Advanced Economics Statistics

FALL 2010

Eighth-Assignment Answer Sheet
by Freddy Rojas Cama

1. These exercises show estimators which attain the Cramer-Rao lower bound.

   (i). Let $x_1, ..., x_n$ be a random sample of size n from a Poisson distribution with parameter We are going to show that the MLE of λ attains the Cramer-Rao bound. First, we need to show the entire expression of the lower bound of Cramer-Rao. Let's begin with the expression in the handout:

$$C = V - \frac{\partial \Psi(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial \Psi(\theta)'}{\partial \theta}$$

Where C is positive definite. And $\frac{\partial \Psi(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial \Psi(\theta)'}{\partial \theta}$ is the Cramer-Rao bound. If $C = 0$ then:

$$V = \frac{\partial \Psi(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial \Psi(\theta)'}{\partial \theta}$$

But in general

$$V \geqq \frac{\partial \Psi(\theta)}{\partial \theta} I^{-1}(\theta) \frac{\partial \Psi(\theta)'}{\partial \theta}$$

as states in Theorem **7.3.9** of Casella and Berger (**CB**). In some texts is common to define the Cramer-Rao as follows

$$V\left(\widetilde{\theta}\right) \geqq \left[-E \left. \frac{\partial^2 \log L}{\partial \theta \partial \theta'} \right|_{\theta_0} \right]^{-1} \tag{1}$$

Where $\widetilde{\theta}$ is an unbiased estimator and θ_0 is the true vector of parameter, thus we evaluate $E\frac{\partial^2 \log L}{\partial \theta \partial \theta'}|_{\theta_0}$ in the true vector of parameters (see Chumacero 2003 and Mittlehammer, et. al. 2000). Where if the estimators are unbiased so $\frac{\partial \Psi(\theta)}{\partial \theta}$ is a identity matrix, in this case V attains the Cramer-Rao bound. Constructing and getting $I(\theta)$ in terms of the problem:

$$I(\theta) = -E \left(\left. \frac{\partial^2 \log L}{\partial \lambda \partial \lambda} \right|_{\overline{\lambda}} \right)$$

where $\overline{\lambda}$ refers to the true vector of parameters. We know that:

$$\begin{aligned} L &= \prod_k \frac{\lambda^k}{k!} \exp(-\lambda), \text{ or} \\ L &= \prod_i \frac{\lambda^{x_i}}{x_i!} \exp(-\lambda) \end{aligned}$$

Log-linearizing the above expression, we have:

$$\log L = \sum_{i}^{N} x_i \log \lambda - n\lambda_i - \log x_i!$$

And by taking the first and second derivatives:

$$\begin{aligned} \frac{\partial \log L}{\partial \lambda} &= \frac{\sum_{i=1}^{N} x_i}{\lambda} - n = 0 \\ \frac{\partial^2 \log L}{\partial \lambda \partial \lambda} &= -\frac{\sum_{i=1}^{N} x_i}{\lambda^2} \end{aligned}$$

And the estimator for λ is given by $\hat{\lambda} = \frac{\sum_{i}^{N} x_i}{n}$. First at all, we need to know if the estimator is unbiased or not, so:

$$\begin{aligned} E\left(\hat{\lambda}\right) &= \frac{nE(x_i)}{n} \\ &= E(x_i) \end{aligned}$$

so, taking expectations to x

$$\begin{aligned} E(x) &= \sum_{i=0}^{\infty} x_i \frac{\lambda^{x_i}}{x_i!} \exp(-\lambda) \\ &= \exp(-\lambda) \sum_{i=0}^{\infty} x_i \frac{\lambda^{x_i}}{x_i!} \\ &= \exp(-\lambda) \sum_{i=1}^{\infty} \frac{\lambda^{x_i}}{(x_i - 1)!} \\ &= \exp(-\lambda) \left[\lambda + \lambda^2 + \frac{\lambda^3}{2} + \frac{\lambda^4}{6} + \frac{\lambda^5}{24} + ...\right] \\ &= \lambda \exp(-\lambda) \left[1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{6} + \frac{\lambda^4}{24} + ...\right] \\ &= \lambda \left[\exp(-\lambda) \left(1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{6} + \frac{\lambda^4}{24} + ...\right)\right] \\ &= \lambda \sum_{i=0}^{\infty} \frac{\exp(-\lambda) \lambda^{x_i}}{(x_i)!} \\ &= \lambda \end{aligned}$$

Can't you think of a simpler way?

