MATRIX SHIFT OPERATORS

APPLIED TO MARKOV CHAINS

by

PETER J. R. BOYLE

B. Sc., University of Pretoria, 1965

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1968

Approved by:

[signature]

LD
2668
T4
1968
B69
c.2

TABLE OF CONTENTS

CHAPTER I

PRELIMINARIES

1.1 Introduction

The orientation of this work is a little different from most writings of this nature in that it is mostly a setting forth of preliminary excursions into what appears to be a promising off-shoot of the applications of Markov chains. Some results of previous authors, primarily Conover [1965] and Gani [1965], are presented in a slightly less procedure-oriented light from their original settings. Whenever interesting points were found, they were developed to some degree and, if they seemed worthwile in any sense, included here. No one of these points was developed to any large extent due to limitations in the time available. Instead, an attempt was made to present a direction in which development may prove of value.

The origins of this work lie in the theory of dams, queues, and provisioning. At this stage it touches them all as well as other related subjects. It is for this reason that examples used to illustrate the results obtained were somewhat simplified selections from these fields.

The remainder of this chapter will present a short resume of the subjects upon which it is hoped this work may build.

1.2 Storage theory

The theory of storage concerns itself with a storage function $S(t)$, defined on a discrete or continuous parameter t (most frequently time) of the following type:

$$S(t) = I(t) - D(t) - F(t) \qquad t \geq 0$$

where $I(t)$ is a random feeding function, $D(t)$ is a depleting function which may or may not be a random variable, and $F(t)$ is a limiting function determining the maximum size of $S(t)$. All of these functions are positive and non-decreasing. Thus they are cumulative functions in time representing the total feed, depletion and 'overflow' respectively.

An early problem of this type concerned the problem of ruin in an insurance company (Segerdahl, 1939) where $S(t)$ represents the company capital, $I(t)$ the premiums, and $D(t)$ the claims. In general one finds problems in storage theory arising under the subject of provisioning or under dam theory. Gani [1955] has pointed out that, with $S(t)$ the deficit in provisioning, or the total content in dam theory, the two problems are direct analogues and can be simultaneously handled.

In dam theory, the major objective appears to be to obtain stationary distributions for the dam content, $S(t)$, under given input distributions and for given release rules. A model, first proposed by Moran in 1954, has provided a major tool for this work and has been developed by many workers. Among them are Prabhu, Yeo, Gani, and Ghosal. For an informal review of the results up to 1957, the reader is referred to the papers by Gani [1957] and Kendall [1957]. Prabhu reviews some later results in 1965.

Gani, referring to storage problems in general, had occasion to say, "a systematic method of attack has yet to emerge from the known solutions". As far as can be ascertained the same situation exists today. This is a motivation for the attempt at some generalization in this work.

Most usually, writers on the subject of storage theory have assumed that the inputs giving rise to the feeding function, $I(t)$, are independent over time, that is that $I(t)$ has (at least) stationary independent increments. With this restriction, a wide class of discrete storage functions may be classed as Markov chains and have relevance to this work.

1.3 Flooding theory

Interest in the statistical treatment of flooding dates back to the early part of this century. Fuller in 1914, was perhaps one of the first. It was not until sometime later that any level of sophistication was reached however, for in 1936, Slade was apparently justified in deriding the methods employed up to that date. During this period "ad hoc" solutions based on interpolation formulae and empirical distributions found favor. Papers on these methods, for example that by Geyer in 1940, can be found in profusion. Then in 1942 Gumbel wrote a paper employing known distributions to obtain theoretical return periods of floods, and with it opened up a whole new approach.

This activity notwithstanding, theories offering solutions to flooding in terms of dams are rare to date. The results by Gani [1965], referred to throughout this work, are far from complete. These results, despite the timely development of the shift operator by Conover [1965], are too complex to offer any immediate solution to real problems of a practical nature. The ground work is laid, however, and in time, practical results will no doubt evolve.

1.4 Gani's flooding model

Consider a river system with a network of m tributaries feeding a main stream. Let $I_n(t)$, $n = 1, 2, \ldots, m$; $t \geq 0$ be the input (in volume units during the t^{th} time unit) supplied by the n^{th} tributary to the main stream.

The following assumptions will be made for this model:

(i) It will be assumed that the input $I_n(t)$ takes on only discrete values from the finite set $\{0, 1, 2, \ldots, L_n\}$; $n = 1, 2, \ldots, m$.

In much of the work it will be necessary to assume that there exists a finite bound, s, on the L_n:

$$s = \text{Max}\ \{L_n\} \qquad n = 1, 2, \ldots, m$$

(ii) It will be assumed that the flow magnitude in the main stream is affected only by the inputs and any dams that may be placed on the river. Thus, for example, there is no incidental loss or gain between one tributary junction and the next.

(iii) It will also be assumed that the flow speed in the river is independent of the flow magnitude, at least during the flood period. This assumption allows for a simplification as follows:

Let τ_n be the time it takes for a particle to travel from the source of the main stream to the junction of the n^{th} tributary. Then make a linear transformation on $I_n(t)$:

$$I_n'(t) = I_n(t + \tau_n) \qquad n = 1, 2, \ldots, m .$$

Now the time variable, t, refers to the same water particle as it arrives at each of the tributary junctions, and thus explicit mention of time is no longer necessary. $I_n'(t)$ will be referred to simply as I_n.

(iv) It will be assumed that I_n, $n = 1, 2, \ldots, m$, is a random variable such that

(a) I_n is independent over time. That is to say that an observation of I_n for one time interval is independent of any observations of I_n for other time intervals. This may be true for example, if a time unit is taken long enough.

(b) The stochastic process $\{I_n;\ n = 1, 2, \ldots, m\}$ is a non-homogeneous Markov chain.

Under these assumptions it is possible to obtain the probability distribution, R_n, (a row vector) of the flow magnitude, S_n, on the n^{th} section, that is between the $(n-1)^{st}$ and n^{th} tributary junctions:

$$R_n = (r_0^{(n)}, r_1^{(n)}, r_2^{(n)}, \ldots)$$

where

$$r_i^{(n)} = P(S_n = i)$$

Since the flow is independent over time (by assumption) this probability distribution fully describes the flow during flood times.

The probability of flooding on the n^{th} section is obtainable as follows: Let F_n be the level of flow magnitude on the n^{th} section which just causes flooding. That is flooding occurs on the n^{th} section if $S_n \geq F_n$. Then define a row vector Z_n such that it has unity in the first F_n positions and zeros elsewhere.

Then:

$$P(\text{flooding on } n^{th} \text{ section}) = 1 - R_n Z_n' \tag{1.1}$$

This model will be referred to in the body of this work and R_n will be obtained under different dam systems.

CHAPTER 2

THE MATRIX SHIFT MULTIPLICATION OPERATORS

2.1 Introduction

The shift multiplication operation is a matrix operation first used by Conover [1965] in order to express the distribution of the sum of the first m states of a non-homogeneous finite state Markov chain in a mathematically manipulable form.

Later, Gani [1965] used a similar operator developed from Conover's operator. He used it to sum the states of a Markov chain in his flooding models. This new operator offered some advantages but in return tended to aggravate some problems in mathematical properties which arise when sequential shift operations are performed repeatedly.

Both of the above authors worked with the finite state case only. The shift operations needed to consider the infinite case are added in paragraph 2.3.

Since both operators have advantages, both will be considered in this chapter. A more detailed discussion of the properties and relative advantages of the two operators can be found in appendix I.

2.2 The definitions of the shift operators

2.2.1 Notation

We will use the following matrices for definition purposes:
Let A be a $(p+1) \times (q+1)$ matrix of the form;

$$\begin{vmatrix} a_{o,o} & \cdots & a_{o,q} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ a_{p,o} & \cdots & a_{p,q} \end{vmatrix}$$

and let B be a (q+1) x (r+1) matrix of the form;

$$\begin{vmatrix} b_{o,o} & \cdots & b_{o,r} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ b_{q,o} & \cdots & b_{q,r} \end{vmatrix}$$

Then let C = AB, the ordinary matrix product with dimensions (p+1) x (r+1) and elements given by:

$$c_{i,j} = \sum_{k=0}^{q} a_{i,k} b_{k,j} \qquad \begin{cases} i = 0, 1, 2, \ldots p \\ j = 0, 1, 2, \ldots r \end{cases}$$

The subscripting from 0 is used throughout this work rather than the more common subscripting from unity because the dimensions of the product matrix of a shift operation are more easily expressed in terms of p,q and r than in terms of full dimensions. Further, their application to Markov chains is under consideration and this notation is significant in the interpretation of the expressions in this context.

2.2.2 Conover's shift operation

DEFINITION I: Conover's shift operation is an operation defined for two matrices, A and B with a common dimension, (q+1), which delivers a matrix D of dimensions (p+r+1) x (r+1) as follows.

The element of D, namely $d_{i,j}$ is given by

$$\begin{aligned} d_{i,j} &= c_{i-j,j} \text{ if } 0 \le i-j \le p \\ &= 0 \text{ otherwise} \end{aligned} \qquad \begin{cases} i = 0, 1, 2, \ldots p+r \\ j = 0, 1, 2, \ldots r \end{cases} \qquad (2.1)$$

The operation is written $D = A * B$.[1]

Example 2.1. Some feeling for this operation can be obtained from a simple example, such as that given, as well as a consideration of the general term in paragraph 2.2.4.

Consider the matrices

$$A = \begin{vmatrix} 1 & 2 & 0 \\ 0 & 1 & 4 \end{vmatrix}$$

and

$$B = \begin{vmatrix} 1 & 2 & 2 \\ 2 & 0 & 1 \\ 3 & 2 & 0 \end{vmatrix}$$

then the ordinary matrix product AB gives us:

$$AB = \begin{vmatrix} 5 & 2 & 4 \\ 14 & 8 & 1 \end{vmatrix}$$

So that by sequentially shifting the columns of AB downward we get:

$$A * B = \begin{vmatrix} 5 & 0 & 0 \\ 14 & 2 & 0 \\ 0 & 8 & 4 \\ 0 & 0 & 1 \end{vmatrix}$$

Dimensions are in order: $(1+2+1) \times (2+1) = 4 \times 3$

2.2.3 Gani's shift operator

DEFINITION II: Gani's shift operator is an operation defined for two matrices, A and B, with a common dimension, $(q+1)$ which delivers a matrix G of dimensions $(p+1) \times (q+r+1)$ as follows:

---

[1] The notation using the asterisk is in accordance with the original notation by Conover [1965]. It is probably intentionally the same as that used by Feller [1965] to indicate a convolution. While the two are related (paragraph 2.4), Gani's operator actually bears closer resemblance to the convolution operator.

