Exercise 1: Maximum Likelihood Estimation: Bernoulli Case

Suppose that y_i takes values 0 and 1 and is Bernoulli distributed with mean $\theta \in [0,1]$, i.e.

$$P(y_i = 1) = \theta \qquad\qquad P(y_i = 0) = 1 - \theta$$

We observe a random sample y_i, $i = 1, ..., n$. In our sample we observe m times the outcome $y_i = 1$, and $(n - m)$ times the outcome $y_i = 0$. The log-likelihood function of the model is given by $Q_n(\theta) = \frac{1}{n} log \prod_{i=1}^{n} f(y_i|\theta)$ wherer $f(y_i|\theta) = \theta^{y_i}(1-\theta)^{1-y_i}$.

(a) Show that $Q_n(\theta)$ can be expressed as a function of n, m and θ only.

(b) Calculate the MLE $\hat{\theta}$ (this should be a function of n and m only).

(c) Calculate an estimator for the variance of $\hat{\theta}$. This should be a function of n and m only. (Hint: Derive the Hessian $H(y_i, \theta) = \frac{\partial^2 \log f(y_i|\theta)}{\partial \theta^2}$ in a first step. In a second, calcuate the expeted value of the Hessian which should be a function of θ only. Now plug into this expression the estimator for θ from (b).)

Exercise 2

Consider the binary probit model as discussed in the lecture.

$$Y^* = X\beta_0 + \epsilon$$
$$Y = 1[Y^* \geq 0]$$

where ϵ is independent of X, and $\epsilon \sim N(0,1)$. Thus

$$p(x, \beta_0) \equiv P(Y = 1|x) = \Phi(x\beta_0).$$

Suppose that the joint distribution of Y and X is known, which implies knowledge of $P(Y = 1|x)$ for any x, and $E[X'X]$.

(a) Show that if $E[X'X]$ is invertible, β_0 can be solved as a function of known quantities. Thus β_0 is identified.

(b) Suppose now that the variance of ϵ is denoted σ_ϵ^2, not necessarily equal to 1, so that $\epsilon \sim N(0, \sigma_\epsilon^2)$. Derive $P(Y = 1|x)$ as a function of x and β for $\beta_0 = \beta$.

Exercise 3

Now consider the following probit model

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

where q is independent of $z=[z_1,z_2]$ and distributed as $N(0,1)$, the vector z is observed but the scalar q is not.

(a) Write the model as an equivalent threshold-crossing model.

(b) Derive the partial effect of z_2 on the response probability, namley,

$$\frac{\partial P(y=1|z,q)}{\partial z_2}$$

Does the partial effect vary with z_1, z_2 and q? Describe three common approaches how to evaluate the partial effect.

(c) Show that

$$P(y=1|z)=\Phi\left(\frac{z_1\delta_1}{(1+\gamma_1^2 z_2^2)^{\frac{1}{2}}}\right)$$

Exercise 4: Applied problem

For this exercise you want to download the following data set provided by Wooldridge for teaching purposes: `http://fmwww.bc.edu/ec-p/data/wooldridge/affairs.dta`. A definition of each variable can be found here: `http://fmwww.bc.edu/ec-p/data/wooldridge/affairs.des`.

(a) Estimate a linear probability model (LPM) using *affair* as your binary outcome and the variables *yrsmarr, relig, ratemarr* as controls. What is the marginal effect of an additional year of marriage on the probability of having an affair? At *yrsmarr*= 1 and *relig*= 4, what is the estimated difference in the probability of having an affair for someone who is very happy married compared to someone who is very unhappy married? What is the predicted probability of having an affair for someone who is very happy married, very religious and married for one year?

(b) Discuss the advantages and disadvantages of the LPM? If possible refer to your results of exercise (a).

(c) Apply Logit and Probit regressions using the commands logit and probit. Interpret the coefficients. What can they tell us and what not?

(d) Use your Probit (Logit) regression outcomes of excersise (c) and calculate the estimated difference in the probability of having an affair for someone who is very happy married compared to someone who is very unhappy married at *yrsmarr*= 1 and *relig*= 4. How do the estimated probablities compare to each other and to your results of exercise (b).

(e) Discuss the advantages and disadvantages of the Probit (Logit) model.

(f) Use the margins command and estimate the marginal effects of *yrsmarr*. That is calculate AME and MEM.