Transition Models

Outline

- Model Specification
- Fitting Transition Models
- Transition Models for Binary Responses Data
- Marginalized Likelihood Models
- Example

Transition Models

- The distribution of the observed response at time j, Y_{ij}, is modeled conditionally as an *explicit* function of the past responses $\mathcal{H}_{ij} = (Y_{i1}, \ldots, Y_{ij-1})$ and covariates $\boldsymbol{X}_{ij}$.
- Typically, a Markov model is assumed, that is, Y_{ij} only depends on q (the *order* of the Markov process) previous responses

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

- For notational convenience, we assume that the observational times are equally spaced. If they aren't, 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:

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

- Conditional mean $\mu_{ij}^C = \mathrm{E}(Y_{ij} \,|\, \mathcal{H}_{ij}) = b'(\theta_{ij})$ satisfies

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

for some functions $f_r(\,\cdot\,)$.

- Conditional variance

$$v_{ij}^C = \mathrm{Var}(Y_{ij} \,|\, \mathcal{H}_{ij}) = b''(\theta_{ij})\phi$$

satisfies

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

Examples

- Continuous response: linear regression with autoregressive errors.

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

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

  – $E[Y_{ij}] = \boldsymbol{X}_{ij}^T\boldsymbol{\beta}$ no matter what q is.

- Binary responses:

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

  The interpretation of the regression coefficients depends on the order q (i.e. $\boldsymbol{\beta} = \boldsymbol{\beta}_q$).

- Count responses: $q = 1$

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

  where

$$y_{ij-1}^* = \max(y_{ij-1}, c); 0 < c < 1$$

  which leads to

$$\mu_{ij}^c = e^{\boldsymbol{X}_{ij}^T\boldsymbol{\beta}} \left(\frac{y_{ij-1}^*}{\exp(\boldsymbol{X}_{ij-1}^T\beta)} \right)^{\alpha}.$$

  – The constant c prevents $y_{i,j-1} = 0$ from being an absorbing state (otherwise $Y_{ij-1} = 0 \Rightarrow Y_{ik} = 0$ for all $k \geq j$).

  – For $\alpha < 0$, a response at time $t-1$ greater than $e^{\boldsymbol{X}_{t-1}^T\boldsymbol{\beta}}$ (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 $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in})$ can be fully determined from the conditional model without additional unknown parameters.

- When the marginal distribution of $\boldsymbol{Y}_i$ is not fully specified by the conditional model, we can estimate $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ by maximizing the conditional likelihood, which is (for one subject i)

$$\begin{aligned}\mathcal{L}_i^C(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= f(Y_{iq+1}, \ldots, Y_{in_i} \,|\, Y_{i1}, \ldots, Y_{iq}; \boldsymbol{\beta}, \boldsymbol{\alpha}) \\ &= \prod_{j=q+1}^{n_i} f(Y_{ij} \,|\, Y_{ij-1}, \ldots, Y_{ij-q}; \boldsymbol{\beta}, \boldsymbol{\alpha}).\end{aligned}$$

- If $f_r(\mathcal{H}_{ij}; \boldsymbol{\alpha}) = \alpha_r f_r(\mathcal{H}_{ij})$ where f_r is known (does not depend on unknown parameters $\boldsymbol{\beta}$), we can simply regress Y_{ij} on the $(p+q)$-dimensional variables $(\boldsymbol{X}_{ij}, f_1(\mathcal{H}_{ij}), \ldots, f_r(\mathcal{H}_{ij}))$.

- In general, $f_r(\mathcal{H}_{ij}; \boldsymbol{\alpha})$ may include both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The conditional score function is

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

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

- Intuitively we can use an iterative algorithm to estimate $\boldsymbol{\delta}$.
  - Given current estimate of $\boldsymbol{\delta}$, calculate $\partial \mu_{ij}^C / \partial \boldsymbol{\delta}$ and v_{ij}^C.
  - Update $\boldsymbol{\delta}$ 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 $\partial \hat{\mu}_{ij}^C / \partial \boldsymbol{\delta}$ are recursive and need to be carried out in turn for $j = q+1, \ldots, n$.

- 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 $\boldsymbol{\delta}$.
- Interestingly, when the Markov assumption does not hold, we can still get consistent confidence intervals for $\hat{\boldsymbol{\delta}}$. However, the interpretation of $\hat{\boldsymbol{\delta}}$ is questionable because $\mu_{ij}^C(\hat{\boldsymbol{\delta}})$ is not the conditional mean anymore.