Here you need to note that we refer to $x_i = i$. So we can conclude that: $E\left(\hat{\lambda}\right) =$

$E(x) = \lambda$.The above can be demonstrated in an elegant way as well:

$$\begin{aligned} E(x) &= \sum_{i=0}^{\infty} x_i \frac{\lambda^{x_i}}{x_i!} \exp(-\lambda) \\ &= \lambda \sum_{i=1}^{\infty} \frac{\lambda^{x_i - 1}}{(x_i - 1)!} \exp(-\lambda) \\ &= \lambda \sum_{i=0}^{\infty} \frac{\lambda^{x_i}}{(x_i)!} \exp(-\lambda) = \lambda \end{aligned}$$

And now for the variance:

$$\begin{aligned} V\left(\hat{\lambda}\right) &= V\left(\frac{\sum x_i}{n}\right) \\ &= \frac{n}{n^2} V(x_i) \\ &= \frac{1}{n} V(x_i) = \frac{1}{n}\widehat{\lambda} \end{aligned}$$

In the appendix we show an elegant (and matricial) way to derive $V\left(\hat{\lambda}\right) = \frac{1}{n}\widehat{\lambda}$. The last equation is demonstrated as follows:

$$\begin{aligned} V(\lambda) &= E\left(\hat{\lambda} - E(\lambda)\right)^2 \\ &= E\left(\hat{\lambda}\right)^2 - \lambda^2 \\ &= E\left(\frac{\sum x_i}{n}\right)^2 - \lambda^2 \\ &= \frac{1}{n^2} E\left(\sum x_i\right)^2 - \lambda^2 \\ &= \frac{n^2}{n^2} E(x)^2 - \lambda^2 \\ &= E(x)^2 - \lambda^2 \end{aligned}$$

As follows we demonstrate which $E(x)^2$ is equal to:

$$\begin{aligned}
E(x)^2 &= \sum_{i=0}^{\infty}(x_i)^2\frac{\lambda^{x_i}}{x_i!}\exp(-\lambda) \\
&= \sum_{i=0}^{\infty}x_ix_i\frac{\lambda^{x_i}}{x_i!}\exp(-\lambda) \\
&= \lambda\sum_{i=1}^{\infty}x_i\frac{\lambda^{x_i-1}}{(x_i-1)!}\exp(-\lambda) \\
&= \lambda\sum_{i=1}^{\infty}\left((x_i-1)\frac{\lambda^{x_i-1}}{(x_i-1)!}\exp(-\lambda)+\frac{\lambda^{x_i-1}}{(x_i-1)!}\exp(-\lambda)\right) \\
&= \lambda\left(\sum_{i=1}^{\infty}(x_i-1)\frac{\lambda^{x_i-1}}{(x_i-1)!}\exp(-\lambda)+\sum_{i=1}^{\infty}\frac{\lambda^{x_i-1}}{(x_i-1)!}\exp(-\lambda)\right) \\
&= \lambda\left(\sum_{i=0}^{\infty}x_i\frac{\lambda^{x_i}}{x_i!}\exp(-\lambda)+\sum_{i=0}^{\infty}\frac{\lambda^{x_i}}{x_i!}\exp(-\lambda)\right) \\
&= \lambda(\lambda+1) \\
&= \lambda^2+\lambda
\end{aligned}$$

