MULTIVARIATE SEMI-MARKOV MATRICES*

by

Marcel F. Neuts*

and

Peter Purdue**

Department of Statistics

Division of Mathematical Sciences

A revised version of Mimeograph Series No. 200

August 1970

*This research was partly supported by the Office of Naval Research Contract NONR 1100(26) at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government. The author was on sabbatical leave at Cornell University during the academic year 1968-1969.

**This author is a David Ross Fellow at Purdue University.

Multivariate Semi-Markov Matrices

by

Marcel F. Neuts* and Peter Purdue**

Purdue University and Cornell University

Abstract

Finite matrices with entries $p_{ij}\ F_{ij}(x_1,\ldots,x_k)$, where $\{p_{ij}\}$ is Stochastic and $F_{ij}(\cdot)$ is a k-variate probability distribution are discussed. It is shown that the matrix of k-fold Laplace-Stieltjes transforms of the $p_{ij}\ F_{ij}(x_1,\ldots,x_k)$ has a Perron-Frobenius eigenvalue which is a convex function in k variables in a suitably defined region. The values of the partial derivatives near the origin of this maximal eigenvalue are exhibited. They are quantities of interest in a variety of applications in Probability theory.

* The research of this author was partly supported by the Office of Naval Research Contract NONR 1100(26) at Purdue University. He was on sabbatical leave at Cornell University during the academic year 1968-1969.

** The research of this author was partly supported by a David Ross Grant #PRF 6012 at Purdue University.

1. Introduction

A natural combination of the theories of stochastic matrices and of distribution functions, which arises in a large number of problems of analytic Probability theory, is the theory of semi-Markov matrices.

In this paper we wish to consider properties of semi-Markov matrices involving multivariate distributions.

Definition: k-variate semi-Markov matrix.

Let $Q(\underline{x})$ be an $m \times m$ matrix, whose entries are real valued functions defined on R^k such that each entry $Q_{ij}(\underline{x})$ may be written as:

$$Q_{ij}(\underline{x}) = p_{ij} F_{ij}(x_1,\ldots,x_k) , \tag{1}$$

where $F_{ij}(x_1,\ldots,x_k)$ is a k-variate probability distribution and where $p_{ij} \geq 0$, $\sum_{j=1}^{m} p_{ij} = 1$, $i = 1,\ldots,m$, then $Q(\underline{x})$ is a k-variate semi-Markov matrix.

We note that if $p_{ij} = 0$, the probability distribution $F_{ij}(\cdot)$ may be arbitrarily chosen.

Definition: Irreducible semi-Markov matrix.

The semi-Markov matrix $Q(\underline{x})$ is called irreducible if and only if the stochastic matrix $P = \{p_{ij}\}$ is irreducible.

Definition: Nondegenerate k-variate semi-Markov matrix.

The semi-Markov matrix $Q(\underline{x})$ is nondegenerate k-variate if and only if for every $\nu = 1,\ldots,k$ there exists a pair of indices (i,j) such that $p_{ij} > 0$ and the corresponding distribution $F_{ij}(x_1,\ldots,x_k)$

has a marginal distribution $F_{ij}(+\infty,\ldots, x_\nu,\ldots,+\infty)$ which is not degenerate at zero.

The nondegeneracy condition eliminates the case where one or more of the k-variables $x_1,\ldots,x_k$ are actually redundant.

Henceforth we assume that $Q(\underline{x})$ is an irreducible and nondegenerate k-variate semi-Markov matrix.

We now consider the k-dimensional Lebesgue-Stieltjes integrals:

$$q_{ij}(\xi_1,\ldots,\xi_k) = q_{ij}(\underline{\xi}) = \int_{R^k} \exp[-\sum_{\nu=1}^{k} \xi_\nu x_\nu] \, d_{x_1,\ldots,x_k} Q_{ij}(x_1,\ldots,x_k) \quad , \tag{2}$$

which we refer to as the Laplace-Stieltjes transforms of the entries $Q_{ij}(x_1,\ldots,x_k)$ of $Q(\underline{x})$.

The functions $q_{ij}(\xi_1,\ldots,\xi_k)$ are obviously defined for $\mathrm{Re}\ \xi_1 = 0,\ldots,$ $\mathrm{Re}\ \xi_k = 0$, but they may not be defined anywhere else. We are mainly interested in the cases where the domain of definition of the $q_{ij}(\underline{\xi})$ is larger, as is the case in most applications.

We distinguish the unilateral and the bilateral cases.

