Logit and Probit

Brad Jones[1]

[1]Department of Political Science
University of California, Davis

April 21, 2009

Today: Binary Response Models

Logit, redux

- Logit resolves the functional form problem (in terms of the response function in the probabilities.
- Note that in the probabilities, logit *is* a non-linear model.
- Suppose $E(y) = P(y = 1 \mid x) = \beta_k x_{ik}$, and y is binary.

$$\Pr(y = 1 \mid x) = \frac{1}{1 + \exp(-\sum \beta_k x_{ik})}$$

- Let $Z = \sum \beta_k x_{ik}$, then

$$\Pr(y = 1 \mid x) = \frac{1}{1 + \exp(-Z)} = \frac{\exp(Z)}{1 + \exp(Z)}$$

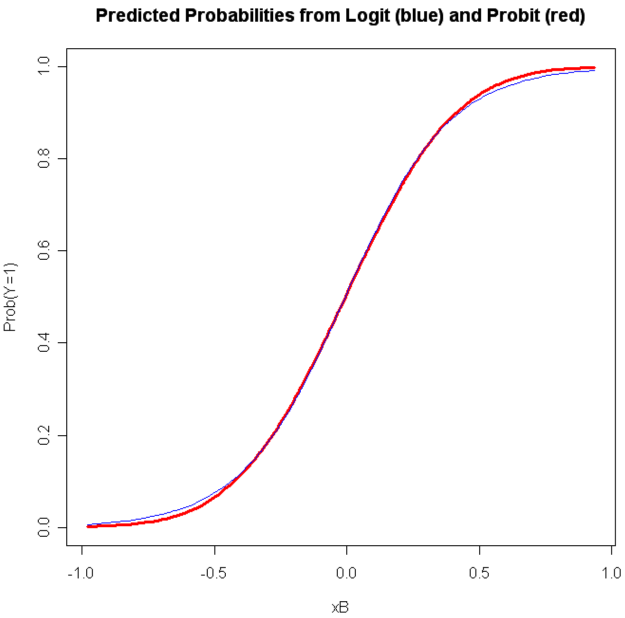
- This is the c.d.f. for the logistic distribution.
- Problems Solved:
  Z is unbounded; P_i must stay in unit interval.
  P_i is nonlinearly related to parameters (though logit is linear!)
  Prediction of "1" or "0" impossible.

The Probit Model

- Often will be similar to binary logit.
- Derived from binomial family.
- But CDF is derived from normal distribution.
- And it looks a little something like this:

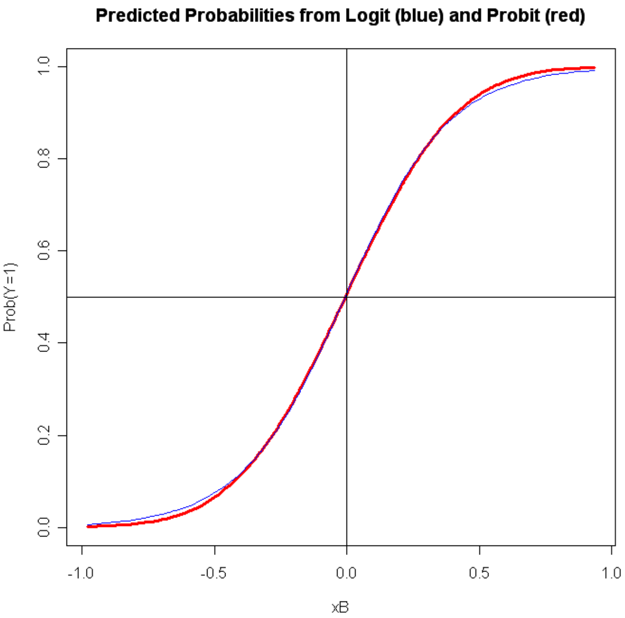
$$F(x) = \Pr(y = 1 \mid x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x)^2\right) dx \qquad (1)$$

- You've seen this before: it's the CDF for the standard normal.
- $\Pr(y = 1)$ is symmetrical about .50 (same is true for logit).
- Quick Fit


Predicted Probabilities from Logit (blue) and Probit (red)
Prob(Y=1)
0.0
0.2
0.4
0.6
0.8
1.0
-1.0
-0.5
0.0
0.5
1.0
xB

Comparisons

- Immediately we see several things.
- The coefficients are not "directly comparable."
- Why? They are scaled differently.
- However, signs and significance are identical (this is typical).
- In the probabilities, the models are virtually indistinguishable.
- For many problems (esp. in social sciences), this result is generally true.
- Logit has "fatter" tails.
- Note also the probabilities are symmetrical about .5.


