 2013).

Motivated by the above arguments and by the properties of the HC distribution, the aim of this study is to evaluate, through simulation, the suitability of assigning the $(\chi^{-2}_{v,s})$ and HC prior distributions to the variance components in Bayesian Ridge regression. The results of both distributions were compared with the estimations using the classic frequentist method REML. The Bayesian learning capacity of the respective models was evaluated using the Hellinger distance (HD) between the marginal prior distribution assigned to the variance component and the corresponding posterior density. An HD nearer to one indicates that the posterior distribution moves away from the prior distribution, thus providing evidence of a robust Bayesian learning process. In contrast, an HD closer to zero will indicate a lack of Bayesian learning due to the influence of the prior distribution and its hyperparameters.

MATERIALS AND METHODS

The following section presents the sampling distribution and the structure of the three models evaluated: a) linear mixed model, b) a Bayesian model with a $(\chi^{-2}_{v,s})$ prior assigned to each variance component, and c) a Bayesian model with an HC prior assigned to the positive square root of each variance component.

Sampling model and prior distributions

Consider the standard linear model:

$$y = 1\mu + X\beta + \epsilon, \tag{1}$$

where y is a vector of dimension $n \times 1$, 1 is a vector of ones of dimension $n \times 1$, μ is an intercept, $X = \{x_{ij}\}$ is the incidence matrix of dimension $n \times p$, β is a vector of dimension $p \times 1$, and ϵ is a vector of errors of dimension $n \times 1$. The errors are assumed to be independent and normally distributed with mean zero and variance ϕ_ϵ, that is, $\epsilon \sim N(0, \phi_\epsilon I)$, where I is the identity matrix of order $n \times n$. The above elements define the following conditional distribution of the data:

$$\mathrm{p}(\boldsymbol{y}|\mu, \boldsymbol{\beta}, \phi_\epsilon) = \prod_{i=1}^{n} N\left(y_i|\mu + \sum_{j=1}^{p} x_{ij}\,\beta_j, \phi_\epsilon\right), \tag{2}$$

where the expression $N\left(y_i|\mu + \sum_{j=1}^{p} x_{ij}\,\beta_j, \phi_\epsilon\right)$ denotes the density function of a normal random variable for y_i, given the mean a and variance b, that is, $N(y_i|a, \mathrm{b})$.

Linear mixed model

In equation (1), if $\mathrm{u} = X\beta$, where $\beta \sim N(0, \phi_\beta I)$, by the properties of the multivariate normal distribution, the following model is obtained:

$$y = 1\mu + \mathrm{u} + \epsilon, \tag{3}$$

where $\mathrm{u} \sim \mathrm{N}(0, \phi_\beta \mathrm{XX}')$ and $\epsilon \sim N(0, \phi_\epsilon I)$. In the context of genomic selection, X is the matrix of markers coded into numerical values (for example, 0, 1, 2 to denote the number of times the major allele appears), XX' is proportional to the variance and covariance matrix of the random term u or what is known as the linear kernel in the RKHS non-parametric regression models (de los Campos *et al*., 2010). Equation (3) defines a standard linear mixed model where 1μ is a fixed term and u is a random effects vector.

Prior density for the intercept

The intercept was assigned a normal prior distribution with a mean zero and a variance $k = 1 \times 10^{10}$. This assignment is similar to assigning a flat prior distribution to the intercept.

Prior density for the regression coefficients

In the Ridge Bayesian regression model, the following prior distribution is assigned to the regression coefficients:

$$\mathrm{p}(\beta|\phi_\beta) = \prod_{j=1}^{p} \mathrm{N}(\beta_j|0, \phi_\beta) \tag{4}$$

Notice that the shrinking of the regression coefficients towards zero depends on variance component ϕ_β; the smaller variance ϕ_β, the greater the concentration of the prior density will be around zero.

A) Scaled inverse chi-squared prior to the variance components. Variance parameter ϕ_β is assigned a prior density $\chi^{-2}_{v_\beta, S_\beta}$, that is, $p(\phi_\beta|v_\beta, S_\beta) = \chi^{-2}(\phi_\beta|v_\beta, S_\beta)$, where $v_\beta > 0$ and $S_\beta > 0$ denote the degree of freedom and the scale parameter, respectively. This study uses the parametrization of prior distribution $\chi^{-2}(\phi_\beta|v_\beta, S_\beta)$ so that $E(\phi_\beta) = S_\beta/(v_\beta - 2)$ and $mode(\phi_\beta) = S_\beta/(S_\beta + 2)$. Similarly, ϕ_ϵ is assigned the prior distribution $\mathrm{p}(\phi_\epsilon|v_\epsilon, S_\epsilon) = \chi^{-2}(\phi_\epsilon|v_\epsilon, S_\epsilon)$. The joint prior distribution is expressed as follows (using the prior for the intercept, equation (4) and the prior for the variance components):

$$\mathrm{p}(\mu, \beta, \phi_\beta, \phi_\epsilon|H_1) = \mathrm{p}(\mu)\mathrm{p}(\beta|\phi_\beta)\mathrm{p}(\phi_\beta|v_\beta, S_\beta)\mathrm{p}(\phi_\epsilon|v_\epsilon, S_\epsilon), \tag{5}$$

where $H_1 = \{k, v_\beta, S_\beta, v_\epsilon, S_\epsilon\}$ is a set of values of the hyperparameters for the prior distributions. By applying Bayes' theorem, the posterior distribution is proportional to the product of the prior distribution (equation 5) and the likelihood (equation 2):

$$\begin{aligned} \mathrm{p}(\mu, \beta, \phi_\beta, \phi_\epsilon | y) \;\; \propto \;\; & \mathrm{N}(\mu|0,k)\left\{\prod_{j=1}^{p} \mathrm{N}(\beta_j|0,\phi_\beta)\right\} \chi^{-2}(\phi_\beta|v_\beta, S_\beta)\, \chi^{-2}(\phi_\epsilon|v_\epsilon, S_\epsilon) \\ \times & \left\{\prod_{i=1}^{n} N\left(y_i|\mu + \sum_{j=1}^{p} x_{ij}\,\beta_j, \phi_\epsilon\right)\right\}. \end{aligned} \quad (6)$$

The kernel of the function of the posterior distribution in equation (6) does not correspond to that of any known distribution, but samples can be obtained using simulation procedures based on MCMC algorithms. The full conditional distributions are known, and it is therefore possible to obtain samples from the posterior distribution using the Gibbs sampler (Casella and George, 1992).

B) Half Cauchy distribution assigned to the standard deviation of the variance component. Following Gelman (2006), standard deviation $\psi_\beta = \sqrt{\phi_\beta}$ is assigned the standard HC prior density:

$$\mathrm{p}(\psi_\beta) = 2\pi^{-1}\left(1 + \psi_\beta^2\right)^{-1}. \quad (7)$$

Notice that $\phi_\beta = \psi_\beta^2$ is the variance component. Using the transformation method (Casella and Berger, 2002), to transform from the standard deviation to the variance, the prior density for variance parameter ϕ_β is equal to:

$$\mathrm{p}(\phi_\beta) = \pi^{-1}\phi_\beta^{-1/2}\left(1 + \phi_\beta\right)^{-1}, \quad (8)$$

for $\phi_\beta > 0$. Notice that ψ_β and ϕ_β have the same domain. Similarly, the HC prior density is obtained for variance component ϕ_ε, which is given by:

$$\mathrm{p}(\phi_\varepsilon) = \pi^{-1}\phi_\varepsilon^{-1/2}(1 + \phi_\varepsilon)^{-1}, \quad (9)$$

for $\phi_\varepsilon > 0$. In addition, assuming independence, the joint prior distribution can be represented as follows (using equations 7, 8 and 9):

$$\mathrm{p}(\mu, \beta, \phi_\beta, \phi_\epsilon | H_2) = \mathrm{p}(\mu)\mathrm{p}(\beta|\phi_\beta)\mathrm{p}(\phi_\beta)\mathrm{p}(\phi_\epsilon). \quad (10)$$

