Mixture Poisson Point Process:

Assessing Heterogeneity in EMA Analysis

by

Nat Kulvanich

(Under the Direction of Stephen L. Rathbun)

Abstract

The times of repeated behavioral events can be viewed as a realization of a temporal point process. Rathbun, Shiffman, and Gwaltney (2007) used a Poisson process (Cox 1972) for modeling repeated behavioral events impacted by time-varying covariates. Taking an inspiration from the techniques of generalized linear mixed models, and the EM algorithm (Dempster et al. 1977) for finite mixture model estimation, we will further extend models to handle data arising from a heterogeneous population. In Chapter 2, we present a finite mixture model for Poisson point processes, classifying subjects into clusters sharing identical responses to time-varying covariates within clusters. In Chapter 3, a mixture mixed-effect model is presented which accommodates variation among subjects within clusters with respect to their responses to the time-varying covariates. In Chapter 4, we discuss some issues we encountered in the research and point out potential topics for future research. All the approaches in this dissertation are illustrated using data from an ecological momentary assessment of smoking.

Index words: Poisson Process; Mixture models; Point process; Ecological Momentary Assessment; smoking; Generalized Linear Mixed Models; hierarchical likelihood; Maximum likelihood estimation.

Mixture Poisson Point Process:
Assessing Heterogeneity in EMA Analysis

by

Nat Kulvanich

B.S., Chulalongkorn University, 2002

M.S., University of Michigan, 2006

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2013

Mixture Poisson Point Process:
Assessing Heterogeneity in EMA Analysis

by

Nat Kulvanich

Approved:

Major Professor: Stephen L. Rathbun

Committee: Daniel Hall
Lynne Seymour
T.N. Sriram
Ye Shen

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2013

Dedication

To my family, and my beloved Anny.

Acknowledgments

First of all, I wish to express my appreciation to my major professor, Dr. Stephen L. Rathbun, for his valuable support in the direction of this dissertation. He gave me an interesting research topic and advised me at any time I seek. He is such a wonderful advisor: not only knowledgeable, but also truly care for his student. I am truly grateful for his encouragement and for the amount of time he spent helping me. It has been a pleasure to work under him.

Furthermore, I would like to thank to my committee members, Dr. Daniel Hall, Dr. Lynne Seymour, Dr. T.N. Sriram, and Dr. Ye Shen for their careful reading and suggestions. I also would like to show my appreciation to the faculty, staff and graduate students of Department of Statistics. They gave me a chance to pursue higher degree, and gave me the best place to learn.

I also wish to thank my Thai friends here in Athens (P' Ple, N' Air, N'Aun, N' Nui, N'Tum, and N' Ree) for making these 4 years a great experience for me.

Last but not least, I would like to thank my family in Thailand, especially my parents, Manas and Amornrat, for mentally support me from far away and for having faith in me.

Table of Contents

List of Tables

Chapter 1

Introduction

1.1 Introduction

The availability of portable electronic devices within the past 20 years has promoted the development of Ecological Momentary Assessment (EMA) (Shiffman & Stone 1998; Stone & Shiffman 2002), a collection of research methods in the behavioral sciences focused upon obtaining information about subjects' current psychological state in their everyday environments. EMA is particularly well suited for investigating recurrent addictive behavioral events such as alcohol consumption (Collins et al. 2007), eating disorders (Smyth et al. 2001), the use of tobacco (Shiffman 2005), and stress coping (Stone et al. 1998).

The collection of recurrent event data is often accomplished through the use of electronic devices, such as handheld personal digital assistants (PDAs), that are provided to participants. Recently, Collins et al. (2003) suggested a new system for collecting EMA data using cellular telephones linked to an interactive voice response (IVR) system. By allowing subjects to record event data as it occurs naturally in their environments, this method of data collection reduces some of the biases associated with retrospective recall and increases the "ecological validity" of the data by asking subjects questions regarding current state and by not placing subjects in an artificial laboratory environment (Shiffman and Stone 1998). Furthermore, by sampling subjects' psychological states using a probability-based scheme, unbiased estimates of the means of these states can be computed. If subjects are also assessed at event times, this method also allows one to estimate the effects of the covariates' mean states upon the timing of these events (Rathbun et al. 2007).

In an EMA cigarette smoking cessation, for example, Shiffman et al. (2002) asked subjects to record each cigarette on a Personal Digital Assistant (PDA) both before and after a designated quit time. The PDA prompted the smokers to answer questions regarding their current mood and environment for randomly selected cigarettes and at times selected according to a probability-based sampling design. Both subject-level covariates, such as gender and nicotine dependence, and time-varying covariates, such as restlessness, were assessed during this study. Analyzing these data will provide insight into the impact of mood and environment on cigarette consumption rates. This understanding may lead to improvements in intervention programs for smoking cessation.

It's been recommended that Hierarchical Linear Models (HLMs) be used to analyze EMA data (Bolger et al. 2003; Schwartz and Stone 2007). This technique can capture the variability among subjects in their responses to time-varying and time-invariant covariate values. However, HLMs are not designed to model the timing of repeated discrete events over an interval.

The times of repeated behavioral events can be viewed as a realization of a temporal point process (Rathbun et al. 2006). Rathbun, Shiffman, and Gwaltney (2007) proposed models for the analysis of EMA smoking data using a Poisson process, and found that smoking rate increased with the increasing restlessness of the smoker. Their model, however, has the limitation that it implicitly assumes that all subjects have equal responses to changes in time-varying covariates.

Results from Shiffman and Rathbun (2011) suggest that there are interactions between gender and time-varying covariates, suggesting that gender-specific treatments may be more effective for smoking cessation. However, gender is not the only factor that might affect subjects' responses to time-varying covariates. More effective treatments for smoking cessation might be obtained if subjects can be classified into groups with like responses to time-varying covariates. Thus, taking inspiration from the implementation of mixture models for the impact of time-varying covariates on recurrent events, we will further extend Rathbun et al. (2007) to describe variation among subjects with respect to the effects of time-varying covariates from which clusters of subjects identified based on shared response to time-varying covariates in an EMA cigarette cessation study. Therefore, treatments may be further customized to each smoker based upon their responses to time-varying covariates and their cluster of smoking patterns. This dissertation will propose two different mixture models that expand upon previous research for the use of Poisson process to analyze repeated event data. The proposed estimators for model parameters will be shown to have the desirable asymptotic properties.

1.2 Motivating Dataset

The motivating dataset for this research comes from Shiffman et al.'s (2002) multi-phase EMA study on smoking habits. This EMA study consisted of a sample of 304 smokers, each of whom was given an electronic diary and was instructed to record on the diary any time that they smoked a cigarette (hereafter referred to as an "event"). Subjects recorded the exact times that they smoked and answered questions regarding their mood and environment (e.g. whether other smokers were present) at randomly selected smoking times and other randomly selected times during the day. According to a Bernoulli process with known probability, the subjects were prompted to answer a series of assessment questions at randomly selected smoking events.

Subjects were also prompted to complete assessments at random times throughout the day. These random assessments at non-event times were chosen according to a stratified (by subject and day) sampling design, under the restriction that a non-event assessment could not occur within the 10 minutes following an event, nor when the diary was off or in "stand-by" mode. An expected 5 random assessments were prompted each day over the time during which the diary was on and available for assessing. Subjects were allowed to turn off the diaries or put them on stand-by when subjects were in a meeting, sleeping, or driving. Among the variables recorded during assessment were Negative Affect (composite of subjects' responses to negative adjectives such as "miserable" and "irritated"), Arousal (reactions to words such as "energetic"), Attention (based on subjects' responses to questions regarding their ease or difficulty concentrating), Restlessness (based on a single item that did not load into other factors), and whether others smokers were presented.

This is the same dataset that was analyzed by Rathbun et al. (2007). They found that only restlessness has a significant impact on the timing of smoking events

1.3 Modulated Poison Process

1.3.1 Poisson Process Intensity

Point processes are frequently used to model the occurrence of events in time, space, or space-time. They have been used to model such events as the timing and location of rainfall (Northrop 1998), neural activity (Johnson 1996), or the location and diameter of tree (Stoyan & Penttinen 2000). The times of repeated behavioral events may be viewed as a realiztion of a temporal point process (Rathbun, Shiffman, & Gwaltney 2006). Rathbun et al. (2007) proposed models for the analysis of EMA longitudinal repeated-event data using a Poisson process.

A point process can be partially described by its intensity, which describes the expected rate of events over time for a specific instant. The intensity $\lambda(t)$ is defined as

$$\lambda(t) = \lim_{\delta \to \infty} \frac{E\left(N\left[t, t+\delta\right]\right)}{\delta},$$

where $N\left[t, t+\delta\right]$ is the number of events in the interval $[t, t+\delta]$. When the point process is a Poisson process (Cox 1972), $N(A)$ is Poisson distributed with mean $\Lambda(A) = \int_A \lambda(t)dt$ for all Borel sets A. Moreover, conditional on $N(A)$, the event locations are independently sampled from the probability density function proportional to $\lambda(t)$.

A modulated Poisson process (Cox 1972), is a special case of a Poisson process where the intensity function takes the form

$$\lambda(t; \beta) = \exp\left\{\beta^T x(t)\right\}, \quad t \in A. \tag{1.1}$$

Here $\lambda(t;\beta)$ is a function of the vector of the time-varying covariates, $x(t)$ is a vector of time-varying covariates (e.g. ratings of mood), and β is a vector of coefficient parameter describing the effects of these covariates on the intensity of the process, and so on the frequency of events. This particular case of a Poisson process is well suited for modeling the influences of the covariates on a spatial point pattern (Rathbun et al. 2007).

Let $\{t_i : i = 1, .., N\}$ denote the times of N events in the study domain A. For the Poisson process with intensity (1.1), the log-likelihood of the parameter vector β over the study interval A is

$$\ell_A(\beta) = \beta^T \sum_{i=1}^{N} x(t_i) - \Lambda(A;\beta), \quad t_i \in A,$$

where $x(t)$ is the covariate vector, and $\Lambda(A;\beta) = \int_A \exp\left\{\beta^T x(t)\right\} dt$ is the integrated intensity. Then the score equation for the parameter β is

$$\Psi_A(\beta) = \sum_{i=1}^{N} x(t_i) - \Lambda^{(1)}(A;\beta), \tag{1.2}$$

where $\Lambda^{(1)}(A;\beta) = (\partial/\partial\beta)\,\Lambda(A;\beta) = \int_A x(t)\exp\left\{\beta^T x(t)\right\} dt$.

Rathbun & Cressie (1994) demonstrated that the maximum likelihood estimator $\hat{\beta}$, which satisfies the score equations $\Psi_A(\beta) = 0$, is consistent, and asymptotically Gaussian under increasing domain asymtotics with asymptotic variance-covariance matrix

$$var(\hat{\beta}) = \left[\int_A x(t)x(t)^T \exp\left\{\beta^T x(t)\right\} dt\right]^{-1}.$$

1.3.2 Integral Estimation

To obtain the solution to (1.2), maximum likelihood estimation requires that the covariates, $x(t)$, must be known functions of time for the entire interval A. In many cases, constant observation of the covariates is unrealistic or impossible to obtain. Therefore, methods of estimating $\Lambda^{(1)}(A;\beta)$ are necessary. Wulfsohn and Tsiatis (1997) introduce a joint modeling approach for lifetimes of events and time-varying covariates. Zhang et al. (2008) and Liu and

Huang (2009) extend this approach to recurrent events data. The estimators for covariate effects are consistent under correctly specified models, but they may be biased if the covariate model is misspecified.

If we assumes that random assessments were taken according to a known probability sampling design, we can obtain design unbiased estimators of these integrals. Rathbun et al. (2007) and Waagepetersen (2008) propose approaches based on sampling the covariates at random times $\{u_j : j = 1, .., m\}$ in the study interval A with inclusion densities $\pi(u)$. Then an unbiased estimator for $\Lambda^{(1)}(A; \beta)$ may be substituted into the log-likelihood, yielding an estimated score function of the form

$$\hat{\Psi}_A(\beta) = \sum_{i=1}^{N} x(t_i) - \hat{\Lambda}^{(1)}(A; \beta),$$

where $\hat{\Lambda}^{(1)}(A; \beta)$ is a design-unbiased estimator of $\Lambda^{(1)}(A; \beta)$. Rathbun et al. (2007) proposed

$$\hat{\Lambda}^{(1)}(A; \beta) = \sum_{j=1}^{m} \frac{x(u_j) \exp\left\{\beta^T x(u_j)\right\}}{\pi(u_j)}.$$

Waagepetersen (2008) suggested another unbiased integral estimator that uses all of the available covariate information:

$$\hat{\Lambda}^{(1)}(A; \beta) = \sum_{i=1}^{N} \frac{x(t_i) \exp\left\{\beta^T x(t_i)\right\}}{\pi(t_i) + \exp\left\{\beta^T x(t_i)\right\}} + \sum_{j=1}^{m} \frac{x(u_j) \exp\left\{\beta^T x(u_j)\right\}}{\pi(u_j) + \exp\left\{\beta^T x(u_j)\right\}}.$$

The authors each show that the estimators obtained from solving these equations are consistent and have asymptotically normal estimates of β, under increasing domain and infill asymptotics.

1.3.3 Adding Mixture components

In the context of our motivating dataset, Shiffman and Rathbun (2011) suggest that there are interactions between gender and time-varying covariates, which suggests that gender-specific treatments may be more effective for smoking cessation. Therefore, more effective treatments for smoking cessation might be obtained if subjects can be classified into groups with like responses to time-varying covariates. This leads to an application of the mixture model where smokers classified into groups or clusters. Thus, the modulated Poisson process (Cox 1972) can be further extended to describe variation among subjects with respect to the effects of time-varying covariates, and from which clusters of subjects showing similar characteristics may be discovered.

In our study of the application of mixture models to an EMA of cigarette smoking cessation, the number of clusters is assumed to be fixed. Suppose that the smokers can be classified into one of g groups or categories with like responses to time-varying covariates, and let p_k denote the probability that a subject belongs to group k. When group membership is unknown, contribution of a subject to the log-likelihood of the parameter vector $\theta = \left(p_1, .., p_{g-1}, \beta_1^T, .., \beta_g^T\right)$ over the study interval A is

$$\sum_{k=1}^{g} p_k \exp\left\{\sum_{i=1}^{N_i} x(t_i) - \int_{A_i} \Lambda(A; \beta_k) dt\right\},$$

where $p_g = 1 - \sum_{k=1}^{g-1} p_k$.

The difficulty in computing parameter estimators in mixture models is well known. However, with the application of EM algorithm (Dempster et al. 1977), this mixture Poisson process model can be fitted by iteratively fitting weighted versions of the component models. Therefore, to facilitate computation, the EM algorithm will be employed for solving parameter estimation problems in this dissertation.

1.3.4 Adding Inter-Subject variability

Rathbun et al.'s approach (2007) assumes that all subjects have identical event rates. However, such an assumption may be unrealistic in many applications, particularly when dealing with behavioral data. In the context of the motivating dataset, some subjects may have inherent genetic factors that lead to higher or lower smoking rates. If we can identify what factors impact individual smoking behaviors, those specific factors may be used to find more effective individual-based smoking cessation treatment. We can develop models of evaluating repeated event data that allow for inter-subject variability by adding random effect terms. This mixed-effects version of modulated Poisson process allows the covariate coefficients to vary randomly among subjects.

Suppose that n subjects are independently sampled, and that subject i is observed at over a set of N_i recurrent event times t_{ij} where $j = 1, .., N_i$. Conditional on $q \times 1$ vector of random effects β_i, the intensity for each subject i takes the form

$$\lambda_i(t) = \exp\left\{\alpha^T x_i(t) + \beta_i^T z_i(t)\right\}.$$

Here $x_i(t)$ and $z_i(t)$ are vectors of fixed and random effects, respectively, α is a vector of fixed coefficients, and β_i is sampled from some probability distribution. In this dissertation, we assume that β_i is sampled from a mixture of normal distributions.

There is little literature that is directly related to Poisson processes with random effects. Lawless (1987) and Thall (1988) proposed Poisson point process models with gamma frailties to account for variation among subjects. However, time-varying covariates were not considered in their models. Most of the previous research focused on an extension of Generalized Linear Models (GLMs) (Nelder & Wedderburn 1972). Hierarchical Generalized Linear models (HGLMs) (Lee & Nelder 1996), which are a further extension of GLMs, may be considered in this case. They model the responses y_i by a GLM with some randomly distributed components. These random components may be assumed to have a variety of distributions, though the distribution conjugate to the response distribution is often selected; thus in the case that the response are Poisson distributed, the random effects are often gamma distributed. Generalized Linear Mixed Models (GLMMS) are a particular type of HGLM, where the random components are assumed to follow a normal distribution. Typically, GLMMs can be evaluated using computationally intensive approaches, such as Monte Carlo Expectation-Maximization (MCEM), or Penalized Quasi Likelihood (PQL) (Breslow & Clayton 1993), since the log-likelihood and the score equations have no closed form.

Lee and Nelder (1996) also suggest maximum hierarchical likelihood (or h-likelihood) as an alternative form of parameter estimation when the random effect is a function of the canonical parameter. However, when it is not, this approach is known as Extended Likelihood estimation, in which the estimators are well known to be biased downward (Lee & Nelder 1996).

1.4 Proposed Models

In the following chapters of this dissertation, two different methods of mixture point process modeling will be proposed. Chapter 2 suggests the implementation of finite mixture models for point processes to describe heterogeneity among subjects. Assuming that group membership is unknown, the contribution to the likelihood for the ith subject is

$$\sum_{k=1}^{g} p_k \exp\left\{ \sum_{l=1}^{N_i} x_i(t_l) - \int_{A_i} \exp\left\{\beta_k^T x_i(t)\right\} dt \right\},$$

where A_i is the ith subject's study interval, and β_k is the vector of the parameters describing the effects of the covariates on the rate of random events for the kth cluster or group. It will also been shown in Chapter 2 that the proposed estimators are consistent and asymptotically normal.

In Chapter 3, a more flexible model, that allows for inter-subject variability and handles heterogeneous populations, will be proposed. We will present a more general class of models consisting of finite mixtures of the mixed effects model for the impact of time-varying covariates on recurrent events data. In this case, the mixture mixed-effect intensity is defined as

$$\lambda_i(t) = \exp\left\{\alpha^T x_i(t) + \beta_i^T z_i(t)\right\}, \quad t \in A_i,$$

where $x_i(t)$ and $z_i(t)$ are fixed and random effects, respectively, α is the parameter vector of coefficients for the fixed effects, and β_i is a vector of subject-specific random effects and is assumed to be sampled from a mixture of g normal distributions with different means and common covariance matrix Σ. That is

$$f(\beta_i; \mu, \Sigma) = \sum_{k=1}^{g} p_k \phi(\beta_i; \mu_k, \Sigma),$$

where $\phi(\beta_i; \mu_k, \Sigma)$ is the normal distribution with mean μ_k, and covariance matrix Σ. This mixture model with random effects poses some challenges to parameter estimation since the marginal likelihood involves an integral that cannot be evaluated in closed form. So, a set of estimating equations that are related to Laplace approximations of the log-likelihood will be derived and the asymptotic properties of both α and β_i will be further examined.

This dissertation extends the current methods of analyzing modulated Poisson process models by using the concept of a mixture model to allow for the heterogeneity among subjects. This research should improve statistical methods for analyzing EMA and other forms of repeated-event data.

Chapter 2

Mixture Poisson Processes

2.1 Introduction

In some epidemiological investigations it is of interest to describe the impact of time-varying covariates on the pattern of recurrent events. One example is the effects of environment and weather on asthma attacks and epileptic seizures (Kalbfleisch and Prentice 1980; Schwartz et al. 1993,1994; Sheppard et al. 1999; Moolgavkar et al. 2000). The random times of these events are customarily modeled as a temporal point process with an inhomogeneous intensity surface (Duan et al. 2009). Wulfsohn and Tsiatis (1997) introduce a joint modeling approach for lifetimes of events and time-varying covariates. Zhang et al. (2008) and Liu and Huang (2009) extend this to recurrent events data. Soyer and Tarimcilar (2008) proposed an alternative model using a Bayesian approach, and presented its extension to describe potential heterogeneity in intensity patterns. These investigations have allowed for random variation in baseline event rates, but with few exceptions (eg. Ha et al. 2001) have not allowed for variation among subjects regarding the impact of time-varying covariates on patterns of recurrent events.

Investigation of the impact of time-varying covariates on recurrent events can be implemented by using Ecological Momentary Assessment (EMA), under which research participants may enter events at the times they occur and report on current symptoms, affect and behavior when prompted to do so on electronic devices at random times or at the times of recurrent events. EMA has frequently been implemented in addictive behavioral research, e.g. alcohol consumption in moderate drinkers (Carney et al. 1998) and dietary temptation in obese women (Carels et al. 2004) . The alcohol dependent participants showed good conformity with EMA recording system of craving and mood states at the start of each drinking episode by using handheld computer. Collins et al. (2003) implemented a new system for collecting EMA data using cellular telephone linked to the interactive voice response (IVR) system.

For an EMA cigarette smoking cessation study, Shiffman et al. (2002) instructed 304 smokers, desiring to quit smoking, to record each cigarette on palm-top computers both before and after a designated quit time. The electronic diary prompted subjects to answer questions regarding their current mood and environment for randomly selected cigarettes, and at times selected according to a probability-based sampling design. Results indicate that there are interactions between gender and time-varying covariates (Shiffman and Rathbun 2011), suggesting that gender-specific treatments may be more effective for smoking cessation. Gender, however, is not the only factor that might affect subjects' responses to time-varying covariates. More effective treatments for smoking cessation might be obtained if subjects can be classified into groups with like responses to time-varying covariates.

We will further extend the Poisson process model to describe variation among subjects with respect to the effects of time-varying covariates, and time of day, from which clusters of subjects showing similar smoking behaviors may be discovered. This leads naturally to an application of the finite mixture model, where subjects are classified into clusters, and then subject-specific data are realized from a modulated Poisson process with cluster specific effects.

Finite mixture models have been widely used since Pearson (1894) proposed a mixture of two Gaussian densities and estimated its parameters by the method of moments. Fisher (1922) introduced a likelihood-based estimation approach to the fitting of the finite mixture model. Since it is common to think of mixture modeling as a missing data problem, maximum likelihood estimation (MLE) with EM algorithm (Dempster et al. 1977) has played a major role in finite mixture model estimation. There are some advantages using this approach:

First, it is more efficient than method of moments used in Pearson (1894) which involved solution of 9th order equations. Second, the EM algorithm is simpler to implement than other direct optimization methods, e.g., Newton methods, or gradient descent (Couvreur 1997). Third, the optimization process of MLE is also faster than that of other estimation methods (Guindon and Gascuel 2003). Fourth, the MLE is asymptotically efficient when the model is correctly specified. Fifth, although the model may be somewhat misspecified, the MLE is still consistent as long as the mixture components come from the linear exponential family (Sundberg 1974, Kaufmann and Fahrmeir 1985). As an alternative approach to maximum likelihood, the finite mixture model parameters can be determined by using posterior sampling in a Bayesian analysis. This is still considered an incomplete data problem. Methods such as reversible jump (Green 1995), and Gibbs sampling (Geman 1984) can be applied here.

Minimum distance estimators have been developed to obtain more robust estimators for mixture models. Beran (1977) developed the minimum Hellinger distance (MHD) estimator, which removes the instability of estimates from perturbations in the data, and still preserve asymptotic efficiency under specified parametric densities. However, it requires the use of nonparametric kernel density estimators. To avoid this computational complication, Scott (1998, 1999, 2001, and 2004) introduced the L2E estimation method based on the integrated squared error criterion. This approach also has the advantage that a key integral can be expressed in closed form (Scott 2001).

Although minimum distance estimators are robust and resist contaminated data better than the MLE at the cost of efficiency (Lu et al. 2003), they cannot readily accommodate longitudinal data with unequal members of repeated observations. The maximum likelihood estimator, on the contrary, is more appropriate to use in longitudinal repeated discrete-event data, in which the variation among individuals should be investigated. The application of a finite mixture model with maximum likelihood estimation via the EM algorithm for longitudinal data can be founded in Dietz (1992) and Dietz and Bohning (1993). For the

EMA of smoking (Shiffman et al. 2002), the current paper proposes a new algorithm for estimating the mixture intensity function in the modulated Poisson process for both fully-observed covariates and partially-observed covariates.

In section 2, the modulated Poisson process will be briefly described, and the mixture of Poisson point processes will be defined. Simple estimating equations for fully observed covariates will be explained. In section 3, parameter estimation using random assessments of covariates will be addressed. Section 4 will give the details of the EM algorithm for mixture Poisson processes. Section 5 will give the conditions under which the estimators are consistent and asymtotically normal. Section 6 will describe criteria for selecting the number of groups in a mixture Poisson process. Section 7 will describe the motivating dataset and demonstrate the use of the proposed estimators on these data. The Appendices contain proofs of consistency and asymptotic normality of the estimators.

