Statistics and Probability Letters 130 (2017) 49–56

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

# A sufficient condition for a unique invariant distribution of a higher-order Markov chain

Bernhard C. Geiger

*Institute for Communications Engineering, Technical University of Munich, Theresienstr. 90, 80333 Munich, Germany*

ARTICLE INFO

*Article history:*
Received 17 November 2016
Received in revised form 29 June 2017
Accepted 12 July 2017
Available online 26 July 2017

*MSC:*
60J10

*Keywords:*
Markov chains
Invariant distribution
Function of Markov chains

ABSTRACT

We derive a sufficient condition for a $k$th order homogeneous Markov chain $\mathbf{Z}$ with finite alphabet $\mathcal{Z}$ to have a unique invariant distribution on $\mathcal{Z}^k$. Specifically, let $\mathbf{X}$ be a first-order, stationary Markov chain with finite alphabet $\mathcal{X}$ and a single recurrent class, let $g : \mathcal{X} \to \mathcal{Z}$ be non-injective, and define the (possibly non-Markovian) process $\mathbf{Y} := g(\mathbf{X})$ (where $g$ is applied coordinate-wise). If $\mathbf{Z}$ is the $k$th order Markov approximation of $\mathbf{Y}$, its invariant distribution is unique. We generalize this to non-Markovian processes $\mathbf{X}$.



## 1. Introduction

We consider invariant distributions of a $k$th order (i.e., "multiple") Markov chain $\mathbf{Z} := (Z_n)_{n\in\mathbb{N}_0}$ on a finite alphabet $\mathcal{Z}$. For $k = 1$, the invariant distribution $\pi$ is a probability distribution on $\mathcal{Z}$ and is unique if the Markov chain has a single recurrent class (Gallager, 2013, Thm. 4.4.2). If, in addition, the Markov chain is aperiodic, then the distribution of $Z_n$ converges to this invariant distribution as $n \to \infty$ (Gallager, 2013, Thm. 4.3.7). A fortiori, uniqueness and convergence are ensured if the Markov chain is *regular*, i.e., irreducible and aperiodic (Kemeny and Snell, 1976, Thm. 4.1.6).

For $k$th order Markov chains, $k > 1$, two types of invariant distributions can be considered: A distribution $\pi$ on $\mathcal{Z}$ and a distribution $\mu$ on $\mathcal{Z}^k$.

The invariant distribution $\pi$ on $\mathcal{Z}$ is related to the eigenvector problem of nonnegative tensors. Chang et al. showed that the eigenvector associated with the largest eigenvalue of an irreducible tensor is positive, but not necessarily simple (Chang et al., 2008, Thm. 1.4). The results were used by Li and Ng in Li and Ng (2014) to derive conditions under which there exists a unique distribution $\pi$ on $\mathcal{Z}$ such that, for a $k$th order Markov chain $\mathbf{Z}$,

$$\forall z \in \mathcal{Z} : \quad \pi_z = \sum_{z_0,\dots,z_{k-1}\in\mathcal{Z}} \mathbb{P}(Z_k = z|Z_{k-1} = z_{k-1}, \dots, Z_0 = z_0)\pi_{z_{k-1}} \cdots \pi_{z_0}. \tag{1}$$

Regarding convergence, Vladimirescu showed that if $\mathbf{Z}$ is a regular second-order Markov chain, then there exists a distribution $\pi$ on $\mathcal{Z}$ such that (Kalpazidou, 2006, eq. (7.1.3))

$$\forall z, z_0, \dots, z_{k-1} \in \mathcal{Z} : \quad \pi_z = \lim_{n\to\infty} \mathbb{P}(Z_n = z|Z_{k-1} = z_{k-1}, \dots, Z_0 = z_0). \tag{2}$$

The more interesting case concerns invariant distributions $\mu$ on $\mathcal{Z}^k$. Following Doob (1990, p. 89), every $k$th order Markov chain $\mathbf{Z}$ on $\mathcal{Z}$ can be converted to a first-order Markov chain $\mathbf{Z}^{(k)} := (Z_n, \dots, Z_{n+k-1})_{n\in\mathbb{N}_0}$ on $\mathcal{Z}^k$. Kalpazidou used this fact to

