Likelihood-based method for longitudinal binary data

Jan Irvahn

April 5, 2012

The Paper

- A likelihood-based method for analysing longitudinal binary responses
- Authors: Fitzmaurice and Laird
- Published in Biometrika in 1993

The Data

- Longitudinal Binary Responses
- $Y_i = (Y_{i1,\ldots,Y_{iT}})$
- i ranges from 1 to n
- n is the total number of clusters/individuals
- each individual has a $J \times 1$ covariate vector, x_{it}, at time t

The Goal

- maximum likelihood estimates of the marginal mean parameters
- find $\hat{\beta}$ such that $E[Y_{it}|X_{it}] = g(X_{it}^T \beta)$
- find Ψ_i and Ω_i such that $f(y_i, \Psi_i, \Omega_i)$ is the joint pdf of Y_i

GEE vs Likelihood

GEE

$$\sum_{i=1}^{n} \frac{\partial g(x_i\beta)}{\partial \beta^T} V_i^{-1}(Y_i - g(x_i\beta)) = 0$$

Likelihood

$$\sum_{i=1}^{n} X_i^T \Delta_i V_{i11}^{-1}(y_i - \mu_i) = 0$$

$$\sum_{i=1}^{n} Z_i^T \left(w_i - \nu_i - V_{i21} V_{i11}^{-1}(y_i - \mu_i)\right) = 0$$

Likelihood Approach

The joint distribution of Y_u has the following distribution.

$$
\begin{aligned}
f(y_i, \Psi_i, \Omega_i) &= \exp(\Psi_i^T y_i + \Omega_i^T w_i - A(\Psi_i, \Omega_i)) \\
W_i &= (Y_{i1} Y_{i2}, \ldots, Y_{iT-1} Y_{iT}, \ldots, Y_{i1} Y_{i2} \ldots Y_{iT})^T \\
&\quad \text{two and higher-way cross products of } Y_i \\
\Omega_i &= (\omega_{i12}, \ldots, \omega_{iT-1T}, \ldots, \omega_{i12\ldots T}) \\
A(\Psi_i, \Omega_i) &\text{ is a normalizing constant} \\
&\quad \exp(A(\Psi_i, \Omega_i)) = \sum \exp(\Psi_i^T y_i + \Omega_i^T w_i) \\
&\quad \text{(the sum is over all } 2^T \text{ possible values of } Y_i)
\end{aligned}
$$

Parameter Interpretation

Ψ_{ir} is a conditional probability.

$$\Psi_{ir} = \text{logit}(\Pr(Y_{ir} = 1 | Y_{is} = 0, \forall s \neq r))$$

ω_{irs} can be interpreted in terms of conditional log odds-ratios.

$$\exp(\omega_{irs}) = \frac{\left(\frac{\Pr(Y_{ir}=1, Y_{is}=1 | Y_{it}=0, \forall t \neq r,s)}{\Pr(Y_{ir}=0, Y_{is}=1 | Y_{it}=0, \forall t \neq r,s)}\right)}{\left(\frac{\Pr(Y_{ir}=1, Y_{is}=0 | Y_{it}=0, \forall t \neq r,s)}{\Pr(Y_{ir}=0, Y_{is}=0 | Y_{it}=0, \forall t \neq r,s)}\right)}$$

Higher order ω terms are more complicated.

Likelihood Equations

$$l_i = \Psi_i^T y_i + \Omega_i^T w_i - A(\Psi_i, \Omega_i)$$

(a 1-1 transformation from (Ψ_i, Ω_i) to (μ_i, Λ_i) is used)

$$0 = \begin{pmatrix} \partial l_i / \partial \beta \\ \partial l_i / \partial \alpha \end{pmatrix}$$

$$= \begin{pmatrix} \partial \mu_i / \partial \beta & \partial \mu_i / \partial \alpha \\ \partial \nu_i / \partial \beta & \partial \nu_i / \partial \alpha \end{pmatrix}^T \begin{pmatrix} V_{i11} & V_{i12} \\ V_{i21} & V_{i22} \end{pmatrix}^{-1} \begin{pmatrix} y_i - \mu_i \\ w_i - \nu_i \end{pmatrix}$$

$$\begin{pmatrix} V_{i11} & V_{i12} \\ V_{i21} & V_{i22} \end{pmatrix} = \begin{pmatrix} \mathrm{cov}(Y_i) & \mathrm{cov}(Y_i, W_i) \\ \mathrm{cov}(W_i, Y_i) & \mathrm{cov}(W_i) \end{pmatrix}$$

$$\nu_i = E(W_i)$$

Likelihood Equations

$$\sum_{i=1}^{n} X_i^T \Delta_i V_{i11}^{-1}(y_i - \mu_i) = 0$$

$$\sum_{i=1}^{n} Z_i^T \left(w_i - \nu_i - V_{i21} V_{i11}^{-1}(y_i - \mu_i)\right) = 0$$

$\Omega_i = Z_i \alpha$

Z_i is a K $\times$ Q design matrix

α is a Q $\times$ 1 parameter vector

$\Delta_i = \mathrm{diag}(\mathrm{var}(Y_{it}))$ is a T $\times$ T diagonal matrix

Prior Work

This method is similar to a pseudo-maximum likelihood approach taken by Zhao and Prentice in 1990.

