Transition Models

Outline

- Model Specification
- Fitting Transition Models
- Transition Models for Binary Responses Data
- Transition Models example: ICHS data

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 from an exponential family distribution:

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

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

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

  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

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

- 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 $e^{\boldsymbol{X}^T_{t-1}\boldsymbol{\beta}}$ (not its expected value) decreases the expectation for the current response. When $\alpha > 0$ the opposite occurs (positive correlation).

Fitting Transitional Models

- The likelihood of $\boldsymbol{Y}_i$ is not always fully specified.
  - In a first order Markov model, the likelihood contribution for the ith subject

$$\begin{aligned}\mathcal{L}_i(Y_{i1}, \ldots, Y_{in_i}) &= f(Y_{i1}) f(Y_{i2}|Y_{i1}) \cdots, f(Y_{in_i}|Y_{in_i-1}) \\ &= f(Y_{i1}) \prod_{j=2}^{n_i} f(Y_{ij} \mid Y_{ij-1}).\end{aligned}$$

  - In a Markov model of order q,

$$\mathcal{L}_i(Y_{i1}, \ldots, Y_{in_i}) = f(Y_{i1}, \ldots, Y_{iq}) \prod_{j=q+1}^{n_i} f(Y_{ij} \mid Y_{ij-1}, \ldots, Y_{ij-q}).$$

    The assumed model only specifies the conditional distribution $f(y_{ij} \mid \mathcal{H}_{ij})$; $f(Y_{i1}, \ldots, Y_{iq})$ is not specified directly.

- 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)

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

- When maximizing the conditional likelihood, there are two distinct cases to consider:
  - 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 $\boldsymbol{\beta}$), i.e.

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

    which is linear in both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)$, then we can simply regress Y_{ij} on the $(p + q)$-dimensional variables $(\boldsymbol{X}_{ij}, f_1(\mathcal{H}_{ij}), \ldots, f_r(\mathcal{H}_{ij}))$. The estimation proceeds as in GLMs for independent data.

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

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

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 an iterative weighted least squares method to solve $\boldsymbol{\delta}$.

(1) Given current estimate of $\hat{\boldsymbol{\delta}}^{(l)}$. Let $\boldsymbol{X}_i^*$ be an $(n_i-q)\times(p+s)$ matrix with the k$^{\text{th}}$ row $= (\partial\mu_{i,q+k}^C)/\partial\boldsymbol{\delta}, k = 1,\ldots,n_i-q$. Let $\boldsymbol{W}_i = \text{diag}(1/v_{i,q+k}^C, k=1,\ldots,n_i-q)$ be a diagonal weighting matrix. Finally, let $\boldsymbol{Z} = \boldsymbol{X}_i^*\hat{\boldsymbol{\delta}}^{(l)} + (\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i^C)$.

(2) Regress $\boldsymbol{Z}$ on $\boldsymbol{X}_i^*$ using weights $\boldsymbol{W}$ to get $\hat{\boldsymbol{\delta}}^{(l+1)}$.

$< Why? >$

- When both the conditional mean and conditional variance are corrected specified, the variance of $\hat{\boldsymbol{\delta}}$ is $\left(\sum_{i=1}^{m} \boldsymbol{X}_i^{*T} \boldsymbol{W}_i \boldsymbol{X}_i^{*}\right)^{-1}$.

- If the conditional mean is correctly specified but the conditional variance is not, we can use empirical variance estimates to get consistent inferences about $\boldsymbol{\delta}$ (see the Weighted Least Squares lecture handout).

- 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.

- 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.

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

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 $\pi_{ab} = \Pr(Y_{ij} = b \,|\, Y_{ij-1} = a)$: transition probability from state a to state b, where $a, b = 0, 1$.

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

$$\text{logit}\,\Pr(Y_{ij} = 1 \,|\, Y_{ij-1} = 0, \boldsymbol{x}_{ij}) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_0,$$

$$\text{logit}\,\Pr(Y_{ij} = 1 \,|\, Y_{ij-1} = 1, \boldsymbol{x}_{ij}) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_1.$$

- This is equivalent to the transition model

$$\text{logit}\,\Pr(Y_{ij} = 1 \,|\, y_{ij-1}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + y_{ij-1}\boldsymbol{x}_{ij}^T\boldsymbol{\alpha} \tag{1}$$

  where

$$\boldsymbol{\beta} = \boldsymbol{\beta}_0 \qquad \text{and} \qquad \boldsymbol{\alpha} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_0.$$

- The transition probabilities are

$$\pi_{01} = \frac{e^{\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_0}}{1 + e^{\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_0}}, \qquad \pi_{00} = 1 - \pi_{01}$$
$$\pi_{11} = \frac{e^{\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_1}}{1 + e^{\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_1}}, \qquad \pi_{10} = 1 - \pi_{11}$$

- We can test whether certain covariates have effects on the transition probabilities by testing whether all items in $\boldsymbol{\alpha}$ except for the intercept term are zero: $\boldsymbol{\alpha} = (\alpha_0, \mathbf{0})$, so that $y_{ij-1}\boldsymbol{x}_{ij}^T\boldsymbol{\alpha} = \alpha_0 y_{ij-1}$, which implies that the covariates have the same effect on the response probability no matter $y_{ij-1} = 0$ or 1.

