Memoirs of the Faculty of Engineering and Design
Kyoto Institute of Technology 43 (March, 1995)

Auto-Regressive Representations of a Stationary Markov Chain with Finite States

Junichi NAKAYAMA

(Received July 22, 1994 ; Accepted September 9, 1994)

Abstract

This study deals with a nonlinear feedback system that transforms an independent stochastic sequence into a stationary Markov chain with finite states. As a nonlinear system, a stochastic difference equation is proposed with a nonlinear system function that is defined by the transition probability. Several types of autoregressive (AR) representations for such stochastic system are then introduced. First, a non-linear AR equation is derived by expanding the system function into a power series. The Markov chain is then represented by a K-dimensional vector which enjoys a linear discrete-valued AR equation, where K is the number of states. Third, the Markov chain is represented by a unit vector sequence which satisfies another linear discrete-valued AR equation. Further, the Markov chain is regarded as a $(K-1)$-dimensional vector sequence which satisfies a linear AR equation with a constant coefficient matrix and white noise excitation. Relationships between these representations are discussed and formulas for spectrum matrix, correlation matrix and joint probability are obtained.

Key Words: *Stochastic process; stationary Markov chain; auto-regressive representations; discrete-valued equation; spectrum matrix.*

1. Introduction

Stationary Markov chains with finite states are widely used as mathematical models of random phenomena with discrete events in physics, engineering and system science. They are important as signal models in the theory of digital communication and signal processing. Mathematically, Markov chains are discussed with the transition probability, in terms of which statistics such as average, moments, correlation and spectra have been studied extensively.[1)-3)]

However, we are interested in such a Markov chain as a discrete approximation of natural signals on the continuous time axis, particularly for a non-Gaussian signal. This is because continuous-valued natural signals are commonly approximated by a discrete-valued sequence obtained by the periodic sampling and quantization. Even though such a discrete approximation is employed in practical applications, natural signals are often regarded as continuous-valued stationary sequences, which are represented by dynamic system equations such as linear auto-regressive (AR) equations and auto-regressive moving average (ARMA) models with white noise excitations in the theory of digital signal processing. A Markov process on a continuous time axis also has the dynamic system representation such as the stochastic differential equation and Langevin equation. On the other hand, little work has been reported on dynamic system representations for finite Markov chains. Therefore, there is a wide gap in mathematical representations of continuous-valued stationary sequences and discrete-valued finite Markov chains.

To bridge this gap, we present a physical system that transforms an independent stochastic sequence into a finite Markov chain in this paper. We introduce a stochastic difference equation that is a nonlinear feedback system with a system function related to the transition probability. Expanding the system function into a power series, we derive a discrete-valued auto-regressive (AR) equation that is nonlinear with respect to the Markov sequence, where the discrete-valued AR equation is an AR equation with discrete-valued random coefficients and discrete-valued white noise. We demonstrate that such a nonlinear AR equation is reduced to a linear equation if K-dimensional vector representation is introduced, where K is the number of states. Further, we represent the Markov chain by a unit vector sequence which enjoys another linear discrete-valued AR equation. Removing the redundancy in the K-dimensional vector representation, we next represent the Markov chain as a $(K-1)$-dimensional vector sequence which satisfies a linear AR equation with constant coefficient matrix and white noise excitation. Relationships between these representations and formulas for stationary probability, correlation matrix, and spectrum matrix are discussed.

Finally, we give a simple example where the non-linear discrete-valued equation becomes a linear equation, and then prove that a discrete-valued stationary sequence with any probability distribution and an exponential correlation function can be generated by such a linear discrete-valued equation.

2. Finite Markov Chain and Nonlinear Discrete-Valued AR Equation

Let us consider a stationary Markov chain y_n, $n=0, \pm 1, \pm 2, \cdots$, with K numerical states:

$$v_1 < v_2 < v_3 < \cdots < v_K. \tag{1}$$

We define the transition probability matrix $\mathbf{t}=[t_{k,l}]$ as

$$t_{k,l}=Prob.\{y_n=v_k|y_{n-1}=v_l\}, \quad t_{k,l}\geq 0, \quad (k, \ l=1, 2, 3, \cdots, K) \tag{2}$$

$$\sum_{k=1}^{K} t_{k,l}=1, \quad (l=1, 2, \cdots, K), \tag{3}$$

which is the probability of $y_n=v_k$ when $y_{n-1}=v_l$. We assume that there exists a stationary probability $\mathbf{q}=[q_k]$:

$$\mathbf{q}=\mathbf{t}\cdot\mathbf{q}, \quad \mathbf{q}=[q_1, \ q_2, \ \cdots, \ q^K]^t, \tag{4}$$

$$q_k\geq 0, \quad \sum_{k=1}^{K} q_k=1, \tag{5}$$

where superscript t denotes the transpose. The stationary probability $\mathbf{q}$ is an eigen vector of $\mathbf{t}=[t_{k,l}]$ with an eigen value of 1. We further assume that the multiplicity of the eigen value 1 is simple and any other eigen values of $\mathbf{t}$ are less than unity in modulus in the complex plane.

Under these conditions, the Markov chain becomes aperiodic and has a stationary probability.[4)]

