BAYESIAN ALGORITHMS FOR ANALYSIS OF CATEGORICAL ORDINAL DATA

Fábio Mathias CORRÊA[1]
José Waldemar da SILVA[2]
Daniel Furtado FERREIRA[3]
Júlio Silvio de Sousa BUENO FILHO[3]

■ *ABSTRACT: This study describes and evaluates a package that implements extensions of the algorithm first presented by Nandram and Chen (1996), replacing Gaussian distribution (NCG) with Student's t distribution (NCt) for Bayesian analysis of ordinal categorical data using mixed models. The algorithms described by Albert and Chib (1993) and Cowles (1996) were also implemented. Comparison was carried on using two different designs. A Steiner triple system with seven treatments used mostly to estimate fixed effects and a 10x10 square lattice designed to rank and select among random effects. Different situations for intraclass correlations were also considered. We reported the total number of iterations required for convergence diagnostics, and the mean square error (MSE) on posterior estimates of both random and fixed effects as well as posterior estimates of intraclass correlation. NCG and NCt algorithms resulted in lower MSE for both designs. This algorithm has also shown faster convergence rates. For the square lattice, NCG and NCt algorithms overestimated the intraclass correlation when the simulated value was large (0.8). But the bias on MSE relative to the other designs did not increase. A real experiment from plant breeding is given as an example of package use, an Incomplete Block Design to evaluate resistance of tomato varieties to late blight (caused by Phytophthora infestans). Gaussian distribution was the parcimonious choice for the latent trait. Algorithms are consistent with regard to the ranking of varieties.*

■ *KEYWORDS: MCMC; bayesthresh; threshold models.*

[1]Universidade Estadual de Santa Cruz – UESC, DCET, Ilhéus, BA, Brazil. E-mail: *fmcorrea@uesc.br*
[2]Universidade Federal de Uberlândia – UFU, FAMAT, Uberlândia, MG, Brazil. E-mail: *zewaldemar@famat.ufu.br*
[3]Universidade Federal de Lavras – UFLA, DES, Lavras, MG, Brazil. E-mail: *danielff@des.ufla.br*, *jssbueno@des.ufla.br*

1 Introduction

Ordinal categorical data are widely used in different areas of research and often come from subjective measures. Many examples can be found in the literature; for example, in opinion studies (GOB; MCCOLLIN; RAMALHOTO, 2007), in food science with the use of hedonic scales to evaluate sensory attributes (PIEPHO; KALKA, 2003), in quantitative genetics and breeding (SORENSEN et al., 1995), in phytopathology with the use of diagrammatic scales for disease quantification (CORRÊA; BUENO-FILHO; CARMO, 2009) and in the physical sciences with Mohs hardness scale (TABOR, 1954).

A number of methods have been proposed to analyze categorical data. We will focus on threshold models, that have appealing properties specially in regard to interpretability of intraclass correlation based on ratios of linear combinations on variance components and has been preferred in quantitative genetics. Other distributions have been proposed for the latent trait, specially logit link (MCCULLAGH, 1980; AGRESTI, 2013).

Threshold models use a latent variable with a continuous distribution in such a way that the response observed for one category is within the limits that define these categories (ALBERT; CHIB, 1993; SORENSEN et al., 1995; MCCULLOGH; SEARLE, 2001; KIZILKAYA et al., 2003; PIEPHO; KALKA, 2003). Bayesian implementations of threshold models were first proposed by Albert and Chib (1993) for the analyses of fixed-effect models. However, their algorithm lead to strong autocorrelation between samples and slower convergence rate in Gibbs sampling when analyzing mixed-effect models (SORENSEN et al., 1995) .

To accelerate the convergence process, Cowles (1996) proposed an algorithm that used a blocking step to accept or reject the entire vector of threshold parameters. On the other hand, Nandram and Chen (1996) proposed a reparameterization of the threshold parameters by considering the Dirichlet distribution as the generator of threshold parameter candidates for fixed effect models. In all of those algorithms it is straightforward to use Student's t distribution to replace the Gaussian distribution. This allows to investigate the need of a more robust (heavy tailed) distribution for the latent trait.

We didn't find any studies that compare the efficiency of those original algorithms and their extensions for mixed models and Student's t distribution as proposed by Albert and Chib (1993) (ACG and ACt), Cowles (1996) (MCG and MCt) and Nandram and Chen (1996) (NCG and NCt). The extended version by Silva and Bueno-Filho (2010) was not compared in simulation studies as well.

The consistency and speed of the algorithms are compared in a simulation study using two designs and four intraclass correlation specifications as well as simulating from assymetric distributiions for the latent traits. The objective of those comparisons is to acess if Student's t or original Gausian (probit) link are affected by asymetric or heavy tailed distributions on latent traits. A real experiment in plant breeding is also provided. In this example an IBD in which 66 tomato families were compared with respect to their tolerance to the leaf blight disease (caused by *Phytophthora infestans*). In the following section are presented the extensions for mixed models of the algorithm proposed by Nandram and Chen (1996), using Gaussian and Student's t distributions. The third section presents the designs, methods and results of the simulation study. Section four presents the tomato

breeding experiment, followed by the conclusions from our study in the final section.

An additional contribution, the Bayesthresh package was developed to perform the analyzes presented in the article. Being available at ⟨https://cran.r-project.org/web/packages/Bayesthresh/index.html⟩

2 Extensions for mixed models of Nandram and Chen (1996) algorithm

2.1 NC: Nandram and Chen (1996) algorithm for the analyses of mixed models, using a cumulative Gaussian distribution (probit link)

Let $\mathbf{Y}$ be the $n \times 1$ vector of realized values for the ordinal data, with range 1 to K. Let γ be the vector of threshold parameters that partitions the real line into K disjoint intervals as follows: $(\gamma_0, \gamma_1); [\gamma_1, \gamma_2); \ldots; [\gamma_{K-1}, \gamma_K)$ with $\gamma_0 = -\infty$ and $\gamma_K = +\infty$. Nandram and Chen (1996) proposed the use of the Dirichlet distribution to generate candidates for γ and adopted $\gamma_1 = 0$. In this manner, for every class k, γ_k^* will be the distance between γ_k and γ_1, which is obtained from the transformation $\gamma_k^* = \gamma_k - \gamma_1$.

Let $\mathbf{L}$ be the vector of latent variables and L_i^* be the difference of the latent variable within the k class.

$L_i^* = L_i - \gamma_i, i = 1, ..., n$, and then $Y_i = k$ if $\gamma_{k-1} - \gamma_1 \leq L_i^* \leq \gamma_k - \gamma_1$.

Let δ be an auxiliary variable, and γ_k^{**} be the new vector of threshold parameters, let θ^* be the new vector of fixed and random effects in the linear predictor, and L^* be the reparameterized latent variable (1).

$$\begin{aligned} &\delta = 1/\gamma_{K-1}^*, \\ &\gamma_k^{**} = \delta_{\gamma_k^*},\ k = 0, 1, 2, ..., K, \\ &\theta^* = \delta\theta \text{ e } L^* = \delta L^*; \end{aligned} \tag{1}$$

The Jacobian of the δ transformation is given as $[\delta^2]^{-\frac{1}{2}(n+m+K)}$ Nandram and Chen (1996), in which m is the number of explanatory variables. This is a proxy to the variance of the latent trait. Assuming an inverse gamma prior for δ^2 and using a cumulative Gaussian distribution as the link function, we have the joint posterior distribution, given as:

$$\begin{aligned} p\left(\theta^*, \gamma_k^{**}, \delta^2, L^* | y\right) \propto &\left[\prod_{i=1}^{n} \Phi\left(L_i^*, w_i\theta^*, \delta^2\right) I_{[\gamma_k^{**}, \gamma_{k+1}^{**})}(L_i^{**})\right] \\ &\times \Phi(\theta^*; \mathbf{0}, \delta^2 V)(\delta^2)^{-k/2} p(\sigma_u^2) p(\delta^2); \end{aligned} \tag{2}$$

Let (σ_u^2) represent the variance component of the $\mathbf{u}$ vector of random effects. Let V be the covariance matrix of the linear effects (both fixed and random). For the random effects this is often called $\mathbf{A}$, that is a $m \times m$ matrix of known constants, Henderson (1976). Some other measure of genetic similarity obtained from molecular markers could be used instead Wang and Da (2014).

The full conditional distribution of θ^* is given as:

$$\theta^* | L^*, \delta^2, Y, \sigma_u^2 \sim N\left(B^{-1}W'L^*, \delta^2 B^{-1}\right); \tag{3}$$

in which $W = [X|Z]$, X and Z are design matrices for the fixed and random effects, respectively, and $B = \delta^2 V^{-1} + W'W$.