Predicted Probabilities from Logit (blue) and Probit (red)
Prob(Y=1)
0.0
0.2
0.4
0.6
0.8
1.0
-1.0
-0.5
0.0
0.5
1.0
xB

Probit: The Details

- The probit probabilities are nonlinear in x.
- We see this in the figures.
- The model we estimate, however, is a linear model.
- How do we get to it?
- Under probit we have:

$$\Pr(y = 1 \mid x) = F(\mathbf{x}\beta) = \mathbf{\Phi}(\mathbf{x}\beta) \tag{2}$$

- $\mathbf{\Phi}$ is the cdf for the standard normal.
- For a given covariate profile, this function gives you the probability area in the standard normal distribution.
- When you looked up z-scores, you were doing the same basic thing.
- We still don't have a model.

Probit: The Details

- Again, the probabilities are nonlinear functions of $\mathbf{x}\boldsymbol{\beta}$.
- We can linearize this.
- As probabilities:
  $\Pr(y = 1 \mid x) = \mathbf{\Phi} \sum \hat{\beta}_k x_{ik}$
- (Yes, still nonlinear)
- Suppose we take the inverse of $\mathbf{\Phi}$?

$$\mathbf{\Phi}^{-1}(p_i) = \sum \hat{\beta}_k x_{ik} \quad (3)$$

- We obtain a linear model.
- This is the *probit model.*
- As with logit, the coefficients are *not* scaled as probabilities.
- Here, they are scaled in terms of the inverse of the standard normal distribution.

Probit: The Details

- We don't usually think in terms of this scale.
- However, the signs are informative (like log-odds).
- As x increases, we move up the probit scale.
- The likelihood of responding "1" concomitantly increase.
- There is no odds ratio interpretation here.
- Since we know the probability function, all we need to do to derive $\Pr(y = 1)$ is to compute the linear prediction and find the corresponding probability.
- Illustration

Probit: The Details

- Model from before:

$$\mathbf{\Phi}^{-1}(p_i) = .0189 + 3.0568x_i$$

- Suppose $x_i = \overline{x} = .01$.
- Linear Prediction: $.0189 + 3.0568 * .01 = .05$
- Probability: $\mathbf{\Phi}.05 = .52$.
- How'd I do that??
- The "hard" way.

Probit: The Details

- You see, the linear prediction from a probit model is . . .
- a *z*-score.
- If you've got a *z*-table and a linear prediction, you can get your probabilities.
- If you have a computer, you can do this too.
- Many ways in `R`. My way (using `VGAM` library): probit(.05, inverse=TRUE) which returns: 0.52.
- In `Stata`: display normal(.05) which returns `.51993881`
- This is all really hard, isn't it?
- Multiple covariates means you'll account for more terms in the linear prediction.
- Appreciably no different from logit.

Logit comparison

- Logit model (in log-odds): $.033 + 5.094x_i$
- Linear prediction at mean of x_i: .084
- Probability: $e^{.084}/(1 + e^{.084}) = .52$
- Equivalency: $1/(1 + e^{-.084}) = .52$
- How'd I do that?
- Utilize the cdf for the logistic distribution.
- Distribution function differs from probit; probabilities are identical.
- Either way, probabilities are easy to compute.
- Just make sure you're computing probabilities for reasonable profiles.
- Nonsensical profiles=nonsensical probabilities.

Interpretation: Extended Illustration

- Data: California Field Poll (Feb. 2006)
- Response Variable: "Would you support creating a temporary worker program in Ca.?"
- Covariates: Party affiliation (-1,0,1); Education Level (1-10); Income Level (1-5); Latino origin (1,0)
- First estimate logit.

Logit Results from California Field Poll Data

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.1620	0.3645	0.44	0.6568
pid	0.3900	0.1453	2.68	0.0073
education	−0.0367	0.0521	−0.71	0.4804
income	0.1668	0.0905	1.84	0.0653
latino	1.3348	0.3023	4.42	0.0000

