A review of “Marginalized Transition Models and Likelihood Inference for Longitudinal Categorical Data” by Patrick Heagerty.

Elizabeth Johnson
Department of Biostatistics at the Johns Hopkins University Bloomberg School of Public Health.

Summary: Heagerty proposes a marginalized transition model for binary longitudinal data that permits a likelihood based approach to associating a mean response to a set of covariates with specification of a serial dependence structure via a q-order Markov model. He shows that his method is a natural extension of the first-order Markov models proposed by Azzalini and compares the efficiency of his method to that of GEE.
1 Introduction

Studies in which data are collected over time or over repeated experiments are becoming increasingly popular. For example, in the Indonesian Children’s Health Study, over 3000 children were medically examined every three months for up to six visits. At each visit, it was determined if the child suffered from respiratory or diarrhoeal infection or xerophthalmia (Sommer, 1982). We may be interested in modeling the probability of infection as a function of gender, age, vitamin usage and other dietary indicators. It may also be the case that once a child becomes infected, they tend to stay infected or that those with no infection may have a natural defense against the infection and will tend to never become infected. Therefore, we believe that infection status recorded at any visit depends not only on the covariates for the child but also on the child’s previous infection status. In situations like this, where the response is assumed to be a function of past responses and a set of covariates, the transition model, an extension of the generalized linear model, is appropriate. This paper reviews a new methodology proposed by Patrick Heagerty in his paper entitled “Marginalized Transition Models and Likelihood Inference for Longitudinal Categorical Data” (Heagerty, 2002).

This paper will i) give a brief summary of the specification and estimation procedures for transition models, ii) briefly describe the first-order Markov model proposed by Azzalini (1994), iii) define the marginalized multilevel models proposed by Heagerty and Zeger (2000), iv) define the marginalized transition model of order $q$ proposed by Heagerty and describe its interpretation, estimation, and efficiency relative to GEE methods.

*Note, I realize that I was only supposed to review the one paper, but I had to read the three papers referenced above and am probably giving a brief review of the three papers instead of a long review of the one, I didn’t know how to bring all the information together otherwise.
2 Transition Models

Transition models are used to estimate a function of the mean response conditional on past responses and set of covariates, see Diggle, Heagerty, Liang and Zeger (2003, chapter 10). Let $y_{ij}$ be the binary response and $x_{ij}$ the vector of (possibly time-varying) covariates for subject $i$ at time $j$, where $i = 1, 2, ..., n$ and $j = 1, 2, ..., n_i$. For the remainder of the review, we assume that the times $j$ are equally spaced. Transition models of order $q$ condition the current response on the previous $q$ responses and the set of covariates as follows:

$$g(\mu_{ij}) = x_{ij}^{T}\beta_q + \sum_{l=1}^{q} y_{ij-l} \alpha_{j-l}$$  \hspace{1cm} (1)$$

$$\mu_{ij} = E(Y_{ij}|x_{ij}, Y_{i1}, ..., Y_{ij-q})$$  \hspace{1cm} (2)$$

where $g(.)$ is an appropriate link function, $\beta_q$ relates the average response to the covariates after adjusting for the previous $q$ responses, and the $\alpha$’s describe the dependence of the current response on the previous $q$ responses. For instance, $\hat{\beta}_{q,1}$ is the estimated change in the log odds of $Y_{ij}=1$ per unit change in $x_{ij,1}$ given response history $(Y_{i1}, Y_{i2}, ..., Y_{ij-q})$. It is important to note here that the interpretations of the regression coefficients are specific to the order of the model and are conditional on the previous $q$ responses. This model is not sufficient if we wish to make marginal inferences, i.e. inferences that describe the association of $Y_{ij}$ on $x_{ij}$ alone.

Estimates of the regression coefficients $(\hat{\beta}_q, \hat{\alpha})$ can be obtained via iterative weighted least squares. If the correct model is specified, we obtain consistent inferences for $(\beta_q, \alpha)$. If the conditional mean is correctly specified but we mis-specify the dependence structure, we can sometimes obtain consistent inferences for $(\beta_q, \alpha)$ by using a robust variance estimator, but the
interpretation of $\hat{\beta}_q$ is then questionable.

