Summer 6-22-2011

Estimation Algorithm for Mixture of Experts Recurrent Event Model

Timesha U. Brooks
Georgia State University

ESTIMATION ALGORITHM FOR MIXTURE OF EXPERTS RECURRENT EVENT MODEL

by

TIMESHA BROOKS

UNDER THE DIRECTION OF DR. JUN HAN

Abstract

This paper proposes a mixture of experts recurrent events model. This general model accommodates an unobservable frailty variable, intervention effect, influence of accumulating event occurrences, and covariate effects. A latent class variable is utilized to deal with a heterogeneous population and associated covariates. A homogeneous nonparametric baseline hazard and heterogeneous parametric covariate effects are assumed. Maximum likelihood principle is employed to obtain parameter estimates. Since the frailty variable and latent classes are unobserved, an estimation procedure is derived through the EM algorithm. A simulated data set is generated to illustrate the data structure of recurrent events for a heterogeneous population.

INDEX WORDS: Recurrent events, Mixture of experts, Accumulating events, Frailty, Effective age

ESTIMATION ALGORITHM FOR MIXTURE OF EXPERTS RECURRENT EVENT MODEL

by

TIMESHA BROOKS

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2011

ESTIMATION ALGORITHM FOR MIXTURE OF EXPERTS RECURRENT EVENT MODEL

by

TIMESHA BROOKS

Committee Chair: Dr. Jun Han

Committee: Dr. Yuanhui Xiao

Dr. Yichuan Zhao

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

June 2011

ACKNOWLEDGEMENTS

I would like to thank God for sending his son Jesus, for without him none of this would be possible. I thank Him for blessing me with an advisor who has been both patient, understanding, and a great help, Dr. Jun Han, where would I be without you. It has been both a struggle and a personal victory to get through this process, but you were there for me every step of the way and for that I am eternally grateful.

I would also like to thank Dr. Yuanhui Xiao, and Dr. Yichuan Zhao for taking the time out of your busy schedules to read and give feedback on my thesis.

To my husband who dealt with all of my gripes and groans through this process, thanks for letting me vent. You have been the driving force and an inspiration, I only want to make you proud. My family and friends who believed in me, even when I didn't believe in myself, I love you all.

Last but certainly not least to my fellow classmates who struggled with me through the years and made it all bearable. Amy Fomo, your wisdom and intellect in the arena of math is something I strive for, I don't know how I would have made it through some of these classes without your help. Garrick and Ben my partners in crime, thanks for enduring with me..."give us FREE"!!! To all my other classmates...if I can do it, so can you, keep on pushing.

Contents

List of Figures

Chapter 1

Introduction

1.1 Recurrent Events

Recurrent events happen in various areas of life such as public health, engineering, reliability studies, et al. Examples of recurrent events in public health settings are the re-occurrence of a low t-cell count in an HIV patient after being medically treated, relapse of a drug or alcohol abuser after going to rehab, recurring migraines, or onset of depression. From an engineering or reliability studies perspective, examples of recurrent events could be the breakdown of a machine, the failure of an operating system, or malfunctioning parts on an assembly line. The deteriorating episodes of visual acuity, and the turnover rate for a company are also examples of recurrent events in other settings.

As recurrent events pervade many studies in a variety of statistical fields it is important to develop statistical models and methods for such data. We have to consider several attributes such as the effects of intervention, the effect on the subject of accumulating event occurrences, and the potential correlation among event occurrences within each subject. We also consider that the number of events observed over the study period as well as the censoring mechanism associated with inter-event time is informative about the event occurrence process.

1.2 Literature Review

Recurrent event data has several models and methods of analysis. However in, Peña et al. [2007], the need for a more general and flexible class of models was observed, therefore they took the idea of some of the more current procedures for estimating the parameters of the general class of semi-parametric models for recurrent events and continued to develop them. Their model is inclusive of an effective age function, the impact of accumulating event occurrences on the unit, a link function in which the effect of possibly time-dependent covariates are incorporated, and allows the incorporation of unobservable frailty components which induce dependencies among the inter-event times for each unit. They developed a procedure for estimating the parameters of the general class of semiparamtric models for recurrent events proposed by Peña and Hollander [2004]. This model allows us to further examine the impact of interventions after each event occurrence, which in turn provides a safeguard for analyzing recurrent event data.

The difficulties of nonparametric estimation for the distribution function governing the time to event occurrence of a recurrent event is discussed in Peña, Strawderman, and Hollander [2001]. In order to deal with these difficulties, they derived a Kaplan-Meier and a Nelson-Aalen type estimator for the survival function and the cumulative hazard function respectively through counting processes of both calendar time and gap time index. Subjects under study were observed for random periods of time and the informative sum-quota nature of the data accrual scheme was explicitly accounted for. These types of concessions and heterogeneity of population, made it more technically difficult by invalidating the use of martingale methods. However we find that Han and Gonzalez

[2010] provided an approach to dealing with such issues by considering conditional doubly indexed processes, so that it becomes possible to apply the counting process theory to intervened recurrent events for a heterogeneous population.

Han et al. [2007] used the EM algorithm for maximum likelihood estimation of a latent class joint model of a longitudinal biomarker and recurrent events, given a parametrically specified baseline hazard and utilizing a penalized likelihood measure to determine the number of latent classes. This latent class model accommodated a heterogenous population and allowed them to involve a subpopulation structure as well. The direct maximization approach for latent class model led to difficulties in maximization thus creating the need to use the complete data in order to utilize the much simpler log-likelihood in helping to estimate the parameters [Han , 2005].

Chapter 2

M-O-E Recurrent Event Model

Various models and methods of analysis are used for recurrent event data, however this paper focuses on a method that satisfies the need for a general and flexible class of models that simultaneously incorporate the effects of covariates, the impact on the unit of accumulating event occurrences, the effect of interventions after each event occurrence, and the effect of unobserved variables, for a heterogeneous population. Thus our model will be an extension of the general class of models for recurrent events for a homogeneous population, developed by Peña et al. [2007].

In order to estimate the parameters of a general class of semiparamtric models for recurrent events we need to consider a model which will accommodate the various parameters that we are interested in. This general model should accommodate the effects of accumulating event occurrences and the effects caused by each intervention after an event occurs. A semiparametric maximum likelihood will be employed to estimate the model since we have a homogeneous baseline hazard and our coefficients for covariates are heterogeneous.

An EM algorithm is used to perform maximum likelihood estimation due to the fact that the latent group label and frailty variable are missing and the simplicity of complete data likelihood.

In the expectation step we look at the conditional expectation of the complete data log-likelihood and in the maximization step we will maximize the expected complete data likelihood. Once the estimation of theses parameters have been done, we will leave finding the standard errors of parameter estimates and simulation study of properties of parameter estimates to future research.

