

## Assessing Convergence of the Markov Chain Monte Carlo Method in Multivariate Case

<sup>1</sup>Denismar Alves Nogueira, <sup>2</sup>Thelma Safadi,  
<sup>2</sup>Daniel Furtado Ferreira and <sup>1</sup>Eric Batista Ferreira

<sup>1</sup>Exact Sciences Institute, Federal University of Alfenas,  
Rua Gabriel Monteiro da Silva, 700, 37130-000, Alfenas, Minas Gerais, Brazil  
<sup>2</sup>Department Exact Sciences, Federal University of Lavras, Lavras, Minas Gerais, Brazil

Received 2012-08-08, Revised 2012-09-05; Accepted 2012-09-05

### ABSTRACT

The formal convergence diagnosis of the Markov Chain Monte Carlo (MCMC) is made using univariate and multivariate criteria. In 1998, a multivariate extension of the univariate criterion of multiple sequences was proposed. However, due to some problems of that multivariate criterion, an alternative form of calculation was proposed in addition to the two new alternatives for multivariate convergence criteria. In this study, two models were used, one related to time series with two interventions and ARMA (2, 2) error and another related to a trivariate normal distribution, considering three different cases for the covariance matrix. In both the cases, the Gibbs sampler and the proposed criteria to monitor the convergence were used. Results revealed the proposed criteria to be adequate, besides being easy to implement.

**Keywords:** Convergence Criterion, Gibbs Sampler, Bayesian Inference, Simulation

### 1. INTRODUCTION

Nowadays, Bayesian inference is a matter of extreme interest, despite having been developed long before frequentist statistics. In some cases, Bayesian inference requires Markov Chain Monte Carlo (MCMC) methods. The Gibbs sampler is one of the major classes of stochastic simulation schemes proposed, which is being used in many situations (Gamerman and Lopes, 2006). The quality of the simulation methods re-lies on good-quality uniform random number generators, an issue recently discussed by Luiz *et al.* (2010). However, a great difficulty is the empirical diagnosis of convergence to the stationary distribution. Several techniques in the literature help in identifying and monitoring convergence (Heidelberger and Welch, 1986; Geweke, 1992; Raftery and Lewis, 1992; Cowles and Carlin, 1996; Brooks and Roberts, 1998; Brooks and Giudici, 2000).

Besides Bayesian analysis being increasingly used, the results are often questioned because researchers do

not use or do not clearly address the implemented criteria to check for convergence. Moreover, in cases of complicated models, Bayesian inference requires a great computational effort. This effort can be minimized by monitoring the chain convergence, thus avoiding iterations beyond the necessity.

In the literature, there are univariate and multivariate criteria for monitoring the convergence of MCMC output to the stationary distribution, where the Gelman and Rubin (1992) criterion is a univariate representative. This criterion uses parallel chains from different starting points, i.e., different arbitrary initial values and the idea that the trajectories of the chains should be the same after convergence has been formalized. This criterion is based on the use of analysis of variance techniques, seeking to verify whether the dispersion within the chains is greater than that between the chains. By analogy, this process has been extended to the multivariate form by Brooks and Gelman (1998).

When dealing with many parameters, the convergence of the distribution will only occur when all

**Corresponding Author:** Denismar Alves Nogueira, Exact Sciences Institute, Federal University of Alfenas,  
Rua Gabriel Monteiro da Silva, 700, 37130-000, Alfenas, Minas Gerais, Brazil

the parameters converge. This is a practical problem because it turns out to be impractical for a large number of parameters. The multivariate criterion is based on a single value for assessing the convergence of the MCMC output for all the parameters.

The possible issues of the multivariate criteria and the cases where they are impossible to compute have been pointed out in Brooks and Gelman (1998). Therefore, this study presents an alternative way for obtaining the Brooks and Gelman criterion, as well as two new alternatives of multivariate convergence criteria and evaluates the performance of the three methods for convergence under two models.

The study is outlined as follows: Section 2 presents the original convergence criterion (Brooks and Gelman, 1998), along with the research problem and motivation (2.1), an alternative computation (2.2) and two proposed criteria (2.3). In Section 3 the methodology and application of the criteria to the two models are presented and in Section 4, they have been compared. Lastly, Section 5 presents the conclusions.

### 1.1. Convergence Criterion of Brooks and Gelman

The original criterion of Brooks and Gelman (1998) is an extension of the criterion of Gelman and Rubin (1992). According to Gelman and Rubin (1992), in many cases, the convergence of chains to the stationary distribution can be easily determined using multiple independent chains, in parallel, but cannot be diagnosed using the simulation result coming from any single chain. They proposed a method using multiple replications of chains to determine if the stationary state was reached in the second half of each sample (chain). The method assumes that  $m$  chains have been simulated in parallel, each from a different starting point. After a starting point belonging to the parameter space of the posterior distribution has been obtained, the chains are run for  $2n$  iterations, of which the first  $n$  are discarded to avoid the period of heating (burn-in) and the influence of initial values. The  $m$  chains yield  $m$  possible statistics. If those statistics are quite similar, it is an indication that convergence has been reached or is close. These authors have also suggested comparing these statistics with those obtained from the union of the chains, i.e., union of the  $nm$  values. The convergence indicator Gelman and Rubin (1992) is named Potential Scale Reduction Factor (PSRF).