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 functions of covariates using separate regressions
$$\begin{aligned} \operatorname{logit} \Pr(Y_{ij} = 1 \mid Y_{ij-1} = 0, \boldsymbol{x}_{ij}) &= \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_0, \\ \operatorname{logit} \Pr(Y_{ij} = 1 \mid Y_{ij-1} = 1, \boldsymbol{x}_{ij}) &= \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_1. \end{aligned}$$

- This is equivalent to the transition model
$$\operatorname{logit} \Pr(Y_{ij} = 1 \mid y_{ij-1}) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + y_{ij-1} \boldsymbol{x}_{ij}^T \boldsymbol{\alpha}$$
where
$$\boldsymbol{\beta} = \boldsymbol{\beta}_0 \qquad \text{and} \qquad \boldsymbol{\alpha} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_0.$$

- The transition probabilities are
$$\begin{aligned} \pi_{01} &= \frac{e^{\boldsymbol{x}_{ij}^T \boldsymbol{\beta}_0}}{1 + e^{\boldsymbol{x}_{ij}^T \boldsymbol{\beta}_0}}, \qquad \pi_{00} = 1 - \pi_{01} \\ \pi_{11} &= \frac{e^{\boldsymbol{x}_{ij}^T \boldsymbol{\beta}_1}}{1 + e^{\boldsymbol{x}_{ij}^T \boldsymbol{\beta}_1}}, \qquad \pi_{10} = 1 - \pi_{11} \end{aligned}$$

- We can test whether certain covariates have effects on the transition probabilities by testing $H_0 : \boldsymbol{\alpha} = (\boldsymbol{\alpha}_0, \boldsymbol{0})$.

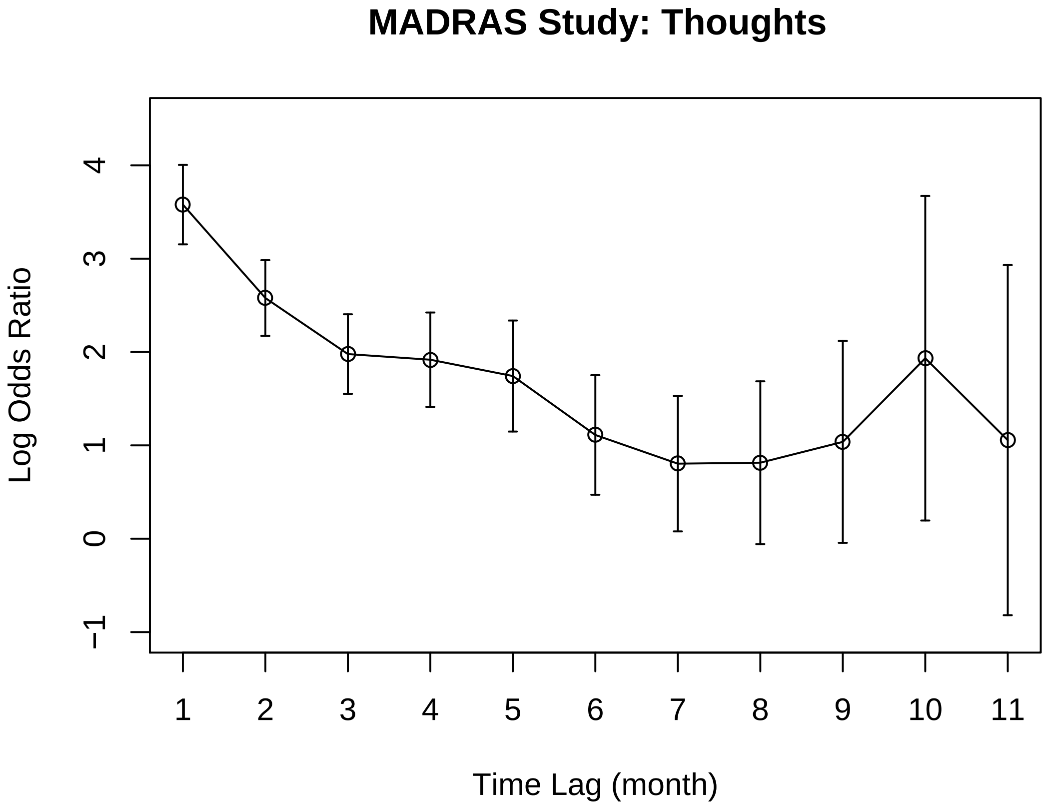
Marginalized Likelihood Models

- In marginal models, the interpretation of the marginal regression coefficients $\boldsymbol{\beta}^M$ 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 $\boldsymbol{\beta}^M$ 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
  $$g\Big(\mathrm{E}(Y_{ij} \,|\, \mathsf{X}_i)\Big) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}^M.$$
  - Dependence model: for some variable $\boldsymbol{A}_{ij}$,
  $$g\left\{\mathrm{E}(Y_{ij} \,|\, \mathsf{X}_i, \boldsymbol{A}_{ij})\right\} = \Delta_{ij}(\mathsf{X}_i) + \boldsymbol{\gamma}_{ij}^T \boldsymbol{A}_{ij}.$$

