THE CONDITION OF REGULARITY IN SIMPLE MARKOV CHAINS

By J. GANI*

[*Manuscript received December* 29, 1955]

Summary

The paper generalizes a proof, and outlines an alternative to it, for the well-known theorem on the conditions of regularity in a simple Markov chain ; this is that the necessary and sufficient conditions for a chain to be regular are that the latent root 1 of the stochastic matrix for the chain must be simple, and the remaining roots have moduli less than 1.

I. INTRODUCTION

A simple Markov chain with s possible states $E_1, \ldots, E_s$, is defined by the stochastic matrix of transition probabilities

$$\mathbf{p}=\{p_{ij}\} \quad (i, j=1, \ldots, s),$$

where $p_{ij}=Pr(E_j \mid E_i)\geqslant 0$. Since, in a one-step transition from E_i, one of the states $E_i, \ldots, E_s$ must be reached, it follows that the probabilities in any row i of the stochastic matrix must sum up to 1, so that

$$\sum_{j=1}^{s} p_{ij}=1.$$

This property leads to the interesting result that 1 is a latent root of any stochastic matrix $\mathbf{p}$; for if we write

$$D(\mu)=\mid \mu\mathbf{I}-\mathbf{p} \mid, \qquad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

we find that on adding all columns of the determinant to the first,

$$D(\mu)=\begin{vmatrix} \mu-1 & -p_{12} & \ldots & -p_{1s} \\ \mu-1 & \mu-p_{22} & \ldots & -p_{2s} \\ \cdot & \cdot & \cdots & \cdot \\ \mu-1 & -p_{s2} & \ldots & \mu-p_{ss} \end{vmatrix},$$

from which it is clear that $\mu_1=1$ is a latent root of the characteristic equation $D(\mu)=0$. It is, in addition, possible to show that all the latent roots μ_r $(r=1, \ldots, s)$ of $\mathbf{p}$ have moduli $\mid \mu_r \mid\leqslant 1$; for if $D(\mu)=0$, there exists at least one non-trivial vector solution $\mathbf{R}(\mu)$ of the matrix equation

$$(\mu\mathbf{I}-\mathbf{p})\mathbf{R}(\mu)=0, \qquad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

* Australian National University, Canberra, A.C.T.; present address: Statistical Laboratory, University of Manchester, England.

where $\mathbf{R}(\mu)$ may have real or complex elements. If $R_i(\mu)$ is the element in $\mathbf{R}(\mu)$ with the largest modulus M, then from (**2**), since $\sum_{j=1}^{s} p_{ij}R_j=\mu R_i$, it follows that

$$|\mu| M \leqslant \sum_{j=1}^{s} p_{ij} |R_j| \leqslant M,$$

so that for all values of $r=1, 2, \ldots, s$, the modulus $|\mu_r| \leqslant 1$. Both these results are given by Fréchet (1938, p. 105).

Let us now consider the probability $p_{ij}^{(2)}$ of a transition from E_i to E_j in exactly two steps; it is easy to see, by taking into account all the possible ways in which this can occur, that

$$p_{ij}^{(2)}=\sum_{r=1}^{s} p_{ir}p_{rj}.$$

Similarly, it will follow that if $p_{ij}^{(n)}$ is the higher transition probability that E_j be reached from E_i in exactly n steps, that

$$p_{ij}^{(n)}=\sum_{r=1}^{s} p_{ir}p_{rj}^{(n-1)} \quad (n=2, \ldots).$$

Clearly, since one of the states $E_1, \ldots, E_s$ must necessarily be reached starting from E_i, then

$$\sum_{j=1}^{s} p_{ij}^{(n)}=1,$$

and the $p_{ij}^{(n)}$ are (i, j)th elements of a stochastic matrix, itself the nth power of the stochastic matrix $\mathbf{p}$, so that

$$\mathbf{p}^n=\{p_{ij}^{(n)}\}.$$

The Markov chain known as regular is that for which

$$\lim_{n\to\infty} p_{ij}^{(n)}=P_j \geqslant 0 \quad (j=1, \ldots, s), \quad \ldots\ldots\ldots\ldots\ldots \quad (3)$$