Here, the set of values of the hyperparameters for the prior distributions has only one element, $H_2 = \{k\}$. By applying Bayes' theorem, the joint posterior distribution is proportional to the product of the prior distribution (equation 10) and the likelihood (equation 2):

$$\begin{aligned} \mathrm{p}(\mu,\beta,\phi_\beta,\phi_\epsilon|y) \quad &\propto \quad \mathrm{N}(\mu|0,k)\left\{\prod_{i=1}^{p}\mathrm{N}(\beta_i|0,\phi_\beta)\right\}\phi_\beta^{-\frac{1}{2}}\phi_\varepsilon^{-\frac{1}{2}}\left(1+\phi_\beta\right)^{-1}(1+\phi_\varepsilon)^{-1} \\ &\times \quad \left\{\prod_{i=1}^{n} N\left(y_i|\mu+\sum_{j=1}^{p}x_{ij}\,\beta_j,\phi_\epsilon\right)\right\}. \end{aligned} \tag{11}$$

The kernel of the function given in equation (11) does not correspond with any known distribution, but samples can be obtained using simulation procedures based on MCMC algorithms. The full conditional distributions for the intercept and the regression coefficients can be verified to be normal. In this context, to perform an inference from the posterior distribution, a hybrid mechanism was implemented, using the Gibbs sampler for sampling the intercept and regression coefficients, and the Metropolis algorithm with a random walk for the variance components.

Selection of hyperparameters for prior distributions $\chi^{-2}(\phi_\beta|v_\beta,S_\beta)$ and $\chi^{-2}(\phi_\epsilon|v_\epsilon,S_\epsilon)$

To select the hyperparameters of prior densities $\mathrm{p}(\phi_\beta|v_\beta,S_\beta)$ and $\mathrm{p}(\phi_\epsilon|v_\epsilon,S_\epsilon)$, the guidelines recommended by Pérez-Rodríguez and de los Campos (2014) and Pérez-Rodríguez *et al*. (2018) were followed. The recommendations establish that the variance of response variable y_i can be divided into two components: the one of the linear predictor and the one of the error:

$$\begin{aligned} Var(y_i) \quad &= \quad Var\left(\sum_{j=1}^{p}x_{ij}\beta_j+\epsilon_i\right)=Var(\mathrm{g}_i)+Var(\epsilon_i) \\ &= \quad \sum_{j=1}^{p}x_{ij}^2 Var(\beta_j)+Var(\epsilon_i)=\phi_\beta\sum_{j=1}^{p}x_{ij}^2+\phi_\epsilon, \end{aligned} \tag{12}$$

for $i = 1, ..., n$, where $\mathrm{g}_i=\sum_{j=1}^{p}x_{ij}\beta_j$. On the other hand, the total variance due to the linear predictor is equal to $\sum_{i=1}^{n}Var(g_i)=\phi_\beta\sum_{i=1}^{n}\sum_{j=1}^{p}x_{ij}^2$. Thus, Pérez-Rodríguez *et al.* (2018) showed that the average prior variance due to the linear predictor in equation (12) is equal to:

$$V_g = \frac{1}{n}\sum_{i=1}^{n} Var(g_i) = \frac{1}{n}\phi_\beta \sum_{i=1}^{n}\sum_{j=1}^{p} x_{ij}^2 = \phi_\beta MS_x$$

where $MS_x = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p} x_{ij}^2$ and the variance of response variable V_y is equal to

$$V_y = V_g + V_\epsilon = \phi_\beta MS_x + \phi_\epsilon, \tag{13}$$

where $V_\epsilon = var(\epsilon_i) = \phi_\epsilon$. From equation (13), it follows that

$$\phi_\beta = \frac{V_g}{MS_x}. \tag{14}$$

By replacing ϕ_β by the mode of the density $\chi^{-2}(\phi_\beta|v_\beta, S_\beta)$ in equation (14) and solving for scale parameter S_β, $S_\beta = V_g(v_\beta + 2)/MS_x$. Moreover, using the concept of heritability in the context of quantitative genetics, $h^2 = V_g/V_y$, it follows that $V_g = h^2 V_y$ (Falconer and Mackay, 1996). Therefore,

$$S_\beta = h^2 V_y(v_\beta + 2)/MS_x = \frac{R^2 V_y\left(v_\beta + 2\right)}{MS_x}$$

where R^2 is the proportion of the variance that is explained a priori by the predictor variables. By default, similar to Pérez-Rodríguez *et al.* (2018), it is considered that $R^2 = 0.5$. Therefore, for fixed v_β and given the y data vector and matrix X, the terms V_y, MS_x, S_β are calculated.

A similar procedure (v_β, S_β) was used to obtain the hyperparameters (v_ϵ, S_ϵ) (Pérez-Rodríguez *et al.*, 2018). The result of the procedure is the hyperparameters $(v_\epsilon, S_\epsilon) = (v_\epsilon, \left(1 - R^2\right)V_y(v_\epsilon + 2))$. On the other hand, to determine the respective degrees of freedom v_β and $v_\in$, the recommendations by Pérez-Rodríguez and de los Campos (2014) were followed, of taking small degrees of freedom to reduce the influence of the prior density on the posterior density. This study considered $v_\beta = v_\epsilon = 5$, since they are the minimum degrees of freedom that ensure finite prior mean and variance.

Discrepancy in the prior and posterior distributions

In Bayesian learning, it is important to evaluate how much useful information is obtained as progress is made in the learning process; in other words, what the discrepancy is between the prior and posterior densities as new evidence, generated with uncertainty under the sampling model, is incorporated. To evaluate the sensitivity of the models, the HD was used. This distance, unlike other measurements used to evaluate the influence of the *f* prior distribution on the posterior *g*, such as the

Kullback-Leibler divergence measurement, has the characteristic of being completely defined in situations in which one of the densities reaches the value of zero and the other one does not. This property, for example, does not appear in the measurement based on the Kullback-Leibler divergence, because when one of the densities reaches the value of zero and the other one does not, the Kullback-Leibler measurement becomes infinite (Roos and Held, 2011). The HD between densities f and g, which is denoted by $H(f, g)$, is defined by:

$$H(f,g) = \sqrt{\frac{1}{2}\int_{-\infty}^{\infty}\left(\sqrt{f(u)} - \sqrt{g(u)}\right)^2 du} = \sqrt{1 - \int_{-\infty}^{\infty}\sqrt{f(u)g(u)}du}$$

It is possible to verify that $0 \leq H(f,g) \leq 1$; in addition, $H(f,g) = 0$ if, and only if, $f = g$ (Roos and Held, 2011). On the other hand, $H(f,g) = 1$ if density f assigns a probability of 0 to each set that g assigns a positive probability and vice-versa.

When the posterior distribution is highly influenced by the election of the hyperparameters that are assigned to the prior distributions, the posterior distribution does not deviate from the prior distribution, which results in a lack of Bayesian learning (Lehermeier *et al*., 2013). In this case, HD will tend towards zero. By contrast, if the distance $H(f, g)$ between prior density f and posterior density g tend towards one, the knowledge of the state of nature is continually modified as new data is acquired. The HD is small if the perfect prior density is used. In this case, a small HD is not indicative of the lack of Bayesian learning. However, Lehermeier *et al*. (2013) claim that selecting the perfect prior density has a probability close to zero, particularly if there is a lack of prior information.