3 Marginal First-Order Markov Model for Binary Responses by Azzalini

Azzalini (1994) proposed a marginal first order Markov model for binary response data which allows estimation of $\mu_{ij}^M = E(Y_{ij}|x_{ij})$ while modeling the serial dependence of $Y_{ij}$ on $Y_{ij-1}$ via the odds ratio. Let the marginal mean model be specified as:

$$g(\mu_{ij}^M) = x_{ij}'\beta^M$$

(3)

The first order Markov model assumes that the current response is only dependent on the previous response. Define the two transition probabilities as: $p_{ij,0} = E(Y_{ij}|Y_{ij-1} = 0)$ and $p_{ij,1} = E(Y_{ij}|Y_{ij-1} = 1)$. The two assumptions required in the Azzalini model are: i) the model for $\mu_{ij}^M$ is defined so that the transition probabilities satisfy

$$\mu_{ij}^M = p_{ij,1}\mu_{ij-1}^M + p_{ij,0}(1 - \mu_{ij-1}^M)$$

(4)

and ii) the transition probabilities satisfy the assumption on the pairwise odds ratio used to describe the strength of the serial dependence.

$$\Psi_{ij} = \frac{p_{ij,1}/(1 - p_{ij,1})}{p_{ij,0}/(1 - p_{ij,0})}$$

(5)

The simplest form for the serial dependence assumes a constant $\Psi_{ij}$ for all $i,j$, but a model for the $\Psi_{ij}$ as a function of $j$ or the covariates may also be specified.

The transition probabilities can be expressed as a function of $\mu_{ij}^M$ and $\gamma_{ij} = \log(\Psi_{ij})$ and hence we can derive the likelihood function. Azzalini provides the details on the calculations for
the maximum likelihood estimation and establishes the independence of the marginal mean and
odds ratio for the model that specifies a constant odds ratio across all times.

Azzalini also addresses the problem of missing data in his proposed model. He suggests
replacing the 1-step transition probabilities with m-step transition probabilities when responses
are missing between times t and t-m.

4 Marginalized Multilevel Models by Heagerty and Zeger

In the paper entitled “Marginalized Multilevel Models and Likelihood Inference” by Heagerty and
Zeger (2000), hierarchical regression models are proposed that allow for a marginal model of
the mean response conditional on covariates only and a dependence model which defines the
association of observations within clusters or individuals. Consider the general case of a random
effects model; the traditional model is one where the mean response is conditioned on a function
of covariates and a set of unobserved factors represented by the random effect. Heagerty and
Zeger proposed the following multilevel model where estimation of the mean response is only
conditional on the covariates and the dependence model is conditional on the random effect. The
mean model is defined by (3) in the previous section and the dependence model is given by:

\[ g(\mu_{ij}) = \Delta(x_{ij}) + b_{ij} \]  \hspace{1cm} (6)

where \( b_{ij} \) is the random effect or the representation of the unobservable factors representing
heterogeneity in the population, and is typically specified to follow a normal distribution with
mean 0 and variance \( \sigma^2 \). The \( \Delta(x_{ij}) \) is an implicit conditional function which can be solved for
by relating the marginal mean to the random effects distribution via:
\[ g^{-1}(x_{ij}^\prime \beta) = \int g^{-1}(\Delta(x_{ij}) + b_{ij}) dF_\alpha(b_{ij}) \] (7)

The likelihood function can be specified under the assumption of conditional independence of \( Y \) and \( b_{ij} \) and the assumption that \( b_{ij} \) follows a mixing distribution known up to a finite parameter \( \sigma^2 \).

5 Marginal Transition Models of order \( q \) by Heagerty

The marginalized transition models of order \( q \) (MTM(\( q \))) are a special case of the marginalized multilevel models described in the previous section and in the case where \( q = 1 \), these models reduce to the model proposed by Azzalini. The marginal mean model is the same as (3), with the assumption that the regression model properly specifies the full covariate conditional mean such that \( \mu_{ij}^M = E(Y_{ij}|x_{ij}) = E(Y_{ij}|x_{i1}, ..., x_{im}) \) for time-varying covariates. For \( q = 1 \), the dependence model is given by:

\[ g(\mu_{ij}^C) = \Delta_{ij} + \gamma_{ij,1} y_{ij-1} \] (8)

where \( \gamma_{ij,1} = log \Psi_{ij} \) as defined in section 3, hence the Azzalini model. \( \Delta_{ij} \) equals \( logit(p_{ij,0}) \) and is determined implicitly by \( \beta \) and \( \gamma_{ij,1} \) through equations (3) and (8).

