Title

A Random-effects Markov Transition Model for Poisson-distributed Repeated Measures with Nonignorable Missing Values

Authors

Jinhui Li Xiaowei Yang
Yingnian Wu
Steven Shoptaw

Publication Date

2011-10-25

Title:

A Random-effects Markov Transition Model for Poisson-Distributed Repeated Measures with Nonignorable Missing Values

Short Running Title

Markov Transition Model for Incomplete Count Measures

Authors:

Jinhui Li[1,2] Xiaowei Yang[1,3], Yingnian Wu[2], and Steven Shoptaw[3]

1. BayesSoft, Inc.,
   3641 Midvale Avenue, #207, Los Angeles, CA 90034.
2. UCLA-Department of Statistics
   PO Box 951554, Los Angeles, CA 90095-1554.
3. UCLA-Integrated Substance Abuse Programs
   11075 Santa Monica Blvd, Suite 200, Los Angeles, CA 90025.

Contact Author:

Xiaowei Yang, Ph.D., BayesSoft, Inc.,
3641 Midvale Avenue, #207, Los Angeles, CA 90034, U.S.A.
Tel 310-600-4368, Fax 310-836-5851
E-mail: XYang@BayesSoft.com

SUMMARY

In biomedical research with longitudinal designs, missing values due to intermittent nonresponse or premature withdrawal are usually 'nonignorable' in the sense that unobserved values are related to the patterns of missingness. When missing values are simply ignored, analyses based on observed-data likelihood may yield biased estimates or invalid inferences. By drawing the framework of a shared-parameter mechanism, the process yielding the repeated count measures and the process yielding missing values can be modelled separately, conditionally on a group of shared parameters. For chronic diseases, Markov transition models can be used to study the transitional features of the pathologic processes. In this paper, Markov chain Monte Carlo (MCMC) algorithms are developed to fit a random-effects Markov transition model (REMTM) for incomplete count repeated measures, within which random effects are shared by the counting process and the missing-data mechanism. Assuming a Poisson distribution for the count measures, the transition probabilities are estimated using a Poisson linear regression model. The missingness mechanism is modeled with a multinomial-logit regression to calculate the transition probabilities of the missingness indicators. The method is demonstrated using both simulated data sets and a practical data set from a smoking cessation clinical trial.

Key words: Repeated Measures, Markov Transition Models, Nonignorable Missing Values, Poisson Regression Model, Shared-Parameter Missingness

1 Introduction

1.1 Background

Longitudinal designs are commonly used to conduct biomedical research, especially for clinical trials. The defining feature of a longitudinal study is that large numbers of repeated measures are collected on study participants. Special statistical methods are required for the data analysis because observations collected on the same participant are correlated to each other [5]. For complete data sets or for those with ignorable missing values, three longitudinal strategies are most popular in biomedical research: (i) generalized linear mixed models (GLMM; [2]), where intra-subject correlations are introduced via random effects; (ii) marginal models using generalized estimating equations (GEEs) [10] where parameters on group means are estimated using quasi-likelihood method assuming a working correlation structure; and (iii) Markov transition models (MTM; [23]) where current measures are modelled by conditioning on the previous observations and covariates. In social-economical sciences, there are other forms of longitudinal modelling, e.g., hierarchical models, latent variable models, and structure equation models [19, 3, 9]. Compared to time-naive analysis using aggregate markers of outcomes(e.g., T-test, ANOVA, and GLM), longitudinal models are capable of modelling intra-correlation structures and drawing inferences on time-trends, thus more closely reflecting the nature of longitudinal design. Longitudinal models are also able to handle certain types of missing data problems encountered in

practical data sets. Many statistical packages (e.g., SAS, SPSS, STATA, S-Plus and R) have implemented the above models for longitudinal data analysis.

In longitudinal studies for chronic health problems, such as drug dependence, respiratory diseases, and cancers, there are usually large vectors of Poisson-distributed repeated measures that count the numbers of adverse events. For a participant of such a study, the current states are usually dependent on the previous observations in addition to the explanatory variables of interest, e.g., dummy variables indicating the treatment assignment in a clinical trial. For such a data set, a Markov transition model could be applied for statistical analysis, since it models dynamic features of transition patterns of the counting process. Markov chains can describe the phenomena that evolves through time, with applications ranging from biomedical research to many other scientific fields such as physics, engineering, sociology, and economics [6].

A noticeable problem with longitudinal data analysis is introduced by missing values. Within certain areas of biomedical research, e.g., drug dependence, HIV, and cancer, the statistical analysis is plagued by a large amount of missing items. The feature of incompleteness is related to the chaotic nature of the clinical disorders. For example, drug abusers in a study frequently missed their scheduled clinic visits or dropped out of studies prematurely, leading to proportions of missing values as large as 70% toward the termination of the study period [7]. High levels of incompleteness usually falsify assumptions that missing data may be ignored [22]. Even in a randomized-controlled clinical trial, the presence of missing values after randomiza-

tion can complicate standard complete-data analysis approaches; missing responses can occur at different rates and with different reasons for different conditions. Data analyses that ignore missing data are apt to introduce biases during significance testing.

In Albert and Follmann [1], an extended version of the above Markov transition model was proposed to handle nonignorable missing values in a binary longitudinal data set. This model introduced shared random-effects in order to link the propensity of transition between measurement states and the probability of being observed, intermittently missed, or dropped out. By jointly modelling the transitional features for observed binary repeated measures and the 3-category missingness indicators, random-effects Markov transition models (REMTMs) provide meaningful clinical interpretations on the dynamic change of cocaine dependence and useful inferences on the patterns and mechanisms of missing data. Recently, we conducted some simulation studies that jointly supported the superior performance of REMTM over traditional,yet inadequate, Markov transition models in analyzing binary longitudinal data with nonignorable missing values. For certain practical binary data sets with large amounts of missing data, REMTM seems to be the only applicable choice with acceptable performance. Considering that Poisson-distributed repeated measures are also frequently encountered in biomedical research, REMTMs are sufficient for analyzing incomplete count repeated measures in the presence of nonignorable missing values. In drug dependence studies, for example, such count measures indicate numbers of

drug use within a certain period.

1.2 A Motivating Study

Before describing the fitting algorithms for REMTM, a brief description of a motivating example is offered. In Shoptaw et al. [18], a smoking cessation clinical trial was conducted to study the relative efficacy of contingency management and relapse prevention types of behavioral therapies when optimizing outcomes using nicotine replacement therapy. In this 2 (contingency management or not) $\times$ 2 (relapse prevention or not) 12-week study, 175 methadone-maintained tobacco smokers were randomized to receive one of the four resulting conditions; all received nicotine replacement therapy.The number of smoking episodes during the previous week was evaluated through self report. Thus, there are at most 12 weekly-reported counts for each participant. Main analyses using carbon monoxide levels from breath samples established that there was a significant treatment effect for contingency management but no effect fore relapse prevention and no interaction during the study period [22]. The four conditions were then collapsed into two: the contingency management group ($n = 90$) with smokers who received contingency management and the control group ($n = 85$) with smokers who did not receive contingency management. As a secondary statistical analysis in this article, we compared the numbers of smoking episodes between the treatment and control groups. As noticeable from the data matrix, there is moderate amount of missing values due to dropout: 43 subjects dropped out during

the study period (24.6%), causing 296 missing values. The proportion of intermittent missingness is very low: 20 total intermittent missing values (about 1%) observed on six subjects (3.4%).

