XLIII CILAMCE
21–25, November, 2022
FOZ DO IGUAÇU – PR – BRAZIL

Analysis of Convergence Diagnostics for MCMC methods

Douglas S. de Albuquerque[1], Renato S. Silva[1]

[1]*Laboratório Nacional de Computação Científica*
Av. Getúlio Vargas 333, Petrópolis RJ, Brasil
dougalbu@posgrad.lncc.br, rssr@lncc.br

Abstract. The most used method in Bayesian inference is the Markov Chain Monte Carlo (MCMC). But they are expensive, not only because they require a lot of simulations of the model to explore the posterior distribution but also because there are no clear criteria to determine if the method has converged to ensure the quality of the obtained parameters and error estimates. This work objective is to explore a set of convergence diagnostics or MCMC methods and their strategies for a convection-diffusion model and an epidemiological model, a SIR-type model.

Keywords: Convergence, MCMC, Hellinger distance, Strategies

1 Introduction

The Bayesian inference is a very powerful tool to calibrate, fit or identify a model to a data set. The most used method is the Markov Chain Monte Carlo (MCMC), which uses a proposal distribution to draw samples from a posterior distribution. Even with the increase in computing power, MCMC methods are expensive, not only because they require a lot of simulations of the model to explore the posterior distribution but also because there are no clear criteria to determine if the method has converged to ensure the quality of the obtained parameters and error estimates which forces the modeler to overestimate the number of iterations.

There are several diagnostics in the literature as presented by Brooks and Roberts [1], Roy [2], Gilks et al. [3]. In this work, we explore a set of convergence diagnostics or MCMC methods. We will present in section 2 the Geweke, Gelman, and Boone methods and their strategies, in section 3 we preset two mathematical models utilized, the first is a convection-diffusion model in which for some scenarios are not identifiable. The second one is an epidemiological model, a compartmental model with loss of immunity, that is a helpful tool in predicting many contagious illnesses.

2 MCMC Convergence Diagnostics

Markov Chain Monte Carlo Method (MCMC) and its variations Andrieu and Thoms [4] prove to be a good tool for Bayesian calibration. However, due to its stochastic nature, there is no well-established convergence diagnostic. Frequently, the objective in MCMC simulations is the moments of the marginal distributions of the parameters, so a traditional way to check the convergence is to use the visual examination of the expectations estimate for each parameter based on the samples obtained as mentioned by Roy [2]. However, this procedure is far from an accurate method for evaluating convergence.

The methods used in this work are created based on the fact that the Markov Chain that is generated has achieved a steady-state or very close to a stationary distribution, Brooks and Roberts [1]. These are sometimes called burn-in methods due to the fact that they can be used to determine the length of the chain that can be discarded to get an unbiased expectations estimate.

2.1 Geweke

The Geweke diagnostic is originally developed for a single long chain in Geweke et al. [5], and the idea is to compare if an initial portion of the chain is similar to a later portion of the chain. It is constructed by dividing the

long chain of the parameter θ_i, into two subchains, of sizes n_A and n_B. The location parameter of the distribution of θ_i is evaluated for the two subchains and if they are similar this indicates that the two subchains are drawing samples from the same stationary distribution which indicates that the MCMC converges.

The Geweke diagnostic uses the Two-sample T test of the location parameters with different variances as:

$$T = \frac{\bar{\theta}_B - \bar{\theta}_A}{\sqrt{n_A^{-1} s_A^2 + n_B^{-1} s_B^2}} \tag{1}$$

where $\bar{\theta}$ are defined as:

$$\bar{\theta}^A = n_A^{-1} \sum_{j=1}^{n_A} \theta_j \qquad \bar{\theta}^B = n_B^{-1} \sum_{j=n^*}^{n_B} \theta_j \tag{2}$$

