Estimating the transitions of a Markov chain from incompletely observed paths in the presence of predictors

Daphné Aurouet   Valentin Patilea

Univ Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France
{daphne.aurouet, valentin.patilea}@ensai.fr

Résumé. Nous nous intéressons à l'estimation de la matrice de transition d'une chaîne de Markov à temps discret avec un nombre fini d'états en présence de covariables. Les données sont obtenues par la réalisation de trajectoires indépendantes de la chaîne, qui sont observées à des moments aléatoires. Cette étude est motivée par la circulation des billets de banque où une trajectoire correspond à un billet. En posant des hypothèses appropriées d'indépendance, nous construisons des estimateurs simples de la matrice de transition conditionnelle, sachant les valeurs des covariables, à l'aide de l'estimateur empirique et des racines ℓ-ièmes des matrices de transition. L'effet de prédicteurs continus est pris en compte par un lissage à noyau et celui des covariables discrètes par une stratification. L'estimateur conditionnel de la matrice de transition peut être facilement appliqué à des jeux de données volimineux et transmis en continu, et celui-ci requiert peu de ressources informatiques. La performance de notre approche est illustrée sur des données simulées.

Mots-clés. Circulation des billets, Lissage à noyau, Chaînes de Markov, Racine $\ell-$ième d'une matrice stochastique

Abstract. We are interested in estimating the transition matrix of a discrete time, finite Markov chain in the presence of covariates. The data points are obtained from independent sample paths of the chain which are observed at random times. This study was motivated by the circulation of banknotes where a sample path corresponds to a banknote. Based on suitable independence assumptions, we build simple conditional transition matrix estimates, given the covariate vector value, using empirical estimators of transition probabilities and $\ell-$th roots of the transition matrices. The effect of continuous covariates is accounted for by kernel smoothing and the one of discrete covariates via stratification. The conditional transition matrix estimator can be easily applied to large streaming data and requires low computer resources. The performance of our approach is illustrated using simulated data.

Keywords. Banknote circulation, Kernel smoothing, Markov chains, $\ell-$root of a stochastic matrix

1 Introduction

In this work, we consider discrete-time Markov chains with finite state space. Such models are commonly found in real-world applications such as statistical mechanics (Seneta, 2016), finance (Israel et al., 2001), predictive maintenance (Tamaloussi and Bouzaouit, 2020) *etc.* Their appeal comes from their memory-less property, ease of interpretation, and simple computations. The systems under consideration in this study are homogeneous, irreducible, and aperiodic, but are observed at random times only. The last characteristic has been little studied in theory, although it is commonly encountered in practice.

Observing the state of a Markovian system at every unit of time, as is usually assumed for Markov chains, can be costly or impossible. We therefore take a more pragmatic approach and allow observations to occur at random discrete times. Consequently, some jumps between successive observations may occur and go unobserved. We also assume that the system is influenced by additional external factors (or covariates), gathered in a vector with continuous or discrete components that do not change over time. Conditionally on the value of the covariate vector, the Markov chain is homogeneous. Our objective is thus to estimate the conditional transition matrix from incompletely observed sample paths of a discrete-time Markov chain given the values of the covariate vector. The lengths of the sample paths of the chain are bounded, but for the applications we have in mind, we can consider that a large number of independent sample paths are generated. While this setting looks similar to discrete-time semi-Markov chains, it adds the difficulty of unobserved transitions.

A related problem has been studied by Barsotti et al. (2014). The authors considered a method to build an estimator of the (unconditional) transition matrix under the assumption that it has some zero entries. Here, in the presence of covariates, we propose a kernel smoothing estimator that remains simple and flexible. Our method is constructed at the cost of more restrictive assumptions on the observation mechanism. However, our assumptions appear to be well suited to the study of the banknote circulation. We leverage the homogeneity of our process given the covariate, and consider the $\ell-$power of the conditional transition matrix. The elements of this matrix can be simply estimated by smooth empirical probability estimators, yielding a matrix estimate which is a stochastic matrix. Under certain conditions, we can then compute the $\ell-$root of the estimated $\ell-$power of the conditional transition matrix. The existence and stochasticity of the $\ell-$root is a challenging aspect of the proposed approach, which will be pointed out below. It is related to a well known, and still open problem in probability theory and matrix algebra. Several positive integer values ℓ can be considered, and we propose to aggregate the resulting conditional transition matrix estimators by a weighted average.

