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

<table>
<thead>
<tr>
<th>No maternal smoking</th>
<th>Age 10</th>
<th>Maternal smoking</th>
<th>Age 10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age 10</td>
<td></td>
<td>Age 10</td>
</tr>
<tr>
<td></td>
<td>Age 10</td>
<td></td>
<td>Age 10</td>
</tr>
<tr>
<td>Age 7</td>
<td>Age 8</td>
<td>Age 9</td>
<td>No</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>No</td>
<td>237</td>
</tr>
<tr>
<td>Yes</td>
<td>15</td>
<td>4</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>16</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>7</td>
<td>3</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>24</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>3</td>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>6</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>5</td>
<td>11</td>
<td>Yes</td>
</tr>
</tbody>
</table>
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

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

\[ W_i = (Y_{i1} Y_{i2}, ..., Y_{iT-1} Y_{iT}, ..., Y_{i1} Y_{i2} ... Y_{iT})^T \]

two and higher-way cross products of \( Y \)

\[ \Omega_i = (\omega_{i12}, ..., \omega_{iT-1T}, ..., \omega_{i12...T}) \]

The length of \( W \) and \( \Omega \) 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

\[ l_i = \psi_i^T y_i + \Omega_i^T w_i - A(\psi_i, \Omega_i) \]

\[
\left( \begin{array}{c}
\frac{\partial l_i}{\partial \psi_i} \\
\frac{\partial l_i}{\partial \Omega_i}
\end{array} \right) = \left( \begin{array}{c}
y_i - \mu_i \\
w_i - \nu_i
\end{array} \right)
\]

\( \mu_i \) is the expected value of \( y_i \).

\( \nu_i \) is the expected value of \( w_i \).

These are nice score equations but we are interested in the parameters \((\beta, \alpha)\), not \((\psi_i, \Omega_i)\).
Important new parameters, $\mu$ and $\Gamma$

$$\mu_{it} = \Pr(Y_{it} = 1) = \expit(x_i^T \beta)$$

$$\Gamma_i = \Omega_i = Z_i \alpha$$

$$\mu_{it} = \frac{\sum_{Y_i | Y_{it}=1} \exp(\Psi_i^T y_i + \Omega_i w_i)}{\sum Y_i \exp(\Psi_i^T y_i + \Omega_i w_i)}$$

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

two-way interaction terms: conditional log odds-ratios

higher-way interaction terms: more complicated
Log Likelihood derivatives, again

This time using the chain rule.

\[
\begin{pmatrix}
  \frac{\partial l_i}{\partial \Psi_i} \\
  \frac{\partial l_i}{\partial \Omega_i}
\end{pmatrix}
= \begin{pmatrix}
  \frac{\partial \mu_i}{\partial \Psi_i} & \frac{\partial \Gamma_i}{\partial \Psi_i} \\
  \frac{\partial \mu_i}{\partial \Omega_i} & \frac{\partial \Gamma_i}{\partial \Omega_i}
\end{pmatrix}
\begin{pmatrix}
  \frac{\partial l_i}{\partial \mu_i} \\
  \frac{\partial l_i}{\partial \Gamma_i}
\end{pmatrix}
= \begin{pmatrix}
  \text{cov}(Y_i) & 0 \\
  \text{cov}(Y_i, W_i) & I
\end{pmatrix}
\begin{pmatrix}
  \frac{\partial l_i}{\partial \mu_i} \\
  \frac{\partial l_i}{\partial \Gamma_i}
\end{pmatrix}
= \begin{pmatrix}
  V_{i11} & 0 \\
  V_{i21} & I
\end{pmatrix}
\begin{pmatrix}
  \frac{\partial l_i}{\partial \mu_i} \\
  \frac{\partial l_i}{\partial \Gamma_i}
\end{pmatrix}
\]
The score we want uses the parameters $\alpha$ and $\beta$.

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

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

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

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

\[I \approx \sum_{i=1}^{n} E \left[ \left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \alpha} \right)^T \right] = \left( \sum_{i=1}^{n} X_i^T \Delta_i V_{i11}^{-1} \Delta_i X_i \right)^{-1} \left( \sum_{i=1}^{n} Z_i^T (V_{i22} - V_{i21} V_{i11}^{-1} V_{i21}^T) Z_i \right)
\]

\[V_{i22} = \text{cov}(W_i)\]
Fisher scoring algorithm

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

\[
\hat{\alpha}^{(J+1)} = \hat{\alpha}^{(J)} + \left( \sum_{i=1}^{n} Z_i^T (V_{i22} - V_{i21} V_{i11}^{-1} V_{i21}^T) Z_i \right)^{-1} \\
\times \left( \sum_{i=1}^{n} Z_i^T \left( w_i - \nu_i - V_{i21} V_{i11}^{-1} (y_i - \mu_i) \right) \right)
\]
How to implement Fisher scoring algorithm

- given \((\beta, \alpha)\) we can calculate \((\mu_i, \Omega_i)\)
  - \(\mu_i = \text{expit}(X_i^T \beta)\)
  - \(\Omega_i = Z_i \alpha\)
- given \(\nu_i\) we can calculate \(V_{i11}, V_{i21}, \text{and } V_{i22}\)
  - \(V_{i11} = \text{cov}(Y_i)\)
  - \(V_{i21} = \text{cov}(W_i, Y_i)\)
  - \(V_{i22} = \text{cov}(W_i)\)
- challenge: given \((\mu_i, \Omega_i)\), calculate \(\nu_i\)
Iterative proportional fitting

- make $m$, an array with $2^T$ cells containing probabilities
- set $T+1$ cells arbitrarily
- use $\Omega_i$'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 $\mu_{i1}$
- normalize the ‘0’ cells so that they sum to $1 - \mu_{i1}$
- repeat for the remaining $T - 1$ dimensions
- repeat normalization to $\mu_i$ until convergence
### Example Array

<table>
<thead>
<tr>
<th>$Y_{i1}$</th>
<th>0</th>
<th>$Y_{i2}$</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$m_{00}$</td>
<td>$m_{01}$</td>
<td>1 $- \mu_{i1}$</td>
</tr>
<tr>
<td>1</td>
<td>$m_{10}$</td>
<td>$e^{\omega_{12}} m_{10} m_{01}/m_{00}$</td>
<td>$\mu_{i1}$</td>
</tr>
<tr>
<td></td>
<td>$1 - \mu_{i2}$</td>
<td>$\mu_{i2}$</td>
<td>1</td>
</tr>
</tbody>
</table>