Maximum Likelihood

- In ML, we first need to specify the data generating process for the dependent variable under examination.
- In other words, we need to specify the probability distribution that generated the dependent variable; e.g., the normal for continuous variable, logit or probit for dichotomous, poisson for count data, etc.

Maximum Likelihood

- Then, we specify the likelihood for case i:

$$L_i = L(\theta \mid y_i)$$
$$L(\theta \mid y_i) \propto p(y_i \mid \theta)$$

- The likelihood for the entire sample is simply the product of individual likelihoods: $L = \prod_{i=1}^{N} L_i$

- MLEs are the values of the parameters for which the likelihood of observing the sample is maximized.

MLE: Normal Regression

- Y ~ N(μ, s²)
- pdf: $f(y_i \mid \mu_i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(y_i-\mu_i)^2}{\sigma^2}}$
- Reparameterize $\mu_i = x_i\beta$
- Likelihood for case i: $L_i = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(y_i-x_i\beta)^2}{\sigma^2}}$
- Log-likelihood for case i

$$\ln L_i = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2}\left[\frac{(y_i - x_i\beta)^2}{\sigma^2}\right]$$

MLE: Normal Regression

- The likelihood for the entire sample is simply the product of the individual likelihoods:

$$L = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(y_i - x_i\beta)^2}{\sigma^2}}$$

- And the log-likelihood for the entire sample is simply:

$$\ln L = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\left[\frac{(y_i - x_i\beta)^2}{\sigma^2}\right]$$

MLE: Logit and Probit

- In binary response models, we want to model the probability of "success" for case i, i.e., $\Pr(y_i=1) = ?_i$
- We parameterize $?_i$ as a cumulative distribution function (cdf) of a particular distribution, i.e., $F(x_i\beta)$
  - For logit, we use the logistic cdf:

$$\Pr(y_i = 1) = F(x_i\beta) = \frac{\exp(x_i\beta)}{1+\exp(x_i\beta)}$$

  - For probit, we use the normal cdf:

$$\Pr(y_i = 1) = F(x_i\beta) = \Phi(x_i\beta)$$

MLE: Logit and Probit

- The likelihood for case i is:

$$L_i = [F(x_i\beta)]^{y_i}[1 - F(x_i\beta)]^{1-y_i}$$

- The log-likelihood for case i is:

$$\ln L_i = y_i \ln[F(x_i\beta)] + (1 - y_i)\ln[1 - F(x_i\beta)]$$

- For logit, we'll replace $F(x_i\beta)$ with the logistic cdf, and for probit, the normal cdf.

Interpretation

- Logit and probit coefficients:
  - Sign and statistical significance.

- For size and substantive interpretations:
  - Predicted probabilities and changes in those probabilities as variable of interest changes from low to high (averaging over other variables).

- Hanmer and Kalkan, 2013 *AJPS*.

Goodness of fit

- Likelihood ratio (LR) test
  - Analogous to F test in OLS
  - Compare fully specified model to null model (with no indep. variables)

- Lots of "Pseudo R-sq" measures: Comparing "full model" to "null model"

- Percent correctly predicted:
  - Using .5 threshold
  - But relative to what?

Goodness of fit

- **Proportional reduction in error**:
  - How much better you do with full model compared with the mean of the dependent variable (null model with no IVs).

- Error from null model: $\boldsymbol{\theta}_N$
- Error from full model: $\boldsymbol{\theta}_F$