Section 2 details the model framework and necessary assumptions, and Section 3 introduces the proposed estimator. Section 4 describes the performance of a finite sample.

2 Markov Chains with randomly observed paths

Let $\mathbf{X} = (X_t)_{t\geq 1}$ by a Markov chain with values in the finite set $\mathcal{S} = \{1, 2, \ldots, S\}$. Let $\mathbf{Z}$ be a random vector of covariates (predictors) which for simplicity we suppose independent of the time t.

2.1 The model

The covariate vector can be decomposed into a sub-vector of continuous predictors $\mathbf{Z}^c \in \mathbb{R}^{\boldsymbol{d}_c}$ and a sub-vector of discrete ones $\mathbf{Z}^d \in \mathbb{R}^{\boldsymbol{d}_d}$. In the following, $\mathbf{z} = (\mathbf{z}^c, \mathbf{z}^d) \in \mathbb{R}^{\boldsymbol{d}_c} \times \mathbb{R}^{\boldsymbol{d}_d}$ is any value in the support of $\mathbf{Z}$. We impose the following assumption on the conditional distribution of $\mathbf{X}$ given the covariate vector value.

(H1) Conditionally given $\mathbf{Z} = \mathbf{z}$, the discrete time process $\mathbf{X}$ is a homogeneous, irreducible and aperiodic Markov chain with transition probabilities

$$p_{ij}(\mathbf{z}) = \mathbb{P}(X_{t+1} = j \mid X_t = i, \mathbf{Z} = \mathbf{z}), \qquad i, j \in \mathcal{S}.$$

Let $\boldsymbol{P}(\mathbf{z})$ be the conditional transition matrix $\mathbf{X}$ given $\mathbf{Z} = \mathbf{z}$.

For the applications we have in mind, we observe several independent paths of $\mathbf{X} = (X_t)_{t\geq 1}$, and for each path we also observe a random draw of $\mathbf{Z}$. Unfortunately, the independent paths of $\mathbf{X}$ are incompletely observed. More precisely, instead of being observed at each integer $t \geq 1$, a sample paths is observed at the integer random times $1 \leq T_0 < T_1 < \ldots < T_k < T_{k+1} < \ldots \leq t_{\max}$. The non-random integer $t_{\max}$ defines the observation window for the transitions of a sample paths, and this value is given. Let

$$\tau_k = T_k - T_{k-1} \qquad \text{and} \qquad Y_k = X_{T_k}, \qquad k \geq 1.$$

The observed sample consists of independent realizations of $\mathbf{Y} = (Y_k, \tau_k, \mathbf{Z})_{k\geq 1}$, observed as long as $\tau_1 + \cdots + \tau_k \leq t_{\max}$. Therefore, we can impose $\tau_k \in \{1, 2, \ldots, L\}$ for some integer $L \geq 1$. Some additional assumptions on the distribution of $\mathbf{Y}$ are considered.

(H2) For any $i, j \in \mathcal{S}$, and $\ell, \ell' \in \{1, 2, \ldots, L\}$,

$$\begin{aligned}\mathbb{P}(Y_{k+1} = j, \tau_{k+1} = \ell \mid Y_k = i, \tau_k = \ell', \mathbf{Z} = \mathbf{z}, Y_{k-1}, \tau_{k-1}, \ldots, Y_1, \tau_1, T_0)&\\ = \mathbb{P}(Y_{k+1} = j, \tau_{k+1} = \ell \mid Y_k = i, \mathbf{Z} = \mathbf{z}), \qquad \forall k \geq 1.&\end{aligned} \tag{1}$$

(H3) For any $i, j \in \mathcal{S}$ and $t' > t \geq 1$,

$$\mathbb{P}(X_{T_{k+1}} = j \mid X_{T_k} = i, T_{k+1} = t', T_k = t, \mathbf{Z} = \mathbf{z}) = \left(\boldsymbol{P}^{t'-t}(\mathbf{z})\right)_{ij}, \qquad \forall k \geq 1. \tag{2}$$

Assumption (1) imposes a lack-of-memory condition for $(Y_k, \tau_k)_{k\geq 1}$, given $\mathbf{Z} = \mathbf{z}$. In particular, it implies that conditionally given $\mathbf{Z} = \mathbf{z}$, the process $(Y_k, \tau_k)_{k\geq 1}$ is a homogeneous Markov chain with a finite state space. Condition (2) can hold true even when the random times (T_k) are not independent of $\mathbf{X}$.

