Microeconometrics Tutorial 7

Karolin Süß

July 2, 2019

Exercise 1: Maximum Likelihood Estimation: Bernoulli Case

(a)

Let us define $y = \{0,1\}$, $\theta \in [0,1]$, $y_i \backsim B(\theta)$ with y_i and y_j independent. We can plug in $f(y_i|\theta)$ into the log-likelihood function and rewrite:

$$\begin{aligned} Q_n(\theta) &= \frac{1}{n} \log \left(\prod_{i=1}^{n} \theta^{y_i} (1-\theta)^{(1-y_i)} \right) \\ &= \frac{1}{n} \sum_{i=1}^{n} \left(\left(\log \left(\theta^{y_i} (1-\theta)^{1-y_i} \right) \right) \right. \\ &= \frac{1}{n} \sum_{i=1}^{n} \left(y_i \log(\theta) + (1-y_i) \log(1-\theta) \right) \end{aligned}$$

Now use $m = \sum_{i=1}^{n} y_i$:

$$Q_n(\theta, n, m) = \frac{1}{n} \left[m \log(\theta) + (n-m) \log(1-\theta) \right]$$

(b)

Maximize Q_n with respect to θ to receive $\hat{\theta}$:

$$\begin{aligned} && \frac{\partial Q_n}{\partial \theta} &= \frac{1}{n} \left[\frac{m}{\theta} - \frac{n-m}{1-\theta} \right] \stackrel{!}{=} 0 \\ \Leftrightarrow && \frac{m}{\theta} &= \frac{n-m}{1-\theta} \\ \Leftrightarrow && m(1-\theta) &= (n-m)\theta \\ \Leftrightarrow && m &= n\theta \\ \Leftrightarrow && \hat{\theta}(n,m) &= \frac{m}{n} \end{aligned}$$

Check for maximum:

$$\frac{\partial^2 Q_n}{\partial \theta^2} = \frac{1}{n} (\underbrace{-\frac{m}{\theta^2}}_{(-)} \underbrace{- \frac{n-m}{(1-\theta)^2}}_{(-)}) < 0$$

(c)

As mentioned in the hint, we want to derive the Hessian matrix $H(\theta)$ from $f(y_i|\theta)$. We get as first derivative:

$$\begin{aligned}\frac{\partial \log\left(\theta^{y_i}(1-\theta)^{(1-y_i)}\right)}{\partial\theta} &= \frac{\partial\left(y_i\log(\theta) + (1-y_i)\log(1-\theta)\right)}{\partial\theta} \\ &= \frac{y_i}{\theta} - \frac{1-y_i}{1-\theta}\end{aligned}$$

We can take the second derivative to obtain $H(\theta)$:

$$H(\theta) = -\frac{y_i}{\theta^2} - \frac{1-y_i}{(1-\theta)^2}$$

Taking the expected value of the Hessian we receive:

$$\begin{aligned}\mathbb{E}[H(\theta)] &= \mathbb{E}[-\frac{y_i}{\theta^2} - \frac{1-y_i}{(1-\theta)^2}] \\ &= -\frac{\theta}{\theta^2} - \frac{1-\theta}{(1-\theta)^2} \\ &= -\frac{1}{\theta} - \frac{1}{1-\theta} \\ &= -\frac{1}{\theta(1-\theta)}\end{aligned}$$

Plug in $\hat{\theta}$ from question **(b)**:

$$\begin{aligned}\mathbb{E}[H(\hat{\theta})] &= -\frac{1}{\frac{m}{n}(1-\frac{m}{n})} \\ &= -\frac{n^2}{m(n-m)}\end{aligned}$$

An estimator of the variance of $\hat{\theta}$ is then given by the negative of the inverse of the second derivative. Thus:

$$V(\hat{\theta}) = \frac{m(n-m)}{n^2}$$

The finite sample variance is given by $\frac{1}{n}V(\hat{\theta})$ which is analogously derived from

$$H(\theta|y) = -\frac{\sum_{i=1}^n y_i}{\theta^2} - \frac{\sum_{i=1}^n (1-y_i)}{(1-\theta)^2}$$

Exercise 2

(a)

The log-likelihood function is given by

$$\begin{aligned}\mathcal{L}_n(\beta) &= \Sigma_{i=1}^{n}\{y_i \log(F(x_i'\beta)) + (1-y_i)\log(1-F(x_i'\beta))\} \\ &= \Sigma_{i=1}^{n}\{y_i \log(\Phi(x_i'\beta)) + (1-y_i)\log(1-\Phi(x_i'\beta))\}\end{aligned}$$