In the unilateral case, we assume that all $F_{ij}(x_1,\ldots,x_k)$ corresponding to indices i,j such that $p_{ij} > 0$, concentrate all their mass on the positive orthant $x_1 \geq 0,\ldots,x_k \geq 0$. In this case all integrals in (2) exist for all $\underline{\xi}$ with $\mathrm{Re}\ \xi_1 \geq 0,\ldots,\mathrm{Re}\ \xi_k \geq 0$. Moreover all the functions $q_{ij}(\xi_1,\ldots,\xi_k)$ are jointly analytic in $\mathrm{Re}\ \xi_1 > 0,\ldots,$ $\mathrm{Re}\ \xi_k > 0$ and any function obtained by setting some but not all of its variables equal to zero is analytic inside the corresponding part of the boundary of the set $\mathrm{Re}\ \xi_1 > 0, \ldots, \mathrm{Re}\ \xi_k > 0$. The latter statement is obvious if we realize that setting one or more, but not all of the ξ-variables equal to zero, corresponds to taking the Laplace-Stieltjes transforms of suitable "marginal" distributions of $Q_{ij}(x_1, \ldots, x_k)$.

The bilateral case encompasses all distributions not in the unilateral case.

In our discussion of the bilateral case we shall assume that there exist $2k$ real numbers ξ_i' and ξ_i'', $i = 1, \ldots, k$ such that:

(3) $$-\infty \leq \xi_i'' < 0 < \xi_i' \leq +\infty , \quad i = 1, \ldots, k$$

and such that in the "box":

(4) $$\xi_i'' \leq \mathrm{Re}\ \xi_i \leq \xi_i' , \quad i = 1, \ldots, k ,$$

all functions $q_{ij}(\xi_1, \ldots, \xi_k)$ are analytic in $\xi_1, \ldots, \xi_k$.

In order to discuss both cases simultaneously, we shall refer to the domain D in the unilateral case as the open positive orthant $\xi_1 > 0, \ldots, \xi_k > 0$ and in the bilateral case as the box $\xi_1'' \leq \xi_1 \leq \xi_1', \ldots, \xi_k'' \leq \xi_k \leq \xi_k'$.

2. The Perron-Frobenius Eigenvalue of $q(\underline{\xi})$.

The matrix $q(\underline{\xi})$ with entries $q_{ij}(\xi_1, \ldots, \xi_k)$ is an irreducible, nonnegative matrix for every real point $\underline{\xi}$ in the domain D or on its boundary. It follows from the classical theory of nonnegative matrices, [1,4], that

$q(\underline{\xi})$ has an eigenvalue of maximum modulus, which is real, positive and of geometric and algebraic multiplicity one. Denoting this, the Perron-Frobenius eigenvalue, by $\rho(\underline{\xi}) = \rho(\xi_1,\ldots,\xi_k)$, we set out to discuss the properties of $\rho(\underline{\xi})$ as a function of $\underline{\xi}$ over the domain D. In the simpler case where $k = 1$, this was done by H. D. Miller [3].

We shall assume that the reader is familiar with the basic properties of nonnegative matrices as discussed in the references listed above.

<u>Lemma 1</u>

All functions $q_{ij}(\underline{\xi})$, $i,j = 1,\ldots,m$ are convex functions over the domain D and its boundary, i.e. for $\underline{\xi}$ and $\underline{\eta}$ in the closure $\bar{D}$, we have:

(5) $$q_{ij}[\alpha\underline{\xi} + (1-\alpha)\underline{\eta}] \leq \alpha q_{ij}(\underline{\xi}) + (1-\alpha)\, q_{ij}(\underline{\eta})$$

for all $0 \leq \alpha \leq 1$, and all $i,j = 1,\ldots,m$.

Moreover if $\underline{\xi} \neq \underline{\eta}$ and $0 < \alpha < 1$, strict inequality must hold in (5) for at least one pair (i,j).

<u>Proof</u>:

Since for all real k-tuples $(x_1,\ldots,x_k)$, the function $\exp[-\sum_{\nu=1}^{k} \xi_\nu x_\nu]$ is strictly convex over the domain $\bar{D}$, the inequality (5) follows immediately from the definition of $q_{ij}(\underline{\xi})$.

To prove the next statement we must clearly consider only those pairs (i,j) for which $p_{ij} > 0$. The corresponding Laplace-Stieltjes transform $q_{ij}(\xi_1,\ldots,\xi_k)$ is strictly convex with respect to all the variables which explicitly occur in it. The variables ξ_r which do not explicitly occur in $q_{ij}(\xi_1,\ldots,\xi_k)$ correspond to variables x_r in $F_{ij}(x_1,\ldots,x_k)$ with respect to which the marginal distributions are degenerate at zero.

The nondegeneracy assumption may be restated as saying that every variable ξ_ν, $\nu = 1,\ldots,k$ must occur explicitly in at least one of the functions $q_{ij}(\xi_1,\ldots,\xi_k)$.

Let now $\underline{\xi} \neq \underline{\eta}$. In particular $\xi_\nu \neq \eta_\nu$. Let (i,j) be a pair such that $q_{ij}(\xi_1,\ldots,\xi_k)$ contains ξ_ν explicitly, then for $0 < \alpha < 1$

$$q_{ij}[(1-\alpha)\underline{\eta} + \alpha\underline{\xi}] < \alpha q_{ij}(\underline{\xi}) + (1-\alpha)\, q_{ij}(\underline{\eta}) \ ,$$

since $q_{ij}(\cdot)$ is jointly strictly convex in all variables upon which it explicitly depends.

Superconvex Matrices.

Let f be a positive function defined on the convex set $\Gamma \in K$. Then f is superconvex if log f is a convex function on Γ. Clearly, f is superconvex if and only if for each $\underline{\xi}$, $\underline{\eta} \in \Gamma$,

$$f(\alpha\underline{\xi} + \beta\underline{\eta}) \leq [f(\underline{\xi})]^{\alpha}[f(\underline{\eta})]^{\beta}; \quad \alpha + \beta = 1$$

$$\alpha \geq 0, \ \beta \geq 0.$$

Definition:

A matrix $A(\underline{\xi}) = [A_{ij}(\underline{\xi})]$ is superconvex if for each (i,j), $A_{ij}(\underline{\xi})$ is superconvex on Γ.

The proofs of the following lemmas can be found in reference (2) or (3).

Lemma 2:

If f is superconvex on Γ, then it is convex there.

Lemma 3:

Let $\gamma(\underline{\xi})$ be any non constant positive linear function on Γ. Then $\gamma(\underline{\xi})$ is not superconvex.

Following Kingman (2) we let C denote the class of all superconvex functions along with the function which is identically zero on Γ.

Lemma 4:

$\mathbf{C}$ is closed under addition, multiplication and raising to any positive power. If for each n, $f_n \varepsilon \mathbf{C}$, so does $\limsup_{n \to \infty} f_n$.

Lemma 5:

Let $A(\underline{\xi})$ be a superconvex matrix on Γ and let $\rho(\underline{\xi})$ denote its largest eigenvalue. Then $\rho(\underline{\xi}) \varepsilon \mathbf{C}$.

Lemma 6:

Let $A(\underline{\xi})$ be a superconvex matrix on Γ and suppose $\rho(\underline{\xi})$ is not a constant function. Then $\rho(\underline{\xi})$ is strictly convex on Γ.

Proof:

By lemma's 2 and 5, $\rho(\underline{\xi})$ is convex on Γ. Suppose now that $\rho(\underline{\xi})$ is in fact linear. Then by lemma 3, since ρ is not constant, $\rho(\underline{\xi})$ is not superconvex. This contradiction implies that $\rho(\underline{\xi})$ is strictly convex on Γ.

Theorem 1:

Let $\underline{\xi} = \underline{\sigma} + i\ \underline{\tau}$ where $\underline{\xi}\ \varepsilon$ D.

(a) The Perron Frobenius eigenvalue, $\rho(\underline{\xi})$ is analytic at $\underline{\xi} = \underline{\sigma}$ in the domain D.

(b) $\rho(\underline{\sigma})$ is a strictly convex function of $\underline{\sigma}$ in $\overline{D}$, suitably continues on the boundary.

Proof:

(a) As in the univariate case, Miller [5], for each real $\underline{\sigma}$, $\rho(\underline{\sigma})$ is a simple root of the determinantal equation $|zI-g(\underline{\sigma})| = 0$. Since $|zI-g(\underline{\sigma})|$ is an analytic function of the k+1 complex variables, $z, \sigma_1,\ldots,\sigma_k$, the result follows from the implicit functions theorem for analytic functions.