Let us consider a nonlinear system that generates the Markov chain y_n from an independent stochastic sequence x_n distributed uniformly over the interval [0, 1]. In terms of the transition probability $\mathbf{t}$, we first introduce a random transition matrix $\mathbf{T}(x_n)=[T_{k,l}(x_n)]$,

$$T_{k,l}(x)=\begin{cases} 0, & x\leq\sum_{k'=1}^{k-1}t_{k',l},\ x>\sum_{k'=1}^{k}t_{k',l} \\ 1, & \sum_{k'=1}^{k-1}t_{k',l}<x\leq\sum_{k'=1}^{k}t_{k',l} \end{cases} \quad (k,\ l=1,\ 2,\ \cdots,\ K), \tag{6}$$

$$T_{k,l}^2(x)=T_{k,l}(x),\ \sum_{k=1}^{K}T_{k,l}(x)=1, \tag{7}$$

which is a binary function of x taking only 1 or 0. Since x_n is uniformly distributed over the interval [0, 1], the average of the random transition matrix $\mathbf{T}(x_n)$ becomes the transition probability:

$$\langle\mathbf{T}(x_n)\rangle=\mathbf{t}, \tag{8}$$

where the angle brackets denote the ensemble average. Then, we write the nonlinear system by a stochastic difference equation,[5),6)]

$$y_n=F(x_n,\ y_{n-1}),\ n=0,\ \pm1,\ \pm2,\ \cdots, \tag{9}$$

where $F(x_n,\ y_{n-1})$ is the deterministic system function of x_n and y_{n-1}. The system function cannot be determined in a unique sense, but it may be defined as,

$$F(x,\ v_l)=\sum_{k=1}^{K}v_k\cdot T_{k,l}(x)=\begin{cases} v_1, & 0<x\leq t_{1,l} \\ v_2, & t_{1,l}<x\leq t_{1,l}+t_{2,l} \\ \cdot & \cdots \\ \cdot & \cdots \\ v_K, & 1-t_{K,l}<x\leq 1 \end{cases} \tag{10}$$

Physically, (9) describe a nonlinear feedback system, where the feedback is carried out by switching the system function $F(x,\ y_{n-1})$ with a one-step past output y_{n-1}.

Let us obtain an AR representation of the nonlinear system. Since y_{n-1} takes only K different values in (1), any function of y_{n-1} may be written by a linear combination of K base functions:[7)] 1, y_{n-1}, y_{n-1}^2, $\cdots$, y_{n-1}^{K-1}. In other words, any function of y_{n-1} is written by an inner product of a coefficient vector and a random vector $\mathbf{Y}_{n-1}$:

$$\mathbf{Y}_{n-1}=[1,\ y_{n-1},\ y_{n-1}^2,\ \cdots,\ y_{n-1}^{K-1}]^t, \tag{11}$$

which takes only the K-dimensional vector values:

$$\mathbf{v}_k=[1,\ v_k,\ v_k^2,\ \cdots,\ v_k^{K-1}]^t,\ (k=1,\ 2,\ 3,\ \cdots,\ K). \tag{12}$$

Taking this into consideration, we may expand $F(x,\ y_{n-1})$ by the base function to obtain a non-linear AR equation for the Markov chain y_n,

$$\begin{aligned} y_n&=F(x_n,\ y_{n-1}) \\ &=D_{1,0}(x_n)+D_{1,1}(x_n)\cdot y_{n-1}+D_{1,2}(x_n)\cdot y_{n-1}^2+\cdots+D_{1,K-1}(x_n)\cdot y_{n-1}^{K-1} \\ &=[D_{1,0}(x_n),\ D_{1,1}(x_n),\ \cdots,\ D_{1,K-1}(x_n)]\cdot\mathbf{Y}_{n-1}, \end{aligned} \tag{13}$$

where $D_{1,0}(x_n)$ is a white noise sequence and $D_{1,p}(x_n)$ with $p\geq1$ is a random coefficient. We call (13) a discrete-valued AR equation, because the solution y_n, the coefficients, and the white noise are all random and discrete-valued. When $K=2$ and y_n is a binary sequence, however, (13) becomes a linear discrete-valued AR equation, which was first introduced in a previous paper.[8] This means that a stationary binary sequence of a simple Markov chain is always generated by a linear discrete-valued AR equation. In what follows, we implicitly assume $K \geq 3$.

Putting $y_{n-1}=v_l$ with $l=1, 2, 3, \cdots, K$ in (13), we obtain the relationship between $F(x, v_l)$ and the coefficient function $D_{1,p}(x)$ as,

$$[F(x, v_1), F(x, v_2), \cdots, F(x, v_K)]=[D_{1,0}(x), D_{1,1}(x), \cdots, D_{1,K-1}(x)] \cdot \mathbf{V}, \tag{14}$$

where $\mathbf{V}$ is the Vandermonde matrix

$$\mathbf{V}=[\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_K]=\begin{bmatrix} 1 & 1 & \cdots & 1 \\ v_1 & v_2 & \cdots & v_K \\ v_1^2 & v_2^2 & \cdots & v_K^2 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ v_1^{K-1} & v_2^{K-1} & \cdots & v_K^{K-1} \end{bmatrix}. \tag{15}$$

Since the inverse matrix $\mathbf{V}^{-1}$ always exists under the condition (1), we may determine $D_{1,p}(x)$ when $F(x, v_k)$ is given. Since $\mathbf{V} \cdot \mathbf{V}^{-1}=\mathbf{I}$ (unit matrix), (15) yields a useful relationship,

$$[v_1^k, v_2^k, \cdots, v_K^k] \cdot \mathbf{V}^{-1}=[\delta_{0,k}, \delta_{1,k}, \cdots, \delta_{K-1,k}], \quad (k=0, 1, 2, \cdots, K-1). \tag{16}$$

3. Product Representation of Vector Random Sequence

We have demonstrated that a finite Markov chain can be represented by an AR equation. Since (13) is nonlinear, it is difficult to obtain an analytical solution y_n. In this section, however, we consider that an event is a K-dimensional vector in (12), which corresponds to a numerical state in (1). This makes it possible to obtain another AR equation which is linear with respect to the vector-valued Markov chain.

As described above, any function of y_{n-1} can be represented by an inner product with the base vector (11). Applying this again to the vector $\mathbf{Y}_n$:

$$\mathbf{Y}_n=\begin{bmatrix} 1 \\ y_n \\ y_n^2 \\ \cdot \\ \cdot \\ y_n^{K-1} \end{bmatrix}=\begin{bmatrix} 1 \\ F(x_n, y_{n-1}) \\ F^2(x_n, y_{n-1}) \\ \cdot \\ \cdot \\ F^{K-1}(x_n, y_{n-1}) \end{bmatrix}, \tag{17}$$

and a power of $F(x_n, y_{n-1})$, we obtain a linear AR equation for the K-dimensional vector sequence $\mathbf{Y}_n$:

$$\mathbf{Y}_n=\mathbf{D}(x_n) \cdot \mathbf{Y}_{n-1}, \tag{18}$$

where the coefficient matrix function $\mathbf{D}(x)$ is given by

$$\mathbf{D}(x)=\begin{bmatrix} 1 & 0 & \cdots & 0 \\ D_{1,0}(x) & D_{1,1}(x) & \cdots & D_{1,K-1}(x) \\ D_{2,0}(x) & D_{2,1}(x) & \cdots & D_{2,K-1}(x) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ D_{K-1,0}(x) & D_{K-1,1}(x) & \cdots & D_{K-1,K-1}(x) \end{bmatrix}. \tag{19}$$

Putting $\mathbf{Y}_{n-1}=\mathbf{v}_l$ with $l=1,\ 2,\ 3,\ \cdots,\ K$ in (18) and (17), we obtain the relationship between $D_{m,p}(x)$ and the system function $F(x, v_l)$ as,

$$\mathbf{F}(x)=\mathbf{D}(x)\cdot\mathbf{V}=\mathbf{V}\cdot\mathbf{T}(x), \tag{20}$$

where the second equality is obtained by (10), and $\mathbf{F}(x)$ is a $K\times K$ matrix function,

$$\mathbf{F}(x)=\begin{bmatrix} 1 & 1 & \cdots & 1 \\ F(x,\ v_1) & F(x,\ v_2) & \cdots & F(x,\ v_K) \\ F^2(x,\ v_1) & F^2(x,\ v_2) & \cdots & F^2(x,\ v_K) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ F^{K-1}(x,\ v_1) & F^{K-1}(x,\ v_2) & \cdots & F^{K-1}(x,\ v_K) \end{bmatrix}. \tag{21}$$

It is important to note that, even when y_n is a stationary Markov chain, its power, such as y_n^2, is stationary but is not always a Markov chain. Since $\mathbf{D}(x_n)$ is an independent stochastic sequence, however, (18) means that the vector $\mathbf{Y}_n$ again becomes a Markov sequence.

Let us calculate the average values of the coefficient $\mathbf{D}(x_n)$ and $\mathbf{F}(x_n)$ to obtain statistical properties of the Markov chain. Since x_n is an independent stationary random sequence with uniform distribution over the interval $[0,\ 1]$, we obtain from (6), (8) and (10),

$$\langle F^m(x_n,\ v_l)\rangle=\sum_{k=1}^{K} v_k^m\cdot t_{k,l},\quad (m=0,\ 1,\ 2,\ \cdots). \tag{22}$$

By (22) and (20) we then obtain the average of $\mathbf{F}(x_n)$ as

$$\langle\mathbf{F}(x_n)\rangle=\mathbf{V}\cdot\mathbf{t}=\mathbf{d}\cdot\mathbf{V}, \tag{23}$$

where $\mathbf{d}$ is the average value of $\mathbf{D}(x_n)$,

$$\mathbf{d}=\langle\mathbf{D}(x_n)\rangle=\mathbf{V}\cdot\mathbf{t}\cdot\mathbf{V}^{-1}. \tag{24}$$

By (24), we obtain the characteristic equation for the eigen value Λ,

$$\|\Lambda\cdot\mathbf{I}-\mathbf{t}\|=\|\Lambda\cdot\mathbf{I}-\mathbf{d}\|=0, \tag{25}$$

where $\|\mathbf{A}\|$ denotes the determinant of $\mathbf{A}$. This equation means that $\mathbf{t}$ and $\mathbf{d}$ have the same eigen values. Since $\mathbf{t}$ has an eigen value of 1 with multiplicity one and any other eigen values are less than unity in modulus in the complex plane, so does the matrix $\mathbf{d}$. Note that random matrix $\mathbf{D}(x_n)$ and $\mathbf{T}(x_n)$ also have eigen values of 1. Since $\langle\mathbf{Y}_n\rangle=\mathbf{V}\cdot\mathbf{q}$ by definition of the average, we obtain from (18),

$$\langle\mathbf{Y}_n\rangle=\mathbf{d}\cdot\langle\mathbf{Y}_n\rangle,\quad \mathbf{q}=\mathbf{V}^{-1}\cdot\langle\mathbf{Y}_n\rangle, \tag{26}$$

which means that the average vector $\langle\mathbf{Y}_n\rangle$ is an eigen vector of $\mathbf{d}$ with eigen value 1; a formula will be given below for calculating the average. This also implies that the stationary probability $\mathbf{q}$ can be uniquely determined from the statistical moment $\langle\mathbf{Y}_n\rangle$. This property holds because the Markov chain has finite states. However, it is well known that probability distributions cannot be determined uniquely from statistical moments in general.

Next, we consider the $K \times K$ correlation matrix $\mathbf{R}(m)$,

$$\mathbf{R}(m) = \langle (\mathbf{Y}_n - \langle \mathbf{Y}_n \rangle) \otimes (\mathbf{Y}^t_{n-m} - \langle \mathbf{Y}^t_{n-m} \rangle) \rangle = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}(m) \end{bmatrix}, \tag{27}$$

where $\otimes$ indicates the Kronecker product and $\mathbf{C}(m)$ is the $(K-1) \times (K-1)$ correlation matrix defined in section 6 below. The first column and row of $\mathbf{R}(m)$ are zero, because the first element of $\mathbf{Y}_n$ is 1. Solving (18) by iteration, we obtain the product representation of the vector sequence,

$$\mathbf{Y}_n = \mathbf{D}(x_n) \cdot \mathbf{D}(x_{n-1}) \cdots \mathbf{D}(x_1) \cdot \mathbf{Y}_0, \tag{28}$$

where $\mathbf{D}(x_n)$ is an independent sequence. This equation directly represents the causality such that vector $\mathbf{Y}_n$ depends on the present and past inputs x_{n-k} with $k \geq 0$. From (28) and (26), we easily obtain the $K \times K$ correlation matrix $\mathbf{R}(m)$,

$$\mathbf{R}(m) = \begin{cases} \mathbf{d}^m \cdot \mathbf{R}(0) & m \geq 0 \\ \mathbf{R}(0) \cdot [\mathbf{d}^t]^m & m \leq 0 \end{cases}. \tag{29}$$

