THE UNIVERSITY OF MANITOBA

APPLYING A REGIONAL FLOW MODEL TO HEALTH CARE PLANNING USING UNIVERSAL HEALTH INSURANCE DATA

BY

FRANK FREDERICK ROSS MAYHOOD

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

DEPARTMENT ....COMPUTER SCIENCE...

WINNIPEG, MANITOBA

October 1977

APPLYING A REGIONAL FLOW MODEL TO HEALTH CARE PLANNING USING UNIVERSAL HEALTH INSURANCE DATA.

by

Frank F. R. Mayhood

A dissertation submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

Master of Science

ABSTRACT

A model of Manitoba's health care delivery system, based on Markov process theory, is discussed in the context of assessing the adequacy of care in various regions. Data processing problems encountered in the course of estimating the model from the universal health insurance data provided by the Manitoba Health Services Commission, are noted. The methods of estimation suggested by S.Zahl [Human Biology 1955 p90] were found to be inadequate for this data and alternatives were developed. The implications of the statistics collected, both for estimation of the model and derived from the model's properties, are explored, and the potential use of this class of model for health care planning is examined.

ACKNOWLEDGEMENTS

I wish to thank the Manitoba Health Services Commission for providing access to their files, and particular thanks to Dennis Roch of the Commission for his helpful suggestions, clarifications, and encouragement throughout the project. I am indebted to Dr. Jana Mossey of the Department of Social and Preventive Medicine and Dr. W. Hoskins of the Department of Computer Science for their perserverance in reading this work and for their helpful comments. Ms. Peg Burh was of great assistance in the preparation of the text. The financial support of the National Research Council through a postgraduate scholarship was much needed and appreciated. For the invaluable advice and council of my advisor, Dr. A. Neil Arnason, I am eternally grateful. Last, but far from least, I wish to acknowledge the efforts of my wife, Linda, whose patience, devotion, and plain hard work made the completion of this thesis possible.

TABLE OF CONTENTS

INDEX OF TABLES AND FIGURES

EQUATION DEFINITIONS

CHAPTER 1. ISSUES

The work presented in this thesis was motivated by three major concerns. The bulk of the theory of Markov processes fills many volumes, and there are oft repeated claims for their abilities as models in the health care field [Zahl 1955, Chiang 1973, Navarro 1970]. However, reports on large scale attempts to apply the theory are almost non-existent. The attempt to apply the Markov process model to a practical situation, solving along the way the many computational problems which arise in such applications, was the primary motivation for this thesis.

The second most important concern was an attempt to provide a workable and useful tool for health care planning in the context of universal health insurance. Direct costs to individuals on a fee for service basis are eliminated in areas where universal health insurance has been accepted. As a result, costs tend to be negotiated between the government regulatory agency and the operational health care facilities, rather than being determined by actual cost considerations. Many of the standard economic models become useless in such a context because they depend upon meaningful and flexible cost structures. In addition, the vast amount of data which is available as a result of the claims procedures generated by the insurance operation should provide the information to substantially improve the quality of modeling tools for health care planning. Many types of model which could not be estimated efficiently before universal health insurance, can now be explored with relative ease.

The third major area of interest was the attempt to provide insight into the functioning of the health care delivery system in Manitoba, with particular interest in the regional variations in access to care. The data provided by the Manitoba Health Sevices Commission (M.H.S.C.) from the operation of the universal health insurance scheme in 1972, provides a unique opportunity to examine the variations in health care usage by a variety of factors. Regional variations are of particular interest because they may indicate the need for additional facilities or for different kinds of facilities in the various areas.

The remainder of this chapter is devoted to enlarging on some specific themes which are derived from one or more of the above concerns and which recur periodically throughout the rest of the dissertation. Chapter 2 contains a discussion of the relevant theory and introduces the notation commonly used for the class of models which we use in the study. In chapter 3, the computational problems of data reduction for estimation of the model and of applying the theory of chapter 2 are discussed. The strengths and weaknesses of the statistical results and their application to the above issues are discussed in chapter 4. The final chapter contains an examination of some possible uses of the model for planning the evolution of the health care system, or for monitoring its performance.

