11-1-2011

Higher Order Markov Structure-Based Logistic Model and Likelihood Inference for Ordinal Data

Soma Chowdhury Biswas
University of Chittagong, Chittagong, Bangladesh, soma_chow@hotmail.com

M. Ataharul Islam
University of Dhaka, Bangladesh, mataharul@yahoo.com

Jamal Nazrul Islam
University of Chittagong, Chittagong, Bangladesh, jnislam@yahoo.com

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Higher Order Markov Structure-Based Logistic Model and Likelihood Inference for Ordinal Data

Soma Chowdhury Biswas	M. Ataharul Islam	Jamal Nazrul Islam
University of Chittagong, Chittagong, Bangladesh	University of Dhaka, Dhaka, Bangladesh	University of Chittagong, Chittagong, Bangladesh

Azzalini (1994) proposed a first order Markov chain for binary data. Azzalini's model is extended for ordinal data and introduces a second order model. Further, the test statistics are developed and the power of the test is determined. An application using real data is also presented.

Key words: Markov chain, serial correlation, longitudinal data, ordinal data, covariate dependence, repeated measures.

Introduction

The Markov chain model is one of the most important and effective model classes for the assessment of probability for time dependent processes. A number of models have been proposed for analyzing repeated categorical and ordinal data. Muenz and Rubinstein (1985) employed a logistic regression model to analyze the transitional probabilities from one state to another. Azzalini (1994) introduced a Markov chain model that incorporated serial dependence and facilitated expression of covariate effects on marginal features. Raftery & Tavare (1994) suggested a Markov chain model of order higher than one that involves only one parameter for each extra lag variable: Heagerty and Zeger (2000) and Heagerty (2002) extended that work to a qth-order marginalized transition model (MTM). These models are based on binary data and do not address the more general issue of ordinal data that arises in many biomedical

studies. Islam and Chowdhury (2007) reviewed the first order model of Muenz and Rubinstem (1985) and developed a general procedure based on the Chapman-Kolmogorov equation for transition, reverse transition and repeated transition. Lee and Daniels (2007) extended Heagerty's (2002) MTM to accommodate longitudinal ordinal data. Ching, Fung and Ng (2004) generalized the Raftery and Tavare (1994) model by allowing $Q = \{q_{ij}\}$ to vary with different lags; they also developed an efficient method to estimate the model parameters. Ching, Ng and Fung (2007) extended their 2004 results (Ching, Fung & Ng, 2004) and proposed a higher-order multivariate Markov model for multiple categorical data sequences.

Azzalini's (1994) Markov structure based regression model for ordinal data is extended here, and a second order model is proposed. Likelihood based inferences are possible because the model is fully specified so that resulting estimators are consistent and fully efficient. The proposed methods are applied to real data collected at successive time points from diabetic patients registered at Bangladesh Institute of Research and Rehabilitation in Diabetes, Endocrine and Metabolic disorders (BIRDEM) in Bangladesh.

First Order Covariate Dependent Markov Model
Consider a stationary process $\{Y_{ij}\}$ for individual i (i = 1, 2, …, n) at follow-up j (j = 1, 2, …, n) representing past and present responses

Soma Chowdhury Biswas is a Professor in the Department of Statistics. Email him at: soma_chow@hotmail.com. M. Ataharul Islam is a Professor in the Department of Statistics, Biostatistics and Informatics. Email him at: mataharul@yahoo.com. Jamal Nazrul Islam is Professor Emeritus, RCMPS. Email him at: jnislam@yahoo.com.

where at time t_j the response $Y_j = k \, (k = 0, 1, 2)$. If the transition models for which the conditional distribution of Y_{ij} given the prior observations $Y_{ij-1} \ldots Y_{ij-r}$ is considered the model of order r, then the first order Markov model can be expressed as:

$$\Pr\left(Y_{ij} \middle| Y_{ij-1}\right) = \Pr\left(Y_{ij} \middle| Y_{i,j-r}, \ldots, Y_{i,j-1}\right).$$

For a three state Markov chain the corresponding transition probability matrix is given by

$$P = \begin{pmatrix} p_{00} & p_{01} & p_{02} \\ p_{10} & p_{11} & p_{12} \\ p_{20} & p_{21} & p_{22} \end{pmatrix}$$

and the transition probabilities are:

$$P_0 = P_{it,0} = P\left(Y_{it} \middle| Y_{it=1} = 0\right)$$

$$P_1 = P_{it,1} = P\left(Y_{it} \middle| Y_{it-1} = 1\right)$$

$$P_2 = P_{it,2} = P\left(Y_{it} \middle| Y_{it-1} = 2\right).$$

Consider a vector of covariates for the ith person $X'_{ij} = [1, X_{i1}, \ldots, X_{ip}]$ and the corresponding vector of parameters $\beta'_k = [\beta_{k0}, \beta_{k1}, \ldots, \beta_{kp}]$. The transition probabilities can be expressed in terms of conditional probabilities (Hosmer & Lemeshow, 1989) as follows:

