A COMPLETELY STATIONARY MARKOV CHAIN WITH INFINITE STATE SPACE*

Y. H. Wang and Linggi Tang

Abstract. We introduce an infinite dimension completely stationary ergodic Markov chain; investigate its properties; establish a sufficient condition that the Markov property of a Markov chain is preserved by its induced multi-state sequence and finally derive the limiting distribution of the sum of the Markov chain. It is the simplest, nontrivial, infinite dimension Markov chain which possesses all the nice properties - stationarity, irreducibility, positive recurrence, aperiodicity, ..., etc.

1. Introduction

The completely stationary Markov Bernoulli model introduced by Edwards (1960) was extended to a $(k+1)$-dimension, $k \geq 1$, multi-state one by Wang and Yang (1995). It was shown in that paper that the whole $(k+1)$-dimension Markov chain could be completely specified by the initial probability and the correlation coefficient of two consecutive variables.

In this paper we extend this Markov model to the infinite state space case, investigate its properties and derive some relevant theorems. In Section 2, we present a completely stationary infinite dimension Markov chain which depends only on its initial probability and the correlation coefficient of two consecutive variables, and prove that it is ergodic. In Section 3, we obtain a sufficient condition that the Markov property of a Markov chain is preserved by its induced multi-state sequence. Finally, we derive the limiting distribution of the partial sum of the sequence.

2. A Completely Stationary Infinite Dimension Markov Chain

2.1. The model and its properties

Denote $\mathcal{X} = \{X_i; i = 1, 2, \ldots\}$ to be a Markov chain with state space $S = \{0, 1, 2, \ldots\}$ and stationary transition matrix $\mathbb{P} = (p_{ij})$ defined by

$$(1) \qquad p_{ij} = \pi \delta_{ij} + (1 - \theta)p_j, \qquad (i, j \in S),$$

where δ_{ij} is the Kronecker delta, $\boldsymbol{p} = \{p_i : p_i \geq 0, \ i \in S\}$, with $\sum_{i \in S} p_i = 1$, is its initial probability. (In the sequel we shall call this Markov chain "model (1)".)

This Markov chain has the following properties:

(a) Because $\boldsymbol{p} = \boldsymbol{p}\mathbb{P}$, it is completely stationary. That is

$$(2) \qquad P(X_i = j) = p_j, \qquad (i \in S; \ i \geq 1).$$

(b) It can be shown that the parameter π is the correlation coefficient of X_i and X_{i+1}, for all $i \geq 1$. Thus this Markov chain is completely specified by its initial probability $\boldsymbol{p}$ and the correlation coefficient π. It is a natural extension of a given sequence of iid discrete random variables to a Markov dependent one with a specified correlation coefficient between two consecutive variables.

(c) The n-th power $\mathbb{P}^n = (p_{ij}^{(n)})$ of $\mathbb{P}$ is of the form

$$(3) \qquad p_{ij}^{(n)} = \pi^n \delta_{ij} + (1 - \pi^n)p_j, \qquad (i, j \in S).$$

It follows that the correlation coefficient of X_i and X_{i+m} is π^m, for all i, and $m \geq 1$. Hence if $\pi < 1$ the Markov chain is asymptotically independent.

(d) It follows from (a) or (c) that the limit

$$\lim_{n \to \infty} \mathbb{P}^n = P,$$

exists with all the rows of P identical to the initial probability $\boldsymbol{p}$. Thus the mean recurrent time μ_j of state j is $1/p_j$, for all $j \in S$.

(e) We shall show later that all the truncated sequences

$$X_n' = \min\{X_n, \ k\}, \ k \geq 1,$$

are also completely stationary Markov chains.

Proposition. *If the parameters in the transition matrix $\mathbb{P}$ are such that $0 \leq \pi < 1$, and $p_i > 0$ for all $i \geq 0$, then the Markov chain is ergodic.*

Proof. Evidently, it is irreducible and aperiodic. Fix $j \in S$. With $\mathbb{P}^n = (p_{ij}^{(n)})$ having the form (3), it follows that

$$\sum_{n=0}^{\infty} p_{jj}^{(n)} \geq p_j \sum_{n=1}^{\infty} (1 - \pi) = \infty.$$

Thus, by Corollary (4) on page 202 in Grimmett and Stirzaker (1992), state j is persistent. Now

(4) $$p_{jj}^{(n)} = \pi^n + (1 - \pi^n)p_j \to p_j > 0, \text{ as } n \to \infty.$$

By Theorem (9) on page 203 in Grimmett and Stirzaker (1992), state j is positive persistent, which proves that state j is ergodic.