For the multivariate case, Brooks and Gelman (1998) proposed to replace the scalar estimators by  $p \times p$  covariance matrices  $B$  and  $W$  (between and within chains, respectively) of the vector of parameter  $\theta$ , whose

elements are  $\theta_{ji}$ , where  $\theta_{ji}^{(p)}$  is the  $p$ -th element of the vector of the parameters of the chain  $j$  at iteration  $i$ .

For large dimension, one should estimate the covariance matrix of the a posteriori chains of the parameters by:

$$\bar{V} = \frac{n-1}{n}W + \left(1 + \frac{1}{m}\right)\frac{B}{n}$$

Where:

$$W = \frac{1}{m(n-1)} \sum_{j=1}^m \sum_{i=1}^n (\theta_{ji} - \bar{\theta}_j)(\theta_{ji} - \bar{\theta}_j)^T$$

And:

$$\frac{B}{n} = \frac{1}{m-1} \sum_{j=1}^m (\bar{\theta}_j - \bar{\theta}_{..})(\bar{\theta}_j - \bar{\theta}_{..})^T$$

are  $p$ -dimensional matrices estimated from chains of the  $p$  parameters. Thus, researchers could monitor the convergence by using the covariance matrices  $\bar{V}$  and  $W$ .

The distance between  $\bar{V}$  and  $W$  is summarized as a scalar measure that should be close to 1 when the convergence is achieved. One way to do this is by seeking the maximum of the characteristic root  $\lambda$  of the product  $W^{-1}V$ , which is also the maximum of the PSRF of any linear projection of  $\theta$ .

The maximum is given by differentiating the ratio of quadratic forms with respect to the vector  $a$ , by setting it to zero, i.e.,  $\frac{\partial \lambda}{\partial a} = 0$  and adopting the restriction given by  $a^T W a = 1$ .

Then:

$$\begin{aligned} \frac{\partial \lambda}{\partial a} &= \frac{2V\alpha\alpha^T W\alpha - 2W\alpha\alpha^T V\alpha}{(\alpha^T W\alpha)^2} \\ &= \frac{2}{\alpha^T W\alpha} [V\alpha - \lambda W\alpha] = (V - \lambda W)\alpha = 0 \end{aligned}$$

The homogeneous system of equations has a nontrivial solution if and only if,  $|V - \lambda W| = 0$ .

The solution can be obtained by taking the eigenvalues of  $W^{-1}V$ , but in some cases, it is not straightforward to obtain the inverse of  $W$  and the eigenvalues of the matrix  $W^{-1}V$ , because this new matrix is non-symmetric.

The convergence indicator Brooks and Gelman (1998)-named Multivariate Potential Scale Reduction Factor (MPSRF)-is based on following quadratic forms' maximization Equation 1:

$$\begin{aligned} \bar{R}^p &= \max_{\alpha} \frac{\alpha^T \bar{V} \alpha}{\alpha^T W \alpha} \\ &= \frac{n-1}{n} + \left( \frac{m+1}{mn} \right) \max_{\alpha} \frac{(\alpha^T B \alpha) / n}{\alpha^T W \alpha} \quad (1) \\ &= \frac{n-1}{n} + \left( \frac{m+1}{mn} \right) \lambda_1 \end{aligned}$$

Under equality of the average of the chains,  $\lambda_1 \rightarrow 0$ . Is it the maximum eigenvalue of  $W^{-1}B$ .  $\bar{R}^p \rightarrow 1$  when  $n$  is large enough, where  $p$  is the number of parameters.

According to Brooks and Gelman (1998) if both  $W$  and  $B$  are both singular we cannot calculate the MPSRF. This can occur if two or more parameters are correlated or one parameter is obtained by linear combinations of the other. If only  $W$  is singular, then this problem can be circumvented as the iterations go on. Another problem that can arise is the time elapsed to perform the inversion of  $W$ , because it could be large in many circumstances. The methods proposed in this study come to solve such issues.

### 1.2. Alternative Computation for Maximizing Quadratic Forms

The maximization of quadratic forms is widely used for circumstances in which we want to get a value that represents a direction of the greater variability of the system. In maximizing quadratic forms, a homogeneous system of equations given by  $(V-\lambda_i I) a_i = 0$  occurs. Maximizing quadratic forms ratios yields a homogeneous system given by  $(V-\lambda_i W) a_i = 0$  and obtaining the characteristic roots and characteristic vectors of the second case is not a trivial task. Therefore, this study proposes rotating the axis on the Cartesian system, seeking for new directions of greater variability, thereby reducing the system of the ratio of two quadratic forms to a system of one quadratic form.

Let the ratio of quadratic forms of the Brooks and Gelman criterion (1) be the one to be maximized. In the Bayesian literature, it is common to find the maximization given by obtaining the eigenvalues and eigenvectors of  $W^{-1}\bar{V}$ . This study proposes an alternative that is described hereafter. For this, the matrix  $W$  is factored (Cholesky factor) as  $W = SS^T$ . Setting  $z$  as the linear transformation of the vector  $a$  by  $z = S^T a$  gives  $a = (S^{-1})^T z$ , because, in accordance with the properties of the Cholesky factor,  $SS^{-1} = S^T (S^{-1})^T = I$ . If the equation  $(\bar{V} - \lambda_i W)a_i = 0$  is premultiplied by  $S^{-1}$ , then  $S^{-1}W (S^{-1})^T = I$ . Setting  $H = S^{-1} \bar{V} (S^{-1})^T$ , gives  $(H$

$-\lambda_i I)z_i = 0$ ; therefore, we achieve the same solution for the case of maximizing a quadratic form, except that  $a = (S^{-1})^T z$  must be recovered, because the eigenvectors  $z$  are changed by non-singular transformation. The eigenvalues are invariant to the non-singular transformation performed.

