Topic 4. Marginal Models

Binary responses –

(a) The log-linear model:

$t_{ij} = t_j$, $i = 1,\ldots,n$, $j = 1,\ldots,m$

$$P(Y_i = y_i) = C(\theta)\exp(\sum_{j=1}^{m}\theta_j^{(1)}y_{ij} + \sum_{j_1<j_2}\theta_{j_1j_2}^{(2)}y_{ij_1}y_{ij_2} + \cdots + \theta_{1\cdots m}^{(m)}y_{i1}\cdots y_{im}),$$

where $\theta = (\theta_1^{(1)},\ldots,\theta_m^{(1)},\theta_{12}^{(2)},\ldots,\theta_{m-1m}^{(2)},\ldots,\theta_{12\cdots m}^{(m)})$ is the canonical parameters and $C(\theta)$ is a function of θ which normalizes the p.d.f. to sum to one.

Remark.
1. This model requires the responses of each subject occurring at the common times. Otherwise, the interpretation and value of the canonical parameters will change.
2. The canonical parameters facilitate the calculation of cell probabilities but are less useful for describing the probability of Y as a function of the covariates X . This is because θ_{jk} describes the association between Y_{ij} and Y_{ik} given $Y_{il} = 0\ \forall l \neq j,k.$ Since X_i may depend on Y_{il}, it would be inappropriate to consider the conditional association.

(b) Log-linear models for marginal means:

$E[Y_j] = \mu_j$, $j = 1,\ldots,m$.

Remark. The saturated log-linear model for $Y = (Y_1,\ldots,Y_m)^T$ has $(2^m - 1)$ free parameters.

(b.1) Log-linear model: $P(Y = y) = C(\theta_1,\theta_2)\exp(Y^T\theta_1 + W^T\theta_2)$, where $W = (Y_1Y_2, Y_1Y_3, \cdots, Y_{m-1}Y_m, \cdots, Y_1Y_2\cdots Y_m)^T$, $\theta_1 = (\theta_1^{(1)},\cdots,\theta_m^{(1)})$ and $\theta_2 = (\theta_{12}^{(2)},\cdots,\theta_{m-1m}^{(2)},\cdots,\theta_{12\cdots m}^{(m)})$.

Transformation: $(\theta_1,\theta_2) \to (\mu,\theta_2)$, $\mu = (\mu_1,\ldots,\mu_m) \triangleq \mu(\theta_1,\theta_2)$.

Model assumption: $\text{logit}(\mu_j) = X_j^{\ T}\beta$.

The score equation for β under this parameterization takes the GEE form:
$(\frac{\partial\mu}{\partial\beta})^T[V(Y)]^{-1}(Y-\mu)=0$, where $\frac{\partial\mu}{\partial\beta}=(\frac{\partial\mu_1}{\partial\beta},\ldots,\frac{\partial\mu_m}{\partial\beta})^T$.

Remark. The conditional odds ratios are not easily interpreted when the association among responses is itself a focus of the study.

Properties:

1. From $M_Y(t)=E[e^{t^TY}]=\sum_y C(\theta_1,\theta_2)\exp(y^T(t+\theta_1)+w^T\theta_2)=\frac{C(\theta_1,\theta_2)}{C(\theta_1+t,\theta_2)}$, one has

$$\mu=\frac{\partial M_Y(t)}{\partial t}|_{t=0}=-\cancel{(}\frac{\frac{\partial}{\partial\theta_1}C(\theta_1,\theta_2)}{C(\theta_1,\theta_2)}\cancel{)},\ E[YY^T]=\frac{\partial^2 M_Y(t)}{\partial t\partial t^T}|_{t=0}=-\frac{\frac{\partial^2}{\partial\theta_1\partial\theta_1^T}C(\theta_1,\theta_2)}{C(\theta_1,\theta_2)}+2\mu\mu^T,$$

and $V(Y)=-\frac{\frac{\partial^2}{\partial\theta_1\partial\theta_1^T}C(\theta_1,\theta_2)}{C(\theta_1,\theta_2)}+\mu\mu^T$.

2. Let $l(\theta_1,\theta_2)=\ln P(Y=y)=\ln C(\theta_1,\theta_2)+(y^T\overset{\theta_1}{\cancel{(t+\theta_1)}}+w^T\theta_2)$. We can derive that

$$\frac{\partial l(\theta_1,\theta_2)}{\partial\theta_1}=\frac{\frac{\partial}{\partial\theta_1}C(\theta_1,\theta_2)}{C(\theta_1,\theta_2)}+Y=(Y-\mu)\cancel{,}\text{ and, hence,}$$

$$\frac{\partial l(\theta_1,\theta_2)}{\partial\beta}=(\frac{\partial\mu}{\partial\beta})^T\frac{\partial\theta_1}{\partial\mu}(\frac{\partial l(\theta_1,\theta_2)}{\partial\theta_1})=(\frac{\partial\mu}{\partial\beta})^T(\frac{\partial\mu}{\partial\theta_1})^{-1}(Y-\mu)=(\frac{\partial\mu}{\partial\beta})^T(V(Y))^{-1}(Y-\mu).$$

(b.2) The Bahadur representation:

Let $r_j=\frac{Y_j-\mu_j}{\sqrt{\mu_j(1-\mu_j)}}$, $j=1,\cdots,m$, $\rho_{jk}=E[r_jr_k]$, $\rho_{jkl}=E[r_jr_kr_l],\cdots,\rho_{12\cdots m}=E[r_1r_2\cdots r_m]$.

$$P(Y=y)=\prod_{j=1}^m\mu_j^{y_j}(1-\mu_j)^{1-y_j}(1+\sum_{j<k}\rho_{jk}r_jr_k+\sum_{j<k<l}\rho_{jkl}r_jr_kr_l+\cdots+\rho_{12\cdots m}r_1r_2\cdots r_m).$$

Remark.

1. the joint probability density function is expressed in terms of the marginal means, pairwise correlations, and higher moments of the standardized variables r_j.

2. The correlations among binary responses are constrained in complicated ways by the marginal means.

Appendix

Let $P_{[1]}(Y=y)=\prod_{j=1}^{m}\mu_j^{\,y_j}(1-\mu_j)^{1-y_j}$, $g(y)=P(Y=y)/P_{[1]}(Y=y)$, and V be a vector space of real-valued functions f on Y_1 (2^m possible values of y). Here, V is regarded as an inner-product space with $<f_1,f_2>\triangleq E_{P_{[1]}}[f_1f_2]=\sum_{y\in Y_1}f_1(y)f_2(y)P_{[1]}(y)$.

It follows easily that the set of functions $S=\{1,r_1,\ldots,r_m;r_1r_2,\ldots,r_{m-1}r_m;\ldots,r_1r_2\cdots r_m\}$ on Y_1 is orthonormal and thus, is a basis in V. Since $g(y)$ is a function on Y_1, there exists a unique representation as a linear combination of functions in S, namely,

$$g(y)=\sum_{f\in S}<g,f>f\,.$$

$$\because <g,f>=\sum_{y\in Y_1}g(y)f(y)P_{[1]}(y)=\sum_{y\in Y_1}f(y)P(Y=y)=E_P[f]\quad\forall f\text{ , and}$$

$$E_P[1]=1,\ E_P[r_j]=0,\ E_P[r_jr_k]=\rho_{jk},\cdots,\text{and } E_P[r_1\cdots r_m]=\rho_{12\cdots m}\,.$$

$$\therefore g(y)=(1+\sum_{j<k}\rho_{jk}r_jr_k+\sum_{j<k<l}\rho_{jkl}r_jr_kr_l+\cdots+\rho_{12\cdots m}r_1r_2\cdots r_m)\,.$$

(b.3) Marginal odds ratios: A compromise between <u>conditional odds ratios</u> (interpretations that depend on m), and <u>correlations</u> (seriously constrained by the means).

