TA session2# 1

Jun Sakamoto

October 11,2018

Contents

1 Optimization

Example
We try to apply Newton's method for MLE of binominal distribution.
A coin is thrown 100 times and a table has gone out 52 times. Then what is most suitable for parameter p?
Likelifood function of binominal distribution is

$$\log L = \log({}_n\mathrm{C}_m) + m \log p + (n-m)\log(1-p). \tag{1}$$

First derivative of likelihood function is

$$\frac{m}{p} - \frac{n-m}{1-p}. \tag{2}$$

Also second derivative of likelihood function is

$$-\frac{m(1-p)^2 + (n-m)p^2}{p^2(1-p)^2}. \tag{3}$$

Thus, optimization procedeure is

$$0 = \frac{m}{p} - \frac{n-m}{1-p} - \frac{m(1-p)^2 + (n-m)p^2}{p^2(1-p)^2}(p^* - p). \tag{4}$$

Then $\bar{H}^{-1}f'_p$ is ($\bar{H}$ is a hessian of $\log L$.)

$$-\frac{p(1-p)(m-np)}{m(1-p)^2+(n-m)p^2}. \tag{5}$$

We set a default is $p_0 = 0.1$. Then, computational result are,

$$0.1 \longrightarrow 0.19 \longrightarrow 0.33 \longrightarrow 0.48 \longrightarrow 0.52$$

Also we set a default is $p_0 = 0.8$. Then, computational result are,

$$0.8 \longrightarrow 0.66 \longrightarrow 0.54 \longrightarrow 0.52$$

Newton's method has some problems. We consider the function as bellow.

$$f(x) = x^3 - 3x^2 + x + 3 \tag{6}$$

We set a default is $x = 1$. Then x does not converge and repeat two value. (If you had an interest, please try.) Also calculation value diverge in another case. Thus a cuculation result may not convergent in Newton's method. Also Even if result converge on the value, that isn't always the global extreme value.

2 Descreat choice model

Linear probability model

Dependent variable y is binary values as below.

$$y_i = \begin{cases} 1 & w.p.P_i \\ 0 & w.p.1 - P_i \end{cases} \tag{7}$$

Next we consider a regression model when data are binary.

$$y_i = x_i'\beta + u_i \tag{8}$$

The expectation of y_i is given by this

$$E[y_i] = x_i'\beta \tag{9}$$

Thus,$E[y_i] = P_i$

This model has two difficulties. First, u_i is not normally distributed. Second P_i might not be in [0,1].To overcome these difficulties, we have to use non linear probability models.

Logit model

If we use logistic distribution for $F()$, we call this model logistic model or logistic regression amd formulated as follows

$$P_i = E[y_i] = \frac{1}{1 + \exp^{-x_i'\beta}}. \tag{10}$$

Similarly

$$1 - P_i = \frac{\exp^{-x_i'\beta}}{1 + \exp^{-x_i'\beta}} \tag{11}$$

Taking the ratio P_i to $1 - P_i$, we get

$$\frac{P_i}{1 - P_i} = \exp^{x_i'\beta}$$
$$\log \frac{P_i}{1 - P_i} = x_i'\beta \tag{12}$$

and we call this odds ratio. If we can observe P_i directry, we estimate β by using following regressiojn model

$$\log \frac{P_i}{1 - P_i} = x_i'\beta + u_i \tag{13}$$

If use normal distribution for $F()$, the we call this model probit model.

Estimation

It is rare to observe P_i directly. Therefore, it is more realistic to estimate β by using maximum likelihood method. Suppose that error term are independent. Then the likelihood function for above model is given by this

$$L = \prod_i (1 - F(-x_i'\beta))^{y_i} F(-x_i'\beta)^{(1-y_i)} \tag{14}$$

In the case of logit model, l could be given by

$$\log L = \sum_i (1 - y_i) \log(\frac{1}{1 + \exp(x_i'\beta)}) + \sum_i y_i \log(\frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)}). \tag{15}$$

The FOC is

$$\sum_i \frac{\exp(x_i'\beta)}{(1 + \exp(x_i'\beta))^2} x_i (y_i - F(x_i\beta))(\frac{\exp(x_i'\beta)}{(1 + \exp(x_i'\beta))^2})^{-1} = 0 \tag{16}$$

Empirical Example

"An empirical analysis of a firm ' s decision to franchise" Pin.Minkler, 1990, Economic letters,34 p77-82 Hypotheses from monitoring and search cost explanations for the existence of franchising are discussed and tested.