According to Brooks and Gelman (1998), the maximum eigenvalue  $\lambda_1$  is the  $R^p$  itself, where  $R^p$  is the  $p$ -dimensional PSRF, given by (1). Thus, the transformation gives the system:

$$(H - R^p I)z_i = 0$$

From this equation, we can determine  $R^p$  and avoid the problems mentioned by Brooks and Gelman (1998), because there is no need to invert  $W$  to obtain the solution of maximizing the ratio of two quadratic forms. The prerequisite is that  $W$  must be positive definite for the condition of existence of the Cholesky factor to be satisfied.

Although the Cholesky factor can be used on positive semi-definite matrices, it does not handle issues of null determinants. Therefore, two new criteria have been proposed, which are presented as follows. In the next section, it can be noted that the trace criterion is efficient to handle null determinant issues.

### 1.3. Two Proposed Convergence Criteria

The original multivariate criterion (Brooks and Gelman, 1998) is theoretically considered efficient if there are high correlations between the parameters. The extreme case is the circumstance of perfect correlation between the parameters. In that case, the multivariate criterion would be equivalent to the univariate criterion of any of the  $p$  parameters, because monitoring the correlation of one parameter reflects what happens in the other parameters. In the case of low correlations between the parameters, or their subgroups, this criterion may fail to monitor the convergence, because it considers only the direction of greatest variability in  $p$ -dimensional hyperspace. In the extreme case of no correlation, this criterion will only monitor the parameter of greatest disturbance (variance) for the system.

Such limitations allow further reflection on this method. This reflection has allowed two new multivariate alternatives. Both consider the variability in all directions of hyperspace, which are linear orthogonal transformations (rotations) of the parameters axes.

The first alternative proposed is based on replacing the scalars  $a^T \bar{V} a$  and  $a^T W a$  from the original criteria by new

scalars trace ( $\bar{V}$ ) and trace (W), respectively, where W is the covariance matrix within the chain given by (1) and  $\bar{V}$  is the estimated total covariance matrix given by (1).

Let us consider a simple case where each Markov chain is an i.i.d. sample and each chain has one mean  $\mu_j$ . Consider m as fixed and let the number of iterations  $n \rightarrow \infty$ .

Then:

$$W \rightarrow I \text{ and } \frac{1}{n}B \rightarrow \frac{1}{m=1}$$

$$\sum_{j=1}^m \left( \mu_j - \frac{1}{m} \sum_{j=1}^m \mu_j \right) \left( \mu_j - \frac{1}{m} \sum_{j=1}^m \mu_j \right)^T = E$$

Thus:

$$W^{-1}\bar{V} \rightarrow I + \left( 1 + \frac{1}{m} \right) E$$

And:

$$\frac{\text{trace}(\bar{V})}{\text{trace}(W)} \rightarrow 1 \text{ since } E = 0$$

The second criterion proposed is the product of non-zero eigenvalues of the matrix  $W^{-1}\bar{V}$ ,  $\prod_{i=1}^r R_i^p = |\bar{V}|/|W|$ , where r is the number of non-zero eigenvalues of (1) and equality holds only when W is invertible. Both the criteria are univariate measures of variability from the process. Thus, the following criteria can be defined:

$$\bar{R}_{\text{trace}}^p = \frac{\text{trace}(\bar{V})}{\text{trace}(W)}$$

$$\bar{R}_{\text{det.}}^p = \prod_{i=1}^r R_i^p = \prod_{i=1}^r \lambda_i$$

These methods do not necessarily need the Bayesian set up. They work for MCMC in general.

## 2. MATERIALS AND METHODS

For assessing the convergence criteria of the MCMC two models were considered. For this, specific cases were simulated for the values of the parameters and the Gibbs sampler was used to generate values for the models. Samples of the posterior joint distribution and the marginal distributions of the parameters were obtained.

The first model used was a time series model with two interventions and correlated errors. The fitted intervention model was an ARMA(2, 2) (Morettin and Toloi, 2008).

The intervention model with autoregressive moving average error of order p and q, denoted by ARMA(p,q), is given by:

$$y_t = X_t^T \beta + \frac{1 + \theta_1 \alpha_{t-1} + \dots + \theta_q \alpha_{t-q}}{1 - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p}}, t = 1, 2, \dots, n$$

where  $\alpha_t$  is the residue, considered as white noise, which is a sequence of random variables i.i.d  $\sim N(0, \tau^{-1})$ , where  $\tau$  is the precision and  $\tau^{-1} = \sigma^2$  is the variance,  $X_t^T = [\xi_{t,1}, \xi_{t,2}, \dots, \xi_{t,w}]$  is a matrix ( $n \times w$ ) of binary variables in which each element is a vector and w is the number of interventions and  $\beta^T = [\beta_1 \beta_2 \dots \beta_w]$  is a vector of intervention parameters. This model was characterized by Diaz (1988). The Bayesian analysis with prior and full conditional distributions for each parameter was developed by Milani (2000).

The parametric values used for the simulation were  $\theta_1 = -0.3$ ;  $\theta_2 = 0.5$ ;  $\beta_1 = -30$ ;  $\beta_2 = 20$ ;  $\phi_1 = 0.5$ ;  $\phi_2 = -0.3$  and  $\tau = 1$ .

