Ordinal and Nominal Dependent Variables

Applying MLE

Jan Rovny

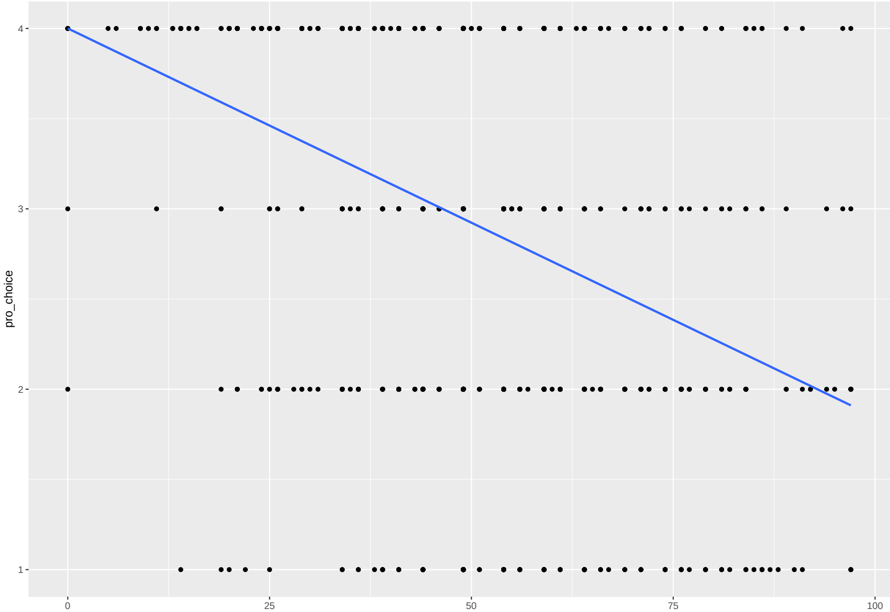
January 10, 2020

Ordinal Outcomes

- Ordinal variables have multiple categories
- They are ordered, but do not have meaningful distances between them – the numeric values have no inherent meaning
- We can imagine
  - a latent continuous variable y^*
  - an observed variable y
  - y provides incomplete information about the underlying y^*
  - different thresholds τ on y^* translate to categories of y

    $y_i = 1$ if $\tau_0 = -\infty \leq y_i^* < \tau_1$
    $y_i = 2$ if $\tau_1 \leq y_i^* < \tau_2$
    $y_i = 3$ if $\tau_2 \leq y_i^* < \tau_3$
    $y_i = 4$ if $\tau_3 \leq y_i^* < \tau_4 = \infty$

Ordinal Outcomes – With OLS line



Ordinal Outcomes – Assumptions

- To estimate with MLE, we must assume form of error distribution
- For order logit, we assume logistic distribution: mean=0, variance=$\pi^2/3$
- Calculate the probability that $y = 1$:

$$Pr(y_i = 1|x_i) = Pr(\epsilon < \tau_1 - \boldsymbol{x_i}\boldsymbol{\beta}|\boldsymbol{x_i}) - Pr(\epsilon_i \leq \tau_0 - \boldsymbol{x_i}\boldsymbol{\beta}|\boldsymbol{x_i})$$

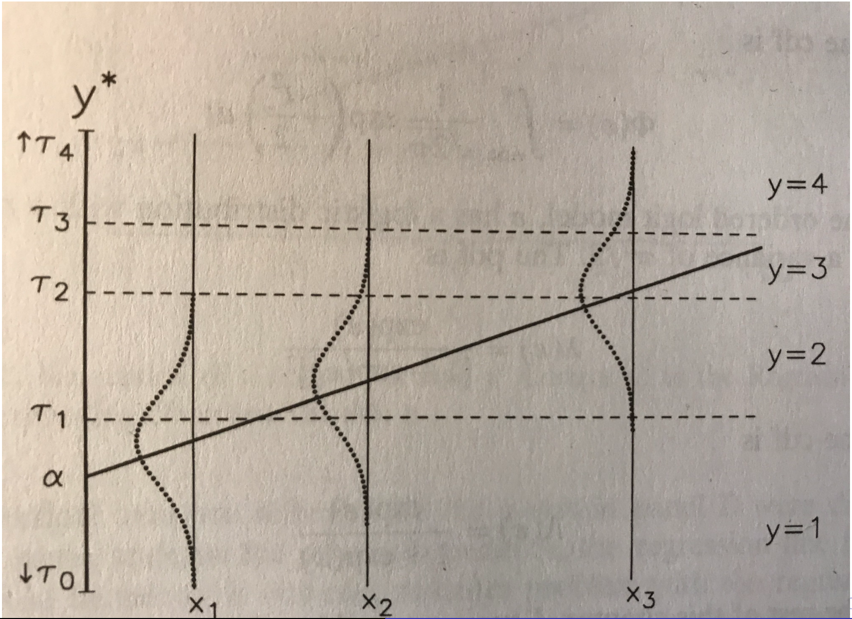
$$= F(\tau_1 - \boldsymbol{x_i}\boldsymbol{\beta}) - F(\tau_0 - \boldsymbol{x_i}\boldsymbol{\beta})$$

