Distribution of a Markov chain in reverse-time with cluster observations in the extremes of a finite time window

Daniel A. Gutierrez-Pachas, Eduardo F. Costa and Alessandro N. Vargas *

Abstract

In this brief note, we find formulas for the distribution and the transition probability matrices of a stochastic process described as a time-reversion in a finite time window of a Markov chain, with cluster observation of the Markov state in the extremes of that window.

Problem formulation and solution

Consider a Markov chain \{\eta(k), k = 0, 1, \ldots\} with distribution and transition probabilities given respectively by

\[ s_{ij}(k) = \Pr(\eta(k+1) = j | \eta(k) = i), \]
\[ v_i(k) = \Pr(\eta(k) = i), \quad i, j \in \mathbb{N}, k \geq 0, \]

where \( \mathbb{N} = \{1, \ldots, N\} \) is the state space. For Markov chains and basic probability rules employed in this note, the reader is referred to [1, 2]. Consider the process \{\theta(k), k = 0, 1, \ldots, \ell\}, defined as follows, based on the Markov chain in a finite time window,

\[ \theta(k) = \eta(\ell - k), \quad k = 0, 1, \ldots, \ell. \]

Assume that a cluster observation is available in the form \{\theta(0) \in C_0\}, \( C_0 \subseteq \mathbb{N} \), as well as a cluster observation with anticipation \{\theta(\ell) \in C_{\ell}\}, \( C_{\ell} \subseteq \mathbb{N} \). \( E_0 \in \mathcal{F} \) stands for the associated \( \sigma \)-algebra.

The adopted process \( \theta \) and information structure may appear in real-world problems, including problems involving a first-in last-out queue. For example, suppose the Markov chain characterise the presence of certain irregularities in \( \ell \) items arriving at a store, and that the first item in goes through an inspection of these irregularities, so that we know the state of this item (when it enters the queue); following the queue rule, and counting the items when leaving the queue, we have information “with anticipation” on the state of the \( \ell \)-th item. Note, via this example, that observation with anticipation does not necessarily require prediction of future events.

In this note we seek for formulas for the conditional transition probabilities and distributions defined in (1), based on the Markov chain parameters \( v_i \) and \( s_{ij}(k) \) and the cluster observations. We shall need some additional notation. For each \( i, j \in \mathbb{N} \) and \( k = 0, 1, \ldots, \ell - 1 \), we define

\[ p_{ij} = \Pr(\theta(k) = j | \theta(k+1) = i, E_0), \]
\[ \pi_i(k) = \Pr(\theta(k) = i | E_0), \quad i, j \in \mathbb{N}, 0 \leq k \leq \ell. \]

We write \( P(k) \) to represent a matrix of dimension \( N \times N \), whose components are \( p_{ij}(k) \), therefore satisfying the Chapman-Kolmogorov equation

\[ \pi(k+1) = P(k)\pi(k). \]

We denote, for \( i \in \mathbb{N} \), and \( k = 1, \ldots, \ell - 1 \),

\[ e = \sum_{r \in C_\ell, j \in C_0} |S_r^{0j}v_j(0), g_i(k) = \sum_{r \in C_0} |S_r^{k+1}v_i(0), \]

Lemma 1. The following is valid for \( 1 \leq k \leq \ell - 2 \) and \( i, j \in \mathbb{N} \); if \( i \) is such that \( v_i(\ell - k - 1) > 0 \), then

\[ p_{ij}(k) = \begin{cases} (g_i(k))^{-1} \sum_{r \in C_0} |S_r^{k}v_j(0), & g_i(k) \neq 0, \\ \text{arbitrary,} & \text{otherwise.} \end{cases} \]

