Probability in the Engineering and Informational Sciences (2026), **40:**3 400–415
doi:10.1017/S0269964826100230

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Can a higher-order Markov chain be treated as a first-order Markov chain?

Jianhong Xu

School of Mathematical and Statistical Sciences, Southern Illinois University Carbondale, Carbondale, IL, USA
Email: jhxu@siu.edu

Keywords: classification of states; ergodicity; ever-reaching probability; higher-order Markov chain; k-step transition probability; limiting probability distribution; mean first passage time; recurrence; regularity; stationary distribution; transience; transition tensor

AMS Subject Classification: 60A05; 60J10; 60J99; 15A69; 15B51

It is well known that any higher-order Markov chain can be associated with a first-order Markov chain. In this primarily expository article, we present the first fairly comprehensive analysis of the relationship between higher-order and first-order Markov chains, together with illustrative examples. Our main objective is to address the central question as posed in the title.

1. Introduction

A first-order or simple Markov chain, or just Markov chain as is often called in literature, is a widely used stochastic model in which the future state of a system is determined by its current state alone via some probabilistic rules, not by any of its past states. In many real-world problems, however, it often arises that the future state of a system depends on not only its current state but also a number of its past states following certain probabilistic rules. A higher-order or multiple Markov chain is exactly a stochastic model for such systems and has been adopted in many applied areas [1, 2, 6, 13, 18, 21, 22, 25, 27, 28, 30, 32]. It is a generalization of the notion of first-order Markov chains.

Formally, first-order and higher-order Markov chains can be defined in a unified way as follows. Let $m \geq 2$. Then, an $(m-1)$th order Markov chain is a stochastic process $\{X_t : t = 1, 2, \ldots\}$, where t can be regarded as time and X_t is a random variable representing the state of a system at t while taking values in the state space $S = \{1, 2, \ldots, n\}$ with $n \geq 2$, such that

$$\begin{aligned} &\Pr(X_{t+1} = i_1 | X_t = i_2, \ldots, X_{t-m+2} = i_m, \ldots, X_1 = i_{t+1}) \\ &= \Pr(X_{t+1} = i_1 | X_t = i_2, \ldots, X_{t-m+2} = i_m) \end{aligned} \tag{1.1}$$

for any $t \geq m-1$ and $i_1, i_2, \ldots, i_m, \ldots, i_{t+1} \in S$. In particular, this chain is called higher order when $m \geq 3$. Incidentally, in this definition, X_{t+1} can be interpreted as the future state of the system, X_t the current state, and X_{t-1} through X_{t-m+2} the $m-2$ past states. By reason of the latter, an $(m-1)$th order Markov chain is also said to have memory $m-2$ and, especially, to be memoryless if it is of first-order with $m = 2$.

In addition, throughout this article, a Markov chain is always assumed to be homogeneous, that is, the probability in (1.1) depends only on $i_1, i_2, \ldots, i_m$ but is independent of t. Accordingly, we can write there

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0269964826100230 Published online by Cambridge University Press

$$\Pr(X_{t+1} = i_1 | X_t = i_2, \ldots, X_{t-m+2} = i_m) = p_{i_1 i_2 \ldots i_m},$$

which is called the transition probability from states $(i_2, \ldots, i_m)$ to i_1. These probabilities form an mth order, n dimensional tensor, which is denoted by $\mathcal{P} = [p_{i_1 i_2 \ldots i_m}]$ and called the transition tensor. Such a $\mathcal{P}$ is stochastic in the sense that $0 \le p_{i_1 i_2 \ldots i_m} \le 1$ for all $i_1, i_2, \ldots, i_m \in S$ and $\sum_{i_1 \in S} p_{i_1 i_2 \ldots i_m} = 1$ for all $i_2, \ldots, i_m \in S$. For a first-order Markov chain, $\mathcal{P}$ is reduced to an $n \times n$ column stochastic transition matrix, denoted as P (or Q) in this article. It should be mentioned that in the literature on first-order Markov chains, the transition matrix is typically taken to be row stochastic. In the higher-order case, however, the above "column stochastic" setting with the summation over the first index tends to be more convenient and hence is often adopted; see, for instance, [3, 8, 29] and the references therein.

For brevity, in what follows, we shall refer to a Markov chain of any order simply as a chain.

From the foregoing unified definition, it is natural to anticipate any result for higher-order chains to extend to the first-order case as well. The converse, nevertheless, is not so clear. Specifically, a key question is whether a problem regarding higher-order chains can be solved by way of first-order chains. As far as we can see, it has been a prevailing perception that the answer to this question is to a large extent affirmative. For one thing, it is well known that a higher-order chain can be associated with a first-order chain [5, 16]. Detail in this respect will be given in the next section.

The main thrust of this article is to shed light on the above key question. The importance of this question lies in the fact that it significantly impacts the methodology of examining higher-order chains. Even though such chains have lately been the focus of many studies, see [3, 4, 7, 8, 15, 23, 24, 29] and the references therein, this question remains open to a great degree. In light of this situation, we shall provide, for the first time, a rather comprehensive answer by synthesizing relevant recent results from [9–12]. We shall also give necessary examples to illustrate the results. In particular, we shall highlight the primary differences between higher-order and first-order chains. These differences, to the best of our knowledge, have yet to be presented in such a collective and systematic way.

To help verify the examples in this article or related works in [9–12], the reader may use the HOMC MATLAB package [31], which is publicly available at

https://neumann.math.siu.edu/homc.

2. Reduced first-order chains

Let $m \ge 3$. Suppose that $X = \{X_t : t = 1, 2, \ldots\}$ is an $(m-1)$th order chain on state space $S = \{1, 2, \ldots, n\}$ whose transition tensor is $\mathcal{P} = [p_{i_1 i_2 \ldots i_m}]$. It is well known [5, 16] that this higher-order chain X can be associated with a first-order chain as follows.

First, define the set T of multi-indices of length $m-1$, ordered by linear indexing [26], as

$$T = \{i_1 i_2 \ldots i_{m-1} : i_1, i_2, \ldots, i_{m-1} \in S\}.$$

Note that the cardinality of T is $N = n^{m-1}$. Next, for $t \ge m-1$, introduce vector-valued random variables

$$Y_t = [X_t \;\; X_{t-1} \;\; \ldots \;\; X_{t-m+2}]^T.$$

Let us denote $Y_t = i_1 i_2 \ldots i_{m-1}$ when $X_t = i_1, X_{t-1} = i_2, \ldots, X_{t-m+2} = i_{m-1}$. Then, the vector-valued stochastic process $Y = \{Y_t : t = m-1, m, \ldots\}$ is a first-order chain on state space T, which we shall refer to as the reduced first-order chain obtained from X.

