Singular value distribution of random matrices with block Markovian dependence

Sanders, Jaron and Van Werde, Alexander
Department of Mathematics and Computer Science

Sequential data, such as text or DNA, are naturally modelled by Markov chains. There has been recent interest in spectral algorithms for community detection in Markov chains [1, 2]. For instance, one can use the singular value decomposition of the empirical transition matrix $\widehat{P}$. The resulting reduction in the size of the state space then allows for extraction of actionable insight from the data.

Block Markov chains

A *Block Markov chain* (BMC) is the Markov chain analogue of the classical stochastic block model [3]. Let $\mathcal{V} := \{1, \dots, n\}$ be a large state space with partition $\mathcal{V} =: \mathcal{V}_1 \cup \dots \cup \mathcal{V}_K$ and consider a $K \times K$ transition matrix p of an irreducible acyclic Markov chain on $\{1, \dots, K\}$. A BMC with cluster transition matrix p is a Markov chain with transition matrix $P_{x,y} = p_{i,j}/\#\mathcal{V}_j$ for $(x,y) \in \mathcal{V}_i \times \mathcal{V}_j$. We let n tend to infinity and consider a BMC sample path $X := (X_t)_{t=0}^{\ell}$ with $\ell = \lambda n^2 + o(n^2)$ and $\#\mathcal{V}_j = \alpha_j n + o(n)$.

Associated to the sample path X we have two $n \times n$ random matrices, denoted $\widehat{N}$ and $\widehat{P}$, given by:

$$\widehat{N}_{ij} = \text{Number of transitions } i \to j; \quad \widehat{P}_{ij} = \widehat{N}_{ij} / \sum_{k=1}^{n} \widehat{N}_{ik}.$$

It has recently been shown that $\widehat{N}$ has K informative singular values of size $\Theta(\ell/n)$ whereas the remaining bulk has magnitude $O(\sqrt{\ell/n})$ [4].

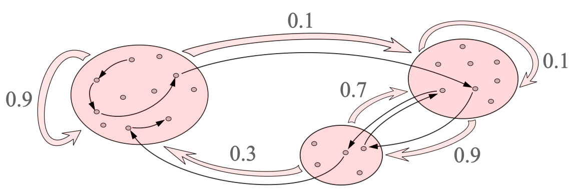

Figure 1: A BMC with cluster transition matrix $p = [[0.9, 0.1, 0], [0, 0.1, 0.9], [0.3, 0, 0.7]]$.

Dependence and variance profiles

We are concerned with asymptotics of the bulk of the singular values of $\widehat{N}$ and $\widehat{P}$. These random matrices introduce two main challenges:

1. The entries are dependent.
2. The entries do not have identical variance.

In this context we define a class of dependent random matrices, called *almost uncorrelated random matrices with a variance profile*, whose asymptotic spectral distributions can be determined using the moment method. This generalizes a result in [5] to include the possibility of variance profiles.

Coupling argument

Establishing limiting singular value distributions for $\widehat{N}$ and $\widehat{P}$ may be reduced to the statement that $M := \widehat{N} - \mathbb{E}[\widehat{N}]$ is approximately uncorrelated. A coupling argument is used to construct a pair of chains (X, Y) with $M_{X,i_1 i_2}$ independent of Y. Then,

$$\begin{aligned}\mathbb{E}[M_{X,i_1 i_2}^{m_1} \cdots M_{X,i_R j_R}^{m_R}] &\approx \mathbb{E}[M_{X,i_1 j_1}^{m_1} M_{Y,i_2 j_2}^{m_2} \cdots M_{Y,i_R j_R}^{m_R}] \\ &= \mathbb{E}[M_{X,i_1 j_1}^{m_1}]\mathbb{E}[M_{Y,i_2 j_2}^{m_2} \cdots M_{Y,i_R j_R}^{m_R}]\end{aligned}$$

which is precisely what it means to be approximately uncorrelated.

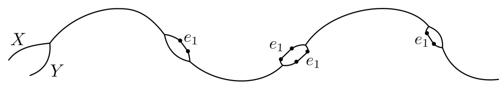

Figure 2: Construction of the pair (X, Y): the chains are allowed to diverge for a short period of time whenever either one uses $e_1 := i_1 j_1$.

Main result: singular values of $\widehat{P}$ and $\widehat{N}$

Theorem. *The empirical singular value distribution $\nu_{\sqrt{n}\widehat{P}}$ converges weakly in probability to a measure ν whose symmetrization $\mathrm{sym}(\nu)$ has Stieltjes transform $s(z) = \sum_{i=1}^{K} \alpha_i (a_i(z) + a_{K+i}(z))/2$ where the $a_i(z)$ satisfy the following system of equations*

$$a_i(z)^{-1} = z - \sum_{j=1}^{K} \lambda^{-1} \pi(i)^{-1} \alpha_i p_{i,j} a_{K+j}(z)$$

$$a_{i+K}^{-1}(z) = z - \sum_{j=1}^{K} \lambda^{-1} \pi(j)^{-1} \alpha_i^{-1} \alpha_j^2 p_{j,i} a_j(z)$$

for $i = 1, \dots, K$. Here π is the equilibrium distribution of p. A similar system of equations with different coefficients describes the Stieltjes transform of $\nu_{\widehat{N}/\sqrt{n}}$.

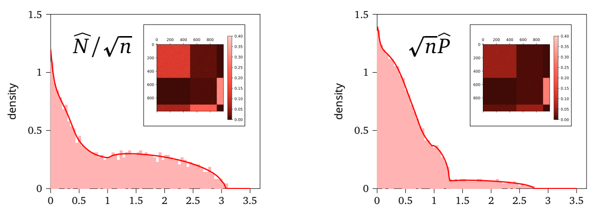

Figure 3: $\widehat{N}/\sqrt{n}$ and $\sqrt{n}\widehat{P}$ and frequency-based histograms of their singular values when $\lambda = 2$, $\alpha = (0.5, 0.4, 0.1)$ and $n = 1000$ with p as in Figure 1. Observe that our theoretical result, depicted as the continous curve, matches well with the empirical distribution.

Acknowledgements

This work is part of the project *Clustering and Spectral Concentration in Markov Chains* with project number OCENW.KLEIN.324 of the research programme *Open Competition Domain Science – M* which is partly financed by the Dutch Research Council (NWO).

References

[1] Sanders, J. & Proutière, A. & Yun, S.Y. (2020) Clustering in block Markov chains *The Annals of Statistics*, 48(6), 3488-3512
[2] Zhang, A. & Wang, M. (2019). Spectral state compression of Markov processes. *IEEE transactions on information theory*, 66(5), 3202-3231.
[3] Holland, P. W. & Laskey, K. B. & Leinhardt, S. (1983). Stochastic blockmodels: First steps. Social networks, 5(2), 109-137.
[4] Sanders, J., & Senen-Cerda, A. (2021). Spectral norm bounds for block Markov chain random matrices. *arXiv preprint* arXiv:2111.06201.
[5] Hochstättler, W. & Kirsch, W., & Warzel, S. (2016). Semicircle law for a matrix ensemble with dependent entries. *Journal of Theoretical Probability*, 29(3), 1047-1068.