8. Transition Models

- The distribution of the observed response at time j, Y_{ij}, is modeled conditionally as an explicit function of the past responses $\mathcal{H}_{ij} = (Y_{i1}, \cdots, Y_{ij-1})$ and covariates X_{ij}.

- Typically, a Markov model is assumed, that is, Y_{ij} only depends on q (the order of the Markov process) previous responses

$$P(Y_{ij}|\mathcal{H}_{ij}) = P(Y_{ij}|Y_{ij-1}, \cdots, Y_{ij-q}).$$

- For notational convenience, we assume that the observational times are equally spaced. If they are not, we need stronger assumptions about the functional form of the time dependence.

Model Specification

- $Y_{ij}|\mathcal{H}_{ij}$ is assumed to be independent from the exponential family:

$$f(y_{ij}|\mathcal{H}_{ij}) = \exp\left\{[y_{ij}\theta_{ij} - b(\theta_{ij})]/\phi + c(y_{ij}, \phi)\right\}.$$

- Conditional mean $\mu_{ij}^c = E(Y_{ij}|\mathcal{H}_{ij}) = \dot{b}(\theta_{ij})$ satisfies

$$g(\mu_{ij}^c) = X_{ij}^T\beta + \sum_{r=1}^{q} f_r(\mathcal{H}_{ij};\alpha)$$

for some functions $f_r(\cdot)$.

- Conditional variance

$$v_{ij}^c = var(Y_{ij}|\mathcal{H}_{ij}) = \ddot{b}(\theta_{ij})\phi$$

satisfies

$$v_{ij}^c = v(\mu_{ij}^c)\phi.$$

Examples

- Continuous response: linear regression with autoregressive errors.

$$Y_{ij} = X_{ij}^T\beta + \sum_{r=1}^{q} \alpha_r (y_{ij-r} - X_{ij-r}^T\beta) + \epsilon_{ij},$$

  where ϵ_{ij} are iid zero-mean Gaussian r.v.'s.

- Binary responses:

$$g(\mu_{ij}^c) = \mathsf{logit}(\mu_{ij}^c) = X_{ij}^T\beta + \sum_{r=1}^{q} \alpha_r y_{ij-r}.$$

  The interpretation of the regression coefficients depends on the order q.

- Count responses: $q = 1$

$$\log(\mu_{ij}^c) = X_{ij}^T\beta + \alpha(\log y_{ij-1}^* - X_{ij-1}^T\beta)$$

where

$$y^*_{ij-1} = \mathsf{max}(y_{ij-1}, c), \quad 0 < c < 1$$

which leads to

$$\mu^c_{ij} = e^{X^T_{ij}\beta} \left(\frac{y^*_{ij-1}}{\exp(X^T_{ij-1}\beta)} \right)^{\alpha}.$$

– The constant c prevents $y_{ij-1} = 0$ from being an absorbing state (otherwise $Y_{ij-1} = 0 \;=>\; Y_{ik} = 0$ for all $k \geq j$).
– For $\alpha < 0$, a response at time $t-1$ greater than $e^{X^T_{t-1}\beta}$ (not its expected value) decreases the expectation for the current response. When $\alpha > 0$ the opposite occurs (positive correlation).

Fitting Transitional Models

- For weak stationary Gaussian process, the marginal distribution of $Y_i = (Y_{i1}, \cdots, Y_{in})$ can be fully determined from the conditional model without additional unknown parameters.

- When the marginal distribution of Y_i is not fully specified by the conditional model, we can estimate β and α by maximizing the conditional likelihood, which is (for one subject i)

$$\begin{aligned}\mathcal{L}_i^c(\beta, \alpha) &= f(Y_{iq+1}, \cdots, Y_{in} | Y_{i1}, \cdots, Y_{iq}; \beta, \alpha) \\ &= \prod_{j=q+1}^{n} f(Y_{ij} | Y_{ij-1}, \cdots, Y_{ij-q}; \beta, \alpha).\end{aligned}$$

- If $f_r(\mathcal{H}_{ij}; \alpha) = \alpha_r f_r(\mathcal{H}_{ij})$ where f_r is known (does not depend on unknown parameters β or α), we can simply regress Y_{ij} on $(X_{ij}, f_1(\mathcal{H}_{ij}), \cdots, f_r(\mathcal{H}_{ij}))$.

- In general, $f_r(\mathcal{H}_{ij}; \alpha)$ may include α and (perhaps implicitly) β. The conditional score function is

$$S^c(\delta) = \frac{\partial \mathcal{L}^c(\delta)}{\partial \delta} = \sum_{i=1}^{m} \prod_{j=q+1}^{n} \frac{\partial \mu_{ij}^c}{\partial \delta} \left(v_{ij}^c\right)^{-1} (y_{ij} - \mu_{ij}^c)$$

where $\delta = (\beta, \alpha)$. The derivative $\partial \mu_{ij}^c / \partial \delta$ depends on both β and α.