To evaluate the effect of the prior density of each variance component, the HD between the prior marginal density and the posterior density was determined numerically. The posterior marginal density was estimated from the MCMC samples with an estimation of non-parametric densities with a Gaussian kernel and a bandwidth parameter given by Silverman's criterion (Silverman, 1998).

Simulated data and fitting of the models

Both the response variable and the predictor variables were simulated, creating complex scenarios similar to what is done in genomic selection (Meuwissen *et al.*, 2001; Pérez-Rodríguez and de los Campos, 2014) in the field of quantitative genetics, where the prediction of a continuous variable (phenotype) based on thousands of covariables (molecular markers) is of interest. The data were simulated using the XSim software implemented in the Julia package (Bezanson *et al*., 2017), based on the method proposed by Cheng *et al*. (2018), which records the positions and the origins of the founders of each chromosome segment.

First, the chromosomes of the founders are labelled using identifiers, and each chromosome is represented by two vectors: one that indicates the origin of the founders and the other, the crossing positions. The states of the alleles of the founding genomes are generated, whether from map positions defined by the user and the allele frequency or from real haplotypes or data sequences. In meiosis, the new gamete is formed with chromosomic segments of the paternal and maternal gametes of the parents, with segments introduced in both sides of the crossing.

The simulation considered a chromosome length of 0.1 cM, 10 QTLs for each chromosome, and a population of 500 unrelated founders. The 10 QTLs distributed at random in each chromosome made up the significant regression coefficients in the Bayesian Ridge regression model. In independent simulations, chromosome numbers 4, 5, 7, 9, 12, 24, and 48 were taken. These chromosome numbers were chosen in order to observe the behavior of the fittings as the number of predictors gradually increases. Each predictor variable takes the values of 0, 1, or 2, according to the frequency of the major allele. The predictor variables with no variation were eliminated from their respective set. This framework produced sets of predictor variables with 349, 456, 616, 785, 1037, 2059, and 4090 variables, respectively. Additionally, the sample sizes considered were 200, 400, 600, 800, 1000, 2000, 4000, 6000, and 8000 (Figure 1), which enables the evaluation of the models in scenarios $n > p$, and $n < p$, including $n << p$.

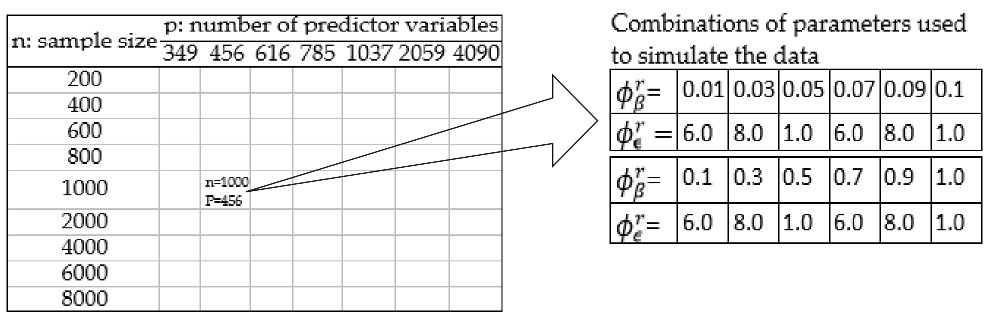

Figure 1. Simulation grid. Sample size (*n*) × number of predictor variables (*p*), and combinations of parameters used to simulate the data in each point of the grid $n \times p$.

To evaluate the influence of the prior densities against small and large variations in the variance components, in each point of the $n \times p$ grid, the response variable was simulated with the combination of the pairs of fixed values (ϕ_β^r, ϕ_ϵ^r) = (0.01, 6.0), (0.03, 8.0), (0.05, 1.0), (0.07, 6.0), (0.09, 8.0), (0.1, 1.0), (0.1, 6.0), (0.3, 8.0), (0.5, 1.0), (0.7, 6.0), (0.9, 8.0), (1.0, 1.0). Values of $0.01 \leq \phi_\beta^r < 0.1$ helped observe the performance of the fittings against ϕ_β^r values close to zero, whereas with values of $0.1 \leq \phi_\beta^r \leq 1,$ showed the behavior of the fittings against values of ϕ_β^r far away from zero. On the other hand, the value of $\phi_\epsilon^r = 1.0$ helps observe the performance of the fittings against little noise in the error, whereas the values of $\phi_\epsilon = 6.0, 8.0$ display the performance of the adjustments in the face of much noise in the error. The combination of parameters (ϕ_β^r, ϕ_ϵ^r) considered to simulate the data in each point of the $n \times p$ grid is illustrated

(Figure 1). Only additive genetic effects were simulated, and an absence of dominance was assumed.

The models were fitted with the simulated response and predictor variables. The Bayesian models were fitted using routines written in the C and R programming languages that implement the algorithms of the Gibbs sampler and random walk. This allowed the optimization of the process to obtain MCMC samples, making it faster, combining compiled and interpreted code. On the other hand, the estimation with the REML method of the variance of the linear mixed model defined in equation (3) was carried out with the *mixed.solve* function of R's rrBLUP package (Endelman, 2011).

The Gelman-Rubin diagnosis (Gelman and Rubin, 1992) verified the convergence of the MCMC chains towards their stationary distribution. For this purpose, two 100 000-sized chains were used, considering half of each chain as burning and implementing a thinning every 10 observations. If the statistical value of the Gelman-Rubin (DG) test is lower than 1.2, it is concluded that the convergence is acceptable. The summary of statistics of the posterior estimated densities was calculated with the average of the corresponding statistics given for each one of the two MCMC chains. To facilitate the discussion of results, fixed values were used to simulate the data, which were denoted by ϕ_β^r y ϕ_ϵ^r, called true parameters.

RESULTS AND DISCUSSION

Scenario $n > p$

The performance of the fittings in scenarios $n > p$ is exposed in detail using the combination of $n = 8000$ and $p = 349$ (Table 1) as an example. For other scenarios, similar results were obtained (not shown). The MCMC chains of each variance component quickly converged, except with the pair of values $\phi_\beta^r = 0.01$ and $\phi_\epsilon^r = 6.0$ and $\phi_\beta^r = 0.03$ and $\phi_\epsilon^r = 8.0$, for which chain convergence was slow. For example, when the data were generated with $\phi_\beta^r = 0.03$ and $\phi_\epsilon^r = 8.0$, $n = 8000$ and $p = 349$, 100 000 realizations of the chain under prior HC were needed to reach their stationary distribution; in contrast, under the same prior distribution and with data generated with real parameters $\phi_\beta^r =$ 0.01 and $\phi_\epsilon^r = 1.0$, $n = 8000$ and $p = 349$, only 10 000 iterations of the chain were needed to reach its stationary distribution. With 100 000 observations, the statistical value of the DG test was always lower than 1.2, which indicates convergence in each one of the chains generated (Table 1). Due to this, the generation of 100 000 observations of each chain was considered so that all were of the same size and their convergences were ensured at the same time.

Under prior density $(\chi_{v,S}^{-2})$ the credibility intervals (CI) of o.95 for 0.95 for ϕ_ϵ generally contained the true values of parameters ϕ_ϵ (Table 1, Figures 2B and 2D). These results contrast with the CI obtained for ϕ_β. In the intervals built for ϕ_β four behaviors are observed, linked to the values of ϕ_β^r and ϕ_ϵ^r. First, when value ϕ_β^r is near zero ($\phi_\beta^r \leq 0.3$) and ϕ_ϵ^r is large ($\phi_\epsilon^r \geq 6.0$), the posterior distribution of ϕ_β is biased towards

Table 1. Summary of statistics after fitting the models with scaled inverse chi-squared ($\chi^{-2}_{v,S}$) and Half Cauchy (HC) priors, with n = 8000 and p = 349 (four chromosomes). Fixed values (ϕ^r_β, ϕ^r_ϵ) to simulate the data (Pars).