Let $\overset{\theta}{B}$ be the matrix of elements $\overset{\theta}{b}_{i,j}$ such that;

$$\left.\begin{aligned} \overset{\theta}{b}_{i,j} &= b_{i,j-i} \text{ if } 0 \leq j-i \leq r \\ &= 0 \text{ otherwise} \end{aligned}\right. \qquad \begin{cases} i = 0, 1, \ldots p \\ j = 0, 1, \ldots q+r \end{cases} \tag{2.2}$$

so that $\overset{\theta}{B}$ has dimension $(q+1) \times (q+r+1)$

then $G = A \theta B \stackrel{\text{def}}{=} A\overset{\theta}{B}$

Example 2.2. Consider the matrix

$$B = \begin{vmatrix} 1 & 2 & 2 \\ 2 & 0 & 1 \\ 3 & 2 & 0 \end{vmatrix}$$

From B we obtain $\overset{\theta}{B}$ by sequentially shifting the rows of B to the right:

$$\overset{\theta}{B} = \begin{vmatrix} 1 & 2 & 2 & 0 & 0 \\ 0 & 2 & 0 & 1 & 0 \\ 0 & 0 & 3 & 2 & 0 \end{vmatrix}$$

so that if

$$A = \begin{vmatrix} 1 & 2 & 0 \\ 0 & 1 & 4 \end{vmatrix}$$

then

$$G = A \theta B = \begin{vmatrix} 1 & 6 & 2 & 2 & 0 \\ 0 & 2 & 12 & 9 & 0 \end{vmatrix}$$

Dimensions are in order:

for $\overset{\theta}{B}$: $(2+1) \times (2+2+1) = 3 \times 5$

for G: $(1+1) \times (2+2+1) = 2 \times 5$

2.2.4 The general elements

It can be easily seen from their definitions that the following expressions give the general elements of the two shift operators:

Conover: Let $A = \{a_{i,j}\}_{(p+1) \times (q+1)}$

$$B = \{b_{i,j}\}_{(q+1) \times (r+1)}$$

then if $D = \{d_{i,j}\} = A * B$ then D is $(p+r+1) \times (r+1)$ with general element:

$$d_{i,j} = \sum_{k=0}^{q} a_{i-j,k} b_{k,j} \text{ if } 0 \leq i-j \leq p \quad \begin{cases} i = 0, 1, \ldots p+1 \\ j = 0, 1, \ldots r \end{cases} \qquad (2.3)$$
$$= 0 \text{ otherwise}$$

Gani; for A and B as above:

If $G = A \ominus B$ then G is $(p+1) \times (q+r+1)$ with the general element given by:

$$g_{i,j} = \sum_{k=0}^{q} a_{i,k} b_{k,j-k} \qquad \begin{cases} i = 0, 1, \ldots p \\ j = 0, 1, \ldots q+r \end{cases} \qquad (2.4)$$

2.3 <u>The case of infinite matrices</u>

2.3.1 Motivation

Gani and Conover considered the finite state case. Thus the shift products were not defined for the case where some or all of the dimensions of the operand matrices are infinite.

However, there does not appear to be much difficulty in expanding these conclusions to include the infinite state space case. Since many problems, including Gani's flooding model, can benefit from an infinite state space as an approximation to the continuous case and since many problems, including the queueing models, customarily use an infinite though discrete state space it is rewarding to develop the infinite state space case whenever possible.

2.3.2 Definitions

If the matrices which are to be shift multiplied are infinite in some or all dimensions, the definitions I and II are still applicable with the appropriate dimensions (p,q or r) taken to be infinite.

The general elements undergo slight changes, however. It was for this reason that the operations were not defined in terms of their general elements.

2.3.3 General elements

A consideration of the definitions I and II lead directly to the following general elements in the infinite case.

Gani shift product: The general element of $G = A \theta B$ where A and B are both infinite in both dimensions is given by;

$$g_{i,j} = \sum_{k=0}^{j} a_{i,k} b_{k,j-k} \qquad \begin{cases} i = 0, 1, 2, \ldots \\ j = 0, 1, 2, \ldots \end{cases} \qquad (2.5)$$

In this case G is an infinite matrix with no elements necessarily zero.

If B is infinite in both dimensions while A is an infinite row vector:

$$A = [a_0, a_1, a_2, \ldots]$$

then the general element of $G = A \theta B$ is given by:

$$g_j = \sum_{k=0}^{j} a_k b_{k,j-k} \qquad j = 0, 1, 2, \ldots \qquad (2.6)$$

In this case G is an infinite row vector.

The summations in equations 2.5 and 2.6 extend from 0 to j rather than, as may be expected, to ∞. This is because the Gani shift operator involves the sequential movement of the rows of B to the right to give $\overset{\theta}{B}$ before multiplication. $\overset{\theta}{B}$ is then an upper triangular matrix and the summations need only extend to the diagonal elements to process all the non-zero elements of each column.

Conover shift product: To minimize the problems associated with the double definition of the element of the Conover shift product (2.1) we introduce a degenerate function as follows:

$$\gamma_{i,j} = 1 \text{ if } i \geq j \qquad = 0 \text{ otherwise} \tag{2.7}$$

Then the matrix $\{\gamma_{i,j}\}$ is lower triangular with unity on and below the main diagonal.

We will also have uses for Kronecker's delta function (Hohn [1958], namely;

$$\delta_{i,j} = 1 \text{ if } i = j \qquad = 0 \text{ otherwise}$$

Then the general element of D = A * B where A and B are both infinite in both dimensions is given by:

$$d_{i,j} = \gamma_{i,j} \sum_{k=0}^{\infty} a_{i-j,k} b_{k,j} \qquad \begin{cases} i = 0, 1, 2, \ldots \\ j = 0, 1, 2, \ldots \end{cases} \tag{2.8}$$

Here we find D to be a lower triangular matrix.

However, if B is infinite in both dimensions while A is an infinite row vector as before, then the general element of D = A * B becomes:

$$d_{i,j} = \delta_{i,j} \sum_{k=0}^{\infty} a_k b_{k,j} \qquad \begin{cases} i = 0, 1, 2, \ldots \\ j = 0, 1, 2, \ldots \end{cases} \tag{2.9}$$

where in this case D is not a row vector, but rather a diagonal matrix.

The subscript on a in (2.9) is k rather than a function i or j because the ordinary matrix product of A and B is a row vector which the shifting simply expands into a diagonal matrix. Thus the moving choice of rows of A implied by the subscript i-j on a in (2.8) does not apply here.

2.4 Relation to convolutions

The purpose of the shift operators that have been presented in this chapter is not unlike the purpose of the well known convolution operator as used, for example, by Feller [1965]. The similarity between the simple form of the convolution operator and the general form of a shift product element is immediately striking. For this reason it may be of value to pursue this comparison with the intention of relating these two concepts, and by this means, further clarifying the operation performed by a shift-multiplication. The reader is referred to appendix II for this comparison.

CHAPTER 3

SUMS OF STATES OF MARKOV CHAINS

3.1 Introduction

3.1.1 Background and content

The derivation of the probability distribution of the partial sum of the first r states of a non-homogeneous Markov chain with a finite state space was first obtained by Conover [1965]. Later Gani [1965], using his form of shift multiplication, asserted a similar expression which was more compact. Unfortunately his form required sequential shift multiplication from the right which presents a mathematical complication. Gani did not prove his assertion in his paper.

This chapter will then present these results with a proof of Gani's assertion for completion. Some related results which were found in the course of the preparation of this work will be presented and the infinite state space case will be mentioned.

3.1.2 Notation

Consider a realisation $\omega = \{y_1, y_2, \ldots\}$ of a non-homogeneous Markov chain $\{Y_t\}$. Let $I_t = f(Y_t)$, $t = 1, 2, \ldots$ be a one to one mapping of the states of the Markov chain onto the set $A = \{a + bk\}$, $k = 0, 1, 2, \ldots$ where a and b are real constants. As is customary with such chains (Parzen [1962]) this work will, without loss of generality, deal simply with the function I_t where $a = 0$ and $b = 1$.

Define the transition probabilities $p^t_{i,j} = P(I_t = j/I_{t-1} = i)$ for $i = 0, 1, 2, \ldots$ and $j = 0, 1, 2, \ldots$ and let $P_t = \{p^t_{i,j}\}$ be the transition matrix. Define further the initial probabilities

p_i^1 as $p_i^1 = P(I_1 = i)$, $i = 0, 1, 2, \ldots$ and let $R_1 = \{p_i^1\}$ be the probability distribution of the initial state expressed as a row vector.

It is now the purpose of this chapter to discuss the probability distribution of the partial sum of the first r states, namely S_r where:

$$S_r = \sum_{t=1}^{r} I_t \qquad r = 1, 2, \ldots$$

Let the probabilities p_i^r be defined by: $p_i^r = P(S_r = i)$ for $i = 0, 1, 2, \ldots$ and $r = 1, 2, \ldots$ and also let $R_r = \{p_i^r\}$ be the probability distribution of S_r.

3.2 <u>The known results</u>

3.2.1 The assumption of a bounded state space

Solutions have been obtained for R_r by both Conover and Gani in the case where the state space is bounded by a finite quantity, s such that $I_r \leq s$ for all positive integers r. They are presented in this paragraph.

3.2.2 Conover's form

Given the notation of paragraph 3.1.2 it was proved by Conover [1965] that:

$$R_m' = [R_1 * I * P_2 * P_3 * \ldots * P_m]\, E \tag{3.1}$$

where the shift multiplications are understood to be performed sequentially from the left and where E is a unit column vector of dimension (ms + 1) x 1.

Example 3.1. Consider a homogeneous Markov chain with the state space {0,1,2} and transition matrix:

$$P = \begin{vmatrix} p_{0,0} & p_{0,1} & p_{0,2} \\ p_{1,0} & p_{1,1} & p_{1,2} \\ p_{2,0} & p_{2,1} & p_{2,2} \end{vmatrix}$$

with initial probability distribution $R_1 = [p_0, p_1, p_2]$

Then if we wish to obtain R_2 by applying (3.1) we find: $R_2' = [R_1 * I * P] E$

The equation is evaluated as follows:

$$R_1 * I = \begin{vmatrix} p_0 & 0 & 0 \\ 0 & p_1 & 0 \\ 0 & 0 & p_2 \end{vmatrix}$$

so that

$$(R_1 * I)P = \begin{vmatrix} p_0p_{0,0} & p_0p_{0,1} & p_0p_{0,2} \\ p_1p_{1,0} & p_1p_{1,1} & p_1p_{1,2} \\ p_2p_{2,0} & p_2p_{2,1} & p_2p_{2,2} \end{vmatrix}$$

and so

$$R_1 * I * P = \begin{vmatrix} p_0p_{0,0} & 0 & 0 \\ p_1p_{1,0} & p_0p_{0,1} & 0 \\ p_2p_{2,0} & p_1p_{1,1} & p_0p_{0,2} \\ 0 & p_2p_{2,1} & p_1p_{1,2} \\ 0 & 0 & p_2p_{2,2} \end{vmatrix}$$

Thus we have

$$R_2' = \begin{vmatrix} p_0p_{0,0} \\ p_1p_{1,0} + p_0p_{0,1} \\ p_2p_{2,0} + p_1p_{1,1} + p_0p_{0,2} \\ p_2p_{2,1} + p_1p_{1,2} \\ p_2p_{2,2} \end{vmatrix}$$

which is intuitively acceptable.

3.2.3 Gani's form

Given the notation of paragraph 3.1.2 it was stated by Gani [1965] that R_m, the probability distribution of S_m, is given by:

$$R_m = R_1 \ominus P_2 \ominus P_3 \ominus \ldots \ominus P_m \tag{3.2}$$

Here the shift multiplications are understood to be performed sequentially from the right hand side.

Example 3.2. If we wish to repeat example 3.1 using equation (3.2) we find: $R_2 = R_1 \ominus P$ where

$$R_1 \overset{\ominus}{P} = [p_0, p_1, p_2] \begin{vmatrix} p_{0,0} & p_{0,1} & p_{0,2} & 0 & 0 \\ 0 & p_{1,0} & p_{1,1} & p_{1,2} & 0 \\ 0 & 0 & p_{2,0} & p_{2,1} & p_{2,2} \end{vmatrix}$$

which gives:

$$[p_0p_{0,0}, p_0p_{0,1} + p_1p_{1,0}, p_0p_{0,2} + p_1p_{1,1} + p_2p_{2,0}, p_1p_{1,2} + p_2p_{2,1}, p_2p_{2,2}]$$

This result is in agreement with that obtained in example 3.1.

