Asian Journal of Probability and Statistics
X(X): XX-XX, 20YY, Article no.AJPAS.20YY.0XX
ISSN: 2456-477X

SCIENCEDOMAIN *international*
www.sciencedomain.org

Empirical Convergence Rate of a Markov Transition Matrix

Steven T. Garren[*1]

[1] *Department of Mathematics and Statistics, James Madison University, Harrisonburg, VA 22807, USA.*

Article Information
Editor(s):
(1)
(2)
Reviewers:
(1)
(2)
(3)
(4)
Complete Peer review History:

Original Research Article

Received: XX May 2019
Accepted: XX month 20XX
Online Ready: XX month 20XX

Abstract

The convergence rate of a Markov transition matrix is governed by the second largest eigenvalue, where the first largest eigenvalue is unity, under general regularity conditions. Garren and Smith (2000) constructed confidence intervals on this second largest eigenvalue, based on asymptotic normality theory, and performed simulations, which were somewhat limited in scope due to the reduced computing power of that time period. Herein we focus on simulating coverage intervals, using the advanced computing power of our current time period. Thus, we compare our simulated coverage intervals to the theoretical confidence intervals from Garren and Smith (2000).

Keywords: Markov chain Monte Carlo; Gibbs sampling; Hilbert-Schmidt operator; eigenvalue
2010 Mathematics Subject Classification: 62F15, 62F25

1 Introduction

We consider a Markov chain governed by a Hilbert-Schmidt operator. The convergence rate is determined by the second largest eigenvalue of the Markov chain, noting that the largest eigenvalue is one. Under general regularity conditions, this Markov chain is ergodic; i.e., aperiodic and irreducible.

When estimating the convergence rate, the least-squares estimator defined herein is the same one used by Garren and Smith (2000) [1] (G-S), and we adopt their notation as well. Additional theoretical

**Corresponding author: E-mail: garrenst@jmu.edu*

details may be found in G-S, and we focus on simulations for an applied example herein.

An overview of Markov chain Monte Carlo (MCMC) is provided by [2], and an elementary introduction is provided by [3]. Convergence diagnostics for MCMC is analyzed by [4]. Subsampling techniques for hastening convergence of the MCMC are discussed by [5]. A clever R package for performing parallel runs of MCMC is introduced by [6], and a method for accelerating the MCMC is discussed by [7]. In an application to genetics, [8] discussed the difficulty in concluding convergence of MCMC using graphical techniques. An application of MCMC to astronomy is given by [9].

The least-squares estimation of the second largest eigenvalue, along with two nuisance parameters, is discussed in section 2. As an example, we analyze the hierarchical Poisson model in section 3. We end with a brief conclusion in section 4.

2 Least-squares estimation

The Markov chain, as governed by a Hilbert-Schmidt operator, is allowed M burn-in iterations and terminates after a total of N iterations. Then, L independent runs of the Markov chain are performed. Modern computers allow us to select L to be quite huge, especially in comparison to the values of L selected by G-S when the computing power was much less efficient.

For each independent run, the Markov chain is given an initial state. Then, a set D is selected, so that for each iterate we determine whether or not $X_n^{(l)}$, the state of the Markov chain after n iterates of the lth run, is in set D. Hence, we define the indicator variable

$$Z_n^{(l)} = I(X_n^{(l)} \in D), \quad 0 \le n \le N,\ 1 \le l \le L,$$

and we also define

$$\bar{Z}_n = \frac{1}{L} \sum_{l=1}^{L} Z_n^{(l)}$$

to be the average of the $Z_n^{(l)}$ values among the independent runs. The asymptotic behavior of $\bar{Z}_n$ may be written as

$$\bar{Z}_n = \rho + a_2\ \lambda_2^n + o_P(\lambda_2^n), \qquad \text{as } n \to \infty.$$

Note that ρ depends on D; a_2 depends on the initial state and D, whereas λ_2 depends on neither the initial state nor D.

The joint least-squares estimators of (ρ, a_2, λ_2) are defined to be the values of $(\theta_1, \theta_2, \theta_3)$ which minimize

$$\sum_{n=M+1}^{N} \left[\bar{Z}_n - (\theta_1 + \theta_2 \theta_3^n)\right]^2,$$

and are found numerically. G-S showed that the least-squares estimators are consistent and asymptotically normal as M, N, and L go to infinity under certain regularity conditions. They further derived the variance of the asymptotic distribution.

