Exercise 1- MLE

You have a sample $\boldsymbol{y} = \{y_1, ..., y_n\}$ coming from a random variable Y with a vector of parameters $\boldsymbol{\theta} \in R^K$ whose true value is $\boldsymbol{\theta}_0$. You don't know the true value $\boldsymbol{\theta}_0$. Consider that Y follows a Poisson distribution $\mathcal{P}(\lambda)$, i.e. X is a discrete random variable such that

$$P(Y = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

where $\lambda > 0$.

1. What is $\boldsymbol{\theta}$ in the case of Poisson distribution ?
2. Construct the log-likelihood function $log\mathcal{L}(\boldsymbol{\theta})$.
3. What is the maximum likelihood estimator of $\boldsymbol{\theta}$?
4. Compute the second derivative of the log-likelihood function.
5. Noting that $E(Y) = \lambda$, compute $\mathbf{I}(\boldsymbol{\theta}_0) = \mathcal{I}_{y_1,...,y_n}(\boldsymbol{\theta}_0)$.
6. Show that the maximum likelihood estimator of $\boldsymbol{\theta}$ is consistent and efficient.
7. Find the asymptotic distribution of $\boldsymbol{\theta}_{MLE}$.

Exercise 2- Binary Logit

We consider a **trust game** in which there are two players: sender and receiver. The sender decides first and has two possible actions: **trust** and **untrust**.

- If the sender decides to untrust then he and the receiver will each win 10 euros.
- If the sender decides to trust then the payoffs are determined by the receiver. The receiver in this case has two possible actions: **reciprocate** and **unreciprocate**. If the receiver decides to **reciprocate**, then he and the sender will each earn 15 euros. If the receiver decides to **unreciprocate**, then he will earn 22 euros and the sender will earn 8 euros.

The database **binary_choices.dta** contains the data of 100 individuals (indexed $i = 1, 2, ..., 100$) who play the role of sender. The database contains three variables :

- **y**: variable that takes the value 1 if the sender has decided to "trust" and 0 if he has decided to "not trust".
- **r**: variable indicating the risk aversion of the sender
- **p**: variable indicating the (subjective) probability of the sender that the receiver is "benevolent

We consider a latent variable $y_i^* = a_0 + a_1 r_i + a_2 p_i + u_i$ such that :

$$y_i = \begin{cases} 1 & \text{if} \quad y_i^* > 0 \\ 0 & \text{if} \quad y_i^* \leq 0 \end{cases} \tag{1}$$

with u_i a continuous random variable with cumulative distribution function $F(.)$ and density function $f(.)$. Assume in addition that distribution of u_i is symetric around 0, *i.e.* $F(-x) = 1 - F(x)$ and $f(-x) = f(x)$ for all $x \in R$.

1. Show that the probability that sender i decides to **trust** is given by $P(y_i = 1 | r_i, p_i) = F(a_0 + a_1 r_i + a_2 p_i)$.

2. Maximum likelihood estimation of a_0, a_1 and a_2:

   - Show that the log-likelihood on the sample is given by

$$l_N(a_0, a_1, a_2; y|r, p) = \sum_{i=1}^{100} y_i log\big[F(a_0+a_1 r_i+a_2 p_i)\big]+\sum_{i=1}^{100}(1-y_i)log\big[1-F(a_0+a_1 r_i+a_2 p_i)\big]$$

   - How do you find the maximum likelihood estimator of a_0, a_1, and a_2? (Explain or just state the maximization program that yields a_0, a_1, and a_2).

3. Marginal effects

   - Show that the marginal effect of r_i (i.e. risk aversion) on the probability that sender i decides to "trust" is given by:

$$a_1 \times f(a_0 + a_1 r_i + a_2 p_i)$$

   - Explain the concept of *Marginal Effect at the Mean* (MEM)
   - Explain the concept of *Average Marginal Effect* (AME)

4. In question 4, we assume that u_i follows a logistic distribution

   - Make a descriptive statistic of the binary variable **y** and interpret the mean of this variable.
   - Provide the estimated values of a_0, a_1 et a_2 from the logit model
   - Interpret a_1 et a_2
   - Compute the MEM and interpret
   - Compute the AME and interpret

5. Provide the expression and the result of the probability of an individual i to trust when $(r_i, p_i) = (0, 0.5)$

6. How would you evaluate the quality of fit of this model ?

Exercise 3- Ordered Logit

This exercise want to model a variable **health**, which contains individuals' self-assessments of their overall health status on a five-point scale: 1 for "poor", 2 for "fair", 3 for "good", 4 for "very good", and 5 for "excellent". Our model will include three explanatory variables for the mean function: **age**, **femal** (gender), **bmi** (body mass index), and **exercise**. The database **multiple_choices.dta** contains these variables.

1. What kind of model can be used ?
2. Using a latent variable and assuming normal distribution of the error term defined on the latent variable, provide the expression of the probability of each modality of the variable **health** given the explanatory variables.
3. Write the log-likelihood function of the model
4. Provide a list of parameter that can be estimated
5. Marginal effects
   - Provide the expressions of the marginal effect
   - Explain the concept of *Marginal Effect at the Mean* (MEM)
   - Explain the concept of *Average Marginal Effect* (AME)
6. Estimation results
   - Provide the estimation results of the ordered probit model
   - Interpret the coefficient
   - Compute the MEM and AME and interpret
7. Provide the expression and the result of the probability of an individual i to have **good** health status when the age, gender, bmi and exercise are respectively 50, femal, 30 and yes.
8. How would you evaluate the quality of fit of this model ?

Exercise 4- MNL, CL and Nested Logit

1. Present Multinomial logit (MNL) and Conditional logit (CL) as well as their link with random utility models
2. Present and discuss the assumption of *Independence of Irrelevant Alternatives* (IIA) underlying MNL and CL
3. Present Nested Logit (NL)
4. Find a working paper or published paper that uses MNL, CL or NL
   - Provide a brief explanation of the paper and interpret the results of estimation