Using (1) and (2) with $\ell = t' - t$, we get,

$$\begin{aligned}\mathbb{P}(Y_{k+1} = j \mid Y_k = i, \tau_{k+1} = \ell, \mathbf{Z} = \mathbf{z}) &= \mathbb{P}(Y_{k+1} = j \mid Y_k = i, \tau_{k+1} = \ell, T_k, \mathbf{Z} = \mathbf{z})\\ &= \mathbb{P}(X_{T_k+\ell} = j \mid X_{T_k} = i, T_{k+1} = T_k + \ell, T_k, \mathbf{Z} = \mathbf{z}) = \left(\boldsymbol{P}^{\ell}(\mathbf{z})\right)_{ij}.\end{aligned}$$

Finally, we obtain

$$\mathbb{P}(Y_{k+1} = j, \tau_{k+1} = \ell \mid Y_k = i, \mathbf{Z} = \mathbf{z}) = \left(\boldsymbol{P}^\ell(\mathbf{z})\right)_{ij} \mathbb{P}(\tau_{k+1} = \ell \mid Y_k = i, \mathbf{Z} = \mathbf{z}). \tag{3}$$

From (3), we rewrite for $i, j \in \mathcal{S}$, $\mathbf{z} = (\mathbf{z}_c, \mathbf{z}_d)$ and $1 \leq \ell \leq L$,

$$\boldsymbol{P}^\ell(\mathbf{z}) = \boldsymbol{A}_\ell(\mathbf{z}) \quad \text{where} \quad (\boldsymbol{A}_\ell(\mathbf{z}))_{ij} = \frac{\mathbb{P}(Y_{k+1} = j, \tau_{k+1} = \ell \mid Y_k = i, \mathbf{Z} = \mathbf{z})}{\mathbb{P}(\tau_{k+1} = \ell \mid Y_k = i, \mathbf{Z} = \mathbf{z})}.$$

The matrix $\boldsymbol{A}_\ell(\mathbf{z})$ is thus the $\ell-$th power of the conditional transition matrix we want to estimate. It is worth noting that $\boldsymbol{A}_\ell(\mathbf{z})$ is a stochastic matrix, that is a square matrix with non-negative elements with each row summing to 1. Moreover, only observable variables are used to define the elements of $\boldsymbol{A}_\ell(\mathbf{z})$. A natural idea is then to estimate $\boldsymbol{A}_\ell(\mathbf{z})$ and next define an estimator of $\boldsymbol{P}(\mathbf{z})$ as the $\ell-$roots of $\boldsymbol{A}_\ell(\mathbf{z})$, provided that such matrix root exists and is a stochastic matrix.

2.2 The transition matrix and the $\ell-$roots of stochastic matrices

For any squared matrix $\boldsymbol{Q}$, the series $\sum_{k=0}^{\infty} \boldsymbol{Q}^k/k!$, converges with respect to any matrix norm. The limit is denoted $\exp(\boldsymbol{Q})$ and defines the exponential of $\boldsymbol{Q}$.

