STUDY ON SOME STOCHASTIC MODELS IN SURVIVAL AND CLINICAL TRIALS

Thesis submitted to the

University of Calicut

for the award of the Degree of

DOCTOR OF PHILOSOPHY

under the Faculty of Science

by

XAVIER, T. D.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALICUT
KERALA – 673 635
INDIA

March 2008

DECLARATION

I hereby declare that this Thesis entitled 'STUDY ON SOME STOCHASTIC MODELS IN SURVIVAL AND CLINICAL TRIALS' submitted to the University of Calicut, for the award of the Degree of Doctor of Philosophy under the Faculty of Science, is an independent work done by me under the supervision and guidance of Dr. M. Manoharan in the Department of Statistics, University of Calicut.

I also declare that this Thesis contains no material which has been accepted for the award of any other degree or diploma of any University or Institution and to the best of my knowledge and belief, it contains no material previously published by any other person, except where due references are made in the text of the Thesis.

Calicut University Campus **XAVIER, T. D.**
March 22, 2008

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALICUT

Dr. M. MANOHARAN, M.Sc., Ph.D.
Professor

CALICUT UNIVERSITY (P. O.)
KERALA, INDIA 673 635.
Phone : 0494 – 2401144 extn. 340 (O)
0494 – 2403403 (H)
Cell : 09447424043
Email : mano30@rediffmail.com

…….

March 22, 2008

Certificate

This is to certify that the work reported in this Thesis entitled **'STUDY ON SOME STOCHASTIC MODELS IN SURVIVAL AND CLINICAL TRIALS'** that is being submitted by Sri. Xavier, T. D. for the award of the **Degree of Doctor of Philosophy**, to the University of Calicut, is based on the bonafide research work carried out by him under my supervision and guidance in the Department of Statistics, University of Calicut. The results embodied in this Thesis have not been included in any other Thesis submitted previously for the award of any degree or diploma of any other University or Institution.

Dr. M. Manoharan

(Supervising Teacher)

ACKNOWLEDGEMENTS

I express my great indebtedness to my supervising teacher, **Dr M.Manoharan,** Professor, Department of Statistics, University of Calicut, for his scholarly guidance, continual supervision, constructive criticism and inspiration throughout the work that enabled me to complete this research. I am grateful to him for his cooperation and willingness for thoughtful discussions that made this difficult task, easy and feasible.

I wish to extend my sincere thanks to Dr. C. Chandran, Head of the Department and all the faculty members for the constant motivation, valuable support and help bestowed upon me during my research period. I gratefully acknowledge their benevolence

I am extremely thankful to the management, principal, my teachers, colleagues and office staff of St. Thomas College, Thrissur and friends who helped me in numerous ways.

I am thankful to the University Grants Commission for the award of Teacher Fellowship under FIP during the Xth Plan Period. Also I thank, the College Development Council, University of Calicut, for their timely help during the period of study. I remember with gratitude, librarians and members of office staff, Department of Statistics, University of Calicut, for providing me, the necessary facilities during my research days.

I express sincere gratitude to my parents, brother and other family members who were the constant source of

inspiration all the time. Also, I am much grateful to my wife Ancy, for her unstinting patience, cooperation & encouragement and to my children Alwin and Aleena, for their enthusiasm which acted as catalyst to me.

I remember with gratitude, all fellow FIP research scholars, other Ph. D scholars and M. Sc students of the Department of Statistics, University of Calicut.

Lastly, but not least, I express my hearty thanks to all those who helped me to complete this thesis in time.

Xavier,T. D.

Contents

Chapter 1

INTRODUCTION

1.1 Introduction

Stochastic models have become an indispensable tool for understanding many real world situations and they have been increasingly realized as an important branch of study in all fields. In the study of any system, we seek a model describing the reality by explaining the responses and outputs in terms of input variables as well as time. Almost all these models are stochastic because most of the variables are random and most of the measurements of the responses are subjected to random measurement errors.

Stochastic models have application in many well-known areas of Physical, Biological/Medical, Engineering, Social, and Economical sciences as well as in other non-trivial areas. To study these systems, it is important to understand the probability laws governing the behavior of these systems. New methodologies are researched and implemented for all these complex problems whose solutions, sometimes are extremely difficult to carry out practically. But the availability of high speed computers have added new dimensions to it. Also it is vital to check whether the new technologies possess their optimal properties so as to ensure their proper use.

Enormous progress have been achieved in the development of the science of survival/clinical trials during the last century. In this progress, several methods have been developed, implemented and refined that enable reliable, efficient and ethical evaluation of the benefits and risks of interventions that target the treatments and prevention of human diseases. The formation of censored data survival analysis meth-

ods, is probably, one of the most important components of this development. As the analysis of clinical trial data with time-to-death outcome provided the original motivation for this new statistical methodology, the field have become known as survival analysis.

The time, life and risk are three basic elements in the empirical process studied in biomedical research. Risk of birth, risk of illness and risk of death and other risks continuously act on human beings with varying degree of intensity and varying degree of frequency. Recent advances in stochastic process have made to study systematically these risks in human populations from a probabilistic point of view. Many have contributed to the theoretical development of stochastic process to the high level of sophistication they enjoy today. Among then we cherish the works of Markov, Kolmogorov Feller, Doob and others.

Stochastic models can be very effectively used for the interpretation of clinical trials. Some disease can be characterized by the patient being in one of a finite number of states; eg: relapse, remissive, toxic etc. These states may be both transient and absorbing. Different investigators have proposed Markovian models to describe data dealing with the time-dependent phenomena associated with these diseases. We mention in particular some early works of Fix and Neyman(1951), Weiss and Zelen (1963) on cancer, the works of Marshal and Goldhamer (1955) on epidemiology of mental disease, Alling(1958) on tuberculosis. All these assume that the distribution of time spent in an occurrence of a particular phase is negative exponential. But it has become necessary to consider any distribution for stay in a given phase as the Markov

Theory was found to be insufficient for the study statistics relating to many diseases, for example leukemia. In certain situation it has been found that gamma density has a convenient representation of the probability density of the stay in a phase. Hence the Semi-Markov model has wide applicability to many clinical situations.

When we look at the progression of some diseases, there we observe different stages of disease and a staging process. Development of many chronic conditions, especially, is characterized by stages. Generally disease advance with time from a mild stage though intermediate and severe stages to death. The process is irreversible, but a patient may die while being in any of the stages. In the natural progression of cancer for example, there are stages of disease determined by the size of tumor and metastasis. AIDS too, can be classified by stages.

Birth order and child spacing are another example of staging process. Here the process begins when the couple decided to start a family; stages are developed by the parities of the woman, from parity zero(no children), to parity one(one child), to parity two(two children) and so on. The process is clearly irreversible and it terminates when the couple decides to stop reproducing. we can find staging phenomena in many other areas such as metamorphosis in biology, foraging process in wildlife and cascade process for statistical studies of chronical illness(see Chiang(1979)).

Almost every theoretical development in the area of stochastic process is applied sooner or later, in survival/clinical trials or biomedical sciences. In this thesis we describe some relatively recent such applications.

In the new millennium, the lime light of advancement of knowledge has been stolen primarily by the spectacular advent of information technology (IT). Modern electronics and computers have invaded each and every corner of Globe and touched all walks of life, science, technology, and society. And yet major challenges have erupted from almost every sphere of life on earth, most noticeably, in the sectors of biomedical sciences.

Longitudinal studies, employing repeated measurement of subject over time, play prominent role in the medical and health sciences as well as in the pharmaceutical studies. As an important strategy in modern clinical research, they provide valuable insights into both development and persistence of disease and those factors alter the course of disease development.

Research on statistical method for design and analysis of human investigations expanded in the second half of the twentieth century. Beginning in the early 1950s, the major developed countries shifted a substantial part of its research support from military to biomedical research. The national institute of health(NIH) of US government gre3w rapidly throughout the period 1950-1970. The NIH sponsored many of the important epidemiological studies and clinical trials of that period. The typical focus of these early studies was morbidity and especially, mortality. Investigators sought to identify the causes of early death and to evaluate the effectiveness of treatments for delaying death and morbidity. In such studies usually participants were seen at specific time intervals (for example one year or two years). Survival outcomes during the successive time periods were treated as independent events and modeled

using multiple logistic regression. The successful use of multiple logistic regression in this setting, and the recognition that it could be applied to case-control data, led to wide spread use of this methodology beginning in the 1960s.The analysis of time-to-event data was revolutionized by the seminal paper of Cox (1972) describing the proportional hazard model. This paper was followed by a rich and important body of work that established the conceptual basis and the computational tools for modern survival analysis.

As the research advanced, however, investigators began to follow populations of all ages over time, both in observational studies and clinical trials, to understand the development and persistence of disease and to identify factors that alter the course of disease development. This interest in the temporal pattern of change in human characteristics came at a period when advances in computing power made new and more computationally intensive approaches to statistical analysis is available at the desktop. Thus in the early 1980s, Laird and Ware proposed the use of the EM algorithm to fit a class of linear mixed effect models appropriate for the analysis of repeated measurements. Laird and Ware(1982), Jenrich and Schluchter(1986) proposed a variety of algorithms, including Fisher scoring and Newton-Raphson algorithms. Later in the decade, Liang and Zeger introduced the generalized estimating equation (GEE) in the biostatistics literature and proposed a family of generalized linear model for fitting repeated observations of binary and counted data.

Many other investigators contributed to the rapid development of methodology for the analysis of these longitudinal data. The past 25 years have seen considerable

progress in the development of statistical methods for the analysis of longitudinal data. Despite these important advances, methods for analysis of longitudinal data have been somewhat slow to move into the mainstream. In this thesis we present a comprehensive description of some stochastic models in survival/clinical trials. Our main emphasis is on the theoretical as well as practical aspects of subject matter. Although the methods are applied to problems drawn from the health sciences they apply equally to the other areas of application, for example education, psychology, and other branches of the behavioral and social sciences.

Clinical trials are prospective studies which often have time to a clinical outcome as the principal response. The dependence of this univariate measure on treatment and other factors is the subject of survival analysis. The problem that we discuss in this thesis have a direct bearing on both Survival analysis and Clinical trials

In clinical trials for assessing a medical treatment data are often collected over multiple visits of participant patients. Despite a thoughtful and well defined study protocol, frequently patients dropout before the completion of study. Resultantly censoring will be an essential part of it. In many situations like follow-up studies in organ transplant, chemotherapy and/or surgical treatment for various cancers etc., the patients are examined only at fixed regular intervals or when reporting for checkup so that the medical practitioner can observe the patient only at that specified points of time.

In type II progressive interval censoring with random removal(Type II PICR),

the individual are examined at fixed regular intervals, at each examination the number of both dropouts and failed individuals are recorded, the study will be terminated when a pre-specified number of failed individuals are observed. It inherits wonderful features of type II censoring, interval censoring and progressive censoring with the provision to discard the subjects at end of any interval at will. Sometime the removal of subjects from a clinical study become necessary when they are not suitable further. For instance, this may be due to patient's uninterest in the present treatment or due to infection of some other contiguous disease.

Practically it is difficult to initiate a study of responses to a rare treatment strategy for a large group simultaneously. The Type II PICR scheme can accommodate a patient following the same treatment strategy at any point of time. In other words, there is no restriction that they are all have to get into the study at a single point time. An investigator has the freedom even to include cases satisfying the set of criteria from old medical records. This advantage gives the censoring scheme great adoptability in medical research problems.

In this thesis we propose a clinical model based on probability structure of Generalized Exponential Distribution. It has some interesting features very similar to those of Weibull family and gamma family but a nice alternative to them in many situations. Although Weibull distribution is a popular life time distribution on account of its several advantages, the maximum likelihood estimates of the Weibull parameters may not behave properly for all parametric values even when location parameter is zero. (see Bain (1978)). Also the monotonicity of Weibull hazard function reaching

an infinite value when the shape parameter is greater than one, may not be appropriate in many situations. The Weibull family does not enjoy likelihood ratio ordering property like gamma family, making the problem of one sided hypothesis testing extremely difficult. Further the distribution of the mean of random sample from the Weibull distribution is not simple to compute though its distribution function has a single form.

Even today in medical practice, a majority of treatment decisions are made using ad hoc or heuristic strategies. There is a growing feeling among medical practitioners that the treatment decisions are too complicated to solve accurately by intuition alone. In many medical treatment, decision must be made sequential in an uncertain environment. A physician determining a course of treatment must consider patient's health as well as the best treatment decision in the future. Often decisions are to be taken in a dynamic environment. Physiological as well as physical changes in patients, may sometime contribute to the changes of the environment. Uncertain environment arises mainly due to patients respond differently even to same treatment for a disease.

Physicians always need to make subjective judgement about the treatment strategies. However a mathematical decision model that provide insight into the nature of optimal decision can aid the treatment. This is necessitated by the fact that the subject/patient often lives in varying environments during which they are subjected to varying environment conditions with significant effects on performance/health status. During a treatment period whole environment of the patient may change due to occurrence of other contagious diseases, hypertension, high blood pressure, cardiac

problems, severe climatic/seasonal changes or adopting entirely new treatment strategy on medical team's advice. When environment changes, the state of patient also changes. The deterioration and failure process therefore depends on the environment. This makes it crucial to identify an optimal treatment strategy especially for a range of multi-state disease process.

In this thesis we propose a method for determining an optimal treatment strategy using semi-Markov decision process. It is a complex survival model according to a semi-Markov process that lives in a randomly changing environment according to a semi-Markov process which affect model parameters.

In MDP models the treatment decision are taken at each of a sequence of unit time intervals or fixed epochs and the sojourn time in states has no effect on rewards or incurring costs for patient. However in health-care and other applications, decisions are taken over continuous time intervals. For instance, the decision may be administering various treatments. The sojourn time in states may depend on the duration of his/her current health status and the treatments. The MDP models might not be suitable to model such disease progression instead Semi-Markov Decision Process (SMDP) models are more appropriate. In SMDP models allow patients' state transition to occur in continuous time and allow to assume any probability distribution for sojourn time in a state.

Semi-Markov stochastic model is a useful tool for predicting the evolution of infection of infectious diseases and the probability of an infected patients survival.

This model, when compared to the most common epidemiologic data analyzes, has the following advantages:

(i) We can consider, the randomness in the different states in which the infection can evolve and also the time elapsed in each state as random;

(ii) since all the states are interrelated, any improvements are also can be considered;

(iii) a large number of disease states can be considered;

(iv) fewer and less rigid working hypotheses are needed;

(v) only raw data obtained from observations are needed, with no strong assumptions about any standard probability functions regarding the random variables analysed;

(vi) the conclusions are simply based on a list of all computed probabilities derived directly from raw data.

Semi-Markov processes were defined in the fifties independently of each other by Levy (1956) and Smith (1955). A detailed theoretical analysis of semi-Markov processes is given in Howard(1971(2)). Since then, they have been applied in a number of scientific fields including: engineering applications (systems reliability) by Howard (1971(2)), Limnios and Oprisan(2001), Janssen and Manca (2006), finance by G Di Biase *et al.*(2005), insurance, actuarial and demographic sciences by Janssen and Manca, (2006,1997), D'Amico et al (2006), respectively. Semi-Markov models have

also been employed in the field of biomedicine, for example, in applications to prevent, screen, and design cancer prevention trials, in Davidov(1999), and Davidov and Zelen(2000), respectively.

As regards the statistical analysis of semi-Markov processes, the fundamental references are Gill(1980), Andersen *et al.* (1993), Ouhbi and Limnios(1999) and, more the recent, Limnios and Ouhbi(2005) and Dabrowska and Ho(2006).

1.2 Survival Analysis- An Overview.

Survival analysis is a traditional statistical theme, but the recent surge of interest in this area is mainly due to its applications in biomedical sciences. In survival and event history analysis, one studies the time to occurrence of certain events. The field of survival analysis emerged in the 20^{th} century and experienced tremendous growth during the latter half of the century. The early efforts in development of survival analysis methodology were predominantly focussed at the estimation of the hazard function $\lambda(t)$ and the survival function $S(t)$. The developments in the field of survival analysis that have had the most profound impact on various application field are the Kaplan-Meier(1958) method for estimating the survival function, the log-rank statistic by Mantel(1966) for comparing two survival distributions and the Cox(1972) proportional hazard model for quantifying the effects of covariates on the survival time. Survival data is a term used for data measuring the time to some event. In the simplest case, the event is death, but the term also covers other events,

like occurrence of diseases, germination time of seeds etc. In industrial applications, it is typically time to failure of a unit or some component in a unit. In economics, it can be time acceptance of a job offer for an unemployed person. In demography, the event can be entering marriage.

Lifetime (survival time) data often come with a feature that creates problems called censoring in the analysis of the data. In broad sense it occurs when exact lifetime are known for only a portion of the individuals under study, the remaining lifetimes are known only to exceed certain values. Some times experiments are run over a fixed time period in such a way that an individual's lifetime will be known exactly only if it is less than some predetermined value. In such situation the data are said to be type I censored. The situation in which only the $'r'$ smallest observations in a random sample of n item are observed $(1 \leq r \leq n)$, the sample collected is said to be type II censored. It is to be noted that with type I censoring the number of exact lifetimes observed is random, in contrast to the case of type II censoring where it is fixed. Generalized type II censoring, random censoring etc., are some other types of censoring existing in literature. Lawless(2003) gives a detailed illustration on various types of censoring on survival (lifetime) data and also the parametric/non-parametric methodology for lifetime data analysis.

Some Concepts in Lifetime Distribution:

Let T be a non-negative continuous random variable representing the lifetime of individuals in some population. Let $f(t)$ denote the probability density function (p.d.f.) of T and let the distribution function be

$$F(t) = Pr(T \leq t) = \int_o^t f(x)dx$$

and the survival function

$$S(t) = 1 - F(t)$$

In case of lifetimes of manufactured items, $S(t)$ has been referred to as the reliability function.

The p.d.f.(or p.m.f.) and the distribution and the survival function are common representation of a probability distribution, the hazard function (called the failure rate earlier) is a function which is particularly useful with lifetime distributions. They describe the way in which the instantaneous probability of death for an individual with time. Often, in applications there may be qualitative information about the hazard function, which can help in selecting a life distribution model. For example, there may be reason to restrict consideration to models with non decreasing hazard functions or with hazard function having some other well-defined characteristic. In short, the hazard function represents an aspect of a distribution that has direct physical meaning and that information about the nature of the hazard function is helpful in selecting a model.

The hazard function of a life distribution which specifies the instantaneous rate of death or failure at time t, given that the individual survives up till t, is denoted by $h(t)$ and is defined as,

$$h(t) = \lim_{\triangle t->0} Pr[\frac{t \leq T < t + \triangle t / T \geq t}{\triangle t}] \quad (1.2.1)$$

$$\Rightarrow h(t) = \frac{f(t)}{S(t)}, \quad \text{provided f(t) exists.} \quad (1.2.2)$$

It now follows that,

$$log_e S(t) = -\int_0^t h(x)dx.$$

$\int_0^t h(x)dx$ is the cumulative hazard function and is denoted by $H(t)$. Then,

$$S(t) = exp(-H(t))$$

$$\text{Also } f(t) = h(t)S(t)$$

$$= h(t)e^{-H(t)}.$$

When lifetime are grouped or when 'lifetime' refers to an integral number of cycles of some sort, it may be desired to treat T as a discrete random variable. Suppose T can taken on values $t_1, t_2, \ldots$ with $0 \leq t_1 \leq t_2 \leq \ldots$ and let the probability function be $p_j = Pr(T = t_j)$ for $j = 1, 2, \ldots$. Then the survival function is,

$$S(t) = Pr(T \geq t) = \sum_{j=t_j \geq t} p_j$$

$$\text{hazard function } h(t_j) = Pr(T = t_j / T \geq t_j)$$

$$= \frac{p_j}{S(t_j)}, j = 1, 2, \ldots$$

As in continuous case, the probability, survival, and the hazard functions give equiv-

alent specifications of the distribution of T. Since $p_j = S(t_j) - S(t_{j+1})$,

$$h(t_j) = 1 - \frac{S(t_{j+1})}{S(t_j)}; j = 1, 2, \ldots$$

and,

$$S(t) = \prod_{j:t_j<t} [1 - h(t_j)].$$

Occasionally situations arises in which one would like T to have both discrete and continuous components. Special notation or definitions will not be introduced for such situations, which will be handled as they occur. No real difficulties are encountered with mixed distributions, especially if one works primarily with the survival function, which, as usual, is a monotone decreasing left-continuous function on $[0, \infty)$.

1.3 Clinical Trials-An Overview.

Treatment Effect in Clinical Trials with Dropouts

In clinical trials for assessing a medical treatment, data are often collected over multiple visits from participated patients. Statistical inference is typically carried out by first defining a measure of treatment efficacy, called treatment effect, and then testing a suitably formulated hypothesis regarding the treatment effect. When all patients complete the entire trial, one commonly used treatment effect is the mean of a patient's primary response at the end of the study or the mean of change of efficacy from baseline to the end of the study. This is motivated by the fact that

many drug products are expected to produce the maximum or clinically meaningful treatment effect at the designed end of the study (last visit). Statistical inference on this treatment effect can be either a longitudinal study (under certain statistical models) or an endpoint analysis.

Despite a thoughtful and well-designed study protocol, it is frequently the case in a clinical trial that patients drop out before the completion of the study (Heyting *et al.*, 1992). In the presence of dropout, how should the treatment effect be defined? Many people think that the mean response at the end of the study can still be defined as the treatment effect, regardless of whether there is a dropout or not. Although this definition may be suitable in some problems, it is not adequate in many situations; for example, the mean response at the end of the study is not well defined for a patient who dropped out because of death. Because dropout is frequently related to the medical treatment, the definition of treatment effect is not straightforward and its importance is often overlooked.

Definition of treatment effect in a clinical trial with dropout, is of primary importance, because how to define the treatment effect drives the statistical analysis, not vice versa. One should first carefully think about what is a relevant treatment effect and then choose a valid statistical method to assess the treatment effect. Two popular definitions of the treatment effect are study-end treatment effect and the last-observed treatment effect. In the following sections, we describe the appropriateness and the pros and cons of these definitions.

Study-end Treatment Effect

First of all, any treatment effect should be defined over the population of all randomized patients, which is the intention-to-treat principle required by all regulatory agencies. The study-end treatment effect is defined the same as in the case of no dropout, i.e., the mean response of all patients at the end of the study. If all dropout patients are still under their assigned treatments until the designed end of the study (so that the only thing missing for a dropout patient is the observation at the end of the study), this treatment effect is meaningful and clearly acceptable. Let $\mu_{s,t}^{(j)}$ denote the mean response at visit t under treatment j of a patient dropping out after visit s; where $s, t = 1, 2, ..., T$; T is the designed number of visits and $s = T$ means that a patient completes the study. Then, the study-end treatment effect is

$$\mu_{end}^{(j)} = \sum_{s=1}^{T} p_s^{(j)} \mu_{s,T}^{(j)} \tag{1.3.1}$$

where $p_s^{(j)}$ is the probability of a patient under treatment j dropping out after visit s.

There are practical situations in which it makes sense to assume that patients are under their assigned treatments after dropout. For example, treatments are still effective after patients drop out. In vaccine studies, some patients may not come back for follow-up evaluations. However, because the vaccine has been injected into their bodies, the treatment is continuing until a specific time (the end of the study).

Another example is that patients continue their assigned treatments until the end of the study although they discontinue their scheduled visits. In many problems, unfortunately, patients do not receive the assigned treatment after dropping out. If we still define the study-end treatment effect as the mean response of all patients at the end of the study, assuming that dropout patients had remained on assigned treatment until the end of the study, then this study-end treatment effect is counterfactual and is a hypothetical parameter. Thus, it will be referred to as the causal study-end treatment effect. The causal study-end treatment parameter is still given by Eq. (1.3.1), but $\mu_{s,t}^{(j)}$, $s < T$ are hypothetical parameters. The causal study-end treatment effect is viewed by many as the cornerstone of drug evaluation so far and causal inference is the main goal of many studies.

In some problems, it is of some scientific relevance; see an example in Scharfstein *et al.* (2003). It is reasonable to believe that the causal study-end treatment effect is relevant when a drug's long-term effect is of primary interest. Even though dropout patients are treated for a time period shorter than the designed time period, the effectiveness of the drug should be evaluated for the entire study period, because the drug product is expected to produce maximum clinical treatment effect at the end of the study. For example, the long-term effect is relevant if the drug product is for lowering cholesterol.

Last-observed Treatment Effect

The study-end treatment effect may be suitable when the long-term drug effect is the only concern, but not in the case where a portion of the patient population cannot wait to see the long-term effect. Claiming a drug product is efficacious in a long term is not relevant to these patients. An alternative to the study-end treatment effect is the last-observed treatment effect

$$\mu_{last}^{(j)} = \sum_{s=1}^{T} p_s^{(j)} \mu_{s,s}^{(j)} \tag{1.3.2}$$

where, according to the notation in Eq.(1.3.1), $\mu_{s,s}^{(j)}$ is the last-observed mean response for a patient under treatment j dropping out after visit s. For a patient dropping out after visit s, $\mu_{s,s}^{(j)}$ is a good summary measure for this patient if we care about his/her response at the last time he/she is treated. The parameter $\mu_{last}^{(j)}$ is also referred to as the global mean in Shao, J., Zhong, B.(2003) because it is a weighted average of stratum means when each subpopulation of patients dropping out after a particular visit is treated as a stratum.

Note that the difference between $\mu_{last}^{(j)}$ in Eq.(1.3.2) and the causal parameter $\mu_{end}^{(j)}$ in Eq. (1.3.1) is that $\mu_{s,T}^{(j)}$ is replaced by $\mu_{s,T}^{(j)}, s < T$. Unlike parameters $\mu_{s,T}^{(j)}, s < T$ parameters $\mu_{s,s}^{(j)}, s < T$ are always meaningful and can be estimated based on what we can observe. When parameters $\mu_{s,T}^{(j)}, s < T$ are not meaningful, a strong case can be made that the last-observed treatment effect is far more real and more relevant than the causal study-end treatment effect. In the case where dropout is

related to death,it is reasonable to treat the last visit prior to death as the actual study end for a dropout patient. Another situation in which the last visit prior to dropout can be treated as the actual study end is when patients recovered and stopped their medication prior to the designed study end. In these cases, an analysis not focusing on a fixed time point (visit), such as the last visit, is more reasonable. When dropout is not caused by reasons such as death or recovery, the last-observed treatment effect is a weighted average of short-term and long-term treatment effects, and it makes sense whenever both types of effects are relevant. In particular, it makes sense when we do not need to make inference on what patients do after dropout. When the long-term treatment effect is the only focus, however, the last-observed treatment effect may be inappropriate. It may be reasonable to treat the two approaches, the one focusing on the study-end parameter and the one focusing on the last-observed treatment effect, as complements to each other, rather than competitors, because one of them mainly concerns long-term treatment effects and the other is useful when short-term treatment effects are also relevant.

Dropout Pattern

In most applications, the dropout pattern, i.e., $(p_1^{(j)}, p_2^{(j)}, ..., p_T^{(j)})$ with $p_s^{(j)}$ being the proportion of dropout patients after visit s, varies with treatment j. A criticism to the use of the last-observed treatment effect $\mu_{last}^{(j)}$ last defined in Eq.(1.3.2) is its dependence on the dropout pattern. In this section, we discuss the appropriate-

ness of the use of a treatment effect parameter depending on the dropout pattern. First, note that the same criticism also applies to the study-end parameters $\mu_{end}^{(j)}$, because it depends on the dropout pattern as indicated in Eq.(1.3.1). Does it make sense to compare parameters $\mu_{last}^{(j)}$ or $\mu_{end}^{(j)}$ for different j values (treatments) when $(p_1^{(j)}, p_2^{(j)}, ..., p_T^{(j)})$ varies with j? Let us divide the entire patient population into T strata with the s^{th} stratum consisting of patients dropping out after visit s. Because sampling is not stratified $\mu_{last}^{(j)}$ and $\mu_{end}^{(j)}$ are poststratified means (McHugh, R. and Matts, J. (1983); Valliant, R. (1993)). In many problems, comparing poststratified means is of interest, although stratum patterns are different.

Another point of view is that the $p_s^{(j)}$ are intrinsic population parameters, and even if we can define a treatment effect parameter not depending on the dropout pattern, one may still question about the appropriateness of concluding the effectiveness (or noneffectiveness) of a medical treatment when dropout patterns are different for different treatments. In many statistical applications, we compare two populations using a chosen parameter, knowing that there are other parameters taking different values for different populations. In the simple two-sample problem with no missing data, for example, two treatments may be compared by using the difference of two population means, even though we know that the two populations may be different in terms of their variances or other population characteristics. Thus, if the response means of two drug products are the same but the variances of the responses are different, can we conclude that the two drug products have the same efficacy? In other words, can we compare treatment effects when dropout patterns are different?

Of course, an analysis on the dropout pattern itself may be important in comparing treatments, in addition to the inference on $\mu_{last}^{(j)}$ and $\mu_{end}^{(j)}$.

Last Observation Carry Forward.

The last observation carry forward (LOCF) has a long history of application and, in many cases, it is used mainly because there is no other established methods available. In LOCF, the missing (or nonexisting) observation at visit T for a dropout patient is imputed by "carrying forward" the patient's last observation prior to dropout, and then a standard endpoint analysis is applied by treating imputed values as if they were observed at visit T.