<INSERT FIGURE 1 HERE>

In Figure 1, the numbers of smoking episodes in the treatment (i.e., contingency management) and control groups are plotted separately. The two groups have similar distributions of the response variable at the beginning period of the study, however the average number of smoking episodes in the treatment group decreased more quickly and to a lower level than the control group. This typical clinical trial data set shows the treatment assignment is an important predictor variable. The repeated count measures and incomplete observations are all common in clinical trials when testing treatment effects. The REMTM will be applied to this 2-group set of data.

2 Method and Model

2.1 Analysis of Incomplete Longitudinal Data

Given a longitudinal data set, the repeated measures are denoted by a matrix $Y = [y_{it}]$ where y_{it} indicates the t^{th} measure ($t = 1, \ldots, T$) collected on the i^{th} subject ($i = 1, .., n$). For the discussion, we restrict the longitudinal data to a balanced design with time-independent covariates (i.e., those measured at baseline). The matrix of

covariates, thus, can be denoted by $X = [x_{ij}]$ where $x_{i1}, ..., x_{ip}$ indicate p predictors collected at baseline for the i^{th} subject. In the presence of missing values, missingness patterns are denoted as $R = [r_{it}]$, a matrix with elements:

$$r_{it} = \begin{cases} 0 & \text{if } y_{it} \text{ observed} \\ 1 & \text{if } y_{it} \text{ intermittently missing} \\ 2 & \text{if } y_{it} \text{ missing after dropout} \end{cases}$$

Further, θ is used to denote the parameters modeling the repeated measures and ϕ is used to represent the parameters modeling the missingness mechanism. For each subject, the full likelihood function is the joint distribution of observed repeated measures (i.e., $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})^T$) and the vector of missingness indicators (i.e., $\mathbf{r}_i = (x_{i1}, \ldots, x_{ip})^T$), i.e.,

$$L(\theta, \phi | \mathbf{y}_i, \mathbf{r}_i, X_i) \propto P(\mathbf{y}_i, \mathbf{r}_i | X_i, \theta, \phi)$$

When determining the influence of missing data, a primary interest is to identify missingness patterns and missingness mechanisms, and their potential relationships with treatment conditions or other baseline factors. While missingness patterns indicate which data points are missing, missingness mechanisms explain why they are missing. In practice, missingness mechanisms refer to the underlying processes yielding missing values. Such mechanisms are usually partially known or completely hid-

den to investigators. By partitioning Y into (Y_{obs}, Y_{mis}), which respectively represent the observed values and the values that would be observed if they were not missing, missingness mechanisms reflect the association between (Y_{obs}, Y_{mis}) and R. When the missingness pattern (R) is not associated with the values of the underlying potentially missing data (i.e., Y_{mis}) – a condition that we call *ignorability* – it is possible to obtain correct inferences without modelling the missingness mechanisms.

Within the framework of outcome-dependent missingness (see Figure 2(A)), the joint distribution of $(\mathbf{y}_i, \mathbf{r}_i)$ is factored into the marginal distribution of $\mathbf{y}_i$ and the conditional distribution of $\mathbf{r}_i$ given $\mathbf{y}_i$, i.e., $P(\mathbf{y}_i, \mathbf{r}_i|X_i, \theta, \phi) = P(\mathbf{y}_i|X_i, \theta)P(\mathbf{r}_i|\mathbf{y}_i, X_i, \phi)$. Within this framework, the definition of ignorability was extensively discussed in the statistical community; see [16, 11, 17]. More specifically, when missing data are "missing at random" (MAR; i.e., $P[R|Y] = P[R|Y_{obs}]$) and the parameters of data (i.e., θ) are distinct from those of the missingness mechanism (i.e., ϕ), the missingness mechanisms can be ignored for likelihood-based inferences about θ. This is because the joint likelihood function $L(\theta, \phi)$ can then be factored as the product of the likelihood function for ϕ and the observed-data likelihood function for θ, i.e., $L(\theta, \phi|Y_{obs}, R) = L_1(\theta|Y_{obs})L_2(\phi|R, Y_{obs})$. As mentioned earlier, most longitudinal models are based on observed-data likelihood (i.e., $L_1(\theta|Y_{obs})) = P(R|Y_{obs}, \phi)$, thus requiring the condition of ignorability. More specifically, generalized linear mixed models require the assumption of MAR, Markov transition models require a special case of MAR where r_{it} depends on $(y_{it-1}, .., y_{i1})$, and marginal models with GEE

assume covariate-dependent MAR (i.e., $P[R|Y,X] = P[R|X]$) [22]. Time-naive methods usually assume that missing data are missing completely at random (MCAR; i.e., $P[R|Y] = P[R]$), which is usually too rigorous for practical data sets.

<INSERT FIGURE 2 HERE>

As seen in Figure 2, there are two other ways in defining the missingness mechanisms: shared-parameter missingness (Figure 2(B)) and pattern-mixture missingness (Figure 2(C)). Contrary to outcome-dependent missingness, pattern-mixture models assumes that the joint distribution of $(\mathbf{y}_i, \mathbf{r}_i)$ is factored into the marginal distribution of $\mathbf{r}_i$ and the conditional distribution of $\mathbf{y}_i$ given $\mathbf{r}_i$, i.e., $P(\mathbf{y}_i, \mathbf{r}_i|X_i, \theta, \phi) = P(\mathbf{r}_i|X_i, \phi)P(\mathbf{y}_i|\mathbf{r}_i, X_i, \theta)$. In other words, different distributions are assumed for repeated measures on subjects within different missingness patterns. For example, in cancer studies, individuals who have died during the study should be treated differently than those who are still alive at the end of the study . By sharing a common vector of parameters (i.e., ξ), a shared-parameter model assumes that the data $\mathbf{y}_i$ and missingness indicators $\mathbf{r}_i$ are conditionally independent of each other given ξ, i.e., $P(\mathbf{y}_i, \mathbf{r}_i|X_i, \theta, \phi) = \int P(\mathbf{r}_i|\xi_i, X_i, \phi)P(\mathbf{y}_i|\xi_i, X_i, \theta)d\xi_i$. In the case of shared-parameter missingness, the shared parameters can be either observed covariates or unobserved latent variables. For example, in a cancer study, we may observe that $\mathbf{y}_i$ and $\mathbf{r}_i$ are independent of each other within each age category, but are dependent of each other across all the age groups. In this case, age can be viewed as a confounder in determin-

ing the relationship between $\mathbf{y}_i$ and $\mathbf{r}_i$. When ξ corresponds to an latent variable, such as random-effects, the missingness mechanism is also called informative [22], which is a special case of nonignorability. For informative missingness, structure equation models [9] potentially provide a tool for analysis.

It has become intuitive that among each of the three missingness modeling setting, ignorability can be achieved so long as there are no marginal association between Y_{mis} and R conditionally on Y_{obs}. Among the three cases, outcome-dependent models and pattern-mixture model have been studied intensively. In this article, we use the shared-parameters model, which will be implemented by a Markov transition models with shared random effects, to analyze the effects of contingency management on reducing cigarette smoking.

2.2 Random-Effects Markov Transition Model for Repeated Count Measures

For longitudinal data with Poisson-distributed count measures and informative missing values, REMTM offers a strategy for implementing the shared-parameter models where random effects are the shared parameters. This model can be viewed as a natural extension of the REMTM for binary longitudinal data [21]. REMTM first assumes that complete data, ($\mathbf{y}_i$, $\mathbf{r}_i$), are identically independently distributed across subjects ($i = 1, \ldots, n$), and for each subject i, the repeated count measures $\mathbf{y}_i$ are

conditionally independent of the missingness indicators $\mathbf{r}_i$ given the random effects ξ_i. Therefore, we can separately model the counting process $p(\mathbf{y}_i|\theta, \xi_i)$ and the missingness mechanism $p(\mathbf{r}_i|\phi, \xi_i)$.

2.2.1 Modelling the Counting Process

To model the counting process, the first-order Markov chain is assumed for $Bfy_i = (y_{i1}, \ldots, y_{iT})^T$, where on any time point, y_{it} is independent of $(y_{i1}, \ldots, y_{it-2})^T$ given the previous observation y_{it-1}. A random intercept effect is used to capture the baseline heterogeneity across subjects. In this random-intercept Markov transition model, we are interested in the transition probability of the Poisson-distributed count measures. Such probability depends on the covariates under investigation and a random intercept, i.e.,

$$P(y_{it}|x_{it}, y_{it-1}, \xi_i) = \frac{\lambda_{it}^{y_{it}}}{y_{it}!} e^{-\lambda_{it}}$$

with $\lambda_{it} = E(y_{it}|x_{it}, y_{it-1}, \xi_i)$ connected to covariates x_{it} and random effect ξ_i $N(0, \sigma^2)$ through a linear regression model using a link function $\log(\cdot)$,

$$\log(\lambda_{it}) = x_{it}\beta + (\log(\max(1, y_{it-1}) - x_{it-1}\beta)\alpha + \xi_i.$$

This article solely deals with baseline covariates that do not change with time, i.e., $x_{it} = (x_{i1}, \ldots, x_{ip}))$. Time-varying covariates can be easily implemented into the above Poisson regression model. Here, β contains the fixed parameters, which are

of the most interest in making inferences on the covariates effect (e.g., treatment efficacy). The parameter α indicates the influence of the previous counts through the logarithm of the residual, $(max(1, y_{it-1}) - x_{it-1}\beta)$, where $max(1, y_{it-1})$ is used to ensure a positive value for logarithm.

2.2.2 Modelling the Missingness Mechanism

By viewing the missingness indicator matrix R as a special form of categorical responses, we can model the missingness mechanism by a multinomial-logit Markov transition model. Still, we adopt the first-order Markov chain assumption to calibrate the transitional probabilities $P_{ij} = Pr(r_{it} = j | r_{it-1} = i)$ between any consecutive 3-category missingness indicators, r_{it-1} and r_{it} $(i = 0, 1, 2;\ j = 0, 1, 2)$. Determined by certain restrictions, the following transitional probabilities would be always equal to zero: $P_{12},\ P_{20},\ P_{21}$. For other combinations of r_{it-1} and r_{it}, the transitional probabilities are calculated in the following manner. First, if the previous count measure is observed (i.e., $r_{it-1} = 0$), then the "current' one could be observed, intermittently missing, or missing due to dropout. In this case, the 3-category multinomial-logit regression model can be used to calculate the transitional probabilities, i.e.,

$$P(r_{it} = j | \xi_i, x_{it}, r_{it-1} = 0) = \begin{cases} \frac{1}{1+\sum_{l=1}^{2} \exp(x_{it}\eta_l + \xi_i\gamma_l)} & \text{if } j = 0, \\ \frac{\exp(x_{it}\eta_j + \xi_{it}\gamma_j)}{1+\sum_{l=1}^{2} \exp(x_{it}\eta_l + \xi_{it}\gamma_l)} & \text{if } j = 1 \textit{ or } 2. \end{cases}$$

Second, if the previous count measure is intermittent missing, then the current one

may only be either observed or intermittently missing. Correspondingly, a logistic regression model can be used for calculating P_{10} and P_{11}, i.e.,

$$P(r_{it} = j|\xi_i, x_{it}, r_{it-1} = 1) = \begin{cases} \frac{1}{1+\exp(x_{it}\eta_1+\xi_i\gamma_1)} & \text{if } j = 0, \\ \frac{\exp(x_{it}\eta_1+\xi_{it}\gamma_1)}{1+\exp(x_{it}\eta_1+\xi_{it}\gamma_1)} & \text{if } j = 1. \end{cases}$$

Third, for the absorbing state 2, we would always have $P(r_{it} = 2|\xi_i, x_{it}, r_{it-1} = 2) = 1$. By denoting T_i as the time for the last observed measurement for subject i, special considerations should be paid to the last observed measures y_{iT_i}, for which we always have $P(r_{iT_i} = 0|r_{iT_i-1} = 1) = 1$. This is because for any individual, if the measure at time $T_i - 1$ is intermittently missing, the one at time T_i must be observed. In the above logit and logistics models, regression coefficients η_1 and η_2 respectively indicate whether intermittent missingness and dropout depend on covariates, while coefficients γ_1 and γ_2 respectively indicate whether intermittent missingness and dropout are informative (i.e., nonignorable).