The second model used was the trivariate normal distribution. The Gibbs sampler was used to generate Monte Carlo samples for the three variates. Furthermore, the property of the multivariate normal distribution, stating that all the subsets of X are also multivariate normally distributed, was used. By taking up a partition

$${}_p X_1 = \begin{bmatrix} {}_q X_1 \\ (p-q) X_1 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \text{ and its corresponding partitions of}$$

the mean vector and covariance matrix, one can obtain:

$${}_p \mu_1 = \begin{bmatrix} {}_p \mu_1 \\ (p-q) \mu_1 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \text{ and } \sum \begin{bmatrix} {}_q \sum_{11q} & {}_q \sum_{12(p-q)} \\ (p-q) \sum_{21q} & (p-q) \sum_{22(p-q)} \end{bmatrix}$$

where,  $X_1 \sim N_p(\mu_1, \Sigma_{11})$ ,  $X_2 \sim N_p(\mu_2, \Sigma_{22})$  for  $q < p$  and  ${}_q \sum_{11q}$  and  $(p-q) \sum_{22(p-q)}$  are the covariance matrices of  $X_1$  and  $X_2$ , respectively.

Under such partition, the conditional distribution of  $X_1|X_2$  is given by:

$$X_1 | X_2 \sim N_p(\mu_c, \Sigma_c)$$

where the vector of means  $\mu_c = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2)$  and covariance matrix is given by  $\Sigma_c = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$  (Bock, 1985).

The trivariate normal model was simulated by using the distribution given in (2), by considering three cases of correlation between the variates with distinct variances (1, 10 and 100). The correlation matrices,  $\rho_i$  and the resulting covariance matrices,  $\Sigma_i$ , adopted were:

$$\rho_N = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \Sigma_N = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 100 \end{bmatrix},$$

$$\rho_M = \begin{bmatrix} 1 & 0.25 & 0.52 \\ 0.50 & 1 & 0.48 \\ 0.52 & 0.48 & 1 \end{bmatrix} \text{ and } \Sigma_M = \begin{bmatrix} 1 & 1.58 & 5.20 \\ 1.58 & 10 & 15.18 \\ 5.20 & 15.18 & 100 \end{bmatrix},$$

$$\rho_H = \begin{bmatrix} 1 & 0.98 & 0.98 \\ 0.98 & 1 & 0.99 \\ 0.98 & 0.99 & 1 \end{bmatrix} \text{ and } \Sigma_H = \begin{bmatrix} 1 & 3.10 & 9.70 \\ 3.10 & 10 & 31.31 \\ 9.70 & 31.31 & 100 \end{bmatrix}$$

where,  $\Sigma_i = V_i^{-\frac{1}{2}} \rho_i V_i^{-\frac{1}{2}}$ ,  $V_i = \text{diag}(\Sigma_i)$  and  $i = N, M$  and  $H$ . This distinction allows cases of high, medium and null correlation between the conditionals to be simulated, coinciding with the extreme case, where the conditionals are same as the marginals.

The criterion of Brooks and Gelman (1998) was adopted to assess both the convergence of the seven parameters in the intervention model with ARMA(2,2) error and the variates of the trivariate normal model. The criterion was performed iteratively along with the Gibbs sampler. At 20 iterations, the criterion was calculated for the first time and then every two iterations, always considering a burn-in of 50%. For the times series model, 3,671 iterations were performed when considering  $m = 2$  chains in parallel. For 3, 5 and 7 chains, 5,000 iterations were calculated and performed in parallel. For the trivariate normal model, 5,000 iterations were calculated for 2, 3, 5 and 7 chains in parallel. Additionally, the two new criteria proposed and described earlier were implemented.

### 3. RESULTS

#### 3.1. Intervention Model with Error ARMA (2, 2)

The number of non-zero eigenvalues of the H matrix is  $r = \min(m-1; p)$ , where  $m$  is the number of chains and  $p$  is the number of parameters. Hence, the number of non-zero eigenvalues is directly related to the number of chains used. Therefore, the cases with 2, 3, 5 and 7 chains in parallel were studied. In the subsequent section, only the evaluation of the performance of the

convergence criteria is addressed and the estimation of the model parameters (2) will not be mentioned.

The trace criterion presented somewhat smoother behavior and lower values than the others ( $m = 2$ ). From **Fig. 1**, it can be observed that the determinant criterion agrees with the original Brooks and Gelman, indicating convergence at  $R = 1.2$  with 396 and 458 iterations, but has a small fluctuation around  $R = 1.05$  with 1,760 and 1,770 iterations, respectively. The trace criterion, as reviewed, characterizes the convergence at  $R = 1.2$  with 384 iterations. The univariate criterion of Gelman and Rubin was obtained and the value of convergence was reached at 380 iterations. It can be noted that the trace criterion detected the convergence as fast as the univariate criterion.

The results of 7 chains are presented in **Fig. 2**. When considering more chains, there is greater precision in the estimation of the covariance matrix and thereby reducing the fluctuations.

For 3 chains, convergence at  $R = 1.2$  with 486 iterations for the trace criterion, 526 iterations for the determinant criterion and 524 iterations for the original criterion of Brooks and Gelman was achieved. When only 5 chains were used in parallel, the convergence was detected with 352, 370 and 364 iterations, respectively. For 7 chains, the convergence was detected with 422, 880 and 850 iterations, respectively (**Fig. 2**). It can be observed that the trace criterion always characterized the convergence before the others. Another interesting point is that the iterations that characterized the convergence did not vary much as the number of chains increased for the trace criterion. The results for the determinant criterion were very close to the original Brooks and Gelman criterion in all circumstances.