2.2 Mixture- Modulated Poisson Process

A temporal point process is a random process whose realizations consist of the times of discrete events scattered in time (Brillinger et al 2002). The point process has a random counting measure N, where $N(A)$ is the number of events in the set $A \subset R$. This measure can be partially described by its intensity,

$$\lambda(t) = \lim_{\delta \to \infty} \frac{E\left(N\left[t, t+\delta\right]\right)}{\delta},$$

where $N\left[t, t+\delta\right]$ is the number of events in the interval $[t, t+\delta]$.

For the Poisson process, $N(A)$ is Poisson distributed with mean $\int_A \lambda(t)dt$ for any interval A. A modulated Poisson process (Cox 1972), is a special case of a Poisson process where the intensity function takes the form

$$\lambda(t;\beta) = \exp\left\{\beta^T x(t)\right\}, \quad t \in A, \tag{2.1}$$

where $\lambda(t;\beta)$ is a function of the vector of the time-varying covariates, $x(t)$, and β is a p-dimensional vector of the unknown parameters describing the effects of the covariates on the rate of random events. This particular case of a Poisson process is well suited for modeling the influences of the covariates on a spatial point pattern (Rathbun et al. 2007).

Let $\{t_i : i = 1, .., N\}$ denote the times of N events in the study domain A, the log-likelihood function over the study domain for the Poisson process (Cressie 1991) with intensity (2.1) can be written as

$$L_A(\beta) = \beta^T \sum_{i=1}^{N} x(t_i) - \Lambda(A;\beta), \quad t_i \in A,$$

where $\Lambda(A;\beta) = \int_A \exp\left\{\beta^T x(t)\right\} dt$, is the integrated intensity. The maximum likelihood estimator satisfies the score equations $\Psi_A(\beta) = 0$, where

$$\Psi_A(\beta) = \sum_{i=1}^{N} x(t_i) - \int_A x(t) \exp\left\{\beta^T x(t)\right\} dt.$$

Rathbun & Cressie (1994) demonstrated that $\hat{\beta}$ is consistent, and asymptotically Gaussian under increasing domain asymptotics with asymptotic variance-covariance matrix

$$var(\hat{\beta}) = \left[\int_A x(t)x(t)^T \exp\left\{\beta^T x(t)\right\} dt\right]^{-1}.$$

For n independent samples, each subject i is observed over a set of times belonging to the set $A_i \subset R$. Let $x_i(t) : t \in A_i$ denote a $p \times 1$ vector of covariates, that are assumed to be known functions of time t for all $t \in A_i$ and all $i = 1, .., n$. Suppose that the subjects can each be classified into one of k groups or categories with like responses to time-varying covariates. Let π_j denote the probability that a subject belongs to group j. Given that individual i is in category j, suppose further that the times of repeated behavioral events for that subject are realized from a Poisson point process with intensity function written as

$$\lambda_i(t;\beta) = \exp\left\{\beta_j^T x_i(t)\right\}. \quad t \in A_i \tag{2.2}$$

When group membership is unknown, then the contribution of subject i to the incomplete-data likelihood is

$$\sum_{j=1}^{k} \pi_j \exp\left\{\sum_{l=1}^{N_i} x_i(t_l) - \int_{A_i} \exp\left\{\beta_j^T x_i(t)\right\} dt\right\},$$

where $\theta = \left(\pi_1, .., \pi_{k-1}, \beta_1^T, .., \beta_k^T\right)$ and $\pi_k = 1 - \sum_{j=1}^{k-1} \pi_j$.

Statistical inference for a mixture model can be regarded as an incomplete data problem, where the classification of subjects into groups is not observed. Define the indicator $z_{ij} = 1$ when subject i comes from the category j and equal to zero if otherwise. Then conditional on z_{ij}, $j = 1, .., k$, the distribution of the data for individual i is

$$\prod_{j=1}^{k} \left(\exp\left\{ \sum_{l=1}^{N_i} x_i(t_l) - \int_{A_i} \exp\left\{\beta_j^T x_i(t)\right\} dt \right\} \right)^{z_{ij}}.$$

The marginal distribution of z_i is $\prod_{j=1}^{k} \pi_j^{z_{ij}}$, where $\pi_j \in (0,1)$ and $\sum_{j=1}^{k} \pi_j = 1$. Then

$$\prod_{j=1}^{k} \pi_j^{z_{ij}} \left(\exp\left\{ \sum_{l=1}^{N_i} x_i(t_l) - \int_{A_i} \exp\left\{\beta_j^T x_i(t)\right\} dt \right\} \right)^{z_{ij}}$$

is the complete-data likelihood.

For simplicity of presentation, the number of categories is initially set to be two. Then the marginal log-likelihood for $\theta = \left(\beta_1^T, \beta_2^T, \pi\right)^T$ can be written as

$$\begin{aligned} L_n(\theta) &= \sum_{i=1}^{n} \ln(\pi \exp\left\{ \sum_{j=1}^{N(A_i)} \beta_1^T x_i(t_{ij}) - \Lambda_i(A_i, \beta_1) \right\} \\ &\quad + (1-\pi) \exp\left\{ \sum_{j=1}^{N(A_i)} \beta_2^T x_i(t_{ij}) - \Lambda_i(A_i, \beta_2) \right\}), \end{aligned}$$

where $\Lambda_i(A_i, \beta) = \int_{A_i} \exp\left\{\beta^T x_i(t)\right\} dt$, is the integrated intensity, and $N(A_i)$ is the number of events in the study interval. Based on this log-likelihood, the score functions $\Psi_n(\theta)$ for θ are

$$\Psi_n(\theta) = n^{-1} \sum_{i=1}^{n} \begin{bmatrix} v_i(A_i, \beta_1) T_i(w_i(A_i, \beta)) \\ v_i(A_i, \beta_2)\left(1 - T_i(w_i(A_i, \beta))\right) \\ \frac{T_i(w_i(A_i,\beta))}{\pi} - \frac{1 - T_i(w_i(A_i,\beta))}{1-\pi} \end{bmatrix}, \tag{2.3}$$

where

$$\Lambda_i^{(1)}(A_i, \beta) = \frac{\partial \Lambda_i(A_i, \beta)}{\partial \beta^T} = \int_{A_i} x_i(t) \exp\left\{\beta^T x_i(t)\right\} dt,$$

$$v_i(A_i, \beta) = \sum_{j=1}^{N(A_i)} x_i(t_{ij}) - \Lambda_i^{(1)}(A_i, \beta),$$

$$T(w) = \frac{\pi}{\pi + (1-\pi)\exp(w)},$$

and

$$w_i\left(A_i, \beta\right) = \left(\sum_{j=1}^{N(A_i)} \beta_2^T x_i(t_{ij}) - \Lambda_i(A_i, \beta_2)\right) - \left(\sum_{j=1}^{N(A_i)} \beta_1^T x_i(t_{ij}) - \Lambda_i(A_i, \beta_1)\right).$$

Here $T_i(w_i\left(A_i, \beta\right))$ is the conditional expectation of z_i given the data and the model parameter θ, where $z_i = 1$ if the subject belongs to group 1 and equal to zero if otherwise.

These score functions are unbiased in the sense that the expected values of the left-hand side of (2.3) are zero under the true value of θ. Moreover, under the restriction that $0 < \pi < 1$, $T_i(w_i\left(A_i, \beta\right))$ must take a value between 0 and 1. This property is very useful to prove the consistency and asymptotic normality of the maximum likelihood estimator.

2.3 Parameter Estimation for Partially-Observed Covariates

To evaluate $\Lambda_i(A_i, \beta)$ and $\Lambda_i^{(1)}(A_i, \beta)$ in the mixture model score functions, $\Psi_n(\theta)$, maximum likelihood estimation requires that the values of the covariates be known functions of time $t \in A_i$ for all $i = 1, .., n$. However, this is generally not the case, since the covariates are observed only at sample locations in many applications.

Parameter estimation requires evaluation of function of the form

$$g(A) = \int_A g(s)ds.$$

In this context, this includes $\Lambda(A, \beta)$ and $\Lambda^{(1)}(A, \beta)$ for example. Cordy (1993) suggested a design-unbiased estimator for $g(A)$, which may be obtained from a probability sample of points $u_1, .., u_m$ in the study interval A. For a sample of size m, define the inclusion density $p(u) = \sum_{j=1}^{m} f_j(u)$, where $f_j(u)$ is the known marginal probability density function for sites u_j, $j = 1, .., m$. Alternatively, the sites may be sampled according to a Poisson point process with known intensity $p(u)$. The Horvitz-Thompson estimator is

$$\hat{g}(A) = \sum_{i=1}^{m} \frac{g(u_i)}{p(u_i)},$$

which has variance-covariance matrix

$$var\left\{\hat{g}(A)\right\} = \int_A \frac{g(s,\beta)g^T(s,\beta)}{p(s)}ds + \int_A\int_A g(s,\beta)g^T(u,\beta)\left(\frac{p(s,u)-p(s)p(u)}{p(s)p(u)}\right)dsdu. \quad (2.4)$$

Here, $\hat{g}(A)$ is an unbiased estimator for $g(A)$ and

$$p(s,u) = \sum_{i\neq j} f_{ij}(s,u), \quad s,u \in A,$$

is the pairwise inclusion density, where $f_{ij}(\cdot,\cdot)$ is the joint probability density function for sites i and j. Under the simple random sampling design, the Horvitz-Thompson estimator takes the form

$$\hat{g}(A) = \frac{|A|}{m}\sum_{i=1}^{m} g(u_i), \quad (2.5)$$

where $|A|$ denotes the size of study region.

Substituting (2.5) into (2.3) yields an estimating function of the form

$$\tilde{\Psi}_n(\theta) = n^{-1}\sum_{i=1}^{n}\left[\begin{array}{c} \hat{v}_i(A_i,\beta_1)T_i(\hat{w}_i\left(A_i,\beta\right)) \\ \hat{v}_i(A_i,\beta_2)\left(1-T_i(\hat{w}_i\left(A_i,\beta\right))\right) \\ \frac{T_i(\hat{w}_i(A_i,\beta))}{\pi} - \frac{1-T_i(\hat{w}_i(A_i,\beta))}{1-\pi} \end{array}\right], \quad (2.6)$$

where

$$\hat{v}_i(A_i,\beta) = \sum_{j=1}^{N(A_i)} x_i(t_{ij}) - \hat{\Lambda}_i^{(1)}(A_i,\beta),$$

and

$$\hat{w}_i\left(A_i,\beta\right) = \left(\sum_{j=1}^{N(A_i)} \beta_2^T x_i(t_{ij}) - \hat{\Lambda}_i(A_i,\beta_2)\right) - \left(\sum_{j=1}^{N(A_i)} \beta_1^T x_i(t_{ij}) - \hat{\Lambda}_i(A_i,\beta_1)\right).$$

The proposed estimator $\hat{\theta}$ is obtained by solving $\tilde{\Psi}_n(\theta) = 0$. The next section describes an EM algorithm that may be used to obtain this solution.

2.4 The EM -algorithm for Mixture Poisson Process

Suppose the group membership z_i can be observed, then the complete log-likelihood for the mixture Poisson process with intensity (2.2) can be written as

$$\begin{aligned} L_n(\theta) &= \sum_{i=1}^{n} z_i \left\{ \ln(\pi) + \left(\sum_{j=1}^{N(A_i)} \beta_1^T x_i(t_{ij}) - \Lambda_i(A_i, \beta_1) \right) \right\} \\ &\quad + \sum_{i=1}^{n} (1 - z_i) \left\{ \ln(1-\pi) + \left(\sum_{j=1}^{N(A_i)} \beta_2^T x_i(t_{ij}) - \Lambda_i(A_i, \beta_2) \right) \right\} \end{aligned} \tag{2.7}$$

To estimate θ, one could consider z_i in (2.7) as missing data as in Dempster et al. (1977), and maximize (2.7) by the EM Algorithm. Let $\theta^{(r)}$ denote the value for θ for the rth iterate of the EM algorithm, with $\theta^{(0)}$ be the initial specified value. Then, the EM algorithm may be implemented as follows:

E-STEP: Compute the conditional expectation of z_i given the observed data,

$$w_i^{(r)} = \Pr\left\{ z_i = 1 | \theta^{(r)}, x_i(\cdot) \right\} = \frac{\pi^{(r)} \exp\left\{ v_i(A_i, \beta_1^{(r)}) \right\}}{\pi^{(r)} \exp\left\{ v_i(A_i, \beta_1^{(r)}) \right\} + (1 - \pi^{(r)}) \exp\left\{ v_i(A_i, \beta_2^{(r)}) \right\}},$$

M-STEP: The parameter vector θ is estimated by finding θ that maximize

$$\begin{aligned} L_n(\theta) &= \sum_{i=1}^{n} w_i^{(r)} \left\{ \ln(\pi) + \left(\sum_{j=1}^{N(A_i)} \beta_1^T x_i(t_{ij}) - \Lambda_i(A_i, \beta_1) \right) \right\} \\ &\quad + \sum_{i=1}^{n} (1 - w_i^{(r)}) \left\{ \ln(1-\pi) + \left(\sum_{j=1}^{N(A_i)} \beta_2^T x_i(t_{ij}) - \Lambda_i(A_i, \beta_2) \right) \right\}. \end{aligned}$$

which is easily done using the Newton-Raphson algorithm. The solution for π in the M-step can be written in closed form as:

$$\pi^{(r+1)} = n^{-1} \sum_{i=1}^{n} w_i^{(r)}.$$

To obtain the large-sample information matrix of the maximum likelihood estimator of θ from EM-algorithm, Louis (1982) shows that the observed information matrix when using EM-algorithm can be written in the following form

$$\Sigma(\theta) = \Sigma_1(\theta) - \Sigma_2(\theta),$$

where in the current context

$$\Sigma_1(\theta) = \sum_{i=1}^{n} \begin{bmatrix} \frac{T_i(w_i(A_i,\beta))}{\pi^2} - \frac{1-T_i(w_i(A_i,\beta))}{(1-\pi)^2} & 0 & 0 \\ 0 & T_i(w_i(A_i,\beta))\Lambda_i^{(2)}(A_i,\beta_1) & 0 \\ 0 & 0 & (1-T_i(w_i(A_i,\beta)))\Lambda_i^{(2)}(A_i,\beta_2) \end{bmatrix},$$

$$\Sigma_2(\theta) = \sum_{i=1}^{n} R_i(\theta)(1-R_i(\theta)) \begin{bmatrix} \frac{1}{(\pi(1-\pi))^2} & \frac{\Lambda_i^{(1)}(A_i,\beta_1)^T}{\pi(1-\pi)} & \frac{\Lambda_i^{(1)}(A_i,\beta_2)^T}{\pi(1-\pi)} \\ \frac{\Lambda_i^{(1)}(A_i,\beta_1)}{\pi(1-\pi)} & \Delta_i(\beta_1,\beta_1) & -\Delta_i(\beta_1,\beta_2) \\ \frac{\Lambda_i^{(1)}(A_i,\beta_2)}{\pi(1-\pi)} & -\Delta_i(\beta_1,\beta_2) & \Delta_i(\beta_2,\beta_2) \end{bmatrix},$$

$$\Delta_i(\beta_j,\beta_k) = \Lambda_i^{(1)}(A_i,\beta_j)\Lambda_i^{(1)}(A_i,\beta_k)^T,$$

and

$$\Lambda_i^{(2)}(A_i,\beta) = \frac{\partial\Lambda_i^{(1)}(A_i,\beta)}{\partial\beta^T} = \int_{A_i} x_i(t)x_i(t)^T \exp\left\{\beta^T x_i(t)\right\} dt.$$

2.5 Asymptotic Properties

In spatial statistics, two asymptotic frameworks, increasing domain asymptotics and infill asymptotics, have been applied for obtaining limiting distribution of estimators. Under increasing domain asymptotics, the size of study region increases while the number of samples per unit area remains constant. Kutoyants (1984), Rathbun and Cressie (1994), Rathbun et al. (2007), and Waagepetersen (2008) demonstrated the consistency and asymptotic normality of the maximum likelihood estimators for the Poisson process under increasing domain asymtotics. Under infill asymptotics, the size of the study region is fixed while the number of samples increases. Infill asymptotics is frequently applied in geostatistics (Stein, 1989, 1993, 1999), and combined increasing domain/infill asymptotics has been applied by Hall and Patil (1994) for nonparametric estimator of the autocovariance function of a random field, and Lahiri et al. (1999) for spatial cumulative distribution functions. Under combined increasing domain/infill asymptotic paradigm, $|A| \to \infty$ and $m \,/\, |A| \to \infty$, where m is the number of observations.

Here we will study the large-sample properties of the proposed estimator as the number of subjects $n \to \infty$ in effect of increasing domain asymptotics, and the size $|A_i|$ of the study intervals for each subject remains fixed. Moreover, take an infill asymptotics approach, the number of covariate samples for each subject will also increase with increasing n.

The following assumptions regarding the covariates and events are required to prove the consistency and asymptotic normality of the proposed estimator:

Assumption 1. The subjects $i = 1, .., n$ are independently sampled, and all event times are independent of all assessment times, both within and between subjects.

Assumption 2. There exists an upper bound M such that $|x_i(t)| < M$ for almost all $t \in R^d$ and all $i = 1, .., n$.

Assumption 3. The study intervals are bounded above and below by some positive constants, a_1 and a_2. That is $0 < a_1 \leq |A_i| \leq a_2 < \infty$.

Assumption 4. The inclusion densities $p_i(u)$ and pairwise inclusion densities $p_i(u, s)$ are bounded below by some constants greater than zero for all $i = 1, .., n$.

Assumption 5.

$$n^{-1} \sum_{i=1}^{n} \int_{A_i} x_i(t) x_i(t)^T dt \to B,$$

as $n \to \infty$, where the minimum eigenvalue of B is strictly positive.

Assumption 6. For each u, the covariate sampling design is such that there exists the positive bounds, b_1 and b_2 such that

$$0 < n^{\alpha} b_1 \leq m_i \leq n^{\alpha} b_2 < \infty,$$

where

$$m_i = \int_{A_i} p_i(s) ds,$$

for $\alpha > 2$.

Assumption 7. The probability of belonging to the first group

$$\pi \in (0, 1),$$

for each subject $i = 1, .., n$.

Assumption 8. For all bounded functions $f(.)$, the covariate sampling design is such that

$$Var(\hat{f}(A)) = O(m^{-1}),$$

where $\hat{f}(A)$ is the Horvitz-Thompson estimator for

$$f(A) = \int_A f(s)ds.$$

These assumptions are similar to those posed by Rathbun et al. (1994), and are sufficient for the maximum likelihood estimator for the modulated Poisson process to be consistent and asymptotic normal as $n \to \infty$. Assumption 1 is required to ensure that there is no correlation among the subjects. Assumption 2 ensures that the estimating functions and their derivatives with respect to the model parameters have bounded moments. Assumption 3 requires that all the study intervals are bounded both from above and below. Assumption 4 is required to ensure that the random assessments have adequate coverage of the study region, and also ensure that the $Var(\hat{\Lambda}_i(A_i, \beta))$ is bounded. Assumption 5 is necessary to ensure that the derivatives of the estimating functions are positive definite at the true values of the model parameters. Assumption 6 ensures that the sampling intensity increases with increasing number of subjects at a controlled rate. Assumption 7 ensures that parameter is in the interior of the parameter space. Finally, Assumption 8 ensures that the sampling design is such that the proposed estimators are consistent under infill asymptotics.

The following lemmas are also required to obtain a consistent solution $\tilde{\theta}_n$ to $\tilde{\Psi}_n(\theta) = 0$:

Lemma 1. Under Assumption 1-8 and for each neighborhood $S(\theta_0, \varepsilon)$ of θ_0, the estimating function $\tilde{\Psi}_n(\theta)$ converge uniformly to a nonstochastic limit $\Psi_\infty(\theta)$ that is equal to zero with probability one at θ_0 as $n \to \infty$.

Lemma 2. Under Assumption 1-8 and for each neighborhood $S(\theta_0, \varepsilon)$ of θ_0, the Jacobians $\tilde{\Psi}_n^{(1)}(\theta)$ converge uniformly in θ to a nonstochastic limit that is nonsingular at θ_0 as. $n \to \infty$.

The proof of both lemmas can be found in Appendix A. The consistency of $\tilde{\theta}_n$ then follows from Theorem 3 of Yuan and Jenrich (1998).

Theorem 1. Suppose that assumptions 1-8 are satisfied, then $\tilde{\theta}_n \to \theta_0$ almost surely as $n \to \infty$.

To obtain the large-sample variance of $\tilde{\theta}_n$, define the information matrix

$$J_n(\theta) = -E_\theta \left\{ \frac{\partial}{\partial \theta} \Psi_n(\theta) \right\},$$

and the variance of the estimating function

$$\Sigma_n(\theta) = var_\theta \left\{ \tilde{\Psi}_n(\theta) \right\}.$$

The full form of $\frac{\partial}{\partial \theta}\Psi_n(\theta)$ can be found in Appendix A, and the matrix $J_n(\theta)$ is nonsingular. The details of attaining the element terms of $\Sigma_n(\theta)$ can be found in Appendix B. To prove the asymptotic normality of $\tilde{\theta}_n$, the asymptotic normality of the estimating function $\Psi_n(\theta_0)$ is required.

Lemma 3. Under Assumption 1-8, $\sqrt{n}\Psi_n(\theta_0)$ converges in distribution to a Gaussian random vector with mean 0 and finite variance $\Sigma_\infty(\theta)$, where

$$\Sigma_\infty(\theta) = \lim_{n \to \infty} \frac{1}{n} \Sigma_n(\theta)$$

The proof of this lemma is also shown in Appendix A. The asymptotic normality of $\tilde{\theta}_n$ then follows from Theorem 4 of Yuan and Jenrich (1998):

Theorem 2. Under Assumption 1-8,

$$\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) \to N(0, V(\theta_0)),$$

where

$$V(\theta_0) = \{J_\infty(\theta_0)\}^{-1} \Sigma_\infty(\theta_0) \{J_\infty(\theta_0)\},$$

and

$$J_\infty(\theta_0) = - \lim_{n \to \infty} E \left\{ \frac{\partial}{\partial \theta} \Psi_n(\theta_0) \right\}.$$

Here, the variance of $\tilde{\theta}_n$ is partitioned into two sources, that is due to the Poisson point process, and due to the random sampling of the covariate.

2.6 Assessing the number of groups

Since the log-likelihood for $\theta = \left(\pi_1, .., \pi_{k-1}, \beta_1^T, .., \beta_k^T\right)^T$

$$L_n(K, \beta) = \sum_{i=1}^{n} \ln \left(\sum_{j=1}^{K} \pi_j \exp \left\{ \beta_j^T \sum_{l=1}^{N_i} x_i(t_l) - \int_{A_i} \exp \left\{ \beta_j^T x_i(t) \right\} dt \right\} \right)$$

is an increasing function of the number of clusters, K, it is not suitable as a selection criterion for the number of components in mixture Poisson process. The Akaike information criterion (Akaike 1974) has been used in mixture model context (Bozdogan and Sclove 1984). This criterion can be written as

$$AIC(K) = 2L_n(K, \beta) - 2\nu(K)$$

where $\nu(K)$ denotes the number of free parameters in the mixture model with K groups.

However, the AIC may overestimate the true number of components (Koehler and Murphrec 1988). The Bayesian information criterion (Schwarz 1978) has been proposed to solve this problem. It takes the form

$$BIC(K) = 2L_n(K, \beta) - \nu(K) \ln n.$$

Banfield and Raftery (1992) suggested another Bayesian criterion, which penalizes more drastically high order models than BIC, to choose the number of components in the mixture. In the mixture context, the approximate weight of evidence (AWE) takes the form

$$AWE(K) = 2L_n(K, \beta) - 2\nu(K) \left(\ln n + \frac{3}{2} \right).$$

Finally, the minimum information ratio (MIR) has been proposed to be a criterion for mixture model selection when using the EM algorithm (Windham and Cutler 1991). The value of MIR is between 0 and 1. A large value suggests good mixture structure in the model. MIR can be estimated by

$$MIR = 1 - \frac{||\theta^{m+1} - \theta^m||}{||\theta^m - \theta^{m-1}||}, \quad \text{for large enough } m$$

where θ^m is the parameter estimated at the mth iteration of EM. The next section will demonstrate the use of these criteria for a real dataset.

2.7 Ecological Momentary Assessment of Smoking

The proposed approach is illustrated using data from an EMA of smoking by Shiffman et al. (2002). A total of 304 smokers were given electronic diaries (PDAs) and were instructed to record each time when they smoked a cigarette. The covariates include both mood affect and environmental setting. Participants reported their smoking location, whether smoking was permitted (forbidden, discouraged, or allowed), whether they were in the company of others, and whether others were smoking in view of the participant. The random assessments were sampled according to a stratified sampling design, where days were treated as strata. To reduce the burden on the study smokers, the covariates were not assessed for every cigarette. The electronic diary targets subjects to record covariate information for 4 or 5 cigarette assessments per day. For subject i on day j, each cigarette was independently chosen to be assessed with probability

$$p_{ij} = \min\left\{5N_{i,j-1}^{-1}, 1\right\},$$

where $N_{i,j}$ denotes the number of events recorded on day j by subject i. Estimates of the parameter $\hat{\theta}_n$ can be obtained by solving the estimating equation $\hat{\Psi}_n(\theta) = 0$, where

$$\hat{\Lambda}_i(A_i, \beta) = \sum_{j=1}^{d_i} \left\{ \frac{p_{ij}\,|L_{ij}|}{m_{ij}} \sum_{k=1}^{m_{ij}} \exp\left\{\beta^T x_i(u_{ijk})\right\} \right\},$$

and

$$\hat{\Lambda}_i^{(1)}(A_i, \beta) = \sum_{j=1}^{d_i} \left\{ \frac{p_{ij}\,|L_{ij}|}{m_{ij}} \sum_{k=1}^{m_{ij}} x_i(u_{ijk}) \exp\left\{\beta^T x_i(u_{ijk})\right\} \right\}$$

with $n = 304$ subjects, d_i is the number of days on which subject i participated in the study, N_{ij} denotes the number of cigarettes for subject i on day j, m_{ij} is the number of random assessments for subject i on day j, and $|L_{ij}|$ denotes the length of time that the diary was active.

