J. Appl. Prob. **53,** 216–230 (2016)
*doi:*10.1017/jpr.2015.20
© *Applied Probability Trust* 2016

CONTINUITY PROPERTIES OF A FACTOR OF MARKOV CHAINS

WALTER A. F. DE CARVALHO,* *Centro Universitário Adventista de São Paulo*
SANDRO GALLO,** *Universidade Federal de São Carlos*
NANCY L. GARCIA,*** *Universidade Estadual de Campinas*

Abstract

Starting from a Markov chain with a finite or a countable infinite alphabet, we consider the chain obtained when all but one symbol are indistinguishable for the practitioner. We study conditions on the transition matrix of the Markov chain ensuring that the image chain has continuous or discontinuous transition probabilities with respect to the past.

Keywords: Factor map; continuity rate; chains of infinite order

2010 Mathematics Subject Classification: Primary 60K15
Secondary 60J10; 60G10

1. Introduction

Consider a positive recurrent, stationary Markov chain $\boldsymbol{X} = \{X_t, t \in \mathbb{Z}\}$ with a finite (or countable) alphabet A with transition matrix $P^{\boldsymbol{X}}$ and a factor map, $\pi : A \to B$ with $\text{card}(B) < \text{card}(A)$. Hidden Markov models obtained through factor mappings of Markov chains have received considerable attention over many years and the literature is exhaustive. We therefore mention only the papers that are most directly related to our concern. In general, (order 1) Markov property is not preserved. This problem was addressed by many authors, we refer the reader, for instance, to Burke and Rosenblatt (1958) who gave necessary and sufficient conditions under which the image process $\boldsymbol{Y} = \{Y_t = \pi(X_t), t \in \mathbb{Z}\}$ is still a Markov process. Harris (1955) proved that, in general, the process $\boldsymbol{Y}$ is a chain with infinite order with continuous transition probabilities decaying exponentially fast if $P^{\boldsymbol{X}}$ is strictly positive (i.e. $P^{\boldsymbol{X}}_{ij} > 0$ for all $i, j \in A$). This result was then extended in Verbitskiy (2011a), (2011b), to the case where the original chain $\boldsymbol{X}$ is not necessarily a Markov chain, but only needs to be a chain with strictly positive and 'strongly' continuous transition probabilities (rapidly decaying continuity rate). Chazottes and Ugalde (2003) and Yoo (2010) studied a related problem in the statistical physics and dynamical systems contexts. They found conditions under which the image process of a Markov process satisfies (Bowen's) Gibbsianess. A nice reference reviewing several aspects of hidden Markov models in symbolic dynamic systems is Boyle and Petersen (2011).

Received 23 September 2014; revision received 1 June 2015.
* Postal address: Alameda Pitangueiras, 231, Engenheiro Coelho, São Paulo, 13.165-000, Brazil.
Email address: walter.carvalho@unasp.edu.br
** Postal address: Departamento de Estatística - UFSCar, Rodovia Washington Luís, s/n - Jardim Guanabara, São Carlos, São Paulo, 13565-905, Brazil. Email address: sandro.gallo@ufscar.br
*** Postal address: Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas, Rua Sergio Buarque de Holanda, 651, Campinas, São Paulo, 13083-859, Brazil. Email address: nancy@ime.unicamp.br

https://doi.org/10.1017/jpr.2015.20 Published online by Cambridge University Press

In this paper we focus our efforts on the special case where $B = \{0, 1\}$ and $\pi^{-1}(1)$ is a singleton. Without loss of generality, we take $A = \{1, \ldots, m\}$ (or $A = \{1, 2, \ldots\}$) and $\pi(x) = \mathbf{1}_{\{x=1\}}$. In this case, it is well known that the image process $\boldsymbol{Y} = \{Y_t, t \in \mathbb{Z}\}$ is always a discrete-time binary renewal chain, and it is an easy matter to obtain an explicit matricial form for its transition probabilities. This explicitness then allows us to search not only for sufficient conditions (as in most of the previously cited articles), but also, in some cases, for necessary conditions under which the image process has continuous transition probabilities.

Consider the following decomposition of the transition matrix $P^{\boldsymbol{X}}$:

$$P^{\boldsymbol{X}} = \begin{pmatrix} P_{11} & V \\ W^\top & P \end{pmatrix},$$

where V and W are row vectors and $W^\top$ denotes the transpose of W, and P is the sub-matrix $(P^{\boldsymbol{X}}_{ij})_{i,j\neq 1}$. To avoid trivial cases, we will always consider $V \neq \mathbf{0}$ and $W \neq \mathbf{0}$. Our conditions will be given by assumptions on the sub-matrix P. We consider three cases.

1. *When A is finite and P is irreducible.* In this case, we can always verify if the image chain has continuous (or discontinuous) transition probabilities.
   - If P is primitive (period equals 1) then $\boldsymbol{Y}$ has a continuous kernel and we give an upper bound for the continuity rate.
   - If P has period $h \geq 2$ then $\boldsymbol{Y}$ has continuous transition probabilities if and only if

     $$\frac{VG^*P^rW^\top}{VG^*P^r\mathbf{1}^\top} = \frac{VG^*W^\top}{VG^*\mathbf{1}^\top} \quad \text{for all } r = 1, \ldots, h-1,$$

     where $G^* = \phi^*(\psi^*)^\top > 0$, and ϕ^* and ψ^* are eigenvectors to the largest eigenvalue from P^h and $(P^h)^\top$, respectively.
2. *When A is finite and P is reducible.* Using the same tools for the irreducible cases, we obtain necessary conditions for the domination blocks matrices to ensure continuity.
3. *When A is infinite and P is irreducible.* If P is aperiodic and positive recurrent then $\boldsymbol{Y}$ has continuous transition probabilities.

Section 3 contains a more detailed explanation of our assumptions and results.

The paper is organized as follows. In Section 2 we introduce the notation and basic definitions. The main results are given in Section 3. The proofs are presented in Section 4. We conclude with a section containing some examples and a discussion comparing our results to those of the related literature.