where P_j is some fixed finite stationary probability independent of the initial state E_i. This result can be written in matrix form as

$$\lim_{n\to\infty} \mathbf{p}^n=\mathbf{1}\mathbf{P}', \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

where $\mathbf{P}'$ is the row vector of probabilities P_j, and $\mathbf{1}$ is the column vector of unit elements. It is obvious that $\sum_{j=1}^{s} P_j=1$, so that at least one of these stationary probabilities is non-zero. Since

$$\mathbf{1}\mathbf{P}'=\lim_{n\to\infty} \mathbf{p}^{n+1}=\lim_{n\to\infty} \mathbf{p}^n\mathbf{p}=\mathbf{1}\mathbf{P}'\mathbf{p},$$

the row vector $\mathbf{P}'$ of probabilities P_j is a solution of the matrix equation

$$\mathbf{P}'=\mathbf{P}'\mathbf{p}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

and such a vector can easily be shown to be unique.

It is the purpose of the present paper to extend a known but not quite general proof, and to give an elementary alternative to this generalization, for the well-known theorem on the regularity of a Markov chain:

The necessary and sufficient conditions for a simple Markov chain to be regular, that is, for $\lim_{n\to\infty} p_{ij}^{(n)}=P_j$, are that the latent root 1 of the stochastic matrix for the chain be simple, and the remaining roots have moduli less than 1.

It is comparatively simple to prove this by considering higher transition probabilities in the case where the latent roots of $\mathbf{p}$ are assumed to be all simple, whether as in Bartlett (1955, Section 2.2) by the method of spectral resolution, or as in Feller (1950, Section 16.1) by using a partial fractions expansion. Fréchet has given a proof of the theorem in the general case when some of the latent roots may be multiple (1938, pp. 109–14), but this is not direct, arising as it does in the study of higher transition probabilities for various types of Markov chains. We briefly outline Bartlett's proof, give a generalization of it in the case where the latent roots of $\mathbf{p}$ may be multiple, which depends on a little-used theorem in canonical matrices, and finally provide an elementary alternative to this.

II. The Generalization of Bartlett's Proof

In essence, Bartlett's (1955) proof of the necessary conditions for the regularity of a Markov chain depends on the fact that, if the stochastic matrix $\mathbf{p}$ of the chain has latent roots, all of which are simple, it can be expressed as

$$\mathbf{p}=\mathbf{T}^{-1}\mathbf{u}\mathbf{T},$$

where $\mathbf{T}$ is a non-singular matrix, and $\mathbf{u}=\{\mu_j\delta_{ij}\}$ is the diagonal matrix whose elements are the latent roots μ_j $(j=1, \ldots, s)$ of $\mathbf{p}$. It is clear then, if equation (3) or its matrix equivalent (4) are to hold, that

$$\mathbf{p}^n=\mathbf{T}^{-1}\mathbf{u}^n\mathbf{T}\to\mathbf{1}\mathbf{P}',$$

so that

$$\lim_{n\to\infty}\mathbf{u}^n=\mathbf{T}\mathbf{1}\mathbf{P}'\mathbf{T}^{-1}, \qquad (6)$$

where the elements of $\mathbf{T}\mathbf{1}\mathbf{P}'\mathbf{T}^{-1}$ have fixed finite values. If therefore the unit root is $\mu_1=1$ say, it is clear that the remaining roots μ_r $(r=2, \ldots, s)$ are necessarily such that $|\mu_r|<1$.

The generalization of this in the case where the roots of $\mathbf{p}$ may be multiple depends on the theorem in canonical matrices (Turnbull and Aitken 1932, Section 6) which states that if $\mu_1, \ldots, \mu_t$ are latent roots of $\mathbf{p}$ of multiplicity $m_1, \ldots, m_t$ respectively, where $\sum_{r=1}^{t} m_r=s$, then the stochastic matrix can be expressed as

$$\mathbf{p}=\mathbf{T}^{-1}\mathbf{M}\mathbf{T}=\mathbf{T}^{-1}\begin{bmatrix}\mathbf{M}_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ & & & & \mathbf{M}_t\end{bmatrix}\mathbf{T},$$