Let $\mathbf{z}$ be fixed. If the matrix $\boldsymbol{Q}$ is such that $\boldsymbol{P}(\mathbf{z}) = \exp(\boldsymbol{Q})$, then $\boldsymbol{P}^\ell(\mathbf{z}) = \exp(\ell\boldsymbol{Q}) = \boldsymbol{A}_\ell(\mathbf{z})$, for all $\ell \geq 1$. Moreover, by definition $\ell\boldsymbol{Q}$ is a logarithm of $\boldsymbol{A}_\ell(\mathbf{z})$, denoted $\log(\boldsymbol{A}_\ell(\mathbf{z}))$. Conversely, we can write

$$\boldsymbol{P}(\mathbf{z}) = \exp(\{\ell\boldsymbol{Q}\}/\ell) = \exp(\log(\boldsymbol{A}_\ell(\mathbf{z}))/\ell). \tag{4}$$

See, for example, Norris (1998) and Higham (2008) for the formal definitions of the exponential and logarithm of a matrix. The existence of the matrix $\boldsymbol{Q}$ is a classical but not completely solved problem in matrix algebra and numerical analysis, related to the so-called embedding problem in Markov chains theory. See Kingman (1962).

To discuss the existence of the representation (4) for the conditional transition matrix, we can start by searching conditions under which $\log(\boldsymbol{A}_\ell(\mathbf{z}))$ is well defined. Let $\boldsymbol{A}$ be a real matrix. The following properties are stated in (Higham, 2008, Theorem 1.31 and 11.2).

- If $\boldsymbol{A}$ has no eigenvalues in $(-\infty, 0]$, there exists a unique logarithm of A which is real. This matrix, say $\boldsymbol{X}$, is referred to as the *principal logarithm of* $\boldsymbol{A}$ and $\exp(\boldsymbol{X}) = \boldsymbol{A}$.
- If $\boldsymbol{A}$ has no eigenvalues in $(-\infty, 0]$, then for any $\alpha \in [-1, 1]$ we have $\log(\boldsymbol{A}^\alpha) = \alpha \log(\boldsymbol{A})$.

Since the exponential of a squared matrix always converge, the remaining and most difficult problem is to guarantee that $\exp(\log(\boldsymbol{A})/\ell)$ is a stochastic matrix. Several sufficient conditions have been provided, see for example Cuthbert (1972, 1973), Higham and Lin (2011) and the references therein.

In addition to the problem of the validity of the representation (4) for the theoretical matrix $\boldsymbol{P}(\mathbf{z})$, there is also the question whether $\exp(\log(\boldsymbol{A}_\ell(\mathbf{z}))/\ell)$ remains a stochastic matrix when $\boldsymbol{A}_\ell(\mathbf{z})$ is replaced by an estimator. Israel et al. (2001) and Bladt and Sørensen (2005) address this aspect and provide some partial answers and remedies.

The problem becomes even more complex in our context in the presence of covariates. For some values $\mathbf{z}$ the matrix $\exp(\log(\boldsymbol{A}_\ell(\mathbf{z}))/\ell)$ may be a stochastic matrix when $\boldsymbol{A}_\ell(\mathbf{z})$ is replaced by an estimate, but this may not happen for other values $\mathbf{z}$.

3 Estimating the conditional transition matrix

The sample consists of the data points corresponding to N independent sample paths of $(Y_k, \tau_k)_{k\geq 1}$ and N independent realizations of the covariate vector $\mathbf{Z}$. More precisely, for each $1 \leq m \leq N$, the vectors $(Y_{m,k}, \tau_{m,k}, \mathbf{Z}_m)$, $1 \leq k \leq M_m$, are observed. The integers M_m are bounded by L.

Let us fix $1 \leq \ell \leq L$. For any $\mathbf{z} = (\mathbf{z}^c, \mathbf{z}^d)$ in the support of $\mathbf{Z}$, we define the estimator of the element (i,j) of the matrix $\boldsymbol{A}_\ell(\mathbf{z})$ as

$$\left(\widehat{\boldsymbol{A}}_\ell(\mathbf{z})\right)_{ij} = \frac{\sum_{m=1}^N \sum_{k=1}^{M_m} \mathbb{1}\left\{Y_{m,k+1} = j, \tau_{m,k+1} = \ell, Y_{m,k} = i, \mathbf{Z}_m^d = \mathbf{z}^d\right\} \boldsymbol{K}_{\boldsymbol{h}}\left(\mathbf{Z}_m^c - \mathbf{z}^c\right)}{\sum_{m=1}^N \sum_{k=1}^{M_m} \mathbb{1}\left\{\tau_{m,k+1} = \ell, Y_{m,k} = i, \mathbf{Z}_m^d = \mathbf{z}^d\right\} \boldsymbol{K}_{\boldsymbol{h}}\left(\mathbf{Z}_m^c - \mathbf{z}^c\right)}. \quad (5)$$