if \( v_i(\ell - k - 1) = 0 \), then \( p_{ij}(k) \) is arbitrary. \( \Box \) is also valid for \( k = \ell - 1 \), \( i \in C_\ell \) and \( j \in \mathbb{N} \), as well as for \( k = 0 \), \( i \in \mathbb{N} \) and \( j \in C_0 \). The remaining cases regarding \( p_{ij}(k) \) are: \( p_{ij}(\ell - 1) \) is arbitrary for \( i \notin C_\ell \) and \( j \in \mathbb{N} \); if \( i \in C_\ell \) and there is no Markov state \( \eta(\ell) \in C_0 \) that can be reached from \( i \) (in the sense that \( \Pr(\eta(\ell) \in C_0 | \eta(0) = i) = 0 \) then \( p_{ij}(\ell - 1) \) is arbitrary; \( p_{ij}(0) \) is arbitrary for \( i \in \mathbb{N} \) and \( j \notin C_0 \). Regarding \( \pi_i(\ell) \), for \( i \in C_\ell \),

\[ \pi_i(\ell) = \begin{cases} e^{-1} \sum_{j \in C_0} |S_r^{0j}v_j(0), & e \neq 0, \\ \text{arbitrary,} & \text{otherwise.} \end{cases} \]

For \( i \notin C_\ell \), \( \pi_i(\ell) = 0 \). Finally, \( \pi(k), 1 \leq k \leq \ell - 1 \), is given by \( \Box \) and the formulas above.

Proof: for \( i \in \mathbb{N} \) by definition we have

\[ \pi_i(\ell) = \Pr(\theta(\ell) = i | \eta(0) \in C_\ell, \eta(\ell) \in C_0) \]
\[ = \Pr(\eta(0) = i | \eta(0) \in C_\ell, \eta(\ell) \in C_0). \]

Of course, \( \pi_i(\ell) = 0 \) whenever \( i \notin C_\ell \), as it is given that \( \theta(\ell) = \eta(0) \) is in \( C_\ell \). If \( i \in C_\ell \), and assuming that \( \Pr(\eta(0) \in C_\ell, \eta(\ell) \in C_0) > 0 \), we may write

\[ \pi_i(\ell) = \Pr(\eta(0) = i | \eta(0) \in C_\ell, \eta(\ell) \in C_0) \]
\[ = \frac{\Pr(\eta(\ell) \in C_0, \eta(0) = i)}{\Pr(\eta(\ell) \in C_0, \eta(0) \in C_\ell)} \]
\[ = \sum_{r \in C_\ell} \Pr(\eta(\ell) = j, \eta(0) = i) \]
\[ = \sum_{r \in C_\ell} \sum_{j \in C_0} \Pr(\eta(\ell) = j, \eta(0) = r) \]

