Within-cases for binary response data using non-linear mixed models[1]

[1] This slide show is an open-source document. See last slide for copyright information.

Vocabulary: Linear vs. non-linear models

- In a linear model, $E(y|\mathbf{x})$ is a linear function of the parameters.
- Ordinary regression is linear:

$$E(y|\mathbf{x}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_{p-1} x_{p-1}$$

- Logistic regression is non-linear:

$$E(y|\mathbf{x}) = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_{p-1} x_{p-1}}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_{p-1} x_{p-1}}}$$

Within-cases for binary data: The idea

- There are several binary responses for each case.
- Like was the person employed right after graduation, 6 months after, one year after . . . Yes or No
- Or did the consumer purchase at least one computer in 2016, 2017, 2018 . . .
- Or did the patient have a seizure on day 1, day 2, . . . after treatment.
- Binary choices in laboratory studies can be repeated measures.
- Model: Logistic regression with a random shock for case, pushing all the log odds values for that case up and down by the same amount.
- Random shock is added to the regression equation for the log odds.
- Usually the random shock is normal — what else?

A random intercept model

For $i = 1, \ldots, n$ and $j = 1, \ldots, k$

- $\Delta_1, \ldots, \Delta_n \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$
- Conditionally on $\Delta_i = \delta_i$ for $i = 1, \ldots, n$, binary responses y_{ij} are independent with

$$\begin{aligned} \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) &= (\beta_0 + \delta_i) + \beta_1 x_{i,j,1} + \ldots + \beta_{p-1} x_{i,j,p-1} \\ &= \mathbf{x}'_{ij}\boldsymbol{\beta} + \delta_i, \text{ so that} \end{aligned}$$

$$\pi_{ij} = \frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\delta_i}}{1 + e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\delta_i}}$$

where $\pi_{ij} = P\{y_{ij} = 1 | \Delta_i = \delta_i\}$.

Some of the $x_{ij\ell}$ could be dummy variables for time period or within-case treatment, different for $j = 1, \ldots, k$ within case i.

Maximum likelihood

- Parameter vector is $\boldsymbol{\theta} = (\beta_0, \beta_1, \ldots, \beta_{p-1}, \sigma^2)'$.
- Vector of binary observations $\mathbf{y}_i = (y_{i1}, \ldots, y_{ik})'$ for each case.
- Likelihood function is $L(\boldsymbol{\theta}) = \prod_{i=1}^{n} p_{\boldsymbol{\theta}}(\mathbf{y}_i)$
- Where $p_{\boldsymbol{\theta}}(\mathbf{y}_i)$ is the probability of observing the vector $\mathbf{y}_i$.
- Need to calculate $p_{\boldsymbol{\theta}}(\mathbf{y}_i)$ as a function of $\boldsymbol{\theta}$ and maximize the likelihood.

Model gives us a *conditional* probability

And we need the unconditional probability $p_{\boldsymbol{\theta}}(\mathbf{y}_i)$

- Given $\Delta_i = \delta_i$, the y_{ij} are independent, so

$$p_{\boldsymbol{\theta}}(\mathbf{y}_i|\Delta_i = \delta_i) = \prod_{j=1}^{k} \left(\frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\delta_i}}{1+e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\delta_i}} \right)^{y_{ij}} \left(1 - \frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\delta_i}}{1+e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\delta_i}} \right)^{1-y_{ij}}$$

- This is a conditional probability.
- Conditional on $\mathbf{x}_{ij}$ as well as δ_i.
- It's okay to treat $\mathbf{x}_{ij}$ as known constants because they are observed.
- But δ_i are unobservable (latent random variables).
- Integrate them out using the law of total probability.

Law of total probability

Double expectation

$$\begin{aligned} p_{\boldsymbol{\theta}}(\mathbf{y}_i) &= \int_{-\infty}^{\infty} p_{\boldsymbol{\theta}}(\mathbf{y}_i|\Delta_i = \delta_i) f(\delta|\sigma^2)\, d\delta_i \\ &= \int_{-\infty}^{\infty} \prod_{j=1}^{k} \left(\frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\delta_i}}{1+e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\delta_i}} \right)^{y_{ij}} \left(1 - \frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\delta_i}}{1+e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\delta_i}} \right)^{1-y_{ij}} f(\delta|\sigma^2)\, d\delta_i \end{aligned}$$

where $f(\delta|\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{\delta^2}{2\sigma^2})$.

- The likelihood is a product of n terms like this.
- Nobody can do the integral.
- It has to be done numerically, n times.
- Numerical integration as well as a numerical search.

State of the art

Contemporary, not just modern

- The theory is mainstream large-sample maximum likelihood.
- Computation is a bit bleeding edge.
- Methods for finding parameter estimates are iterative.
- Convergence problems are common.
- R and SAS give similar results for all the examples I've seen.
- In R, use the `glmer` function in the `lme4` package.
- In SAS, use `proc nlmixed`.
- It's not at all like `proc mixed`.

Copyright Information