- $\boldsymbol{A}_{ij}$ is introduced to account for the dependence.
  - *Marginalized log-linear model*:
$$\boldsymbol{A}_{ij} = \{Y_{ik} : k \neq j\}.$$
  - *Marginalized latent variable (random effects) model*:
$$\boldsymbol{A}_{ij} = \boldsymbol{U}_i.$$
  - *Marginalized transition model*:
$$\boldsymbol{A}_{ij} = \{Y_{ik} : k < j\} = \mathcal{H}_{ij}.$$
- $\Delta_{ij}(\mathsf{X}_i)$ is a function of the marginal means μ_{ij}^M and dependence parameters $\boldsymbol{\gamma}_{ij}$. It is chosen such that
$$\begin{aligned}\mu_{ij}^M &= \mathrm{E}_{\boldsymbol{A}_{ij}}\left[\mathrm{E}(Y_{ij} \mid \mathsf{X}_i, \boldsymbol{A}_{ij})\right] \\ &= \mathrm{E}_{\boldsymbol{A}_{ij}}\left[g^{-1}(\Delta_{ij}(\mathsf{X}_i) + \boldsymbol{\gamma}_{ij}^T \boldsymbol{A}_{ij})\right] \\ &= g^{-1}(\boldsymbol{x}_{ij}^T \boldsymbol{\beta}^M)\end{aligned}$$

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 drop-out is not missing completely at random (MCAR).
- We are interested in factors that correlate with the course of illness, in particular, the interactions "time × age-at-onset" and "time × gender".
- For "thoughts", the serial correlation decays with time interval.



Calculation of (crude) lorelogram:

```
> madras <- read.table ("../data/madras.data",
+                        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)                           # for oddsratio()
> 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:

$$\mathrm{logit}(\mu_{ij}^C) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}^C + b_{0,i}$$
$$b_{0,i} \sim \mathcal{N}(0, G)$$

- GLMM with random intercept and random slope for time:

$$\mathrm{logit}(\mu_{ij}^C) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}^C + b_{0,i} + b_{1,i} t_{ij}$$
$$\begin{pmatrix} b_{0,i} \\ b_{1,i} \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} G_{11} & R \\ R & G_{22} \end{pmatrix}\right)$$

- GLMM with autocorrelated random effects:

$$\mathrm{logit}(\mu_{ij}^C) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}^C + U_{ij}$$
$$U_{ij} \sim \mathcal{N}(0, G)$$
$$\mathrm{Cor}(U_{ij}, U_{ik}) = \rho^{|t_{ij} - t_{ik}|}$$

  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.

$$\mathrm{logit}(\mu_{ij}^M) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}^M.$$

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

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

For dependence:

  – MTM(1): First order transition model:

$$\begin{aligned}\text{logit}(\text{E}(Y_{ij} \mid \boldsymbol{x}_{ij}^T, \mathcal{H}_{ij})) &= \Delta_{ij} + \gamma_{ij,1} y_{ij-1} \\ \gamma_{ij,1} &= \alpha_{1,0}\end{aligned}$$

  – MTM(2): Second order transition model:

$$\begin{aligned}\text{logit}(\text{E}(Y_{ij} \mid \boldsymbol{x}_{ij}^T, \mathcal{H}_{ij})) &= \Delta_{ij} + \gamma_{ij,1} y_{ij-1} + \gamma_{ij,2} y_{ij-2} \\ \gamma_{ij,1} &= \alpha_{1,0} + \alpha_{1,1} 1_{j=1}\end{aligned}$$

  or

$$\begin{aligned}\gamma_{ij,1} &= \alpha_{1,0} + \alpha_{1,1} 1_{j=1} + \alpha_{1,2} t \\ \gamma_{ij,2} &= \alpha_{2,0}\end{aligned}$$

Further Reading

- Chapter 10 of DHLZ.