There are many criticisms to the LOCF method (e.g., Mallinckrodt et al.(2003a,b), Ting, N. (2000), Verbeke, G.and Mohlenberghs, G. (2000)). Now it is of interest to revisit the issue of when and why the LOCF is wrong. First, assume that the study-end treatment effect $\mu_{end}^{(j)}$ in Eq.(1.3.1) is chosen as the treatment effect parameter, regardless of whether it is a causal parameter or not. Then, treating patients' last observations prior to dropout as their observations at the end of the study is a biased imputation. The sample mean based on LOCF is in fact an unbiased estimator of the last-observed treatment effect $\mu_{last}^{(j)}$ defined in Eq.(1.3.2) (Shao, J., and Zhong, B. (2003)), $\mu_{end}^{(j)}$ defined in Eq.(1.3.1) unless $\mu_{end}^{(j)}$=$\mu_{last}^{(j)}$, which rarely occurs. Of course, statistical inference based on biased estimators is also biased.

One argument to support the LOCF method is that it provides a conservative inference when $\mu_{end}^{(j)} > \mu_{last}^{(j)}$, assuming that larger mean response indicates effectiveness of the medical treatment. There are two problems here. The first is that $\mu_{end}^{(j)} > \mu_{last}^{(j)}$ is not always true. The second is that treating imputed values as observed data may produce a bias in assessing the variability that leads to a biased inference even if $\mu_{end}^{(j)}$ $= \mu_{last}^{(j)}$ last, the standard formula of calculating the variability for statistical inference does not take into account of the fact that some data are imputed.

Next, assume that the last-observed treatment effect $\mu_{last}^{(j)}$ in Eq.(1.3.2) is chosen as the treatment effect parameter. In this case, the LOCF sample mean is an unbiased estimator of $\mu_{last}^{(j)}$. This is due to the fact that the LOCF sample mean is identical to the poststratified sample mean when the population is strati.ed according to whether patients drop out after visit $s, s = 1, ..., T$. Inference based on LOCF may still be biased, because of the previously discussed issue of not correctly assessing the variability. In some special cases, it is shown in Shao, J., and Zhong, B. (2003) and Cheng *et al.* (2005) that some statistical tests based on LOCF are asymptotically correct, because certain balance structure of data eliminates the problem of not correctly assessing the variability. In general, however, it is advised that the LOCF method should be replaced by the last observation analysis (LOAN) (Cheng *et al.*, 2005; Dawson,J.D. 1994; Dawson, J.D. and Lagakos, S.W.(1994); Shao, J., and Zhong, B. (2003); Shih, W., Quan, H. (1998)). When the parameters $\mu_{s,s}^{(j)}, s \leq T$ are of interest, we should analyze last observed data from patients rather than "carrying them forward" to the end of the study. Therefore the LOCF is entirely wrong when

the study-end parameters are used as the treatment effect. When the last-observed treatment effect is adopted, it may be nearly correct in some special cases, but the LOAN is preferred.

Missing Data and Dropout

Presence of missing observations is a problem associated with many longitudinal studies and is more acute than that in cross-sectional studies, since non response can occur at any occasion. An individual's response can be missing at one time and then be measured at a later time, resulting in a large number of distinct missingness patterns. The problem of drop out also is likely as an individual may withdraw from the study before its completion. The term missing data indicates that an intended measurement that could not be obtained and includes either of the above cases.

Missing data have three important implications. First, it makes the data unbalanced over time, since not all individuals have the same number of repeated measurements at a common set of occasions. This feature of missingness will not be of any concern for the methods described later. Second, there will be a loss of information and it causes a reduction in efficiency or a drop in precision with which changes in mean responses over time can be estimated, due to the associated loss of information. This reduction in precision is directly related to the amount of missing data and will also be influenced to a certain extend by how the analysis handles the missing data. Finally, under certain circumstances, missing data can introduce bias and thereby

lead to misleading inferences about changes in the mean response. It is this last factor, the potential for serious bias, that complicates the analysis. Hence the reason for missing data or the missing data mechanism must be carefully considered. Some reasons for missing data are relatively benign and do not complicate the analysis, where as others are not and can potentially introduce bias in the estimates of parameters.

There can be more than a single reason for missing data and these reasons may or may not be related to the response variable of interest. For example in a longitudinal study designed to study pulmonary function in school children in a district, if a child changes school district during the study because of employment relocation of parents, the missing data mechanism is unrelated to the child's pulmonary function. On the other hand, if the child moved out of school district because he developed respiratory problems, then missing is related to child's pulmonary function. When missing data mechanism is not related to the response variable of interest, the impact of missing data is relatively benign and does not complicate the analysis. In the other case greater care is required because there is potential for bias.

We review three general types of missing data mechanisms and illustrate the main distinctions between them. The mechanisms differ in terms of assumptions concerning whether missingness is related to observed and unobserved responses.

Recently there has been a great deal of attention on modelling longitudinal data subject to missingness. Modern missing data terminology is largely due to Rubin (1976) and Little and Rubin (2001). The taxonomy of missing data mechanisms de-

veloped by them for describing the assumptions concerning the dependence of the missingness process on observed and unobserved responses is widely used. Recently, Little (1993, 1995) advocate pattern-mixture models as a valuable alternative to selection models. An early reference is Glynn *et al.*(1986). The EM algorithm, a general technique for ML estimation with incomplete data, is introduced in the seminal paper by Dempster *et al.* (1977).

A useful discussion of methods for handling dropout in longitudinal studies can be found in Heyting *et al.* (1992), Little and Rubin (2001) and Schafer (1997). Ware (2003) gives "last value carried forward" imputations that are widely used to handle dropouts. Laird (1988), Little (1995) and Kenward and Molenberghs (1999) discuss various aspects of missing data issues in longitudinal studies.

Inverse probability weighted methods were first proposed in the sample survey literature by Horvitz and Thompson (1952). Robins *et al.* (1995) developed an inverse probability weighted estimating equations approach for handling missing data in longitudinal studies. Propensity score methods are described in Rosenbaum and Rubin (1983). A comprehensive description of imputation methods can be found in Rubin (1987). Molenberghs *et al.* (1998) shows that the classical taxonomy of missing data models namely MCAR, MAR, and informative missingness, which has been exclusively within a selection modelling framework, can also be applied to pattern-mixture models.

Wu and Caroll (1988) uses a probit model for informative censoring in con-

junction with linear random effects. Modifications of this approach were proposed by Degruttola and Tu (1995) and Schulchter (1992). Diggle and Kenward (1994) propose likelihood based methods for longitudinal data subject to non-ignorable missingness. Troxel *et al.* (1998) gives an extension of this method that can be used when missing data are both non-ignorable and non-monotone.

Follmann and Wu (1995) show that informative missingness is a special case of a nonignorable missing data mechanism. Various authors have proposed shared random effect models for longitudinal data subject to informative missingness. Wu and Carroll (1988) and Wu and Bailey (1989) propose methodology for repeated Gaussian data; Follmann and Wu (1995), Ten Have *et al.* (1998), and Pulkstenis*et al.*(1998) propose models for binary longitudinal responses; and Albert and Follmann (2000) propose modeling approaches for longitudinal count data. All these approaches account for informative missingness by introducing random effects that are shared between the response and missing-data processes.

Diggle and Kenward (1994) propose likelihood based methods for longitudinal data subject to non-ignorable missingness. Their model allows the missingness probability to depend on previous and current values of the longitudinal variable, thus moving beyond the MAR assumption.

Albert *et al.* (2002) present a latent autoregressive model for longitudinal binary data subject to informative missingness. In this model, a Gaussian autoregressive process is shared between the binary response and missing data processes, thereby

inducing informative missingness. The approach extends the work of Follmann and Wu (1995) and Ten Have *et al.* (1998), by developing a model for longitudinal binary data in which a Gaussian autoregressive process rather than a random effect is shared between the response and missing data mechanism.

Hierarchy of Missing Data Mechanisms

Ordinarily the missing data mechanism is not under the control of the investigators and often is not well understood. But, as it decides the validity of the inferences, assumptions are made about the missing data mechanism. The validity of the analysis depends on whether these assumptions hold.

The missing data mechanism can be thought of as a probability model for the distribution of a set of response indicator variables taking values 1 if the response is obtained and the vale 0 if otherwise. We denote the vector of response indicators by

$$\boldsymbol{R}_i = (R_{i1}, R_{i2}. \ldots, R_{in})' \tag{1.3.3}$$

with $R_{ij} = 1$ if Y_{ij} is observed and $R_{ij} = 0$ if Y_{ij} is missing. We do not consider missingness in the covariates. Given $\boldsymbol{R}_i$, $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in})'$ can be partitioned into two components $\boldsymbol{Y}_i^O$ and $\boldsymbol{Y}_i^M$, where $\boldsymbol{Y}_i^O$ denotes the vector of observed responses and $\boldsymbol{Y}_i^M$ denotes the vector of missing responses. Considering how $\boldsymbol{R}_i$ is related to $\boldsymbol{Y}_i$, three types of missing data mechanisms can be distinguished. They are

1. Missing Completely at Random (MCAR)

2. Missing at Random (MAR), and

3. Not Missing at Random (NMAR).

The type of missing data mechanism determines the appropriateness of different methods of analyses, for example, maximum likelihood,generalized least squares (GLS) or the generalized estimation equations(GEE).

Missing Completely at Random (MCAR)

Data are said to be MCAR when the probability that responses are missing is unrelated to both the observed responses and the values that would have been obtained. That is longitudinal data are MCAR when missingness in $\boldsymbol{Y}_i$ is simply the result of a chance mechanism that does not depend on either observed or unobserved components of $\boldsymbol{Y}_i$. That is $\boldsymbol{R}_i$ is independent of both $\boldsymbol{Y}_i^O$ and $\boldsymbol{Y}_i^M$ or

$$P(R_{ij} = 1/Y_{11}, Y_{12}, \ldots, Y_{1n}, \boldsymbol{X}_i) = P(R_{ij} = 1/\boldsymbol{X}_i). \tag{1.3.4}$$

That is missingness in Y_{ij} is simply the result of a chance mechanism that does not depend on observed or unobserved components of $\boldsymbol{Y}_i$.

If the data are MCAR, the observed data can be thought of as a random sample of the complete data. As a result, the moments of observed data do not differ from the corresponding moments of the complete data. Thus completers(*i.e.*, those with no missing data) can be regarded a random sample from the target population, albeit

with a small sample size than intended. Thus the subjects with missing data can be removed from the analysis. Therefore an MCAR mechanism does not require any special method of analysis.

Missing at Random (MAR)

Data are said to be MAR when the probability that responses are missing depends on the set of observed responses, but is unrelated to the specific missing values that would have been obtained. That is

$$P(\boldsymbol{R}_i/\boldsymbol{Y}_i^O, \boldsymbol{Y}_i^M, \boldsymbol{X}_i) = P(\boldsymbol{R}_i/\boldsymbol{Y}_i^O, \boldsymbol{X}_i). \tag{1.3.5}$$

For example, in the longitudinal study designed to study pulmonary function of school children in a district, suppose children moved out of school district because they developed respiratory problems. Then if the decision to relocate could be predicted based only on the recorded history of pulmonary function measurements, the missing data are MAR. However, data will not be MAR if the decision to relocate was based on some extraneous variable, unavailable to the investigator, that was predictive of the future unobserved pulmonary function measure.

If data are MAR, the distribution of $\boldsymbol{Y}_i$ in the subpopulation defined by missing data patterns is not the same as that in the target population. As a consequence, the analysis restricted to the completers is not valid or the completers are a biased sample from the target population and a complete case analysis produces biased estimates of change in mean response over time. Further, the distribution of $\boldsymbol{Y}_i^O$, the observed

components of $\boldsymbol{Y}_i$, does not coincide with the distribution of same components of $\boldsymbol{Y}_i$, in the target population. Therefore the sample means, variances and covariances based on the available data are biased estimates of the corresponding parameters in the target population. As a result, GLS no longer provides valid estimates of mean response $\boldsymbol{\beta}$ without making correct assumptions about the joint distribution of the longitudinal responses. On the other hand, ML estimation of $\boldsymbol{\beta}$ is valid when data are MAR provided the multivariate normal distribution has been correctly specified. The distinction between the MCAR and MAR mechanisms determines the appropriateness of ML estimation under the assumption of a multivariate normal distribution for the responses and GLS without requiring assumptions about the shape of the distribution. Properties of GLS require that either the data are complete or that the missing data are MCAR. If the data are MAR, GLS based only on the moments of available data can yield biased estimates of $\boldsymbol{\beta}$. In contrast, ML estimation yields valid estimates of $\boldsymbol{\beta}$ when data are MCAR or MAR, but for the latter mechanism at the cost of requiring that the joint distribution of the responses is correctly specified.

If data are MAR, the observed data cannot be regarded as a random sample of the complete data. The distribution of $\boldsymbol{Y}_i^M$ conditional on $\boldsymbol{Y}_i^O$ is the same as the conditional distribution of the corresponding observations among the complete cases, conditional on those complete cases having the same value as $\boldsymbol{Y}_i^O$. As a result, the missing values can be validly predicted using the observed data (and a model for the joint distribution).

With MAR, the missing values can be predicted using the observed data and

a model for joint distribution of $\boldsymbol{Y}_i$. But one does not need to use the model for $P(\boldsymbol{R}_i/\boldsymbol{Y}_i^O, \boldsymbol{X}_i)$ as a function $\boldsymbol{X}_i$ and $\boldsymbol{Y}_i^O$, only a model for $\boldsymbol{Y}_i$ given $\boldsymbol{X}_i$. The same is true for MCAR as well as it is a special case of MAR. That is one does not need to use the model for $P(R_i/\boldsymbol{Y}_i^O, \boldsymbol{X}_i)$ to obtain valid likelihood based inferences, only a model for $f(\boldsymbol{Y}_i/\boldsymbol{X}_i)$. Since it is common to use a model for $f(\boldsymbol{Y}_i/\boldsymbol{X}_i)$, valid likelihood based analyses can be obtained with MAR or MCAR data with no extra assumptions, other than the general statement of MAR or MCAR. For this reason MAR and MCAR are often referred to as ignorable mechanisms, the ignorability referring to the fact that $P(\boldsymbol{R}_i/\boldsymbol{Y}_i, \boldsymbol{X}_i)$ does not depend on missing observations.

The MAR assumption is far less restrictive on $P(\boldsymbol{R}_i)$ than MCAR and may be considered to be a more plausible assumption about missing data in many applications.

Not Missing at Random (NMAR)

Missing data are said to be NMAR when the probability that responses are missing is related to the specific values that would have been obtained. That is $P(\boldsymbol{R}_i/\boldsymbol{Y}_i^O, \boldsymbol{Y}_i^M, \boldsymbol{X}_i)$ depends on at least some elements of $\boldsymbol{Y}_i^M$.

An NMAR mechanism is referred to as a nonignorable missingness as the missing data mechanism cannot be ignored when the goal is to make inferences about the distribution of the complete longitudinal responses. The term informative is also

sometimes used in the sense that missingness informs us about the distribution of missing observations. Specifically the distribution of $\boldsymbol{Y}_i^M$ conditional on $\boldsymbol{Y}_i^O$ is not same as that in the completers as in the target population, but rather the distribution of $\boldsymbol{Y}_i^M$ depends on $\boldsymbol{Y}_i^O$ and on $P(\boldsymbol{R}_i/\boldsymbol{Y}_i, \boldsymbol{X}_i)$.

Common Approaches For Handling Dropout.

In this section we present a short review of some of the most commonly used methods for handling dropout in longitudinal analysis. We also discuss the assumptions about dropout required for each of the methods to yield valid inferences. We note that many traditional methods for handling missing data (e.g., complete−case analysis, imputation) became popular when the only approaches for analyzing data were ones based on complete and balanced data.

Complete-Case Analysis.

One approach to handling dropout is to simply exclude all data from the analysis on any subject who drops out. That is, a so-called complete-case analysis can be performed by excluding any subjects that do not have data at all intended measurement occasions. We must stress that this methods is very problematic and is rarely an acceptable approach to the analysis. It will yield unbiased estimates of mean response trends only when it can be assumed that dropout is MCAR. Recall that when dropout

is MCAR, the study completers are a random subsample of the original sample from the population. However, even in cases where the MCAR assumption might be tenable, a complete-case analysis is very unappealing because of the reduction in the number of subjects contributing to the analysis. A complete-case can be immensely inefficient, leading to an analysis with reduced statistical power.

Available-Data Analysis.

Another approach for handling dropout is the available−data method. This is not a single method, but a very general term that refers to a wide collection of techniques that can readily incorporate vectors of repeated measures of unequal length in the analysis. For example, standard applications of GLS or GEE approach can be considered available−data methods, since these approaches base the analysis on all of the available observations. In general, available−data methods are more efficient than complete−case methods because they incorporate the partial information obtained from those who dropout. However many available-data methods will yield valid analyses only if the conditional (i.e., conditional on $\boldsymbol{X}_i$) means and covariances of the observed components of $\boldsymbol{Y}_i$) among those who dropout coincide with the corresponding conditional means and covariances of $\boldsymbol{Y}_i$) in the target population. As a result, available−data methods will yield biased estimates of mean response trends unless dropout is MCAR. In general, for available−data (and complete-case) methods to be valid we require that dropout is MCAR.

Imputation.

A third approach, and one that is widely used in practice, is some form of imputation for the missing responses following dropout. The idea behind imputation is very simple: substitute or fill-in the values that were not recorded with imputed values. One of the chief attractions of imputation methods is that, once a filled−in data set has been constructed, a standard methods for complete data can be applied. However, methods that rely on just a single imputation, creating only a single filled-in data set, fail to acknowledge the uncertainty inherent in the imputation of the unobserved responses. Multiple imputation circumvents this difficulty. In multiple imputation the missing values are replaced by a set of m plausible values, thereby acknowledging the uncertainty about what values to impute for the missing responses. Typically, a small number of imputations, for instance, $5 \leq m \leq 10$, is sufficient to obtain realistic estimates of the sampling variability. With multiple imputation, m filled in data sets are created, producing m different sets of parameter estimates and their standard errors. These are then appropriately combined to provide a single estimate of the parameters of interest, together with standard errors that reflect the uncertainty inherent in the imputation of the unobserved responses. Specifically, a single estimate of the regression parameters is obtained by taking the arithmetic average of the estimates obtained from the m filled-in data sets. Letting $\hat{\boldsymbol{\beta}}^{(k)}$ and $\widehat{\boldsymbol{Cov}}(\hat{\boldsymbol{\beta}})^{(k)}$ denote the estimate of $\boldsymbol{\beta}$ is given by

$$\bar{\boldsymbol{\beta}} = \frac{\boldsymbol{i}}{\boldsymbol{m}} \sum_{\boldsymbol{k=1}}^{\boldsymbol{m}} \hat{\boldsymbol{\beta}}^{(k)}, \tag{1.3.6}$$

and the estimated covariance of $\bar{\beta}$ is given by

$$\frac{1}{m}\sum_{k=1}^{m}\widehat{Cov}(\hat{\beta}^{(k)}) + \left(1+\frac{1}{m}\right)\frac{1}{m-1}\sum_{k=1}^{m}\left(\hat{\beta}^{(k)}-\bar{\beta}\right)\left(\hat{\beta}^{(k)}-\bar{\beta}'\right) \qquad (1.3.7)$$

Although the latter expression for calculating the standard errors appears somewhat complicated, it simply combines two inherent sources of variability : the within-imputation variance and the between-imputation variance. The main idea behind multiple imputation is very simple; what is less clear-cut is how to produce the imputed values for the missing responses. Next, we consider some of the commonly used methods for imputing missing data.

One widely used imputation method, especially in longitudinal clinical trials, is "last value carried forward"(LVCF), occasionally referred to as "last observation carried forward" (LOCF). This is a single imputation method that fills-in or imputes the missing values following dropout with the last observed values for the subject. Despite its widespread use, it should be recognized that LVCF makes a strong, and often very unrealistic, assumption that the responses following dropout remain constant at the last observed value prior to dropout. Perhaps the only setting where this assumption might conceivably be appropriate is when dropout is due to recovery or cure. In the context of placebo-controlled longitudinal clinical trials, there appears to be some statistical folklore that LVCF yields a conservative estimate of the comparison of an active treatment versus the control. However, this is a gross misconception, and will only be true to the extent that the active treatment prior to dropout has

carry-over effects following dropout.

In many clinical trials, this is unlikely to be the case; instead, dropout from the active treatment (e.g. due to adverse side effects) might very well result in a deterioration of the response. Despite frequent and well-founded criticisms by statisticians, LVCF is still widely used to handle dropouts in clinical trials. Regulatory agencies such as the US Food and Drug Administration (FDA) seem to encourage the continuing use of LVCF as a method for handling dropouts, despite all of its obvious shortcomings. Except in very rare cases (as mentioned above), the use of LVCF is not recommend as a method for handling dropout.

Variations on the LVCF theme include baseline value carried forward and worst value carried forward. Worst value carried forward is most often used in comparisons of an active treatment to a placebo, since it is assumed to be conservative in that setting' However, both of these alternatives suffer the same difficulties as LVCF and cannot be counted on to give unbiased treatment estimates. In addition, all of the methods suffer from optimistic standard error estimates. It is easy to see that these analysis give smaller standard errors than complete-case , or even available-data estimates because they assume completes data on everyone. However they will generally give smaller standard errors than what we would expect if we had been fortunate enough to have complete data on everyones. This is because the variability of baseline measurement is usually smaller because of selection criteria in to the study, and as we move out in time, the observations tend to become more variable. Hence substituting baseline or intermediate values for final values that can be expected to give

a less variable data set. It is also true if we use worst value, since worst values are often similar especially for responses based on a scale.

1.4 Longitudinal Analysis.

A response sequence is the set of sequentially observed response variables on an individual or unit involved in a study. The response variables are outcomes measured during the course of the trial. A response may be total mortality, death from a specific cause, incidence of a disease, a complication or specific adverse effect of disease, symptomatic relief, a clinical finding, a laboratory measurement, or the cost and ease of administering the intervention. If the primary question concerns total mortality, the occurrence of deaths in the trial clearly answers the question. If the primary question involves severity of arthritis, on the other hand, extent of mobility or a measure of freedom from pain may be reasonably good indicators. In other circumstances, a specific response variable may only partially reflect the overall question, the response variable may show a change from one discrete state(living) to another (dead), from one discrete state to any of several other states (changing from one stages of disease to another), or from one level of a continuous variable to another. If the question can be appropriately defined using a continuous variable, the required sample size may be reduced .

However, the investigators need to be careful with this variable and any observed differences are clinically meaningful and relevant and that the use of a continuous

variable is not simply a device to reduce sample size.

In general, a single response variable should be identified to answer the primary question. If more than one are used, the probability of getting a nominally significant result by chance alone is increased. If several response variables give inconsistent result, interpretation becomes difficult. The investigators would then need to consider which outcome is most important. There may be circumstance when more than one primary response variables need to be looked at. This may be the case when investigators truly cannot state which of the several response variables relates most closely to the primary question.

Combining event to make up a response variable might be useful if any one event occurs too infrequently. However, that the combined events should be capable of meaningful interpretation such as being related through a common underlying condition. One kind of combination response variable involves two kinds of events. In a study of heart disease, combined events might be death from coronary heart disease plus nonfatal myocardial infraction. This is clinically meaningful since death from coronary heart disease and nonfatal myocardial infraction might together represent a measure of coronary heart disease. Difficulties in interpretation can arise if the results of each of the components in such a response variable are inconsistent. In the physicians' health study report of aspirin to prevent cardiovascular disease, there was no difference between intervention and control groups in mortality, a large reduction in myocardial infarction in the aspirin-treated group, and an increase in stroke , primarily hemorrhagic. In this case, cardiovascular mortality was the primary response

variable, rather than a combination. If it had been a combination, the interpretation of the results would have been even more difficult than it was. When a combination response variable is used, and more than one event may occur in an individual, the rules for establishing a hierarchy of events should be established in advance. Thus a fatal event would take precedence over a nonfatal event, and only the fatal event would be counted, or, in the case of two nonfatal events, the first to occur would be counted.

Regardless of whether the investigators are measuring a primary or secondary response variable, certain rules apply. First, they should define and write the questions in advance, being as specific as possible. It is essential for planning of study design and calculation of sample size. Specifying response variables and anticipated benefit in advance eliminates the possibility of legitimate criticism that can be made if the investigator looked at the data until they found a statistically significant result and then decided that the response variable what they really had in mind for the entire time.

Second, the primary response variable must be capable of being assessed in all participants. Selecting one response variable to answer the primary question in some participants and another response variable to answer the same primary question in other participants is not a legitimate practice.

Third, unless there is a combination primary response variable in which the participant remains at risk of having additional events, participation generally ends

when the primary response variable occurs. In other words, unless death is the primary response variable, the investigator may well be interested in certain events subsequent to the occurrence of the primary response variable. These events will not change the analysis of the primary response variable but may effect the interpretation of results. For example, deaths occurring after a nonfatal primary response variable, but before the official end of the trial as a whole, may be of interests.

Fourth, response variables should be capable of unbiased assessment. Truly double-blind studies gave a distinct advantage over other studies in this regard. If a trial is not double blinded , then, whenever possible, response variable assessment should be done by people who are not involved in participant follow-up and who are blinded to the identity of the study group,. Independent reviewers are often helpful. Of course, the use of blinded or independent reviewers does not entirely solve the problem of bias. Unblinded investigators sometimes fill out forms, and the participants may be influenced by these investigators,

Fifth, it is important to have response variables that can be ascertained as completely as possible. A hazard of long-term studies is that participants may fail to return for follow-up appointments.

If the response variable is one that depends on an interview or an examination and participants fail to return for follow-up appointments information will be lost. Not only it be lost, but it may be differentially lost in the intervention and control groups.

A common criticism of clinical trials is that they are expensive and of long duration. This is particularly true for trials that use the occurrences of clinical events as the primary response variable. It has been suggested that response variables that are intermediate or continuous in nature might substitute for the clinical outcomes. Thus, instead of monitoring cardiovascular mortality or myocardial infarction an investigator could examine progress of atherosclerosis by means of angiography, ultrasound imaging, or change in cardiac arrhythmia by means of ambulatory electrocardiograms. In the cancer field, change in tumor size might replace mortality. In AIDS trials, change in CD-4 lymphocyte level has been used as a response to treatment instead of incidence of AIDS or mortality in HIV-positive patients. Osteoporosis has been used as a surrogate for bone fractures.

An argument for use of these surrogate response variables is that since the variables are continuous, the sample size can be smaller and the study less expensive than otherwise. Also, changes in the variables are likely to occur before the clinical event, shortening the time required for the trial.

1.5 Censoring of Observations

Right censoring

Right censoring whereby only lower bounds on lifetime are available for some individuals, can occur for various reasons. It may be planned, as when a decision is made to terminate a life test before all items have failed, or unplanned, as when a person in prospective study is " lost to follow up" because they move away from the region where the study takes place. To obtain a likelihood function or the properties of statistical procedures based on censored data it is necessary to consider the process by which both lifetimes and censoring times arise. To do this we apparently need a probability model for the censoring mechanism. Interestingly, it turns out that the observed likelihood function for lifetime parameters takes the same form under a wide variety of mechanisms. We consider some specific types of censoring in the next section and then give a general formulation. We first introduce some notation for censored data. Suppose that n individual have lifetimes represented by random variables $T_i ... T_n$. Instead of the observed values for each lifetime, however, we have a time t_I which we know is either the lifetime or a censoring time,. Let us define a variable $= I(T_i = t_i)$ that equals 1 if $T_i = t_I$ and 0 if $T_i > t_i$ this is called the censoring or status indicator for t_i The observed data then consist of $(t_i, i), i = 1...n$. With this notation we occasionally let t_i represent either a random variable or a realized value. For a variety of censoring mechanisms the observed likelihood function takes the form

$$L \quad = \quad \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i+)^{1-\delta_i}. \tag{1.5.1}$$

The expression is derived for the most basic type of censoring, and subsequently for some other censoring mechanisms in the following sections.

Some types of Right Censoring

Several censoring mechanisms and the likelihood function obtained for each are described in this section. For simplicity we ignore covariates and assume that lifetimes T_i are independent and identically distributed,. Extensions to allow covariates are straight forward.

Type I censoring

A type I censoring mechanism is said to apply when each individual has a fixed potential censoring time $C_I > 0$ such that T_I is observed if $Ti \leq Ci$; otherwise, we know only that $T_I > C_i$ Type I censoring often arises when a study is conducted over a specified time period.

In our general notation, we have

$$t_i = min(T_i, C_i); \qquad \delta_i = 1(T_i \leq C_i) \tag{1.5.2}$$

for type I censoring. The likelihood function for a type I censored sample is based on

probability distribution of $(t_i, \delta_i), i = 1, 2, \ldots, n$. Both t_i and δ_i are random variables in(1.1.1) and their joint probability density function

$$f(t_i)^{\delta_i} Pr(T_i > C_i)^{(1-\delta_i)}. \tag{1.5.3}$$

To see this,note that the C_i are fixed constants that $t_i$can take on values $\geq C_i$, with

$$Pr(t_i = C_i, \delta_i = 0) = Pr(T_i > C_i)$$

$$Pr(t_i, \delta_i = 1) = f(t_i) \qquad t_i \leq C_i,$$

where pr in the second expression denotes either a p.d.f or p.m.f according to whether the distribution is continuous or discrete at t_i . assuming that the life times $T_i, ..., T_n$ are statistically independent, we obtain likelihood function from(1.1.2)

$$L \quad = \quad \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i+)^{1-\delta_i} \tag{1.5.4}$$

The term $S(t_i+)$ appears in Eqn.(1.5.4) because it equals $Pr(T_i > t_i)$. In general; if $S(t)$ is continuous at t_i then $S(t_i+) = S(t_i)$.

Exact sampling properties of estimates or tests based on a likelihood function of the form Eqn.(1.5.4) are generally in tractable mathematically, but standard large sample results for maximum likelihood apply and finite sample properties can be investigated by simulation.