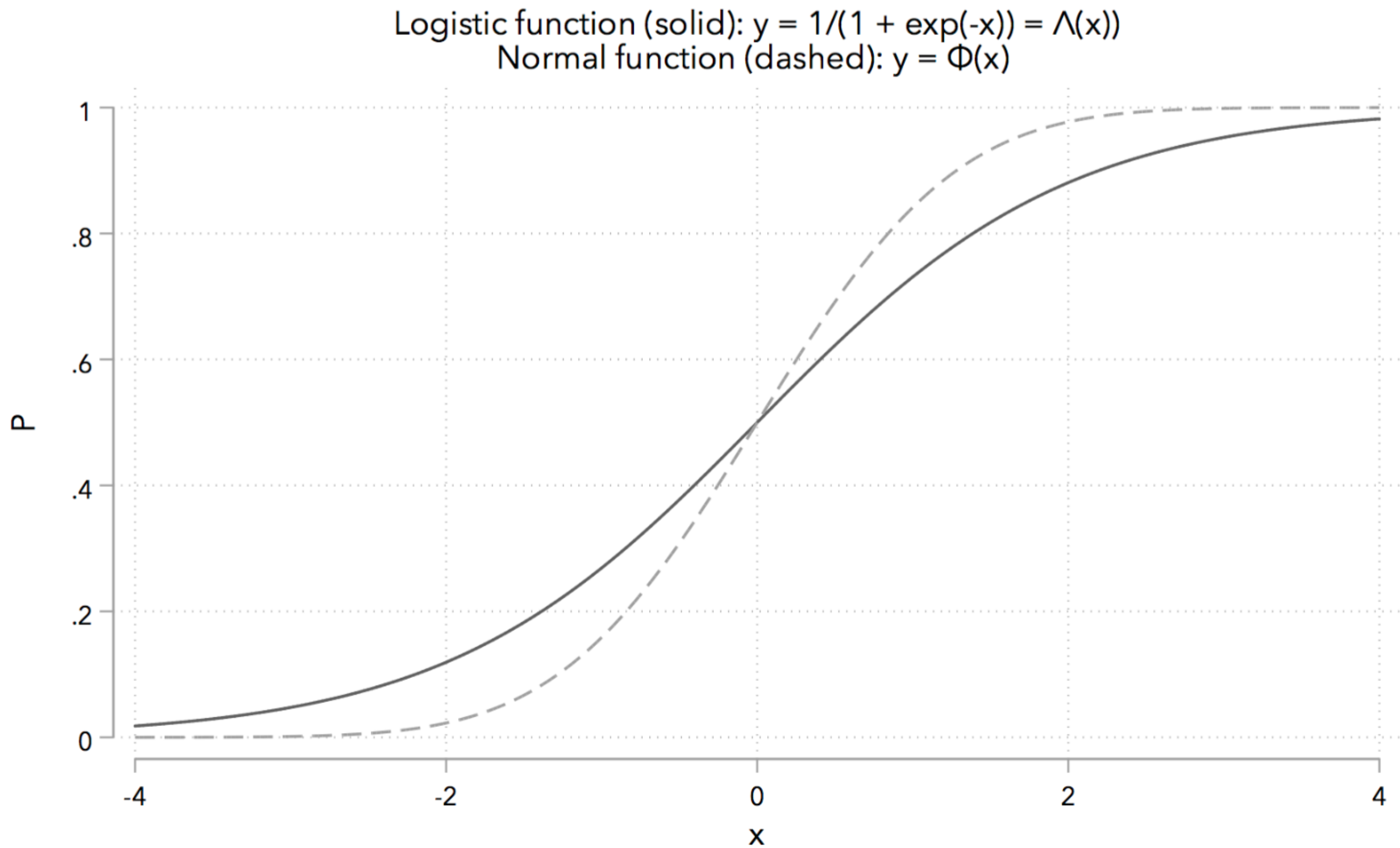
- Proportional reduction in error: $\frac{\theta_N - \theta_F}{\theta_N}$

- Null model: "proportion modal category" (PMC); mean of the binary dependent variable.
- Full model: "proportion correctly predicted" (PCP).
- "Error" in each represented by: (1-PMC) and (1-PCP)

- Plug in to equation above: $\frac{(1-PMC)-(1-PCP)}{(1-PMC)} = \frac{PCP-PMC}{1-PMC}$

Logit v. Probit

Logit/probit: $y = F(x)$



Predicted Probabilities

- Observed value v. average case approach,
  - Hanmer and Kalkan

Data

Vote (DV)	Econ	Party	Relig
1	4	0	3
1	3	5	0
0	5	4	2
1	1	6	4
0	2	2	3
0	5	4	1
1	4	1	5

Average Case Approach

Vote Clinton	Econ	Party	Relig
1	4	0	3
1	3	5	0
0	5	4	2
1	1	6	4
0	2	2	3
0	5	4	1
1	4	1	5

Prob(Vote Clinton) for all five values of econ perceptions (1 to 5) while holding other vars constant at their means.