2. Notation and basic definitions

Let A be a countable alphabet. Given two integers $m \leq n$, we denote by a_m^n the string $a_m \cdots a_n$ of symbols in A. For any $m \leq n$, the length of the string a_m^n is denoted by $|a_m^n|$ and defined by $n - m + 1$. Given two strings v and v', we denote by vv' the string of length $|v| + |v'|$ obtained by concatenating the two strings. The concatenation of strings is also extended to the case where $v = \cdots a_{-2}a_{-1}$ is a semi-infinite sequence of symbols. Let $A^{-\mathbb{N}} = A^{\{\ldots,-2,-1\}}$ be the set of all infinite strings of past symbols. Finally, we denote by $\underline{a} = \cdots a_{-2}a_{-1}$ the elements of $A^{-\mathbb{N}}$.

https://doi.org/10.1017/jpr.2015.20 Published online by Cambridge University Press

2.1. Kernel, compatibility, and continuity

Definition 2.1. (*Kernel.*) A *transition probability kernel* (or simply *kernel* in the sequel) on a countable alphabet A is a (measurable) function

$$P\colon A \times A^{-\mathbb{N}} \to [0,1], \qquad (a, \underline{a}) \mapsto P(a \mid \underline{a})$$

such that $\sum_{a\in A} P(a \mid \underline{a}) = 1$ for any $\underline{a} \in A^{-\mathbb{N}}$.

For instance, a given kernel P is Markovian of order $k \geq 1$ if $P(a \mid \underline{a}) = P(a \mid \underline{b})$ for any $\underline{a}$ and $\underline{b}$ such that $a_{-k}^{-1} = b_{-k}^{-1}$. So the above definition is a natural extension of the concept of a transition matrix, the latter being a terminology commonly used for k-step Markov chains.

A kernel is sometimes called a set of transition probabilities, a terminology that we used in the introduction. It is important to note that these are transition probabilities with respect to the 'past'. When considering transition probabilities with respect to both past and future, as we will do in Section 5 when comparing our results to the concept of Gibbsianess, we will use the terminology *specification*.

In this paper stochastic chains are regarded as a bi-infinite sequence of random variables whose dependence information is given by a kernel. We therefore need to define what we mean by a stochastic chain being compatible with a given kernel.

Definition 2.2. (*Compatibility.*) A stationary stochastic chain $\boldsymbol{X} = (X_n)_{n\in\mathbb{Z}}$ is said to be *compatible* with a kernel P if the later is a regular version of the conditional probabilities of the former, that is

$$\mathbb{P}(X_0 = a_0 \mid X_{-\infty}^{-1} = a_{-\infty}^{-1}) = P(a \mid a_{-\infty}^{-1})$$

for every $a_0 \in A$ and $\mathbb{P}$-almost everywhere (a.e.) $a_{-\infty}^{-1}$ in $A^{-\mathbb{N}}$.

Let us emphasize that stationarity here means that for any $n \geq 1$ and any a_0^{n-1},

$$\mathbb{P}(X_0^{n-1} = a_0^{n-1}) = \mathbb{P}(X_k^{k+n-1} = a_0^{n-1}) \quad \text{for any } k \in \mathbb{Z}.$$

As an example for Definition 2.2, a stationary k-step Markov chain satisfies $\mathbb{P}(X_0 = a_0 \mid X_{-k}^{-1} = a_{-k}^{-1}) = P(a_0 \mid \underline{a})$, since this later does not depend on $a_{-\infty}^{-k-1}$. In what follows, $P_{ij}, i, j \in A$, will always denote the entries of a transition matrix on A.

We finally need to introduce what we mean by a continuous kernel, the main concept of interest in this paper.

Definition 2.3. (*Continuity.*) A kernel P is *continuous* (with respect to the product topology) at some point $\underline{a}$ if $P(a \mid a_{-i}^{-1}\underline{z}) \to P(a \mid \underline{a})$ as $i \to +\infty$ for any $\underline{z}$. Naturally, a kernel P is *continuous* if it is continuous at every point.

In the above definition, $P(a \mid a_{-i}^{-1}\underline{z})$ stands for the transition probability from the (concatenated) past $\underline{z}a_{-i}^{-1}$ to the symbol a, we reversed time in the conditioning.

It is clear that k-step Markov kernels are continuous since for them,

$$P(a \mid a_{-i}^{-1}\underline{z}) = P(a \mid a_{-k}^{-1}\underline{z}) \quad \text{for any } i \geq k.$$

So continuous kernels constitute a natural extension of k-step Markov kernels. The next definition aims to quantify *to what extent* a given kernel is continuous.

https://doi.org/10.1017/jpr.2015.20 Published online by Cambridge University Press

Definition 2.4. (*Continuity rate.*) The continuity rate of a kernel P is defined, for any $k \geq 1$, by

$$\beta(k) := \sup_{a_{-k}^0} \sup_{\underline{b},\underline{c}} |P(a_0 \mid a_{-k}^{-1}\underline{b}) - P(a_0 \mid a_{-k}^{-1}\underline{c})|.$$

Observe that if the alphabet A is finite, the compactness of $A^{-\mathbb{N}}$ implies that $P(\cdot \mid \cdot)$ is continuous if and only if its continuity rate converges to 0.

3. Results

3.1. Preliminary considerations

One of the main assumptions of this paper is that we restrict our considerations to a particular case of mapping.

Definition 3.1. (*Aggregation map.*) A map $\pi : A \to \{0, 1\}$ defined by $\pi(a) = \mathbf{1}_{\{a=1\}}$ will be called an aggregation map.

Notation Alert. We will use the same notation π to denote the map $\pi : A^n \to \{0, 1\}^n$ defined by $\pi(a_1^n)_i = \mathbf{1}_{\{a_i=1\}}$. The generalization for infinite sequences and processes is analogous.

In what follows, we will essentially study two stochastic chains, the original Markov chain $\boldsymbol{X}$ and its image $\boldsymbol{Y}$. In order to avoid confusion, we will index quantities by the related stochastic chains: $P^{\boldsymbol{X}}$ and $P^{\boldsymbol{Y}}$ for the kernels, and $\beta^{\boldsymbol{X}}(k)$ and $\beta^{\boldsymbol{Y}}$ for the continuity rates. Also, the P will denote a sub-matrix of $P^{\boldsymbol{X}}$.

Focussing on aggregation maps (see Definition 3.1) allows us to obtain a closed-form equation for the continuity rate of the image process, as stated in the next proposition.

