Topic 3: Generalized Linear Models for Longitudinal Data

Marginal Models -

Assumptions:

(a) $E[Y_{ij} \mid x_{ij}] = h(x_{ij}^T \beta)$, where $h(\cdot)$ is a known ~~link~~ response function.

(b) $Var(Y_{ij} \mid x_{ij}) = \phi \upsilon(h(x_{ij}^T \beta))$, where ϕ is a scale parameter and $\upsilon(\cdot)$ is a known function.

$Corr(Y_{ij}, Y_{ik} \mid x_{ij}, x_{ik}) = \rho(h(x_{ij}^T \beta), h(x_{ik}^T \beta); \alpha)$, where ρ is a known function and α is a parameter vector.

Example: Binary response ($Y_{ij} = 0 \text{ or } 1$)

Logistic regression model:

$$E[Y_{ij} \mid x_{ij}] = \frac{e^{\beta^T x_{ij}}}{1 + e^{\beta^T x_{ij}}}, Var(Y_{ij} \mid x_{ij}) = E[Y_{ij} \mid x_{ij}](1 - E[Y_{ij} \mid x_{ij}]), Corr(Y_{ij}, Y_{ik} \mid x_{ij}, x_{ik}) = \alpha .$$

Another way for modeling the association among binary data using the odds ratio:

$$OR(Y_{ij}, Y_{ik}) = \frac{P(Y_{ij} = 1, Y_{ik} = 1)P(Y_{ij} = 0, Y_{ik} = 0)}{P(Y_{ij} = 1, Y_{ik} = 0)P(Y_{ij} = 0, Y_{ik} = 1)}$$, not constrained by the means.

Random Effects Models - (Given the actual coefficients for a subject, the random effects model further assumes that the repeated measurements for each individual are independent.)

Assumptions:

(a) Given $(U_i, X_{i1}, \cdots, X_{im_i}) = (u_i, x_{i1}, \cdots, x_{im_i})$'s, Y_{ij}'s are mutually independent and follow a GLM with a density function $f(y_{ij} \mid u_i, x_{ij}) = \exp(\frac{y_{ij}\theta_{ij} - \psi(\theta_{ij})}{\phi} + c(y_{ij}, \phi))$.

(b) U_i's $\overset{i.i.d.}{\sim} F_u(\cdot)$.

Facts:

$\mu_{ij} = E[Y_{ij} \mid u_i, x_{ij}] = \psi'(\theta_{ij})$ and $\upsilon_{ij} = Var(Y_{ij} \mid u_i, x_{ij}) = \phi\psi''(\theta_{ij})$ satisfy $h(\mu_{ij}) = x_{ij}^T\beta +$ $d_{ij}^T u_i$ and $\upsilon_{ij} = \phi\upsilon(\mu_{ij})$, where $h(\cdot)$ and $\upsilon(\cdot)$ are known link functions, and d_{ij} is a subset of x_{ij}.

Transition (Markov) models -

Assumptions:

(a) $E[Y_{ij} \mid x_{ij}, H_{ij}] = h(x_{ij}^T\beta + \sum_{r=1}^{s} f_r(H_{ij};\alpha))$, where $H_{ij} = \{y_{i1}, \cdots, y_{ij-1}\}$.

(b) $Var(Y_{ij} \mid x_{ij}, H_{ij}) = \phi\upsilon(E[Y_{ij} \mid x_{ij}, H_{ij}])$.

Statistical Inferences:

Random effects model -

The likelihood function is $L(\beta, \phi, \alpha \mid \underset{\sim}{y}) = \prod_{i=1}^{n} \int (\prod_{j=1}^{m_i} f(y_{ij} \mid u_i, \beta, \phi)) dF_U(u_i;\alpha)$.

Find a maximizer of the likelihood function $L(\beta, \phi, \alpha \mid \underset{\sim}{y})$.

Transition model -

Facts: $f(y_{i1}, \cdots, y_{im_i}) = f(y_{im_i} \big| y_{im_i-1}, \cdots, y_{i1}) f(y_{im_i-1} \big| y_{im_i-2}, \cdots, y_{i1}) \cdots f(y_{i2} \big| y_{i1}) f(y_{i1})$.