- Intuitively we can use an iterative algorithm to estimate δ.
  - Given current estimate of δ, calculate $\partial \mu_{ij}^c / \partial \delta$ and v_{ij}^c.
  - Update δ by solving the estimating equation.

- Statistical package developed for GEE of marginal models can be utilized, and this approach shares the same robustness property enjoyed by GEE for marginal models.

- The calculations of $\hat{\mu}_{ij}^c$ and $\partial \mu_{ij}^c / \partial \delta$ are recursive and need to be carried out in turn for $j = q + 1, \cdots, n$

- If q is large relative to n_i, the use of transitional models with conditional likelihood could be inefficient.

- If the conditional mean is correctly specified but the conditional variance is not, we can use empirical variance estimates to get consistent inferences about δ.

- When the Markov assumption does not hold, remarkably we can still get consistent estimate of β but that is a "right answer to the wrong question".

Transition models for Binary Responses data

- A first-order Markov chain is characterized by the transition matrix

$$\begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix}.$$

  Two possible states: 1 (disease), 0 (no disease) and π_{ab}: transition probability from state a to state b.

- We can model the transition probabilities as function of covariates using separate regressions

$$\text{logit}P(Y_{ij} = 1|Y_{ij-1} = 0, x_{ij}) = x_{ij}^T\beta_0,$$
$$\text{logit}P(Y_{ij} = 1|Y_{ij-1} = 1, x_{ij}) = x_{ij}^T\beta_1.$$

- This is equivalent to the transition model

$$\text{logit}P(Y_{ij} = 1|y_{ij-1}) = x_{ij}^T\beta + y_{ij-1}x_{ij}^T\alpha$$

  where $\beta = \beta_0$ and $\alpha = \beta_1 - \beta_0$.

- The transition probabilities are

$$\pi_{01} = \frac{e^{x_{ij}^T \beta_0}}{1 + e^{x_{ij}^T \beta_0}}, \quad \pi_{00} = 1 - \pi_{01}$$

$$\pi_{11} = \frac{e^{x_{ij}^T \beta_1}}{1 + e^{x_{ij}^T \beta_1}}, \quad \pi_{10} = 1 - \pi_{11}$$

- We can test whether certain covariates have effects on the transition probabilities by testing $H_0 : \alpha = (\alpha_0, 0)$.

Marginalized Likelihood Models

- In marginal models, the interpretation of the marginal regression coefficients β^M does not depend on the specification of the dependence structure.

- We have been using GEE for estimation in marginal models.
  - GEE yields consistent estimator for β^M even when the dependence model is misspecified.
  - Valid inference is achieved by using empirical variance estimates.
  - GEE for marginalized models is computationally efficient.

- Likelihood-based inference is still attractive.
  - MLE can be more efficient.
  - The likelihood can be used for comparing models.
  - The existence of likelihood allows flexible modeling of missing at random (MAR).

- The idea of marginalized likelihood models is to use a random effects/latent variable/transition model only for the dependence structure. It allows likelihood-based inference and retains the advantage of marginal models.

- A marginalized likelihood model is appropriate when the dependence structure and subject specific effects are not of interest.

- A marginalized model has two parts:

  – Marginal regression model

$$g(E(Y_{ij}|X_i)) = x_{ij}^T \beta^M.$$

  – Dependence model: for some variable A_{ij},

$$g\left\{E(Y_{ij}|X_i, A_{ij})\right\} = \Delta_{ij}(X_i) + \gamma_{ij}^T A_{ij},$$

- A_{ij} is introduced to account for the dependence.

  – **Marginalized log-linear model**:

$$A_{ij} = \left\{Y_{ij} : k \neq j\right\}.$$

- **Marginalized latent variable (random effects) model**:

$$A_{ij} = U_i.$$

- **Marginalized transition model**:

$$A_{ij} = \{Y_{ik} : k < j\} = \mathcal{H}_{ij}.$$

- $\Delta_{ij}(X_i)$ is a function of the marginal means μ_{ij}^M and dependence parameters γ_{ij}. It is chosen such that

$$\begin{aligned} \mu_{ij}^M &= E_{A_{ij}}\left[E(Y_{ij}|X_i, A_{ij})\right] \\ &= E_{A_{ij}}\left[g^{-1}\left(\Delta_{ij}(X_i) + \gamma_{ij}^T A_{ij}\right)\right] \\ &= g^{-1}(x_{ij}^T \beta^M). \end{aligned}$$

- We need solve the above integral equation for $\Delta_{ij}(X_i)$ to evaluate to the likelihood for (β^M, γ).