Since ergodicity is a class property, we conclude the proof.

2.2. Remarks:

1). It is well known that for an irreducible aperiodic Markov chain with stationary transition matrix $\mathbb{Q}$, there is a unique solution to the equation $\boldsymbol{p} = \boldsymbol{q}\mathbb{Q}$, where $\boldsymbol{q} = \{q_i; \; q_i \geq 0, \; i \in S\}$, with $\sum_{i \in S} q_i = 1$. If the process is initiated with probability $\boldsymbol{q}$ then the resultant Markov chain is completely stationary. But the model (1) is quite different. For one, in the model (1), the initial probability p is incorporated in the transition matrix $\mathbb{P}$, not simply a solution to the equation $\boldsymbol{p} = \boldsymbol{p}\mathbb{P}$. For the other, it involves less parameters than a regular stationary model. For example, let us take Edwards' Markov Bernoulli model as an illustration. In his model, the whole chain is completely specified by only two parameters; $p = P(X_0 = 1)$ and π. While a two-state stationary Markov chain usually involves three parameters. In fact, the model (1) is the smallest class of stationary Markov chain containing the iid model.

2). In the literature, the Markov chain is often regarded as the simplest generalization of the independence model. But it is interesting to note that Aki (1985) came up with a non-Markovian generalization of the iid Bernoulli model.

3). If we let $p_i = e^{-\lambda}\lambda^i/i!$, for all $i \geq 0$, in the model (1), then we have an ergodic "Markov-Poisson" process with pre-assigned correlation coefficient between two consecutive variables. Likewise, we can easily construct ergodic "Markov-geometric" process, "Markov-Pascal" (Markov-negative binomial) process, etc. based on the model (1). Thus, as already stated in section 2.1 - c), that the model (1) can be regarded as an extention of an i.i.d.

sequence of discrete random variables to a Markov dependent sequence with specific correlation coefficient between two consecutive variables.

3. An Induced Markov Chain

Given a sequence of Markov chain $\mathcal{X} = \{X_i\}$ with state space S, a natural question to raise is whether "the sequence of indicator functions $\{I_{\{X_i \in B\}}\}$ of a proper subset B of S is also a Markov chain?". The answer in general is no. (See Wang and Yang (1995).) For example, consider the transition matrix

$$
\begin{array}{c}
\\
0 \\
1 \\
2
\end{array}
\begin{array}{ccc}
0 & 1 & 2 \\
\left(\begin{array}{ccc}
a & \alpha & b \\
c & d & e \\
f & \alpha & g
\end{array}\right)
\end{array} .
$$

(All letters represent distinct positive numbers with row sums equal 1.) It is easy to check that the only sequences of indicator variables $\{I_{\{X_i \in B\}}\}$ which preserve the Markovian property are when $B = \{1\}$ or $\{0, 2\}$, no matter what is the initial probability.

In this section we consider the problem in a more general setting than in Wang and Yang (1995) and a sufficient condition is found for a sequence of induced random variables of a Markov chain to preserve the Markovian property.

Let $\mathcal{B} = \{B_0, \ B_1, \ B_2, \ldots\}$ be a sequence, finite or infinite, of non-empty proper subsets of the state space S. We say that $\mathcal{B}$ constitutes a partition of S, if B_j are mutually exclusive with union equal to S. Define a sequence of multi-state random variables $\mathcal{T} = \{T_i; \ i \geq 1\}$ by

$$
(5) \qquad\qquad T_i = \sum_{j=0}^{\infty} j 1_{\{X_i \in B_J\}},
$$

where 1_A is the indicator variable of A. Thus T_i takes values $0, 1, 2, \ldots$.