Indonesian Children Health Study (ICHS)

Transition probabilty $\Pr(Y_{ij} = 1 \,|\, y_{ij-1})$ for ICHS data: Table 10.1 in DHZL.

Table 1: Number (frequency) of transitions from respiratory disease status Y_{ij-1} at visit $j-1$ to disease status Y_{ij} at visit j for ICHS data

	Y_{ij}		
Y_{ij-1}	0	1	
0	721	60	781
	(0.923)	(0.077)	(1.0)
1	64	10	74
	(0.865)	(0.135)	(1.0)
			855

```
> xerop <- read.table ("../data/xerop.data",
+                      col.names = c("id", "RI", "intercept", "age",
+                      "xero", "cos.time", "sin.time", "sex",
+                      "height.age", "stunted", "time", "base.age", "season",
+                      "time.time"))


> #create data for transition analysis
> xeropw <- reshape (xerop, direction = "wide",
+                    v.names = c("RI", "xero", "age", "season"),
+                    idvar = c("id"), timevar = "time",
+                    drop = c("intercept", "cos.time", "sin.time","height.age",
+                    "stunted", "base.age",  "time.time","ageyr"))
> #creat previous responses
> xeropw$RIpy.2 <- xeropw$RI.1
> xeropw$RIpy.3 <- xeropw$RI.2
> xeropw$RIpy.4 <- xeropw$RI.3
> xeropw$RIpy.5 <- xeropw$RI.4
> xeropw$RIpy.6 <- xeropw$RI.5
> xeropL <- reshape (xeropw[,-(3:6)], direction = "long",idvar = c("id"),varying=3:27)
> #delete the missing values
> xeropL <- xeropL[!(is.na(xeropL$RI)|is.na(xeropL$RIpy)),]
> xeropL<- xeropL[order(xeropL$id,xeropL$time),]
> nrow(xeropL)
[1] 855

> #creat indicator for 2nd season
```

```
> xeropL$sea2 <- ifelse(xeropL$season==2,1,0)

#table 10.1
> freq.tab10.1 <- table(xeropL$RIpy,xeropL$RI)
> prop.tab10.1 <- round(prop.table(freq.tab10.1,1),3)
> marg.tab10.1 <- margin.table(freq.tab10.1,1)

> cbind(freq.tab10.1,tot=marg.tab10.1)

    0  1 tot
0 721 60 781
1  64 10  74

> prop.tab10.1

       0     1
  0 0.923 0.077
  1 0.865 0.135
```

Table 2: DHZL Table 10.2 Cross tabulation of respiratory disease Y_{ij} against xerophthalmia status x_{ij} for ICHS data

	Y_{ij}		
x_{ij}	0	1	
0	748	65	813
	(0.920)	(0.080)	(1.0)
1	37	5	42
	(0.881)	(0.119)	(1.0)
			855

An important question is whether vitA was associated with a higher prevalence of respiratory infection. Let's recreate Table 10.2 in DHZL (there is discrepancy between the table and what I got from the data).

```
> freq.tab10.2 <- table(xeropL$xero,xeropL$RI)
> prop.tab10.2 <- round(prop.table(freq.tab10.2,1),3)
> marg.tab10.2 <- margin.table(freq.tab10.2,1)
> cbind(freq.tab10.2,tot=marg.tab10.2)

    0  1 tot
0 749 65 814
1  36  5  41
```

```
> prop.tab10.2

       0     1
  0 0.920 0.080
  1 0.878 0.122

> #log odds of infection among children without xerophthalmia
> log(prop.tab10.2[1,2]/prop.tab10.2[1,1])
[1] -2.442347

> #log odds ratio of infection with vs. without xerophthalmia
> log((prop.tab10.2[2,2]/prop.tab10.2[2,1])/(prop.tab10.2[1,2]/prop.tab10.2[1,1]))
[1] 0.4687215
```

Table 3: Cross-tabulation of current respiratory disease status Y_{ij} against xerophthalmia x_{ij} and previous respiratory disease status Y_{ij-1} for ICHS data