3.2.4 Proof of Gani's form.

Mathematical induction will be used to prove (3.2).

Consider $P_{r-1} \ominus P_r = \{P(I_{r-1} = j \,/\, I_{r-2} = i\} \ominus \{P(I_r = j \,/\, I_{r-1} = i)\}$

where $i = 0, 1, 2, \ldots, s$ and $j = 0, 1, 2, \ldots, s$.

The shifting of the rows of P_r results in:

$$P_{r-1} \ominus P_r = \{P(I_{r-1} = j \,/\, I_{r-2} = i)\} \cdot \{P(I_r = j-i \,/\, I_{r-1} = i)\}$$

$$= \{ \sum_{k=0}^{s} P(I_{r-1} = k \,/\, I_{r-2} = i) P(I_r = j-k \,/\, I_{r-1} = k)\}$$

Due to the Markov property we may write this as:

$$= \{ \sum_{k=0}^{s} P(I_{r-1} = k \,/\, I_{r-2} = i) P(I_r = j-k \,/\, I_{r-1} = k,\, I_{r-2} = i) \}$$

$$= \{ \sum_{k=0}^{s} \frac{P(I_{r-1} = k,\, I_{r-2} = i)}{P(I_{r-2} = i)} \cdot \frac{P(I_r = j-k,\, I_{r-1} = k,\, I_{r-2} = i)}{P(I_{r-1} = k,\, I_{r-2} = i)} \}$$

$$= \{ \sum_{k=0}^{s} P(I_r = j-k,\, I_{r-1} = k \,/\, I_{r-2} = i) \}$$

$$= \{ P(I_r + I_{r-1} = j \,/\, I_{r-2} = i) \} \tag{3.3}$$

So assuming that:

$$P_r \,\theta\, . . . \,\theta\, P_m = \{ P(I_m + I_{m-1} + . . . + I_r = j \,/\, I_{r-1} = i) \} \tag{3.4}$$

where $r \leq m$, it can be shown that:

$$P_{r-1} \,\theta\, . . . \,\theta\, P_m = \{ P(I_m + . . . + I_{r-1} = j \,/\, I_{r-2} = i) \}$$

Following the same procedure as that used to obtain (3.3) we get:

$$P_{r-1} \,\theta\, (P_r \,\theta\, \dots \,\theta\, P_m) = \{ P(I_{r-1} = j \mid I_{r-2} = i) \} \,\theta\, \{ P(I_m + \dots + I_r = j / I_{r-1} = i) \}$$

$$= \{ P(I_{r-1} = j \,/\, I_{r-2} = i) \} \cdot \{ P(I_m + \dots + I_r = j-i \,/\, I_{r-1} = i) \}$$

$$= \{ \sum_{k=0}^{s} P(I_{r-1} = k \,/\, I_{r-2} = i) P(I_m + \dots + I_r = j-k \mid I_{r-1} = k) \}$$

$$= \{ P(I_m + \dots + I_{r-1} = j \,/\, I_{r-2} = i) \} \tag{3.5}$$

Thus since (3.3) is true we can use (3.5) to proceed, step by step to expand until we have shown (3.4) true for any value of $m = 1, 2, \ldots$ and $r = 1, 2, \ldots, m$. Thus by mathematical induction we have that (3.4) is true for any such m and r. So that we have:

$$P_2 \ominus \ldots \ominus P_m = \{P(I_2 + I_3 + \ldots + I_m = j \,/\, I_1 = i)\}$$

and $R_1 \ominus P_2 \ominus \ldots \ominus P_m$ can be written as:

$$\{P(I_1 = i)\} \ominus \{P(I_2 + I_3 + \ldots + I_m = j \,/\, I_1 = i)\}$$

$$= \{P(I_1 = i)\} \cdot \{P(I_2 + \ldots + I_m = j-i \,/\, I_1 = i)\}$$

$$= \{\sum_{k=0}^{s} P(I_1 = k)P(I_2 + \ldots + I_m = j-k \,/\, I_1 = k)\}$$

$$= \{\sum_{k=0}^{s} P(I_2 + I_3 + \ldots + I_m = j-k,\ I_1 = k)\}$$

$$= \{P(I_1 + I_2 + \ldots + I_m = j)\}$$

$$= R_m$$

Since P_i has dimension $(s+1) \times (s+1)$ it follows that $P_i \ominus P_{i+1}$ has dimension $(s+1) \times (2s+1)$, and finally that $P_2 \ominus P_3 \ominus \ldots \ominus P_m$ has the dimension $(s+1) \times ([m-1]s+1)$ so that $R_1 \ominus P_2 \ominus \ldots \ominus P_m$ has the dimension given by $1 \times (ms+1)$.

This completes the proof of (3.2). An attempt has been made to relate (3.2) and (3.1) directly but because of the difference in direction of operation for the two equations only a complete evaluation would suffice. This would be equivalent to the proof above combined with Conover's published proof.

3.3 An alternative

The probability distribution of S_r can be obtained in an alternative form which does not involve shift operations. The advantage of such a form lies in algebraic convenience of ordinary matrix multiplication. Particularly its familiarity and associativity are desirable. Unfortunately, however, this alternative form has associated with it increased complexity. Also, this form cannot be generalized to include the infinite state space case.

To obtain the same end result as that of (3.1) or (3.2) in any alternative form it is necessary to define special matrices which, with ordinary matrix multiplication, will result in a shifting process analogous to that performed by the shift operator. Given a Markov chain such as that defined in 3.1.2, with the state space bounded by s as in 3.2.1 a basic matrix which can be used to perform this shifting process is given by the $[(r-1)s+1] \times (rs+1]$ matrix U_k^r:

$$U_k^r = [Z_k^1 \;\vdots\; I \;\vdots\; Z_k^2] \qquad \begin{cases} r = 1, 2, 3, \ldots \\ k = 0, 1, 2, \ldots, s \end{cases}$$

where Z_k^1 and Z_k^2 are zero matrices of dimensions $[(r-1)s+1] \times k$ and $[(r-1)s+1] \times [s-k]$ respectively. Here U_k^r is a matrix whose size is determined by r, the iteration number, and which contains the identity matrix placed after the k^{th} column of zeros. Hence, for example, in the case where the state space is the set $\{0,1,2,3\}$:

$$U_2^2 = \left|\begin{array}{cc|cccc|c} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right|$$

while

$$U_3^1 = [0 \quad 0 \quad 0 \quad 1]$$

Now from these matrices the iteration matrices A^r can be constructed using the transition probabilities $\{p_{i,j}^r\}$ of the Markov chain:

$$A^r = \left|\begin{array}{c|c|c} p_{0,0}^{r+1} \quad U_0^r & \cdots & p_{0,s}^{r+1} \quad U_0^r \\ \hline \hline \vdots & \vdots & \vdots \\ \hline p_{s,0}^{r+1} \quad U_s^r & \cdots & p_{s,s}^{r+1} \quad U_s^r \end{array}\right|$$

Here A^r has dimension $(s+1)[(r-1)s+1] \times (s+1)(rs+1)$ so that A^r becomes an expanded shifted form of the transition probability matrix P_{r+1} filled in with zeros.

Lastly there is a need for an accumulator matrix U^r defined as follows:

$$U^r = \left|\begin{array}{c} U_0^r \\ \hline \hline \vdots \\ U_s^r \end{array}\right|$$

of dimension $(s+1)[(r-1)s+1] \times [rs+1]$

Then the probability distribution of S_m is:

$$R_m = R_1 A^1 A^2 \ldots A^{m-1} U^m \qquad (3.6)$$

The proof of this, requiring results from appendices I and II, will be left for appendix III.

Example 3.3 If we wish to repeat examples (3.1) and (3.2) for equation (3.6) we have $R_2 = R_1 A^1 U^2$

where

$$U^2 = \left|\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right|$$

$$A^1 = \left|\begin{array}{ccc|ccc|ccc} p_{0,0} & 0 & 0 & p_{0,1} & 0 & 0 & p_{0,2} & 0 & 0 \\ \hline 0 & p_{1,0} & 0 & 0 & p_{1,1} & 0 & 0 & p_{1,2} & 0 \\ \hline 0 & 0 & p_{2,0} & 0 & 0 & p_{2,1} & 0 & 0 & p_{2,2} \end{array}\right|$$

So that $R_1 A^1 =$

$[p_0\, p_{0,0}, p_1\, p_{1,0}, p_2\, p_{2,0}\; p_0\, p_{0,1}, p_1\, p_{1,1}, p_2\, p_{2,1}\; p_0\, p_{0,2}, p_1\, p_{1,2}, p_2\, p_{2,2}]$

thus $R_1 A^1 U^2 =$

$$[p_0\, p_{0,0}, p_1\, p_{1,0} + p_0\, p_{0,1}, p_2\, p_{2,0} + p_1\, p_{1,1} + p_0\, p_{0,2}, p_2\, p_{2,1} + p_1\, p_{1,2}, p_2\, p_{2,2}]$$

which is in agreement with (3.1) and (3.2).

3.4 Sums of states of Markov chains with infinite state spaces

Using the definitions of the shift products for matrices with infinite dimensions, the case of the Markov chain whose state space is not bounded as in paragraph 3.2.1, can still be encompassed by the results using the shift operators (3.1) and (3.2). The proofs follow closely the proofs for the finite state case and may be found in appendix III.

Unfortunately, however, the alternative form of paragraph 3.3 cannot be generalized to include the infinite state case. This is because the alternative form involves the matrices A^r and U^r which are formed from submatrices. In the infinite state case these submatrices would have to be infinite, which would result in an impossible structure.

3.5 Sums of states of Markov chains as Markov chains

3.5.1 Introduction

Consider a realization $\{y_1, y_2, \ldots\}$ of a stochastic process $\{Y_t;\ t = 1, 2, \ldots\}$ defined on the countably infinite sample space $\{E_0, E_1, E_2, \ldots\}$ such that a probability measure of the form $p_r^k = P(Y_k = E_r)$ is defined for $k = 1, 2, \ldots$ and $r = 0, 1, 2, \ldots$, where the events $\{E_0, E_1, E_2, \ldots\}$ are mutually exclusive while the random variables Y_t, $t = 1, 2, 3, \ldots$ are statistically mutually independent.