That work was, in turn, an extension of work done on multivariate binary data by Cox in 1972.

To solve the Likelihood equations Fitzmaurice and Laird use an iterative proportional fitting procedure developed by Deming and Stephan in 1940.

Prior Work

Zhao and Prentice parameterized their model in terms of correlations. Fitzmaurice and Laird parameterized their model in terms of conditional log-odds-ratios. This allows the incorporation of higher order associations.

"One important advantage of this parameterization is that the maximum likelihood estimates of the marginal mean parameters are robust to misspecification of the time dependence."

example:

$$\text{logit}(\mu) = \beta_0 + \beta_1 \text{Treatment} + \beta_2 \text{Sequence}$$

parameterized either by

(i) ρ_{12}, the pairwise correlation, or

(ii) ω_{12}, the log odds-ratio

Graphical Comparisons

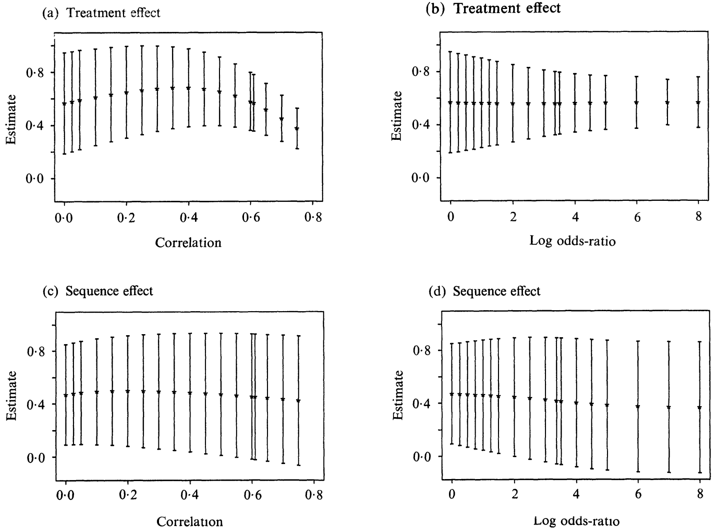

Fig. 1. Maximum likelihood estimates and model standard error bands for fixed time dependence parameter.

Graphical Comparisons

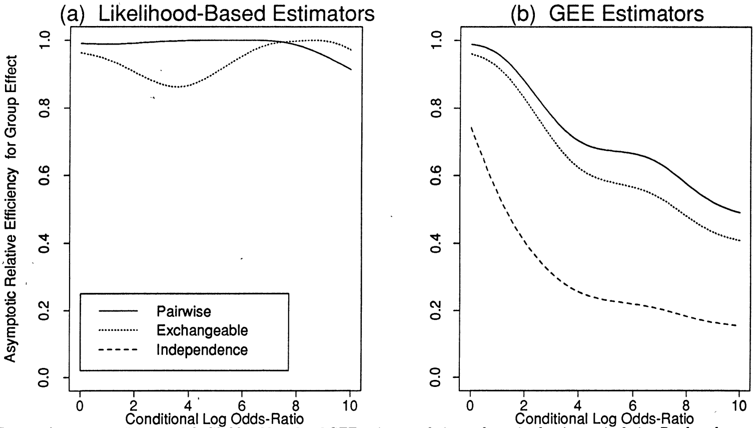

FIG. 1. *Asymptotic efficiency of the likelihood-based and GEE estimatorsrelative to the optimal estimator in design B, when the true underlying joint distribution has a log-linear representation.*

from Fitzmaurice, Laird, and Rotnitsky 1993

Graphical Comparisons

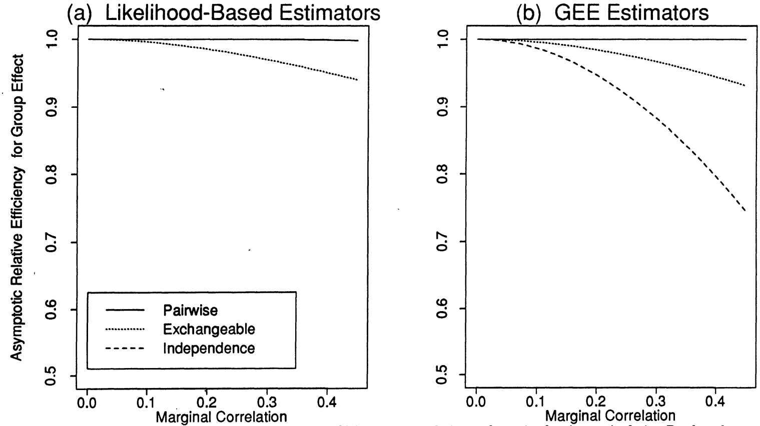

FIG. 2. *Asymptotic efficiency of the likelihood-based and GEE estimatorsrelative to the optimal estimator in design B, when the true underlying joint distribution has Bahadur's representation.*

from Fitzmaurice, Laird, and Rotnitsky 1993

Why this method is good (and bad)

- β is unbiased even when the time dependence is misspecified
- parameter space is orthogonal, the asymptotic variance of β is the same whether α is known or estimated
- the conditional association parameters are not constrained
- interpreting the parameters is difficult
- the method is not appropriate for clusters of different sizes