Chapter 12: Marginal approaches to categorical data

Timothy Hanson

Department of Statistics, University of South Carolina

Stat 770: Categorical Data Analysis

Example of repeated measures:

- Data are comprised of several repeated measurements on the same individual over time, e.g. Y_{ij} might indicate an acne outbreak for patient i in month j.

- Data are recorded in clusters, e.g. Y_{ij} might indicate the presence of tooth decay for tooth j in patient i.

- Data are from naturally associated groups, e.g. Y_{ij} might denote a successful treatment of patient j at clinic i.

In all of these examples, the repeated measurements are (typically positively) correlated within an individual or group.

12.2 Marginal GLM modeling of multiple categorical responses

Let T_i binary responses $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iT_i})$ come from the ith cluster (individual, litter, clinic, etc.) Let $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{iT_i})$ where $\mu_{ij} = E(Y_{ij})$. Let $\mathbf{x}_{ij}$ be a $p \times 1$ vector of explanatory variables.

We assume the vectors $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ are independent, but that elements of $\mathbf{Y}_i$ are correlated. Common choices are

$$\mathbf{R}(\alpha) = \text{corr}(\mathbf{Y}_i) = \begin{bmatrix} 1 & \alpha & \alpha & \cdots & \alpha \\ \alpha & 1 & \alpha & \cdots & \alpha \\ \alpha & \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \alpha & \cdots & 1 \end{bmatrix}_{T_i \times T_i} \quad \text{exchangeable,}$$

$$\text{and } \mathbf{R}(\alpha) = \text{corr}(\mathbf{Y}_i) = \begin{bmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{T_i-1} \\ \alpha & 1 & \alpha & \cdots & \alpha^{T_i-2} \\ \alpha^2 & \alpha & 1 & \cdots & \alpha^{T_i-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{T_i-1} & \alpha^{T_i-2} & \alpha^{T_i-3} & \cdots & 1 \end{bmatrix}_{T_i \times T_i} \quad \text{AR(1).}$$

$$\mathbf{R}(\boldsymbol{\alpha}) = \text{corr}(\mathbf{Y}_i) = \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1T} \\ \alpha_{12} & 1 & \alpha_{23} & \cdots & \alpha_{2T} \\ \alpha_{13} & \alpha_{23} & 1 & \cdots & \alpha_{3T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{1T} & \alpha_{2T} & \alpha_{3T} & \cdots & 1 \end{bmatrix}_{T \times T} \quad \text{unstructured,}$$

$$\text{and } \mathbf{R} = \text{corr}(\mathbf{Y}_i) = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{T_i \times T_i} \quad \text{independence.}$$

Comments

You can also specify a fixed, known **R** as well as MDEP(m) which yields $\mathbf{R}(\boldsymbol{\alpha})$ as

$$\text{corr}(Y_{ij}, Y_{i,j+t}) = \left\{ \begin{array}{ll} 1 & t = 0 \\ \alpha_t & t = 1, \ldots, m \\ 0 & t > m \end{array} \right\}.$$

- Unstructured most general; often a default choice. However, need balance: $T_i = T$ for all i.
- Exchangeable useful when time is not important and correlations thought to be approximately equal, e.g. repeated measurements on individual in crossover study, measurements across several individuals from clinic i.
- AR(1) useful when serial correlation plausible, e.g. repeated measurements across equally spaced time points on individual.

More comments

- These correlation matrices are used in a GEE algorithm (sketched below) in PROC GENMOD.
- Repeated measures are accounted for via REPEATED statement.
- The order of $(Y_{i1}, \ldots, Y_{iT})$ makes a difference with some $\mathbf{R}(\boldsymbol{\alpha})$. If ordering is different to that defined in the DATA step, one can use the WITHIN subcommand in the REPEATED statement to tell SAS what the ordering is. Also used when missing some measurements in $(Y_{i1}, \ldots, Y_{iT})$.
- CORRW in the REPEATED statement gives the final working correlation matrix estimate.
- Elements of $\boldsymbol{\beta}$ are interpreted as usual, but *averaged over clusters*. This is a *marginal* interpretation.

GEE approach

Let $\mu_{ij} = g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta})$ be the *marginal* mean. We assume Y_{ij} is from an exponential family

$$Y_{ij} \sim f(y_{ij}; \theta_{ij}, \phi) = \exp\{[y_{ij}\theta_{ij} - b(\theta_{ij})]/\phi + c(y_{ij}, \phi)\},$$

where the dispersion ϕ is known. The GEE approach requires some notation:

- $\mu_{ij} = b'(\theta_{ij})$ and $v(\mu_{ij}) = \text{var}(Y_{ij}) = b''(\theta_{ij})\phi$.
- $\mathbf{R}(\boldsymbol{\alpha})$ is "working correlation matrix," reflecting our best guess at the true correlation structure among the elements of $\mathbf{Y}_i$. See the previous slide. Choice of $\mathbf{R}(\boldsymbol{\alpha})$ can be made based on QIC (Pan, 2001).
- $\mathbf{B}_i = \text{diag}(b''(\theta_{i1}), \ldots, b''(\theta_{iT}))$ is a diagonal matrix with $\text{var}(Y_{ij})/\phi$ along the diagonal.
- $\mathbf{V}_i = \mathbf{B}_i^{1/2}\mathbf{R}(\boldsymbol{\alpha})\mathbf{B}_i^{1/2}\phi$ is the working covariance matrix.

GEE approach

Let $\mathbf{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} = \mathbf{B}_i \boldsymbol{\Delta}_i \mathbf{X}_i$ be the $T_i \times p$ matrix of first partial derivatives where $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\beta}) = (g^{-1}(\mathbf{x}'_{i1}\boldsymbol{\beta}), \ldots, g^{-1}(\mathbf{x}'_{iT_i}\boldsymbol{\beta}))$,

$\boldsymbol{\Delta}_i = \text{diag}(\frac{\partial \theta_{i1}}{\partial \eta_{iT_i}}, \ldots, \frac{\partial \theta_{i1}}{\partial \eta_{iT_i}})$, $\eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta}$, and $\mathbf{X}_i = \begin{bmatrix} \mathbf{x}'_{i1} \\ \vdots \\ \mathbf{x}'_{iT_i} \end{bmatrix}$.

The generalized estimating equations (GEE) are

$$\mathbf{u}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{D}'_i \mathbf{V}_i^{-1} [\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})] = \mathbf{0}.$$

These correspond to likelihood (score) equations, but are *not* derived from a proper likelihood. However, the $\hat{\boldsymbol{\beta}}$ that solves them *is consistent*, even when the correlation assumption is *wrong*. Roughly speaking, this is because consistency is a first moment (mean) property.

Basic idea

For correlated *normal* data where $E(\mathbf{Y}_i) = \mathbf{X}_i\beta$ and $\text{cov}(\mathbf{Y}_i) = \mathbf{\Sigma}_i$), the maximum likelihood estimate of β is

$$\hat{\beta} = \left[\sum_{i=1}^{n} \mathbf{X}_i' \widehat{\mathbf{\Sigma}}_i^{-1} \mathbf{X}_i \right]^{-1} \sum_{i=1}^{n} \mathbf{X}_i \widehat{\mathbf{\Sigma}}_i^{-1} \mathbf{Y}_i,$$

which solves the score equations

$$\sum_{i=1}^{n} \mathbf{X}_i' \widehat{\mathbf{\Sigma}}_i^{-1} \left(\mathbf{Y}_i - \mathbf{X}_i\beta \right) = \mathbf{0}.$$

For our (uncorrelated) data, the score equations are

$$\sum_{i=1}^{n} \underbrace{\mathbf{D}_i'}_{\mathbf{B}_i \mathbf{\Delta}_i \mathbf{X}_i} \mathbf{B}_i^{-1} \left(\mathbf{Y}_i - \mu_i(\beta) \right) = \mathbf{0}.$$

GEE simply replaces $\mathbf{B}_i$ by $\mathbf{V}_i = \mathbf{B}_i^{1/2} \mathbf{R}(\alpha) \mathbf{B}_i^{1/2}$. Note that $\mathbf{V}_i = \widehat{\mathbf{\Sigma}}_i$ in the correlated normal model.

