Likelihood-based method for longitudinal binary data

Jan Irvahn

May 22, 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

Example data

Six Cities data set: child's wheeze status

No maternal smoking

			Age 10	
Age 7	Age 8	Age 9	No	Yes
No	No	No	237	10
		Yes	15	4
	Yes	No	16	2
		Yes	7	3
Yes	No	No	24	3
		Yes	3	2
	Yes	No	6	2
		Yes	5	11

Maternal smoking

			Age 10	
Age 7	Age 8	Age 9	No	Yes
No	No	No	118	6
		Yes	8	2
	Yes	No	11	1
		Yes	6	4
Yes	No	No	7	3
		Yes	3	1
	Yes	No	4	2
		Yes	4	7

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 $\hat{\alpha}$, (additional likelihood parameters)

Framework

- Derive estimating equations.
- Use Fisher scoring algorithm to find MLEs
- Figure out iterative proportional fitting to implement Fisher scoring algorithm

Likelihood

$$f(y_i, \Psi_i, \Omega_i) = \exp\left(\Psi_i^T y_i + \Omega_i w_i - A(\Psi_i, \Omega_i)\right)$$

$$W_i = (Y_{i1} Y_{i2}, \ldots, Y_{iT-1} Y_{iT}, \ldots, Y_{i1} Y_{i2} \ldots Y_{iT})^T$$

two and higher-way cross products of Y

$$\Omega_i = (\omega_{i12}, \ldots, \omega_{iT-1T}, \ldots, \omega_{i12\ldots T})$$

The length of W and Ω is $K = 2^T - (T+1)$.

$A(\Psi_i, \Omega_i)$ is a normalizing constant

$$\exp(A(\Psi_i, \Omega_i)) = \sum \exp(\Psi_i^T y_i + \Omega_i^T w_i)$$

(the sum is over all 2^T possible values of Y_i)

Log Likelihood

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

$$\begin{pmatrix} \partial l_i / \partial \Psi_i \\ \partial l_i / \partial \Omega_i \end{pmatrix} = \begin{pmatrix} y_i - \mu_i \\ w_i - \nu_i \end{pmatrix}$$

μ_i is the expected value of y_i.

ν_i is the expected value of w_i.

These are nice score equations but we are interested in the parameters (β, α), not (Ψ_i, Ω_i).

Important new parameters, μ and Γ

$$\mu_{it} = \Pr(Y_{it} = 1) = \text{expit}(x_i^T \beta)$$
$$\Gamma_i = \Omega_i = Z_i \alpha$$

$$\mu_{it} = \frac{\sum_{Y_i | Y_{it}=1} \exp(\Psi_i^T y_i + \Omega_i w_i)}{\sum_{Y_i} \exp(\Psi_i^T y_i + \Omega_i w_i)}$$
$$\exp(\omega_{irs}) = \frac{\left(\frac{\Pr(Y_{ir}=1, Y_{is}=1 | Y_{it}=0, t \neq r,s)}{\Pr(Y_{ir}=0, Y_{is}=1 | Y_{it}=0, t \neq r,s)}\right)}{\left(\frac{\Pr(Y_{ir}=1, Y_{is}=0 | Y_{it}=0, t \neq r,s)}{\Pr(Y_{ir}=0, Y_{is}=0 | Y_{it}=0, t \neq r,s)}\right)}$$

two-way interaction terms: conditional log odds-ratios

higher-way interaction terms: more complicated

Log Likelihood derivatives, again

This time using the chain rule.

$$
\begin{aligned}
\begin{pmatrix} \partial l_i/\partial \Psi_i \\ \partial l_i/\partial \Omega_i \end{pmatrix} &= \begin{pmatrix} \partial \mu_i/\partial \Psi_i & \partial \Gamma_i/\partial \Psi_i \\ \partial \mu_i/\partial \Omega_i & \partial \Gamma_i/\partial \Omega_i \end{pmatrix} \begin{pmatrix} \partial l_i/\partial \mu_i \\ \partial l_i/\partial \Gamma_i \end{pmatrix} \\
&= \begin{pmatrix} \mathrm{cov}(Y_i) & 0 \\ \mathrm{cov}(Y_i, W_i) & I \end{pmatrix} \begin{pmatrix} \partial l_i/\partial \mu_i \\ \partial l_i/\partial \Gamma_i \end{pmatrix} \\
&= \begin{pmatrix} V_{i11} & 0 \\ V_{i21} & I \end{pmatrix} \begin{pmatrix} \partial l_i/\partial \mu_i \\ \partial l_i/\partial \Gamma_i \end{pmatrix}
\end{aligned}
$$