where $\mathbf{T}$ is again non-singular, and the canonical matrix $\mathbf{M}$ is partitioned into t diagonal matrices $\mathbf{M}_1, \ldots, \mathbf{M}_t$ of orders $m_1, \ldots, m_t$ respectively, each matrix $\mathbf{M}_i$ consisting of one or more non-overlapping diagonal submatrices $\mathbf{C}_i^{(v)}$ of the "simple classical" type

$$\mathbf{C}_i^{(v)} = \begin{bmatrix} \mu_i & 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \mu_i & 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \mu_i & 1 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \mu_i \end{bmatrix},$$

where $v \leqslant m_i$ indicates the order. For example, if the root μ_i has multiplicity $m_i = 6$, a possible form for the matrix $\mathbf{M}_i$ could be

$$\mathbf{M}_i = \begin{bmatrix} [\mu_i] & & \\ & \begin{bmatrix} \mu_i & 1 & 0 \\ 0 & \mu_i & 1 \\ 0 & 0 & \mu_i \end{bmatrix} & \\ & & \begin{bmatrix} \mu_i & 1 \\ 0 & \mu_i \end{bmatrix} \end{bmatrix}.$$

If $m_i = 1$, the matrix $\mathbf{M}_i$ must consist of the single diagonal element μ_i.

In this case again

$$\mathbf{p}^n = \mathbf{T}^{-1}\mathbf{M}^n\mathbf{T} \to \mathbf{1P}',$$

so that

$$\lim_{n\to\infty} \mathbf{M}^n = \mathbf{T1P}'\mathbf{T}^{-1}, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (7)$$

where, just as before, the elements of $\mathbf{T1P}'\mathbf{T}^{-1}$ are fixed finite values. Now suppose we raise a simple classical matrix $\mathbf{C}_i^{(v)}$ to the nth power; we obtain for $n \geqslant v-1$

$$\{\mathbf{C}_i^{(v)}\}^n = \begin{bmatrix} \mu_i^n & n\mu_i^{n-1} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \binom{n}{v-1}\mu_i^{n-v+1} \\ \cdot & \mu_i^n & n\mu_i^{n-1} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \mu_i^n & n\mu_i^{n-1} \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \mu_i^n \end{bmatrix},$$

where the elements in the main diagonal, and the diagonal above it are respectively

$$\mu_i^n, \binom{n}{1}\mu_i^{n-1}, \binom{n}{2}\mu_i^{n-2}, \ldots, \binom{n}{v-1}\mu_i^{n-v+1},$$

and the remaining elements are zero.

If the unit root is $\mu_1 = 1$ say, then it is clear again that if (7) is to hold, $|\mu_r| < 1$ $(r = 2, \ldots, t)$; further, even if the multiplicity of μ_1 is $m_1 > 1$, it is not possible that $\mathbf{M}_1$ consist of simple classical matrices of order greater than 1, so that at the most the form of $\mathbf{M}_1$ is

$$\mathbf{M}_1 = \begin{bmatrix} [1] & & & & \\ & [1] & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & [1] \end{bmatrix} = \mathbf{I},$$

of order $m_1 > 1$. In this case $\mathbf{M}_1^n = \mathbf{I}$, and we have from (7) that

$$\lim_{n\to\infty} \mathbf{M}^n = \begin{bmatrix} \mathbf{I} & & & & \\ & \mathbf{O} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \mathbf{O} \end{bmatrix} = \mathbf{T} \begin{bmatrix} P_1 & \dots & P_s \\ P_1 & \dots & P_s \\ \cdot & \cdot\;\cdot\;\cdot & \cdot \\ P_1 & \dots & P_s \end{bmatrix} \mathbf{T}^{-1}.$$

Now the rank of $\mathbf{I}$ is m_1; but since $\mathbf{T}$ is a non-singular matrix, the rank of the right-hand matrix will be the same as that of $\mathbf{1P}'$, that is 1. It follows that $m_1 = 1$, so that the root $\mu_1 = 1$ is simple.