Sandwich estimator of cov($\hat{\boldsymbol{\beta}}$)

Liang and Zeger (1986) show $\hat{\boldsymbol{\beta}} \stackrel{\bullet}{\sim} N_p(\boldsymbol{\beta}, \mathbf{V}_G)$ where

$$\mathbf{V}_G = \left[\sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i^{-1} \mathbf{D}_i \right]^{-1} \left[\sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i^{-1} \text{cov}(\mathbf{Y}_i) \mathbf{V}_i^{-1} \mathbf{D}_i \right] \left[\sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i^{-1} \mathbf{D}_i \right]^{-1}.$$

Here $\boldsymbol{\beta}$ is replaced by $\hat{\boldsymbol{\beta}}$, ϕ replaced with $\hat{\phi}$ ($\phi = 1$ for binomial and Poisson models), and $\boldsymbol{\alpha}$ replaced by $\hat{\boldsymbol{\alpha}}$. $\text{cov}(\mathbf{Y}_i)$ is replaced by $[\mathbf{y}_i - \boldsymbol{\mu}_i(\hat{\beta})][\mathbf{y}_i - \boldsymbol{\mu}_i(\hat{\beta})]'$.

This *sandwich estimator* sandwiches an empirical estimate between the theoretical (working guess) $\left[\sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i^{-1} \mathbf{D}_i \right]^{-1}$. If we know for certain (we don't) that $\text{corr}(\mathbf{Y}_i) = \mathbf{R}(\boldsymbol{\alpha})$, then we can use this instead (MODELSE in the REPEATED statement).

GEE approach summary

To reiterate, the ingredients for the marginal GEE approach are

- A marginal model where Y_{ij} is binomial, Poisson, normal, gamma, etc. with mean $\mu_{ij} = g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta})$.

  Note that often for repeated measures, $\mathbf{x}_{ij} = \mathbf{x}_i$ for $j = 1, \ldots, T_i$; e.g. gender and weight are not apt to change over a 6 month study.

- An assumption on how the elements of $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iT_i})$ are correlated, $\mathrm{corr}(\mathbf{Y}_i) = \mathbf{R}(\boldsymbol{\alpha})$.

Mental depression example

Table 12.1 (p. 456) houses data from a longitudinal study comparing a new drug with a standard drug for treatment of subjects suffering mental depression. $n = 340$ Patients were either mildly or severely depressed upon admission into the study. At weeks 1, 2, and 4, corresponding to $j = 1, 2, 3$, patient i's suffering Y_{ij} was classified as normal $Y_{ij} = 1$ or abnormal $Y_{ij} = 0$. Let $s_i = 0, 1$ be the severity of the diagnosis (mild, severe) and $d_i = 0, 1$ denote the drug (standard, new).

We treat time as a categorical predictor and fit a marginal logit model with an exchangeable correlation structure; note $T = 3$:

$$\text{corr}(\mathbf{Y}_i) = \text{corr} \left(\left[\begin{array}{c} Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{array} \right] \right) = \left[\begin{array}{ccc} 1 & \alpha & \alpha \\ \alpha & 1 & \alpha \\ \alpha & \alpha & 1 \end{array} \right].$$

SAS code and output

```
data depress;
 infile "c:/tim/cat/depress.txt";
 input case diagnose treat time outcome; time=time+1;
proc genmod descending;  class case time;
  model outcome = diagnose treat time treat*time / dist=bin link=logit type3;
  repeated subject=case / type=exch corrw;
```

Fit of independence model to get initial estimate of β:

```
                        Analysis Of Initial Parameter Estimates

                                     Standard    Wald 95% Confidence     Chi-
Parameter         DF    Estimate       Error           Limits          Square   Pr > ChiSq
Intercept          1      0.9812      0.1809      0.6267      1.3356     29.43      <.0001
diagnose           1     -1.3116      0.1462     -1.5981     -1.0251     80.50      <.0001
treat              1      2.0429      0.3056      1.4439      2.6420     44.68      <.0001
time        1      1     -0.9600      0.2290     -1.4088     -0.5112     17.58      <.0001
time        2      1     -0.6206      0.2245     -1.0607     -0.1806      7.64      0.0057
time        3      0      0.0000      0.0000      0.0000      0.0000       .          .
treat*time  1      1     -2.0980      0.3893     -2.8610     -1.3351     29.05      <.0001
treat*time  2      1     -1.0961      0.3838     -1.8482     -0.3439      8.16      0.0043
treat*time  3      0      0.0000      0.0000      0.0000      0.0000       .          .

                              GEE Model Information

               Correlation Structure          Exchangeable
               Subject Effect              case (340 levels)
               Number of Clusters                       340
               Correlation Matrix Dimension               3
```

```
                        Working Correlation Matrix

                         Col1        Col2        Col3
            Row1        1.0000     -0.0034     -0.0034
            Row2       -0.0034      1.0000     -0.0034
            Row3       -0.0034     -0.0034      1.0000

                       Exchangeable Working
                            Correlation

                 Correlation    -0.003436171

                  Analysis Of GEE Parameter Estimates
                  Empirical Standard Error Estimates

                            Standard    95% Confidence
Parameter      Estimate      Error        Limits           Z Pr > |Z|
Intercept       0.9812      0.1841     0.6203    1.3421    5.33  <.0001
diagnose       -1.3117      0.1453    -1.5964   -1.0269   -9.03  <.0001
treat           2.0427      0.3061     1.4428    2.6426    6.67  <.0001
time      1    -0.9601      0.2379    -1.4265   -0.4938   -4.04  <.0001
time      2    -0.6207      0.2372    -1.0855   -0.1559   -2.62  0.0089
time      3     0.0000      0.0000     0.0000    0.0000      .       .
treat*time 1   -2.0975      0.3923    -2.8663   -1.3287   -5.35  <.0001
treat*time 2   -1.0958      0.3900    -1.8602   -0.3314   -2.81  0.0050
treat*time 3    0.0000      0.0000     0.0000    0.0000      .       .
```

```
Score Statistics For Type 3 GEE Analysis

                        Chi-
Source            DF    Square    Pr > ChiSq
diagnose          1     70.83     <.0001
treat             1     40.38     <.0001
time              2     15.73     0.0004
treat*time        2     29.52     <.0001
```