In the rth-order Markov model,

$$f(y_{i1}, \cdots, y_{im_i}; \beta, \alpha^*) = f(y_{i1}; \beta, \alpha_1) \cdots f(y_{ir} \mid y_{ir-1}, \cdots, y_{i1}; \beta, \alpha_r)$$
$$\cdot \prod_{j=r+1}^{m_i} f(y_{ij} \mid y_{ij-1}, \cdots, y_{ij-r}; \beta, \alpha), \text{ where } \alpha^* = (\alpha_1, \cdots, \alpha_r, \alpha).$$

Find the maximizer of $\prod_{i=1}^{n} [\prod_{j=r+1}^{m_i} f(y_{ij} \mid y_{ij-1}, \cdots, y_{ij-r}; \beta, \alpha)]$,

where $\prod_{j=r+1}^{m_i} f(y_{ij} \mid y_{ij-1}, \cdots, y_{ij-r}; \beta, \alpha) = f(y_{ir+1}, \cdots, y_{im_i} \mid y_{i1}, \cdots, y_{ir})$

Marginal model -

Generalized Estimating Equations (GEE), which is a multivariate analogue of quasi-likelihood.

Generalized Estimating Equations (GEE):

$$S_\beta(\beta,\alpha)=\sum_{i=1}^n\left(\frac{\partial\mu_i}{\partial\beta}\right)^T[Var(Y_i)]^{-1}(Y_i-\mu_i)\text{ ,where }\mu_i=h(x_{ij}^T\beta)\,,Var(Y_i)=Var(Y_i;\beta,\alpha)$$

.

$$S_\alpha(\beta,\alpha)=\sum_{i=1}^n\left(\frac{\partial\eta_i}{\partial\alpha}\right)^T H_i^{-1}(\omega_i-\eta_i)\ ,\quad\text{where }\ \omega_i=(R_{i1}R_{i2},\cdots,R_{i1}R_{im_i},\cdots,R_{i1}^2,\cdots,R_{im_i}^2)\ ,$$

$$\eta_i=E[\omega_i|(\beta,\alpha)]\,,\ \text{and}\ H_i=Var(\omega_i)\ \cancel{\text{and}}\text{ with }R_{ij}=\frac{Y_{ij}-\mu_{ij}}{\sqrt{Var(Y_{ij})}}\,.$$

The estimator, say $(\hat\beta,\hat\alpha)$, of (β,α) is defined to be the solution of the above equations, i.e. $S_\beta(\hat\beta,\hat\alpha)=0$ and $S_\alpha(\hat\beta,\hat\alpha)=0$.

<u>Theorem 3.1.</u> Under the regularity conditions, $n^{\frac{-1}{2}}\left[\begin{pmatrix}\hat\beta\\ \hat\alpha\end{pmatrix}-\begin{pmatrix}\beta\\ \alpha\end{pmatrix}\right]\xrightarrow{d}N(0,\Sigma)$, where Σ can be estimated by $(\frac{1}{n}\sum_{i=1}^n C_i^T B_i^{-1}D_i)^{-1}(\frac{1}{n}\sum_{i=1}^n C_i^T B_i^{-1}V_{0i}D_i B_i^{-1}C_i)(\frac{1}{n}\sum_{i=1}^n D_i^T B_i^{-1}C_i)^{-1}$,

where $C_i=\begin{pmatrix}\frac{\partial\mu_i}{\partial\beta} & 0\\ 0 & \frac{\partial\eta_i}{\partial\alpha}\end{pmatrix}$, $B_i=\begin{pmatrix}Var(Y_i) & 0\\ 0 & H_i\end{pmatrix}$, $D_i=\begin{pmatrix}\frac{\partial\mu_i}{\partial\beta} & \frac{\partial\mu_i}{\partial\alpha}\\ \frac{\partial\eta_i}{\partial\beta} & \frac{\partial\eta_i}{\partial\alpha}\end{pmatrix}$, and $V_{0i}=\begin{pmatrix}y_i-\mu_i\\ \omega_i-\eta_i\end{pmatrix}^{\otimes 2}$.