	$Y_{ij}\|Y_{ij-1}=0$			$Y_{ij}\|Y_{ij-1}=1$		
x_{ij}	0	1		0	1	
0	688	56	744	60	9	69
	(0.925)	(0.075)	(1.0)	(0.870)	(0.130)	(1.0)
1	33	4	37	4	1	5
	(0.892)	(0.108)	(1.0)	(0.800)	(0.200)	(1.0)
			855			74

```
> #table 10.3
> mytrantab <- function(x,y)
+ freqtab<- table(x,y)
+ proptab <- round(prop.table(freqtab,1),3)
+ margtab <- margin.table(freqtab,1)
+ return(list(frequence=cbind(freqtab,tot=margtab),proption=proptab))
+
> pr0 <- xeropL$RIpy==0
> # no. infection in previous visit
> mytrantab(xeropL$xero[pr0],xeropL$RI[pr0])
$frequence
    0  1 tot
```

```
0 689 56 745
1  32  4  36
$proption
        0     1
  0 0.925 0.075
  1 0.889 0.111
> log((0.111/.889)/(.075/.925))
[1] 0.4317386

> #infection in previous visit
> mytrantab(xeropL$xero[!pr0],xeropL$RI[!pr0])
$frequence
   0 1 tot
0 60 9  69
1  4 1   5
$proption
       0    1
  0 0.87 0.13
  1 0.80 0.20
> log((0.20/.80)/(.13/.87))
[1] 0.5146644
```

The previous results suggest a model

$$\text{logit } \Pr(Y_{ij} = 1 | Y_{ij-1} = y_{ij-1}) = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \alpha y_{ij-1}.$$

First, let's fit several logistic regressions for the 855 respiratory disease transitions in ICHS data: Table 10.4 in DHZL.

```
> library(gee)
> mod1 <- gee (RI ~ xero,
+              scale.fix = TRUE, cor = "independent",
+              id = id, data = xeropL, family = "binomial")

Coefficients:
             Estimate Naive S.E.    Naive z Robust S.E.    Robust z
(Intercept) -2.4443517  0.1293048 -18.903799   0.1357392 -18.0077053
xero         0.4702707  0.4944669   0.951066   0.5413819   0.8686487


> mod2 <- gee (RI ~ xero+RIpy+RIpy*xero,
+              scale.fix = TRUE, cor = "independence",
+              id = id, data = xeropL, family = "binomial")

Coefficients:
              Estimate Naive S.E.      Naive z Robust S.E.     Robust z
(Intercept) -2.50988958  0.1389551 -18.06259331   0.1405040 -17.86347668
xero         0.43044804  0.5482322   0.78515646   0.5076427   0.84793510
```

```
RIpy         0.61276960  0.3835183   1.59775836   0.4054718   1.51125072
xero:RIpy    0.08037759  1.2955062   0.06204338   1.1035169   0.07283766
```

```
> mod3 <- gee (RI ~ xero+RIpy,
+              scale.fix = TRUE, cor = "independence",
+              id = id, data = xeropL, family = "binomial")

Coefficients:
             Estimate Naive S.E.     Naive z Robust S.E.   Robust z
(Intercept) -2.5108031  0.1382267 -18.1643818   0.1407673 -17.836551
xero         0.4445944  0.4962070   0.8959858   0.5338671   0.832781
RIpy         0.6197126  0.3661443   1.6925366   0.3847436   1.610716
```

```
> mod4 <- gee (RI ~ xero+age+sea2+RIpy+RIpy*xero+RIpy*age+RIpy*sea2,
+              scale.fix = TRUE, cor = "independence",
+              id = id, data = xeropL, family = "binomial")

Coefficients:
                Estimate  Naive S.E.      Naive z Robust S.E.     Robust z
(Intercept) -2.8462484299 0.181559204 -15.67669593 0.175640463 -16.20496998
xero         0.8124232191 0.567694732   1.43109170 0.493330064   1.64681474
age         -0.0237589967 0.007496458  -3.16936293 0.007002374  -3.39299171
sea2         1.2317354199 0.287880788   4.27863016 0.277189207   4.44366298
```

```
RIpy         0.8277026434 0.457623654   1.80869725 0.444357904   1.86269364
xero:RIpy   -0.1389588693 1.329988869  -0.10448123 1.112323923  -0.12492662
age:RIpy     0.0006205091 0.028164087   0.02203193 0.023899760   0.02596298
sea2:RIpy   -1.2376343514 1.190985488  -1.03916829 1.185491429  -1.04398423
```

```
> mod5 <- gee (RI ~ xero+age+sea2+RIpy,
+              scale.fix = TRUE, cor = "independence",
+              id = id, data = xeropL, family = "binomial")

Coefficients:
               Estimate Naive S.E.    Naive z Robust S.E.   Robust z
(Intercept) -2.81006738 0.17481819 -16.074227 0.169882045 -16.541285
xero         0.80140086 0.51204568   1.565096 0.532535939   1.504877
age         -0.02314326 0.00718448  -3.221285 0.006508378  -3.555918
sea2         1.13668651 0.27564425   4.123745 0.270214068   4.206615
RIpy         0.62233686 0.37747929   1.648665 0.404181119   1.539747

> The naive S.E.'s from gee, when a working indep. corr. is assumed, are the same as those from GLM
> glm.mod1 <- glm (RI ~ xero, data = xeropL, family = "binomial")
> summary (glm.mod1)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.4444     0.1293 -18.904   <2e-16 ***
xero          0.4703     0.4945   0.951    0.342
```

Further Reading

- Chapter 10 of DHLZ.