Clearly, there is an important interaction between time and the treatment. The initial diagnosis is also important. Fitting two more models shows that there is no evidence of interaction between diagnosis and treatment or diagnosis and time.

We see a severe diagnosis ($s = 1$) significantly decreases the odds of a normal classification by a factor of $e^{-1.31} = 0.27$. The odds (for normal classification) ratio comparing the new drug to the standard drug changes with time because of the interaction. At 1 week it's $e^{2.04-2.09} = 0.95$, and week 2 it's $e^{2.04-1.10} = 2.6$, and at 4 weeks it's $e^{2.04-0} = 7.7$. The new drug is better, but takes time to work.

Here, the focus is on whole populations of patients at 1, 2, and 4 weeks, and on the new drug versus the standard drug. These interpretations are not within the individual, as one would make for a conditional analysis, coming up in Chapter 13.

Look at the estimate of the working correlation matrix. What does this tell you? In fact, if "comment out" the REPEATED statement and assume independent observations across individuals, i.e. Y_{i1}, Y_{i2}, Y_{i3} independent, regression coefficients and standard errors change negligibly.

Which working correlation?

When to use which correlation structure $\mathbf{R}(\boldsymbol{\alpha})$?

Because GENMOD automatically uses the "sandwich" estimate of the variance, adjusting the working correlation with an empirical (but yet model-based from mean estimates!) estimate of $\text{cov}(\hat{\boldsymbol{\beta}})$, this GEE is robust to misspecification of $\mathbf{R}(\boldsymbol{\alpha})$. However, it's nice to have a formal tool for choosing.

Pan (2001) proposes a measure analogous to AIC for quasi-likelihood termed the QIC. When $\phi = 1$ it reduces to

$$QIC = -2L(\boldsymbol{\mu}(\hat{\boldsymbol{\beta}}); \mathbf{y}_1, \ldots, \mathbf{y}_n) + 2\text{trace}(\hat{\boldsymbol{\Omega}}\mathbf{V}_G),$$

where $\hat{\boldsymbol{\Omega}} = \sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i \mathbf{D}_i$; see Pan (2001).

QIC is automatically included in version 9.2 and above.

There are two versions: QIC and QIC_u.

QIC_u replaces $2\text{trace}(\hat{\mathbf{\Omega}}\mathbf{V}_G)$ with $2p$ and should only be used to choose among regression models (with fixed working correlation), whereas QIC can be used to choose among both regression models and working correlation structure...just use QIC to be safe.