*Daniel A. Gutierrez-Pachas is with the Department of Computer Science. Universidade Católica San Pablo, Arequipa, Peru daniel.gutierrez@inv@ucap.edu.pe. Eduardo F. Costa is with the Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, Brazil efcosta@icmc.usp.br. Alessandro N. Vargas is with the Departamento de Engenharia Elétrica, Universidade Tecnológica Federal do Paraná, Brazil avargas@utfpr.edu.br. This work was supported by CNPq Grant 310877/2017-2, CNPq-Universal 421486/2016-3, FAPESP 2017/20934-9 and FAPESP-CEPID 2013/07375-0.
Note that
\[ \Pr(\eta(\ell) = j, \eta(0) = r) = \Pr(\eta(\ell) = j|\eta(0) = r) \cdot \Pr(\eta(0) = r) = [S^t]_{ij} v_r(0), \]
and substituting in the above
\[ \pi_t(\ell) = \sum_{j \in \mathbb{C}_\ell} [S^t]_{ij} v_i(0) \left( \sum_{r \in \mathbb{C}_\ell} \sum_{j \in \mathbb{C}_\ell} [S^t]_{rj} v_r(0) \right)^{-1} = e^{-1} \sum_{j \in \mathbb{C}_\ell} [S^t]_{ij} v_i(0) \]
If \( i \in \mathbb{C}_\ell \) and \( \Pr(\eta(0) = \mathbb{C}_\ell, \eta(\ell) = \mathbb{C}_\ell) = 0 \). (in which case, the inverse in the above equation does not exist), then \( \pi_t(\ell) \) is conditioned on an event of probability zero therefore \( \pi_t(\ell) \) is arbitrary, thus completing the demonstration of (1). Now we turn our attention to \( \mathbf{P} \).
\[ p_{ij}(\ell - 1) = \Pr(\theta(\ell - 1) = j|\theta(\ell) = i, \eta(0) \in \mathbb{C}_\ell, \eta(\ell) \in \mathbb{C}_\ell) = \Pr(\eta(1) = j|\eta(0) = i, \eta(0) \in \mathbb{C}_\ell, \eta(\ell) \in \mathbb{C}_\ell). \]
It is clear that, if \( i \notin \mathbb{C}_\ell \), then \( p_{ij}(\ell - 1) \) is conditioned on an empty set, therefore an event of probability zero, making \( p_{ij}(\ell - 1) \) arbitrary, for any \( j \in \mathbb{N} \). If \( i \in \mathbb{C}_\ell \) and the event \( \{\eta(\ell) \in \mathbb{C}_\ell | \eta(0) = i\} \) is not of probability zero, then using the total probability law we write
\[ p_{ij}(\ell - 1) = \frac{\Pr(\eta(\ell) \in \mathbb{C}_\ell | \eta(1) = j, \eta(0) = i) \cdot s_{ij}(k)}{\Pr(\eta(\ell) \in \mathbb{C}_\ell | \eta(0) = i)} = \frac{\Pr(\eta(\ell) \in \mathbb{C}_\ell | \eta(1) = j) \cdot s_{ij}(k)}{\Pr(\eta(\ell) \in \mathbb{C}_\ell | \eta(0) = i)} = \sum_{r \in \mathbb{C}_\ell} [S^{t-1}]_{jr} s_{ij}(k) \left( \sum_{r \in \mathbb{C}_\ell} [S^t]_{jr} \right)^{-1}, j \in \mathbb{N}, \]
where the second inequality is due to the Markov property. If \( i \in \mathbb{C}_\ell \) and the event \( \{\eta(\ell) \in \mathbb{C}_\ell | \eta(0) = i\} \) is an empty set, then \( p_{ij}(\ell - 1) \) is conditional on an event of zero probability, hence it is arbitrary. Regarding \( p_{ij}(k) \) with \( 1 \leq k \leq \ell - 2 \) and \( i, j \in \mathbb{N} \), denoting \( A = \{\eta(\ell - k - 1) = i\} \) for a better visual diagramming of the next equation, assuming \( \Pr(\eta(\ell) \in \mathbb{C}_\ell, A) > 0 \) and using the Markov property, we write
\[ p_{ij}(k) = \Pr(\eta(\ell - k) = j|A, \eta(0) \in \mathbb{C}_\ell, \eta(\ell) \in \mathbb{C}_\ell) = \Pr(\eta(\ell - k) = j|A, \eta(\ell) \in \mathbb{C}_\ell, \eta(\ell) \in \mathbb{C}_\ell) \]
\[ = \frac{\Pr(\eta(\ell) \in \mathbb{C}_\ell | \eta(\ell - k) = j, A) \Pr(\eta(\ell - k) = j | A) P(A)}{\Pr(\eta(\ell) \in \mathbb{C}_\ell | A)} \]
\[ = \frac{\Pr(\eta(\ell) \in \mathbb{C}_\ell | \eta(\ell - k) = j) P(\eta(\ell - k) = j | A)}{\Pr(\eta(\ell) \in \mathbb{C}_\ell | A)}, \]
yielding
\[ p_{ij}(k) = \sum_{r \in \mathbb{C}_\ell} [S^k]_{jr} s_{ij}(k) \left( \sum_{r \in \mathbb{C}_\ell} [S^{k+1}]_{jr} \right)^{-1} = (g_i(k))^{-1} \sum_{r \in \mathbb{C}_\ell} [S^k]_{jr} s_{ij}(k). \]
Note that the requirement \( \Pr(\eta(\ell) \in \mathbb{C}_0, A) > 0 \) is equivalent to say: (i) considering the Markov chain \( \{\eta_k, k \geq 0\} \), the set \( \mathbb{C}_0 \) is reachable from state \( i \) in \( k + 1 \) steps (in which case \( g_i(k) = 0 \), and (ii) \( \Pr(\eta(\ell - k - 1) = i) = v_i(\ell - k - 1) > 0 \). If any of (i) or (ii) is false, or both are false, then \( p_{ij}(k) \) is conditional on an event of probability zero, hence it is arbitrary. It only remains to find the formula for \( p_{ij}(0) \).
\[ p_{ij}(0) = \Pr(\theta(0) = j|\theta(1) = i, \eta(0) \in \mathbb{C}_\ell, \eta(\ell) \in \mathbb{C}_\ell) \]
\[ = \Pr(\eta(1) = j|\eta(\ell - 1) = i, \eta(0) \in \mathbb{C}_\ell, \eta(\ell) \in \mathbb{C}_\ell). \]
If \( j \notin \mathbb{C}_\ell \) then \( p_{ij}(0) = 0 \) because it is given that \( \eta(0) \in \mathbb{C}_\ell \); otherwise, we have
\[ p_{ij}(0) = \Pr(\eta(\ell) = j | \eta(\ell - 1) = i, \eta(\ell) \in \mathbb{C}_\ell) \]
so that, when the conditional event is not of probability zero,
\[ p_{ij}(0) = \frac{\Pr(\eta(\ell) \in \mathbb{C}_\ell | \eta(\ell - 1) = i)}{\Pr(\eta(\ell) \in \mathbb{C}_\ell | \eta(\ell - 1) = i)} \]
\[ = \frac{\Pr(\eta(\ell) = j | \eta(\ell - 1) = i)}{\Pr(\eta(\ell) \in \mathbb{C}_\ell | \eta(\ell - 1) = i)} \]
\[ = s_{ij}(k) \left( \sum_{r \in \mathbb{C}_\ell} s_{ir} \right)^{-1}, \forall i \in \mathbb{N} \text{ and } j \in \mathbb{C}_\ell, \]
and, when the conditional event \( \{\eta(\ell - 1) = i, \eta(\ell) \in \mathbb{C}_\ell\} \) is of probability zero (in which case the inverse in the above equation does not exist), then \( p_{ij}(0) \) is arbitrary.
Finally, \( \pi(k), 1 \leq k \leq \ell - 1 \), is given by (3) and the formulas above.

**Remark 1.** All arbitrary values in Lemma 1 can be set to zero. This is a good choice in some cases, as in (3), where a Riccati-like equation is computed for every \( i, k \) such that \( \pi_k(0) > 0 \), so that, choosing \( \pi_t(k) = 0 \) avoids unnecessary computations.

**Remark 2.** In view of (3), the transition probabilities of the process \( \theta \) depend on \( k \) even if the Markov chain is time-homogeneous. For \( \mathbf{P} \) to be irrespective of \( k \), it would be necessary that the Markov chain is time-homogeneous and there is no observation of \( \theta(0) \), that is, \( \mathbb{C}_0 = \mathbb{N} \); in this case, (3) reduces to \( p_{ij} = s_{ij} \).

**References**

[1] E. Cinlar, *Introduction to Stochastic Processes*. Prentice-Hall, 1975.

[2] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1984.

[3] D. A. Gutierrez-Pachas and E. F. Costa, “On the linear quadratic problem for systems with time-reversed Markov jump parameters and the duality with filtering of Markov jump linear systems,” *IEEE Transactions on Automatic Control*, vol. 63, no. 9, pp. 3040–3045, 2018.

[4] D. A. Gutierrez-Pachas, E. F. Costa, and A. N. Vargas, “Linear quadratic control problem of systems with Markov jumps in reverse time and observation with anticipation of the jumps,” *Submitted to the IEEE Control Systems Letters*. 