This implies that the correlation matrix of a Markov chain can be written by an exponential function. Using this relationship, we will calculate the spectrum matrix in the next section. The correlation matrix enjoys the symmetry,

$$\mathbf{R}(-m) = \mathbf{R}^t(m), \quad \mathbf{R}(0) = \mathbf{R}^t(0), \tag{30}$$

and $\mathbf{R}(0)$ is given explicitly as,

$$\mathbf{R}(0) = \mathbf{V} \cdot \begin{bmatrix} q_1 & 0 & 0 & \cdots & 0 \\ 0 & q_2 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & q_K \end{bmatrix} \cdot \mathbf{V}^t - \mathbf{V} \cdot \mathbf{q} \otimes \mathbf{q}^t \cdot \mathbf{V}^t, \tag{31}$$

where the diagonal matrix in the right hand side is the joint probability distribution $\mathbf{P}_0$, as is discussed in Section 5.

4. Spectrum Matrix

We obtained a formula for the correlation matrix $\mathbf{R}(m)$. As a Fourier transformation of the correlation matrix, we derive the spectrum matrix in this section.

We define $K \times K$ spectrum matrix $\mathbf{S}(\lambda)$ as the Fourier transformation of $\mathbf{R}(m)$,

$$\mathbf{S}(\lambda) = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{s}(\lambda) \end{bmatrix} = \sum_{m=-\infty}^{\infty} \mathbf{R}(m) \cdot e^{2\pi i m \lambda}, \tag{32}$$

where $\mathbf{s}(\lambda)$ is the $(K-1) \times (K-1)$ spectrum matrix defined later. Inserting (29) into this, the spectrum matrix $\mathbf{S}(\lambda)$ can be calculated as follows:

$$\mathbf{S}(\lambda) = -\mathbf{R}(0) + \sum_{m=0}^{\infty} \mathbf{d}^m \cdot e^{2\pi i m \lambda} \cdot \mathbf{R}(0) + \mathbf{R}(0) \cdot \sum_{m=0}^{\infty} [\mathbf{d}^t]^m \cdot e^{2\pi i m \lambda} \tag{33}$$

$$= -\mathbf{R}(0) + [\mathbf{I} - e^{2\pi i \lambda} \cdot \mathbf{d}]^{-1} \cdot \mathbf{R}(0) + \mathbf{R}(0) \cdot [\mathbf{I} - e^{-2\pi i \lambda} \cdot \mathbf{d}^t]^{-1} \tag{34}$$

$$=-\mathbf{R}(0)+\mathbf{V}\cdot[\mathbf{I}-e^{2\pi i\lambda}\cdot\mathbf{t}]^{-1}\cdot\mathbf{V}^{-1}\cdot\mathbf{R}(0)$$
$$+\mathbf{R}(0)\cdot[\mathbf{V}^{t}]^{-1}\cdot[\mathbf{I}-e^{-2\pi i\lambda}\cdot\mathbf{t}^{t}]^{-1}\cdot\mathbf{V}^{t}, \tag{35}$$

where we have used (24) to drive (35). Our formula (35) is convenient for evaluating the spectrum matrix when the transition matrix $\mathbf{t}$ is given. Since the first column and row of $\mathbf{R}(m)$ are 0, the two series in (33) converge for real λ. However, it should be noted that $[\mathbf{I}-e^{2\pi i\lambda}\cdot\mathbf{t}]^{-1}$ diverges at $\lambda=0$, because $\mathbf{t}$ has an eigen value equal to 1. Therefore, (35) is valid except at $\lambda=0$.

Now, let us obtain a formula for an element of the matrix $\mathbf{S}(\lambda)$. To calculate an element, we introduce K-dimensional unit vectors $\mathbf{u}_k$ by the relationship,

$$\mathbf{u}_k=[\delta_{k,1},\ \delta_{k,2},\ \cdots,\ \delta_{k,K}]^t,\ (k=1,\ 2,\ \cdots,\ K). \tag{36}$$

Using (36), we get $S_{m,n}(\lambda)$ as

$$S_{m,n}(\lambda)=\mathbf{u}_m^t\cdot\mathbf{S}(\lambda)\cdot\mathbf{u}_n \tag{37}$$
$$=-R_{m,n}(0)+\mathbf{u}_m^t\cdot\mathbf{V}\cdot[\mathbf{I}-e^{2\pi i\lambda}\cdot\mathbf{t}]^{-1}\cdot\mathbf{V}^{-1}\cdot\mathbf{R}(0)\cdot\mathbf{u}_n$$
$$+\mathbf{u}_m^t\cdot\mathbf{R}(0)\cdot[\mathbf{V}^t]^{-1}\cdot[\mathbf{I}-e^{-2\pi i\lambda}\cdot\mathbf{t}^t]^{-1}\cdot\mathbf{V}^t\cdot\mathbf{u}_n. \tag{38}$$

