Another way to generalize item response and ideal point models is to go beyond the logistic distribution, for example using a robit model as described in Section 6.6 that allows for occasional mispredictions.

14.4 Non-nested overdispersed model for death sentence reversals

So far in this chapter we have presented logistic regression for binary data points y_i that can equal 0 or 1. The model can also be used for proportions, in which each data point y_i equals the number of "successes" out of n_i chances. For example, Section 6.3 describes data on death penalty reversals, in which i indexes state-years (for example, Alabama in 1983), n_i is the number of death sentences given out in that particular state in that particular year, and y_i is the number of these death sentences that were reversed by a higher court. We now describe how we added multilevel structure to this model.

Non-nested model for state and year coefficients

The death penalty model had several predictors in X, including measures of the frequency that the death sentence was imposed, the backlog of capital cases in the appeals courts, the level of political pressure on judges, and other variables at the state-year level.

In addition, we included indicators for the years from 1973 to 1995 and the 34 states (all of those in this time span that had death penalty laws). The regression model with all these predictors can be written as

$$\begin{aligned} y_i &\sim \mathrm{Bin}(n_i, p_i) \\ p_i &= \mathrm{logit}^{-1}(X_i\beta + \alpha_{j[i]} + \gamma_{t[i]}), \end{aligned} \tag{14.15}$$

where j indexes states and t indexes years. We complete the multilevel model with distributions for the state and year coefficients,

$$\begin{aligned} \alpha_j &\sim \mathrm{N}(0, \sigma_\alpha^2) \\ \gamma_t &\sim \mathrm{N}(a + bt, \sigma_\gamma^2). \end{aligned}$$

The coefficients for year include a linear time trend to capture the overall increase in reversal rates during the period under study. The model for the γ_t's also includes an intercept, and so we do not need to include a constant term in the model for the α_j's or in the matrix X of individual-level predictors in (14.15).

In this particular example, we are not particularly interested in the coefficients for individual states or years; rather, we want to include these sources of variability into the model in order to get appropriate uncertainty estimates for the coefficients of interest, β.

Multilevel overdispersed binomial regression

Testing for overdispersion. Model (14.15) is inappropriate for the death penalty data because the data are overdispersed, as discussed in Section 6.3. To measure the overdispersion, we compute the standardized residuals, $z_i = (y_i - p_i)/\sqrt{p_i(1-p_i)/n_i}$ with p_i as defined in (14.15). Under the binomial model, the residuals should have mean 0 and standard deviation 1, and so $\sum_i z_i^2$ should look like a random draw from a χ^2 distribution with degrees of freedom equal to 520 (the number of state-years in the data).

Testing for overdispersion in a classical binomial regression is described in Section

6.3, where the z_i's are computed based on estimated probabilities $\hat{p}_i$, and $\sum_i z_i^2$ is compared to a χ^2 distribution with degrees of freedom adjusted for the number of coefficients estimated in the model.

Beta-binomial model. There are two natural overdispersed generalizations of the multilevel binomial regression (14.15). The first approach uses the beta-binomial distribution:

$$y_i \sim \text{beta-binomial}(n_i, p_i, \omega),$$

where $\omega \geq 1$ is the overdispersion parameter (and the model with $\omega = 1$ reduces to the binomial).

Binomial-normal model. The other direct way to construct an overdispersed binomial distribution is to add normal errors on the logistic scale, keeping the binomial model but adding a data-level error ξ_i to the linear predictor in (14.15):

$$p_i = \text{logit}^{-1}(X_i\beta + \alpha_{j[i]} + \gamma_{t[i]} + \xi_i),$$

with these errors having their own normal distribution:

$$\xi_i \sim \text{N}(0, \sigma_\xi^2).$$

The resulting model reduces to the binomial when $\sigma_\xi = 0$; otherwise it is overdispersed.

With moderate sample sizes, it is typically difficult to distinguish between the beta-binomial and binomial-normal models, and the choice between them is one of convenience. The beta-binomial model adds only one new parameter and so can be easier to fit; however, the binomial-normal model has the advantage that the new error term ξ_i is on the same scale as the group-level predictors, α_j and γ_t, which can make the fitted model easier to understand.

14.5 Bibliographic note

Multilevel logistic regression has a long history in the statistical and applied literature which we do not attempt to trace here: the basic ideas are the same as in multilevel linear models (see references in Sections 12.10 and 13.8) but with complications arising from the discreteness of the data and the nonlinearity of some of the computational steps.

The example of state-level opinions from national polls comes from Gelman and Little (1997) and Park, Gelman, and Bafumi (2004). The analysis of income and voting comes from Gelman, Shor, et al. (2005); see also Wright (1989), Ansolabehere, Rodden, and Snyder (2005), and McCarty, Poole, and Rosenthal (2005) for related work. Figure 14.10, which simultaneously displays patterns within and between groups, is related to the "B-K plot" (discussed by Wainer, 2002, and named after Baker and Kramer, 2001).

The multilevel framework for item-response and ideal-point models appears in Bafumi, Gelman, and Park (2005). See Lord and Novick (1968) and van der Linden and Hambleton (1997) for more on item-response models, and Poole and Rosenthal (1997), Jackman (2001), and Martin and Quinn (2002a) for more on ideal-point models. Loken (2004) discusses identifiability problems in models with aliasing.

The death sentencing example comes from Gelman, Liebman, et al. (2004). See Donohue and Wolfers (2006) for an overview of some of the research literature on death sentencing.