Topic 6. Transition Models

t_{ij}'s are assumed to be equally spaced.

Let $H_{ij} = \{y_{ik}, k = 1, \cdots, j-1\}$.

Consider $f(y_{ij} \mid H_{ij}, \alpha, \beta) = \exp\{\frac{y_{ij}\theta_{ij} - \psi(\theta_{ij})}{\phi} + c(y_{ij}, \phi)\}$, where $\psi(\theta_{ij})$ and $c(y_{ij}, \phi)$ are known functions. One has $\mu_{ij}^c = E[y_{ij} \mid H_{ij}] = \psi'(\theta_{ij})$ and $v_{ij}^c = V[y_{ij} \mid H_{ij}] = \psi''(\theta_{ij})\phi$ with $h(\mu_{ij}^c) = x_{ij}{}^T\beta + \sum_{r=1}^{s} f_r(H_{ij};\alpha)$ for suitable functions $f_r(\cdot)$'s and $v_{ij}^c = v(\mu_{ij}^c)\phi$.

Examples:

(1) Linear link - linear regression with autoregressive errors for Gaussian data.

$Y_{ij} = x_{ij}{}^T\beta + \sum_{r=1}^{s}\alpha_r(Y_{i\,j-r} - x_{i\,j-r}^T\beta) + z_{ij}$, where $z_{ij} \overset{iid}{\sim} N(0,\phi)$.

$h(\mu_{ij}^c) = \mu_{ij}^c$, $v(\mu_{ij}^c) = 1$, and $f_r = \alpha_r(Y_{i\,j-r} - x_{i\,j-r}^T\beta)$

(2) Logit link -

$\operatorname{logit} P(Y_{ij} \mid H_{ij}) = x_{ij}{}^T\beta + \sum_{r=1}^{q}\alpha_r Y_{i\,j-r}$.

$h(\mu_{ij}^c) = \ln(\frac{\mu_{ij}^c}{1-\mu_{ij}^c})$, $v(\mu_{ij}^c) = \mu_{ij}^c(1-\mu_{ij}^c)$, and $f_r = \alpha_r Y_{i\,j-1}$.

(3) Log link -

$\ln Y_{ij} = x_{ij}{}^T\beta + \alpha\{\ln(Y_{i\,j-1}^*) - x_{i\,j-1}^T\beta\}$, $Y_{ij}^* = \max\{Y_{ij}, d\}$ with $0 < d < 1$. (Note that the constant d is set to prevent $Y_{ij} = 0$.

$$\mu_{ij}^c = e^{x_{ij}{}^T\beta}\left(\frac{y_{i\,j-1}^*}{\exp(x_{i\,j-1}^T\beta)}\right)^{\alpha}$$

Fitting transition models: (A markov model of order q)

By $L_i(y_{i1},\cdots,y_{im_i}) = f(y_{i1},\cdots,y_{iq})\prod_{j=q+1}^{m_i} f(y_{ij}\,|\,y_{i\,j-1},\cdots,y_{i\,j-q}),\ i=1,\cdots,n,$ one can get the likelihood function

$$\mathrm{L}(\alpha,\beta)=\prod_{i=1}^{n} f(y_{i1},\ldots,y_{iq})\prod_{j=q+1}^{n_i} f(y_{ij}\,|\,H_{ij},\alpha,\beta) \quad , \quad \text{where } H_{ij}=\{y_{i\,j-1},\cdots,y_{i\,j-q}\} \ .$$

Since the term $f(y_{i1},\cdots,y_{iq})$ is always unavailable, the estimators of (α,β) are obtained via maximizing the conditional likelihood $\prod_{i=1}^{n}\prod_{j=q+1}^{m_i} f(y_{ij}\,|\,H_{ij},\alpha,\beta)$.

Let $\delta=(\alpha,\beta)$. The log-conditional likelihood or conditional score function has the form $S^c(\delta)=\sum_{i=1}^{n}\sum_{j=(q+1)}^{m_i}\frac{\partial\mu_{ij}^c}{\partial\delta}v_{ij}^{c^{-1}}(y_{ij}-\mu_{ij}^c)=0$.