Generally we can use mixture of experts models [Jacobs et al. , 1991] for regression or classification problems of a heterogeneous population. Mixture of experts models attempt to solve problems using a "divide and conquer" technique in which complex problems are decomposed into a set of simpler sub-problems. These sub-problems are then defined over a covariate space that sometimes have overlapping regions. Using the divide and conquer technique, different experts are appropriate in different regions of the covariate space. Thus in general ME provide a richer class of models than ordinary generalized linear models [Rosen and Tanner , 1999].

2.1 Recurrent Events Data

Consider a subject that is being observed for the happening of a recurrent event over a study period which we define as $[0, \tau]$ where τ represents our study time or right censored variable for the object. We will represent consecutive times of event occurrences as $S_0 \equiv 0 < S_1 < S_2 < S_3 \dots$ and let $T_1, T_2, T_3, \dots$ be the times between successive occurrences. Thus for $j = 1, 2, 3 \dots$, $T_j = S_j - S_{j-1}$ and $S_j = T_1 + T_2 + \cdots + T_j$. The number of event occurrences over the observation period $[0, \tau]$, is $K = \max\{k \in \{0, 1, 2, \dots\} : S_k \leq \tau\}$ which is a random variable whose distribution depends on the distribution properties of τ and $[T_j]$. Thus with regards to the distributional properties of event occurrences, K is informative [Peña et al., 2007].

In this section we will discuss the general class of models for recurrent events that was proposed in Peña et al. [2007]. $[0\ ,\ \tau_i]$ represents our study period in which we are monitoring the i^{th} subject of n independent subjects for the occurrence of a recurrent event, where $i = 1, \cdots, n$, $\{T_{ij}, j = 1, 2, \cdots\}$ represents the gap times for the ith subject, and the successive calendar times of event occurrences are denoted by $\{S_{ij}, j = 1, 2, \cdots\}$. We will define $N_i^{\dagger}(s) = \sum_{j=1}^{\infty} I(S_{ij} \leq s, S_{ij} \leq \tau_i)$ as the number of event occurrences over the time period $[0, \tau_i]$, $R_i^{\dagger}(s) = I(s \leq \tau_i)$ as the at-risk indicator at time s, and $x_i(s)$ as the possibly time-dependent covariate. Instead of representing the recurrent event time data in terms of observation times, we can alternatively represent data in terms of a stochastic counting process

$$\{(N_i^{\dagger}(s), R_i^{\dagger}(s), x_i(s)) : 0 \leq s \leq \tau_i)\}, i = 1, \cdots, n. \tag{2.1}$$

For intervened recurrent event data, intervention information at observed event times S_{ij} such as mode of intervention and growth fashion like perfect repair, should be included as well.

2.2 MOE Model

The Mixtures of Experts model [Jacobs et al. , 1991] is used to provide a more flexible approach to modeling survival data. Rosen and Tanner [1999] discussed a mixture model which combined features of the classical Cox proportional hazards model with mixtures of expert models. They discovered that M-O-E model provides a "generic approach for relaxing the parametric assumptions in the systematic component of **any** statistical model". In this paper we will utilize the approach of M-O-E to represent a group membership submodel and a recurrent event submodel. Han and Gonzalez [2010] used the ME model for submodels using a heterogeneous baseline hazard, and homogeneous coefficients for covariates. However this paper takes into consideration a homogeneous baseline hazard and heterogeneous coefficients for covaratiates. Our approach and Han and Gonzalez [2010] reflect heterogeneity in two different ways. This modeling approach improves upon the model presented in Peña et al. [2007] in that our approach and [Han and Gonzalez , 2010] reflect the heterogeneity in the underlying population. This can be important in the area of survival analysis in that there are situations in which the heterogeniety is caused by some unobservable factors.

Latent class models have the ability to accommodate heterogeneous populations and take advantage of subpopulation structures. Here, we propose a latent class model, which is actually a finite mixture of experts model where each observation can be viewed as arising from a population which is a mixture of a finite number of subpopulations. The M-O-E model assumes that there are a finite number g of latent groups corresponding to different patterns of recurrent events. As mentioned before our model has two submodels, class membership and intervened recurrent events submodels.

2.3 Class Membership Sub-model

Let the group membership submodel consist of g latent groups, n subjects, and u covariates. Let $c_i = (c_{i1} \ldots, c_{ig})$ be the latent group vector that has a multinomial distribution with c_{ik} the indicator variable for subject i in group k. The probability $P(c_{ik} = 1)$ that subject i falls into group k, is modeled by a multinomial logit regression in which

$$\pi_{ik} = P(c_{ik} = 1) = \frac{\exp(w_i^T \gamma_k)}{\sum_{j=1}^{g} \exp(w_i^T \gamma_j)}, k = 1, \cdots, g, \tag{2.2}$$

where the covariate vector and the associated group-specific coefficient vector for subject i in group k are denoted by $w_i = (w_{i1}, \cdots, w_{iu})^T$, and $\gamma_k = (\gamma_{1k}, \cdots, \gamma_{uk})^T$ respectively, and $\gamma_1 = 0$ is imposed to guarantee model identifiability.

2.4 Recurrent Events Sub-model

The conditional intensity process for subject i falling in the kth class is given by

$$\xi_i(s|c_{ik} = 1, \zeta_i) = \zeta_i R_i^\dagger(s) h_{ik}(s), \tag{2.3}$$

$$h_{ik}(s) = \lambda(\mathcal{E}_i(s))\rho(N_i^\dagger(s-); \alpha_k)\varphi(\beta_k^T x_i(s)), \tag{2.4}$$

where ζ_i is the frailty variable, $\lambda(\cdot)$ is a homogeneous baseline hazard function, $\mathcal{E}_i(s)$ is the effective age [Peña and Hollander , 2004, Han et al. , 2007] of the subject i at calendar time s, $\rho(j; \alpha_k)$ is the event accumulation effect function of known form, α_k is class-specific accumulation effect coefficient, $N_i^\dagger(s-)$ is the number of accumulated events just before time s, $\varphi(\cdot)$ is a nonnegative link function of know form, x_i is the covariate vector, and β_k is the class-specific covariate coefficient vector.

In order for the model to be identifiable we will let ζ_i be a frailty variable with a mean of 1 and a variance of θ. The frailty variable is an unobserved random effect shared by subjects within a group. The effective age expresses the effect of repairs such as fixing a physical structure, and repairs to an operating system. It is an observable, predictable, nonnegative piecewise differentiable, and piecewise nondecreasing process [Peña and Hollander , 2004]. We will denote this process as $\mathcal{E}_i(s)$. This process can be a piecewise linear or nonlinear function of calendar time. If the effective age process is linear it is given by a piecewise linear function $\mathcal{E}_i(s) = b_{iN_i^\dagger(s-)} + m_{iN_i^\dagger(s-)}(s - S_{iN_i\dagger(s-)})$ where b_{ij} is the intercept at the repair time and m_{ij} is the aging rate, which can be determined by the person performing the repair/intervention [Peña et al., 2007].

