Likelihood-Based Methods for Repeated Binary Data

Outline

- Joint multinomial distribution
- Log-linear model
- A hybrid model
- Bahadur model
- Modeling marginal odds ratios

Joint Multinomial Distribution

- An n-vector of binary variables $\boldsymbol{Y}$ has an *exact* joint multinomial distribution with 2^n points in its sample space. Let the π's be the probability of each point in the space and $\boldsymbol{\pi}$ be the vector of all 2^n π's.

- In the most general case, the multinomial distribution has $2^n - 1$ number of parameters.

- A subset of the n-vector $\boldsymbol{Y}$, say $\boldsymbol{Y}_s$, also has a multinomial distribution. The parameters of $\Pr(\boldsymbol{Y}_s)$ are sums of the parameters of $\Pr(\boldsymbol{Y})$.

- The variances are functions of the means.

- To relate covariates to the means $\boldsymbol{\mu}$, a nonlinear link function is typically used (logit, probit).

Issues with Modeling Repeated Binary Data

- Parsimony: constrains higher-order associations to be zero.
- Flexibility: allows dependence on covariates.
- Interpretability: e.g., odds ratio is more natural than correlation.

The Log-Linear Model

- Log-linear models (Bishop et al, 1975) have been popular in studying multiple correlated categorical (binary) variables.

- The saturated log-linear model is:

$$\log \Pr(\boldsymbol{Y} = \boldsymbol{y}) = c(\boldsymbol{\theta}) + \sum_{j=1}^{n} \theta_j y_j + \sum_{j_1<j_2} \theta_{j_1,j_2} y_{j1} y_{j2} + \cdots + \theta_{1,\ldots,n} y_1 \cdots y_n, \quad (1)$$

  where $c(\boldsymbol{\theta})$ is a normalizing constant.

- $\boldsymbol{\theta}$ is a $(2^n - 1)$-vector of canonical parameters:

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n, \theta_{12}, \ldots, \theta_{n-1,n}, \ldots, \theta_{1,\ldots,n})^T .$$

- $\boldsymbol{\theta}$ can be viewed as a log-linear transformation of the multinomial cell probabilities $\boldsymbol{\pi}$ (an 2^n vector),

$$\boldsymbol{\theta} = C_1^T \log \boldsymbol{\pi},$$

where C_1 is an $2^n \times (2^n - 1)$ matrix.

- The elements of $\boldsymbol{\theta}$ can be partitioned as:

$$\begin{array}{ccc}
\text{main effects} & \theta_1, \ldots, \theta_n & n \\
\text{2-way effects} & \theta_{12}, \theta_{13}, \ldots, \theta_{n-1,n} & \binom{n}{2} \\
\text{3-way effects} & \theta_{123}, \theta_{124}, \ldots, \theta_{n-2,n-1,n} & \binom{n}{3} \\
\vdots & \vdots & \vdots \\
n\text{-way effect} & \theta_{1,\ldots,n} & 1
\end{array}$$

- When $n = 1$, the distribution in equation (1) degenerates to binomial distribution.

Interpretation

For $n = 3$,

- The interpretation of θ_1 is:

$$\theta_1 = \log \frac{\pi_{100}}{\pi_{000}},$$

- The interpretation of θ_{12} is:

$$\theta_{12} = \log \frac{\pi_{110}\pi_{000}}{\pi_{100}\pi_{010}},$$

- The interpretation of θ_{123} is:

$$\begin{aligned}
\theta_{123} &= \log\left\{\frac{\pi_{111}\pi_{001}}{\pi_{101}\pi_{011}} \div \frac{\pi_{110}\pi_{000}}{\pi_{100}\pi_{110}}\right\}. \\
&= \log\left\{\frac{\Pr(Y_1 = 1, Y_2 = 1 \,|\, Y_3 = 1)\Pr(Y_1 = 0, Y_2 = 0 \,|\, Y_3 = 1)}{\Pr(Y_1 = 1, Y_2 = 0 \,|\, Y_3 = 1)\Pr(Y_1 = 0, Y_2 = 1 \,|\, Y_3 = 1)}\right\} \\
&\quad - \log\left\{\frac{\Pr(Y_1 = 1, Y_2 = 1 \,|\, Y_3 = 0)\Pr(Y_1 = 0, Y_2 = 0 \,|\, Y_3 = 0)}{\Pr(Y_1 = 1, Y_2 = 0 \,|\, Y_3 = 0)\Pr(Y_1 = 0, Y_2 = 1 \,|\, Y_3 = 0)}\right\}.
\end{aligned}$$

So

$$\theta_{123} = \log(\mathrm{OR}(Y_1, Y_2 \,|\, Y_3 = 1)) - \log(\mathrm{OR}(Y_1, Y_2 \,|\, Y_3 = 0)).$$

- When $\theta_{123} = 0$, the 2nd order parameters can be directly interpreted as log of the *conditional* odds ratios. Take θ_{12} as an example,

$$\theta_{12} = \log \mathrm{OR}(Y_1, Y_2 \,|\, Y_3).$$

- Each θ is a linear combination of $\log \pi$.

- The θ's can be interpreted as log odds ratios and differences of log odds ratios and so on.

Pros and Cons of Log-Linear Models

- By setting higher order parameters to 0, we get reduced parsimonious models.
- It is easy to characterize and compute the MLEs for $\boldsymbol{\theta}$.
- $\boldsymbol{\theta}$ is not constrained by the means, i.e., the log odds ratios do not depend on the marginal means (variation independent).
- The log-linear model is not convenient to model the marginal means as a function of the covariates because the marginal means are not simple functions of $\boldsymbol{\theta}$.
- The interpretation of the canonical parameters depends on the number of responses. Hence this formulation is not suitable for unbalanced data.

A Hybrid Model by Fitzmaurice and Laird (1993)

- A compromise is to use the marginal means $\mu_j = \Pr(Y_j = 1) = \mathrm{E}(Y_j)$, and the second- and higher order canonical parameters.

- The joint distribution of $\boldsymbol{Y}$ is given by:

$$\Pr(\boldsymbol{Y} = \boldsymbol{y}) = c(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) \exp(\boldsymbol{y}^T \boldsymbol{\theta_1} + \boldsymbol{w}^T \boldsymbol{\theta_2}). \tag{2}$$

where $\boldsymbol{w} = (y_1 y_2, y_1 y_3, \ldots, y_{n-1} y_n, y_1 y_2 y_3, \ldots, y_1 y_2 \ldots y_n)$, $\boldsymbol{\theta_1} = (\theta_1, \ldots, \theta_n)$, and $\boldsymbol{\theta_2} = (\theta_{12}, \theta_{13}, \ldots, \theta_{n-1,n}, \ldots, \theta_{1,\ldots,n})$ are the second- and higher-order canonical parameters.

- The parameters of interest are $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n) = \boldsymbol{\mu}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$, so we can make a transformation from the canonical parameters, $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ to $(\boldsymbol{\mu}, \boldsymbol{\theta}_2)$.

- Assume that the marginal means satisfy

$$\text{logit}(\boldsymbol{\mu}) = X_i^T \boldsymbol{\beta}$$

and the higher order parameters satisfy

$$\boldsymbol{\theta}_2 = Z_i \boldsymbol{\alpha}$$

- Interestingly, the score function for $\boldsymbol{\beta}$ under this parameterization takes the GEE form

$$\left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}}\right)^T \text{Var}(\boldsymbol{Y})^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}) = 0,$$

where $\text{Var}(\boldsymbol{Y})$ is a function of both $\boldsymbol{\beta}$ and $\boldsymbol{\theta}_2$. The solution of GEE is the maximum likelihood estimate under the log-linear model when the covariance assumption, $\text{Var}(\boldsymbol{Y_i})$, is correct for all i.

- We can get a consistent estimate of $\boldsymbol{\beta}$ even if the model for $\boldsymbol{\theta}_2$ is wrong. When the $\boldsymbol{\alpha}$ part is misspecified, the "robust" variance should be used.

- If we set the third- and higher effects to zero, we get a *quadratic exponential family* distribution.

- $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are orthogonal (i.e. $\mathrm{Var}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})$, estimated by the Fisher information matrix, is a diagonal matrix)

- Conditional odds ratios are not easily interpreted.

- The interpretation of $\boldsymbol{\theta}_2$ depends on the number of responses in a cluster. Hence the formulation in (2) is most useful when the data are balanced.

Bahadur Model

- The Bahadur model (Bahadur 1961) uses marginal means, pairwise correlations and higher-order moments to parameterize the multinomial distribution.

- Let $\mu_j = \mathrm{E}(Y_j), j = 1, \ldots, n,$

$$
\begin{aligned}
R_j &= \frac{Y_j - \mu_j}{[\mu_j(1-\mu_j)]^{1/2}}, \\
\rho_{jk} &= \mathrm{Cor}(Y_j, Y_k) = \mathrm{E}(R_j R_k), \\
\rho_{jkl} &= \mathrm{E}(R_j R_k R_l), \\
\cdots \\
\rho_{1,\ldots,n} &= \mathrm{E}(R_1 R_2 \cdots R_n).
\end{aligned}
$$

Then the probability distribution of $\Pr(\boldsymbol{Y} = \boldsymbol{y})$ is:

$$\begin{aligned}\Pr(\boldsymbol{Y} = \boldsymbol{y}) &= \prod_{j=1}^{n} \mu_j^{y_j}(1-\mu_j)^{(1-y_j)} \\ &\times \left(1 + \sum_{j<k} \rho_{jk} r_j r_k + \sum_{j<k<l} \rho_{jkl} r_j r_k r_l + \cdots + \rho_{1,\ldots,n} r_1 r_2 \cdots r_n\right)\end{aligned}$$

- Uses marginal means (parameters of interest) and correlations (familiar from continuous variables).
- The correlations are constrained by the marginal means (not variation independent) in a complicated manner. If we assume that the marginal means depend on $\boldsymbol{x}$, it may not be correct to assume that the ρ's are independent of $\boldsymbol{x}$.