Theorem 1. *Let $\mathcal{X} = \{X_i, \ i = 1, 2, \ldots\}$ be a Markov chain with state space $S = \{0, 1, 2, \ldots\}$ stationary transition matrix $\mathbb{P} = (p_{ij})$ and initial probabilities $p_j = P(X_1 = j)$, $i, j \in S$. Let $\mathcal{B} = \{B_0, B_1, B_2, \ldots\}$ be a partition of S. Then the sequence of random variables $\mathcal{T} = \{T_i; i \geq 1\}$ defined by (5) is a Markov chain if*

$$
(6) \qquad p_{ij} = p_{\ell j} \qquad (i, \ell \in B_S, \ j \in B_t; \ s, t = 0, 1, 2, \ldots, s \neq t).
$$

Moreover, the transition matrix $\mathbb{Q} = (q_{st})$ of $\mathcal{T}$ is $q_{st} = \sum_{j \in B_t} p_{\ell j}, \ \ell \in B_S,$ with initial probabilities $q_s = \sum_{i \in B_s} p_i, \ s, \ t = 0, 1, 2, \ldots,$ and, if $\mathcal{X}$ is completely stationary, so is $\mathcal{T}$.

Proof. For $s \neq t, s, t, j_{i-1}, \ldots, j_1 \in \{0, 1, 2, \ldots, \}$, $i \geq 1$, condition (6) implies

$$
\begin{aligned}
&P\{T_{i+1} = t | T_i = s, T_{i-1} = j_{i-1}, \ldots, T_1 = j_1\} \\
&= P\{X_{i+1} \in B_t | X_i \in B_S, X_{i-1} \in B_{j_{i-1}}, \ldots, X_1 \in B_{j_1}\} \\
\text{(7)} \quad &= \sum_{\ell \in B_S} P\{X_i = \ell | X_i \in B_S, X_{i-1} \in B_{j_{i-1}}, \ldots, X_1 \in B_{j_1}\} P\{X_{i+1} \in B_t | X_i = \ell\} \\
&= P\{X_{i+1} \in B_t | X_i = \ell\} \quad (\ell \in B_S).
\end{aligned}
$$

For $s = t$, condition (6) leads to

$$
\begin{aligned}
&P\{T_{i+1} = s | T_i = s, \ T_{i-1} = j_{i-1}, \ldots, \ T_1 = j_1\} \\
&= 1 - \sum_{t \neq S} P\{T_{i+1} = t | T_i = s, \ T_{i-1} = j_{i-1}, \ldots, T_1 = j_1\} \\
\text{(8)} \quad &= 1 - \sum_{t \neq S} P\{X_{i+1} \in B_t | X_i = \ell\} \quad (\ell \in B_S) \\
&= P\{X_{i+1} \in B_S | X_i = \ell\} \quad (\ell \in B_S).
\end{aligned}
$$

Now, for $\ell \in B_S$, the last expressions in (7) and (8) equal

$$
P\{X_{i+1} \in B_t | X_i = \ell\} = \sum_{V \in B_t} p_{\ell V} = q_{st} = P\{T_{i+1} = t | T_i = s\}.
$$

And the initial probabilities

$$
q_s = P(T_1 = s) = P(X_1 \in B_S) = \sum_{i \in B_S} p_i,
$$

as required.

This proves the first two parts of the theorem. The last part is obvious and its proof is skipped.

A special case of transition matrices satisfying condition (6) is

$$
\text{(9)} \qquad p_{ij} = \alpha_j \delta_{ij} + \beta_j (1 - \delta_{ij}) \qquad (i, \ j \in S).
$$

That is the off diagonal entries of each column are all identical. Thus, in this particular case, for any proper subset B of S, the sequence of indicator functions $\{I_{\{X_i \in B\}}\}$ is also a Markov chain. Evidently, the model (1) is a special case of (9), and hence of (6).

The next corollary shows that if we start with an infinite dimensional completely stationary Markov chain, we can induce a sequence of finite dimensional completely stationary Markov sequence.

Corollary 1. *Let* $\mathcal{X} = \{X_i;\ i = 1, 2, \ldots\}$ *be the completely stationary Markov chain defined by the transition matrix (1). Let* $\mathcal{B} = \{B_0, B_1, \ldots, B_k\}$, $k \geq 1$, *be a partition of* S. *Then the sequence of* $(k+1)$-*dimensional random variables* $\mathcal{T} = \{T_i;\ i \geq 1\}$ *defined by (5) is again a completely stationary Markov chain with transition matrix* $\mathbb{P}_k = (q_{ij})$

$$(10) \qquad\qquad q_{ij} = \pi\delta_{ij} + (1 - \pi)q_j, \qquad (i, j = 0, 1, \ldots, k),$$

and initial probability $\boldsymbol{q} = \{q_i; i = 0, 1, \ldots, k\}$, *where* $q_i = \sum_{j \in B_i} p_j$.