The increase in the number of chains increased the number of eigenvalues. By evaluating the results of these eigenvalues, we can see that there is always an eigenvalue representing over 70% of the variation of the system due to the parameters of the intervention model being very much correlated.

#### 3.2. Trivariate Normal distribution

By using a trivariate normal model and considering no correlation, there is, as commented earlier, only Monte Carlo simulation. Therefore, the process is expected to suffer only the influence of the arbitrary initial value. For this situation, there are results considering 2, 3, 5 and 7 chains in parallel. **Figure 3** presents the criteria for 7 chains.

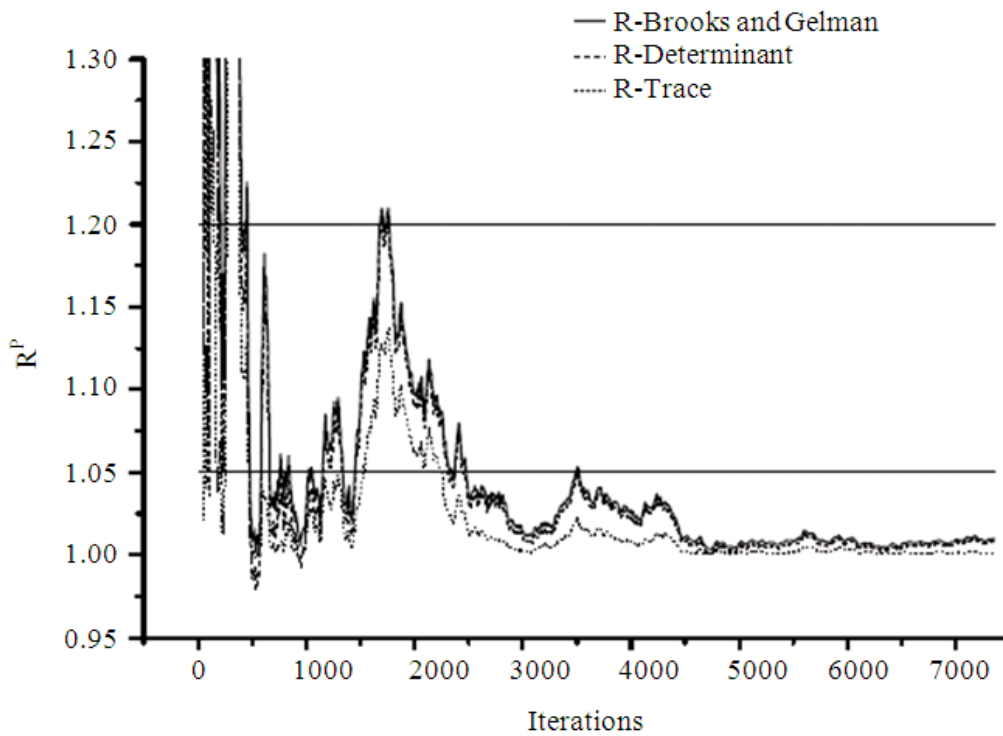


Fig. 1. Graphical representations of multivariate criteria whereas 2 chains, emphasizes the scale of factor R

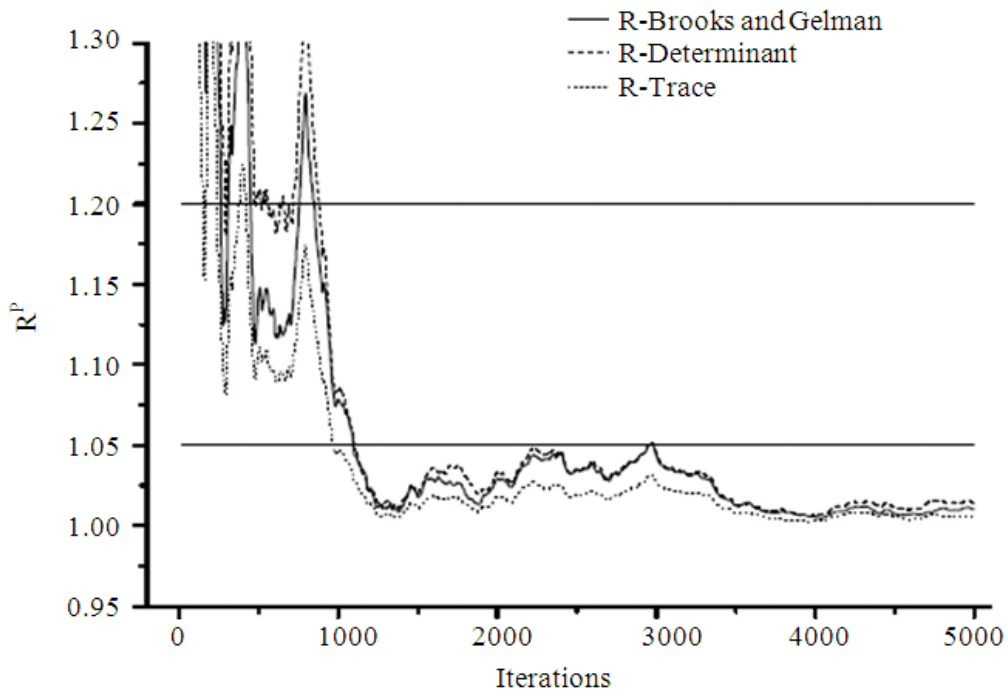


Fig. 2. Graphical representation of multivariate criteria for 7 chains in parallel for the intervention