Pars		REML		Prior	Posterior ϕ_β							Posterior ϕ_ϵ						
ϕ^r_β	ϕ^r_ϵ	$\hat{\phi}_\beta$	$\hat{\phi}_\epsilon$		M	Me	SD	Q0.025	Q0.975	HD	DG	M	Me	SD	Q0.025	Q0.975	HD	DG
0.01	6	0.00005	5.9361	($\chi^{-2}_{v,S}$)	0.2944	0.2908	0.0387	0.2250	0.3759	1.0000	1.02	5.9028	5.9034	0.0921	5.7254	6.0724	0.8921	1.00
				HC	0.0110	0.0090	0.0089	0.0008	0.0332	0.8207	1.03	5.9351	5.9334	0.0959	5.7532	6.1350	0.9524	1.00
0.1	6	0.00063	5.9347	($\chi^{-2}_{v,S}$)	0.3368	0.3330	0.0444	0.2610	0.4295	1.0000	1.00	5.9115	5.9077	0.0946	5.7305	6.0954	0.8887	1.00
				HC	0.0776	0.0750	0.0229	0.0407	0.1301	0.8077	1.00	5.9428	5.9472	0.0921	5.7648	6.1275	0.9521	1.00
0.03	8	0.00020	7.9649	($\chi^{-2}_{v,S}$)	0.4025	0.3985	0.0526	0.3141	0.5147	1.0000	1.02	7.9212	7.9231	0.1277	7.6787	8.1660	0.8843	1.01
				HC	0.0162	0.0104	0.0169	0.0002	0.0607	0.7958	1.01	7.9717	7.9706	0.1239	7.7264	8.2203	0.9546	1.00
0.3	8	0.00251	7.9662	($\chi^{-2}_{v,S}$)	0.6107	0.6078	0.0794	0.4635	0.7758	0.9996	1.00	7.9397	7.9389	0.1279	7.6974	8.1974	0.8880	1.00
				HC	0.2972	0.2911	0.0583	0.1972	0.4227	0.8002	1.01	7.9830	7.9838	0.1279	7.7310	8.2297	0.9538	1.00
0.05	1	0.00045	0.9965	($\chi^{-2}_{v,S}$)	0.0909	0.0904	0.0123	0.0685	0.1164	0.9987	1.02	0.9952	0.9950	0.0165	0.9627	1.0271	0.9755	1.00
				HC	0.0558	0.0553	0.0096	0.0386	0.0767	0.8655	1.01	0.9977	0.9978	0.0163	0.9676	1.0301	0.9406	1.00
0.5	1	0.00400	0.9968	($\chi^{-2}_{v,S}$)	0.5700	0.5681	0.0571	0.4737	0.6909	0.7742	1.00	0.9993	0.9993	0.0165	0.9668	1.0323	0.8797	1.01
				HC	0.5524	0.5502	0.0579	0.4553	0.6712	0.8481	1.00	0.9980	0.9973	0.0154	0.9693	1.0299	0.9430	1.01
1.0	1	0.0078	0.9975	($\chi^{-2}_{v,S}$)	1.1223	1.1180	0.1064	0.9284	1.3343	0.6950	1.01	0.9995	0.9990	0.0162	0.9667	1.0300	0.8890	1.01
				HC	1.1173	1.1108	1.1062	0.9307	1.3362	0.8508	1.01	0.9984	0.9985	0.0157	0.9682	1.0293	0.9421	1.00
0.07	6	0.00043	5.9346	($\chi^{-2}_{v,S}$)	0.3189	0.3176	0.0410	0.2506	0.4079	1.0000	1.01	5.9075	5.9087	0.0946	5.7330	6.0931	0.8901	1.00
				HC	0.0515	0.0494	0.0184	0.0200	0.0932	0.8032	1.01	5.9381	5.9357	0.0957	5.7543	6.1309	0.9507	1.01
0.7	6	0.00449	5.9362	($\chi^{-2}_{v,S}$)	0.7378	0.7302	0.0939	0.5744	0.9400	0.9935	1.00	5.9281	5.9282	0.0914	5.7574	6.0969	0.8883	1.00
				HC	0.5612	0.5553	0.0800	0.4192	0.7271	0.8170	1.00	5.9401	5.9428	0.0959	5.7568	6.1229	0.9515	1.00
0.09	8	0.00066	7.9666	($\chi^{-2}_{v,S}$)	0.4449	0.4420	0.0590	0.3409	0.5673	1.0000	1.00	7.9348	7.9338	0.1240	7.7079	8.1819	0.8824	1.01
				HC	0.0756	0.0742	0.0286	0.0246	0.1345	0.7711	1.01	7.9740	7.9782	0.1229	7.7358	8.2187	0.9545	1.00
0.9	8	0.00808	7.9612	($\chi^{-2}_{v,S}$)	1.2335	1.2274	0.1516	0.9789	1.5465	0.9815	1.00	7.9637	7.9594	0.1269	7.7252	8.2152	0.8868	1.00
				HC	1.0044	0.9865	0.1384	0.7639	1.3088	0.8177	1.00	7.9676	7.9676	0.1269	7.7242	8.2350	0.9538	1.00

REML: restricted maximum likelihood estimations; M: mean; Me: median; SD: standard deviation; Q0.025: quantile 0.025; Q0.975: quantile 0.975; HD: Hellinger distance; DG: Gelman-Rubin diagnosis.

the right of the true value ϕ^r_β , which exposes its overestimation (Figures 2A and 2C). This shows the difficulty of estimating the variance component ϕ_β when the variability of the error is large and the effects of the predictor variables are small. In scenarios where ($\phi^r_\beta \leq 0.3$) and ($\phi^r_\epsilon \geq 6.0$), the HD distances under the prior ($\chi^{-2}_{v,S}$) are near 1.0, indicating a discrepancy between the prior and posterior densities. However, Bayesian learning is poor because the posterior densities fail to capture the true values of the parameters ϕ^r_β (Figures 2A and 2C).

Second, when the variance component ϕ_β tends towards the value of 1.0 ($\phi^r_\beta \geq 0.7$), the value ϕ^r_ϵ = 6.0 (ϕ^r_ϵ = 8.0) leads parameter ϕ^r_β to be overestimated if the size of the sample is insufficient (n = 8000). Under this scenario, at least n = 8000 registers are needed for the posterior distribution of ϕ_β to center its mass around the true value ϕ^r_β. This performance was also verified when calculating the HD distance, whose value close to 1.0 indicates a complete distancing between the prior and posterior densities.