Interpretation: Extended Illustration

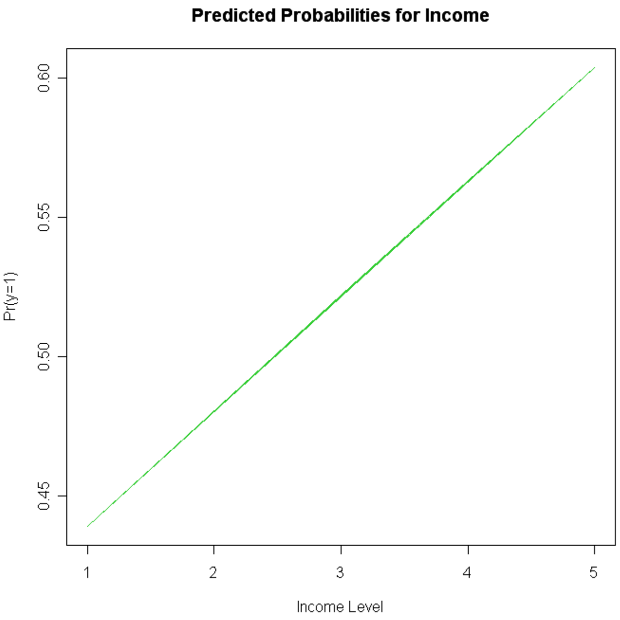
- Signs? Related to log-odds ratios.
- Hypothesis testing: z-ratio is $\hat{\beta}/s.e.(\hat{\beta})$.
- 95 percent confidence interval approximately $\hat{\beta} + -1.96(s.e.)$
- Usual rules apply (and usual caveats).
- I advise against using stars. Why?
- Turn to interpretation

Interpreting the Model

- Consider probabilities first.
- Useful to compute tables of probabilities.
- e.g. what is $\Pr(y = 1 \mid \text{profile})$?
- Non-latino, low-education, low-income, Republican?
- $x_1 = -1$, $x_2 = 3$, $x_3 = 1$, $x_4 = 0$
- $\Pr(y = 1 \mid Z) = .46$
- Same scenario, but for a Democrat?
- $\Pr(y = 1 \mid Z) = .65$
- Almost a .20 difference attributable to partisanship.
- But be careful.

Probabilities

- You can imagine there are a lot of probabilities you could compute.
- Specifically, $10 * 5 * 3 * 2 = 300$. Not all profiles exist, however!
- You could graph them as well. Consider income.
- $\Pr(y = 1)$ given a non-Latino Republican having mean-level education.
- Let income assume each value and then plot it.


Predicted Probabilities for Income
Pr(y=1)
0.45
0.50
0.55
0.60
1
2
3
4
5
Income Level

Probabilities

- You could overlay different probability scenarios on the graph.
- Marginal changes in $\Pr(y)$ are another way to interpret the model.
- For logit:

$$\begin{aligned}\frac{\partial \Pr(y=1 \mid \mathbf{x})}{\partial x_k} &= \frac{\exp(\mathbf{x}\boldsymbol{\beta})}{[1+\exp(\mathbf{x}\boldsymbol{\beta})]^2}\beta_k \\ &= \Pr(y=1 \mid \mathbf{x})[1-\Pr(y=1 \mid \mathbf{x})]\beta_k \quad (4)\end{aligned}$$

- Analogy to regression: the marginal change is just β.
- Here, quantity of interest (probability) is nonlinear in x.
- Interpretation? It's the slope of the probability curve holding all other variables constant at some value.

Probabilities

- In probit, marginal change is given by:

$$\frac{\partial \Pr(y = 1 \mid \mathbf{x})}{\partial x_k} = \phi(\mathbf{x}\boldsymbol{\beta})\beta_k \tag{5}$$

- Same kind of interpretation is forthcoming from it.
- These are easy quantities to derive. If you use `Stata`, `spost` will help you out here.
- But if we can compute probabilities (which we just have!), we can compute marginal effects.
- Computing the marginal effects at means is sometimes useful.

Odds Ratios

- In logit, odds ratios may be useful.
- The odds ratio is just $\exp(\hat{\beta}_k)$.
- For a binary covariate, the interpretation is simple.
- Consider Latino origin variable. Coefficient is 1.335.
- Odds ratio? $\exp(1.335 * 1) = 3.8$.
- Interpretation: respondents of Latino origin are nearly 4 times more likely to claim support for temporary worker program than when compared to non-Latino origin respondents.
- Nice interpretation.

Odds Ratios

- Continuous variables
- Odds are proportional between adjacent categories.
- Percentage change in odds:

$$\%\Delta \frac{p_i}{1 - p_i} = \left[\frac{\exp(\beta_k)x - \exp(\beta_k)x'}{\exp(\beta_k)x'} \right] * 100 \qquad (6)$$

- For income covariate, the odds ratio when income=5 is: 2.3
- For income=4, odds ratio is: 1.9
- Change in odds of moving from category 4 to 5? About 18 percent.
- For binary covariates, this is simple to compute.
- Odds for Latinos: 3.8; percentage change?
- $(3.8 - 1)/1$=2.8 (or about 280 percent).

Probit

- Consider probit model on same data.
- Coefficients will be in "*z*-score" metric.
- Signage should be the same.
- As should significance.
- Only exceptions would be when extreme probabilities are of interest.