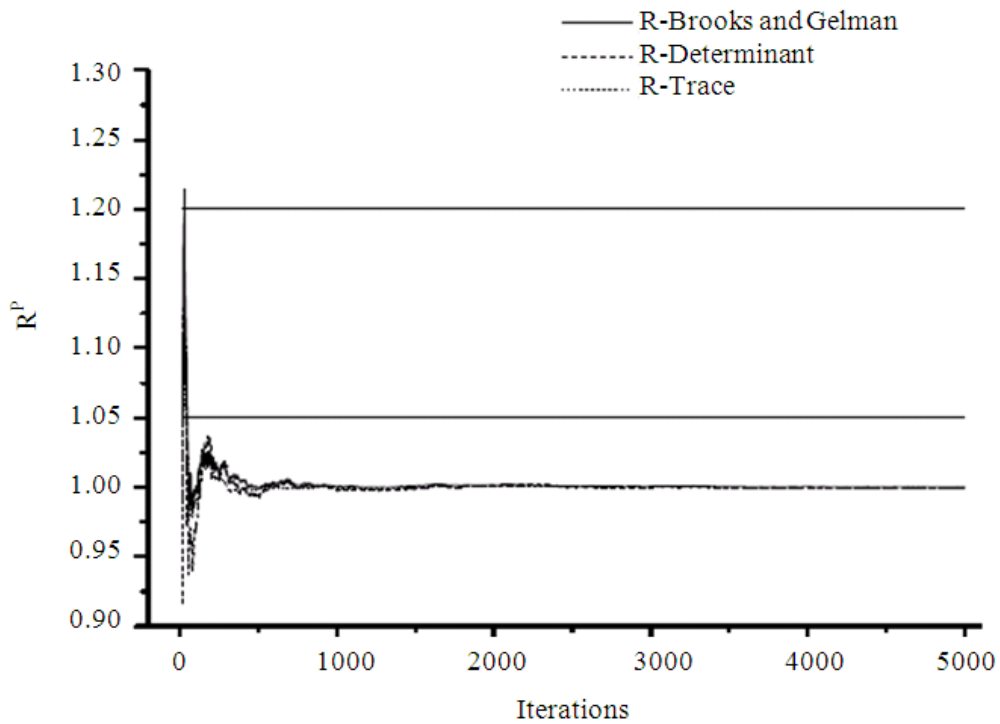


Fig. 3. Graphical representation of multivariate criteria for 7 chains in the situation of lack of correlation

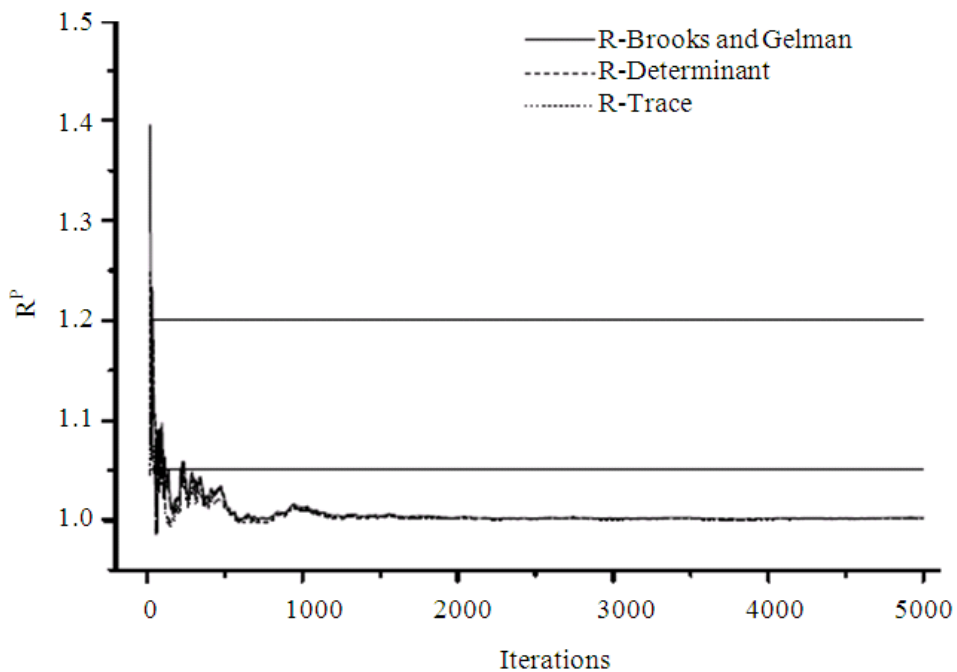


Fig. 4. Graphical representation of multivariate criteria for 7 chains in parallel for the trivariate normal model under medium correlation