For any $\underline{a} \in \{0, 1\}^{-\mathbb{N}}$, let

$$\ell(\underline{a}) := \inf\{i \geq 0 \colon a_{-i-1} = 1\},$$

using the convention that $\ell(\cdots 00) = \infty$. Also let us introduce the $[0, 1]$-valued sequence $p_k, k \geq 0$, defined by $p_0 := P_{11} = \mathbb{P}(X_0 = 1 \mid X_{-1} = 1)$ and for any $k \geq 1$,

$$p_k := \frac{\sum_{b_{-k}^{-1} \in \pi^{-1}(0_{-k}^{-1})} \mathbb{P}(X_{-k}^{-1} = b_{-k}^{-1}, X_0 = 1 \mid X_{-k-1} = 1)}{\sum_{b_{-k}^{-1} \in \pi^{-1}(0_{-k}^{-1})} \mathbb{P}(X_{-k}^{-1} = b_{-k}^{-1} \mid X_{-k-1} = 1)}. \tag{3.1}$$

Proposition 3.1. *If $\boldsymbol{X}$ is a stationary positive recurrent Markov chain then the image process $\boldsymbol{Y}$ is such that*

$$\mathbb{P}(Y_0 = 1 \mid Y_{-k}^{-1} = 0_{-k}^{-1}, Y_{-k-1} = 1, Y_{-i-k-1}^{-k-2} = a_{-i}^{-1}) = p_k \tag{3.2}$$

for any $k \geq 0$, $i \geq 1$ and any $a_{-i}^{-1} \in \{0, 1\}^i$. Moreover,

$$\beta^{\boldsymbol{Y}}(k) \leq \sup_{l,m \geq k} |p_l - p_m|$$

and thus $\boldsymbol{Y}$ has a continuous kernel if and only if p_k converges.

Remark 3.1. Equation (3.2) implies, in particula,r that in the realization of the process $\boldsymbol{Y}$, the distances between consecutive occurrences of 1s are independent and identically distributed (i.i.d.) with distribution $(f_n)_{n\geq 1}$, where $f_n = p_{n-1} \prod_{i=0}^{n-2}(1 - p_i)$ (with the convention that $\prod_{i=0}^{-1}(1 - q_i) = 1$). Such processes are called 'renewal processes' in the literature (the process 'starts afresh' at each occurrence of a 1) and the distribution $(f_n)_{n\in\mathbb{N}}$ is called the interarrival distribution of the process.

https://doi.org/10.1017/jpr.2015.20 Published online by Cambridge University Press

Remark 3.2. In the proof of this proposition (see Section 4), we will see that the value of $P^{\boldsymbol{Y}}(1 \mid \cdots 00)$ can be chosen arbitrarily. For instance, if the sequence $(p_k)_{k\geq 1}$ converges, we could have taken a different value than the limit of p_k, creating a discontinuity at $\cdots 00$ according to Definition 2.3. But this is an artificial discontinuity, because the regular version $P^{\boldsymbol{Y}}$ is defined up to sets of null measure. Thus, in this case, we will always choose the continuous one.

In certain cases however, it is impossible to make $P^{\boldsymbol{Y}}$ continuous just by changing its value on a set of null measure. For instance, let $\boldsymbol{X}$ be a Markov chain with alphabet $A = \{1, 2, 3\}$ and transition matrix given by

$$P^{\boldsymbol{X}} = \begin{pmatrix} \alpha & 1-\alpha & 0 \\ \beta & 0 & 1-\beta \\ \gamma & 1-\gamma & 0 \end{pmatrix}.$$

Let $\pi : A \to \{0, 1\}$ be such that $\pi(1) = 1, \pi(2) = \pi(3) = 0$. In this case, simple calculations show that $p_0 = \alpha$ for $k \geq 1$ odd, $p_k = \beta$ and for $k \geq 2$ even, $p_k = \gamma$. What happens then is, borrowing the terminology used in the Gibbs/non-Gibbs literature, that the kernel $P^{\boldsymbol{Y}}$ has an *essential discontinuity* at the point $\cdots 00$, independently of the value $P(1 \mid \cdots 00)$ we choose. We refer the reader to Fernández *et al.* (2011) for a complete discussion.

The main objective of this paper is to understand which condition for $P^{\boldsymbol{X}}$ leads to a renewal process with a continuous/discontinuous kernel. To do so, Proposition 3.1 says that it suffices to obtain sufficient conditions for p_k to converge (or not). To study p_k, we are going to use a decomposition of the transition matrix $P^{\boldsymbol{X}}$ similar to the one used by Darroch and Seneta (1967) in their study of quasi-stationary measures. This decomposition will allow us to use the Perron–Frobenius theory.

Recall that $P^{\boldsymbol{X}}$ is an $m \times m$ (m could be infinite) stochastic matrix. Let

$$P^{\boldsymbol{X}} = \begin{pmatrix} P_{11} & V \\ W^{\top} & P \end{pmatrix}, \tag{3.3}$$

where

- V is the row vector $(P_{12} \cdots P_{1m})$;
- W is the row vector $(P_{21} \cdots P_{m1})$, and $W^{\top}$ denotes the transpose of W;
- P is the $(m-1) \times (m-1)$ sub-matrix P_{ij}, $i, j \in \{2, \ldots, m\}$.

Denote by $\mathbf{1}$ and $\mathbf{0}$ the row vectors formed by 1s and 0s, respectively. To avoid trivial cases, we will always consider $V \neq \mathbf{0}$ and $W \neq \mathbf{0}$.

Notation alert. In what follows, P will always denote the sub-matrix of $P^{\boldsymbol{X}}$. The kernels will always be indexed by the corresponding process.

Recall that $p_0 = P_{11}$. We can now write $p_k, k \geq 1$ (see (3.1)) in the following form:

$$p_{k+1} = \frac{V P^k W^{\top}}{V P^k \mathbf{1}^{\top}} \quad \text{if } k \geq 0$$

with $P^0 = I$ being the identity $(m-1) \times (m-1)$ matrix.

Before we come to the statement of the results, let us recall the following very basic definitions concerning matrices.

https://doi.org/10.1017/jpr.2015.20 Published online by Cambridge University Press

Definition 3.2. A square matrix (P_{ij}) is called *irreducible* if, for any i, j there exists an $n(i, j)$ such that $P_{ij}^n > 0$. Otherwise, it is called *reducible*.

Definition 3.3. A square matrix (P_{ij}) is called *primitive* if it is nonnegative and if there is an $n > 0$ such that $P_{ij}^n \neq 0$ for all i, j.

Definition 3.4. Let (P_{ij}) be nonnegative. Fix a state i and define the *period of state* i as

$$h(i) := \text{gcm}\{n\colon P_{ii}^n > 0\}.$$

In the case of irreducible matrices, every state has the same period h and this number is called the *period of* P. If $h = 1$, P is called *aperiodic*.

Thus, primitive matrices are irreducible aperiodic nonnegative matrices.

3.2. The finite irreducible case

Let $\boldsymbol{X}$ be a stationary Markov process with finite space state $A = \{1, 2, \ldots, m\}$ and its transition matrix $P^{\boldsymbol{X}}$ decomposed as (3.3). Let $\boldsymbol{Y} = \pi(\boldsymbol{X})$, where π is an aggregation map. If P is irreducible, Theorem 3.1 states which are the conditions for the process $\boldsymbol{Y}$ to have a continuous kernel.