Given the last expression, we can conclude that:

$$\begin{aligned}
V(\lambda) &= E(x)^2-\lambda^2 \\
&= \lambda^2+\lambda-\lambda^2 \\
&= \lambda
\end{aligned}$$

So $I^{-1}(\theta)\leq V$

$$I(\theta)=-E\left(\frac{\partial^2\log L}{\partial\lambda\partial\lambda}\bigg|_{\overline{\lambda}}\right)=E\left(\frac{\sum x_i}{\lambda^2}\right)=n\frac{E(x_i)}{\lambda^2}=n\frac{\lambda}{\lambda^2}=n\frac{1}{\lambda}$$

So:

$$I^{-1}(\theta)=\frac{\lambda}{n}$$

here we use λ and $\overline{\lambda}$ as the true value of parameter.

(ii). Let's begin with the linear function

$$Y=X\beta+\varepsilon$$

where:

$$y=\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix},\ X=\begin{pmatrix}x_{11}&x_{12}&\cdots&x_{1k}\\x_{21}&\cdots&\cdots&x_{2k}\\\vdots&\ddots&\ddots&\vdots\\x_{n1}&x_{n2}&\cdots&x_{nk}\end{pmatrix},\ \beta=\begin{pmatrix}\beta_1\\\beta_2\\\vdots\\\beta_k\end{pmatrix},\ \varepsilon=\begin{pmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_n\end{pmatrix}$$

Assume that x_{ij}'s are nonstochastic and $\varepsilon \sim N(0,\sigma^2 I)$. Let β be the MLE of where $\theta = (\beta, \sigma^2)'$.Does β attain the Cramer-Rao bound?. Now againg, we need to check out if the estimators are unbiased or not:

$$P(x|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)$$

We have that the likelihood is given by:

$$L = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\sum \frac{\varepsilon_i^2}{2\sigma^2}\right)$$

And now, by taking logarithm:

$$\log L = n \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \sum \frac{\varepsilon_i^2}{2\sigma^2}$$

And matricial terms:

$$\log L = -n \log\left(\sqrt{2\pi\sigma^2}\right) - \frac{\varepsilon'\varepsilon}{2\sigma^2}$$

By knowing that $\varepsilon \sim N(0,\sigma^2 I)$ and replacing this into the last equation, therefore:

$$\begin{aligned}
\log L &= -n \log\left(\sqrt{2\pi\sigma^2}\right) - \frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2} \\
&= -n \log\left(\sqrt{2\pi\sigma^2}\right) - \frac{\left(Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta\right)}{2\sigma^2} \\
&= -n \log\left(\sqrt{2\pi\sigma^2}\right) - \frac{\left(Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta\right)}{2\sigma^2}
\end{aligned}$$

By taking first derivates, we have that:

$$\begin{aligned}
\frac{\partial \log L}{\partial \beta'} &= -\frac{\left(-2Y'X + 2\beta'X'X\right)}{2\sigma^2} = 0 \text{ or} \\
\frac{\partial \log L}{\partial \beta} &= -\frac{\left(-2X'Y + 2X'X\beta\right)}{2\sigma^2} = 0 \\
\frac{\partial \log L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^4} = 0
\end{aligned}$$

So by replacing the estimators, we have the following expressions:

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1} X'Y \\
\hat{\sigma}^2 &= \frac{\left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right)}{n} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n}
\end{aligned}$$

Are they unbiased?, let's see;

$$\begin{aligned}
\hat{\beta} &= \left(X'X\right)^{-1} X'\left(X\beta + \varepsilon\right) \\
&= \left(X'X\right)^{-1} X'X\beta + \left(X'X\right)^{-1} X'\varepsilon \\
&= \beta + \left(X'X\right)^{-1} X'\varepsilon \\
E\left(\hat{\beta}\right) &= \beta + \left(X'X\right)^{-1} X'E\left(\varepsilon\right) \\
E\left(\hat{\beta}\right) &= \beta
\end{aligned}$$

$$\begin{aligned}
E\left(\hat{\sigma}^2\right) &= E\left(\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n}\right) \\
&= E\left(\frac{\sum \hat{\varepsilon}_i^2}{n}\right) \\
&= \frac{\sum E\left(\hat{\varepsilon}_i^2\right)}{n} \\
&= \frac{(n-k)\,\sigma^2}{n}
\end{aligned}$$