For example, the lifetimes T_i are iid exponential with p.d.f.

$$f(t) = \frac{1}{\lambda} \exp\left(-\frac{t}{\lambda}\right); \quad \theta, t > 0$$

The cumulative distribution function F(t) is given by

$$F(t) = 1 - \exp\left(-\frac{t}{\lambda}\right); \quad t > 0$$

and survival function

$$S(t) = \exp\left(-\frac{t}{\lambda}\right)$$

Then

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^{n} (\lambda e^{-\lambda t})^{\delta_i} (e^{-\lambda t})^{1-\delta_i}. \\ &= \lambda^r \exp\left(-\lambda \sum_{i=1}^{n} t_i\right) \end{aligned} \tag{1.5.5}$$

where the $r = \sum \delta_i$ is the observed number of uncensored lifetimes, or failures. The log-likelihood function $\ell(\lambda) = logL(\lambda)$ is

$$\ell(\lambda) = r\log\lambda - \lambda \sum_{i=1}^{n} t_i. \tag{1.5.6}$$

The maximum likelihood estimate is given by solving $\frac{d\ell}{d\lambda} = 0$, and is $\hat{\lambda} = \frac{r}{\sum_{i=1}^{n} t_i}$ The exact distribution of $\hat{\lambda}$ is rather intractable, as is the distribution of the minimal sufficient statistic $(r, \sum t_i)$. For the Type I censoring scheme the censoring time C_i are specified fixed values. In many settings they are actually random. For example, in clinical trial of leukemia, individuals entered the study in a more or less random fashion according to their time of diagnosis with leukemia, so their censoring times were effectively random. In fact, the study was actually terminated early, based on the accumulating data, thus altering the original censoring times.

Independent random Censoring

A very simple random censoring process that is often realistic is one in which each individual is assumed to have a lifetime T and a censoring time C, with T and C independent continuous random variables, with survivor functions $S(t)$ and $G(t)$ respectively. All lifetimes and censoring times are assumed mutually independent, and it is assumed that $G(t)$ does not depend on any of the parameters of $S(t)$ As in the case of Type-1 censoring, $t_i = min(T_i, C_i)$ The data from observations of n individuals is assumed to consist of the pairs $(t_i, \delta_i), i = 1, 2, ...n$; the same final result is obtained if C_i is available for all $i = 1, 2, ..., n$. The p.d.f of (t_i, δ_i) is easily obtained: if $f(t)$ and $g(t)$ are the p.d.f's for T_i and C_i then

$$\begin{aligned} Pr(t_i = t, \delta_i = 0) &= Pr(C_i = t, T_i > C_i) \\ &= g(t)S(t) \\ Pr(t_i = t, \delta_i = 1) &= Pr(T_i = t, T_i \leq C_i) \\ &= f(t)G(t) \end{aligned}$$

These can be combined into single expression

$$Pr(t_i = t, \delta_i) = [f(t)G(t)]^{\delta_i}[g(t)S(t)]^{1-\delta_i}.$$

and thus the distribution of $(t_i, \delta_i), i = 1, ..., n$ is

$$\prod_{i=1}^{n}[f(t)G(t)]^{\delta_i}[g(t)S(t)]^{1-\delta_i}.$$

Since $G(t)$ and $g(t)$ do not involve any of the parameter in $f(t)$, they can be neglected and the likelihood function taken to be

$$L = \prod_{i=1}^{n} f(t)^{\delta_i} S(t)^{1-\delta_i}$$

which is of the same form as Eqn.(1.5.4). The earlier result for Type-I censoring can in fact be considered as a special case of this if we allow the C_i to have degenerate distributions, each with mass at one fixed point. Another approach that leads directly to this likelihood function is to argue that if $G(t)$ and $g(t)$ do not involve any parameters of $f(t)$, then $C_1, ..., C_n$ are ancillary and One should condition on the realized censoring times when making inferences about the distribution of T. This takes us back to the Type I censoring framework. A point to note is that although it may be desirable to make inferences conditional on the C_i in any given situation, the properties of the procedures average over the distribution of the C_i may be of interest when planning studies, and in some applications.

Although the independent random-censorship model is often reasonable, in many situations the censoring process is linked to the failure time process. Suppose, for example, that the termination date for a medical trial is not fixed before the study commences, but is chosen later, with the choice influenced by the results of the study up to that time. In such instances, it may be difficult to write down a model that fully represents the process under study. Fortunately, the likelihood function Eqn.(1.5.4) is still applicable in many such complicated situations.

Type II Censoring

In Type-II censoring only the r smallest lifetimes $t_{(1)} < t_{(2)} < ... < t_{(r)}$ in a random sample of n are observed; $1 \leq r \leq n$. This censoring scheme when n individuals start on study at the same time, with the study terminating once r failures (or lifetimes) have been observed.

For continuous distributions we can ignore the possibility of ties and denote the r smallest lifetimes as $T_{(1)} < T_{(2)} < ... < T_{(r)}$. If the T_i have p.d.f $f(t)$ and survival function $S(t)$ then from general results on order statistics,the joint p.d.f of $T_{(1)}, T_{(2)}, ..., T_{(r)}$. is

$$\frac{n!}{(n-r)!}\left\{\prod_{i=1}^{r} f(t_{(i)})\right\} S(t_{(r)})^{n-r} \tag{1.5.7}$$

The likelihood function is based on Expn.(1.5.7). By dropping the constant $\frac{n!}{(n-r)!}$ and noting that in terms of the (δ_i, t_i) notation we have $\delta_i = 0$ and $t_i = t_(r)$. For those individuals whose lifetimes are censored. We see that Expn.(1.5.7) gives a likelihood of the same form Eqn.(1.5.4) as for Type I censoring. The sampling properties are, however, different in finite samples.

When we consider type II censored exponential lifetimes, the log-likelihood is same as Eqn.(1.5.6)

$$\ell(\lambda) = r\log\lambda - \lambda\left[\sum_{i=1}^{n} t_{(i)} + (n-r)t_{(r)}\right]$$

and likelihood estimate for λ can be written as $\hat{\lambda} = r/W$ where

$$W = \sum_{i=1}^{n} t_{(i)} + (n-r)t_{(r)}$$

Since r is fixed, the statistic W is sufficient for λ and it is shown that $2\lambda W = 2r\lambda/\hat{\lambda} \sim \chi^2_{(2r)}$, a chi-squared distribution with $2r$ degree of freedom. This allows exact confidence intervals and tests for λ to be developed.

Progressive Type II Censoring

Progressive Type II censoring is a generalization of Type II censoring. In this case, the first r_1 failures in a life test of n items are observed; then n_1 of the remaining $n - r_i$ unfailed items are removed from the experiment, leaving $n - r_1 - n_1$ items still present. When a further r_2 items have failed,n_2 of the still unfailed items are removed, and so on. The experiment terminates after some prearranged series of repetitions of this procedure.

This scheme is of more theoretical than practical interest, but let us obtain the likelihood function assuming that lifetimes are independent and identically distributed (i.i.d) with p.d.f.$f(t)$ and survivor function $S(t)$ For simplicity we suppose the censoring has only two stages: at the times of the r_1th failure, n_1 of the remaining $n - r_1$ unfailed items are randomly selected and removed. The experiment then terminates when a further r_2 items have failed. At this point there will be $n - r_1 - n_1 - r_2$ items still unfailed. The observations in this case are the r_1 failure times $T_{(1)} < T_{(2)} < ... < T_{(r_1)}$ in first stages of the experiment and the r_2 failure times

in the second stage of the experiment which we will denote by $T^*_{(1)} < T^*_{(2)} < ... < T^*_{(r_2)}$

.

The distribution of the data can be written as

$$g_1(t_1, ..., t_{r_1}) g_2(t^*_1, ..., t^*_{r_1} | t_1, ..., t_{r_1}) \tag{1.5.8}$$

where g_1 and g_2 represent p.d.f.'s of the variable indicated. The joint p.d.f. $g_1(t_1, ..., t_{r_1})$ is given by Expn.(1.5.7) with $r = r_1$. To write down the second term in Expn.(1.5.8), we observe that given $t_1, ..., t_{r_1}$, the lifetimes of the items left in the experiment have a left-truncated distribution with p.d.f. and survival functions as

$$f_1(t) = \frac{f(t)}{S(t_{(r_1)})}, \qquad S_1(t) = \frac{S(t)}{S(t_{(r_1)})} \qquad t \geq t_{(r_1)}$$

respectively. Thus $T^*_{(1)}, ..., T^*_{(r_2)}$ are the $r_2$smallest observations in the random sample of size $n - n_1 - r_1$ from this truncated distribution by Expn.(1.5.7), the second term in Expn.(1.5.8) is therefore

$$\frac{(n - r_1 - n_1)!}{(n - r_1 - n_1 - r_2)!} \left\{ \prod_{i=1}^{r_2} f^*_1(t_{(i)}) \right\} \left[S_1(t^*_{(r_2)}) \right]^{n - r_1 - n_1 - r_2} \tag{1.5.9}$$

Combining two parts of the Expn.(1.5.8), we obtain the likelihood function as

$$c \left\{ \prod_{i=1}^{r_1} f(t_{(i)}) \right\} \left[S(t_{(r_1)}) \right]^{n_1} \left\{ \prod_{i=1}^{r_2} f^*(t_{(i)}) \right\} \left[S(t^*_{(r_2)}) \right]^{n - r_1 - n_1 - r_2} \tag{1.5.10}$$

where $c = n!(n - r_1 - n_1)!/[(n - r_1)!(n - r_1 - n_1 - r_2)!]$.

The above likelihood function then will be form the base for the further inferential procedures in progressive type II censoring case.

Progressive Type II Interval Censoring

In survival analysis, clinicians are often interested in estimating the corresponding lifetime distribution after a treatment is administered to test subjects. In setting up these life tests, censoring is often adopted due to time and cost constraints. Type-II censoring is used to ensure a pre-assigned number of failures, say m, observed at the end of the test. However, practical applications may dictate that it is impossible or undesirable to conduct this type of continuous monitoring. For instance, patients or test subjects come back for treatment or diagnosis at regular intervals. Thus, the observed data are interval censored. For interval censoring, the length of the successive inspections and the number of inspection k are pre-determined. Therefore, unlike Type-II censoring, one would have no idea of how many failures would occur at the end of the kth inspection. Too few failures would provide insufficient information about the tail performance of the lifetime distribution. Details on these censoring schemes can be found in Bain *et al.*(1991),Lawless(1982)and Meeker(1998). Some recent developments in the use of interval censoring to lifetime analysis can be found in Aggarwala(2002),Sun, J.G(2001), Lim *et al.*(2002).

Tse *et al.*(2002) studied a new type of censoring scheme, namely Type-II interval censoring, which integrates the specific features of Type-II censoring and interval censoring. In particular, suppose that n subjects are selected in a life test. Inspections are conducted at pre-determined intervals and the number of "failures" occurring between two successive inspection times are recorded. The test will be terminated

when the total number of failures is greater than or equal to a pre-assigned number m. Adopting this censoring scheme, an experimenter can assure that there is at least m failures observed before the test is terminated.

We assume that the survival time T follows a exponential distribution with parameter θ The probability density function of T is given by

$$f(t) = \frac{1}{\theta}\exp\left(-\frac{t}{\theta}\right); \quad \theta, t > 0$$

The cumulative distribution function F(t) is given by

$$F(t) = 1 - \exp\left(-\frac{t}{\theta}\right); \quad t > 0$$

Suppose there are n subjects are randomly selected for the study and when specified number or percentage of total m (say) or more subjects are failed, the study will be terminated. Let $t_i, t_2, ...t_k, ...$ be the predetermined inspection times and t_0=0. Under a integrated Type II interval censoring scheme, the study is terminated after the k^{th} inspection time if the total number of failed subjects is equal to or more than m. At the i^{th} inspection, d_i failed subjects are observed . In other words, d_i is the number of failed subjects between any two successive inspections at t_{i-1} and t_i. Thus, d_i's are random variables obtained from the study. Let us denote $Y_j = \sum_{i=1}^{j} d_i$ and $t_{k+1} = \infty$ the total number of failed subjects observed upto the j^{th} inspection time t_j. If $Y_{k-1} < m$ and $Y_k \geq m$, for the predetermined integer m , $0 \leq m \leq n$; the test is terminated at the k^{th} inspection time t_k . Denote $D = (d_1, d_2, , d_k)$ and where k is random and corresponds to the last inspection time t_k. the likelihood function of D

is given by

$$L(d_1, d_2, , d_k) = \frac{n!}{r_1! r_2! ... r_{k+1}!} \prod_{i=1}^{k+1} (p_i - P_{i-1})^{r_1} \tag{1.5.11}$$

where $p_0 = 0; \quad p_{k+1} = 1; \quad p_i = p(T \leq t_i) = 1 - \exp\left(-\frac{t_i}{\theta}\right) \quad and$

$r_{k+1} = n - \sum_{i=1}^{k} r_1$. Note that k should be treated as a random variable in Eqn.(1.5.11).

The MLE of θ is found by maximizing the likelihood function L given in Eq. 1.5.11. After taking logarithm of L, the likelihood equation is given by

$$\frac{dlnL}{d\theta} = \sum_{i=1}^{k+1} \frac{r_i}{p_i - P_{i-1}} \left(\frac{dp_i}{d\theta} - \frac{dp_{i-1}}{d\theta} \right) \tag{1.5.12}$$

where $\frac{dp_i}{d\theta} = -\frac{t_i}{\theta^2} \exp\left(-\frac{t_i}{\theta}\right)$.

Standard numerical methods such as Newton-Raphson can be used to find the solution of this equation. Furthermore, the Fisher information, denoted by $I(\theta)$ is given by

$$I(\theta) = E\left(\frac{-d^2 lnL}{d\theta^2} \right) \tag{1.5.13}$$

1.6 An Overview of The Main Contribution of The Thesis.

Some stochastic models in survival and clinical trials are described in this thesis. It is organized into five chapters and each chapter is divided into different sections. In the *introductory Chapter*, a general introduction to the thesis is given. Section 2 provides an overview of survival analysis and basic ideas. Next two sections give

general idea about clinical trials and longitudinal analysis. The section 5 gives brief account of important right censoring schemes. A special discussion on progressive type II censoring schemes is also given here. In clinical trials medical treatment data are often collected over multiple visits of participant patients. Despite a thoughtful and well defined study protocol frequently patients dropout before the completion of study. Resultantly censoring will be an essential part of it. In each censoring schemes the corresponding likelihood function is indicated for further inference procedure. The last section of this chapter provides an overview of the main contribution of thesis

In *Chapter 2* we describe the analytical tools used in the thesis, viz. Markov Decision Process(MDP), Markov Renewal process (MRP), semi-Markov process(SMP) and semi-Markov decision process(SMDP). The Markov decision processes are simple yet powerful models for sequential decision problems. In these models, we assume that there is a state space; at each time the system occupies a certain state, and the decision maker, or controller, has a set of feasible actions for that state that can be applied. Semi-Markov stochastic model is a useful tool for predicting the evolution of infection of infectious diseases and the probability of an infected patients survival. In SMDP models allow patients' state transition to occur in continuous time and allow to assume any probability distribution for sojourn time in a state.

In *Chapter 3* we present a method for analysing longitudinal data that imposes minimal structure or restrictions on the mean responses over time and on the covariance among the repeated measures. The method focusses on analysing response profiles and can be applied to longitudinal data when the design is balanced, with

the timing of the repeat measures common to all individuals in the study. This is method of modeling mean response. Although the method is applied to problems drawn from the health sciences they apply equally to the other areas of application , for example finance, education, psychology, and other branches of the behavioral and social sciences.

We propose a clinical study model based on probability structure of Generalized Exponential Distribution in *Chapter 4*. The Generalized Exponential Distribution has some interesting features very similar to those of Weibull family and gamma family but a nice alternative to them in many situations. Although Weibull distribution is a popular life time distribution on account of its several advantages, the maximum likelihood estimates of the Weibull parameters may not behave properly for all parametric values even when location parameter is zero(see Bain (1978)). Also the monotonicity of Weibull hazard function reaching an infinite value when the shape parameter is greater than one, may not be appropriate in many situations. The Weibull family does not enjoy likelihood ratio ordering property like gamma family, making the problem of one sided hypothesis testing extremely difficult. Further the distribution of the mean of random sample from the Weibull distribution is not simple to compute though its distribution function has a single form. This is a Generalized Exponential Model under a more flexible and practical censoring scheme namely Type II progressive interval censoring with random removals (PICR). In type II progressive interval censoring with random removal(Type II PICR), the individual are examined at fixed regular intervals, at each examination the number of both dropouts and failed

individuals are recorded, the study will be terminated when a pre-specified number of failed individuals are observed. It inherits wonderful features of type II censoring, interval censoring and progressive censoring with the provision to discard the subjects at end of any interval at will. Sometime the removal of subjects from a clinical study become necessary when they are not suitable further. Maximum likelihood estimation of parameters of the generalized exponential model is discussed and their properties are studied in this Chapter. An illustrative example towards the application of the model is also given here. The generalized exponential model is suggested as a better alternative for analysis of life time data.

In *Chapter 5* we describe a Semi-Markov decision process(SMDP) modeling in the contest of medical treatment wherein the decisions are often sequential and uncertain. The subject/patient lives in varying random environments, imparting significant effects on performance/health status. The environment is modelled as a Semi-Markov Process and in each environment state, the patient goes through several states of disease according to a Semi-Markov Process. In an environment $'k'$ when the patient state is $'i'$, one of the following two actions are available: continue the present treatment strategy (C) with a given cost rate $h^k(i)$ or initiate a rejuvenating treatment strategy (R) with a cost rate $c^k(i)$, In this complex model the optimal strategy is found out minimizing the expected discounted total cost. In section 2, stage wise prognosis of three diseases; smallpox, liver Disease and Alzheimer's Disease is discussed. A special case of Markov environment is discussed indicating the feasibility of the computation of optimal policy. A numerical illustration is also provided to

support the viability of the analysis and results. The model provide a useful and flexible representation of acute and chronic events and can be used to explore the economic impact of changes in therapy. Semi-Markov stochastic model is a useful tool for predicting the evolution of infection of infectious diseases and the probability of an infected patients survival. In SMDP models allow patients' state transition to occur in continuous time and allow to assume any probability distribution for sojourn time in a state. The method given can be utilized to make optimal decision in variety of problem in diversified field because its generic nature.

Many mathematical illustrations appeared in this thesis are performed using MATHLAB. Recursive computations are by C++ programme. A computer programme in C++ used to determine the optimal treatment strategy in numerical example given in Chapter 5, is appended at the end of the thesis.

We conclude the thesis pointing out the salient features of our study and scope for further work. A fairly comprehensive bibliography on the topic of interest is given at the end.

Chapter 2

SOME ANALYTICAL TOOLS USED IN THE THESIS

2.1 Introduction

In this chapter we describe the analytical tools used in this thesis. They are Markov Decision Processes(MDP), Markov Renewal process (MRP), semi-Markov process(SMP) and semi-Markov decision process(SMDP). The Markov decision processes are simple yet powerful models for sequential decision problems. In these models, we assume that there is a state space; at each time the system occupies a certain state, and the decision maker, or controller, has a set of feasible actions for that state that can be applied. At the next time, the state changes according to some probability distribution which depends only on the current state and action, and does not depend on the past. MDPs are also called controlled Markov chains in the literature, and have a wide range of application areas.

2.2 Markov Decision Processes (MDP)

At the first phase we describe the theory and computational methods developed for MDPs with a finite number of states and actions, and in a discrete-time context, by which we mean that the state transitions occur and the actions are applied at integer times 0, 1, ... Now, we will focus on the probabilistic aspect of our problem formulation: how is the uncertainty modeled?; what is the corresponding mathematical object that we are dealing with?; what is the form of the optimization problems? and what does the Markovian assumption on the state evolution imply?.

Model Parameters

The model of an MDP specifies parameters relating to system dynamics and cost:

- a state space S ;
- for every state a set of feasible actions, which can be jointly represented by the set

  $$\{(s, U(s)), \forall s \in S\},$$

  where $U(s)$ is the set of feasible actions at state s; we define

  $$U = \bigcup_{s \in S} U(s)$$

  and call it the action or control space;
- a set of state transition probabilities,

  $$\{p_{ij}(u), \forall\, i \in S, u \in U(i)\};$$

  where $\sum_{j \in S} U(j) = 1$ and
- a per-stage cost function, $c_i(u)$.

Here the state space S and the action space $U(s)$ are assumed to be finite. The economic consequences of the decisions taken at the decision epochs are reflected in receiving a lump sum reward (or pays a cost). This controlled dynamic system is called a discrete-time Markov decision model when the Markov property is satisfied. Note that the one-step costs $c_i(u)$ and the one-step transition probabilities $p_{ij}(u)$ are assumed to be time homogeneous.

Stationary policies

A policy or rule for controlling a system is a prescription for taking actions at each decision epoch. A stationary policy π is a rule that always prescribe a single action π_i whenever the system is found in state i at a decision epoch.

We define for $n = 0, 1, ...$

X_n = the state of the system at the nth decision epoch.

Under he given stationary policy π, We have

$$P\{X_{n+1} = j / X_{n+1} = i\} = p_{ij}(\pi_i),$$

regardless of past history of the system up to time n. Hence under a given stationary policy π the stochastic process $\{x_n\}$ is a discrete-time Markov chain with one step transition probabilities $p_{ij}(\pi_i)$. This Markov chain incurs a cost $c_i(\pi_i)$ each time the system visits the state i. Thus we can invoke results from Markov chain theory to specify the long-run average cost per unit time under a given stationary policy.

Average cost for a given stationary policy

For a given stationary policy π, we denote the n-step transition probabilities of the corresponding Markov chain $\{x_n\}$ by

$$p_{ij}^n(\pi) = P\{X_n = j / X_0 = i\}, \quad i, j \quad \in S \quad and \quad n = 1, 2, ...$$

where $p_{ij}^1(\pi) = p_{ij}(\pi)$.

By Chapman-Kolmogorov equations,

$$p_{ij}^n(\pi) = \sum_{k \in S} p_{ik}^{n-1}(\pi) p_{kj}(\pi_k), \quad n = 2, 3....$$

Also we define the expected cost function $V_n(i, \pi)$ by

$V_n(i, \pi)$ = the total expected costs over the first n decision epochs when the initial state is i and the policy π is used.

Thus, we have

$$V_n(i, \pi) = \sum_{t=0}^{n-1} \sum_{j \in S} p_{ij}^t(\pi) c_j(\pi_j) \tag{2.2.1}$$

where $p_{ij}^0(\pi)$=1 for $j = i$ and $p_{ij}^0(\pi)$=0 for $j \neq i$. Next we define the average cost function $g_i(\pi)$ by

$$g_i(\pi) = lim_{n \to \infty} \frac{1}{n} V_n(i, \pi), i \in S$$

The long run average expected cost per unit time is independent of initial state i when it is assumed that the Markov chain $\{X_n\}$ corresponding to policy π has no two disjoint closed sets.

In the unichain case we can write

$$g_i(\pi) = g(\pi), i \in S$$

Then it follows by the ergodic theorem,

$$g(\pi) = \sum_{j \in S} c_j(\pi_j) E_j(\pi),$$

where $\{E_j(\pi), j \in S\}$ is the unique equilibrium distribution of Markov chain $\{X_n\}$. The $E_j(\pi)$'s are the unique solution to the system of linear equations

$$E_j(\pi) = \sum_{k \in S} p_{kj}(\pi_k) E_k(\pi), j \in S$$

$$\sum_{j \in S} E_j(\pi) = 1$$

Moreover, for any $j \in S$,

$$E_j(\pi) = lim_{m \to \infty} \frac{1}{m} \sum_{n=1}^{m} p_{ij}^n(\pi) \quad \forall \quad i \in S \tag{2.2.2}$$

We have that $g(\pi)$ is the expected value. Also with probability 1,the long-run actual average cost per unit time $= g(\pi)$ independently of the initial state.

Average cost optimal policy

The optimization problem is now to find a policy with the minimum cost, with respect to the chosen cost criterion, for a given or every initial state.

A stationary policy π^* is said to be average cost optimal if

$$g_i(\pi^*) \leq g_i(\pi)$$

for each stationary policy π uniformly in the state i. It is stated without proof that an average cost minimal policy π^* always exists. Moreover, policy π^* is not only optimal among the class of stationary policies but it is also optimal among the class of all conceivable policies (see ref: Derman(1970)).

2.3 Markov Renewal Process and Semi-Markov Processes

Some systems follows a Markov property not at all point of time but only for a special increasing stopping times. These times are the state changes times of the considered stochastic process.

we present below basic results and definitions of a Markov renewal process.

Consider an at most countable set S say, a two-dimensional stochastic process $(X,T) = (X_n, T_n, n \in \mathbb{N})$ where the random variable (r.v.) X_n takes values in S and the r.v. T_n takes values in $\mathbb{R}_+$ and satisfies $0 = S_0 \leq S_1 \leq S_2 \leq \ldots$

Definition 2.3.1 *The stochastic process (X,T) is called a Markov Renewal Process (MRP) if it satisfies the following relation*

$$\begin{aligned} P[X_{n+1} = j, T_{n+1} - T_n \leq t | X_0, ..., X_{n-i}, X_n = i; T_0, ..., T_n] \\ = \quad P[X_{n+1} = j, T_{n+1} - T_n \leq t | X_n = i] \\ = \quad Q_{i,j}(t) \end{aligned}$$

for all $n \in \mathbb{N}$,$j \in S$ and $t \in \mathbb{R}_+$

The set S is called state space of the MRP the function $Q_{i,j}(t)$ form semi-Markov kernel.

From these relations it is clear that (X_n) is a Markov chain with state space S

and transition probability $P(i,j) = Q_{i,j}(\infty)$. It is called embedded Markov chain. For every $i \in S$, we have $P(i,i) = 0$.

Now, we define, the counting process $\{N(t), t \geq 0\}$ associated to the point process $\{T_n, n \geq 0\}$ ie. for each time $t \geq 0$ the r.v $N(t)$ is

$$N(t) := sup\{n : T_n \leq t\}$$

and define continuous time process $Z = \{Z(t), t \in \mathbb{R}\}$ by

$$Z(t) := X_{N(t)}.$$

Then the process Z is called semi-Markov process (SMP). Also define

$$\begin{aligned} P_{i,j}(t) = &\quad P[Z(t) = j | Z(t) = 0] \\ H_i(t) = &\quad \sum_{j \in S} Q_{i,j}(t), \\ m_i = &\quad \int_0^\infty [1 - H_i(u)] du. \end{aligned}$$

We have the following particular classes of the MRP

(1) Discrete time Markov chain

$$Q_{i,j}(t) = P(i,j) \quad I\{t \geq 0\}, \quad \text{for all } i,j \in S, \text{ and } t \geq 0.$$

(2) Continuous time Markov chain

$$Q_{i,j}(t) = P(i,j) \quad (1 - e^{-\lambda(i)t}) \quad \text{for all } i,j \in S, \text{ and } t \geq 0.$$

(3) Renewal Process :

(a) Ordinary: It is an MRP with two states $S = \{0, 1\}$,

P(0,1)=P(1,0)=1 and $Q_{01}(.) = F(.)$,where F is the common distribution function of the inter-arrival times of the renewal process.

(4) Modified or delayed : It is an MRP with three states $S = \{0, 1, 2\}$,

P(0,1)=1,P(1,2)=P(2,1)=1 and 0 elsewhere and $Q_{01}(.) = F_0(.)$, $Q_{12}(.) = Q_{21}(.) = F(.)$, where F_0 is the common distribution function of the first arrival time and F is the common distribution function of the inter-arrival times of the renewal process.

(5) Alternating: It is an MRP with two states $S = \{0, 1\}$,

P(0,1)=P(1,0)=1 and 0 elsewhere and $Q_{01}(.) = F(.)$, $Q_{10}(.) = G(.)$ where F and G are the common distribution function corresponding to the odd and even inter-arrival times .

For a systematic account of these topic, covering the Markov renewal equation and its solution one may refer to Cinlar(1975). As an extension of the basic classical limit theorems in probability theory to the semi-Markov setting those are available in the literature, some limit theorems useful for reliability/survival analysis.

As noted above, in the SMP environment, two random variables run simultaneously.

$$X_n : \Omega \to S, \quad T_n; \Omega \to R, n \in N.$$

X_n with state space, say, $S = \{S_1, ..., S_m\}$ represents the state at the nth transition.

In the health care environment, the elements of S represent all the possible stages in which the disease may show level of seriousness. T_n, with state space equal to $\mathbb{R}$, represents the time of the nth transition. In this way, we can not only consider the randomness of the states but also the randomness of the time elapsed in each state. The process (X_n,T_n) is assumed to be a homogeneous Markovian renewal process.

Furthermore, it is necessary to introduce the probability that the process will leave state i in a time t as

$$H_i(t) = P[T_{n+1} - T_n \leq t | X_n = i].$$

Obiviously,

$$H_i(t) = \sum_{j=1}^{m} Q_{i,j}(t).$$

It is now possible to define the distribution function of the waiting time in each state i, given that the state successively occupied is known,

$$G_{i,j}(t) = P[T_{n+1} - T_n \leq t | X_n = i, X_{n+1} = j]$$

Obviously, the related probabilities can be obtained by means of the following formula:

$$G_{ij}(t) = \begin{cases} \frac{Q_{i,j}(t)}{P(i,j)}, & \text{if } P(i,j) = 0 \\ 1, & \text{if } P(i,j) \neq 0 \end{cases}$$

The main difference between a continuous time Markov process and a semi-Markov process lies in the distribution functions $G_{ij}(t)$. In a Markov environment this function must be a negative exponential function. On the other hand, in the semi-Markov case, the distribution functions $G_{ij}(t)$ can be of any type. This means that the transition intensity can be decreasing or increasing.

If we apply the semi-Markov model in the health care environment, we can consider, by means of the $G_{ij}(t)$, the problem given by the duration of the time spent inside one of the possible disease states.