- Generally, the probability that $y = m$ (m is the category of y: 1,2,3,4)

$$Pr(y_i = m) = F(\tau_m - \boldsymbol{x_i}\boldsymbol{\beta}) - F(\tau_{m-1} - \boldsymbol{x_i}\boldsymbol{\beta})$$

(F is the logistic or normal CDF)

Ordinal Outcome Probabilities



Ordinal Outcomes – Identification

- To estimate the model, we need to make further assumptions
- Note that we cannot estimate the βs and τs simultaneously – the model is unidentified
- To identify the model, we assume that $\beta_0 = 0$ in order logit or that $\tau_1 = 0$ in ordered probit
- This choice is arbitrary and does not affect the other βs or the probabilities

Ordinal Outcomes – Estimation

- To estimate the model, we maximize:

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\tau} | \boldsymbol{y}, \boldsymbol{X}) = \prod_{m=1} \prod_{y_i = m} Pr(y_i = m | \boldsymbol{x_i}, \boldsymbol{\beta}, \boldsymbol{\tau})$$

$$= \prod_{m=1} \prod_{y_i = m} [F(\tau_m - \boldsymbol{x_i}\beta) - F(\tau_{m-1} - \boldsymbol{x_i}\beta)]$$

(F is the logistic or normal CDF)

Nominal Outcomes – Multinomial Logit (MNL)

- Dependent variable is a set of nominal categories, with no order.
- The approach is to simultaneously estimates binary logits for all possible comparisons among the outcome categories.
- The advantage of the MNL is that the sample is used more efficiently.
- To do this, it is necessary to establish a baseline choice category (usually m=1), whose parameters are constrained to be 0.

Nominal Outcomes – Assumptions

- Assume a logit function to have probabilities of outcomes constrained between [0;1]
- And constrain β_1 to 0
- The probability of observing outcome $y = 1$ is:

$$Pr(y_i = 1|x) = \pi_{i1} = \frac{\exp(x_i 0)}{\exp(x_i 0) + \sum_{k=2}^{M} \exp(x_i \beta_k)} =$$

$$= \frac{1}{1 + \sum_{k=2}^{M} \exp(x_i \beta_k)}$$

(where k are the remaining, unconstrained outcomes)

- Then for the remaining (unconstrained) outcomes m:

$$Pr(y_i = m|x) = \pi_{im} = \frac{\exp(x_i \beta_m)}{1 + \sum_{k=2}^{M} \exp(x_i \beta_k)}$$

Nominal Outcomes – Estimation

- Maximize the likelihood function:

$$\mathcal{L} = \prod_{y_{i1}=1} \pi_{i1} \prod_{y_{i2}=1} \pi_{i2} \cdots \prod_{y_{iM}=1} \pi_{iM} \quad (1)$$

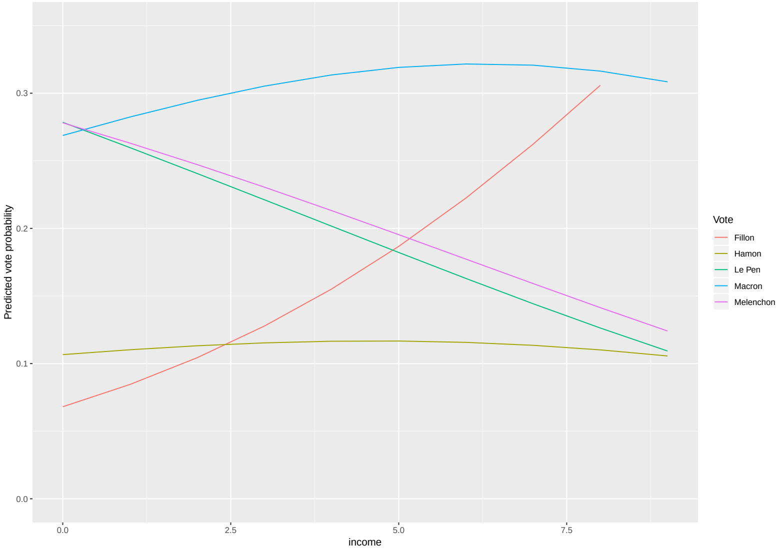
$$\mathcal{L} = \prod_{i=1}^{N} \prod_{m=1}^{M} \pi_{im}^{y_{im}} \quad (2)$$

Predicted Probabilities

- The neatest way to communicate our results for continuous predictors is to demonstrate the predicted probabilities of a given predictor,
- Other predictors are held constant (usually at their mean or other logical value)
- In R, we define a data frame, mandating the values of the xs at which we wish to estimate the probabilities
- We keep all but one constant. We vary the predictor of interest logically
  - From its minimum to maximum
  - From `mean-sd` to `mean` to `mean+sd`
  - Or over some key values of interest

Example

Explaining vote in 2017 French presidential election



Warning: Independence of Irrelevant Alternatives

- MNL is based on an assumption that the decision between alternative 1 and alternative 2, for example, is independent on all other alternatives.
- In our case, whether a voter supports say Macron over Hamon is assumed independent on Melenchon.
- This is called the Independence of Irrelevant Alternatives or IIA assumption.