Let the inverse gamma prior for δ^2 have hyperparameters c and d as follows:

$$p(\delta^2) \propto (\delta^2)^{-(c+1)} exp\left\{-\frac{d}{\delta^2}\right\}; \tag{4}$$

Thus, the full conditional posterior distribution for δ^2 will be:

$$\delta^2|\theta^*, Y, \sigma_u^2 \sim IG(a_\delta, b_\delta); \tag{5}$$

where a_δ is the shape parameter and b_δ is the scale parameter, presented in equations (6) and (7):

$$a_\delta = \frac{n + m + K + 2c}{2}; \tag{6}$$

$$b_\delta = \frac{(L^* - W\theta^*)'(L^* - W\theta^*) + \theta^{*'} V^{-1}\theta^* + 2d}{2}. \tag{7}$$

The full conditional distribution for the variance of the random effects (σ_u^2) is given as:

$$p(\sigma_u^2|\theta^*, L^*, Y) = (\sigma_u^2)^{-(\frac{q}{2}+a+1)} exp\left\{\frac{-1}{2\sigma_u^2}(u^{'}u + 2b)\right\}; \tag{8}$$

That can be seen as an inverse gamma distribution:

$$\sigma_u^2|\theta^*, L^*, y \sim IG\left(\frac{q + 2a}{2}, \frac{u^{'}u + 2b}{2}\right); \tag{9}$$

With the reparameterization in (1), the conditional distribution of $(\gamma^{**}|\theta^*, \delta^2, Y)$ becomes:

$$\begin{aligned}\pi(\gamma^{**}|\theta^*, \delta^2, Y) \propto & \prod_{Y_i=2}\left[\Phi\left(\frac{\gamma_2^{**} - w_i^{'}\theta^*}{\delta}\right) - \Phi\left(\frac{-w_i^{'}\theta^*}{\delta}\right)\right] \\ & \times \prod_{Y_i=3}\left[\Phi\left(\frac{\gamma_3^{**} - w_i^{'}\theta^*}{\delta}\right) - \Phi\left(\frac{\gamma_2^{**} - w_i^{'}\theta^*}{\delta}\right)\right] \cdots \\ & \times \prod_{Y_i=K-1}\left[\Phi\left(\frac{1 - w_i^{'}\theta^*}{\delta}\right) - \Phi\left(\frac{\gamma_{K-2}^{**} - w_i^{'}\theta^*}{\delta}\right)\right];\end{aligned} \tag{10}$$

and the full conditional posterior distribution for $(L^*|\gamma^{**}, \theta^*, \delta^2, Y)$ is given as:

$$L^*|\gamma^{**}, \theta^*, \delta^2, Y = k \sim N(W\theta^*, \delta^2); \tag{11}$$

As for γ^{**} sampling, it is performed by means of the Metropolis-Hastings algorithm using an auxiliary vector $\mathbf{p}$ whose elements are described as follows:

$$p_{k-1} = \gamma_k^{**} - \gamma_{k-1}^{**},\ k = 2, ..., K-1; \tag{12}$$

where:

$$p = (p_1, p_2, \ldots, p_{k-2})',\ p_k \geq 0,\ k = 1, 2, \ldots, K-2 \text{ and } \sum_{k=1}^{K-2} p_k = 1; \qquad (13)$$

According to Nandram and Chen (1996), the mean value theorem gives:

$$\Phi\left(\frac{\gamma_k^{**} - w_i^{'}\theta^*}{\delta}\right) - \Phi\left(\frac{\gamma_{k-1}^{**} - w_i^{'}\theta^*}{\delta}\right) = \frac{1}{\delta}\Phi\left(\frac{\xi_{k-1} - w_i^{'}\theta^{**}}{\delta}\right) p_{k-1}; \qquad (14)$$

where $\xi_{k-1} \in (\gamma_k^{**}; \gamma_{k-1}^{**}), k = 2, 3, ..., K-1,$ and $\Phi(.)$ is the cumulative standard normal density function. Thus, the conditional distribution of γ^{**} is given as the product of $h_1 \times h_2$, where h_2 is the core of a Dirichlet distribution with parameters $n = (n_2 + 1, ..., n_{k-1} + 1)^{'}$ and does not depend either on θ^* or on δ. We proceed as follows:

$$h_1(\xi) = \prod_{k=1}^{K-2} \prod_{i=1}^{n_k} \Phi\left(\frac{\xi_k - w^{'}\theta^*}{\delta}\right);$$

and

$$h_2(\mathbf{p}) = \prod_{k=1}^{K-2} P_k^{n_k+1}.$$

The acceptance probability for the new vector γ is given by the $min\{1, \alpha\}$, where

$$\alpha = w(\gamma^{**(j)}, \mathbf{p}^j)/w(\gamma^{**(j-1)}, \mathbf{p}^{j-1}); \qquad (15)$$

and j represents the j^{th} iteration of the algorithm and:

$$w(\gamma^{**}, \mathbf{p}) = P(\gamma^{**}|\theta^*, \delta^2)/P(\mathbf{p}|n, \theta^*, \delta^2). \qquad (16)$$

The expression of $P(\gamma^{**}|\theta^*, \delta^2)$ is given in (10), and $P(\mathbf{p}|n, \theta^*, \delta^2)$ is the Dirichlet distribution:

$$p(\mathbf{p}) = \frac{1}{Z(n)} \prod_{k=1}^{K-2} p_k^{n_{k+1}-1}, \qquad (17)$$

where $p_1, ..., p_{K-2} \geq 0; \sum_{k=1}^{K-2} p_k = 1;\ n_2, ..., n_{K-1} > 0$. and $Z(n)$ is the normalizing constant given as:

$$Z(n) = \frac{\prod_{k=1}^{K-2} \Gamma_{(n_{k+1})}}{\Gamma\left(\sum_{k=1}^{K-2} n_{k+1}\right)}; \qquad (18)$$

As we want the joint distribution of L^* and γ^{**}, we sample for $\mathbf{p}$, and this is used to build γ_k^{**}. If the new threshold parameter vector is accepted, the values of L^* are used; otherwise we keep the previous sample of the latent variable.

2.2 NCt: Nandram and Chen (1996) modified algorithm, using the cumulative Student's t distribution as a link function for mixed models

The likelihood is given by:

$$P(Y_i = k|\theta^*, \delta, v, \gamma^{**}) = F_v\left(\frac{\gamma_k^{**} - w_i^{'}\theta^*}{\delta}\right) - F_v\left(\frac{\gamma_{k-1}^{**} - w_i^{'}\theta^*}{\delta}\right), k = 1, 2, ..., K; \tag{19}$$

where F_v is the function of cumulative Student's t distribution with v degrees of freedom. To promote the algebraic ease of obtaining full conditional posterior distributions, Student's t distribution will be written in two stages (equations 20 and 21) as a mixture of Gaussian distributions with inverse-gamma distributions for the variance parameters, as described in Sorensen and Gianola (2002) .

$$\lambda_i|v \sim Gamma(\frac{v}{2}, \frac{v}{2}); \tag{20}$$

$$L_i^*|\theta^*, \delta^2, \lambda_i \sim N\left(w_i^{'}\theta^*, \frac{\delta^2}{\lambda_i}\right); \tag{21}$$

Thus, the model in (19) may be rewritten as:

$$P(Y_i = k|\theta^*, \delta_i, v, \gamma^{**}) = F_v\left(\frac{\gamma_k^{**} - w_i^{'}}{\frac{\delta}{\sqrt{\lambda_i}}}\theta^*\right) - F_v\left(\frac{\gamma_{k-1}^{**} - w_i^{'}}{\frac{\delta}{\sqrt{\lambda_i}}}\theta^*\right) k = 1, 2, ..., K; \tag{22}$$

The prior distribution for v follows Kizilkaya et al. (2003), with $p(v) = 1/(1+v)^2$.

Assuming a uniform prior for θ^*, the joint posterior for all parameters is:

$$\begin{aligned} p(\theta^*, \gamma_k^{**}, L^{**}, \lambda|y) \propto (\delta^2)^{-\frac{k}{2}} & \left[\prod_{i=1}^{n} \phi\left(L_i^*, w_i\theta^*, \frac{\delta^2}{\lambda_i}\right) I_{[\gamma_k^{**}, \gamma_{k+1}^{**})}(L_k^*)\right] \\ & \times \left[\prod_{i=1}^{n} \lambda_i^{(\frac{v}{2})-1} exp\left(-\frac{\lambda_i}{2}v\right)\right] \Phi(\theta^*, \mathbf{0}, \delta^2 V)\frac{1}{(1+v)^2} \\ & \times p(\sigma_u^2)p(\delta^2)p(v); \end{aligned} \tag{23}$$

where $\lambda = \{\lambda_i\}_{i=1}^n$, and V was defined in (2).