A Markov chain governed by a Hilbert-Schmidt operator allows analysis of a continuous distribution. Thus, G-S generalizes and improves the approach of [10], who estimated the number of iterations needed for just a two-state Markov chain, and G-S also can handle higher-state discrete Markov chains. In terms of real-world applications, estimating the number of iterations needed for convergence allows a researcher to discern the amount of computing time needed when analyzing data from a Bayesian model based on a Hilbert-Schmidt

Table 1: Number of pump failures at a nuclear power plant

Number of failures (y_i)	Time (t_i)
5	94.320
1	15.720
5	62.880
14	125.760
3	5.240
19	31.440
1	1.048
1	1.048
4	2.096
22	10.480

operator. Convergence is rapid when λ_2 is near zero, but is slow when λ_2 is near unity, although λ_2 is more difficult to estimate when it is near zero.

3 Hierarchical Poisson model

The example we analyze herein was studied by [11], [12], and also G-S. Let y_i, the number of failures at a nuclear power plant, be modeled as a Poisson distribution with mean $\omega_i t_i$ for time t_i. We model the parameter $\omega_i \sim \Gamma(\alpha, \beta)$, whose density is $\omega_i^{\alpha-1} \exp\{-\omega_i/\beta\}/\beta^\alpha \Gamma(\alpha)$, for $i = 1, \ldots, 10$.

The data for y_i and t_i are shown in Table 1. We set $\alpha = 1.802$, based on method of moments estimates, as suggested by [11]. Furthermore, we model $1/\beta$ as a $\Gamma(\gamma = 0.01, \delta = 1)$ distribution. These values of α, γ, and δ were used by [11], [12], and G-S.

To set up the Gibbs sampler, we use the conditional distributions

$$[\, \omega_i \mid \beta,\, \omega_j,\, j \neq i;\, \mathbf{y} \,] \sim \Gamma\left(\alpha + y_i,\, (t_i + 1/\beta)^{-1}\right), \quad i = 1, \ldots, 10,$$

and

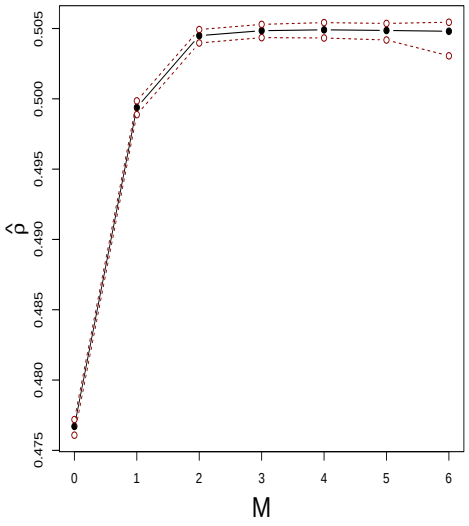
$$[\, 1/\beta \mid \boldsymbol{\omega},\, \mathbf{y} \,] \sim \Gamma\left(\gamma + 10\alpha,\, \left\{1/\delta + \sum_{i=1}^{10} \omega_i\right\}^{-1}\right).$$

This Gibbs sampler is reversible and is produced by a Hilbert-Schmidt operator; see G-S.

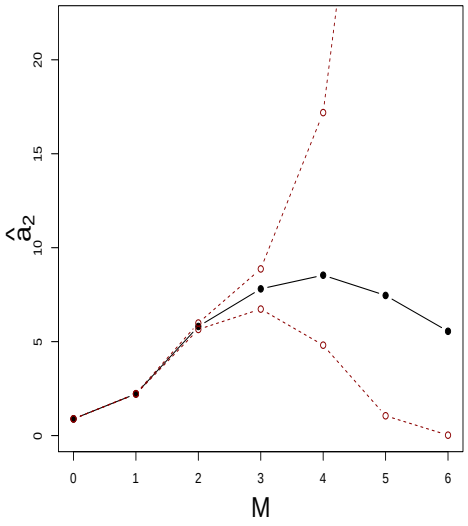
G-S selected $M = 0, \ldots, 6$, $N = 12$, and $L = 5000$. Hence, for each of the seven values of M, G-S obtained one joint least-squares estimator of $(\rho,\, a_2,\, \lambda_2)$, error bounds based on the information sandwich approach, and hence 95% confidence intervals as well. Graphs of their estimates, along with 95% confidence intervals, are shown in Figures 1, 2, and 3 of G-S.