Such a process may be regarded as a Markov chain in which the state space E can be mapped by a one to one mapping onto the set of non-negative integers. Let $I_k = f(Y_k)$ be such a mapping. Let this mapping be chosen such that

$$p_r^k = P(I_k = r) \qquad \begin{cases} k = 1, 2, \ldots \\ r = 0, 1, 2, \ldots \end{cases}$$

Then the transition probabilities, $p_{i,j}^k$ of this Markov chain are given by:

$$\begin{aligned} p_{i,j}^k &= P(I_k = j \,/\, I_{k-1} = i) \\ &= P(I_k = j) \\ &= p_j^k \end{aligned} \qquad \begin{cases} i = 0, 1, 2, \ldots \\ j = 0, 1, 2, \ldots \\ k = 1, 2, \quad \ldots \end{cases}$$

Now we have a realization of a Markov chain $\{I_1, \ldots\}$ where all the I_t, $t = 1, 2, \ldots$ are mutually independent. This being so, we may observe that the sum of the first r of these:

$$S_r = \sum_{k=1}^{r} I_k \qquad r = 1, 2, \ldots$$

forms a stochastic process $\{S_1, S_2, \ldots\}$ which is a Markov chain. This can be seen as follows:

$$P(S_r = j \,/\, S_{r-1} = i_{r-1}, S_{r-2} = i_{r-2}, \ldots, S_1 = i_1)$$

$$= P(S_r = j \,/\, I_1 = i_1, I_2 = i_2 - i_1, \ldots, I_{r-1} = i_{r-1} - i_{r-2})$$

which is equivalent to:

$$= P(I_r = j - i_{r-1} \,/\, I_1 = i_1, I_2 = i_2 - i_1, \ldots, I_{r-1} = i_{r-1} - i_{r-2})$$

which, by the independence assumption, gives:

$$= P(I_r = j - i_{r-1})$$

so that we may write:

$$= P(S_r = j \;/\; S_{r-1} = i_{r-1})$$

Thus if the states of a Markov chain are mutually independent then the partial sums of states also form a Markov chain. The reverse of this is not proved. The above is merely a setting forth of a very common special case of a Markov chain. (Parzen, 1962)

3.5.2 The transition matrices in a special case

Under the conditions of paragraph 3.5.1, that is, if the partial sums of states of a Markov chain form a Markov chain, then the transition matrices of the new Markov chain bear a simple relationship to the transition matrices of the original Markov chain.

Let, for $p^k_{i,j}$ the transition probabilities as in paragraph 3.5.1, the transition matrices of the Markov chain be $T_k = \{p^k_{i,j}\}$.

Then let P_k be the transition matrices of the Markov chain formed by the partial sums of states:

$$P_k = \{P(S_{k+1} = j \;/\; S_k = i)\} \qquad k = 1, 2, \ldots$$

Then it can be shown that

$$P_k = (I * T_k')' \tag{3.7}$$

Since the proof of (3.7) requires some results from appendix I, it will be left for appendix II. It is, however, intuitively obvious as can be seen from the following example.

Example 3.4: The content of these last two paragraphs is essentially simple and their content becomes immediately clear if (3.7) is written out.

The transition probabilities of the original independent stochastic process from 2.5.1 can be written:

$$p^k_{i,j} = p^k_j \qquad \begin{cases} i = 0, 1, 2, \ldots \\ j = 0, 1, 2, \ldots \\ k = 1, 2, \quad \ldots \end{cases}$$

so that

$$T_k = \begin{vmatrix} p^k_0 & p^k_1 & p^k_2 & \cdots \\ p^k_0 & p^k_1 & p^k_2 & \cdots \\ p^k_0 & p^k_1 & p^k_2 & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{vmatrix}$$

thus by (3.7)

$$P_k = \begin{vmatrix} p^k_0 & p^k_1 & p^k_2 & \cdots \\ 0 & p^k_0 & p^k_1 & \cdots \\ 0 & 0 & p^k_0 & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{vmatrix}$$

where validity is intuitively obvious.

3.5.3 An example - a Poisson counting process

Consider a stochastic process $\{N_t;\ t = 1, 2, \ldots\}$ where each N_t is distributed independently according to the Poisson probability law. That is:

$$P(N_t = x) = \frac{e^{-\nu}\,\nu^x}{x!} \qquad x = 0, 1, 2, \ldots \tag{3.8}$$

$$= 0 \qquad \text{otherwise}$$

It is known then that the stochastic process $\{S_1, S_2, \ldots\}$ where $S_r = \sum_{k=1}^{r} N_k$ is a Poisson counting process with

$$P(S_r = x) = \frac{e^{-r\nu}(r\nu)^x}{x!} \qquad x = 0, 1, 2, \ldots \tag{3.9}$$

This follows from the reproductivity of the Poisson distribution (Wilks, 1962).

It may be noted, however, that any of the equations (3.1), (3.2), or (3.6) would agree with this result. For example consider (3.2). Observe that the above process $\{N_1, N_2, \ldots\}$ qualifies as in paragraph 3.5.1 as a Markov chain. It has, setting $P(N_t = x) = p_x^t$; $t = 1, 2, \ldots$; $x = 0, 1, 2, \ldots$, the initial probability distribution:

$$R_1 = [p_0^1, p_1^1, p_2^1, \ldots]$$

with the stationary transition probabilities given by:

$$p_{i,j} = P(N_t = j \,/\, N_{t-1} = i) \quad \text{for all } t = 2, 3, \ldots$$

$$= P(N_t = j) = p_j^t$$

so that if $T = \{p_{i,j}\}$, then the general element of $T \theta T$ is given by $t_{i,j}^2$ using (2.5):

$$t_{i,j}^2 = \sum_{k=0}^{j} p_k \, p_{j-k} \qquad \begin{cases} i = 0, 1, 2, \ldots \\ j = 0, 1, 2, \ldots \end{cases}$$

$$= \sum_{k=0}^{j} \frac{e^{-\nu} \nu^k}{k!} \cdot \frac{e^{-\nu} e^{j-k}}{(j-k)!}$$

$$= e^{-2\nu} \nu^j \sum_{k=0}^{j} \frac{1}{k!(j-k)!}$$

which, using the binomial expansion gives:

$$= \frac{e^{-2\nu}(2\nu)^j}{j!}$$

Continuing in this manner evaluation of $T \theta T \theta \ldots \theta T$ to (r-1) terms results in the general element:

$$t^{r-1}_{i,j} = \frac{e^{-(r-1)\nu}([r-1]\nu)^j}{j!} \qquad \begin{cases} i = 0, 1, 2, \ldots \\ j = 0, 1, 2, \ldots \end{cases}$$

then, using (2.6) we find that $R_1 \theta (T \theta T \theta \ldots \theta T)$ results in (3.9) as expected.

3.5.4 An example - flooding models

Gani [1965] has suggested that the inputs supplied to a river by its tributaries during a flood season may be Markovian. Under this assumption (3.1) or (3.2) will supply the probability distribution of the flow in the main stream between the r^{th} and $r+1^{st}$ tributary junction directly. It is necessary for assumptions (i) to (iv) of Gani's flooding model (paragraph 1.4) to hold.

Let $\{I_1, I_2, I_3, \ldots\}$ be the Markov chain describing the tributary inputs. Let $D_1 = \{P(I_1 = i)\}$ be the probability distribution of the source stream and let T_r be the transition matrix from the tributary (r-1) to tributary r. Then

$$S_m = \sum_{r=1}^{m} I_r$$

is the flow in the main stream between the junction of tributary m and tributary (m+1). Let $R_m = \{P(S_m = i)\}$ be the probability distribution of S_m. Then R_m is given by (3.1) or (3.2). If the flow in all tributaries is bounded by a maximum value, say s, then (3.6) could be used.

CHAPTER 4

A FILTERING PROCESS

4.1 Introduction

4.1.1 A definition of filtering

A filtering process is not a concept which is standard throughout the literature. The term as it is used in this work is different from that used in at least one major reference (Parzen [1962]). In this work the term is used because it seems to be intuitively appropriate. One may consider a 'black-box' type of process where input, consisting of one realization of a random variable, results in an output which is, in turn, one realization of another random variable whose distribution is determined by the input. Then the output may be taken to be a filtering of the input. More specifically, the following definition will be used:

DEFINITION III: Let $\{I_t; t = 1, 2, . . .\}$ and $\{0_t; t = 1, 2, . . .\}$ be discrete parameter stochastic processes defined on the state space $\{0, 1, 2, . . .\}$. Then $\{0_t\}$ will be said to be a filtering of $\{I_t\}$ if 0_t is a discrete random variable with

$$P(0_t = j/I_t = i) = P(0_t=j/I_t=i,0_{t-1}=k) \qquad \begin{cases} i,j,k = 0, 1, 2, . . . \\ t = 2, 3, . . . \end{cases} \qquad (3.10)$$

Thus 0_t, given I_t, is not dependent upon the output, 0_{t-1}, of the previous stage. For notation convenience let

$$q^t_{i,j} = P(0_t = j / I_t = i)$$

It may be noted here that 0_t is not independent of 0_{t-1}, simply that given the value of I no further information is given by the value of 0_{t-1}.

It will also be necessary in paragraph 4.4.3 to assume

$$P(O_t = j/I_t=i) = P(O_t=j/I_t=i, I_{t+1}=k) \qquad \begin{cases} i,j,k = 0,1,2,\ldots \\ t = 1,2,\ldots \end{cases} \tag{3.11}$$

This last requirement will be of some significance in deterministic filtering. This type of filtering is one which, it is felt, may have wide application. There follow a few examples.

(i) If we consider, as input, the arrival process of a queue, then losses and defection (refusing to join the queue) (Cohen, 1957, et. al.) can be handled by a filtering of this class, to produce as output the resultant stochastic process of components that actually do join the queue. This type of situation could arise, for example, when fresh commodities are subject to spoiling losses before sale.

(ii) If we consider as input the stochastic process describing the length of a queue then reneging (impatiently leaving) (Palm, 1937, et. al.) as well as some service processes, both independent and not, could be described in terms of a filtering of this class.

(iii) Tandem service procedures arise very frequently (Burke, 1956, et. al.) for example in traffic problems, sequential processing situations and in multiple stage storage procedures. In such situations the input to a stage is the output from the previous stage and this type of process could also be described in terms of a filtering of the type used here, provided that requirement (3.11) was not necessary in the intended application.

Much work has been done for situations like these. The reader is referred to Saaty (1961) for bibliography of related work.

4.1.2 Notation

Define the following matrices for t = 1, 2, . . .:

The probability distribution of the input I_t as the row:

$$D_t = \{d_i^t = P(I_t = i)\} \qquad i = 0, 1, 2, \ldots$$

The probability distribution of the output 0_t as the row:

$$A_t = \{a_i^t\} = P(0_t = i)\} \qquad i = 0, 1, 2, \ldots$$

The response matrix of the filtering:

$$Q_t = \{q_{i,j}^t = P(0_t = j \mid I_t = i)\} \qquad \begin{cases} i = 0, 1, 2, \ldots \\ j = 0, 1, 2, \ldots \end{cases}$$

Then it is easily verified that $A_t = D_t\, Q_t$ (4.2)

The general element of $D_t\, Q_t$ is given by:

$$\sum_{k=0}^{\infty} d_k^t\, q_{k,i}^t = \sum_{k=0}^{\infty} P(I_t = k)\, P(0_t = i \;/\; I_t = k)$$

$$= P(0_t = i)$$

$$= a_i^t$$

For Markov chains the notation of paragraph 3.1.2 will be used.

4.1.3 A hypothetical example

Consider a situation where a service is offered repeatedly at discrete times. An example might be a cinema with showings at times τ_t; $t = 1, 2, \ldots$. Assume that the probability distribution, R_t, of the number of persons aware of the service at time τ_t is obtainable. If, for example, the persons became aware independently at a constant rate, ν, then R_t in (3.9) of the example in paragraph 3.5.3 would supply the required distribution. Then one may regard the persons who, being aware of the service at time τ_t, do not make use of it at time τ_t as defectors.