Table 2.1: The criteria for assessing the number of groups in a mixture modulated Poisson process model.

K	AIC	BIC	AWE	MIR
2	29535.18	29579.78	29684.39	0.68
3	26383.21	27698.23	27043.54	0.54
4	26212.32	26823.43	26887.24	0.47
5	24321.66	26421.44	25345.65	0.32

Table 2.1 displays the value of each criterion AIC, BIC, AWE, and MIR for different numbers of clusters, K. From Table 1, the two clusters solution has the highest values of all criteria. Therefore, two-component mixture is the best solution here for these data.

Table 2.2: Parameter estimates from mixture modulated Poisson process model.

	first cluster				second cluster			
Covariate	Estimate	SE	z	$p-$value	Estimate	SE	z	$p-$value
Intercept	-0.0270	0.0457			0.0298	0.0105		
Negative Affect	0.0144	0.0102	1.40	0.14	-0.0140	0.0077	-1.81	0.077
Arousal	-0.0144	0.0261	-0.55	0.342	0.0102	0.0131	0.77	0.296
Attention	0.0114	0.0324	0.351	0.375	-0.0129	0.0148	-0.87	0.273
Restlessness	0.3316	0.0230	14.36	<0.0001	0.0475	0.0268	1.77	0.083
Discouraged	0.0124	0.0112	1.113	0.2147	0.0154	0.0148	1.04	0.2321
Allowed	0.2342	0.0315	7.434	< 0.0001	0.0125	0.0402	0.31	0.380
Alone	-0.0117	0.0558	-0.209	0.3902	-0.0201	0.0671	-0.299	0.3814
Other	0.0154	0.0689	0.224	0.389	0.2559	0.0211	12.128	< 0.0001
SE, standard error.								

Table 2.2 shows the results of fitting a two-component mixture modulated Poisson process for both subject's clusters with an estimated that $\hat{\pi} = 0.6518$ of the subjects belong to the first cluster. From these results, we can see that only Restlessness and Allowed have a significant effect on the smoking rate in the first cluster, while Other smokers was statistically significant in the second cluster. Therefore, given that subject is in the first cluster, the smoking rate increases by a multiplicative factor of $\exp(0.33158) = 1.393$ for each unit increase in Restlessness with all remaining variables held constant. Moreover, when the

smoking is allowed, the smoking rate is also increased from when smoking is prohibited by a multiplicative factor of $\exp(0.2342) = 1.26$. Finally, given that subject is in the second cluster, the smoking rate is increased by a multiplicative factor of $\exp(0.2559) = 1.29$ when there are others smoking in view of the participant.

2.8 Appendix A

The methods of Yuan and Jennrich (1997) were used to demonstrate the consistency and the asymptotic normality of solutions to estimating functions. If the estimating functions $\Psi_n(\theta)$ satisfy the following conditions, then the solution $\hat{\theta}_n$ to $\Psi_n(\theta) = 0$ is consistent estimator of the true value θ_0, and also asymptotically normal as $n \to \infty$:

Condition A1. $\tilde{\Psi}_n(\theta_0) \to 0$ with probability one

Condition A2. There is a neighborhood K of θ_0 on which with probability one all $\tilde{\Psi}_n(\theta)$ are continuously differentiable and the Jacobian $\tilde{\Psi}_n^{(1)}(\theta)$ converge uniformly to the nonstochastic limit which is nonnegative definite at θ_0.

Condition A3. $\sqrt{n}\tilde{\Psi}_n(\theta_0) \to N(0, V)$ in distribution.

Proof of Lemma 1. Suppose that the data consists of the locations $t_1, .., t_N$ of N events in study region A . Then, the likelihood function for the Poisson process whose intensity takes the form $\exp\left\{\beta^T x(t)\right\}$ is given by the Jenossy density (Daley and Vere-Jones, 1988)

$$l(N, \beta) = \frac{1}{N!} \exp \sum_{i=1}^{N} \beta^T x(t_i) - \Lambda(A, \beta).$$

This likelihood function has discrete and continuous components. The discrete component is the number of events. Continuous components are the event locations. Then, consider summing the Jenossy density over the number of events N, and integrating it over the possible event locations, we can see that

$$\begin{aligned}
\sum_{N=0}^{\infty} \int_{A_N} l(N, \beta) dt_1..dt_N &= \sum_{N=0}^{\infty} \int_{A_N} \frac{1}{N!} \exp\left\{\sum_{i=1}^{N} \beta^T x(t_i) - \Lambda(A, \beta)\right\} dt_1..dt_N \\
&= \exp\left(-\Lambda(A, \beta)\right) \sum_{N=0}^{\infty} \frac{1}{N!} \prod_{i=1}^{N} \left\{\int_A \exp\left\{\beta^T x(t_i)\right\} dt_i\right\} \\
&= \exp\left(-\Lambda(A, \beta)\right) \sum_{N=0}^{\infty} \frac{\left(\Lambda(A, \beta)\right)^{N_i}}{N!} = 1.
\end{aligned}$$

When $l(N, \beta) = \pi \exp\left\{v(A, \beta_1)\right\} + (1 - \pi) \exp\left\{v(A, \beta_2)\right\}$ for the mixture model case, It can still be demonstrated that $\sum_{N=0}^{\infty} \int_{A_N} l(N, \beta) dt_1..dt_N = 1$. Therefore,

$$E\left(\frac{\partial \log l(N, \beta)}{\partial \beta_1}\right) = 0,$$

which implies,

$$E\left\{\sum_{i=1}^{n} v_i(A_i,\beta_1)T_i(w_i\left(A_i,\beta\right))\right\} = E\left\{\sum_{i=1}^{n} v_i(A_i,\beta_2)\left(1-T_i(w_i\left(A_i,\beta\right))\right)\right\} = 0.$$

In order to prove $E\left\{\hat{v}_i(A_i,\beta_1)T_i(\hat{w}_i\left(A_i,\beta\right))\right\} = E\left\{\hat{v}_i(A_i,\beta_2)\left(1-T_i(\hat{w}_i\left(A_i,\beta\right))\right)\right\} \to 0$ for the partially-observed covariates case, we consider the Taylor series around $v_i(A_i,\beta_1)$ and $w_i\left(A_i,\beta\right)$ for the first element of $\tilde{\Psi}_n(\theta)$:

$$\begin{aligned}\sum_{i=1}^{n} \hat{v}_i(A_i,\beta_1)T_i(\hat{w}_i\left(A_i,\beta\right)) \;=\;& \sum_{i=1}^{n} v_i(A_i,\beta_1)T_i(w_i\left(A_i,\beta\right)) + \sum_{i=1}^{n} T_i(w_i\left(A_i,\beta\right))\left[\hat{v}_i(A_i,\beta_1)-v_i(A_i,\beta_1)\right] \\ &+\sum_{i=1}^{n} v_i(A_i,\beta_1)T_i^{(1)}(w_i\left(A_i,\beta\right))\left[\hat{w}_i\left(A_i,\beta\right)-w_i\left(A_i,\beta\right)\right] \\ &+O((\hat{w}_i\left(A_i,\beta\right)-w_i\left(A_i,\beta\right))+(\hat{v}_i(A_i,\beta_1)-v_i(A_i,\beta_1)))\ , \qquad (A.1)\end{aligned}$$

where

$$T^{(1)}(w) = \frac{\partial T(w)}{\partial w} = -\frac{\pi(1-\pi)\exp(w)}{\left(\pi+(1-\pi)\exp(w)\right)^2} = T(w)\left(1-T(w)\right).$$

Under Assumption 8, it can be seen that

$$\hat{v}_i(A_i,\beta_1)-v_i(A_i,\beta_1) = o(m^{-1/2}) = o(n^{-\alpha/2}).$$

So, under infill asymptotics,

$$n^{-1}\sum_{i=1}^{n}\left\{\hat{v}_i(A_i,\beta_1)-v_i(A_i,\beta_1)\right\} \to 0$$

in probability. Before considering the expectation of the second term on the RHS of $(A.1)$, we consider the properties applied to the elements of the vector $\hat{v}_i(A_i,\beta_1)-v_i(A_i,\beta_1)$:

(1) since $E\left(\hat{v}_i(A_i,\beta_1)-v_i(A_i,\beta_1)\right)=0$, we have

$$E\left(T_i(w_i\left(A_i,\beta\right))\left[\hat{v}_i(A_i,\beta_1)-v_i(A_i,\beta_1)\right]\right) = Cov\left(T_i(w_i\left(A_i,\beta\right)),\left[\hat{v}_i(A_i,\beta_1)-v_i(A_i,\beta_1)\right]\right)$$

(2) Moreover,

$$\begin{aligned}-\sqrt{VarT_i(w_i\left(A_i,\beta\right))Var\left[\hat{v}_i(A_i,\beta_1)-v_i(A_i,\beta_1)\right]} \;\leq\;& Cov\left(T_i(w_i\left(A_i,\beta\right)),\left[\hat{v}_i(A_i,\beta_1)-v_i(A_i,\beta_1)\right]\right) \\ \leq\;& \sqrt{VarT_i(w_i\left(A_i,\beta\right))Var\left[\hat{v}_i(A_i,\beta_1)-v_i(A_i,\beta_1)\right]}\,.\end{aligned}$$

Since $0 < T_i(\hat{w}_i(A_i, \beta)) < 1$, so is $Var(T_i(\hat{w}_i(A_i, \beta)))$.

(3) It suffices to show that

$$Var[\hat{v}_i(A_i, \beta_1) - v_i(A_i, \beta_1)] \to 0.$$

Therefore the expectation of the second term on the RHS of (A.1) converges in probability to zero.

Using Holder's inequality and the property that $0 < T_i^{(1)}(\hat{w}_i(A_i, \beta)) < 1$, the expectation of the third term in $(A.1)$ can be calculated as

$$\begin{aligned} E\left(v_i T_i^{(1)}(w_i(A_i, \beta))\left[\hat{w}_i - w_i\right]\right) &\leq \left(E\left(v_i(A_i, \beta_1) T_i^{(1)}(w_i(A_i, \beta))\right)^2\right)^{1/2} \left(E\left(\hat{w}_i(A_i, \beta) - w_i(A_i, \beta)\right)^2\right)^{1/2} \\ &\leq \left(E\left(v_i(A_i, \beta_1)\right)^2\right)^{1/2} \left(E\left(\hat{w}_i(A_i, \beta) - w_i(A_i, \beta)\right)^2\right)^{1/2}. \end{aligned}$$

where

$$\left(E\left(v_i(A_i, \beta_1)\right)^2\right)^{1/2} \left(E\left(\hat{w}_i(A_i, \beta) - w_i(A_i, \beta)\right)^2\right)^{1/2} = o_p\left(|A_i| m_i^{-1/2}\right) = o_p\left(n^{-\alpha/2}\right).$$

So, under infill asymptotics and Assumption 8, $\sum_{i=1}^{n} \hat{v}_i(A_i, \beta_1) T_i(\hat{w}_i(A_i, \beta)$ converges in probability to zero. A similar approach can be used for the remaining two terms in $\tilde{\Psi}_n(\theta)$.

To prove uniformly convergence of the estimating equations, let $Var_{jk}(B)$ denote the element in the jth row and kth column of matrix $Var(B)$. Using the property of $0 < \hat{T}_i(\hat{w}_i(A_i, \beta)) < 1$, it can be seen that

$$Var_{jk}\left\{\hat{v}_i(A_i, \beta_1) T_i(\hat{w}_i(A_i, \beta))\right\} \leq 2 Var_{jk}\left\{\hat{v}_i(A_i, \beta_1)\right\},$$

$$Var_{jk}\left\{\hat{v}_i(A_i, \beta_2)\left(1 - T_i(\hat{w}_i(A_i, \beta))\right)\right\} \leq 2 Var_{jk}\left\{\hat{v}_i(A_i, \beta_2)\right\},$$

and

$$Var\left\{\frac{T_i(\hat{w}_i(A_i, \beta))}{\pi} - \frac{1 - T_i(\hat{w}_i(A_i, \beta))}{1 - \pi}\right\} = \frac{Var\left\{T_i(\hat{w}_i(A_i, \beta))\right\}}{(\pi(1 - \pi))^2} \leq \frac{1}{\pi(1 - \pi)},$$

where the matrix $Var(\hat{v}_i(A_i, \beta))$ is

$$Var(\hat{v}_i(A_i, \beta)) = Var\left(\sum_{j=1}^{N(A_i)} x_i(t_{ij})\right) + Var\left(\hat{\Lambda}_i^{(1)}(A_i, \beta)\right), \qquad (A.2)$$

$$
\begin{aligned}
Var\left(\sum_{j=1}^{N(A_i)} x_i(t_{ij})\right) &= \pi\Lambda_i^{(2)}(A_i,\beta_1)+(1-\pi)\Lambda_i^{(2)}(A_i,\beta_2)\\
&\quad +\pi(1-\pi)\left(\hat{\Lambda}_i^{(1)}(A_i,\beta_1)-\hat{\Lambda}_i^{(1)}(A_i,\beta_2)\right)\left(\hat{\Lambda}_i^{(1)}(A_i,\beta_1)-\hat{\Lambda}_i^{(1)}(A_i,\beta_2)\right)^T.
\end{aligned}
$$

and $Var\left(\hat{\Lambda}_i^{(1)}(A_i,\beta)\right)$ can be calculated by (2.4).

By Assumption 2, 4, 5,and 6, $(A.2)$ is bounded. Therefore, the variances and the second moments of the independent random variables contained in $\tilde{\Psi}_n(\theta)$ share a common bound. So, Theorem 5.1.1 (Chung K.L. 1974, p. 103) and Kolmogorov's condition (Feller 1968, p. 259) are satisfied. $\tilde{\Psi}_n(\theta)$ converges pointwise to a nonstochastic limit for each $\theta\in\Theta$ with probability 1. In addition, since $\tilde{\Psi}_n(\theta)$ is unbiased estimating function, $\tilde{\Psi}_n(\theta_0)$ converges to zero with probability one.

Proof of Lemma 2. The negative of the derivative of $\tilde{\Psi}_n(\theta)$ with respect to θ is

$$
\begin{aligned}
-\tilde{\Psi}_n^{(1)}(\theta) &= -n^{-1}\sum_{i=1}^{n}\frac{1}{h_i(A_i,\theta)}\begin{bmatrix}\pi f_i^{(2)}(A_i,\beta_1) & 0 & f_i^{(1)}(A_i,\beta_1)^T\\ 0 & (1-\pi)f_i^{(2)}(A_i,\beta_2) & -f_i^{(2)}(A_i,\beta_2)\\ f_i^{(1)}(A_i,\beta_1) & -f_i^{(2)}(A_i,\beta_2)^T & 0\end{bmatrix}\\
&\quad +n^{-1}\sum_{i=1}^{n}\frac{1}{|h_i(A_i,\theta)|^2}d_i(A_i,\theta)d_i(A_i,\theta)^T\ , \qquad (A.3)
\end{aligned}
$$

where

$$
h_i(A_i,\theta)=\pi\exp\left\{\sum_{j=1}^{N_i}\beta_1^Tx_i(t_{ij})-\tilde{\Lambda}_i(A_i,\beta_1)\right\}+(1-\pi)\exp\left\{\sum_{j=1}^{N_i}\beta_2^Tx_i(t_{ij})-\tilde{\Lambda}_i(A_i,\beta_2)\right\},
$$

$$
f_i(A_i,\beta)=\exp\left\{\sum_{j=1}^{N_i}\beta^Tx_i(t_{ij})-\tilde{\Lambda}_i(A_i,\beta)\right\},
$$

$$
f_i^{(1)}(A_i,\beta)=\frac{\partial f_i(A_i,\beta)}{\partial\beta^T}=\hat{v}_i(A_i,\beta)\exp\left\{\sum_{j=1}^{N_i}\beta^Tx_i(t_{ij})-\tilde{\Lambda}_i(A_i,\beta)\right\},
$$

$$
f_i^{(2)}(N_i,\beta)=\frac{\partial f_i^{(1)}(A_i,\beta)}{\partial\beta^T}=\left(\hat{v}_i(A_i,\beta)\hat{v}_i(A_i,\beta)^T-\hat{\Lambda}_i^{(2)}(A_i,\beta)\right)\exp\left\{\sum_{j=1}^{N_i}\beta^Tx_i(t_{ij})-\tilde{\Lambda}_i(A_i,\beta)\right\},
$$

and

$$
d_i(A_i,\beta)=\begin{pmatrix}\pi f_i^{(1)}(A_i,\beta_1)\\ (1-\pi)f_i^{(1)}(A_i,\beta_2)\\ f_i(A_i,\beta_1)-f_i(A_i,\beta_2)\end{pmatrix}.
$$

By Assumption 1-8, the first term on the RHS of $(A.3)$ converges to zero with probability 1, while the second term is positive definite. In addition, $\tilde{\Psi}_n^{(1)}(\theta)$ is also the sample mean of independent random variables whose variance share a common upper bound under Assumption 1-8. Therefore, Komogorov's condition (Feller 1968, p. 259) is satisfied, and hence $\tilde{\Psi}_n^{(1)}(\theta)$ converges pointwise to a nonstochastic limit for each $\theta \in \Theta$ with probability 1. Moreover, the derivative of $\tilde{\Psi}_n^{(1)}(\theta)$ includes the function of $\hat{\Lambda}(A, \beta)$ and its derivatives. By Assumption 2-6, all of which are bounded above by a positive constant. Therefore, $\tilde{\Psi}_n^{(1)}(\theta)$ converges uniformly to a nonstochastic limit as $n \to \infty$. Specifically, $\tilde{\Psi}_n^{(1)}(\theta_0)$ converges to a nonsingular matrix.

Proof of Lemma 3. First, consider the first element of $\tilde{\Psi}_n(\theta)$.Since $T_i(\hat{w}_i\left(A_i, \beta\right)) < 1$, it can be seen that

$$|\hat{v}_i(A_i, \beta_1) T_i(\hat{w}_i\left(A_i, \beta\right))|^3 \leq |\hat{v}_i(A_i, \beta_1)|^3 .$$

$$E\left|\hat{v}_i(A_i, \beta_1) T_i(\hat{w}_i\left(A_i, \beta\right))\right|^3 \leq E\left|\hat{v}_i(A_i, \beta_1)\right|^3 .$$

Let a denote any unit vector in R^p. Let $y(t) = a^T x(t) \in R$, and let $|y(t)|$ denotes the absolute value of $y(t)$element by element. By straightforward calculation,

$$\begin{aligned}
E\left(\hat{v}_i(A_i, \beta_1)\right)^3 &= E\left(\sum_{j=1}^{N_i} y_i(t_{ij}) - \sum_{t \in D} \frac{y_i(t) \exp\left\{\beta_1^T x_i(t)\right\}}{p_i(t)}\right)^3 \\
&\leq E\left|\sum_{j=1}^{N_i} y_i(t_{ij})\right|^3 + 3E\left|\sum_{j=1}^{N_i} y_i(t_{ij})\right| E\left|\sum_{t \in D} \frac{|y_i(t)| \exp\left\{\beta_1^T x_i(t)\right\}}{p_i(t)}\right|^2 \\
&\quad +3E\left|\sum_{j=1}^{N_i} y_i(t_{ij})\right|^2 E\left|\sum_{t \in D} \frac{|y_i(t)| \exp\left\{\beta_1^T x_i(t)\right\}}{p_i(t)}\right| + E\left|\sum_{t \in D} \frac{|y_i(t)| \exp\left\{\beta_1^T x_i(t)\right\}}{p_i(t)}\right|^3, \quad (A.4)
\end{aligned}$$

where the terms on the RHS of $(A.4)$ can be written as

$$\begin{aligned}
E\left|\sum_{j=1}^{N_i} y_i(t_{ij})\right|^3 &= \int_{A_i} y_i^3(t) \exp\left\{\beta_1^T x_i(t)\right\} + \left(\int_{A_i} y_i(t) \exp\left\{\beta_1^T x_i(t)\right\}\right)^3 \\
&\quad +3\left(\int_{A_i} y_i^2(t) \exp\left\{\beta_1^T x_i(t)\right\}\right)\left(\int_{A_i} y_i(t) \exp\left\{\beta_1^T x_i(t)\right\}\right),
\end{aligned}$$

$$E\left|\sum_{t \in D} \frac{|y_i(t)| \exp\left\{\beta_1^T x_i(t)\right\}}{p_i(t)}\right| = \int_{A_i} |y_i(t)| \exp\left\{\beta_1^T x_i(t)\right\},$$

$$E\left|\sum_{t\in D}\frac{|y_i(t)|\exp\left\{\beta_1^T x_i(t)\right\}}{p_i(t)}\right|^2 = \int_{A_i}\frac{y_i^2(t)\exp\left\{2\beta_1^T x_i(t)\right\}}{p_i(t)}dt + \int_{A_i}\int_{A_i}\frac{y_i(t)y_i(s)}{p_i(t)p_i(s)}\exp\left\{\beta_1^T\left(x_i(t)+x_i(s)\right)\right\}dtds,$$

and

$$\begin{aligned} E\left|\sum_{t\in D}\frac{|y_i(t)|\exp\left\{\beta_1^T x_i(t)\right\}}{p_i(t)}\right|^3 &= \int_{A_i}\frac{y_i^3(t)\exp\left\{3\beta_1^T x_i(t)\right\}}{p_i^2(t)}dt \\ &+3\int_{A_i}\int_{A_i}\frac{y_i(t)y_i(s)}{p_i(t)p_i(s)}\exp\left\{\beta_1^T\left(x_i(t)+x_i(s)\right)\right\}dtds \\ &+\int_{A_i}\int_{A_i}\int_{A_i}\frac{y_i(t)y_i(s)y_i(t)y_i(u)}{p_i(t)p_i(s)p_i(u)}\exp\left\{\beta_1^T\left(x_i(t)+x_i(s)+x_i(u)\right)\right\}dtds \\ &= O(|A_i|^3). \end{aligned}$$

Under Assumption 2, $y_i(t_{ij})$ is bounded above, and $E\left(\hat{v}_i(A_i,\beta_1)\right)^3 = O(|A_i|^3)$.

Then, since we know that $Var(\hat{v}_i(A_i,\beta_1)\hat{T}_i(\hat{w}_i\left(A_i,\beta\right)) = O(|A_i|^2)$, we have

$$\begin{aligned} \frac{\sum_{i=1}^n E\left|\hat{v}_i(A_i,\beta_1)\hat{T}_i(\hat{w}_i\left(A_i,\beta\right))\right|^3}{\left(\sum_{i=1}^n Var(\hat{v}_i(A_i,\beta_1)\hat{T}_i(\hat{w}_i\left(A_i,\beta\right)\right)^{3/2}} &\le \frac{\sum_{i=1}^n E\left|\hat{v}_i(A_i,\beta_1)\right|^3}{\left(\sum_{i=1}^n Var(\hat{v}_i(A_i,\beta_1)\hat{T}_i(\hat{w}_i\left(A_i,\beta\right)\right)^{3/2}} \\ &= \frac{\sum_{i=1}^n O(|A_i|^3)}{\left(\sum_{i=1}^n O(|A_i|^2)\right)^{3/2}} \\ &\le \frac{n\times O(\max|A_i|^3)}{n^{3/2}\times O(\min|A_i|^3)}\ . \qquad (A.5) \end{aligned}$$

By Assumption 3, which ensures that all the study intervals are in a well-behaved fashion, $(A.5)$ will converge to zero, as $n\to\infty$. Hence, Liapounov's condition (Ash 1972, p. 337) holds with $\delta = 1$. The remaining terms of $\tilde{\Psi}_n(\theta)$ can be proven in a similar approach. Therefore, $\tilde{\Psi}_n(\theta_0)$ converges to a Gaussian random vector with finite variance.

2.9 Appendix B

The first part of this Appendix displays the variance components of the estimating function $\Sigma_n(\theta) = var_\theta\left\{\tilde{\Psi}_n(\theta)\right\}$ on which some terms involve the variance and covariance terms between $\hat{v}_i(A_i,\beta)$ and $\hat{w}_i(A_i,\beta)$. The calculation details will also be shown in this part. The second part of this Appendix will demonstrate the uniform convergence of $\tilde{\Psi}_n^{(1)}(\theta)$.