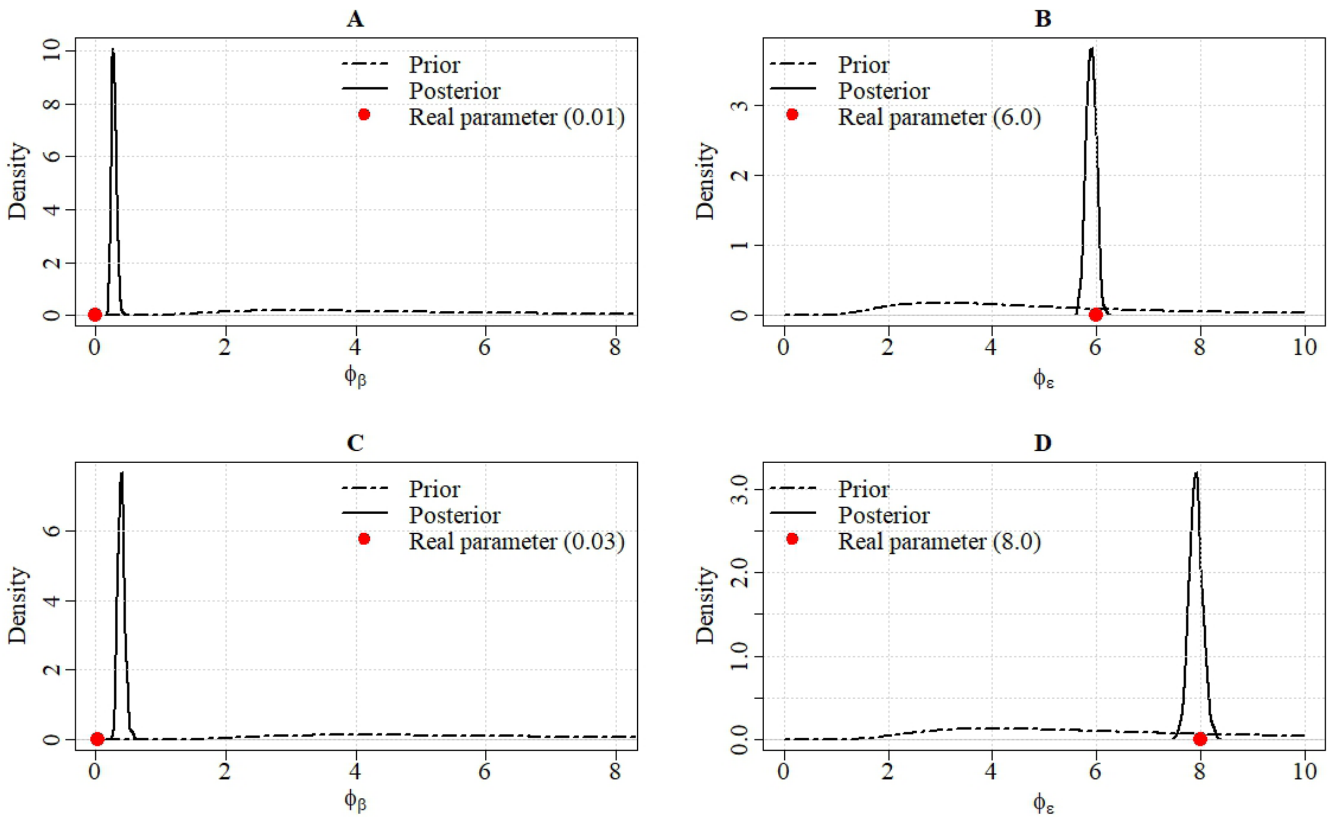

Figure 2. Posterior densities for ϕ_β and ϕ_ϵ with scaled inverse chi-squared $(\chi^{-2}_{v,S})$ prior, $n > p$ (n = 8000, p = 349) under two scenarios. A, B: (ϕ^r_β = 0.01, ϕ^r_ϵ = 6.0); C, D: (ϕ^r_β = 0.03, ϕ^r_ϵ = 8.0).

Third, when ϕ^r_ϵ = 1.0, the posterior distribution of ϕ_β was found around the true value of parameter ϕ^r_β (ϕ^r_β = 0.5, 1.0). However, this behavior does not express a substantial Bayesian learning because the prior distribution is found around the true values of the parameters (ϕ^r_β = 0.5, 1.0), which is why the HD distance only reached values below 0.8 (0.7742 and 0.6950). In this context, a relevant prior density is being used, in which case an HD distance relatively far from the value of 1.0 does not imply a lack of Bayesian learning. Finally, it was observed that although ϕ^r_β moved away from the origin (ϕ^r_β = 0.9), the high noise induced by the variance of the error (ϕ^r_ϵ = 8.0) generally causes the overestimation of parameter ϕ^r_β.

When standard HC prior distribution is used, almost all CI at a probability of 0.95 for ϕ_ϵ captured the different values assigned to parameter ϕ^r_ϵ (Table 1, Figures 3B and 3D). These results are similar to those obtained under prior $(\chi^{-2}_{v,S})$. However, for component ϕ_β, the performance of the posterior distribution contrasts with the performance of the posterior density when considering $(\chi^{-2}_{v,S})$ prior. In general, each one of the CI at 0.95 probability for ϕ_β contains the true value of the parameter (Figures 3A and

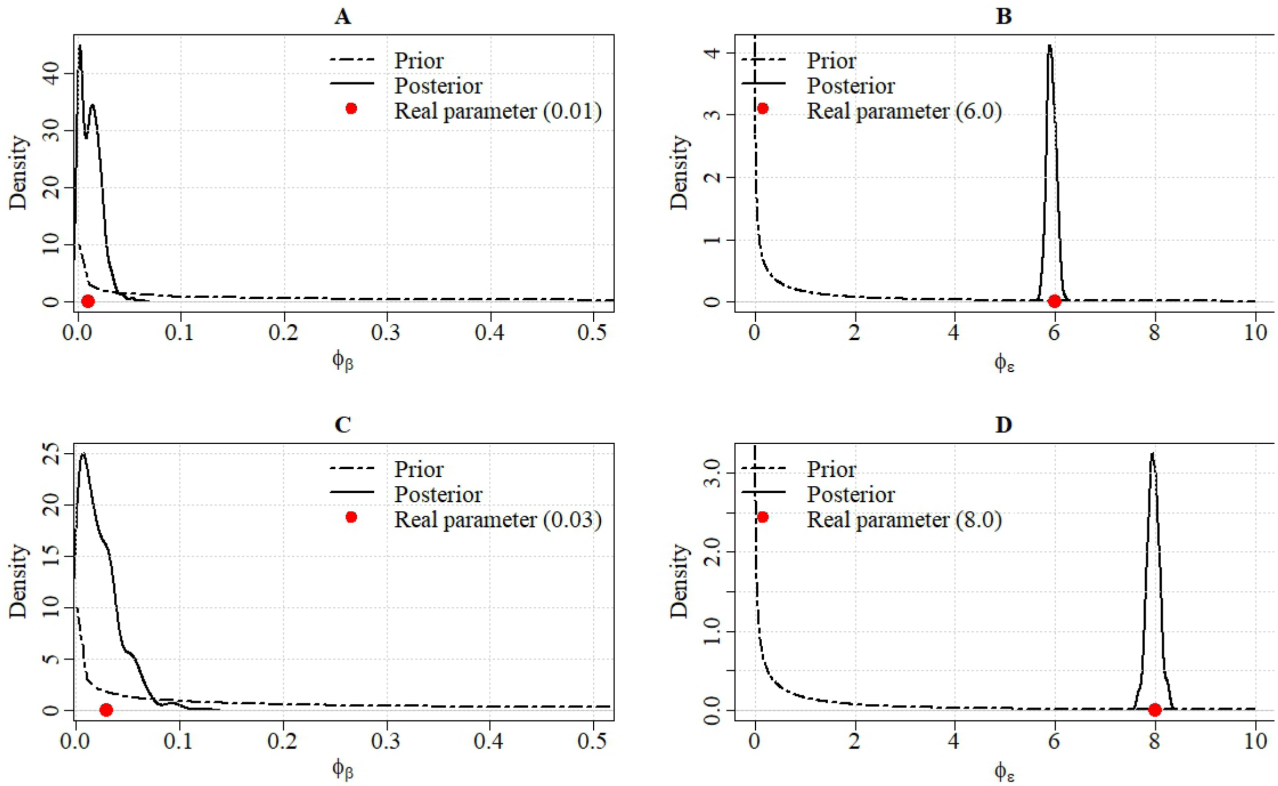

Figure 3. Posterior densities for ϕ_β and ϕ_ϵ with Half Cauchy prior (HC), $n > p$ (n = 8000, p = 349) under two scenarios. A, B: (ϕ_β^r = 0.01, ϕ_ϵ^r = 6.0); C, D: (ϕ_β^r = 0.03, ϕ_ϵ^r = 8.0).