The defection probability is frequently taken as a function of time, in this case perhaps of the number of showings, t. Not uncommonly (Riordan, 1953) the defection probability is taken to be negative exponential with the rate $\lambda\tau_t$. That is:

$$P(\text{defection at time } \tau_t) = e^{-\lambda(\tau_{t-1}-\tau_1)} - e^{-\lambda(\tau_t-\tau_1)} \qquad \begin{cases} \lambda > 0 \\ \tau_t \geq 0 \end{cases}$$

Then for a given time the defection probability is constant and can be used to obtain the probability of i-j defectors out of i persons by the binomial probability law. These probabilities $\{q^t_{i,j};\ i,j = 0,1,2,\ldots$ and $t = 1,2,\ldots\}$ used as elements of the matrix Q_t form a response matrix for the filtering applied by the defection process. Then $R_t\, Q_t$ is the probability distribution of the number of persons making use of the service at time τ_t. If one defined a column vector, F, such that it had its first M elements unity and the rest zero where M is the number of customers required for a success then P(t^{th} performance a success)=1-R_tQ_tF

4.2 Deterministic filtering of Markov chains

4.2.1 Introduction

Let $\{I_1, I_2, \ldots\}$ be a Markov chain as in paragraph 3.1.2. Then define a deterministic filtering as one such that, for all $I_t = 0,1,2,\ldots$ and $t = 1,2,\ldots$

$$P(O_t = j \,/\, I_t = i) = 1 \quad \text{for some non-negative integer } j$$
$$= 0 \quad \text{otherwise}$$

In this case then the output process is simply a 'many to one' or a 'one to one' mapping of the state space {0, 1, 2, . . .} onto itself. Thus the response matrix Q_t has, in each row, all zeros except for one unit element.

Gani [1965] applied this particular case to his flooding model but there is some doubt as to the validity of one of his results (paragraph 4.2.2).

It is obvious here that the requirement (3.11) is fulfilled by a filtering of this class.

4.2.2. A disagreement with Gani

Further to the notation of paragraph 4.1.2 let R_m be the probability distribution, expressed as a row vector, of the partial sums of outputs:

$$\rho_m = \sum_{t=1}^{m} O_t \qquad r = 1, 2, \ldots .$$

Let T_n be the matrix of transition probabilities $p^n_{i,j}$ of the Markov chain $\{I_1, I_2, \ldots\}$:

$$p^n_{i,j} = P(I_n = j \,/\, I_{n-1} = i) \qquad \begin{cases} i,j = 0,1,2,\ldots \\ n = 2,3,\ldots \end{cases}$$

Then Gani states that

$$R_m = D_1Q_1 \,\theta\, T_2Q_2 \,\theta\, \ldots \,\theta\, T_mQ_m \tag{4.3}$$

A simple example follows which will show that this interpretation of Gani's result is not acceptable. Under the assumption that $\{O_1, O_2, \ldots\}$ forms a Markov chain it has been found possible to obtain the transition matrices of this chain in terms of T_n and Q_n and thus, by (3.1), (3.2) or (3.6) obtain an expression for R_m. To date however the conditions under which $\{O_1, O_2, \ldots\}$ does form a Markov chain have not been investigated.

Example 4.1

Take $D_1 = [d_0, d_1, d_2]$

and

$$T_1 = T_2 = \begin{vmatrix} p_{0,0} & p_{0,1} & p_{0,2} \\ p_{1,0} & p_{1,1} & p_{1,2} \\ p_{2,0} & p_{2,1} & p_{2,2} \end{vmatrix}$$

with

$$Q_1 = Q_2 = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{vmatrix}$$

Applying (4.3) we have

$$T_2Q_2 = \begin{vmatrix} p_{0,0} & p_{0,1}+p_{0,2} & 0 \\ p_{1,0} & p_{1,1}+p_{1,2} & 0 \\ p_{2,0} & p_{2,1}+p_{2,2} & 0 \end{vmatrix}$$

and

$$D_1Q_1 = [d_0,\ d_1+d_2,\ \ 0]$$

thus

$$R_2 = D_1Q_1 \theta T_2Q_2 = [d_0,\ d_1+d_2,\ 0] \begin{vmatrix} p_{0,0} & p_{0,1}+p_{0,2} & 0 & 0 & 0 \\ 0 & p_{0,1} & p_{0,1}+p_{1,2} & 0 & 0 \\ 0 & 0 & p_{2,0} & p_{2,1}+p_{2,2} & 0 \end{vmatrix}$$

$$= [d_0p_{0,0}, d_0(p_{0,1}+p_{0,2})+p_{1,0}(d_1+d_2), (d_1+d_2)(p_{1,1}+p_{1,2}), 0, 0]$$

which disagrees with the anticipated result, namely:

$$R_2 = [d_0p_{0,0}, d_0(p_{0,1}+p_{0,2})+d_1p_{1,0}+d_2p_{2,0}, d_1(p_{1,1}+p_{1,2})+d_2(p_{2,1}+p_{2,2}), 0, 0] \tag{4.4}$$

The latter form is intuitively valid, for example take the second element, $P(S_2 = 1)$. Observing Q we have:

$$\begin{aligned} P(S_2 = 1) &= P(O_1 + O_2 = 1) \\ &= P(O_1 = 0,\ O_2 = 1) + P(O_1 = 1,\ O_2 = 0) \\ &= P(I_1 = 0,\ I_2 = 1 \text{ or } 2) + P(I_1 = 1 \text{ or } 2,\ I_2 = 0) \\ &= (d_0p_{01} + d_0p_{02}) + (d_1p_{10} + d_2p_{20}) \end{aligned}$$

as in (4.4).

4.3 Filtering the sums of states of a Markov chain

4.3.1 Introduction

If $\{I_1, I_2, \ldots\}$ is a Markov chain and S_r is the partial sum of the first r states of this chain as in paragraph 3.1.2, then this section presents some consequences of filtering (in the sense of paragraph 4.1.1) the stochastic process $\{S_1, S_2, \ldots\}$ as input.

A solution for the probability distribution of the output of the filter, generalized slightly from a result from Gani, will be given and proved.

It is not necessary in this section of the work to assume that the filtering is of a deterministic nature. The only limitations are those mentioned in definition III, namely that both the input and output processes be defined on the state space $\{0, 1, 2, \ldots\}$ and that the output given the input does not depend on the previous output.

4.3.2 Notation

Let $\{I_1', I_2', \ldots\}$ be a Markov chain with state space $\{0,1,2,\ldots,s\}$ such as that defined in paragraph 3.1.2, with the initial distribution

$R_1 = \{p_i^1\}$ and with the transition matrices $P_t = \{p_{i,j}^t\}$. Define $I_r = 0_{r-1} + I_r'$; $r = 1, 2, \ldots$. I_r is now the filtered sum of the states $1, 2, \ldots, r-1$ plus state r. That is, a filtering is applied to every sum of states (The lack of filtering is considered as a special case of filtering). As before $R_r = \{p_i^r\}$ is the probability distribution of I_r.

Now let the response matrix of the r^{th} filtering be $Q_r = \{q_{i,j}^r\}$ where

$$q_{i,j}^r = P(0_r = j \,/\, 0_{r-1} + I_r' = i \qquad \begin{cases} i = 0, 1, 2, \ldots, s \\ j = 0, 1, 2, \ldots, s \\ r = 1, 2, \quad \ldots \end{cases}$$

in keeping with the definition of filtering. Here Q_r is [rs+1] x [rs+1].

Now let

$$u_{i,j,(k)}^r = P(0_r = j \,/\, 0_{r-1} = i, I_r' = k) \quad k = 0, 1, 2, \ldots, s$$

$$= P(0_r = j \,/\, 0_{r-1} = i, 0_{r-1} + I_r' = i + k)$$

$$= P(0_r = j \,/\, 0_{r-1} = i, I_r = i + k)$$

so that by the definition of the filtering

$$= P(0_r = j \,/\, I_r = i + k)$$

$$= q_{i+k,j}^r$$

Then define the $[(r-1)\cdot s+1] \times [rs+1]$ matrix $U_k^r = \{u_{i,j,(k)}^r\}$. This matrix is then the submatrix of Q_r:

$$
Q_r = \begin{array}{c}
\\ \\ \\ \\ \\ \\ \\ \\ \\
\\ \\
U_k^r\left\{ \begin{array}{c} \\ \\ \\ \\ \\ \\ \end{array} \right. \\
\\ \\ \\ \\ \\
\end{array}
\left|\begin{array}{ccccccc}
q_{0,0} & q_{0,1} & q_{0,2} & \cdot & \cdot & \cdot & q_{0,rs} \\
q_{1,0} & q_{1,1} & q_{1,2} & \cdot & \cdot & \cdot & q_{1,rs} \\
\cdot & & & & & & \cdot \\
\cdot & & & & & & \cdot \\
\cdot & & & & & & \cdot \\
\hline
q_{k,0} & q_{k,1} & q_{k,2} & \cdot & \cdot & \cdot & q_{k,rs} \\
\cdot & & & & & & \cdot \\
\cdot & & & & & & \cdot \\
\cdot & & & & & & \cdot \\
q_{(r-1)s+k,0} & \cdot & \cdot & \cdot & \cdot & \cdot & q_{(r-1)s+k,rs} \\
\hline
\cdot & & & & & & \cdot \\
\cdot & & & & & & \cdot \\
\cdot & & & & & & \cdot \\
q_{rs,0} & \cdot & \cdot & \cdot & \cdot & \cdot & q_{rs,rs}
\end{array}\right|
$$

Now define a matrix A_r in terms of the U_k^r and the transition probabilities:

$$
p_{i,j}^r = P(I_r' = j \,/\, I_{r-1}' = i) \qquad \begin{cases} i = 0, 1, 2, \ldots, s \\ j = 0, 1, 2, \ldots, s \\ r = 2, \ldots \end{cases}
$$

as follows; if $r \geq 2$:

$$A^r = \begin{vmatrix} p_{0,0}^{r+1}\, U_0^r & p_{0,1}^{r+1}\, U_0^r & \cdot & \cdot & \cdot & p_{0,s}^{r+1}\, U_0^r \\ p_{1,0}^{r+1}\, U_1^r & p_{1,1}^{r+1}\, U_1^r & \cdot & \cdot & \cdot & p_{1,s}^{r+1}\, U_1^r \\ \vdots & \vdots & & & & \vdots \\ p_{s,0}^{r+1}\, U_s^r & p_{s,1}^{r+1}\, U_s^r & \cdot & \cdot & \cdot & p_{s,s}^{r+1}\, U_s^r \end{vmatrix}$$

here A_r is a $(s+1)[(r-1)s+1] \times (s+1)[rs+1]$ matrix which will be used to perform a 'shifting-type' of operation.

If $r = 1$ then the $(s+1) \times (s+1)^2$ matrix A^1 is formed:

$$A^1 = [A_0^1 \quad A_1^1 \quad . \quad . \quad . \quad A_s^1]$$

where

$$A_r^1 = \begin{vmatrix} p_{0,r}^2\, u_{0,0,(0)}^1 & p_{0,r}^2\, u_{0,1,(0)}^1 & \cdot & \cdot & \cdot & p_{0,r}^2\, u_{0,s,(0)}^1 \\ p_{1,r}^2\, u_{0,0,(1)}^1 & p_{1,r}^2\, u_{0,1,(1)}^1 & \cdot & \cdot & \cdot & p_{1,r}^2\, u_{0,s,(1)}^1 \\ \vdots & \vdots & & & & \vdots \\ p_{s,r}^2\, u_{0,0,(s)}^1 & p_{s,r}^2\, u_{0,1,(s)}^1 & \cdot & \cdot & \cdot & p_{s,r}^2\, u_{0,s,(s)}^1 \end{vmatrix}$$

The elements of this matrix may have a direct probabilistic interpretation, namely the i,j^{th} element is:

$$P(I_2' = r \,/\, I_1' = i)\; P(O_1 = j \,/\, I_1' = i)$$

$$= P(I_2' = r \,/\, I_1' = i)\; P(O_1 = j \,/\, I_1' = i,\; I_2' = r)$$

if condition (3.11) is met

$$= P(I_2' = r,\ O_1 = j \ /\ I_1' = i) \qquad (4.5)$$

which will make available probabilities like $P(O_1 + I_2 = j \ /\ I_1 = i)$. The elements of A^r in general are not as clear.