Ownership in franchising is utilized the notion of costly monitoring. The first hypothesis is that ownership tenure is related to distance form monitoring headquarters. Outlets close to monitoring headquarters are relatively cheap to monitor and thus are more likely to be companyowned.

The second monitoring hypothesis suggests that locations along freeways are more likely to be company-owned [Klein (1980)]. Outlets which have a high percentage of non-repeat customers are more likely to be company-owned to reduce the cost of cheating by store operators.

The last hypothesis is that ownership tenure is related to the density of outlets. Clusters imply that monitoring costs per individual outlet are low because a monitor can travel among dense locations cheaply. Hence, outlets in clusters are more likely to be company-owned, according to the monitoring theory.

The following model was estimated using three different techniques (1) linear probability, (2) logit, and (3) probit

$$Tenure = B_0 + B_1 AGE + B_2 DIST + B_3 CLUST + B_4 HWY + e \tag{17}$$

where
Tenure = 1 if outlet franchised, 0 if not
AGE = number of years to outlet ' s opening from first outlet ' s opening (1965)
DIST = distance of outlet from monitoring headquarters
CUST = number of outlets within five miles
HWY = 1 if outlet within one mile of highway, 0 if not.

Furthermore, they estimated same equation using subsample of Sacrament.

Table 2
Full sample estimates. [a]

Variable	Linear probability [b]	Logit	Probit
AGE	−0.0087 **	−0.1043 **	−0.0507 **
	(−2.3996)	(−2.3248)	(−2.2941)
DIST	0.0008 **	0.0135 **	0.0068 **
	(2.0069)	(2.1759)	(2.1517)
CLUST	0.0054	0.0882	0.0520
	(0.7842)	(1.1353)	(1.2376)
HWY	−0.0415	−0.5954	−0.2287
	(−0.8055)	(−1.0589)	(−0.7815)
Intercept	0.9063 ***	2.3731 **	1.2222 **
	(10.6320)	(2.2678)	(2.3442)
N	154	154	154
R^2	−0.0790	0.1374	0.1323
% Predicted correctly	–	88 %	88 %

[a] t-statistics are in parentheses. * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level. McFadden R^2s are reported for the logit and probit estimates to allow comparability with the linear probability model.
[b] 14 predicted values were outside the [0,1] interval.

Figure 1:

Table 3
Sacramento estimates. [a]

Variable	Linear probability [b]	Logit	Probit
AGE	0.0272 **	−0.2075 *	−0.1230 **
	(−2.2225)	(−1.9446)	(−2.0365)
DIST	0.0087	0.1160	0.0697
	(0.5672)	(0.9483)	(0.9572)
CLUST	0.0111	0.2106	0.1153
	(0.4231)	(0.9914)	(0.9242)
HWY	−0.2468 *	−2.1787 *	−1.2703 *
	(−1.7952)	(−1.8693)	(−1.9213)
Intercept	0.9996 **	3.0738	1.8797
	(2.4891)	(1.1092)	(1.1514)
N	34	34	34
R^2	0.2271	0.2861	0.2955
% Predicted Correctly	–	82 %	76 %

[a] t-statistics are in parentheses. * significant at the 10% level, ** significant at the 5% level, *** 'significant at the 1% level. McFadden R^2s are reported for the logit and probit estimates to allow comparability with the linear probability model.
[b] 4 predicted values were outside the [0,1] interval.

[5] A Chow test was performed to see if the Sacramento and non-Sacramento data can be pooled. The F-statistic with 5 and 144 degrees of freedom is 1.846., significant at the 11% level.

Figure 2: