STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

AFDELING MATHEMATISCHE STATISTIEK

Report S 304 (VP 18)

On Markov chains, the transition function of which is a finite sum of products of functions of one variable

by

J.Th. Runnenburg
and
F.W. Steutel

Preliminary report

August 1962

1. Introduction

In most literature on waiting time and renewal theory it is assumed that the intervals $\underline{x}_1, \underline{x}_2, \ldots$ [1] between successive arrivals (renewals) are independent random variables having the same distribution function G(x). More generally one may consider a sequence of Markov-dependent random variables $\underline{x}_1^*, \underline{x}_2^*, \ldots$, having G(x) as invariant distribution. Some information about this matter is contained in Runnenburg [1960] and [1961] . In Runnenburg [1960] as an example the following type of stationary Markov chain is considered: the chain starts in some initial position $\underline{x}_o$, while the transition function

(1.1) $$A(y|x) \overset{\text{def}}{=} P\{\underline{x}_{n+1} \leqslant y | \underline{x}_n = x\} \qquad (n \geqslant 0)$$

has the form

(1.2) $$A(y\|x) = \sum_{j=1}^{r} A_j(x) B_j(y),$$

where r is finite.

Markov chains of this type have the advantage of being more general than finite Markov chains without involving greater computational difficulties.

In the sections 2 and 3 of this paper some general aspects of these Markov chains are studied. In section 4 for a given invariant distribution function the effect of variation of the transition function on the correlation coefficient is considered for r=2.

----

1) Underlined symbols will denote random variables.

2. Definitions and elementary properties

Concerning the functions $A_j(x)$ and $B_j(y)$ we make the following assumptions:

(a) the $A_j(x)$ are complex-valued functions defined on a Borel-measurable set B on the real axis. They have bounded variation on B, i.e. for each j $(1 \leq j \leq r)$ we have: if $x_1, x_2, \ldots, x_{N+1} \in B$, $x_1 \leq x_2 \leq \ldots \leq x_{N+1}$, then $\sum_{n=1}^{N} \left| A_j(x_{n+1}) - A_j(x_n) \right|$ is bounded for all N.

(b) the $B_j(y)$ are complex-valued functions defined for all real y. They have bounded variation on $(-\infty, \infty)$ and are continuous from the right.

(c) $A(y|x)$ is for all $x \in B$ a probability measure [1] on a Borel field on B containing all sets of the form $(-\infty, y] \cap B$. This may be interpreted as: for all $x \in B$ $A(y|x)$ is a distribution function with $\int_B dA(y|x) \leq 1$.

(d) the representation 1.2 of $A(y|x)$ is minimal, i.e. $A(y|x)$ cannot be represented as a sum of less than r terms.

Lemma 2.1: the representation 1.2 is minimal if and only if both the $A_j(x)$ and the $B_j(y)$ are linearly independent.

Proof: clearly if either the $A_j(x)$ or the $B_j(y)$ are linearly dependent $A(y|x)$ can be written as a sum of less than r terms.

Conversely, suppose that $A(y|x)$ can be written as a sum of r-1 terms, then

$$(2.1) \qquad \sum_{j=1}^{r} A_j(x)\, B_j(y) = \sum_{j=1}^{r-1} A_j^*(x)\, B_j^*(y)$$

---

1) It follows from (a) that the usual condition of measurability of $A(y|x)$ for fixed y with respect to this measure is satisfied. For a general definition see Doob [1953] page 190.

for all $x \in B$ and all y. Nor for every r-tuple $x_1, x_2, \ldots, x_r \in B$ there exist complex numbers $c_1, c_2, \ldots, c_r$ not all zero, such that $\sum_{k=1}^{r} c_k A_j^*(x_k) = 0$ $(j=1,2,\ldots,r-1)$. It follows from (2.1) that

$$\sum_{j=1}^{r} \sum_{k=1}^{r} c_k A_j(x_k) B_j(y) = 0 \tag{2.2}$$

for all y, which means that either the $B_j(y)$ are linearly dependent or that

$$\sum_{k=1}^{r} c_k A_j(x_k) = 0 \qquad (j=1,2,\ldots,r). \tag{2.3}$$

The equalities (2.3) imply that $\det\left[A_j(x_k)\right] = 0$. If the $B_j(y)$ are linearly independent then for every r-tuple $x_1, x_2, \ldots, x_r \in B$ we must have $\det\left[A_j(x_k)\right] = 0$, from which it immediately follows that the $A_j(x)$ are linearly dependent.

Lemma 2.2: it is no restriction to assume that both the $A_j(x)$ and the $B_j(y)$ are real-valued. We may even assume without loss of generality, that the $B_j(y)$ are distribution functions.

Proof: as the $A_j(x)$ are linearly independent there exist $x_1, x_2, \ldots, x_r \in B$ such that $\det\left[A_j(x_k)\right] \neq 0$. If we define the linear transformation T by $T_{kj} = A_j(x_k)$ and write

$$\begin{cases} A_k^*(x) = \sum_{j=1}^{r} A_j(x) (T^{-1})_{jk} \\ B_k^*(y) = \sum_{j=1}^{r} T_{kj} B_j(y) \quad , \end{cases} \tag{2.4}$$

we have

$$A(y|x) = \sum_{k=1}^{r} A_k^*(x)\; B_k^*(y), \tag{2.5}$$

where

$$B_k^*(y) = A(y|x_k) \tag{2.6}$$

is a distribution function. It now follows immediately from the linear independence of the $B_k^{*}(y)$, that the $A_k^{*}(x)$ must be real.

In the sequel we assume the $A_j(x)$ and $B_j(y)$ to be real-valued.

Remark: by applying the transformation S with elements $S_{jk}=B_j(y_k)$ we may transform the $A_k(x)$ into $A_k^{**}(x) = A(y_k \| x)$, such that $0 \leq A_k^{**}(x) \leq 1$. The resulting $B_k^{**}(y)$ are not necessarily distribution functions.

By assumption (b)

$$(2.7) \qquad \begin{cases} a_j \overset{\text{def}}{=} \lim\limits_{y \to -\infty} B_j(y) \\ b_j \overset{\text{def}}{=} \lim\limits_{y \to \infty} B_j(y) \end{cases}$$

exist, while it follows from the linear independence of the $A_j(x)$ that

$$(2.8) \qquad a_j = 0 \qquad (j=1,2,\ldots,r).$$

Clearly

$$(2.9) \qquad \sum_{j=1}^{r} b_j A_j(x) = 1$$

for all $x \in B$.

We now introduce the matrix C with elements c_{jk} defined by

$$(2.10) \qquad c_{jk} = \int A_k(x)\, dB_j(y) \;^{1)} .$$

It is easily seen that every finite Markov matrix P can be represented by (2.10): if P is a Markov matrix with elements p_{jk} $(1 \leq j \leq r;\ 1 \leq k \leq r)$ we define

$$(2.11) \qquad \begin{cases} B = \{x_1, x_2, \ldots, x_r\} \\ A_j(x_k) = p_{kj} \\ B_j(y) = \delta(y - x_j) \end{cases} ,$$

----

1) When not otherwise indicated, integration is over the set B.

where

$$\epsilon(x) = \begin{cases} 0 & \text{for } x \leqslant 0 \\ 1 & \text{for } x \geqslant 0 \end{cases} . \tag{2.12}$$

We now find from (2.10) that

$$c_{jk} = p_{jk} . \tag{2.13}$$

It is noted however, that in the case of det P = 0 the functions $A_j(x)$ as defined by (2.11) are linearly dependent, i.e. the representation (1.2) is not minimal.

Returning to the general case we define

$$\begin{cases} A^{(1)}(y|x) = A(y|x) \\ A^{(n+1)}(y|x) = \int A^{(n)}(y|z)dA(z|x) & (n \geqslant 1), \end{cases} \tag{2.14}$$

from which it follows that

$$A^{(n+1)}(y|x) = \sum_{j=1}^{r} \sum_{k=1}^{r} A_j(x) c_{jk}^{(n)} B_k(y) \qquad (n \geqslant 0), \tag{2.15}$$

$c_{jk}^{(n)}$ denoting the elements of C^n (C^o=I, the unit matrix).

<u>Remark</u>: the linear transformation $A_k^*(x) = \sum_{j=1}^{r} A_j(x)(T^{-1})_{jk}$ and $B_k^*(y) = \sum_{j=1}^{r} T_{kj}B_j(y)$ results in

$$C^* = T\,C\,T^{-1}. \tag{2.16}$$

It now follows from Lemma 2.2 that C may be transformed in such a way that the c_{jk}^* satisfy

$$\sum_{k=1}^{r} c_{jk}^{*(n)} = 1 \qquad (j=1,2,\dots,r; n \geqslant 0). \tag{2.17}$$

3. General results

From (2.15) it is seen that the asymptotic behaviour of $A^{(n)}(y|x)$ for large n is governed by the asymptotic behaviour of C^n, which is determined by the eigenvalues of C. We therefore will now study the properties of these eigenvalues.

Theorem 3.1: the matrix C has an eigenvalue equal to 1 and no eigenvalues with modulus exceeding unity.

Proof: as it is no restriction to assume that $\sum_{k=1}^{r} c_{jk} = 1$ for all j $(1 \leqslant j \leqslant r)$, C admits of an eigenvalue 1.
Generally if λ is an eigenvalue and $\vartheta_1, \vartheta_2, \ldots, \vartheta_r$ a corresponding eigenvector we define an eigenfunction $\varphi(x)$ by

(3.1) $$\varphi(x) \overset{\text{def}}{=} \sum_{j=1}^{r} \vartheta_j A_j(x),$$

which is easily seen to satisfy

(3.2) $$\lambda \varphi(x) = \int \varphi(y)\, dA(y|x).$$

From (2.19) it follows that $|\lambda|\,|\varphi(x)| \leqslant \sup |\varphi(y)|$, whence by the boundedness of $\varphi(x)$ we have $|\lambda| \leqslant 1$.

Remark: Theorem 3.1 might suggest that for every matrix C there exists a non-singular linear transformation T such that $T\,C\,T^{-1}$ is a Markov matrix. The following example shows that this is not true: if we take $B = \{x_1, x_2, x_3, x_4\}$, for the matrix with elements $A_j(x_k)$ $(1 \leqslant j \leqslant 3;\ 1 \leqslant k \leqslant 4)$

$$\begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix},$$

and $B_1(y) = \frac{1}{2}\{\iota(y-x_1) + \iota(y-x_n)\}$, $B_2 = \frac{1}{2}\{\iota(y-x_1) + \iota(y-x_2)\}$, $B_3 = \frac{1}{2}\{\iota(y-x_2) + \iota(y-x_3)\}$, we find

$$C = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & 1 \end{pmatrix} ,$$

which has eigenvalues 1, $\frac{1+i}{2}$, while for a Markov matrix with 3 rows and 3 columns the non-real eigenvalues are restricted to the domain bounded by the triangle with vertices 1, $e^{\pm\frac{2\pi i}{3}}$. This example has been constructed by considering a Markov matrix of 4 rows and 4 columns having one eigenvalue zero and an eigenvalue outside the abovementioned triangle. For more information concerning this matter we refer to Dmitriev and Dynkin [1946].

As in the case of finite Markov matrices the eigenvalues of modulus 1 are of special interest. It can be proved that the number of independent eigenvectors corresponding to an eigenvalue of modulus 1 is equal to the multiplicity of that eigenvalue (for a proof in the case of a finite Markov matrix, which is easily generalized, we refer to Van Dantzig [1956] page 38.

We first consider the case of a multiple eigenvalue 1: Now there exists an eigenfunction $\varphi(x)$ not identically equal to a constant. As $\varphi(x)$ is bounded we may assume that

(3.3) $$\sup_{x \in B} |\varphi(x)| = 1 .$$

We now assume[1)] that to every eigenfunction $\varphi(x)$ corresponding to an eigenvalue of modulus 1 there exists a value $x_o \in B$ with $|\varphi(x_o)| = 1$. This is trivially true for instance if B is finite or if B is a bounded closed interval, on which the $A_j(x)$ are continuous.

Without loss of generality we may now assume that $\varphi(x_o) = 1$. We define

(3.4) $$B_o \overset{\text{def}}{=} \left\{ x \mid \varphi(x_o) = 1 \right\} .$$

For all $x_o \in B_o$ we now have

(3.5) $$\int \left\{ 1 - \varphi(y) \right\} dA(y \mid x_o) = 0 ,$$

----

1) See footnote on page 9.

from which it follows that for all $x_o \in B_o$

(3.6) $$\int_{B_o} dA(y|x_o) = 1,$$

i.e. B_o is an absorbing set.

Considering the eigenfunction $\psi(x) = 1-\varphi(x)$ we find in the same way (after norming $\psi(x)$ in such a way that $\sup|\psi(x)|=1$) an other absorbing set B_o', which is disjoint from B_o.

Next we consider the case $|\lambda|=1$, $\lambda\neq 1$:

As before we assume that a value $x_o \in B$ exists with $\varphi(x_o) = 1$.[1] For all $x_o \in B_o$ (as defined by (3.4)) we now have

(3.7) $$\int \left(1 - \frac{\varphi(y)}{\lambda}\right) dA(y|x_o) = 0,$$

from which it follows that

(3.8) $$B_1 \overset{\text{def}}{=} \left\{x \,|\, \varphi(x) = \lambda\right\}$$

satisfies

(3.9) $$\int_{B_1} dA(y|x_o) = 1$$

for all $x_o \in B_o$. Further evidently B_o and B_1 are disjoint. Defining generally

$$B_k \overset{\text{def}}{=} \left\{x \,|\, \varphi(x) = \lambda^k\right\}$$

and supposing

(3.10) $$\lambda^k \neq 1 \qquad \text{for} \quad 1 \leq k \leq N,$$

we find sets $B_o, B_1, \ldots, B_N$ with the properties

$$\begin{cases} \text{(i)} & B_o, B_1, \ldots, B_N \text{ are disjoint} \\ \text{(ii)} & \int_{B_{k+1}} dA(y|x_k) = 1 \quad \text{for all } x_k \in B_k \quad (k=0,1,\ldots,N-1) \\ \text{(iii)} & \int_{B_o} dA(y|x) \text{ is not identically zero.} \end{cases}$$

----

1) See footnote on page 9.

Remark: (iii) follows from (3.9), as we might (by proper norming of $\varphi(x)$) have taken B_1 instead of B_o.

We shall prove that N cannot exceed r-1. Suppose $N \geq r$, then we have by (i) and (ii) that the functions $A(y|x_k)$ (x_k an arbitrary element of B_k, $k=0,1,\ldots,r-1$) are linearly independent. From this it follows, that $\det[A_j(x_k)] \neq 0$. By (ii) however we have

$$\int_{B_o} dA(y|x_k) = \sum_{j=1}^{r} A_j(x_k) \int_{B_o} dB_j(y) = 0 \qquad (k=0,1,\ldots,r-1), \tag{3.11}$$

where by (iii) $\int_{B_o} dB_j(y)$ is not zero for all j. Thus (3.11) requires $\det[A_j(x_k)] = 0$. From this contradiction it follows that $N \leq r-1$, i.e. $\lambda^k=1$ for some $k \leq r$. From (ii) we now see that B contains k cyclically moving subsets $B_o \rightarrow B_1 \rightarrow \ldots \rightarrow B_{k-1} \rightarrow B_o$.

Summarizing we have

Theorem 3.2[1)]: If for every eigenfunction $\varphi(x)$ (of $A(y|x)$) corresponding to an eigenvalue of modulus 1 $\max_{x \in B} |\varphi(x)|$ exists we have

I If 1 is a multiple eigenvalue then B contains (at least) two disjoint absorbing sets.

II An eigenvalue of modulus one is a root of unity of an order $k \leq r$ and implies the existence of k cyclically moving subsets.

The following theorem has been proved in Runnenburg [1960] with the use of generating functions. Here we give a proof analogous to the proof of the analogous theorem for finite Markov chains.

Theorem 3.3: if C has a single eigenvalue 1 and no other eigenvalues of modulus 1, then

$$G(y) = \lim_{n \rightarrow \infty} A^{(n)}(y|x) \tag{3.12}$$

exists and is independent of x. G(y) is the only invariant distribution function, i.e. the only distribution function satisfying

---

1) After the manuscript had been typed it was seen that this condition is always satisfied.

(3.13) $$G(y) = \int A(y|x)\,dG(x)$$

and is given by

(3.14) $$G(y) = \sum_{j=1}^{r} \alpha_j B_j(y),$$

where the α_j are the (unique) solution of

(3.15) $$\sum_{j=1}^{r} \alpha_j c_{jk} = \alpha_k$$

$$\sum_{j=1}^{r} \alpha_j b_j = 1 \ .$$

Proof: as in the case of a finite Markov matrix the existence of $\Gamma \overset{\text{def}}{=} \lim_{n\to\infty} C^n$ follows from the well-known matrix decomposition

(3.16) $$C = \Gamma_o + \sum_{j=1}^{k} (\lambda_j \Gamma_j + C_j),$$

where $\lambda_1, \lambda_2, \ldots, \lambda_r$ are the distinct eigenvalues $\neq 1$ of C, the Γ_j are idempotent and orthogonal, the C_j are nilpotent and satisfy $\Gamma_j C_j = C_j \Gamma_j = \Gamma_j$ (see e.g. Wedderburn [1934] page 29). Now it follows from (3.16) that $\Gamma = \Gamma_o$.

For Γ_{jk} we evidently have

(3.17) $$\begin{cases} \Gamma_{jk} = \sum_{l=1}^{r} \Gamma_{jl} c_{lk} \\ \sum_{k=1}^{r} \Gamma_{jk} b_k = b_j \end{cases} \qquad (j=1,2,\ldots,r) \ .$$

From (3.15) and (3.17) it follows that

(3.18) $$\Gamma_{jk} = b_j \alpha_k \ .$$

From the existence of $\lim_{n\to\infty} C^n$ the existence of $G(y) = \lim_{n\to\infty} A^{(n)}(y|x)$ follows by (2.15). Clearly we have

$$G(y) = \sum_{j=1}^{r} \sum_{k=1}^{r} A_j(x) b_j \alpha_k B_k(y) = \sum_{k=1}^{r} \alpha_k B_k(y).$$

It is easily seen that G(y) is a distribution function satisfying (3.13), while its uniqueness as an invariant distribution follows from the fact that every distribution function satisfying (3.13) must satisfy (3.14) and (3.15).

4. Correlation coefficients for r=2

In order to compare the behaviour of sequences $\underline{x}_n$ (n=1,2,...) and $\underline{y}_n$ (n=1,2,...), where the $\underline{x}_n$ are independent random variables all having the same distribution function $G(x) = P\{\underline{x} \leqslant x\}$ and the $\underline{y}_n$ are Markov-dependent random variables with transition function $A(y|x)$ and invariant distribution G(y), we have to search first for those transition functions admitting a given G(y) as invariant distribution. Or, if we compute

$$(4.1) \qquad H(x,y) = P\{\underline{x} \leqslant x,\ \underline{y} \leqslant y\} = \int_{-\infty}^{x} A(y|t)\,dG(t),$$

we have to search for those bivariate distribution functions of random variables $\underline{x}$ and $\underline{y}$, which have G(y) as marginal distribution function for both $\underline{x}$ and $\underline{y}$. The latter question has been considered in Fréchet [1951], Gumbel [1958] .

It has been shown by Fréchet that all bivariate distribution functions H(x,y) with prescribed marginal distribution functions $F(x) = H(x,\infty)$ and $G(y) = H(\infty,y)$ satisfy

$$(4.2) \qquad \overline{C}(x,y) \overset{\text{def}}{=} \max(F(x)+G(y) - 1,0) \leqslant H(x,y) \leqslant \overline{D}(x,y) \overset{\text{def}}{=} \min(F(x),G(y))$$

for all x and y, whatever F(x) and G(y). Moreover, $\overline{C}(x,y)$ and $\overline{D}(x,y)$ are always bivariate distribution functions. If we assume that

$$(4.3) \qquad F(0-) = G(0-) = 0, \quad \mathcal{E}\,\underline{x}^2 < \infty\ , \quad \mathcal{E}\,\underline{y}^2 < \infty,$$

then $\mathcal{E}\,\underline{x}\,\underline{y}$ always exists and we have by Fubini's theorem

$$(4.4) \qquad \mathcal{E}\,\underline{x}\,\underline{y} = \int_0^\infty \int_0^\infty \{1 - F(x) - G(y) + H(x,y)\}\, dx\, dy$$

and also

$$(4.5) \qquad \mathcal{E}\,\underline{x}\,\underline{y} - \mathcal{E}\,\underline{x}\; \mathcal{E}\,\underline{y} = \int_0^\infty \int_0^\infty \{H(x,y) - F(x)G(y)\}\, dx\, dy.$$

Hence the bivariate distribution function H(x,y) with prescribed F(x) and G(y) having the largest correlation coefficient is given by $\overline{D}(x,y)$, while that with the smallest correlation coefficient is given by $\overline{C}(x,y)$.

We can use the conditional distribution function $A(y|x)$ to a given bivariate distribution function with equal marginal distribution functions F(y) = G(y) as transition function for a Markov chain $\underline{y}_1$, $\underline{y}_2$,... with invariant distribution G(y), i.e. from

$$(4.6) \qquad D(x,y) = \min(G(x), G(y))$$

we may take as a conditional distribution function

$$(4.7) \qquad A_D(y|x) = \ell(y-x),$$

where

$$(4.8) \qquad \ell(x) = \begin{cases} 1 & \text{if } x \geqslant 0, \\ 0 & \text{if } x < 0. \end{cases}$$

In this case all mass is distributed on the line x=y, i.e. the Markov chain obtained is trivial. For

$$(4.9) \qquad C(x,y) = \max(G(x) + G(y) - 1, 0)$$

we may take

$$(4.10) \qquad A_C(y|x) = \ell(G(x) + G(y) - 1),$$

i.e. all mass now lies on the curve

(4.11) $$G(x) + G(y) = 1.$$

Here again the Markov chain obtained is trivial. If the chain starts with $\underline{y}_1=x$, then there is exactly one possible value y for $\underline{y}_2$, i.e. $\underline{y}_2=y$, from which $\underline{y}_3=x$, $\underline{y}_4=y$, etc. are obtained. More interesting transition functions are those for which we start from a mixture of the bivariate distribution functions (Fréchet)

$$C(x,y),\ G(x)G(y) \text{ and } D(x,y), \text{ say}$$

(4.12) $$A(y|x) = p_1A_C(y|x) + p_2G(y) + p_3A_D(y|x)$$

with $p_1,p_2,p_3 \geq 0$ and $p_1+p_2+p_3=1$.

Consider the class $H_r(G)$ of bivariate distribution functions $H_r(x,y)$ for which we have

(4.13) $$H_r(x,y) = \sum_{j=1}^{r} \bar{B}_j(x)\ B_j(y),\quad H_r(x,\infty) = G(x),\ H_r(\infty,y) = G(y),$$

where G(y) is a distribution function and the $\bar{B}_j(x)$ (as well as the $B_j(y)$) are linearly independent. Then for fixed $r \geq 2$ it is impossible to find $C_r(x,y) \in H_r(G)$ and $D_r(x,y) \in H_r(G)$ with

(4.14) $$C_r(x,y) \leq H_r(x,y) \leq D_r(x,y) \text{ for all } x,y \text{ and all } H_r(x,y) \in H_r(G),$$

because the restrictions on $H_r(x,y)$ are such that we must have $C_r(x,y) = C(x,y)$ and $D_r(x,y) = D(x,y)$, but $C(x,y)$ and $D(x,y)$ do not belong to $H_r(G)$.
For r=1 there is no problem, because $H_1(G)$ only contains

(4.15) $$H_1(x,y) = G(x)G(y).$$

However, for $r \geq 2$ one may ask for those $C_r(x,y) \in H_r(G)$ (or those $D_r(x,y) \in H_r(G)$) which have smallest (largest) correlation coefficient. We shall solve this problem for r=2 in order to get more bivariate distribution functions to make a mixture from and also to see in how far the correlation coefficients obtained under the restriction r=2 fall short of the extreme correlation coefficients supplied by C(x,y) and D(x,y).

We assume in the remaining part of this section that G(y) is an arbitrary continuous distribution function with finite second moment.

We first show that any

$$A(y|x) = \sum_{j=1}^{2} A_j(x)\ B_j(y) \tag{4.16}$$

satisfying the conditions (a),(b),(c) and (d) of section 2 and having invariant distribution G(y) may be written

$$A(y|x) = G(y) + R(x)S(y), \tag{4.17}$$

where $S(-\infty) = S(\infty) = 0$ and R(x) and S(y) are of bounded variation. It is known that

$$G(y) = \alpha_1 B_1(y) + \alpha_2 B_2(y) \tag{4.18}$$

and hence

$$A(y|x) = \alpha_1 B_1(y) + \alpha_2 B_2(y) + (A_1(x) - \alpha_1)B_1(y) + (A_2(x) - \alpha_2)B_2(y), \tag{4.19}$$

where

$$\begin{cases} 1 = \alpha_1 B_1(\infty) + \alpha_2 B_2(\infty) \ , \\ 1 = A_1(x)B_1(\infty) + A_2(x)B_2(\infty) \end{cases} \tag{4.20}$$

indicating that $A_1(x) - \alpha_1$ and $A_2(x) - \alpha_2$ are linearly dependent (because $B_1(\infty) = B_2(\infty) = 0$ is impossible). We may assume $B_1(\infty) \neq 0$, i.e.

$$A(y|x) = G(y) + (A_2(x) - \alpha_2)(B_2(y) - \frac{B_2(\infty)}{B_1(\infty)} B_1(y)). \tag{4.21}$$

Thus

$$\begin{cases} R(x) \overset{\text{def}}{=} A_2(x) - \alpha_2 \\ S(y) \overset{\text{def}}{=} B_2(y) - \frac{B_2(\infty)}{B_1(\infty)} B_1(y) \end{cases} \tag{4.22}$$

satisfy our conditions.

As

(4.23) $$\int_{-\infty}^{\infty} A(y|x)\,dG(x) = G(y),$$

we have

(4.24) $$S(y) \int_{-\infty}^{\infty} R(x)\,dG(x) = 0 \quad \text{for all } y.$$

The functions G(y) and S(y) are linearly independent, hence $S(y) \neq 0$ for some y and

(4.25) $$\int_{-\infty}^{\infty} R(x)\,dG(x) = 0.$$

Extreme values for the correlation coefficient occur for those R(x) and S(y) for which

(4.26) $$\begin{aligned} \mathcal{E}\,\underline{x}\,\underline{y} - \mathcal{E}\,\underline{x}\,\mathcal{E}\,\underline{y} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\; d_y A(y|x)\,dG(x) - \mathcal{E}\,\underline{x}\,\mathcal{E}\,\underline{y} = \\ &= \int_{-\infty}^{\infty} x\,R(x)\,dG(x)\,.\int_{-\infty}^{\infty} y\,d\,S(y) \end{aligned}$$

has an extreme value. One can easily indicate functions R(x) and S(y) for which

(4.27) $$\begin{cases} I_R \overset{\text{def}}{=} \int_{-\infty}^{\infty} x\,R(x)\,dG(x) \\ I_S \overset{\text{def}}{=} \int_{-\infty}^{\infty} y\,d\,S(y). \end{cases}$$

are both unequal zero and of the same sign or unequal zero and of different sign. As we are interested in the product of I_R and I_S, it is no restriction to assume that $I_R > 0$. Now let us assume that S(y) is known and make I_R as large as possible. We know that R(x) is bounded, say

(4.28) $$-R_1 \leqslant R(x) \leqslant R_2 \quad \text{for all } x$$

with finite nonnegative R_1 and R_2, because $-1 \leqslant R(x)S(y) \leqslant 1$ for all x and y and $S(y) \neq 0$ for some y. We take R_1 and R_2 as large as possible: If $G(y) + R\,S(y)$ is a distribution function, then we can prove that there exists an $R_2 \geqslant 0$ such that $G(y) + R_2\, S(y)$ is a distribution function while $G(y) + (R_2+\varepsilon)S(y)$ is not, whatever $\varepsilon > 0$ we consider. We already noticed that R is bounded from above. If $G(y)+R\,S(y)$ is a distribution function for some $R > 0$, the same is true for all $G(y) + \Theta R\, S(y)$ with $0 \leqslant \Theta \leqslant 1$, because

$$G(y) + \Theta R\, S(y) = (1-\Theta)G(y) + \Theta \left\{ G(y) + R\, S(y) \right\} . \tag{4.29}$$

If $G(y) + (R_2-\varepsilon)S(y)$ is a distribution function for all $\varepsilon > 0$ (such that $R_2 - \varepsilon \geqslant 0$) and not for any $\varepsilon < 0$, then $G(y) + R_2 S(y)$ is also a distribution function, as can easily be verified. Hence a largest $R_2 \geqslant 0$ exists. In the same way we can prove the existence of a largest $R_1 \geqslant 0$.

Now the problem: "If S(y) is given, for which R(x) satisfying (4.25) and (4.28) has I_R (as given by (4.27)) its largest value" has as solution

$$R^*(x) = \begin{cases} -R_1 & \text{for } x \leqslant \xi , \\ R_2 & \text{for } x > \xi , \end{cases} \tag{4.30}$$

where from (4.25) we have that ξ satisfies

$$-R_1 \int_{-\infty}^{\xi} dG(x) + R_2 \int_{\xi}^{\infty} dG(x) = 0 \tag{4.31}$$

or (as $R_1+R_2 = 0$ implies r=1 instead of r=2, hence $R_1+R_2 > 0$).

$$G(\xi) = \frac{R_2}{R_1+R_2} \text{ [1]} . \tag{4.32}$$

Indeed for any other function R(x) satisfying (4.25) we have

$$\begin{aligned} \int_{-\infty}^{\infty} x \left\{ R^*(x)-R(x) \right\} dG(x) &= \int_{-\infty}^{\xi} x \left\{ R^*(x)-R(x) \right\} dG(x) + \\ &+ \int_{\xi}^{\infty} x \left\{ R^*(x)-R(x) \right\} dG(x) \geqslant \\ \geqslant \xi \int_{-\infty}^{\xi} \left\{ R^*(x)-R(x) \right\} dG(x) &+ \xi \int_{\xi}^{\infty} \left\{ R^*(x)-R(x) \right\} dG(x) = 0. \end{aligned} \tag{4.33}$$

---

1) Of course equation (4.32) need not determine ξ uniquely.

Because $R_1=0$ implies $I_R=0$, while we assumed $I_R > 0$, we must have $R_1 > 0$. In the same way $R_2 > 0$. It is thus no restriction to assume

(4.34) $$R_1 R_2 = 1,$$

because we can replace $R^*(x)$ by $\alpha^{-1}R^*(x)$ and $S(y)$ by $\alpha S(y)$, without altering $A(y|x)$ and take $\alpha = \sqrt{R_1 R_2} > 0$.

From the foregoing it is clear that extreme correlation coefficients occur for

(4.35) $$A(y|x) = \begin{cases} G(y) - R_1 S(y) & \text{for } x \leqslant \xi , \\ G(y) + R_2 S(y) & \text{for } x > \xi , \end{cases}$$

where $A(y|x)$ is a distribution function for all values of x, while R_1 and R_2 are extreme values corresponding to $S(y)$ and ξ follows from (4.32). Therefore we must certainly have

(4.36) $$0 \leqslant G(y) - R_1 S(y) \leqslant 1 \quad \text{and} \quad 0 \leqslant G(y) + R_2 S(y) \leqslant 1$$

or

(4.37) $$-\min\left(\frac{1-G(y)}{R_1} , \frac{G(y)}{R_2}\right) \leqslant S(y) \leqslant \min\left(\frac{1-G(y)}{R_2} , \frac{G(y)}{R_1}\right).$$

It is not hard to verify (say by partial integration) that we obtain from the $S(y)$ (of bounded variation, satisfying (4.37) and having R_1 and R_2 as extreme values in the sense of (4.28)) the largest value for I_s, if we take

(4.38) $$S^*(y) \overset{\text{def}}{=} -\min\left(\frac{1-G(y)}{R_1} , \frac{G(y)}{R_2}\right)$$

and the smallest value, if we take

(4.39) $$S_*(y) \overset{\text{def}}{=} \min\left(\frac{1-G(y)}{R_2} , \frac{G(y)}{R_1}\right).$$

Now $G(y) - R_1 S^*(y)$, $G(y) + R_2 S^*(y)$, $G(y) - R_1 S_*(y)$ and $G(y) + R_2 S_*(y)$ are all distribution functions.

Apart from trivial calculations we have now proved the following

Theorem 4.1.[1] Among the $A(y|x) \in H_2(G)$ the one with largest correlation coefficient (if it exists) is to be found among those for which

$$A^*(y|x) = \begin{cases} G(y) - R_1 S^*(y) & \text{for } x \leq \xi , \\ G(y) + R_1^{-1} S^*(y) & \text{for } x > \xi \end{cases} \tag{4.40}$$

holds, where $R_1 > 0$ and ξ satisfies

$$G(\xi) = \frac{1}{R_1^2+1} . \tag{4.41}$$

The corresponding bivariate distribution function is

$$H^*(x,y) = P\{\underline{x}^* \leq x, \underline{y}^* \leq y\} = G(x)G(y) + S^*(x)S^*(y). \tag{4.42}$$

Assuming the random variables $\underline{x}^*$ and $\underline{y}^*$ to have zero expectation and unit variance, the corresponding correlation coefficient is given by

$$\rho(\underline{x}^*, \underline{y}^*) = \frac{\left\{\int_{-\infty}^{\xi} x \, dG(x)\right\}^2}{G(\xi)\left\{1-G(\xi)\right\}} . \tag{4.43}$$

The $A(y|x)$ yielding the smallest correlation coefficient (if it exists) is to be found among those for which

$$A_*(y|x) = \begin{cases} G(y) - R_1 S_*(y) & \text{for } x \leq \xi , \\ G(y) + R_1^{-1} S_*(y) & \text{for } x > \xi \end{cases} \tag{4.44}$$

holds, where $R_1 > 0$ and ξ satisfies (4.41). The corresponding bivariate distribution function is

$$H_*(x,y) = P\{\underline{x}_* \leq x, \underline{y}_* \leq y\} = G(x)G(y) + S^*(x)S_*(y). \tag{4.45}$$

Assuming the random variables $\underline{x}_*$ and $\underline{y}_*$ to have zero expectation and unit variance, the corresponding correlation coefficient is given by

---

1) G(y) is continuous with finite second moment.

(4.46) $$\rho(\underline{x}_*, \underline{y}_*) = - \frac{\int_{-\infty}^{\xi} x dG(x) \int_{-\infty}^{\eta} y dG(y)}{G(\xi)\, G(\eta)} ,$$

where η satisfies

(4.47) $$G(\eta) = \frac{R_1^2}{R_1^2+1} = 1-G(\xi) .$$

Examples

A) Rectangular distribution G(y). Here

(4.48) $$G(y) = \frac{y+\sqrt{3}}{2\sqrt{3}} \quad \text{for} \quad -\sqrt{3} \leq y \leq \sqrt{3}.$$

From (4.43) we have

(4.49) $$\rho(\underline{x}^*, \underline{y}^*) = \tfrac{1}{4}(3-\xi^2)$$

and hence $\xi = 0$ leads to a largest correlation coefficient

(4.50) $$\rho^* = \frac{3}{4} .$$

From (4.46), (4.41) and (4.47) we obtain

(4.51) $$\begin{cases} \rho(\underline{x}_*, \underline{y}_*) = -\tfrac{1}{4}(3-\eta^2) \\ \xi + \eta = 0 \end{cases}$$

and hence $\xi = \eta = 0$ leads to a smallest correlation coefficient

(4.52) $$\rho_* = -\frac{3}{4} .$$

Under Fréchet's restrictions, i.e. using C(x,y) and D(x,y) as bivariate distribution functions, we find a largest correlation coefficient ρ_{max} and a smallest ρ_{min} with