The elements of $\Sigma_n(\theta)$

The elements of $\Sigma_n(\theta)$ comprises of the following variance and covariance terms :

$$\begin{aligned} cov(\hat{v}_i(A_i,\beta_1),\hat{w}_i(A_i,\beta)) &= \left[\beta_2^T Var(\sum_{j=1}^{N(A_i)} x_i(t_{ij})) + cov\left(\hat{\Lambda}_i^{(1)}(A_i,\beta_1),\hat{\Lambda}_i(A_i,\beta_2)\right)\right] \\ &\quad - \left[\beta_1^T Var(\sum_{j=1}^{N(A_i)} x_i(t_{ij})) + cov\left(\hat{\Lambda}_i^{(1)}(A_i,\beta_1),\hat{\Lambda}_i(A_i,\beta_1)\right)\right]. \end{aligned}$$

$$\begin{aligned} cov(\hat{v}_i(A_i,\beta_2),\hat{w}_i(A_i,\beta)) &= \left[\beta_2^T Var(\sum_{j=1}^{N(A_i)} x_i(t_{ij})) + cov\left(\hat{\Lambda}_i^{(1)}(A_i,\beta_2),\hat{\Lambda}_i(A_i,\beta_2)\right)\right] \\ &\quad - \left[\beta_1^T Var(\sum_{j=1}^{N(A_i)} x_i(t_{ij})) + cov\left(\hat{\Lambda}_i^{(1)}(A_i,\beta_2),\hat{\Lambda}_i(A_i,\beta_1)\right)\right]. \end{aligned}$$

and

$$cov(\hat{v}_i(A_i,\beta_1),\hat{v}_i(A_i,\beta_2)) = Var(\sum_{j=1}^{N(A_i)} x_i(t_{ij})) + cov\left(\hat{\Lambda}_i^{(1)}(A_i,\beta_1),\hat{\Lambda}_i^{(1)}(A_i,\beta_2)\right) \qquad (B.1).$$

The term $cov\left(\hat{\Lambda}_i^{(1)}(A_i,\beta_1),\hat{\Lambda}_i^{(1)}(A_i,\beta_2)\right)$ in $(B.1)$ can be calculated by a straighforward extension of Theorem 2 of Cordy (1993)

$$\begin{aligned} cov\left(f(t),z(t)\right) &= Cov\left(\sum_{t\in D_i}\frac{f(t)}{p(t)},\sum_{s\in D_i}\frac{z(s)}{p(s)}\right) \\ &= \int_{A_i}\frac{f(t)f(t)^T}{p(t)}dt + \int_{A_i}\int_{A_i} f(t)z(s)^T\left(\frac{p(s,t)-p(t)p(s)}{p(t)p(s)}\right)dsdt. \end{aligned}$$

The uniform convergence of $\tilde{\Psi}_n^{(1)}(\theta)$

To prove the uniform convergenge of $\tilde{\Psi}_n^{(1)}(\theta)$, first consider the following terms:

$$\begin{aligned} v_i^{(1)}(A_i,\beta) &= \frac{\partial v_i(A_i,\beta)}{\partial \beta} = -\frac{\partial}{\partial \beta}\Lambda_i^{(1)}(A_i,\beta) \\ &= -\Lambda_i^{(2)}(A_i,\beta). \qquad (B.2) \end{aligned}$$

Then, consider the first element of the derivative of $\tilde{\Psi}_n^{(1)}(\theta)$:

$$\frac{\partial \tilde{\Psi}_n^{(1)}(\theta)}{\partial \beta_1} \leq 2n^{-1}\pi^2 \sum_{i=1}^{n} s_i(A_i,\beta_1)\left(\hat{v}_i^{(1)}(A_i,\beta_1) + \hat{v}_i^3(A_i,\beta_1)\right),$$

where

$$s_i(A_i,\beta) = \exp\left\{2\left(\sum_{j=1}^{N_i}\beta_1^T x_i(t_{ij}) - \hat{\Lambda}_i(A_i,\beta_1)\right)\right\}.$$

By Assumption 2 - 5, $\hat{v}_i^{(1)}(A_i,\beta_1)$ and $\hat{v}_i^3(A_i,\beta_1)$ are bounded and can be proved in the similar manner as in the proof of Lemma 3. So, $\frac{\partial \tilde{\Psi}_n^{(1)}(\theta)}{\partial \beta_1}$are also bounded above by some positive value. The remaining terms of the derivative of $\tilde{\Psi}_n^{(1)}(\theta)$ can be proved in the similar manner. Therefore, $\tilde{\Psi}_n^{(1)}(\theta)$ converges unifromly to a nonstochastic limit as $n \to \infty$.

Chapter 3

Mixture Mixed Point Processes

3.1 Introduction

Investigation of the impact of time-varying covariates on Poisson process intensity can be implemented using Ecological Momentary Assessment (EMA), which is a method of data collection that often uses electronic devices to gather information about human behavior in the subject's environment. This method of data collection reduces recall bias and increases the "ecological validity" of the data by asking subjects questions regarding current state and by not placing subjects in an artificial labortary environment (Shiffman and Stone 1998). EMA has been used to study numerous addictive behavioral events, including alcohol (Collins et al. 2007), eating disorders (Smyth et al. 2001), the use of tobacco (Shiffman 2005), and stress coping (Stone et al. 1998).

In an EMA cigarette smoking cessation study, Shiffman et al. (2002) instructed 304 smokers, desiring to quit smoking, to record each cigarette on a Personal Digital Assistant (PDA) both before and after a designated quit time. The electronic diary prompted subjects to answer questions regarding their current mood and environment for randomly selected cigarettes, and at times selected according to a probability-based sampling design. By analyzing these data, we hope to describe the impact of the covariates of mood and environment on cigarette consumption rates in smokers. This understanding may lead to improvements in intervention programs for smoking cessation.

Schwartz and Stone (2007) recommended Hierarchical Linear Models (HLMs) for the analysis of EMA data, since these models can capture the variability among subjects in their responses to time-varying and time-invariant covariates. HLMs, nevertheless, assume that the response is normally distributed, and are not suitable to modeling the timing of repeated discrete events over an interval. To relax the normality assumption of the response in HLMs, Lee and Nelder (1996) suggested Hierarchical Generalized Linear Models (HGLMs), which allow the response and random components to have any distribution in the exponential family. Generalized Linear Mixed Models (GLMMs) (Breslow & Clayton 1993) are a particular type of HGLMs, where the random components are assumed to follow a normal distribution.

The random times of repeated behavioral events are customarily modeled as a temporal point process with an inhomogeneous intensity surface (Duan et al. 2009). For example Rathbun et al (2007) fit a Poisson point process of to EMA smoking data whose log intensity is a linear function of time-varying covariates. However, their model implicitly assumes that all subjects have equal responses to changes in time-varying covariates. In the context of the motivating dataset, some subjects may have inherent genetic factors that lead to differential response to time-varying covariates. If we can identify what factors impact individual smoking behaviors, those specific factors may be focused for more effective individual based smoking cessation treatment. To achieve this objective, mixed-effect models for the impact of time-varying covariates on the recurrent events data may be considered.

Results from Shiffman and Rathbun (2011) further suggest that there are interactions between gender and time-varying covariates, suggesting that gender-specific treatments may be more effective for smoking cessation. Gender, however, is not the only factor that might affect subjects' responses to time-varying covariates. More effective individual-based treatments for smoking cessation might be obtained if subjects can be classified into groups with like responses to time-varying covariates. This suggests the implementation of finite mixture model for random effects describing the impact of time-varying covariates on recurrent events data.

Finite mixture models with regression structure have been widely used in many fields such as epidemiology (Hardelid et al. 2008), medicine (Kelvin et al. 2003), and physiology (Colder et al. 2002). An example of fitting mixture of Guassian regression model can be found in Grun and Leisch (2006). Finite mixtures of generalized linear models (Jansen 1993; Dietz and Bohning 1997) have proven useful for modeling data arising from a heterogeneous population. However, these papers considered regression structure in the linear predictor only. Recently, Hall and Wang (2005) developed a new class of regression models consisting of two component mixtures of generalized linear mixed effect models (two-component GLMMs). This class can be viewed as an extension of finite mixtures of generalized linear models obtained by adding random effects to each component.

Mixed effect models (Breslow and Clayton 1993) are often used to handle correlation arising in longitudinal or clustered data. The marginal distribution for the mixed effect model of recurrent events can be obtained by integrating the joint distribution of the events and the random effects with respect to the random effects. However, closed form solutions for this marginal distribution can not be obtained except for the models with gamma frailties (Lawless 1987). Hierarchical likelihood (h-likelihood), which does not require evaluation of high dimensional integrals, has been proposed for statistical inference from model with fixed and random effects by Lee and Nelder (1996). This approach has been further studied in Lee and Nelder (2001) and Lee et al. (2006) and for non-linear mixed effects model by Noh and Lee (2008). Ha et al. (2001, 2003) also extend the use of h-likelihood estimation to model time-invariant and time-varying covariates on the life-times to events. The main idea of h-likelihood is to treat random effects as parameters and to find the estimates of all the parameters by maximizing a log likelihood function which is conditional on the random effects minus a penalty term. Penalized likelihood (Green 1987) has been further applied for function estimation. This approach provides a simple framework since it does not require numeric integration.

Taking inspiration from the techniques of GLMMs, h-likelihood ,and the EM algorithm (Dempster et al. 1977) for finite mixture model estimation, we will modify the models of Rathbun et al. (2007) to handle both correlation and heterogeneity simultaneously. In this paper, we will describe a mixture mixed-effects version of a modulated Poisson process, where the intensity will be modified to allow the covariate coefficients to vary randomly among subjects, with respect to the effects of time-covariates, from which cluster of subjects showing similar smoking behaviors may be discovered. The random effects terms are assumed to vary according to a mixture of Normal distributions with different means.

The paper is organized as follows: we introduce the concept of Mixture Mixed point process in section 2. In section 3, the Hierarchical likelihood of Mixture Mixed point process will be briefly described. In section 4, estimation of parameters using random assessments of covariates will be addressed. Section 5 will give the details of the estimation of variance component. Section 6 will give the details of the EM algorithm for mixture mixed Point process. Section 7 will discuss the parameter estimation using random assessment of the covariates. Section 8 will provide some asymptotic properties of the proposed estimators. Section 9 will illustrate our methods using data from an EMA study of smoking.

3.2 Mixture Mixed Point Process

A temporal point process is a random process whose realizations consist of the times of discrete events scattered in time (Brillinger et al 2002). A point process has random counting measure N, where $N(A)$ is the number of events in any Borel set $A \subset R$. This measure can be described by the corresponding intensity,

$$\lambda(t) = \lim_{\delta \to \infty} \frac{E\left(N\left[t, t+\delta\right]\right)}{\delta},$$

where $N\left[t, t+\delta\right]$ is the number of events in the interval $[t, t+\delta]$. For a Poisson process (Cox 1972), $N(A)$ is Poisson distributed with mean $\Lambda(A) = \int_A \lambda(t)dt$ for all intervals A.

The modulated Poisson process (Cox 1972) is a special case of a Poisson process where the intensity function takes the log-linear form

$$\lambda(t; \alpha) = \exp\left\{\alpha^T x(t)\right\}, \quad t \in A. \tag{3.1}$$

α is a model parameter and $x(t)$ is a $r \times 1$ vector of time-varying covariates. This model assumes that the responses in all subjects' event rates due to the covariates are identical. To create more flexible model that allows variability among subjects with respect to covariate effects, we consider a mixed effects version of (3.1). Conditional on $q \times 1$ vector of random effects β_i, the intensity for each subject i takes the form

$$\lambda_i(t; \alpha, \beta_i) = \exp\left\{\alpha^T x_i(t) + \beta_i^T z_i(t)\right\}, \quad t \in A_i \tag{3.2}$$

where A_i is the study interval for subject i, α is a $r \times 1$ vector of fixed effects, β_i is a $q \times 1$ vector of random effects that vary by subject, and $z(t)$ is a subset of $x(t)$. We assume further that each subject can be classified into one of k groups or categories with like responses to time-varying covariates. Let p_j denote the probability that a subject belongs to group j. The random effects β_i are assumed to be sampled from a mixture of k normal distributions with different means and common covariance matrix Σ. The marginal density of β_i can be written in the form

$$f(\beta_i; \mu, \Sigma) = \sum_{j=1}^{k} p_j f(\beta_i; \mu_j, \Sigma),$$

where $f(\beta_i; \mu_j, \Sigma)$ is the normal distribution with mean μ_j, and covariance matrix Σ.

In the marginal maximum likelihood procedure (MML), random effect β_i are integrated out and only fixed effects α and the variance component Σ are retained in the maximized function. For each subject i, N_i recurrent events are observed at times $t \in X_i$ in the Borel set A_i, where X_i is the collection of event times. Then, the marginal likelihood of Mixture Mixed Point process for $\theta = (\alpha, \mu, \Sigma)^T$ can be written as

$$L_M(\theta; N) = \prod_{i=1}^{n} \int_{R^q} f\left(\alpha; N_i | \beta_i\right) f(\beta_i; \mu, \Sigma) d\beta_i, \tag{3.3}$$

where

$$f\left(\alpha; N_i | \beta_i\right) = \exp\left\{\alpha^T \sum_{t \in N_i} x_i(t) + \beta_i^T \sum_{t \in N_i} z_i(t) - \Lambda_i(\alpha, \beta_i)\right\},$$

is the Janossy density (see Borodin and Soshnikov 2003), and

$$\Lambda_i(\alpha, \beta_i) = \int_{A_i} \exp\left\{\alpha^T x_i(t) + \beta_i^T z_i(t)\right\} dt,$$

is the integrated intensity.

Maximum likelihood estimation requires maximization of (3.3) with respect to θ. However, the evaluation of this integral typically proves difficulty since these models often are nonlinear with respect to the random effect, and thus there is no closed form solution for the marginal distribution. Various techniques have been proposed to solve this problem. Numerical integration may be applied (e.g. Pinheiro and Bates 1995; Liu and Pierce 1994), but it is not suitable for more than two or three random effects owing to the curse of dimensionality (Bellman 1957). Monte Carlo methods have been also suggested, but they are computationally intensive and might not be practical for large EMA data sets. Hierarchical likelihood has been considered to avoid the difficult integration of (3.3) by direct maximization of the approximate joint posterior. This approach has been considered for the HGLMs where the mixing distribution is conjugate, and for the recurrent events by Ha and Neustifer et al. (2012), but has not been considered for mixture models.

3.3 Hierarchical Likelihood Estimation

Hierarchical Linear Models (HLMs) are often used for analyzing EMA data, as they allow researchers to seperate within and between subject variation. Ha et al. (2001) have extended HLMs to recurrent events data. Hierarchical Generalized Linear Models (HGLMs) are a further extension of Generalized Linear Models (GLMs) that are used to model multi-level data where the response may be clustered (Lee & Nelder 1996). Traditionally, the random components are conjugate to the distribution of the response, and the GLM uses the canonical link. Lee and Nelder (1996) suggest the h-likelihood as an alternative approach of parameter estimation when the random effect is a function of the canonical parameter. However, when it is not, it is instead Extended Likelihood estimation.

Since $\lambda_i(t; \alpha, \beta_i)$ has the general form of log-linear on both fixed and random effects , β_i is the canonical parameter of the response (Poisson) distribution. Then the hierarchical likelihood (Lee and Nelder 1996) for the mixture mixed Poisson process model is defined as

$$\begin{aligned} h(\mu, \alpha, \beta, \Sigma; N) &= \log \left(\prod_{i=1}^{n} f(\alpha, N_i|\beta_i) f(\beta_i; \mu, \Sigma) \right) \\ &= l(\alpha; N|\beta) + l(\mu, \Sigma; \beta), \end{aligned}$$

where in the current context

$$\begin{aligned} l(\alpha; N|\beta) &= \sum_{i=1}^{n} \{\log f(\alpha; N_i|\beta_i)\} \\ &= \sum_{i=1}^{n} \left\{ \alpha^T \sum_{t \in N_i} x_i(t) + \beta_i^T \sum_{t \in N_i} z_i(t) - \Lambda_i(\alpha, \beta_i) \right\}, \end{aligned}$$

is the logarithm of the distribution of the replicated point process model given the realization of the random effects β, and $l(\mu, \Sigma; \beta) = \sum_{i=1}^{n} \log f(\beta_i; \mu, \Sigma)$ is the logarithm of the distribution of the random effects.

The next section will show how to obtain the maximum h-likelihood estimation (MHLE) by jointly maximizing Penalized Quasi-Likelihood (Green 1987) of $h(\mu, \alpha, \beta, \Sigma; N)$.

3.4 Fixed and Random Effect Estimation

Lee and Nelder (1996) introduced hierarchical generalized linear models (HGLMs) by including random components in the linear predictor with arbitrary distributions in generalized linear models (GLMs). Their main idea is using the joint likelihood of response and random effects, which was called hierarchical likelihood (H-likelihood), to substitute the marginal likelihood for inference of HGLMs. This section will demonstrate how the estimating equations for both fixed and random effects can be derived using a hierarchical likelihood. This method has an advantage that integrating out of random effects is not required.

For simplicity in calculation and proof, the number of categories is initially set to be two. Let p denotes the probability that a subject belongs to the first group. Since group membership is unknown, the contribution of subject i to the marginal likelihood from incomplete data is

$$\int_{R^q} f(\alpha; N_i|\beta_i)\left\{pf(\beta_i;\mu_1,\Sigma) + (1-p)f(\beta_i;\mu_2,\Sigma)\right\} d\beta_i,$$

The main issue here is missing information regarding the realization of random effects and group membership of subjects. With respect to the latter, we define the indicator $y_i = 1$ when subject i comes from the first category with probability p. Then conditional on y_i, the distribution of the data for individual i is

$$\int_{R^q} f(\alpha; N_i|\beta_i)\left[f(\beta_i;\mu_1,\Sigma)\right]^{y_i}\left[f(\beta_i;\mu_2,\Sigma)\right]^{1-y_i} d\beta_i,$$

The marginal density of y_i is $p^{y_i}(1-p)^{1-y_i}$, where $p \in (0,1)$. If we could observe y_i

$$\int_{R^q} f(\alpha; N_i|\beta_i)\left[pf(\beta_i;\mu_1,\Sigma)\right]^{y_i}\left[(1-p)f(\beta_i;\mu_2,\Sigma)\right]^{1-y_i} d\beta_i,$$

is the complete-data marginal likelihood.

To obtain estimating equations for $\theta = (\alpha, \mu_1, \mu_2, p, \Sigma)^T$, we consider methods similar to Breslow and Clayton's (1993) version of the Penalized Quasi-Likelihood. (PQL). First, the complete-data marginal likelihood may be written as

$$L_c(\theta; N) = (2\pi)^{-\frac{nq}{2}}|\Sigma|^{-\frac{nq}{2}}\prod_{i=1}^{n}\left(p\int_{R^q} e^{-j_i(\alpha,\beta_i,\Sigma,\mu_1)}d\beta_i\right)^{y_i}\left((1-p)\int_{R^q} e^{-j_i(\alpha,\beta_i,\Sigma,\mu_2)}d\beta_i\right)^{1-y_i} \tag{3.4}$$

where

$$j_i(\alpha,\beta_i,\Sigma,\mu) = -\left\{\alpha^T\sum_{t\in N_i} x_i(t) + \beta_i^T\sum_{t\in N_i} z_i(t)\right\} + \Lambda_i(\alpha,\beta_i) + \frac{1}{2}(\beta_i-\mu)^T\Sigma^{-1}(\beta_i-\mu).$$

The difficulty with this likelihood is that the random effects enter nonlinearly. Consequently, there is no closed form expression for this marginal likelihood or the corresponding score equations. However, Laplace's method (Tierney & Kadane, 1986) may be applied to the complete-data marginal log-likelihood to obtain an approximation.

For a function $F(u)$, define $D_u F(u) = \frac{\partial}{\partial u} F(u)$ and $D_{uu^T} F(u) = \frac{\partial}{\partial u \partial u^T} F(u)$. Then, the Laplace approximation of (3.4) can be written as

$$\begin{aligned}
\log L_c(\theta; N) \quad \simeq \quad & -\frac{nq}{2} \log|\Sigma| + \sum_{i=1}^{n} y_i \left(\log(p) - \frac{1}{2} \log \left| D_{\beta_i \beta_i^T} j_i(\alpha, \tilde{\beta}_i, \Sigma, \mu_1) \right| - j_i(\alpha, \tilde{\beta}_i, \Sigma, \mu_1) \right) \\
& + \sum_{i=1}^{n} (1 - y_i) \left(\log(1-p) - \frac{1}{2} \log \left| D_{\beta_i \beta_i^T} j_i(\alpha, \tilde{\beta}_i, \Sigma, \mu_2) \right| - j_i(\alpha, \tilde{\beta}_i, \Sigma, \mu_2) \right) \\
& + c \qquad (3.5)
\end{aligned}$$

where the constant c does not depend on any model parameter, and $\tilde{\beta}_i$ denotes the solution to

$$D_{\beta_i} j_i(\alpha, \beta_i, \Sigma, \mu) = -\sum_{t \in N_i} z_i(t) + D_{\beta_i} \Lambda_i(\alpha, \beta_i) + \Sigma^{-1} (\beta_i - \mu) = 0,$$

which minimizes $j_i(\alpha, \beta_i, \Sigma, \mu)$. Differentiating with respect to β_i again, we get

$$D_{\beta_i \beta_i^T} j_i(\alpha, \beta_i, \Sigma, \mu) = D_{\beta_i \beta_i^T} \Lambda_i(\alpha, \beta_i) + \Sigma^{-1}.$$

Then, substituting into (3.5), we obtain the approximation

$$\begin{aligned}
\log L_c(\theta; N) \quad \simeq \quad & \sum_{i=1}^{n} y_i \left\{ \log(p) + r_i(\alpha, \tilde{\beta}_i, \Sigma, \mu_1) \right\} \\
& + \sum_{i=1}^{n} (1 - y_i) \left\{ \log(1-p) + r_i(\alpha, \tilde{\beta}_i, \Sigma, \mu_2) \right\} + c,
\end{aligned}$$

where

$$\begin{aligned}
r_i(\alpha, \beta_i, \Sigma, \mu) \quad = \quad & -\frac{1}{2} \log \left| I + \Sigma D_{\beta_i \beta_i^T} \Lambda_i(\alpha, \beta_i) \right| \\
& + \sum_{t \in N_i} \left\{ \alpha^T x_i(t) + \beta_i^T z_i(t) \right\} - \Lambda_i(\alpha, \beta_i) \\
& - \frac{1}{2} (\beta_i - \mu)^T \Sigma^{-1} (\beta_i - \mu).
\end{aligned}$$

Assuming that the first term in $r_i(\alpha, \beta_i, \Sigma, \mu)$ varies slowly as a function of α (Breslow & Clayton 1993), we can estimate (α, β_i, μ) by jointly maximizing the following Green's (1987) Penalized Quasi-Likelihood (PQL), which may also be regarded as the complete-data

H-likelihood:

$$h_c(\mu,\alpha,\beta,\Sigma;N) = \sum_{i=1}^{n}\left\{\sum_{t\in N_i}\left\{\alpha^T x_i(t)+\beta_i^T z_i(t)\right\}-\Lambda_i(\alpha,\beta_i)\right\}+\sum_{i=1}^{n}\left\{y_i\log(p)+(1-y_i)\log(1-p)\right\}$$
$$-\sum_{i=1}^{n}\left\{\frac{y_i}{2}(\beta_i-\mu_1)^T\Sigma^{-1}(\beta_i-\mu_1)+\frac{(1-y_i)}{2}(\beta_i-\mu_2)^T\Sigma^{-1}(\beta_i-\mu_2)\right\}. \quad (3.6)$$

Lee et al. (2007) also show that the marginal (log-) likelihood can be typically approximated from H-likelihood via

$$\log L(\theta;N)\simeq\log\left(\int_{R^q}\exp(h(\theta;N))d\beta\right).$$

Differentiating (3.6) with respect to α, β_i, μ_1, and μ_2 respectively, gives the h-score equations

$$q_{MHLE}(\alpha;\beta)=\sum_{i=1}^{n}\left\{\sum_{t\in N_i}x_i(t)-D_\alpha\Lambda_i(\alpha,\beta_i)\right\}=0,$$

$$q_{MHLE}(\beta_i;\alpha,\mu,\Sigma)=\sum_{t\in N_i}z_i(t)-D_{\beta_i}\Lambda_i(\alpha,\beta_i)-y_i\Sigma^{-1}(\beta_i-\mu_1)-(1-y_i)\Sigma^{-1}(\beta_i-\mu_2)=0,$$

$$q_{MHLE}(\mu_1;\beta_i,\Sigma)=\sum_{i=1}^{n}\left\{y_i(\beta_i-\mu_1)\Sigma^{-1}\right\}=0,$$

and

$$q_{MHLE}(\mu_2;\beta_i,\Sigma)=\sum_{i=1}^{n}\left\{(1-y_i)(\beta_i-\mu_2)\Sigma^{-1}\right\}=0.$$

The obtained estimators derived from these score equations are termed as maximum H-likelihood estimators (MHLEs). Solving the last two equations yields the MHLE for μ_1, and μ_2

$$\hat{\mu}_1=\frac{\sum_{i=1}^{n}y_i\beta_i}{\sum_{i=1}^{n}y_i},$$

and

$$\hat{\mu}_2=\frac{\sum_{i=1}^{n}(1-y_i)\beta_i}{\sum_{i=1}^{n}(1-y_i)}.$$

Since $y_i's$ are not typically observed, the EM algorithm can be used to estimate model parameters. However, this is not strictly an EM algorithm since the E-step of the proposed algorithm does not involve estimation $E\left[\partial h(\theta;N)/\partial\theta|N,\hat{\theta}^{(p)}\right]$, but rather the computation of an approximation to this conditional expectation as in the hierarchical likelihood approach of Lee and Nelder (1996). The EM type algorithm can be applied to the alternation between the estimating equations, in which

(i) Estimating $y_i's$ in the E-step

(ii) Estimating α, β, μ, and Σ in the M-step

Section 3.6 will show complete details of an application of EM algorithm for the parameter estimation.

