Non-parametric copula-based estimator of transition matrices for high-order Markov chains

Xiaoxiao Ma* Kaveh Salehzadeh Nobari†

University of Leeds Durham University

September 19, 2020

This is a preliminary draft. Please do not cite or distribute without the permission of the authors.

*Institute for Transport Studies (ITS), University of Leeds. Address and email: Institute for Transport Studies (ITS), University of Leeds, 34-40 University Road, Leeds, LS2 9JT

†Department of Economics and Finance, Durham University. Address and email: Department of Economics and Finance, Durham University Business School, Mill Hill Lane, Durham, DH1 3LB.

Non-parametric copula-based estimator of transition matrices for high-order Markov chains

ABSTRACT

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Journal of Economic Classification: C12, C13, C14

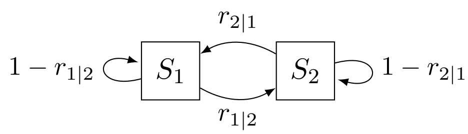
Keywords: Causality measures; Non-parametric estimation; Vine decomposition; Markov chains; Bernstein copula density; Local bootstrap.

Contents

1 Introduction

2 Motivation

Consider the simple case of a two-state first-order stationary Markov chain, such as the one depicted below



with an *unknown* transition matrix

$$\mathbf{P} = \begin{bmatrix} 1 - r_{1|2} & r_{1|2} \\ r_{2|1} & 1 - r_{2|1} \end{bmatrix}$$

where

$$r_{i|j} = P[X_t = i \mid X_{t-1} = j], \quad i, j \in \{1, 2\} \tag{1}$$

In such scenarios, estimating the conditional probabilities $r_{i|j}$ is rather straightforward. Consider observing a sequence of states for individual entities at times $t-1$ and t. The probability of transitioning from one state to another can be estimated by calculating the ratio of the entities that transitioned from, say, $S1$ to $S2$ in the period $t-1$ to t to the total number of entities that began at state $S1$ at $t-1$. In other words:

$$r_{i|j} = \frac{n_{ij}}{\sum_j n_{ij}} \tag{2}$$

Now, instead let us consider that the Markov chain under consideration is of order p with a transition probability

$$r_{j_0|j_1,\cdots,j_p} = P[X_t = j_0 \mid X_{t-1} = j_1, \cdots, X_{t-p} = j_p], \quad j_k \in \{1, 2\}, \quad \text{for } k = 0, 1, \cdots, p \tag{3}$$

Evidently, estimating the transition probabilities is no longer trivial. In what follows, we propose methods for estimating the transition probabilities for Markov chains of high-order.

3 Copula-based transition matrices

First, let us consider the case of a third-order stationary Markov chain with a transition matrix

$$r_{j_0|j_1,\cdots,j_3} = P[X_t = j_0 \mid X_{t-1} = j_1, X_{t-2} = j_2, X_{t-3} = j_3]$$

we may express this transition matrix as

$$r_{j_0|j_1,\cdots,j_3} = P[X_t = j_0, X_{t-1} = j_1, X_{t-2} = j_2 \mid X_{t-3} = j_3] \Big/ P[X_{t-1} = j_1, X_{t-2} = j_2 \mid X_{t-3} = j_3] \quad (4)$$

Furthermore, note that

$$P[X_t = j_0, X_{t-1} = j_1, X_{t-2} = j_2 \mid X_{t-3} = j_3] = \sum_{l=0,1} \sum_{m=0,1} \sum_{n=0,1} (-1)^{(l+m+n)} \times P[X_t \leq j_0 - l, X_{t-1} \leq j_1 - m, X_{t-2} \leq j_2 - n \mid X_{t-3} = j_3] \quad (5)$$

where we know from the Theorem of Sklar (1959) that the joint conditional distribution function on the right hand side of relationship (4) can be expressed by its marginal conditional distributions and a copula function capturing their respective dependence, - i.e.

$$\begin{aligned} P[X_t \leq j_0 - l, X_{t-1} \leq j_1 - m, X_{t-2} \leq j_2 - n \mid X_{t-3} = j_3] &= F(j_0 - l, j_1 - m, j_2 - n \mid j_3) \\ &= C(F(j_0 - l \mid j_3), F(j_1 - m \mid j_3), F(j_2 - n \mid j_3)) \end{aligned} \quad (6)$$

Therefore, using relationship (6), we may express (5) as follows

$$P[X_t = j_0, X_{t-1} = j_1, X_{t-2} = j_2 \mid X_{t-3} = j_3] = \sum_{l=0,1} \sum_{m=0,1} \sum_{n=0,1} (-1)^{(l+m+n)} \times C(F(j_0 - l \mid j_3), F(j_1 - m \mid j_3), F(j_2 - n \mid j_3)) \quad (7)$$

Therefore, the transition matrix (4) can alternatively be expressed as

$$r_{j_0|j_1,\cdots,j_3} = \frac{\sum\limits_{l=0,1} \sum\limits_{m=0,1} \sum\limits_{n=0,1} (-1)^{(l+m+n)} C(F(j_0 - l \mid j_3), F(j_1 - m \mid j_3), F(j_2 - n \mid j_3))}{\sum\limits_{m=0,1} \sum\limits_{n=0,1} (-1)^{(m+n)} C(F(j_1 - m \mid j_3), F(j_2 - n \mid j_3))} \quad (8)$$

In the following Proposition these results are extended to a Markov chain of order p.

Proposition 1 *For a Markov chain of order p, the copula-based transition matrix $r_{j_0|j_1,\cdots,j_p}$ is expressed as follows*

$$r_{j_0|j_1,\cdots,j_p} = \frac{\sum\limits_{l_0=0,1} \cdots \sum\limits_{l_p=0,1} (-1)^{(l_0+\cdots+l_p)} C(F(j_0 - l_0 \mid j_p), \cdots, F(j_{p-1} - l_{p-1} \mid j_p))}{\sum\limits_{l_1=0,1} \cdots \sum\limits_{l_p=0,1} (-1)^{(l_1+\cdots+l_p)} C(F(j_1 - l_1 \mid j_p), \cdots, F(j_{p-1} - l_{p-1} \mid j_p))} \quad (9)$$

where $F(. \mid .)$ is a conditional distribution function and $C(u_0, \cdot, u_p)$ is a copula function in $[0,1]^{p+1}$.

4 Estimation and inference

4.1 Estimation

4.2 Inference

5 Measuring causality in high-dimensional Markov chains

6 Monte Carlo simulations

6.1 Bootstrap bias-corrected estimator of Granger causality measures

6.1.1 Bootstrap bias-correction

6.1.2 Simulation study

6.2 Empirical size and power

7 Empirical application

8 Conclusion

References

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.

9 Appendix: Proofs