Now the homogeneous SMP, $Z = \{Z(t), t \in \mathbb{R}\}$ represents, for each waiting time, the state occupied by the process

$$Z(t) = X_{N(t)}, \quad \text{where} \quad N_t = sup\{n : T_n \leq t \} .$$

The transition probabilities are defined in the following way:

$$\phi_{ij}(t) = P[Z(t) = j | Z(0) = i]$$

. They are obtained by solving the following evolution equations:

$$\phi_{ij}(t) = \delta_{i,j}(1 - H_i(t)) + \sum_{\beta=1}^{m} \int_0^t Q'_{i\beta}(\vartheta)\phi_{\beta j}(t - \vartheta)d\vartheta, \tag{2.3.1}$$

where $\delta_{i,j}$ represents the Kronecker delta.

The first addendum of formula (2.3.1) gives the probability that the system does not undergo transitions up to time t given that it was in state i at an initial time 0. In predicting the disease evolution model, it represents the probability that the infected patient does not shift to any new stage in a time t. In the second addendum, $Q'_{i\beta}(\vartheta)$ is the derivative at a time ϑ of $Q_{i,\beta}(\vartheta)$ and it represents the probability that the system remained in a state i up to the time ϑ and that it shifted to state β exactly at a time ϑ. After the transition, the system will shift to state j following one of all the possible trajectories from state β to state j within a time $t - \vartheta$. In disease evolution model, it means that up to a time an infected subject remains in the state i. At the

time ϑ, the patient moves into a new stage β and then reaches state j following one of the possible trajectories in some time $t - \vartheta$.

2.4 Semi-Markov Decision Processes (SMDPs)

In Markov decision processes (MDPs) the decisions are taken at each of a sequence of unit time intervals. The Semi-Markov decision processes (SMDPs) generalize MDPs by allowing the decision maker to choose actions whenever the system state changes, modeling the system evolution in continuous time and allowing the sojourn time in a particular state to follow an arbitrary probability distribution. The system state may change several times between decision epochs; only the state at a decision epoch is relevant to the decision maker. If transition times between states are distributed exponentially, we refer to the process as continuous-time Markov decision process (CTMDP)

The model of an SMDP specifies parameters relating to system dynamics and cost are specified similarly as in MDP:

- At a decision epoch the system occupies a state $s \in S$, S is the state space ;
- for every state s a set of feasible actions called action or control space $U(s)$, which can be jointly represented by the set

$$\{(s, U(s)), \forall s \in S\},$$

Also we define

$$U = \bigcup_{s \in S} U(s);$$

- a set of state transition probabilities,

$$\{Q_{ij}(t|u), \forall\ s \in S, u \in U(s)\};$$

In most applications $Q_{ij}(t|u)$ is not provided directly, but instead $F_i(t|u)$ and $p_{ij}(t|u)$ are used. The $F_i(t|u)$denotes the probability that the next decision epoch occurs within t time units, given that action $u \in U$ is chosen in state i. The quantity $p_{ij}(t|u)$ denotes the probability that the system occupies state j in t time units after the decision epoch given i and u. (If the natural process does not change state until the next decision epoch, $p_{ij}(t|u) = 1$ for all t).

The following assumption is needed to guarantee that there will not be an infinite number of decision epochs within finite time:

There exists $\epsilon > 0$ and $\delta > 0$ such that $F_i(t|u) \leq 1 - \epsilon$ for all $u \in U$ and $i \in S$

Two special cases for$F_i(t|u)$:

- When $F_i(t|u) = 1 - e^{-\beta(i,u)t}$

  we refer to this as a continuous-time Markov decision process.

- When $F_i(t|u)$=0 ,if $t \leq t'$; =1 if $t > t'$

  for some fixed t for all i and u, we obtain a discrete-time MDP.

When decision maker chooses action u in state i, he receives a lump sum reward (or pays a cost) $g(i,u)$ In addition to that, he accrues a reward (or incurs a cost) at rate $c(j',i,u)$ as long as the natural process occupies state j, and action u was chosen in state i at preceding decision epoch.

Decision rules and policies

In SMDP, the decision rules may be deterministic or randomized, Markovian or history dependent. Let us define a policy $\pi = (d_1, d_2...)$ decision rule d_1 is used at $t_0 = 0$

In discounted model for a policy $\pi \in \Pi$, let us denote $\upsilon_\alpha^\pi(s)$ as the expected infinite-horizon discounted reward with the discount factor α, given that the process occupies state s in the first decision epoch:

$$\upsilon_\alpha^\pi(s) \equiv E_s^\pi \left(\sum_{n=0}^{\infty} e^{-\alpha T_n} \left[g(X_n, Y_n) + \int_{T_n}^{T_{n+1}} e^{-\alpha(t-T_n)} c(w_t, X_n, Y_n) dt \right] \right)$$

where $T_0, T_1, ...$ represent the times of successive decision epochs, w_t denotes the state of the natural process at time t. Define the value of a discounted SMDP by

$$\upsilon_\alpha^*(s) \equiv sup_{\pi \in \Pi} \upsilon_\alpha^\pi(s)$$

The goal is to find a policy π^* for which $\upsilon_\alpha^{\pi^*}(s) = \upsilon_\alpha^*(s)$

In other words, the objectives in these problems, is to maximize the expected

discounted reward or minimize the expected discounted total costs and obtain the best policy. We confine ourselves to the optimality criterion of the total expected discounted cost. This criterion is found to be more appropriate in many applications, particularly in biomedical and reliability studies. An alternative criterion is the long run average cost per unit time which is more appropriate when many state transitions occur within a relatively short time just as in the stochastic control problems in telecommunication application.

Chapter 3

ANALYSIS OF RESPONSE PROFILES

3.1 Introduction

In this chapter we present a method for analysing longitudinal data that imposes minimal structure or restrictions on the mean responses over time and on the covariance among the repeated measures. The method focusses on analysing response profiles and can be applied to longitudinal data when the design is balanced, with the timing of the repeat measures common to all individuals in the study. However, the method can also handle incomplete longitudinal studies with balanced designs. The method is appealing when there is a single categorical covariate (denoting exposure groups) and when no specific *a priori* pattern of the difference in response profiles between groups can be specified.

If $E(Y_{ij})$ is expressed exclusively in terms of the p explanatory variables X_{ij1}, $X_{ij2}, \ldots, X_{ijp}$, the model is said to be unconditional. If among these, time is simply interpreted as a discrete variable indicating the order in which the response is observed within each sample unit in a balanced design the corresponding models are known as profile models and are equivalent to those usually considered in ANOVA or ANCOVA problems. The profile model may be expressed as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{M} + \boldsymbol{E},$$

where $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_N)', \boldsymbol{X} = (X_1, X_2, \ldots, X_p), \boldsymbol{M}$ is a $p \times n$ matrix of unknown parameters and $\boldsymbol{E} = (e_1, e_2, \ldots, e_n)'$. Typically, the elements of $\boldsymbol{M}$ denote expected responses for units under the different treatments at the different instants.

At any given level of the group factor, the sequence of means over time is referred to as the mean response profile. For example, the mean response profiles for the two groups randomized to a drug and its placebo

The main goal in the analysis of response profiles is to characterise the patterns of change in the mean response over time in the groups and to determine whether the shapes of the mean response profiles differ for the two groups.

The methods for analysing response profiles can be extended in a straightforward way to handle the case where there is more than a single group factor and when there are baseline covariances that need to be adjusted for. For example, in an observational study the groups might be defined by characteristics of the study subjects, such as age, gender or exposure level.

Rowel and Walters (1976) and Bryant and Gillings (1985) discuss the analysis strategies using standard parametric or non-parametric ANOVA or regression procedures on some kind of univariate or bivariate summary measure for the individual response profiles. The article by Rowell and Walters (1976) is widely cited for popularising the analysis of summary measures of growth in many different disciplines.

Potthoff and Roy (1964) gives an extension of the profile models, which in the literature is often referred to as growth curves models. The paper gives a typical example that involves the linear relation between an anatomic distance and age in orthodontical study.

A detailed discussion of issues surrounding adjustment for baseline response in the analysis of change can be found in Lord (1967), Laird (1983) and Fitzmaurice (2001).

MANOVA models may also be employed to analyse problems under the growth curve model. Khatri (1966) show that the estimators proposed in Potthoff and Roy's (1964) paper are ML estimators.

Strengths and Weaknesses of Analysing Response Profiles

The analysis of response profiles is a straightforward way to analyse data from a longitudinal study when the design is balanced, with the timings of the repeated measures common to all individuals in the study, and when all the covariates are discrete (e.g., representing different treatments, interventions, or characteristics of the study subjects). It allows arbitrary patterns in the mean response over time and in the covariance of the responses. As a result, this method has certain robustness since the potential risk of bias due to misspecification of the models for mean and covariance are minimal. Although the method requires that data arise from a balanced design, it can be applied when the data are incomplete due to missing response data. The analysis of response profiles can be extended in a straightforward way to handle the case where individuals can be grouped according to more than a single factor. For example, if there are two covariates that are discrete (e.g., treatment group and gender), the analysis will include tests of the 3-way and 2-way interactions among

these two factors and time (in addition to their main effect.)

The analysis of response profiles does have a number of potential drawbacks. First, the requirement that the longitudinal design be balanced implies that the method cannot be applied when the vectors of repeated measures are obtained at different sequences of time except by moving an observation to the nearest planned measurement time. As a result, the method is not well suited to handle mistimed measurements, a common problem in many longitudinal studies. Note, however, that the method can handle unbalanced patterns of observations due to missing response data. Second, the analysis ignores the time ordering of the repeated measures. Third, because the analysis of response profiles produces an overall omnibus test of effects, it may have low power to detect group differences in specific trends in the mean response over time. (e.g., linear trends in the mean response). Single degree of freedom tests of specific time trends are more powerful. Finally, in the analysis of response profiles, the number of estimated parameters ($G \times n$ mean parameters and $\frac{n(n+1)}{2}$ covariance parameters) grows rapidly with the number of measurement occasions.

3.2 Hypotheses Concerning Response Profiles

Initially we focus on two group designs. Generalizations to more than two groups are straightforward. Given a sequence of n repeated measures on a number of distinct groups of individuals, three main questions concerning response profiles can be posed.

1. Are the two mean response profiles similar in the two groups, in the sense that the mean response profiles are parallel? That is whether the patterns of changes in the response over time are same across the groups. This question concerns the group $\times$ time interaction effects.

2. Assuming that the population mean responses are parallel, are the means constant over time, in the sense that the mean response profiles are flat. This question concerns the time effect.

3. Assuming that the population mean response profiles are parallel, are they also same at the same level in the sense that mean response profiles for the two groups coincide?

The first question is of main scientific interest. This question is the *raison d' être* of a longitudinal study. Whether the second and third questions have scientific interest depends upon the longitudinal study design. An affirmative answer to the first question is assumed while considering the second and third questions, consistent with the general principle that the main effects are not of interest when there is an interaction among them.

The choice of appropriate hypothesis deserves considerations regarding whether the longitudinal data arise from a randomised trial or from an observational study. In the former case, the study participants are randomised to treatment groups and the baseline values of the response are obtained prior to scientific interventions, *i.e.*, the mean response at occasion 1 is independent of treatment assignment. That is, by

design, the group means are equal at baseline. In contrast, in an observational study, there is no *a priori* reason to assume that the groups have the same mean at baseline unless the groups are selected by matching on baseline response.

In a randomised longitudinal clinical trial the only question of scientific interest is the first because it addresses whether changes in the mean response over time are the same in all groups. The second question is of less importance as it does not involve a direct comparison of groups. The second question concerns the time effect, where focus is on the comparison of the mean response at each occasion averaged over the groups. Thus hypothesis concerning the main effect of time translate into the question whether overall (*i.e.,* averaged over groups) mean response has changed from baseline. In a randomised trial, the third question is of no interest. Actually, the test of group effect is subsumed within the test of group $\times$ time interaction, since the absence of group $\times$ time interaction implies that groups have same pattern of change over time and their mean response profiles must necessarily coincide.

In an observational study, the first question is usually of primary interest. It addresses whether the patterns of change over time in the mean response vary by group. In contrast to randomised trial, however, the second and third questions may also be of substantive interest. For example, in a longitudinal study of growth or ageing, there may be interest in the pattern of change in the mean response over time, even when the pattern of change is same in all the groups. This concerns the main effect of time and is addressed by the second question. Ordinarily, when time effect is of interest, the trend in mean response is described with a parametric curve.

In an observational study, there may also be interest in group comparisons of the mean response averaged over time. This concerns the group effect and is addressed in the third question.

To highlight the main features of analysis of response profiles, consider an example from a two-group study comparing a novel treatment and a control. Assume that the two groups have repeated measurements at the same set of n occasions. The analysis of response profiles is based on comparing the mean response profiles in the two groups. Let $\boldsymbol{\mu}(T) = \{\mu_1(T), \mu_2(T), \ldots, \mu_n(T)\}'$ denote the mean response profile for the treatment group and $\boldsymbol{\mu}(C) = \{\mu_1(C), \mu_2(C). \ldots, \mu_n(C)\}'$ denote that for the control group. Let $\Delta_j = \mu_j(T) - \mu_j(C),\ j = 1, 2, \ldots, n$. The comparison can be made by considering the null hypothesis of no group $\times$ time interaction, that is, the hypothesis that the mean response profiles are parallel, for which the requirement is that the differences Δ_j's are constant over time. Thus, in terms of Δ_j's, the null hypothesis is

$$H_{01} : \Delta_1 = \Delta_2 = \cdots = \Delta_n.$$

The test has $n - 1$ degrees of freedom as the number of constraints on the mean responses under this hypothesis is $n - 1$.

When there are G groups with repeated measurements at the same set of n occasions, let $\boldsymbol{\mu}(g) = \{\mu_1(g), \mu_2(g), \ldots, \mu_n(g)\}'$ denote the mean response profile for the g^{th} group ($g = 1, 2, \ldots, G$). With G groups, there are $G - 1$ redundant comparisons. Define $\Delta_j(g) = \mu_j(g) - \mu_j(G)$, (for $j = 1, 2, \ldots$,n; $g = 1, 2, \ldots, G - 1$). *i.e.,* $\Delta_j(g)$ is

the contrast or comparison of the mean response at the j^{th} occasion for the g^{th} group $(g = 1, 2, \ldots, G-1)$ with the mean response at the j^{th} occasion for the group G. Then, the null hypothesis that the mean response profiles are parallel is

$$H_{01} : \Delta_1(g) = \Delta_2(g) = \cdots = \Delta_n(g), \text{ for } g = 1, 2, \ldots, G-1.$$

The test has $(G-1) \times (n-1)$ degrees of freedom. However, unless the test of group $\times$ time interaction has only one degree of freedom, this test does not help in discerning in what manner the pattern of change over time differ across groups. In the next section we describe a general linear model formulation of the analysis of response profiles.

3.3 General Linear Model Formulation

Consider the general linear regression model

$$E(\boldsymbol{Y}_i / \boldsymbol{X}_i) = \boldsymbol{\mu}_i = \boldsymbol{X}_i \boldsymbol{\beta}, \tag{3.3.1}$$

for appropriate choices of $\boldsymbol{X}_i$. The hypothesis of no group $\times$ time interaction can be expressed in terms of $\boldsymbol{\beta}$. Let 'n' be the number of repeated measures and N be the number of subjects. To express the model for a design with G groups and n occasions of measurement, we require $G \times n$ parameters for the G mean response profiles. We illustrate the main idea with the help of a numerical example with two

groups measured at three occasions. For the first group, let the design matrix be

$$\boldsymbol{X}_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix},$$

while for the second group the design matrix is

$$\boldsymbol{X}_i = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then in terms of the model (3.3.1) where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_6)'$ is a 6×1 vector of regression coefficients,

$$\boldsymbol{\mu}(1) = \begin{pmatrix} \mu_1(1) \\ \mu_2(1) \\ \mu_3(1) \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}.$$

Similarly

$$\boldsymbol{\mu}(2) = \begin{pmatrix} \mu_1(2) \\ \mu_2(2) \\ \mu_3(2) \end{pmatrix} = \begin{pmatrix} \beta_4 \\ \beta_5 \\ \beta_6 \end{pmatrix}.$$

The hypothesis about the mean response profiles in two groups in terms of $\boldsymbol{\mu}(1)$ and $\boldsymbol{\mu}(2)$ can now be expressed in terms of hypotheses about the components of $\boldsymbol{\beta}$. Specifically, the hypothesis of no group $\times$ time interaction effect can be expressed as

$$H_{01} : (\beta_1 - \beta_4) = (\beta_2 - \beta_5) = (\beta_3 - \beta_6) \tag{3.3.2}$$

In this parameterisation, hypothesis about the group $\times$ time interaction cannot be expressed in terms of certain components of $\boldsymbol{\beta}$ being zero; instead, these hypotheses can be expressed in terms of $\boldsymbol{L\beta} = 0$, for particular choices of vectors or matrices $\boldsymbol{L}$. For example (3.3.2) may now be expressed as

$$H_{01} : \boldsymbol{L\beta} = 0,$$

where

$$L = \begin{pmatrix} 1 & -1 & 0 & -1 & 1 & 0 \\ 1 & 0 & -1 & -1 & 0 & 1 \end{pmatrix}$$

An attractive feature of the general linear model (3.3.1) is that it can handle settings where the data for some subjects are missing. For example, suppose that the i^{th} subject belongs to the first group and is missing the response at the third occasion. Then the appropriate design matrix for the subject is obtained by removing the third row of the full data design matrix for the subjects from the first group as

$$\boldsymbol{X}_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

In general, the appropriate design matrix for the i^{th} subject is simply obtained by removing rows of the full data design matrix corresponding to the missing responses. This allows analysis of response profiles to be based on all available observations of the subjects.

The general linear model (3.3.1) for two groups measured at three occasions could have also been expressed in terms of the following matrices.

$$\boldsymbol{X}_i = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{pmatrix},$$

for the first group, and

$$\boldsymbol{X}_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix},$$

for the second group. In that case

$$\boldsymbol{\mu}(2) = \begin{pmatrix} \mu_1(2) \\ \mu_2(2) \\ \mu_3(2) \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_1 + \beta_2 \\ \beta_1 + \beta_3 \end{pmatrix};$$

$$\text{and,} \quad \boldsymbol{\mu}(1) = \begin{pmatrix} \mu_1(1) \\ \mu_2(1) \\ \mu_3(1) \end{pmatrix} = \begin{pmatrix} \beta_1 + \beta_4 \\ (\beta_1 + \beta_4) + (\beta_2 + \beta_5) \\ (\beta_1 + \beta_4) + (\beta_3 + \beta_6) \end{pmatrix}.$$

The choice of the reference group (here the second group) is arbitrary. With this choice of design matrices for the two groups, the interpretation of the regression coefficients has changed. The hypothesis of no group $\times$ time interaction, now, is

$$H_{01} : \beta_5 = \beta_6 = 0.$$

This parameterisation, often called the reference parameterisation, is more convenient since the hypothesis H_{01} is represented by the vanishing of certain components of $\boldsymbol{\beta}$, and is the one that is commonly adopted by many statistical packages

If H_{01} cannot be rejected, the hypotheses concerning mean effects of time and/or group may be of secondary interest. Hypotheses concerning main effects can similarly be represented by the vanishing of certain components of $\boldsymbol{\beta}$. For example hypothesis of no time effect is

$$H_{02} : \beta_2 = \beta_3 = 0,$$

and the hypothesis of no group effect is

$$H_{03} : \beta_4 = 0.$$

For the general case with G groups measured at n occasions, the number of constraints under H_{02} is n-1 and is the same regardless of the number of groups and the test of H_{02} has n-1 degrees of freedom. Similarly the number of constraints under H_{03} is G-1 and is the same regardless of the number of occasions and the test of H_{03} has G-1 degrees of freedom. The tests of main effects are obtained from the reduced models that excludes the group $\times$ time interaction.

Finally, Given that the analysis of response profiles can be expressed in terms of the linear model (3.3.1), where $\boldsymbol{\beta} = (\beta_1, \beta_2 \ldots, \beta_p)'$ is a p $\times$ 1 vector of regression coefficients (with $p = G \times n$), maximum likelihood estimation of β, and the construction of tests of the group $\times$ time interaction (and the main effects of time and group), are possible once the covariance of $\boldsymbol{Y}_i$ has been specified. In the analysis of response profiles, the covariance of $\boldsymbol{Y}_i$ is usually assumed to be unstructured with no constraints on $\frac{n(n+1)}{2}$ covariance parameters other than the requirement that they yield a symmetric matrix that is positive definite. The condition of positive definiteness of the covariance matrix ensures that while repeated measures can be highly correlated, there must be no redundancy in the sense that one of the repeated measures can be expressed as a linear combination of the others. The condition also ensures that no linear combination of responses can have negative variance. Given REML (or ML) estimates of $\boldsymbol{\beta}$, and their standard errors (and the estimated covariance of $\widehat{\boldsymbol{\beta}}$, tests of group $\times$ time interaction (and main effects of time and group), can be conducted using multivariate Wald tests. Alternatively likelihood ratio tests can be constructed, but requires that the model be fit to the data with and without the constraints under the null hypothesis (*i.e.*, fitting the 'reduced' and the 'full' models respectively). Before illustrating the analysis of response profiles, we present a brief review of the 'reference group' parametrisation.

3.4 Reference Group Parameterisation

Consider a group factor with G levels. To present this factor in a linear model we can define a set of dummy or indicator variables,

$$Z_{ig} = \begin{cases} 1, & \text{if the } i^{th} \text{ subject belongs to group } g; \\ 0, & \text{otherwise.} \end{cases}$$

Letting $\boldsymbol{X}_i = (Z_{i1}, Z_{i2}, \ldots, Z_{iG})$, the mean response in G groups, denoted by $\mu_i(1), \mu_i(2), \ldots, \mu_i(G)$, can be expressed in terms of the linear model

$$E(\boldsymbol{Y}_i/\boldsymbol{X}_i) = \boldsymbol{\mu}_i = \boldsymbol{X}_i\boldsymbol{\beta}.$$

In this parameterisation

$$\begin{pmatrix} \mu_i(1) \\ \mu_i(2) \\ \vdots \\ \mu_i(G) \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_G \end{pmatrix}$$

If we wish to introduce an intercept, say β_1, by setting the elements of the first column of $\boldsymbol{X}_i$ to 1 (for all $i = 1, 2, \ldots, N$) then there is redundancy in $\boldsymbol{X}_i$, if all G indicator variables $Z_{i1}, Z_{i2}, \ldots, Z_{iG}$ are also included in the design vector $\boldsymbol{X}_i$. To avoid this over specification, one of the indicator variables must be excluded from $\boldsymbol{X}_i$. Arbitrarily we can drop Z_{iG}. Then with $\boldsymbol{X}_i = (1, Z_{i1}, Z_{i2}, \ldots, Z_{i,G-1})$, the mean response in G groups can be expressed in terms of the linear model

$$E(\boldsymbol{Y}_i/\boldsymbol{X}_i) = \boldsymbol{\mu}_i = \boldsymbol{X}_i\boldsymbol{\beta},$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_G)'$. In this parameterisation,

$$\begin{pmatrix} \mu_i(1) \\ \mu_i(2) \\ \vdots \\ \mu_i(G-1) \\ \mu_i(G) \end{pmatrix} = \begin{pmatrix} \beta_1 + \beta_2 \\ \beta_1 + \beta_3 \\ \vdots \\ \beta_1 + \beta_G \\ \beta_1 \end{pmatrix}$$

Because the intercept term, β_1, is also the mean of group G, and all the remaining components of β represent deviations from the mean of group G, this parameterisation is often referred to as the reference group parameterisation. Here, the last level of the group factor (*i.e.*, group G) is the reference group and it is no coincidence that this is the same group whose indicator variable was excluded from $\boldsymbol{X}_i$.

3.5 One Degree of freedom Tests for Group $\times$ Time Interaction

The tests of group $\times$ time interaction is quite general and it posits no specific pattern for the difference in response profiles between groups. This lack of specificity becomes a problem in studies with a large number of occasions of measurement because the general test for group $\times$ time interaction with $(G-1) \times (n-1)$ degrees of freedom, becomes less sensitive to an interaction with a specific pattern as n increases.

In the typical randomised trial of interventions, subjects are randomised to the intervention groups at baseline and the investigator seeks to determine whether the pattern of response after intervention differs between groups. Randomisation implies that the mean at baseline is independent of treatment group, that is, by design, the groups have the same mean response at baseline. In that setting, analysts frequently specify a single contrast believed to best represent the direction in which the pattern of response will differ most markedly. For example, if we assume the first parameter-

isation described in Section 3.3 with two groups and wish to test for the equality of the difference between the average response at occasion 2 through n and the baseline value in the two groups, we can choose the contrast

$$\boldsymbol{L} = (-\boldsymbol{L_1}, \boldsymbol{L_1}),$$

where

$$\boldsymbol{L}_1 = \left(-1, \frac{1}{n-1}, \frac{1}{n-1}, \ldots, \frac{1}{n-1}\right).$$

Here, $\boldsymbol{L}_1$ computes the mean response from occasion 2 through n and subtracts the mean response at baseline for a single group. The latter can be thought of as the average change over the interval for a single group. Thus $\boldsymbol{L}$ is the group contrast of this average change in the two groups.

A variant of this approach, known as Area Under the Curve Minus Baseline, or sometimes simply AUC, corresponds to calculation of the area under the trapezoidal curve created by connecting the responses plotted at the respective time points and subtracting $y_1(t_n - t_1)$, the area of the rectangle of height y_1 and width $t_n - t_1$.

The AUC (minus baseline) is negative because the responses after intervention begins are smaller than baseline value. The AUC (minus baseline) can be constructed by subtracting the baseline mean μ_1, from each of the means μ_1 through μ_n and calculating area under the trapezoid constructed by connecting these differences. To test for the equality of AUC in two groups, one employs the contrast

$$\boldsymbol{L} = (-\boldsymbol{L_2}, \boldsymbol{L_2}),$$

where

$$\boldsymbol{L}_2 = \frac{1}{2}\left(t_1 + t_2 - 2t_n,\ t_3 - t_1,\ \ldots\ ,\ t_{j+1} - t_{j-1},\ \ldots\ ,\ t_n - t_{n-1}\right)$$

and $\frac{1}{2}(t_{j+1} - t_{j-1})$ is the value of the contrast vector for time points 1 (baseline) or n (the last occasion). These contrast weights are not intuitively obvious, but can be derived from the formula for the area of a trapezoid. Although the curve in the above figure suggests that $\boldsymbol{L}$ is applied to the individual observations, it must be emphasized that the contrast weights are applied to the estimated means, not the individual observations.

A third popular method for constructing a single-degree-of freedom test corresponds to a test of the hypothesis that the trend over time is the same in several treatment groups. This method is a special case of the growth curve analysis.

In many applications, the one-degree-of-freedom test will be statistically significant when the overall test for group $\times$ time interaction is not. For valid application of conventional significant levels, the form of the contrast must be specified prior to data analysis. Otherwise, one would be at risk of seeking the best contrast and testing its significance as if it had been chosen in advance. To guard against this criticism, the protocols for randomised trials usually specify the form of the contrast. This requirement highlights a hazard of the one-degree-of-freedom tests. The added sensitivity comes at the price of reduced generality. If the difference between the treatment groups takes a form quite different from the pattern anticipated by the contrast, one can fail to obtain statistically significant result for a one-degree-of-freedom test even

when the overall test for group × time interaction is statistically significant. Thus one-degree-of-freedom test should be employed only when there is sufficient prior information to specify the contrast with confidence.

3.6 A Numerical Illustration

We shall illustrate the analysis of mean response using recent stock market data. Stock prices of Indian IT major *Infosys Technologies* for the period 26^{th} December 2003 to 31^{st} December 2008 obtained from official web site of National Stock Exchange (NSE) of India. Bombay Stock Exchange (BSE)indices for the corresponding period obtained from the BSE itself.

We arrived at the stock prices of *Infosys Technologies* by considering bonus share issue and stock spit for the last five years. On $1^{st}july2004$ the company issued bonus shares in the ratio 3:1 (3 shares for every 1 share in demat account before the book closure). Evidently every investor will have **four** shares instead of **one** in his demat account. Therefore Net Asset Value (NAV) of one share held previously to $1^{st}july2004$ will be $4 \times close\ price$. Similarly the company made a stock split in the ratio 2:1 on $13^{th}july2006$ (one share with face value Rs.10/- becomes 2 shares with face value Rs.5/- each). Hence NAV of one share held previously to $1^{st}july2004$ will be $8 \times close\ price$. The relevant figures shown in **bold type** and with * in table-3.1. The NAV taken as effective stock price for the share.

Table- 3.1 Stock prices of *Infosys Technologies.* with traded volume and NAV

Date	Prev Close	Open Price	Close Price	Total Traded Quantity	Net Asset Value
⋮	⋮	⋮	⋮	⋮	⋮
28-Jun-04	5483.90	5530.00	5597.90	490516	5597.90
29-Jun-04	5597.90	5605.00	5586.70	687773	5586.70
30-Jun-04	5586.70	5570.00	5524.10	611316	5524.10
1-Jul-04	**5524.10**	**1395.00**	**1408.80***	1767215	@ **5635.20***
2-Jul-04	1408.80	1365.25	1423.35	1558128	5693.40
5-Jul-04	1423.35	1426.30	1429.45	969096	5717.80
6-Jul-04	1429.45	1427.00	1426.50	1201123	5706.00
7-Jul-04	1426.50	1404.15	1392.45	1776746	5569.80
8-Jul-04	1392.45	1400.00	1352.25	2366130	@ 5409.00
9-Jul-04	1352.25	1300.00	1382.50	1424235	5530.00
⋮	⋮	⋮	⋮	⋮	⋮
10-Jul-06	3105.00	3189.00	3188.05	764033	12752.20
11-Jul-06	3188.05	3199.00	3148.20	810381	12592.80
12-Jul-06	3148.20	3300.00	3385.65	4251043	13542.60
13-Jul-06	**3385.65**	**1748.00***	**1681.95***	1423289	@ **13455.60***
14-Jul-06	1681.95	1650.00	1648.75	1737712	13190.00
17-Jul-06	1648.75	1641.00	1605.25	1552356	12842.00
18-Jul-06	1605.25	1610.35	1629.55	1323057	13036.40
19-Jul-06	1629.55	1661.00	1604.55	1448682	12836.40
20-Jul-06	1604.55	1698.70	1647.85	1503490	@ 13182.80
21-Jul-06	1647.85	1634.80	1602.10	1140542	12816.80
⋮	⋮	⋮	⋮	⋮	⋮

@ figures considered for data set

Nowadays, Cash Market of stock exchanges are highly influenced by Derivative Market (Futures & Options Market). Life of monthly contracts of derivatives end on last Thursday (critical day) of the month. Stock prices will have a new lease of life after the critical day. Thus we have considered stock prices on first Thursday (base line), on second Thursday (Week 1), on third Thursday (Week 2)and on the critical day (Week 4), for data set. To make independent entities for the months, We have

taken the deviation of stock price from stock price on critical day of previous month.

Table- 3.2A Increase in Stock Prices and BSE Indices

	Stock Price				BSE Index			
Month	Base line	Week 1	Week 2	Week 4	Base line	Week 1	Week 2	Week 4
2003								
Jan	-26.00	49.95	-138.55	-489.15	-17.58	1.56	-9.12	-162.76
Feb	258.60	-85.15	-7.80	-153.95	85.56	27.63	83.34	57.46
Mar	-7.55	-40.05	255.55	184.40	-86.99	-169.10	-84.41	-160.55
Apr	5.10	-1240.00	-1326.35	-1391.90	34.37	-81.46	-132.29	-179.56
May	-111.25	77.95	-14.25	-128.25	22.56	24.37	75.74	227.02
Jun	90.20	200.30	273.60	469.40	97.95	173.05	290.35	388.15
Jul	-120.40	387.80	224.40	363.20	87.49	127.23	116.51	240.21
Aug	-93.85	-185.55	112.25	296.60	14.22	128.59	302.78	419.68
Sep	310.85	325.75	342.35	608.70	98.22	180.84	-78.14	84.86
Oct	66.40	-60.40	123.15	109.10	157.93	401.53	590.17	483.37
Nov	458.80	130.55	5.25	311.75	267.02	168.64	-9.29	208.52
Dec	133.40	49.45	200.50	495.50	236.86	310.92	465.96	652.88
2004								
Jan	212.50	424.25	137.60	-170.75	273.55	466.62	421.99	160.83
Feb	186.35	100.05	-146.80	-368.00	-82.12	134.21	52.35	-235.63
Mar	399.90	-29.50	52.10	281.90	248.75	82.74	-152.18	-152.68
Apr	-56.00	145.10	169.40	-128.25	326.41	424.01	429.53	253.99
May	258.50	155.00	-110.80	259.00	88.87	-268.96	-736.32	-609.88
Jun	-56.30	-24.60	-98.20	111.40	-240.56	-113.91	-218.67	-350
Jul	226.05	-0.15	364.65	733.25	165.50	135.29	179.64	411.9
Aug	88.80	-147.80	4.60	39.80	132.33	19.32	3.2	15
Sep	228.40	308.40	597.60	599.80	63.27	162.78	342.23	448.16
Oct	40.00	476.00	337.00	1013.20	190.05	129.49	57.45	132.01
Nov	155.60	273.20	456.00	323.20	117.26	238.69	309.85	319.41
Dec	201.20	50.40	363.00	74.20	293.40	269.24	385.35	487.51
2005								
Jan	-17.60	-319.20	-333.60	-384.20	-155.15	-301.48	-339.3	-283.11
Feb	597.00	369.20	887.40	848.80	380.54	338.40	349.86	334.78
Mar	200.60	271.00	67.20	371.60	210.51	333.44	95.31	-81.39
Apr	-275.60	-622.80	-1249.60	-1406.40	52.82	-24.90	-193.62	-208.62
May	358.20	622.20	764.60	1198.40	75.45	172.62	194.74	386.58
Jun	-15.40	126.20	301.00	612.20	-15.22	161.75	229.63	523.07

Table- 3.2B Increase in Stock Prices and BSE Indices

	Stock Price				BSE Index			
Month	Base line	Week 1	Week 2	Week 4	Base line	Week 1	Week 2	Week 4
Jul	-153.40	-889.60	-597.00	-510.60	-48.72	-6.15	110.47	411.18
Aug	331.60	251.20	272.20	164.80	192.05	211.48	206.3	55.39
Sep	615.40	602.40	715.00	1017.60	215.73	392.14	623.34	989.75
Oct	331.60	385.80	49.60	241.20	-121.47	-273.27	-715.05	-577.42
Nov	0.00	269.80	482.20	457.80	0.00	236.18	576.77	671.29
Dec	267.80	481.40	866.80	1146.00	200.74	162.27	426.36	579.21
2006								
Jan	269.80	-598.60	-666.20	-662.00	294.49	57.63	126.59	226.67
Feb	24.60	322.80	-163.40	-109.60	293.95	494.90	574.38	694.13
Mar	239.00	193.80	509.00	993.80	382.73	329.49	634.69	1062.99
Apr	287.00	-96.60	934.40	358.80	439.86	-69.81	732.51	527.98
May	314.40	493.40	-395.00	-1216.60	512.61	600.39	-443.59	-1168.7
Jun	5.20	-499.00	-406.00	660.40	-594.90	-1370.51	-1121.26	-504.16
Jul	680.80	1481.00	1208.20	1292.20	605.81	696.34	190.78	579.43
Aug	116.80	402.00	794.00	1184.40	181.57	407.58	735.89	957.46
Sep	4.00	199.20	253.60	414.00	154.80	273.97	575.22	681.69
Oct	56.40	1295.20	1682.80	1721.60	8.67	157.24	342.85	317.67
Nov	91.20	510.40	1012.80	850.80	392.71	439.08	807.48	997.9
Dec	352.80	131.20	-128.00	548.00	275.72	-209.15	-311.45	150.03
2007								
Jan	296.80	-521.60	-192.40	-57.60	25.37	-215.63	371.41	436.38
Feb	223.60	1061.60	1135.60	372.00	-15.54	369.37	72.83	-261.41
Mar	-1026.80	-1214.00	-1664.80	-2374.40	-861.76	-971.96	-1477.46	-1041.65
Apr	12.80	441.20	393.60	225.20	-123.58	134.15	640.04	1249.22
May	502.40	-357.20	-282.80	-762.80	-150.67	-457.65	70.83	315.58
Jun	269.20	765.60	279.60	12.40	-358.28	-340.74	-45.22	-39.89
Jul	-66.00	-22.80	571.60	874.40	357.32	587.47	1045.56	1271.74
Aug	-1081.20	-800.80	-970.00	-1353.20	-790.61	-676.16	-1418.1	-654.57
Sep	308.80	-247.60	-518.40	405.20	494.57	492.70	1226.21	2028.82
Oct	664.40	488.00	-206.80	-630.40	626.58	1663.51	847.83	1620.33
Nov	295.20	-965.60	-1518.00	-2158.40	953.46	288.04	1014	232.37
Dec	544.80	741.60	1026.40	1881.20	792.61	1101.13	159.31	1213.46

Table- 3.3: Variance-Covariance matrix $\widehat{\mathrm{Cov}}(\widehat{\beta})$

	Stock Price				BSE Index			
Group	Base line	Week 1	Week 2	Week 4	Base line	Week 1	Week 2	Week 4
Stock Price								
Base line	1572.22	1375.29	1505.92	1924.48	1081.81	1124.91	1241.92	1130.25
Week 1	1375.29	4598.29	4818.44	5237.34	880.84	1933.76	1355.10	1401.20
Week 2	1505.92	4818.44	6983.50	7915.94	960.76	1882.83	2360.46	2603.12
Week 4	1924.48	5237.34	7915.94	11416.54	1145.26	1826.03	2558.10	3917.19
BSE Index								
Base line	1081.81	880.84	960.76	1145.26	1625.29	1790.04	1901.68	1655.75
Week 1	1124.91	1933.76	1882.83	1826.03	1790.04	3080.12	2644.77	2664.90
Week 2	1241.92	1355.10	2360.46	2558.10	1901.68	2644.77	4553.41	4326.16
Week 4	1130.25	1401.20	2603.12	3917.19	1655.75	2664.90	4326.16	5807.13

Table-3.4: Mean increase and(Std. deviation) at Base line, Week 1, Week 2, Week 4

Group	Base line	Week 1	Week 2	Week 4
stock Price	141.26	101.53	118.20	161.18
	(307.14)	(525.26)	(647.31)	(827.64)
BSE Index	124.32	135.95	160.04	280.57
	(312.28)	(429.89)	(522.69)	(590.28)

A popular method for constructing a single-degree-of freedom test corresponds to a test of the hypothesis that the trend over time is the same in several groups. This method is a special case of the growth curve analysis.

The test of group $\times$ time interaction is based on multivariate Wald Test. The test provides a simultaneous test of $H_0 : \boldsymbol{L\beta} = 0$ *vs* $H_A : \boldsymbol{L\beta} \neq 0$, for a suitable

choice of $\boldsymbol{L}$ and the test statistic can be constructed as

$$W^2 = \left(\boldsymbol{L}\widehat{\boldsymbol{\beta}}\right)\left\{\boldsymbol{L}\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\beta}})\boldsymbol{L}'\right\}^{-1}\left(\boldsymbol{L}\widehat{\boldsymbol{\beta}}\right), \tag{3.6.1}$$

and compared to a χ^2-distribution with degrees of freedom equal to the number of rows of $\boldsymbol{L}$.

The increase in stock prices are computed for four time points in a month, the vector representing the contrast based on the mean response at time 2 through n minus baseline is given by

$$\boldsymbol{L} = (-\boldsymbol{L_1}, \boldsymbol{L_1}), = \left(-1, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, -1, \frac{1}{3}, \frac{1}{3}\frac{1}{3}\right).$$

The Mean increase and standard deviation at base line, Week 1, Week 2, Week 4 for both the groups are shown in . Its Variance-Covariance matrix shown in table (3.3). Thus from the table (3.3), we can easily determine that the average value of the mean response minus baseline as 14.4167 in the stock price group and 67.6753 in the BSE index group. Thus if we assume the first parameterisation assumed in Section 3.4, then the $\boldsymbol{L}\widehat{\boldsymbol{\beta}} = 82.0920$ and the value of the Wald test statistics is $Z = 1.2441$ (or $W^2 = 1.5478$, with one degree of freedom), indicating no significant difference in the response pattern between the two groups.

The contrast for comparing the AUC (minus baseline) in the two groups is given by

$$\boldsymbol{L} = (-\boldsymbol{L_2}, \boldsymbol{L_2}), = (\ 3.5,\ -1,\ -1.5,\ -1,\ -3.5,\ 1,\ 1.5,\ 1).$$

From Table‘(3.4), the estimated mean AUC is -54.3996 in the stock price group and 221.4614 in the BSE index group. Thus if we assume the first parameterisation assumed in Section 3.4, then the $\boldsymbol{L}\widehat{\boldsymbol{\beta}} = 277.8610$, yielding a Wald statistics of $Z = 1.895$ (or $W^2 = 1.4149$, with one degree of freedom), again not a significant statistic. Thus both the methods of analysis provide a clear signal that the response profile does not differ in the two groups

Chapter 4

GENERALIZED EXPONENTIAL MODEL [1]

[1]This chapter is based on T. D. Xavier and M. Manoharan(2007)

4.1 Introduction

Censoring is a common feature in clinical trials. A Generalized Exponential Model under a more flexible and practical censoring scheme namely Type II progressive interval censoring with random removals (PICR) is appropriate in many situations like follow-up studies in organ transplant, chemotherapy and/or surgical treatment for various cancers etc. The patients are examined only at fixed regular intervals or when reporting for checkup at the hospital so that one can observe the patient only at that specified points of time. The survival time of some of these patients cannot be observed exactly since some of them withdraw from study due to the reasons only known to them.

In type II interval censoring 'n' subjects are selected in a life test, inspections are conducted at pre-determined intervals and the number of 'failures' occurring between two successive inspection times are recorded. The study will be terminated when the total number of failures is greater than or equal to a pre-assigned number 'm'. A handicap of this censoring scheme is that we could not utilize the sensitive information of drop-outs in the study. There are many real life situations, in particular clinical study, where one may come across more complicated censoring scheme combining the features of Type II censoring , interval censoring and progressive censoring with random removals. When the total time of study and the number of failed subjects are random outcomes, such a scheme is highly warranted. Xing and Tse (2005) in a clinical trial investigated a Weibull model under the censoring scheme called Type II

PICR to cope with the setting that patients are examined at fixed regular intervals and dropouts occur during the study period. In Type II PICR, the individual are examined at fixed regular intervals, at each examination the number of both dropouts and failed individuals are recorded, the study will be terminated when a pre-specified number of failed individuals are observed

Although Weibull distribution is a popular life time distribution on account of its several advantages, the maximum likelihood estimates of the Weibull parameters may not behave properly for all parametric values even when location parameter is zero(see Bain (1978)). Also the monotonicity of Weibull hazard function reaching an infinite value when the shape parameter is greater than one, may not be appropriate in many situations. The Weibull family does not enjoy likelihood ratio ordering property like gamma family, making the problem of one sided hypothesis testing extremely difficult. Further the distribution of the mean of random sample from the Weibull distribution is not simple to compute though its distribution function has a simple form.

Gupta and Kundu (1999) introduced generalized exponential(GE) family that has some interesting features very similar to those of Weibull family and gamma family but a nice alternative to them in many situations. The generalized exponential distribution is a two-parameter distribution having distribution function

$$F(x) = (1 - e^{-\beta x})^{\alpha}, \;\; x > 0, \; (\alpha, \beta > 0)$$

where α is the shape parameter and β is the scale parameter. Its density function is

given by

$$f(x;\alpha,\beta) = \alpha\beta(1-e^{-\beta x})^{\alpha-1}e^{-\beta x} \;\; ; x > 0$$

It is interesting to note the similarities of the density and distribution function of GE family with corresponding gamma family and Weibull family. If the shape parameter α =1, then all the three distributions coincide with the one-parameter exponential distribution. Therefore all the three distributions are extensions or generalizations of the exponential distribution, different ways. If X has an exponential distribution with moment generating function(mgf) $M_E(t)$ and distribution function $F_E(x)$, and similarly, subscript symbols G, W and GE respectively represent gamma, weibull and generalized exponential distributions, then it is well known that $M_G(t) = (M_E(t))^\alpha$ and $F_W(x) = F_E(x^\alpha)$. The GE distribution is such that $F_{GE}(x) = (F_E(x))^\alpha$. It can be said that GE distribution is the distribution of maximum of α (an integer) number of i.i.d exponential variables. From the form of density function of the GE distribution, we see that, if $\alpha \leq 1$, the density function is strictly decreasing function, where as if $\alpha > 1$, it is a unimodel skewed density function.

If $X \sim GE(\alpha,\beta)$, the survival function and hazard function are given by

$$S(t;\alpha,\beta) \;\; = 1 - F(t) \;\; = 1-(1-e^{-\beta t})^\alpha \;\; ; t > 0 \tag{4.1.1}$$

$$h(t;\alpha,\beta) \;\; = \tfrac{f(t;\alpha,\beta)}{S(t;\alpha,\beta)} \;\; = \frac{\alpha\beta(1-e^{-\beta t})^{\alpha-1}e^{-\beta t}}{1-(1-e^{-\beta t})^\alpha} \;\; ; t > 0 \tag{4.1.2}$$

If $\alpha = 1$, the hazard function becomes β, independent of x

For the Weibull distribution, if $\alpha > 1$, the hazard function increases from zero to

∞ and if $\alpha < 1$, the hazard function decreases from ∞ to zero. Many authors point out that because for the gamma distribution (for $\alpha > 1$), the hazard function increases from zero to a finite number, the gamma may be more appropriate as a population model when the items in the population are in a regular maintenance program. The hazard rate may increase initially, but after some time the system reaches a stable condition because of maintenance. The same comments hold for the GE distribution. Therefore, if it is known that the data are from regular environment, it may make more sense to gamma distribution or the GE distribution than the Weibull distribution.

Type II Progressive Interval Censoring with Random Removals

We follow the discussion on Type II PICR scheme of chapter I. Suppose that n subjects are randomly selected for the study and when specified number or percentage of total m (say) or more subjects are failed, the study will be terminated. Let $t_i, t_2, ...$ be the predetermined inspection times and t_0=0. Under a Type II PICR censoring scheme, the study is terminated after the k^{th} inspection time if the total number of failed subjects is equal to or more than m. At the i^{th} inspection, d_i failed subjects are observed and R_i subjects are randomly removed from the test. In other words, d_i is the number of failed subjects between any two successive inspections at t_{i-1} and t_i. Thus, R_i and d_i are random variables obtained from the study. Let us denote $Y_j = \sum_{i=1}^{j} d_i$ the total number of failed subjects observed upto the j^{th} inspection

time t_j. If $Y_{k-1} < m$ and $Y_k \geq m$, for the predetermined integer m , $0 \leq m \leq n$; the test is terminated at the k^{th} inspection time t_k . Denote $D = (d_1, d_2, , d_k)$ and $R = (R_1, R_2, ..., R_{k-1})$ where k is random and corresponds to the number of inspections before the termination of the experiment t_k.

Now we discuss exponential, Weibull and generalized exponential model for the survival times under the type II PICR censoring scheme.

Exponential Model:

Assume that the survival time T follows an exponential distribution with parameters β . The probability density function of T is given by

$$f(t) = \beta e^{-\beta t} \ , t > 0 \ (\beta > 0). \tag{4.1.3}$$

The cumulative distribution function F(t) is given by

$$F(t) = 1 - e^{-\beta t}, \ \ t \geq 0$$

In order to derive the joint likelihood function based on the observations under this set up, we first consider the following conditional joint probability density function. The conditional joint probability density function of number of observations d_i and

k, conditional on R_i

$$f(d_1,...,d_k,k|R) = \binom{n}{d_1}\binom{n-d_1-R_1}{d_2}\dots\binom{n-\sum_{j=1}^{k-1}d_j-\sum_{j=1}^{k-1}R_j}{d_k} \times \prod_{i=1}^{k}(p_{i-1}-p_i)^{d_i}(1-p_i)^{R_i} \tag{4.1.4}$$

$$\text{where } R_k = n-\sum_{j=1}^{k}d_j-\sum_{j=1}^{k-1}R_j$$

$$p_0 = 0, \quad p_i = 1-e^{-\beta t_i} \ \ for \ \ i = 1,2,,k-1, \ \ p_{k+1} = 1.$$

Weibull Model:

Assume that the survival time T follows a Weibull distribution with parameters α and β , where α is the scale parameter and β is the shape parameter. The probability density function of T is given by

$$f(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right], \ \ ;\alpha,\beta,t>0 \tag{4.1.5}$$

The cumulative distribution function F(t) is given by

$$F(t) = 1-exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right], \ \ for \ \ t \geq 0$$

Now the joint probability density function of number of observations d_i and k, conditional on R_i

$$f(d_1,...,d_k,k|R) = \binom{n}{d_1}\binom{n-d_1-R_1}{d_2}\dots\binom{n-\sum_{j=1}^{k-1}d_j-\sum_{j=1}^{k-1}R_j}{d_k}$$
$$\times\prod_{i=1}^{k}(p_{i-1}-p_i)^{d_i}(1-p_i)^{R_i} \tag{4.1.6}$$
$$\text{where}\quad R_k = n-\sum_{j=1}^{k}d_j-\sum_{j=1}^{k-1}R_j$$
$$p_0=0,\quad p_i=1-exp\left[-\left(\frac{t_i}{\alpha}\right)^{\beta}\right]\quad for\ \ i=1,2,,k,\ \ p_{k+1}=1.$$

Xiang and Tse(2005) discussed the maximum likelihood estimators of the parameters of this model under progressive interval censoring with random removals.

4.2 The Generalized Exponential Model

We assume that the survival time T follows a generalized exponential distribution with parameters α and β , where α is the shape parameter and β is the scale parameter. The probability density function of T is given by

$$f(t) = \alpha\beta(1-e^{-\beta t})^{\alpha-1}e^{-\beta t}\ \ ;\alpha,\beta,t>0 \tag{4.2.1}$$

The cumulative distribution function F(t) is given by

$$F(t) = (1-e^{-\beta t})^{\alpha},\ \ for\ \ t\geq 0$$

Now the joint probability density function of number of observations d_i and k, conditional on R_i, can be derived inductively following Xiang and Tse (2005).

$$f(d_1,...,d_k,k|R) = \binom{n}{d_1}\binom{n-d_1-R_1}{d_2}\dots\binom{n-\sum_{j=1}^{k-1}d_j-\sum_{j=1}^{k-1}R_j}{d_k} \times\prod_{i=1}^{k}(p_{i-1}-p_i)^{d_i}(1-p_i)^{R_i} \tag{4.2.2}$$

$$\text{where}\quad R_k = n-\sum_{j=1}^{k}d_j-\sum_{j=1}^{k-1}R_j$$

$$p_0=0,\quad p_i=(1-e^{-\beta t_i})^{\alpha}\quad for\quad i=1,2,,k-1,\quad p_{k+1}=1.$$

If R_i is assumed to follow a binomial distribution with parameter λ , the probability of r_i subjects removed from the study at the i^{th} inspection time is given by

$$Pr(R_i=r_i|R_{i-1}=r_{i-1},...,R_1=r_1)=\binom{n_i-m}{r_i}\lambda^{r_i}(1-\lambda)^{r_{i+1}-m} \tag{4.2.3}$$

where $\quad n_i=n-\sum_{j=1}^{i-1}r_j,\quad 0\le r_i\le n_i-m\quad fori=1,2,...,k-1$

We have the joint distribution of $D=(d_1,d_2,...,d_k)$ and $R=(R_1,R_2,...,R_{k-1})$ obtained as follows

$$\begin{aligned}P(R,\lambda) &= Pr(R_{k-1}=r_{k-1}|R_{k-2}=r_{k-2},...,R_1=r_1)\\ &\times Pr(R_{k-2}=r_{k-2}|R_{k-3}=r_{k-3},...,R_1=r_1)\\ &\times\dots\times Pr(R_2=r_2|R_1=r_1)Pr(R_1=r_1)\\ &=\tbinom{n_{k-1}-m}{r_{k-1}}\tbinom{n_{k-2}-m}{r_{k-2}}...\tbinom{n_2-m}{r_2}(\tbinom{n_1-m}{r_1})^{\lambda(\sum_{j-1}^{k-1}r_j)}\\ &\times\lambda^{\sum_{j-1}^{k-1}r_j}(1-\lambda)^{(k-1)(n-m)-\sum_{j=1}^{k-1}(k-j)r_j}\\ &=\frac{(n-m)!}{\prod_{j=1}^{k-1}r_j!(n_k-m)!}\lambda^{\sum_{j-1}^{k-1}r_j}(1-\lambda)^{(k-1)(n-m)-\sum_{j=1}^{k-1}(k-j)r_j}\end{aligned} \tag{4.2.4}$$

Therefore, the joint likelihood function based on the observations $D=(d_1,d_2,\dots,d_k)$

and $R = (R_1, R_2, ..., R_{k-1})$ can be written as

$$L(\theta, \lambda, k, D, R) = f(d_1, ..., d_k, k; \theta|R) \times P(R, \lambda) \tag{4.2.5}$$

Remark 4.2.1 *Type II censoring, interval censoring and progressive censoring are particular cases of Type II PICR censoring scheme because the later combines the features of all the above three with a provision to accommodate dropouts . Thus, the probability density function under Type II progressive censoring is obtained as a special case of Eq.(4.2.2) when all d_i's are fixed to be 1 and $t_i = T_(i)$, where $T_(i)$ is the i^{th} ordered survival time. When $R_i = 0$ for all i, we have the Type II censoring. On the other hand $R_i = 0$ for all i and $m = n$, it will be the interval censoring.*

4.3 Parameter Estimation

The maximum likelihood estimate of λ can be obtained directly by maximizing the Eqn.(4.2.4) because $f(d_1, ..., d_k, k, \theta|R)$ does not depend on λ. Gupta and Kundu (1999, 2000) studied the properties of maximum likelihood estimators(MLEs) of the parameters of GE distribution based on complete sample. They compared the MLEs with the other estimators like method of moments estimators, estimators based on percentages, least square estimators, weighted least square estimators etc. mainly with respect to their biases and mean square errors(MSEs)using extensive simulation technique. Further it is established that, the MLE works the best in almost all cases considered for estimating both α and β . Also the computational complexity

is minimal for MLE. For moderate or large sample sizes MLEs are well preferred to any other method. Hence we follow the maximum likelihood method estimation in the present context of Type II PICR censoring scheme.

Now normal equations are

$$\frac{\partial lnL}{\partial \alpha} = 0 \Rightarrow \quad \sum_{i=1}^{k} \left[\frac{d_i}{p_i - p_{i-1}} \left(\frac{\partial p_i}{\partial \alpha} - \frac{\partial p_{i-1}}{\partial \alpha} \right) - \frac{p_i R_i ln(1-e^{-\beta t_i})}{1-p_i} \right] = 0 \tag{4.3.1}$$

$$\frac{\partial lnL}{\partial \beta} = 0 \Rightarrow \quad \sum_{i=1}^{k} \left[\frac{d_i}{p_{i-1} - p_i} \left(\frac{\partial p_i}{\partial \beta} - \frac{\partial p_{i-1}}{\partial \beta} \right) + \alpha \frac{R_i t_i e^{-\beta t_i} (1-e^{-\beta t_i})^{\alpha-1}}{1-p_i} \right] = 0 \tag{4.3.2}$$

$$\frac{\partial lnL}{\partial \lambda} = 0 \Rightarrow \quad \frac{1}{\lambda} \sum_{j=1}^{k-1} R_j - \frac{1}{1-\lambda} \left[(k-1)(n-m) - \sum_{j=1}^{k-1} (k-j) R_j \right] = 0 \tag{4.3.3}$$

where $\frac{\partial p_i}{\partial \alpha} = p_i log(1 - e^{-\beta t_i})$, $\quad \frac{\partial p_i}{\partial \beta} = \alpha t_i e^{-\beta t_i} (1 - e^{-\beta t_i})^{\alpha-1}$.

The MLE of λ is easily obtained from Eqn.(4.3.3), as

$$\hat{\lambda} = \frac{\sum_{j=1}^{k-1} R_j}{(k-1)(n-m) - \sum_{j=1}^{k-1} (k-j-1) R_j} \tag{4.3.4}$$

On the other hand the MLE of α and β can be solved from Eqn.(4.3.1) and Eqn.(4.3.2) by using iterative algorithms like Newton-Raphson method. Denote the Fisher information matrix associated with α, β and λ by $I(\alpha, \beta, \gamma)$, we write the partitioned form as follows

$$I(\alpha, \beta, \gamma) = E \begin{pmatrix} \frac{\partial^2 lnL}{\partial \alpha^2} & \frac{\partial^2 lnL}{\partial \alpha \partial \beta} & 0 \\ \frac{\partial^2 lnL}{\partial \alpha \partial \beta} & \frac{\partial^2 lnL}{\partial \beta^2} & 0 \\ 0 & 0 & \frac{\partial^2 lnL}{\partial \lambda^2} \end{pmatrix}$$

$$= \begin{pmatrix} I_1(\alpha, \beta) & 0 \\ 0 & I_2(\lambda) \end{pmatrix} \tag{4.3.5}$$

From the Eqns.(4.3.1) to (4.3.3), the second-order partial derivatives are

$$\frac{\partial^2 lnL}{\partial\alpha^2} = \sum_{i=1}^{k}\left[\frac{d_i}{p_i - p_{i-1}}\left(\frac{\partial^2 p_i}{\partial\alpha^2} - \frac{\partial^2 p_{i-1}}{\partial\alpha^2}\right) - \frac{d_i}{(p_i - p_{i-1})^2}\left(\frac{\partial p_i}{\partial\alpha} - \frac{\partial p_{i-1}}{\partial\alpha}\right)^2\right] - \sum_{i=1}^{k}\left[\frac{p_i R_i ln(1-e^{-\beta t_i})^2}{(1-p_i)^2}\right] \tag{4.3.6}$$

$$\frac{\partial^2 lnL}{\partial\beta^2} = \sum_{i=1}^{k}\left[\frac{d_i}{p_i - p_{i-1}}\left(\frac{\partial^2 p_i}{\partial\beta^2} - \frac{\partial^2 p_{i-1}}{\partial\beta^2}\right) - \frac{d_i}{(p_i - p_{i-1})^2}\left(\frac{\partial p_i}{\partial\beta} - \frac{\partial p_{i-1}}{\partial\beta}\right)^2\right] - \alpha\sum_{i=1}^{k}\left[\frac{p_i R_i t_i^2(\alpha - e^{\beta t_i}(1-p_i))}{(1-e^{-\beta t_i})^2(1-p_i)^2}\right] \tag{4.3.7}$$

$$\frac{\partial^2 lnL}{\partial\alpha\partial\beta} = \sum_{i=1}^{k}\left[\frac{d_i}{p_i - p_{i-1}}\left(\frac{\partial^2 p_i}{\partial\alpha\partial\beta} - \frac{\partial^2 p_{i-1}}{\partial\alpha\partial\beta}\right) - \frac{d_i}{(p_i - p_{i-1})^2}\left(\frac{\partial p_i}{\partial\alpha} - \frac{\partial p_{i-1}}{\partial\alpha}\right)\left(\frac{\partial p_i}{\partial\beta} - \frac{\partial p_{i-1}}{\partial\beta}\right)\right] + \sum_{i=1}^{k}\left[\frac{p_i R_i t_i(1 - p_i + \alpha ln(1-e^{-\beta t_i}))}{(1-e^{\beta t_i})(1-p_i)^2}\right] \tag{4.3.8}$$

$$\frac{\partial^2 lnL}{\partial\lambda^2} = -\left(\frac{\sum_{i=1}^{k-1} R_j}{\lambda^2} + \frac{(k-1)(n-m) - \sum_{i=1}^{k-1}(k-j)R_j}{(1-\lambda)^2}\right) \tag{4.3.9}$$

where $\frac{\partial^2 p_i}{\partial\alpha^2} = p_i log(1-e^{-\beta t_i})^2$, $\frac{\partial^2 p_i}{\partial\beta^2} = \frac{\alpha p_i t_i^2(e^{-\beta t_i}-\alpha)}{(e^{-\beta t_i}-1)^2}$ and $\frac{\partial^2 p_i}{\partial\alpha\partial\beta} = \frac{p_i t_i(1+\alpha log(1-e^{-\beta t_i}))}{e^{-\beta t_i}-1}$.

Remark 4.3.1 *The closed form of expression of the expected values of these second order partial derivatives are not readily available . These terms can be evaluated by using numerical method. The standard errors of the estimators can be evaluated by using expressions (4.3.6), (4.3.7) and (4.3.9) The joint assymtotic distribution of the MLE of α and β is multivariate normal and in particular $(\sqrt{n}(\hat{\alpha}-\alpha), \sqrt{n}(\hat{\beta}-\beta)) \sim N_2(\boldsymbol{0}, nI_1(\alpha,\beta))$ where $I_1(\alpha,\beta)$ is given by (4.3.5)*

4.4 An Illustrative Example

We shall illustrate the methodology using a hypothetical data on leukemia patients. Suppose that a group of 70 leukemia patients was considered, end of every month progress of the group was recorded. However during the course of study some of the patients had to be removed from the study because they developed other infections. The study was terminated after majority of them (60 %) died. The following table shows its details

	Month								
t_i	1	2	3	4	5	6	7	8	9
d_i	6	5	8	6	5	4	3	2	3
r_i	2	3	3	2	3	2	2	2	0

where d_i=number of patients died during the i^{th} month and r_i=number of patients removed from the study at the i^{th} month.

The data from the study can be fitted into the GE model. Here n=70 and m=42. From the normal equations 3.1 and 3.2, the maximum likelihood estimate of α and β are obtained as $\hat{\alpha}$ =1.1563 and $\hat{\beta}$ =0.2069 with standard errors 0.618 and 0.146 respectively. The MLE of the removal probability $\hat{\lambda}$ =0.1234 with standard errors 0.0403

Remark 4.4.1 *The figure1 shows graph of Survival function S(t) of both Weibull and GE model for the data given in the example. It may be noted that the two graph shows almost perfect agreement and hence it suggests that the proposed GE-model can be used as an alternative to the Weibull model.*

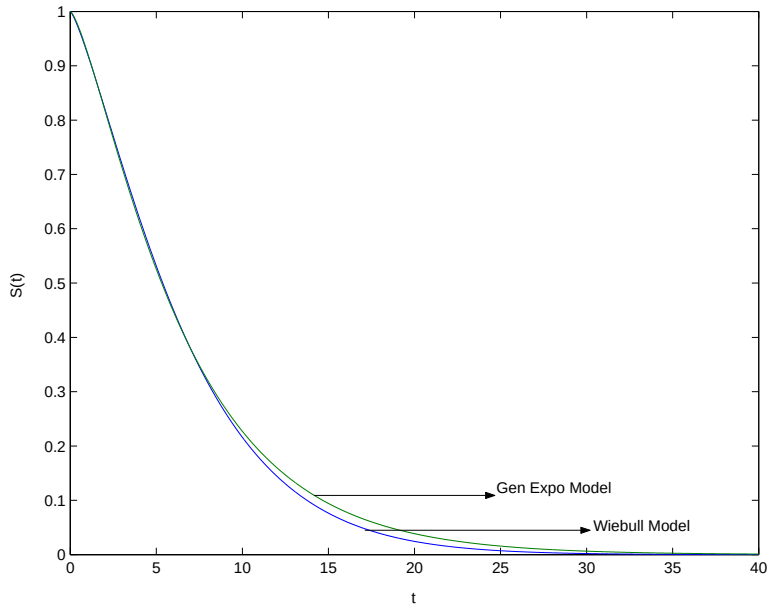

Figure 4.1: Survival Function S(t).

4.5 Discussion

The type II progressive interval censoring with random removals is a more flexible and practical censoring scheme since it integrates the features of Type II censoring, interval censoring and progressive censoring with random removals. The GE distributions are more flexible than gamma and as flexible as Weibull distributions (the distribution function of GE is in a closed form, the inference based on the censored data can be handled more easily than with gamma). Therefore the GE model can be used as a better alternative for analyzing lifetime data.

Chapter 5

OPTIMAL TREATMENT STRATEGY USING SEMI-MARKOV DECISION PROCESS

5.1 Introduction

In some medical treatment, decision must be made sequential and in an uncertain environment. A physician determining a course of treatment must consider patient's health as well as the best treatment decision in the future. Often decisions are to be taken in a dynamic environment. Physiological as well as physical changes in patients, may sometime contribute to the changes of the environment. Uncertain environment arises mainly due to patients respond differently even to same treatment for a disease.

Physicians always need to make subjective judgement about the treatment strategies. However a mathematical decision model that provide insight into the nature of optimal decision can aid the treatment. Markov decision processes (MDPs)are appropriate technique useful in class of problems involving complex, stochastic and dynamic decisions for which it can find optimal solutions. The goal of a MDP is to provide a optimal policy which is a decision strategy to optimize a particular criterion such as maximizing total discounted reward or minimizing the total discounted cost.

MDPs are a general framework for modeling dynamic systems under uncertainty. It binds previous, current and future treatment decision through the proper definition of patient's states defined as variables that contain the relevant information for making future decisions. The treatment model evolves in the following manner. The condition or state of patient is observed (or partially observed), an action is taken, a cost is incurred (or a reward is received) and the patient get into a new state according to a known probability distribution. The state variable defined so that given current state

of patient, the future transitions and rewards are independent of the past. It is the standard assumption of a Markov Process

The broad classes of MDPs are Finite Horizon MDPs and Infinite Horizon MDPs. The number N of decision epochs is finite in the former and it goes to infinite in the latter. For a finite horizon model optimal policy for both the average reward per state and the total reward criterion are equivalent. Infinite horizon models requires a large amount of data hence it is assumed that the data are time homogeneous. As a result the states of infinite horizon MDP must be carefully defined to ensure that the state transition of patients are stationary. When the data are time dependent the time homogeneity assumption can be satisfied by properly augmenting the state definition with the time at which the transition occurs

In most of the medical investigation, state of a patients is decided in the light of a series of medical tests which are subjected to test errors. A modified MDPs called Partially Observed MDPS (POMDP) have been developed to deal the data with imperfect information (Lovejoy (1991), White et al.(1989)) In these models it is assumed that uncertainty exist, in patient's transition and the state he/she truly occupies. Therefore the objective is to find an optimal policy based on the observation of the patient and the previous decision rule applied.

In MDP models the treatment decision are taken at each of a sequence of unit time intervals or fixed epochs and the sojourn time in states has no effect on rewards or incurring costs for patient. However in health-care and other application, decision are

taken over continuous time intervals such as varying treatment can be administered. The sojourn time in states may depend on the duration of his/her current health status. The MDP models might not be suitable to model such disease progression instead Semi-Markov Decision Process (SMDP) models are more appropriate. In SMDP models allow patients' state transition to occur in continuous time and allow to assume any probability distribution for sojourn time in a state.

For a range of multi-state diseases in which the data arises as transition-times and states, Markov and Semi-Markov process are appropriate models. Historically it has been difficult to adopt realistic models for biomedical applications since the likelihood turns out to be prohibitively complicated. With the development of computational methods and aids, the problem of tractability can be overcome. Kay(1986)introduced a k-state Markov model for continuous time processes for analyzing cancer markers. Similar models have been applied to AIDS (Longini et al(1989)), heart transplantation (Sharples(1993)), diabetes(Anderson(1988), Marshall and Jones (1995)), infectious diseases , dementia etc. Extension to these models have been further extended to include fixed time and time-varying covariate information (see for eg: Anderson et al (1991), Gauvreau et al(1994)). Methods for estimation of transition rates are generally numerically based and have usual maximum likelihood sampling schemes such as Metroplois-Hasting method(Sharples (1993), Prevost *et al.*(1998), Richard et al.(1993), or have used population-based approaches akin to weighted least squares(see Chen *et al.* 1996). These approaches have all concentrated on continuous-time Markov processes usually due to unequally spaced observation times.

In this chapter, we consider a complex survival model that lives in a randomly changing environment which affect model parameters. The term 'environment' is used in the generic sense so that it represents any set of conditions that affect the stochastic structure of the model investigated. The concept of 'environment' process in one form or another, has been used in the literature for various purpose.

The use of environmental process to modulate the deterministic and stochastic parameters of Operation Research models can be seen in reliability, inventory and queueing applications. One may refer to Ozekici and Soyer(2003)for an expository coverage in reliability theory. The problem of optimal replacement of a semi-Markov system under semi-Markov environment is studied by Hu and Yue(2003). Ozekici(1996), Ozekici and Parler(1999) discuss other applications in inventory and queueing. A comprehensive discussion on Markov modulated queueing system can be found in Prabu and Zhu(1988).

Although the literature cited above illustrate the use of random environment in reliability, inventory and queueing model, the concept is of paramount interest in survival analysis. It is generally assumed that a patient stays in a given fixed environment. The probability law of his ageing and death process there remain intact throughout his useful life. The life duration and corresponding hazard rate is taken to be the one obtained through statistical life testing procedures that are believed to be under ideal conditions.There has been growing interest in the recent years in lifetime models under random environment.

This is necessitated by the fact that the subject/patient often lives in varying environments during which they are subjected to varying environment conditions with significant effects on performance/health status. During a treatment period whole environment of the patient may change due to occurrence of other contagious diseases, hypertension, high blood pressure, cardiac problems, severe climatic/seasonal changes or adopting entirely new treatment strategy on medical team's advice. When environment changes, the state of patient also changes. The deterioration and failure process therefore depends on the environment. This makes it crucial to identify an optimal treatment strategy especially for a range of multi-state disease processes.

The remainder of this chapter is organized as follows. In section 2 prognosis of three diseases are discussed. In section 3 the model is presented in the frame of MDP. Section 4 deals with the problem of optimal control limit policies. Section 5 addresses a special case of Markov environment in which computationally feasible solution is arrived at. A numerical example is provided in the next section to illustrated the methodology. Followed by a discussion in the final section.

To begin with we shall examine some situations where we observe the progression of disease through stages.

5.2 Prognosis of Some Diseases

Alzheimer's Disease

It is a progressive, degenerative disease that destroys vital brain cells. As each area of the brain is affected, certain functions or abilities can be lost. The losses affect the individual's ability to think, to remember, to understand and to make decisions. In addition to affecting a person's mental abilities, Alzheimer's disease affects moods and emotions. Along with loss of abilities, changes in behaviour occur. Gradually, independence disappears. The progression of Alzheimer's disease varies from person to person and can span three to twenty years (the average length of the disease is between eight and twelve years). The progression can be described as a series of stages, providing a guide to the pattern of the disease, which can help when making care decisions. One staging system explains the disease in three stages: early, middle and late. Another staging system, often used by medical professionals, is the Global Deterioration Scale (also called the Reisberg Scale). This scale divides the disease into seven stages.

SmallPox

Highly contagious disease Smallpox was the most feared epidemic for centuries. Now it is eradicated from the most part of the globe by vaccination. Progression of smallpox has several medically well defined stages of disease.

Stages	Duration	Contagious?
1.Incubation period	7 to 17 days	Not Contagious
2.Initial symptoms (Prodome)	2 to 4 days	Possibly contagious
3.Early rash	About 4 days	Highly contagious
4.Pustular rash	About 5 days	Contagious
5.Pustules and scabs	About 5 days	Contagious
6.Resolving scabs	About 6 days	Contagious
7.Scabs resolved		Not contagious

Incubation period: Exposure to the virus is followed by an incubation period during which people do not have any symptoms and they may feel fine. The average incubation period is 12 to 14 days after exposure to the virus but it can range from 7 to 17 days. At this time the infected person is not contagious.

Initial Symptoms - Prodome: The first symptoms of smallpox include: Fever (38 degrees Celcius), malaise, head and body aches, and sometimes vomiting. At this time people are normally prostrate.

Early rash: The smallpox rash has a characteristic centrifugal distribution. The rash emerges first as small red spots on the tongue and in the mouth. These develop into sores and then the sores break open which releases large amounts of the virus into the mouth and throat. When this happens the person becomes contagious. About the same time that the sores break open a rash appears on the skin. It starts on the face, spreads to the arms and legs and then to the hands and feet. The rash will usually spread to all parts of the body within 24 hours. As the rash appears the fever may fall and the person may feel a bit better. The rash becomes raised bumps by the third day of the rash. By the forth day the bumps fill with a thick opaque fluid and often have a depression in the centre. The fever will often increase again at

this time and may remain high until scabs have formed over the bumps.

Pustular rash: The bumps become pustules, which are sharply raised and usually round and firm to the touch.

Pustules and scabs: The pustules begin to form a crust and then a scab. By the end of the second week after the rash has appeared most of the sores will have scabbed over.

Resolving Scabs: The scabs will begin to fall off but will often leave marks on the skin that will become pitted scars. Most of the scabs will have fallen off three weeks after the rash first appeared. The person is contagious to others until all of the scabs have fallen off. Scabs Resolved: Once the scabs have fallen off the person is no longer contagious.

Liver Disease

The prognosis of liver disease has Various stages namely , Inflammation, Fibrosis, Cirrhosis, Liver cancer and Liver failure. Healthy liver helps fight infections and cleans our blood. It also helps digest food and stores energy for when we need it. It has the amazing ability to grow back, or regenerate, when it is damaged. Anything that keeps the liver from doing its job - or from growing back after injury - may put our life in danger. Whether the liver is infected with a virus, injured by chemicals, or under attack from your own immune system, the basic danger is the same - that

the liver will become so damaged that it can no longer function properly.

Stage 1 . Inflammation: In the early stage of any liver disease, liver may become inflamed. It may become tender and enlarged. Inflammation shows that the body is trying to fight an infection or heal an injury. But if the inflammation continues over time, it can start to hurt liver permanently. When most other parts of our body become inflamed, we can feel it - the area becomes hot and painful. But an inflamed liver may cause no discomfort at all. If the liver disease is diagnosed and treated successfully at this stage, the inflammation may go away.

Stage 2 . Fibrosis: If left untreated, the inflamed liver will start to scar. As excess scar tissue grows, it replaces healthy liver tissue. This process is called fibrosis. (Scar tissue is a kind of fibrous tissue.) Scar tissue cannot do the work that healthy liver tissue can. Moreover, scar tissue can keep blood from flowing through the liver. As more scar tissue builds up, liver may not work as well as it once did. Or, the healthy part of liver has to work harder to make up for the scarred part. If the liver disease is diagnosed and treated successfully at this stage, there's still a chance that the liver can heal itself over time.

Stage 3 . Cirrhosis: But if left untreated, liver may become so seriously scarred that it can no longer heal itself. This stage - when the damage cannot be reversed - is called cirrhosis. Once one been diagnosed with cirrhosis, treatment will focus on keeping his/her condition from getting worse. It may be possible to stop or slow the liver damage. It is important to protect the healthy liver tissue that have

left. Cirrhosis can lead to a number of complications, including liver cancer. In some people, the symptoms of cirrhosis may be the first signs of liver disease.

Stage 4 .Liver cancer: Cancer that starts in the liver is called primary liver cancer. Cirrhosis and hepatitis B are leading risk factors for primary liver cancer. But cancer can develop in the liver at any stage in the progression of liver disease

Stage 5 . Liver failure Liver failure means that your liver is losing or has lost all of its function. It is a life-threatening condition that demands urgent medical care. The first symptoms of liver failure are often nausea, loss of appetite, fatigue, and diarrhea. Because these symptoms can have any number of causes, it may be hard to tell that the liver is failing. But as liver failure progresses, the symptoms become more serious. The patient may become confused and disoriented, and extremely sleepy. There is a risk of coma and death. Immediate treatment is needed. The medical team will try to save whatever part of the liver that still works. If this is not possible, the only option may be a liver transplant. When liver failure occurs as a result of cirrhosis, it usually means that the liver has been failing gradually for some time, possibly for years. This is called chronic liver failure. Chronic liver failure can also be caused by malnutrition. More rarely, liver failure can occur suddenly, in as little as 48 hours. This is called acute liver failure and is usually a reaction to poisoning or a medication overdose.

5.3 Semi-Markov Model for Treatment Strategy

we shall state the general form of the model that represents the foregoing situation may be stated as follows:

1. The patient is in a semi-Markov environment$\{(J_n, L_n), n \geq 0\}$ on a set K of countable environment states, where $J_n$is the state of the environment immediately after its nth transition epoch T_n, and $0 = T_0 < T_1 < T_2 < \cdots$. L_n is the time duration of the patient in the state J_n. Let the state's kernel be

$$G_{kk}(t) = Pr(T_{n-1} - T_n \leq t, J_{n+1} = k'/J_n = k)$$

and let $\psi_{kk'} = G_{kk'}(\infty)$ and $G_k(t) = \sum_{k' \in K} G_{kk'}(t)$.

2. During an environment state k, the patient goes through several states of disease according to a semi-Markov process with a kernel $\{P_{ij}^k(t), i, j \in S\}$ and a set $S = \{0, 1, 2, ...\}$ of countable states, where the state 0 represents disease-free state, and states 1, 2, . . . represent the different adverse disease states of the patient and the bigger the value, the more serious is the condition.

   Let $P_{ij}^k = P_{ij}^k(\infty), T_{ij}^k(t) = P_{ij}^k(t)/P_{ij}^k$ and $T_i^k(t) = \sum_{j \in S} P_{ij}^k(t)$

3. Suppose that the patient is in environment k, then one of the following two actions can be chosen if his state transfers to i :

   (a) Continue the present treatment strategy (denoted by C) with a cost rate $h^k(i)$;

   (b)Initiating a rejuvenating treatment strategy (denoted by R)like chemother-

apy, radiation, surgery, organ transplant and/or admission in ICU etc., with a cost rate $c^k(i)$, and the time of action R is assumed to be a random variable with probability distribution function $F^k(t)$, and the state after rejuvenation will be 0, the disease-free state.

4. When the environment state changes from k to k' if action C or R is chosen then the patient's state will change immediately according to a probability q_{ij}^k and an instantaneous cost $R^k(i, C)$occurs; while if action R is chosen then within no time it completed and an instantaneous cost $R^k(i, R)$ occurs.

5. The objective is to minimize the expected discounted total costs with discount factor $\alpha > 0$.

The above treatment strategy can be modeled by a semi-Markov decision process (SMDP) in a semi- Markov environment, presented and studied by Hu (1997), as follows.

During the environment state k, i.e., $J_n = k$ for some $n \geq 0$, it can be modeled by the following SMDP:

$$SMDP_k := \{S, A, p^k(j|i, a), T^k(.|i, a, j), r^k(i, a, j, u)\} \tag{5.3.1}$$

where S is the state space and A ={C, R}is the action set. The transition probability p^k, the distribution function T^k of the transition time, and the one step cost function

r^k are given, respectively by

$$
\begin{aligned}
P^k(j|i,C) &= P^k_{ij}\ , \qquad P^k(j|i,R) = \delta_{i0} \\
T^k(t|i,C,j) &= T^k_{ij}(t)\ , \quad T^k(t|i,R,0) = F^k(t) \\
r^k(i,C,j,u) &= h^k(i)\int_0^u e^{-\alpha t}dt = h^k(i)\alpha^{-1}(1-e^{-\alpha t}) \\
r^k(i,R,j,u) &= C^k(i)\int_0^u e^{-\alpha t}dt = C^k(i)\alpha^{-1}(1-e^{-\alpha t}) \qquad (5.3.2) \\
\delta_{i0} &= \begin{cases} 1 & \text{if} \quad j=0, \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

For SMDP in a semi-Markov environment, when the environment state changes from k to k' i.e., at L_{n+1} for some $n \geq 0$ with $J_n = k$ and $J_{n+1} = k'$, the patient's state changes immediately to j with a probability $q(j|i,a,k,k')$ if the patient's state is i at $L_{n+1} - 0$ and the last action taken before L_{n+1} is 'a', and in the same time, an instantaneous cost $R^k(i,a)$occurs, where

$$q(j|i,C,k,k') = q^k_{ij}$$

$$q(j|i,R,k,k') = \delta_{j0}$$

To simplify notations, for $k \in K$ and s, t ≥ 0, we let

$$
\begin{aligned}
h^k(s,t) &= \sum_{k'} Pr(T_{n+1} - T_n > t, J_{n+1} = k'/J_n = k)\int_0^t e^{-\alpha u}du \\
&\quad + \int_{s^+}^{s+t}\sum_k Pr(T_{n+1} - T_n \leq u, J_{n+1} = k'/J_n = k)\int_0^{u-s} e^{-\alpha l}dl \\
&= \alpha^{-1}(1-e^{-\alpha t})[1-G_k(t+s)] + \alpha^{-1}\int_{s^+}^{s+t}(1-e^{-\alpha(u-s)}dG_k(u), \\
g^{k.k'}(s,t) &= \int_{s^+}^{s+t} e^{-\alpha(u-s)}dG_{k'}(u), \qquad (5.3.3) \\
g^k(s,t) &= \sum_{k'\in K} g^{k.k'}(s,t) = \int_{s^+}^{s+t} e^{-\alpha(u-s)}dG_k(u),
\end{aligned}
$$

Let $x = (k, s, i) \in \Omega = \{(k, s, i) : k \geq 0, s \geq 0, i \in S\}$ be the mathematical state which means that the environment is in state k just since time s ago and the patient's state just transfers to i . Then, we define Expected discounted cost occurring when the state x is reached and action 'a' is taken,the r(x,a)

$$r(x, C) = h^k(i) \int_0^\infty h^k(s, t)\ dT_i^k(t) + R^k(i, C) \int_0^\infty g^k(s, t)\ dT_i^k(t)$$
$$r(x, R) = c^k(i) \int_0^\infty h^k(s, t)\ dF^k(t) + R^k(i, R) \int_0^\infty g^k(s, t)\ dF^k(t) \tag{5.3.4}$$

and $\beta(x, a, k')$ is corresponding to a discount factor depending on the state x,when the action is 'a' and the next environment state k'.

$$\beta(x, C, k') = \int_0^\infty g^{kk'}(s, t)\ dT_i^k(t)$$
$$\beta(x, R, k') = \int_0^\infty g^{kk'}(s, t)\ dF^k(t)$$

Now, it follows that $V^*(x)$, the minimal expected discounted total cost starting from the initial state x,is the minimal nonnegative solution of the following optimality equation

$$V^*(x) = min\{V^*(x, C),\ \ V^*(x, R)\} \tag{5.3.5}$$

where, $x = (k, s, i) \in \Omega$

$$V^*(x, C) = r(x, C) + \sum_{k' \in K} \beta(x, C, k') \sum_{j \in S} q_{ij}^k\ V^*(k', 0, j)$$
$$+ \sum_{j \in S} P_{ij}^k \int_0^\infty e^{-\alpha t} V^*(k, s + t, j)\ dT_{ij}^k(t) \tag{5.3.6}$$

$$V^*(x, R) = r(x, R) + \sum_{k' \in K} \beta(x, R, k')\ V^*(k', 0, 0)$$
$$+ \int_0^\infty e^{-\alpha t} V^*(k, s + t, 0)\ dF^k(t)$$

are respectively the discounted total cost if action C or R is used in the first horizon with the mathematical state x and then the optimal policy is used in the remaining horizons.

5.4 Optimal Control Limit Policies

From the standard results in discrete time Markov decision processes DTMDP, Eqn.(5.3.5) can be considered as an optimality equation for an adequate DTMDP with state space Ω Thus we can consider its n-horizon problem with the optimality equation

$$V_n^*(x) = min\{V_n^*(x,C), V_n^*(x,R)\}, \text{ where } x = (k,s,i) \in \Omega \tag{5.4.1}$$

where $V_n^*(x)$ is the optimal value from state x for n horizons problem, while

$$\begin{aligned} V_n^*(x,C) &= r(x,C) + \sum_{k\in K} \beta(x,C,k') \sum_{j\in S} q_{ij}^k \; V_{n-1}^*(k',0,j) \\ &\quad + \sum_{j\in S} P_{ij}^k \int_0^\infty e^{-\alpha t} V_{n-1}^*(k,s+t,j) \; dT_{ij}^k(t) \\ V_n^*(x,R) &= r(x,R) + \sum_{k\in K} \beta(x,R,k') \; V_{n-1}^*(k',0,0) \\ &\quad + \int_0^\infty e^{-\alpha t} V_{n-1}^*(k,s+t,0) \; dF^k(t) \end{aligned} \tag{5.4.2}$$

are the values from state x in n horizons if action C or R is used respectively in the first horizon and then an optimal policy in the remaining horizons. The initial conditions are

$$V_0^*(x,C) = V_0^*(x,C) = 0$$

$$Let \quad v_n(x) = V_n^*(x, C) - V_n^*(x, R), \quad v(x) = V^*(x, C) - V^*(x, R)\}$$

$$x = (k, s, i) \in \Omega$$

then it follows from the standard theory of DTMDP that

$$\lim_{n\to\infty} V_n^*(x, a) = V^*(x, a), \quad a = C, R \tag{5.4.3}$$

$$\lim_{n\to\infty} v_n(x) = v(x)$$

while the optimal policies can be depicted as $f_n^*(x) = C \iff v_n(x) < 0, f^*(x) = C \iff v(x) < 0$. So,$(f_N^*, f_{N_1}^*, ..., f_0^*)$ is optimal for N-horizon problem; and f^* is optimal for the infinite horizon discounted criterion. A concept of stochastic order between two distribution functions is needed. For two distribution functions F and G, F is said to be smaller stochastically than G, denoted by $F \preceq G$, if $F(t) \geq G(t)$ for each t

We have the following familiar result on stochastic order.

For two distribution functions F and G, $F \preceq G$ if and only if $\int_{-\infty}^{\infty} f(t)dF(t) \leq \int_{-\infty}^{\infty} f(t)dG(t)$ for each nondecreasing function f .

To obtain some properties of the optimal policies, we introduce the following assumption.

ASSUMPTIONS A

For each $k \in K$,

(A.1) $\sum_{j=m}^{\infty} q_{ij}^k$ nondecreasing in i for each $m \geq 0$;

(A.2) $h^k(i), c^k(i), R^k(i, C)$ and$R^k(i, R)$ are all nonnegative and nondecreasing in i ;

(A.3) both $h^k(i) - c^k(i)$ and $R^k(i, C) - R^k(i, R)$ are nondecreasing in i ;

(A.4) $F \preceq T_0^k \preceq T_1^k \preceq T_2^k \ldots$, i.e., T_i^k is stochastically nondecreasing in i and $F^k(.)$ is the smallest;

(A.5) $\int_0^\infty e^{-at} \sum_{j \in S} V(t, j) p_{ij}^k dT_{ij}^k(t)$ is nondecreasing in i if V (t, j)is nonnegative and nondecreasing in j for each $t \geq 0$.

As usual, Assumption (A.1) that $\sum_{j=m}^\infty q_{ij}^k$ is nondecreasing in i for $m > 0$ means that more serious is the condition of patient, faster the critical state reached by the environment change. Assumption (A.3) that $h^k(i) - c^k(i)$ is nondecreasing in i indicates that the treatment cost increases faster than the rejuvenating cost as the increasing of serious condition of patient , and similar for $R^k(i, C) - R^k(i, R)$. In fact, Assumptions A.1), (A.2) and (A.3) are those in the literature for the discrete time model,while Assumption (A.4) is given for the continuous time case here. Assumption (A.4) means that the sojourn time in a state for patient is nondecreasing as seriousness of his condition increases, and the rejuvenating time is smaller than the sojourn time in any state. Assumption (A.5) follows that if $T_{ij}^k(t)$ is absolutely continuous with probability density function $t_{ij}^k(t)$, and $\sum_{j=m}^\infty p_{ij}^k t_{ij}^k(t)$ is nondecreasing in i for each $t \geq 0, m \geq 0$, which is similar as (A.1). It is easy to see that the latter two conditions are involved in the state definition.

By the earlier stated result on stochastic order and Assumption A, it is easy to see that the $g_{kk'}(t)$ is nondecreasing in i which implies that $\beta(x, C, k')$ is nondecreasing in i.

We have the following well known result on transition probabilities

Let $[r_{ij}]$ *be a transition probability matrix, then the following two are equivalent:*

(1) For each $m \geq 0$, $\sum_{j=m}^{\infty} r_{ij}$ *is nondecreasing in* i ;

(2) For each nonnegative and nondecreasing function $h(\ j\ )$, $\sum_{j=0}^{\infty} r_{ij}h(j)$ *is nondecreasing in* i.

Then by using the induction method it can be shown from the above result and Assumption A that all of $V_n^*(x,C), V_n^*(x,R) and V^*(x)$ are nondecreasing in i.

