

Spectral Gap and Mixing Time in Dynamic Assessment of Transition Probability Matrix Estimators in Markov Chains

1 Rasool Ali Hussein Abbas

rasool.24csp53@student.uomosul.edu.iq

1 Department of Statistics and Informatics – College of Computer Science and Mathematics – University of Mosul – Mosul

2 Prof. Dr. Muthanna Subhi Sulaiman

muthanna.sulaiman@uomosul.edu.iq

Department of Statistics and Informatics – College of Computer Science 2 and Mathematics – University of Mosul – Mosul

الفجوة الطيفية وزمن الخلط في التقييم الديناميكي لمقدرات مصفوفة احتمالية الانتقال في سلاسل ماركوف

رسول علي حسين عباس

قسم الإحصاء والمعلوماتية_ كلية علوم الحاسوب والرياضيات_ جامعة الموصل

أ.د. مثنى صبحي سليمان

قسم الإحصاء والمعلوماتية_ كلية علوم الحاسوب والرياضيات_ جامعة الموصل

Abstract

In this paper, we evaluate transition probability matrix estimators in Markov chains on a dynamic basis. The majority of the previous studies have been about statistical accuracy, whereas the present study concerns the effect of the estimated matrix on convergence and long-run probabilistic behavior. The Maximum Likelihood Estimator (MLE), the Bayesian-Dirichlet estimator, the shrinkage estimator and the Artificial Bee Colony (ABC) estimator are compared. A comparison is made based on a simulation design with dense, moderate and sparse transition-data settings. Performance is measured in terms of mean squared error, element-wise variance, spectral gap and mixing time. The results imply that the general performance of ABC is best especially where the data is sparse, while shrinkage is always second and it is even better than the traditional estimators. The findings demonstrate that statistical superiority does not necessarily imply dynamic superiority and that dynamic criteria provide much-desired additional information on the quality of estimators.

Keywords: Convergence, Dynamics, Estimation, Markov Chains, Optimization, Robustness, Stochastic Models, spectral gap, Transition Matrix, Variance

ملخص

في هذا البحث، نقوم بتقييم مقدرات مصفوفة احتمالية الانتقال في سلاسل ماركوف على أساس ديناميكي. ركزت معظم الدراسات السابقة على الدقة الإحصائية، بينما تتناول هذه الدراسة تأثير المصفوفة المقدرة على التقارب والسلوك الاحتمالي طويل المدى. تتم مقارنة مقدر الاحتمال الأقصى (MLE)، ومقدر بايزي-ديريشلت، ومقدر الانكماش، ومقدر مستعمرة النحل الاصطناعية (ABC) تجرى المقارنة بناءً على تصميم محاكاة مع إعدادات بيانات انتقال كثيفة ومتوسطة ومتفرقة. يُقاس الأداء من حيث متوسط مربع الخطأ، والتباين العنصري، والفجوة الطيفية، وزمن الخلط. تشير النتائج إلى أن الأداء العام لمقدر مستعمرة النحل الاصطناعية هو الأفضل، خاصةً عندما تكون البيانات متفرقة، بينما يأتي مقدر الانكماش في المرتبة الثانية دائماً، وهو أفضل من المقدرات التقليدية. تُظهر النتائج أن التفوق الإحصائي لا يعني بالضرورة التفوق الديناميكي، وأن المعايير الديناميكية تُوفر معلومات إضافية مرغوبة للغاية حول جودة المقدرات. الكلمات المفتاحية: التقارب، الديناميكية، التقدير، سلاسل ماركوف، التحسين، المتانة، النماذج العشوائية، الفجوة الطيفية، مصفوفة الانتقال، التباين