Denote the $N \times N$ transition matrix of Y by Q. It is more convenient to write

$$Q = [q_{i_1 i_2 \ldots i_{m-1} j_2 j_3 \ldots j_m}]$$

https://doi.org/10.1017/S0269964826100230 Published online by Cambridge University Press

in multi-index form with $i_1 i_2 \ldots i_{m-1}, j_2 j_3 \ldots j_m \in T$. Then, the entries of Q can be expressed as, see [12],

$$q_{i_1 i_2 \ldots i_{m-1}, j_2 j_3 \ldots j_m} = \begin{cases} p_{i_1 i_2 \ldots i_{m-1} j_m}, & i_\ell = j_\ell,\ \ell = 2, 3, \ldots, m-1; \\ 0, & \text{otherwise.} \end{cases}$$

In addition, $p_{i_1 i_2 \ldots i_m}$ is located in Q on row

$$i_1 + n(i_2 - 1) + \ldots + n^{m-2}(i_{m-1} - 1)$$

and column

$$i_2 + n(i_3 - 1) + \ldots + n^{m-2}(i_m - 1).$$

For a third-order chain with two states, that is, $m = 4$ and $n = 2$, for example, we have

$$T = \{111, 211, 121, 221, 112, 212, 122, 222\}$$

and

$$Q = \begin{bmatrix} p_{1111} & 0 & 0 & 0 & p_{1112} & 0 & 0 & 0 \\ p_{2111} & 0 & 0 & 0 & p_{2112} & 0 & 0 & 0 \\ 0 & p_{1211} & 0 & 0 & 0 & p_{1212} & 0 & 0 \\ 0 & p_{2211} & 0 & 0 & 0 & p_{2212} & 0 & 0 \\ 0 & 0 & p_{1121} & 0 & 0 & 0 & p_{1122} & 0 \\ 0 & 0 & p_{2121} & 0 & 0 & 0 & p_{2122} & 0 \\ 0 & 0 & 0 & p_{1221} & 0 & 0 & 0 & p_{1222} \\ 0 & 0 & 0 & p_{2221} & 0 & 0 & 0 & p_{2222} \end{bmatrix}. \tag{2.1}$$

In particular, Q coincides with the transition matrix of X when this chain is first order, that is, in the special case $m = 2$.

Speaking of the connection between a higher-order chain and its reduced first-order chain, we mention here the opinion of Joseph Doob, who has left a profound, lasting impact on the field of stochastic processes. In the seminal monograph [5], he stated that "the generalization (from simple Markov processes to multiple ones) is not very significant, because the (vector) process with random variables $\{\hat{x}_n\}$, $\hat{x}_n = (x_n, \ldots, x_{n+v-1})$ (i.e., Y_t in our notation) has the Markov property ... Thus multiple Markov processes can be reduced to simple ones at the small expense of going to vector-valued random variables." This appears to suggest that problems regarding a higher-order chain can be solved through its reduced first-order chain.

From what we can gather in literature, different voices have been few and far between. The most notable argument can be found in the classic treatise [17] by Marius Iosifescu, in which he pointed out first that "In this way with any multiple Markov chain one can associate a simple Markov chain. This is the origin of the rooted prejudice that the study of multiple Markov chains reduces to that of simple ones." and then cited the following example: State i can be recurrent in a second-order chain but $ij \in T$ is not recurrent in the reduced first-order chain for any j. This example, albeit useful, falls short to adequately address why it is a "prejudice" to investigate higher-order chains by reducing them to first-order chains, that is, to fully address the key question as raised in the introductory section. Having laid out the notion of reduced first-order chains, we can now rephrase this key questions as:

https://doi.org/10.1017/S0269964826100230 Published online by Cambridge University Press

Question 2.1. *Can a higher-order chain be studied via its reduced first-order chain?*

This formal statement clarifies what the first-order chain is in the previously stated approach "by way of first-order chains." It is also easier for us to refer to this question in the rest of this article.

3. k-step transition and ever-reaching probabilities

Let us begin here with some background material.

For $m, n \geq 2$, let $\mathcal{A} = [a_{i_1 i_2 \ldots i_m}]$ and $\mathcal{B} = [b_{i_1 i_2 \ldots i_m}]$ be both mth order, n dimensional tensors. The tensor product $\mathcal{A} \boxtimes \mathcal{B}$ is introduced in [10, 12] as an mth order, n dimensional tensor $\mathcal{C} = [c_{i_1 i_2 \ldots i_m}]$ such that

$$c_{i_1 i_2 \ldots i_m} = \sum_{j=1}^{n} a_{i_1 j i_2 \ldots i_{m-1}} b_{j i_2 \ldots i_m}$$

for any $1 \leq i_1, i_2, \ldots, i_m \leq n$. When $m = 2$, in particular, $\mathcal{A} \boxtimes \mathcal{B}$ reduces to a matrix multiplication. Note, however, that the $\boxtimes$ product is in general not associative whenever $m \geq 3$. As a side remark, the product $\mathcal{A} \boxtimes \mathcal{B}$ is originally written as $\mathcal{B} \boxtimes \mathcal{A}$ in [10]. The current notation is consistent with the special case when $\mathcal{A}$ and $\mathcal{B}$ are matrices.

For $k = 2, 3, \ldots$, the kth power of the mth order, n dimensional tensor $\mathcal{A}$ is defined recursively by

$$\mathcal{A}^{k+1} = \mathcal{A}^k \boxtimes \mathcal{A}, \; k = 1, 2, \ldots,$$

with $\mathcal{A}^1 = \mathcal{A}$. By convention, $\mathcal{A}^0 = \mathcal{I} = [\delta_{i_1 i_2 \ldots i_m}]$, the mth order, n dimensional identity tensor whose entries satisfy

$$\delta_{i_1 i_2 \ldots i_m} = \begin{cases} 1, & i_1 = i_2; \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to see that $\mathcal{I} \boxtimes \mathcal{A} = \mathcal{A}$, but in general $\mathcal{A} \boxtimes \mathcal{I} \neq \mathcal{A}$.

Besides the above, the mode-k matricization of the mth order, n dimensional tensor $\mathcal{A}$ is defined to be an n by N, with $N = n^{m-1}$, matrix that has the mode-k fibers $\mathcal{A}(i_1, \ldots, i_{k-1}, :, i_{k+1}, \ldots, i_m)$ as columns. These fibers are arranged in the linear indexing order of $i_1 \ldots i_{k-1} i_{k+1} \ldots i_m$ [20]. Especially, the mode-1 matricization of $\mathcal{A}$ is formed by arranging $\mathcal{A}(:, :, i_3, \ldots, i_m)$, that is, the frontal slices of $\mathcal{A}$, side by side via the linear indexing order of $i_3 \ldots i_m$.

As shown in [9, 10, 12], the $\boxtimes$ product plays a crucial role in the study of higher-order chains. It can be used to formulate various important quantities such as k-step transition probabilities and ever-reaching probabilities, which we shall describe next.