(4.53) $$\begin{cases} \rho_{max} = 1, \\ \rho_{min} = -1. \end{cases}$$

Gumbel's bivariate distribution functions

$$H_a(x,y) = G(x)G(y)\left\{1+a(1-G(x))(1-G(y))\right\} \qquad (-1 \leqslant a \leqslant 1) \tag{4.54}$$

lead to a largest correlation coefficient $\rho_{G_{max}}$ and a smallest $\rho_{G_{min}}$ with

$$\begin{cases} \rho_{G_{max}} = \frac{1}{3} & \text{for } a = 1, \\ \rho_{G_{min}} = -\frac{1}{3} & \text{for } a = -1. \end{cases} \tag{4.55}$$

B) Exponential distribution G(y). Here

$$G(y) = 1-e^{-y-1} \qquad \text{for } y \geqslant -1. \tag{4.56}$$

From (4.43) we have

$$\rho(\underline{x}^*,\underline{y}^*) = \frac{\left\{\int_{-1}^{\xi} xe^{-x}dx\right\}^2}{e^{-\xi-1}(1-e^{-\xi-1})} = \frac{\tau^2 e^{-\tau}}{1-e^{-\tau}}, \tag{4.57}$$

where $\tau = \xi+1 > 0$. The largest value for $\rho(\underline{x}^*,\underline{y}^*)$ occurs for $\tau = \tau_o$, where τ_o is the (only) positive solution of

$$\tau = 2(1-e^{-\tau}).$$

It turns out that

$$\tau_o = 0{,}20325$$

and hence

$$\rho^* = \sup_{\tau > 0} \rho(\underline{x}^*,\underline{y}^*) = \tfrac{1}{4}\,\tau_o(1-\tau_o) = 0{,}648. \tag{4.58}$$

From (4.46), (4.41) and (4.47) we obtain

$$\rho(\underline{x}_*,\underline{y}_*) = -\frac{\int_{-1}^{\xi} xe^{-x-1}dx \, . \int_{-1}^{\eta} ye^{-y-1}dy}{(1-e^{-\xi-1})(1-e^{-\eta-1})} =$$

$$= -\log \Theta . \log(1-\Theta)$$

with

$$\theta \overset{\text{def}}{=} 1-e^{-\xi-1} \quad .$$

Hence the smallest value for $\rho(\underline{x}_*, \underline{y}_*)$ is given by

(4.59) $$\rho_* = -(\log 2)^2 = -0,480.$$

and occurs for $\theta=\frac{1}{2}$, i.e. $\xi = \eta = -1 + \log 2$.

Under Fréchet's restrictions we find

$$\rho_{max} = 1$$

$$\rho_{min} = -\int_0^\infty xe^{-x}\log(1-e^{-x})dx-1 =$$

$$= \sum_{n=1}^\infty \frac{1}{n(n+1)^2} -1 = 1 - \frac{\pi^2}{6} = -0,645.$$

Gumbel's bivariate distribution functions lead to

(4.61) $$\begin{cases} \rho_{G_{max}} = \frac{1}{4} & \text{for } a=1, \\ \rho_{G_{min}} = -\frac{1}{4} & \text{for } a=-1. \end{cases}$$

The foregoing examples suggest that ρ_* always occurs for $\xi = \eta$. This is not true as may be verified by taking

(4.62) $$G(y) = \begin{cases} \dfrac{y+a+\frac{1}{6}}{3a} & \text{for } -a-\frac{1}{6} \leqslant y \leqslant -\frac{1}{6}, \\ y+\frac{1}{2} & \text{for } -\frac{1}{6} \leqslant y \leqslant \frac{1}{6}, \\ \dfrac{y+2a-\frac{1}{6}}{3a} & \text{for } \frac{1}{6} \leqslant y \leqslant a+\frac{1}{6}, \end{cases}$$

where a is a sufficiently large positive constant.

It is interesting to note that ρ^* always occurs for a decomposable Markov chain, while ρ_* sometimes occurs for a periodic Markov chain (if $\xi = \eta$) and sometimes for an aperiodic Markov chain (if $\xi \neq \eta$).

References

Dantzig, D. van, Markov Chains (stencilled notes, in dutch), Report S 156 (C 7) of the Mathematical Centre, 1956.

Dmitriev, N. and E. Dynkin, On characteristic roots of stochastic matrices. Isvestija, Ser.Mat.10, 1946, pp.167-194.

Doob, J.L., Stochastic Processes, John Wiley & Sons, New York, 1953.

Fréchet, M., Sur les tableaux de corrélation dont les marges sont donnés, Ann.Univ.de Lyon III ser.fasc.14 A, pp. 53-77, 1951.

Gumbel, E.J., Multivariate distributions with given margins and analytical examples, 31st Session of the Int.Stat. Inst., Brussels 1958, published in Bull.de l'Inst. de Stat., Tome XXXVII-3, Bruxelles, 1960.

Runnenburg, J.Th., Markov Processes in Waiting-time and Renewal Theory, Poortpers, Amsterdam, 1960.

Runnenburg, J.Th., An example illustrating the possibilities of renewal theory and waiting-time theory for Markov-dependent arrival-intervals. Indag.Math.23, 5, 1961.

Wedderburn, J.H.M., Lectures on Matrices, Am.Math.Soc. Coll.Publ., vol.17, 1934.