Minimal repair is a standard case of linear effective age process, $\mathcal{E}_i(s) = s$, in which the effective age of a unit restarts at an age equal to the age just before the failure after an intervention [Han et al. , 2007]. For example upon the failure of an auto transmission it may just be restored to the state in which it was just before the failure, giving it an effective age equal to that of which it was prior to the transmission failing. Perfect repair would be another usual case of linear effective age process where, $\mathcal{E}_i(s) = s - S_{iN_i^\dagger(s-)}$, in which the effective age of the unit restarts at the age of zero after an intervention. In the case of the auto transmission this would be the equivalent of replacing the transmission with a new identical one.

The effect of accumulating event occurrences on the unit is represented by the function $\rho(\cdot, \alpha)$. This function will be increasing if the accrual of the number of event occurrences leads to a weakening of the subject, such as an increasing number of repairs to auto transmission. However if an increasing number of event occurrences lead to an improvement on the subject such as a patient going into rehab, then this function will be decreasing [Peña et al., 2007]. The exponential function with unknown parameter, $\rho(j; \alpha) = \alpha^j, j = 1, 2, 3, \ldots$ is a simple form for this function. We will assume that $\lambda(\cdot)$, the underlying baseline hazard, is nonparametrically specified.

A pictorial representation of the effective age process is given in Figure 1. We see that the subject started with an effective age of 0 at time $t = 0$, and grew linearly. At time $t = 0.9$ it received a perfect repair and at time $t = 2$ it received a minimal repair, which both grew linearly. At time $t = 2.8$ it received a perfect repair and its effective age grew in a linear manner. At time $t = 3.8$ the subject received a perfect repair and its effect age increased in a nonlinear concave style and censored at time $t = 5.2$.

Example of Effective age process with minimal and perfect repair

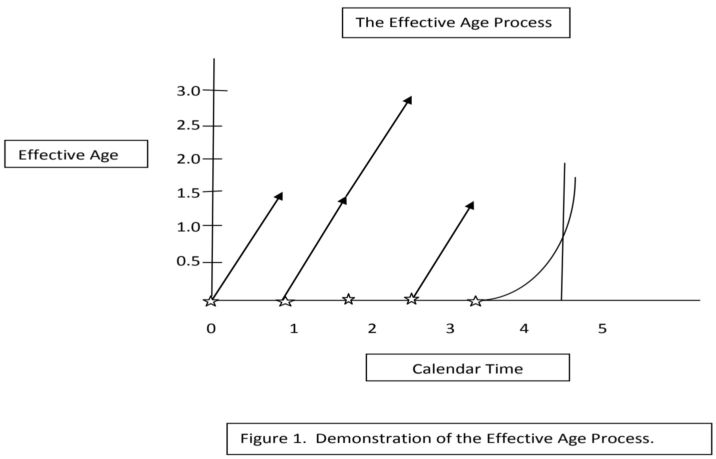

Figure 2.1: Example of Effective age process with minimal and perfect repair

Chapter 3

The Algorithm for Estimation

3.1 Likelihood Function

We will use a semiparametric maximum likelihood principle to estimate the parameters in the ME recurrent event model since the unobserved baseline hazard of the intensity process is non-parametrically specified, and the covariate effects are parametrically specified. For a fixed number of g latent groups, let $\phi = (\{\gamma_k\}, \theta, \Lambda, \{\alpha_k\}, \{\beta_k\})$ be the parameters in the ME model where $\gamma = (\gamma_2, \ldots, \gamma_g), \Lambda_k(t) = \int_0^t \lambda(u)du, \alpha = (\alpha_1, \cdots, \alpha_g)$, and $\beta = (\beta_1, \cdots, \beta_g)$. Let all the covariates $G_i(s) = (w_i, x_i(s))$ be for subject i, and the log-likelihood of the observed data $\{N_i^\dagger(s), R_i^\dagger(s), G_i(s) : s \leq \tau\}$ or simply $\{N_i^\dagger(s), R_i^\dagger(s), G_i\}$ be written as [Han et al. , 2007]

$$l_o(\phi) = \sum_{i=1}^{n} log \sum_{k=1}^{g} [c_{ik} = 1|G_i][N_i^\dagger, R_i^\dagger|c_{ik} = 1, G_i] \tag{3.1}$$

where $[c_{ik} = 1|G_i]$ is given by (2.3) and $[N_i^\dagger, R_i^\dagger|c_{ik} = 1, G_i]$ is given by

$$[N_i^\dagger, R_i^\dagger|c_{ik} = 1, G_i] = \int_0^\infty [N_i^\dagger, R_i^\dagger|G_i, \zeta_i, c_{ik} = 1] g(\zeta_i, \theta) d\zeta_i$$

$$= \frac{\Pi_{j=1}^{N_i^\dagger(s)-1}(1 + j\theta) \prod_{t\in[0,s]} [R_i^\dagger(t)h_{ik}(t)]^{dN_i^\dagger(t)}}{[1 + \theta \int_0^s R_i^\dagger(t)h_{ik}(t)dt]^{\theta^{-1}+N_i^\dagger(s)}}$$

$$= \frac{\prod_{t\in[0,s]}[(1+\theta N_i^\dagger(t-))R_i^\dagger(t)h_{ik}(t)]^{dN_i^\dagger(t)}}{[1+\theta\int_0^s R_i^\dagger(t)h_{ik}(t)dt]^{\theta^{-1}+N_i^\dagger(s)}} \tag{3.2}$$