Back to the $(m-1)$th order chain X with state space $S = \{1, 2, \ldots, n\}$ and transition tensor $\mathcal{P} = [p_{i_1 i_2 \ldots i_m}]$. Denote $\mathcal{P}^k = [p^{(k)}_{i_1 i_2 \ldots i_m}]$. Then, $p^{(k)}_{i_1 i_2 \ldots i_m}$ is the k-step transition probability from states $(i_2, \ldots, i_m)$ to i_1 [9], that is,

$$p^{(k)}_{i_1 i_2 \ldots i_m} = \Pr(X_{t+k} = i_1 | X_t = i_2, \ldots, X_{t-m+2} = i_m).$$

Accordingly, $\mathcal{P}^k$ is called the k-step transition tensor.

Continuing, let

$$\eta_{i_1 i_2 \ldots i_m} = \min\{j \geq 1 : X_{m+j-1} = i_1 | X_{m-1} = i_2, \ldots, X_1 = i_m\} \tag{3.1}$$

be the random variable of the first passage time from states $(i_2, \ldots, i_m)$ to i_1. Denote the probability for this passage to happen at $j = k$ by

$$f^{[k]}_{i_1 i_2 \ldots i_m} = \Pr(\eta_{i_1 i_2 \ldots i_m} = k).$$

https://doi.org/10.1017/S0269964826100230 Published online by Cambridge University Press

In the special case when $k = 1, f^{[1]}_{i_1 i_2 \ldots i_m} = p_{i_1 i_2 \ldots i_m}$. Denote $\mathcal{F}^{[k]} = [f^{[k]}_{i_1 i_2 \ldots i_m}]$, which is called the k-step first passage time probability tensor. It is known [9] that

$$f^{[k+1]}_{i_1 i_2 \ldots i_m} = \sum_{j \in S, j \neq i_1} f^{[k]}_{i_1 j i_2 \ldots i_{m-1}} p_{j i_2 \ldots i_m}, \; k = 1, 2, \ldots, \tag{3.2}$$

To express (3.2) via the $\boxtimes$ product, for any mth order, n dimensional tensor $\mathcal{A} = [a_{i_1 i_2 \ldots i_m}]$, let us define the diagonal tensor $\mathcal{A}_d = [a^{(d)}_{i_1 i_2 \ldots i_m}]$ satisfying

$$a^{(d)}_{i_1 i_2 \ldots i_m} = \begin{cases} a_{i_1 i_1 i_3 \ldots i_m}, & i_1 = i_2; \\ 0, & \text{otherwise.} \end{cases}$$

Clearly, (3.2) can now be written simply as

$$\mathcal{F}^{[k+1]} = (\mathcal{F}^{[k]} - \mathcal{F}^{[k]}_d) \boxtimes \mathcal{P}.$$

The probability of ever reaching i_1 from $(i_2, \ldots, i_m)$ is given by

$$f_{i_1 i_2 \ldots i_m} = \Pr(\eta_{i_1 i_2 \ldots i_m} < \infty) = \sum_{k=1}^{\infty} f^{[k]}_{i_1 i_2 \ldots i_m},$$

or in tensor form, the ever-reaching probability tensor $\mathcal{F} = [f_{i_1 i_2 \ldots i_m}]$ can be expressed as

$$\mathcal{F} = \sum_{k=1}^{\infty} \mathcal{F}^{[k]}.$$

These results regarding k-step transition probabilities and ever-reaching probabilities remain valid for first-order chains. In the first-order case, they can also be written as $P^k = [p^{(k)}_{ij}]$, the k-step transition matrix, and $F = [f_{ij}]$, the ever-reaching probability matrix, respectively [16, 19].

In passing, we also give one example of the mode-1 matricization of the transition tensor $\mathcal{P}$, denoted by P. Take a third-order chain with two states just as the one in the previous section, that is, $m = 4$ and $n = 2$. Then,

$$P = \left[\begin{array}{cc|cc|cc|cc} p_{1111} & p_{1211} & p_{1121} & p_{1221} & p_{1112} & p_{1212} & p_{1122} & p_{1222} \\ p_{2111} & p_{2211} & p_{2121} & p_{2221} & p_{2112} & p_{2212} & p_{2122} & p_{2222} \end{array}\right].$$

Clearly, P coincides with the transition matrix when the chain is first order. For $k = 0, 1, \ldots$, the mode-1 matricization of $\mathcal{P}^k$ is formed in a similar way and denoted by $P^{(k)}$.

In view of Question 2.1, we now raise the question of whether the k-step transition tensor or ever-reaching probability tensor of a higher-order chain can be obtained from their respective counterpart of the reduced first-order chain. Given an $(m-1)$th order, $m \geq 3$, chain with transition tensor $\mathcal{P}$, it is shown in [12] that

$$P^{(k)} = P^{(0)} Q^k, \; k = 1, 2, \ldots, \tag{3.3}$$

where $P^{(k)}$ and $P^{(0)}$ stand for the mode-1 matricizations of $\mathcal{P}^k$ and $\mathcal{P}^0 = \mathcal{I}$, respectively, and Q^k denotes the k-step transition matrix of the associated reduced first-order chain. The formula (3.3) provides a connection between the k-step transition probabilities of the higher-order chain and those of its reduced

https://doi.org/10.1017/S0269964826100230 Published online by Cambridge University Press

first-order chain. It is worth noting, nevertheless, that these k-step transition probabilities of the higher-order case can be obtained directly via $\mathcal{P}^k$ without resorting to Q^k.

The question concerning ever-reaching probabilities will be considered in the next section after the introduction of the notion of ergodicity.

4. Irreducibility, ergodicity, and regularity

Let us begin with several well-known concepts for first-order chains [16, 19]. Suppose that $P = [p_{ij}]$ is the transition matrix of a first-order chain on state space S. Then:

- The chain is irreducible if for any $\emptyset \neq K \subsetneq S$, there exists $p_{ij} > 0$ for some $i \in K$ and some $j \in K^c$. In matrix theory, its transition matrix P is also said to be irreducible. The chain is said to be reducible if it is not irreducible.
- The chain is ergodic if for any $i, j \in S$, there exists $k \geq 1$, which may depend on i and j, such that $p_{ij}^{(k)} > 0$.
- The chain is regular if there exists some $k \geq 1$ such that $p_{ij}^{(k)} > 0$ for all $i, j \in S$, that is, $P^k > 0$. In matrix theory, such P is said to be primitive.

For a first-order chain, it is well known [14] that irreducibility is equivalent to ergodicity.

The above definitions can be extended as follows to a higher-order chain on state space S, whose transition tensor is $\mathcal{P} = [p_{i_1 i_2 \ldots i_m}]$.

- The chain is said to be irreducible if for each $\emptyset \neq K \subsetneq S$, there exists $p_{i_1 i_2 \ldots i_m} > 0$ for some $i_1 \in K$ and some $i_2, \ldots, i_m \in K^c$ [24].
- The chain is said to be ergodic if for any $i_1, i_2, \ldots, i_m \in S$, there exists $k \geq 1$, which may depend on $i_1, i_2, \ldots, i_m$, such that $p_{i_1 i_2 \ldots i_m}^{(k)} > 0$ [9].
- The chain is said to be regular if there exists $k \geq 1$ such that for any $i_1, i_2, \ldots, i_m \in S$, $p_{i_1 i_2 \ldots i_m}^{(k)} > 0$, that is, $\mathcal{P}^k > 0$ [12].

Unlike the first-order case, for a higher-order chain, ergodicity is a stronger condition than irreducibility [9], that is, an ergodic higher-order chain must be irreducible, but the converse is not true in general. Consider, for example, a second-order chain with three states and transition tensor

$$\mathcal{P}(:,:,1) = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \ \mathcal{P}(:,:,2) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \ \mathcal{P}(:,:,3) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

This chain is irreducible, but not ergodic since $p_{2 i_2 i_3}^{(k)} = 0$ for all $1 \leq i_2, i_3 \leq 3$ and $k \geq 2$.

In order to guarantee the mean first passage times to be well defined and finite, for example, ergodicity is an important condition no matter the order of a chain [10, 19]. Following Question 2.1, it is natural to ask whether the mean first passage times of a higher-order chain can be obtained from those of its reduced first-order chain. First and foremost, however, it makes sense to ask whether ergodicity of a higher-order chain carries over to its reduced first-order chain. The answer to the latter question, unfortunately, turns out to be negative as shown by the following example.

Let $\mathcal{P} = [p_{i_1 i_2 i_3}]$ be the transition tensor of a second-order chain of three states such that all $p_{i_1 i_2 i_3}$ are positive except $p_{311} = p_{312} = p_{313} = 0$. Let us simply take

$$\mathcal{P}(:,:,1) = \mathcal{P}(:,:,2) = \mathcal{P}(:,:,3) = \begin{bmatrix} 1/2 & 1/3 & 1/3 \\ 1/2 & 1/3 & 1/3 \\ 0 & 1/3 & 1/3 \end{bmatrix}.$$

https://doi.org/10.1017/S0269964826100230 Published online by Cambridge University Press

Since $\mathcal{P}^2 > 0$, this chain is regular and thus is also ergodic. Its reduced first-order chain has transition matrix

$$Q = \begin{bmatrix} 1/2 & 0 & 0 & 1/2 & 0 & 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 1/3 & 0 & 0 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 1/3 & 0 & 0 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 1/3 & 0 & 0 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 1/3 & 0 & 0 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 1/3 & 0 & 0 & 1/3 & 0 & 0 & 1/3 \end{bmatrix},$$

which is clearly not even irreducible due to its third row of all zeros. Hence, the reduced first-order chain is not ergodic.

The above example also justifies that, like ergodicity, regularity does not carry over either from a higher-order chain to its reduced first-order chain.

What this example demonstrates is, in fact, the first obstacle to studying a higher-order chain through its reduced first-order chain. We shall demonstrate later in Section 6 that other difficulties may arise even if ergodicity or regularity happens to carry over to the reduced first-order chain. The issue of mean first passage times will be addressed there too.

Returning to the question raised in the preceding section concerning the relationship between the ever-reaching probabilities of a higher-order chain and those of its reduced first-order chain, we consider below a second-order chain with four states and transition tensor

$$\mathcal{P}(:,:,1) = \begin{bmatrix} 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \ \mathcal{P}(:,:,2) = \begin{bmatrix} 0 & 0 & 1/2 & 1 \\ 0 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix},$$

$$\mathcal{P}(:,:,3) = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \ \mathcal{P}(:,:,4) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 \end{bmatrix}.$$

It is easy to verify $\mathcal{P}^{10} > 0$. Clearly, this chain is regular and, consequently, ergodic. It is shown in [10] that for such an ergodic chain, its ever-reaching probabilities $f_{i_1 i_2 i_3} = 1$ for all $1 \le i_1, i_2, i_3 \le 4$. The associated reduced first-order chain, however, exhibits a noticeably different dynamics as illustrated by the digraph in Figure 1 that represents its transition matrix Q. Starting, for example, from state 12, the reduced first-order chain has zero probability of reaching state 11. This, nevertheless, does not lead to the conclusion that state 1 cannot be reached from states $(1, 2)$ in the second-order chain since $f_{112} = 1$. This illustrates that the answer to Question 2.1 is negative when dealing with the problem of ever-reaching probabilities.

It is quite straightforward, though a bit tedious, to show that the ever-reaching probabilities of a higher-order chain cannot be obtained via those of its reduced first-order chain in a similar manner as neat as (3.3). Again, we consider the general third-order chain with two states in Section 2. Let $G^{[2]} = [g^{[2]}_{i_1 i_2 i_3, j_1 j_2 j_3}]$ be the 2-step first passage time probability matrix of the reduced first-order chain.

https://doi.org/10.1017/S0269964826100230 Published online by Cambridge University Press

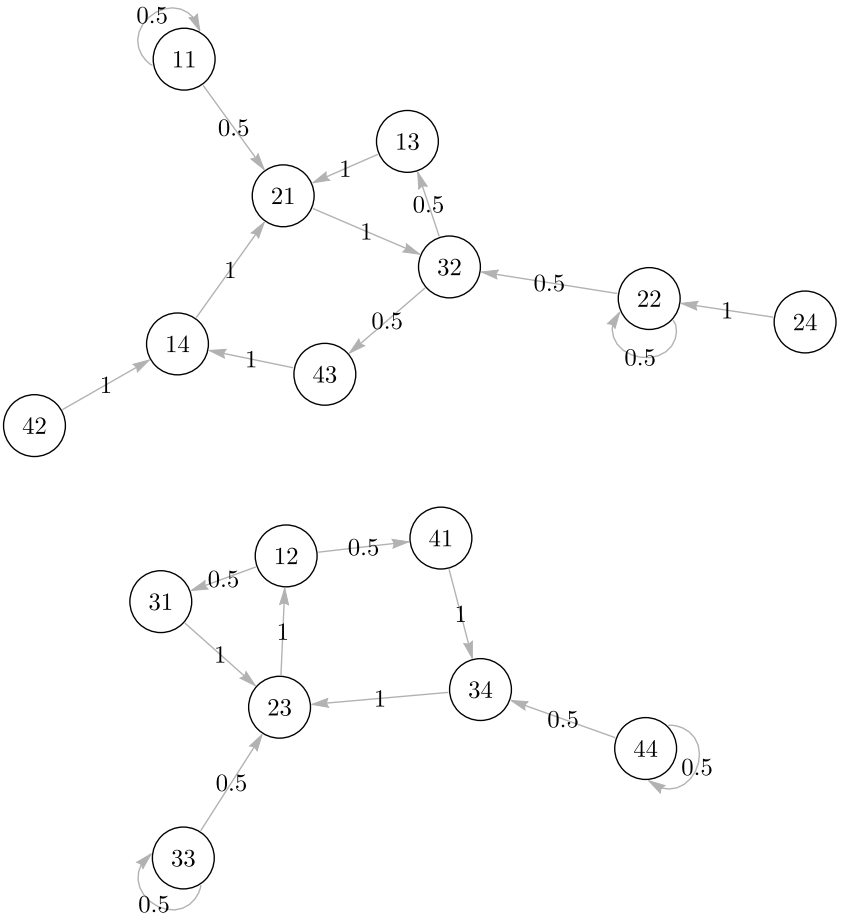


Figure 1. *Digraph of reduced first-order chain.*

Then, it is easy to verify that

$$g^{[2]}_{i_1i_2i_3,j_1j_2j_3} = p_{i_1i_2i_3j_2}p_{i_2i_3j_2j_3},\ i_1i_2i_3, j_1j_2j_3 \in T, \tag{4.1}$$