The sample variances,s_A^2 and s_B^2 need to be adjusted because the samples are not independent, so Geweke utilizes a spectral density at frequency zero $\hat{S}(0)$ as a estimate of s^2, which implies in:

$$T = \frac{\bar{\theta}_B - \bar{\theta}_A}{\sqrt{n_A^{-1} \hat{S}_A(0) + n_B^{-1} \hat{S}_B(0)}} \tag{3}$$

If the ratio n_A/n and n_B/n are constant and with

$$\frac{n_A + n_B}{n} < 1 \tag{4}$$

we have that $T \to \mathcal{N}(0,1)$ if $n \to \infty$, making possible to use a hypothesis test to check if the chain has converged. There not a optimal choice for n_A/n and n_B/n as mention in Brooks and Roberts [1], the only restriction is Eq.4. In this work we follow the Geweke et al. [5] suggestion, that is $n_A = 0.1n$ and $n_B = 0.5n$.

2.2 Gelman

The Gelman method presented in Gilks et al. [3], Boone et al. [6], Gelman and Rubin [7] is created to analyze the convergence based on the use of multiple chains. This diagnostic is based on the Analysis of the variance among the chains and in each chain. It is based on the fact that if all chains that was initiated from a different start point, and at the convergence, they are close to each other and the variation inside the chain is small this means that the MCMC achieved a steady state or converged.

$$\bar{\theta}_{..} = \frac{1}{m} \sum_{j=1}^{m} \bar{\theta}_{.j} \tag{5}$$

$$s_j^2 = \frac{1}{(n-1)} \sum_{j=1}^{n} (\theta_{ij} - \bar{\theta}_{.j})^2 \qquad \bar{\theta}_{.j} = \frac{1}{n} \sum_{j=1}^{n} \theta_{ij} \tag{6}$$

$$B = \frac{n}{n-m} \sum_{j=1}^{m} (\bar{\theta}_{.j} - \bar{\theta}_{..})^2 \qquad W = \frac{1}{m} \sum_{j=1}^{m} s_j^2 \tag{7}$$

$$\hat{R} = \sqrt{\hat{var}^+(\Theta|D)} \tag{8}$$

with

$$\hat{var}^+(\Theta|D) = \frac{n-1}{n} W + \frac{1}{n} B \tag{9}$$

$\hat{R}$ close to 1

2.3 Hellinger Distance

Both diagnostics present in previous sections and others in the literature as in Brooks and Roberts [1] are based only on the moments and other summary statistics of the posterior, as noted in Boone et al. [6], even if these diagnostics detect a convergence, the posterior can be different in shape. To overcome this limitation Boone et al. [6], present a diagnostic that uses the Hellinger distance.

The Hellinger distance between two distributions f and g is defined as:

$$H(f,g) = \sqrt{\frac{1}{2}\int \left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 dx} \tag{10}$$

Many distances can be used, but they choose this because it is bounded, so $0 < H(f,g) < 1$, meaning $H(f,g) = 0$ there are no divergence between then and with $H(f,g) = 1$ the distributions do not share a common support as stated by Boone et al. [6]. There are also other distances and methods to compare distributions, but they are not bounded and in the same cases, they are sensitive to the presence of tiles.

Since we have only n samples of f and g, to estimate $H(f,g)$ the distributions are approximate by a kernel density. In this work we use the traditional Gaussian kernel with the bandwidth of the kernel given by Silverman's Rule of Thumb so we have:

$$\hat{f} = \frac{1}{nh\sqrt{2\pi}} \sum_{i=1}^{n} \exp\left(-\frac{(x-x_i)^2}{2h^2}\right) \tag{11}$$

with the bandwidth defined as

$$h = \frac{0.9\min\left(\hat{s}, IQR\right)}{1.34} n^{1/5} \tag{12}$$

where $\hat{s}$ is the sample standard deviation and IQR is the interquartile range.

This diagnostic is used by dividing the chain for θ_i into two parts ($n_A = 0.5n$ and $n_B = 0.5n$) and evaluating the Hellinger distance. For a stop criterium we utilize $\hat{H}(\hat{f}, \hat{g}) < C$ where C is a constant and will choose in function of the accuracy and computational cost in this work we uses the recomendation in Boone et al. [6], where can C can take any of this values $C = \{0.05, 0.075, 0.1\}$.