The prior for δ^2 is as in (4) and for σ_u^2 is an inverse gamma (24). The full conditional posterior distributions for all parameters are presented below. The full conditional distribution of θ^* is:

$$P(\sigma_u^2) = (\sigma_u^2)^{-(a+1)} exp\left(\frac{-b}{\sigma_u^2}\right); \tag{24}$$

$$\theta^*|L^*, \delta^2, v, \lambda \sim N(M^{-1}W^{'}R^{-1}L^*, \delta^2 M^{-1}); \tag{25}$$

where $M = W^{'}R^{-1}W + \delta^2 V^{-1}$

The full conditional distribution for λ_i follows Kizilkaya et al. (2003):

$$p(\lambda_i|\lambda_{-i}, L^*, \theta^*, v) \propto \lambda_i^{\left(\frac{v+1}{2}\right)-1} exp\left(-\frac{\lambda_i}{2}\left(\left(L_i^* - w_i^{'}\theta^*\right)^2 + v\right)\right); \tag{26}$$

where λ_{-1} represents the elements of λ, except λ_i. The above distribution is proportional to a gamma distribution with parameters $(v+1)/2$ and $(L_i^* - w_i^{'}\theta^*)^2 + v$; that is,

$$\lambda_i|\lambda_{-i}, L_i^*, \theta^*, v, \delta^2 \sim Gamma\left(\frac{v+1}{2}, \frac{1}{2}\left(\left(L_i^{**} - w_i^{'}\theta^*\right)^2 + v\right)\right); \tag{27}$$

and

$$L_i^*|\lambda_i, \theta^*, v, \delta^2 \sim N\left(w_i^{'}\theta^*, \frac{\delta^2}{\lambda_i}\right); \tag{28}$$

Sampling from conditional distribution of v (equation 29) is performed via Metropolis-Hastings algorithm Kizilkaya et al. (2003) .

$$p(v|\theta^*, L^*, \lambda, \delta^2) \propto \left(\frac{(\frac{v}{2})^{(v/2)}}{\Gamma(v/2)}\right)^n \left(\prod_{i=1}^{n} \lambda_i^{\frac{v}{2}-1} exp\left(\frac{v}{2}\lambda_i\right)\right) \frac{1}{(1+v)^2}; \tag{29}$$

Sampling δ^2 is done using the same distribution of equation (5).

The joint distribution of L^{**} and γ^{**} is the product of $(\gamma^{**}|\theta^*, \delta^2, \lambda, Y)$ and $(L^*|\gamma^{**}, \theta^*, \delta^2, \lambda, Y)$. The full distribution of the latent variable is as shown in (28). For γ^{**}, the full conditional distribution is given as:

$$\begin{aligned} p(\gamma^{**}|\delta^2, \lambda, \theta^*, Y) \propto & \prod_{Y_i=2}\left[\Phi\left(\frac{\gamma_2^{**} - w_{,}^{'}\theta^*}{\frac{\delta}{\sqrt{\gamma_i}}}\right) - \Phi\left(\frac{-w_i^{'}\theta^{**}}{\frac{\delta}{\sqrt{\lambda_i}}}\right)\right] \\ & \times \prod_{Y_i=3}\left[\Phi\left(\frac{\gamma_3^{**} - w_{,}^{'}\theta^*}{\frac{\delta}{\sqrt{\lambda_i}}}\right) - \Phi\left(\frac{\gamma_2^{**} - w_i^{'}\theta^*}{\frac{\delta}{\sqrt{\lambda_i}}}\right)\right] \\ & \cdots \prod_{Y_i=K-1}\left[\Phi\left(\frac{1 - w_{,}^{'}\theta^{**}}{\frac{\delta}{\sqrt{\lambda_i}}}\right) - \Phi\left(\frac{\gamma_{K-2}^{**} - w_i^{'}\theta^*}{\frac{\delta}{\sqrt{\lambda_i}}}\right)\right]; \end{aligned} \tag{30}$$

In the same manner as in (14), according to the mean value theorem, we have:

$$\Phi\left(\frac{\gamma_k^{**} - w_i^{'}\theta^*}{\frac{\delta}{\sqrt{\lambda_i}}}\right) - \Phi\left(\frac{\gamma_{k-1}^{**} - w_i^{'}\theta^*}{\frac{\delta}{\sqrt{\lambda_i}}}\right) = \frac{1}{\frac{\delta}{\sqrt{\lambda_i}}}\Phi\left(\frac{\xi_{k-1} - x_i^{'}\theta^*}{\frac{\delta}{\sqrt{\lambda_i}}}\right) p_{k-1}; \tag{31}$$

where $\xi_{k-1} \in (\gamma_k^{**}; \gamma_{k-1}^{**}), k = 2, ..., K-1$, and $\Phi(.)$ is the standard normal density function. And from (31), we obtain:

$$\pi(\gamma^{**}|\theta^*, \delta^2, Y) \propto h_3(\xi)h_4(p); \tag{32}$$

where h_3 and h_4 are from (32), respectively:

$$\begin{aligned} h_3(\xi) &= \prod_{k=1}^{K-2}\prod_{i=1}^{n_k}\left(\frac{\xi_k - w_i^{'}\theta^*}{\frac{\delta}{\sqrt{\lambda_i}}}\right); \\ h_4(p) &= \prod_{k=1}^{K-2} p_k^{n_k+1}; \end{aligned} \tag{33}$$

The other procedures (sampling latent variable and threshold parameters) are the same as in previous section.

3 Simulation study

The simulation study was conducted to evaluate the consistency of implemented routines under different scenarios. Mean Squared Errors (MSE) were used as a guide to compare the precision of posterior means. Chain size to achieve ergodic sampling and processing time were also monitored.

Two incomplete block designs were considered: a small design, represented by a Steinerś Triple System (STS) with seven treatments, and a large partially balanced design, represented by a Simple Square Lattice (SSL) with 100 treatments. The treatments will be described in the next subsections.

3.1 Steiner's Triple System (STS)

STS was constructed with three treatments, two replicates and seven blocks of three plots. The linear model used to simulate the effects and realizations of a continuous variable was:

$$u = \mu + Xb + Za + \epsilon; \tag{34}$$

where u is the vector of realizations, μ is the overall mean (equal to five), b is the fixed effects vector, with $b \sim N(0,1)$, and X is the design matrix of the fixed effects; a is the random effects vector, simulated from a normal distribution, $a \sim (0, \sigma_a^2)$, and Z is the design matrix of the random effects.

Four values of σ_a^2 (0.1, 0.2, 0.5 and 0.8) were adopted for the random effects. The sampling error was considered as $\epsilon \sim (0, \sigma_e^2)$, where $\sigma_e^2 = (0.9, 0.8, 0.5 \text{ and } 0.2)$. From the combinations of σ_a^2 and σ_e^2, we obtained four parameter values (0.1, 0.2, 0.5 and 0.8) for the intraclass correlation $\rho = \frac{\sigma_a^2}{\sigma_a^2+\sigma_e^2}$.

The translation of vector u into an ordinal categorical variable was performed using four threshold values, resulting in five categories (named 1 to 5).

Three sets of threshold values were chosen to emulate symmetric, asymmetric and uniform distributions for the latent variable as follows:

$$\eta_i = \begin{cases} 1,\ u_i \leq Q_1; \\ 2,\ Q_1 < u_i \leq Q_2; \\ 3,\ Q_2 < u_i \leq Q_3; \\ 4,\ Q_3 < u_i \leq Q_4; \\ 5,\ u_i > Q_4; \end{cases} \tag{35}$$

Symmetric choice was based on the quantiles 0.0001, 0.15, 0.50 and 0.85; asymmetric was based on quantiles 0.4, 0.5, 0.75 and 0.90; and uniform on quantiles 0.017, 0.28, 0.525 and 0.775. Figure (1) shows examples of the distribution for realizations of categorical variable using the quantiles specification.