Score

$$\begin{pmatrix} \partial l_i/\partial \mu_i \\ \partial_i l/\partial \Gamma_i \end{pmatrix} = \begin{pmatrix} V_{i11}^{-1} & 0 \\ -V_{i21}V_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} y_i - \mu_i \\ w_i - \nu_i \end{pmatrix}$$

The score we want uses the parameters α and β.

$$\begin{aligned} \begin{pmatrix} \partial l_i/\partial \beta \\ \partial l_i/\partial \alpha \end{pmatrix} &= \begin{pmatrix} \partial \mu_i/\partial \beta & \partial \Gamma_i/\partial \beta \\ \partial \mu_i/\partial \alpha & \partial \Gamma_i/\partial \alpha \end{pmatrix} \begin{pmatrix} \partial l_i/\partial \mu_i \\ \partial l_i/\partial \Gamma_i \end{pmatrix} \\ &= \begin{pmatrix} X_i^T \Delta_i & 0 \\ 0 & Z_i^T \end{pmatrix} \begin{pmatrix} V_{i11}^{-1} & 0 \\ -V_{i21}V_{i11}^{-1} & I \end{pmatrix} \begin{pmatrix} y_i - \mu_i \\ w_i - \nu_i \end{pmatrix} \end{aligned}$$

$\Delta_i = \text{diag}(\text{var}(Y_{it}))$

Score and Information

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

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

$$I \approx \sum_{i=1}^{n} E\left[\begin{pmatrix} \partial l_i/\partial\beta \\ \partial l_i/\partial\alpha \end{pmatrix}\begin{pmatrix} \partial l_i/\partial\beta \\ \partial l_i/\partial\alpha \end{pmatrix}^T\right]$$

$$= \begin{pmatrix} \sum_{i=1}^{n} X_i^T \Delta_i V_{i11}^{-1} \Delta_i X_i & 0 \\ 0 & \sum_{i=1}^{n} Z_i^T (V_{i22} - V_{i21} V_{i11}^{-1} V_{i21}^T) Z_i \end{pmatrix}$$

$$V_{i22} = \text{cov}(W_i)$$

Fisher scoring algorithm

$$\hat{\beta}^{(J+1)} = \hat{\beta}^{(J)} + \left(\sum_{i=1}^{n} X_i^T \Delta_i V_{i11}^{-1} \Delta_i X_i\right)^{-1} \Big(X_i^T \Delta_i V_{i11}^{-1}(y_i - \mu_i)\Big)$$

$$\hat{\alpha}^{(J+1)} = \hat{\alpha}^{(J)} + \left(\sum_{i=1}^{n} Z_i^T (V_{i22} - V_{i21} V_{i11}^{-1} V_{i21}^T) Z_i\right)^{-1} \times \left(\sum_{i=1}^{n} Z_i^T \Big(w_i - \nu_i - V_{i21} V_{i11}^{-1}(y_i - \mu_i)\Big)\right)$$

How to implement Fisher scoring algorithm

- given (β, α) we can calculate (μ_i, Ω_i)
    - $\mu_i = \text{expit}(X_i^T \beta)$
    - $\Omega_i = Z_i \alpha$
- given ν_i we can calculate V_{i11}, V_{i21}, and V_{i22}
    - $V_{i11} = \text{cov}(Y_i)$
    - $V_{i21} = \text{cov}(W_i, Y_i)$
    - $V_{i22} = \text{cov}(W_i)$
- challenge: given (μ_i, Ω_i), calculate ν_i

Iterative proportional fitting

- make m, an array with 2^T cells containing probabilities
- set T+1 cells arbitrarily
- use Ω_i's $2^T - (T+1)$ constraints to complete m
- normalize m using another constraint

- split m into two sets of cells, cells where the first dimension is '0' and cells where the first dimension is '1'
- normalize the '1' cells so that they sum to μ_{i1}
- normalize the '0' cells so that they sum to $1 - \mu_{i1}$
- repeat for the remaining $T - 1$ dimensions
- repeat normalization to μ_i until convergence

Example Array

		Y_{i2}		
		0	1	
Y_{i1}	0	m_{00}	m_{01}	$1 - \mu_{i1}$
	1	m_{10}	$e^{\omega_{12}} m_{10} m_{01} / m_{00}$	μ_{i1}
		$1 - \mu_{i2}$	μ_{i2}	1