Since observed data likelihood is complicated to maximize, for the sake of simplicity we will work with the complete data $\{N_i^\dagger, R_i^\dagger, c_i, \zeta_i, G_i\}$. Now the much simpler log-likelihood is given by [Han and Gonzalez , 2010]

$$l_c(\phi) = \sum_{i=1}^{n} \log[c_i|G_i] + \log[\zeta_i|G_i] + \log[N_i^\dagger, R_i^\dagger|c_i, \zeta_i, G_i] \tag{3.3}$$

where

$$\sum_{i=1}^{n} \log[c_i|G_i] = \sum_{i=1}^{n}\sum_{k=1}^{g} c_{ik} w_i^T \gamma_k - \sum_{i=1}^{n} \log\{\sum_{j=1}^{g} \exp(w_i^T \gamma_j)\}, \tag{3.4}$$

$$\sum_{i=1}^{n} \log[\zeta_i|G_i] = -n\log\Gamma(\theta^{-1}) - n\theta^{-1}\log(\theta) + (\theta^{-1}-1)\sum_{i=1}^{n}\log\zeta_i - \theta^{-1}\sum_{i=1}^{n}\zeta_i, \tag{3.5}$$

$$\sum_{i=1}^{n} \log[N_i^\dagger, R_i^\dagger|c_i, \zeta_i, G_i]$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{g} c_{ik}\{N_i^\dagger(s)\log\zeta_i + \sum_{i=1}^{N_i^\dagger(s)} \log(R_i^\dagger(S_{ij})h_{ik}(S_{ij}))\Delta N_i^\dagger(S_{ij}) - \zeta_i\int_0^s R_i^\dagger(v)h_{ik}(v)dv\} \tag{3.6}$$

3.2 Estimation of Baseline Hazard

The outline of the counting process for a nonparametric estimator for the baseline hazard $\Lambda(s)$ will be established in this section. Let $\tau = \max \tau_i$ and $F = \{F : 0 \leq s \leq \tau\}$ be the accumulated knowledge about what has happened up to time T or the history on some probability space (Ω, F, P) such that $N_i^\dagger(s)$ is a counting process and $R_i^\dagger$ is a predictable process with respect to F. Given the conditional intensity process in (2.4), the cumulative intensity process of $N_i^\dagger(s)$ can be written as [Han and Gonzalez , 2010]

$$A_i^\dagger(s|c_{ik}=1, \zeta_i, x_i) = \int_0^s \zeta_i R_i^\dagger(v)h_{ik}(v)dv \tag{3.7}$$

The residual process defined by

$$M_i^\dagger(s|c_{ik}=1,\zeta_i,x_i) = N_i^\dagger - A_i^\dagger(s|c_{ik}=1,\zeta_i,x_i) \tag{3.8}$$

is a square-integrable F-martingale, and hence the vector of processes $(M_1^\dagger, \cdots, M_n^\dagger)$ consists of orthogonal square-integrable martingales with predictable quadratic covariation processes

$$\langle M_i^\dagger, M_j^\dagger\rangle(s|c_{ik}=1,\zeta_i,x_i) = A_i^\dagger(s|c_{ik}=1,\zeta_i,x_i)I\{i=j\}.$$

Despite the martingale quality of the process defined in terms of calendar time s, the processes indexed by gap times between the event occurrences no more have martingale properties [Peña, Strawderman, and Hollander , 2001]. So we will exploit a doubly indexed process $Z_i(s,t) = I\{\mathcal{E}_i(s) \le t\}$ [Peña et al., 2007], $i = 1, \cdots, n$. Calendar times and gap times are denoted by the indices s, and t respectively. This process signifies whether the effective age of the ith subject at calendar time s is no more than t. Following Han and Gonzalez [2010], we can define the following conditional doubly indexed processes

$$N_i(s,t) = \int_0^s Z_i(v,t)N_i^\dagger(dv), A_i(s,t|c_{ik}=1,\zeta_i,x_i) = \int_0^s Z_i(v,t)A_i^\dagger(dv|c_{ik}=1,\zeta_i,x_i),$$

$$M_i(s,t|c_{ik}=1,\zeta_i,x_i) = N_i(s,t) - A_i(s,t|c_{ik}=1,\zeta_i,x_i) = \int_0^s Z_i(v,t)M_i^\dagger(dv|c_{ik}=1,\zeta_i,x_i)$$

The number of events for the ith subject that occurred over $[0,s]$ with effective ages at most t are counted by the doubly indexed process $N_i^\dagger(s,t)$. Each doubly indexed processes can be looked upon as single calendar time indexed processes constrained by effective age of t. The process $M_i(\cdot,t)$ is a square-integrable zero-mean martingale by the martingale property of $M_i^\dagger$ and predictability of $Z_i(\cdot,t)$, given a gap time t. The process $M_i(s,\cdot)$ is not a martingale for a fixed calendar time s, nevertheless it also has mean zero. Now it is possible to apply the counting process theory to intervened recurrent events via these doubly indexed processes.

Keeping in mind our goal is to estimate $\Lambda(t)$ for a given t, and considering $\lambda(\cdot)$ appearing in $A_i^\dagger(\cdot)$ is time transformed by the effective age process $\mathcal{E}(\cdot)$, we need to figure out the best way to produce $A_i(s,t)$ in such a way that $\Lambda(t)$ is represented directly rather than in its time-transformed version. To accomplish this, for $j = 1, \cdots, N_i^\dagger(\tau_i) + 1$ and $s \in [0, \tau_i]$, similar to Han and Gonzalez [2010] we define,

$$\mathcal{E}_{ij}(s) = \mathcal{E}_i(s) I\{S_{ij-1} < s \leq S_{ij}\}, \tag{3.9}$$

$$\varphi_{ij}(s; \alpha_k, \beta_k) = \rho[N_i^\dagger(s-); \alpha_k]\phi[\beta_k^T x_i(s)][\mathcal{E}'_{ij}(s)]^{-1}, \tag{3.10}$$

where $S_{iN_i^\dagger(\tau_i)+1} \equiv \tau_i$ and $\mathcal{E}'_{ij}(s) = (d/ds)\mathcal{E}_{ij}(s)$. The following proposition will provide an alternative expression for $A_i(s,t)$.

Proposition 1. For each $i = 1 \cdots, n, A_i(s,t)|c_{ik} = 1, \zeta_i, x_i) = \int_0^t \zeta R_i^{(k)}(s,u)\lambda(u)du$, where

$$R_i^{(k)}(s,u) \equiv R_i^{(k)}(s,u|\alpha_k,\beta_k) = \sum_{j=1}^{N_i^\dagger(s-)+1} I_{[E_{ij-1}+, E_{ij}]}(u)\varphi_{ij}^{(k)}(\mathcal{E}_{ij}^{-1}(u); \alpha_k, \beta_k),$$

$$E_{ij-1}+ = \mathcal{E}_i(S_{ij-1}+), E_{ij} = \mathcal{E}_i(S_{ij}), \quad and \quad E_{iN_i^\dagger(s-)+1} = \mathcal{E}_{iN_i^\dagger(s-)}(s\Lambda\tau_i).$$

Proposition 1 can be proven in an analogous manner as Proposition 1 in Han and Gonzalez [2010]. As we can see in our model, our A_i is based on coefficient covariates α, β that are both class specific and our baseline hazard λ is non-class specific. It is also important to note that a major underlying difference between our proposition 1 and that of Han and Gonzalez [2010] is the $\phi_{ij}s$ are defined differently. Because of the effectiveness of proposition 1 it follows that

$$M_i(s,t|c_{ik} = 1, \zeta_i, x_i) = N_i(s,t) - \int_0^t \zeta_i R_i^{(k)}(s,u)\Lambda(du).$$