provided $j_1 = i_3$ and either $i_1 \neq i_2$ or $i_2 \neq i_3$ or $i_3 \neq j_2$. As a comparison, it follows from (2.1) that

$$q^{(2)}_{i_1i_2i_3,i_3j_2j_3} = p_{i_1i_2i_3j_2}p_{i_2i_3j_2j_3}. \tag{4.2}$$

Clearly, (4.1) is much more complicated than (4.2). This, together with the preceding example as illustrated in Figure 1, offers an additional perspective showing that the answer to Question 2.1 is negative as far as ever-reaching probabilities are concerned.

5. Classification of states

Given a first-order chain with transition matrix $P = [p_{ij}]$ and ever-reaching probability matrix $F = [f_{ij}]$, it is well known [16] that:

- A state i is recurrent if $f_{ii} = 1$. A state i is recurrent iff $\sum_{k=1}^{\infty} p_{ii}^{(k)} = \infty$.

https://doi.org/10.1017/S0269964826100230 Published online by Cambridge University Press

- A state i is transient if $f_{ii} < 1$. A state i is transient iff $\sum_{k=1}^{\infty} p_{ii}^{(k)} < \infty$.
- A state i is absorbing if $p_{ii} = 1$.

The example that is cited by [17], see the introductory section, indicates that the states of a higher-order chain cannot be classified according to the states of its reduced first-order chain. A more complete classification scheme for states of higher-order chains is recently proposed in [11]. Here is a brief description of this scheme. Consider a higher-order chain whose state space is S, transition tensor is $\mathcal{P} = [p_{i_1 i_2 \ldots i_m}]$, and ever-reaching probability tensor is $\mathcal{F} = [f_{i_1 i_2 \ldots i_m}]$. Then,

- A state i is said to be recurrent if $f_{iii_3 \ldots i_m} = 1$ for all $i_3, \ldots, i_m \in S$. If state i is recurrent, then $\sum_{k=1}^{\infty} p_{iii_3 \ldots i_m}^{(k)} = \infty$ for all $i_3, \ldots, i_m \in S$.
- A state i is said to be transient if $f_{iii_3 \ldots i_m} < 1$ for some $i_3, \ldots, i_m \in S$. Especially, a state i is said to be fully transient if $f_{iii_3 \ldots i_m} < 1$ for any $i_3, \ldots, i_m \in S$. When state i is fully transient, $\sum_{k=1}^{\infty} p_{iii_3 \ldots i_m}^{(k)} < \infty$ for all $i_3, \ldots, i_m \in S$.
- A state i is said to be absorbing if $p_{iii_3 \ldots i_m} = 1$ for all $i_3, \ldots, i_m \in S$.

These results from [11] seem to be parallel to their first-order counterparts. We, however, point out some differences between the higher-order and first-order cases. Specifically, for a higher-order chain:

- There may exist the scenario where a state i is transient but not fully transient, that is, $f_{iii_3 \ldots i_m}$ is smaller than 1 for some $i_3, \ldots, i_m \in S$ but is equal to 1 for other $i_3, \ldots, i_m \in S$. This is indeed at the center of the key differences between the higher-order and first-order chains when it comes to the classification of states.
- The condition of whether $\sum_{k=1}^{\infty} p_{iii_3 \ldots i_m}^{(k)}$ is infinite or finite becomes necessary, but not sufficient, for characterizing recurrence or full transience; see the next example.

Before proceeding, we need to introduce a few terms. A state j is said to be reachable from state i, denoted as $i \to j$, if there exists $k \geq 0$, which may depend on $j, i, i_3, \ldots, i_m$, such that $p_{jii_3 \ldots i_m}^{(k)} > 0$ or, equivalently, $f_{jii_3 \ldots i_m} > 0$, for all $i_3, \ldots, i_m \in S$. In addition, states i and j are said to communicate, denoted by $i \leftrightarrow j$, if $i \to j$ and $j \to i$. The $\leftrightarrow$ relationship is an equivalence relationship and thus it partitions the state space S into equivalence classes [11].

As illustrated in [11], there are quite a few other differences when we go from first-order to higher-order chains regarding the classification of states. For example,

- A first-order chain must have a recurrent state, but not so for a higher-order chain. Take a second-order chain on three states with transition tensor

$$\mathcal{P}(:,:,1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}, \ \mathcal{P}(:,:,2) = \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}, \ \mathcal{P}(:,:,3) = \begin{bmatrix} 0 & 1/2 & 1 \\ 0 & 1/2 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

  Then,

$$\mathcal{F}(:,:,1) = \begin{bmatrix} 1 & 1/2 & 3/4 \\ 0 & 1 & 1 \\ 0 & 1/2 & 1/2 \end{bmatrix}, \ \mathcal{F}(:,:,2) = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \ \mathcal{F}(:,:,3) = \begin{bmatrix} 3/4 & 1/2 & 1 \\ 1 & 1/2 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

https://doi.org/10.1017/S0269964826100230 Published online by Cambridge University Press

For each $1 \le i \le 3$, its ever-reaching probability f_{iii_3} equals 1 for some i_3 and is less than 1 for some other i_3. Consequently, none of the states in this example are recurrent. In addition, we see

$$\sum_{k=1}^{\infty} p_{11i_3}^{(k)} = \infty \text{ and } \sum_{k=1}^{\infty} p_{33i_3}^{(k)} < \infty$$

for all $1 \le i_3 \le 3$, implying that whether $\sum_{k=1}^{\infty} p_{iii_3\ldots i_m}^{(k)}$ is infinite or not for all $i_3, \ldots, i_m \in S$ is only a necessary condition for recurrence or full transience, respectively.

- In a first-order chain, a state i is recurrent provided $j \to i$ whenever $i \to j$. Such a characterization does not apply to a higher-order chain. Consider a second-order chain with two states and transition tensor

$$\mathcal{P}(:,:,1) = \begin{bmatrix} 1 & 1/2 \\ 0 & 1/2 \end{bmatrix}, \ \mathcal{P}(:,:,2) = \begin{bmatrix} 0 & 1/2 \\ 1 & 1/2 \end{bmatrix}.$$

It follows

$$\mathcal{F}(:,:,1) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \ \mathcal{F}(:,:,2) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Both states are recurrent since $f_{iii_3} = 1$ for all $i, i_3 = 1, 2$. Meanwhile, $2 \to 1$ yet $1 \nrightarrow 2$ since $p_{i_1 i_2 i_3}^{(k)} > 0$ for all $1 \le i_1, i_2 \le 2$ and $k \ge 2$ but $p_{211}^{(k)} = 0$ for all $k \ge 1$.

- In a first-order chain, both recurrence and transience are class properties, that is, the states in an equivalence class must be either all recurrent or all transient. When it comes to a higher-order chain, recurrence and transience are not class properties anymore. To see this, we examine a second-order chain with three states and transition tensor

$$\mathcal{P}(:,:,1) = \begin{bmatrix} 1/2 & 1/3 & 1/2 \\ 1/2 & 1/2 & 0 \\ 0 & 1/3 & 1/2 \end{bmatrix}, \ \mathcal{P}(:,:,2) = \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{bmatrix}, \ \mathcal{P}(:,:,3) = \begin{bmatrix} 0 & 0 & 1/2 \\ 0 & 0 & 1/2 \\ 1 & 1 & 0 \end{bmatrix}.$$

For this chain, we have

$$\mathcal{F}(:,:,1) = \begin{bmatrix} 5/6 & 2/3 & 1 \\ 1 & 1 & 1 \\ 1/2 & 1/2 & 1 \end{bmatrix}, \ \mathcal{F}(:,:,2) = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1/2 & 0 & 1 \end{bmatrix}, \ \mathcal{F}(:,:,3) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

By checking its ever-reaching probabilities f_{iii_3}, state 1 is transient but not fully transient, and state 3 is recurrent. These states, however, are in the same equivalence class, that is, $1 \leftrightarrow 3$.

Such differences constitute yet another roadblock to studying a higher-order chain via its reduced first-order chain. Thus, as far as classification of states is concerned, the answer to Question 2.1 is again negative.

The following was conjectured in [11]. To end this section, we shall prove it here.

Theorem 5.1. *Let $K \subset S$ be an equivalence class. Then, it is impossible to have $i, j \in K$ such that i is recurrent while j is fully transient.*

https://doi.org/10.1017/S0269964826100230 Published online by Cambridge University Press

Proof. Let $i, j \in K$, $i \neq j$, with j being recurrent. Then, it suffices to show that i cannot be fully transient.

Let us fix $i_3, \ldots, i_m$. Since $i \to j$ and $i \neq j$, there exists $\alpha \geq 1$ such that $p^{(\alpha)}_{jii_3\ldots i_m} > 0$. Denote the event "the chain takes α steps to go from $ii_3 \ldots i_m$ to j" by A.

Next, denote the event "starting from $jj_3 \ldots j_m$, the chain takes β steps to return to j" by B, where $j_3, \ldots, j_m$ are given by:

- when $\alpha = 1$: $j_3 = i, j_4 = i_3, \ldots, j_m = i_{m-1}$
- when $\alpha = 2$: $j_4 = i, j_5 = i_3, \ldots, j_m = i_{m-2}$, any j_3
- $\ldots$
- when $\alpha = m - 2$: $j_m = i$, any $j_3, \ldots, j_{m-1}$
- $\ldots$

Fix $j_3, \ldots, j_m$ from now on.

Next, denote the event "the chain takes γ steps to go from $jk_3 \ldots k_m$ to i" by C, where $k_3, \ldots, k_m$ are determined in the same way as $j_3, \ldots, j_m$ above.

From this point on, let us assume $\beta \geq m - 1$. Thus, we can fix $k_3, \ldots, k_m$. Accordingly, by $j \to i$ and $i \neq j$, there exists $\gamma \geq 1$ such that $p^{(\gamma)}_{ijk_3\ldots k_m} > 0$.

Finally, we observe

$$\begin{aligned} p^{(\gamma+\beta+\alpha)}_{iii_3\ldots i_m} &\geq \Pr(ABC) \\ &= \Pr(A)\Pr(B|A)\Pr(C|AB) \\ &= p^{(\alpha)}_{jii_3\ldots i_m} p^{(\beta)}_{jjj_3\ldots j_m} p^{(\gamma)}_{ijk_3\ldots k_m}. \end{aligned}$$

It follows that

$$\sum_{\beta=m-1}^{\infty} p^{(\gamma+\beta+\alpha)}_{iii_3\ldots i_m} \geq p^{(\alpha)}_{jii_3\ldots i_m} p^{(\gamma)}_{ijk_3\ldots k_m} \sum_{\beta=m-1}^{\infty} p^{(\beta)}_{jjj_3\ldots j_m} = \infty$$

since j is recurrent. The above clearly leads to

$$\sum_{\beta=1}^{\infty} p^{(\beta)}_{iii_3\ldots i_m} = \infty.$$

The proof is now complete. □

6. Mean first passage times

For convenience, let us use a framework that includes both higher-order and first-order chains.

Recall $\eta_{i_1 i_2 \ldots i_m}$ in (3.1). For an $(m-1)$th order, $m \geq 2$, chain with state space $S = \{1, 2, \ldots, n\}$, the mean first passage time from states $(i_2, \ldots, i_m)$ to i_1 is defined by

$$\mu_{i_1 i_2 \ldots i_m} = \mathrm{E}(\eta_{i_1 i_2 \ldots i_m}) = \sum_{k=1}^{\infty} k \Pr(\eta_{i_1 i_2 \ldots i_m} = k).$$

Accordingly, the mean first passage time tensor is the mth order, n dimensional tensor $\mu = [\mu_{i_1 i_2 \ldots i_m}]$.

As shown in [10], for an ergodic higher-order chain of order $m - 1$ whose mth order, n dimensional transition tensor is $\mathcal{P}$, its mean first passage times are finite and are uniquely determined by the following tensor equation:

https://doi.org/10.1017/S0269964826100230 Published online by Cambridge University Press

$$\mu = \mathcal{E} + (\mu - \mu_d) \boxtimes \mathcal{P}, \tag{6.1}$$

where $\mathcal{E}$ is the mth order, n dimensional tensor of all ones. Furthermore, as a linear system, (6.1) is nonsingular if and only if the chain is ergodic.

Comparing with an ergodic first-order chain, the mean first passage time matrix M satisfies [19]

$$M = E + (M - M_d)P, \tag{6.2}$$

where E is the matrix of all ones and $M_d = [m_{ij}^{(d)}]$ is such that

$$m_{ij}^{(d)} = \begin{cases} m_{ii}, & i = j; \\ 0, & \text{otherwise.} \end{cases}$$

Clearly, (6.2) is a special case of (6.1). Because of the resemblance between (6.1) and (6.2), and in the spirit of Question 2.1, it is an intriguing question whether (6.1) can be derived from (6.2) using the reduced first-order chain. The answer, however, turns out to be negative, as illustrated by the example below.