*E-mail address:* bernhard.geiger@tum.de.

http://dx.doi.org/10.1016/j.spl.2017.07.006
0167-7152/

characterize invariant distributions for Markov chains $\mathbf{Z}$ derived from *weighted circuits* (Kalpazidou, 2006, Prop. 7.2.2). If $\mathbf{Z}^{(k)}$ is a regular first-order Markov chain, then $\mathbf{Z}$ is a regular $k$th order Markov chain (Kalpazidou, 2006, Prop. 7.1.6). Moreover, since in this case $\mathbf{Z}^{(k)}$ has a unique invariant distribution on $\mathcal{Z}^k$, so has $\mathbf{Z}$.

The converse, however, is not true: The regularity of $\mathbf{Z}^{(k)}$ does *not* follow from the regularity of $\mathbf{Z}$, and hence the uniqueness of an invariant distribution $\mu$ on $\mathcal{Z}^k$ is not guaranteed even for a regular $k$th order Markov chain $\mathbf{Z}$. To address this problem, Herkenrath discussed the uniform ergodicity of a second-order Markov process $\mathbf{Z}$ on a general alphabet $\mathcal{Z}$. Specifically, he defined the second-order Markov process $\mathbf{Z}$ to be uniformly ergodic if the first-order Markov process $\mathbf{Z}^{[2]} := (Z_{2n}, Z_{2n+1})_{n\in\mathbb{N}_0}$ on $\mathcal{Z}^2$ is uniformly ergodic (Herkenrath, 2003, Def. 4). Herkenrath showed how sufficient and/or necessary conditions for uniform ergodicity (such as a strengthened Doeblin condition) carry over from $\mathbf{Z}$ to $\mathbf{Z}^{[2]}$ (Herkenrath, 2003, Lem. 3). He moreover presented sufficient conditions for uniform ergodicity of certain classes of second-order Markov processes, such as nonlinear autoregressive time series with absolutely continuous noise processes (Herkenrath, 2003, Thms. 2-5). Herkenrath moreover showed a relation between the invariant distributions $\pi$ and $\mu$ on $\mathcal{Z}$ and $\mathcal{Z}^2$, respectively (Herkenrath, 2003, Cor. 2):

$$\forall z \in \mathcal{Z}: \quad \pi_z = \sum_{z'\in\mathcal{Z}} \mu_{z,z'} = \sum_{z'\in\mathcal{Z}} \mu_{z',z}. \tag{3}$$

In this work, we present a sufficient condition for a $k$th order homogeneous Markov chain $\mathbf{Z}$ on a finite alphabet $\mathcal{Z}$ to have a unique invariant distribution $\mu$ on $\mathcal{Z}^k$. The condition is formulated via a function of a first-order Markov chain $\mathbf{X}$ with a single recurrent class. Since functions of Markov chains, so-called *lumpings*, usually do not possess the Markov property, one may need to approximate this lumping by a Markov chain with a given order. Assuming that this Markov approximation satisfies certain conditions, it can be shown that its invariant distribution is unique. We moreover generalize this result by letting $\mathbf{X}$ be a higher-order Markov chain and a non-Markovian process, respectively.

## 2. Problem setting

We abbreviate vectors as $z_1^k := (z_1, \dots, z_k)$ and random vectors as $Z_1^k := (Z_1, \dots, Z_k)$. If the length of the vector is clear from the context, we omit indices. The probability that $Z = z$ is written as $p_Z(z) := \mathbb{P}(Z = z)$; the conditional probability that $Z_1 = z_1$ given that $Z_2 = z_2$ is written as $p_{Z_1|Z_2}(z_1|z_2)$. Stochastic processes are written as bold-faced letters, e.g., $\mathbf{Z} := (Z_n)_{n\in\mathbb{N}_0}$. We write sets with calligraphic letters. For example, the alphabet of $\mathbf{Z}$ is $\mathcal{Z}$. All processes and random variables are assumed to live on a finite alphabet, i.e., $|\mathcal{Z}| < \infty$. The complement of a set $\mathcal{A} \subseteq \mathcal{Z}$ is $\mathcal{A}^c := \mathcal{Z} \setminus \mathcal{A}$. Transition probability matrices are written in bold-face, too; whether a symbol is a matrix or a stochastic process will always be clear from the context. We naturally extend a function $g: \mathcal{Z} \to \mathcal{W}$ from scalars to vectors by applying it coordinate-wise, i.e., $g(z_1^k) := (g(z_1), \dots, g(z_k))$. Similarly, the preimage of a vector is the Cartesian product of the preimages, i.e., $g^{-1}(w_1^k) := g^{-1}(w_1) \times \cdots \times g^{-1}(w_k)$.