$$p_{ij} = P\left(Y_{i,j} = j \middle| Y_{i,j-1} = j-1, X_{i,j-1}\right)$$

$$= \frac{e^{g_j X}}{\sum_{k=0}^{m-1} e^{g_j X}},$$

where

$$g_k(X) = \text{In}\left[\frac{p(Y = k / X)}{p(Y = 0 / X)}\right]$$

$$= \beta_{k0} + \beta_{k1} X_1 + \ldots + \beta_{kp} X_p \tag{1}$$

Azzalini (1994) searched for a parameterization such that $\theta = E(Y_t)$, which is free from a parameter and that regulates serial dependence. The odds ratio Ψ is a quantity that measures the dependence between successive observations. A technical reason in favor of this choice was provided by Fitzmaurice and Laird (1993) who stated that, when the association between observations is modeled using an odds ratio, the estimates of the mean are relatively insensitive to changes of the association parameter. Moreover, the range of feasible values for Ψ is independent of the θ value.

The above three stated Markov models for non-stationary cases are parameterized as

$$\theta_t = P_1 \theta_{t-1} + P_0(1 - \theta_{t-1}) \tag{2}$$

and

$$\theta_t = P_2 \theta_{t-1} + P_0(1 - \theta_{t-1}) \tag{3}$$

where, P_0, P_1 and P_2 will vary with t and (t = 2, 3, 4, 5, ..., T). For t = 1, $E(Y_1) = \Pr(Y_1 = 1) = \theta_1$. and odds ratios:

$$\psi_1 = \frac{P_1|(1 - P_1)}{P_0|(1 - P_0)} \tag{4}$$

$$\psi_2 = \frac{P_2|(1 - P_2)}{P_0|(1 - P_0)} \tag{5}$$

For a given value of β the sequence of θ can be determined by (1) and solving (1) and (3); (2) and (4) and after algebraic manipulation, results in,

$$P_{\cdot} = \begin{cases} \theta_t, & \text{for } \psi_i = 1 \\ \dfrac{\delta - 1 + (\psi_i - 1)(\theta_t - \theta_{t-1})}{2(\psi_i - 1)(1 - \theta_t - 1)} \\ + j\dfrac{1 - \delta + (\psi_i - 1)(\theta_t + \theta_{t-1} - 2\theta_t \theta_{t-1})}{2(\psi_i - 1)\theta_{t-1}(1 - \theta_{t-1})} \end{cases} ; \psi \neq 1,$$

(6)

where

$$\delta^2 = 1 + (\Psi_i - 1)\left\{ \begin{array}{c} (\theta_t - \theta_{t-1})^2 \Psi_i \\ -(\theta_t - \theta_{t-1})^2 + 2(\theta_t + \theta_{t-1}) \end{array} \right\}$$

and

$$P_{y_{t-1}} = \left\{ \begin{array}{c} \dfrac{\delta - 1 + (\psi - 1)(\theta_t - \theta_{t-1})}{2(\psi - 1)(1 - \theta_{t-1})} \\ + y_{t-1}\dfrac{1 - \delta + (\psi - 1)(\theta_t + \theta_{t-1} - 2\theta_t \theta_{t-1})}{2(\psi - 1)\theta_{t-1}(1 - \theta_{t-1})} \end{array} \right\}$$

for $t > 1$ and for $t = 1$, $p_{y_{it}} = \theta_1$.

Second Order Covariate Dependent Markov Model

A second order Markov model assumes that the current response variable is dependent on the history not only through the immediate previous response but also on the previous two responses, that is,

$$\Pr\left(Y_{it} | Y_{it-2}, Y_{ij=1}\right) = \Pr\left(Y_{it} | Y_{it-1}, Y_{it-2}, ..., Y_{it-n}\right).$$

The transition probabilities for the three state second order Markov Chain can be written as:

$$P_0 = p_{it,0} = \Pr\left(y_{it} | y_{it-1} = 0, y_{t-2} = 0\right)$$

$$P_1 = p_{it,1} = \Pr\left(y_{it} | y_{it-1} = 0, y_{t-2} = 1\right)$$

$$P_2 = p_{it,2} = \Pr\left(y_{it} | y_{it-1} = 0, y_{t-2} = 2\right).$$

Let the parameterization of mean and odds ratio for second order can be extended as

$$\theta_{it} = p_{it,1}\,\theta_{it-2} + p_{it,0}(1 - \theta_{it-2})$$
$$= \left(\begin{array}{c} P_{001} + P_{101} + P_{201} + P_{011} + P_{111} \\ + P_{211} + P_{021} + P_{121} + P_{221} \end{array} \right)\theta_{t-2}$$
$$+ \left(\begin{array}{c} P_{000} + P_{100} + P_{200} + P_{010} + P_{110} \\ + P_{210} + P_{020} + P_{120} + P_{220} \end{array} \right)(1 - \theta_{t-2})$$

(7)

$$\theta_{it} = p_{it,2}\,\theta_{it-2} + p_{it,0}(1 - \theta_{it-2})$$
$$= \left(\begin{array}{c} P_{002} + P_{102} + P_{202} + P_{012} + P_{112} \\ + P_{212} + P_{022} + P_{122} + P_{222} \end{array} \right)\theta_{t-2}$$
$$+ \left(\begin{array}{c} P_{000} + P_{100} + P_{200} + P_{010} + P_{110} \\ + P_{210} + P_{020} + P_{120} + P_{220} \end{array} \right)(1 - \theta_{t-2})$$