Proof: (Hint)

$$\begin{pmatrix}0\\0\end{pmatrix}=\begin{pmatrix}S_\beta(\hat\beta,\hat\alpha)\\ S_\alpha(\hat\beta,\hat\alpha)\end{pmatrix}=\begin{pmatrix}S_\beta(\beta,\alpha)\\ S_\alpha(\beta,\alpha)\end{pmatrix}+\left.\begin{pmatrix}\frac{\partial S_\beta(\beta,\alpha)}{\partial\beta} & \frac{\partial S_\beta(\beta,\alpha)}{\partial\alpha}\\ \frac{\partial S_\alpha(\beta,\alpha)}{\partial\beta} & \frac{\partial S_\alpha(\beta,\alpha)}{\partial\alpha}\end{pmatrix}\right|_{\begin{pmatrix}\hat\beta^*\\ \hat\alpha^*\end{pmatrix}}\left(\begin{pmatrix}\hat\beta\\ \hat\alpha\end{pmatrix}-\begin{pmatrix}\beta\\ \alpha\end{pmatrix}\right),$$

where $\begin{pmatrix}\hat\beta^*\\ \hat\alpha^*\end{pmatrix}$ lies on the line segment between $\begin{pmatrix}\beta\\ \alpha\end{pmatrix}$ and $\begin{pmatrix}\hat\beta\\ \hat\alpha\end{pmatrix}$.

Appendix:

1. Structure of the generalized linear model (GLM) –

GLM consists of three components:

a. The random component, which specifies the p.d.f. of the response $Y_i, i = 1, \cdots, n$.

b. The systematic component, which specifies a linear function of the explanatory variables x_i, i.e. $x_i^T \beta$.

c. The link function, which describes a function relationship between the systematic component $x_i^T \beta$ and the expectation of the random component, i.e. $E[Y_i|X_i]$.

$f(y_i|\theta_i, \phi) = \exp\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i;\phi)\}$, where ϕ is called the scaling parameter (or the dispersion parameter) and θ_i is called the natural parameter.

Remark: When ϕ is known, the above model represents a linear exponential family. If ϕ is unknown, the above model is called an exponential dispersion model.

Example: $Gamma(\alpha, \beta)$ and $Normal(\mu, \sigma^2)$

Properties:

Let $L(\underset{\sim}{\theta}, \phi) = \prod_{i=1}^n f(y_i|\theta_i, \phi) = \exp\{\frac{\sum_{i=1}^n y_i\theta_i - b(\theta_i)}{a(\phi)} + \sum_{i=1}^n c(y_i;\phi)\}$, $\underset{\sim}{\theta} = (\theta_1, \cdots, \theta_n)$,

and $l(\underset{\sim}{\theta}, \phi) = \frac{\sum_{i=1}^n y_i\theta_i - b(\theta_i)}{a(\phi)} + \sum_{i=1}^n c(y_i;\phi)$ with $l_i = \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i;\phi)$.

(a) $\frac{\partial l_i}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{a(\phi)}$ and $\frac{\partial^2 l_i}{\partial \theta_i^2} = \frac{-b''(\theta_i)}{a(\phi)}$.

(b) Under the regularity conditions, $E\left[\frac{\partial l_i}{\partial \theta_i}\right] = 0$ and $-E\left[\frac{\partial^2 l_i}{\partial \theta_i^2}\right] = E\left[\left(\frac{\partial l_i}{\partial \theta_i}\right)^2\right]$

From (a) and (b), one has

(c) $E[Y_i] = b'(\theta_i) \triangleq \mu_i$ and $\frac{Var(Y_i)}{a^2(\phi)} = \frac{b''(\theta_i)}{a(\phi)}$, i.e. $Var(Y_i) = b''(\theta_i)a(\phi) = \frac{\partial \mu_i}{\partial \theta_i} a(\phi)$.

Thus,

$$E[Y]=\frac{\partial b(\theta)}{\partial\theta}=\left(\frac{\partial b(\theta_1)}{\partial\theta},\cdots,\frac{\partial b(\theta_n)}{\partial\theta_n}\right)^T$$ and $Cov(Y)=a(\phi)diag(b''(\theta_1),\cdots,b''(\theta_n))$, where

$Y=(Y_1,\cdots,Y_n)^T$.

