5 Maximum Likelihood Estimation

A sample of N decision makers is obtained for the purpose of estimation.

5.1 Log Likelihood function

Since the logit probabilities take a closed form, the traditional maximum-likelihood procedures can be applied. Show that the log likelihood function is then

$$LL(\beta) = \sum_{n=1}^{N} \sum_{i} y_{ni} \ln P_{ni} \qquad (2)$$

$$\mathcal{L}(\beta) = f(Y|X,\beta) \overset{iid}{=} \prod_{n=1}^{N} \prod_{i=1}^{J} P_{ni}^{y_{ni}}$$

if individual n chose i, the P_{ni} is used
" " " " $j \neq i$, " " is not used. ($P_{ni}^0 = 1$)

$$\mathcal{LL}(\beta) = \sum_{n=1}^{N} \sum_{i=1}^{J} y_{ni} \ln P_{ni}$$

5.2 First order condition

(H) Show that the first order condition for the problem of $\max_\beta LL(\beta)$ is given by:

$$\sum_{n=1}^{N} \sum_{i} (y_{ni} - P_{ni}) x_{ni} = 0 \qquad (3)$$

What is the interpretation of this equation? Show that the maximum likelihood estimates of β are those that make the predicted average of each explanatory variable equal to the observed average in the sample.

$$\mathcal{LL}(\beta) = \sum_{n=1}^{N} \sum_{i=1}^{J} y_{ni} \ln P_{ni} \xrightarrow{\nabla_\beta} \sum_{n=1}^{N} \sum_{i=1}^{J} y_{ni} \nabla_\beta \ln P_{ni} = 0$$

$$\sum_{i=1}^{J} P_{ni} = 0 \rightarrow \sum_{i=1}^{J} \nabla_\beta P_{ni} = 0 \rightarrow \sum_{i=1}^{J} P_{ni} \nabla_\beta \ln P_{ni} = 0$$

$$\boxed{\sum_{n=1}^{N} \sum_{i=1}^{J} (y_{ni} - P_{ni}) \nabla_\beta \ln P_{ni} = 0} \quad (1)$$

General Discrete Choice

$$P_{ni} = \frac{e^{x_{ni}'\beta}}{\sum_j e^{x_{nj}'\beta}} \longrightarrow \nabla_\beta P_{ni} = x_{ni} P_{ni} - \frac{P_{ni}}{\sum_j e^{x_{nj}'\beta}} \sum_j x_{nj} e^{x_{nj}'\beta} \longrightarrow \nabla_\beta \ln P_{ni} = x_{ni} - \frac{\sum_{j=1}^{J} x_{nj} e^{x_{nj}'\beta}}{\sum_{j=1}^{J} e^{x_{nj}'\beta}} \quad (2)$$

$$(1)(2) \rightarrow \sum_{n=1}^{N} \sum_{i=1}^{J} (y_{ni} - P_{ni}) \left(x_{ni} - \frac{\sum_{j=1}^{J} x_{nj} e^{x_{nj}'\beta}}{\sum_{j=1}^{J} e^{x_{nj}'\beta}} \right) = \sum_{n=1}^{N} \sum_{i=1}^{J} (y_{ni} - P_{ni}) \left(x_{ni} - \sum_{j=1}^{J} x_{nj} P_{nj} \right) = 0 \quad (3)$$

$$\sum_{n=1}^{N} \sum_{i=1}^{J} (y_{ni} - P_{ni}) \sum_{j=1}^{J} x_{nj} P_{nj} = \sum_{n=1}^{N} \sum_{j=1}^{J} x_{nj} P_{nj} \underbrace{\sum_{i=1}^{J} (y_{ni} - P_{ni})}_{\sum_i y_{ni} = 1,\ \sum_i P_{ni} = 1} = 0$$

$$\text{Logit} \leftarrow \boxed{\sum_{n=1}^{N} \sum_{i=1}^{J} (y_{ni} - P_{ni}) x_{ni} = 0}$$

The fitted predicted average of each explanatory variable ($\sum\sum P_{ni} x_{ni}$) equals to the sample average of each x_{ni} ($\sum_n \sum_i y_{ni} x_{ni}$).

So MLE makes these averages equal.

5.3 Goodness of fit

The likelihood ratio index is defined as

$$\rho = 1 - \frac{LL(\hat{\beta})}{LL(0)}$$

In two sentences interpret this index. Is it similar to R^2 in linear regression? For instance, can you compare likelihood ratio index from two models on two datasets to each other? (Hint: the answer is NO! Explain why.)

$\mathcal{L}(\hat{\beta}) < 1 \rightarrow LL(\hat{\beta}) < \log(1) = 0$

$L(0) < 1 \rightarrow L(0) < \log(1) = 0$

if $\hat{\beta}$ is better fit than 0 $\Leftrightarrow L(\hat{\beta}) > L(0) \Leftrightarrow LL(\hat{\beta}) > LL(0) \xleftrightarrow{LL(0)<0} \frac{LL(\hat{\beta})}{LL(0)} < 1$

" 0 " " " " $\hat{\beta}$ $\Leftrightarrow$ " $<$ " $\Leftrightarrow$ " $<$ " $\xleftrightarrow{''}$ $\frac{LL(\hat{\beta})}{LL(0)} > 1$

if $\hat{\beta}$ is better fit than 0 $\longleftrightarrow 1 \geq 1 - \frac{LL(\hat{\beta})}{LL(0)} > 0$

" 0 " " " " $\hat{\beta}$ $\longleftrightarrow 1 - \frac{LL(\hat{\beta})}{LL(0)} < 0$

As $\hat{\beta}$ improves, $1 - \frac{LL(\hat{\beta})}{LL(0)}$ increases & gets closer to 1

this index is a comparison between $\hat{\beta}$ & $\beta = 0$, so comparing 2 $\hat{\beta}$ within same model & data is possible with this index, but not possible between different models or data sets.