1.1 Quantity and Quality

The advent of universal health insurance under the control of a government monopoly has shifted the burden of planning from the various institutions and individuals involved in delivering the care, to the government who now pays for it. This shift has provided for a wider outlook, longer term goals, and a political character to decision making which all demand a more complete approach to understanding and to modeling the health care system.

Fortunately, the very nature of the insurance type of implementation has provided a rich source of raw data from which this better understanding could be built. The first step is to isolate those factors which are important ones in regards to the functioning of the system. For health care, there can be little doubt that the quality and quantity of care provided are important outputs of the system. These factors will intimately affect the overall health of the population served, which in turn will affect that population's productivity. These factors will also affect the public's perception of the performance of the health care system, and ultimately of the politicians who control it. Thus the ability to measure the quality and quantity of care delivered, particularly as it relates to the population's perception of its needs, is a desirable characteristic of a model of the health care system.

The overall health status of a population is difficult, if not impossible, to quantify. Various manifestation of ill health are measurable. Such measures as mortality rates, morbidty rates, and counts of services rendered give some indication of the underlying "healthiness" of the population, but these only scratch the surface of deviations from perfect health (whatever that may be). The shortcomings of these statistics or measures of overall "health" have been widely discussed [Goldsmith, 1972]. An alternative index suggested by C.L. Chiang and R. Cohen [1975] is a bi-product of the type of model we discuss in later chapters, and its performance in our situation was examined.

The quantity issue may also be viewed from the point of view of the various facilities which comprise the health care system. These characteristically require a long lead time for either expansion or contraction. Thus accurate forecasts of future requirements, particularly where the population is growing or changing rapidly, are necessary on a medium to long term basis. A model of the health care system should assist in the provision of such forecasts.

1.2 Equality of Access

One of the consequences of the control of health care planning by the political process is the interjection of a new set of goals and priorities. Normally, the health care system would be almost entirely concerned with the provision of adequate care. But because of the political process and our understanding of fair government practice, the issue of equality of access to health care becomes important. Each region in a province such as Manitoba must perceive itself to be receiving fair and adequate health care services, relative to other regions.

Measurement of the degree of access to care is almost impossible. However, manifestations of inequality of access can be anticipated and looked for in the data. Lack of adequate access to a particular type of facility in a particular region is likely to manifest itself in two ways, before it affects the overall health of that region's population. First, those in need of types of care which are inadequately supplied in a particular region can be expected to travel to other regions in search of care. Large numbers of people observed seeking care in other regions would be evidence of a lack of access to that type of care locally. If, in addition to large flows of people seeking care elsewhere, there are no appreciable flows of people into that region to receive care, then there is evidence of a difference in the level of access to health care among regions. Second, those in need of a particular type of care would be expected to substitute other types of care which are more readily available in the region for those with more limited access in that region.

Both of these measurements depend upon the ability to compare regions on the basis of the health care delivered to those resident in that region, rather than just the health care delivered by the region's various health care facilities. By examining care regionally in this way, instead of the more usual facility "catchment area" approach, and by measuring and interpreting the inter-regional delivery of care, many advantages are realized. Regions may be precisely defined, based on administrative or geographical considerations, without conflict with modeling requirements. This makes the statistics produced more readily interpretable in terms of the physical system modeled, and allows for the treatment of issues such as equality of access as well.

1.3 Data Requirements and Computational Aspects

The types of problems which arise in using data captured for administrative purposes, such as the claim data from a universal health insurance scheme, for research purposes, are well known, if not understood, by those who call themselves Computer Scientists. Any use of data for a purpose other than the one for which it was collected and maintained is fraught with peril. In examining the M.H.S.C. files for the purposes of this study, many of these problems were encountered, some of which were circumvented successfully and some of which were only noted. All have had their impact on the utility of the results. If further attempts to use this type of data in a research context are made, some valuable lessons may be taken from the struggles presented in chapter 3. Most of the examples are particular to the form of the data maintained by the M.H.S.C., but some should be useful whenever the use of universal health insurance data is contemplated.

