Day 5: Everything Else

Introduction to Multilevel Models
EUI Short Course 22–27, 2011
Prof. Kenneth Benoit

May 27, 2011

Generalized linear model specification

As in models for continuous responses, we are interested in the expectation (mean) of the response as a function of the covariate. The expectation of a binary (0 or 1) response is just the probability that the response is 1:

$$E(y_i|x_i) = \Pr(y_i = 1|x_i)$$

In linear regression, the conditional expectation of the response is modeled as a linear function $E(y_i|x_i) = \beta_1 + \beta_2 x_i$ of the covariate (see sec. 1.5). For dichotomous responses, this approach may be problematic because the probability must lie between 0 and 1, whereas regression lines increase (or decrease) indefinitely as the covariate increases (or decreases). Instead, a nonlinear function is specified in one of two ways:

$$\Pr(y_i = 1|x_i) = h(\beta_1 + \beta_2 x_i)$$

or

$$g\{\Pr(y_i = 1|x_i)\} = \beta_1 + \beta_2 x_i = \nu_i$$

where ν_i is referred to as the *linear predictor*. These two formulations are equivalent if the function $h(\cdot)$ is the inverse of the function $g(\cdot)$. Here $g(\cdot)$ is known as the *link function* and $h(\cdot)$ as the *inverse link function*, sometimes written as $g^{-1}(\cdot)$.

Link function for binary response

$$\Pr(y_i = 1|x_i) = \text{logit}^{-1}(\beta_1 + \beta_2 x_i) \equiv \frac{\exp(\beta_1 + \beta_2 x_i)}{1 + \exp(\beta_1 + \beta_2 x_i)}$$

or

$$\text{logit}\left\{\Pr(y_i = 1|x_i)\right\} \equiv \ln \underbrace{\left\{ \frac{\Pr(y_i = 1|x_i)}{1 - \Pr(y_i = 1|x_i)} \right\}}_{\text{Odds}_{(y_i=1|x_i)}} = \beta_1 + \beta_2 x_i$$

Logit models as latent response models

The logistic regression model and other models for dichotomous responses can also be viewed as latent-response models. Underlying the observed dichotomous response y_i (whether the woman works or not), there is an unobserved or latent continuous response y_i^*, representing the propensity to work or the excess utility of working as compared with not working. If this latent response is greater than 0, the observed response is 1:

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

For simplicity, we will assume that there is one covariate x_i. A linear regression model is then specified for the latent response y_i^*

$$y_i^* = \beta_1 + \beta_2 x_i + \epsilon_i$$

where ϵ_i is a residual error term with $E(\epsilon_i|x_i) = 0$ and the error terms of different women i are independent.

"logit" regression v. "progit" regression

- Logit regression:

In logistic regression, ϵ_i is assumed to have a logistic cumulative density function given x_i,

$$\Pr(\epsilon_i < \tau | x_i) = \frac{\exp(\tau)}{1 + \exp(\tau)}$$

which has mean zero and variance $\pi^2/3 \approx 3.29$ (note that π here represents the famous mathematical constant 'pi').

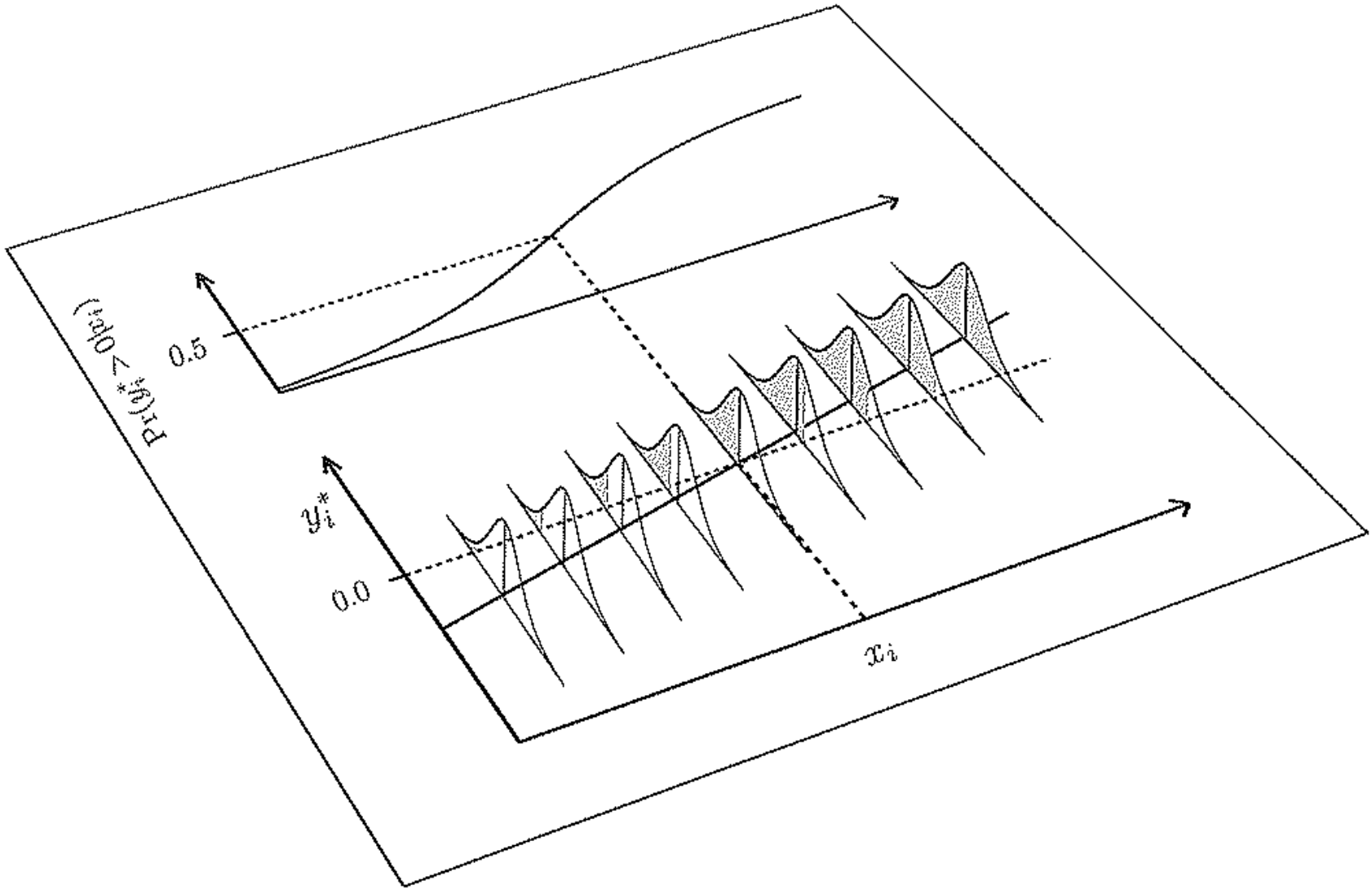
- Probit regression:

When a latent-response formulation is used, it seems natural to assume that ϵ_i has a normal distribution given x_i, as is usually done in linear regression. If a standard (mean 0 and variance 1) normal distribution is assumed, the model becomes a probit model

$$\begin{aligned} \Pr(y_i = 1 | x_i) &= \Pr(y_i^* > 0 | x_i) = \Pr(\beta_1 + \beta_2 x_i + \epsilon_i > 0) \\ &= \Pr\{\epsilon_i > -(\beta_1 + \beta_2 x_i)\} = \Pr(-\epsilon_i \leq \beta_1 + \beta_2 x_i) \\ &= \Pr(\epsilon_i \leq \beta_1 + \beta_2 x_i) = \Phi(\beta_1 + \beta_2 x_i) \end{aligned} \tag{6.4}$$

Here $\Phi(\cdot)$ is the standard normal cumulative distribution function, the probability that a standard normally distributed random variable (here ϵ_i) is less than the argument. $\Phi(\cdot)$ is the inverse link function $h(\cdot)$, whereas the link function $g(\cdot)$ is $\Phi^{-1}(\cdot)$, the inverse standard normal cumulative distribution function, called the *probit link* function. The penultimate equality in (6.4) exploits the symmetry of the normal distribution.

Illustration of logit GLM and latent response formulations



Random-intercept logistic regression

- To relax the assumption of conditional independence we add a group-specific random intercept ζ_j to the linear predictor:

$$\text{logit}\{\Pr(y_{ij}=1|\mathbf{x}_{ij},\zeta_j)\} = \beta_1 + \beta_2 x_{2j} + \beta_3 x_{3ij} + \beta_4 x_{2j} x_{3ij} + \zeta_j$$

- We assume that $y_{ij}|\pi_{ij} \sim \text{binomial}(1, \pi_{ij})$, given that $\pi ij \equiv \Pr(y_{ij}|\mathbf{x}_{ij}, \zeta_j)$
- We can estimate this model using `xtlogit`
- Alternatively we can use the `xtmelogit` command, but we must specify the number of integration points

Poisson models

- This is also a GLM, but with different link and error functions
- Focus here is on a constant incidence rate λ defined as the instantaneous probability of a new event per time interval
- The number of events y that occur in time t has a Poisson distribution:

$$\Pr(y|\mu) = \frac{\exp(-\mu)\mu^y}{y!}$$

where $\mathrm{E}(y) = \mu$ and is given by

$$\mu = \lambda t$$

Specification for single-level Poisson regression

The expected number of visits μ_{ij} at occasion i for subject j is specifie
linear model

$$\ln(\mu_{ij}) = \nu_{ij} = \beta_1 + \beta_2 x_{2i} + \cdots + \beta_7 x_{7ij}$$

or equivalently as an exponential model for the expected number of visits:

$$\mu_{ij} = \exp(\nu_{ij})$$

Random intercept model for multilevel Poisson regression

One way to address the dependence within persons is to include a person-spec intercept ζ_{1j} in the Poisson regression model

$$\begin{aligned}\mu_{ij} \equiv E(y_{ij}|\mathbf{x}_{ij}, \zeta_{1j}) &= \exp(\beta_1 + \beta_2 x_{2i} + \cdots + \beta_7 x_{7ij} + \zeta_{1j}) \\ &= \exp\{(\beta_1 + \zeta_{1j}) + \beta_2 x_{2i} + \cdots + \beta_7 x_{7ij}\} \\ &= \exp(\zeta_{1j}) \exp(\beta_1 + \beta_2 x_{2i} + \cdots + \beta_7 x_{7ij})\end{aligned}$$

where $\zeta_{1j}|\mathbf{x}_{ij} \sim \mathrm{N}(0, \psi_{11})$ and the ζ_{1j} are independent across persons j. The of the random intercept, $\exp(\zeta_{1j})$, is sometimes called a *frailty*. The number

- Commands are `xtpoisson` and `xtmepoisson`
- It is also possible to run random coefficients models for Poisson