A more general form for \( \gamma_{ij,1} \) may be specified as:

\[ \gamma_{ij,1} = z_{ij,1} \alpha_1 \] (9)

where \( z_{ij,1} \) is a subset of the covariates \( x_{ij} \) and \( \alpha_1 \) defines how \( Y_{ij} \) and \( Y_{ij-1} \) relate to \( z_{ij,1} \).

The more general MTM(\( q \)) dependence model is specified as:

6
\[ g(\mu_{ij}^C) = \Delta_{ij} + \sum_{l=1}^{q} \gamma_{ij,l} y_{ij-l} \quad (10) \]

\[ \gamma_{ij,l} = z_{ij,l} \alpha_l \text{ for } l = 1, \ldots, q \quad (11) \]

In the MTM(q), the \( \beta^M \) describe changes in \( \mu_{ij}^M \) as a function of the covariates ignoring the history of responses and \( \gamma_{ij}^s \) describe how the current response is predicted by previous responses. Therefore, if we change the order of the model, our interpretation of the \( \beta^M \) do not change.

The likelihood function for the MTM(q) model factors into the distribution of the first q responses and the subsequent Bernoulli contributions with parameters \( \mu_{ij}^C \) for \( j > q \). The algorithm for finding the MLEs starts with a model for the distribution of the first q responses using lower order marginalized transition model assumptions. Then the subsequent transition probabilities \( \mu_{ij}^C \) can be determined using the relationship between the \( \beta^M \) and \( (\alpha_1, \ldots, \alpha_q) \). Heagerty provides the specific details of the estimation procedure for the MTM(2) model.

Several simple procedures are available with the MTM(q) models which allow for assessing the dependence model assumptions. Heagerty presents one procedure for testing for an additional lagged response. The test statistic for testing \( \alpha_{q+1} = \gamma_{ij,q+1} = 0 \) is given by

\[ U_{q+1} = \sum_{i=1}^{N} \sum_{j=q+1}^{m} Y_{ij-(q+1)} (\hat{\mu}_{ij}^C - \hat{\mu}_{ij}^C) \quad (12) \]

where \( \hat{\mu}_{ij}^C \) is the fitted conditional mean from the transition model of order q described in section 2. The statistic is an approximation to the likelihood score statistic and is fairly intuitive since it estimates the association of the lagged value \( Y_{ij-(q+1)} \) and the conditional residual from the model of order q.

Heagerty presents two settings where he compares the performance of the MTM(q) models to models fit using GEE with a variety of dependence structures. He generates data from a
group-by-time design and a cross-over design that have order 2 dependence with a fraction of data missing completely at random. In both settings, he shows that a MTM(1) and the AR(1) GEE model are both highly efficient for $\beta^M$ relative to the correct MTM(2) model (efficiency ranging from 0.95 to 0.99); the GEE models with the dependence structure specified as either working independence or exchangeable perform poorly (efficiency ranging from 0.60 to 0.89).

Heagerty proposes a marginalized transition model for binary longitudinal data that permits a likelihood based approach to associating a mean response to a set of covariates with specification of a serial dependence structure via a q-order Markov model. Advantages of these models are: the interpretation of the marginal regression coefficients do not change when the dependence structure is changed; the proposed class of models offer model checking of the specified serial dependence structure; these model should extend to ordinal or nominal categorical response models, possibly by using a global odds ratio to characterize the serial dependence; and they appear to be relatively efficient even to misspecifications of the order of the serial dependence (although this was only investigated for the order 1 and 2 models). Some possible drawbacks and areas for future research include: the proposed models are for equally spaced data only and need to be flexible to handle unequally spaced data; it was assumed that $\mu_{ij}^M = E(Y_{ij}|x_{ij}) = E(Y_{ij}|x_{i1}, ..., x_{in})$ for time-varying covariates, but there may be interest in partly conditional means where $\mu_{ij}^M = E(Y_{ij}|x_{ij}, j \leq t) = E(Y_{ij}|x_{i1}, ..., x_{in})$; this would require a model for the covariate process and incorporating this into the likelihood maximization.
References