3C), alongside the fact that in each simulation scenario, the mean of the posterior distribution is relatively near the real value of the parameter.

Regarding the effectiveness of Bayesian learning, the prior distribution for ϕ_β accumulated its greatest mass near the origin. One could suspect that this property allows the posterior density to effectively capture the true value of parameter ϕ_β^r when it is near zero. However, when the real value of parameter ϕ_β^r in the data-generating model moves away from the origin (ϕ_β^r = 1.0), the greatest discrepancy is created between prior and posterior densities (Table 1). The posterior distribution of ϕ_β concentrates its mass around the true value of the parameter (ϕ_β^r = 1.0). This behavior is also reflected in the posterior distribution of ϕ_ϵ, because when the standard HC prior distribution is assigned to the positive square root of variance component ϕ_ϵ, the prior density for ϕ_ϵ also accumulates its largest mass near zero.

This characteristic of the prior distribution for ϕ_ϵ has no influence in the learning of the posterior distribution, which concentrates around the true value of parameter ϕ_ϵ^r, regardless of it moving away from the origin (Figures 3B and 3D). Thus, the posterior density is effectively updated towards the true value of the parameter, whether ϕ_ϵ^r

= 1.0, 6.0, or 8.0, as applicable (Table 1). This behavior is also expressed in the HD distances between the prior and posterior densities, which surpass the value of 0.943, indicating an efficient and effective Bayesian learning.

The results of fitting the linear mixed model were reported in equation (3) using the REML method (Table 1). The adjustment shows the inability of the method to correctly estimate the variance component ϕ_{β}^{r}, which was underestimated in each simulation scenario. An important positive aspect of the method is its precise estimation of variance component ϕ_{ϵ}^{r}, which should be emphasized.

Scenario *n* << *p*

The performance of the fittings in scenarios *n* << *p* is exposed in detail with *n* = 200 and *p* = 349 (Table 2). Similar results were obtained for other scenarios (not shown). The value of the DG statistic obtained in each chain diagnosis was always near the value of

Table 2. Summary of posterior statistics when fitting the models with the scaled inverse chi-squared ($\chi_{v,S}^{-2}$) and Half Cauchy (HC) priors, with *n* = 200 and *p* = 349 (four chromosomes). Fixed values (ϕ_{β}^{r}, ϕ_{ϵ}^{r}) to simulate the data (Pars).

Pars		REML		Prior	Posterior ϕ_{β}							Posterior ϕ_{ϵ}						
ϕ_{β}^{r}	ϕ_{ϵ}^{r}	$\hat{\phi}_{\beta}$	$\hat{\phi}_{\epsilon}$		M	Me	SD	Q0.025	Q0.975	HD	DG	M	Me	SD	Q0.025	Q0.975	HD	DG
0.01	6	0.2275	6.3076	$(\chi_{v,S}^{-2})$	1.9487	1.8772	0.5930	1.0405	3.3452	0.7308	1.00	5.4021	5.3246	0.6638	4.2412	6.8612	0.6395	1.00
				HC	0.4022	0.2618	0.4129	0.0021	1.4026	0.5410	1.00	6.2458	6.1744	0.7396	5.0096	7.8028	0.8620	1.00
0.1	6	<0.0001	6.2279	$(\chi_{v,S}^{-2})$	1.1901	1.1414	0.3177	0.7245	1.9346	0.8306	1.00	4.5427	4.5210	0.5110	3.6001	5.6098	0.6623	1.00
				HC	0.1403	0.0716	0.1886	0.0003	0.6513	0.6248	1.01	4.9041	4.8810	0.5195	3.9745	6.0137	0.8652	1.00
0.03	8	0.2237	7.7108	$(\chi_{v,S}^{-2})$	2.2131	2.1204	0.6317	1.2347	3.7102	0.7650	1.01	6.7496	6.6938	0.8207	5.3649	8.5626	0.6447	1.00
				HC	0.3566	0.2165	0.3828	0.0064	1.3704	0.5614	1.11	7.6705	7.6344	0.8350	6.1926	9.4141	0.8747	1.00
0.3	8	<0.0001	8.2110	$(\chi_{v,S}^{-2})$	1.8041	1.7300	0.5050	1.0513	2.9742	0.8541	1.00	7.7970	7.7091	0.8790	6.3354	9.7150	0.6680	1.00
				HC	0.1184	0.0601	0.1484	0.0006	0.5081	0.6499	1.06	8.2006	8.1574	0.8104	6.7311	10.062	0.8818	1.01
0.05	1	0.0449	0.9737	$(\chi_{v,S}^{-2})$	0.2887	0.2811	0.0795	0.1633	0.4543	0.7601	1.00	0.8601	0.8570	0.1014	0.6764	1.0677	0.6476	1.00
				HC	0.0678	0.0498	0.0669	0.0017	0.2433	0.6963	1.09	0.9726	0.9660	0.1122	0.7606	1.2089	0.8316	1.00
0.1	1	0.2430	0.8681	$(\chi_{v,S}^{-2})$	0.4012	0.3900	0.1170	0.2168	0.6854	0.6452	1.01	0.8119	0.8029	0.1092	0.6325	1.0573	0.6136	1.00
				HC	0.2268	0.2169	0.1328	0.0160	0.5136	0.6425	1.00	0.9069	0.8965	0.1365	0.6574	1.1926	0.6048	1.00
0.5	1	0.6857	0.8070	$(\chi_{v,S}^{-2})$	0.7564	0.7360	0.1946	0.4431	1.2107	0.4998	1.00	0.8186	0.8139	0.1212	0.6074	1.0710	0.6055	1.00
				HC	0.7082	0.7068	0.2111	0.3266	1.1693	0.7093	1.03	0.8308	0.8200	0.1370	0.5983	1.1174	0.7977	1.01
1.0	1	0.9288	1.3627	$(\chi_{v,S}^{-2})$	1.0629	1.0447	0.2629	0.6050	1.6496	0.5635	1.00	1.3618	1.3558	0.1886	1.0209	1.7529	0.6107	1.00
				HC	0.9401	0.9079	0.2861	0.4600	1.5646	0.7065	1.02	1.3845	1.3663	0.2024	1.0380	1.8006	0.8124	1.00
0.07	6	0.2507	5.7746	$(\chi_{v,S}^{-2})$	1.6935	1.6122	0.4811	0.9593	2.7936	0.7603	1.00	5.1111	5.0594	0.6036	4.0211	6.5185	0.6480	1.00
				HC	0.2998	0.1975	0.3135	0.0033	1.0564	0.5760	1.09	5.8007	5.7680	0.6313	4.6230	7.0731	0.8656	1.01
0.7	6	1.2383	4.7563	$(\chi_{v,S}^{-2})$	2.1189	2.0505	0.6153	1.1409	3.5246	0.6610	1.00	4.4630	4.4246	0.5587	3.4763	5.6572	0.6264	1.01
				HC	1.1893	1.0960	0.6333	0.1965	2.6282	0.5699	1.00	4.8899	4.8401	0.6927	3.6580	6.3870	0.8404	1.02
0.09	8	<0.0001	8.4050	$(\chi_{v,S}^{-2})$	2.3169	2.1960	0.7028	1.2862	4.0000	0.7581	1.00	7.2775	7.2057	0.8767	5.8100	9.0892	0.6471	1.00
				HC	0.1344	0.0306	0.2366	0.0007	0.7694	0.6410	1.03	8.4020	8.3493	0.8498	6.8836	10.243	0.8807	1.04
0.9	8	0.5396	8.5494	$(\chi_{v,S}^{-2})$	2.9830	2.8282	0.9100	1.6091	5.1945	0.6838	1.00	7.3021	7.2452	0.9541	5.5731	9.2495	0.6168	1.00
				HC	0.5606	0.3302	0.6740	0.0009	2.4129	0.5199	1.00	8.6582	8.6184	1.0601	6.6583	10.799	0.8661	1.00

REML: restricted maximum likelihood estimations; M: mean; Me: median; SD: standard deviation; Q0.025: quantile 0.025; Q0.975: quantile 0.975; HD: Hellinger distance; DG: Gelman-Rubin diagnosis.