Due to increased computing speeds in the statistical software R [13], we increased L to 500,000 and computed 20,000 least-squares estimates rather than just one. By producing 20,000 least-squares estimates, we empirically constructed the coverage intervals, rather than use the information sandwich approach based on just one estimate. Therefore, we are able to evaluate the theoretical

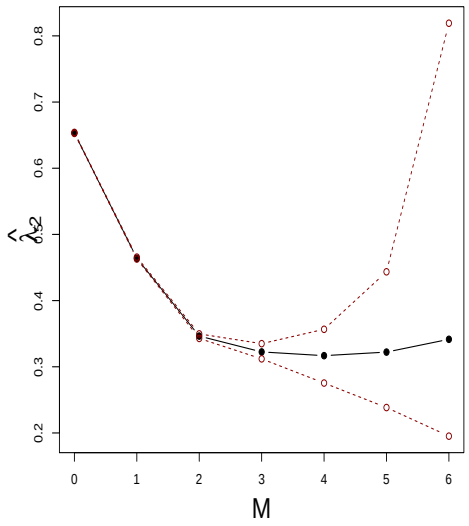
information sandwich approach of G-S by simulating the coverage intervals. Our estimates of ρ, a_2, and λ_2, are shown in Figures 1a, 1b, and 1c, respectively, where the inner line segments represent the median of the 20,000 least-squares estimates and the two outer line segments represent the 95% coverage intervals.



(a) Estimating ρ



(b) Estimating a_2



(b) Estimating λ

Figure 1: 95% coverage intervals on ρ, a_2, and λ, for $M = 0, \ldots, 6$.

Estimates of ρ seem most stable for $2 \leq M \leq 6$, with tight coverage intervals for $0 \leq M \leq 5$, as seen in Figure 1a. The median of $\widehat{\rho}$ is approximately 0.505 for $2 \leq M \leq 6$. This figure hints at the importance of allowing at least a small amount of burn-in, implying a preference of $M > 0$ when performing this least-squares estimation.

The coverage intervals on a_2 are quite narrow for $0 \leq M \leq 2$ but get quite a bit wider as M increases, as shown in Figure 1b. These median values of $\widehat{a}_2$ range between 0.89 and 8.06 for $0 \leq M \leq 6$.

The coverage intervals on λ_2 are somewhat narrow for $0 \leq M \leq 3$ but get a lot wider as M increases, as shown in Figure 1c. The median estimate of λ_2 tends to stabilize around 0.33 for $2 \leq M \leq 6$ despite the widening of the coverage intervals.

This widening of coverage intervals on both a_2 and λ_2 as M increases is anticipated, since a small value of λ produces rapid convergence of the Gibbs sampler, causing increased difficulty in estimating both a_2 and λ_2. G-S tended to obtain even wider coverage intervals, which were calculated by the information sandwich approach, but this is not at all surprising since their value of L was much smaller than ours.

Next, we increase L by a factor of 100, so that the new value is $L = 50,000,000$, but we obtain only one least-squares estimate of $(\rho,\ a_2,\ \lambda_2)$. In the color red, we plot $\widehat{\rho}_n = \widehat{\rho} + \widehat{a}_2\widehat{\lambda}_2^n$, $n = 1, \ldots, 12$, where the least-squares estimates are based on $M = 0, \ldots, 6$ in Figure 2. Also in those seven figures, in the color black, we plot what $\widehat{\rho}_n$ is estimating; i.e., $\bar{Z}_n$. The standard errors on $\bar{Z}_n$ are no more than $0.5/\sqrt{L} = 0.00007$, so these standard errors are quite negligible and in fact non-detectable in Figure 2.

4 Conclusion

The simulations herein exemplify the huge difficulty in estimating the second largest eigenvalue, which is heavily tied to the convergence rate of the Gibbs sampler. Using no burn-in of the chain tends to confound the impact of the second largest eigenvalue with the remaining eigenvalues. However, as the amount of burn-in increases, the impact of all eigenvalues, including the second largest eigenvalue though excluding the largest eigenvalue of unity, decreases substantially, again increasing the difficulty in estimating the second largest eigenvalue. Therefore, reasonable estimation of the second largest eigenvalue requires a large number of replications and a small amount of burn-in.

Acknowledgements

The author is very thankful to three anonymous reviewers and an editor, whose helpful suggestions added greatly to the quality of this paper.

Competing Interests

The author declares that no competing interests exist.

References

1. Garren ST, Smith RL. Estimating the second largest eigenvalue of a Markov transition matrix. Bernoulli 2000;6(2):215-242.

2. Gamerman D, Lopes HF. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, 2nd edition, Chapman and Hall, New York, 2006.

3. Van Ravenzwaaij D, Cassey P, Brown SD. A simple introduction to Markov chain Monte Carlo sampling. Psychonomic Bulletin & Review 2018;25(1)143-154.

4. Cowles MK, Carlin BP. Markov chain Monte Carlo convergence diagnostics: a comparative review. Journal of the American Statistical Association 1996;91(434)883-904.

5. Quiroz M, Kohn R, Villani M, Tran M-N. Speeding up MCMC by efficient data subsampling. Journal of the American Statistical Association 2018;1-35.

6. Denwood MJ. (2016). runjags: an R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. Journal of Statistical Software 2016;71(9):1-25.