3.5 Estimation of Variance Component

Now consider the construction of estimating equations for the variance components Σ. First, differentiating the h-score functions with respect to θ, the negative Hessian matrix can be written as

$$H_n(\mu,\alpha,\beta,\Sigma;N) = \begin{bmatrix} W_n(\alpha,\beta) & V_n^T(\alpha,\beta) & 0 \\ V_n(\alpha,\beta) & U_n(\alpha,\beta,\Sigma) & T_n^T(\Sigma) \\ 0 & T_n(\Sigma) & P_n(\Sigma) \end{bmatrix}, \tag{3.7}$$

where

$$W_n(\alpha,\beta) = \sum_{i=1}^{n} D_{\alpha\alpha^T}\Lambda_i(\alpha,\beta_i),$$

$$V_n(\alpha,\beta) = \begin{bmatrix} D_{\alpha\beta_1}\Lambda_i(\alpha,\beta_1) \\ \vdots \\ D_{\alpha\beta_n}\Lambda_i(\alpha,\beta_n) \end{bmatrix},$$

$$T_n(\Sigma) = \begin{bmatrix} y_1\Sigma^{-1} & \dots & y_n\Sigma^{-1} \\ (1-y_1)\Sigma^{-1} & \dots & (1-y_n)\Sigma^{-1} \end{bmatrix},$$

$$P_n(\Sigma) = \begin{bmatrix} \sum_{i=1}^{n} y_i\Sigma^{-1} & 0 \\ 0 & \sum_{i=1}^{n}(1-y_i)\Sigma^{-1} \end{bmatrix},$$

and the block diagonal matrix

$$U_n(\alpha, \beta, \Sigma) = diag\left[D_{\beta_i\beta_i^T}\Lambda_i(\alpha, \beta_i) + \Sigma^{-1}\right].$$

For estimation of the variance components, we suppose further that $\Sigma = \Sigma(\gamma)$, which is a function of parameter γ. Lee and Nelder (1996) considered the following adjusted profile h-likelihood for the estimation of variance components in Σ :

$$h_A\left(\mu, \alpha, \beta, \Sigma(\gamma); N\right) = h\left(\mu, \alpha, \beta, \Sigma(\gamma); N\right) - \frac{1}{2}\log\left|H(\mu, \alpha, \beta, \Sigma(\gamma); N)\right|, \tag{3.8}$$

evaluated at $\hat{\phi} = \left(\hat{\alpha}, \hat{\mu}, \hat{\beta}\right)$, where $H_n(\mu, \alpha, \beta, \Sigma(\gamma); N)$ is defined as in (3.7). Differentiating (3.8) with respect to γ yields the score equations for γ

$$\begin{aligned}
D_\gamma h_A(\alpha, \mu, \beta, \Sigma(\gamma))|_{\hat{\phi}} \quad = \quad & -\frac{n}{2}\text{trace}\Sigma^{-1}(\gamma)\frac{\partial}{\partial\gamma}\Sigma(\gamma) \\
& +\sum_{i=1}^{n}\frac{y_i}{2}\left(\hat{\beta}_i - \hat{\mu}_1\right)^T\Sigma^{-1}(\gamma)\frac{\partial}{\partial\gamma}\Sigma(\gamma)\Sigma^{-1}(\gamma)\left(\hat{\beta}_i - \hat{\mu}_1\right) \\
& +\frac{1}{2}\sum_{i=1}^{n}(1-y_i)\left(\hat{\beta}_i - \hat{\mu}_2\right)^T\Sigma^{-1}(\gamma)\frac{\partial}{\partial\gamma}\Sigma(\gamma)\Sigma^{-1}(\gamma)\left(\hat{\beta}_i - \hat{\mu}_2\right) \\
& -\frac{n}{2}\text{trace}H^{-1}(\mu, \alpha, \beta, \Sigma(\gamma); N)\frac{\partial}{\partial\gamma}H(\mu, \alpha, \beta, \Sigma(\gamma); N).
\end{aligned}$$

These score equations can be solved using the Newton-Raphson method. The details of Newton-Raphson method for the estimation of variance component can be found in Appendix B.

3.6 EM-Algorithm for Mixture Mixed Point Process

In this section, the approximated EM-algorithm of Maximum h-likelihood estimation (MHLE) approach for a mixture Poisson process will be outlined. Suppose the random variable y_i can be observed, and β_i is treated as an unknown parameter to be estimated.

Then, the complete h-likelihood for the mixture mixed Point process can be written as

$$
\begin{aligned}
h_c(\mu,\alpha,\beta,\Sigma;N) &= \sum_{i=1}^{n}\left(\sum_{t\in N_i}\alpha^T x_i(t) - \Lambda_i(\alpha,\beta_i) - \frac{1}{2}\log|\Sigma|\right) \\
&\quad + \sum_{i=1}^{n} y_i\left(\log(p) - \frac{1}{2}(\beta_i-\mu_1)^T\Sigma^{-1}(\beta_i-\mu_1)\right) \\
&\quad + \sum_{i=1}^{n}(1-y_i)\left(\log(1-p) - \frac{1}{2}(\beta_i-\mu_2)^T\Sigma^{-1}(\beta_i-\mu_2)\right), \qquad (3.9)
\end{aligned}
$$

To estimate $\theta = (\mu,\alpha,\beta,p,\Sigma)^T$, one could consider y_i in (3.9) as missing data as in Dempster et al. (1977), and maximize (3.9) by an EM Algorithm. Let $\theta^{(r)}$ denotes the value for θ on the rth iteration of the EM algorithm, with $\theta^{(0)}$ be the initial specified value. The EM algorithm applies estimation (E-step) and maximization (M-step) steps to obtain maximum h-likelihood estimates of model parameters. In the current application, the E-step is

E-STEP: The conditional expectations of the complete data log likelihood based on observed data,$E\left(y_i|\theta^{(r)}\right)$ is

$$
w_i^{(r)} = \frac{p^{(r)}\exp\left\{-\frac{1}{2}(\beta_i^{(r)}-\mu_1^{(r)})^T\Sigma^{-1}(\beta_i^{(r)}-\mu_1^{(r)})\right\}}{p^{(r)}\exp\left\{-\frac{1}{2}(\beta_i^{(r)}-\mu_1^{(r)})^T\Sigma^{-1}(\beta_i^{(r)}-\mu_1^{(r)})\right\} + (1-p^{(r)})\exp\left\{-\frac{1}{2}(\beta_i^{(r)}-\mu_2^{(r)})^T\Sigma^{-1}(\beta_i^{(r)}-\mu_2^{(r)})\right\}}.
$$

M-STEP: The parameter for fixed effect α is estimated by finding α that solves the estimating equation

$$
\sum_{i=1}^{n}\left\{\sum_{t\in N_i} x_i(t) - \int_{A_i} x_i(t)\exp\left\{\alpha^T x_i(t) + \beta_i^T z_i(t)\right\}dt\right\} = 0
$$

The parameter vector β_i is estimated by solving

$$
\sum_{i=1}^{n}\left\{\sum_{t\in N_i} z_i(t) - \int_{A_i} z_i(t)\exp\left\{\alpha^T x_i(t) + \beta_i^T z_i(t)\right\}dt - w_i\Sigma^{-1}(\beta_i-\mu_1) - (1-w_i)\Sigma^{-1}(\beta_i-\mu_2)\right\} = 0.
$$

The estimating equation for the variance component is

$$
\begin{aligned}
&-\frac{n}{2}\mathrm{trace}\Sigma^{-1}(\gamma)\frac{\partial}{\partial\gamma}\Sigma(\gamma) + \sum_{i=1}^{n}\frac{w_i}{2}\left(\hat\beta_i-\hat\mu_1\right)^T\Sigma^{-1}(\gamma)\frac{\partial}{\partial\gamma}\Sigma(\gamma)\Sigma^{-1}(\gamma)\left(\hat\beta_i-\hat\mu_1\right) \\
&+\frac{1}{2}\sum_{i=1}^{n}(1-w_i)\left(\hat\beta_i-\hat\mu_2\right)^T\Sigma^{-1}(\gamma)\frac{\partial}{\partial\gamma}\Sigma(\gamma)\Sigma^{-1}(\gamma)\left(\hat\beta_i-\hat\mu_2\right) \\
&-\frac{1}{2}\mathrm{trace}H^{-1}(\mu,\alpha,\beta,\Sigma(\gamma);N)\frac{\partial}{\partial\gamma}H(\mu,\alpha,\beta,\Sigma(\gamma);N).
\end{aligned}
$$

All these fixed effect, α, random effects, β, and γ can be estimated by using Newton-Raphson method. Finally, the estimate of μ_1, μ_2, and p can be updated by

$$p^{(r+1)} = n^{-1} \sum_{i=1}^{n} w_i^{(r)},$$

$$\mu_1^{(r+1)} = \frac{\sum_{i=1}^{n} w_i^{(r)} \beta_i^{(r)}}{\sum_{i=1}^{n} w_i^{(r)}},$$

and

$$\mu_2^{(r+1)} = \frac{\sum_{i=1}^{n} (1 - w_i^{(r)}) \beta_i^{(r)}}{\sum_{i=1}^{n} (1 - w_i^{(r)})}.$$

3.7 Partially-Observed Covariates

The above procedures all require the evaluation of $\Lambda_i(A_i, \alpha, \beta_i)$ and its derivatives with respect to α and β_i. That is parameter estimating requires evaluation of integral of the form

$$g(A) = \int_A g(s) ds.$$

Since the integrands are functions of time-varying covariates, their evaluation requires that the values of varying covariates be known functions of time $t \in A$. This is generally not the case, since the covariates are often observed only at sample locations in many applications. Cordy (1993) suggested a design-unbiased estimator for $g(A)$, which may be obtained from a probability sample of points $\{u; u \in S\}$ in the study interval A, where S is the total set of such sampled "dummy" points. Suppose further that this is obtained from a known probability-based sampling with inclusion density

$$\pi(u) = \sum_{u \in S} f_u(u), \quad u \in A,$$

where $f_u(u)$ is the known marginal probability density function for site $u \in S$. Alternatively, the sites may be sampled according to point process with known conditional intensity $\pi(u)$.

The Horvitz-Thompson estimator is

$$\hat{g}(A) = \sum_{u \in S} \frac{g(u)}{\pi(u)}, \tag{3.10}$$

which has variance-covariance matrix

$$var\left\{\hat{g}(A)\right\} = \int_A \frac{g(s,\beta)g^T(s,\beta)}{\pi(s)} ds + \int_A \int_A g(s,\beta)g^T(u,\beta)\left(\frac{\pi(s,u) - \pi(s)\pi(u)}{\pi(s)\pi(u)}\right) dsdu.$$

where

$$\pi(u,t) = \sum_{u \neq t \in S} f_{ut}(u,t), \quad u,t \in \ A,$$

is the pairwise inclusion density, and $f_{ut}(u,t)$ is the joint probability density function for sites $u \neq t \in S$. The second term in $var\left\{\hat{g}(A)\right\}$ is zero if sites are sampled according to a point process with conditional intensity $\pi(u)$. Then, substituting (3.10) into the h-score functions of α and β_i yields the approximate h-score functions of the form

$$\hat{q}_{MHLE}(\alpha;\beta) = \sum_{i=1}^{n} \sum_{t \in N_i} x_i(t) - \sum_{u \in S_i} \frac{x_i(u) \exp\left\{\alpha^T x_i(u) + \beta_i^T z_i(u)\right\}}{\pi(u)},$$

and

$$\hat{q}_{MHLE}(\beta_i;\alpha,\mu,\Sigma) = \sum_{t \in N_i} z_i(t) - \sum_{u \in S_i} \frac{z_i(u) \exp\left\{\alpha^T x_i(u) + \beta_1^T z_i(u)\right\}}{\pi(u)} - y_1 \Sigma^{-1}\left(\beta_1 - \mu_1\right) - (1-y_1)\Sigma^{-1}\left(\beta_1 - \mu_2\right)$$

In the following section, we will show the asymptotic properties of MHLE of α, β_i, μ, and p obtained through solving these score functions.

3.8 Asymptotic Properties of MHLE

The maximum likelihood estimator is a well defined method since Barlett identities are satisfied. In the Appendix A, we demonstrate that the first and second Barrett identity also hold for the h-likelihood of Mixture Mixed point process. That is

$$E_\theta\left[\frac{\partial h\left(\phi;N\right)}{\partial \phi}\right] = 0,$$

and

$$E\left[\frac{\partial^2 h\left(\phi;N\right)}{\partial \phi \partial \phi^T}\right] = -E\left[\left(\frac{\partial h\left(\phi;N\right))}{\partial \phi}\right)\left(\frac{\partial h\left(\phi;N\right)}{\partial \phi}\right)^T\right],$$

where $h(\phi; N)$ is a h-likelihood of Mixture Mixed point process, and $\phi = (\mu, p, \alpha, \beta)^T$. The first identity ensures that the estimating equations for ϕ are unbiased under true values for ϕ. However, it is not sufficient for consistency of the estimators.

Next, we will study the asymptotic properties of the estimators from this h-likelihood. The h-score equation may be solved to obtain joint estimates $\left(\hat{\alpha}, \hat{\beta}\right)$ of the fixed effects and the random effects. There are some problems with this approach. The asymptotic distribution of the estimators of the fixed parameters and the consistency of MHLE have not been proved yet. However, under some conditions, the asymptotic distribution and the consistency of MHLE for the fixed effects can be achieved. Xia et al. (2006) demonstrated the strong consistency and convergence rate of MHLE in the Poisson-Gamma cases. Commenges et al. (2011) also proposed the asymptotic normal distribution of the MHLE for fixed effects in the HIV dynamics model.

Although, it is typical to let the number of subjects $n \to \infty$, two temporal asymptotic frameworks borrowed from spatial statistics when investigating the properties of the proposed estimators. Under infill asymptotic framework which is frequently applied in geostatistics (Stein, 1989, 1993, 1999), the size of the study region, $|A|$, is fixed for each subject while the number of covariate samples, m, tends to infinity. So, under infill asymptotic paradigm, $m \ / \ |A| \to \infty$. Under increasing domain asymptotics (Cressie 1991), domain of each subject increases, that is $|A| \to \infty$. Kutoyants (1984), Rathbun and Cressie (1994), Rathbun et al. (2007), and Waagepetersen (2008) have demonstrated the consistency and asymptotic normality of the parameter estimators for the Poisson process under increasing domain asymptotics. However, the size of sampling domain is normally fixed for each subject in practical application. So, we will demonstrate the asymptotic properties of the proposed estimator for the fixed and random effects obtained under increasing domain asymptotics under which $a_{\min} = \min\{|A_i|,\ i = 1, .., n\} \to \infty$, and $n \to \infty$.

Let $\tilde{\beta}_i$ denote a solution to

$$q_{MHLE}(\beta) = \sum_{t \in N_i} z_i(t) - D_{\beta_i}\Lambda_i(\alpha, \beta_i) - \Sigma^{-1}(\beta_i - \mu) = 0.$$

For GLMM, Lee and Nelder (1996) demonstrate that given α, μ, and Σ, $\tilde{\beta}_i$ is approximately equal to the conditional expectation of β_i given the data. We will demonstrate that this is also true for mixture mixed point processed. The following assumptions regarding the covariates and events are required to prove the properties of our proposed estimator.

Assumption 1. The subjects $i = 1, .., n$ are independently sampled, and all event times are independent of all assessment times, both within and between subject.

Assumption 2. There exists an upper bound M such that $|x_i(t)| < M$ and $|z_i(t)| < M$ for almost all $t \in R^d$ and all $i = 1, .., n$.

Assumption 3

$$n^{-1} \sum_{i=1}^{n} \int_{A_i} x_i(t) x_i(t)^T dt \to B,$$

as $n \to \infty$, where the minimum eigenvalue of B is strictly positive.

Assumption 4 Define

$$J_\infty(\alpha) = -\lim_{n\to\infty} E\left\{\frac{\partial}{\partial\alpha}\hat{q}_{MHLE}(\alpha, \beta, \mu; \Sigma)\right\},$$

The matrix $J_\infty(\alpha)$ is nonsingular.

Assumption 5 It is assumed that

$$\lim_{n\to\infty} \inf \pi_n(t) = c$$

for $c > 0$

Assumption 6 Define $a_{\max} = \max\{|A_i|,\ i = 1, .., n\}$. Assume that

$$\frac{a_{\min}}{a_{\max}} = O(1).$$

Assumption 7. It is assumed that

$$|A|^{-\frac{1}{2}} \{\hat{g}(A) - g(A)\} \to N(0, G_\infty),$$

in the distribution as $|A| \to \infty$, where G_∞ is a nonnegative definite matrix.

These assumptions are similar to those posed by Rathbun et al.(1994). Assumption 1 is required to ensure that there is no correlation among the subjects. Assumption 2 ensures that the estimating functions and their derivative with respect to the model parameters have bounded moments. This assumption is required for application of laws of large numbers to estimating equations. Assumption 3 is necessary to ensure that the derivative of the estimating functions are positive definite at the true values of the model parameters. Assumption 4 is required for the information matrix to be invertible. Assumption 5 ensures that design unbiased estimates may be obtained for the integrated intensity and its derivatives with respect to the model parameters (Cordy 1993). Assumption 6 prevents a finite number of individuals from dominating the remaining subjects. Moreover, we also require asymptotic normality for function g to prove the asymptotic property of the approximate h-score equations for partially observed covariates. Assumption 7 is satisfied for simple random sampling designs and when covariates are sampled from a point process with intensity $\pi(u)$. Thus, following from this central limit theorem, we have

$$\hat{g}(A) = g(A) + O_p\left(|A|^{1/2}\right).$$

Under these assumptions, it is possible the prove the following properties of our MHLE estimator

(a) $\tilde{\beta}_i = E\left(\beta_i|N_i\right) + O_p\left(|A_i|^{-1}\right).$

(b) $var\left(\beta_i|N_i\right) = U_i^{-1}(\alpha, \beta, \Sigma)\left(1 + O_p\left(|A_i|^{-1}\right)\right).$

(c) The difference between the approximate h-score equations and the h-score equations for fixed effect is $O_p(a_{\min}^{-1/2})$.

(d) Let $\hat{\alpha}_{MLE}$ be the maximum likelihood estimator, and let $\hat{\alpha}_{MHLE}$ be the MHLE for the fixed effects. For fixed sampling domain A,

$$(\hat{\alpha}_{MLE} - \hat{\alpha}_{MHLE}) = O_p(a_{\min}^{-1})$$

(e) Let $\hat{\alpha}_{MHLE}$ be the MHLE for the fixed effect. It can be shown that

$$(\hat{\alpha}_{MHLE} - \alpha) = O_p\left\{\max\left(n^{-\frac{1}{2}}, a_{\min}^{-1/2}\right)\right\}$$

(f) Let $\hat{\alpha}_{MOM}$ be the estimator for the fixed effect from method of moments. It can be shown that

$$(\hat{\alpha}_{MHLE} - \hat{\alpha}_{MOM}) = O_p(a_{\min}^{-1})$$

Properties (a), (b), and (d) are analogous to properties of section 5.3 of Lee and Nelder (1996). The proofs of these properties can be found in Appendix A.

To discuss the asymptotic variance of the fixed effect estimator, let $q_{MHLE}(\alpha; \tilde{\beta}) = D_\alpha h(\alpha, \mu, \Sigma, \beta; N)|_{\beta=\tilde{\beta}}$ denotes the fixed effect score equation for partially observed covartiates case, and $\tilde{\beta}$ is the approximate h-likelihood estimator for the random effect β; that is,

$$\begin{aligned}
\hat{q}_{MHLE}(\alpha; \tilde{\beta}) &= n^{-1} \sum_{i=1}^{n} \left(\sum_{t \in N_i} x_i(t) - D_\alpha \hat{\Lambda}_i(\alpha, \tilde{\beta}_i) \right) \\
&= n^{-1} \sum_{i=1}^{n} \left(\sum_{t \in N_i} x_i(t) - \sum_{u \in S_i} \frac{x_i(u) \exp\left\{ \alpha^T x_i(u) + \tilde{\beta}_i^T z_i(u) \right\}}{\pi_i(u)} \right),
\end{aligned}$$

and define $W_\infty(\alpha)$, $V_\infty(\alpha)$, and $U_\infty(\alpha)$ to be the respective limits of $n^{-1}W_n(\alpha)$, $n^{-1}V_n(\alpha)$, and $n^{-1}U_n(\alpha)$ as $n \to \infty$. For example, the quantity $W_\infty(\alpha)$ is

$$\begin{aligned}
W_\infty(\alpha) &= \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} E \left\{ \int_{A_i} x_i(t) x_i(t)^T \exp\left\{ \alpha^T x_i(t) + \beta_i^T z_i(t) \right\} dt \right\} \\
&= \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \int_{A_{ii}} x_i(t) x_i(t)^T \lambda_i(t; \alpha) M_{\beta_i}(z_i(t); \mu, \Sigma)\, dt,
\end{aligned}$$

where

$$\lambda_i(t; \alpha) = \exp\left\{ \alpha^T x_i(t) \right\},$$

and

$$\begin{aligned}
M_{\beta_i}(z_i(t); \mu, \Sigma) &= E\left\{ \exp\left(\beta_i^T z_i(t) \right) \right\} \\
&= p \exp\left\{ z_i(t)^T \mu_1 + \frac{1}{2} z_i(t)^T \Sigma z_i(t) \right\} + (1-p) \exp\left\{ z_i(t)^T \mu_2 + \frac{1}{2} z_i(t)^T \Sigma z_i(t) \right\}.
\end{aligned}$$

The remaining terms $V_\infty(\alpha)$, and $U_\infty(\alpha)$ can be written in a similar manner. The full expressions for these quantities and all variance components are provided in Appendix B. Then,

$Var\left(\hat{\alpha}_{MHLE}\right)$ is asymptotically

$$\left\{J_\infty(\alpha)\right\}^{-1} Q_\infty\left(\alpha\right)\left\{J_\infty(\alpha)\right\}^{-1},$$

where $Q_\infty\left(\alpha\right) = S^X_\infty\left(\alpha\right) + S^Y_\infty\left(\alpha\right) - V^T_\infty(\alpha)U^{-1}_\infty(\alpha)V_\infty(\alpha)$, and

$$J_\infty(\alpha) = W_\infty(\alpha) - V^T_\infty(\alpha)U^{-1}_\infty(\alpha)V_\infty(\alpha), \tag{3.11}$$

Here $J_\infty(\alpha)$ is a consistent estimate of $-E\left\{\frac{\partial}{\partial\alpha}\hat{q}_{MHLE}(\alpha,\beta,\mu;\Sigma)\right\}$ evaluated at $\hat{\alpha}_{MHLE}$, and $S^X_\infty\left(\alpha\right)$ and $S^Y_\infty\left(\alpha\right)$ are the asymptotic variances of the estimating equation due to the event assessments and the random assessments. These quantities can be written as

$$\begin{aligned}
S^X_\infty\left(\alpha\right) &= \lim_{n\to\infty} Var_\theta\left\{n^{-\frac{1}{2}}\sum_{i=1}^n\left(\sum_{t\in N_i} x_i(t)\right)\right\}\\
&= n^{-1}\sum_{i=1}^n\Big(\int\limits_{A_i} x_i(t)x_i(t)^T\lambda_i(t;\alpha)M_{\beta_i}\left(z_i(t);\mu,\Sigma\right)dt\\
&\quad+\int\limits_{A_i}\int\limits_{A_i} x_i(s)x_i(t)^T\lambda_i(s;\alpha)\lambda_i(t;\alpha)\\
&\quad*\left[M_{\beta_i}\left(z_i(s)+z_i(t);\mu,\Sigma\right) - M_{\beta_i}\left(z_i(s);\mu,\Sigma\right)M_{\beta_i}\left(z_i(t);\mu,\Sigma\right)\right]dsdt\ \Big),
\end{aligned}$$

$$\begin{aligned}
S^Y_\infty\left(\alpha\right) &= \lim_{n\to\infty} Var_\theta\left(n^{-\frac{1}{2}}\sum_{i=1}^n\left(\sum_{u\in S_i}\pi_i(u)^{-1}x_i(u)\exp\left\{\alpha^T x_i(u)+\beta_i^T z_i(u)\right\}\right)\right)\\
&= n^{-1}\sum_{i=1}^n\Big(\int\limits_{A_{ii}}\frac{x_i(u)x_i(u)^T\lambda_i^2(u;\alpha)M_{\beta_i}\left(z_i(u);\mu,\Sigma\right)M_{\beta_i}\left(z_i(u);\mu,\Sigma\right)^T}{\pi_i(u)}du\\
&\quad+\int\limits_{A_{ii}}\int\limits_{A_{ii}}\left(\frac{\pi_i(u,v)-\pi_i(u)\pi_i(v)}{\pi_i(u)\pi_i(v)}\right)x_i(u)x_i(v)^T\lambda_i(u;\alpha)\lambda_i(v;\alpha)M_{\beta_i}\left(z_i(u);\mu,\Sigma\right)M_{\beta_i}\left(z_i(v);\mu,\Sigma\right)^T
\end{aligned}$$

Note that many terms in $Var\left(\hat{\alpha}_{MHLE}\right)$ need to be estimated by integral estimators. For example, $W_\infty(\alpha)$ can be approximated by using Rathbun et al.'s (2001) estimation technique, or

$$\hat{W}_n(\alpha) = n^{-1}\sum_{i=1}^n\left(\sum_{u\in S_i}\frac{x_i(u)x_i(u)^T\lambda_i(t;\alpha)M_{\beta_i}\left(z_i(t);\mu,\Sigma\right)}{\pi_i(u)}\right).$$

Instead, this estimation may be simplified by utilizing

$$E\left\{\sum_{t\in N_i} x_i(u)\right\} = \int\limits_{A_i} x_i(t)\lambda_i(t;\alpha)M_{\beta_i}\left(z_i(t);\mu,\Sigma\right)dt.$$

So, under Rathbun et al.'s estimation methods,

$$\hat{W}_n(\alpha) = n^{-1} \sum_{i=1}^{n} \left(\sum_{u \in S_i} x_i(u) x_i(u)^T \right)$$

is an unbiased estimator of $W_\infty(\alpha)$. The remaining terms in (3.11), as well much of S_∞ matrices may be estimated in a similar manner by exploiting the expected intensity of the process.

