MLE Summary

Wonjae

1. What is MLE?

: The ML estimate is that value of the parameter that makes the observed data most likely. The maximum occurs when **the derivative** of the likelihood function equals **0** (the tangent on the top of the curve with slope 0). The value that maximizes the likelihood function also maximizes **the log** of the likelihood.

That is, $\boldsymbol{\partial \ln L(\pi | s = ?, N = ??)/\partial\pi = 0}$

(**N**: a sample size *s*: a particular outcome *π*: the probability of success
ϕ: a standardized normal variable with mean 0 and variance 1
ϒ: the threshold or cutpoint)

- The **properties** of ML estimators: *consistent*, *asymptotically* (as the sample size approaches infinity) *efficient, asymptotically normally distributed*

2. When to use: When the dependent variable is censored, truncated, binary, ordinal, nominal or count.

- Binary variable→ “Binary logit and probit model”,“The linear probability model”
- Ordered variable→ “The ordered logit and probit model”
- Nominal variable→ “Mutinomial logit “
- ***Censored*** variable— When the value of a variable is unknown over some range of the variable → “The tobit model”
- ***Count*** variable—Indicating the number of times that some event has occurred. → “Poisson and Negative binomial regression model”
  “ebb model (Palmquist)”

➔ **Important similarities** among the models:

a. *Nonlinear* models—The effect if a unit change in a variable (x) depends on the values of all variables in the model and is no longer simply equal to a parameter of the model.
b. The same *systematic component*—the independent variables as a linear combination
   (nonlinearities can be introduced by transforming the independent variables)
c. Estimated by maximum likelihood
d. The same general ideas used for interpreting each model—“*expected values*”, “*marginal effects*”, “*discrete changes*”
   ((In nonlinear models, the partial change (derivative) does not necessarily equal to the discrete change.))

3. Binary Outcomes: The Linear Probability, Probit, and Logit Models

- Problems with the Linear Probability Model: *Heteroscedasticity, Normality, Nonsensical Predictions,* ***Functional Form*** (the effects of independent variables will have diminishing returns as the predicted probability approaches 0 or 1.)
- In the probit model, the variance is 1. In the logit model, the variance is $\pi^2/3 \cong 3.29$
- The probability of observing an event given x is *the cumulative density* evaluated at XB.
- "**Identification**" is an issue that is essential for understanding models with *latent* variables.
- The **probability** of an event is **unaffected** by the identifying assumption regarding variance. The probabilities can be interpreted **without concern about the arbitrary assumption** that is made to identify the model.
- **Interpretation** using 4 methods

  A. Predicted Probabilities:

  a. Examining the range of predicted probabilities within the sample

  b. Determining the extent to which change in a variable affects the predicted Probability

  In STATA) **"probpred Y X1, from(min) to(max) adj(x2 x3,,,*) logit*"**

  "probpred Y X1, from(min) to(max) adj(x2 x3,,,*)* "

  "probpred Y X1, from(min) to(max) adj(x2 x3,,,) zero(x2) one(x3) *logit*"

  "probpred Y X1, from(min) to(max) adj(x2 x3,,,) zero(x2) one(x3) "

  B. Partial change in y* and ***in the probability of an event given x (marginal effect)***

  "dprobit Y X1 X2 X3, robust"

  → It gives you the estimated effect of an independent variable at the mean values of the covariates and the discrete change of dummy variable from 0 to 1.

"matrix x= (2, 2, 3, 3, *1*, 2)"

"matrix colnames x = X1 X2 X3 X4 X5 X6"

"dprobit Y X1 X2 X3 X4 X5 X6, at(x)"

→ If we vary X5 from 1 to 2,

"matrix x= (2, 2, 3, 3, *2*, 2)"

"matrix colnames x = X1 X2 X3 X4 X5 X6"

"dprobit Y X1 X2 X3 X4 X5 X6, at(x)"

C. Discrete change in the probability

(to summarize the effects if each variable → We can get discrete change by using "probpred" commands above. Or, in the "dprobit" command, the predicted probability is given automatically.)

D. Using Odds ratios

4. *Multinomial outcomes*

- **Multinomial Logit Model**

.mlogit Y X1 X2 X3 X4, base(0)

- **lrtest** – This test is for choosing appropriate variables. That is, we can test a model including some variables and test another model excluding those variables.

.mlogit Y X1 *X2 X3* X4 X5 X6, base(0)

.lrtest, saving(0)

.mlogit Y X1 X4 X5 X6, base(0)

→ If the result of likelihood-ratio test is significant, we can choose the later model excluding X2 X3 variables. However, even though the result is NOT significant, we can prefer the constraint model, if the X2 and X3 variables are not significant or you have a theoretically strong reason to exclude those variables.