Let $\gamma_{jk}=OR(Y_j,Y_k)\triangleq\dfrac{P(Y_j=1,Y_k=1)P(Y_j=0,Y_k=0)}{P(Y_j=1,Y_k=0)P(Y_j=0,Y_k=1)}$.

$$\begin{aligned}
\eta_{jkl}&=\ln(OR(Y_j,Y_k\mid Y_l=1))-\ln(OR(Y_j,Y_k\mid Y_l=0)).\\
\eta_{jklm}&=\ln(OR(Y_j,Y_k\mid Y_l=1,Y_m=1))+\ln(OR(Y_j,Y_k\mid Y_l=0,Y_m=0))\\
&\quad-\ln(OR(Y_j,Y_k\mid Y_l=1,Y_m=0))-\ln(OR(Y_j,Y_k\mid Y_l=0,Y_m=1)).\\
&\vdots\\
\eta_{j_1j_2\cdots j_m}&=\sum_{y_{j_3}\cdots y_{j_m}=0,1}(-1)^{\left[\sum_{l=3}^{m}y_{j_l}+m-2\right]}\ln(OR(Y_{j_1},Y_{j_2}\mid Y_{j_3}=y_{j_3},\ldots,Y_{j_m}=y_{j_m})).
\end{aligned}$$

The joint p.d.f. of Y may be specified by μ, η_{jk}'s, η_{jkl}'s, …, and $\eta_{j_1j_2\cdots j_m}$.

Note: γ_{jk} can be modeled as is done for the marginal expectation.

Generalized estimating equations -

Let $E[Y_{ij}] = \mu_{ij}$ with $\text{logit}(\mu_{ij}) = X_{ij}^{\ T} \beta$.

Estimate (β, α) by solving the GEE:

$S_\beta(\beta, \alpha) = \sum_{i=1}^{n} (\frac{\partial \mu_i}{\partial \beta})^T (Var(Y_i))^{-1}(Y_i - \mu_i) = 0$, where $Var(Y_i) \triangleq Var(Y_i; \beta, \alpha)$ and

$S_\alpha(\beta, \alpha) = \sum_{i=1}^{n} (\frac{\partial \eta_i}{\partial \alpha})^T \mathrm{H}_i^{-1}(W_i - \eta_i) = 0$, where $W_i = (r_{i1}r_{i2}, r_{i1}r_{i3}, \ldots, r_{im_i-1}r_{im_i}, r_{i1}^2, \ldots, r_{im_i}^2)$,

$\eta_i = E[W_i]$, and $\mathrm{H}_i = Diag\{Var(r_{i1}r_{i2}), \cdots, Var(r_{im_i-1}r_{im_i}), Var(r_{i1}^2), \cdots, Var(r_{im_i}^2)\}$,

Counted Responses -

$Y \sim Poisson(\mu)$, $\mu = E[Y] = V[Y]$.

<u>Over-dispersion model</u> or <u>under-dispersion model</u>

$V[Y] > E[Y]$ $\qquad$ $V[Y] < E[Y]$

Examples: (Random-effects model)

$$\begin{matrix} Y_{ij} \big| \mu_i \overset{iid}{\sim} Poisson(\mu_i) \\ \mu_i \overset{iid}{\sim} Gamma(\mu, \phi\mu^2) \end{matrix} \Rightarrow \begin{matrix} E[Y_{ij}] = \mu \\ V[Y_{ij}] = \mu + \phi\mu^2,\ \phi > 0 \end{matrix} .$$

Log-linear model:

Common unit time: $\ln(E[Y_{ij}]) = X_{ij}^{\ T} \beta (\triangleq \lambda_{ij})$, i.e., $E[Y_{ij}] = e^{X_{ij}^{\ T} \beta}$.

Different time units: $\ln(E[Y_{ij}]) = \ln(t_{ij}) + X_{ij}^{\ T} \beta$, where $\ln(t_{ij})$ is the offset.

$V(Y_{ij}) = \phi_{ij} E[Y_{ij}]$: over-dispersion model when $\phi_{ij} > 1$. A regression model is needed

for ϕ_{ij} , i.e., $\phi_{ij} = \phi(\alpha_1)$.

Generalized estimating equation approach:

$S_\beta(\beta, \alpha) = \sum_{i=1}^{n} (\frac{\partial \mu_i}{\partial \beta})^T (V(Y; \alpha))^{-1}(Y_i - \mu_i) = 0$, where $\alpha = (\alpha_1, \alpha_2)$ with α_1 ~~being~~ related

to dispersion parameters and $S_\alpha(\beta, \alpha) = \sum_{i=1}^{n} (\frac{\partial \eta_i}{\partial \alpha})^T (W_i - \eta_i) = 0$.