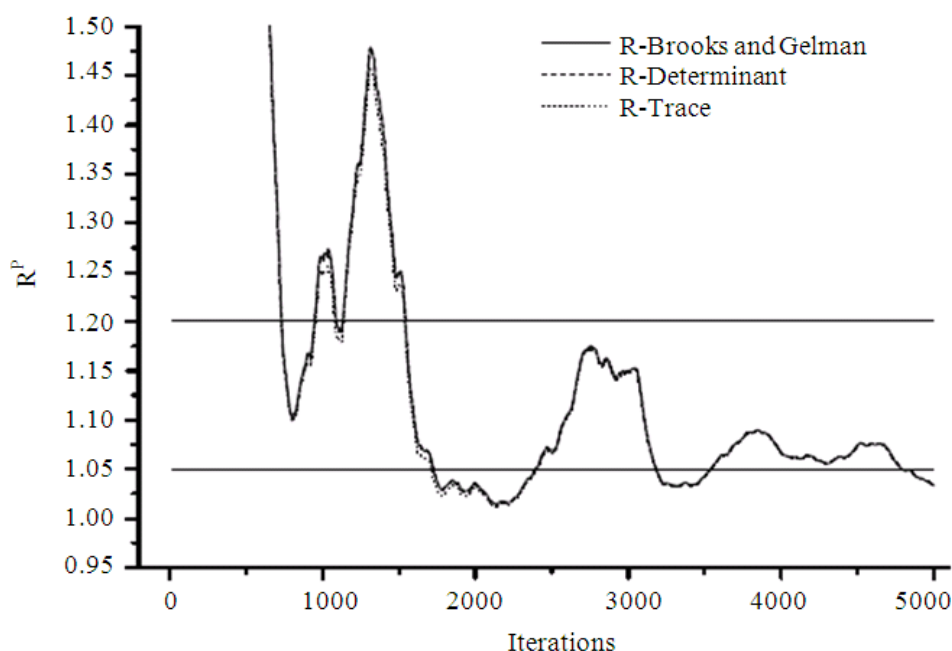


Fig. 5. Graphical representations of multivariate criteria considering 2 chains for the case of high

The characterization of convergence used 2 chains for  $R = 1.2$ , when it occurred with 28 iterations for the original criterion of Brooks and Gelman, with 26 iterations for the determinant criterion and with less than 20 iterations to the trace criterion. When 7 chains were used, the convergence was characterized at very close values: 30, 30 and less than 20 iterations, respectively. In all cases, the convergence was achieved with fewer iterations using the trace criterion. It appears that the other two methods overestimated the convergence time because, as already reported, the chains genuinely originated from a Monte Carlo process and were serially uncorrelated. Except for the initial value, the sample was already in equilibrium from the second iteration.

Similar to the time series model, one can notice that there are few differences when using a different number of chains. However, only for the trace criterion, such possible differences were not apparent, because this criterion showed the best results for the lack of a correlation situation. This can also be explained by the fact that the model has only three parameters and hence, there is no gain in increasing the number of chains from 3. Now, with regard to the intermediate correlation, a similar behavior of the criteria for 7 chains was observed, as shown in Fig. 4. The characterization of convergence when 2 chains were used for  $R = 1.2$  occurred with 54 iterations for the original criterion of Brooks and Gelman,

with 40 iterations for the determinant criterion and with 40 iterations for the trace criterion.

The characterization of convergence when 7 chains were used showed little difference between the previous values, i.e., 40, 32 and less than 20 iterations, respectively.

It must be noted that high correlation between the variables allows a single variable to explain the other two, i.e., the sampling process becomes slow due to the dependence of the full conditional, which exhibit the same behavior (Gamerman and Lopes, 2006). Therefore, it has an eigenvalue that explains virtually 100% of the variation. From Fig. 5 and 6, one can observe, as expected, the equality of the criteria even with the increase in the number of chains. The characterization of convergence when 2 chains were used for  $R = 1.2$  occurred at 1,544 iterations for the original criterion of Brooks and Gelman, at 1,544 iterations for the determinant criterion and at 1,538 iterations for the trace criterion. Furthermore, the characterization of convergence when 7 chains were used showed the values of 1,760, 1,762 and 1,754 iterations, respectively.

It is clear that in the presence of high correlation, the process mixes slowly, requiring more iterations. One way to accelerate the convergence would be reparametrization, but in some circumstances, this is not desired, requiring greater attention.



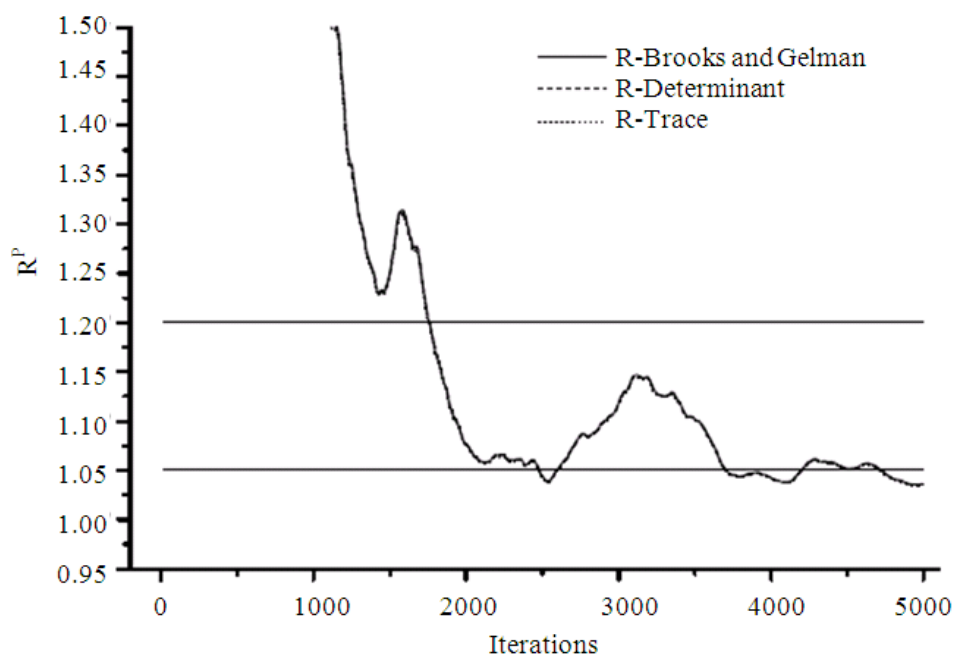


Fig. 6. Graphical representations of multivariate criteria considering 7 chains for the case of high correlation