Following the idea of [Han and Gonzalez , 2010], by multiplying c_{ik} and summing over i and k yields,

$$\sum_{k=1}^{g}\sum_{i=1}^{n} c_{ik} M_i(s,du|c_{ik} = 1, \zeta_i, x_i) = \sum_{k=1}^{g}\sum_{i=1}^{n} c_{ik} N_i(s,du) - \sum_{k=1}^{g}\sum_{i=1}^{n} c_{ik}\zeta_i R_i^{(k)}(s,u)\Lambda(du).$$

Since $E[\sum_{k=1}^{g}\sum_{i=1}^{n} c_{ik} M_i] = 0$, an application of Method of Moments yields,

$$0 = \sum_{i=1}^{n}(\sum_{k=1}^{g} c_{ik})N_i(s, du) - \sum_{k=1}^{g}\sum_{i=1}^{n} \zeta_i c_{ik} R_i^{(k)}(s, u)\Lambda(du)$$

$$0 = \sum_{i=1}^{n} N_i(s, du) - [\sum_{i=1}^{n} \zeta_i(\sum_{k=1}^{g} c_{ik} R_i^{(k)}(s, u)]\Lambda(du).$$

Consequently,

$$\hat{\Lambda}(du) = \frac{\sum_{i=1}^{n} N_i(s, du)}{[\sum_{i=1}^{n} \zeta_i(\sum_{k=1}^{g} c_{ik} R_i^{(k)}(s, u))]}$$

Integration gives,

$$\hat{\Lambda}(s, t) = \int_{o}^{t} \frac{J(s, u|\alpha, \beta)}{R(s, u)} \sum_{i=1}^{n} N_i(s, du). \tag{3.11}$$

where $J(s, u) = I\{R(s, u) > 0\}$, $R(s, u) = \sum_{i=1}^{n}(\sum_{k=1}^{g} c_{ik}\zeta R_i^{(k)}(s, u))$. By the product-integral representation and substitution principle, the estimator of baseline survivor function defined by $S(t) = \exp[-\Lambda_k(t)]$ can be written as

$$\hat{S}(s, t) = \prod_{u=0}^{t}[1 - \hat{\Lambda}(s, du)] = \prod_{u=0}^{t}\left[1 - \frac{\sum_{i=1}^{n} N_i(s, du)}{R(s, t)}\right] \tag{3.12}$$

3.3 The Expectation Step of the EM algorithm

Because of the use and simplicity of complete data likelihood and the absence of latent group label and frailty, we can use an EM algorithm [Dempster et al., 1977] for estimation. We will look at both steps of the EM algorithm, first, the Expectation step, and then the Maximization step. With the Expectation step we must take the conditional expectation of the complete data log-likelihood (3.3) given the observed data $\{N_i^{\dagger}, R_i^{\dagger}, G_i\}$ with the current parameter estimates. Then, we will maximize the expected complete data log-likelihood with respect to the components of parameter Φ during the Maximization step. These steps will alternate until some convergence criterion is met.

We will assume that the jth parameter estimate $\phi^{(j)}$ in the EM iteration is available. Let $\tilde{a}_i = E(a_i|N_i^\dagger, R_i^\dagger, G_i; \phi^{(j)}) \equiv \tilde{E}(a_i)$ denote the conditional expectation of a random variable a_i using the jth parameter estimate and given the ith observed data. In the expectation step, we use Bayes formula which will allow us to derive the conditional expectation of class membership.

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{3.13}$$

The expectation of group membership c_{ik} given the observed data $\{N_i^\dagger, R_i^\dagger, G_i\}$ where,

$$\tilde{c}_{ik} \equiv E(c_{ik}|N_i^\dagger, R_i^\dagger, G_i; \phi^{(j)}),$$

is given by

$$\tilde{c}_{ik} = \frac{\pi_{ik}[N_i^\dagger, R_i^\dagger|c_{ik} = 1, G_i; \phi^{(j)}]}{\sum_{j=1}^{g} \pi_{ij}[N_i^\dagger, R_i^\dagger|c_{ij} = 1, G_i; \phi^{(j)}]} \tag{3.14}$$

Once again we note the difference between our compensator and the one proposed by [Han and Gonzalez , 2010] is that our α and β are both class specific. Now we will let $\psi(\cdot)$ be the digamma function and $h_{ik}^{(j)}(u)$ be the evaluation of $h_{ik}(u)$ using the jth parameter value $\Phi^{(j)}$. The conditional expectations in the second and third parts of (3.3) are computed as [Han et al. , 2007],

$$\tilde{\zeta}_i \equiv \sum_{k=1}^{g} \tilde{c}_{ik} \frac{1 + \theta^{(j)} N_i^\dagger(s)}{1 + \theta^{(j)} \int_0^s R_i^\dagger(u) h_{ik}^{(j)}(u)du}, \tag{3.15}$$

$$\widetilde{c_{ik}\zeta_i} \equiv \frac{c_{ik}[1 + \theta^{(j)} N_i^\dagger(s)]}{1 + \theta \int_0^s R_i^\dagger(u) h_{ik}^{(j)}(u)du}, \tag{3.16}$$

$$\widetilde{\log \zeta_i} \equiv \sum_{k=1}^{g} \tilde{c}_{ik}[\psi(\theta^{(j)^{(-1)}} + N_i^\dagger(s)) - \log(\theta^{(j)^{(-1)}}) + \int_0^s R_i^\dagger(u) h_{ik}^{(j)}(u)du)], \tag{3.17}$$

The evaluation of the compensator $A_i^\dagger(s|c_{ik} = 1, \zeta_i = 1, x_i)$ is involved in each of the above four conditional expectations. Let $t_{(1)} < t_{(2)}, \cdots < t_{(d)}$ be the d distinct jump times of $\Lambda^{(j)}(s, \cdot)$ and $\lambda^{(j)}(s, t_{(l)}) = \Lambda^{(j)}(s, t_{(l)}) - \Lambda^{(j)}(s, t_{(l)}-)$, we can write this compensator as

$$\int_0^s R_i^\dagger(u) h_{ik}^{(j)}(u)du = \sum_{l=1}^{d} R_i^{(k)}(s, t_{(l)}|\alpha_k^{(j)}, \beta_k^{(j)}) \lambda^{(j)}(s, t_{(l)}). \tag{3.18}$$

Now we can substitute the above conditional expectation into (3.3) and obtain the complete data log-likelihood by

$$Q(\phi|\phi^{(j)}) = E_{\phi^{(j)}}\{l_c(\phi)|N_i^{\dagger}, R_i^{\dagger}, G_i\}.$$