(8)

$$\psi_1 = \frac{p_{it,1}\left|(1 - P_{it,1})\right|}{p_{it,0}\left|(1 - P_{it,0})\right|} \quad (9)$$

$$\psi_2 = \frac{p_{it,2}\,|(1 - P_{it,2})}{p_{it,0}\,|(1 - P_{it,0})} \quad (10)$$

and

$$P_j = P_{y_{t-2}} = \frac{\delta - 1 + (\Psi_i - 1)(\theta_i - \theta_{t-2})}{2(\Psi_i - 1)(1 - \theta_{t-2})}$$
$$+ j\frac{(\delta - 1) + (\Psi_i - 1)(\theta_i + \theta_{t-2} - 2\theta_t \theta_{t-2})}{2(\Psi_i - 1)\theta_{t-2}(1 - \theta_{t-2})}$$

for $\Psi_i \neq 1$,

where

$$\delta^2 = 1 + (\Psi_i - 1)\left\{ (\theta_t - \theta_{t-2})^2 \Psi_i - (\theta_t - \theta_{t-2})^2 + 2(\theta_t + \theta_{t-2}) \right\}$$

and $\log \Psi_i = \lambda_1$ and $\log \Psi_2 = \lambda_2$.

These relationships generate a process having the desired properties. Upon taking $P_r(y_t = 1) = \theta_1$ and generating $y_2, ..., y_t$ via a non-homogeneous Markov chain with transition probabilities P_j, a sequence is obtained such that

$E(y_t) = \theta_t$ for $t = 1, 2, ..., t$ and the odds ratios for (y_{t-2}, y_t) are equal to ψ.

Estimation

The conditional likelihood function for a sample of n independent observations is:

$$L(\beta) = \prod_{i=1}^{n} \left[[P(0|x)]^{y_{0i}} [P(1|x)]^{y_{1i}} [P(2|x)]^{y_{2i}} \right]$$

(11)

The log-likelihood function can be written as

$$l_t(\beta) = \sum_{i=1}^{n} \begin{bmatrix} y_{01} \log P(0|x) \\ +y_{1i} \log P(1|x) \\ +y_{2i} \log P(2|x) \end{bmatrix}$$

$$= \sum_{i=1}^{n} \begin{bmatrix} y_{0i} \log \left(\dfrac{P(y=0)}{(1-P(y=0))} \right) \\ +y_{1i} \log \dfrac{P(y=1)}{1-P(y=1)} \\ +y_{2i} \log \dfrac{P(y=2)}{1-P(y=2)} \end{bmatrix}$$

(12)

and

$$= \sum_{i=1}^{n} \left[y_{0i} g_0(x) + y_{1i} g_1(x) + y_{2i} g_0(x) \right]$$

$$= \sum_{i=1}^{n} \begin{bmatrix} y_{1i} g_1(x) + y_{2i} g_2(x) \\ -\log(1+y_{1i} g_1(x) + y_{2i} g_2(x)) \end{bmatrix}$$

(13)

where

$$(y_{0i} + y_{1i} + y_{2i} = 1)]$$

and

$$\log \psi_1 = \lambda_1 \Rightarrow \psi_1 = e^{\lambda_1},$$

$$\log \psi_1 = \lambda_2 \Rightarrow \psi_2 = e^{\lambda_2}.$$

In the case of repeated measures, because dependence between successive observations on the same individual must be taken into account, it is plausible that adjacent data are more strongly correlated than data that are separated by time and that different individuals behave independently on the log likelihood, this is given by:

$$l(\beta, \lambda) = \sum_{T=1}^{4} l_t(\beta, \lambda)$$

where score vectors are:

$$\frac{\partial l}{\partial \beta} = \sum_{t=1}^{T} \sum_{i=1}^{n} \frac{\partial l_{it}}{\partial \beta}$$

and

$$\frac{\partial l}{\partial \lambda} = \sum_{t=1}^{T} \sum_{i=1}^{n} \frac{\partial l_{it}}{\partial \lambda}$$

and the variance of the estimate is approximated by

$$V\left(\hat{\beta}, \hat{\lambda}\right) = \left[\sum_{i=1}^{n} \begin{pmatrix} \dfrac{\partial l_t}{\partial \beta} \\ \dfrac{\partial l_t}{\partial \lambda} \end{pmatrix} \begin{pmatrix} \dfrac{\partial l_t}{\partial \beta} \\ \dfrac{\partial l_t}{\partial \lambda} \end{pmatrix}^{T} \right]^{-1}_{\beta=\hat{\beta}, \lambda=\hat{\lambda}}$$

The quantity inside the square brackets approximates the Fisher Information for large n. Similarly, the model can be generalized to third, fourth and up to n^{th} order.