Specifications of variance components for the continuous variable resulted in four different configurations of intraclass correlations ρ. The 12 experimental situations were analyzed considering for each one two different values for hyperparameters of the inverse

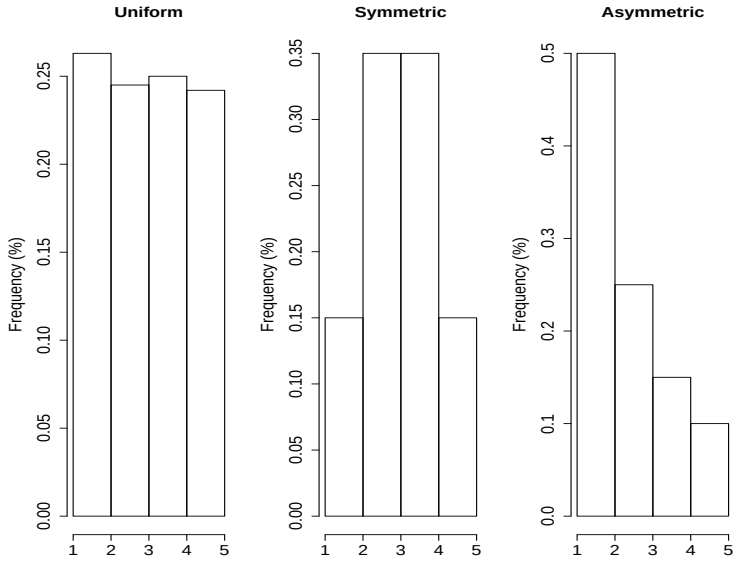

Figure 1 - Example of distribution for realizations of categorical variable using the quantiles specification for the different experimental configurations. Steiner's Triple System - simulation study.

gamma priors for the variance of random effects: a less informative prior inverse-gamma $IG(3,\ 5)$ and a more informative prior $IG(8,\ 5)$. In the algorithms NCG and NCt for σ_e^2, an inverse-gamma prior $IG(20,\ 5)$ was used.

For the analysis of STS the following linear model was used:

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \tag{36}$$

where y_{ij} is the vector of pseudo-realizations of the latent variable from the i^{th} treatment in the j^{th} block, μ is a common experimental constant, τ_i is the effect of i^{th} treatment, taken as fixed (assuming prior proportional to a constant), $\beta_j \sim N(0, \sigma_\beta^2)$ is the effect of k^{th} block, and $\epsilon_{ij} \sim N(0, \sigma_e^2)$ is the experimental error in the latent variable. A total of 1000 experiments were simulated for each experimental situation.

3.2 Square Lattice Design (SLD)

A 10x10 square lattice design (SLD) with 20 blocks, grouped in two repetitions, and 100 treatments was used in this simulation. The model used to generate the observations was the same described in equation (34). The simulated response vector was categorized into nine classes, expanding the scheme presented in (35) using the quantiles 0.005, 0.075, 0.185, 0.325, 0.50, 0.675, 0.825, and 0.925, for a symmetric distribution. The

same simulation procedure described in section (3.1) was used, giving a total of four experimental situations. Like STS, the SLD was analyzed by considering a less informative prior $IG(3,\ 5)$ and a more informative prior $IG(10,\ 2)$. The prior for residual variance in NCG and NCt was the same as that used in the analysis of STS: $IG(20,\ 5)$. For SLD, 500 experiments were simulated for each experimental situation.

The analysis used the same model described in (36), but changing the prior specification for random and fixed effects: β_j is the effect of j^{th} block, taken as fixed (assuming prior proportional to a constant) and $\tau_i \sim N(0, \sigma^2_\tau)$ is the effect of i^{th} treatment, representing a random additive genetic effect.

Figure (2) illustrates sampling from the prior distributions used and how they reflect on intraclass correlation (meaning heritability, for SLD in genetic context). The first situation would be the more realistic to a genetic experiment using a SLD and the last is what you would expect in a nearly perfect blocking in the STS experiment. Note that NCG and NCt have different prior specifications due to a prior specification for σ^2_e. Note also that if a mild informative prior like $IG(3, 5)$ were used for both σ^2_τ and σ^2_e it would result in a uniform prior distribution for intraclass correlation in NCG and NCt. However, this could not be achieved for the other algorithms.

3.3 MCMC simulation and diagnostics procedure

An initial chain of 4000 samples was drawn. The diagnostic proposed by Raftery and Lewis (1992) was used to assess the size of initial sample to discard (”burn in”) and the sampling step size to avoid autocorrelation within the chain (”jump”). Two independent chains of 4000 samples were them drawn with those values of burn in and jump. The convergence for variance components was evaluated with the test of Gelman and Rubin (1992). The analyses were performed with the aid of Bayesthresh package Correa and Bueno-Filho (2012), and the chain diagnosis was performed with the coda package Plummer et al. (2006). Both analyses were implemented in R software (R Development Core Team, 2012).

3.4 Results

The NCG and NCt algorithm had a lower MSE for the estimates of fixed effects when compared to the other algorithms, regardless of the design used (Table 1). The prior for variance components had no effect on the MSE of fixed effects for the NCG and NCt algorithms. For the other algorithms, the increase in intraclass correlation resulted in a larger bias for either informative or non-informative priors and for all distributions. It worths to notice that the more informative prior on σ^2_β the smaller MSE (Table 1). This is expected as the prior would be ”correct” in those situations.

The link function did not affect the posterior of fixed effects, with both the Gaussian and Student's t distributions showing very similar MSE behavior, regardless the shape of true distributions. The MSE for random effects were qualitatively the same for fixed effects using NCG and NCt algorithms (Figure 2).Similar MSE values were found regardless of prior specification and latent trait simulated distribution. When an informative prior on σ^2_β was used, the algorithms ACG, ACt, MCG, and MCt showed a smaller MSE, similar to

Table 1 - Mean squared error (MSE) and bias of the posterior mean of the fixed effects in the STS design.

Prior	Algorithm	Intraclass correlation	Asymmetric		Symmetric		Uniform	
			Bias	MSE	Bias	MSE	Bias	MSE
IG (3,5)	ACG	0.1	5.71	8.62	12.58	9.64	135.49	9.53
		0.2	5.81	1.73	13.45	10.61	15.77	10.84
		0.5	7.32	1.70	14.17	7.93	14.52	8.50
		0.8	16.21	1.12	32.29	20.63	35.49	20.02
	ACt	0.1	8.97	11.44	18.15	12.71	189.03	12.33
		0.2	8.25	12.38	18.52	13.41	20.47	13.58
		0.5	10.62	9.20	20.38	10.41	17.35	10.23
		0.8	16.06	20.04	32.72	20.63	35.92	20.25
	MCG	0.1	3.94	7.75	11.24	8.90	122.15	8.93
		0.2	4.18	8.67	12.17	9.91	14.87	10.17
		0.5	5.76	6.14	12.83	7.17	11.15	7.16
		0.8	13.54	18.36	28.86	19.32	31.54	18.78
	MCt	0.1	8.21	10.65	16.81	11.91	175.34	11.68
		0.2	7.48	11.54	17.15	12.55	19.37	13.06
		0.5	8.77	8.22	17.58	9.28	15.61	9.26
		0.8	14.16	18.50	29.29	19.37	32.63	18.71
	NCG	0.1	-2.15	3.02	1.51	1.73	40.35	4.23
		0.2	-3.90	3.25	1.67	1.70	5.84	4.35
		0.5	-1.17	1.95	1.65	1.12	2.79	2.42
		0.8	-3.40	2.93	2.26	1.91	6.16	4.24
	NCt	0.1	-1.55	1.99	1.25	1.59	23.69	3.14
		0.2	-2.04	2.15	1.79	1.60	5.58	3.34
		0.5	-0.46	1.40	1.64	1.09	3.03	2.03
		0.8	-2.01	2.05	2.47	1.85	5.73	3.29
IG (10,2)	ACG	0.1	4.84	7.74	10.86	8.52	113.27	8.60
		0.2	4.95	8.74	11.00	9.31	13.76	9.68
		0.5	3.58	4.44	8.28	4.42	7.20	4.83
		0.8	5.35	9.52	14.59	11.23	15.81	10.38
	ACt	0.1	7.65	9.92	15.27	11.13	153.35	10.86
		0.2	6.78	10.94	15.41	11.79	18.03	12.04
		0.5	6.05	6.02	13.26	6.58	10.29	6.72
		0.8	6.16	10.34	16.16	12.10	17.02	11.09
	MCG	0.1	2.79	6.62	9.54	7.87	104.18	8.13
		0.2	2.52	7.38	9.88	8.62	13.12	9.01
		0.5	2.62	3.93	7.84	4.29	6.97	4.63
		0.8	4.06	8.58	13.48	10.59	15.21	9.80
	MCt	0.1	6.66	9.20	13.62	10.42	144.01	10.31
		0.2	5.57	9.94	14.12	10.96	17.24	11.36
		0.5	5.04	5.51	12.14	6.14	9.41	6.26
		0.8	5.08	9.55	15.34	11.43	16.14	10.62
	NCG	0.1	-2.63	3.00	1.34	1.67	40.06	4.23
		0.2	-4.02	3.27	1.50	1.65	6.03	4.37
		0.5	-1.22	1.94	1.57	1.09	2.65	2.37
		0.8	-3.38	2.90	2.32	1.95	5.92	4.14
	NCt	0.1	-1.48	1.93	1.49	1.57	22.75	3.06
		0.2	-1.96	2.07	1.53	1.59	4.87	3.14
		0.5	-0.25	1.39	1.56	1.06	2.73	1.93
		0.8	-1.75	1.91	2.57	1.89	5.78	3.30

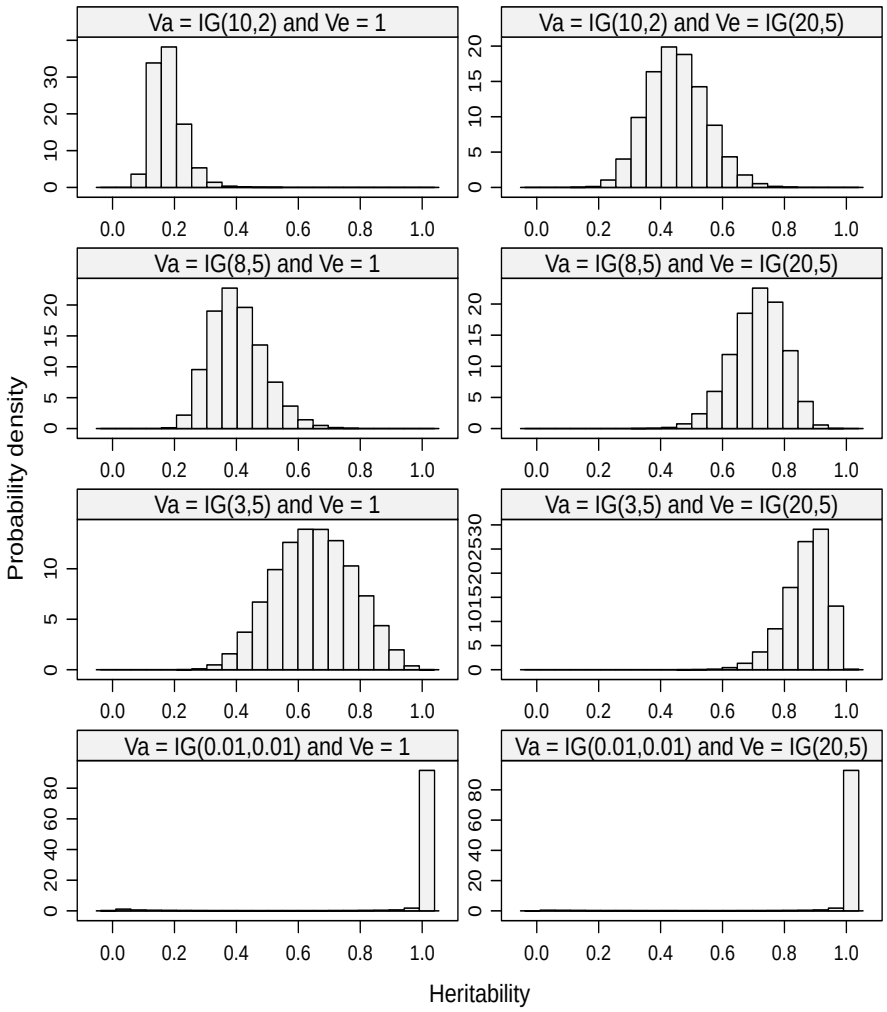

Figure 2 - Histogram with samples from the four specifications of prior distributions for the intraclass correlation (heritablity) used in the square lattice analysis. "Va" stands for the variance of random effects and "Ve" for the experimental error (in the latent trait scale)

that obtained with the NCG and NCt algorithms. However it is likely that in this case the bias is low because a "correct" and informative prior was used. One would expect changes if the prior were not on variance components or if dispersion of latent true trait were too different from the fixed value $\sigma_e^2 = 1$.

Strandén and Gianola (1998) suggested that Student's t distribution might be more flexible than the Gaussian one and that the former might be more robust when the response variable is asymmetric distributed. However, this was not observed in our study. Posterior inferences using Student's t distribution were very similar to Gaussian, not affected either by assimetric or very heavy tailed distribution for latent trait. In other words, degrees of freedom for Student's t were always high (results not shown). Our findings in some way agree with Kizilkaya et al. (2003) that highlighted the need of more studies that use Student's t distribution in threshold models to assess its properties. Our conclusion is that allocation of threshold parameters is enough to get a robust analysis using Gaussian models both for assimetric and heavy tailed latent trait distributions.

Despite the smaller MSE observed in the NCG and NCt for the fixed and random effects, these algorithms were more sensitive to the prior used for the variance components. Using less informative prior $(IG(3,\ 5))$ was more accurate only for intraclass correlation equal to 0.1 and 0.2, with a tendency to overestimate the ρ values when it was greater than or equal to 0.5 (Figure 3). In this case, the informative prior resulted in a smaller MSE for ρ. This is due to a smaller bias and was expected because this inference was done using a "correct" and informative prior. The other algorithms showed a smaller MSE for ρ when using the informative prior, but the MSE values of these algorithms (ACG, ACt, MCG and MCt) were larger than the ones of NCG and NCt when $\rho = 0.8$. The average correlation between predicted and observed values was approximately 0.99 for $\rho = 0.8$ and varied from 0.75 to 0.90 for $\rho < 0.8$.

Among the algorithms, NCG and NCt distinguished themselves, being fast and flexible for the experimental situations simulated. This reparameterization resulted in a smaller MSE for the estimates of fixed and random effects. On the other hand, they were more sensitive to the priors used for the variance components, when they reflect informative priors to ρ. Nevertheless, they are also more flexible from the viewpoint of prior specification, making possible to get an effectively uniform prior for intraclass correlation, instead of the other algorithms.

Using the algorithms NCG ans NCt also yielded better behaved MCMC chains, with smaller dispersion and early reaching of stationary distribution (Figure 3). This not only reduced the MSE of fixed and random effects but also accelerated the chain convergence process. Nandram and Chen (1996), when comparing their algorithm to AC and MC, observed an improvement in the convergence process. In the case of mixed models, a reduction in the number of iterations necessary for convergence was also observed. Moreover, we have shown that there were no differences in the convergence process between the Gaussian and Student's t distributions (Table 4). This means that a routine test can be done to see if Student's t distribution should be used, just checking posterior inferences by both models or degrees of freedom in NCt.

For the variance components, the informative prior accelerates the convergence process of the algorithms (Table 5) but does not improve their overall performance. In other words, the NCG and NCt algorithms showed a greater convergence speed than the others, regardless of the prior used. It is also noticeable that a design having large number of random effects to be estimated, such as SLD, and a design with very few random effects, such as STS, have almost the same effect on the dependency of MCMC chains. The

Table 2 - Mean squared error (MSE) and bias of the posterior mean of the random effects in the STS design.

Prior	Algorithm	Intraclass correlation	Asymmetric		Symmetric		Uniform	
			Bias	MSE	Bias	MSE	Bias	MSE
IG (3,5)	ACG	0.1	0.06	1.38	0.13	1.42	0.13	1.32
		0.2	0.08	1.42	0.13	1.51	0.10	1.39
		0.5	0.56	2.79	1.61	3.57	1.10	3.72
		0.8	3.46	9.95	3.44	7.88	3.39	8.24
	ACt	0.1	0.09	1.70	0.14	1.67	0.14	1.62
		0.2	0.11	1.77	0.17	1.75	0.13	1.65
		0.5	0.57	3.49	1.54	3.89	1.12	3.57
		0.8	3.39	9.77	3.17	7.66	3.40	8.17
	MCG	0.1	0.06	1.31	0.08	1.33	0.08	1.25
		0.2	0.07	1.36	0.12	1.45	0.07	1.33
		0.5	0.36	2.38	0.88	2.92	0.62	2.56
		0.8	2.40	8.66	2.89	7.21	2.61	7.38
	MCt	0.1	0.07	1.56	0.11	1.57	0.11	1.54
		0.2	0.07	1.59	0.14	1.65	0.11	1.56
		0.5	0.44	2.97	0.86	3.20	0.77	3.05
		0.8	2.52	8.57	2.61	6.93	2.50	7.20
	NCG	0.1	0.008	0.32	0.004	0.26	0.004	0.32
		0.2	0.01	0.32	0.006	0.28	0.008	0.32
		0.5	0.04	0.71	0.03	0.75	0.02	0.71
		0.8	-0.11	0.46	0.04	0.49	-0.12	0.46
	NCt	0.1	0.008	0.31	0.003	0.28	0.003	0.31
		0.2	0.01	0.32	0.005	0.30	0.005	0.33
		0.5	0.03	0.72	0.03	0.75	0.02	0.72
		0.8	-0.10	0.49	0.04	0.49	-0.13	0.49
IG (10,2)	ACG	0.1	0.01	0.64	0.01	0.63	0.01	0.63
		0.2	0.02	0.63	0.01	0.64	0.01	0.63
		0.5	0.04	0.71	0.03	0.74	0.03	0.74
		0.8	-0.10	0.84	0.04	0.78	-0.11	0.81
	ACt	0.1	0.01	0.56	0.01	0.55	0.01	0.56
		0.2	0.02	0.55	0.01	0.55	0.01	0.55
		0.5	0.04	0.69	0.04	0.72	0.03	0.71
		0.8	-0.09	0.70	0.05	0.64	-0.11	0.67
	MCG	0.1	0.01	0.62	0.01	0.62	0.01	0.62
		0.2	0.01	0.61	0.01	0.62	0.01	0.62
		0.5	0.04	0.70	0.03	0.73	0.02	0.73
		0.8	-0.10	0.78	0.05	0.74	-0.11	0.78
	MCt	0.1	0.01	0.55	0.01	0.54	0.01	0.55
		0.2	0.01	0.54	0.01	0.54	0.01	0.54
		0.5	0.04	0.68	0.04	0.71	0.03	0.70
		0.8	-0.10	0.64	0.05	0.60	-0.11	0.63
	NCG	0.1	0.007	0.29	0.005	0.23	0.005	0.28
		0.2	0.01	0.29	0.004	0.26	0.006	0.29
		0.5	0.04	0.72	0.03	0.77	0.02	0.73
		0.8	-0.10	0.48	0.03	0.51	-0.12	0.48
	NCt	0.1	0.006	0.26	0.001	0.24	0.001	0.26
		0.2	0.01	0.28	0.004	0.27	0.003	0.28
		0.5	0.04	0.74	0.03	0.77	0.001	0.74
		0.8	-0.10	0.51	0.03	0.51	-0.12	0.51