2.4 Strategies

In standard MCMC implementation, the number of iterations is fixed and given as a parameter. Which makes the user overestimate it. Using a convergence diagnostic, it is possible to make the number of iterations vary by controlling the size of an increment of the chains, reducing the computational cost and controlling the convergence.

In this work, we utilize the Hellinger diagnostic to implement a simple strategy to verify its potential for use in more complex problems. We assume that the initial number of iterations is lower than usual and utilize the following strategy.

$$\begin{cases} \text{IF} \quad \hat{H} \geq 0.07 \quad \text{THEN} \qquad increment = 2.0 \times size; \\ \text{IF} \quad \hat{H} < 0.07 \quad \text{THEN} \qquad increment = 0.02 \times size^* - 0.01 \times size \end{cases} \tag{13}$$

where, *size** is the last chains's size before the $\hat{H}$ value be less than 0.07. To guarantee that the increment always takes positive values, the $size^*$ will be updated as follows:

$$\text{IF} \quad size > 2 \times size^* \quad \text{THEN} \quad size^* = 1.5 \times size^* \tag{14}$$

furthermore, If $increment = 0$ then the chain's size will be increased by 10.

3 Results

We use the standard MCMC method with Metropolis-Hastings sampler as suggested in Gilks et al. [3]. Where each sample of the proposal is a random perturbation of the current position, and it is drawn from a Gaussian distribution where the variance matrix is obtained as indicated by Smith [8]. We apply the diagnostics to two problems: the first is the convection-diffusion problem and the second is an epidemiological problem for an infectious disease where the immunity lasts for a limited period before decay making the individuals once again susceptible to the disease.

The tolerance used in Eq.13 has to be satisfied for all variables. All runs have the same initial point and the same seed for the random number generator. In the Geweke diagnostic, we utilize the same chain partition suggested in Geweke et al. [5]. In Gelman and Hellinger we divide the chain in half to evaluate the diagnostics.

3.1 Convection-Diffusion Problem

The one dimensional transport of a scalar-value passive species concentration $\varphi(x)$ by the velocity field u in a region is modeled in Brooks and Hughes [9] as:

$$-K\frac{d^2\varphi}{dx^2} + u\frac{d\varphi}{dx} = 0 \tag{15}$$

with the following boundary conditions as $\varphi(0) = 0$ and $\varphi(1) = 1$, K is the diffusion coefficient.

We utilize this problem because it has an analytical solution and when we calibrate the two parameters, k and u, with uniform distributions as priors, this model is non-identifiable. A synthetic data set with eleven points equally spaced, is generated in the domain $[0; 1]$ by running the model, eq. (15), with $k = 0.5\ m^2/s$ and $u = 3.0$ m/s and adding an Gaussian error given by $\mathcal{N}(0, 0.1^2)$.

The priors used are $u \sim U(0.01, 5)$ and $k \sim U(0.01, 2)$. Table 1 shows that the MAPs are very close. The Geweke and Gelman take fewer iterations, but the Hellinger diagnostic is comparable with the other in terms of accuracy. One point that is important to note here is that the model has identifiability problems. (Table 1).