Using the formula for a regular matrix $\mathbf{A}$ and column vectors $\mathbf{x}$ and $\mathbf{y}$[3),9)],

$$\mathbf{x}^t\cdot\mathbf{A}^{-1}\cdot\mathbf{y}=-\frac{\begin{Vmatrix}0 & \mathbf{x}^t\\ \mathbf{y} & \mathbf{A}\end{Vmatrix}}{\|\mathbf{A}\|}=-\frac{\begin{Vmatrix}0 & \mathbf{y}^t\\ \mathbf{x} & \mathbf{A}^t\end{Vmatrix}}{\|\mathbf{A}^t\|}, \tag{39}$$

we obtain a closed form expression of an element of the spectrum matrix,

$$S_{m,n}(\lambda)=-R_{m,n}(0)-\frac{\begin{Vmatrix}0 & \mathbf{u}_m^t\cdot\mathbf{V}\\ \mathbf{V}^{-1}\cdot\mathbf{R}(0)\cdot\mathbf{u}_n & \mathbf{I}-\exp(2\pi i\lambda)\cdot\mathbf{t}\end{Vmatrix}}{\|\mathbf{I}-\exp(2\pi i\lambda)\cdot\mathbf{t}\|}$$
$$-\frac{\begin{Vmatrix}0 & \mathbf{u}_n^t\cdot\mathbf{V}\\ \mathbf{V}^{-1}\cdot\mathbf{R}(0)\cdot\mathbf{u}_m & \mathbf{I}-\exp(-2\pi i\lambda)\cdot\mathbf{t}\end{Vmatrix}}{\|\mathbf{I}-\exp(-2\pi i\lambda)\cdot\mathbf{t}\|}, \tag{40}$$

where

$$\mathbf{V}^t\cdot\mathbf{u}_m=\begin{bmatrix}v_1^{m-1}\\ v_2^{m-1}\\ \cdot\\ \cdot\\ v_K^{m-1}\end{bmatrix},\quad \mathbf{V}^{-1}\cdot\mathbf{R}(0)\cdot\mathbf{u}_m=\begin{bmatrix}(v_1^{m-1}-\langle y_n^{m-1}\rangle)\cdot q_1\\ (v_2^{m-1}-\langle y_n^{m-1}\rangle)\cdot q_2\\ \cdot\\ \cdot\\ (v_K^{m-1}-\langle y_n^{m-1}\rangle)\cdot q_K\end{bmatrix}. \tag{41}$$

When $m=n=2$, our equation (40) agrees with a known formula[3)] for the power spectrum of y_n. Therefore, (40) is regarded as a generalization of the known result. Our formulas (40) and (35) are convenient for evaluating the spectrum matrix when the transition matrix $\mathbf{t}$ is given. As $\mathbf{d}$ and $\mathbf{t}$ commonly have eigen values equal to 1, the numerator and denominator in (40) vanish at $\lambda=0$. However, $\lim_{\lambda\to0}S_{m,n}(\lambda)$ gives the correct value for $S_{m,n}(0)$. This means that special caution is needed in numerical calculations. In what follows, we derive another formula for the spectrum matrix that is free from such a drawback and is suitable for numerical calculation.

5. Representation by Unit Random Vectors

We first consider the Markov chain that takes the numerical states in (1). Then, making a mapping from the numerical states in (1) to vector states in (12), we obtain the equation (18) for the vector $\mathbf{Y}_n$. Let us consider another transformation from the vector states (12) to unit vectors in the K-dimensional space.

Since $\mathbf{V}^{-1}\cdot\mathbf{V}=\mathbf{I}$, we obtain the relationship between vector states $\mathbf{v}_k$ in (12) and unit vectors $\mathbf{u}_k$,

$$\mathbf{V}^{-1}\cdot\mathbf{v}_k=\mathbf{u}_k. \tag{42}$$

By this transformation, $\mathbf{Y}_n$ is transformed into a K-dimensional unit vector $\mathbf{Q}_n$,

$$\mathbf{Q}_n=[Q_n(1),\ Q_n(2),\ \cdots,\ Q_n(K)]^t=\mathbf{V}^{-1}\cdot\mathbf{Y}_n,$$
$$\sum_{k=1}^{K}Q_n(k)=1, \tag{43}$$

where an element of $\mathbf{Q}_n$ is a binary sequence by (42),

$$Q_n(k)=\begin{cases}0, & y_n\neq v_k\\ 1, & y_n=v_k\end{cases}, \tag{44}$$

which is one only when $y_n=v_k$. By the transformation (43), the AR equation (18) becomes a linear AR equation for unit vector sequence,

$$\mathbf{Q}_n=\mathbf{T}(x_n)\cdot\mathbf{Q}_{n-1}. \tag{45}$$

Taking the ensemble average of this equation and using (43), we obtain another representation of the stationary probability,

$$\mathbf{q}=\langle\mathbf{Q}_n\rangle=\mathbf{V}^{-1}\cdot\langle\mathbf{Y}_n\rangle. \tag{46}$$

In terms of $\mathbf{Q}_n$, we may define the joint probability distribution $\mathbf{P}_n=[P_n(k,\ l)]$, where $P_n(k,\ l)=Prob\{y_m=v_k,\ y_{m-n}=v_l\}$, as

$$\begin{aligned}\mathbf{P}_n&=\langle\mathbf{Q}_m\otimes\mathbf{Q}^t_{m-n}\rangle\\ &=\mathbf{V}^{-1}\cdot\langle\mathbf{Y}_m\otimes\mathbf{Y}^t_{m-n}\rangle\cdot[\mathbf{V}^t]^{-1}\\ &=\mathbf{V}^{-1}\cdot\mathbf{R}(n)\cdot[\mathbf{V}^t]^{-1}+\mathbf{q}\otimes\mathbf{q}^t.\end{aligned} \tag{47}$$

Inserting (29) and (31) into this, we obtain the well known formula :

$$\mathbf{P}_n=\mathbf{t}^n\cdot\begin{bmatrix}q_1 & 0 & 0 & \cdots & 0\\ 0 & q_2 & 0 & \cdots & 0\\ \cdot & \cdot & \cdot & \cdots & \cdot\\ \cdot & \cdot & \cdot & \cdots & \cdot\\ 0 & 0 & 0 & \cdots & q_K\end{bmatrix}, \tag{48}$$

where the diagonal matrix is $\mathbf{P}_0$, which appeared in (31) above. The relation (47) means that the joint probability distribution $\mathbf{P}_n$ can be constructed from the correlation matrix $\mathbf{R}(n)$ and the stationary probability $\mathbf{q}$. This may be useful for designing a Markov signal with known probability distribution and correlation function.