#### 4. DISCUSSION

Multivariate methods are essential even in circumstances with few parameters, taking into account the variation and correlation in hyperspace. The criterion of Brooks and Gelman (1998) presented results consistent with the convergence of simulated models, but as a generalization of the criterion of Gelman and Rubin, it also has the feature of only monitoring of convergence rather than with the quality of the sample. The proposed alternative for computing such criterion was easily implemented and found to be numerically robust during the simulation. Two alternative criteria were proposed to cover the whole range of parameters in the chains. The trace criterion was easily implemented and showed consistent results and in some cases, was more consistent than the other competitors.

In some circumstances, such as when one is interested in linear combinations of the parameters, the matrix of covariances within chain ( $W$ ) will present linearly dependent columns. Its determinant is zero and hence it does not allow the use of the determinant criterion. The algorithms built in software such as the SAS and R allow obtaining the Cholesky factor of positive semi-definite matrices. In some cases, due to numerical problems, these algorithms result in negative

eigenvalues, not allowing the computation of the criterion (Brooks and Gelman, 1998). As the trace criterion did not present any problem of this type of situation, it is considered more robust. Moreover, the case of lack of correlation allowed the conclusion that this criterion provides a more precise time of convergence, measured in the number of iterations. Thus, the two competing criteria tend to overestimate the number of iterations needed for convergence to equilibrium.

#### 5. CONCLUSION

The alternative for the calculation of the criterion of Brooks and Gelman (1998) was feasible to be applied and was found to be numerically robust.

Two new criteria for monitoring the convergence of multiple Monte Carlo chains were successfully proposed.

The multivariate criterion, based on the ratio of traces of covariance matrices  $\hat{V}$  and  $W$ , gave rise to suitable and more accurate results and was found to be easy to implement and achieve.

In further studies, other univariate criteria Heidelberger and Welch (1986); Geweke (1992) and Raftery and Lewis (1992), can be generalized to their multivariate versions.

## 6. REFERENCES

- Bock, R.D., 1985. *Multivariate Statistical Methods in Behavioral Research*. 2nd Edn., Scientific Software International, ISBN-10: 0894980149, pp: 658.
- Brooks, S.P. and A. Gelman, 1998. General methods for monitoring convergence of iterative simulations. *J. Comput. Graphical Stat.*, 7: 434-455. DOI: 10.1080/10618600.1998.10474787
- Brooks, S.P. and G. Roberts, 1998. Assessment convergence of Markov chain Monte Carlo algorithms. *Stat. Comput. Dordrecht*, 8: 319-335.
- Brooks, S.P. and P. Giudici, 2000. Markov Chain Monte Carlo convergence assessment via two-way analysis of variance. *J. Comput. Graph. Stat.*, 9: 266-285.
- Cowles, M.K. and B.P. Carlin, 1996. Markov Chain Monte Carlo convergence diagnostics: A comparative review. *J. Am. Stat. Assoc.*, 91: 883-904. DOI: 10.1080/01621459.1996.10476956
- Diaz, J., 1988. Bayesian Analysis of Regression Models with Time Series Errors. In: *Bayesian analysis of Time Series and Dynamic Models*, Spall, J. (Ed.), CRC Press, New York, ISBN-10: 0824779363, pp: 347-358.
- Gamerman, D. and H.F. Lopes, 2006. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. 2nd Edn., Chapman and Hall, London, ISBN-10: 1584885874, pp: 323.
- Gelman, A. and D.B. Rubin, 1992. Inference from iterative simulation using multiple sequence. *Stat. Sci.*, 7: 457-511. DOI: 10.1214/ss/1177011136
- Geweke, J.M., 1992. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In: *Bayesian Statistics 4*, Bernardo, J.M., J.O. Berger, A.P. Dawid and A.F.M. Smith (Eds.), Oxford University Press, USA., ISBN-10: 0198522665.
- Heidelberger, P. and P. Welch, 1986. Simulation run length control in the presence of an initial transient. *Operat. Res. Landing*, 31: 1109-1144.
- Luizi, P.C.S., J. Van De Graaf and F.R.B. Cruz, 2010. Assessing the quality of pseudo-random number generators. *Comput. Econ.*, 36: 57-67. DOI: 10.1007/s10614-010-9210-6
- Milani, L.L., 2000. *Análise Bayesiana do Modelo de Intervenção Com Erro Arma*. 1st Edn., Universidade Federal de Lavras, Lavras, Brazil, pp: 142.
- Morettin, P.A. and C.M.C. Toloi, 2008. *Análise de Series Temporais*. 2nd Edn., Edgard Blucher, Sao Paulo, ISBN-10: 8521203896, pp: 538.
- Raftery, A.E. and S. Lewis, 1992. How Many Iterations in the Gibbs Sampler? In: *Bayesian Statistics 4*, Bernardo, J.M., J.O. Berger, A.P. Dawid and A.F.M. Smith (Eds.), Oxford University Press, USA., ISBN-10: 0198522665, pp: 763-773.