1. Introduction

Markov chains have been widely used in the modelling of systems that evolve over time, but in an unpredictable way. They are significant because of the Markov property that says the future state depends on the present state and not the whole past history (Kemeny & Snell, 1976). The property has made Markov chains useful in probability, statistics, economics, engineering, operations research and most applied sciences (Ross, 2014) (Norris, 1998) (Billingsley, 1961). The transition probability matrix is the most important aspect in these models since it determines the movement between the states and convergence, stationarity, and long-run stability (Seneta, 2006) (Levin et al., 2017). This has rendered the estimation of the transition probability matrix the primary concern in the area of stochastic modeling (Anderson & Goodman, 1957). The classical benchmark is the Maximum Likelihood Estimator, the result of the row-wise normalization of the observed counts of transitions. This was followed with the introduction of Bayesian-Dirichlet estimators to introduce some smoothing in the sparse data case, and shrinkage estimates and optimization techniques were introduced to increase robustness in the sparse or irregular transition data case. These developments suggest that estimators are practically needed that are not sensitive to transition data being non-ideal (Anderson & Goodman, 1957) (Minka, 2000) (Ledoit & Wolf, 2004). Despite this being done, a lot of comparative analysis continues to address estimators founded on statistical measures such as the mean squared error or likelihood measures. These are informative criteria however, are not sufficient in the instance where the estimated matrix is utilized as a dynamic mechanism. Two estimators may appear to be the same in terms of their numerical values, but imply different rates of convergence, different long-run persistence of the initial state, and different long-run probabilistic behavior (Gelman et al., 2013) (Levin et al., 2017). In practice, these dynamic differences may be most helpful to interpret and forecast. This paper deals with that limitation by discussing transition probability matrix estimators in a dynamic view. The general aim is to compare four estimators, including MLE, Bayesian-Dirichlet, shrinkage, and Artificial Bee Colony (ABC) in terms of spectral gap and mixing time as the primary dynamic indicators and mean squared error and element-wise variance as the secondary statistical and stability indicators. The study question of interest is whether the dynamically best estimator is also the statistically best estimator, especially in cases where transition data become sparse. The novelty of the paper is not that it presents a new dynamic concept, but rather that it repositions the dynamic criteria as a major point of comparison amongst estimators within one simulation framework. This is how the paper provides the more detailed picture of the quality of estimators and clarifies the difference between the numerical accuracy and dynamic adequacy.

2. Research Gap and Literature Review

The empirical estimator emerged as a typical standard benchmark in the estimation of transition probabilities in the early inferential literature on Markov chains. It is a simple, transparent and computationally friendly estimator, but can do badly with small or highly unbalanced transition counts. Finite samples, especially those with sparse observations, may cause the empirical probabilities to become unstable, and it may be excessively sensitive to the effects of chance sampling (Anderson & Goodman, 1957) (Billingsley, 1961).

To reduce this instability, the Bayesian-Dirichlet estimation was suggested to provide prior smoothing. It eradicates extreme empirical probabilities, and tends to provide more regular estimates, by accumulating transition counts with the prior mass. But its effectiveness depends upon prior specification, and its improvement may not be much when structural sparsity occurs, as opposed to incidental sparsity. In such cases, smoothing may not be sufficient to regain desirable matrix behavior (Minka, 2000) (Bernardo & Smith, 1994). Shrinkage methodology offers a second way out in terms of combining an empirical estimator with an organized target matrix. The idea is to reduce variance at the cost of a constant level of bias, thereby improving stability in the case of noisy or inadequate data. Shrinkage can be especially useful in the case of transition matrices as it introduces regularity without distorting the information in the observed counts. This makes it attractive in moderate and sparse settings (Ledoit & Wolf, 2004) (Schäfer & Strimmer, 2005). Another important methodological direction is the one that is based on optimization. It is possible to estimate the transition matrix by solving a constrained global optimization problem on the space of row-stochastic matrices, instead of estimating the individual matrices. The Artificial Bee Colony algorithm is a metaheuristic algorithm of particular interest in this respect as it provides flexible search under stochastic constraints, and can be adapted to objective functions where a variety of desirable properties are balanced

(Karaboga, 2005) (Karaboga & Basturk, 2007) (Akay et al., 2021). However, the literature trend is largely to compare the statistical accuracy as the main basis of comparison. The dynamic aspects of the estimated matrix are well known in the theoretical analysis of Markov-chains, but are rarely viewed as of importance in estimator comparison. This creates a gap in research: transition probability matrix estimators have not been sufficiently compared with regard to the convergence behaviour they cause, especially across different levels of the transition-data density (Levin et al., 2017) (Montenegro & Tetali, 2006). The present paper closes this gap by taking into account the quality of estimators as it would depend on the dynamic criteria rather than merely depending on the statistical closeness to the true matrix. It also serves to position both traditional and proposed estimators within a single coherent simulation framework and to highlight sparse-data situations, where practical distinctions between estimators are most apparent, and where dynamic evaluation is especially useful.