(b) We need only show that $q_{ij}(\underline{\sigma})$ is a superconvex function for each (i,j). This follows at once since

$$\int_D e^{(\alpha\underline{\sigma} + \beta\ \ \underline{\sigma}')\cdot\underline{X}}\, d\, Q(\underline{X})$$

$$\leq [\int_D e^{\underline{\sigma}\cdot\underline{X}}\, d\, Q(\underline{X})]^{\alpha}\ [\int_D e^{-\underline{\sigma}'\cdot\underline{X}}\, d\, Q(\underline{X})]^{\beta}$$

for $\underline{\xi} = \underline{\sigma} + i\ \underline{\tau}$, $\underline{\xi}' = \underline{\sigma}' + i\ \underline{\tau}'$, $\underline{\xi}$, $\underline{\xi}'\ \varepsilon$ D, and $\underline{\sigma}\cdot\underline{X} = \sigma_1 X_1 + \ldots + \sigma_k X_k$. This is just Holdens inequality for a Banach space with a finite measure. Consequently $g(\underline{\sigma})$ is a superconvex matrix and so $\rho(\underline{\sigma})$ is convex. By lemma 1 $\rho(\underline{\sigma})$ is not constant and so by lemma 6 $\rho(\underline{\sigma})$ is strictly convex on D.

By suitably convex an the boundary $\overline{D}$ we mean that if $\underline{\xi}^* = \underline{\sigma}^* + i\ \underline{\tau}^*\ \varepsilon\ \overline{D}$ and if $\underline{\xi}_n \to \xi^*$ where $\underline{\xi}_n\ \varepsilon$ D then $\rho(\underline{\sigma}_n) \to \rho(\underline{\sigma}^*)$. Hence we have $\rho(\underline{\sigma})$ is strictly convex on $\overline{D}$.

Q.E.D.

The entries of $q(\underline{\xi})$ are all suitably continuous on the boundary and hence $\rho(\underline{\xi})$ is suitably continuous on the boundary, since convergence of a sequence of positive matrices entails convergence of their Perron-Frobenius eigenvalues to that of the limit matrix.

The theorem 1 implies in particular that $\rho(\underline{\xi})$ is a continuously differentiable function of $\underline{\xi}$ in D. In the unilateral case one may easily verify that $\rho(\underline{\xi})$ is also suitably differentiable at all boundary points of the positive orthant D, with the possible exception of the origin.

In many applications, see Neuts [6], the quantities

(11) $$M_j = \left[\frac{\partial}{\partial \xi_j} \rho(\xi_1,\ldots,\xi_k)\right]_{\underline{\xi}=\underline{0}}$$

play a fundamental role. In the unilateral case, the derivatives at $\underline{0}$ are to be understood in the same "suitable" sense as in theorem 1.

We denote by $\alpha_i^{(\nu)}$, the mean with respect to the variable x_ν of the probability distribution $H_i(x_1,\ldots,x_k)$ defined by:

(12) $$H_i(x_1,\ldots,x_k) = \sum_{j=1}^{m} p_{ij} F_{ij}(x_1,\ldots,x_k), \quad i = 1,\ldots,m$$

i.e. $\alpha_i^{(\nu)}$ is given by:

(13) $$\alpha_i^{(\nu)} = \int_{R^k} x_\nu \, d_{x_1,\ldots,x_k} H_i(x_1,\ldots,x_k) \quad ,$$

provided the integral (13) converges absolutely. In this case $\alpha_i^{(\nu)}$ is also given by:

(14) $$\alpha_i^{(\nu)} = -[\frac{\partial}{\partial \xi_\nu} \sum_{j=1}^{m} q_{ij} (\xi_1,\ldots,\xi_k)]_{\underline{\xi}=\underline{0}}$$

where the derivative is in the suitable sense in the unilateral case.

Furthermore, let $\pi_1,\ldots,\pi_m$ be the stationary probabilities associated with the matrix P, i.e. the row-vector $\underline{\pi} = (\pi_1,\ldots,\pi_m)$ is the unique solution to the equations:

(15) $$\underline{\pi} = \underline{\pi}P, \quad \underline{\pi}\cdot\underline{e} = 1 ,$$

where $\underline{e}$ is the columnvector with all its components equal to one.

<u>Theorem 2</u>

The quantities M_j are given by:

(16) $$M_j = -\sum_{i=1}^{m} \pi_i \alpha_i^{(j)} .$$

In the unilateral case, this is provided the means $\alpha_i^{(j)}$, $i = 1,\ldots,m$ exist. In the bilateral case, our earlier assumptions encompass the existence of these means.