Table 3 - MSE and bias for the intraclass correlation in the SLD.

Prior	Algorithm	Intraclass correlation	Asymmetric	
			Bias	MSE
IG (3,5)	ACG	0.1	0.30	0.090
		0.2	0.30	0.040
		0.5	0.10	0.010
		0.8	0.04	0.030
	ACt	0.1	0.41	0.170
		0.2	0.41	0.110
		0.5	0.27	0.080
		0.8	0.05	0.030
	MCG	0.1	0.29	0.080
		0.2	0.29	0.040
		0.5	0.08	0.010
		0.8	0.03	0.030
	MCt	0.1	0.39	0.150
		0.2	0.38	0.090
		0.5	0.22	0.060
		0.8	0.03	0.030
	NCG	0.1	0.29	0.080
		0.2	0.29	0.030
		0.5	-0.08	0.006
		0.8	-0.10	0.133
	NCt	0.1	0.30	0.090
		0.2	0.30	0.040
		0.5	-0.07	0.005
		0.8	-0.10	0.132
IG (8,5)	ACG	0.1	0.07	0.006
		0.2	0.08	0.004
		0.5	-0.12	0.020
		0.8	0.04	0.020
	ACt	0.1	0.09	0.008
		0.2	0.09	0.005
		0.5	-0.06	0.026
		0.8	0.04	0.032
	MCG	0.1	0.07	0.006
		0.2	0.07	0.010
		0.5	-0.13	0.030
		0.8	0.02	0.025
	MCt	0.1	0.08	0.007
		0.2	0.08	0.007
		0.5	-0.10	0.030
		0.8	0.03	0.030
	NCG	0.1	0.10	0.010
		0.2	0.10	0.002
		0.5	-0.27	0.080
		0.8	-0.11	0.169
	NCt	0.1	0.10	0.011
		0.2	0.10	0.004
		0.5	-0.27	0.080
		0.8	-0.11	0.168

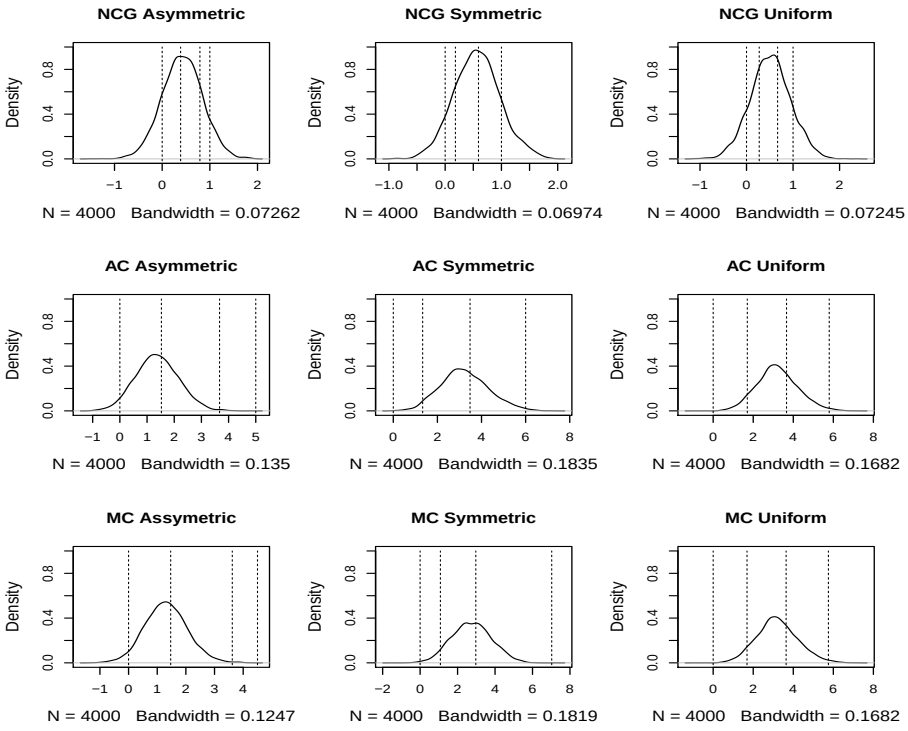

Figure 3 - Posterior density for an example simulated in the STS design with $\rho = 0.2$, burn=0, and jump=1 and a final chain sample size equal to 4000 with symmetric, asymmetric and uniform distributions of the random variable. Dotted lines show the posterior means of the threshold positions.

performance of NCG and NCt algorithms was superior to that of the others, regardless the number of random effects to be estimated.

Kizilkaya et al. (2003) suggest that the rejection sampling for threshold parameter reduced the independence between samples, but not necessarily leads to a faster convergence of chains. Our results from the simulated study indicates at least that NCG and NCt converges more often for a planned chain size (Table 6).

One advantage of NCG and NCt algorithms is the generation of candidates for the threshold parameters in the $[0, 1]$ range. This is because Dirichlet distribution is conjugated to the multinomial. The acceleration of convergence process achieved by this procedure can be better seen in Figure (4), which shows the trajectory of MCMC chains for the second threshold (STS design).

Figure (5) depicts the autocorrelation in MCMC chains for the second threshold by using different algorithms. This parameter generally have the highest autocorrelation during the sampling process. Clearly, the algorithm adapted from Nandram and Chen

Table 4 - Average processing time, burn-in, jump, and total number of iterations for the six algorithms assessed for the STS and SSL designs, according to the Raftery and Lewis diagnostic.

Algorithm	Time(min.)	Burnin	Jump	Total number of iteractions
		STS		
ACG	3.57	78.36	17.59	67032.08
ACt	4.12	89.23	22.21	83281.63
MCG	1.97	61.87	15.38	57742.86
MCt	4.33	75.05	18.72	70173.76
NCG	0.91	23.79	6.65	25255.95
NCt	1.06	23.48	6.84	25959.31
		LQS		
ACG	23.74	60.74	15.17	56869.24
ACt	41.40	94.38	20.34	76290.67
MCG	11.39	27.00	8.09	30317.97
MCt	39.20	47.25	13.76	51562.65
NCG	9.74	14.74	4.57	17150.79
NCt	16.20	16.43	5.15	19226.35

Table 5 - Average processing time, burn-in, jump and total number of iterations for the two priors used to estimate the variance components for the STS and SL designs, according to the Raftery and Lewis diagnostic.

Prior	Time (min.)	Burnin	Jump	Total number of iteractions
		STS		
More informative	2.35	52.58	13.48	50485.61
Less informative	2.97	64.68	15.65	59309.06
		LQS		
More informative	19.08	36.49	10.57	39591.30
Less informative	28.08	50.36	11.79	44214.49

Table 6 - Convergence rate of the samples for the STS and SLD designs for the six algorithms, according to the Gelman and Rubin test.

Algorithm	Design	
	STS	SLD
ACG	0.77	0.81
ACt	0.74	0.75
MCG	0.61	0.65
MCt	0.50	0.42
NCG	0.89	0.96
NCt	0.72	0.96

(1996) needs fewer iterations to obtain a stationary and independent chain.

Simulations has shown that the NCG and NCt algorithms are more consistent and fast, but the others can also be used for inference on fixed and random effects, although for variance components some care is needed. NCG and NCt are also better on allowing specification of uniform priors for intraclass correlation, being a much more flexible choice.