Take a second-order chain with three states and

$$\mathcal{P}(:,:,1) = \mathcal{P}(:,:,2) = \mathcal{P}(:,:,3) = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}.$$

By solving (6.1),

$$\mu(:,:,1) = \mu(:,:,2) = \mu(:,:,3) = \begin{bmatrix} 3 & 3 & 3 \\ 3 & 3 & 3 \\ 3 & 3 & 3 \end{bmatrix}.$$

On the other hand, observing that the regularity of this second-order chain carries over to its reduced first-order chain since $Q^2 = (1/9)E > 0$, where E is the matrix of all ones, and using (6.2) for the reduced first-order chain, we obtain

$$M = \begin{bmatrix} 9 & 12 & 12 & 9 & 12 & 12 & 9 & 12 & 12 \\ 6 & 9 & 9 & 6 & 9 & 9 & 6 & 9 & 9 \\ 6 & 9 & 9 & 6 & 9 & 9 & 6 & 9 & 9 \\ 9 & 6 & 9 & 9 & 6 & 9 & 9 & 6 & 9 \\ 12 & 9 & 12 & 12 & 9 & 12 & 12 & 9 & 12 \\ 9 & 6 & 9 & 9 & 6 & 9 & 9 & 6 & 9 \\ 9 & 9 & 6 & 9 & 9 & 6 & 9 & 9 & 6 \\ 9 & 9 & 6 & 9 & 9 & 6 & 9 & 9 & 6 \\ 12 & 12 & 9 & 12 & 12 & 9 & 12 & 12 & 9 \end{bmatrix}.$$

Obviously, in this example, μ does not follow directly from M. If we look at m_{11}, for instance, it represents the mean first passage time from state 11, in multi-index form, to itself. This quantity, however, is different from μ_{111}. In other words, while addressing Question 2.1, other difficulties may still arise even if ergodicity or regularity does carry forward from a higher-order chain to its reduced first-order chain.

https://doi.org/10.1017/S0269964826100230 Published online by Cambridge University Press

One way of working around the discrepancy as shown above between the mean first passage times of a higher-order chain and those of its reduced first-order chain is to resort to certain matricization of the transition tensor [10]. This approach is not as convenient as the direct tensor treatment leading to (6.1). Besides, its result can be easily obtained by matricizing the mean first passage time tensor from (6.1).

7. Limiting probability distributions

Having presented a range of differences between higher-order and first-order chains in previous sections, we now come to a point where these two appear to converge to a certain level. Again, we use here a framework that includes both higher-order and first-order chains.

With $m \geq 2$, let $X = \{X_t : t = 1, 2, \ldots\}$ be an $(m-1)$th order chain on state space $S = \{1, 2, \ldots, n\}$ whose transition tensor is $\mathcal{P} = [p_{i_1 i_2 \ldots i_m}]$. The probability distribution of X at t is given by

$$x_t = [\Pr(X_t = 1) \ \Pr(X_t = 2) \ \ldots \ \Pr(X_t = n)]^T.$$

If there exists an n-vector π, which is independent of the initial probability distributions $x_1, x_2, \ldots, x_{m-1}$, such that

$$\lim_{t \to \infty} x_t = \pi,$$

then π is called the limiting probability distribution of the chain X[12]. It is clear that π satisfies $\pi \geq 0$ and $\sum_{i \in S} \pi_i = 1$.

On the other hand, let Y be the reduced first-order chain associated with X, whose transition matrix is given by Q. Then, an N-vector ξ is said to be a stationary distribution of both the chains X and Y if ξ satisfies $\xi \geq 0$, $\sum_{i_1 i_2 \ldots i_{m-1} \in T} \xi_{i_1 i_2 \ldots i_{m-1}} = 1$, and $Q\xi = \xi$. The subscript for the entries of ξ is written here in multi-index form. Consequently, $Q\xi = \xi$ translates to

$$\sum_{j \in S} p_{i_1 i_2 \ldots i_{m-1} j} \xi_{i_2 \ldots i_{m-1} j} = \xi_{i_1 i_2 \ldots i_{m-1}}, \ i_1, i_2, \ldots, i_{m-1} \in S,$$

which is precisely how a stationary distribution is equivalently defined in [8] when ξ is treated as an $(m-1)$th order, n dimensional tensor. Obviously, as a vector, ξ is just a normalized right eigenvector of Q corresponding to the dominant eigenvalue $\lambda = 1$, sometimes also called a Perron vector [14].

It is shown in [12] that if the chain X is regular, then its limiting probability distribution π is guaranteed to exist. In this case, $\pi > 0$ and

$$\lim_{k \to \infty} \mathcal{P}^k = \pi \otimes \underbrace{e \otimes \cdots \otimes e}_{m-1},$$

where, and in what follows, e is the n-vector of all ones and $\otimes$ is the outer product. Without the regularity condition on X, π may not exist and therefore such a condition is tight.

Regarding a stationary distribution of the chains X and Y, its existence is guaranteed following the celebrated Perron–Frobenius theorem, yet such a distribution may not be unique. If, in addition, Y is a regular chain, then its stationary distribution ξ is unique and $\xi > 0$.

The preceding results regarding the limiting probability distribution and stationary distribution continue to hold when the chain X is first order and regular [16, 19]. In this scenario, the limiting probability distribution is the same as the stationary distribution. To be more specific, on letting P be the transition matrix, then there exists a unique n-vector $\pi > 0$ with $\sum_{i \in S} \pi_i = 1$ such that $\lim_{k \to \infty} P^k = \pi e^T$ and, at the same

https://doi.org/10.1017/S0269964826100230 Published online by Cambridge University Press

time, $P\pi = \pi$ [19]. This appears to be the cause of why the nomenclatures "limiting probability distribution" and "stationary distribution" are at times used interchangeably in works on higher-order chains. Our foregoing description of these concepts clarifies the difference between them in the higher-order case.

Despite the differences, there is a principal connection between them too. As shown in [12], when the chain X is higher order and regular, its limiting probability distribution π can be determined via any stationary distribution ξ through

$$\pi = P^{(0)}\xi, \tag{7.1}$$

where $P^{(0)}$ stands for the mode-1 matricization of the mth order, n dimensional identity tensor. This result does not require the chain Y to be regular.

To illustrate the usage of (7.1), let us take the same regular second-order chain as in Section 4 with four states and transition tensor

$$\mathcal{P}(:,:,1) = \begin{bmatrix} 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \ \mathcal{P}(:,:,2) = \begin{bmatrix} 0 & 0 & 1/2 & 1 \\ 0 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix},$$

$$\mathcal{P}(:,:,3) = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \ \mathcal{P}(:,:,4) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 1/2 \end{bmatrix}.$$

The associated reduced first-order chain is reducible, which can also be seen from the fact that the dominant eigenvalue $\lambda = 1$ of its transition matrix Q has multiplicity 2. This, of course, implies that the reduced first-order chain is not regular. It is easy to verify that

$$z = \frac{1}{7}[0\ \ 0\ \ 1\ \ 1\ \ 2\ \ 0\ \ 0\ \ 0\ \ 0\ \ 2\ \ 0\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0]^T$$

is a normalized right eigenvector corresponding to that dominant eigenvalue above. Hence, the limiting probability distribution of the second-order chain is

$$\pi = P^{(0)}z = \frac{1}{7}[2\ \ 2\ \ 2\ \ 1]^T.$$

Regarding Question 2.1, when it comes to the limiting probability distribution problem of a higher-order chain, (7.1) represents its close tie with a stationary distribution of the corresponding reduced first-order chain. This, along with the k-step transition probabilities in Section 3, is another case in which Question 2.1 yields a positive answer.