One other matrix needs to be defined as a 'collector-of-terms', namely:

$$U^r = \begin{vmatrix} U_0^r \\ \cdot \\ \cdot \\ \cdot \\ U_s^r \end{vmatrix}$$

which is a (s+1)[(r-1)s+1] x [rs+1] matrix.

4.3.3 A known result

It was stated by Gani that if Q_r is defined as in paragraph 4.3.2, then, in the notation of the previous paragraph:

$$R_m = D_1 A^1 A^2 \ . \ . \ . \ A^{m-1} U^m \qquad (4.6)$$

however, this result will hold for filtering of the type defined in paragraph 4.1.1, provided that condition (3.11) is net.

Proof of (4.6):

The matrices used have definitions which differ in the three cases: $m = 1, 2, \geq 3$ so each will be considered separately.

(i) If $m = 1$:

U_k^1 is a 1 x (s+1) matrix with elements

$$u^1_{0,j,(k)} = P(O_1 = j \ /\ I_1' = k) \qquad \begin{cases} j = 0, 1, 2, \ . \ . \ . \\ k = 0, 1, 2, \ . \ . \ . \end{cases}$$

so that $D_1 U^1$ is 1 x (s+1) by (s+1) x (s+1) thus 1 x (s+1) matrix with elements:

$$\sum_{k=0}^{s} P(0_1 = j \,/\, I_1' = k)\; P(I_1' = k) = P(0_1 = j)$$

thus $D_1 U^1 = R_1$

(ii) If m = 2:

By the definitions and (4.11) $D_1 A_r^1$ is a 1 x (s+1) matrix with elements:

$$\sum_{k=0}^{s} P(I_1' = k)\; P(I_2' = r,\, 0_1 = j \,/\, I_1' = k) \qquad \begin{cases} j = 0, 1, 2, \ldots D \\ r = 0, 1, 2, \ldots D \end{cases}$$

$$= P(I_2' = r,\, 0_1 = j)$$

so that $D_1 A_r^1 U_r^2$ is a 1 x (2s+1) matrix with elements:

$$\sum_{k=0}^{s} P(I_2' = r,\, 0_1 = k)\; P(0_2 = j \,/\, 0_1 = k,\, I_2' = r)$$

$$= P(0_2 = j,\, I_2' = r)$$

Consequently

$$D_1 A^1 U^2 = D_1 [A_0^1 \; A_1^1 \; . \;\; . \;\; . \;\; A_s^1] \begin{vmatrix} U_0^2 \\ . \\ . \\ . \\ U_s^2 \end{vmatrix}$$

$$= \sum_{r=0}^{s} D_1 A_r^1 U_r^2$$

Thus $D_1 A^1 U^2$ is a 1 x (2s+1) matrix with elements:

$$\sum_{r=0}^{s} P(O_2 = j, I_2' = r) = P(I_2' = r)$$

which is an agreement with (4.6).

(iii) If $m \geq 3$:

For simplicity partition $D_1 A^1 \ldots A^n = [\zeta_0^n \quad \zeta_1^n \quad \ldots \quad \zeta_s^n]$ in accordance with the partitioning of A^n. Then ζ_r^n is a $1 \times [ns+1]$ vector. Consider then $\zeta_r^1 = D_1 A_r^1$ $\qquad r = 0, 1, 2, \ldots, s$

$$= \{\sum_{k=0}^{s} P(I_1' = k)\ P(I_2' = r, O_1 = j / I_1' = k)\}$$

$$= \{P(I_2' = r, O_1 = j)\} \qquad \text{(a)}$$

which is a $1 \times (s+1)$ vector. By mathematical induction with (a) as starting point, it will be shown that:

$$\zeta_k^r = \{P(I_{r+1}' = k, O_r = j)\} \qquad k = 0, 1, 2, \ldots, s \qquad \text{(b)}$$

Assume then (b) to hold for some fixed r, and consider ζ_k^{r+1}. From (b) we have the subvectors of $D_1 A^1 \ldots A^r$, consequently, since ζ_k^{r+1} is a subvector of $(D_1 A^1 \ldots A^r)A^{r+1}$ we observe that:

$$\zeta_k^{r+1} = [\sum_{n=0}^{s} p_{n,k}^{r+2}\ \zeta_n^r\ U_n^{r+1}] \qquad \text{(c)}$$

where

$$\zeta_k^r\ U_n^{r+1} = \{\sum_{j=0}^{s} P(I_{r+1}' = n, O_r = j)\ P(O_{r+1} = i / O_r = j, I_{r+1}' = n)\}$$

$$= \{P(I_{r+1}' = k, O_{r+1} = i)\}$$

which is a $1 \times [(r+1)s+1]$ vector. Note that:

$$P(I'_{1+2} = k \;/\; I'_{r+1} = n,\; 0_{r+1} = i) = \frac{P(I'_{1+2} = k,\; 0_{r+1} = i \;/\; I'_{r+1} = n)}{P(0_{r+1} = i \;/\; I'_{r+1} = n)}$$

$$= \frac{P(I'_{r+2} = k,\; 0_{r+1} = i \;/\; I'_{r+1} = n)}{P(0_{r+1} = i \;/\; I'_{r+1} = n,\; I'_{r+2} = k)}$$

$$= \frac{P(I'_{r+1} = n,\; I'_{r+2} = k)}{P(I'_{r+1} = n)} = P(I'_{r+2} = k \;/\; I'_{r+1} = n)$$

So that;

$$p^{r+2}_{n,k}\, \zeta^{r}_{n}\, U^{r+1}_{n} = \{P(I'_{r+2} = k \;/\; I'_{r+1} = n)\; P(I'_{r+1} = n,\; 0_{r+1} = i)\}$$

$$= \{P(I'_{r+2} = k \;/\; I'_{r+1} = n,\; 0_{r+1} = i)\; P(I'_{r+1} = n,\; 0_{r+1} = i)\}$$

$$= \{P(I'_{r+2} = k,\; I'_{r+1} = n,\; 0_{r+1} = i)\} \qquad \text{(d)}$$

Thus by (c) $\zeta^{r+1}_{k} = \{P(I_{r+2} = k,\; 0_{r+1} = i)\}$ which completes the proof of (b).

That being so we may proceed to evaluate:

$$D_1 A^1 \ldots A^{m-1} U^m = [\zeta_0^{m-1} \ldots \zeta_s^{m-1}]U^m$$

$$= [\sum_{k=0}^{s} \zeta_k^{m-1} U_k^m]$$

where

$$\zeta_k^{m-1} U_k^m = \{\sum_{k=0}^{s} P(I'_m = k,\ O_{m-1} = j)\ P(O_m = i/O_{m-1} = j, I'_m = k)\}$$

$$= \{P(I'_m = k,\ O_m = i)\}$$

so that $D_1 A^1 \ldots A^{m-1} U^m$ has elements $P(O_m = i)$ $i = 0, 1, \ldots, ms$ which completes the proof of (4.6).

4.3.4 Application to flooding models

Continuing the development of Gani's flooding model (paragraphs 1.4 and 3.5.4); the case where the dams on a river, whose tributaries supply Markovian inputs, are placed on the main stream after each tributary junction, will be considered.

Suppose that the inputs $\{I_1, I_2, \ldots\}$, where I_n is the input supplied by the n^{th} tributary, form a Markov chain as in paragraph 3.5.4. It is necessary, in order to apply (4.6), to add the restriction of an upper bound, s, on the values of the inputs. That is $I_n \leq s$ for all $n = 1, 2, \ldots$. Suppose further that after each tributary junction there is a dam on the main stream which, for the n^{th} junction, has the response matrix Q_n. This Q_n need not be of a deterministic nature, however requirement (3.11) is met by a deterministic filtering.

From each Q_n one may obtain the matrices U^n of paragraph 4.3.2, and then, given the transition matrices of the Markov chain $\{I_1, I_2, \ldots\}$, obtain the A^n matrices. Then the probability distribution of the flow in the main stream between the m^{th} filtering, is given by (4.6).

4.3.5 Application to queues

Consider a queue whose length, O_t, is observed at the discrete times τ_t; $t = 0, 1, 2, \ldots$. Let the arrivals to this queue be Markovian, that is if I_t is the number of arrivals during the t^{th} time interval $[\tau_{t-1}, \tau_t]$ then $\{I_1, I_2, \ldots\}$ is a Markov chain. Further, assume that the state space of this chain is bounded. That is $I_t \leq s$, $t = 1, 2, \ldots$.

Now $\{O_t; t = 1, 2, \ldots\}$ is the stochastic process where $O_t = O_{t-1} - L_t$. Here L_t is the number of persons who leave the queue during the t^{th} time interval. This quantity L_t will then be taken to include persons who leave after service as well as persons who defect, for whatever reason. In order to regard this leaving process as a filtering on the process $\{O_t\}$ it will be necessary to be able to obtain the probabilities:

$$u^t_{i,j,(k)} = P(O_t = j \,/\, O_{t-1} = i, I_t = k) \qquad \begin{cases} t = 1, 2, \ldots \\ i = 0, 1, \ldots (t-1)s \\ j = 0, 1, \ldots ts \\ k = 0, 1, \ldots s \end{cases}$$

Where in this situation these are the probabilities that, given a previous queue length of i and an arrival of k, (i+k-j) persons leave. Consequently the leaving process may be dependent upon the queue length, so that situations where balking, defection and service time are influenced by queue length or recent arrivals can be handled. This type of situation has attracted the attention of many workers in the field of queue theory, among the most important are Haight (1957) and Finch (1959).

In the above situation it is now possible to determine the probability distribution, R_m, of the queue length, O_m, at the end of time interval m by the use of (4.6). At this time no attempt has been made to obtain steady state probabilities in terms of filtering models.

A simplified case is the following. Let τ_t be the time when the t^{th} person leaves after service. Assume no defection and exponential service times with parameter μ. Assume also independent Poisson arrivals with rate λ. These assumptions are common and lead to the simplest form of queue namely the M/M/1 queue (Parzen [1962]). In this case;

$$\begin{aligned} u^t_{i,j,(k)} &= P(O_t = j \,/\, O_{t-1} = i,\ I_t = k) \\ &= 1 \text{ if } i+k = j+1 \text{ or if } i = j = k = 0 \\ &= 0 \quad \text{otherwise} \end{aligned} \qquad \begin{cases} t = 1, 2, \ldots \\ i = 0, 1, 2, \ldots, (t-1)s \\ j = 0, 1, 2, \ldots, ts \\ k = 0, 1, 2, \ldots, s \end{cases}$$

Also $P(I_t = j \,/\, I_{t-1} = i) = p^t_{i,j} = qp^j \qquad j = 0, 1, 2, \ldots$

with $p = \frac{\lambda}{\mu\lambda}$.

$$\text{because } p^t_{i,j} = P(I_t = j) = \int_0^\infty \mu e^{-\mu t} \cdot \frac{e^{-\lambda t}\,(\lambda t)^j}{j!}\,dt$$

$$= \left(\frac{\mu}{\mu+\lambda}\right) \cdot \left(\frac{\lambda}{\mu+\lambda}\right)^j = qp^j$$

In order to place a bound on the number of arrivals during any one service it will be necessary to assume that, for all $j \geq s$, qp^j is small enough to be disregarded.