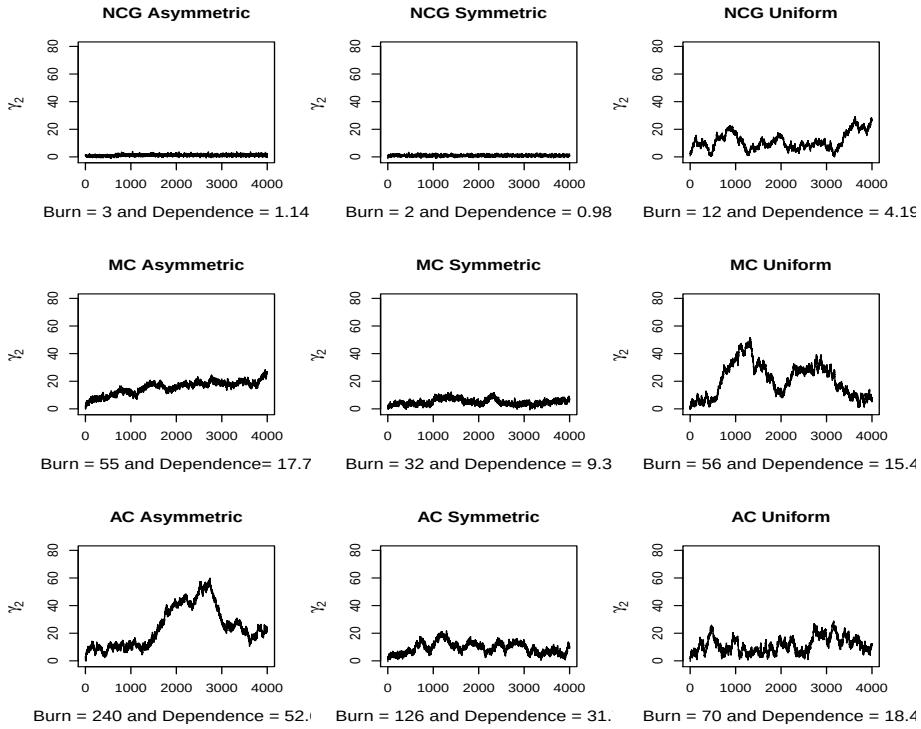

Figure 4 - Trajectory of the chain for a simulated example in the STS design with $\rho = 0.2$, burn=0, jump=0, and a final chain size of 4000 for the threshold 2 with symmetric, asymmetric, and uniform distributions of the response variable. γ_2 stands for the second threshold parameter, sampled in all models, rescaled acordingly.

4 An example applied to plant breeding for disease resistance

This was a plant breeding experiment on resistence to late blight (caused by *Phytophthora infestans*) in tomato. Treatments were 66 full sib families of tomato (different genotypes). An incomplete block design (IBD) with 20 blocks having 33 plants grouped (resolved) in 10 repetitions per family. Disease severity was assessed weekly for 54 days

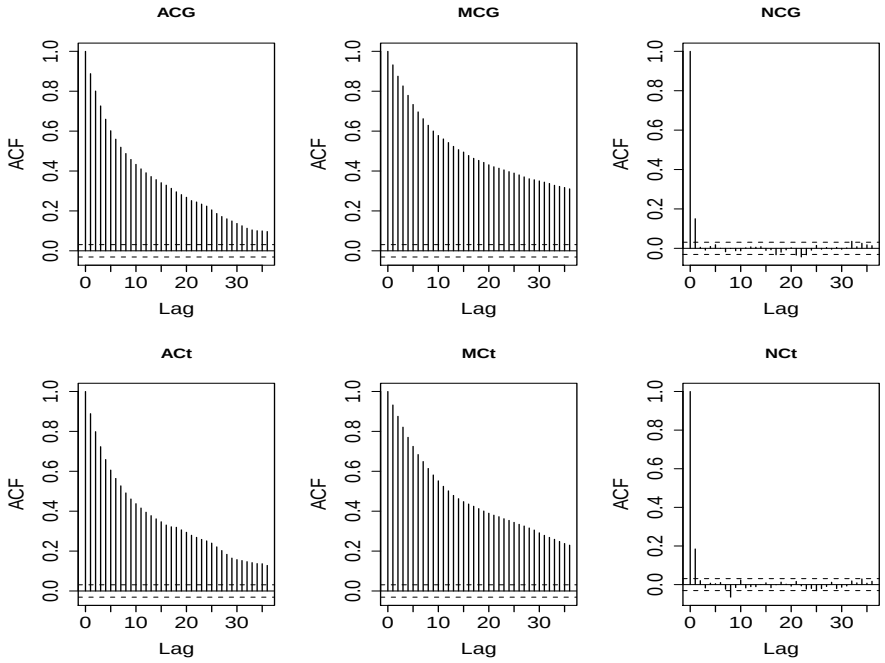

Figure 5 - Auto-correlation (ACF) plot of the chain for a simulated example in the STS design with $\rho = 0.2$, burn=0, jump=0, and a final chain size of 4000 for the second threshold with symmetric distribution of the response variable.

using a scale whith six categories, roughly representing 1%, 5%, 10%, 16%, 32%, and 50% disease severity Corrêa, Bueno-Filho and Carmo (2009). The model adopted for the analysis was a split-plot in time, where the plots were a single tomato plant randomized to the IBD and subplots were their repeated evaluations in time.

The initial analysis used all of the six algorithms described before to provide initial MCMC chains of 4000 iterations. After Raftery and Lewis diagnostic, estimated sample sizes were used to plan final MCMC. Two chains were sampled to evaluate the convergence by Gelman and Rubin's diagnostic.

Bayes factors (JEFFREYS, 1961) were applied to compare models with cumulative Normal and Student's t distributions to assess the evidence in favor of the former. According to Gelman et al. (2003), the Bayes factor was obtained from the mean estimate of the log-likelihood (from posterior samples). The experiment was analyzed using a personal computer with a Core(TM)i7-2600 3.4GHz processor and 16 Gb of RAM. Table(7) shows the estimated parameters and the controling MCMC sizes used for the sampling process, such as the burn-in (Burn), jump (Jump), and the effective number of iterations (Iter).

Table 7 - Parameters of the iteration process after the application of the Raftery and Lewis diagnostic to obtain the ideal sample size, and the model parameters from the ideal sample.

	ACG	ACt	MCG	MCt	NCG	NCt
LogVeros.	-4813.46	-5072.754	-4302.46.16	-4701.44	-7095.68	-6801.88
σ^2_{gen}	0.137	0.127	0.198	0.141	0.059	0.059
σ^2_{bl}	0.028	0.052	0.031	0.029	0.036	0.026
σ^2_{res}	1.00	1.00	1.00	1.00	0.16	0.11
ρ_{gen}	0.12	0.11	0.16	0.12	0.26	0.34
Burnin	688	698	333	31	6	194
Jump	37	45	96	1	9	66
Time (min.)	2.17	192.68	1.68	41.31	2.60	194.76

We observed that the algorithms proposed by Kizilkaya et al. (2003) had the smallest processing time and that the ACt algorithm took more time, but there were great differences among the algorithms with regard to burn-in, jump, and the effective number of iterations. Although the simulation process indicated a smaller processing time for the NCG and NCt algorithms, only the ACt algorithm had a noticeably greater time than the others in present study (about 93-fold). The NCG and NCt algorithms gave smaller genotype variance estimates but also greater estimated intraclass correlation for the genotypes (Table 7). As stated before this is the comparison in which we expect algorithms to differ, with a greater prior influence on intraclass correlation. This suggests prior sensitiveness and the worthy of another analysis using near-uniform priors to intraclass correlation.

Estimates for the genotype effects and time trajectories were very similar among all algorithms, even though the prior used for the variance components was the same as used in the simulations. Bayes factors were approximately equal to 1 in the comparison of models with a cumulative Normal distribution and those with a Student's t, indicating that both distributions have a similar performance.

There were no differences in the rankings of genotypes with regard to disease resistance, indicating that the algorithms had similar predictive power. Figure (6) shows the predictions and highest posterior difference (HPD) intervals for the most resistant genotype (62), a moderately resistant genotype (10) and a susceptible genotype (2).