Now for each $k \in K$ and $s \geq 0$

$$\begin{aligned} r(x,C) - r(x,R) = h^k(i) \int_0^\infty h^k(s,t)\ [dT_i^k(t) - dF^k(t)] \\ + [h^k(i) - c^k(i)] \int_0^\infty h^k(s,t)\ dF^k(t) \\ + R^k(i,C) \int_0^\infty g^k(s,t)\ [dT_i^k(t) - dF^k(t)] \\ + [R^k(i,C) - R^k(i,R)] \int_0^\infty g^k(s,t)\ dF^k(t) \end{aligned}$$

is also nondecreasing in i due to Assumption A, result on stochastic order and the fact that both $h^k(s,t)$ and $g^k(s,t)$ are nondecreasing in t for each $k \in K, s \geq 0$. It should be noted that the latter two terms in $V_n^*(x,R)$ of Eqn.(5.4.2) are independent of i . So $v_n(x)$ is nondecreasing in i and thus the following result

Theorem 5.4.1 *Under Assumption A, both* $V_n^*(k,s,i)$ *and* $v_n(k,s,i)$ *are nondecreas-*

ing in i for each $n \geq 0, k \in K, s \geq 0,$ *so*

$$\begin{aligned} V_n^*(k,s,i) &= V_n^*(k,s,i,c), 0 \leq i \leq i_n^*(k,s) \\ &= V_n^*(k,s,i,R), i \geq i_n^*(k,s) \end{aligned} \tag{5.4.4}$$

$$where \;\; i_n^*(k,s) := min(i|v_n(k,s,i) \geq 0).$$

Similarly, both $V_(^*k,s,i)$ *and* $v(k,s,i)$ *are also nondecreasing in* i *and*

$$\begin{aligned} V^*(k,s,i) &= V^*(k,s,i,c), 0 \leq i \leq i^*(k,s) \\ &= V^*(k,s,i,R), i \geq i^*(k,s) \end{aligned} \tag{5.4.5}$$

$$where \;\; i^*(k,s) := min(i|v(k,s,i) \geq 0).$$

The above Theorem states that there exists a state limit $i^*(k,s)$ just since time s ago for each $k \in K$ and $s \geq 0$ such that if the patient enters a state i while the environment is in state k, then the optimal action is to replace the patient by a new one if and only if the deteriorative degree of the patient is over the limit $i^*(k,s)$, i.e., $i \leq i^*(k,s)$. Such a policy is called a control limit policy. So the Theorem shows that there exists optimal control limit policies for both finite and infinite-horizon problems. In the next section, we will discuss a special case of Markov environment and get some better results.

5.5 Markov Environment-A Special Case.

In this section, we consider that the environment is Markov as follows:

$$G_{kk'}(t) = \psi_{kk'} G_k(t), \quad G_k(t) = 1 - e^{-\lambda_k t}, t \geq 0, k \; and \; k' \; \in \; K. \tag{5.5.1}$$

In this case, it will be shown that the variable s in state x = (k, s, i) can be deleted. Let

$$\begin{aligned}
t_F^k &= \int_0^\infty [1 - e^{-(\lambda_k + \alpha)t}] dF^k(t) \\
t_{ij}^k &= \int_0^\infty [1 - e^{-(\lambda_k + \alpha)t}] dT_{ij}^k(t)] \\
t_t^k &= \sum_{j \in S} P_{ij}^k t_{ij}^k, \;\; \alpha_F^k = 1 - t_F^k, \;\; \alpha_{ij}^k = 1 - t_{ij}^k, \;\; \alpha_i^k = 1 - t_i^k
\end{aligned} \tag{5.5.2}$$

where $F^k(t)$ and$T_{ij}^k(t)$ are defined in Section 1. Then it can be calculated, due to Eqn.(5.3.4), that

$$\begin{aligned}
r(x, C) = r'(k, i, C) e^{-(\lambda_k s)} &= \frac{t_i^k}{\lambda_k + \alpha} [h^k(i) + \lambda_k R^k(i, C)] e^{-(\lambda_k s)} \\
r(x, R) = r'(k, i, R) e^{-(\lambda_k s)} &= \frac{t_F^k}{\lambda_k + \alpha} [c^k(i) + \lambda_k R^k(i, R)] e^{-(\lambda_k s)} \\
\beta(x, C, k') &= \frac{\lambda_k t_i^k}{\lambda_k + \alpha} \psi_{kk'} e^{-\lambda_k s} \\
\beta(x, R, k') &= \frac{\lambda_k t_F^k}{\lambda_k + \alpha} \psi_{kk'} e^{-\lambda_k s}
\end{aligned} \tag{5.5.3}$$

Based on Eqn.(5.5.3), it can be shown that $e^{\lambda_k s} V^*(k, s, i)$ and therefore $e^{\lambda_k s} V^*(k, s, i, C)$, $e^{\lambda_k s} V^*(k, s, i, R)$ are independent of s and thus

$$e^{\lambda_k s} V^*(k, s, i) = V^*(k, 0, i)$$

$$e^{\lambda_k s} V^*(k, s, i, C) = V^*(k, 0, i, C)$$

$$e^{\lambda_k s} V^*(k, s, i, R) = V^*(k, 0, i, R)$$

We denote by

$$V^*(k,i) := V^*(k,0,i), \;\; V^*(k,i,C) := V^*(k,0,i,C), \;\; V^*(k,i,R) := V^*(k,0,i,R)$$

and

$$v(k,i) = V^*(k,i,C) - V^*(k,i,R)$$

Then $V^*(k,i)$ is the minimal nonnegative solution of the following optimality equation

$$V^*(k,i) = min\{V^*(k,i,C), \;\; V^*(k,i,R)\} \tag{5.5.4}$$

with corresponding

$$V^*(k,i,C) = r'(k,i,C) + \frac{\lambda_k t_i^k}{\lambda_k + \alpha} \sum_{k' \in K} \psi_{kk'} \sum_{j \in S} q_{ij}^k V^*(k',j) + \sum_{j \in S} p_{ij}^k \alpha_{ij}^k V^*(k,j)$$

$$V^*(k,i,R) = r'(k,i,R) + \frac{\lambda_k t_F^k}{\lambda_k + \alpha} \sum_{k' \in K} \psi_{kk'} V^*(k',0) + \alpha_F^k V^*(k,0) \tag{5.5.5}$$

Now, the problem is simplified by deleting the time variable s, and we can solve for $V^*(k,i)$ only. From the standard results in DTMDP, Eqn.(5.5.4) can also be considered as the optimality equation of an adequate defined DTMDP with state space $S' = \{(k,i) : k \in K, i \in S\}$ and action set A ={C, R}.

In the case of Markov environment, Assumptions (A.4) and (A.5) can be replaced, respectively, by the following weaker ones:

(A.4′) $t_F^k \le t_0^k \le t_1^k \le t_2^k \le \ldots$ for each $k \in K$;

(A.5′) $\sum_{j=m}^{\infty} p_{ij}^k \alpha_{ij}^k$ is nondecreasing in i for each $k \in K$ and $m \ge 0$.

The following corollary can be proved exactly as that of Theorem 5.4.1 from the above discussions

Corollary 5.5.1 *For the Markov environment case, suppose that Assumptions (A.1), (A.2), (A.3), (A.4′), (A.5′) hold, then* $v(k,i) = V^*(k,i,C) - V^*(k,i,R)$ *is nondecreasing in* i *and*

$$\begin{aligned} V^*(k,i) &= V^*(k,i,O), \quad i < i^*(k) \\ &= V^*(k,i,R), \quad i \geqslant i^*(k) \end{aligned} \tag{5.5.6}$$

where $i^*(k,i) = min\{i | v(k,i) \geq 0\}$

The corollary says that the state limit is also independent of the time variable s, that is, $i^*(k,s) = i^*(k)$.

Remark 5.5.1 *(1) If* $t_F^k \leq t_0^k$ *is not true, then it can be shown similarly that Corollary 5.5.1 holds in* $i \geq J_k := min\{i | t_i^k \geq t_F^k\}$ *and* $i^*(k)$ *should be redefined by* $i^*(k) := min\{i \geq J_k | v(k,i) \geq 0\}$. *In this case, the optimal policy is to operate if* $J_k \leq i < i^*(k)$ *and to replace if* $i \geq i^*(k)$, *while it is not known what optimal action is when* $0 \leq i < J_k$. *We call such a policy an extended control limit policy.*

(2) Due to the expressions of r (x, a) of Eqn. 5.5.3, one can know that both the optimal value and the optimal policies depend on $T_{ij}^k(t)$ *only through* t_{ij}^k. *This is to say that the model with a Markov environment is robust with respect to the distribution function* $T_{ij}^k(t)$ *of the time of state transition for the patient.*

Moreover, if

$$P_{ij}^k(t) = P_{ij}^k T_i^k(t), \quad \forall\, i,j,k$$

k then, we can assume that the patient is Markov, i.e.,

$$T_i^k(t) = 1 - e^{-\mu_i^k t}$$

where ^k_i and t_i^k are determined by each other with

$$t_i^k = \frac{\lambda_k + \alpha}{\lambda_k + \alpha + \mu_i^k}$$

$$\mu_i^k = (\lambda_k + \alpha)\frac{1 - t_i^k}{t_i^k}$$

. In Assumption (A.4′), t_i^k is nondecreasing in i, so $\sum_{j=m}^{\infty} P_{ij}^k \alpha_{ij}^k = \sum_{j=m}^{\infty} P_{ij}^k (1 - t_{ij}^k)$ may not be nondecreasing. The following lemma gives a sufficient condition for it.

Lemma 5.5.1 *Suppose that $T_{ij}^k(t) = T_i^k(t)$ for all $i, j \in S$, $k \in K$ and $m \geq 0$, then $\sum_{j=m}^{\infty} p_{ij}^k \alpha_{ij}^k$ is nondecreasing in i if and only if*

$$\frac{t_{i+1}^k - t_i^k}{1 - t_i^k} \leq \frac{\sum_{j=m}^{\infty} P_{i+1,j}^k - \sum_{j=m}^{\infty} P_{i,j}^k}{\sum_{j=m}^{\infty} P_{ij}^k} \tag{5.5.7}$$

Eqn. (5.5.7) means that the increasing speed of $\sum_{j=m}^{\infty} P_{ij}^k$ in i for each $m \geq 0$ is larger than or equals to the decreasing speed of $(1 - t_i^k)$.

Proof: It follows the given condition that

$$t_{i,j}^k = t_i^k, \;\; \alpha_{i,j}^k = \alpha_i^k = 1 - t_i^k, \;\; \sum_{j=m}^{\infty} P_{ij}^k \alpha_{ij}^k = (1 - t_i^k) \sum_{j=m}^{\infty} P_{ij}^k$$

It is obvious that for two nonnegative functions h(i) and g(i), if h(i) is non-increasing while g(i) is nondecreasing, then h(i),g(i) is nondecreasing if and only if

$$\frac{h(i)}{h(i+1)} \leq \frac{g(i+1)}{g(i} \quad or \quad \frac{h(i) - h(i+1)}{h(i+1)} \leq \frac{g(i+1) - g(i)}{g(i}$$

which immediately implies the Lemma 4.1. The optimal policies f_n^* and f^* are characterized by $i_n^*(k)$ and $i^*(k)$, respectively.We have the following result about the upper bound of these numbers, which is useful for the state reduction problem discussed below.

Lemma 5.5.2 *Under the conditions given in Corollary 5.5.1, if* $t_0^k = t_F^k$, *then* $i_n^*(k) \leq i_0^*(k) := min\{i|\triangle r(k,i) \geq 0\}$ *and* $i^*(k) \leq i_0^*(k)$ *where* $\triangle r(k,i) = r'(k,i,C) - r'(k,i,R)$.

Proof: If $t_0^k = t_F^k$, then $\alpha_0^k = \alpha_F^k$. So it follows from Theorem 5.4.1 that

$$t_i^k \sum_{j \in S} q_{ij}^k V_n(k',j) - t_F^k V_n(k',0) \geq t_F^k \sum_{j \in S} q_{ij}^k V_n(k',0) - t_F^k V_n(k',0) \;=\; 0$$

$$\sum_{j \in S} p_{ij}^k \alpha_{ij}^k V_n(k,j) - \alpha_F^k V_n(k,0) \geq \sum_{j \in S} p_{0j}^k \alpha_{0j}^k V_n(k,j) - \alpha_F^k V_n(k,0)$$

$$= \alpha_0^k V_n(k,0) - \alpha_F^k V_n(k,0) = 0$$

So, we can get that $v_n(k,i) \geq \triangle r(k,i)$, which implies the lemma immediately. Assumption (A.3) is about the cost rate, we now replace it by a new one about the expected total cost in a state. (A.3′) for each k ∈ K, both $h^k(i)t_i^k - c^k(i)t_F^k$ and $R^k(i,C)t_i^k - R^k(i,R)t_F^k$ are nondecreasing in i . Here, $h^k(i)t_i^k$ and $c^k(i)t_F^k$ are respectively the expected treatment and rejuvenating treatment costs in state i when the environment state is k. So the nondecreasingness of $h^k(i)t_i^k - c^k(i)t_F^k$ means that the expected treatment cost increases faster than the expected rejuvenating treatment cost as the patient's state increases. The nondecreasingness of $R^k(i,C)t_i^k - R^k(i,R)t_F^k$ has a similar meaning.

Theorem 5.5.1 *Under Assumptions (A.1), (A.2), (A.3,), (A.4′) and (A.5′), for each* $k \in K$ *and* $n \geq 1, v_n(k,i) := V_n^*(k,i,C) - V_n^*(k,i,R)$ *is nondecreasing in* i *, so* $v_n(k,i) < 0$ *iff* $i < i_n^*(k) := min\{i|v_n(k,i) \geq 0\}$*; moreover,* $v(k,i) := V^*(k,i,C) - V^*(k,i,R)$ *is also nondecreasing in* i *, and* $v(k,i) < 0$ *iff* $i < i^*(k) := min\{i|v(k,i) \geq 0\}$*. Thus, there exist optimal control limit policies.*

Proof: It should be noted first that under the given conditions,

$$(\lambda_k + \alpha)\triangle r(k,i) = [h^k(i)t_i^k - c^k(i)t_F^k] + \lambda_k[R^k(i,C)t_i^k - R^k(i,R)t_F^k]$$

is nondecreasing in i . Then the theorem can be proved exactly as that of Theorem 5.4.1.

Remark 5.5.2 *The above theorem shows the existence of optimal control limit policies whose state limit* $i^*(k)$ *depends only on the environment state* k*. Thus, the Markov environment case is more simpler than the semi-Markov environment case.*

Now we reduce the number of states of the patient under the Markov environment (see Eqn.(5.5.1)). First, we suppose that

$$i_n^*(k) \leq j(k), \quad n \geq 0, \;\; k \in K \tag{5.5.8}$$

for some j (k), where $i_n^*(k)$ is defined in Theorem 5.5.1 .

By Theorem 5.5.1, we have

$$[V^*(k,i) = V^*(k,i,R) = r'(k,i,R) + V_0(k), \;\; i \geq j(k), \;\; k \in K \tag{5.5.9}$$

Where

$$V_0(k) = \frac{\lambda_k t_F^k}{\lambda_k + \alpha} \sum_{k' \in K} \psi_{kk'} V^*(k', 0) + \alpha_F^k V^*(k, 0).$$

Thus one can get by Eqn.(5.5.4) for i ≥ 0 as follows:

$$\begin{aligned}
V^*(k,i,C) &= r'(k,i,C) + \frac{\lambda_k t_i^k}{\lambda_k + \alpha} \sum_{k' \in K} \psi_{kk'} \{ \sum_{j=0}^{j(k')-1} q_{ij}^k V^*(k', j) \\
&+ \sum_{j=j(k')}^{\infty} q_{ij}^k [r'(k,i,R) + V_0(k')] \} \\
&+ \sum_{j=0}^{j(k)-1} p_{ij}^k \alpha_{ij}^k V^*(k,j) + \sum_{j=j(k)}^{\infty} p_{ij}^k \alpha_{ij}^k [r'(k,i,R) + V_0(k)] \\
&= r'(k,i,C) \frac{\lambda_k t_i^k}{\lambda_k + \alpha} \sum_{k' \in K} \psi_{kk'} \sum_{j=j(k')}^{\infty} q_{ij}^k [r'(k',j,R) - r'(k', j(k'), R)] \\
&+ \sum_{j=j(k')}^{\infty} p_{ij}^k \alpha_{ij}^k [r'(k,j,R) - r'(k, j(k), R)] \\
&+ \frac{\lambda_k t_i^k}{\lambda_k + \alpha} \sum_{k' \in K} \psi_{kk'} [\sum_{j=0}^{j(k')-1} q_{ij}^k V^*(k', j) + \sum_{j=j(k')}^{\infty} q_{ij}^k V^*(k', j(k'))] \\
&+ \sum_{j=0}^{j(k)-1} p_{ij}^k \alpha_{ij}^k V^*(k,j) + \sum_{j=j(k)}^{\infty} p_{ij}^k \alpha_{ij}^k V^*(k, j(k))
\end{aligned} \tag{5.5.10}$$

We define that

$$\begin{aligned}
\widetilde{q}_{ij}^{kk'} &= \begin{cases} q_{ij}^k, & \text{When} \quad j < j(k') \\ \sum_{j=j(k')}^{\infty} q_{ij}^k & \text{When} \quad j = j(k') \end{cases} \\
\widetilde{p}_{ij}^k &= \begin{cases} p_{ij}^k, & \text{When} \quad j < j(k') \\ \sum_{j=j(k')}^{\infty} p_{ij}^k & \text{When} \quad j = j(k') \end{cases} \\
\widetilde{T}_{ij}^k(t) &= \begin{cases} T_{ij}^k(t), & \text{When} \quad j < j(k') \\ \sum_{j=j(k)}^{\infty} p_{ij}^k T_{ij}^k(t) / \widetilde{p}_{i,j(k)}^k & \text{When} \quad j = j(k') \end{cases}
\end{aligned} \tag{5.5.11}$$

Thus

$$\sum_{j=j(k)}^{\infty} p_{ij}^k \alpha_{ij}^k = \widetilde{p}_{i,j(k)}^k \widetilde{\alpha}_{i,j(k)}^k$$

Where $\widetilde{\alpha}^k_{i,j(k)}$ is defined as $\alpha^k_{i,j(k)}$ with $T^k_{ij}(t)$ being replaced by $\widetilde{T}^k_{ij}(t)$. Let

$$\begin{aligned}
\widetilde{h}^k(i) &= h^k(i) + \lambda_k \sum_{k' \in K} \psi_{kk'} \sum_{j=j(K')}^{\infty} q^k_{ij} \frac{t^{k'}_F}{\lambda_{k'} + \alpha} [c^{k'}(j) - c^{k'}(j(k'))] \\
&\quad + (t^k_i)^{-1} t^k_F \sum_{j=j(K')}^{\infty} p^k_{ij} \alpha^k_{ij} [c^{k'}(j) - c^{k'}(j(k'))] \\
\widetilde{R}^k(i, C) &= R^k(i, C) + \sum_{k' \in K} \psi_{kk'} \sum_{j=j(k')}^{\infty} q^k_{ij} \frac{t^{k'}_F}{\lambda_{k'} + \alpha} \lambda_{k'} [R^{k'}(j, R) - R^{k'}(j(k'), R)] \\
&\quad + (t^k_i)^{-1} t^k_F \sum_{j=j(K)}^{\infty} p^k_{ij} \alpha^k_{ij} [R^k(j, R) - R^k(j(k), R)] \\
\widetilde{r}(k, i, C) &= \frac{t^k_i}{\lambda_k + \alpha} [\widetilde{h}^k(i) + \lambda_k R^k(i, C)]
\end{aligned}$$

It is easy to see that $\widetilde{r}(k, i, C)$ is still nondecreasing in i for each k under Assumption A. Then for $i \geq 0$,

$$V^*(k, i, C) = \widetilde{r}(k, i, C) + \frac{\lambda_k t^k_i}{\lambda_{k'} \alpha} \sum_{k' \in K} \psi_{kk'} \sum_{j=0}^{j(k')} \widetilde{q}^{kk'}_{ij} V^*(k', j) + \sum_{j=0}^{j(k)} \widetilde{p}^k_{ij} \widetilde{\alpha}^k_{ij} V^*(k, j). \tag{5.5.12}$$

Now, we construct a new rejuvenating model (NRM), which is similar as the original rejuvenating model (ORM) excepts that

1 the state set of the patient in environment k is $S_k = \{0, 1, ..., j(k)\}$ for k $\in$ K;

2 the parameters p^k_{ij}, $T^k_{ij}(t)$, q^k_{ij}, $h^k(i)$ and $R^k(i, C)$ are replaced by $\widetilde{p}^k_{ij}$, $\widetilde{T}^k_{ij}(t)$, $\widetilde{q}^k_{ij}$, $\widetilde{h}^k(i)$ and $\widetilde{R}^k(i, C)$, respectively, which are defined the above;

3 the patient must be given rejuvenating treatment in state $j(k)$ during environment state k (due to Eqn.(5.5.8)).

From the above discussions, we know that the NRM and the ORM are equivalent under the meanings that the optimal objective values are identical and their optimality equations are equivalent for both the finite- and infinite-horizon problems. So their optimal policies are identical. The difference between them is that the number of patient's states is finite for NRM. Certainly, the problem with finite states is simpler than that with infinite states, e.g., the computation for the case of finite states is feasible while that for the case of infinite states should be approximated.

When $j(k) \leq j^*$ for some j^*, it can take the state set as $S_k = \{0, 1, ..., j^*\}$, which is irrespective of k. In the remaining of this section, we consider two further special cases. The first is that the state of patient itself is Markov, i.e.,

$$T_{ij}^k(t) = 1 - e^{-\mu_{k.i}t}, \qquad F^k(t) = 1 - e^{-\mu_F(t} \tag{5.5.13}$$

$$then \quad \alpha_{ij}^k = \frac{\mu_{k.i}}{\lambda_k + \mu_{k.i} + \alpha}, \qquad t_{ij}^k = \frac{\lambda_k + \alpha}{\lambda_k + \mu_{k.i} + \alpha}$$

The second further special case is that the environment is a Poisson process with rate , i.e., the Markov environment (see Eqn.(5.5.1) with

$$\psi_{k,k'} = 1 \quad G_k(t) = 1 - e^{-\lambda t}, \;\; t \geq 0, \;\; k \in K. \tag{5.5.14}$$

Moreover, it is assumed that each adverse factor increases the degree of the seriousness in the condition of the patient with a probability distribution $\{q_j, j \geq 0\}$ as follows:

$$q_{ij}^k = 0 \;\; for \; j < 0 \quad and \;\; q_{ij}^k = q_{j-1} \;\; for \; j \geq i \tag{5.5.15}$$

Furthermore, all $p_{ij}^k, T_{ij}^k, h^k(i), c^k(i), R^k(i, C)$ and $R^k(i, R)$ are independent of k and will be denoted by p_{ij} , $T_{ij}(t)$ and so on, by only deleting k in the original notations.

Then $t_{ij}^k, t_i^k, \alpha_{ij}^k, \alpha_i^k, t_F^k, \alpha_F^k$ are also independent of k and will be denoted by t_{ij} , t_i and so on.

Under these conditions, it can be shown that $V^*(k,i)$ and therefore $V^*(k,i,C)$ and $V^*(k,i,R)$ are independent of k. So $i^*(k) = i^*$ is also independent of k.

5.6 Numerical Example

Consider a numerical example where the environment is a Markov process having two states with parameters as follows

$$\psi_{kk'} = \begin{pmatrix} 0.64 & 0.36 \\ 0.57 & 0.43 \end{pmatrix}, \quad \lambda_1 = 0.076, \quad \lambda_2 = 0.093$$

and the state transition probabilities for the system are

$$(P_{ij}^1) = \begin{pmatrix} 0.68 & 0.26 & 0.06 & 0.0 & 0.0 \\ 0.04 & 0.68 & 0.22 & 0.00 & 0.0 \\ 0.0 & 0.07 & 0.65 & 0.28 & 0.00 \\ 0.0 & 0.0 & 0.03 & 0.65 & 0.32 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix}, \quad (P_{ij}^2) = \begin{pmatrix} 0.83 & 0.15 & 0.02 & 0.0 & 0.0 \\ 0.03 & 0.72 & 0.18 & 0.07 & 0.0 \\ 0.0 & 0.04 & 0.67 & 0.18 & 0.11 \\ 0.0 & 0.0 & 0.05 & 0.61 & 0.34 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix}$$

while two probability systems caused by the environment changes are

$$(q_{ij}^1) = \begin{pmatrix} 0.58 & 0.28 & 0.14 & 0.00 & 0.0 \\ 0.06 & 0.48 & 0.28 & 0.18 & 0.0 \\ 0.0 & 0.04 & 0.41 & 0.27 & 0.18 \\ 0.0 & 0.0 & 0.0 & 0.42 & 0.58 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix}, \quad (q_{ij}^2) = \begin{pmatrix} 0.64 & 0.26 & 0.10 & 0.0 & 0.0 \\ 0.06 & 0.55 & 0.27 & 0.12 & 0.0 \\ 0.0 & 0.06 & 0.47 & 0.29 & 0.18 \\ 0.0 & 0.0 & 0.0 & 0.48 & 0.52 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix}$$

The cost rate funcitons are as follows:

$$h^1(i) = 18 + 2i, \quad c^1(i) = 46 + i, \quad R^1(i,C) = 55 + i, \quad R^1(i,R) = 0$$

$$h^2(i) = 15 + 2i, \quad c^2(i) = 41 + i, \quad R^2(i,C) = 50 + i, \quad R^2(i,R) = 0$$

Now, it is assumed that the continuous discount factor is $\alpha = 0.45$, and

$$(t_F^1, t_1^1, t_2^1, t_3^1, t_4^1) = (0.46, 0.73, 0.74, 0.76, 0.78, 0.80),$$

$$(t_F^2, t_1^2, t_2^2, t_3^2, t_4^2) = (0.57, 0.78, 0.79, 0.80, 0.82, 0.86).$$

Thus for i=0,1,2,3,4

$$r'(1,i,C) = \frac{t_i^1}{\lambda_1+\alpha}(21.496 + 2.076i), \qquad r'(2,i,C) = \frac{t_i^1}{\lambda_1+\alpha}(18.731 + 2.093i),$$
$$r'(1,i,R) = \frac{t_F^1}{\lambda_1+\alpha}(46 + i), \qquad r'(2,i,R) = \frac{t_F^1}{\lambda_1+\alpha}(41 + i),$$

Now we compute the finite-horizon optimal values $V_n(k,i)$ iteratively by

$$V_{n+1}(k,i,C) = r'(k,i,C) + \frac{\lambda_k t_i^k}{\lambda_{k'}+\alpha}\sum_{k'\in K}\psi_{kk'}\sum_{j\in S}q_{ij}^k V_n(k',j) + \sum_{j\in S}p_{ij}^k(1-t_j^k)V_n(k,j),$$

$$V_{n+1}(k,i,R) = r'(k,i,R) + \frac{\lambda_k t_i^F}{\lambda_{k'}+\alpha}\sum_{k'\in K}\psi_{kk'}V_n(k',0) + (1-t_F^k)V_n(k,0),$$

$$V_{n+1}(k,i) = min\{V_{n+1}(k,i,C) - V_{n+1}(k,i,R)\}$$

$$for\ n\ \geq\ 0\ with\ V_0(k,i,C) = V_0(k,i,R) = 0,\ \forall\ k,i \tag{5.6.1}$$

The numerical results are shown in Table 1 and Table 2 when n=26

$|V_{n+1}(k,i,a) - V_n(k,i,a)|\ \leq\ 0.01$ for all (k, i, a),

so we take the optimal value$V^*(k,i) = V_{27}(k,i)$. Now $v(k,i)$ is shown in the last line of Table (5.1) and thus the optimal limits for both the environments 1 and 2 is 3. Namely,

$$v^*(1) = 3, \quad v^*(2) = 3.$$

The optimal policy in this example, is to initiating a rejuvenating treatment strategy if and only if the state of the patient reaches or exceeds 3 in both the environments 1 and 2.

Table-5.1 Computed Values of $V_n(k,i)$

	$V_n(1,i)$ (Environment.1)					$V_n(2,i)$ (Environment.2)				
n	$i=0$	1	2	3	4	$i=0$	1	2	3	4
1	129.69	144.16	161.09	178.72	197.02	105.87	119.20	132.83	148.57	168.85
2	225.48	247.43	270.47	291.30	295.11	195.97	219.98	246.56	273.52	284.67
3	297.28	324.65	346.80	384.85	402.44	289.54	324.08	358.84	389.73	412.60
4	355.19	384.76	407.90	459.72	477.68	350.36	389.05	428.46	460.69	464.82
5	398.77	429.02	451.97	510.15	515.60	391.23	431.43	470.73	494.62	498.75
6	430.30	459.77	480.78	538.91	542.71	418.94	458.48	496.05	518.92	523.05
7	452.28	480.17	499.12	557.72	561.52	437.53	475.89	512.38	535.75	539.89
8	467.18	493.55	511.09	570.46	574.26	449.91	487.25	523.23	547.14	551.27
9	477.14	502.36	519.05	578.96	582.76	458.13	494.74	530.46	554.73	558.86
10	483.76	508.19	524.34	584.61	588.42	463.59	499.70	535.29	559.78	563.91
11	488.15	512.05	527.86	588.37	592.17	467.21	502.99	538.50	563.14	567.27
12	491.07	514.62	530.21	590.86	594.66	469.61	505.17	540.63	565.36	569.49
13	493.01	516.32	531.76	592.52	596.32	471.21	506.62	542.05	566.84	570.97
14	494.30	517.45	532.79	593.61	597.42	472.27	507.59	542.99	567.82	571.95
15	495.15	518.20	533.48	594.34	598.15	472.97	508.23	543.61	568.48	572.61
16	495.72	518.70	533.94	594.83	598.63	473.44	508.65	544.03	568.91	573.04
17	496.10	519.03	534.24	595.15	598.95	473.75	508.93	544.30	569.20	573.33
18	496.35	519.25	534.44	595.36	599.17	473.95	509.12	544.49	569.39	573.52
19	496.51	519.40	534.57	595.51	599.31	474.09	509.25	544.61	569.51	573.64
20	496.62	519.50	534.66	595.60	599.40	474.18	509.33	544.69	569.60	573.73
21	496.70	519.56	534.72	595.66	599.46	474.24	509.38	544.74	569.65	573.78
22	496.74	519.61	534.76	595.70	599.51	474.28	509.42	544.78	569.69	573.82
23	496.78	519.63	534.79	595.73	599.53	474.31	509.44	544.80	569.72	573.85
24	496.80	519.65	534.80	595.75	599.55	474.33	509.46	544.82	569.73	573.86
25	496.81	519.67	534.81	595.76	599.56	474.34	509.47	544.83	569.74	573.87
26	496.82	519.67	534.82	595.77	599.57	474.34	509.48	544.84	569.75	573.88
$v(k,i)$	-87.54	-68.49	-57.15	2.28	13.99	-83.01	-52.01	-20.78	5.58	16.40

5.7 Discussion

In this chapter, we studied an optimal treatment strategy of a patient in a semi-Markov environment. It considered the performance/health status both by the patient

itself in a semi-Markov setup and by the influence of the environment to the patient. For both the finite- and infinite-horizon discounted criterions, it was shown that there exist optimal control limit policies. A special case for a Markov environment was discussed.When the control limits are bounded for each environment state, the countable states of the patient was simplified equivalently to a finite one. Finally, a numerical example was illustrated to prove the correctness and validity of the analysis.