Theorem 3.1. *Let P be irreducible.*

1. *If P is primitive then $\boldsymbol{Y}$ has a continuous kernel. Moreover,*

$$\beta^{\boldsymbol{Y}}(n) = O\left(n^{d_2-1}\left(\frac{|\lambda_2|}{\lambda_1}\right)^n\right),$$

   where λ_1 is the maximal eigenvalue of P, λ_2 and d_2 are, respectively, the value and multiplicity of the second largest eigenvalue in absolute value.

2. *If P has period $h \geq 2$ then the process $\boldsymbol{Y}$ has a continuous kernel if and only if*

$$\frac{VG^*P^rW^\top}{VG^*P^r\mathbf{1}^\top} = \frac{VG^*W^\top}{VG^*\mathbf{1}^\top} \quad \text{for all } r = 1, \ldots, h-1, \tag{3.4}$$

   where $G^* = \phi^*.(\psi^*)^\top > 0$, and ϕ^* and ψ^* are the Perron vectors corresponding to the Perron value λ_1^* from P^h and $(P^h)^\top$, respectively.

In other words, Theorem 3.1 states that, in the case where P is irreducible, the if and only if condition for the image process to have a continuous kernel is that either P is primitive or (3.4) holds for $r = 1, \ldots, h-1$.

In some cases the aggregated process has finite memory, for example it is known that if the sum of all rows of P are equal then the resulting process is Markovian; see Burke and Rosenblatt (1958). It is possible to obtain higher order Markov processes as shown by the following proposition.

Proposition 3.2. *If P is irreducible with a single nonnull eigenvalue then*

(i) *the process $\boldsymbol{Y}$ is a Markov process with order at most $m-1$;*

(ii) *the process $\boldsymbol{Y}$ is a Markov process with order $k+1$ if P has k nonnull rows in its triangular form.*

https://doi.org/10.1017/jpr.2015.20 Published online by Cambridge University Press

Note that this proposition does not cover the usual order 1 Markov case since Δ cannot have 0 nonnull rows (the second largest eigenvalue is equal to 0, therefore at least one row of Δ is equal to 0).

3.3. The finite reducible case

Without loss of generality, consider that the transition submatrix P can be written as

$$P = \begin{pmatrix} T_{11} & T_{12} & T_{13} & \cdots & T_{1c} \\ \mathbf{0} & T_{22} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & T_{33} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & T_{cc} \end{pmatrix}, \tag{3.5}$$

where the submatrices $T_{22}, T_{33}, \ldots, T_{cc}$ are irreducible.

The reducible case is more difficult to analyse in full generality since the matrices $T_{22}, T_{33}, \ldots, T_{cc}$ can be periodic with different periods. However, this situation can be studied case by case using the same tools as described before and in some special cases, we can obtain sufficient conditions for continuity.

Let $\lambda_{1,i}$ be the Perron value of T_{ii} and define $\lambda_{\max} = \max_i \lambda_{1,i}$.

Definition 3.5. The matrices $T_{ii}, i \in \{2, 3, \ldots, c\}$ with $\lambda_i = \lambda_{\max}$ are called *dominating blocks*.

Let

$$\Delta = \{i \in \{2, \ldots, c\} : T_{ii} \text{ is a dominating block}\}.$$

The proof of the following theorem is omitted as it is analogous to the proof of Theorem 3.1.

Theorem 3.2. *If the matrix P is reducible and written as (3.5) then sufficient conditions for the process $\mathbf{Y}$ to have a continuous kernel are either that all dominating blocks (T_{ii} for $i \in \Delta$) are irreducible, or that they are all periodic with the same period h and, for all $i \in \Delta$,*

$$\frac{V_i G_i (T_{ii})^r W_i^\top}{V_i G_i (T_{ii})^r \mathbf{1}_i^\top} = \frac{V_i G_i W_i^\top}{V_i G_i \mathbf{1}_i^\top} \quad \text{for all } r = 0, \ldots, h-1$$

with $G_i = \phi_i \psi_i^\top > 0$. Here ϕ_i and ψ_i are the right and left Perron vectors corresponding to $\lambda_{1,i}$ from T_{ii}^h.

3.4. Countable alphabet

The Perron–Frobenius theory can be used for countable, real, nonnegative matrices. The definitions of irreducibility and periodicity extend easily to such matrices. Moreover, we can define the Perron-value of a denumerable matrix using the following lemma.

Lemma 3.1. (Kitchens (1998, Lemma 7.1.1).) *Let P be a countable, real, nonnegative, irreducible, and aperiodic matrix. For a fixed state i:*

(i) *there exists a k such that $(P^n)_{ii} > 0$ for all $n \geq k$;*

(ii) $(P^{n+m})_{ii} \geq (P^n)_{ii} \cdot (P^m)_{ii}$;

(iii) $\lim_{n\to\infty} \sqrt{n}(P^n)_{ii}$ *exists and equals* $\sup_n \sqrt{n}(P^n)_{ii}$;

(iv) *if* $\lambda = \lim_{n\to\infty} \sqrt{n}(P^n)_{ii}$ *then* $(P^n)_{ii}/\lambda^n \leq 1$ *for all n.*

https://doi.org/10.1017/jpr.2015.20 Published online by Cambridge University Press

Then, λ is the *Perron value* of P. It can be infinite. For this section, we will assume it is finite. However, this is not enough to have the desired results in terms of continuity. We will need positive recurrence. This concept is well understood for stochastic matrices in terms of the mean return time to the states. However, for general nonnegative matrices the following definitions are necessary. Let

$$t_{ij}(0) = \delta_{ij}, \qquad t_{ij}(1) = P_{ij}, \quad \text{and} \quad t_{ij}(n) = (P^n)_{ij}.$$

Define the generating function as

$$T_{ij}(z) = \sum_{n=0}^{\infty} t_{ij}(n) z^n. \tag{3.6}$$

Now, let

$$\ell_{ij}(0) = 0, \qquad \ell_{ij}(1) = P_{ij}, \quad \text{and} \quad \ell_{ij}(n+1) = \sum_{r \neq i} \ell_{ir}(n) t_{rj}.$$

Observe that the radius of convergence of $T_{ij}(z)$ is $1/\lambda$ for all i, j.