Example: Madras Schizophrenia Study

- A longitudinal study where schizophrenia symptoms (e.g., thoughts disorder presence yes/no) were recorded monthly in the first year following hospitalization.

- 86 subjects: covariates include age, gender and time.

- 17 subjects only have partial follow-up. There is evidence suggesting the dropout is not missing completely at random (MCAR).

- We are interested in factos that correlate with the course of illness, in particular, the interactions "time $\times$ age-at-onset" and "time $\times$ gender".

- For "thoughts", the serial correlation decays with time interval.

Figure 1: MADRAS Study: Thoughts

Calculation of (crude) lorelogram:

```
setwd("d:/course/SKKU/Longitudinal_Data_Analysis/2016Fall/R-codes")

madras <- read.table('madras.dat',col.names=c("id","thoughts","month",
                     "age","gender","month.age","month.gender"))
madras.w <-reshape(madras[,1:5],direction="wide",
                   v.names="thoughts",timevar="month",idvar="id")
thoughts <-madras.w[,-c(1:3)]
n <-ncol(thoughts)
ttall <-array(0,dim=c(nrow(thoughts),2,2,n-1))
tt <- matrix(0,2,2)
for(i in 1:nrow(thoughts)){
  y <-as.numeric(thoughts[i,])
  for(lag in 1:(n-1)){
```

```
    tmp <-cbind(y[1:(n-lag)],y[(lag+1):n])
    tmp <-na.omit(tmp)
    tt[1,1] <-sum(tmp[,1]+tmp[,2]==0)
    tt[2,2] <-sum(tmp[,1]+tmp[,2]==2)
    tt[1,2] <-sum(tmp[,2]+tmp[,1]==1)
    tt[2,1] <-sum(tmp[,2]+tmp[,1]==-1)
    ttall[i,,,lag] <-tt
  }
}
ttacross <- apply(ttall,c(2,3,4),sum)

library(vcd)
plot(oddsratio(ttacross),ylim=c(-1,4.5),xlab="Time Lag (month)",
     main="MADRAS Study: Thoughts")
```

Madras Study: Models

- Covariates: age at enrollment, time (t, months after follow-up), gender, time by gender, time by age.
- GLMM with random intercept:

$$\text{logit}(\mu_{ij}^c) = x_{ij}^c \beta^c + b_{0i},$$
$$b_{0i} \sim N(0, G).$$

- GLMM with random intercept and random slope for time:

$$\text{logit}(\mu_{ij}^c) = x_{ij}^T \beta^c + b_{0i} + b_{1i} t_{ij},$$
$$\gamma_{ij1} = \alpha_{1,0} \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N\left(0, \begin{pmatrix} G_{11} & R \\ R & G_{22} \end{pmatrix}\right).$$

- GLMM with autocorrelated random effects:

$$\begin{aligned} \text{logit}(\mu_{ij}^c) &= x_{ij}^T \beta^c + U_{ij}, \\ U_{ij} &\sim N(0, G), \\ Cor(U_{ij}, U_{ik}) &= \rho^{|t_{ij} - t_{ik}|}. \end{aligned}$$

  There are $n_i = 12$ random effects. When $\rho = 1$, reduced to a single random intercept model.

- GEE with independent, exchangeable or AR(1) working variance.

$$\text{logit}(\mu_{ij}^M) = x_{ij}^T \beta^M.$$

- MTM: The marginalized transition models have the same mean model:

$$\text{logit}(\mu_{ij}^M) = x_{ij}^T \beta^M.$$

  For dependence:

  – MTM(1): First order transition model:

$$\begin{aligned} \text{logit}(E(Y_{ij}|x_{ij}, \mathcal{H}_{ij})) &= \Delta_{ij} + \gamma_{ij1} y_{ij-1}, \\ \gamma_{ij1} &= \alpha_{10}. \end{aligned}$$

  – MTM(2): Second order transition model:

$$\begin{aligned} \text{logit}(E(Y_{ij}|x_{ij}, \mathcal{H}_{ij})) &= \Delta_{ij} + \gamma_{ij1} y_{ij-1} + \gamma_{ij2} y_{ij-2}, \\ \gamma_{ij1} &= \alpha_{10} + \alpha_{11} 1_{j=1} \\ \text{or} & \\ \gamma_{ij1} &= \alpha_{10} + \alpha_{11} 1_{j=1} + \alpha_{12} t, \\ \gamma_{ij2} &= \alpha_{20}. \end{aligned}$$