Probit Results from California Field Poll Data

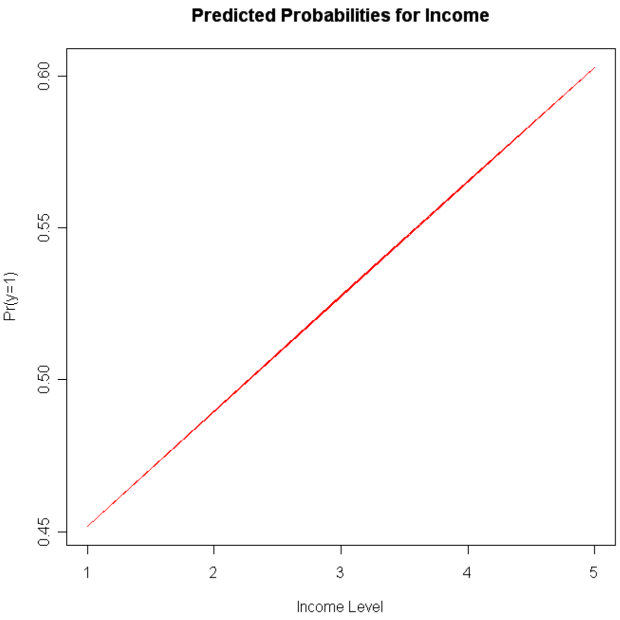
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.1358	0.2184	0.62	0.5340
pid	0.2333	0.0871	2.68	0.0074
education	−0.0239	0.0310	−0.77	0.4411
income	0.0955	0.0540	1.77	0.0771
latino	0.7629	0.1707	4.47	0.0000

Interpretation: Extended Illustration

- Signs? Related to z scores.
- Hypothesis testing: z-ratio is $\hat{\beta}/s.e.(\hat{\beta})$.
- 95 percent confidence interval approximately $\hat{\beta} + -1.96(s.e.)$
- Usual rules apply (and usual caveats).
- I advise against using stars. Why?
- Turn to interpretation.

Interpreting the Model

- Consider probabilities first.
- Useful to compute tables of probabilities.
- e.g. what is $\Pr(y = 1 \mid \text{profile})$?
- Non-latino, low-education, low-income, Republican?
- $x_1 = -1$, $x_2 = 3$, $x_3 = 1$, $x_4 = 0$
- $\Pr(y = 1 \mid Z) = .47$ $(z = -.074)$
- Same scenario, but for a Democrat?
- $\Pr(y = 1 \mid Z) = .65$ $(z = .39)$
- Almost a .20 difference attributable to partisanship.
- But be careful.


Predicted Probabilities for Income
Pr(y=1)
0.45
0.50
0.55
0.60
1
2
3
4
5
Income Level

Likelihood

- Assume binary response is Bernoulli distributed.
- The PMF:

$$f(y_i \mid \mathbf{x}_i) = p_i^{y_i}(1-p_i)^{1-y_i}$$

- $p_i = F(\mathbf{x}_i'\beta)$ (i.e. the probabilities assume some distribution function.)
- Since we like to work with log-likelihoods, the "generic" log-likelihood is:

$$\log \mathcal{L}(\beta) = \sum_{i=1}^{N}\{y_i \log F(\mathbf{x}_i'\beta) + (1-y_i)\log(1-F(\mathbf{x}_i'\beta))\}.$$

- MLE:

$$\frac{\partial \log \mathcal{L}}{\partial \beta} = \sum_{i=1}^{N}\left\{\frac{y_i}{F_i}F_i'\mathbf{x}_i - \frac{1-y_i}{1-F_i}F_i'\mathbf{x}_i\right\} = 0 \qquad (7)$$

Likelihood

- Often with binary response, we center on logit or probit.
- Log-likelihood redux:

$$\log \mathcal{L}(\beta) = \sum_{i=1}^{N} \{y_i \log \hat{p}_i + (1 - y_i) \log \hat{p}_i\}.$$

- For logit, $\hat{p}_i = \Lambda(\mathbf{x}_i' \hat{\beta}_{\text{Logit}})$
- For probit, $\hat{p}_i = \Phi(\mathbf{x}_i' \hat{\beta}_{\text{Probit}})$
- Λ and Φ were previously defined.

Likelihood

- First order conditions (logit):

$$\sum_{i=1}^{N}(y_i - \Lambda(\mathbf{x}_i'\beta))\mathbf{x}_i = 0$$

- First order conditions (probit):

$$\sum_{i=1}^{N} w_i(y_i - \Phi(\mathbf{x}_i'\beta))\mathbf{x}_i = 0$$

- w_i is a weighting factor necessary for probit.

Next Time

- Interpretation and more on likelihood.
- Goodness-of-fit.