APPENDIX I

A. Some Shift Multiplication Algebra

I. Properties of Conover's shift operator

(a) It is transitive, but not in general associative or commutative.

$$A * (B+C) = (A*B) + (A*C) \tag{i}$$

$$(B+C) * A = (B*A) + (C*A) \tag{ii}$$

(b) $$(AB) * C = A * (BC) \tag{iii}$$

however

$$(A*B)C \neq A * (BC)$$

(c) For suitably dimensioned identity matrices I_1 and I_2

$$I_1 * A = A * I_2 \tag{iv}$$

where if A is square $I_1 = I_2$.

also $Q \quad QA = I * A$ (v)

(d) If A is a row vector and E a unit column

$$[(A*I) * B]\, E = (I*B')A' \tag{vi}$$

which allows for a way of dropping E as in (3.1) for example.

II. Properties of Gani's shift operator

(a) Neither transitive, associative nor commutative

(b) $$A \ominus B = A(I \ominus B) \tag{vii}$$

$$= A\overset{\theta}{B} \text{ where } \overset{\theta}{B} = I \ominus B. \tag{viii}$$

(c) If A is a row vector then

$$A \ominus B = E' \ominus (B' \ominus A') \tag{ix}$$

where E is a unit column vector of suitable length.

III. Some relations between the shift operators

(a) $$A * B = (I \ominus B'A')' \tag{x}$$

(b) $$A \ominus B = A(I*B')' \tag{xi}$$

(c) If A is a row vector then

$$(A\Theta B)' = [(A*I) * B]E \qquad \text{(xii)}$$

B. Proofs of equations (i) through (xii)

Use the notation $A = \{a_{i,j}\}$ for a matrix A with elements $a_{i,j}$, $i = 1, 2, \ldots n$ and $j = 1, 2, \ldots m$. Let E be a unit column vector of suitable length and let $I = \delta_{i,j}$ be the identity matrix with elements defined by Kroneker's delta:

$$\delta_{i,j} = 1 \text{ if } i = j$$
$$= 0 \text{ otherwise}$$

The equations are not proved in numerical sequence.

(i) and (ii):

These have been proved by Conover [1965].

(iii): Proof of

$$(AB) * C = A * (BC)$$

This equation follows from the fact that the shifting of the columns in Conover's shift operation takes place after the formation of a common matrix product so that both (AB) * C and A * (BC) are shiftings of the product ABC.

Dimensions:

If A is $(p+1) \times (q+1)$, B is $(q+1) \times (r+1)$ with C then $(p+1) \times (r+1)$ then both (AB) * C and A * (BC) are $(p+s+1) \times (s+1)$.

(iv): This follows directly from (iii).

(v): Proof that

$$\exists Q \ni QA = I * A$$

Observe that if A is $(p+1) \times (q+1)$ and I is $(p+1) \times (p+1)$ then I * A is a $(p+q+1) \times (q+1)$ matrix containing columns of

A shifted sequentially downward. The same result could be obtained by multiplying A on the left by a $(p+q+1) \times (p+1)$ matrix Q whose elements are solution to the consistent set of $(2q+1)(q+1)$ equations in as many unknowns, represented by:

$$QA = I * A$$

(vii) and (viii):

These follow directly from the definition of Gani's shift operator.

(x): Proof that

$$A * B = (I \theta B'A')'$$

This result follows from a consideration of their general elements, using paragraph 2.2.4. It is also easy to see intuitively. Since $B'A' = (AB)'$, the right hand side is simply the transpose of $(AB)'$ after its rows have been shifted. This is equivalent to shifting its columns, which is the left hand side.

Dimensions: If A is $(p+1) \times (q+1)$ and B is $(q+1) \times (r+1)$ then AB is $(p+1) \times (r+1)$ so that $I \theta B'A'$ is $(r+1) \times (r+p+1)$ which when transposed is the dimension of $A * B$.

(xi): Proof that

$$A \theta B = A(I * B')' :$$

$$A \theta B = A(I \theta B) \qquad \text{by (vii)}$$

but $\quad I * B' = (I \theta B)' \qquad$ by (x)

so that $A \theta B = A(I * B')'$

(xii): Proof that

$(A \theta B)' = [(A * I) * B]E$ if A is a $1 \times (q+1)$ row and E is a $(r+1) \times 1$ column where B is $(q+1) \times (r+1)$:

Here $A * I = \{a_i \, \delta_{i,j}\}$ by (2.3)

so that $(A * I) * B = \{a_i \, \delta_{i,j}\} * \{b_{i,j}\}$

$$= \{\sum_k a_{i-j} \, \delta_{i-j,k} \, b_{k,j}\} \quad \text{by (2.3)}$$

$$= \{a_{i-j} \, b_{i-j,j}\}$$

thus $[(A * I) * B]E = \{\sum_j a_{i-j} \, b_{i-j,j}\}$

while by (2.4), $(A \theta B)'$ is the same.

Dimensions: $A * I$ is $(q+1) \times (q+1)$ so that

$(A * I) * B$ is $(q+r+1) \times (r+1)$. Thus $[(A * I) * B]E$ is

$(q+r+1) \times 1$ as is $(A \theta B)'$

(vi): This follows directly from (xii) and (xi)

(ix): Proof of

$A \theta B = E' \theta (B' \theta A')$ where A is a row vector and E a unit column matrix.

Note that $(B' \theta A')' = (I * A)B$ by (xi)

so that $(A * I_1) * B = I(A * I_1) * B$

$$= I * (A * I_1)B \quad \text{by (iv)}$$

$$= I * (I * A)B \quad \text{by (iv)}$$

$$= I * (B' \theta A')'$$

Thus $A \theta B = E' ([A * I] * B)'$ by (vi)

$$= E' (I * [B' \theta A'])'$$

$$= E' \theta (B' \theta A') \quad \text{by (xi)}$$

APPENDIX II

Shift Operators And Convolutions

A. A general discussion

Convolutions are defined in several different ways in the literature. The most simple form is that used by Feller in volume I of his book [1950]. To avoid confusion, this definition is reproduced here.

Definition: Let $\{a_k\}$ and $\{b_k\}$, $k = 0, 1, 2, \ldots$ be any two number sequences (not necessarily probability distributions). The new sequence $\{c_r\}$ $r = 0, 1, 2, \ldots$ defined by

$$c_r = a_0 b_r + a_1 b_{r-1} + \ldots + a_{r-1} b_1 + a_r b_0 \qquad \text{(i)}$$

is called the convolution of $\{a_k\}$ and $\{b_k\}$.

Thus $c_r = \sum_{k=0}^{r} a_k b_{r-k}$ which bears resemblance to the general term of a shift product. If we let the sequences $\{a_k\}$, $\{b_k\}$ and $\{c_k\}$ form the elements of the row vectors A, B, and C respectively, then we have the following minor theorems.

Theorem 1. If A and B are row vectors whose convolution is C then

$$C = A \ominus EB \qquad \text{(ii)}$$

where E is a unit column vector.

Proof: Since E is a unit column vector, the product EB produces a matrix with the row vector B repeated as each row of EB. So that if $x_{i,j}$ is the general element of EB then:

$$x_{i,j} = b_j \begin{cases} i = 0, 1, 2, \ldots \\ j = 0, 1, 2, \ldots \end{cases}$$

Thus c_j, the general element of $A \ominus EB$ is, by (2.6):

$$c_j = \sum_{k=0}^{j} a_k \, x_{k,j-k} \qquad j = 0, 1, 2, \ldots$$

$$= \sum_{k=0}^{j} a_k \, b_{j-k}$$

So that the vector $\{C_j\}$ is the convolution of A and B.

The result (ii) has equivalent forms in terms of Conover's operator:

Theorem 2. If A and B are row vectors whose convolution is C then

$$C = E'[A * I) * EB]' \qquad \text{(iii)}$$

$$\text{and} \quad C = A(I * B'E')' \qquad \text{(iv)}$$

Proof: These results follow from (ii) directly by using equations (xii) and (vi) of appendix I.

Since the theory of convolutions is fairly well developed, the relationships expressed above could be of value. One significant difference between shift operators and convolutions is worthy of comment. The convolution operation is associative (Feller [1950]) while, as has been mentioned, neither Gani's nor Conover's shift operator is. This leads to the conclusion that:

Theorem 3. If A, B, and C are row vectors then

$$A \,\theta\, E(B \,\theta\, EC) = (A \,\theta\, EB) \theta EC = A \,\theta (EB \,\theta\, EC) \qquad \text{(v)}$$

Proof: It has been shown that

$$B \,\theta\, EC = \{ \sum_{k=0}^{j} b_k \, C_{j-k} \}$$

$$\text{so that } A \,\theta\, E(B \,\theta\, EC) = \{ \sum_{n=0}^{j} a_n \sum_{k=0}^{j-n} b_k \, C_{j-n-k} \}$$

$$= \{ \sum_{n=0}^{j} \sum_{k=0}^{j-n} a_n \, b_k \, C_{j-n-k} \} \qquad \text{(vi)}$$

$$A \ominus EB = \{\sum_{k=0}^{j} a_k\, b_{j-k}\}$$

so that $$(A \ominus EB) \ominus EC = \{\sum_{n=0}^{j} \sum_{k=0}^{n} a_k\, b_{n-k}\, c_{j-n}\} \qquad \text{(vii)}$$

Similarly $$A \ominus (EB \ominus EC) = \{\sum_{n=0}^{j} \sum_{k=0}^{j-n} a_n\, b_k\, c_{j-n-k}\} \qquad \text{(viii)}$$

Obviously (viii) is the same as (vi). That (vii) is an equivalent form follows from a consideration of the terms involved. Each term of (vi) has a similar term in the expansion of (vii).

B. Relevance to sums as Markov chains

In paragraph 3.5 the sum of independent variables was considered as a Markov chain. A statement was left unproved, namely:

$$P_k = (I * T_k')' \qquad \text{(ix)}$$

where T_k is the transition matrix of a Markov chain in which all states are independent and P_k is the transition matrix of the corresponding sum of states Markov chain. (See 3.5.1)

Proof of (ix): If all the states of the Markov chain $\{I_1, I_2, \ldots\}$ are independent with:

$$p_i = P(I_n = i) = P(I_n = i \,/\, I_{n-1} = j) \qquad i = 0, 1, 2, \ldots$$

then clearly the elements, $t^k_{i,j}$ of T_k are

$$t^k_{i,j} = p_j \qquad \begin{cases} i = 0, 1, 2, \ldots \\ j = 0, 1, 2, \ldots \end{cases}$$

so that if $R_1 = \{p_i\}$ then

$$T_k = ER_1 \qquad \text{(x)}$$

where E is a unit column matrix.

From (3.1) $\quad R_2' = [R_1 * I*T_2]E$

where R_i is the probability distribution of

$$S_i = \sum_{k=0}^{i} I_k; \quad i = 1, 2, \ldots$$

This can be generalized to:

$$R_k' = [R_{k-1} * I * T_k]E$$

$$= (I * T_k')R_{k-1}' \qquad \text{by (vi) of appendix I}$$

thus $R_k = R_{k-1}(I * T_k')'$

so that by analogy to the Markov chain of sums of states

for which $R_k = R_{k-1} P_k$

it follows that $P_k = (I * T_k')'$ is the transition matrix here

This completes the proof of (ix).

It is of value to note that, by (3.2)

$$R_m = R_1 \theta T_2 \theta T_3 \theta \ldots \theta T_m$$

where operation from the right is understood. However, since (x):

$$R_m = R_1 \theta E R_1 \theta \ldots \theta E R_1 = R_1(\theta E R_1)^{m-1}$$

It follows from (v) that in this case the direction of operation is not significant.

APPENDIX III

A. Proof of alternative (3.6)

It has been mentioned in the body of this work that, at any stage the lack of a filtering can be represented by a response matrix which is simply the identity matrix. That is a filtering may be applied which has as the elements of its response matrix:

$$q^t_{i,j} = P(O_t = j \,/\, I_t = i) \qquad \begin{cases} i = 0, 1, 2, \ldots \\ j = 0, 1, 2, \ldots \\ t = 1, 2, \ldots \end{cases}$$
$$= 1 \text{ if } i = j$$
$$= 0 \quad \text{otherwise}$$

Thus the response probability is certain if the output is equal to the input and zero otherwise. By using the concepts of paragraph 4.3 it is possible to consider the sum of states of a Markov chain as the stochastic process resulting from the filtered sums of states of a Markov chain, with the identity as response matrix. So that equation (4.6) results in (3.6) directly.

B. Proof of Conover's form for the case of an infinite state space

In paragraph 3.2.2 it was stated that

$$R'_m = [R_i * I * P_2 * P_3 * \ldots * P_m]E$$

where, if we have an infinite state space, all of these matrices are infinite. R_i; $i = 1, 2, \ldots$ is the probability distribution of the partial sum of the first i states, P_i is the transition matrix from the $(i-1)^{st}$ to the i^{th} state and I is an identity matrix of infinite dimensions. E is an infinite unit column. The elements of these matrices will be denoted as

$$R_r = \{p_i^{(r)}\}; \quad P_r = \{p_{i,j}^{(r)}\}$$

(i) By (2.9) $R_1 * I = \{\delta_{i,j} \sum_{k=0}^{\infty} p_k^{(1)} \delta_{k,j}\}$

$$= \{\delta_{i,j}\ p_j^{(1)}\}$$

where $\delta_{i,j}$ is Kroneker's delta, defined earlier.

(ii) Then $R * I * P_2 = \{\sum_{k=0}^{\infty} \delta_{i-j,k}\ p_{i-j}^{(1)}\ p_{k,j}^{(2)}\}$ by (2.9)

$$= \{p_{i-j}^{(1)}\ p_{i-j,j}^{(2)}\}$$

$= \{P(I_1 = i-j)\ P(I_2 = j\ /\ I_1 = i-j)\}$

$= \{P(I_2 = j,\ I_1 = i-j)\} = P(S_2 = i,\ I_2 = j)\}$ (a)

if $S_r = \sum_{i=1}^{r} I_i$; $i = 1, 2, \ldots$

(iii) Now by using (a) as a starting point it will be shown, by mathematical induction, that

$$R * I * P_2 * \ldots * P_r = \{P(S_r=i,\ I_r=j)\} \overset{\text{let}}{=} \{q_{i,j}^r\} \qquad \text{(b)}$$

Assume (b) to be true for some r. Then by (2.8)

$$(R_1 * I_1 * P_2 * \ldots * P_r) * P_{r+1} = \{q_{i,j}^{r+1}\}$$

$$= \{\gamma_{i,j} \sum_{k=0}^{\infty} q_{i-j,k}^{r}\ p_{k,j}^{r+1}\}$$

where $\gamma_{i,j}$ is, as before, a function which is zero only when j exceeds i and is otherwise unity. So now:

$$\{q_{i,j}^{r+1}\} = \{\gamma_{i,j} \sum_{k=0}^{\infty} P(S_r=i-j,\ I_r=k)\ P(I_{r+1}=j/I_r=k)\}$$

$$= \{\gamma_{i,j} \sum_{k=0}^{\infty} P(S_{r-1}=i-j-k,\ I_r=k)\ P(I_{r+1}=j/I_r=k, S_{r-1}=i-j-k)\}$$

$$= \{\gamma_{i,j} \sum_{k=0}^{\infty} P(I_{r+1}=j,\ I_r=k,\ S_{r-1}= i-j-k)\}$$

$$= \{P(S_{r+1} = i,\ I_{r+1} = j)\}$$

So that (b) holds true for all r = 1, 2, . . .

(iv) Thus $[R_1 * I * P_2 * \dots * P_r]E = \{ \sum_{k=0}^{\infty} q_{i,k}^{r} \}$

$$= \{ \sum_{k=0}^{\infty} P(S_r = i,\ I_r = k)\}$$

$$= \{P(S_r = i)\}$$

C. Proof of Gani's form for the case of an infinite state space

If the reader will compare the general element of Gani's shift product for the finite case (2.4) and for the infinite case (2.5 and 2.6) he will observe that the difference lies only in that the summation for the infinite case extends to j whereas for the finite case it extends to q (where the common dimension of the operand matrices is q+1).

Reading through the proof given for the finite case in paragraph 3.2.4, it will be clear that, if the summations used extend to j rather than to s, the entire set of non-zero probabilities is covered for the infinite case so that (3.2) holds equally for the unbounded state space.

ACKNOWLEDGEMENT

The writer wishes to express his sincere appreciation to his major professor, Dr. W.J. Conover for his invaluable assistance and encouragement during the preparation of this thesis.

He also wishes to thank all of the many individuals who assisted, in one way or another, to make this thesis possible; in particular he wishes to thank Miss Beverly Old for her many hours of typing.

Acknowledgement is also due to all of the writers' instructors, both here and at his home university for the efforts that they have made to bring him thus far.

REFERENCES

Burke, P. J. [1956]. The output of a queueing system. J. Opns. Res. Soc. Amer., 4, pp. 699-704.

Cohen, J. W. [1957]. Basic problems of telephone traffic theory and the influence of repeated calls. Phillips Tel. Rev., 18 p. 49-100.

Conover, W. J. [1965]. The distribution of $\sum_0^m f(y_t)$, where $(y_0, Y_1, \ldots)$ is a realisation of a non-homogeneous finite state Markov chain. Biometrika, 52 p. 277-279.

Feller, W. [1950]. An introduction to probability theory and its applications. Vol. I, Wiley and Sons, New York.

Finch, P. D. [1959]. Balking in the queueing system GI/M/1. Acta Math. Acad. Sci. Hung., 10 pp. 241-247.

Fuller, L. E. [1962]. Basic Matrix Theory. Prentice Hall.

Fuller, W. E. [1914]. Flood flows. Trans. Am. Soc. Civ. Eng., V 77.

Gani, J. [1955]. Some problems in the theory of provisioning and of dams. Biometrika, V 42, pp. 179-200.

Gani, J. [1957]. Problems in the probability theory of storage systems. J. Roy. Stat. Soc., B 19, pp. 181-233.

Gani, J. and Prabhu, N. U. [1957]. Stationary distributions of the exponential type for the infinite dam. J. Roy. Stat. Soc., B 19, pp. 342-351.

Gani, J. [1965]. Flooding models. Presented at the Reservoir Yield symposium, Paper 4, Water Res. Assoc. Medmenham, Marlow, Bucks.

Geyer, J. C. [1940]. New curve fitting methods for the analysis of flood records. Tr. Am. Geo. Union, Part II, pp. 660-668.

Ghosal, A. [1960]. Emptiness in the finite dam. Ann. Math. Stat., V 31, pp. 803-808.

Gumbel, E. J. [1942]. The return period of flood flows. Ann. Math. Stat., V 12, pp. 163-189.

Haight, F. A. [1957]. Queueing with balking. Biometrika, pp. 360-369.

Hohn, F. E. [1958]. Elementary Matrix Algebra, MacMillan, New York.

Kendall, D. G. [1957]. Problems in the theory of dams. J. Roy. Stat. Soc. B 19, pp. 181-233.

Moran, P. A. P. [1954]. A probability theory of dams and storage systems, Aust. J. Appl. Sci. Vol 5 pp. 116-124.

Palm, C. [1937]. Inhomogeneous telephone traffic in full availability groups. Erricsson Tech., V 5, pp. 3-36.

Parzen, E. [1962]. Stochastic Processes. Holden - Day, Inc., San Francisco.

Prabhu, N. U. [1958]. Some exact results for the finite dam. Ann. Math. Stat., V 29, pp. 1234-1243.

Prabhu, N. U. [1965]. Queues and Inventories. Wiley and Sons, New York.

Riordan, J. [1953]. Delay curves for calls served at random. Bell System Tech. J. V 32, pp. 100-119.

Riordan, J. [1962]. Stochastic server systems. Wiley and Sons, New York.

Sasty, T. L. [1961]. Elements of queueing theory, McGraw - Hill, New York.

Segerdahl, C. O. [1939]. On homogeneous random processes and collective risk theory. Uppsala: Almqvist and Wiksells.

Slade, J. J. [1936]. The reliability of statistical methods in the determination of flood frequencies. Geo. Sur. Water Supply paper, 771. Wash.

Wilks, S. [1962]. Mathematical Statistics. Wiley and Sons, New York.

Yeo, G. F. [1961]. The time dependent solution for an infinite dam with discrete additive inputs. J. Roy. Stat. Soc., B 23, pp. 173-179.

MATRIX SHIFT OPERATORS

APPLIED TO MARKOV CHAINS

by

PETER J. R. BOYLE

B. Sc., University of Pretoria, 1965

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1968

ABSTRACT

In the investigation of the distribution of the partial sums of states, $S_m = \sum_{i=0}^{m} I_i$, m = 1, 2, . . . , of the non-homogeneous finite state Markov chain $\{I_1, I_2, . . .\}$, Conover developed a matrix operation which he called 'shift multiplication'. It was in terms of this operator that he was able to express the probability distribution of S_m. Later Gani used a similar operator, developed from Conover's operator, to perform the same function. The two operators were, however, quite different in their algebraic properties.

In the first part of this thesis, these two operators were defined and the general elements of their products obtained. Since neither Conover nor Gani had occasion to generalize their operations to include infinite matrices, this was done. Some of their more obvious algebraic properties were given and their relation to the convolution operator from probability theory was discussed.

Then the probability distribution of the partial sums of states of a Markov chain was presented in the two forms obtained by Conover and Gani. Gani's form was proved and an alternative method for obtaining this distribution without recourse to shift operators was presented and proved. The case where the state space of the Markov chain is not bounded, was given and proved. Some consequences of the case where the sums of states of a stochastic process form a Markov chain, were mentioned. A Poisson counting process was used to illustrate this.

Gani used his operator to obtain the probability distribution of the flow magnitude in the main stream of a river system with tributaries supplying Markovian inputs. Using a dam model slightly modified from

Moran [1954] he obtained this distribution for two dam systems.

In the thesis, generalizing from Gani's results, a "black-box" type of filtering was defined in which the output is a random variable whose distribution is determined by the realization of the input variable. The probability distribution of this output was discussed for the case where the states of a Markov chain are filtered for the case where the partial sums of states of a Markov chain are filtered.

These results were applied to Gani's flooding model and one of his conclusions confirmed. The other, concerning a solution for the case where dams are placed on the tributaries, as interpreted in this thesis, was shown to be in error.

Some brief examples of applications of these results to other situations were given.