Taking the first derivative w.r.t β yields the FOC

$$\sum_{i=1}^{N} \frac{(y_i - \Phi(x_i'\hat{\beta}))}{\Phi(x_i'\hat{\beta})\ (1-\Phi(x_i'\hat{\beta}))}\phi(x_i'\hat{\beta})x_i \overset{!}{=} 0$$

Thus, the MLE estimator $\hat{\beta}$ can be solved as a function of known quantities x_i and y_i.
In order to ensure uniqueness of the MLE estimator (and identification), global concavity of the likelihood function w.r.t. β is pivotal. Generally, concavity can be proved by evaluating the Hessian matrix of a log-likelihood function. If the determinant of the Hessian is negative definite, the likelihood function is strictly concave granting identification of β.
To evade the cumbersome calculation of the Hessian, we can instead evaluate the asymptotic variance matrix, which is just the negative inverse of the Hessian (assuming it is non-singular) and is provided by Cameron & Trivedi (2005, p. 469) and the lecture (chapter 7, slide 19):

$$\begin{aligned}\hat{V}\left[\hat{\beta}_{ML}\right] &= \left(\sum_{i=1}^{N} \frac{(F'(x_i'\hat{\beta}))^2\ x_i x_i'}{F(x_i'\hat{\beta})\ (1-F(x_i'\hat{\beta}))}\right)^{-1} \\ &= \left(\sum_{i=1}^{N} \frac{(\phi(x_i'\hat{\beta}))^2\ x_i x_i'}{\Phi(x_i'\hat{\beta})\ (1-\Phi(x_i'\hat{\beta}))}\right)^{-1}\end{aligned}$$

Please note that in matrix algebra, the corresponding expression to $\sum_{i=1}^{N} x_i x_i'$ actually is $X'X$. In order to obtain a positive definite variance matrix (resulting in a negative definite Hessian), $\sum_{i=1}^{N} x_i x_i'$ clearly needs to be positive definite as well. The remaining expressions consist of density and distribution functions. By the model assumptions, theses functions take values within the open (0,1) interval (if not, we might not be able to observe a binary outcome, it would be singular instead).
As $\sum_{i=1}^{N} x_i x_i'$ is symmetric, we know that it has to be positive semidefinite. In the univariate case we would have $\sum_{i=1}^{N} x_i^2$ which is clearly positive. By further assuming invertibility, the case of $det(\sum_{i=1}^{N} x_i x_i') = 0$ is ruled out. Hence what remains is a positive definite variance matrix converting into a negative definite Hessian. This again grants global concavity and uniqueness.

(b)

Use the known (or estimated) σ_ϵ to normalize observations:

$$\begin{aligned}P(Y=1 \mid x) &= P(y^* > 0 \mid x) \\ &= P(x'\beta + \epsilon > 0 \mid x)\end{aligned}$$

$$
\begin{aligned}
&= P(\frac{x'\beta + \epsilon}{\sigma_\epsilon} > 0 \mid x) \\
&= P(\frac{x'\beta}{\sigma_\epsilon} > -\frac{\epsilon}{\sigma_\epsilon} \mid x) \\
&= F(\frac{x'\beta}{\sigma_\epsilon}) \\
&= \Phi(\frac{x'\beta}{\sigma_\epsilon}) \quad \text{since } \frac{\epsilon}{\sigma_\epsilon} \sim \mathcal{N}(0,1).
\end{aligned}
$$

Exercise 3

(a)

I follow the slides p.13 here. We have:

$$P(y = 1|z, q) = \Phi(z_1\delta_1 + \gamma_1 z_2 q) \tag{1}$$

An equivalent threshold-crossing model would be:

$$y^* = z_1\delta_1 + \gamma_1 z_2 q + u \text{ with } u \sim \mathcal{N}(0,1),\ \ y = 1[y^* > 0] \tag{2}$$

Proof:

$$
\begin{aligned}
P(y = 1|z, q) &= P(z_1\delta_1 + \gamma_1 z_2 q + u > 0) \\
&= P(z_1\delta_1 + \gamma_1 z_2 q > -u) \\
&= \Phi(z_1\delta_1 + \gamma_1 z_2 q)
\end{aligned}
$$

(b)

Marginal effect:

$$\frac{\partial\Phi(z_1\delta_1 + \gamma_1 z_2 q)}{\partial z_2} = \phi(z_1\delta_1 + \gamma_1 z_2 q)\gamma_1 q, \tag{3}$$

where ϕ is the standard normal density function. The magnitude of the marginal effect varies with z_1, z_2 and q. The sign of the effect, however, depends only on q. As $\Phi(\cdot)$ is strictly increasing in z_1, z_2 and q, $\phi(\cdot) > 0$ for all z_1, z_2 and q.