- **Wald test**: This is for testing significance of individual variable in both models.

.test X2

.test X3

→ If the test is significant for a variable, the variable can be included in the constraint model.

- *Marginal Effects*

: For the purpose of illustrating the effects of the variables, we obtain the marginal effects by fixing variables on some values. And, by varying a variable, we can obtain predict marginal effects indifferent points

.matrix x= (2, 2, 3, 3, *1*, 2)

.matrix colnames x = X1 X2 X3 X4 X5 X6

.dmlogit2 Y X1 X2 X3 X4 X5 X6, at(x) base(0)

→ If we vary X5 from 1 to 2,

.matrix x= (2, 2, 3, 3, *2*, 2)

.matrix colnames x = X1 X2 X3 X4 X5 X6

.dmlogit2 Y X1 X2 X3 X4 X5 X6, at(x) base(0)

Cf. You must install the command file (dmlogit) first from "help" manual in order to get this result.

- There is a way to get the "marginal effect" (that is partial change), factor change, discrete change (that is "first difference"), and percentage change.

  → Scott Long's "listcoef" and "prchange" (or "prval") commands provide an easy way to make these calculations.

- You need to install the files first.

  → Go to http://www.indiana.edu/~jsl650/stat.htm and install "cdaado.pkg" file

- Ex: In Stata,

  .nbreg Y X1 X2 X3 X4 X5, nolog

  .listcoef

  (.set matsize 500)

  .prvalue, x(X1 1 X2 1 X3 5) rest(mean)

  . prchange, x(X1 1 X2 1 X3 1) rest(mean)

- **Multinomial Probit Model**

- The best way of running multinomial probit model is to use a Bayesian analysis with marginal data augmentation method (Imai & Dyk 2004). You can find documentation and other information from the following web site.

  http://www.princeton.edu/~kimai/research/MNP.html

- You have to download "R" program first and then install "MNP" packages into R by typing: install.packages("MNP")

- *Predicted Probability*

: There are a few ways to obtain predicted probability: By hand (see "Long, Scott's Regression Models for Categorical and Limited Dependent Variables, pp.164-165"), By inputting some commands in STATA, or By using CLARIFY file (http://Gking.Harvard.Edu). Since the simplest way is to use Clarify file, here I will explain it.

Cf. The Clarify package contains eight main files. Thus, you have all eight before installing the software. To install Clarify under Stata 6.0, launch Stata and they type "sysdir" at the command prompt. And copy all files you need in an appropriate directory.

.clear

.set mem 10m → In order to increase memory span

.estsimp mlogit (or oprobit) Y X1 X2 X3 X4 X5 X6 X7, base(3)

→ When Y has 3 different values and you want to fix base on 3

.setx X1 mean X2 median X3 min X4 max X5 p25 X6 ln(20) X7 2.5

→ When to set X1 at it mean, X2 at its median, X3 at its minimum, X4 at its maximum, X5 at its 25th percentile, X6 at ln(20), X7 at 2.5

.simqi prval(1 2 3) → for the probabilities that the dependent variable takes on a values of 1, 2, or 3

.simqi, fd(prval (1 2 3)) changex(x7 2.5 3.5) → For the changes in Pr(Y = 1, 2, 3) caused by raising X7 from 2.5 to 3.5 when other variables are held at their values

Cf. In order to get the mode of a variable (X4), type:

.modes X4

- **Definition**

"***Marginal Effect***"(Partial derivative)—The ratio of the change in y to the change in x when the change in x is infinitely small holding all other variables constant (Scott, p5) → Thus, the value of the marginal effect depends on the *values* of all independent variables and on *the coefficients* of each outcome.

→ For **continuous** independent variables!

"***Discrete Changes***"— The change in the predicted probability when Xk changes from Xs (the starting value) to XE (the ending value), holding all other variables constant.

→ The amount and even the direction of the discrete change depends on the values at which the independent variables are being held constant.

((Limitations: At different levels of the variables, the changes will be different. Measures of discrete change do not indicate the dynamics among the dependent outcomes))

→ For **continuous and dummy** independent variables!

“***Fixed Effects***”**—**This approach takes α (intercept) to be a group specific constant term in the regression model.

“***Random Effects***”**—**This approach specifies that α (intercept) is a group specific disturbance, similar to disturbance except that for each group there is but a single draw that enters the regression identically in each period.

→ The fixed effects model is a reasonable approach when we can be confident that the differences between units can be viewed as parametric shifts of the regression function. In other settings it might be more appropriate to view individual specific constant terms as randomly distributed across cross-sectional units.