Definition 3.6. The irreducible matrix P corresponding to the generating function (3.6) is:

(i) *recurrent* if $T_{ii}(1/\lambda) = \infty$;

(ii) *transient* if $T_{ii}(1/\lambda) < \infty$;

(iii) *positive recurrent* if $T_{ii}(1/\lambda) = \infty$ and $\sum_n n\ell_{ii}(n)/\lambda^n < \infty$;

(iv) *null recurrent* if $T_{ii}(1/\lambda) = \infty$ and $\sum_n n\ell_{ii}(n)/\lambda^n = \infty$.

Note that the definition is independent of the choice of i.

The generalized Perron–Frobenius theorem states the following.

Theorem 3.3. (Kitchens (1998, Theorem 7.1.3).) *Suppose that P is a countable, nonnegative, irreducible, aperiodic, and positive recurrent. Then there exists a Perron value $\lambda > 0$ (assumed to be finite) such that $\lim_{n\to\infty} P^n/\lambda^n = \phi \cdot \psi^\top > 0$, and ϕ and ψ are the Perron vectors corresponding to λ from P and $P^\top$, respectively.*

Therefore, we obtain the following result in the countably infinite case.

Theorem 3.4. *If the matrix P is countable, nonnegative, irreducible, aperiodic, and positive recurrent matrix, then the process $\boldsymbol{Y}$ has a continuous kernel.*

4. Proofs

Proof of Proposition 3.1. Equation (3.2) follows easily by the definition of conditioning, noting that π is the aggregating function of Definition 3.1 and that $\boldsymbol{X}$ is Markovian. In particular, it means that, for any $\underline{a} \in \{0,1\}^{-\mathbb{N}} \setminus \{\cdots 00\}$, $\mathbb{P}(Y_0 = 1 \mid Y_{-k}^{-1} = a_{-k}^{-1})$ converges, since it is constant for any $k \geq \ell(\underline{a}) + 1$. By Kalikow (1990), $\boldsymbol{Y}$ has a continuous kernel $P^{\boldsymbol{Y}}$ if and only if $\mathbb{P}(Y_0 = 1 \mid Y_{-k}^{-1} = a_{-k}^{-1})$ converges uniformly on $\{0,1\}^{-\mathbb{N}}$, which in the present case, due to the compactness of $\{0,1\}^{-\mathbb{N}}$, amounts to saying that $P^{\boldsymbol{Y}}$ is continuous if and only if $\mathbb{P}(Y_0 = 1 \mid Y_{-k}^{-1} = a_{-k}^{-1})$ converges for any $\underline{a} \in \{0,1\}^{-\mathbb{N}}$. Since this quantity converges for any $\underline{a} \in \{0,1\}^{-\mathbb{N}} \setminus \{\cdots 00\}$, we have continuity if and only if the convergence occurs at $\cdots 00$.

https://doi.org/10.1017/jpr.2015.20 Published online by Cambridge University Press

Now, by the definition of compatibility, $p_\infty = P^{\mathbf{Y}}(1 \mid \cdots 00)$ can be defined arbitrarily. We therefore define $p_\infty = \lim_k p_k$ (which is natural) if the limit exists and $p_\infty = 0$ otherwise (which is arbitrary).

Then we can compute the continuity rate of $P^{\mathbf{Y}}$:

$$\begin{aligned}\beta^{\mathbf{Y}}(k) &:= \sup_{a_{-k}^{-1}} \sup_{\underline{b},\underline{c}} |P^{\mathbf{Y}}(1 \mid a_{-k}^{-1}\underline{b}) - P^{\mathbf{Y}}(1 \mid a_{-k}^{-1}\underline{c})| \\ &= \sup_{\underline{b},\underline{c}} |P^{\mathbf{Y}}(1 \mid 0_{-k}^{-1}\underline{b}) - P^{\mathbf{Y}}(1 \mid 0_{-k}^{-1}\underline{c})| \\ &= \sup_{\underline{b},\underline{c}} |p_{\ell(\underline{b})+k} - p_{\ell(\underline{c})+k}| \\ &= \sup_{l,m\geq k} |p_l - p_m|,\end{aligned}$$

where in the second equality we used the fact that if $a_{-i} = 1$ for some $i \in \{1, \ldots, k\}$, then as we said above $P^{\mathbf{Y}}(1 \mid a_{-k}^{-1}\underline{b}) = P^{\mathbf{Y}}(1 \mid a_{-k}^{-1}\underline{c}) = p_i$, and in the third equality, we used the fact that the transition probability $P^{\mathbf{Y}}(1 \mid \underline{a})$ only depends on $\ell(\underline{a})$. □

Proof of Theorem 3.1 when P is primitive. There exists a unique maximal eigenvalue λ_1 (Perron–Frobenius eigenvalue or Perron root), which is strictly greater in absolute value than all other eigenvalues and λ_1 is the spectral radius of P. Therefore,

$$\lim_{n\to\infty}\left(\frac{P}{\lambda_1}\right)^n = G, \qquad p_\infty = \lim_{n\to\infty} p_n = \frac{VGW^\top}{VG\mathbf{1}^\top},$$

where $G = \phi.\psi^\top > 0$, and ϕ and ψ are the Perron vectors corresponding to λ_1 from P and $P^\top$, respectively.

Moreover, let λ_2 be the second largest eigenvalue in absolute value and d_2 its multiplicity. Then, it is known (Seneta (2006, Theorem 1.2, p. 9)) that

$$\left(\frac{P}{\lambda_1}\right)^n = G + O\left(n^{d_2-1}\left(\frac{|\lambda_2|}{\lambda_1}\right)^n\right).$$

Therefore,

$$\beta^{\mathbf{Y}}(n) = |p_n - p_\infty| = \left|\frac{VP^nW^\top}{VP^n\mathbf{1}^\top} - \frac{VGW^\top}{VG\mathbf{1}^\top}\right| = O\left(n^{d_2-1}\left(\frac{|\lambda_2|}{\lambda_1}\right)^n\right). \qquad \square$$

Proof of Theorem 3.1 when P is not primitive. Without loss of generality we can relabel the symbols on A such that P can be written in the Frobenius form decomposed into primitive blocks as

$$P = \begin{pmatrix} \mathbf{0} & B_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & B_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & B_{h-1} \\ B_h & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix}.$$