6. Linear AR Equation with Constant Coefficient Matrix

As can be seen in (17), the vector $\mathbf{Y}_n$ is a redundant expression of the Markov chain, because the first element of $\mathbf{Y}_n$ is 1. Removing such redundancy, we represent the Markov chain by the $(K-1)$-dimension vector $\mathbf{y}_n$,

$$\mathbf{y}_n=[y_n,\ y_n^2,\ \cdots,\ y_n^{K-1}]^t, \tag{49}$$

which is related with $\mathbf{Y}_n=[1,\ \mathbf{y}_n]^t$. Then, we rewrite (18) as a linear AR equation for the vector sequence

$$\mathbf{y}_n=\mathbf{E}(x_n)+\mathbf{G}(x_n)\cdot\mathbf{y}_{n-1}, \tag{50}$$

which is a vector extension of the discrete-valued AR equation introduced previously[5),8)]. Here, $\mathbf{E}(x)$ is a $(K-1)$ vector function and $\mathbf{G}(x)$ is a $(K-1)\times(K-1)$ matrix futction,

$$\mathbf{E}(x)=\begin{bmatrix} D_{1,0}(x) \\ D_{2,0}(x) \\ \cdot \\ \cdot \\ D_{K-1,0}(x) \end{bmatrix},\quad \mathbf{G}(x)=\begin{bmatrix} D_{1,1}(x) & D_{1,2}(x) & \cdots & D_{1,K-1}(x) \\ D_{2,1}(x) & D_{2,2}(x) & \cdots & D_{2,K-1}(x) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ D_{K-1,1}(x) & D_{K-1,2}(x) & \cdots & D_{K-1,K-1}(x) \end{bmatrix}. \tag{51}$$

From (51) and (21) one easily finds the relationship,

$$\mathbf{D}(x)=\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{E}(x) & \mathbf{G}(x) \end{bmatrix}. \tag{52}$$

Since $\mathbf{E}(x_n)$ is an independent stochastic sequence, however, (50) means that the vector $\mathbf{y}_n$ again becomes a Markov sequence.

Let us calculate the average vector $\langle\mathbf{y}_n\rangle$, the $(K-1)\times(K-1)$ correlation matrix $\mathbf{C}(m)$ and the $(K-1)\times(K-1)$ spectrum matrix $\mathbf{s}(\lambda)$ from (50).

Writing the averages

$$\mathbf{e}=\langle\mathbf{E}(x_n)\rangle,\quad \mathbf{g}=\langle\mathbf{G}(x_n)\rangle,\quad \mathbf{d}=\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{e} & \mathbf{g} \end{bmatrix}, \tag{53}$$

we obtain from (23),

$$\mathbf{t}=[t_{k,l}]=\mathbf{V}^{-1}\cdot\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{e} & \mathbf{g} \end{bmatrix}\cdot\mathbf{V}, \tag{54}$$

which gives a characteristic equation for eigen value Λ,

$$\|\Lambda\cdot\mathbf{I}-\mathbf{t}\|=(\Lambda-1)\cdot\|\Lambda\cdot\mathbf{I}-\mathbf{g}\|=0. \tag{55}$$

We have assumed that $\mathbf{t}$ has an eigen value of 1 with simple multiplicity, and that any other eigen values are less than unity in modulus. Thus, any eigen value of $\mathbf{g}$ becomes less than unity in modulus. This will be used below.

Solving (50) by iteration, we obtain a series representation

$$\mathbf{y}_n=\mathbf{E}(x_n)+\mathbf{G}(x_n)\cdot\mathbf{E}(x_{n-1})+\mathbf{G}(x_n)\cdot\mathbf{G}(x_{n-1})\cdot\mathbf{E}(x_{n-2})+\cdots, \tag{56}$$

which again shows that $\mathbf{y}_n$ depends on the present input x_n and the past inputs x_{n-m} with $m>0$. Since x_n is an independent sequence, term by term averaging gives the average vector $\langle\mathbf{y}_n\rangle$,

$$\langle\mathbf{y}_n\rangle=\mathbf{e}+\mathbf{g}\cdot\mathbf{e}+\mathbf{g}^2\cdot\mathbf{e}+\mathbf{g}^3\cdot\mathbf{e}+\cdots \tag{57}$$

$$=[\mathbf{I}-\mathbf{g}]^{-1}\cdot\mathbf{e}. \tag{58}$$

As any eigen value of $\mathbf{g}$ is less than unity in modulus, the right-hand side of (57) always converges to a constant vector (58). Equation (58) is useful for calculating the average vector numerically.

Let us define the $(K-1)\times(K-1)$ correlation matrix $\mathbf{C}(m)$. Subtracting the average value, we put

$$\mathbf{Z}_n=\mathbf{y}_n-\langle\mathbf{y}_n\rangle, \tag{59}$$

to define the $(K-1)\times(K-1)$ correlation matrix $\mathbf{C}(m)$ by,

$$\mathbf{C}(n)=\langle\mathbf{Z}_k\otimes\mathbf{Z}^t_{k-n}\rangle, \tag{60}$$

$$\mathbf{C}(n)=\mathbf{C}^t(-n),\quad \mathbf{C}(0)=\mathbf{C}^t(0), \tag{61}$$

which is related to $\mathbf{R}(n)$ by (27) and $\mathbf{C}(0)$ is obtained from (31) above.