Modeling Marginal Odds Ratios

- Motivation.
  - The conditional odds ratios are not constrained by the means but have interpretations that depend on n.
  - The correlations are seriously constrained by the means.
  - A compromise is to parameterize the likelihood in terms of **marginal odds ratios**,

$$\gamma_{jk} = \frac{\Pr(Y_j = 1, Y_k = 1)\Pr(Y_j = 0, Y_k = 0)}{\Pr(Y_j = 1, Y_k = 0)\Pr(Y_j = 0, Y_k = 1)},$$

which is unconstrained on the log scale and whose interpretations are independent of n:

- Now, the likelihood of $\boldsymbol{Y}$ can be specified in terms of $\boldsymbol{\mu}$, $\boldsymbol{\gamma} = c(\gamma_{12}, \ldots, \gamma_{n-1,n})$, and **contrasts of odds ratios**, namely

$$\begin{aligned}
\zeta_{123} &= \log(\mathrm{OR}(Y_1, Y_2 \mid Y_3 = 1)) - \log(\mathrm{OR}(Y_1, Y_2 \mid Y_3 = 0)) \\
\zeta_{1234} &= \log(\mathrm{OR}(Y_1, Y_2 \mid Y_3 = 1, Y_4 = 1)) - \log(\mathrm{OR}(Y_1, Y_2 \mid Y_3 = 1, Y_4 = 0)) \\
&\quad - \log(\mathrm{OR}(Y_1, Y_2 \mid Y_3 = 0, Y_4 = 1)) + \log(\mathrm{OR}(Y_1, Y_2 \mid Y_3 = 0, Y_4 = 0))
\end{aligned}$$

...

The general form is $\zeta_{j_1,\ldots,j_n} = \sum_{y_{j_3},\ldots,y_{j_n} \in (0,1)} (-1)^{b(\boldsymbol{y})} \log(\mathrm{OR}(Y_{j_1}, Y_{j_2} \mid Y_{j_3}, \ldots, Y_{j_n}))$, where $b(\boldsymbol{y}) = \sum_{l=3}^{n} y_{j_l} + n - 2$.

- The evaluation of the likelihood function in terms of those odds ratios and their contrasts is not simple except in cases with a small number of observations per person (Liang et al., 1992).

- When the subject i has n_i observations, there are $n_i(n_i - 1)/2$ pairwise odds ratios per subject, and this number becomes substantial when n_i increases.

- Possible models for 2nd order parameters, $\boldsymbol{\gamma}$ include the exchangeable model

$$\gamma_{ijk} = \gamma$$

or alternatively

$$\log \gamma_{ijk} = \alpha_0 + \alpha_1 |t_{ij} - t_{ik}|^{-1}.$$

In general, we write $\gamma_{ijk} = \gamma(\boldsymbol{\alpha})$.

- We often don't have sensible models for third- and higher-order parameters regardless of which formulation we adopt.

- Even when a probability model is fully specified, the likelihood can be complicated to evaluate except with small and constant n_i.

- For these reasons, people use the GEE approach for mutivariate binary data.

Further Reading

- Chapter 8.2 of DHLZ.

References

- Bahadur RR (1961). A representation of the joint distribution of responses to n dichotomous items. In *Studies on item analysis and prediction* (ed. H. Solomon), pp.158-68. Stanford Mathematical Studies in the Social Sciences VI, Stanford University Press, Stanford, California.

- Bishop YMM, Finberg SE, and Holland PW (1975) Discrete multivariate analysis: theory and practice. MIT Press, Cambridge, MA.

- Fitzmaurice GM and Laird NM (1993) A likelihood-based method for analysing longitudinal binary responses. *Biometrika* **80**:141151.

- Laird N (2004) Analysis of longitudinal and cluster-correlated data, vol. 8 of NSF-CBMS Regional Conference Series in Probability and Statistics. Institute of Mathematical Statistics.

- Liang K-Y, Zeger SL, and Qaqish B (1992). Multivariate regression analyses for categorical data (with discussion). *JRSSB* **54**:3-40.