Therefore, P^h will be a primitive matrix given by

$$P^h = \begin{pmatrix} B_1.B_2 \ldots B_h & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & B_2.B_3 \ldots B_h, B_1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & B_h.B_1 \ldots B_{h-1} \end{pmatrix}$$

https://doi.org/10.1017/jpr.2015.20 Published online by Cambridge University Press

and, we have

$$\lim_{n\to\infty}\left(\frac{P^h}{\lambda_1^*}\right)^n = G^*$$

with $G^* = \phi.\psi^\top > 0$, and ϕ and ψ are the right and left Perron vectors corresponding to λ_1 from P^h, respectively. Since

$$p_{n.h+1} = \frac{V P^{n.h} W^\top}{V P^{nh}\mathbf{1}^\top} = \frac{V(P^h/\lambda_1^*)^n W^\top}{V(P^h/\lambda_1^*)^n\mathbf{1}^\top}$$

and

$$p_{n.h+r+1} = \frac{V P^{n.h+r} W^\top}{V P^{n.h+r}\mathbf{1}^\top} = \frac{V(P^h/\lambda_1^*)^n P^r W^\top}{V(P^h/\lambda_1^*)^n P^r\mathbf{1}^\top},$$

it follows that

$$\lim_{n\to\infty} p_{n.h+r+1} = \lim_{n\to\infty}\frac{V(P^h/\lambda_1^*)^n P^r W^\top}{V(P^h/\lambda_1^*)^n P^r\mathbf{1}^\top} = \frac{V G^* P^r W^\top}{V G^* P^r\mathbf{1}^\top}. \tag{4.1}$$

Therefore, in order to have continuity of the kernel, we need (4.1) to be constant for all $r = 0,\ldots,h-1$. □

Proof of Proposition 3.2. Under the stated conditions then λ_1 is the only nonnull eigenvalue and by (Seneta (2006, Theorem 1.2, p. 9)), we have

$$P^n = \lambda_1^n\phi\psi^\top \quad \text{for all } n \geq m-2.$$

Therefore, for any $n \geq m-2$,

$$p_{n+1} = \frac{V P^n W^\top}{V P^n\mathbf{1}^\top} = \frac{V\lambda_1^n\phi\psi^\top W^\top}{V\lambda_1^n\phi\psi^\top\mathbf{1}^\top} = \frac{V G W^\top}{V G\mathbf{1}^\top}.$$

Moreover, according to Schur's triangularization theorem, if all eigenvalues of P are real then P can be triangularized, that is, there exist Q and Δ, such that

$$P = Q\Delta Q^\top \quad \text{and} \quad P^n = Q\Delta^n Q^\top,$$

where Δ is the triangular superior and Q is orthogonal. Also, Δ is composed of the eigenvalues in its diagonal. Suppose that Δ has k nonnull rows. Then it is easy to see that, for $n \geq k$, $\Delta^n = \lambda_1^{n-k}.\Delta^k$. Now, for $n \geq k$,

$$p_{n+1} = \frac{V Q\Delta^n Q^\top W^\top}{V Q\Delta^n Q^\top\mathbf{1}^\top} = \frac{\lambda_1^{n-k} V Q\Delta^k Q^\top W^\top}{\lambda_1^{n-k} V Q\Delta^k Q^\top\mathbf{1}^\top} = \frac{V Q\Delta^k Q^\top W^\top}{V Q\Delta^k Q^\top\mathbf{1}^\top} = p_{k+1}.$$

Therefore, the process $\boldsymbol{Y}$ is an order $k+1$ Markov chain. □

5. Discussion

5.1. Examples

Example 5.1. (*All entries positive—comparing with Harris' rate of convergence.*) Consider the Markov chain $\boldsymbol{X}$ with alphabet $A = \{1,2,3\}$ and transition matrix given by

$$P^X = \begin{pmatrix} 0.10 & 0.30 & 0.60 \\ 0.20 & 0.30 & 0.50 \\ 0.05 & 0.70 & 0.25 \end{pmatrix}.$$

https://doi.org/10.1017/jpr.2015.20 Published online by Cambridge University Press

Note that the process $\boldsymbol{Y}$ is not Markov since W is not constant. In this case, it is easy to compute the eigenvalues and eigenvectors of P and P' corresponding to the largest eigenvalue obtaining that $p_\infty = \lim_{n\to\infty} p_n = 0.132\,864$ and the rate of convergence is bounded by

$$\beta^{\boldsymbol{Y}}(n) = O\left(\left(n\frac{|\lambda_2|}{\lambda_1}\right)^n\right) = O(n0.366^n).$$

Note that all entries of $P^{\boldsymbol{X}}$ are positive and the continuity rate given in Harris (1955, Equation (6.4)) gives the following upper bound for the continuity rate:

$$\beta^{\boldsymbol{Y}}(n) \le (1-\lambda)^{n-1},$$

where $\lambda = \min_{i,j,k,l}(p_{kj}\,p_{il}/|A|^2 p_{ij}\,p_{kl})$. In this case, $\lambda = 0.012$ leading to a substantially worse estimation of the continuity rate.

Example 5.2. (*Where P is reducible and $\boldsymbol{Y}$ has a continuous kernel.*) Consider the Markov chain $\boldsymbol{X}$ with alphabet $A = \{1, 2, \ldots, 7\}$ and transition matrix given by

$$P^{\boldsymbol{X}} = \begin{pmatrix} 0.2 & 0.3 & 0.1 & 0.2 & 0.1 & 0.0 & 0.1 \\ 0.2 & 0.3 & 0.2 & 0.1 & 0.0 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.2 & 0.0 & 0.3 & 0.1 & 0.2 \\ 0.0 & 0.0 & 0.0 & 0.4 & 0.6 & 0.0 & 0.0 \\ 0.3 & 0.0 & 0.0 & 0.2 & 0.5 & 0.0 & 0.0 \\ 0.7 & 0.0 & 0.0 & 0.0 & 0.0 & 0.3 & 0.0 \\ 0.6 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.4 \end{pmatrix}.$$

Since P is reducible, we have to find the dominant blocks and the maximum eigenvalue of the dominant blocks. In this case, the dominant block is

$$T_{22} = \begin{pmatrix} 0.4 & 0.6 \\ 0.2 & 0.5 \end{pmatrix},$$

which has eigenvalues 0.8 and 0.1 and left and right normalized eigenvectors $\psi = (0.515\,078\,7,\ 1.030\,157\,4)'$ and $\phi = (0.832\,050\,3,\ 0.554\,700\,2)'$, respectively, yielding $p_\infty = \lim_{n\to\infty} p_n = 0.2$ and the rate of convergence is bounded by

$$\beta^{\boldsymbol{Y}}(n) = O\left(\left(n\frac{|\lambda_2|}{\lambda_1}\right)^n\right) = O\left(n\left(\frac{1}{8}\right)^n\right).$$

Example 5.3. (*Where P is periodic with constant sum on its columns.*) Consider the Markov chain $\boldsymbol{X}$ with alphabet $A = \{1, 2, \ldots, m\}$ and transition matrix such that

$$P = \begin{pmatrix} \mathbf{0} & B_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & B_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & B_{h-1} \\ B_h & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix}$$

with $\sum_i P_{ij} = k$ for all j.

https://doi.org/10.1017/jpr.2015.20 Published online by Cambridge University Press

Note that, in this case, k is the largest eigenvalue of P and $\mathbf{1}$ is a left eigenvector of P^r, for all r. Therefore, (4.1) can be written explicitly as

$$\frac{VG^*P^rW^\top}{VG^*P^r\mathbf{1}^\top} = \frac{V\phi\mathbf{1}P^rW^\top}{V\phi\mathbf{1}P^r\mathbf{1}^\top} = \frac{k^rV\phi\mathbf{1}W^\top}{k^rV\phi\mathbf{1}\mathbf{1}^\top},$$

which is constant for all r.