First, calculate "linear predictions" (plugging in betas and x's) for logit:

Econ=1: $\hat{y} = \beta_0 + \beta_1(1) + \beta_2(3.83) + \beta_3(1.7) =$
7.76 - 1.06(1) - 1.07(3.83) - 0.17(1.7)

Econ=2: 7.76 - 1.06(2) - 1.07(3.83) - 0.17(1.7)

Econ=3: 7.76 - 1.06(3) - 1.07(3.83) - 0.17(1.7)

Econ=4: 7.76 - 1.06(4) - 1.07(3.83) - 0.17(1.7)

Econ=5: 7.76 - 1.06(5) - 1.07(3.83) - 0.17(1.7)

Average Case Approach

Prob(Vote Clinton) for all five values of econ perceptions (1 to 5) while holding other vars constant at their means.

First, calculate "linear predictions" (plugging in betas and x's) for logit:

Econ=1: $\hat{y} = \beta_0 + \beta_1(1) + \beta_2(3.83) + \beta_3(1.7) =$
7.76 - 1.06(1) - 1.07(3.83) - 0.17(1.7)
Econ=2: 7.76 - 1.06(2) - 1.07(3.83) - 0.17(1.7)
Econ=3: 7.76 - 1.06(3) - 1.07(3.83) - 0.17(1.7)
Econ=4: 7.76 - 1.06(4) - 1.07(3.83) - 0.17(1.7)
Econ=5: 7.76 - 1.06(5) - 1.07(3.83) - 0.17(1.7)

Plug in "linear predictions" for each of the five values above into cdf equations.

Logit:

Predicted Prob(Vote Clinton) = $\frac{\exp(\hat{y})}{1+\exp(\hat{y})}$

Observed Value Approach

Vote Clinton	Econ	Party	Relig
1	4	0	3
1	3	5	0
0	5	4	2
1	1	6	4
0	2	2	3
0	5	4	1
1	4	1	5

Econ=1, other vars maintain values

Vote Clinton	Econ	Party	Relig	exp(xb)/ (1+exp(xb))
1	1	0	3	.65
1	1	5	0	.88
0	1	4	2	.46
1	1	6	4	.72
0	1	2	3	.50
0	1	4	1	.12
1	1	1	5	.55
			Mean Prob	**.74**

Observed Value Approach

Econ=2, other vars maintain values

Vote Clinton	Econ	Party	Relig	exp(xb)/ (1+exp(xb))
1	2	0	3	.62
1	2	5	0	.79
0	2	4	2	.33
1	2	6	4	.66
0	2	2	3	.35
0	2	4	1	.08
1	2	1	5	.54
			Mean Prob	**.63**

Observed Value Approach

Econ=3, other vars maintain values

Vote Clinton	Econ	Party	Relig	exp(xb)/ (1+exp(xb))
1	3	0	3	.59
1	3	5	0	.70
0	3	4	2	.26
1	3	6	4	.60
0	3	2	3	.30
0	3	4	1	.07
1	3	1	5	.51
			Mean Prob	**.53**

Observed Value Approach

Econ=4, other vars maintain values

Vote Clinton	Econ	Party	Relig	exp(xb)/ (1+exp(xb))
1	4	0	3	.55
1	4	5	0	.60
0	4	4	2	.20
1	4	6	4	.55
0	4	2	3	.22
0	4	4	1	.04
1	4	1	5	.46
			Mean Prob	**.42**

Observed Value Approach

Econ=5, other vars maintain values

Vote Clinton	Econ	Party	Relig	exp(xb)/ (1+exp(xb))
1	5	0	3	.50
1	5	5	0	.55
0	5	4	2	.15
1	5	6	4	.40
0	5	2	3	.15
0	5	4	1	.02
1	5	1	5	.33
			Mean Prob	**.30**

Linear Probability Model

- Using OLS with robust standard errors.

- Issues: Making out-of-bounds predictions (>1 or <0)

- Benefits: Not prone to bias when lots of fixed effects (with clustered data).

- Robustness checks

- Lean toward logit/probit unless good reason to use LPM.

COURSE EVALUATIONS

https://gwu.smartevals.com