Methodology

Test Procedure

To test the null hypothesis that all slope coefficients are simultaneously equal to zero is the usual likelihood ratio statistic used and it follows the χ^2 distribution with degrees of freedom (df) equal to the number of explanatory variable(s). For an rth order Markov model the df for Chi square is $m^r(m-1)p$, where m is the number of states. Therefore, to test the null hypothesis $H_0 : \beta = 0$, the usual likelihood ratio test can be employed.

$$-2[\ln L(\beta_0) - \ln L(\beta)] \approx X^2_{m^r(m-1)p}$$

531

where the vectors $m^r(m-1)$ set a parameters for the rth order Markov model.

For a first order Markov model with dichotomous transition outcomes and an independent variable, the likelihood ratio Chi square is

$$-2\left[\ln L(\beta_0) - \ln L(\beta)\right] \approx \chi^2_{3^1(3-1)p} \approx \chi^2_{6p}$$

where

$$\ln L(\beta) = \log l_t(\beta, \lambda)$$
$$= \sum_{t-2}^{T}\left\{y_t\, logit\left(P_{y_{t=t2}}\right) + \log\left(1 - P_{y_{t-2}}\right)\right\}.$$

Similarly, for a second order model with binary outcomes and p independent variables, the null hypothesis of the null parameter vector can be tested by using the test statistic:

$$-2\left[\ln L(\beta_0) - \ln L(\beta)\right] \approx \chi^2_{3^2(3-1)p} = \chi^2_{18p}.$$

The Wald test statistic for the null hypothesis H_o: $\beta_j = 0$ can be written as the multivariate analogue of this test, which follows a Chi square distribution and is given by

$$W_i = \left[(\hat{\beta} - \beta_0)\right]'\left[I(\hat{\beta}, o)\right]^{-1}\left[(\hat{\beta} - \beta_0)\right].$$

Data and Variables

The proposed model is illustrated using Diabetes mellitus data. This data was collected by the Bangladesh Institute of Research and Rehabilitation in Diabetes, Endocrine and Metabolic disorders (BIRDEM). After registration, a patient visits the BIRDEM for regular check-ups and treatment. During registration each patient answers a detailed questionnaire and a comprehensive record sheet is maintained for each patient until death of the patient or loss of follow up over time. The patients experience impaired glucose tolerance (IGT) levels at that time of registration, so the number of follow-ups for each patient is not equal. For convenience of analysis 999 patients were randomly selected for this study that had four consecutive follow-up visits to BIRDEM. Consider a few selected variables such as, age (AGE), sex (SEX: coded as 0 = female, 1= male) and blood glucose level in each visit. Some clinical variables were not used in the study due to a high percentage of missing values over the four consecutive visits. In the record sheet, the age of first registration of the patient is logged. The age at different consecutive follow-ups visits from date of registration of the patients was calculated using SPSS, thus making age a continuous variable that was used in the study directly to observe the time effect.

The blood glucose level of each patient (two hours after 75 grams of glucose load) was assumed to be indicative of the patient's diabetic health status and was therefore considered the dependent variable. Specifically, a person having blood glucose concentration level in venous plasma after 2 hours of 75 grams glucose load greater than or equal to 11.1 mmol/liter was considered a confirmed diabetic and coded as 2; a person with a blood glucose level between 4.4 and 11.1 mmol/l was considered as border line diabetic and coded as 1; and a person with a blood glucose concentration level less than 4.4 mmol/liter was considered a controlled diabetic and coded as 0. The response variable Y_{it} of interest can be defined as:

$$Y_{it} = \begin{cases} 0, \text{Controlled diabetic} \\ \quad if \text{ blood glucose level} < 4.4 \\ 1, \text{Borderline diabetic} \\ \quad if \ 4.4 \leq \text{blood glucose level} < 11.1 \\ 2, \text{Confirmed diabetic} \\ \quad if \text{ blood glucose level} \geq 11.1 \end{cases}$$

where $t = 1, 2, 3, 4$. The response Y_{it} to be generated by a ordinal Markov chain with values 0, 1 and 2 and with transition probabilities $p_{ij} = \Pr(Y_{it} = i | Y_{it-1} = j)$ for j = 0, 1, 2 for the first order, and $p_{ijk} = \Pr(y_{it} = i / y_{it-1} = j, y_{t-2} = k)$ for i, j, k = 0, 1, 2 for the second order. This study concentrates on modeling the mean value θ_{it} via a covariate, which can be obtained by using (1).

Results

First Order Markov Model

The dependence between successive observations y_{it} $t = 1, 2, 3, 4$ is measured by the odds ratio Ψ_{it} which is defined in (4) and (5). A sequence of mean values of the process, such that, $\theta_{it} = E(y_{it})$ for $t = 1, 2, 3, 4$ and odds ratio Ψ_{it} for $(y_{i(t-1)}y_{it})$ could be obtained by taking $Pr(y_{i1} = 1) = \theta_{i1}$ and $Pr(y_{i2} = 2) = \theta_{i2}$, generating y_{i2}, y_{i3}, y_{i4} via a homogeneous Markov Chain with transition probabilities p_{t0}, p_{t1} and p_{t2} obtained from (6). Because an objective is to determine the effects of covariates on the risk of confirmed and borderline diabetics, the marginal probabilities of confirmed and borderline diabetic for each given value of the covariates is the quantity of interest. The log-likelihood function for β and $\lambda = \log \Psi t$ for $T = 4$ time points can be described by (12) and also for repeated data. The model in (1) for the marginal probability of the event is fitted to the data. The parameters of the model were estimated using the maximum likelihood method of estimation and the Newton-Raphson iteration method. All the calculations were performed by programming in R. Table 1 summarizes the results of fitting of the 1st order model.