1.0, suggesting that both chains originate from the marginal posterior distribution of each variable, regardless of the prior density considered (Table 2). On the other hand, when n = 200 and p = 349, every HD distance between the prior and posterior densities was less than its corresponding HD when n = 800 and p = 349.

This result indicates the consistency of the Bayesian learning as the model is updated, adding new available information (Tables 1 and 2). The posterior density of ϕ_ϵ registered more learning under the HC prior than under ($\chi^{-2}_{v,S}$) prior. In contrast, for ϕ_β, the learning under prior density HC, was only greater in the data generated with parameters (ϕ^r_β = 0.5, ϕ^r_ϵ = 1.0) and (ϕ^r_β = 1.0, ϕ^r_ϵ = 1.0). These greater HD distances expose the versatility of the prior HC density to efficiently learn in high dimensions, even when the true value of parameter ϕ^r_β moves away from zero, which is the region in which the prior density accumulates its greatest mass.

On the other hand, the CIs built from the posterior density for ϕ_β with prior ($\chi^{-2}_{v,S}$) in contexts of great uncertainty, which is induced by ϕ^r_ϵ = 6.0 (8.0), tended to overestimate parameter ϕ^r_β. Nevertheless, when the variance of the error uncertainty reduced (ϕ^r_ϵ = 1.0), the CI at a probability of 0.95 correctly captured true value ϕ^r_β when true parameters ϕ^r_β were relatively far from the origin (ϕ^r_β = 0.5, 1.0). This result shows the difficulty of correctly estimating parameter ϕ^r_β if its real value is near the origin, regardless of the noise induced by component ϕ^r_ϵ. Results also indicate that parameter ϕ^r_ϵ tended to be correctly estimated, regardless of its value ϕ^r_ϵ = 1.0, 6.0, 8.0.

In contrast with the CIs built under ($\chi^{-2}_{v,S}$), the CIs built under HC prior contained the real values of the parameters, regardless of the real values of parameters ϕ^r_β or ϕ^r_ϵ (Table 2, Figure 4). However, in accordance with the true parameters (ϕ^r_β, ϕ^r_ϵ) used in the data-generating mechanism, three substantial behaviors can be distinguished. First, when the variance of the error is large (ϕ^r_ϵ = 6.0, 8.0) and variance component ϕ^r_β tends towards zero, the posterior distribution of ϕ^r_β was observed to present a positive bias, taking value ϕ^r_β as a reference, which shows the influence of parameter ϕ^r_ϵ to attract the distribution of ϕ^r_β towards value ϕ^r_ϵ (Figures 4A and 4B). Second, when the variance of the error decreases (ϕ^r_ϵ = 1), the posterior distribution of ϕ_β usually centers its mass around true value ϕ^r_β; under these circumstances, the posterior means and medians of ϕ_β practically coincide with true values ϕ^r_β used to simulate the data (Table 2). Third, when parameter ϕ^r_β moves away from the origin (ϕ^r_β = 0.5, 0.7, 0.9, 1.0), the posterior distribution of ϕ_β concentrates its mass around the value ϕ^r_β, regardless of the location of the posterior distribution of ϕ_ϵ (Figure 4). Regardless of the density of the prior density assigned to the variance components and the noise induced by ϕ^r_ϵ, the CIs tended to be shorter as the sample size increased.

Regarding the estimation of the variance components according to the linear mixed model via REML, two substantial behaviors can be distinguished. In scenarios in which ϕ^r_β is relatively near zero (ϕ^r_β = 0.01, 0.03, 0.07, 0.09) and ϕ^r_ϵ is relatively large (ϕ^r_β = 6.0, 8.0), component ϕ^r_β was significantly overestimated, whereas ϕ^r_ϵ was estimated correctly (Table 2). In general terms, the total variability (ϕ^r_β + ϕ^r_ϵ) was appropriately captured. However, the method is incapable of breaking the total variability down

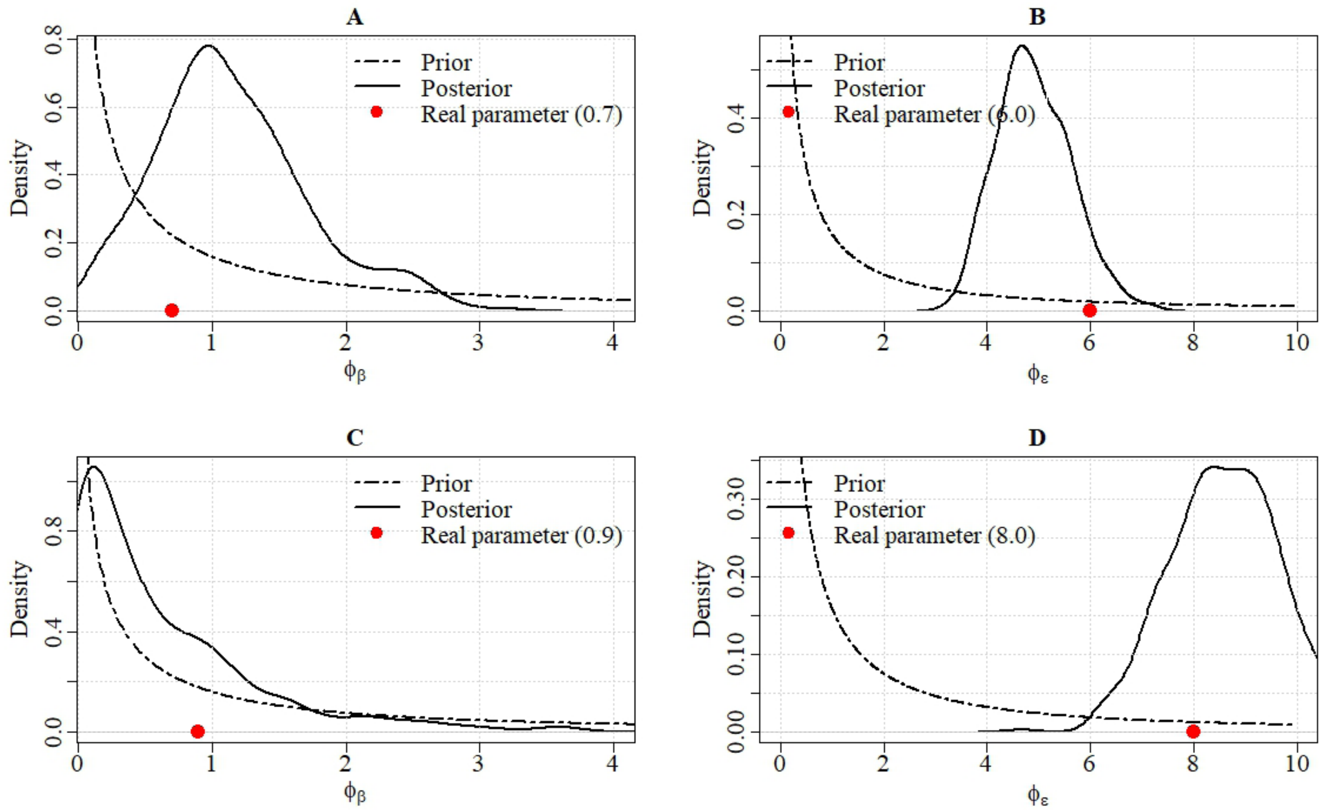

Figure 4. Posterior densities for ϕ_β and ϕ_ϵ with a Half Cauchy prior (HC), $n << p$ (n = 200, p = 349) under two scenarios. A, B: (ϕ_β^r = 0.07, ϕ_ϵ^r = 6.0); C, D: (ϕ_β^r = 0.09, ϕ_ϵ^r = 8.0).