7. Robert CP, Elvira V, Tawn N, Wu C. Accelerating MCMC algorithms. WIREs Computational Statistics 2018;10:e1345.

8. Nylander JAA, Wilgenbusch JC, Warren DL, Swofford DL. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 2008, 24(4):581-583.

9. Fulton BJ, Petigura EA, Blunt S, Sinukoff E. RadVel: The radial velocity modeling toolkit. Publications of the Astronomical Society of the Pacific 2018;130(986):044504.

10. Raftery AE, Lewis SM. How many iterations in the Gibbs sampler? In Bernardo JM, Berger JO, Dawid AP, and Smith AFM (eds), Bayesian Statistics 4, pp. 763-773. New York: Oxford University Press, 1992.

11. Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. Journal of American Statistical Association 1990;85:398-409.

12. Tierney L. Markov chains for exploring posterior distributions (with discussion). Annals of Statistics 1994;22:1701-1762.

13. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/; 2019.

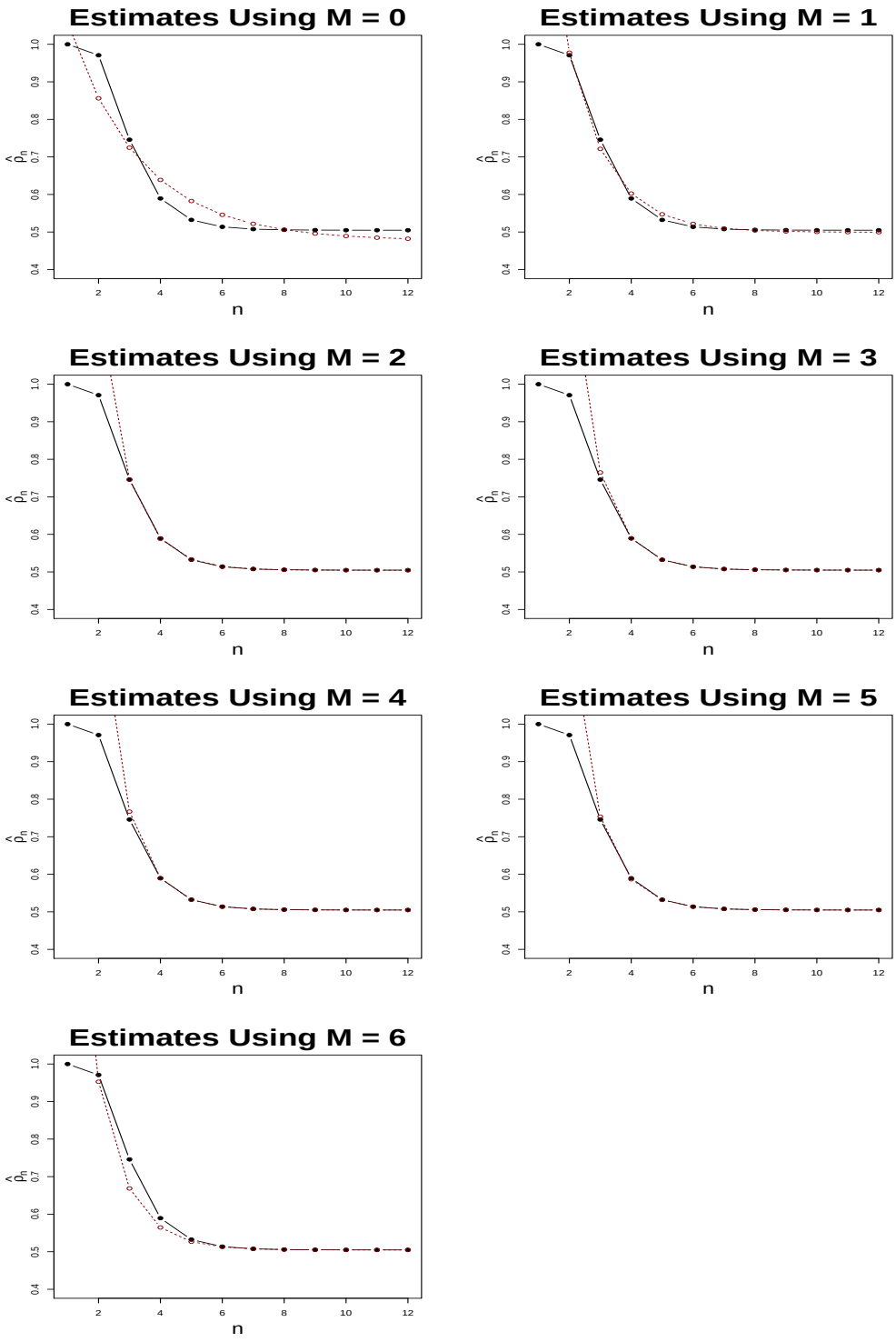


Figure 2: Estimating ρ_n for $M = 0, \ldots, 6$.