2.3 Bayesian Inference using MCMC

After setting up models for the counting process and the missingness mechanism, we need to estimate the parameters in the above models. In [22], a maximum likelihood method was adopted to estimate the parameters of REMTM for binary longitudinal data. Similarly, we can optimize the REMTM likelihood function for the count data,

i.e.,

$$L(\theta, \phi | Y_{obs}, X, R) \propto \int \prod_{i=1}^{n} \{ \prod_{t=1}^{T_i} p(y_{it} | x_{it}, y_{it-1}, \xi_i, \theta) \} \{ \prod_{t=1}^{T} p(r_{it} | \xi_i, x_{it}, r_{it-1}, \phi) \} p(\xi_i) d\xi_i$$

where $\theta = (\alpha, \beta)$ representing the parameters of the Poisson regression model (see equations (2)-(3)), $\phi = (\eta_1, \eta_2, \gamma_1, \gamma_2)$ representing the parameters related to missingness mechanisms (see equations (4)-(5)), and $p(\xi_i)$ is the pdf of $N(0, \sigma^2)$. A problem with the maximum likelihood estimation comes from the optimization procedure that requires numerical integration and computationally-expensive calculation of transitional probabilities in the presence of intermittent missing values. In this article, we adopt the approach of Bayesian inference based on MCMC. By specifying various prior distributions and starting with different initial points for the parameter vector, we can loosen the analytical computation burden and verify whether the posterior distribution has multiple modes.

2.3.1 Bayesian Inference and Prior Specification

In the application of the Bayesian inference to the REMTM for count data, each parameter of $\psi = (\theta, \phi, \xi_i, \sigma^2)$ is viewed as a variable instead of a constant, certain prior distributions of the parameters are assumed, and the posterior distribution of

parameters is obtained using Bayes' theorem, i.e.,

$$P(\psi|Y,X,R) \propto P(Y|\psi)P(R|\psi)P(\psi)$$
$$= \prod_{i=1}^{n}\{\prod_{t=1}^{T_i} p(y_{it}|x_{it},y_{it-1},\xi_i,\theta)\}\{\prod_{t=1}^{T} p(r_{it}|\xi_i,x_{it},r_{it-1},\phi)\}P(\psi).$$

Using MCMC, we can sample the parameters from the posterior distribution and draw inferences using the center (median or mean) and variance of the stationary distribution. The prior distribution of each parameter can be a diffusion normal distribution (i.e., normal distribution with large variance) or a uniform distribution on certain intervals (i.e. flat prior). We assume flat priors for all the parameters except σ^2, which is specified by an inverse Gamma(a,b), i.e., $f(\sigma^2) = \frac{b^a \exp(-b/\sigma^2)}{(\sigma^2)^{a+1}\Gamma(a)}$. According to Bayes' theories, with enough data, the difference on the posterior distribution introduced by varying prior specification would be negligible as long as all the priors cover the whole range of the possible values of the parameters.

2.3.2 An Augmented Gibbs Sampler

As seen from the likelihood function in (6) and the posterior distribution in (7), it is required that the product of the transition probabilities (i.e., $L_i^y = \prod_{t=1}^{T_i} p(y_{it}|x_{it},y_{it-1},\xi_i,\theta)$) be computed. In the presence of intermittent missing values, the corresponding Markov chain $(y_{i1},\ldots,y_{iT_i})$ is broken into segments of observed values. Thus, the calculation of the whole product would require either integration or imputation. For

example, if $T_i = 4$ and y_{i2} is the only intermittent missing value, then we can obtain L_i^y from integration (i.e., $L_i^y = \int p(y_{i2}|y_{i1})p(y_{i3}|y_{i2})p(y_{i4}|y_{i3})dy_2$) or imputation (i.e., $L_i^y = \frac{1}{m}\sum_{l=1}^{m} p(y_{i2}^l|y_{i1})p(y_{i3}|y_{i2})p(y_{i4}|y_{i3})$ with m imputed values for y_{i2}). When there are more than one consecutive intermittent missing values, (e.g., $r_{i2} = r_{i3} = 1$), the option of integration becomes too complicated to be feasible and the method of imputation becomes more appealing. Considering that integration can be realized through Monte Carlo computation, which itself depends on imputation to deal with missing values, we propose to adopt the imputation approach to deal with the problem of intermittent missingness.

Once intermittent missing values are imputed, then standard MCMC steps can be used to simulate samples of the parameters ψ. Similar ideas were adopted by [**?**, 17] when they used data augmentation [**?**]in their multiple imputation . If we treat missing values as another group of parameters in an approximate sense, an augmented Gibbs sampler can be developed to draw samples for the parameters and missing values. This algorithms consist of two iterative steps. The first step can be called *imputation step*, where the missing values are updated by drawing from the conditional predictive distribution, i.e.,

$$y_i^{mis} \sim f(\mathbf{y}_i^{mis}|\mathbf{y}_i^{obs}, \mathbf{x_i}, \mathbf{r_i}, \psi).$$

The second step is called *parameter estimation step*, where the parameters are drawn

from the posterior distribution

$$\psi \sim P(\psi|Y, X, R).$$

This second step can be replaced by further Gibbs steps. Roughly speaking, at each iteration, we sample and update one parameter or one random effect from the distribution conditional on all the others and the imputed data. If this conditional distribution has a closed form, we sample from it directly. Otherwise, we have two approaches: the *adaptive rejection sampling* [8] and the *griddy Gibbs sampler* [13]. The basic ideas of the two sapling methods are illustrated by Figure 3.

<INSERT FIGURE 3 HERE>

When the conditional density function has a log-concave from, we use the adaptive rejection sampling to get efficient sampling results in the following two steps. First, we set the upper hull $u(x)$ and lower hall $l(x)$, which are piecewise linear functions respectively consisting of tangent lines and cords of the logarithm of the density function, $h(x) = \log(f(x))$, at selected points. Second, we sample a point from the cumulative density function determined by $u(x)$, and then update the upper and lower halls depending on whether the sampled value is accepted or rejected.

When the conditional density function does not have a log-concave form and there are no more efficient sampling methods available, we resort to the simple and intuitive griddy Gibbs sampler, which is based on the empirical distribution method. As

depicted by Figure 3(b), three steps are required to create a griddy Gibbs sampler. First, the range of the conditional density function should be decided up to a constant. Then, we divide the range with or without respect to the probability change to form a grid. Finally, we sample from the grid points by simple inverse sampling [**?**] or a more sophisticated method.