So far, we have described a discrete-valued AR equations with random coefficients. However, we can obtain a linear AR equation for $\mathbf{Z}_n$, where the cofficient matrix is constant. Rewriting (50), we obtain such a linear AR equation,

$$\mathbf{Z}_n=\mathbf{g}\cdot\mathbf{Z}_{n-1}+\mathbf{W}_n, \tag{62}$$

where $\mathbf{W}_n$ is a $(K-1)$ vector random sequence,

$$\mathbf{W}_n=[\mathbf{G}(x_n)-\mathbf{g}]\cdot\mathbf{y}_{n-1}+\mathbf{E}(x_n)-\mathbf{e}. \tag{63}$$

It can be shown that the vector $\mathbf{W}_n$ is white noise with zero average,

$$\langle\mathbf{W}_n\rangle=0,\quad \langle\mathbf{W}_m\otimes\mathbf{W}^t_n\rangle=\delta_{m,n}\hat{\sigma}, \tag{64}$$

where $\hat{\sigma}$ is the covariance matrix of $\mathbf{W}_n$,

$$\hat{\sigma}=\hat{\sigma}^t=\mathbf{C}(0)-\mathbf{g}\cdot\mathbf{C}(0)\cdot\mathbf{g}^t. \tag{65}$$

Even though $\mathbf{W}_n$ explicitly depends on $\mathbf{y}_{n-1}$ in (63), one can easily find from (63),

$$\langle\mathbf{W}_n\otimes\mathbf{y}^t_{n-m}\rangle=\langle\mathbf{W}_n\otimes\mathbf{Z}^t_{n-m}\rangle=0,\quad m\geq 1. \tag{66}$$

These relationships mean that (62) is a conventional AR equation with discrete-valued white noise excitation $\mathbf{W}_n$. In other words, (62) is a linear prediction formula for the Markov chain. We note that this is the first time such a formula has been introduced.

Multiplying $\mathbf{Z}^t_{n-m}$ with a positive m to the both sides of (62) and taking the ensemble average, we obtain an equation for $\mathbf{C}(m)$,

$$\mathbf{C}(m)=\mathbf{g}\cdot\mathbf{C}(m-1),\quad m\geq 1, \tag{67}$$

which along with (61) yields

$$\mathbf{C}(m)=\begin{cases}\mathbf{g}^m\cdot\mathbf{C}(0), & m\geq 1\\ \mathbf{C}(0)\cdot[\mathbf{g}^t]^{-m}, & m<0\end{cases}. \tag{68}$$

Next, let us calculate the spectrum matrix $\mathbf{s}(\lambda)$ as a Fourier transformation of the correlation matrix. Using (65), we obtain

$$\mathbf{s}(\lambda)=\sum_{m=-\infty}^{\infty}\mathbf{C}(m)\cdot e^{2\pi im\lambda} \tag{69}$$

$$=\sum_{m=0}^{\infty}\mathbf{g}^m e^{2\pi im\lambda}\mathbf{C}(0)+\sum_{m=1}^{\infty}\mathbf{C}(0)\cdot[\mathbf{g}^t]^m\cdot e^{-2\pi im\lambda} \tag{70}$$

$$=[\mathbf{I}-e^{2\pi i\lambda}\mathbf{g}]^{-1}\cdot\mathbf{C}(0)+e^{-2\pi i\lambda}\mathbf{C}(0)\cdot\mathbf{g}^t\cdot[\mathbf{I}-e^{-2\pi i\lambda}\mathbf{g}^t]^{-1} \tag{71}$$

$$=[\mathbf{I}-e^{2\pi i\lambda}\mathbf{g}]^{-1}\cdot\hat{\sigma}\cdot[\mathbf{I}-e^{-2\pi i\lambda}\mathbf{g}^t]^{-1}, \tag{72}$$

where two series in (70) converge uniformly in λ because any eigen values of $\mathbf{g}$ are less than unity in modulus. Therefore, (72) is a useful formula for calculating the spectrum matrix numerically.

7. Function of Markov Chains

In some applications, a random sequence given by a function of a finite Markov chain is a subject of interest. We include here brief remarks on such a case. We consider a random sequence β_n generated by a Markov chain y_n,

$$\beta_n=h(y_n)=h_0+\mathbf{h}^t\cdot\mathbf{y}_n, \tag{73}$$

where h_0 and $\mathbf{h}$ are scalar and vector constants, respectively,

$$\mathbf{h}=[h_1,\ h_2,\ \cdots,\ h_{K-1}]^t. \tag{74}$$

Even when y_n is a K-valued Markov chain, β_n is not always a Markov chain and takes K different values at most. After simple calculation, we obtain the average $\langle\beta_n\rangle$, correlation $C_\beta(n)$ and spectrum $S_\beta(\lambda)$ as,

$$\langle\beta_n\rangle=h_0+\mathbf{h}^t\cdot\langle\mathbf{y}_n\rangle, \tag{75}$$

$$C_\beta(n)=\langle(\beta_n-\langle\beta_n\rangle)\cdot(\beta_0-\langle\beta_0\rangle)\rangle=\mathbf{h}^t\cdot\mathbf{C}(n)\cdot\mathbf{h}, \tag{76}$$

$$S_\beta(\lambda)=\sum_{m=-\infty}^{\infty}C_\beta(m)\cdot e^{2\pi im\lambda}=\mathbf{h}^t\cdot\mathbf{s}(\lambda)\cdot\mathbf{h}. \tag{77}$$

This relationship implies that the correlation matrix $\mathbf{s}(\lambda)$ is completely in the sense that the power spectrum of a stationary sequence generated by any function of the Markov chain y_n is represented by $\mathbf{s}(\lambda)$.

8. Simple Application to an Inverse Problem

In the theory of finite Markov chains, the transition probability is always given apriori. However, an inverse problem[10] that determines the transition probability from a given stationry probability, correlation function and other statistical properties has not been solved. For a special case where the correlation function is exponential, however, we gave a simple solution,[11] where the transition probability is determined by a given stationary probability and correlation length. The solution, which was found intuitively without theoretical background, makes it possible to design a discrete-valued sequence with an arbitrary probability distribution.

In this section, however, we clarify that such a simple solution is related to a linear discrete-valued equation, which is a special case of the non-linear discrete-valued AR equation (13).