3.9 Ecological Momentary Assessment of Smoking

Mixture Mixed effects versions of the modulated Poisson process are illustrated using data from an EMA of smoking (Shiffman et al. 2002). A total of 304 smokers were given electronic diaries (PDAs) and were instructed to record on the diary each cigarette that was smoked. The covariates include both mood affect and environmental setting. Negative Affect (composite of subjects' responses to negative adjectives such as "miserable" and "irritated"), Arousal (reactions to words such as "energetic"), Attention (based on subjects' responses to questions regarding their ease or difficulty concentrating), and Restlessness (based on a single item that did not load into other factors) are the four of these variables analyzed by Rathbun et al. (2007).

The random assessments were sampled according to a stratified sampling design, where days were treated as strata. To reduce the burden on the study smokers, not all cigarettes were assessed for the covariates. The electronic diary was prompted to target 4 or 5 cigarette assessments per day according to a Bernoulli sampling scheme. For subject i on day j, each cigarette was independently chosen to be assessed with probability

$$p_{ij} = \min \left\{ 5N_{i,j-1}^{-1}, 1 \right\},$$

where $N_{i,j}$ denotes the number of events recorded on day j by subject i. Estimates of the parameter $\tilde{\theta}_n$ were obtained by solving modified versions of (12) accounted for the process intensity from both stratified and Bernoulli sampling of events. So, the estimating equation for the fixed effect can be modified as

$$n^{-1}\sum_{i=1}^{n}\left[\sum_{j=1}^{d_i}\left\{\sum_{k=1}^{N_{ij}}x_i(t_{ijk})-\frac{p_{ij}\left|L_{ij}\right|}{m_{ij}}\sum_{k=1}^{m_{ij}}x_i(u_{ijk})\exp\left\{\alpha^T x_i(u_{ijk})+\beta_i^T z_i(u_{ijk})\right\}\right\}\right]=0,$$

the estimating equation for the random effect, β_i is

$$\begin{aligned}&\sum_{j=1}^{d_i}\left\{\sum_{k=1}^{N_{ij}}z_i(t_{ijk})-\frac{p_{ij}\left|L_{ij}\right|}{m_{ij}}\sum_{k=1}^{m_{ij}}z_i(u_{ijk})\exp\left\{\alpha^T x_i(u_{ijk})+\beta_i^T z_i(u_{ijk})\right\}\right\}\\&-w_i\Sigma^{-1}\left(\beta_i-\mu_1\right)-(1-w_i)\Sigma^{-1}\left(\beta_i-\mu_2\right)=0,\end{aligned}$$

and the estimating equation for the variance component is

$$\begin{aligned}&-\frac{n}{2}\mathrm{trace}\Sigma^{-1}(\gamma)\frac{\partial}{\partial\gamma}\Sigma(\gamma)+\sum_{i=1}^{n}\frac{w_i}{2}\left(\beta_i-\mu_1\right)^T\Sigma^{-1}(\gamma)\frac{\partial}{\partial\gamma}\Sigma(\gamma)\Sigma^{-1}(\gamma)\left(\beta_i-\mu_1\right)\\&+\frac{1}{2}\sum_{i=1}^{n}(1-w_i)\left(\beta_i-\mu_2\right)^T\Sigma^{-1}(\gamma)\frac{\partial}{\partial\gamma}\Sigma(\gamma)\Sigma^{-1}(\gamma)\left(\beta_i-\mu_2\right)\\&-\frac{1}{2}\mathrm{trace}H^{-1}(\mu,\alpha,\beta,\Sigma(\gamma);N)\frac{\partial}{\partial\gamma}H(\mu,\alpha,\beta,\Sigma(\gamma);N)=0,\end{aligned}$$

where w_i is the conditional expectations of the complete log likelihood based on observed data. Here with $n=304$ subjects, d_i is the number of days on which subject i participated in the study, N_{ij} denotes the number of cigarettes for subject i on day j, m_{ij} is the number of random assessments, and $|L_{ij}|$ denotes the length of time that the diary was active for subject i on day j.

Table 3.1-3.3 contain the results of fitting a mixture mixed Poisson process for the analysis of the EMA data using "Age" as a fixed-effect covariate, and "Restlessness" and "Attention" as random-effect covariates. Subjects were partitioned into two groups. We estimate that $\hat{p}=0.6177$ ($se=0.094$) of the subjects belong to group 1 while remaining subjects belong to group 2.

Table 3.1: Fixed effect estimates for EMA data

Covariate	Estimate	SE	$p-$value
Intercept	-0.2431	0.049	<0.0001
Age	0.0060	0.0025	0.0243

Table 3.1 contains the estimates, estimated errors, and significance tests (versus $\alpha = 0$) for the fixed effects. A subject-level covariate, "Age", is found to be significantly different from 0, suggesting that smoking rate increases with increasing age, with an estimated rate ratio equal to $\exp(0.00608) = 1.0061$ (95% confidence interval, 1.001043, 1.011179).

Table 3.2: Mean parameters for random effects

	μ_1			μ_2		
Covariate	Estimate	SE	$p-$value	Estimate	SE	$p-$value
Restlessness	0.30102	0.0781	0.0002	0.0065	0.0514	0.3957
Attention	0.0013	0.0668	0.3988	0.0009	0.0089	0.3968
Negative affect	0.0099	0.0077	0.1732	-0.0084	0.0238	0.3746
Arousal	-0.0104	0.0294	0.3747	0.0099	0.0114	0.2749

Table 3.2 shows the mean parameters μ of mixture components for the random-effect covariates. We found that only restlessness is shown to have a significant effect on the smoking rate in the first cluster. Therefore, given that subject is in the first cluster, the smoking rate is increased by a multiplicative factor of $\exp(0.301024) = 1.351242$ for each unit increase in Restlessness with all remaining variables. Moreover, the mean parameter estimates are all found to not be significantly different from 0 in the second group. The results from the first cluster agree with Rathbun et al.'s results as well as previous literature (Shiffman et al. 1996) by having restlessness significantly impact smoking timing in a positive manner. However, the results of Rathbun et al. consider subjects event for covariates as fixed effects.

Table 3.3: Variance/covariance parameters for random effects

Covariate	Restlessness	Attention	Arousal	Negative affect
Restlessness	0.0188			
Attention	-0.0024	0.0175		
Arousal	-0.0134	0.0020	0.0221	
Negative affect	0.0096	0.0010	-0.0014	0.0118

Table 3.3 contains the estimates of the elements of the common covariance matrix, Σ, explaining the correlations between random-effect covariates. There is a positive correlation between Negative affect and Restlessness (cor=0.6466), and there is a negative correlation between Arousal and Restlessness (cor=-0.656). The remaining correlations are all found to be less than 0.2.

3.10 Appendix A

This Appendix demonstrates that the first and second Bartlett identities hold for the h-likelihood of the mixture mixed point process model and proves that properties (a)-(f) in section 3.8 are satisfied. To make this discussion more clear, Table 3.4 provides a summary of the notation used in this paper.

Table 3.4: Summary of the notation

Notation	**Meaning**
θ	$(\alpha, p, \mu, \Sigma)^T$
ϕ	$(\mu, \alpha, \beta)^T$
$\lambda_i(t; \alpha, \beta_i)$	the intensity for each subject i: $\lambda_i(t; \alpha, \beta_i) = \exp\left\{\alpha^T x_i(t) + \beta_i^T z_i(t)\right\}$
$\Lambda_i(\alpha, \beta_i)$	the integrated intensity: $\Lambda_i(\alpha, \beta_i) = \int_{A_i} \exp\left\{\alpha^T x_i(t) + \beta_i^T z_i(t)\right\} dt$
$f(\beta_i; \mu, \Sigma)$	The marginal density of β_i, which is the mixture of normal distributions
$f(\alpha; N_i \vert \beta_i)$	The Janossy density
$L_M(\theta; N)$	The marginal likelihood: $L_M(\theta; N) = \prod_{i=1}^{n} \int_{R^q} f(\alpha; N_i \vert \beta_i) f(\beta_i; \mu, \Sigma) d\beta_i$
$h(\mu, \alpha, \beta, \Sigma; N)$	h-likelihood, where $\log\left(L_M(\theta; N)\right) \simeq \log\left[\int_{R^q} \exp(h(\mu, \alpha, \beta, \Sigma; N)) d\beta\right]$
$h_c(\mu, \alpha, \beta, \Sigma; N)$	complete data h-likelihood (y_i involved)
$h_A(\mu, \alpha, \beta, \Sigma(\gamma); N)$	adjusted profile h-likelihood
$D_\alpha h(\alpha, \mu, \Sigma, \beta; N)$	the marginal score equations for the fixed effect:
$q_{MHLE}(\alpha; \tilde{\beta})$	the h-score equation for fixed effect: $q_{MHLE}(\alpha; \tilde{\beta}) = D_\alpha h(\alpha, \mu, \Sigma, \beta; N)\vert_{\beta=\tilde{\beta}}$
$\hat{q}_{MHLE}(\alpha; \tilde{\beta})$	the h-score equation for fixed effect (partially observed covariates)
$q_{MOM}(\alpha; \beta)$	the estimating equation for fixed effect from the method of moment
$\hat{\alpha}_{MLE}$	the MLE (Maximum likelihood Estimator) for fixed effect
$\hat{\alpha}_{MHLE}$	the MHLE for the fixed effects
$\hat{\alpha}_{MOM}$	the estimaor for the fixed effects from the method of moment

Bartlett identities of Mixture Point Process H-likelihood

For the n independent samples, each subject i is observed over a set of times belonging to the Borel set $A_i \subset R$. The hierarchical likelihood (Lee and Nelder 1996) for the mixture mixed Poisson process model is defined as

$$\begin{aligned} h(\mu, \alpha, \beta, \Sigma; N) &= \sum_{i=1}^{n} h_i(\mu, \alpha, \beta_i, \Sigma; N_i) = \sum_{i=1}^{n} \log f\left(N_i, \beta_i\right) \\ &= \sum_{i=1}^{n} \left\{\log f\left(\alpha; N_i; |\beta_i\right) + \log f(\beta_i; \mu, \Sigma)\right\}, \qquad (A.1) \end{aligned}$$

where $f\left(N_i, \beta_i\right)$ is simply the product of the marginal density of the random effect, β_i and the Janossy density of the point process. The maximum h-likelihood estimation (MHLE) can be obtained by maximizing $h(\theta; N)$ jointly over $\{\theta, \beta\}$. We let for $\phi = (\mu, \alpha, \beta)^T$, and consider the properties of both terms in $(A.1)$ separately. For the first term, suppose that the data consists of the locations of N events in study region A. Then,

$$f\left(\alpha; N|\beta\right) = \frac{1}{N!} \exp\left\{\sum_{t \in N} \left\{\alpha^T x(t) + \beta^T z(t)\right\} - \Lambda(\alpha, \beta)\right\}.$$

This likelihood function has discrete and continuous components. The discrete component is the number of events and the continuous components are the event locations. Summing this density over the number of events N, and integration over possible event locations, it can be seen that

$$\begin{aligned} \sum_{N=0}^{\infty} \int_A f\left(\alpha; N; |\beta\right) dt &= \sum_{N=0}^{\infty} \int_A \frac{1}{N!} \exp\left\{\sum_{t \in N} \left\{\alpha^T x(t) + \beta^T z(t)\right\} - \Lambda(\alpha, \beta)\right\} dt \\ &= \exp\left(-\Lambda(\alpha, \beta)\right) \sum_{N=0}^{\infty} \frac{1}{N!} \prod_{i=1}^{N} \left\{\int_A \left\{\alpha^T x(t) + \beta^T z(t)\right\} dt\right\} \\ &= \exp\left(-\Lambda(A, \beta)\right) \sum_{N=0}^{\infty} \frac{\left(\Lambda(\alpha, \beta)\right)^N}{N!} = 1. \end{aligned}$$

This result implies that

$$E\left[\frac{\partial \log f\left(\alpha; N_i; |\beta_i\right)}{\partial \phi} |\beta\right] = 0,$$

and

$$E\left[\frac{\partial^2 \log f\left(\alpha; N_i; |\beta_i\right)}{\partial \phi \partial \phi^T} |\beta\right] = -E\left[\left(\frac{\partial \log f\left(\alpha; N_i; |\beta_i\right)}{\partial \phi}\right) \left(\frac{\partial \log f\left(\alpha; N_i; |\beta_i\right)}{\partial \phi}\right)^T |\beta\right].$$

Clearly, for the second term in (1), we still have

$$\sum_{j=1}^{k} p_j \int_{R^q} (2\pi)^{-1/2} |\Sigma|^{-1} \exp\left\{-\frac{1}{2}(\beta-\mu_j)^T \Sigma^{-1}(\beta-\mu_j))\right\} d\beta = 1.$$

These results show that the first and second Bartlett identities hold for the h-likelihood of Mixture Mixed Poisson process; that is

$$E_\theta\left[\frac{\partial h(\phi; N)}{\partial \phi}\right] = 0,$$

and

$$E\left[\frac{\partial^2 h(\phi; N)}{\partial \phi \partial \phi^T}\right] = -E\left[\left(\frac{\partial h(\phi; N))}{\partial \phi}\right)\left(\frac{\partial h(\phi; N)}{\partial \phi}\right)^T\right].$$

However, having the Bartlett identities can not guarantee consistent estimators from MHLE.

Laplace Approximation for the marginal likelihood

For simplicity of notation, suppose that the random effects β_i are scalar. Define

$$c_k = h_i^{(k)}(\alpha, \mu, \Sigma, \beta_i; N_i) = \frac{\partial^k}{\partial \beta^k}(h_i(\alpha, \mu, \Sigma, \beta_i; N_i))|_{\beta_i=\tilde{\beta}_i}.$$

where $\tilde{\beta}_i$ is a solution of the equation $\partial(h_i(\alpha, \mu, \Sigma, \beta_i; N_i))/\partial\beta = 0$ given α. Set $c_2 = -1/\sigma^2$, where $\sigma^2 = U_i^{-1}(\alpha, \beta_i, \Sigma)|_{\beta_i=\tilde{\beta}_i}$, and $U_i(\alpha, \beta_i, \Sigma)$ is as defined in (3.8). It can also be seen that $c_k = D_{\beta^{(k)}}\Lambda_i(\alpha, \beta_i)$ for $k \geq 3$.

Compute the Taylor expansion for $\exp\{h_i(\alpha, \mu, \Sigma, \beta_i; N_i)\}$ at $\tilde{\beta}_i$

$$\begin{aligned}\int \exp\{h_i(\alpha, \mu, \Sigma, \beta_i; N_i)\} d\beta_i &= \sqrt{2\pi\sigma^2} \exp\left\{h_i(\alpha, \mu, \Sigma, \tilde{\beta}_i; N_i)\right\} \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\left(\beta_i - \tilde{\beta}_i\right)^2}{2\sigma^2}\right\} [1 \\ &\quad + \frac{1}{6}c_3\left(\beta_i - \tilde{\beta}_i\right)^3 + \frac{1}{24}c_4\left(\beta_i - \tilde{\beta}_i\right)^4 + \frac{1}{120}c_5\left(\beta_i - \tilde{\beta}_i\right)^5 \\ &\quad + \left(\frac{c_6}{720} + \frac{c_3^2}{72}\right)\left(\beta_i - \tilde{\beta}_i\right)^6 + ...]d\beta_i \\ &= \sqrt{2\pi\sigma^2} \exp\left\{h_i(\alpha, \mu, \Sigma, \tilde{\beta}_i; N_i)\right\} [1 + \frac{c_4\sigma^4}{8} + 15\left(\frac{c_6}{720} + \frac{c_3^2}{72}\right)\sigma^6 + ...]\end{aligned}$$

By Assumption 2, $c_k = O_p\left(|A_i|\right)$ for $k \geq 3$, and $\sigma^2 = O_p\left(|A^{-1}|\right)$. This produces an approximation

$$
\begin{aligned}
\int \exp\left\{h_i(\alpha,\mu,\Sigma,\beta_i;N_i)\right\} d\beta_i &= \exp\left\{h_i(\alpha,\mu,\Sigma,\tilde{\beta}_i;N_i)\right\}\sqrt{2\pi\sigma^2}\left[1+\frac{c_4\sigma^4}{8}+\frac{15}{72}c_3^2\sigma^6+O_p\left(\left|A_i^{-2}\right|\right)\right] \\
&= \exp\left\{h_i(\alpha,\mu,\Sigma,\tilde{\beta}_i;N_i)\right\}\sqrt{2\pi\sigma^2}\left[1+O_p\left(\left|A_i^{-1}\right|\right)\right].
\end{aligned}
$$

Proof of property (a)

The following proves that the h-likelihood estimator for the random effect β_i is approximately equal to the conditional expectation of β_i given the data, N_i:

$$
\begin{aligned}
E\left(\beta_i|N_i\right) &= \frac{\int \beta_i \exp\left\{h_i(\alpha,\mu,\Sigma,\beta_i;N_i)\right\} d\beta_i}{\int \exp\left\{h_i(\alpha,\mu,\Sigma,\beta_i;N_i)\right\} d\beta_i} \\
&= \tilde{\beta}_i + \frac{\int \left(\beta_i-\tilde{\beta}_i\right)\exp\left\{h_i(\alpha,\mu,\Sigma,\beta_i;N_i)\right\} d\beta_i}{\int \exp\left\{h_i(\alpha,\mu,\Sigma,\beta_i;N_i)\right\} d\beta_i}. \qquad (A.2)
\end{aligned}
$$

We consider the second term in (A.2),

$$
\begin{aligned}
\int \left(\beta_i-\tilde{\beta}_i\right)\exp\left\{h_i(\alpha,\mu,\Sigma,\beta_i;N_i)\right\} d\beta_i &= \exp\left\{h_i(\alpha,\mu,\Sigma,\tilde{\beta}_i;N_i)\right\}\sqrt{2\pi\sigma^2} \\
&\quad \int \frac{\left(\beta_i-\tilde{\beta}_i\right)}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{\left(\beta_i-\tilde{\beta}_i\right)^2}{2\sigma^2}\right\} \\
&\quad [1+\frac{1}{6}c_3\left(\beta_i-\tilde{\beta}_i\right)^3+\frac{1}{24}c_4\left(\beta_i-\tilde{\beta}_i\right)^4+... \\
&\quad +\frac{1}{2}\left(\frac{1}{6}c_3\left(\beta_i-\tilde{\beta}_i\right)^3+\frac{1}{24}c_4\left(\beta_i-\tilde{\beta}_i\right)^4+...\right)^2+...]d\beta_i \\
&= \exp\left\{h_i(\alpha,\mu,\Sigma,\tilde{\beta}_i;N_i)\right\}\sqrt{2\pi\sigma^2}[\frac{1}{2}c_3\sigma^4+\frac{1}{8}c_5\sigma^6+...]
\end{aligned}
$$

Then $E(\beta_i|N_i)$ can be written as

$$
\begin{aligned}
E(\beta_i|N_i) &= \tilde{\beta}_i + \left[\frac{\frac{1}{2}c_3\sigma^4 + \frac{1}{8}c_5\sigma^6 + ...}{1 + \frac{c_4\sigma^4}{8} + 15\left(\frac{c_6}{720} + \frac{c_3^2}{72}\right)\sigma^6 + ...}\right] \\
&= \tilde{\beta}_i + \frac{1}{2}c_3\sigma^4\left[\frac{1 + \frac{1}{8}c_5\sigma^6 + ...}{1 + \frac{c_4\sigma^4}{8} + 15\left(\frac{c_6}{720} + \frac{c_3^2}{72}\right)\sigma^6 + ...}\right] \\
&= \tilde{\beta}_i + \frac{1}{2}c_3\sigma^4\left[\frac{1 + \frac{1}{4}\frac{c_5}{c_3}\sigma^2 + \frac{1}{24}\frac{c_7}{c_3}\sigma^4 + ..}{1 + \frac{c_4\sigma^4}{8} + 15\left(\frac{c_6}{720} + \frac{c_3^2}{72}\right)\sigma^6 + ..}\right] \\
&= \tilde{\beta}_i + \frac{c_3}{2}\sigma^4 + \frac{c_5}{8}\sigma^6 + \frac{c_7}{48}\sigma^8 + ... \\
&= \tilde{\beta}_i + \frac{c_3}{2}\sigma^4 + O_p\left(|A_i|^{-2}\right) \\
&= \tilde{\beta}_i + O_p\left(|A_i|^{-1}\right).
\end{aligned}
$$

see Abramowitz and Stegun 1972, page 15 for computation of ratios of series.