The rule $0/0 = 0$ applies. Here, for any $\boldsymbol{u} \in \mathbb{R}^{\boldsymbol{d}_c}$ with components $u^{(1)}, \ldots, u^{(\boldsymbol{d}_c)}$, we define

$$\boldsymbol{K}_{\boldsymbol{h}}(\boldsymbol{u}) = \prod_{l=1}^{\boldsymbol{d}_c} K(u^{(l)}/h^{(l)}),$$

with $\boldsymbol{h} = (h^{(1)}, \ldots, h^{(\boldsymbol{d}_c)})$ a vector of bandwidths and $K(\cdot)$ an univariate kernel, for instance the standard Gaussian density. The components of the bandwidth vector $\boldsymbol{h}$ are defined as

$$h^{(l)} = \{M_1 + \cdots + M_N\}^{-\alpha}\, \widehat{\sigma}_l, \qquad l = 1, \cdots, \boldsymbol{d}_c, \qquad \text{with} \qquad \alpha = 1/(4 + \boldsymbol{d}_c), \quad (6)$$

and $\widehat{\sigma}_l$ the empirical standard deviation of the $l-$th component of the sub-vector $\mathbf{Z}^c$ gathering the continuous components of the covariate vector.

Assume that $\widehat{\boldsymbol{A}}_\ell(\mathbf{z})$ has no eigenvalues in $(-\infty, 0]$, and there exists a unique logarithm of $\widehat{\boldsymbol{A}}_\ell(\mathbf{z})$ which is a matrix with real elements. Then we can define an estimator of $\widehat{\boldsymbol{P}}(\mathbf{z})$ as the $\ell-$root of $\widehat{\boldsymbol{A}}_\ell(\mathbf{z})$ computed as

$$\widehat{\boldsymbol{P}}_\ell(\mathbf{z}) = \exp\left\{\frac{1}{\ell} \log\left(\widehat{\boldsymbol{A}}_\ell(\mathbf{z})\right)\right\}.$$

Whenever this estimator $\widehat{\boldsymbol{P}}(\mathbf{z})$ exists and is a stochastic matrix, we get an estimator of the conditional transition matrix $\boldsymbol{P}(\mathbf{z})$ given $\mathbf{Z} = \mathbf{z}$ based on the transitions observed after ℓ periods of time. In order to better exploit the information carried by the sample, we can consider several ℓ values, typically those with the largest frequencies. Let $[\underline{L}, \overline{L}]$ a range of ℓ and let

$$\widehat{\boldsymbol{P}}(\mathbf{z}) = \frac{1}{\sum_{\ell=\underline{L}}^{\overline{L}} \widehat{\pi}_l} \sum_{\ell=\underline{L}}^{\overline{L}} \widehat{\pi}_l \widehat{\boldsymbol{P}}_\ell(\mathbf{z}),$$

be an aggregated estimator of the conditional transition matrix $\boldsymbol{P}(\mathbf{z})$ given $\mathbf{Z} = \mathbf{z}$. Here, $\widehat{\pi}_l$ is the empirical estimator of the marginal probability $\mathbb{P}(\tau_k = \ell)$.

It is worth noting that our estimation approach allows for very large datasets, in particular for streaming data. More precisely, few more observations can be added on each sample path, and, more important, many other sample paths can observed. The numerator and the denominator in (5) can be easily updated in such situations, without low memory resources and computational complexity. The bandwidth rule (6) guarantees the necessary gradual decrease with the new observations.

The asymptotic behavior of our estimators $\widehat{\boldsymbol{P}}_\ell(\mathbf{z})$, and the aggregated version $\widehat{\boldsymbol{P}}(\mathbf{z})$ can be derived from the asymptotic behavior of the estimators $\widehat{\boldsymbol{A}}_\ell(\mathbf{z})$. The asymptotic of these latter estimators can be studied by the standard tools used for the kernel regression with discrete responses. Theoretical grounds for the recursive versions of $\widehat{\boldsymbol{A}}_\ell(\mathbf{z})$, which are better suited with streaming data, can be obtained from the existing theoretical results on recursive kernel regression.