3.4 The Maximization Step of the EM algorithm

The objective function relevant to γ in $Q(\phi|\phi^{(j)})$ and its gradient are

$$\tilde{b}(\gamma) = \sum_{i=1}^{n}\sum_{k=1}^{g} \tilde{c}_{ik} w_i^T \gamma_k - \sum_{i=1}^{n} \log\{\sum_{j=1}^{g} \exp(w_i^T \gamma_j)\},$$

$$\frac{\partial \tilde{b}(\gamma)}{\partial \gamma_k} = \sum_{i=1}^{n} w_i(\tilde{c}_{ik} - \pi_{ik}) \tag{3.19}$$

In the maximization step, we will use the Newton-Raphson procedure to obtain the $(j+1)$th update $\gamma_k^{(j+1)}$ for coefficient γ_k, which is given by

$$\gamma_k^{(j+1)} = \gamma_k^{(j)} + [\sum_{i=1}^{n} \pi_{ik}(1-\pi_{ik}) w_i w_i^T]^{-1} [\sum_{i=1}^{n} w_i(\tilde{c}_{ik} - \pi_{ik})] \tag{3.20}$$

By maximizing the following objective function we can achieve the updated estimate for θ^{-1} [Han et al. , 2007].

$$\tilde{b}(\theta) = -n \log \Gamma(\theta^{-1}) + n\theta^{-1} + (\theta^{-1} - 1) \sum_{i=1}^{n} \widetilde{\log \zeta_i} - \theta^{-1} \sum_{i=1}^{n} \tilde{\zeta}_i. \tag{3.21}$$

The derivative of $\tilde{b}(\theta)$ with respect to θ^{-1} is given by

$$\frac{\partial \tilde{b}(\theta)}{\partial \theta^{-1}} = n + \sum_{i=1}^{n} (\widetilde{\log \zeta_i} - \tilde{\zeta}_i) - n\psi(\theta^{-1}) + n \log(\theta^{-1}). \tag{3.22}$$

where $\psi(\cdot)$ is the digamma function. The derivative of $\tilde{b}(\theta)$ with respect to θ is given by

$$\frac{\partial \tilde{b}(\theta)}{\partial \theta} = -\theta^{-2}[n + \sum_{i=1}^{n} (\widetilde{\log \zeta_i} - \tilde{\zeta}_i) - n\psi(\theta^{-1}) + n \log(\theta^{-1})]. \tag{3.23}$$

In order to perform the optimization we can do it either in terms of θ or θ^{-1}, whichever is convenient. Given the current estimate of $\{\alpha_k\}$ and $\{\beta_k\}$ the updated estimate of baseline cumulative hazard is given by

$$\Lambda^{(j+1)}(s,t) = \int_0^t \left\{ \frac{\tilde{J}(s,u|\alpha^{(j)},\beta^{(j)})}{\tilde{R}(s,u|\alpha^{(j)},\beta^{(j)})} \right\} \left\{ \sum_{i=1}^n N_i(s,du) \right\} \tag{3.24}$$

where $\tilde{R}(s,u|\alpha,\beta) = \sum_{i=1}^n \sum_{k=1}^g \widetilde{c_{ik}\zeta_i} R_i^{(k)}(s,u)$, and $\tilde{J}(s,u|\alpha,\beta) = I\{\tilde{R}(s,u|\alpha,\beta) > 0\}$. In accordance with Han and Gonzalez [2010], the third portion in the right hand side of (3.6) can be represented by

$$\sum_{i=1}^n \sum_{k=1}^g c_{ik}\zeta_i \int_0^s R_i^\dagger(v) h_{ik}(v) dv = \sum_{k=1}^g \int_0^\infty R_{(k)}(s,u)\Lambda(s,du) \tag{3.25}$$

where Proposition 1 is used. The above expression becomes $\sum_{i=1}^n \sum_{k=1}^g c_{ik} N_i(s,\infty)$ if substituting $\tilde{\Lambda}(s,du)$ in (3.11) for $\Lambda(s,du)$, which is independent of (α_k,β_k). Because of this, the conditional expectation of this term will not contribute to the profile expected complete data log-likelihood for (α,β).

We obtain the relevant portion of the profile expected complete data log-likelihood in integral form by inserting $\Lambda^{(j+1)}(s,t)$ for $\Lambda(t)$ in the second term of (3.6), which takes the form of

$$\tilde{b}_p(\alpha,\beta|s) = \sum_{i=1}^n \sum_{k=1}^g \tilde{c}_{ik} \int_0^s [\log \rho(N_i^\dagger(v-);\alpha_k) + \log \phi(\beta_k^T x_i(v)) - \log \tilde{R}(s,\mathcal{E}_i(v)|\alpha,\beta)] N_i^\dagger(dv). \tag{3.26}$$

The first derivatives of the profile expected log-likelihood function with respect to (α,β) are given by

$$\frac{\partial \tilde{b}}{\partial \alpha_k} = \sum_{i=1}^n \tilde{c}_{ik} \int_0^s \left\{ \frac{\frac{\partial}{\partial \alpha_k}\rho[N_i^\dagger(v-);\alpha_k]}{\rho[N_i^\dagger(v-);\alpha_k]} - \sum_{k=1}^g \frac{\frac{\partial}{\partial \alpha_k}\tilde{R}(s,\mathcal{E}_i(v)|\alpha_k,\beta)}{\tilde{R}(s,\mathcal{E}_i(v)|\alpha_k,\beta)} \right\} N_i^\dagger(dv). \tag{3.27}$$

$$\frac{\partial \tilde{b}}{\partial \beta_k} = \sum_{i=1}^n \tilde{c}_{ik} \int_0^s \left\{ \frac{\frac{\partial}{\partial \beta_k}\phi(\beta_k^T x_i(v))}{\phi(\beta_k^T x_i(v))} - \sum_{k=1}^g \frac{\frac{\partial}{\partial \beta_k}\tilde{R}(s,\mathcal{E}_i(v)|\alpha,\beta_k)}{\tilde{R}(s,\mathcal{E}_i(v)|\alpha,\beta_k)} \right\} N_i^\dagger(dv). \tag{3.28}$$

Numerical iterative procedure can be used to estimate α_k, β_k.