Where we must take into account the k degrees of freedom of $\hat{\varepsilon}$ and we see that $\hat{\sigma}^2$ is a biased estimator, and on the other hand $\hat{\beta}$ is unbiased, but asymptotically they are unbiased and consistent. Now getting:

$$I\left(\theta\right) = -E\left(\left.\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right|_{\overline{\theta}}\right)$$

We get the hessian as follows:

$$\begin{aligned}
\frac{\partial^2 \log L}{\partial\beta\partial\beta'} &= -\frac{\left(X'X\right)}{\sigma^2} \\
\frac{\partial^2 \log L}{\partial\beta\partial\sigma^2} &= \frac{\left(-X'Y + X'X\beta\right)}{\sigma^4} \\
\frac{\partial^2 \log L}{\partial\sigma^2\partial\beta} &= \frac{\left(-X'Y + X'X\beta\right)}{\sigma^4} \\
\frac{\partial^2 \log L}{\partial\sigma^2\partial\sigma^2} &= \frac{n}{2\sigma^4} - 2\frac{\left(Y - X\beta\right)'\left(Y - X\beta\right)}{2\sigma^6}
\end{aligned}$$

$$\begin{aligned}
I\left(\theta\right) &= -E\left(\left.\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right|_{\overline{\theta}}\right) = -E\begin{bmatrix} -\frac{(X'X)}{\hat{\sigma}^2} & \frac{\left(-Y'X+X'X\hat{\beta}\right)}{\hat{\sigma}^4} \\ \frac{\left(-X'Y+X'X\hat{\beta}\right)}{\hat{\sigma}^4} & \frac{n}{2\hat{\sigma}^4} - 2\frac{\left(Y-X\hat{\beta}\right)'\left(Y-X\hat{\beta}\right)}{2\hat{\sigma}^6} \end{bmatrix} \\
&= -E\left(\left.\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right|_{\overline{\theta}}\right) = \begin{bmatrix} \frac{(X'X)}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{n}{2\hat{\sigma}^4} + \frac{n\hat{\sigma}^2}{\hat{\sigma}^6} \end{bmatrix} \\
&= -E\left(\left.\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right|_{\overline{\theta}}\right) = \begin{bmatrix} \frac{(X'X)}{\hat{\sigma}^2} & 0 \\ 0 & \frac{n}{2\hat{\sigma}^4} \end{bmatrix}
\end{aligned}$$

note

$E(-X'Y + X'X\hat{\beta})$
$E(-X'(Y - X\beta))$
$E(-X'\varepsilon) = 0$ &
$E(-Y'X + \beta'X'X)$
$E(-(Y - X\beta)'X)$
$E(-\varepsilon'X) = 0$

Where the inverse of the information matrix is given by:

$$I^{-1}(\theta) = \left(-E\left(\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\bigg|_{\overline{\theta}}\right)\right)^{-1} = \begin{bmatrix} \overline{\sigma}^2 (X'X)^{-1} & 0 \\ 0 & \frac{2\overline{\sigma}^4}{n} \end{bmatrix}$$