A.

Concluding Remarks

Modeling of any system, physical or biological, involve several variables which are random and most of the measurements of these variables are subjected to random measurement errors. Hence almost all the models describing the real life situations are stochastic models. The stochastic models have become an important branch of study in many scientific areas. They can be very effectively used for the improved understanding and interpretation of clinical trials as well as for life studies. The work reported in this thesis augment great deal towards the development of modern methodologies in stochastic models on clinical/survival trials.

We presented a method for analyzing longitudinal data that imposes minimal structure or restrictions on the mean responses over time and on the covariance among the repeated measures in *Chapter 3*. The method focusses on analyzing response profiles and can be applied to longitudinal data when the design is balanced, with the timing of the repeat measures common to all individuals in the study. This is method of modeling mean response. Although the method is applied to problems drawn from the health sciences they apply equally to the other areas of application. A numerical illustration with recent stock market data, emphasis the point. Application of method in problems in education, psychology, and other branches of the behavioral and social sciences is an interesting research problem in this regard.

A distinctive characteristic of survival data is that the event of interest may not be observed on every observational unit. This feature is known as censoring. In clinical and epidemiological studies, censoring is mainly caused by time restriction of the study. *Chapter 4* contributes a clinical study model under a more flexible

and practical censoring scheme namely Type II progressive interval censoring with random removals(PICR), based on probability structure of Generalized Exponential Distribution. The Type II PICR inherits wonderful features of type II censoring, interval censoring and progressive censoring with the provision to discard the subjects at end of any interval at will. The Generalized Exponential Distribution has some interesting features very similar to those of Weibull family and gamma family but a nice alternative to them in many situations. Maximum likelihood estimation of parameters of the generalized exponential model is discussed and their properties are studied in this Chapter. An illustrative example towards the application of the model is also given. The generalized exponential model is suggested as a better alternative for analysis of life time data.Thus we showed that one can assume an underlying random effect model with a parametric distribution such as generalized exponential and apply this methodology in a clinical or biological setting. One may set out to do this with added proviso that assuming prior information, one can then further extend the methodological application to the Bayesian framework.

The Markov decision processes are simple yet powerful models for sequential decision problems. In our proposed model of optimum treatment strategy in the *last chapter*, we assume that there is a state space; at each time the system occupies a certain state, and the decision maker, or controller, has a set of feasible actions for that state that can be applied. Semi-Markov stochastic model is a useful tool for predicting the evolution of infection of infectious diseases and the probability of an infected patients survival. In SMDP models allow patients' state transition to occur

in continuous time and allow to assume any probability distribution for sojourn time in a state.

In this chapter we proposed an optimal treatment strategy for subject/patient lives in varying random environments, imparting significant effects on performance/health status; using semi-markov decision process. The environment is modelled as a Semi-Markov Process and in each environment state, the patient goes through several states of disease according to a Semi-Markov Process. A special case for a Markov environment was discussed. When the control limits are bounded for each environment state, the countable states of patient was simplified equivalently to a finite one. Finally, a numerical example was illustrated to prove the correctness and validity of the analysis. Further studies should include the properties of monotone of optimal policies, rejuvenating treatments with varying success rate and so on. It is likely that the semi-Markov models will be more and more applied to epidemiology and this will be facilitated by the development of more flexible estimation methods, increasing power of computing and the availability of data from large cohort studies.

Obviously this model does not show all the potential of the semi-Markov environment. Indeed, by means of the backward recurrence time process it is possible to assess different transition probabilities as a function of the duration inside the states. Moreover, it is possible to attach a reward structure to the process that allows the possibility of doing a cost analysis that considers, for example, the cost of antiretroviral treatment and/or other social costs related to the dynamic evolution of the HIV infection. These features will be the object of future research.

```
#include<stdio.h>
#include<stdlib.h>
#include<conio.h>
#include<math.h>
#include<iostream.h>

#define alpha 0.045
double *v_alloc(int n)
{       double *v;
        v=(double *) calloc(n,sizeof(double));
        if(v==NULL)
        {       fprintf(stderr,"could not allocate memory");
                exit(1);}
        return v;
}
void init_vector(double *vector, int NoOfElements)
{       int i;
        for(i=0; i< NoOfElements; i++)
   vector[i] = 0.0;}

double **m_alloc(int n, int k)
{       int i;
        double **mat;
        mat=(double **) calloc(n,sizeof(double *));
        if(mat == NULL)
        {
                fprintf(stderr,"could not allocate memory");
                exit(1);}
        for(i=0; i<n; i++)
        {
                mat[i]=(double *) calloc(k,sizeof(double));
                if(mat[i] == NULL)
                {       fprintf(stderr,"could not allocate memory");
                        exit(1);}}
        return mat;}
void init_matrix(double **matrix, int dim1, int dim2)
{       int i, j;
        for (i=0; i<dim1; i++)
                for(j=0; j<dim2; j++)
                        matrix[i][j] = 0.0;}
void free_matrix(double **matrix, int dim1)
```

```
{       int i;
        for (i=0; i<dim1; i++)
                free(matrix[i]);
        free(matrix);}


double ***m3_alloc(int n, int k, int v)
{
        int i,j;
        double ***mat;
        mat=(double ***) calloc(n,sizeof(double **));
        if(mat == NULL)
        {       fprintf(stderr,"could not allocate memory");
                exit(1);}
        for(i=0; i<n; i++)
        {       mat[i]=(double **) calloc(k,sizeof(double *));
                if(mat[i] == NULL)
                {       fprintf(stderr,"could not allocate memory");
                        exit(1);}

                for(j=0; j<k; j++)
                {       mat[i][j]=(double *) calloc(v,sizeof(double));
                        if(mat[i][j] == NULL)
                        {       fprintf(stderr,"could not allocate memory");
                                exit(1);}}}

        return mat;}
void init_3dimmatrix(double ***matrix, int dim1, int dim2, int dim3)
{       int i, j, k;
        for (i=0; i<dim1; i++)
           for(j=0; j<dim2; j++)
                        for(k=0; k<dim3; k++)
                                matrix[i][j][k] = 0;}
void free_3dimmatrix(double ***matrix, int dim1, int dim2)
{       int i;
        for (i=0; i<dim1; i++)
           free_matrix(matrix[i],dim2);
        free(matrix);}
double mcm(double *m1, double *m2, int n)
{       double sum=0.0;
        for(int i=0; i<n; i++)
                sum = sum + m1[i] * m2[i];
        return sum;}
```

```
int main()
{       int i, j, k;
        double *O1, *R1, *O2, *R2;

        double phi[2][2]=
        {       {0.72, 0.28},
                {0.57, 0.43}};
        double lamda[2]={0.076, 0.091};
        double P[2][5][5]={{                    {0.66, 0.23, 0.11, 0.0,  0.0},
                                                {0.08, 0.62, 0.26, 0.04, 0.0},
                                                {0.0,  0.13, 0.53, 0.28, 0.06},
                                                {0.0,  0.0,  0.12, 0.65, 0.23},
                                                {0.0,  0.0,  0.0,  0.0,  1.0},},
                                        {       {0.78, 0.2,  0.02, 0.0,  0.0},
                                                {0.06, 0.64, 0.28, 0.02, 0.0},
                                                {0.0,  0.12, 0.58, 0.24, 0.06},
                                                {0.0,  0.0,  0.14, 0.61, 0.25},
                                                {0.0,  0.0,  0.0,  0.0,  1.0},}};

        double Q[2][5][5]={{                    {0.54, 0.24, 0.16, 0.06, 0.0},
                                                {0.11, 0.56, 0.28, 0.05, 0.0},
                                                {0.0,  0.07, 0.51, 0.27, 0.15},
                                                {0.0,  0.0,  0.0,  0.42, 0.58},
                                                {0.0,  0.0,  0.0,  0.0,  1.0},},
                                        {       {0.64, 0.26, 0.1,  0.0,  0.0},
                                                {0.06, 0.55, 0.27, 0.12, 0.0},
                                                {0.0,  0.10, 0.57, 0.29, 0.04},
                                                {0.0,  0.0,  0.0,  0.48, 0.52},
                                                {0.0,  0.0,  0.0,  0.0,  1.0},}};

        double T[2][6]=
        {       { 0.82, 0.83, 0.85, 0.86, 0.90, 0.49},
                { 0.80, 0.81, 0.82, 0.84, 0.88, 0.53}};
        O1=v_alloc(5);
        init_vector(O1, 5);
        O2=v_alloc(5);
        init_vector(O2, 5);
        R1=v_alloc(5);
        init_vector(R1, 5);
        R2=v_alloc(5);
        init_vector(R2, 5);
```

```
clrscr();
        double R[2][5];
        double r[2][5];
        cout<<"\n 1&&";
        for(i=0; i<5; i++)
        {       O1[i] = (T[0][i]/(lamda[0]+alpha))*(25+3.08*i);
                R[0][i] = (T[0][5]/(lamda[0]+alpha))*(55+i);
                  r[0][i] = O1[i];//< R[0][i] ? O1[i]: R[0][i];
                  printf("%6.2f", r[0][i]);
                        cout<<"&";} cout<<"&";
        for(i=0; i<5; i++)
        {       O2[i]=(T[1][i]/(lamda[1]+alpha))*(24.5+3.1*i);

                R[1][i]=(T[1][5]/(lamda[1]+alpha))*(50+i);

                 r[1][i] = O2[i] ;//< R[1][i] ? O2[i]: R[1][i];
                 printf("%6.2f", r[1][i]);
                 cout<<"&";}

        double VO1[5][5];
        double VR1[5][5];
        double v[2][5];
        cout<<"\n 2&";

        for(k=0; k<2; k++)
        {       cout<<"&";
                for(i=0; i<5; i++)
        {       VO1[k][i] = r[k][i] + ((1-T[k][i])*mcm(r[k], P[k][i], 5))
                                        + (lamda[k]*T[k][i])/(lamda[k]+alpha)*(phi[k]
[0]*mcm(r[0], Q[k][i], 5)
                                        + phi[k][1]*mcm(r[1], Q[k][i], 5));

                VR1[k][i] = R[k][i] + (lamda[k]*T[k][5])/(lamda[k]+alpha)*(phi[k][0]*
r[0][0]
                                        + phi[0][1] * r[1][0]) + (1 - T[k][5])* r[0][0] ;

                  v[k][i]=VO1[k][i]<VR1[k][i]? VO1[k][i]:VR1[k][i] ;
                  printf("%6.2f", v[k][i]);
        cout<<"&";}}
```

```
int flag;
                  int n;
                for(n=3;n<30 ;n++)
        {       printf("\n%2d", n);
                for(k=0; k<2; k++)
                {       cout<<"&";
                        for(i=0; i<5; i++)
                        {  VO1[k][i] = r[k][i] + ((1-T[k][i])*mcm(v[k], P[k][i], 5))
                                + (lamda[k]*T[k][i])/(lamda[k]+alpha)*(phi[k]
[0]*mcm(v[0], Q[k][i], 5)
                                + phi[k][1]*mcm(v[1], Q[k][i], 5));
                                VR1[k][i] = R[k][i] + (lamda[k]*T[k][5])/(lamda[k]
+alpha)*(phi[k][0]* v[0][0]
                                + phi[0][1] * v[1][0]) + (1 - T[k][5])* v[0][0] ;

                               v[k][i]=VO1[k][i]<VR1[k][i]? VO1[k][i]:VR1[k][i] ;

                               printf("%6.2f", v[k][i]);
                         cout<<"&";} }
        /*if(flag==10) break;
        flag=0;
                for(k=0; k<2; k++)
                        {for(i=0; i<5; i++)
                             {   int a=v[k][i]*100;
                                 int b=v[k][i]*100; }}
                            int d=a-b;
                            if(d==1 || d==0) flag++;
                                for(k=0; k<2; k++)
                                for(i=0; i<5; i++)
                                        v[k][i]= VO1[k][i];*/  }

                cout<<"\n\n ";
                for(k=0; k<2; k++)
                {       cout<<"& ";
                        for(i=0; i<5; i++)
                                        {printf("%6.2f ", VO1[k][i]-VR1[k][i] );
                                        cout<<"&"; }  }
        cout<<"\n\n     Optimum attained at "<<(n-1)<< "th iteration";
        getch();
        return 0;
}
```

C.

Bibliography

Bibliography

[1] Aggarwala, R. (2001). Progressive interval censoring: Some mathematical results with applications to inference. *Commun. Stat.Theory Methods*, **30**, 1921 1935.

[2] Albert, P. S. and Follmann, D.A. (2000). Modeling repeated count data subject to informative dropout. *Biometrics* **66**, 667 - 677.

[3] Albert, P.S., Follmann, D.A., Wang, S.A., and Suh, E.B.(2002). A latent autoregressive model for longitudinal binary data subject to informative missingness. *Biometrics* **51**, 631 - 642.

[4] Alling, D.(1958). *Biometrics* **14**,527 .

[5] Andersen, P.K., Hansen, L.S., and Keiding, N.(1991). Assessing the influence of reversible disease indicators on survival. *stat. Med.* **10**, 1061 - 1067.

[6] Andersen, P. K., Borgan, Q., Gill, R. D. and Keiding, N.(1993). *Statistical Models Based on Counting Processes*. Springer Series in Statistics, Springer, New York,USA.

[7] Andersen, P.K.(1988). Multi-state model in survival analysis: A study of nephropathy and mortality in diabetes. *Statatics in Medicine.* **7**, 661 - 670.

[8] Bain, L.J.(1978). *Satistical analysis of reliability and life testing models.* New York: Marcel Dekker Inc.

[9] Bain, L.J. and Engelhardt, M.(1991). *Satistical analysis of reliability and life testing models:Theory and Methods.* New York: Marcel Dekker Inc.

[10] Bryant, E. amd Gillings, D.(1985). Statistical analysis of longitudinal repeated measures designs in Biostatistics. *Statistics in Biomedical, Public Heallth and Environmental Sciences,* (Ed., Sen, P.K.), 252 - 282, North Holland.

[11] Cheng, B., Shao, J., Zhong, B. (2005). Last observation analysis in ANOVA and ANCOVA. *Statistica Sinica*

[12] Chiang, C.L.(1979). Survival and stages of disease. *Mathematical Biostatistics* **43**, 159 - 171.

[13] Cinlar, E.(1975). *Introduction to stochastic process.* Printice-Hall Inc, Englewood cliffs,NJ.

[14] Cox, D.R.(1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society* **B 34**, 187 - 200.

[15] D Amico, G., Janssen, J. and Manca, R.(2006). Homogeneous semi-Markov reliabilitymodels for credit risk management. *Decisions in Economics and Finance* **28(2)**, 79 - 93.

[16] Dabrowska, D. M. and Ho, W. T.(2006). Estimation in a semiparametric modulated renewal process. *Statistica Sinica* **16(1)**, 93 - 119.

[17] Davidov, O. (1999). The steady state probabilities for a regenerative semi-Markov processes with application to prevention and screening. *Applied Stochastic Models and Data Analysis* **15(1)**, 55 - 63.

[18] Davidov, O., and Zelen, M.(2000). Designing cancer prevention trials: a stochastic approach. *Statistics in Medicine* **19(15)**, 1983 - 1995.

[19] Dawson, J.D., and Lagakos, S. W.(1994). Size and power of two-sample tests of repeated measures data. *Biometrics***49**, 1022 - 1032.

[20] Dawson, J.D.(1994). Stratification of summary statistic tests according to missing patterns. *Statistics in Medicine*, **13**, 1853 - 1863.

[21] DeGruttola, V. and Tu, X.M.(1995). Modelling progression of CD-4 lymphocyte count and its relationship to survival time. *Biometrics* **50**, 1003 - 1014.

[22] Dempster, A.P., Laird, N.M. and Rubin, D.B.(1977). Maximum likelihood from incomplete data via EM algorithm. *Journal of the Royal Statistical Society* **B 39**, 1 - 38.

[23] Derman, C.(1970). *Finite State Markovian Decision Processes*, Academic press,New York.

[24] Di Biase, G.;Janssen, J. and Manca, R.(2005). Future pricing through homogeneous semi-Markov processes. *Applied Stochastic Models in Business and Industry* **21(3)**, 241 - 249.

[25] Diggle, P.J., Kenward, M.G.(1994). Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics* **43**, 49 - 93.

[26] Fitzmaurice, G.M.(2001). A conundrum in the analysis of change. *Nutrition* **17**, 360 - 361.

[27] Fix, E. and Neymman, J. (1951). *Human biology.* **23** ,205.

[28] Follmann, D.A. and Wu, M.C.(1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51**, 151-168.

[29] Gauvreau, K., Degruttola, V., Pagano, M.(1994). The effect of covariates on the induction time of AIDS using improved imputation of exact seroconversion times. *Statistics in Medicine*, **13**, 2021-2030.

[30] Gill, R. D.(1980). Nonparametric estimation based on censored observations of a Markov renewal process. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **53(1)** 97 - 116.

[31] Glynn, R.J., Laird, N.M. and Rubin, D.B.(1986). *Selection modelling versus mixture modelling with nonignorable nonresponses. In: Drawing Inferences from Self-selected Samples*, (Ed. Wainer, H). Springer Veralag, 115 - 142.

[32] Gupta, R. D. and Kundu, D.(1999). Generalised Exponential Distributions. *Austral. & New zealand J. Statist* **41(2)**, 173 - 188.

[33] Gupta, R. D. and Kundu, D.(2000). Generalised Exponential Distributions: Different Method of Estimations. *J. statist.Comput. Siml.* **00** 01 - 22.

[34] Heyting, A., Tolboom, J. and Essers, J.(1992). Statistical handling of dropouts in longitudinal clinical trials. *Statistics in Medicine* **11**, 2043 - 2061.

[35] Heyting, A., Tolboom, J., Essers, J.(1992). Statistical handling of drop-outs in longitudinal clinical trials. *Statistics in Medicine*, **11**, 2043 - 2061.

[36] Horvitz, D.G. and Thompson, D.J.(1952). A generalisation of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663 - 685.

[37] Howard, R.A.(1971). *Dynamic Probabilistic Systems. Vol. I: Markov Models*,John Wiley & Sons, NY .

[38] Howard, R.A.(1971). *Dynamic Probabilistic Systems. Vol.II: Semi-Markov Decision Processes* . John Wiley & Sons, NY.

[39] Hu, Q.(1997). Discounted semi-Markov decision process in a semi-Markov environment. *Optimization*, **39**, 367 - 382.

[40] Hu,Q. and Yue, W.(2003). Optimal replacement of a system according to a semi-Markov decision process in a semi-Markov environment. *Optimization Methods and Software* **18.2**, 181-196.

[41] Janssen, J. and Manca, R.(1997). A realistic non-homogeneous stochastic pension fund model on scenario basis. *Scandinavian Actuarial Journal* **2**, 113 - 137.

[42] Janssen, J. and Manca, R.(2006). *Applied Semi-Markov Processes*. Springer, New York, NY, USA.

[43] Jennrich, R.I. and Schluchter, M.D.(1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, **42**, 805 - 820.

[44] Kay,R.(1986). A Markov model for analyzing cancer markers and disease states in survival studies. *Biometrics,*, **42**, 855 - 865.

[45] Kenward, M.G. Molenberghs, G.(1999). Parametric models for incomplete continuous and categorical longitudinal data. *Statistical Methods in Medical Research*, **8**, 51 - 83.

[46] Khatri, C.G.(1966). A note on MANOVA model applied to problems in growth curves. *Annals of the Institute of Statistical Mathematics* **18**, 75 - 86.

[47] Laird, N.M. and Ware, J.H.(1982). Random effects models for longitudinal data. *Biometrics* **38**, 963-974.

[48] Laird, N.M.(1983). Further comparative analysis of pretest-posttest research designs. *American Statistician* **37**, 329 - 330.

[49] Laird, N.M.(1988). Missing data in longitudinal studies. *Statistics in Medicine* **7**, 305 - 315.

[50] Lawless, J.F.(2003). *Statistical Models and Methods for Lifetime Data Analysis*. John Wiley & Sons.

[51] Levy, P.(1954).Processus semi-markoviens. *Proceedings of the International Congress of Mathematicians, Amsterdam*, **3**, 416 - 426.

[52] Lim, H., Sun, J.G., Matthews, D.E.(2002). Maximum Likelihood Estimation of a Survival Function with a Change Point for Truncated and Interval-Censored Data. *Stat. Med.* **21**, 743 752.

[53] Limnios, N.and Oprisan, G.(2001). *Semi-Markov Processes and Reliability, Statistics for Industry and Technology*. Birkhauser, Boston, Mass, USA.

[54] Limnios, N. and Oubhi, B.(2005). Nonparametric estimation for semi-Markov processes based on ksample paths with application to reliability. *Proceedings of the 11th Symposium on Applied Stochastic Models and Data Analysis*. Brest, France, May 2005, 1061 - 1068.

[55] Little, R.J.A.(1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88**, 125 - 134.

[56] Little, R.J.A.(1995). Modelling the dropout mechanism in repeated measures studies. *Journal of the American Statistical Association* **90**, 1112 - 1121.

[57] Little, R.J.A. and Rubin, D.B.(2001). *Statistical Analysis with Missing Data*. John Wiley & Sons.

[58] Longini, I.M. Jr,Clark, W.S.,Byers, R.H .(1989). Statistical analysis of the stages of HIV infection using a Markov model. *Stat. Med.* **8**, 831-843.

[59] Lord, F.(1967). A paradox in the interpretation of group comparisons. *Psychometrical Bulletin* **68**, 304 - 305.

[60] Lovejoy, W .S.(1991). A survey of algorithm methods for partially observed Markov decision problems. *Annals of Operation Research* , **28**, 47-66.

[61] Mallinckrodt, C. H., Clark, W. S., Carroll, R. J., Molenberghs, G.(2003a). Assessing response profiles from incomplete longitudinal clinical trial data under regulatory considerations. *Journal of Biopharmaceutical Statistics*, **13**:179-190.

[62] Mallinckrodt, C. H., Sanger, T. M., Dube, S., Debrota, D. J.,Molenberghs, G., Carroll, R. J., Potter, W. M., Tollefson, G. D.(2003b). Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. *Biol. Psychiatry* **53**:754-760.

[63] Marshall, A.W. and Goldhammer, H.(1955). *Journal of the American Statistical Association*, **50**, 99.

[64] Marshall, G. and Jones, R.H.(1995). Multi-state models and diabetic retinopathy. *Statistics in Medicine*, **14**, 1975 - 1983.

[65] McHugh, R. and Matts, J.(1983). Post-stratification in the randomized clinical trial. *Biometrics* **39**, 217 - 225.

[66] Meeker, W.Q., Escobar, L.A.(1998). *Statistical Methods for Reliability Data.* Wiley: New York.

[67] Molenberghs, G., Michiels, B., Kenward, M.G. and Diggle, P.J. (1998). Monotone missing data and pattern-mixed models. *Statistica Neerlandica* **52**, 153 - 161.

[68] Ouhbi, B.and Limnios, N.(1999). Nonparametric estimation for semi-Markov processes based on its hazard rate functions. *Statistical Inference for Stochastic Processes*, **2**, 151 - 173.

[69] Ozekici, S. and Parlar, M.(1999). Inventory Models with Unreliable Suppliers in a Random Environment. *Annals of Operation Research,* **91**, 123 - 136.

[70] Ozekici, S.(1996). Complex Systems in Random Environments. in S.Ozekici, editor, *Reliability and Maintenance of Complex Systems; Berlin: Springer-Verlag,* NATO ASI Services F**154**, 137 - 157.

[71] Potthoff, R.F. and Roy, S.W.(1964). A generalised multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51**, 313 - 326.

[72] Prabu, N. V .and Zhu, Y.(1989). Markov Modulated Queueing System, *Queueing Systems* **5**, 215 - 246.

[73] Prevost, T.C, Launoy, G, Duffy, S.W., Chen, H.H.(1998). Estimating sensitivity and sojourn time in screening colorectal cancer: A comparison of statistical approaches. *Am. J. Epidemiology*, **148**, 609 - 619 .

[74] Pulkstenis, E. P., Ten Have, T. R., and Landis, J. R.(1998). Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication. *Journal of the American Statistical Association*, **93**, 438-450.

[75] Richard, A., Richardson, S.,Maccario, J.(1993). A three-state Markov model of plasmodium falciparum parasitemia. *Math. Biosci.*, **117**, 283 - 300 .

[76] Robins, J.M., Rotnitzky, A. and Zhao, L.P. (1995). Analysis of semi-parametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, **90**, 106 - 121.

[77] Rosenbaum, P.R. and Rubin, D.B.(1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society* **B 45**, 212 - 218.

[78] Rowell, J.G. and Walters, D.E.(1976). Analysing data with repeated observations on each experimental unit. *Journal of Agricultural Science* **87**, 423 - 432.

[79] Rubin, D.B.(1976). Inference and missing data. *Biometrika***63**, 581-592.

[80] Rubin, D.B.(1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.

[81] Schafer, J.(1997). *Analysis of Incomplete Multivariate Data*. Chapman and Hall.

[82] Scharfstein, D. O., Daniels, M. J., Robins,J. M.(2003). Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes. *Biostatistics***4**, 495 - 512.

[83] Schluchter, M.D.(1992). Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine* **11**, 1861 - 1870.

[84] Shao, J., and Zhong, B.(2003). Last observation carry-forward and intention-to-treat analysis. *Statistics in Medicine*, **22**, 2429 - 2441.

[85] Sharples, LD. (1993). Use of gibbs sampler to estimate transition rates between grades of coronary disease following cardiac transplantation. *Statistics in Medicine*, **12** , 1155 - 1169.

[86] Shih, W., Quan, H.(1998). Stratified testing for treatment effects with missing data. *Biometrics*, **54**, 782 - 787.

[87] Smith, W.L.(1955). Regenerative stochastic processes. *Proceedings of the Royal Society of London; Series A.***232**, 1 - 31.

[88] Sun, J.G (2001). Variance Estimation of a Survival Function for Interval-Censored Data. *Statistics in Medicine*, **20**, 1249-1257.

[89] Ten Have, T. R., Kunselman, A. R., Pulkstenis, E. P. and Landis, J.R. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative dropout. *Biometrics*

[90] Tijms, H. C.(1994). *Stochastic Models-An Alogorithmic Approach.*, John Wiley & Sons , New York.

[91] Ting, N.(2000). Carry-forward analysis. In: Chow, S., ed. Encyclopedia of Biopharmaceutical Statistics. *New York: Marcel Dekker,* pp. 103 - 109.

[92] Troxel, A.B.(1998). Analysis of longitudinal data with non-ignorable non-monotone missing values. *Applied Statistics* **47**, 425 - 438.

[93] Tse, S.K.,Yuen, H.K.,Yang, C.(2002). design and Analysis of Survival data Under an Integrated Type-II Interval Censoring Scheme. *Journal of Biopharmaceutical Statistics*,**12(3)**, 333 - 345.

[94] Valliant, R.(1993). Poststratification and conditional variance estimation. *Journal of the American Statistical Association*, **88**: 89-96.

[95] Verbeke, G.and Mohlenberghs, G.(2000). *Linear Mixed Models for Longitudinal Data*, Springer, New York.

[96] Ware, J.H.(2003). Interpreting incomplete data in studies of diet and weight loss. *New England Journal of Medicine* **348**, 2136 - 2137.

[97] Weiss, G.H and Zelen, M.(1963). A stochastic model for the interpretation of clinical trials. *Proc N.A.S*, 988 - 984.

[98] White, C.C and Scherer, W.T.(1989). Solution procedures for partially observed Markov decision processes. *Operation Research* , **37**, 791 - 797.

[99] Wu, M. C. and Bailey, K. R.(1989). Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. *Biometrics* **45**, 939 - 955.

[100] Wu, M. C. and Carroll, R. J.(1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175 - 188.

[101] Xavier,T.D. and Manoharan, M.(2007). Generalized Exponential Model Under Type II Progressive Interval Censoring With Random Removals. peresented in *International conference on statistical science, OR & IT ;* **2007 Jan 7-9**,Thirupati.

[102] Xavier,T.D. and Manoharan, M.(2007). Optimal Treatment Strategy Using Semi-Markov Decision Process. Communicated.

[103] Xiang, L. and Tse, S. K.(2005). Maximum Likelihood Estimation in survival studies under progressive Interval Censoring with Random Removals. *Journal of Biopharmaceutical Statistics*, **15**, 981 - 991.