It follows from Corollary 1 that for the model (1):

(1) All the sequences of indicator variables of non-trivial subsets of the state space S reduce to the Edwards' Markov Bernoulli sequences.

(2) As stated earlier that all the truncated sequences $X'_n = \min(X_n, k)$, for $k \geq 1$, are also completely stationary Markov chains.

In view of Theorem 1, another generalization of the Edwards Markov Bernoulli sequence can be done as follows:

For a sequence of partition $\mathcal{B} = \{B_0, B_1, B_2, \ldots, \}$ of S, finite or infinite, denote

$$Y_{ij} = 1_{\{X_i \in B_j\}} \quad (i = 1, 2, \ldots;\ i = 0, 1, 2, \ldots,).$$

Then $\sum_{j=0}^{\infty} Y_{ij} = 1$ for all $i \geq 1$. Define a sequence of binary random vectors $\mathcal{Y} = \{Y_i; i \geq 1\}$ by $Y_i = (Y_{i0}, Y_{i1}, Y_{i2}, \ldots)$. Then Y_i takes values in the set

$$\mathbb{S} = \{(s_0, s_1, s_2, \ldots,) : s_j = 0 \text{ or } 1 \text{ with } \sum_{j=0}^{\infty} s_j = 1\}.$$

By Theorem 1, if $\mathcal{X}$ is a completely stationary Markov chain defined by (1), then so is $\mathcal{Y}$. Since the two events $\{Y_{ij} = 1\}$ and $\{T_i = j\}$ are identical for all i and j, the next corollary follows right from Theorem 1.

Corollary 2. *The sequence of binary random vectors* $\mathcal{Y}$ *is a completely stationary Markov chain with state space* $\mathbb{S}$.

4. A Limit Theorem

In this section we derive the infinite dimension version of the main theorem in Wang (1981). The limit conditions in this section are

(11) $\qquad np_j \to \lambda_j > 0, \;\; j \geq 1, \;\; \text{and} \;\; \lambda = \sum_{j=1}^{\infty} \lambda_j < \infty, \qquad (n \to \infty).$

Let $1 \leq k < \infty$, and consider the partition $\mathcal{B} = \{B_0, B_1, \ldots, B_k\}$, where $B_i = \{i\}$ for $i = 0, 1, \ldots, k-1$ and $B_k = \{i : i \geq k\}$. Then by Corollary 1, the sequence of (k + 1)-dimensional random variables $\mathcal{T} = \{T_i; i \geq 1\}$ defined by (5) is a truncated sequence and hence is a completely stationary Markov chain with transition matrix $\mathbb{P}_k$ defined by (10) with $q_i = p_i$ for $i = 0, 1, \ldots, k-1$, and $q_k = \sum_{i \geq k} p_i$. (For brevity, the dependence of $\mathcal{T}$ on k is suppressed hereafter.) Denote $S_{nk} = T_1 + \cdots + T_n$. Then by Corollary II in Wang and Yang (1995), see also Hsiau (1997) Lemma 3.2, the pgf $G_{nk}(s) = E(s^{S_{nk}})$ can be written as

$$G_{nk}(s) = (1, s^1, \ldots, s^k) \, R^{n-1}(s) \boldsymbol{q} \qquad (0 \leq s \leq 1),$$

where

$$R(s) = \begin{pmatrix} \pi + (1-\pi)q_0 & (1-\pi)q_1 s & \cdots\cdots & (1-\pi)q_k s^k \\ (1-\pi)q_0 & [\pi + (1-\pi)q_1]s & \cdots\cdots & (1-\pi)q_k s^k \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ (1-\pi)q_0 & (1-\pi)q_1 s & \cdots\cdots & [\pi + (1-\pi)q_k]s^k \end{pmatrix}.$$

The next lemma was proved in Wang and Tang (1997) and we state it here for later reference. If $k = 1$, this lemma reduces to the main theorem in Wang (1981).