As shown in Table 1, the likelihood ratio test value, 6759.67, is highly significant, thus, it can be used to identify the effect of the covariates on the disease status of the patient. For model $0 \rightarrow 1$, age is a significant factor and has a positive association with the transition of the disease from controlled (0) to borderline (1); sex is not significant a factor although it has a positive association with the transition of disease, that is, female patients are less likely to transition from controlled to borderline diabetes compared to male patients. For model $0 \rightarrow 2$, both covariates have a positive association with the transition of the disease from controlled to confirmed diabetes. The risk of transition from controlled (0) to confirmed (2) diabetes increases with the increase of age. Female patients are more likely to transition from borderline diabetes to confirmed diabetes

compared to male patients, although the difference is not significant.

Figure 1 illustrates the power comparison of the Wald test for testing the hypothesis $H_0 : \beta_i = 0$ versus alternative hypothesis $H_1 : \beta_i \neq 0$ for a first order model. Both graphs show that the power obtained from the $0 \rightarrow 2$ model (controlled to confirmed) is higher than that of the $0 \rightarrow 1$ (controlled to borderline) model for the parameters sex and age.

Second Order Markov Model

The dependence between successive observations Y_{it}, $t = 1, 2, 3, 4$ is measured by the odds ratio Ψ_{it} which is defined in (9) and (10). A sequence of mean values of the process, such that $\theta_{it} = E(Yit)$ for $t = 1, 2, 3, 4$, and odds ratio Ψ_{it} for $(Y_{i(t-2)}Y_{it})$ could be obtained by taking $Pr(Y_{it=1} = 1) = \theta_{1t}$ and $Pr(Y_{it-2} = 2) = \theta_{2t}$. This will generate Y_{i3}, Y_{i4} via a non-homogeneous Markov Chain with transition probability $P_{t2} \ and \ P_{t3}$ obtained in (11). The log-likelihood function for β and $\lambda = \log \Psi t$ for $T = 4$ time points is described by (12). Model (13) for the marginal probability of the event was fitted to the data and the parameters of the Markov based second order model were estimated using the maximum likelihood and Newton Raphson Iteration methods; all calculations were performed with R.

Table 2 summarizes the results of the fitted second order model and shows that the likelihood ratio for the overall model is 6780349.580, which is significant and follows a Chi-square distribution with 5 df; thus, the null hypothesis may be rejected and significance for at least one of the covariates may be concluded. To reveal the significance of individual parameters, the Wald test was performed. For model $0 \rightarrow 1$, age and sex show a positive association with the response variable. The risk of transition from controlled (0) to borderline (1) diabetes increases as age increases; both variables have a significant effect on the transition from controlled to borderline diabetes.

Table 1: Estimates and Associated Wald Test from First Order Markov Model

Variable	Estimated Coefficient	Standard Error	Wald χ^2	P-value	Odds Ratio
$0 \rightarrow 1$					
Constant	0.3579	0.1906	3.5247	NA	--
Sex	0.1944	0.01848	110.61	0.0000	1.21
Age	0.1733	0.2079	0.69441	0.4046	1.19
λ	-1.1965	0.00329	13204.62	0.0000	0.31
$0 \rightarrow 2$					
Constant	0.3662	0.22831	2.5731	NA	--
Sex	0.1984	0.00651	927.688	0.0000	1.22
Age	0.1013	0.12993	0.60825	0.4354	1.11
λ	-2.1997	0.00145	22994.55	0.0000	0.11

Likelihood Ratio = 6759.67; p-value = 0.000

Figure 1: Power Curves for Covariates Sex and Age for First Order Markov Model
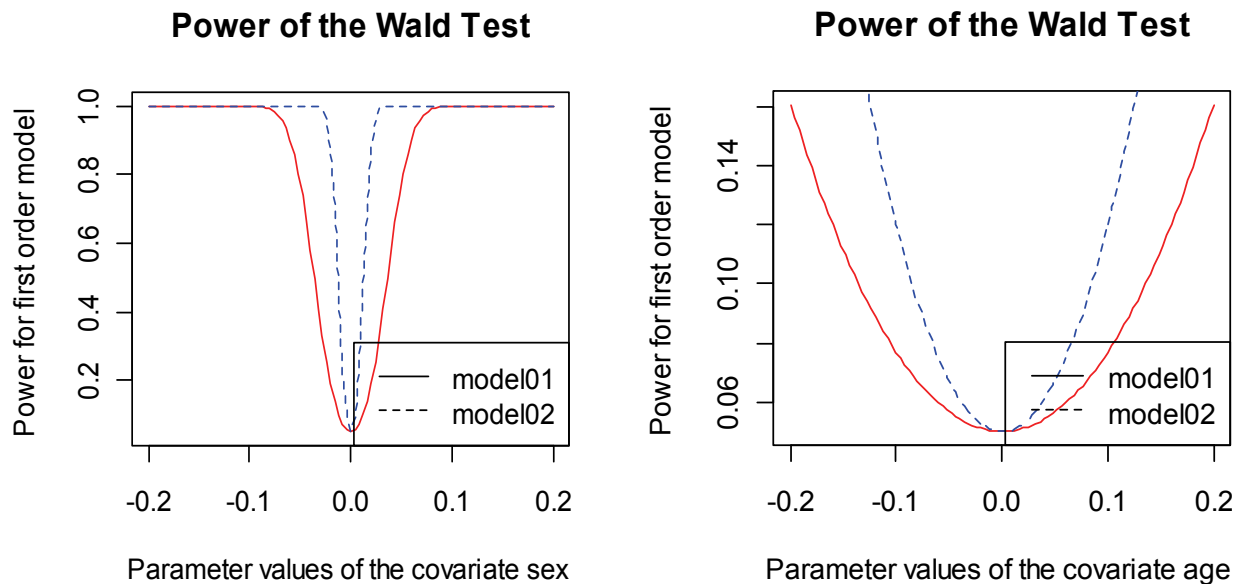
Table 2: Estimates and Associated Test from Second Order Markov Model