More specifically, to impute intermittent missing values in the first step of the augmented Gibbs sampler, the griddy Gibbs sampler would be used. For $i = 1, \ldots, n$ and $t = 1, \ldots, T_i$, if y_{it} is missing, then an imputation would be drawn using conditionally on the observed or imputed y_{it-1} and y_{it+1}, i.e.,

$$f(y_{it}^{mis}|y_{it-1}, y_{it+1}, \psi) \propto \frac{\lambda_{it}^{y_{it}}}{y_{it}!} e^{-\lambda_{it}} \lambda_{it+1}^{y_{it+1}} e^{-\lambda_{it+1}}.$$

For the parameter estimation step, the adaptive rejection sampling would be used to draw parameters ψ in the following order. The counting process parameters α and β of the Poisson regression model are drawn using,

$$\begin{aligned} f(\alpha|\psi_{\setminus\alpha}, Y_{mis}) &\propto \prod_{i=1}^{n} \prod_{t=1}^{T_i} P(y_{it}|x_{it}, y_{it-1}, \xi_i) \\ f(\beta|\psi_{\setminus\beta}, Y_{mis}) &\propto \prod_{i=1}^{n} \prod_{t=1}^{T_i} P(y_{it}|x_{it}, y_{it-1}, \xi_i) \end{aligned}$$

where $\psi_{\setminus\alpha}$ refers to the sub vector of ψ excluding α and "\' has the similar meaning for $\psi_{\setminus\beta}$ and other notations that follow. For the parameters η and γ of the multinomial

logit and the logistic regression models that describe the missingness mechamism, the following conditional distributions are used.

$$\begin{aligned} f(\eta|\psi_{\setminus\eta}, Y_{mis}) &\propto \prod_{i=1}^{n}\prod_{t=1}^{D_i} P(r_{it}|\xi_i, x_{it}, r_{it-1}) \\ f(\gamma|\psi_{\setminus\gamma}, Y_{mis}) &\propto \prod_{i=1}^{n}\prod_{t=1}^{D_i} P(r_{it}|\xi_i, x_{it}, r_{it-1}) \end{aligned}$$

For random intercept effects, ξ_i, and their variance σ^2, we use the following conditional distributions to draw samples:

$$f(\xi_i|\psi_{\setminus\xi_i}, Y_{mis}) \propto \prod_{t=1}^{T_i} P(y_{it}|x_{it}, y_{it-1}, \xi_i) \prod_{t=1}^{D_i} P(r_{it}|\xi_i, x_{it}, r_{it-1}) \exp\{\frac{\xi_i^2}{2\sigma^2}\}$$

$$f(\sigma^2|\psi_{\setminus\sigma^2}, Y_{mis}) \propto (\sigma^2)^{-n/2} \exp\{-\sum_{1}^{n} \frac{\xi_i^2}{2\sigma^2}\} \frac{b^a \exp(-b/\sigma^2)}{\sigma^{2(a+1)}}$$

It is not difficult to verify that the above density functions for α, β, η, γ, and ξ_i are log-concave by computing the second order derivatives of the corresponding log-transformed functions. For example, $(\log(f(\alpha|\psi_{\setminus\eta}, Y_{mis})))'' \propto \sum_{i=1}^{n}\sum_{t=1}^{T_i} -\lambda_{it}(\log(max(1, y_{it-1}) - x_{it-1}\beta)^2 < 0$. By denoting $s = \log(\sigma^2)$, it can be shown that $f(s|\psi_{\setminus s}, Y_{mis}) \propto \exp\{-s\frac{n+2}{2} - \frac{\sum_1^n \xi_i^2}{2\sigma^2}\}$ whose second derivative is also negative.

3 Application

3.1 A Simulation Study

To verify the validity of REMTM and the augmented Gibbs sampler for incomplete Poisson-distributed repeated measures, we performed a simulation study where the sample size was set at three levels: $n = 100$, 300, or 500, to reflect a typical applied scenario seen in Phase II and III clinical trials. We fixed the maximum number of possible repeated measures at $T = 12$, which is the same as the number in the motivating study in Section 1.2, and randomized each of the n subjects to receive an imaginary 'treatment' or 'control' group with 50% probability. Then, to generate an incomplete data set, two steps were applied . First, a complete data matrix was simulated using the Poisson Markov transition model in (2)-(3) with fixed parameters $\alpha = 0.3$ and $\beta = (\beta_0, \beta_1) = (0.5, -0.5)$ (where $x_i = (1, Trt_i)$, $Trt = 1$ or 0 indicating 'treatment' or 'control' condition for subject i) and random effects $\xi_i \sim N(0, 0.36)$ (i.e., $\sigma^2 = 0, 36$). Then, we created intermittent missing values and dropouts using the multinomiial-logit and logistic regression models with parameters $\phi = (\eta_1 = (\eta_{10}, \eta_{11}), \eta_2 = (\eta_{20}, \eta_{21}), \gamma_1, \gamma_2)$ set to reflect 16 scenarios as shown in Table 1, where we fixed $\eta_{10} = \eta 20 = -1.0$ and let each of the rest four parameters vary between two levels. For example, when $\eta_{11}, \eta_{21}, \gamma_1$, and γ_2 are all set to zero, it corresponds to an incomplete data set where both intermittent missingness and dropout are ignorable and not treatment-dependent. The setting, '$\eta_{11} = -0.5, \eta_{21} = 0, \gamma_1 = 0,$

and $\gamma_2 = 0.4$,' corresponds to the scenario where intermittent missingness is ignorable and treatment-dependent, while dropout is nonignorable (or informative) and not treatment-dependent.

Repeating the above steps generated 20 data sets for each sample size and one combination of the missingness parameters ϕ. The augmented Gibbs sampler was then applied to fit an REMTM for each data set. When summarizing the simulation results, we paid more attention to the estimation of β_1, the parameter of most interest. The overall performance of the augmented Gibbs sampler was satisfactory. In Table 1, the averaged biases and variances for β_1 over the estimators for the 20 data sets for sample size $n = 300$ are listed, with the bias within 0.1 for most combinations of the missingness parameters. The worst cases, with the biases at 0.162 and 0.107, occurred when the data sets were generated using $\gamma_1 = \gamma_2 = 0$, corresponding to ignorable intermittent missingness and dropout. In these cases of ignorability, the assumption of shared-parameter missingness was violated, but the biases were still acceptable. This also shows REMTM to be generally robust.

<INSERT TABLE 1 HERE>

During the simulation study, we found that different parameters have different converging speed. We monitored the sampling process by comparing multiple chains and used some statistical criteria as suggested by [4] to judge the convergence. After all of them converge, we used the means of the sampled posterior distributions as the

point estimators. A typical posterior distribution of β_1 and its corresponding multiple chains for the sampling process are plotted in Figure 4, from which it is noticable that the posterior distribution of β_1 is unimodal and the estimated treatment effect size is very close to its true value (i.e., β_1=-0.5).

<INSERT FIGURE 4 HERE>

3.2 Analyze the Motivating Study

The model REMTM with augmented Gibbs sampling algorithm was also used to analyze the incomplete smoking episode data set. As mentioned earlier, there were very few intermittent missing values in the data, which precluded stable estimators of the parameters related to intermittent missingness mechanism. To circumvent this limitation, we decided to create artificial intermittent missing values. By randomly deleting certain amounts of observed values and increasing the levels of intermittent missingness, we aimed to investigate the sensitivity of the estimator of β_i (i.e., treatment efficacy). If we created more missing values and all the estimates suggested the same conclusion, we would have more confidence in accepting the conclusion. The creation of artificial missingness also provided a device to verify the validity of REMTM and the MCMC algorithm. If they provided similar solutions, then the intermittent missing values could be inferred as being ignorable and independent of the treatment condition. Four levels of intermittent missingness were specified: 5%, 10%, 15%, and 20%. At each level, twenty sets of intermittent missing values were

randomly generated by converting that proportion of missingness indicators from 0 to 1. By setting random initial values for the parameters to trigger the augmented Gibbs sampler to analyze each data set, the estimators were averaged to make final inferences.

