Likelihood-based method for longitudinal binary data

Jan Irvahn

April 26, 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

Overall 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

Mini-Goal

- Derive the estimating equations.
- Write down a likelihood
- take a log, take some derivatives, set it equal to zero

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

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

(most of these symbols have not been defined)

Likelihood

Note: I will suppress the index i so $f(y_i, \Psi_i, \Omega_i) = f(y, \Psi, \Omega)$

$$
\begin{aligned}
f(y, \Psi, \Omega) &= \exp\left(\Psi^T y + \Omega w - A(\Psi, \Omega)\right) \\
W &= (Y_1 Y_2, \ldots, Y_{T-1} Y_T, \ldots, Y_1 Y_2 \ldots Y_T)^T \\
&\quad \text{two and higher-way cross products of } Y \\
\Omega &= (\omega_{12}, \ldots, \omega_{T-1\,T}, \ldots, \omega_{12\ldots T})
\end{aligned}
$$

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

$A(\Psi, \Omega)$ is a normalizing constant

$$
\exp(A(\Psi, \Omega)) = \sum \exp(\Psi^T y + \Omega^T w)
$$

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

Example

Suppose we only measure each person at two times points, so $T = 2$. Also, suppose we only have a single covariate at each time point, so $X = (x_1, x_2)$.

$$f(y, \psi_1, \psi_2, \omega_{12}) = \Delta^{-1} \exp(\psi_1 y_1 + \psi_2 y_2 + \omega_{12} y_1 y_2)$$
$$\Delta = 1 + e^{\psi_1} + e^{\psi_2} + e^{\psi_1 + \psi_2 + \omega_{12}}$$

Important new parameters

- μ_t is the expected value of y_t
  - ν is the expected value of w

$$
\begin{aligned}
\mu_t &= \mu_t(\beta) = \Pr(Y_t = 1|x_t, \beta) = \text{expit}(x_t^T \beta) \\
\mu_1 &= \Delta^{-1}(e^{\psi_1} + e^{\psi_1+\psi_2+\omega_{12}}) = \text{expit}(x_1\beta) \\
\nu &= \Delta^{-1} e^{\psi_1+\psi_2+\omega_{12}}
\end{aligned}
$$

- α is a $Q \times 1$ vector parameterizing Ω (a $K \times 1$ vector)

$$\Omega = Z\alpha$$

Z is some $K \times Q$ design matrix. In our example we will let $Z = 1$ so $\omega_{12} = \alpha$.

Log Likelihood

$$l = \Psi^T y + \Omega^T w - \log(\Delta)$$
$$l = \psi_1 y_1 + \psi_2 y_2 + \omega_{12} y_1 y_2 - \log(1 + e^{\psi_1} + e^{\psi_2} + e^{\psi_1+\psi_2+\omega_{12}})$$

A transformation:

$$(\Psi, \Omega) \to (\mu, \Gamma)$$

somewhat underwhelming because $\Gamma = \Omega$

Score Equations

$$\begin{pmatrix} \partial l/\partial \Psi \\ \partial l/\partial \Omega \end{pmatrix} = \begin{pmatrix} y - \mu \\ w - \nu \end{pmatrix}$$

$$\begin{aligned} \begin{pmatrix} \partial l/\partial \Psi \\ \partial l/\partial \Omega \end{pmatrix} &= \begin{pmatrix} \partial \mu/\partial \Psi & \partial \Gamma/\partial \Psi \\ \partial \mu/\partial \Omega & \partial \Gamma/\partial \Omega \end{pmatrix} \begin{pmatrix} \partial l/\partial \mu \\ \partial l/\partial \Gamma \end{pmatrix} \\ &= \begin{pmatrix} V_{11} & 0 \\ V_{21} & I \end{pmatrix} \begin{pmatrix} \partial l/\partial \mu \\ \partial l/\partial \Gamma \end{pmatrix} \end{aligned}$$

V_{11} is $\mathrm{cov}(Y)$ and V_{21} is $\mathrm{cov}(Y, W)$

Example Derivatives

$$E(Y_1) = \mu_1 = \Delta^{-1}(e^{\psi_1} + e^{\psi_1+\psi_2+\omega_{12}})$$
$$\Delta = 1 + e^{\psi_1} + e^{\psi_2} + e^{\psi_1+\psi_2+\omega_{12}}$$

$$\partial\mu_1/\partial\psi_1 = \mu_1 - \mu_1^2 = \mathrm{cov}(Y_1)$$
$$\partial\mu_1/\partial\psi_2 = \nu - \mu_1\mu_2 = \mathrm{cov}(Y_1, Y_2)$$
$$\partial\mu_1/\partial\omega_{12} = \nu - \mu_1\nu = E(Y_1^2 Y_2) - E(Y_1)E(Y_1 Y_2) = \mathrm{cov}(Y_1, Y_1 Y_2)$$

Score Continued

$$\begin{pmatrix} \partial l/\partial \mu \\ \partial l/\partial \Gamma \end{pmatrix} = \begin{pmatrix} V_{11}^{-1} & 0 \\ -V_{21}V_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} y-\mu \\ w-\nu \end{pmatrix}$$

The score we want uses the parameters α and β.

$$\begin{aligned} \begin{pmatrix} \partial l/\partial \beta \\ \partial l/\partial \alpha \end{pmatrix} &= \begin{pmatrix} \partial \mu/\partial \beta & \partial \Gamma/\partial \beta \\ \partial \mu/\partial \alpha & \partial \Gamma/\partial \alpha \end{pmatrix} \begin{pmatrix} \partial l/\partial \mu \\ \partial l/\partial \Gamma \end{pmatrix} \\ &= \begin{pmatrix} X^T D & 0 \\ 0 & Z^T \end{pmatrix} \begin{pmatrix} \partial l/\partial \mu \\ \partial l/\partial \Gamma \end{pmatrix} \end{aligned}$$

$\mu = \text{expit}(X^T\beta)$, $\Gamma = \Omega = Z^T\alpha$, and $D =$diag(var(Y))

More Score

$$\begin{pmatrix} \partial l/\partial\beta \\ \partial l/\partial\alpha \end{pmatrix} = \begin{pmatrix} X^T D & 0 \\ 0 & Z^T \end{pmatrix} \begin{pmatrix} V_{11}^{-1} & 0 \\ -V_{21}V_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} y-\mu \\ w-\nu \end{pmatrix}$$

and seven years ago...

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

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

Next Step

Solve the estimating equations using the Fisher scoring algorithm and an iterative proportional fitting algorithm.