The computational problems associated with applying the theory of Markov processes were found to be substantial. Some of the theoretical formulae were found to be of little computational use in their published forms. The methods described for computing the various estimates of parameters and derived statistics should be of use to anyone who contemplates applying this class of model, whatever the application.

Chapter 2. A Markov Flow Model

2.1 Theoretical Groundwork

The mathematical objects named for A.A. Markov (1856-1922) have received a great deal of attention from mathematicians in the last 50 years. The resulting theory has filled many books (Cox & Miller, Bharucha-Reid, Chiang) some of which are listed in the bibliography. The volume of results makes it impractical to present even the barest outline of them here, but a basic understanding of the concepts is necessary to the discussion which follows, and a brief theoretical discussion is a convenient way to introduce the notational conventions which should facilitate the presention of our conclusions. The attempt is to communicate the essence of the theory which forms a framework for the study, and refer the reader to the appropriate references for a more rigorous approach.

Markov processes are mathematical constructs which arise in the context of the study of objects or systems which change from state to state. As mathematical entities, they may be axiomatized and their properties developed analytically from those axioms. To apply a Markov process to a real world situation is to assume that the axioms are true and use the derived properties to describe, forecast, or analyze the real world. Thus, in an application of the theory of Markov processes to the study of the health care delivery system in Manitoba, one must expend a fair amount of effort in determining the extent to which the mathematical axioms can be said to hold, and the robustness of the derived properties to violations of these axioms. The first step in this process is to identify the set of axioms which are assumed to be true and phrase them in the terms of the real world situation. That task is attempted in the remainder of this chapter.

2.1.1 Basic Definitions

In a modeling effort, the mathematical axioms of the model are used to simplify the complexity of a real world system into a form which yields to mathematical analysis. The results of the analysis can then be used to draw hypotheses about the performance of the real world system. A Markov process simplifies in two primary ways. Firstly, it assumes that there exists a set of mutually exclusive categories or "states" such that at any instant, the system can be said to be <u>in</u> a particular state. The state of a system may be one dimensional (like the number of people in a queue) or multi-dimensional (like the number of people alive and in a particular age cohort for each of a set of cohorts).

In mathematical terms there must exist a set $\mathcal{X}$ of states, a time space (partially ordered) set $\mathcal{T}$ and a function:

$$X : \mathcal{T} \rightarrow \mathcal{X} \qquad \text{(E2.1.1)}$$

The function X(t) gives the state of the system at each point ,t, in the time space, $\mathcal{T}$. This classification function reduces the real world system to movements between states in a well defined set of states, $\mathcal{X}$. But this is still far too general to yield much useful from analysis. The nature of these movements from state to state must be controlled. A Markov process controls the mechanism of movements by assuming that the classification into states is a very special one in that knowledge of the state of the system at a particular time t^0 completely summarizes the past history of the system at all times $t < t^0$ as far as the future evolution of the system is concerned. More precisely:

$$Pr(X(t_1) = x \mid X(t) \quad t < t_0) = Pr(X(t_1) = x \mid X(t_0)) \qquad \text{(E2.1.2)}$$

The conditional probability that the system is in a particular state x at some later time t^1 given the entire history of the system prior to time t^0 is equal to the conditional probability that the system will be in state x at t^1 given only the state $X(t^0)$ which it was in at time t^0. This "memorylessness" of the transitions between states of a Markov process is the cornerstone of the theory which permits statements to be made about the future evolution of the system.

Markov processes can be classified according to the cardinality of their state and time spaces $\mathcal{X}$ and $\mathfrak{J}$. If $\mathcal{X}$ and $\mathfrak{J}$ are both countable, that is they can both be considered as equivalent to subsets of the integers, then the process is called a Markov Chain. These will be useful later [2.1.6] as approximations to other, more complex, processes. The type of process which will be of primary interest in what follows has a finite state space $\mathcal{X}$, and a continuous time space. In this case, one usually replaces by the set of integers {1,2,...NS} where NS is the number of elements in the original state space. The conditional probability of the system being in state j ($1 \le j \le NS$) at time t^1 given that it was in state i ($1 \le i \le NS$) at time $t^0 \le t^1$ is written P (t^1, t^0):