4 Empirical evidence

To construct our simulation design, we first consider a pilot stochastic matrix $\boldsymbol{B}$ for which the $\ell-$roots exist and are stochastic. The conditional transition matrix $\boldsymbol{P}(\mathbf{z})$ given $\mathbf{Z} = \mathbf{z}$ is then constructed by perturbing the pilot matrix $\boldsymbol{B}$, with the perturbation depending on $\mathbf{z}$. The covariate vector $\mathbf{Z}$ has up to three components. The discrete component of $\mathbf{Z}$ is a Bernoulli variable with parameter $p = 0.7$, while the continuous components are generated using a Beta distribution.

Independently for each $1 \leq m \leq N$, we draw the initial value $Y_{m,1}$ from a discrete uniform distribution over $\mathcal{S}$. Next, given $Y_{m,k}$, the $\tau_{m,k+1}$ is obtained as a random draw from a Poisson distribution with parameter λ to which we add the value 1. The parameter λ depends on the value : it is equal to 10 if $Y_{m,k} \in \{1, 2\}$ and equal to 15 if $Y_{m,k} \in \{3, \cdots, S\}$. Finally, given $\mathbf{Z}^{(m)} = \mathbf{z}$, $\tau_{m,\ell} = \ell$ and $Y_{m,k} = i$, we draw $Y_{m,k+1}$ from a multinomial distribution with the vector of parameters equal to the $i-$th row of the matrix $\boldsymbol{P}^\ell(\mathbf{z})$. The procedure is repeated M_m steps where M_m is the smallest integer such that $\sum_{k=1}^{M_m} \tau_{m,k} \geq t_{\max}$. We set $t_{\max} = 20$ in our experiences. The performance of the estimators of the conditional transition matrices is evaluated with the spectral norm of the error $\widehat{\boldsymbol{P}}(\mathbf{z}) - \boldsymbol{P}(\mathbf{z})$ (for any matrix $\boldsymbol{U}$, the square of the spectral norm is defined by $\|\boldsymbol{U}\|_2^2 = \lambda_{\max}\left(\boldsymbol{U}^\top \boldsymbol{U}\right)$, where for any positive semi-definite matrix $\boldsymbol{C}$, $\lambda_{\max}(\boldsymbol{C})$ is the largest eigenvalue of $\boldsymbol{C}$).

Table 1 and Figure 1 illustrate the performance of the estimator over 200 replications of experiments on different sample and state space sizes and different values of the predictor vector. It reveals the convergence of our estimator towards the true transition matrix for most of the different scenarios tested. In very few cases, for some covariate vector values $\mathbf{z}$, the consistency does not seem guaranteed, most likely because of the $\ell-$roots of the empirical estimates of the matrices $\widehat{\boldsymbol{A}}_\ell(\mathbf{z})$ do not exist. See also the discussion in the section 2.2 above.

For larger state spaces, the number of parameters to be estimated is larger, and hence the performance is poorer. It is worth noting that a better accuracy for the estimators is achieved when the diagonal of the matrix $\boldsymbol{P}(\mathbf{z})$ is to a large extent dominant.

Figure 1: Median of the spectral norm of the errors for the multivariate predictor case, over different sample sizes from $N = 200$ to $N = 32400$, for a Markov chain with $S = 3$ states. The vector $\mathbf{Z}$ has a binary component z_d and two continuous components $z_{c,1}$ and $z_{c,2}$. The value $t_{\max}$ is set equal to 20. Results obtained from $R = 200$ replications.