<INSERT TABLE 2 HERE>

Since, the parameter β_1 is of interest, we listed all the four sets of point estimators and standard variance in Table 2. From this table, it is seen that the variances of β_1 across the 20 data sets for each intermittent missingness level becomes larger as the level increases. Nonetheless, the bias remains small across the four levels of intermittent missingness and all the corresponding 95% posterior credible intervals exclude zero. In other words, the estimators jointly supported a significant favorable effect for contingency management, the active treatment condition, in helping the methadone-maintained smokers in this study reduce cigarette smoking.

<INSERT TABLE 3 HERE>

Table 3 depicts all the parameter estimates (averaged across the 20 incomplete data sets) when the fake missingness rate is set at 5%. The point estimators and 95% posterior credible intervals for η, and γ jointly suggest that both intermittent missingness and dropout are ignorable and independent of the covariate. This is coincidental with the fact that intermittent fake missing values were created completely at random. Without direct or auxiliary information at hand to infer the possible

dropout mechanism, we were unable to verify whether the suggested ignorability of dropout by the fitted REMTM was correct or not.

4 Discussion

For various longitudinal data in biomedical research, three groups of modeling strategies are popularly used: generalized linear mixed models, marginal models with GEE, and transition models. In the presence of missing values, which is a common problem in practice, all these models assume that missing data are ignorable and draw inferences solely based on observed data. In many practical data sets, however, ignorability is a very strong assumption and analysis ignoring the patterns and mechanism of missingness would bias hypothesis testing, especially if missing values are nonignorable. As seen from Figure 2, nonignorablility could be further classified into three cases: outcome-dependent, pattern-mixture, and shared-parameter. The first two cases were extensively studied in the statistical community and handed by various forms of selection models and pattern-mixture models (see [22]) for a detailed review). In this article, a specific implementation of shared-parameter model called REMTM was proposed to analyze incomplete Poisson-distributed repeated measures with potentially informative missingn values due to occasional omission or premature withdrawal. A MCMC fitting algorithm called augmented Gibbs sampler was also developed. Both the simulation study and the practical application verified the

satisfactory performance of REMTM and the fitting algorithm.

When designing the simulation study, we tried to mimic the setting of a placebo-controlled clinical trial. The augmented Gibbs sampler for REMTM shows acceptable robustness in estimating treatment efficacy. Under various scenarios of missingness mechanism, the fixed parameter of interest was well estimated. When applying REMTM to the smoking cessation data set which contains self-reported repeated measures counting the number of episodes of tobacco using, the various fitted models consistently support the efficacy of contingency management. In the analysis, we adopted the idea of artificial deletion of observed values in order to obtain intermittent missing values. As more and more observed values became randomly excluded from analysis, the observed information became smaller and smaller. However, the steadfast sign and significance of the estimated treatment effect parameter. This provided very strong evidence for accepting the conclusion of treatment efficacy. This method of artificial deletion potentially has practical value in statistical analysis in that the method is analogous to assaying the strength of the constituents of a solution by using dilution methods. The variable strength (i.e., the intervention) could be measured as the amount of missing data introduced to yield a non-significant result when compared to the control.

REMTM is a case of shared-parameter model. Because of the feature of conditional independency, measurement models and missingness models can be considered conditionally independently. There exists many other ways to implement the shared-

parameters models. Conditioned on the same group of shared-parameters, mixed models and marginal models can be used to study the joint distribution of longitudinal data or missingness mechanisms. When shared parameters include observable covariates, then this shared-parameter model can be used to analyze epidemiological longitudinal data sets with possibly informative missingness. In these models, the observed shared-parameters can be confounders or derived scores such as propensity scores [15] that address the problem of confounding. Adopting the idea of multiple partial imputation [22], we can first impute the intermittent missing values and then manage only dropouts using a simpler version of REMTM where a random-effects logistic regression model can be used to model the dropout mechanism. This model was also used to analyze the smoking cessation data set and supported a consistent conclusion efficacy of contingency management.

A noticeable difficulty of incomplete data analysis comes from model selection. For a practical data set, there is often not enough direct or indirect evidence for choosing a proper model among many choices, e.g., shared-parameter models, selection models, or pattern-mixture models. In this situation, we recommend adopting the strategy of sensitivity analysis, where several models with various assumptions on the missingness mechanism are applied to the same data to investigate the sensitivity of the parameter estimation. Only if all the models provide conclusions that are not conflicting with one another, would we have confidence in accepting them. Otherwise, further investigation is likely needed.

In this article, we restricted our discussion to longitudinal data sets with a balanced design and time-independent covariates, but it is not difficult to modify the model and algorithms to handle uneven repeated measures and time-varying covariates. Considering that other types of repeated measures such as continuous and categorical measures are also popular in practice, we are currently extending our MCMC algorithm and Markov transition models to address them. To conduct sensitivity analysis, various modeling strategies should be developed and implemented into software packages. Recently, a software package called MPI has implemented both selection models and pattern mixtures models [21]. The augmented Gibbs sampler for the REMTM in this paper was programmed in C++ and the codes are available by request.

ACKNOWLEDGEMENT

This work was partially supported by the National Institute of Drug Abuse through an SBIR contract N44 DA35513 and three research grants: R03 DA016721, R01 DA09992 and P50 DA18185. We especially thank Hamutahl Cohen for her editorial assistance.

References

[1] Albert PS,Follman DA. A Random Effects Transition Model for Longitudinal Binary Data with Informative Missingness.*Statistica Neerlandica* (2003) Vol. **57**, 100–111,2003.

[2] Brown H, Prescott R. *Applied Mixed Models in Medicine*. John Wiley and Sons,1999.

[3] Bryk AS, Raudenbush SW. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications; 2nd edition,2001.

[4] Cowles MK, Carlin BP. Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association*, **91**, 883-904,1996.

[5] Diggle PJ, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. Oxford University Press,1994.

[6] Durrett R. *Essentials of Stochastic Processes*. New York: Springer-Verlag,1999.

[7] Follman D, Wu M. An Approximate Generalized Linear Model with Random Effects for Informative Missing Data. *Biometrics* **51**, 151-168,1995.

[8] Gilks WR, Wild P. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* 41, pp 337-348,1992.

[9] Kaplan D. *Structural Equation Modeling: Foundations and Extensions*. Sage Publication,2000.

[10] Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22,1986.

[11] Little RJA, Rubin DB. *Statistical Analysis with Missing Data*, 2^{nd} edition, New York: John Wiley,2002.

[12] McCulloch CE,Searle SR. *Generalized, Linear, and Mixed Models*. John Wiley and Son,2001.

[13] Ritter C, Tanner MA. Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler. *Journal of the American Statistical Society*, 87, 861–868,1992.

[14] Robert CP,Casella G. *Monte Carlo Statistical Methods*, 2nd edition. Springer, New York,2004.

[15] Rosenbaum PR,Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* **70**: 41-55,1983.

[16] Rubin DB. Inference and missing data, *Biometrika* **63**. 581-582.1976.

[17] Schafer JL. *Analysis of Incomplete Multivariate Data*, London: Chapman & Hall,1997.

[18] Shoptaw S, Rotheram-Fuller E, Yang X, Frosch D, Nahom D, Jarvik ME, Rawson RA, Ling W. Smoking cessation in methadone maintenance. *Addiction* **97**, 1317-1328,2002.

[19] Singer SCJ, Willett JB. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford Press,2003.

[20] Tanner MA, Wong WH. The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association* **82**, 528-550,1987.

[21] Yang X, Belin RT, Zhang QJ, Shoptaw S. Incomplete Longitudinal Data Analysis with Multiple Partial Imputation. *Technical Report 2003-1*, BayesSoft, Inc,2003.

[22] Yang X, Shoptaw S. Assessing Missing Data Assumptions in Longitudinal Studies: An Example Using a Smoking Cessation Trial. *Drug and Alcohol Dependence*, **77**, 213-225,2005.

[23] Zeger SL, Liang KY, Self SG. The analysis of binary longitudinal data with time-independent covariates.*Bionetrika*, **72**, 31-38,1985.

η_1	η_2	γ_1	γ_2	Bias($\bar{\hat{\beta}}_1$)	Var($\bar{\hat{\beta}}_1$)($\cdot 10^{-3}$)
0	0	0	0	-0.162	0.268
0	0	0	0.4	-0.044	0.278
0	0	0.5	0	-0.059	0.202
0	0	0.5	0.4	-0.045	0.629
0	-1	0	0	-0.038	0.517
0	-1	0	0.4	0.034	0.288
0	-1	0.5	0	-0.014	0.562
0	-1	0.5	0.4	-0.028	0.102
-0.5	0	0	0	0.028	0.036
-0.5	0	0	0.4	0.059	0.025
-0.5	0	0.5	0	0.029	0.136
-0.5	0	0.5	0.4	0.016	0.149
-0.5	-1	0	0	0.107	0.055
-0.5	-1	0	0.4	0.031	0.500
-0.5	-1	0.5	0	0.086	0.602
-0.5	-1	0.5	0.4	0.025	0.158

Table 1: Biases and Variances of the Averaged Treatment Effect Estimators ($\bar{\hat{\beta}}$) in the Simulation Study.

Statistics	5% missing	10% missing	15% missing	20% missing
From the posterior distribution of β_1				
$\Diamond$ average of 20 means	-0.94	-1.05	-1.00	-1.03
$\Diamond$ average of 20 std's	0.17	0.16	0.12	0.09
$\Diamond$ variance of 20 means	0.07	0.11	0.12	0.14

Table 2: Estimation of Treatment Effect Parameter (β_1) in the Smoking Cessation Study.

Parameter	Estimate	Std. Dev.	95% C. I.
Possion Count Transition			
◊ Dependence(α)	0.42	0.01	(0.40 , 0.43)
◊ Intercept(β_0)	2.19	0.10	(1.99 , 2.40)
◊ Treatment(β_1)	-0.94	0.17	(-1.26 , -0.61)
Intermittent Missing			
◊Intercept(η_{10})	-0.51	0.08	(-0.67 , -0.34)
◊Treatment(η_{11})	-0.09	0.12	(-0.33 , 0.14)
◊random effect(γ_1)	0.03	0.05	(-0.06 , 0.13)
Dropout Missing			
◊Intercept(η_{20})	-2.75	0.24	(-3.21 , -2.28)
◊Treatment(η_{21})	-0.02	0.33	(-0.66 , 0.62)
◊Random effect(γ_2)	0.17	0.15	(-0.12 , 0.47)
variance of random effect σ_2	1.66	0.23	(1.21 , 2.10)

Table 3: Parameter Estimation with Standard Deviations and 95% Posterior Credible Intervals for the Smoking Cessation Data Set with 5% Randomly Generated Intermittent Missing Values.

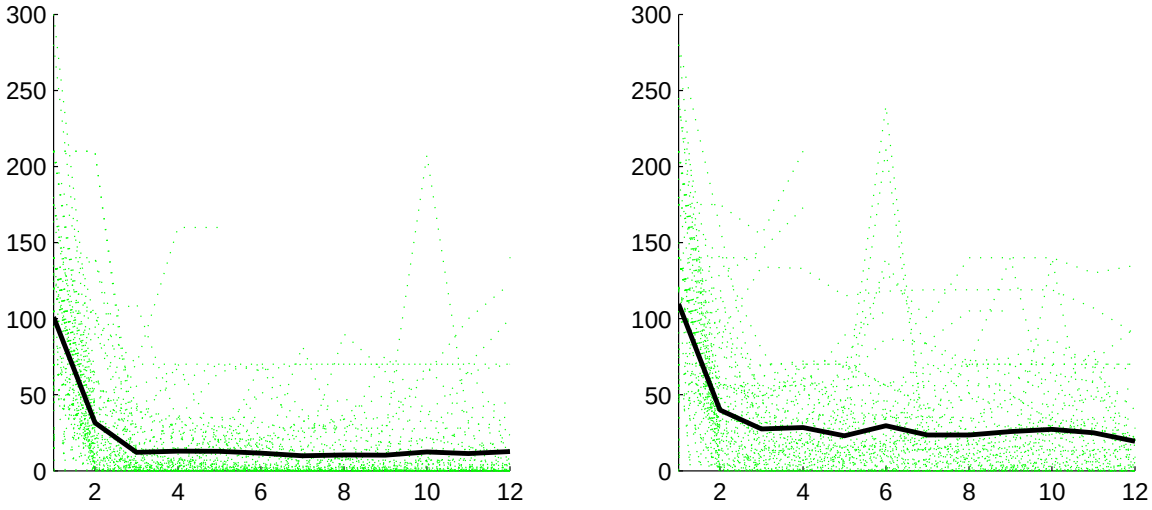


Figure 1: Repeated Count Measures Over the 12-week Study Period for the Smoking Cessation Clinical Trial.

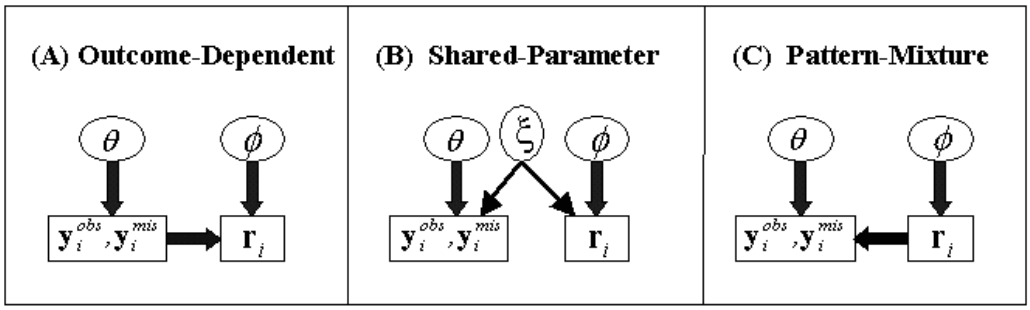
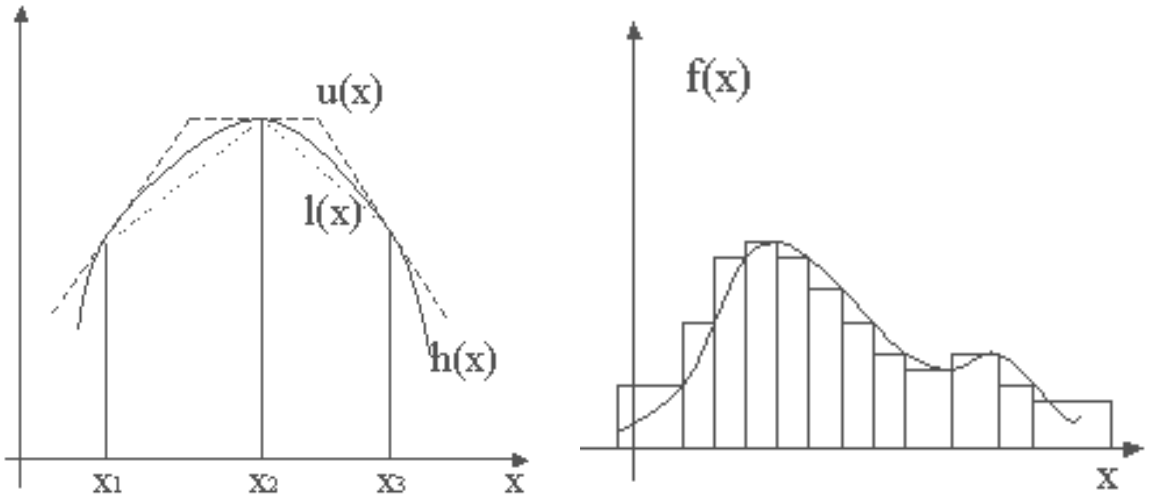
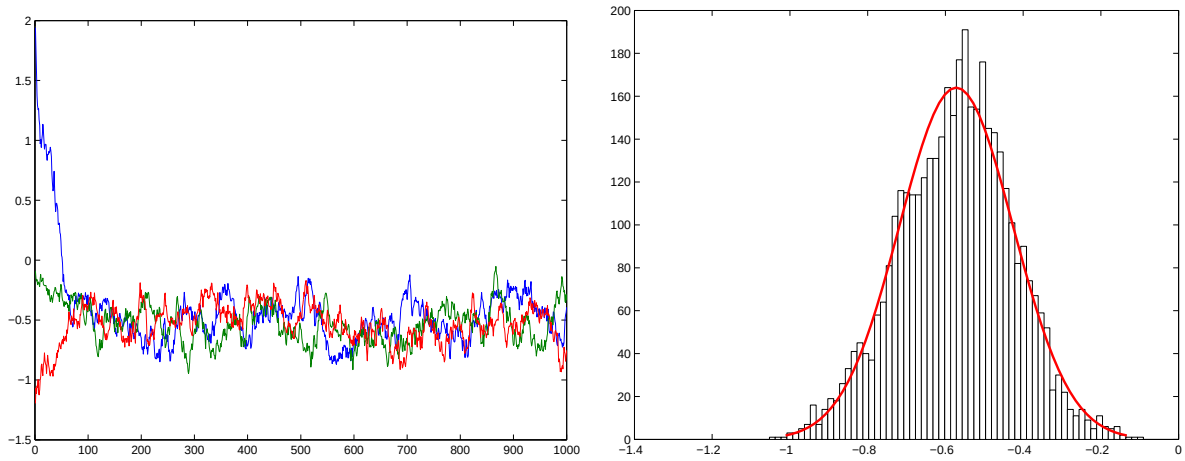


Figure 2: Three Representations of Missingness Mechanisms



(a) (b)

Figure 3: Adaptive Rejection Sampling Method and Griddy Gibbs Sampling Method.



(a) (b)

Figure 4: The Sampling Process and Posterior Distribution of Parameter β_1 in the Simulation Study.

Figure Captions

Figure 1. Repeated Count Measures Over the 12-week Study Period for the Smoking Cessation Clinical Trial.

The left graph plots the repeated measured number of smoking episodes for the 90 smokers in the treatment group who received contingency management. The right graph plots the repeated measured number of smoking episodes for the 85 smokers in the control group who did not received contingency management. In both plots, the y-axis indicates the number or counts of tobacco use for the previous week, the x-axis corresponds to the week numbers (1-12). The thick solid curves depict the mean profiles in the two groups, while the dashed curves represent the individual profiles.

Figure 2. Three Representations of Missingness Mechanisms.

Three ways in modeling incomplete data are depicted here. Parameters and symbols in the figures are defined as: $\mathbf{y}_i = (\mathbf{y}_i^{obs}, \mathbf{y}_i^{mis})$ —-

observed and missing values for subject i; $\mathbf{r}_i$ —- missingness indicator for repeated measures on subject i; θ —- parameters of data; ϕ —- parameters of missingess indicators; and ξ —- parameters shared by data and missingness indicators.

Figure 3. Adaptive Rejection Sampling Method and Griddy Gibbs Sampling Method.

The left side figure shows the idea of Adaptive Rejection Sampling with functions defined as: $h(x) = \log(x)$ — the logarithm of a statistical density function $f(x)$; $u(x)$ — upper hall of $h(x)$; and $l(x)$ — lower hall of $h(x)$. The right side figure depicts the idea of Griddy Gibbs Sampling within which the density function is approximated by rectangles.

Figure 4. The Sampling Process and Posterior Distribution of Parameter β_1 in the Simulation Study.

The left side figure draws three chains of the augmented Gibbs sampler, each starting from randomly selected points. It is seen that after around

100 iterations, the Gibbs sampler converges. The right side histogram depicts the posterior distribution of the parameter of interest (i.e., β_1). From the histogram, we can see that the mean value of the posterior distribution is very close to the true value (i.e., -0.5).