Example 5.4. (*Where P has only one nonnull eigenvalue and Y is a Markov chain of finite order.*) Consider the Markov chain $\boldsymbol{X}$ with alphabet $A = \{1, 2, 3, 4, 5\}$ and transition matrix

$$P^X = \begin{pmatrix} 0.30 & 0.2 & 0.1 & 0.25 & 0.15 \\ 0.20 & 0.5 & 0.3 & 0.00 & 0.00 \\ 0.85 & 0.0 & 0.0 & 0.10 & 0.05 \\ 0.80 & 0.0 & 0.0 & 0.00 & 0.20 \\ 1.00 & 0.0 & 0.0 & 0.00 & 0.00 \end{pmatrix}.$$

Note that $\boldsymbol{Y}$ is not Markov of order 1 since the W is not constant. On the other hand, it is easy to see that P has only one nonnull eigenvalue and its triangular form has three nonnull rows. Therefore, by Proposition 3.2 we have an order 4 Markov chain.

In fact, it is easy to see that

$$\frac{P^k}{\lambda_1^k} = \frac{P^3}{\lambda_1^3} \quad \text{for all } k \geq 3.$$

Example 5.5. (*The 'house-of-cards' process: P countably infinite and transient.*) A house-of-cards process is a Markov chain on the natural numbers with transition matrix Q defined through $Q(n+1 \mid n) = 1 - q_n = 1 - Q(0 \mid n)$, for $n = 0, 1, 2, \ldots$, i.e. the chain can go 'up' with one unit or go 'down' to 0. It is well known that we have positive recurrence if and only if $\sum_{k\geq 1} \prod_{i=0}^{k-1}(1 - q_i) < \infty$.

Note that in this case, the sub-matrix P has all 0s in its diagonal and values $1-q_1, 1-q_2, \ldots$ in the first diagonal above it. Therefore, it does not satisfy the conditions Theorem 3.4 as it is not irreducible. However, it is still possible to compute explicitly (3.1) as

$$p_{k+1} = \frac{VP^kW^\top}{VP^k\mathbf{1}^\top} = \frac{\prod_{\ell=0}^{k+1}(1-q_\ell)q_{k+1}}{\prod_{\ell=0}^{k+1}(1-q_\ell)} = q_{k+1}.$$

As expected, the renewal process associated with a positive recurrent house-of-cards process is continuous if and only if q_k converges as $k \to \infty$.

There is a well-known connection between the house-of-cards and the image renewal process $\boldsymbol{Y}$. In the literature, the house-of-cards is called the 'backward recurrence time chain' of the process $\boldsymbol{Y}$; see, for example, Foss and Tweedie (1998, Section 3.3.1). In the same spirit, we could have considered the 'forward recurrence time chain'. It has transition matrix $\bar{Q}$ defined through $\bar{Q}(n \mid 0) = \bar{q}_n$ and $\bar{Q}(n \mid n+1) = 1$ for $n \geq 0$, i.e. the chain can go 'up' from 0 to any natural number and then 'goes down' step-by-step until it reaches 0 again. We also easily check that positive recurrence holds if and only if $\sum_{i\geq 0}(i+1)\bar{q}_i < \infty$.

Here also, if we make $P^X = Q$ then the sub-matrix P has all 0s in its diagonal and all 1s in the first diagonal below it, so it does not fit into the conditions of Theorem 3.4. However,

https://doi.org/10.1017/jpr.2015.20 Published online by Cambridge University Press

computing directly (3.1), we will obtain

$$p_n = \frac{\bar{q}_{n+1}}{\sum_{j\geq n+1} \bar{q}_j}$$

as expected by the well-known relation between forward and backward recurrences time chains; see Foss and Tweedie (1998, Section 3.3.1). Thus, continuity of the image process $\boldsymbol{Y}$ will hold if and only if the above quotient converges.

Example 5.6. (*The 'birth and death' Markov chain: P countably infinite and irreducible.*) This Markov chain has transition matrix Q defined on the natural numbers through $Q(i+1 \mid i) = r_i$, $Q(i-1 \mid i) = q_i$, and $Q(i \mid i) = 1 - r_i - q_i$ for $i \geq 1$, and $Q(1 \mid 0) = 1 - Q(0 \mid 0) = r_0$. We naturally assume that q_i and r_i are both strictly positives for $i \geq 1$. It is well known that positive recurrence holds if and only if

$$\sum_{k\geq 1} \frac{r_0 \cdots r_{k-1}}{q_1 \cdots q_k} < \infty. \tag{5.1}$$

Now consider $P^{\boldsymbol{X}} = Q$. If we have $q_i + r_i < 1$, we guarantee that P is irreducible. It remains to check whether P is positive recurrent. This can be seen by the fact that P has 0s at the same positions as $P^{\boldsymbol{X}}$ and the nonnull entries are simply obtained by translating the indexes of the sequences r_i, $i \geq 0$ and q_i, $i \geq 0$:

$$P_{i,i+1} = P^{\boldsymbol{X}}_{i+1,i+2} = r_i, \qquad P_{i+1,i} = P^{\boldsymbol{X}}_{i+2,i+1} = q_{i+1} i \geq 1.$$

Thus, positive recurrence of P is equivalent to

$$\sum_{k\geq 1} \frac{r_1 \cdots r_k}{q_2 \cdots q_{k+1}} < \infty,$$

which holds when (5.1) holds. Thus, the image $\boldsymbol{Y}$ process is continuous by Theorem 3.4.