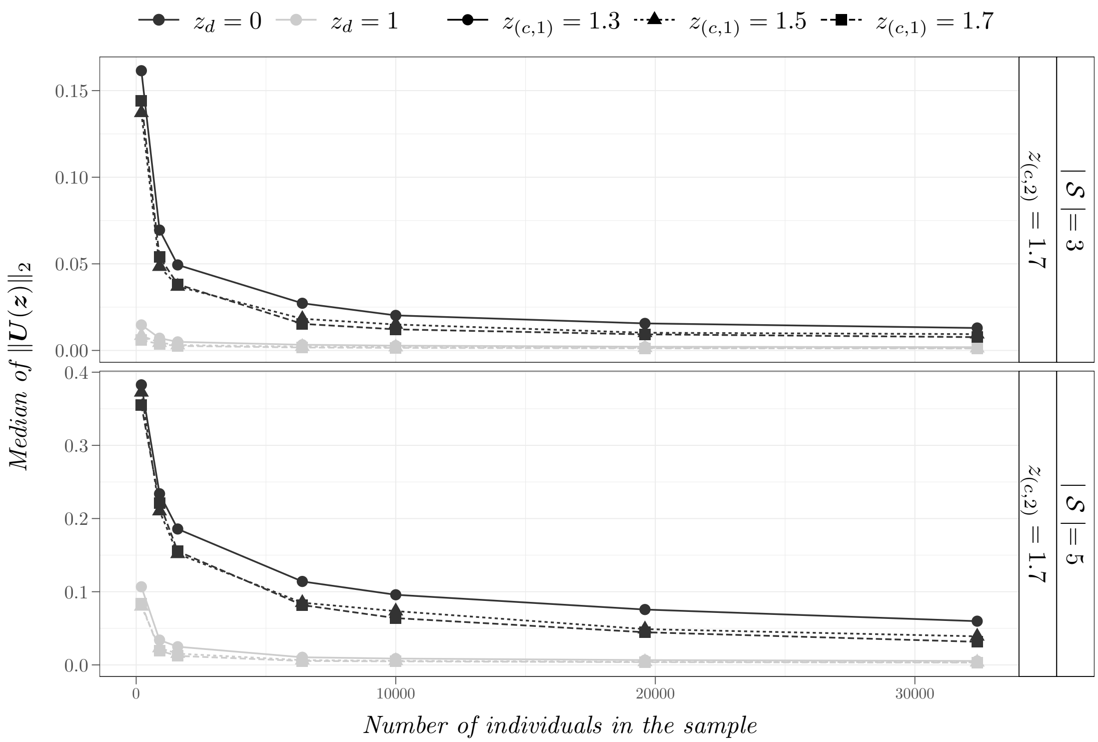

Table 1: Median of the spectral norm of the errors for the multivariate predictor case, over different sample sizes, for a Markov chain with $S \in \{3, 5, 9\}$ states. The cases of no predictor, one predictor, and three predictors (where $\mathbf{Z}$ has two continuous components and a binary component z_d) are considered. The value $t_{\max}$ is set equal to 20. Results obtained from $R = 200$ replications.