3.5 Selection of Number of Classes

In order to fit a finite mixture model we have to choose the number of classes, g, such that g is inferred from the data. Previous studies have shown that when using the penalized log-likelihood, ICL-BIC [Biernacki et al. , 2000], and BIC [Schwarz , 1978] perform the best when approximating the integrated classification likelihood criterion. If Φ represents the true parameters, the magnitude of penalty can be illustrated as such:

$$BIC \leq ICL - BIC$$

where

$$BIC = -2\log L(\hat{\Phi}) + d\log(n) \tag{3.29}$$

$$ICL - BIC = -2\log L(\hat{\Phi}) + 2EN(\hat{z})n + d\log(n) \tag{3.30}$$

where

$d =$ number of known parameters

$g =$ number of latent groups

$n =$ sample size

$$EN(\hat{z}) = \sum_{i=1}^{g}\sum_{j=1}^{n} \hat{z}_{ij}\log\hat{z}_{ij}$$

$z_{ij} = z_i(N_j^\dagger, R_j^\dagger; \Phi) =$ posterior probability of ith component membership of subject j

$z_i(N_j^\dagger, R_j^\dagger) = P[z_{ij} = 1|N_j^\dagger, R_j^\dagger]$

3.6 Estimation Scheme

1. Select the largest possible number of latent groups, defined as M.

2. Iterate from $g = 1, \cdots, M$

   a) do a cluster analysis of n subjects for a given g.

   b) choose some trivial starting values for EM iterations based on the classification in step 2a

   c) compute the conditional expectation in section (3.3)

   d) perform the maximization step in section (3.3)

   e) alternate step (2c) and (2d) until convergence is reached and compute the ICL-BLC for this given g.

3. Choose the model of g components, which minimized the ICL-BLC criterion.

Chapter 4

Generation of Data

4.1 Generation Mechanisms

In order to produce the data we will generate, class membership, recurrent events, and censor time. Our class membership which generated the covariate vector v and the latent class membership for n subjects was set at a default of 3 particular classes. The parameters ϕ of our latent class model were chosen to maintain proportions of about one third for each class. The class membership model generated the shared covariates, measurement and treatment. The measurement covariates are shared by both the class model and the event model and was created using $N(0,1)$. The the option of two treatments was created using a Bernoulli distribution. The probability of being assigned to any 1 of the 3 latent groups was done using a multinomial logit model.

For our censoring times, data for the censoring variables for n subjects were generated by uniform distribution. We used a lower bound of 0 and created an upper bound b and then took the uniform distribution in order to generate this data. The bounds were also created to have at most an average of 10 recurrent events for all latent class subjects.

To Generate the recurrent events we used a Weibull distribution for the baseline hazard. Within this function we were able to generate the data for the number of event occurrences, the number of observations, gap times, frailty variable, calendar times, event indicators, perfect repair indicators and the effective age. In order to create the calendar times for each subject, accumulate intercepts, slopes and perfect repair indicators a while loop was created. The generation of perfect repair or not is determined by a bernoulli distribution. To create the frailty variable we used a mean of one and a shape parameter α. Gap time generation was based on a Weibull baseline hazard.

4.2 Data Structure

In generating the data for one subject the 'share' and 'V' variables gives us the 'measurement' and 'treatment' outcomes. The latent probability of this subject being in one of our three classes was greatest for the third class thus they were assigned to 'category' 3. Their censor time was at $t = 3.5174$ in which we observed 7 events (not inclusive of the origin and the censor time). We are also able to see the calendar times at which each event took place and the amount of time between each of the event occurrences (gap times) for this subject. Based on the perfect repair indicators we find that this subject had five perfect repairs and 2 minimal repairs. By looking at his effective age we can determine the effective age at the time of each event which varied as increasing and decreasing over the study period. Lastly we are able to see the accumulated number of events over study period for this subject.

```
DATA OUTPUT FOR SUBJECT [[1]]

[[1]]

[[1]]$share

[1] 0.9569752 0.0000000

[[1]]$V

[1] 1.0000000 0.9569752 0.0000000

[[1]]$latentprob

[1] 0.1971047 0.3112854 0.4916099

[[1]]$latentmem

[1] 0 0 1

[[1]]$category

[1] 3

[[1]]$censortime

[1] 3.517481

[[1]]$omega

[1] 0.7108873

[[1]]$nevent

[1] 7

[[1]]$k

[1] 8

[[1]]$nobs

[1] 9

[[1]]$tau
```

```
[1] 3.517481

[[1]]$censored

[1] 0.0001184048

[[1]]$caltimes

[1] 0.0000000 0.2608825 1.2579270 1.6995033 2.2121248 2.5122645 3.1755695

[8] 3.5173631 3.5174815

[[1]]$gaptimes

[1] 0.0000000000 0.2608824714 0.9970444986 0.4415762980 0.5126215806

[6] 0.3001396524 0.6633050380 0.3417935409 0.0001184048

[[1]]$eventind

[1] 2 1 1 1 1 1 1 1 0

[[1]]$perrepind

[1] 1 1 0 1 1 1 1 0 0

[[1]]$lastperrep

[1] 1 2 2 4 5 6 7 7 7

[[1]]$effaccum

[1] 0 0 1 0 0 0 0 1 1

[[1]]$accum

[1] 0 1 2 3 4 5 6 7 7

[[1]]$effagebegin

[1] 0.0000000 0.0000000 0.9970445 0.0000000 0.0000000 0.0000000 0.0000000

[8] 0.3417935 0.3419119

[[1]]$effage
```

```
[1] 0.0000000 0.2608825 0.9970445 1.4386208 0.5126216 0.3001397 0.6633050

[8] 0.3417935 0.3419119

[[1]]$intercepts

[1] 0 0 0 0 0 0 0 0 0

[[1]]$slopes

[1] 1 1 1 1 1 1 1 1 1

[1]]$covariate

          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]

[1,] 0.9569752 0.9569752 0.9569752 0.9569752 0.9569752 0.9569752 0.9569752

[2,] 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000

          [,8]      [,9]

[1,] 0.9569752 0.9569752

[2,] 0.0000000 0.0000000
```

The following figure is a representation of the output for 10 generated subjects against the calender time. We see in this pictorial representation the censor time for each subject and the type of repair performed at each intervention, perfect repair or minimal repair.