A stochastic process $\mathbf{Z}$ is a $k$th order Markov chain with alphabet $\mathcal{Z}$ if and only if

$$\forall n \geq k: \quad \forall z_0^n \in \mathcal{Z}^{n+1}: \quad p_{Z_n|Z_0^{n-1}}(z_n|z_0^{n-1}) = p_{Z_n|Z_{n-k}^{n-1}}(z_n|z_{n-k}^{n-1}). \tag{4}$$

If the right-hand side of (4) does not depend on $n$, we can write

$$Q_{z_{n-k}^{n-1}\to z_n} := p_{Z_n|Z_0^{n-1}}(z_n|z_0^{n-1}) \tag{5}$$

and call $\mathbf{Z}$ *homogeneous*. We let $\mathbf{Q}$ be a $|\mathcal{Z}|^k \times |\mathcal{Z}|$ matrix with entries $Q_{z_{n-k}^{n-1}\to z_n}$ and abbreviate $\mathbf{Z} \sim \mathsf{HMC}(k, \mathcal{Z}, \mathbf{Q})$. Similarly, we define

$$Q^{(n)}_{z_0^{k-1}\to z} := p_{Z_{k+n-1}|Z_0^{k-1}}(z|z_0^{k-1}) \tag{6}$$

and collect the corresponding values in the $|\mathcal{Z}|^k \times |\mathcal{Z}|$ matrix $\mathbf{Q}^{(n)}$. Note that $\mathbf{Q}^{(1)} = \mathbf{Q}$.

We recall basic definitions for $k$th order Markov chains $\mathbf{Z} \sim \mathsf{HMC}(k, \mathcal{Z}, \mathbf{Q})$; the details can be found in, e.g., Kalpazidou (2006, Def. 7.1.1-7.1.4) or Gallager (2013, Def. 4.2.2-4.2.7). A state $z \in \mathcal{Z}$ is *accessible* from $z' \in \mathcal{Z}$ (in short: $z' \to z$), if and only if

$$\forall u \in \mathcal{Z}^{k-1}: \quad \exists n = n(u, i, j): \quad Q^{(n)}_{(u,z')\to z} > 0. \tag{7}$$

If $z' \to z$ and $z \to z'$, then $z$ and $z'$ *communicate* (in short: $z \leftrightarrow z'$). A state $z$ is *recurrent* if $z \to z'$ implies $z' \to z$, otherwise $z$ is *transient*. The relation "$\leftrightarrow$" partitions the set $\{z \in \mathcal{Z}: \quad z \leftrightarrow z\}$ into equivalence classes (called recurrent classes). The Markov chain $\mathbf{Z}$ is *irreducible* if and only if $\mathcal{Z}$ is the unique recurrent class; it is *regular* if and only if there exists $n \geq 1$ such that $\mathbf{Q}^{(n)}$ is a positive matrix.

A $k$th order Markov chain $\mathbf{Z} \sim \mathsf{HMC}(k, \mathcal{Z}, \mathbf{Q})$ can be converted to a first-order Markov chain on $\mathcal{Z}^k$. Let $\mathbf{Z}^{(k)} := (Z_n^{n+k-1})_{n\in\mathbb{N}_0}$. Then, $\mathbf{Z}^{(k)} \sim \mathsf{HMC}(1, \mathcal{Z}^k, \mathbf{P})$, where (Doob, 1990, p. 89)

$$\forall z_0^{k-1}, z'^{k-1}_0 \in \mathcal{Z}^k: \quad P_{z'^{k-1}_0\to z_0^{k-1}} = \begin{cases} Q_{z'^{k-1}_0\to z_{k-1}} & z'_1 = z_0, \dots, z'_{k-1} = z_{k-2} \\ 0 & \text{else.} \end{cases} \tag{8}$$