Variable	Estimated Coefficient	Standard Error	Wald χ^2	P-value	Odds Ratio
$0 \rightarrow 1$					
Constant	0.3962	0.0498	63.31	NA	--
Sex	0.4001	0.0187	456.71	0.0000	1.49
Age	0.1864	0.0479	151194	0.0001	1.20
λ	-0.9001	0.00007	1364683.0	0.0000	0.40
$0 \rightarrow 2$					
Constant	0.7127	0.0596	142.56	NA	--
Sex	0.3695	0.0044	702.64	0.0000	1.44
Age	0.0718	0.0394	3.3334	0.0067	1.07
λ	-1.2001	0.0001	121102.9	0.0000	0.30

Overall Chi square = 6780349.580; p-value = 0.0000

Figure 2: Power Curves for Covariates Sex and Age for Second Order Markov Model
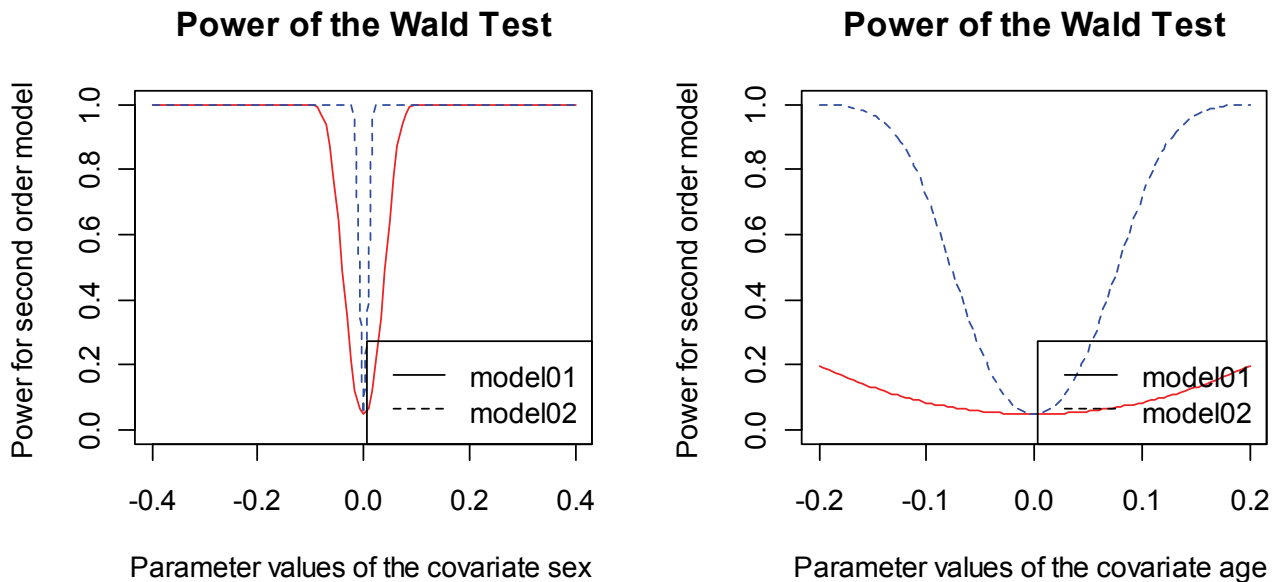
For model $0 \rightarrow 2$, similar to $0 \rightarrow 1$, age and sex show a positive association with the response variable and the association is significant with the subject's transition both from controlled (0) to boarderline (1) and controlled (0) to confirmed (2) diabetes. The risk of transition in both cases increases with a unit increase in age level. Male patients are more likely to transition to borderline (1) and to confirmed (2) diabetes compared to female patients. Finally, it can be concluded that an increase in age increases the risk of transition of the disease to a higher stage. The value of likelihood ratio also has a noticeable increase when considering a higher order Markov model.

Figure 2 shows the power curves of the Wald test for testing the hypothesis $H_0 : \beta_i = 0$ against the alternative hypothesis $H_1 : \beta_i \neq 0$ for a second order model. The charts illustrate that the power obtained from the $0 \rightarrow 2$ model is higher than that of the $0 \rightarrow 1$ model.

Conclusion

This study extended Azzalini's (1994) model for ordinal data up to second order, which can then be generalized to any order in the same setting as Islam and Chowdhury (2007). The proposed model was applied to repeated measures of diabetes mellitus testing and it was observed that the variables age and sex show significant contributions to the diabetes status of a patient. Comparison of the estimates of first and second order and power curve are displayed for the Wald Chi square test, which shows a significant improvement in power for the $0 \rightarrow 2$ transition model. Based on results of this study it is reasonable to conclude that, for analyzing the repeated measures data of diabetes mellitus, a higher order Markov model approach can be conveniently employed for any number of states and for any order with any number of covariates; this may prove valuable for health policy makers. Further research could be conducted using a continuous time Markov model for estimation and testing in other settings.

References

Azzalini, A. (1994). Logistic regression for auto correlated data with application to repeated measures. *Biometrika*, *81*, 767-775.

Ching, W. K., Ng, M. K., & Fung, E. S. (2008). Higher order multivariate Markov chains and their applications. *Linear Algebra and its Applications*, *428*(*2-3*), 492-507.

Ching, W. K., Fung, E. S., & Ng, K. (2004). Higher order Markov chain models for categorical data sequences. *International Journal of Naval Research Logistics*, *51*, 557-574.

Fitzmanurice, G. M., & Laird, N. M. (1993). A likelihood based method for analyzing binary responses. *Biometrika*, *80*, 141-151.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York, NY: John Wiley and Sons.

Heagerty, P. J., & Zeger, S. L. (2000). Marginalized multilevel models and likelihood inference (with discussion). *Statistical Science*, *15*, 1-26.

Heagerty, P. J. (2002). Marginalized transition models and likelihood inference for longitudinal categorical data, *Biometrics 58, 342-35*.

Islam, M. A., & Chowdhury, R. I. (2006). A higher order Markov model of analyzing covariate dependence. *Applied Mathematical Modeling*, *30*, 477-488.

Islam, M. A., & Chowdhury, R. I. (2008). First and higher order transition models with covariate dependence. In *Progress in Applied Mathematical Modeling*, 153-196. New York, NY: Nova Science Publishers, Inc.

Islam, M. A., Chowdhury, R. I., & Huda, S. (2009). *Markov models with covariate dependence for repeated measures*. New York, NY: Nova Science Publishers, Inc.

Lee, K., & Daniels, M. J. (2007). A class of Markov models for longitudinal ordinal data. *Biometrics, 64, 4,1060-1067*.

Muenz, K.R., & Rubinstein, L.V. (1985). Markov chain for covariance dependence of binary sequences. *Biometrics*, *41*, 91-101.

Phillips, P. C. B. (1986). The exact distribution of the Wald statistic. *Econometrica*, *54*(*4*), 881-895.

Raftery, A., & Tavare, S. (1994). Estimating and modeling repeated patterns in higher order Markov chains with the mixture transition distribution model. *Applied Statistics*, *43*(*1*), 179-199.

Rahman, M. S., & Islam, M. A. (2007). Markov structure based logistic regression for repeated measures: An application to diabetes mellitus data. *Statistical Methodology, 4,* 448-460.

Appendix

The derivatives $\dfrac{\partial l}{\partial \beta}$ and $\dfrac{\partial l}{\partial \lambda}$ were computed using a chain rule, giving elements of score vectors of a first order model. The parameters β and λ were estimated by maximum likelihood method and using chain rule of differentiation.

$$\frac{\partial l_t}{\partial \beta} = \frac{\partial l_t}{\partial p_{y_{t-1}}} \left(\frac{\partial p_{yt-1}}{\partial \theta_t} \cdot \frac{\partial \theta_t}{\partial \beta} + \frac{\partial p_{yt-1}}{\partial \theta_{t-1}} \cdot \frac{\partial \theta_{t-1}}{\partial \beta} \right)$$
$$(t = 1, 2, \ldots T)$$

$$\frac{\partial l_t}{\partial \lambda} = \frac{\partial l_t}{\partial p_{y_{t-1}}} \cdot \frac{\partial p_{y_{t-1}}}{\partial \psi} \cdot \frac{\partial \psi}{\partial \lambda}$$
$$(t = 2, 3, \ldots T)$$

where,

$$\log \psi_1 = \lambda_1 \Rightarrow \psi_1 = e^{\lambda_1};$$

$$\log \psi_2 = \lambda_2 \Rightarrow \psi_2 = e^{\lambda_2}$$

$$\frac{\partial Py_{t-1}}{\partial \theta_t} = \frac{1}{A}(-(2y_{t-1} - 1)\frac{\partial \delta}{\partial \theta_t} + \psi - 1)$$

$$\frac{\partial \theta_t}{\partial \beta} = \theta_t(1 - \theta_t)x_t ; \text{where } \theta_t = \frac{e^{x_{t\beta}}}{1 + e^{x_{t\beta}}}$$