Proof of property (b)

The following proves that the conditional variance of β_i given the data, N_i, is approximately equal to $\sigma^2 = U_i^{-1}(\alpha, \beta_i, \Sigma)|_{\beta_i=\tilde{\beta}_i}$. In a similar manner to Proof of property (a), we can show that

$$
\begin{aligned}
var(\beta_i|N_i) &= \frac{\int\left(\beta_i - \tilde{\beta}_i\right)^2 \exp\{h_i(\alpha, \mu, \Sigma, \beta_i; N_i)\}\, d\beta_i}{\int \exp\{h_i(\alpha, \mu, \Sigma, \beta_i; N_i)\}\, d\beta_i} \\
&= \sigma^2\left[\frac{1 + \frac{1}{24}c_3\sigma^4 + \left(\frac{c_6}{720} + \frac{c_3^2}{72}\right)\sigma^6 + ...}{1 + \frac{1}{8}c_4\sigma^4 + 15\left(\frac{c_6}{720} + \frac{c_3^2}{72}\right)\sigma^6 + ...}\right] \\
&= \sigma^2\left(1 + \left(\frac{c_3}{24} + \frac{c_4}{8}\right)\sigma^4 - 14\left(\frac{c_6}{720} + \frac{c_3^2}{72}\right)\sigma^6 + ...\right) \\
&= \sigma^2 + O_p\left(|A_i|^{-1}\right).
\end{aligned}
$$

Proof of property (c)

The following demonstrates that the difference between the h-score equations and the approximate h-score equations for fixed effects is $O_p(a_{\min}^{-1/2})$. First, the h-score equations for fixed effect are written as

$$
\begin{aligned}
q_{MHLE}(\alpha;\tilde{\beta}) &= D_\alpha h(\alpha,\mu,\Sigma,\beta;N)|_{\beta=\tilde{\beta}} \\
&= n^{-1}a_{\min}^{-1}\sum_{i=1}^{n}\left[\sum_{t\in N_i} x_i(t) - D_\alpha \Lambda_i(\alpha,\tilde{\beta}_i)\right],
\end{aligned}
$$

where $\tilde{\beta}_i$ is the estimate of the random effect for subject i given α. These h-score equations for fixed effect are special case of the marginal score equations, $D_\alpha h(\alpha,\mu,\Sigma,\beta;N)$, when $\beta_i = \tilde{\beta}_i$ for all $i = 1,..,n$. For technical reasons, the h-score equations are divided by the area $a_{\min}$ of the study region. Then, the approximate h-score equations for fixed effect are written as

$$
\begin{aligned}
\hat{q}_{MHLE}(\alpha;\tilde{\beta}) &= n^{-1}a_{\min}^{-1}\sum_{i=1}^{n}\left[\sum_{t\in N_i} x_i(t) - D_\alpha \hat{\Lambda}_i(\alpha,\tilde{\beta}_i)\right] \\
&= n^{-1}a_{\min}^{-1}\sum_{i=1}^{n}\left[\sum_{t\in N_i} x_i(t) - \sum_{u\in S_i}\frac{x_i(u)\exp\left\{\alpha^T x_i(u) + \tilde{\beta}_i^T z_i(u)\right\}}{\pi(u)}\right].
\end{aligned}
$$

Consider the first order approximation of $D_\alpha \Lambda_i(\alpha,\tilde{\beta}_i)$ and $D_\alpha \hat{\Lambda}_i(\alpha,\tilde{\beta}_i)$ about β_i, and by Assumption 7, the difference between these score equations can be written as

$$
\begin{aligned}
\hat{q}_{MHLE}(\alpha;\tilde{\beta}) - q_{MHLE}(\alpha;\tilde{\beta}) &= n^{-1}a_{\min}^{-1}\sum_{i=1}^{n}\Big(D_\alpha \Lambda_i(\alpha,\tilde{\beta}_i) - D_\alpha \hat{\Lambda}_i(\alpha,\tilde{\beta}_i)\Big) \\
&= n^{-1}a_{\min}^{-1}\sum_{i=1}^{n}\Big(D_\alpha \Lambda_i(\alpha,\beta_i) - D_\alpha \hat{\Lambda}_i(\alpha,\beta_i)\Big) \\
&\quad + n^{-1}a_{\min}^{-1}\sum_{i=1}^{n}\Big\{(\beta_i - \tilde{\beta}_i)\Big(D_{\alpha\beta}\Lambda_i(\alpha,\beta_i) - D_{\alpha\beta}\hat{\Lambda}_i(\alpha,\beta_i)\Big)\Big\} \\
&= O_p(a_{\min}^{-1/2}) + O_p(a_{\min}^{-3/2}) \\
&= O_p(a_{\min}^{-1/2}).
\end{aligned}
$$

Proof of property (d)

For a fixed sampling domain A, the following demonstrates that the difference between the maximum likelihood estimator (MLE) and MHLE for the fixed effect is $O_p(a_{\min}^{-1})$. To prove this property, we follow the similar approach as in Vonesh (1996). First, α is assumed to be scalar for a simplicity of notation. Compute the Taylor series-expansion of the marginal score equations for the fixed effect, $D_\alpha h(\alpha, \mu, \Sigma, \beta; N)$:

$$D_\alpha h(\alpha, \mu, \Sigma, \beta; N) = q_{MHLE}(\alpha; \tilde{\beta}) + \sum_{i=1}^{n} B_{1i}\left(\beta_i - \tilde{\beta}_i\right) + \sum_{i=1}^{n}\left(\beta_i - \tilde{\beta}_i\right)^T B_{2i}\left(\beta_i - \tilde{\beta}_i\right) + ...,$$

where $q_{MHLE}(\alpha; \tilde{\beta}) = D_\alpha h(\alpha, \mu, \Sigma, \beta; N)|_{\beta=\tilde{\beta}}$ is the h-score equation for fixed effect,

$$B_{1i} = (\partial/\partial\alpha)\left(\partial h(\alpha, \mu, \Sigma, \beta; N)/\partial\beta\right)|_{\beta=\tilde{\beta}} = O_p(|A_i|),$$

and

$$B_{2i} = (\partial/\partial\alpha)\left(\partial^2 h(\alpha, \mu, \Sigma, \beta; N)/\partial\beta\partial\beta^T\right)|_{\beta=\tilde{\beta}} = O_p(|A_i|).$$

Let $q_{MLE}(\alpha)$ denotes the maximum likelihood estimating equations for the fixed effects. Since $E\left(\beta_i|N_i\right) = \tilde{\beta}_i + O_p\left(|A_i|^{-1}\right)$, and $var(\beta_i|N_i) = U_i^{-1}(\alpha, \beta_i, \Sigma)\left(1 + O_p\left(|A_i|^{-1}\right)\right)$, then both $B_{1i}\left(\beta_i - \tilde{\beta}_i\right)$ and $\left(\beta_i - \tilde{\beta}_i\right)^T B_{2i}\left(\beta_i - \tilde{\beta}_i\right)$ are $O_p(1)$. So

$$\begin{aligned} q_{MLE}(\alpha) &= E\left\{D_\alpha h(\alpha, \mu, \Sigma, \beta; N)\right\} \\ &= q_{MHLE}(\alpha; \tilde{\beta}) + O_p(n). \end{aligned}$$

Let $\hat{\alpha}_{MLE}$ denotes the MLE, and $\hat{\alpha}_{MHLE}$ denotes the MHLE for the fixed effects. From the above equation, the marginal ML equations for fixed effects become

$$q_{MLE}(\hat{\alpha}_{MLE}) = q_{MHLE}(\hat{\alpha}_{MLE}; \tilde{\beta}) + O_p(n)$$

Then, compute the first-order Taylor series expansion of $q_{MHLE}(\hat{\alpha}_{MLE}; \tilde{\beta})$ about $\hat{\alpha}_{MHLE}$:

$$q_{MLE}(\hat{\alpha}_{MLE}) \simeq q_{MHLE}(\hat{\alpha}_{MHLE}; \tilde{\beta}) + D_\alpha q_{MHLE}(\hat{\alpha}_{MHLE}; \tilde{\beta})\left(\hat{\alpha}_{MLE} - \hat{\alpha}_{MHLE}\right) + O_p(n).$$

Since both $q_{MLE}(\hat{\alpha}_{MLE})$ and $q_{MHLE}(\hat{\alpha}_{MHLE}; \tilde{\beta})$ are zero, and $D_\alpha q_{MHLE}(\hat{\alpha}_{MHLE}; \tilde{\beta}) = O_p(na_{\min})$, we have

$$\begin{aligned} \hat{\alpha}_{MLE} - \hat{\alpha}_{MHLE} &= \left(D_\alpha q_{MHLE}(\hat{\alpha}_{MHLE}; \tilde{\beta})\right)^{-1} O_p(n) \\ &= O_p(a_{\min}^{-1}). \end{aligned}$$

Proof of property (e)

The following proves that $(\hat{\alpha}_{MHLE} - \alpha) = O_p\left\{\max\left(n^{-\frac{1}{2}}, a_{\min}^{-1}\right)\right\}$. First, in a similar manner to Proof of property (d), it can be shown that

$$\hat{\alpha}_{MLE} - \alpha = O_p(n^{-1/2}).$$

see Nie (2006) for more details. By using the property (d), we have

$$\begin{aligned}
\hat{\alpha}_{MHLE} &= \hat{\alpha}_{MLE} + O_p(a_{\min}^{-1}) \\
&= \alpha + O_p(n^{-1/2}) + O_p(a_{\min}^{-1}) \\
&= \alpha + O_p\left\{\max\left(n^{-\frac{1}{2}}, a_{\min}^{-1}\right)\right\}.
\end{aligned}$$

Therefore,

$$(\hat{\alpha}_{MHLE} - \alpha) = O_p\left\{\max\left(n^{-\frac{1}{2}}, a_{\min}^{-1}\right)\right\}.$$

Proof of property (f)

The following proves that $(\hat{\alpha}_{MHLE} - \hat{\alpha}_{MOM}) = O_p\left(a_{\min}^{-1}\right)$. First, the h-score equations for fixed effect are

$$\begin{aligned}
q_{MHLE}(\alpha) &= D_\alpha h(\alpha, \mu, \Sigma, \beta; N)|_{\beta=\tilde{\beta}} \\
&= \sum_{i=1}^{n}\left[\sum_{t\in N_i} x_i(t) - D_\alpha \Lambda_i(\alpha, \tilde{\beta}_i)\right],
\end{aligned}$$

The estimating equations for fixed effect from the method of moment are

$$q_{MOM}(\alpha; \tilde{\beta}) = \sum_{i=1}^{n}\left[\sum_{t\in N_i} x_i(t) - E\left[D_\alpha \Lambda_i(\alpha, \tilde{\beta}_i)\right]\right],$$

where

$$\begin{aligned}
E\left[D_\alpha \Lambda_i(\alpha,\tilde{\beta}_i)\right] &= E\left[\sum_{t\in N_i} x_i(t)\right] \\
&= E\left[\int x(t)\exp\left\{\alpha^T x(t)+\beta^T z(t)\right\}dt\right] \\
&= p\int x(t)\exp\left\{\alpha^T x(t)+z^T(t)\mu_1+\frac{1}{2}z^T(t)\Sigma z(t)\right\}dt \\
&\quad +(1-p)\int x(t)\exp\left\{\alpha^T x(t)+z^T(t)\mu_2+\frac{1}{2}z^T(t)\Sigma z(t)\right\}dt \\
&= \int x(t)\exp\left\{\alpha^T x(t)+\frac{1}{2}z^T(t)\Sigma z(t)\right\}\left(p\exp\left\{z^T(t)\mu_1\right\}+(1-p)\exp\left\{z^T(t)\mu_2\right\}\right)dt,
\end{aligned}$$

and the difference is

$$n^{-1}q_{MHLE}(\alpha;\tilde{\beta})-n^{-1}q_{MOM}(\alpha;\tilde{\beta})=r_n(\alpha,\beta_i,\tilde{\beta}_i)+k_n(\alpha,p,\mu,\Sigma)$$

where

$$r_n(\alpha,\beta_i,\tilde{\beta}_i)=n^{-1}\sum_{i=1}^{n}\left[D_\alpha\Lambda_i(\alpha,\beta_i)-D_\alpha\Lambda_i(\alpha,\tilde{\beta}_i)\right],$$

and

$$k_n(\alpha,p,\mu,\Sigma)=n^{-1}\sum_{i=1}^{n}\left[E\left[D_\alpha\Lambda_i(\alpha,\tilde{\beta}_i)\right]-D_\alpha\Lambda_i(\alpha,\beta_i)\right].$$

By Assumption 3, which ensures that all the study intervals are in a well-behaved fashion and $a_{\max}/a_{\min}=O_p(1)$, the Liapounov's condition (Ash 1972, p. 337) holds with $\delta=1$. Hence, $k_n(\alpha,p,\mu,\Sigma)$ is $O_p(n^{-1/2})$ by central limit theorem. Moreover, the variances and the second moments of the independent random variables contained in $k_n(\alpha,p,\mu,\Sigma)$ share a common bound by Assumption 1-4. So, Theorem 5.1.1 (Chung K.L. 1974, p. 103) and Kolmogorov's condition (Feller 1968, p. 259) are satisfied. $k_n(\alpha,p,\mu,\Sigma)$converges pointwise to zero for each $\theta\in\Theta$ with probability 1.

Consider the first approximation of $D_\alpha\Lambda_i(\alpha,\beta_i)$ about $\tilde{\beta}_i$ in $r_n(\alpha,\beta_i,\tilde{\beta}_i)$.We obtain

$$\begin{aligned}
r_n(\alpha,\beta_i,\tilde{\beta}_i) &\simeq n^{-1}\sum_{i=1}^{n}\left[D_\alpha\Lambda_i(\alpha,\tilde{\beta}_i)+D_{\alpha\beta_i}\Lambda_i(\alpha,\tilde{\beta}_i)\left(\beta_i-\tilde{\beta}_i\right)-D_\alpha\Lambda_i(\alpha,\tilde{\beta}_i)\right] \\
&= n^{-1}\sum_{i=1}^{n}D_{\alpha\beta_i}\Lambda_i(\alpha,\tilde{\beta}_i)\left(\beta_i-\tilde{\beta}_i\right)=O_p(1)
\end{aligned}$$

So, the difference between the estimating equations of method of mement and maximum h-likelihood for fixed effect is $O_p(1)$. Then, by adding and subtracting $n^{-1}q_{MOM}(\hat{\alpha}_{MHLE};\tilde{\beta})$:

$$\begin{aligned} n^{-1}q_{MHLE}(\hat{\alpha}_{MHLE};\tilde{\beta}) &= n^{-1}q_{MOM}(\hat{\alpha}_{MHLE};\tilde{\beta}) + n^{-1}q_{MHLE}(\hat{\alpha}_{MHLE};\tilde{\beta}) - n^{-1}q_{MOM}(\hat{\alpha}_{MHLE};\tilde{\beta}) \\ &= n^{-1}q_{MOM}(\hat{\alpha}_{MHLE};\tilde{\beta}) + O_p(1). \end{aligned}$$

Compute the first-order Taylor series expansion of $n^{-1}q_{MOM}(\hat{\alpha}_{MHLE};\tilde{\beta})$ about $\hat{\alpha}_{MOM}$:

$$n^{-1}q_{MHLE}(\hat{\alpha}_{MHLE};\tilde{\beta}) \simeq n^{-1}q_{MOM}(\hat{\alpha}_{MOM};\tilde{\beta})+n^{-1}D_\alpha q_{MOM}(\hat{\alpha}_{MOM};\tilde{\beta})\,(\hat{\alpha}_{MHLE}-\hat{\alpha}_{MOM})+O_p(1).$$

Since both $q_{MOM}(\hat{\alpha}_{MOM};\tilde{\beta})$ and $q_{MHLE}(\hat{\alpha}_{MHLE};\tilde{\beta})$ are zero, and $D_\alpha q_{MOM}(\hat{\alpha}_{MOM};\tilde{\beta}) = O_p(na_{\min})$, we have

$$\begin{aligned} \hat{\alpha}_{MHLE}-\hat{\alpha}_{MOM} &= \left(n^{-1}D_\alpha q_{MOM}(\hat{\alpha}_{MOM};\tilde{\beta})\right)^{-1} O_p(1) \\ &= O_p(a_{\min}^{-1}). \end{aligned}$$

3.11 Appendix B

The first part of this Appendix displays the terms of the asymtotic variance components of the fixed effect estimators. The second part will demonstrate how to derive the asymptotic variance of fixed effect estimators. Then, the third part will demonstrate the asymptotic variance of μ derived from incomplete data likelihood. Finally, the fourth part will show the details of Newton method for variance component estimation.

The asymptotic variance components

The complete expansions of the variance terms $S_\infty^X(\theta)$ and $S_\infty^Y(\theta)$ can be derived as

$$
\begin{aligned}
S_\infty^X(\theta) &= \lim_{n\to\infty} Var_\theta \left\{ n^{-\frac{1}{2}} \sum_{i=1}^{n} \left(\sum_{t\in N_i} x_i(t) \right) \right\} \\
&= n^{-1} \sum_{i=1}^{n} (\; E\left\{ \sum_{t\in N_i} x_i(t)x_i(t)^T + \sum_{s\neq t\in N_i} x_i(s)x_i(t) \right\} \\
&\quad - E\left\{ \sum_{t\in N_i} x_i(t) \right\} E\left\{ \sum_{t\in N_i} x_i(t) \right\}^T \;) \\
&= n^{-1} \sum_{i=1}^{n} (\int_{A_i} x_i(t)x_i(t)^T \lambda_i(t;\alpha) M_{\beta_i}(z_i(t);\mu,\Sigma)\, dt \\
&\quad + \int_{A_i}\int_{A_i} x_i(s)x_i(t)^T \lambda_i(s;\alpha)\lambda_i(t;\alpha) M_{\beta_i}(z_i(s)+z_i(t);\mu,\Sigma)\, dsdt \\
&\quad - \int_{A_i}\int_{A_i} x_i(s)x_i(t)^T \lambda_i(s;\alpha)\lambda_i(t;\alpha) M_{\beta_i}(z_i(s);\mu,\Sigma)\, M_{\beta_i}(z_i(t);\mu,\Sigma)\, dsdt \;) \\
&= n^{-1} \sum_{i=1}^{n} (\int_{A_i} x_i(t)x_i(t)^T \lambda_i(t;\alpha) M_{\beta_i}(z_i(t);\mu,\Sigma)\, dt \\
&\quad + \int_{A_i}\int_{A_i} x_i(s)x_i(t)^T \lambda_i(s;\alpha)\lambda_i(t;\alpha) \\
&\quad * \left[M_{\beta_i}(z_i(s)+z_i(t);\mu,\Sigma) - M_{\beta_i}(z_i(s);\mu,\Sigma)\, M_{\beta_i}(z_i(t);\mu,\Sigma) \right] dsdt \;).
\end{aligned}
$$

and

$$\begin{aligned}
S_{\infty}^{Y}(\theta) &= \lim_{n\to\infty} Var_{\theta}\left(n^{-\frac{1}{2}}\sum_{i=1}^{n}\left(\sum_{u\in S_i}\pi_i(u)^{-1}x_i(u)\exp\left\{\alpha^T x_i(u)+\beta_i^T z_i(u)\right\}\right)\right) \\
&= n^{-1}\sum_{i=1}^{n}(\ E(\ \sum_{u\in S_i}\frac{x_i(u)x_i(u)^T\exp\left\{2\left(\alpha^T x_i(u)+\beta_i^T z_i(u)\right)\right\}}{\pi_i^2(u)} \\
&\quad + \sum_{u\neq v\in N_i}\frac{x_i(u)x_i(v)^T\exp\left\{\alpha^T x_i(u)+\beta_i^T z_i(u)\right\}\exp\left\{\alpha^T x_i(v)+\beta_i^T z_i(v)\right\}}{\pi_i(u)\pi_i(v)}\ ) \\
&\quad -E\left\{\frac{x_i(u)\exp\left\{\alpha^T x_i(u)+\beta_i^T z_i(u)\right\}}{\pi_i(u)}\right\}E\left\{\frac{x_i(u)\exp\left\{\alpha^T x_i(u)+\beta_i^T z_i(u)\right\}}{\pi_i(u)}\right\}^T\ ) \\
&= n^{-1}\sum_{i=1}^{n}(\ \int_{A_{ii}}\frac{x_i(u)x_i(u)^T\lambda_i^2(u;\alpha)M_{\beta_i}\left(z_i(u);\mu,\Sigma\right)M_{\beta_i}\left(z_i(u);\mu,\Sigma\right)^T}{\pi_i(u)}du \\
&\quad +\int_{A_{ii}}\int_{A_{ii}}\left(\frac{\pi_i(u,v)-\pi_i(u)\pi_i(v)}{\pi_i(u)\pi_i(v)}\right)x_i(u)x_i(v)^T\lambda_i(u;\alpha)\lambda_i(v;\alpha)M_{\beta_i}\left(z_i(u);\mu,\Sigma\right)M_{\beta_i}\left(z_i(v);\mu,\Sigma\right)^T\ )
\end{aligned}$$

Asymptotic Variance of Fixed effect estimators

Let $\hat{q}_{MHLE}(\alpha)$ be a score equations for the fixed effect for partially observed covariates, that is

$$\begin{aligned}
\hat{q}_{MHLE}(\alpha;\tilde{\beta}) &= \sum_{i=1}^{n}\left(\sum_{t\in N_i}x_i(t)-D_{\alpha}\hat{\Lambda}_i(\alpha,\tilde{\beta}_i)\right) \\
&\quad \sum_{i=1}^{n}\left\{\sum_{t\in N_i}x_i(t)-\sum_{u\in S_i}\frac{x_i(u)\exp\left\{\alpha^T x_i(u)+\tilde{\beta}_i^T z_i(u)\right\}}{\pi(u)}\right\}.
\end{aligned}$$

By the first order approximation of $D_{\alpha}\hat{\Lambda}_i(\alpha,\beta_i)$, we get

$$\begin{aligned}
\hat{q}_{MHLE}(\alpha;\tilde{\beta}) &\simeq \sum_{i=1}^{n}\left\{\sum_{t\in N_i}x_i(t)-\sum_{u\in S_i}\frac{x_i(u)\exp\left\{\alpha^T x_i(u)+\beta_i^T z_i(u)\right\}}{\pi(u)}\right\} \\
&\quad -\sum_{i=1}^{n}\left\{\sum_{u\in S_i}\frac{x_i(u)\exp\left\{\alpha^T x_i(u)+\beta_i^T z_i(u)\right\}}{\pi(u)}\left(\tilde{\beta}_i-\beta_i\right)\right\}.
\end{aligned}$$

Then, $Var\left(\hat{q}_{MHLE}(\alpha;\tilde{\beta})\right)$ is approximately

$$\begin{aligned}
Var\left(\hat{q}_{MHLE}(\alpha;\tilde{\beta})\right) &= Q_{\infty}(\alpha) \\
&= S_{\infty}^{X}(\alpha)+S_{\infty}^{Y}(\alpha)-V_{\infty}^{T}(\alpha)U_{\infty}^{-1}(\alpha)V_{\infty}(\alpha),
\end{aligned}$$

Here, $V(\alpha)$, and $U(\alpha)$ are taken the forms as in (3.7). And, by adding and subtracting $q_{MHLE}(\alpha;\tilde{\beta})$, we have

$$\frac{\partial}{\partial\alpha}\hat{q}_{MHLE}(\alpha;\tilde{\beta}) = \frac{\partial}{\partial\alpha}q_{MHLE}(\alpha;\tilde{\beta}) + \frac{\partial}{\partial\alpha}\left(\hat{q}_{MHLE}(\alpha;\tilde{\beta}) - q_{MHLE}(\alpha;\tilde{\beta})\right) ,$$

where

$$q_{MHLE}(\alpha;\tilde{\beta}) = \sum_{i=1}^{n}\left\{\sum_{t\in N_i} x_i(t) - \int_{A_i} x_i(t)\exp\left\{\alpha^T x_i(t) + \tilde{\beta}_i^T z_i(t)\right\} dt\right\}.$$

It can be seen that $\left(\hat{q}_{MHLE}(\alpha;\tilde{\beta}) - q_{MHLE}(\alpha;\tilde{\beta})\right) \to 0$. Then, using the estimate of $-E\left\{\partial q_{MHLE}(\alpha)/\partial\alpha\right\}$ on p. 654 of Lee and Nelder (1996), we have consistent estimate of $-E\left\{\frac{\partial}{\partial\alpha}\hat{q}_{MHLE}(\alpha,\beta,\mu;\Sigma)\right\}$, which is defined as

$$\begin{aligned} J_\infty(\alpha) &= -\lim_{n\to\infty} E\left\{\partial\hat{q}_{MHLE}(\alpha)/\partial\alpha\right\} \\ &= W_\infty(\alpha) - V_\infty^T(\alpha)U_\infty^{-1}(\alpha)V_\infty(\alpha). \end{aligned}$$

So, $Var\left(\hat{\alpha}_{MHLE}\right)$ is asymptotically $\left\{J_\infty(\alpha)\right\}^{-1} Q_\infty\left(\alpha\right)\left\{J_\infty(\alpha)\right\}^{-1}$..

Asymptotic Variance of μ and p

The incomplete-data h-likelihood subject i can be written as

$$\begin{aligned} h_{ic}(\theta;N) &= \alpha^T\sum_{t\in N_i} x_i(t) + \beta_i^T\sum_{t\in N_i} z_i(t) - \Lambda_i(\alpha,\beta_i) \\ &\quad +\log\left\{pe_{i1} + (1-p)e_{i2}\right\}, \end{aligned}$$

where

$$e_{ij} = \exp\left(-\frac{1}{2}\left(\beta_i-\mu_j\right)^T\Sigma^{-1}\left(\beta_i-\mu_j\right)\right).$$

For each subject i, the incomplete score equations for μ are

$$\begin{aligned} q_i(\mu) &= \frac{\partial h_{ic}(\theta;N)}{\partial\mu}|_{\beta=\tilde{\beta}} \\ &= \left[\begin{array}{c} \frac{p\Sigma^{-1}\left(\tilde{\beta}_i-\mu_1\right)\hat{e}_{i1}}{p\hat{e}_{i1}+(1-p)\hat{e}_{i2}} \\ \frac{(1-p)\Sigma^{-1}\left(\tilde{\beta}_i-\mu_2\right)\hat{e}_{i2}}{p\hat{e}_{i1}+(1-p)\hat{e}_{i2}} \end{array}\right]. \end{aligned}$$

and the incomplete score equation for p is

$$q_i(p) = \frac{\hat{e}_{i1} - \hat{e}_{i2}}{p\hat{e}_{i1} + (1-p)\hat{e}_{i2}}|_{\beta=\tilde{\beta}}.$$

Then $Var\left(\hat{\mu}_{MHLE}\right)$ is asymptotically

$$\left\{J_\infty(\mu)\right\}^{-1} Q_\infty\left(\mu\right)\left\{J_\infty(\mu)\right\}^{-1},$$

Here

$$Q_\infty(\mu) \quad = \quad \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} q_i(\mu) q_i(\mu)^T,$$

is a consistent estimate of $E\left\{q(\mu)q(\mu)^T\right\}$, and

$$J_\infty(\mu) \quad = \quad -\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \left(\partial q_i(\mu)/\partial\mu\right),$$

is a consistent estimate of $-E\left\{\partial q(\mu)/\partial\mu\right\}$.