III. An Elementary Proof of the Theorem

The following proof, which is equivalent to the preceding one, is based on a simple way of obtaining the expression for the higher transition probability $p_{ij}^{(n)}$. To do this, we shall consider for some suitable value of μ, the expansion in powers of $\mathbf{p}$ of the matrix

$$(\mu\mathbf{I} - \mathbf{p})^{-1} = \mu^{-1}(\mathbf{I} + \mu^{-1}\mathbf{p} + \mu^{-2}\mathbf{p}^2 + \dots);$$

this is convergent for $|\mu| > 1$. It is clear that the (i, j)th element of this matrix is

$$\{(\mu\mathbf{I} - \mathbf{p})^{-1}\}_{ij} = \mu^{-1}(\delta_{ij} + \mu^{-1}p_{ij} + \mu^{-2}p_{ij}^{(2)} + \dots),$$

so that, if $\{(\mu\mathbf{I} - \mathbf{p})^{-1}\}_{ij}$ can be expanded in some other way in descending powers of μ, it is possible to obtain an expression for $p_{ij}^{(n)}$ by equating the coefficients of $\mu^{-(n+1)}$ in the two expansions.

Consider the matrix

$$(\mu\mathbf{I} - \mathbf{p}') = \begin{bmatrix} \mu - p_{11} & -p_{21} & \dots & -p_{s1} \\ -p_{12} & \mu - p_{22} & \dots & -p_{s2} \\ \cdot & \cdot & \cdot\;\cdot\;\cdot & \cdot \\ -p_{1s} & \cdot & \dots & \mu - p_{ss} \end{bmatrix},$$

the cofactors $C_{ij}(\mu)$ of each of the elements $(\mu\delta_{ji} - p_{ji})$ of the matrix are of degree no greater than $s-1$ in μ, whereas the degree of $D(\mu) = |\mu\mathbf{I} - \mathbf{p}|$ is always one greater than that of any C_{ij}, that is, no greater than s. It follows that

$$\{(\mu\mathbf{I} - \mathbf{p})^{-1}\}_{ij} = C_{ij}(\mu)\,|\mu\mathbf{I} - \mathbf{p}|^{-1} = C_{ij}(\mu)/\{(\mu - \mu_1)^{m_1} \dots (\mu - \mu_t)^{m_t}\}, \quad \dots\dots \quad (8)$$

where $\mu_1, \ldots, \mu_t$ are the roots of the characteristic equation $D(\mu)=0$, with multiplicity $m_1, \ldots, m_t$ respectively, where $\sum_{r=1}^{t} m_r = s$.

From (8), by the method of partial fractions, we obtain that

$$\{(\mu\mathbf{I}-\mathbf{p})^{-1}\}_{ij} = \sum_{r=1}^{t} B_r^{(ij)}(\mu)/(\mu-\mu_r)^{m_r}, \quad \ldots\ldots\ldots\ldots \quad (9)$$

where the $B_r^{(ij)}(\mu)$ are polynomials in μ of degree no greater than m_r-1, of the form

$$B_r^{(ij)}(\mu) = \sum_{u=0}^{m_r-1} b_{ru}^{(ij)} \mu^u ;$$

obviously, if $m_r=1$, $B_r^{(ij)}(\mu)=b_{r0}^{(ij)}$. Since $|\mu|>1$, and we know that for any stochastic matrix $\mathbf{p}$, $|\mu_r| \leqslant 1$ for all r, we may expand the expression (9) as follows :

$$\{(\mu\mathbf{I}-\mathbf{p})^{-1}\}_{ij} =$$
$$\sum_{r=1} \mu^{-m_r} \left\{ \sum_{u=0}^{m_r-1} b_{ru}^{(ij)} \mu^u \right\} \left\{ 1+m_r\mu^{-1}\mu_r + \ldots + \binom{m_r+n-1}{n} \mu^{-n}\mu_r^n + \ldots \right\}. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (10)$$

Let m be the largest of $m_1, \ldots, m_t$; then, for all values of $n \geqslant m-1$, the coefficient of $\mu^{-(n+1)}$ in the element $\{(\mu\mathbf{I}-\mathbf{p})^{-1}\}_{ij}$ of (10) is equal to

$$\sum_{r=1}^{t} \left\{ b_{r,m_r-1}^{(ij)} \mu_r^n \binom{m_r+n-1}{n} + \ldots + b_{r0}^{(ij)} \mu_r^{n-m_r+1} \binom{n}{n-m_r+1} \right\}, \quad \ldots \quad (11)$$

or for simplicity,

$$\sum_{r=1}^{t} \beta_r^{(ij)}(n)\mu_r^n, \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (12)$$

where $\beta_r^{(ij)}(n)$ are polynomials in n of degree no greater than m_r-1.