8. Conclusions

In this article, we provide a relatively comprehensive answer to Question 2.1 in order to address a still prevailing perception, or "rooted prejudice" in the words of Marius Iosifescu, concerning higher-order chains. Our results show that there can be significant differences between a higher-order chain and its reduced first-order chain in most aspects such as ergodicity, regularity, ever-reaching probabilities, mean first passage times, and classification of states. Owing to such differences, the answer to Question 2.1 is negative in general. In the meantime, the closest, and also probably the most useful, tie between

https://doi.org/10.1017/S0269964826100230 Published online by Cambridge University Press

a higher-order chain and its reduced first-order chain is seen in the limiting probability distribution problem. Overall, it remains a very interesting topic to further explore whether a reduced first-order chain may, or may not, play a useful role in studying other key quantities for its originating higher-order chain. Moreover, the largely tensor theoretic framework employed throughout this article proves to be highly effective for analyzing higher-order chains.

Acknowledgements. We thank the anonymous reviewer for the careful reading and helpful comments, which have led to an improved presentation.

References

[1] Baena-Mirabete, S., & Puig, P. (2018). Parsimonious higher order Markov models for rating transitions. *Journal of the royal statistical society a: statistics in society* 181: 107–131.
[2] Burks, D., & Azad, R. (2020). Higher-order Markov models for metagenomic sequence classification. *Bioinformatics* 36: 4130–4136.
[3] Chang, K., & Zhang, T. (2013). On the uniqueness and non-uniqueness of the positive Z-eigenvector for transition probability tensors. *Journal of mathematical analysis & applications* 408: 525–540.
[4] Culp, J., Pearson, K., & Zhang, T. (2017). On the uniqueness of the Z_1-eigenvector of transition probability tensors. *Linear & multilinear algebra* 65: 891–896.
[5] Doob, J. (1990). *Stochastic processes*. New York: Wiley Classic Library Edition, Wiley.
[6] Flett, G., & Kelly, N. (2016). An occupant-differentiated, higher-order Markov chain method for prediction of domestic occupancy. *Energy & buildings* 125: 219–230.
[7] Geiger, B. (2017). A sufficient condition for a unique invariant distribution of a higher-order Markov chain. *Statistics & probability letters* 130: 49–56.
[8] Gleich, D., Lim, L., & Yu, Y. (2015). Multilinear pagerank. *SIAM Journal on matrix analysis & applications* 36: 1507–1541.
[9] Han, L., Wang, K., & Xu, J. (2022). Higher order ergodic Markov chains and first passage times. *Linear & multilinear algebra* 70: 6772–6779.
[10] Han, L., & Xu, J. (2024). Ever-reaching probabilities and mean first passage times of higher order ergodic Markov chains. *Linear & multilinear algebra* 72: 59–75.
[11] Han, L., & Xu, J. (2024). On classification of states in higher order Markov chains. *Linear algebra & its applications* 685: 24–45.
[12] Han, L., & Xu, J. On limiting probability distributions of higher order Markov chains, under review. doi:10.48550/arXiv.2506.08874.
[13] Ho, L., & Rajapakse, J. (2003). Splice site detection with a higher-order Markov model implemented on a neural network. *Genome informatics* 14: 64–72.
[14] Horn, R., & Johnson, C. (1990). *Matrix analysis*. Cambridge, UK: Cambridge University Press.
[15] Hu, S., & Qi, L. (2014). Convergence of a second order Markov chain. *Applied mathematics & computation* 241: 183–192.
[16] Hunter, J. (1983). *Mathematical techniques of applied probability, volume i: discrete time models: basic theory*. New York: Academic Press.
[17] Iosifescu, M. (2007). *Finite markov processes & their applications*. New York: Dover Publications.
[18] Islam, M., & Chowdhury, R. (2006). A higher order Markov model for analyzing covariate dependence. *Applied mathematical modelling* 30: 477–488.
[19] Kemeny, J., & Snell, J. (1960). *Finite markov chains*. New York: Springer-Verlag.
[20] Kolda, T., & Bader, B. (2009). Tensor decompositions and applications. *SIAM Review* 51: 455–500.
[21] Kwak, J., Lee, C., & Eun, D. (2015). A high-order Markov-chain-based scheduling algorithm for low delay in CSMA networks. *IEEE/ACM Transactions on Networking* 24: 2278–2290.
[22] Lan, J., Li, X., Jilkov, V., & Mu, C. (2013). Second-order Markov chain based multiple-model algorithm for maneuvering target tracking. *IEEE Transactions on Aerospace & Electronic Systems* 49: 3–19.
[23] Li, C., & Zhang, S. (2016). Stationary probability vectors of higher-order Markov chains. *Linear Algebra & Its Applications* 473: 114–125.
[24] Li, W., & Ng, M. (2014). On the limiting probability distribution of a transition probability tensor. *Linear Algebra & Its Applications* 62: 362–385.
[25] Liu, Z., Luo, Y., & Zhu, Y. (2020). State estimation for linear dynamic system with multiple-step random delays using higher-order Markov chain. *IEEE Access* 8: 76218–76227.
[26] Martin, C., Shafer, R., & Larue, B. (2013). An order-p tensor factorization with applications in imaging. *SIAM Journal on Scientific Computing* 35: 474–490.
[27] Masseran, N. (2015). Markov chain model for the stochastic behaviors of wind-direction data. *Energy Conversion & Management* 92: 266–274.
[28] Sanjari, M., & Gooi, H. (2016). Probabilistic forecast of PV power generation based on higher order Markov chain. *IEEE Transactions on Power Systems* 32: 2942–2952.

https://doi.org/10.1017/S0269964826100230 Published online by Cambridge University Press

[29] Wu, S., & Chu, M. (2017). Markov chains with memory, tensor formulation, and the dynamics of power iteration. *Applied Mathematics & Computation* 303: 226–239.
[30] Xiong, H., & Mamon, R. (2019). A higher-order Markov chain-modulated model for electricity spot-price dynamics. *Applied Energy* 233-234: 495–515.
[31] Xu, J. HOMC: a MATLAB package for higher order Markov chains, under review. doi:10.48550/arXiv.2510.02664.
[32] Yang, Y., Jang, H., & Kim, B. (2020). A hybrid recommender system for sequential recommendation: combining similarity models with Markov chains. *IEEE Access* 8: 190136–190146.

Cite this article: Xu J. (2026). Can a higher-order Markov chain be treated as a first-order Markov chain? *Probability in the Engineering and Informational Sciences* 40(3): 400–415. https://doi.org/10.1017/S0269964826100230

https://doi.org/10.1017/S0269964826100230 Published online by Cambridge University Press