<u>Proof</u>

Let $\underline{x}(\underline{\xi})$ and $\underline{y}(\underline{\xi})$ be right and left eigenvectors of $q(\underline{\xi})$ corresponding to $\rho(\underline{\xi})$, normalized such that $\underline{y}(\underline{\xi})\cdot\underline{x}(\underline{\xi}) = 1$, and $\underline{y}(\underline{\xi})\cdot e = 1$. It is known that such a normalization is possible and uniquely determines $\underline{x}$ and $\underline{y}$ for every $\underline{\xi}$. Moreover as $\underline{\xi}$ tends (suitably) to $\underline{0}$, we have that $\underline{y}(\underline{\xi}) \to \underline{\pi}$ and $\underline{x}(\underline{\xi}) \to \underline{e}$, componentwise. The components of $\underline{x}(\underline{\xi})$ and $\underline{y}(\underline{\xi})$ are (suitably) continuously differentiable functions of $\underline{\xi}$ in $\bar{D}$.

We have that:

(17) $$\sum_{j=1}^{m} q_{\nu j}(\xi_1,\dots,\xi_k)\, x_j(\xi_1,\dots,\xi_k) = \rho(\xi_1,\dots,\xi_k)\, x_\nu(\xi_1,\dots,\xi_k),$$

for $\nu = 1,\dots,m$ and all $\underline{\xi}$ in $\bar{D}$.

Differentiation with respect to ξ_i yields:

(18) $$\rho(\xi_1,\dots,\xi_k)\frac{\partial}{\partial\xi_i} x_\nu(\xi_1,\dots,\xi_k) + x_\nu(\xi_1,\dots,\xi_k)\frac{\partial}{\partial\xi_i}\rho(\xi_1,\dots,\xi_k)$$
$$= \sum_{j=1}^{m} x_j(\xi_1,\dots,\xi_k)\frac{\partial}{\partial\xi_i} q_{\nu j}(\xi_1,\dots,\xi_k) + \sum_{j=1}^{m} q_{\nu j}(\xi_1,\dots,\xi_k)\frac{\partial}{\partial\xi_i} x_j(\xi_1,\dots,\xi_k).$$

Upon letting $\underline{\xi} \to \underline{0}$ (suitably) and noting that $\rho(\underline{0}) = 1$, we obtain:

(19) $$\left[\frac{\partial}{\partial\xi_i} x_\nu(\underline{\xi})\right]_{\underline{\xi}=\underline{0}} + M_i = -\alpha_\nu^{(i)} + \sum_{j=1}^{m} p_{\nu j}\left[\frac{\partial}{\partial\xi_i} x_j(\underline{\xi})\right]_{\underline{\xi}=\underline{0}}$$

for $\nu = 1,\dots,m$.

Multiplying by π_ν in (19), summing on ν and applying (15), it follows that:

(20) $$M_i = -\sum_{\nu=1}^{m} \pi_\nu \alpha_\nu^{(i)} .$$

<u>Remark</u>

Formally, the quantities M_i appear in the same manner as the first moment does from the Laplace-Stieltjes transform of a probability distribution. A natural question to ask is whether $\rho(\xi_1,\ldots,\xi_k)$ is itself the transform of a probability distribution. The answer is negative in general. Consider the following example of a 2×2 univariate semi-Markov matrix

$$p_{11} = p_{22} = 0, \qquad p_{12} = p_{21} = 1 \quad .$$

It is easy to see that:

$$\rho(\xi) \;=\; [f_1(\xi)\cdot f_2(\xi)]^{1/2} \quad ,$$

where $f_1(\xi)$ and $f_2(\xi)$ are the Laplace-Stieltjes transforms of the probability distributions $F_{12}(\cdot)$ and $F_{21}(\cdot)$. It is well-known that $f_1(\xi)$ and $f_2(\xi)$ can be chosen so that their product is not the square of a Laplace-Stieltjes transform of a probability distribution, e.g.:

$$f_1(\xi) = e^{-\xi} \quad , \qquad f_2(\xi) = \frac{1}{2} + \frac{1}{2}e^{-\xi} \quad .$$

Bibliography

[1] Karlin, Samuel (1966). A First Course in Stochastic Processes, Academic Press, New York & London.

[2] Kingman, J.F.C. A Convexity Property of Positive Matrics, Quart. J. Math., Oxford (2) 12, 283-4.

[3] Klinger, A. & Mangasariau, O. L. Logarithmic Convexity & Geometric Programming, J. Math. Anal. & Appl. 24, 388-408.

[4] Marcus, Marvin and Minc, Henryk (1964). A Survey of Matrix Theory and Matrix Inequalities, Allyn and Bacon, Boston.

[5] Miller, H. D. (1961). A Convexity Property in the Theory of Random Variables Defined on a Finite Markov Chain. Ann. Math. Stat., 32, 1260-1270.

[6] Neuts, Marcel F. (1966). The Single Server Queue with Poisson Input and Semi-Markov Service Times, J. Appl. Prob., 3, 202-230.