			N						
Predictors	S	$\mathbf{z}$	**200**	**900**	**1600**	**6400**	**10000**	**19600**	**32400**
No predictor	**3**		0.0675	0.0209	0.0135	0.0060	0.0048	0.0033	0.0023
	5		0.1830	0.0800	0.0549	0.0184	0.0132	0.0086	0.0062
	9		0.4915	0.2025	0.1560	0.0784	0.0636	0.0433	0.0330
One predictor	**3**	(1.3)	0.1096	0.0755	0.0702	0.0493	0.0469	0.0396	0.0333
		(1.5)	0.0793	0.0491	0.0442	0.0336	0.0331	0.0262	0.0209
		(1.7)	0.0808	0.0518	0.0413	0.0270	0.0218	0.0158	0.0126
	5	(1.3)	0.2053	0.1790	0.1696	0.1533	0.1499	0.1397	0.1328
		(1.5)	0.1648	0.1025	0.0958	0.0836	0.0779	0.0715	0.0690
		(1.7)	0.2184	0.1117	0.0969	0.0745	0.0698	0.0692	0.0594
	9	(1.3)	0.4493	0.3414	0.3468	0.3408	0.3268	0.3325	0.3246
		(1.5)	0.3971	0.2239	0.1980	0.1876	0.1876	0.1751	0.1764
		(1.7)	0.4785	0.1699	0.1408	0.1193	0.1157	0.1093	0.1132
Three predictor	**3**	$(1.3, 1.3, 0)$	0.1754	0.0730	0.0532	0.0325	0.0265	0.0199	0.0174
		$(1.3, 1.3, 1)$	0.0205	0.0096	0.0077	0.0047	0.0041	0.0034	0.0029
		$(1.3, 2.9, 0)$	0.1467	0.0699	0.0446	0.0131	0.0112	0.0078	0.0065
		$(1.3, 2.9, 1)$	0.0020	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005
		$(1.7, 1.3, 0)$	0.1452	0.0568	0.0427	0.0207	0.0148	0.0118	0.0102
		$(1.7, 1.3, 1)$	0.0083	0.0053	0.0043	0.0029	0.0022	0.0020	0.0017
		$(1.7, 1.7, 0)$	0.1440	0.0540	0.0382	0.0154	0.0122	0.0093	0.0077
		$(1.7, 1.7, 1)$	0.0059	0.0037	0.0026	0.0017	0.0015	0.0012	0.0011
		$(1.7, 2.9, 0)$	0.1258	0.0491	0.0183	0.0072	0.0055	0.0046	0.0040
		$(1.7, 2.9, 1)$	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
	5	$(1.3, 1.3, 0)$	0.3478	0.2313	0.1890	0.1204	0.1088	0.0940	0.0870
		$(1.3, 1.3, 1)$	0.1128	0.0510	0.0391	0.0156	0.0126	0.0091	0.0080
		$(1.3, 1.7, 0)$	0.3827	0.2339	0.1857	0.1140	0.0958	0.0757	0.0597
		$(1.3, 1.7, 1)$	0.1067	0.0340	0.0248	0.0104	0.0086	0.0066	0.0054
		$(1.3, 2.9, 0)$	0.3526	0.2329	0.1725	0.0782	0.0614	0.0359	0.0268
		$(1.3, 2.9, 1)$	0.0817	0.0128	0.0080	0.0012	0.0012	0.0011	0.0013
		$(1.7, 1.3, 0)$	0.3412	0.2464	0.1770	0.0978	0.0809	0.0582	0.0448
		$(1.7, 1.3, 1)$	0.0896	0.0316	0.0187	0.0081	0.0070	0.0055	0.0044
		$(1.7, 1.7, 0)$	0.3551	0.2209	0.1553	0.0818	0.0639	0.0446	0.0315
		$(1.7, 1.7, 1)$	0.0834	0.0193	0.0123	0.0054	0.0048	0.0037	0.0031
		$(1.7, 2.9, 0)$	0.3265	0.1804	0.1335	0.0607	0.0387	0.0210	0.0132
		$(1.7, 2.9, 1)$	0.0676	0.0085	0.0009	0.0005	0.0005	0.0005	0.0005

References

Barsotti, F., De Castro, Y., Espinasse, T., and Rochet, P. (2014). Estimating the transition matrix of a Markov chain observed at random times. *Statistics & Probability Letters*, 94:98–105.

Bladt, M. and Sørensen, M. (2005). Statistical inference for discretely observed Markov jump processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(3):395–410.

Cuthbert, J. R. (1972). On uniqueness of the logarithm for Markov semi-groups. *J. London Math. Soc. (2)*, 4:623–630.

Cuthbert, J. R. (1973). The logarithm function of finite-state Markov semi-groups. *J. London Math. Soc. (2)*, 6:524–532.

Higham, N. J. (2008). *Functions of matrices*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Theory and computation.

Higham, N. J. and Lin, L. (2011). On $p-$th roots of stochastic matrices. *Linear Algebra and its Applications*, 435(3):448–463.

Israel, R. B., Rosenthal, J. S., and Wei, J. Z. (2001). Finding generators for markov chains via empirical transition matrices, with applications to credit ratings. *Mathematical Finance*, 11(2):245–265.

Kingman, J. F. C. (1962). The imbedding problem for finite Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 1:14–24.

Norris, J. R. (1998). *Markov chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

Seneta, E. (2016). Markov chains as models in statistical mechanics. *Statistical Science*, 31(3):399–414.

Tamaloussi, N. and Bouzaouit, A. (2020). Study of reliability in a repairable system by markov chains. *Acta Universitatis Sapientiae, Electrical and Mechanical Engineering*, 12(1):66–76.