into its individual components, ϕ_β^r and ϕ_ϵ^r. When ϕ_ϵ^r = 1.0, the REML method correctly estimated each variance component, ϕ_β^r and ϕ_ϵ^r, regardless of the true value ϕ_β^r used to generate the data. In addition, the results obtained with $n > p$ contrast with those obtained with $n << p$. In contexts in which $n > p$, the REML method did not detect component ϕ_β^r (Table 1), whereas when the number of records is lower than the number of parameters ($n < p$), ϕ_β^r was overestimated when ϕ_ϵ^r = 6.0, 8.0 and correctly estimated when ϕ_ϵ^r = 1.0 (Table 2).

CONCLUSIONS

The influence of $(\chi_{v,S}^{-2})$ prior and standard Half Cauchy (HC) distributions on the posterior density of each variance component in the Bayesian Ridge regression model was evaluated in contexts in which there are more observations than predictor variables and in high dimensions. Based on the results of the genetic evaluation simulation, the following conclusions can be drawn: a) The $(\chi_{v,S}^{-2})$ prior density hyperparameters have a strong influence over the learning of the posterior density of the variance of the predictor variables, particularly when true parameter ϕ_β^r is near the value of zero and the variance of the error is large ($\phi_\epsilon^r \geq 6.0$). b) Under prior density ($\chi_{v,S}^{-2}$), when the variance of the error is large ($\phi_\epsilon^r \geq 6.0$), the CIs at 0.95 overestimate variance component ϕ_β. c) Under prior density ($\chi_{v,S}^{-2}$), the CIs for parameter ϕ_ϵ^r are almost always exact and their accuracy increases with the sample size. d) If the HC prior density is assigned to every variance component, the CIs of 0.95 for ϕ_β and ϕ_ϵ are almost always exact and their accuracy increases with the sample size. e) The problem of the curse of dimensionality does not interfere in the precision of the CIs, neither for ϕ_β^r nor for ϕ_ϵ^r. f) Regardless of the prior density considered in the parametric scenarios explored in this study, the CIs for ϕ_β^r are highly inaccurate when $n \leq 600$. The accuracy of the CIs for ϕ_β^r worsens specifically under prior density ($\chi_{v,S}^{-2}$), where the CIs are also usually inexact. However, the CIs for both ϕ_β^r and ϕ_ϵ^r tend to be shorter as the sample size increases, regardless of the prior density assigned to the variance component.

The results show the suitability of the HC distribution as a prior density to model variance components in the Bayesian Ridge regression model. Its election must be preferred over the REML method or, in Bayesian contexts, over distribution ($\chi_{v,S}^{-2}$) as the prior of the variance components.

ACKNOWLEDGEMENTS

To the National Council for Humanities, Science and Technology (CONAHCYT) for the economic support granted for the doctorate studies of the main author.

REFERENCES

Bernardo JM, Smith AFM. 2000. Bayesian theory. John Wiley and Sons: Hoboken, NJ, USA. 586 p.

Bezanson J, Edelman A, Karpinski S, Shah VB. 2017. Julia: A fresh approach to numerical computing. SIAM Review 59 (1): 65–98. https://doi.org/10.1137/141000671

Browne WJ, Draper D. 2006. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. Bayesian Analysis 1 (3): 473–514. https://doi.org/10.1214/06-ba117

Casella G, Berger, RL. 2002 Statistical Inference (Second edition). Duxbury Press: Belmont, CA, USA. 686 p.

Casella G, George EI. 1992. Explaining the Gibbs sampler. The American Statistician 46 (3): 167–174. https://doi.org/10.1080/00031305.1992.10475878

Cheng H, Garrick D, Fernando R. 2015. XSim: Simulation of descendants from ancestors with sequence data. G3: Genes|Genomes|Genetics 5 (7): 1415–1417. https://doi.org/10.1534/g3.115.016683

de los Campos G, Gianola D, Rosa GJ, Weigel KA, Crossa J. 2010. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. Genetics Research 92 (4): 295–308. https://doi.org/10.1017/s0016672310000285

Endelman JB. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome 4 (3): 250–255. https://doi.org/10.3835/plantgenome2011.08.0024

Falconer DS, Mackay TFC. 1996. Introduction to quantitative genetics (Fourth edition). Longman: Harlow, UK. 448 p.

Fink D. 1997. A Compendium of conjugate priors. Montana State University: Bozeman, MT, USA. 47 p.

Gelman A, Hill J. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press: Cambridge, UK. 625 p.

Gelman A, Rubin DB. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7 (4). https://doi.org/10.1214/ss/1177011136

Gelman A. 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1 (3). https://doi.org/10.1214/06-ba117a

Gianola D, de los Campos G, Hill WG, Manfredi E, Fernando R. 2009. Additive genetic variability and the Bayesian alphabet. Genetics 183 (1): 347–363. https://doi.org/10.1534/genetics.109.103952

Gianola D. 2013. Priors in whole-genome regression: The Bayesian alphabet returns. Genetics 194 (3): 573–596. https://doi.org/10.1534/genetics.113.151753

Giraud C. 2015. Introduction to high-dimensional statistics. Universite Paris Saclay: Paris, France. 345 p.

Lehermeier C, Wimmer V, Albrecht T, Auinger HJ, Gianola D, Schmid VJ, Schön CC. 2013. Sensitivity to prior specification in Bayesian genome-based prediction models. Statistical Applications in Genetics and Molecular Biology 12 (3). https://doi.org/10.1515/sagmb-2012-0042

Meuwissen TH, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157 (4): 1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Pérez-Rodríguez P, Acosta-Pech R, Pérez-Elizalde S, Cruz CV, Espinosa JS, Crossa J. 2018. A Bayesian genomic regression model with skew normal random errors. G3 Genes|Genomes|Genetics 8 (5): 1771–1785. https://doi.org/10.1534/g3.117.300406

Pérez-Rodríguez P, de los Campos G. 2014 Genome-wide regression and prediction with the BGLR statistical package. Genetics 198 (2): 483–495. https://doi.org/10.1534/genetics.114.164442

Polson NG, Scott JG. 2012. On the half-cauchy prior for a global scale parameter. Bayesian Analysis 7 (4). https://doi.org/10.1214/12-ba730

Roos M, Herd L. 2011. Sensitivity analysis in Bayesian generalized linear mixed models for binary data. Bayesian Analysis 6 (2): 259–278. https://doi.org/10.1214/11-ba609

Silverman BW. 1998. Density estimation for statistics and data analysis. Chapman and Hall: London, UK. 22 p.

Spiegelhalter DJ, Thomas A, Best NG, Gilks, WR. 1996. BUGS 0.5 examples. MRC Biostatistics Unit: Cambridge, UK. 75 p.