Here's my SAS code:

```
data abc1;
 input seq a b c count @@;
 datalines;
 1 0 0 0 0 1 0 0 1 2 1 0 1 0 2 1 0 1 1 9 1 1 0 0 0 1 1 0 1 0 1 1 1 0 1 1 1 1 1
 2 0 0 0 2 2 0 0 1 0 2 0 1 0 0 2 0 1 1 9 2 1 0 0 1 2 1 0 1 0 2 1 1 0 0 2 1 1 1 4
 3 0 0 0 0 3 0 0 1 1 3 0 1 0 1 3 0 1 1 8 3 1 0 0 1 3 1 0 1 3 3 1 1 0 0 3 1 1 1 1
 4 0 0 0 0 4 0 0 1 1 4 0 1 0 1 4 0 1 1 8 4 1 0 0 1 4 1 0 1 0 4 1 1 0 0 4 1 1 1 1
 5 0 0 0 3 5 0 0 1 0 5 0 1 0 0 5 0 1 1 7 5 1 0 0 0 5 1 0 1 1 5 1 1 0 2 5 1 1 1 1
 6 0 0 0 1 6 0 0 1 5 6 0 1 0 0 6 0 1 1 4 6 1 0 0 0 6 1 0 1 3 6 1 1 0 1 6 1 1 1 0
 ;
data abc2; set abc1;
 case=0;
 do i=1 to count;
  case=case+1;
  pattern=4*a+2*b+c;
  y=a; treat=1; output;
  y=b; treat=2; output;
  y=c; treat=3; output;
 end;
proc print;
proc genmod descending; class pattern case treat seq;
 model y=treat seq / dist=bin link=logit;
 repeated subject=case(seq*pattern) / type=exch;
 estimate '3 vs 1' treat -1 0 1 / exp;
 estimate '2 vs 1' treat -1 1 0 / exp;
 estimate '3 vs 2' treat 0 -1 1 / exp;
```

SAS output

```
                              GEE Model Information

          Correlation Structure                        Exchangeable
          Subject Effect                  case(pattern*seq) (86 levels)
          Number of Clusters                                86
          Correlation Matrix Dimension                       3
          Maximum Cluster Size                               3
          Minimum Cluster Size                               3

                            Exchangeable Working
                                 Correlation

                         Correlation   -0.04403048

                          Contrast Estimate Results

                        Standard                              Chi-
Label           Estimate   Error    Alpha    Confidence Limits   Square   Pr > ChiSq
3 vs 1            2.5076   0.4141    0.05    1.6959    3.3193    36.66     <.0001
Exp(3 vs 1)     12.2750   5.0836    0.05    5.4513   27.6400
2 vs 1            1.9914   0.3876    0.05    1.2317    2.7511    26.39     <.0001
Exp(2 vs 1)      7.3257   2.8396    0.05    3.4270   15.6599
3 vs 2            0.5162   0.3158    0.05   -0.1029    1.1352     2.67     0.1022
Exp(3 vs 2)      1.6756   0.5292    0.05    0.9023    3.1118
```

SAS output

```
              Empirical Standard Error Estimates

                          Standard   95% Confidence
    Parameter   Estimate    Error        Limits        Z Pr > |Z|
    Intercept     0.9554    0.3282    0.3121   1.5987   2.91  0.0036
    treat    1   -2.5076    0.4141   -3.3193  -1.6959  -6.05  <.0001
    treat    2   -0.5162    0.3158   -1.1352   0.1029  -1.63  0.1022
    treat    3    0.0000    0.0000    0.0000   0.0000    .     .
    seq      1    0.5200    0.3907   -0.2459   1.2858   1.33  0.1833
    seq      2    0.7775    0.5352   -0.2715   1.8265   1.45  0.1463
    seq      3    0.6454    0.3865   -0.1122   1.4029   1.67  0.0950
    seq      4    0.5830    0.4230   -0.2460   1.4121   1.38  0.1681
    seq      5    0.2384    0.5116   -0.7642   1.2410   0.47  0.6412
    seq      6    0.0000    0.0000    0.0000   0.0000    .     .
```

I am nesting the subject (case) index within both the drug
sequence $k = 1, \ldots, 6$ and pattern type $p = 1, \ldots, 8$ for
$(0, 0, 0), (0, 0, 1), \ldots, (1, 1, 1)$. The model looks like

$$\text{logit } P(Y_{i(k*p)j} = 1) = \gamma + \alpha_k + \beta_j,$$

where $\beta_3 = \alpha_6 = 0$ correspond to baseline.

When j indexes time, $Y_{i1}, Y_{i2}, \ldots, Y_{iT_i}$ is a stochastic process, often termed a *time series*. Let's consider $Y_{ij} = 0, 1$ for now.