If we put $D_{1,l}(x)=0$, $(l=2,\ 3,\ \cdots,\ K-1)$ in (13), we obtain a linear discrete-valued AR equation,

$$y_n=D_{1,0}(x_n)+D_{1,1}(x_n)\cdot y_{n-1}, \tag{78}$$

where we assume that the coefficient functions satisfy the conditions,

$$D_{1,0}(x)\cdot D_{1,1}(x_n)=0,\quad D_{1,1}^2(x_n)=D_{1,1}(x_n). \tag{79}$$

Under these conditions, one easily finds

$$y_n^m=D_{1,0}^m(x_n)+D_{1,1}(x_n)\cdot y_{n-1}^m,\ (m=1,\ 2,\ \cdots,\ K-1). \tag{80}$$

This is a linear discrete-valued equation for y_n^m, and implies that y_n^m again becomes a Markov chain for $m=2,\ 3,\ \cdots,\ K-1$. Equation (80) also means that $\mathbf{G}(x)$ in (52) and its average $\mathbf{g}$ in (53) become diagonal.

If we put

$$g=\langle D_{1,1}(x_n)\rangle,\quad 0<g<1, \tag{81}$$

it can be shown from (54) and (58) that the transition probability $\mathbf{t}=[t_{k,l}]$ becomes

$$t_{k,l}=q_k+(\delta_{k,l}-q_k)\cdot g \tag{82}$$

which was found intuitively in a previous paper[11]. As can be easily seen, (82) enjoys (3) and has an eigen vector $\mathbf{q}=[q_1,\ q_2,\ \cdots,\ q_K]^t$ with an eigen value of 1. From (6), (20) and (82), the coefficient functions can be defined as

$$D_{1,1}(x)=\begin{cases}0, & 0<x\leq 1-g\\ 1, & 1-g<x\leq 1,\end{cases} \tag{83}$$

$$D_{1,0}(x)=v_k,\quad (1-g)\sum_{l=1}^{k-1}q_l<x\leq(1-g)\sum_{l=1}^{k}q_l,\quad (k=1,\ 2,\ \cdots,\ K). \tag{84}$$

where $D_{1,0}(x)$ is a binary function with values of only 1 or 0, and $D_{1,1}(x)$ is a multi-step function of x. Here, it is important to say that q_l's are arbitrary positive numbers enjoying (5).

Since x_n is uniformly distributed in the interval $[0,\ 1]$, we obtain from (80) and (84),

$$\langle y_n^m\rangle=\frac{\langle D_{1,0}^m(x_n)\rangle}{1-g}=\sum_{k=1}^{k}v_k^m\cdot q_k,\quad (m=1,\ 2,\ \cdots,\ K-1). \tag{85}$$

This directly indicates that y_n has the probability distribution given by q_l's. From (78) and (80), we obtain the correlation functions,

$$\begin{aligned}C_{k,l}(m)&=\langle(y_n^k-\langle y_n^k\rangle)\cdot(y_{n-m}^l-\langle y_n^l\rangle)\rangle\\&=(\langle y_n^{k+1}\rangle-\langle y_n^k\rangle\cdot\langle y_n^l\rangle)\cdot g^{|m|},\quad (k,\ l=1,\ 2,\ \cdots,\ K-1),\end{aligned} \tag{86}$$

which is an exponential correlation function with a correlation length $-1/ln(g)$. Therefore, the transition probability (82) and the AR equation (78) are uniquely determined by a stationary probability and a correlation length.

It is well known that conventional AR and ARMA models define a continuous-valued random sequence with various power spectra. However, they do not work well for non-Gaussian random sequences. This is essentially due to the central limit theorem in the probability theory. However, the linear discrete-valued equation (78) may define discrete-valued sequence with any probability distributions, because positive q_l's are arbitrary numbers under the condition (5). In fact, stationary random sequences with several probability distributions such as uniform distribution, Laplace distribution with spikes and Weibull distribution were generated in a computer.[11]

9. Conclusions

A finite Markov chain with a stationary probability is commonly discussed in terms of the transition probability on an abstract sample space. In this paper, however, we demonstrated that the Markov chain can be expressed as a K-dimensional vector sequence or unit vector sequence in the K-dimensional space. We studied AR representations for such a vector sequence. Using such AR representations, we have derived an explicit formula for the spectrum matrix, which has a drawback due to the redundancy of the K-dimensional vector expression. Removing the redundancy, we introduced another expression where the Markov chain was regarded as a $(K-1)$-dimensional vector sequence, which enjoys an AR equation with constant coefficient matrix and white noise excitation. Then, we obtained another formula for the average vector, the correlation matrix and the spectrum matrix.

The AR representations and formulas may be applied to the theory of nonlinear prediction and spectrum analysis of Markov signals. These formulas are also useful for solving inverse problems which determine a transition probability from a given probability distribution, known correlation function, and other statistical properties. However, such an inverse problem, which was solved in a special linear case, will be left for future study.

Acknowledgments

The author would like to thank Yasuo Yoshida, Professor of Electronics and Information Science, for his valuable comments. This work was supported in part by the Grant-In-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan.

Department of Electronics and Information Science,
Faculty of Engineering and Design,
Kyoto Institute of Technology,
Matsugasaki, Sakyo-ku, Kyoto 606

References

1) H. Yasuda and H. Inose, *IECE Trans.*, **53-A**, 604 (1970).
2) Y. Yoshida, *IECE Trans.*, **56-A**, 536 (1978).
3) H. Sugiyama, *IECE Trans.*, **E66**, 155 (1983).
4) S. Karlin, 'A first course in stochastic processes', Academic, New York (1966).
5) J. Nakayama, *IEICE Trans. Fundamentals*, **E76-A**, 1031 (1993).
6) J. Nakayama and H. Motoyama, *IEICE Trans. Fundamentals*, **E76-A**, 2114 (1993).
7) H. Ogura, Technical Report of IECE, IT69-29 (1969).
8) J. Nakayama and E. Omori, *J. Math. Phys.*, **29**, 1982 (1988).
9) R. Courant and D. Hilbert: 'Methods of mathematical physics I', Interscience (1953).
10) J. Justesen, *IEEE Trans. Inform. Theory*, **IT-28**, 457 (1982).
11) J. Nakayama, *IEICE Trans. Fundamentals*, **E77-A**, 917 (1994).