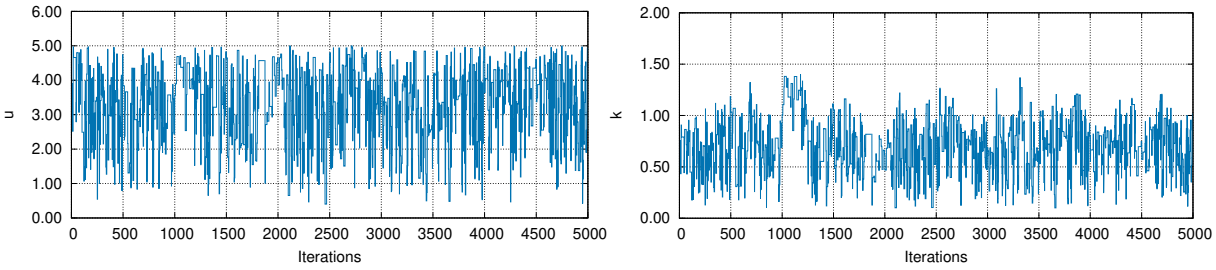
Figure 1, Fig. 2 shows the chains for u and k using the Geweke, Gelman , and Helling diagnostics. All the chains are well mixed and the Hellinger generated a longer chain than the other two diagnostics. This is expected since this diagnostic compares the distribution evolution in the chain and the others compare only averages.

Figure 3 shows that the variable k takes more time to converge. All diagnostics behave well even if the model is non-identifiable.

Figure 3 shows the convergence for each of the variables u and k for the Hellinger diagnostic. We notice that the u converges very fast concerning k. We also notice that in the beginning exists a period with strong oscillations in the convergence curves. After this period, $\hat{H}$ begins to converge. These strong oscillations mark the size of the chain that can be discarded. Which is called the burn-in. Table 1 shows the maximum a posteriori probability (MAP) estimate of the parameters and the number of iterations necessary to converge for each diagnostic.

Table 1. Calibrated parameters (MAP) for each diagnostics

	u	k	Iterations
Geweke	3.39767791	0.70648397	5
Gelman	3.39767791	0.70648397	5
Hellinger	2.43581482	0.50641629	43



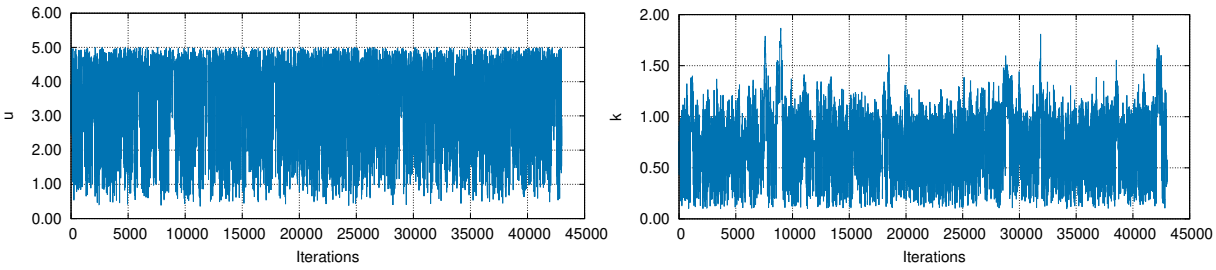
(a) The Chain obtained for u.

(b) The Chain obtained for k.

(c) The Chain obtained for u.

(d) The Chain obtained for k.

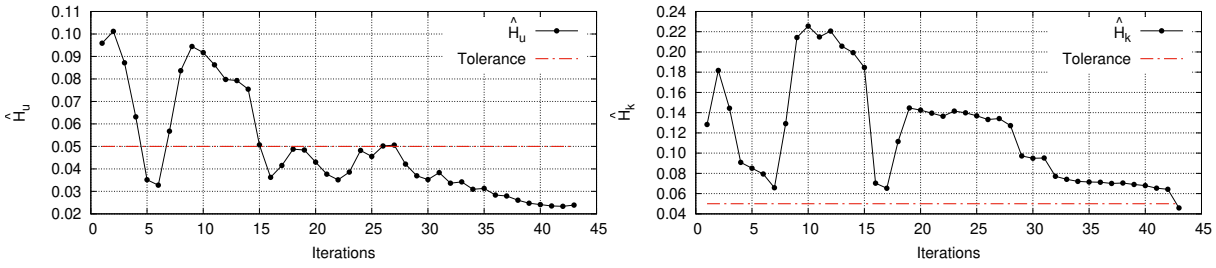
Figure 1. Chains for the convection problem using Geweke and Gelman diagnostic.