The series $Y_{i1}, Y_{i2}, \ldots, Y_{iT_i}$ follows a first-order Markov chain if the distribution of Y_{ij} only cares about the previous value $Y_{i,j-1}$, formally $[Y_{ij}|Y_{i1}, \ldots, Y_{i,j-1}] = [Y_{ij}|Y_{i,j-1}]$.

Time-varying covariates can be included:

$$\text{logit } P(Y_{ij} = 1|Y_{i,j-1}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \gamma_1 Y_{i,j-1},$$

where γ_1 models the effect of the ith subject's previous observation on the probability of a current (time j) success. e^{γ_1} has a nice interpretation in terms of how success odds changes based on what happened at last time point.

Markov chain

Second-order, and in general t-order, Markov chains can be considered by including the most previous t observations $(Y_{i,j-1}, \ldots, Y_{i,j-t})$:

$$\text{logit } P(Y_{ij} = 1 | Y_{i,j-1}, \ldots, Y_{i,j-t}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \sum_{s=1}^{t} \gamma_s Y_{i,j-s}.$$

Interactions between covariates $\mathbf{x}_{ij}$ and previous values can also improve model fit.

Likelihood

For a first order Markov-chain with no interaction the likelihood is written

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} f_1(y_{i1}) f_2(y_{i2}|y_{i1}) f_3(y_{i3}|y_{i2}) \cdots f_{T_i}(y_{iT_i}|y_{i,T_i-1}).$$

if we ignore the marginal contribution of the first observation $f_1(y_{i1})$ we get

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} f_2(y_{i2}|y_{i1}) f_3(y_{i3}|y_{i2}) \cdots f_{T_i}(y_{iT_i}|y_{i,T_i-1}).$$

For each subject i we have the product of $T_i - 1$ conditional logistic regression kernels; the transitional model can be fit in PROC LOGISTIC as usual, but for observation Y_{ij}, treating $Y_{i,j-1}$ as an observed predictor!

Example (p. 476): Children were evaluated every year on whether they had a respiratory illness. A covariate of interest is whether the child's mom smoked at the beginning of the study; $s_i = 0$ indicates not and $s_i = 1$ indicates a smoker.

Each child has a sequence of 4 indicators $(Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$ taken at 7, 8, 9, and 10 years. For each child we have covariates s_i and $t_j = j + 6$. The first order Markov model is fit

$$\text{logit } P(Y_{ij} = 1 | Y_{i,j-1} = y_{i,j-1}) = \beta_0 + \beta_1 s_i + \beta_2 t_j + \beta_3 y_{i,j-1},$$

for $i = 1, \ldots, 537$ and $j = 2, 3, 4$.

SAS code to fit the Markov model

```
data mm1;
 input s y1 y2 y3 y4 count;
 y=y2; yp=y1; sm=s; t=8;  ct=count; output;
 y=y3; yp=y2; sm=s; t=9;  ct=count; output;
 y=y4; yp=y3; sm=s; t=10; ct=count; output;
 datalines;
0 0 0 0 0 237
0 0 0 0 1 10
0 0 0 1 0 15
0 0 0 1 1 4
0 0 1 0 0 16
0 0 1 0 1 2
0 0 1 1 0 7
0 0 1 1 1 3
etc...
1 1 1 0 0 4
1 1 1 0 1 2
1 1 1 1 0 4
1 1 1 1 1 7
;
proc logistic descending;
 freq ct; model y=sm t yp / lackfit;
```

```
                             Analysis of Maximum Likelihood Estimates

                    Parameter    Standard      Wald         Pr >      Standardized     Odds
Variable     DF     Estimate      Error     Chi-Square   Chi-Square     Estimate       Ratio
INTERCPT      1     -0.2926      0.8460      0.1196       0.7295            .             .
SM            1      0.2960      0.1563      3.5837       0.0583         0.077761       1.344
T             1     -0.2428      0.0947      6.5800       0.0103        -0.109336       0.784
YP            1      2.2111      0.1582    195.3589       0.0001         0.450688       9.126

                         Hosmer and Lemeshow Goodness-of-Fit Test

                 Goodness-of-fit Statistic = 1.1723 with 6 DF (p=0.9782)
```

We see both time and whether the child had a respiratory illness
the previous year are important predictors. Smoking is *almost*
significant at the 5% level (and is significant if we perform a
one-sided test). Maternal smoking increases the odds of a
respiratory illness by about 34%. As time goes on the child is less
likely to have a respiratory illness. If a child had a respiratory
illness last year, the odds of having one this year are nine times
greater than if the child did not have one last year.