Lemma. *In the model* (1), *under the limit conditions* (11), *for all* $1 \leq k < \infty$, *we have*

(12) $\quad \lim_{n \to \infty} G_{nk}(s) = \exp\left\{ -(1-\pi) \sum_{j=1}^{k} \lambda_j' \left(1 - \frac{(1-\pi)s^j}{1 - \pi s^j} \right) \right\}, \quad (0 \leq s \leq 1),$

where $\lambda_j' = \lambda_j$ *for* $1 \leq j < k$ *and* $\lambda_k' = \sum_{i=k}^{\infty} \lambda_i$.

Let $S_n = X_1 + X_2 + \cdots + X_n$. Denote $G_n(s) = (s^{S_n})$, the pgf of S_n and $\phi(s)$ a pgf defined by

$$\phi(s) = \sum_{j=1}^{\infty} (\pi_j / \lambda) \left(\frac{(1-\pi)s^j}{1 - \pi s^j} \right),$$

where $\lambda = \sum_{j=1}^{\infty} \lambda_j < \infty$.

Theorem 2. *For the model* (1), *if the parameters p_j satisfy the limit conditions* (11), *then the distribution of S_n converges to a compound Poisson distribution with Poisson parameter $(1 - \pi)\lambda$, and compounding distribution whose pgf is $\phi(s)$ which is the pgf of the mixture of infinitely many geometric-type distributions. That is,*

$$\lim_{n\to\infty} G_n(s) = \exp\{-(1 - \pi)\lambda[1 - \phi(s)]\}.$$

Proof. We first define a truncated sequence $\mathcal{Y} = \{Y_{ik} : i = 1, 2, \ldots\}$ by $Y_{ik} = \min(X_i, k)$, where $k \geq 1$ is a positive integer. Then, by the Lemma, the pgf of the partial sum $S_{nk} = Y_{1k} + \cdots + Y_{nk}$ converges to

$$(13) \qquad \exp\{-(1 - \pi)\lambda_k'\left[1 - \phi_k'(s)\right]\} \exp\left\{-(1 - \pi) \wedge_k \left(1 - \frac{(1 - \pi)s^k}{1 - \pi s^k}\right)\right\},$$

where $\wedge_k = \sum_{j \geq k} \lambda_j$, $\lambda_k' = \sum_{j \leq k-1} \lambda_j$ and

$$\phi_k'(s) = \sum_{j=1}^{k-1} (\lambda_j/\lambda_k') \left(\frac{(1 - \pi)s^j}{1 - \pi s^j}\right).$$

(Note that (13) is the same as (12), but in a seemingly different form.)

Since $\wedge_k \to 0$; $\lambda_k' \to \lambda$ and $\phi_k'(s) \to \phi(s)$ for $0 \leq s \leq 1$, as $k \to \infty$,

$$\lim_{k\to\infty} \lim_{n\to\infty} G_{nk}(s) = \exp\{-(1 - \pi)\lambda[1 - \phi(s)]\}.$$

Finally, the sequence of functions $\{G_{nk}(s)\}$ satisfies $0 \leq G_{nk}(s) \leq 1$ for all n, $k \geq 1$ and $0 \leq s \leq 1$, we conclude the proof by noting

$$\lim_{n\to\infty} G_n(s) = \lim_{n\to\infty} \lim_{k\to\infty} G_{nk}(s) = \lim_{k\to\infty} \lim_{n\to\infty} G_{nk}(s).$$

References

1. Siego Aki, Discrete distributions of order k on a binary sequence, *Ann. Inst. Statist. Math.* **37A** (1985), 205-224.

2. A.W.F. Edwards, The meaning of binomial distribution, *Nature* **186** (1960), 1074.

3. G. R. Grimmett and D. R. Stirzaker, Probability and Random Processes, 2nd edition, Clarendon Press, Oxford, 1992.

4. Shoou-Ren Hsiau, Compound poisson limit theorems for Markov chains, *J. Appl. Probab.* **34** (1997), 24-34.

5. Y. H. Wang, On the limit of the Markov binomial distribution, *J. Appl. Probab.* **18** (1981), 937-942.

6. Y. H. Wang and Zejiang Yang, On a Markov multinomial distribution, *Math. Scientist* **20** (1995), 40-49.

7. Y. H. Wang and Lingqi Tang, An alternative proof for the limit of the Markov multinomial distribution, *Math. Scientist* **22** (1997), 62-64.

Y. H. Wang
Department of Statistics, Tunghai University
Taichung, Taiwan

Lingqi Tang
Department of Statistics, University of Manitoba
Winnipeg, Manitoba, Canada R3T 2N2