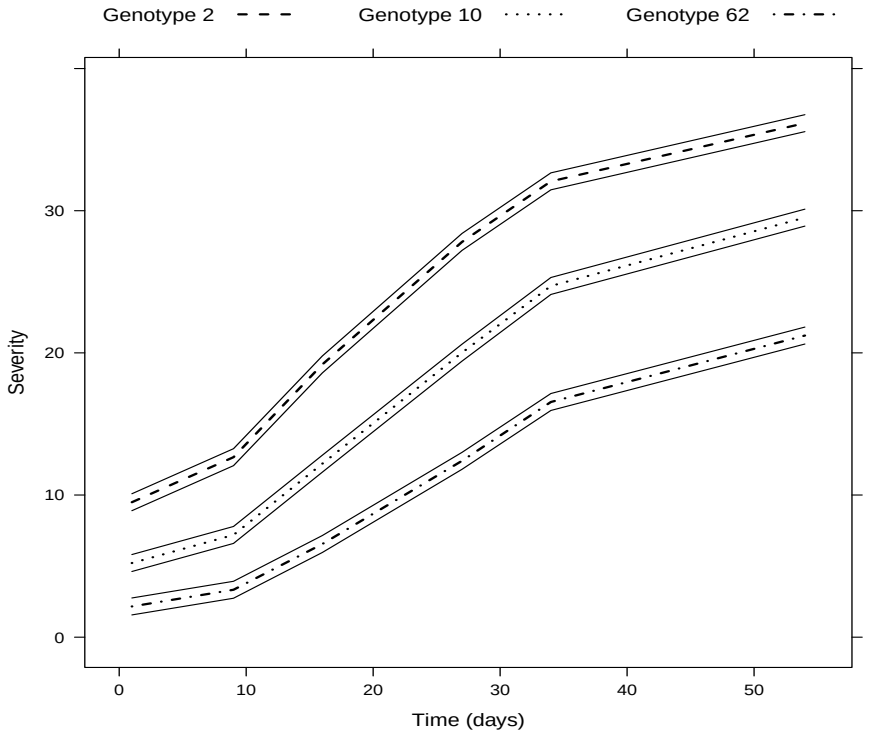

Figure 6 - Predicted values for the three genotypes with different degrees of disease resistance. The solid line indicates the HPD interval with regard to the predicted severity value.

Conclusions

The proposed extension of Nandram and Chen (1996) algorithm for the analysis of mixed models is a faster alternative than other algorithms as shown in experimental situations simulated. Using Student's t cumulative link instead of probit link did not change properties of the analysis, indicating that Gaussian distribution is a robust choice and can be used regardless of the shape of the latent trait, even in experiments with small numbers of samples. Algorithms implemented in the Bayesthresh package can be used as a fast and robust tool for the analysis of ordinal categorical data with mixed models (and threshold approach) using Bayesian inference.

Acknowledgments

Authors would like to thank Brazilian agencies CNPq and FAPEMIG for funding this work. The Universidade Estadual de Santa Cruz with the help of manuscript translation.

CORRÊA, F. M.; SILVA, J. W.; FERREIRA, D. F.; BUENO FILHO, J. S. S. Algoritmos bayesianos para análise de dados categóricos ordinais. *Rev. Bras. Biom.,* Lavras, v.34, n.4, p.597-620, 2016.

- RESUMO: Este estudo descreve e avalia um pacote que implementa as extensões dos algoritmos descritos por Nandram e Chen (1996) utilizando a distribuição Gaussiana (NCG) e a distribuição de Student (NCT) para análise bayesiana de dados categorizados ordinais. Os algoritmos descritos por Albert e Chib (1993) e Cowles (1996) também foram implementados. As comparações foram realizadas utilizando dois modelos diferentes. Um modelo foi o sistema triplo de Steiner com sete tratamentos utilizado para avaliar os efeitos fixos e um delineamento em láttice quadrado 10x10 foi usado para avaliar os efeitos aleatórios. Também foram consideradas diferentes situações para as correlações intraclasse. Nós avaliamos o número total de iterações para a convergência, o erro quadrático médio (EQM) e o viés para as estimativas a posteriori dos efeitos fixos, aleatórios e das correlações intraclasse. Os algoritmos NCG e NCT resultaram em menor EQM e viés para os efeitos avaliados. Também apresentaram maior velocidade para atingir a convergência dos parâmetros. Estes algoritmos apresentaram estimativas para o a correlação intraclasse superestimada quando o valor paramétrico era de 0.8. O viés e o EQM nas demais simulações não foi alterado. Um experimento real de melhoramento genético, foi analisado utilizando os mesmos algoritmos, em delineamento de blocos incompletos para selecionar genótipos de tomate resistêntes a *Phytophthora infestans*. A distribuição Gaussiana foi considerada parcimoniosa para a escolha do traçølatente. Os algoritmos implementados são considerados consistentes quanto a classificação dos genótipos.

- PALAVRAS-CHAVE: MCMC; bayesthresh; modelos de limiar

References

AGRESTI, A. *Categorical data analysis*. [S.l.]: Willey, Hoboken, NJ, 2013. 714 p.

ALBERT, J. H.; CHIB, S. Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*, v. 88, n. 442, p. 669–679, 1993.

CORREA, F. M.; BUENO-FILHO, J. S. S. *Bayesthresh: Bayesian thresholds mixed-effects models for categorical data*. Rod. Ilheus-Itabuna, km 16, Bahia - Brazil, 2012. Disponível em: ⟨http://cran.r-project.org/web/packages/Bayesthresh/index.html⟩.

CORRÊA, F. M.; BUENO-FILHO, J. S. S.; CARMO, M. G. F. Comparison of the three diagrammatic key for the quantification of late blight in tomato leaves. *Plant Pathology*, v. 58, n. 6, p. 1128–1133, 2009.

COWLES, M. K. Accelerating Monte Carlo Markov chain convergence for cumulative link generalized linear models. *Statistics and Computing*, v. 6, n. 2, p. 101–111, 1996.

GELMAN, A. et al. *Bayesian data analysis*. [S.l.]: London: Chapman and Hall, 2003. 668 p.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. *Statistical Science*, v. 7, n. 4, p. 457–511, 1992.

GOB, R.; MCCOLLIN, C.; RAMALHOTO, M. F. Ordinal methodology in the analysis of likert scales. *Quality and Quantity*, v. 41, n. 5, p. 601–626, 2007.

HENDERSON, C. R. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Applied Statistics*, v. 32, n. 1, p. 69–83, 1976.

JEFFREYS, H. *Theory of probability*. [S.l.]: Oxford: Claredon Press, 1961. 470 p.

KIZILKAYA, K. et al. Cumulative t-link threshold models for genetic analysis of calving ease scores. *Statistics and Computing*, v. 35, n. 1, p. 489–512, 2003.

MCCULLAGH, P. Regression models for ordinal data. *J. R. Stat. Soc. - B*, v. 42, n. 1, p. 109–142, 1980.

MCCULLOGH, C. E.; SEARLE, S. R. *Generalized, Linear and Mixed Models*. [S.l.]: Wiley, New York, 2001.

NANDRAM, B.; CHEN, M.-H. Reparameterizing the generalized linear model to accelerate gibbs sample convergence. *Journal of Statistical Computation and Simulation*, v. 54, n. 1, p. 129–144, 1996.

PIEPHO, H.-P.; KALKA, E. Threshold models with fixed and random effects for ordered categorical data. *Food Quality and Preference*, v. 14, n. 1, p. 343–357, 2003.

PLUMMER, M. et al. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, v. 6, n. 1, p. 7–11, 2006. Disponível em: ⟨http://CRAN.R-project.org/doc/Rnews/⟩.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2012. ISBN 3-900051-07-0. Disponível em: ⟨http://www.R-project.org/⟩.

RAFTERY, A. E.; LEWIS, S. M. One long run with diagnostics: Implementation strategies for Markov Chain Monte Carlo. *Statistical Science*, v. 7, n. 4, p. 493–497, 1992.

SILVA, J. W.; BUENO-FILHO, J. S. S. Um algoritmo para modelos de limiar usando as distribuições acumuladas normal e "t" de Student. *Revista Brasileira de Matemática e Estatística*, v. 28, n. 3, p. 59–83, 2010.

SORENSEN, D.; GIANOLA, D. *Likelihood, Bayesian and MCMC methods in quantitative genetics*. [S.l.]: Springer-Verlag New York, 2002. 740 p.

SORENSEN, D. A. et al. Bayesian inference in threshold models using Gibbs sampling. *Genetic Selection Evoution*, v. 27, n. 1, p. 229–249, 1995.

STRANDÉN, I.; GIANOLA, D. Attenuating effects of preferential treatment with Student-t mixed linear models: a simulation study. *Genetics selection evolution*, v. 30, n. 1, p. 25–42, 1998.

TABOR, D. Mohs's hardness scale - A physical interpretation. *Proceedings of the Physical Society. Section B*, v. 67, n. 3, p. 249–257, 1954.

WANG, C.; DA, Y. Quantitative genetics model as the unifying model for defining genomic relationship and inbreeding coefficient. *PloS One*, v. 9, n. 12, p. e114484, 2014. ISSN 1932-6203.

Received in 04.12.2015.

Approved after revised in 18.05.2016.