5.2. Relation to Gibbsianess

Gibbsianess (in the sense of statistical physics) corresponds to the continuity of the 'two-sided set of transition probabilities', or *specification*. We do not enter further into detailed definitions, and refer the interested reader to Fernández and Mallard (2004a) and (2004b). As explained in Fernández *et al.* (2011), the renewal chain $\boldsymbol{Y}$ is Gibbsian (in the sense of statistical physics) if and only if we have convergence, as m and n diverge, of

$$\mathbb{P}(Y_0 = 1 \mid Y_{-m-1} = 1, Y_{-m}^{-1} = 0^m, Y_1^n = 0^n, Y_{n+1} = 1) =: p_{m,n}.$$

Let $r_{m,n}$ denote

$$\sum_{\substack{b_{-m}^{-1} \in \pi^{-1}(0_{-m}^{-1}) \\ b_1^n \in \pi^{-1}(0_1^n)}} \mathbb{P}(X_{-m}^{-1} = b_{-m}^{-1}, X_0 = 1 \mid X_{-m-1} = 1) \mathbb{P}(X_1^n = b_1^n, X_{n+1} = 1 \mid X_0 = 1)$$

and $s_{m,n}$ denote

$$\sum_{b_{-m}^n \in \pi^{-1}(0_{-m}^n)} \mathbb{P}(X_{-m}^n = b_{-m}^n, X_{n+1} = 1 \mid X_{-m-1} = 1).$$

https://doi.org/10.1017/jpr.2015.20 Published online by Cambridge University Press

Using similar calculations as the ones performed in the proof of Proposition 3.1, it is easy to arrive at

$$p_{m,n} = \frac{r_{m,n}}{r_{m,n} + s_{m,n}}.$$

This quantity also has a simple matrix form, which is

$$p_{0,0} = \frac{P_{1,1}^2}{P_{1,0}P_{0,1} + P_{1,1}^2}, \qquad p_{m+1,n+1} = \frac{V P^m W^\top V P^n W^\top}{V P^{m+n+1} \mathbf{1}^\top} \quad \text{if } m, n \geq 0.$$

The conditions for convergence, or not, of this quantity, and consequently, for Gibbsianess or non-Gibbsianess, are the same as for continuity. In particular, this means that, under the conditions assumed in this paper, the image processes are either Gibbsian and have a continuous kernel or non-Gibbsian and have a discontinuous kernel.

In relation to the paper of Chazottes and Ugalde (2003), it is important to note that their concept of Gibbsianess is *not* that of statistical physics. They consider Gibbs measures in the sense of Bowen, which is slightly different. This makes it difficult to make a direct comparison between our results and theirs.

To conclude, let us mention other recent papers with interests related to ours. As already mentioned, Verbitskiy (2011a) and (2011b) prove that, if the original measure has a continuous kernel (that is, not necessarily a Markov measure, but it could be) with summable continuity rate (that is $\sum \beta_k < \infty$), then factors of this measure still have a continuous kernel. Similar results were obtained concerning conservation of Bowen's Gibbsianess (Chazottes and Ugalde (2011), Pollicott and Kempton (2011), Yoo (2010), and Boyle and Peterson (2011)), and of statistical physics Gibbsianess (Redig and Wang (2010)).

Acknowledgements

We would like to thank the anonymous referees for suggestions that greatly improved the presentation of the paper. The authors were partially supported by CNPq grants 462064/2014-0 (SG), 302755/2010-1, and 302598/2014-6 (NLG). This work is part of USP project 'Mathematics, computation, language and the brain' and FAPESP project 'NeuroMat' (grant no. 2013/07699-0).

References

Boyle, M. and Petersen, K. (2011). Hidden Markov processes in the context of symbolic dynamics. In *Entropy of Hidden Markov Processes and Connections to Dynamical Systems* (London Math. Soc. Lecture Note Ser. **385**), Cambridge University Press, pp. 5–71.

Burke, C. J. and Rosenblatt, M. (1958). A Markovian function of a Markov chain. *Ann. Math. Statist.* **29,** 1112–1122.

Chazottes, J.-R. and Ugalde, E. (2003). Projection of Markov measures may be Gibbsian. *J. Statist. Phys.* **111,** 1245–1272.

Chazottes, J.-R. and Ugalde, E. (2011). On the preservation of Gibbsianness under symbol amalgamation. In *Entropy of Hidden Markov Processes and Connections to Dynamical Systems* (London Math. Soc. Lecture Note Ser. **385**), Cambridge University Press, pp. 72–97.

Darroch, J. N. and Seneta, E. (1967). On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *J. Appl. Prob.* **4,** 192–196.

Fernández, R. and Maillard, G. (2004a). Chains and specifications. *Markov Process. Relat. Fields* **10,** 435–456.

Fernández, R. and Maillard, G. (2004b). Chains with complete connections and one-dimensional Gibbs measures. *Electron. J. Prob.* **9,** 145–176.

Fernández, R., Gallo, S. and Maillard, G. (2011). Regular C-measures are not always Gibbsian. *Electron. Commun. Prob.* **16,** 732–740.

Foss, S. G. and Tweedie, R. L. (1998). Perfect simulation and backward coupling. *Commun. Statist. Stoch. Models* **14,** 187–203.

https://doi.org/10.1017/jpr.2015.20 Published online by Cambridge University Press

Harris, T. E. (1955). On chains of infinite order. *Pacific J. Math.* **5,** 707–724.
Kalikow, S. (1990). Random Markov processes and uniform martingales. *Israel J. Math.* **71,** 33–54.
Kitchens, B. P. (1998). *Symbolic Dynamics*. Springer, Berlin.
Pollicott, M. and Kempton, T. (2011). Factors of Gibbs measures for full shifts. In *Entropy of Hidden Markov Processes and Connections to Dynamical Systems* (London Math. Soc. Lecture Note Ser. **385**), Cambridge University Press, pp. 246–257.
Redig, F. and Wang, F. (2010). Transformations of one-dimensional Gibbs measures with infinite range interaction. *Markov Process. Relat. Fields* **16,** 737–752.
Seneta, E. (2006). *Non-Negative Matrices and Markov Chains*, 2nd edn. Springer, New York.
Verbitskiy, E. (2011a). On factors of g-measures. *Indag. Math. (N.S.)* **22,** 315–329.
Verbitskiy, E. (2011b). Thermodynamics of hidden Markov processes. In *Entropy of Hidden Markov Processes and Connections to Dynamical Systems* (London Math. Soc. Lecture Note Ser. **385**), Cambridge University Press, pp. 258–272.
Yoo, J. (2010). On factor maps that send Markov measures to Gibbs measures. *J. Statist. Phys.* **141,** 1055–1070.

https://doi.org/10.1017/jpr.2015.20 Published online by Cambridge University Press