We show the variances of the parameters are defined as follows;

$$\begin{aligned} V\left(\widehat{\beta}\right) &= E\left(\widehat{\beta}-\beta\right)\left(\widehat{\beta}-\beta\right)' \\ &= (X'X)^{-1} X'E\left(\hat{\varepsilon}\hat{\varepsilon}'\right) X (X'X)^{-1} \\ &= (X'X)^{-1} X'\left(\sigma^2 I\right) X (X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1} \end{aligned}$$

where I is an identity matrix. In the case of σ^2 we have the following (see CB appendix and verify that the variance of a `chi-square` variable is $2p$ where p is the degrees of freedom);

$$V\left(\widehat{\sigma}^2\right) = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} \tag{2}$$

$$\begin{aligned} V\left(\widehat{\sigma}^2\right) &= \frac{V\left(\hat{\varepsilon}'\hat{\varepsilon}/\sigma^2\right)}{n^2/\left(\sigma^2\right)^2} \\ V\left(\widehat{\sigma}^2\right) &= \frac{2(n-k)\sigma^4}{n^2} \end{aligned}$$

But, we need to show the variance of some unbiased estimator and do the comparison with the Cramer-Rao bound. Thus, considering the bias-correction in (2)

$$V\left(\widetilde{\sigma}^2\right) = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k} \tag{3}$$

$$\begin{aligned} V\left(\widetilde{\sigma}^2\right) &= \frac{V\left(\hat{\varepsilon}'\hat{\varepsilon}/\sigma^2\right)}{(n-k)^2/\left(\sigma^2\right)^2} \\ &= \frac{2(n-k)\sigma^4}{(n-k)^2} \\ V\left(\widetilde{\sigma}^2\right) &= \frac{2\sigma^4}{n-k} \end{aligned}$$

Thus, $V\left(\widetilde{\sigma}^2\right) = \frac{2\sigma^4}{n-k} \geq \frac{2\sigma^4}{n}$, that means that $\widetilde{\sigma}^2$ does not attain the Cramer-Rao lower bound (CRLB), in fact no unbiased estimator for σ^2 attains this lower bound. On the other hand the estimators β attain the CRLB. (see Chumacero, 2003).

2. According to `wikipedia`, in this exercise the phone calls arrive at a Poisson realization and at an average rate of λ per minute. This rate is not observable, but the numbers X1, ..., Xn of phone calls that arrived during n successive one-minute periods are observed. It is desired to estimate the probability e^{λ} that the next one-minute period passes with no

phone calls. Wikipedia says that an extremely crude estimator of the desired probability is

$$\delta_0 = \begin{matrix} 1 & \text{if } X_1 = 0 \\ 0 & \text{if } X_1 > 0 \end{matrix}$$

i.e $E(\delta_0) = \Pr(\delta_0 = 1) = \Pr(X_1 = 0)$. In fact $\Pr(x_1 = 0|\lambda) = \frac{e^{-\lambda}\lambda^0}{0!} = e^{-\lambda}$.We know that a $\sum_i^n x_i$ is a sufficient statistic for estimate λ, this because $\sum_i^n x_i = n\lambda$. We construct $\delta_1 = E[\delta_0|S_n]$ is binomial, where $S_n = \sum_i^n x_i$. In those terms $\Pr(\delta_0 = 1|S_n)$ is equivalent to $\Pr(x_1 = 0|S_n)$. Thus we have that $x_1|S_n$~$\mathtt{Binomial}\left(S_n, \frac{1}{n}\right)$. So,

$$\begin{aligned} \Pr(x_1 = 0|S_n) &= \begin{pmatrix} S_n \\ 0 \end{pmatrix} P^0 (1-P)^{S_n} \\ &= (1-P)^{S_n} \end{aligned}$$

We know that $P = \frac{1}{n}$, thus we have $\Pr(x_1 = 0|S_n) = \left(1 - \frac{1}{n}\right)^{S_n} = \left(1 - \frac{1}{n}\right)^{n\lambda}$. Since the average number of calls arriving during the first n minutes is $n\lambda$, we have

$$\left(1 - \frac{1}{n}\right)^{n\lambda} \approx e^{-\lambda}$$

3. In this question we use estiamtions using logistic and normality assumptions.

   (i). Let us use the radiotherapy data to estimate the logit and probit models. The binary choice model in latent variable representation is specified by the following model

$$Y = X\beta + \varepsilon$$

We show the estimations in the following table:

Table 3.1 Final Report
(convergence criteria 1e-5)

Algorithm	Logit		Probit	
	Conv/Non-conv	#Iter	Conv/Non-conv	#Iter
1	3.8192,-0.0865	2418	2.3017,-0.0522	1400
2	3.8194,-0.0865	951	2.3018,-0.0522	290
3	3.8192,-0.0865	2251	2.3017,-0.0522	1091
4	3.8194,-0.0865	3	2.3018,-0.0522	3
5	3.8192,-0.0865	2535	2.3017,-0.0522	1163
6	3.8193,-0.0865	2274	2.3017,-0.0522	1206
7	3.8194,-0.0865*	48	2.3018,-0.0522*	29
8	3.8194,-0.0865	13	2.3018,-0.0522	31

Note 1. Algorithm: 1 (steep), 2 (BFGS), 3 (DFP), 4 (Newton-Raphson), 5 (BHHH) 6 (PRCG), 7 (BFGS-S), 8 (DFP-SC)..

Note 2: All results converge under return GAUSS code=0 (normal convergence).

Number maximum of iteration was fixed to 3000 in order to avoid possible divergences

Note 3: the starting values for BFGS-S were fixed to 0.20 and -0.005.

(ii). Of course the coefficients of covariates x_{ij}'s of the logit model do not directly compare to the coefficients of covariates x_{ij}'s of the probit model. To make them comparable, the marginal effects are computed. We derive the marginal effect; the logistic function is (see CB appendix):

$$f\left(x\beta|\mu, \frac{1}{\theta}\right) = \theta \frac{e^{-(x\beta-\mu)\theta}}{\left[1+e^{-(x\beta-\mu)\theta}\right]^2}$$

In our case we simplify the expression fixing $\mu = 0$ and $\theta = 1$. Then, we construct the probability function

$$\begin{aligned} \int_x f(x\beta)\, dx &= \int_x \frac{e^{-x\beta}}{\left[1+e^{-x\beta}\right]^2} dx \\ F(X &= x\beta) = \frac{1}{1+e^{-x\beta}} \end{aligned}$$

hen the marginal effect is computed as follows:

$$\begin{aligned} \frac{\partial F(x\beta)}{\partial x} &= \frac{1}{\left(1+e^{-x\beta}\right)^2} e^{-x\beta}\beta \\ &= \frac{e^{-x\beta}}{\left(1+e^{-x\beta}\right)^2}\beta \\ &= f(x\beta)\cdot\beta \end{aligned}$$

In the case of the probabilistic function using a normal distribution we have the same above steps and we define the marginal effects as follows:

$$\frac{\partial F(x\beta)}{\partial x} = \Lambda(x\beta)\cdot\beta$$

where Λ is the normal distribution function. We report in the following table *(i)* The marginal effect evaluated at the sample mean of x_{ij}'s, *(ii)* The mean of the marginal effects of x_{ij}'s.

Table 3.1: Marginal Effects in Logit and Probit models

Marginal Effect	Logit Model	Probit Model
At the sample mean of x_i	-0.0207	-0.0202
Mean of $\widehat{\eta}_{ij}$	-0.0170	-0.0172
Median of $\widehat{\eta}_{ij}$	-0.0175	-0.0177
Max of $\widehat{\eta}_{ij}$	-0.0091	-0.0101
Min of $\widehat{\eta}_{ij}$	-0.0216	-0.0208

As we see in table 3.1 the marginal effects are similar across methods, we include statistics for the median, maximum and minimum values of η_{ij} for having an idea of the ddistribution of η_{ij} in the sample. The mean and the median are similar but they are greater than the marginal effect evaluated at the sample mean, the range of the η_{ij} across sample goes from -0.010 to -0.020, precisely the lower bound is close to the marginal effect evaluated at the sample mean.