The elements of $J_\infty(\mu)$ comprises of the following second derivative terms of $q_i(\mu)$ with respect to μ:

$$\begin{aligned}\frac{\partial^2 h_{ic}(\theta;N)}{\partial\mu_1\partial\mu_1^T} \quad &= \quad \frac{pe_{i1}\left(\Sigma^{-1}\left(\beta_i-\mu_1\right)\left(\beta_i-\mu_1\right)^T\Sigma^{-1}-\Sigma^{-1}\right)}{\left(pe_{i1}+(1-p)e_{i2}\right)} \\ &\quad + \frac{p^2e_{i1}^2\left(\Sigma^{-1}\left(\beta_i-\mu_1\right)\left(\beta_i-\mu_1\right)^T\Sigma^{-1}\right)}{\left(pe_{i1}+(1-p)e_{i2}\right)^2}.\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 h_{ic}(\theta;N)}{\partial\mu_2\partial\mu_2^T} \quad &= \quad \frac{(1-p)e_{i2}\left(\Sigma^{-1}\left(\beta_i-\mu_2\right)\left(\beta_i-\mu_2\right)^T\Sigma^{-1}-\Sigma^{-1}\right)}{\left(pe_{i1}+(1-p)e_{i2}\right)} \\ &\quad + \frac{(1-p)^2e_{i2}^2\left(\Sigma^{-1}\left(\beta_i-\mu_2\right)\left(\beta_i-\mu_2\right)^T\Sigma^{-1}\right)}{\left(pe_{i1}+(1-p)e_{i2}\right)^2}.\end{aligned}$$

and

$$\frac{\partial^2 h_{ic}(\theta;N)}{\partial\mu_1\partial\mu_2^T} = -\frac{p(1-p)e_{i1}e_{i2}\left(\Sigma^{-1}\left(\beta_i-\mu_1\right)\left(\beta_i-\mu_2\right)^T\Sigma^{-1}\right)}{\left(pe_{i1}+(1-p)e_{i2}\right)^2}.$$

The asymptotic variance of p can also be written as

$$Var\left(\hat{p}_{MHLE}\right) = \left\{J_\infty(p)\right\}^{-1} Q_\infty\left(p\right)\left\{J_\infty(p)\right\}^{-1},$$

Here

$$Q_{\infty}(p) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} q_i(p) q_i(p)^T,$$

and

$$J_{\infty}(p) = -\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \left(\partial q_i(p)/\partial \mu\right),$$

where

$$\frac{\partial q_i(p)}{\partial p} = -\left(\frac{\hat{e}_{i1} - \hat{e}_{i2}}{p\hat{e}_{i1} + (1-p)\hat{e}_{i2}}\right)^2$$

Newton method for Variance component estimation

First, rearrange the $H(\mu, \alpha, \beta, \Sigma; N)$ matrix as defined in (3.8)

$$H(\mu, \alpha, \beta, \Sigma; N) = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix},$$

where

$$H_{11} = \begin{bmatrix} W(\alpha, \beta) & 0 \\ 0 & P(\Sigma) \end{bmatrix},$$

$$H_{12} = H_{21}^T = \begin{bmatrix} V^T(\alpha, \beta) \\ T(\Sigma) \end{bmatrix},$$

and

$$H_{22} = \left[U(\alpha, \beta, \Sigma)\right].$$

Let $Z = H_{11} - H_{12}H_{22}^{-1}H_{21}$, we get

$$H^{-1}(\mu, \alpha, \beta, \Sigma; N) = \begin{bmatrix} Z^{-1} & -Z^{-1}H_{12}H_{22}^{-1} \\ -H_{22}^{-1}H_{21}Z^{-1} & H_{22}^{-1} + H_{22}^{-1}H_{21}Z^{-1}H_{12}H_{22}^{-1} \end{bmatrix}.$$

Since $\text{trace}\Big(H^{-1}(\mu, \alpha, \beta, \Sigma(\gamma); N)\frac{\partial}{\partial\gamma}H(\mu, \alpha, \beta, \Sigma(\gamma); N).\Big) = \text{trace}\Big(K(\mu, \alpha, \beta, \Sigma(\gamma)\left(\frac{\partial U(\alpha,\beta,\Sigma)}{\partial\gamma}\right)\Big)$, where $K(\mu, \alpha, \beta, \Sigma(\gamma)$ is the matrix given by the bottom right-hand of $H^{-1}(\mu, \alpha, \beta, \Sigma; N)$, the score equations for γ can be written as

$$
\begin{aligned}
&D_\gamma h = -\frac{n}{2}\text{trace}\Sigma^{-1}(\gamma)\frac{\partial}{\partial\gamma}\Sigma(\gamma) \\
&+\sum_{i=1}^{n}\frac{w_i}{2}\left(\hat{\beta}_i-\hat{\mu}_1\right)^T\Sigma^{-1}(\gamma)\frac{\partial}{\partial\gamma}\Sigma(\gamma)\Sigma^{-1}(\gamma)\left(\hat{\beta}_i-\hat{\mu}_1\right) \\
&+\frac{1}{2}\sum_{i=1}^{n}(1-w_i)\left(\hat{\beta}_i-\hat{\mu}_2\right)^T\Sigma^{-1}(\gamma)\frac{\partial}{\partial\gamma}\Sigma(\gamma)\Sigma^{-1}(\gamma)\left(\hat{\beta}_i-\hat{\mu}_2\right) \\
&-\frac{n}{2}\text{trace}\left(K(\mu,\alpha,\beta,\Sigma(\gamma)\left(\frac{\partial U(\alpha,\beta,\Sigma)}{\partial\gamma}\right)\right).
\end{aligned}
$$

Using $\left(\hat{\beta}_i-\hat{\mu}_1\right)=(\beta_i-\mu_1)+\left\{\left(\hat{\beta}_i-\hat{\mu}_1\right)-(\beta_i-\mu_1)\right\}$, it can be seen that

$$E\left(D_\gamma h\right)=0.$$

Moreover, the Newton's hessian matrix can be simplified as

$$
\begin{aligned}
E\left(D_{\gamma\gamma^T}h\right) &= -\frac{n}{2}\text{trace}\left(\left(\Sigma^{-1}(\gamma)\frac{\partial}{\partial\gamma}\Sigma(\gamma)\right)\left(\Sigma^{-1}(\gamma)\frac{\partial}{\partial\gamma}\Sigma(\gamma)\right)\right) \\
&\quad -\frac{n}{2}\text{trace}\left(\left(K\left(\frac{\partial U(\alpha,\beta,\Sigma)}{\partial\gamma}\right)\right)\left(K\left(\frac{\partial U(\alpha,\beta,\Sigma)}{\partial\gamma}\right)\right)\right)
\end{aligned}
$$

Chapter 4

Conclusion

4.1 Dissertation review

Ecological momentary assessment (EMA) is an emerging method for collecting real-time data in subjects' environment. This technique uses electronic devices to take repeated assessments over time of the immediate states of subjects in their natural environments. In doing so, EMA allows for unbiased estimation of their mean psychological states through probability-based sampling of their mood and other variables over time. Moreover, this method makes it possible to record the times of repeated behavioral events in subjects. Hierarchical Linear Models (HLMs) have been generally recommended for the analysis of EMA data (Affleck et al. 1999; Bolger, Davis, & Rafaeli 2003; Schwartz & Stone 2007). However, HLMs are not designed to model the timing of these repeated discrete events over an interval, because they are based on regression type models.

Recently, Rathbun, Shiffman, and Gwaltney (2007) proposed models for the analysis of EMA longitudinal repeated-event data using a Poisson process. Their model, however, contains the implicit assumption that all subjects have identical rates of event occurrence. For psychological research, this assumption can be unrealistic, since subjects may have inherent genetic factors that lead to higher or lower intensity functions. Moreover, recent research (Shiffman and Rathbun 2011) also suggests that subjects can also be divided into groups or categories depending on their different responses to changes in time-varying covariates.

This dissertation proposed two different models that account for heterogeneity among subjects. The modulated Poisson process with implementation of the finite mixture model was developed in Chapter 2. The maximum likelihood parameter estimation problem were described and we showed how the EM algorithm can be used for its solution. The ML equations were estimated using Rathbun et al.'s (2007) integral estimation method, and the estimators have been proved to be consistent and asymptotically normal under combined increasing domain/infill asymptotics.

The second model which is more flexible was developed in Chapter 3. A mixture mixed-effects version of a modulated Poisson process, where the intensity was modified to allow for inter-subject variability and handles heterogeneous population. This model resembled the Generalized Linear Mixed Model, since random effects were assumed to be sampled from a mixture of normal distributions with different means and common unknown covariance matrix. Since the likelihood had no closed form, we considered an estimation method based on hierarchical or h-likelihood (Lee and Nelder 1996). Estimating equations were developed from the application of Laplace approximation of the likelihood. Due to the difficulties of parameter estimation, the EM type algorithm were employed in Chapter 3. Furthermore, this chapter also provided the asymptotic properties of the estimators of the fixed parameters in effect of the increasing domain asymptotics, and rates of convergence were discussed.

All of the proposed methods were illustrated using Shiffman et al.'s (2002) multi-phase EMA study on smoking habits. This EMA study consisted of a sample of 304 smokers, each of whom was given an electronic diary and was instructed to record on the diary any time that they smoked a cigarette. The results from both Mixture-model versions of the modulated Poisson process were compared with Rathbun et al.'s (2007) results. The results coincided with both Rathbun et al. (2007), and literature on smoking, that is restlessness was shown to significantly increase smoking rate given that subject is in the one cluster. Therefore, more effective individual-based treatments for smoking cessation related to restlessness might be further applied to subjects in that cluster. Moreover, the model in Chapter 2 also suggested that the smoking rate is increased when there are others smoking in view of the participant given that subject is in the second cluster.

In conclusion, the two models proposed in this dissertation provided new, more flexible methods of analyzing EMA and other repeated-event data. Both models demonstrated a manner in which one may handle heterogeneity among the subjects using a finite-mixture model. For the EMA study on smoking cessation (Shiffman et al. 2002), these models may help us better understand how time-varying covariates impact smoking rate from which clusters of subjects showing similar smoking behaviors. This understanding may lead to improvements in smoking intervention programs.

4.2 Future research

This dissertation focuses on the two new approaches of mixture modulated ppoint process for the analysis of event timing in repeated-event data. Based on our experience, there are potential future research that can be focused on following topics:

4.2.1 Model Diagnostics

Since the asymptotic properties in both proposed models depend upon the point process being Poisson, a method of verifying this assumption such as K-function (Ripley 1976, 1977, 1981) may be used to analyze the second order properties of data. Moreover, with regard to the mixed-effect model in Chapter 4, a conclusive test for the appropriateness of a random effect term would assist a researcher to determine whether a variable should be modelled by a fixed or random effect.

4.2.2 Monte Carlo Approaches

McCulloch (1997) reviews Monte Carlo approaches for maximum likelihood estimation of parameters of generalized linear mixed models including the Monte Carlo EM (Wei and Tanner 1990; McCulloch 1994) and simulated maximum likelihood based on the importance sampling (Geyer and Thompson 1992). With regard to the mixed-effect model in Chapter 4, we can further apply these procedures for the analysis of event timing in repeated-event data. By using Monte Carlo algorithm for statistical inference, we can get an approximate log likelihood, score equations, and hessian matrix. These statistical procedures is particularly useful for relatively small EMA data sets.

Bibliography

[1] Abramovich, F., and Benjamini, Y. (1996). Adaptive thresholding of wavelet coefficients. *Computational Statistics and Data Analysis.* 22, 351-361.

[2] Aitkin, M., and Wilson, G. T. (1980). Mixture Models, Outliers, and the EM Algorithm, *Technometrics* 22, 325-331.

[3] Akaike H. A new look at the statistical model identification. Automatic Control, IEEE Transactions. 19, 716 - 723

[4] Ash, R.B. (1972). Real Analysis and Probability, 1st edition. Academic Press.

[5] Banfield J. D., Raftery A. E. (1992). Model-Based Gaussian and Non-Gaussian Clustering. Biometrics . 49, 803-821 .

[6] Basu, A. et al. (1998).Robust and efficient estimation by minimizing a density power divergence, *Biometrika* 85, 549 - 560.

[7] Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B.* 57, 289-300.

[8] Bolger, N. et al. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology.* 54, 579-616.

[9] Bozdogan H., Sclove SL. (1984) Multi-sample cluster analysis using Akaike's information criterion. Annals of the Institute of Statistical Mathematics. Springer.

[10] Beran, R. (1977). Minimum Hellinger distance estimates for parametric models, *The Annals of Statistics*. 5, 445-463.

[11] Breslow, N.E., and Clayton, D.G. (1993). Approximate inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association* 88, 9-25.

[12] Brillinger D.R. et al. (2002). Point process: temporal. *Encyclopedia of Environmetrics.* 3, 1576-1581.

[13] Carney, K.B. and Correia, C.J. (1998) Drinking motives predict alcohol-related problems in college students. J. Stud. Alcohol 58: 100-105.

[14] Carels R. A. and Musher-Eizenman D.R. (2004). Individual differences and weight bias: Do people with an anti-fat bias have a pro-thin bias?. Body Image. 2004 March; 7(2): 143148.

[15] Chandra, S. et al. (2007).Daily smoking patterns, their determinants, and implications for quitting. *Experimental and Clinical Psychopharmacology.* 15, 67-80.

[16] Chung K.L. (1974). A Course in Probability Theory, 1st edition.Academic Press.

[17] Collins, R.L., Morsheimer, E.T., Shiffman, S., Paty, J.A., Gnys, M., and Papandonators, G.D. (1998). Ecological momentary assessment in a behavioral drinking moderation training program. *Experimental and Clinical Psychopharmacology* 6, 306-315.

[18] Couvreur, C. (1997). The EM Algorithm: A Guided Tour. *To appear in the proceedings of the 2d IEEE European Workshop on Computationaly Intensive Method in Control and Signal Processing, Pragues, Czech Rep., Aug 28-30, 1996.*

[19] Cox, D.R. (1972). The statistical analysis of dependencies in point process. In *Stochastic Point Processes,* P.A.W. Lewis (Ed.), 55-66. New York: Wiley.

[20] Dempster, A.P. et al. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B.* 39, 1-38.

[21] Dietz, E. (1992). Estimation of heterogeneity—a GLM-approach. In: Fahrmeir, L., Francis, F., Gilchrist, R., Tutz, G. (Eds.), *Advances in GLIM and Statistical Modeling. Lecture Notes in Statistics*, Springer, Berlin, 66–72.

[22] Dietz and Böhning (1993). Statistical inference based on a general model of unobserved heterogeneity. In: Fahrmeir, L., Francis, F., Gilchrist, R., Tutz, G. (Eds.), *Advances in GLIM and Statistical Modeling. Lecture Notes in Statistics*, Springer, Berlin, 75–82.

[23] Donoho, D. L., and Johnstone, I.M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*. 81, 425-455.

[24] Duan J.A. et al. (2009). Modeling space-time data using stochastic differential equations, *Bayesian Analysis.* 4, 733-758.

[25] Fahrmeir, L., and Kaufmann, H. (1985). Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models. *Ann. Statist.* 13, 342-368.

[26] Feller W. (1968). An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd Edition. Wiley.

[27] Freedman, M. et al. (2006). Cell phones for Ecological Momentary Assessment with cocain-addicted homeless patients in treatment. *F. Subst. Abuse Treat.* 36, 105-111.

[28] Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London A.* 222, 309-368.

[29] Geman, S.; Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 6, 721-741.

[30] Green, P.J. (1987). Penalized likelihood for general semi-parametric regression models. International Statistical Review / Revue. 55, 245-259.

[31] Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika.*82, 711-732.

[32] Grn B., Leisch F. (2006). Fitting finite mixtures of generalized linear regressions in R. 51, 52475252.

[33] Guindon S, and Gascuel O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 5, 696-704.

[34] Ha ID, Lee Y, Song J-K. (2001). Hierarchical likelihood approach for frailty models. *Biometrika.* 88, 233243.

[35] Ha ID, Park T, and Lee Y. (2003). Joint modelling of repeated measures and survival time data. *Biometrical Journal*, 45, 647-658.

[36] Haberstick, B. C. et al. (2007). Genes, time to first cigarette and nicotine dependence in a general population sample of young adults.

[37] Hardelid P. et al. (2008). Agreement of rubella IgG antibody measured in serum and dried blood spots using two commercial enzyme-linked immunosorbent assays. Journal of Medical Virology. 80, 360-364. *Addiction.* 102, 655-665.

[38] Hall D., Wang L. (2005). Two-component mixtures of generalized linear mixed effects models for cluster correlated data. Statistical Modelling April 2005 vol. 5 no. 1 21-37.

[39] Hall P., Patil P. (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. Probability Theory and Related Fields. 99, 399-424.

[40] Hopper J.W. et al. (2006). Incidence and patterns of polydrug use and craving for ecstasy in regular ecstasy users: an ecological momentary assessment study. *Drug Alcohol Depend. 2006.* 85, 221-35.

[41] Jouhnson, D. H. (1996). Point process models of single-neuron discharges. *Journal of Computational Neuroscience,* 3(4), 275-299.

[42] Kalbfleisch, J.D., and Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data*. John Wiley&Sons, New York.

[43] Kaufmann H. and Fahrmeir L. (1985). Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models. *Ann. Statist.*13, 42-368.

[44] Kelvin K. W. Yau et al. (2003). Zero-Inflated Negative Binomial Mixed Regression Modeling of Over-Dispersed Count Data with Extra Zeros. Biometrical Journal. 45, 437452.

[45] Koehler AB., Murphree ES. (1988). A comparison of results from state space forecasting with forecasts from the Makridakis Competition. International Journal of Forecasting. 4, 4555.

[46] Kutoyants, Y.A. (1984). Statistical Inference for Spatial Poisson Process. New York: Springer.

[47] Lahiri et al. (1999). Prediction of Spatial Cumulative Distribution Functions Using Subsampling. Journal of the American Statistical Association. 94, 86-97.

[48] Lawless, J.F. (1987). Regression methods for Poisson process data. *Journal of the American Statistical Association* 82, 808-815.

[49] Lee, Y., and Nelder, J.A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society B* 58, 619-678.

[50] Liu Q., Pierce R. (1994). A note on GaussHermite quadrature. Biometrika. 81, 624-629.

[51] Liu L., and Huang XL. (2009). Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society Series C-Applied Statistics.* 58, 65-81.

[52] Louis, T.A. (1982). Finding the Observed Information Matrix when Using the EM Algorithm. *Journal of the Royal Statistical Society Series B.* 44, 226-233.

[53] Lu, Z. et al. (2003). Minimum hellinger distance estimation for finite mixtures of poisson regression models and its applications. *Biometrics.* 59, 1016-1026.

[54] Moolgavkar, S.H. et al. (1999a). Air pollution, pollens, and respiratory admissions for chronic obstructive pulmonary disease in King County. *Inhalation Toxicology (In Press).*

[55] Nelder J.A. and Wedderburn R.W.M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society.* 135. 370-384.

[56] Nie L. (2006). Strong consistency of the maximum likelihood estimator in generalized linear and nonlinear mixed-effects models. Metrika. 63, 123-143.

[57] Northrop, P. (1998). A clustered spatial-temporal model of rainfall. *Proceedings of the Royal Society: Mathematical, Physical and Engineering Sciences,* 454(1975), 1875-1888.

[58] Ogata, Y., and Akaike, H. (1982). Onlinear intensity model doubly stochastic Poisson and self exciting point process.*Journal of the Royal Statistical Society Series B.* 44, 102-107.

[59] Pearson (1894). Contributions to the mathematical theory of evolution, *Phil. Trans. Royal Soc.* 185A, 71-110.

[60] Pinheiro J. C., Bates D.M. (1995). Unconstrained parametrizations for variance-covariance matrices. Statistics and Computing. 6, 289-296.

[61] Ramsey, J.O., and Silverman, B.W. (1997) *Functional Data Analysis.* New York: Springer.

[62] Rathbun, S.L., and Cressie, N. (1994). Asymtotic properties of estimators for the parameters of spatial inhomogeneous Poisson point process. *Adv. Appl. Prob.* 26, 122-154.

[63] Rathbun, S.L. et al. (2007). Modeling the effects of partially observed covariates on Poisson process intensity, *Biometrika* 94, 153-165.

[64] Schwartz, J. (1993). Air pollution and daily mortality in Birmingham, Alabama. *The American Journal of Epidemiology.* 137, 1136-1147.

[65] Schwartz, J. (1994). Air pollution and hospital admissions for elderly in Birmingham, Alabama. *The American Journal of Epidemiology.* 139, 589-598.

[66] Schwartz, J.E., and Stone, A.A. (2007). The analysis of real-time momentary data: A practical guide. In *The Science of Real-time Data Capture: Self-report in Health Research,* Stone, A.A., Shiffman, S.S., Atienza, A., and Nebeling, L. (Eds.), 76-113. New York: Oxford University Press USA.

[67] Scott, D. W. (1998).On fitting and adapting of density estimates, *Computing Science and Statistics,* S. Weisberg.30, 124 - 133.

[68] Scott, D.W. (1999). Remarks on fitting and interpreting mixture models, *Computing Science and Statistics*, K. Berk and M. Pourahmadi. 31, 104-109.

[69] Scott, D. W. (2001). Parametric statistical modeling by minimum integrated square error. *Technometrics.* 43, 274-285.

[70] Scott, D.W. (2004). Outlier detection and clustering by partial mixture modeling. *COMPSTAT Symposium,* Physica-Verlag/Springer.

[71] Sheppard, L. et al. (1999). Effexts of ambient air pollution on nonelderly asthma hospital admissions in Seattle, Washington, 1987-1994. *Epidemiology.* 10, 23-30.

[72] Shiffman, S. (1989). Conceptual issues in the study of relapse. In M. Gossop (Ed.), *Relapse and additive behavior*. Kent, England, Croom Helm Lts.

[73] Shiffman, S. et al. (1996a). First Lapses to Smoking: Within-Subjects Analysis of. Real-Time Reports. *Journal of Consulting and Clinical Psychology,* 64, 366-379.

[74] Shiffman, S. et al. (1996b). Temptations to smoke after quitting a comparison of lapsers and maintainers. 15, 455-461.

[75] Shiffman, S. et al. (2002). Immediate antecedents of cigarette smoking: an analysis from ecological momentaryassessment. *J. Abnorm. Psychol.* 111, 531–45.

[76] Shiffman, S., and Rathbun, S. (2011). Point process analyses of variations in smoking rate by setting, mood, gender, and dependence. *Psychol Addict Behav.* 25, 501-510.

[77] Shiffman, S., and Stone,A.A. (1998). Ecological momentary assessment: A new tool for behavioral medicine research. In T*echnology and Methods in Behavioral Medicine*, Krantz, D.S. & Baum, A. (Eds.), 117-132.Philadelphia: Lawrence Erlbuam Associates.

[78] Smyth, J. et al. (2001). The use of ecological momentary assessment approaches in eating disorder research. *International Journal of Eating Disorders.* 30, 83-95.

[79] Soyer R., and Tarimcilar M. (2008). Modeling and Analysis of Call Center Arrival Data. *Management Science.* 54, 226-278.

[80] Stein et al. (1988). Use of soil-map delineations to improve (co)-kriging of point data on moisture deficits Geoderma, 43,163177.

[81] Stein, A., (1993). SPATANAL PC software. Department of Soil Science and Geology. Agricultural University of Wageningen, Wageningen.

[82] Stein M.L. (1999). Interpolation of Spatial Data: Some Theory For Kriging Springer, New York.

[83] Stone,A. A. et al. (1997). The experience of rheumatoid arthritis pain and fatigue: Examining momentary reports and correlates over one week. *Arthritis Care and Research,* 10*,* 185-193

[84] Stone,A. A. et al. (1998). A comparison of coping assessed by Ecological Momentary Assessment and restropective recal. *F. Personal. Soc. Psychol.* 74*,* 1670-1680

[85] Stone,A. A. & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Ann.Behav. Med*. 16, 199–202.

[86] Stoyan D. and Penttinen A. (2000). Recent applications of point process methods in forestry statistics. *Statist. Sci*. 15, 61-78

[87] Sundberg, R. (1974). Maximum likelihood theory for in-complete data from an exponential family. *Scand. J. Statist.*1, 49-58.

[88] Tadesse, M.G. et al. (2005). Wavelet thresholding with Baysian false discovery rate control. *Biometrics.* 61, 25-35.

[89] Waagepetersen, R. (2008). Estimating functions for inhomogeneous spatial point processes with incomplete covariate data. *Biometrika*, 95(2), 351-363.

[90] Windhama M. P., Cutlera M. (1991). Information Ratios for Validating Mixture Analyses. Journal of the American Statistical Association. 87, 1188-1192.

[91] Wulfsohn, M.S., and Tsiatis, A.A. (1997). A joint model for survival and longitudinal data measured with error. *Biometric.* 53, 330-339.

[92] Xia et al. (2006). Convergence Rate of the L-N Estimator in Poisson-Gamma Models. Acta Mathematicae Applicatae Sinica. 22, 639-654.

[93] Yuan, K.-H.(1997). A theorem of uniform convergence of stochastic functions with applications. *J. Multivariate Anal.* 62, 100–109.

[94] Yuan, K.-H., and Jenrich R.I. (1998). Asymptotics of Estimating Equations under Natural Conditions. *J. Multivariate Anal.* 65, 245-260.

[95] Zhang H. P. et al. (2008). Joint modeling of time series measures and recurrent events and analysis of the effects of air quality on respiratory symptoms. *Journal of the American Statistical Association.* 103, 48-60.