One approach to evaluate marginal effects would be to estimate the sample average of the marginal effect (AME). Thus, here we would estimate the marginal effects for every possible values of z_1, z_2 and q and then take the average of this. On the other hand, we could estimate the marginal effect at the sample average of the regressors (MEM). Here, we would evaluate the marginal effects at the mean of z_1, z_2 and q. Finally, we could use the median, minimum, maximum or any other value of z_1, z_2 and q to estimate the marginal effect at a representative point (MER).

(c)

$$\begin{aligned} P(y=1|z,q) &= P(z_1\delta_1 + \gamma_1 z_2 q + u > 0 \mid z, q) \\ &= P(z_1\delta_1 > -(u + \gamma_1 z_2 q) \mid z, q) \\ &= P(z_1\delta_1 > -\epsilon \mid z, q) \end{aligned}$$

Using that q is not known, standard normal and independent of z_1 and z_2, it follows that

$$\begin{aligned} \sigma_\epsilon^2 &= 1 + \gamma_1^2 z_2^2 \\ \epsilon &\sim \mathcal{N}\left(0, 1 + \gamma_1^2 z_2^2\right) \end{aligned}$$

Since $\frac{\epsilon}{\sigma_\epsilon} \sim \mathcal{N}(0,1)$, the conditional probability can then be rewritten analogously to Task 2b):

$$\begin{aligned} P(y=1 \mid z,q) &= P(z_1\delta_1 > -\epsilon \mid z) \\ &= P(\frac{z_1\delta_1}{\sigma_\epsilon} > -\frac{\epsilon}{\sigma_\epsilon} \mid z) \\ &= \Phi(\frac{z_1\delta_1}{\sigma_\epsilon}) \\ &= \Phi\left(\frac{z_1\delta_1}{\sqrt{1+\gamma_1^2 z_2^2}}\right) \end{aligned}$$

Exercise 4: Applied Problem

(a)

The regression results of the Linear Probability Model (LPM) indicate that an additional year of marriage significantly increases the probability of having an affair by roughly 0.9 percentage points.
At $yrsmarr = 1$ and $relig = 4$, a person that is very happy married is roughly 34.7 percentage points more likely to have an affair than someone who is very unhappy married. Note that the levels of $yrsmarr$ and $relig$ do not play a role in the effect, because they difference out in a linear model.
The predicted probability of having an affair for someone who is very happy married, very religious and married for one year amounts to approximately -0.02.

(b)

The LPM does not constrain the predicted probabilities to lie between zero and one which can be seen in the results of Task (a) (negative predicted probability). It ignores the discreteness of the dependent variable and instead treats it as continuous. Thus, the model generally does not fit the CEF perfectly, but it is the best linear predictor. Specifically, the LPM may be preferable if the interest is in analyzing partial effects averaged over the distribution of x. Then, the fact that some predicted values are outside the unit interval may not be important. It might further be preferable if the regressors take on only a few values because then the CEF is close to a linear function of the regressors. In the extreme case when the model is saturated (x contains dummy variables for mutually exclusive and exhaustive categories), the CEF is a linear function.

(c)

Since logit and probit coefficients do not directly measure marginal effects, it is hard to interpret their magnitude. However, signs of the coefficients can be interpreted. The regression results indicate that an increase in the years of marriage increases the probability of having an affair. Being religious and being happy in the marriage have a negative effect on the likeliness of having an affair. Further, the Stata output tells us that the coefficients are significantly different from zero (on a 1%-level).

(d)

At $yrsmarr = 1$ and $relig = 4$, the probit model results indicate that someone who is very happy married is roughly 28.6 percentage points more likely to have an affair than someone who is very unhappy married (logit 27.5 percentage points). Both coefficients are lower than the ones estimated with the LPM (see Task (a)).

(e)

Theoretically, the choice of the model depends on the data generating process, which is not known. Practically,n coefficients in the logit model are simpler to interpret while the latent framework makes the probit model attractive in selection models. Logit has the advantage of generating coefficients that can be interpreted as the marginal effects on the odds ratio (ratio of successes to failures). Empirically, the difference is typically very small when comparing fitted log-likelihood from the different specifications which can also be seen in the results of the given example.

(f)

	Logit	Probit
AME	0.0093967	0.0094125
MEM	0.0097526	0.0098861

The marginal effects of $yrsmarr$ differ only slightly between the two models.