$$P_{ij}(t_1,t_0) = Pr(X(t_1)=j \mid X(t_0)=i) \qquad \text{(E2.1.3)}$$

Such a process is said to be <u>stationary</u> or <u>time homogeneous</u>

if these transition probabilities are functions of $\Delta t = t^1 - t^0$ alone:

$$p'_{ij}(\Delta t) = p_{ij}(t_1, t_0) \qquad \text{(E2.1.4)}$$

This means that the probability of being in j at time t^1 given that the system was in state i at time t^0 does not depend on the absolute location along the time axis of t^0. Associated with the transition probabilities $p_{ij}(t^1, t^0)$ are the instantaneous rate functions $\gamma_{ij}(t)$ defined by:

$$\lim_{\Delta t \to 0} \frac{p_{ij}(t+\Delta t, t)}{\Delta t} = \gamma_{ij}(t) \qquad i \neq j \qquad \text{(E2.1.5)}$$

which implies:

$$p_{ij}(t+\Delta t, t) = \gamma_{ij}(t)\ \Delta t + o(\Delta t) \qquad i \neq j \qquad \text{(E2.1.6)}$$

where : $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0.$

It is logical to require :

$$\lim_{\Delta t \to 0} p_{ij}(t+\Delta t, t) = \begin{matrix} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{matrix} \qquad \text{(E2.1.7)}$$

thus (E2.1.5) cannot be used to define $\gamma_{ii}(t)$. Instead using (E2.1.6) as a model, define:

$$-\gamma_{ii} = \gamma_i = \sum_{j=1}^{NS} \gamma_{ij} \qquad \text{(E2.1.8)}$$

then:

$$p_{ii}(t+\Delta t, t) = (1 + \gamma_{ii}(t)\ )\ \Delta t + o(\Delta t) \qquad \text{(E2.1.9)}$$

since the system must be in one of the states at each instant. For convenience, define the transition probability matrix $P(t^1,t^0)$ and the instantaneous rate matrix $\Gamma(t)$ as:

$$P(t_1,t_0) = \left(P_{ij}(t_1,t_0)\right)$$
$$\Gamma(t) = \left(\gamma_{ij}(t)\right) \qquad (E2.1.10)$$

If the process is stationary (E2.1.4) then $P(t^1,t^0)=P(t)$ and $\Gamma(t)=\Gamma$ is independent of time.

For the processes which will interest us, it is possible to view the system in two ways, each with it's own insights into the real system. These models involve the movements of individuals or particles about a finite set of states. The model of an individual's movements has a one dimensional state space and $X(t)$ gives the state that that individual or particle occupied at time t. One may also be interested in the system as it describes a set of individuals moving about simultaneously in the same state space. This process is multidimensional, its classification function $X(t)$ gives the number of individuals in each state at time t. It is a vector of population levels of dimension NS, the number of states in the one dimensional model. If the individuals or particles move independently of one another then the $P(t^1,t^0)$ and $\Gamma(t)$ matrices will be related in a very simple way, the $P(t^1,t^0)$ for the multidimensional process being the convolution of $P(t^1,t^0)$ for the one dimensional process, and the rate matrices will be essentially related in the same way. Knowledge of either will determine the properties of the other.

This concludes the definition of a Markov process and the notation particular to the class of processes with which we will be concerned. For further details see the references [eg. 6,7] in the bibliography. The next sections present the useful properties that can be derived from these definitions for this class of finite discrete state space, continuous time Markov processes.

2.1.2 Properties: The Chapman-Kolmorgorov Equation

A fundamental property of a Markov process is that the state space exhausts the possible configurations of the system. Thus, in order that the system be in state j at some future time t¹, it must have been in some other state at each intermediate time s. The Chapman-Kolmorgorov equation formalizes this property as:

$$p_{ij}(t_1,t_0) = \sum_{k=1}^{NS} p_{ik}(s,t_0)\, p_{kj}(t_1,s) \qquad t_0 < s < t_1 \qquad \text{(E2.1.11)}$$

If the process is stationary this can be written in matrix form as:

$$P(t + t_0) = P(\ t\ )\, P(\ t_0\ ) \qquad \text{(E2.1.12)}$$

which relates the theory of Markov processes and the theory of semigroups of operators. More importantly, when compared to $a^{x+y}=a^x a^y$, (E2.1.12) strongly hints at the functional form of P(t) in the time homogeneous case.