Remark.

Let $X_i^*=(\frac{\partial\mu_{ij}^c}{\partial\delta})_{(m_i+q)\times(p+s)}$ with $\dim(\beta)=p$ and $\dim(\alpha)=s$, $W_i=diag({v_{i(q+1)}^c}^{-1},\cdots,{v_{im_i}^c}^{-1})$, and $Z_i=X_i^*\hat{\delta}+(Y_i-\hat{\mu}_i^c)$.

Updated $\hat{\delta}$ can be obtained by iteratively regression Z_i on X_i^* using weight W_i as below.

$$\begin{aligned}\hat{\delta}^{(k+1)}&=\sum_{i=1}^{n}(X_i^{*T}W_iX_i)^{-1}(\sum_{i=1}^{n}X_i^{*T}W_iZ_i^{(k)})\\&=\hat{\delta}^{(k)}+(\sum_{i=1}^{n}X_i^{*T}W_iX_i)^{-1}(\sum_{i=1}^{n}X_i^{*T}W_i)(Y_i-\hat{\mu}_i^{c(k)})\end{aligned}$$

. One has $(\hat{\delta}-\delta)\xrightarrow{d}N(0,V_\delta)$, where V_δ is the robust variance with

$$V_R=(\sum_{i=1}^{n}X_i^{*T}W_iX_i)^{-1}(\sum_{i=1}^{n}X_i^{*T}W_iV_iW_iX_i^*)(\sum_{i=1}^{n}X_i^{*T}W_iX_i)^{-1}.$$

Note that δ is still consistent even the conditional variance is not specified correctly.

Transition models for binary response -

Examples:

(1) First-order binary Markov chain -

The transition matrix is $\begin{pmatrix}\pi_{00} & \pi_{01}\\ \pi_{10} & \pi_{11}\end{pmatrix}$ with $\pi_{ab}=P(Y_{ij}=b\,|\,Y_{i\,j-1}=a),\ a,b=0,1.$

Assume that $\log it P(Y_{ij}=1 \mid Y_{i\,j-1}=0)=x_{ij}{}^{T}\beta_0$ and $\log it P(Y_{ij}=1 \mid Y_{i\,j-1}=1)=x_{ij}{}^{T}\beta_1$,

where $x_{ij}=(1,x_{ij1},\cdots,x_{ijp})^T$. One has $\log it P(Y_{ij}=1 \mid Y_{i\,j-1}=y_{i\,j-1})=x_{ij}{}^{T}\beta_0+y_{i\,j-1}x_{ij}{}^{T}\alpha$,

where $\beta_1=\beta_0+\alpha$.

(2) Second-order binary Markov chain -

$Y_{i\,j-2}$	$Y_{i\,j-1}$	Y_{ij}	
		0	1
0	0	π_{000}	π_{001}
0	1	π_{010}	π_{011}
1	0	π_{100}	π_{101}
1	1	π_{110}	π_{111}

The transition probabilities can be modeled as below.

$$\log it P(Y_{ij}=1 \mid Y_{i\,j-2}=y_{i\,j-2},Y_{i\,j-1}=y_{i\,j-1})=x_{ij}{}^{T}\beta+x_{ij}{}^{T}(y_{i\,j-1}\alpha_1+y_{i\,j-2}\alpha_2+y_{i\,j-1}y_{i\,j-2}\alpha_3).$$

(3) The mixture of random effects and transition model -

$$\operatorname{logit} P(Y_{ij}=1 \mid Y_{i\,j-1}=y_{i\,j-1},U_i=u_i)=u_i+x_{ij}{}^{T}\beta+\alpha y_{i\,j-1}, \text{ where } U_i \overset{iid}{\sim} N(0,\sigma^2).$$

Ordered Categorical data -

Y : ordinal response with categories labeled $1,2,\cdots,k$.

Let $F(a \mid x)=P(Y\le a \mid x)$, where $a=1,\cdots,(k-1)$, $x=(x_1,\cdots,x_p)^T$.

Proportional odds model: $\log it F(a \mid x)=\theta_a+x^T\beta,\ \ a=1,\cdots,(k-1).$

Define $\underset{\sim}{Y}^*=(Y_1^*,\cdots,Y_{k-1}^*)$ with $Y_a{}^*=1_{(Y\le a)}$. Then, $\log it F(a \mid x)=\log it P(Y_a{}^*=1 \mid x)$.

Y	1	2	3	$\cdots$	$k-1$	k
Y_1^*	1	0	0	$\cdots$	0	0
Y_2^*	1	1	0	$\cdots$	0	0
$\vdots$	$\vdots$	$\vdots$				$\vdots$
Y_{k-1}^*	1	1	1	$\cdots$	1	0

Example:

Assume that $\operatorname{logit} P(Y_{ij} \le b \mid Y_{i\,j-1} = a) = \theta_{ab} + x_{ij}{}^T \beta_a,\ a,b = 1,\cdots,(k-1)$. It can be derived that $\operatorname{logit} P(Y_{ij} \le b \big| \underset{\sim}{Y}^*_{ij-1} = \underset{\sim}{y}^*_{i\,j-1}) = \theta_b + \sum_{l=1}^{k-1} \alpha_{ab} y^*_{i\,(j-1)l} + x_{ij}{}^T (\beta + \sum_{l=1}^{k-1} r_l y^*_{i\,(j-1)l})$,

where $\begin{cases} \theta_{kb} = \theta_b, \alpha_{ab} = \theta_{lb} - \theta_{k+1\,b} \\ \beta_k = \beta, r_l = \beta_l - \beta_{l+1} \end{cases}$.

Log-linear transition models for count data -

$Y_{ij} \mid (H_{ij}, x_{ij}) \sim Poisson(\mu^c_{ij})$.

Model 1. Wong (1986) proposed that $\mu^c_{ij} = \exp(x_{ij}{}^T \beta)\{1 + \exp(-\alpha_0 - \alpha_1 y_{i\,j-1})\}$, $\alpha_0, \alpha_1 > 0$, where β is the influence of x_{ij} as $y_{i\,j-1} = 0$.

Remark. When $y_{i\,j-1} > 0$, μ^c_{ij} decreases as $y_{i\,j-1}$ increases. A negative association is implied between the prior and current responses.

Model 2. $\mu^c_{ij} = \exp(x_{ij}{}^T \beta + \alpha y_{i\,j-1})$.

Properties: (1) μ^c_{ij} increases as an exponential function of time as $\alpha > 0$.

(2) When $\exp(x_{ij}{}^T \beta) = \mu$ and $\alpha < 0$, it leads to a stationary process.

Model 3. $\mu^c_{ij} = \exp(x_{ij}{}^T \beta + \alpha\{\ln(y^*_{i\,j-1}) - x^T_{i\,j-1}\beta\})$, where $y^*_{i\,j-1} = \max\{y_{i\,j-1}, d\}$, $0 < d < 1$.

Property: $\begin{cases} \alpha = 0: \text{ it reduces to an ordinary log-linear model.} \\ \alpha < 0: \text{ negative correlation between } y_{i\,j-1} \text{ and } y_{ij}. \\ \alpha > 0: \text{ positive correlation between } y_{i\,j-1} \text{ and } y_{ij}. \end{cases}$

Application to a size-independent branching process: $\exp(x_{ij}^{T}\beta) = \mu$

$$\begin{cases} y_{ij}: \text{ the number of individuals in the } i\text{th population at generation } j. \\ Z_k(y_{i\,j-1}): \text{ the number of offspring for person } k \text{ in generation } (j-1). \end{cases}$$

For $y_{i\,j-1} > 0$, $y_{ij} = \sum_{k=1}^{y_{i(j-1)}} Z_k(y_{i(j-1)})$, where $Z_k \overset{iid}{\sim} Poisson((\frac{\mu}{y^*_{i\,(j-1)}})^{1-\alpha})$. One can get

$\mu^c_{ij} = \mu \cdot (\frac{y^*_{i\,j-1}}{\mu})^{\alpha}$.

Property:

$\alpha < 0$: the sample paths oscillate back and forth about their long-term average level.

$\alpha > 0$: the sample paths have sharper peaks and broader valleys.