(a) The Chain obtained for u.

(b) The Chain obtained for k.

Figure 2. Chains for the convection problem using Hellinger diagnostic.



(a) Convergence for u.

(b) Convergence for k.

Figure 3. Hellinger diagnostic convergence for the convection problem.

3.2 Epidemiological Problem

Keeling and Rohani [10], presented an epidemiological model for diseases that do not activate a long time immunity in an organism. The model is defined as a function of the Susceptible (S), infectious (I), and recovered (R) populations by the system of equations:

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI + wR \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I - wR \end{aligned} \tag{16}$$

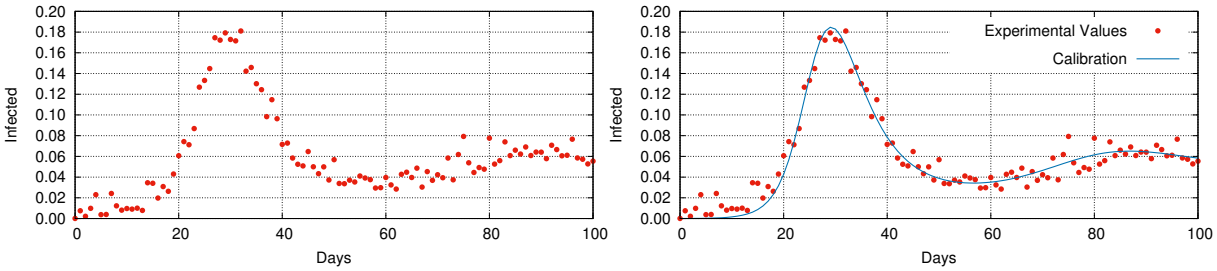
where β is the contact rate, $1/\gamma$ is the infectious period and $1/w$ is the average protected period. The total population is normalized and set to one and the initial condition for the infectious population is $I(0) = 5.0E-05$ and for the recovery recovered population is $R(0) = 0$. In this example, we also utilize a synthetic data set generated with following parameters $\beta = 0.675$, $\gamma = 1/3$, $w = 1/30$ and adding a Gaussian erro generated by $\mathcal{N}(0, 0.015^2)$.

In Table 2 we present the parameters evaluated for all diagnostics and in this model, and all are very close. We utilize in all diagnostics a chain with 10000 samples in the initial step, which is enough for Geweke and Gelman to converge in one iteration. For the Hellinger diagnostic, we show two strategies, one that keeps the size constant and the other presented in Eq. 13 named as Adaptive in Tab.2. Notice that there is a gain in using the adaptive strategy.

Figure 4 presents the data used and the solution for the infectious population using the maximum a posteriori probability estimate of the parameters. The results are in agreement with the data. Figure 5 shows all curves obtained for both strategies and they are equal making the Adaptive strategy save some solutions of the model without affecting the quality of the solution.

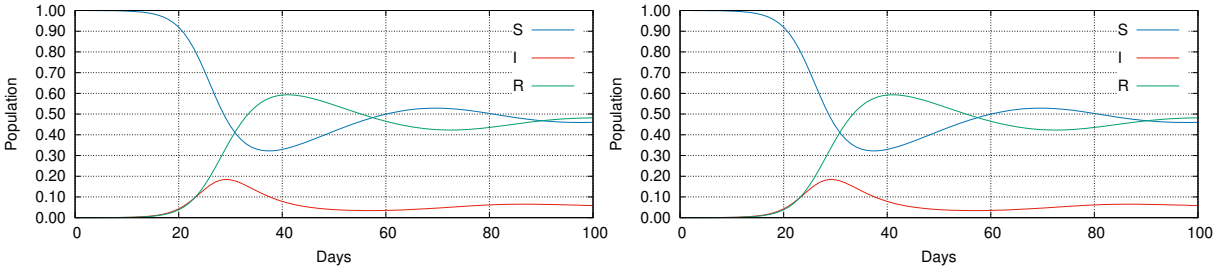
Table 2. Calibrated parameters (MAP) for each diagnostics

	β	γ	w	Iterations	size
Geweke	0.66123972	0.31487478	0.03771673	1	10000
Gelman	0.66123972	0.31487478	0.03771673	1	10000
Hellinger					
Constant	0.66078769	0.31512646	0.03771443	10	60000
Adaptive	0.66078769	0.31512646	0.03771443	16	44899



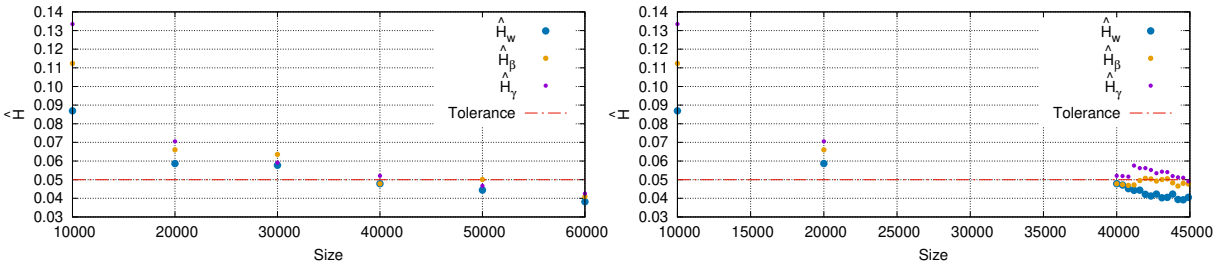
(a) Data (b) Adaptive Strategy

Figure 4. Data and solution obtained with the adaptive strategy for the infectious population



(a) Constant size (b) Adaptive Strategy

Figure 5. Solution obtained with both strategies for the SIR model



(a) Constant size (b) Adaptive Strategy

Figure 6. Convergence for both strategies for thr SIR model

4 Conclusions

In this work, we utilize the Hellinger diagnostics presented by Boone et al. [6], which does not use summary statistics like averages and the variance of the MCMC chains but considers the full distribution, analyzing the Hellinger distance along the chain. It presents good behavior for the two problems we present. Even when one

of the problems is a model with an identifiability problem. Because the diagnostic is dealing with a complete distribution it takes more time to converge. We developed a simple strategy to change the number of samples as a function of the $\hat{H}$ to reduce this difficulty. Of course, a more elaborated strategy can be built but it is the focus of future work. The diagnostic also gives an estimative of the burn-in.

Acknowledgements. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] S. P. Brooks and G. O. Roberts. Convergence assessment techniques for markov chain monte carlo. *Statistics and Computing*, vol. 8, n. 4, pp. 319–335, 1998.
[2] V. Roy. Convergence diagnostics for markov chain monte carlo. *Annual Review of Statistics and Its Application*, vol. 7, pp. 387–412, 2020.
[3] W. R. Gilks, S. R. Richardson, and D. J. Spiegelhater, eds. *Markov Chain Monte Carlo in Practice*. Chapman & Hall / CRC, 1998.
[4] C. Andrieu and J. Thoms. A tutorial on adaptive mcmc. *Statistics and computing*, vol. 18, n. 4, pp. 343–373, 2008.
[5] J. F. Geweke and others. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Technical report, Federal Reserve Bank of Minneapolis, 1991.
[6] E. L. Boone, J. R. Merrick, and M. J. Krachey. A hellinger distance approach to mcmc diagnostics. *Journal of Statistical Computation and Simulation*, vol. 84, n. 4, pp. 833–849, 2014.
[7] A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, vol. 7, n. 4, pp. 457–472, 1992.
[8] R. C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. Society for Industrial and Applied Mathematics, USA, 2013.
[9] A. N. Brooks and T. J. R. Hughes. Streamline upwind Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations. *Computer Methods in Applied Mechanics and Engineering*, vol. 32, pp. 199–259, 1982.
[10] M. J. Keeling and P. Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, 2011.