It is clear that this is the transition probability $p_{ij}^{(n)}$ for values of n greater than or equal to $m-1$; this is all that is required, since we are to consider the limit of $p_{ij}^{(n)}$ as $n\to\infty$. Writing, for sufficiently large n, the result

$$p_{ij}^{(n)} = \sum_{r=1}^{t} \beta_r^{(ij)}(n)\mu_r^n, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (13)$$

we see that, if the chain is regular, so that the condition (3) holds, then if the unit root is $\mu_1=1$ say, all the remaining roots whether simple or multiple must have moduli $|\mu_r|<1$ $(r=2, \ldots, t)$; further μ_1 cannot be multiple, for if it were, then

$$\lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty} \beta_1^{(ij)}(n) = \infty,$$

which contradicts (3). It follows that the root $\mu_1=1$ is simple, the remaining roots having moduli $|\mu_r|<1$ $(r=2, \ldots, t)$.

The sufficient condition for regularity presents no difficulty; it is that, if the unit root of the stochastic matrix $\mathbf{p}$ is simple, and the remaining roots have moduli less than 1, then $\lim_{n\to\infty} p_{ij}^{(n)}=P_j$ as in (**3**). From the result (**13**), it is clear that if $\mu_1=1$ is simple, and $|\mu_r|<1$ $(r=2, \ldots, t)$, then

$$\lim_{n\to\infty} p_{ij}^{(n)}=b_{10}^{(ij)},$$

where $b_{10}^{(ij)}$ is some constant. In matrix form this can be written

$$\lim_{n\to\infty} \mathbf{p}^n=\mathbf{b}=\begin{bmatrix} \mathbf{b}^{(1)} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{b}^{(s)} \end{bmatrix}, \quad \ldots\ldots\ldots\ldots\ldots \quad (\mathbf{14})$$

where $\mathbf{b}$ is the matrix of elements $b_{10}^{(ij)}$, and $\mathbf{b}^{(i)}$ its row vectors with elements such that $\sum_{j=1}^{s} b_{10}^{(ij)}=1$. We show that these are identically equal to a single row vector $\mathbf{P}'$ with elements P_j such that $\sum_{j=1}^{s} P_j=1$. For, from (**14**) we have that

$$\mathbf{b}=\lim_{n\to\infty} \mathbf{p}^{n+1}=\lim_{n\to\infty} \mathbf{p}^n\mathbf{p}=\mathbf{bp},$$

so that for each $\mathbf{b}^{(i)}$ $(i=1, \ldots, s)$,

$$\mathbf{b}^{(i)}\mathbf{p}=\mathbf{b}^{(i)}.$$

But, since the latent root $\mu_1=1$ of $\mathbf{p}$ is simple, then there exists only one row vector $\mathbf{P}'$ of probabilities P_j such that $\sum_{j=1}^{s} P_j=1$, which is a solution of the equation (**5**) (Frazer, Duncan, and Collar 1947, Section 3,5). It follows that

$$\mathbf{b}^{(i)}=\mathbf{P}' \quad (i=1, \ldots, s),$$

so that $\lim_{n\to\infty} p_{ij}^{(n)}=P_j$ as required.

IV. Acknowledgment

The author is indebted to Dr. E. J. Hannan for his helpful criticism and suggestions.

V. References

Bartlett, M. S. (1955).—" An Introduction to Stochastic Processes." (Cambridge Univ. Press.)

Feller, W. (1950).—" An Introduction to Probability Theory and Its Applications." (John Wiley: New York.)

Frazer, R. A., Duncan, W. J., and Collar, A. R. (1947).—" Elementary Matrices." (Cambridge Univ. Press.)

Fréchet, M. (1938).—" Recherches Théoriques Modernes." Vol. 2. (Gauthier-Villars: Paris.)

Turnbull, H. W., and Aitken, A. C. (1932).—" An Introduction to the Theory of Canonical Matrices." (Blackie: London.)