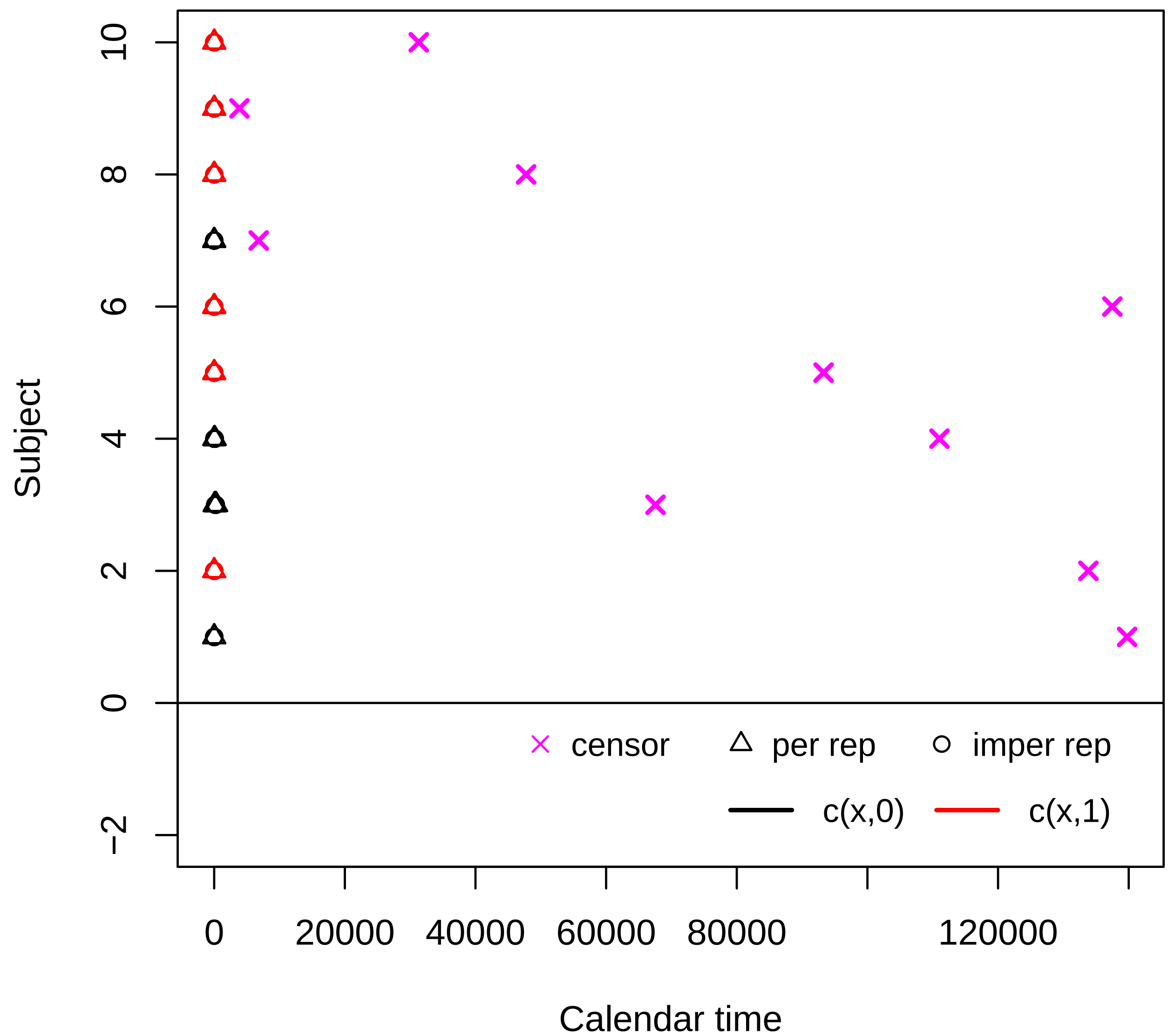

Figure 4.1: Recurrent event times for 10 subjects

4.3 Explanation of Variables

share: measurement and treatment of a subject

measurement: and treatment of a subject inclusive of the intercept, the intercept term accommodates the control group effect.

latentprob: creates the prob of being in each of 3 classes

latentmem: the class that each subject belongs to

category: the class that the subject belongs to (similar to latentmem)

cencortime: the original cencoring calendar time

omega frailty random deviate

numofevents: the number of recurrent events that took place

k: the number of events including origin not including censor time.

nobs:number of observations including origin and tau(not repeated)

tau:type II adjusted censor time, tau may be smaller than original censortime if max number of events is reached

censored: censored gap time

caltimes: include the origin, true event calendar times, and censored calendar time if a censored gap time occurs. The length of the caltimes is k for the first case and k+1 for the second case.

gaptimes: time between each event

eventind: origin=2, event=1, censored=0

perrepind: Indicates whether there was a perfect repair or minimal repair at each event occurrence 0=minimal repair 1=perfect repair

lastperrep: The position of the last perfect repair with respect to the current piece of effective age

effaccum: effective accumulation of events up to time s

accum: accumulation of events up to time s

effagebegin: initially 0

effage: the effective age at time s

intercepts: initially intercept =0

slopes: initially slope =1

covariate: time independent covariates are generated

Chapter 5

Conclusion & Future Work

In this paper an estimation algorithm for mixture of experts recurrent event model was proposed and developed. This model is general and flexible in that it not only takes into account various features such as the effect of accumulating event occurrences, the effect of interventions after each event occurrence,and the effect of possible unobservable frailties, it also considers the effect of class specific covariate effects in a heterogeneous population. Data generation was performed in order to see the intervened recurrent event structure for a heterogeneous population. However we leave simulation studies, application to real data, and model testing for a future paper.

Bibliography

Biernacki, C., Celeux G., Govaert G., 2000. Assessing a mixture model for clustering with the integrated completed likelihood, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(7), 719-725.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum Likelihood from Incomplete Data via EM algorithm, Journal of Royal Statistical Society, 39, 1-38.

Han, J., Slate, E., Peña, E., 2007. Parametric Latent Class Joint Model for Longitudinal Biomarker and Recurrent Events, Statistics in Medicine, 26, 5285-5302.

Han, J., González, J., 2010. Dynamic recurrent event model for heterogeneous population, Unpublished Manuscript.

Han, J., 2005. Parametric Latent Class Joint Model for Longitudinal Markers and Recurrent Events, PhD Dissertation, University of South Carolina.

Jacobs R. A., Jordan M. I., Nowlan S. J., and Hinton G. E., 1991. Adaptive mixtures of local experts, Neural Computation, 3, 79-87.

Peña E.A., Hollander M., 2004. Models for recurrent events in reliability and survival analysis.

In Mathematical Reliability: An Expository Perspective, Soyer R, Mazzuchi T, Singpurwalla N (eds). Kluwer Academic Publishers: Dordrecht, 105-123.

Peña, E.A., Strawderman, M., Hollander, M., 2001. Nonparamtric Estimation with Recurrent Event Data, Journal of American Statistics, 96(456), 1299-1315.

Peña, E.A, Slate, E., Gonzalez, J., 2007. Semiparametric inference for a General Class of Models for recurrent Events, Journal of Statistical Planning and inference, 137, 1727-1747.

Rosen, O., Tanner, M., 1999. Mixtrues of Proportional Hazards Regression Models, Statistics In Medicine, 18, 1119-1131.

Schwarz G., 1978. Estimating the dimension of a model, Annals of Statistics, 461-464.