(d) $$\frac{\partial l_i}{\partial\theta_i}=\frac{y_i-\mu_i}{a(\phi)}=\frac{\partial\mu_i}{\partial\theta_i}\left(\frac{y_i-\mu_i}{Var(y_i)}\right),i=1,\cdots,n.$$

Let $\eta_i=g(\mu_i)$ and $\eta_i=x_i^T\beta$, where $g(\cdot)$ is a monotone and differentiable function. One has

$$\frac{\partial l_i}{\partial\beta_j}=\frac{\partial\eta_i}{\partial\beta_j}\frac{\partial\mu_i}{\partial\eta_i}\frac{\partial\theta_i}{\partial\mu_i}\frac{\partial l_i}{\partial\theta_i}\quad=x_{ij}\cdot\frac{\partial\mu_i}{\partial\eta_i}\cdot\frac{\partial\theta_i}{\partial\mu_i}\cdot\left(\frac{\partial\mu_i}{\partial\theta_i}\left(\frac{y_i-\mu_i}{Var(y_i)}\right)\right)$$

$$=x_{ij}\frac{\partial\mu_i}{\partial\eta_i}\left(\frac{y_i-\mu_i}{Var(y_i)}\right),\ j=1,\cdots,p.$$

Note: $\begin{cases} g(\mu)=\mu \text{ is called the identity link} \\ g(\mu)=Q(\theta_i)=x_i^T\beta \text{ is called the canonical(natural) link}\end{cases}$

2. **Quasi Loglikelihood –**

$$Y=\begin{pmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{pmatrix}\sim(\mu,\phi\upsilon(\mu)),\text{ where }\mu=\begin{pmatrix}\mu_1\\\mu_2\\\vdots\\\mu_n\end{pmatrix}\text{ and }V=diag(\upsilon(\mu_1),\cdots,\upsilon(\mu_n)).$$

Let $U_i=\dfrac{y_i-\mu_i}{\phi\upsilon(\mu_i)}$

Properties:

(a) $E[U_i]=0$ and $Var(U_i)=\dfrac{1}{\phi\upsilon(\mu_i)}$.

$$\frac{\partial U_i}{\partial\mu_i}=\frac{-\upsilon(\mu_i)-(y_i-\mu_i)\dfrac{\partial\upsilon(\mu_i)}{\partial\mu_i}}{\phi\upsilon^2(\mu_i)}\text{ and }-E\left[\frac{\partial U_i}{\partial\mu_i}\right]=\frac{1}{\phi\upsilon(\mu_i)}\left(=Var(U_i)\right).$$

(b) U_i has the same property as the derivative of l_i.

Let $Q(\mu;Y)=\sum_{i=1}^{n} Q_i(\mu_i;y_i)$, where $Q_i(\mu_i;y_i)=\int_{y_i}^{\mu_i}\left(\frac{y_i-t}{\phi\upsilon(t)}\right)dt$ is the analogue of the log-likelihood function.

$Q(\mu;Y)$: the qusi loglikelihood.

The quasi-score function is obtained by differentiating $Q(\mu;Y)$ with $\beta=\begin{pmatrix}\beta_1\\ \vdots\\ \beta_p\end{pmatrix}$, which is equal to

$$U(\beta)=\phi^{-1}D^T V^{-1}(Y-\mu)\text{, where } D=\begin{pmatrix}\frac{\partial\mu_1}{\partial\beta_1} & \cdots & \frac{\partial\mu_1}{\partial\beta_p}\\ \frac{\partial\mu_2}{\partial\beta_1} & \cdots & \frac{\partial\mu_2}{\partial\beta_p}\\ \vdots & & \vdots\\ \frac{\partial\mu_n}{\partial\beta_1} & \cdots & \frac{\partial\mu_n}{\partial\beta_p}\end{pmatrix}.$$

The quasi-likelihood estimator $\hat{\beta}$ is defined to be the solution of $U(\hat{\beta})=0$.