2.1 Nomenclature and Symbols

The absolute eigenvalue of transition matrix second largest is labeled as λ_2 , the stationary distribution is denoted as π , and *spectral gap* is labeled as $Gap(P)$ and *mixing time* is labeled as t_{mix} . These symbols are used in all the formulae that follow in such a manner that the comparison of the estimators is clear, and can be understood easily. In order to be specific, the key notation that will be employed throughout the paper is as follows summarized. K denotes the number of states of the Markov chain, P represents the true transition probability matrix and $\hat{P}$ is an estimated transition probability matrix in general. Where, entry p_{ij} is the probability of transitioning between state i and state j in one step and f_{ij} is the observed transitions between state i and state j and n_i is the transitions observed to have left state i .

3. Methodology

3.1 Markov Chain Framework

Consider a finite discrete-time Markov chain with state space $S = \{1, 2, \dots, K\}$ and transition matrix $P = p_{ij}$ (Neuts, 1981):

Where

$$p_{ij} = \Pr\{X_{t+1} = j \mid X_t = i\} \quad (1)$$

The transition matrix is stochastic and has to comply with two requirements: it has to have nonnegative entries, (i.e., $0 \leq p_{ij} \leq 1$), and the sum of any row equals one

The estimation problem is to recover P using F and ensure the stochastic constraints of the matrix and hopefully, its favorable dynamic properties.

3.2.1 Maximum Likelihood Estimator

Four estimators are taken into consideration.

1. Maximum Likelihood Estimator (Boas, 2006):

$$\hat{p}_{ij}^{(MLE)} = \frac{f_{ij}}{\sum_j f_{ij}} = \frac{f_{ij}}{n_i}, \quad (2)$$

when

$$n_i = \sum_{j=1}^K f_{ij}.$$

3.2.2 Bayesian-Dirichlet Estimator

Bayesian-Dirichlet estimator assumes the Dirichlet prior on each row, and employs the posterior mean (Gelman et al., 2013) (Trendelkamp-Schroer et al., 2015):

$$\hat{p}_{ij}^{(Bayes)} = \frac{q_{ij}^c + f_{ij}}{n_i + \hat{c}}, \quad (3)$$

3.2.3 Shrinkage Estimator

The shrinkage estimator is a combination of the empirical matrix and a structured target $T = (t_{ij})$ (Fiecas et al., 2017) (Zhou et al., 2024):

$$\hat{P}^{(SH)} = (1 - \lambda_{global}^c) \hat{P}^{(MLE)} + \lambda_{global}^c T, \quad (4)$$

Where $0 \leq \lambda_{global}^c \leq 1$.

3.2.4 Artificial Bee Colony (ABC) Estimator

The ABC estimator can be obtained by solving a constrained optimization problem over the row-stochastic matrices, in a manner that

$$\hat{P}^{(ABC)} = \arg \min \phi(P), \quad (5)$$

where ϕ is the objective in the thesis framework.

This formulation is fully consistent with the established ABC literature and with more recent reviews of ABC variants in broader optimization settings (Karaboga *et al.*, 2014) (Akay *et al.*, 2021).

3.3 Evaluation Criteria

There are four criteria used to compare the estimators.

3.3.1 Mean Squared Error

Mean squared error is defined by

$$MSE(\hat{P}) = \frac{1}{K^2} \sum_{i=1}^K \sum_{j=1}^K (\hat{p}_{ij} - p_{ij}^{true})^2. \quad (6)$$

3.3.2 Element-wise Variance

The element-wise variance between simulation replicas is summarized as:

$$VarElem(\hat{P}) = \frac{1}{K^2 - 1} \sum_{i=1}^K \sum_{j=1}^K (\hat{p}_{ij} - \bar{p})^2 \quad (7)$$

3.3.3 Spectral Gap

Suppose that the second largest eigenvalue of P in absolute value is denoted by λ_2 , the spectral gap is:

$$Gap(\hat{P}) = \begin{cases} 1 - |\lambda_2|, & 0 < |\lambda_2| < 1 \\ 0, & |\lambda_2| \geq 1 \end{cases} \quad (8)$$

A bigger spectral gap means that there is a faster convergence and hence a better dynamic behavior.

3.3.4 Mixing Time

The mixing time

$$t_{mix}(\epsilon_{mix}) = \begin{cases} \left\lceil \frac{\log(\epsilon_{mix})}{\log(|\lambda_2|)} \right\rceil, & 0 < |\lambda_2| < 1 \\ \infty, & |\lambda_2| \geq 1 \end{cases} \quad (9)$$

The lower the mixing times, the quicker the approach to stationarity. Spectral gap and mixing time, together, give direct information on the rate of disappearance of the effect of the initial state and the rate at which the chain settles down.

3.4 Simulation Design *The empirical research is based on the simulation framework which is developed in the thesis. Three settings of transition-data density are taken into account, dense, moderate, and sparse. In each setting, 100 transition-frequency matrices are created, and the four estimators are calculated. Though mean squared error is also used as the benchmark statistical criterion, spectral gap and mixing time are given more prominence in the current paper since the focus is to analyze the estimators as dynamic entities and not as mere numerical estimations.*

4. Results and Discussion

4.1 Dense Setting Larger variations in mean squared error with the dense setting would be less than large, which would be expected in scenarios with large counts of transitions. However, the proposed estimators particularly, shrinkage and ABC have better dynamic properties in comparison to the conventional estimators as their spectral gaps are higher and mixing times are lower. This means that as much as the difference between the variables may not appear significant statistically, improvement is dynamic and there is still value of dynamic assessment where there is rich data.

4.2 Moderate Setting The variance between the estimators can be more seen in the moderate setting. ABC does best on the criterion of mean squared error as well as indicates that the true matrix is recoverable statistically in a good way. Among other things the shrinkage gives a somewhat wider spectral gap and somewhat less mixing time than ABC. It is also methodologically important since it demonstrates that the optimal dynamic behaviour of the estimator does not always suggest that the estimator will achieve the smallest estimation error. The purest statistical ranking would thereby be lacking in a big part of the comparative picture.

4.3 Sparse Setting The sparse setting is the most practically informative setting since, in this scenario, it is only the classical estimators that are most susceptible to instability. It is obvious here that it is ABC which

would win the ultimate fight between the competing approaches as the result is in the form of the smallest mean squared error and the smallest element-wise variance, the greatest spectral gap and the smallest mixing time. Shrinkage comes in second place and also does much better than MLE and Bayesian-Dirichlet. The above findings show that the proposed estimators are not only statistically competitive, but also dynamically superior, when the transition data available are small and irregular.

4.4 Discussion of the Results One of the potentially useful interpretations is that the shrinkage enhances performance because it stabilises the empirical structure, and that the ABC enhances performance because it applies full-matrix optimisation with stochastic constraints. This would be one of the reasons why shrinkage may have very strong dynamic behaviour in moderate environments and why ABC is dominant in sparse environments where an even more global strategy of search may be more appropriate for handling instability on the entire matrix.

4.5 General Comparative Insight

The significant overall finding is that dynamic quality and statistical accuracy are related, though different facets of estimator performance. The mean squared error is a measure of how close it is to the true matrix, and spectral gap and mixing time are measures of the behavior that is actually due to the resulting chain. Consequently, total evaluation of transition probability matrix estimators must take into account not only the statistical but also the dynamic aspects of the evaluation and it cannot be rooted in either of the two dimensions.

4.6 Practical implications of the Findings

Such results have large practical implications. When doing convergence analysis, long-run prediction or substantive interpretation, using the estimated matrix may be inadequate, and the selection of an estimator based solely on the mean squared error may be inadequate. It is the dynamic action, which the estimator generates in such applications, which are themselves a component of the inferential objective and need to be thought of as such a criterion, not a diagnostic.

٩. Comparative Summary of the Main Findings

Table 1. Comparative Summary of the Main Findings

Setting	Best MSE	Best Dynamic Performance	Interpretation
Dense	Differences small	ABC and Shrinkage	Dynamic gains remain visible even under rich data
Moderate	ABC	Shrinkage slightly stronger dynamically	Best statistical estimator is not always dynamically best
Sparse	ABC	ABC, followed by Shrinkage	Dynamic evaluation is most informative under sparsity

١٠. Real-world applications to the selection of estimators

In this perspective, ABC would be the most appropriate choice when the sparse transition data are an issue of instability throughout the entire matrix, whilst shrinkage is particularly appealing when the analyst is interested in having a simple, stable and dynamically competitive approach. MLE is a convenient reference point due to its clearness, although the limitations are more pronounced as the transition data get weaker. Bayesian-Dirichlet estimation is superior to MLE in terms of smoothing, but the findings suggest that the proposed estimators have better overall behavior in the contexts that are discussed here. The results of the current paper indicate that the choice of an estimator ought to be based on the practical objective of the analysis. The classic estimators can be satisfactory when the goal is to numerically approximate the transition matrix under rich data. But where the speed of convergence, persistence, long-run stabilization, or system behavior is to be studied with the estimated matrix, dynamic criteria must play a direct role in determining the estimator to be used.

١١. Conclusion Dynamic assessments of transition probability matrix estimators on a dynamic basis were performed in this paper with the element-wise variance as a measure of stability. The above results suggest that the statistical criterion is not enough to provide an overall assessment of the quality of the estimators as even a statistically accurate estimator may have undesirable long-run characteristics. Artificial Bee Colony

estimator has been found to be the most overall effective with regards to performance especially on sparse transition data as opposed to the other four estimators as mentioned. The shrinkage estimator has the second-best overall performance, and in the moderate environment, can in fact exhibit a little better dynamic behavior compared to ABC even though it is not the lowest mean squared error. This implies that statistical superiority does not necessarily have to be dynamic superiority as confirmed through such findings. Research therefore recommends that future applications of transition probability matrix estimators should have dynamic criteria whenever the transition probability matrix is to be used in the analysis of convergence, to provide long-run prediction, or substantive interpretation. This can be generalized to higher-dimensional chains, non-homogeneous Markov models, hidden Markov structures, and more, as well as empirical applications where dynamic interpretability is essential, in future research.

٧.١ Future Research

Research Future The existing framework can develop in various ways in future. One of the methods to do it is to consider more-dimensional transition structures or less connected transition structures, where we might observe even more sensitive convergence behavior to estimator choice. One of them is learning with real-world data, with dynamic interpretability as its emphasis, e.g., sustainability metrics, labor mobility, health trajectory, or reliability systems. The third algorithm is to use hybrid estimators that explicitly find a composite statistical accuracy and dynamic performance goal.

9. References

1. Akay, B., Karaboga, D., Gorkemli, B., & Kaya, E. (2021). A survey on the Artificial Bee Colony algorithm variants for binary, integer and mixed integer programming problems. *Applied Soft Computing*, 106, 107351.
2. Anderson, T. W., & Goodman, L. A. (1957). Statistical inference about Markov chains. *The Annals of Mathematical Statistics*, 28(1), 89–110.
3. Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian Theory*. John Wiley & Sons.
4. Billingsley, P. (1961). Statistical methods in Markov chains. *The Annals of Mathematical Statistics*, 32(1), 12–40.
5. Boas, M. L. (2006). *Mathematical Methods in the Physical Sciences* (3rd ed.). Wiley.
6. Fiecas, M., Ombao, H., van Lunen, D., Baumgartner, R., & Coimbra, A. (2017). Quantifying temporal correlations: A test-retest evaluation of functional connectivity in resting-state fMRI. *NeuroImage*, 65, 231–241.
7. Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). CRC Press.
8. Karaboga, D. (2005). An Idea Based on Honey Bee Swarm for Numerical Optimization. Technical Report TR06, Erciyes University.
9. Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: Artificial Bee Colony algorithm. *Journal of Global Optimization*, 39(3), 459–471.
10. Karaboga, D., Gorkemli, B., Ozturk, C., & Karaboga, N. (2014). A comprehensive survey: Artificial Bee Colony algorithm and applications. *Artificial Intelligence Review*, 42(1), 21–57.
11. Kemeny, J. G., & Snell, J. L. (1976). *Finite Markov Chains*. Springer.
12. Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2), 365–411.
13. Levin, D. A., Peres, Y., & Wilmer, E. L. (2017). *Markov Chains and Mixing Times* (2nd ed.). American Mathematical Society.
14. Minka, T. P. (2000). Estimating a Dirichlet distribution. Technical Report, MIT.
15. Montenegro, R., & Tetali, P. (2006). *Mathematical Aspects of Mixing Times in Markov Chains*. Now Publishers.
16. Neuts, M. F. (1981). *Matrix-Geometric Solutions in Stochastic Models*. Johns Hopkins University Press.
17. Norris, J. R. (1998). *Markov Chains*. Cambridge University Press.
18. Ross, S. M. (2014). *Introduction to Probability Models* (11th ed.). Academic Press.
19. Schäfer, J., & Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation. *Statistical Applications in Genetics and Molecular Biology*, 4(1), Article 32.
20. Seneta, E. (2006). *Non-negative Matrices and Markov Chains*. Springer.

21. Trendelkamp-Schroer, B., Wu, H., Paul, F., & Noé, F. (2015). Estimation and uncertainty of reversible Markov models. *Journal of Chemical Physics*, 143(17), 174101.
22. Zhou, X., Li, Y., & Wang, H. (2024). Advances in shrinkage estimation techniques for high-dimensional stochastic matrices. *Journal of Statistical Theory and Practice*, 18(2), 45–67.