2.1.3 Properties: The Kolmorgorov Differential Equations

Using the definitions of the instantaneous rate functions [E2.1.6,E2.1.9] we may rewrite E2.1.11 as:

$$\begin{aligned} p_{ij}(t + \Delta t,t_0) &= \sum_{k=1}^{NS} p_{ik}(t,t_0)p_{kj}(t + \Delta t,t) \\ &= p_{ij}(t,t_0)\ [\ (1 + \gamma_{jj}(t))\ \Delta t + o(\Delta t)\ ] \\ &+ \sum_{k\neq i}^{NS} p_{ik}(t,t_0)\ [\ \gamma_{kj}(t)\Delta t\ + o(\Delta t)\ ] \end{aligned} \qquad \text{(E2.1.13)}$$

The first term is the probability of being in state i at t^0 and getting to state j by time t, then remaining in state j for at least t time units. The remaining terms are the respective probabilities of transferring into state j from some other state, given that at time t^0 the system was in state i, in the interval $[t, t+\Delta t]$. Subtracting $P \; (t+\Delta t, t^0)$ from both sides, dividing by t and taking the limit as $\Delta t \to 0$ yields the differential equations:

$$\frac{\partial p_{ij}(t,t_0)}{\partial t} = \sum_{k=1}^{NS} \gamma_{kj}(t) \; p_{ik}(t,t_0) \qquad \text{(E2.1.14a)}$$

In matrix form for fixed t^0:

$$\frac{\partial P(t)}{\partial t} = P(t) \; \Gamma(t) \qquad \text{with } P(t_0,t_0) = I \qquad \text{(E2.1.14b)}$$

This system of homogeneous first order differential equations are known as the Kolmorgorov forward differential equations of the Markov process. Unfortunately for the case of general time dependent rates $(\gamma_{ij}(t))$, the solution to these equations may not exist. However, for the stationary case, (ie with constant, time independent rates γ_{ij})the solutions to E2.1.14 can be shown to exist and to be unique under quite general circumstances [Feller, 1940]. Under these conditions the solutions may be written in the form:

$$P(t) = e^{\Gamma t} = \sum_{n=0}^{\infty} \frac{(\Gamma t)^n}{n!} \qquad \text{(E2.1.15)}$$

This equation will be of little direct use as a method for computing the solution [see 3.3.3] because of the alternating nature of the infinite series and the roundoff problems this creates. It does represent a complete solution to the mathematical problem presented by the model.

2.1.4 Properties: Length of Stay

Each time a system enters a particular state,it remains in that state for some interval of time. This "length of stay" is a random variable whose expected value will represent the average amount time spent in a state i before leaving it for some other state. It is natural to expect that this expected value will be closely related to the transition rates from state i. For a stationary Markov process this relationship is particularly simple. Letting L_i be the length of stay in state i we write, following Arnason[1975]:

$$G_i(T) = Pr(\ L_i \leq T \mid \text{currently in state } i)$$

then:

$$G_i(T + \Delta t) = G_i(T)\ Pr(\ \text{no transitions in } [T, T + \Delta t]\ )$$

$$= G_i(T)\ [\ (1 + \gamma_{ii})\ \Delta t + o(\Delta t)]$$

therefore:

$$\frac{\partial G_i(t)}{\partial t} = G_i(t)\ \gamma_{ii}$$

implying:

$$G_i(t) = e^{\gamma_{ii} t} \qquad \text{(E2.1.16)}$$

Since 1-G(t) is the cumulative distribution function of L_i, this implies that L_i is exponentially distributed with parameter $-\gamma_{ii}$. The mean of the exponential distribution is $\mu_i = 1/\gamma_i$ which simply relates the average length of stay to the instantaneous transition rates.

The variance of the length of stay in state i is $1/\gamma_{ii}^2$ by the properties of the exponential distribution. This implies that the standard deviation ($\sqrt{\text{variance}}$) is equal to the mean, a property which will be of use later when we examine the model assumptions.