$$\frac{\partial Py_{t-1}}{\partial \theta_{t-1}} = \frac{1}{A^2} \left\{ \begin{array}{l} (-(2y_{t-1}-1)(\psi - 1 - \frac{\partial \delta}{\partial \theta_{t-1}}) \\ A - 2(\psi - 1)(2y_{t-1} - 1)B \end{array} \right\}$$

where

$$B = (2y_{t-1})\{(1 - \delta + (\psi - 1)\theta_t\}(\psi - 1)\theta_t - 1$$

and

$$A = 2(\psi - 1)(1 - y_{t-1}) + (2y_{t-1} - 1)\theta_t$$

$$\frac{\partial \theta_{t-1}}{\partial \beta} = \theta_{t-1}(1 - \theta_{t-1})x_{t-1}$$

$$\frac{\partial p_{yt-1}}{\partial \theta_{t-1}} = \frac{1}{A^2} \left[\begin{array}{l} ((2y_{t-1}-1)(-\frac{\partial \delta}{\partial \psi} + \theta_{t-1}) + \theta_t\} \\ A - 2B\{1 - y_{t-1} + (2y_{t-1})\theta_{t-1}\} \end{array} \right]$$

$$\frac{\partial s}{\partial \theta} = \frac{1}{\delta} \left[\begin{array}{l} (\psi - 1)\{\psi(\theta_t - \theta_{t-1}) \\ -(\theta_t - \theta_{t-1}) + 1\} \end{array} \right]$$

$$\frac{\partial s}{\partial \theta_{t-1}} = \frac{1}{\delta} \left[\begin{array}{l} (\psi - 1)\{-\psi(\theta_t - \theta_{t-1}) \\ -(\theta_t - \theta_{t-1}) + 1\} \end{array} \right]$$

$$\frac{\partial s}{\partial \psi} = \frac{1}{2\delta} \left[\begin{array}{l} (\theta_t - \theta_{t-1})^2(2\psi - 1) \\ -(\theta_t + \theta_{t-1})^2 + 2(\theta_t + \theta_{t-1})\} \end{array} \right]$$

$$\frac{\partial \psi}{\partial \lambda} = \psi$$

and

$$\frac{\partial p_{y_{t-1}}}{\partial \psi} = \frac{1}{A2} \left[\begin{array}{l} \left\{(2y_{t-1} - 1)\left(-\frac{\partial \delta}{\partial \psi} + \theta_{t-1}\right) + \theta_t\right\} \\ A - 2B\left(1 - y_{t-1} + (2y_{t-1} - 1)\theta_{t-1}\right) \end{array} \right]$$

Second order model elements of score vectors were computed as:

$$\frac{\partial p_{\gamma_{t-2}}}{\partial \theta_t} = \frac{1}{A}\left(-2(2y_{t-2} - 1)\frac{\partial \delta}{\partial \theta_t} + \psi - 1\right)$$

where

$$A = 2(\psi-1) \{(1-y) + (2y-1) \theta_{t-2}\}$$

$$\frac{\partial \theta_t}{\partial \beta} = \theta_t(1 - \theta_t)\chi_t$$

where

$$\theta_t = \frac{e^x\beta}{1 + e^x\beta}$$

$$\frac{\partial p_{y-2}}{\partial \theta_{t-2}} = \frac{1}{A^2} \left\{ \begin{array}{l} (2y_{t-2} - 1)\left(\psi - 1 - \frac{\partial \delta}{\partial \theta_{t-2}}\right) \\ A - 2(\psi - 1)(2y_{t-2} - 1)B \end{array} \right\}$$

where

$$B = (2y - 1)(1 - \delta(\Psi - 1) \times \theta_{t-2} + (\Psi - 1)\theta_t \}$$

$$\frac{\partial \theta_{t-2}}{\partial \beta} = \theta_{t-2}(1 - \theta_{t-2})\chi_{t-2}$$

$$\frac{\partial p_{\gamma t-2}}{\partial \psi} = \frac{1}{A^2} \left[\begin{array}{l} (2y_{t-1} - 1)\left(-\frac{\partial \delta}{\partial \psi} + \theta_{t-2}\right) + \theta_t \quad) \\ A - 2B(1 - y_{t-2}) + (2y_{t-2} - 1)\theta_{t-2} \end{array} \right]$$

$$\frac{\partial \delta}{\partial \theta_t} = \frac{1}{\delta}\left[(\psi - 1)\left\{ \begin{array}{l} \psi(\theta_t - \theta_{t-2}) \\ -(\theta_t + \theta_{t-2}) + 1 \end{array} \right\} \right]$$

$$\frac{\partial \delta}{\partial \theta_{t-2}} = \frac{1}{\delta}\left[(\psi - 1)\left\{ \begin{array}{l} -\psi(\theta_t - \theta_{t-2}) \\ -(\theta_t + \theta_{t-2}) + 1 \end{array} \right\} \right]$$

$$\frac{\partial \delta}{\partial \psi} = \frac{1}{2\delta}\left[\begin{array}{l} (\theta_t - \theta_{t-2})^2 (2\psi - 1) \\ -(\theta_t + \theta_{t-2})^2 + 2(\theta_t + \theta_{t-2}) \} \end{array} \right]$$