4. We generate n = 1,000 random numbers from the Gamma$(2, 3^{-1})$ distribution with the density function We report the following;

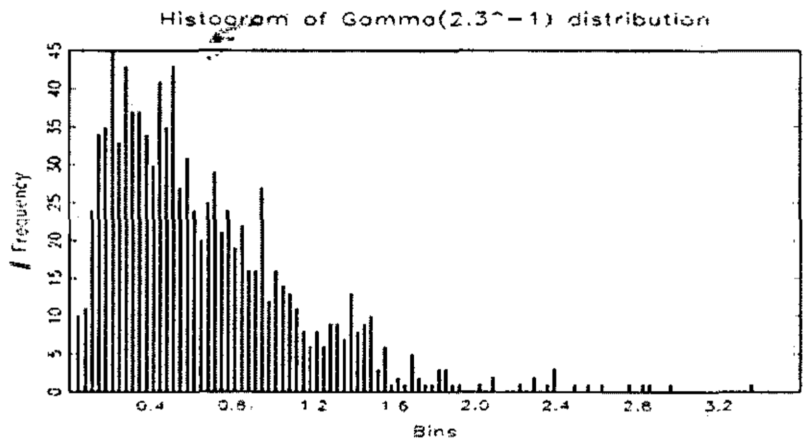
 (i). We get the log-likelihood function in the following steps (we assume independence between realizations)

$$\begin{aligned} f(x|\alpha,\lambda) &= \frac{\lambda^{\alpha}x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)} \\ \prod_{i=1}^{N} f(x_i|\alpha,\lambda) &= \prod_{i=1}^{N}\frac{\lambda^{\alpha}x_i^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)} \\ \log\prod_{i=1}^{N} f(x_i|\alpha,\lambda) &= \log\prod_{i=1}^{N}\frac{\lambda^{\alpha}x_i^{\alpha-1}e^{-\lambda x_i}}{\Gamma(\alpha)} \end{aligned}$$

 finally

$$\log L = n\alpha\log\lambda + (\alpha-1)\sum\log x_i - \lambda\sum x_i - n\Gamma(\alpha)$$

 (ii). In graph 4.1 we show the gamma distribution with parameters $\alpha = 2$ and $\lambda = 3$.



we calculate the exact mean as follows:

$$\begin{aligned} E(x) &= \int_0^{\infty} x\frac{\lambda^{\alpha}x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}dx \\ &= \int_0^{\infty}\frac{\lambda^{\alpha}x^{\alpha}e^{-\lambda x}}{\Gamma(\alpha)}dx \end{aligned}$$

using the following rule $uv|_0^{\infty} = \int_0^{\infty} udv + \int_0^{\infty} vdu$, being in our case $u = \frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha}$ and $dv = e^{-\lambda x}$

$$\begin{aligned} \int_0^{\infty}\frac{\lambda^{\alpha}x^{\alpha}e^{-\lambda x}}{\Gamma(\alpha)}dx &= \frac{\lambda^{\alpha}}{\Gamma(\alpha)}\frac{x^{\alpha}e^{-\lambda x}}{-\lambda}|_0^{\infty} + \frac{\alpha}{\lambda}\int\frac{\lambda^{\alpha}}{\Gamma(\alpha)}e^{-\lambda x}x^{\alpha-1} \qquad (4) \\ &= \frac{\lambda^{\alpha}}{\Gamma(\alpha)}\frac{x^{\alpha}e^{-\lambda x}}{-\lambda}|_0^{\infty} + \frac{\alpha}{\lambda} \end{aligned}$$

evaluating the limits in the following expression:

$$\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\frac{x^{\alpha}e^{-\lambda x}}{-\lambda}|_0^{\infty} = -\frac{\lambda^{\alpha-1}}{\Gamma(\alpha)}\frac{x^{\alpha}}{e^{\lambda x}}|_0^{\infty} = -\frac{\lambda^{\alpha-1}}{\Gamma(\alpha)}\frac{x^{\alpha}}{e^{\lambda x}}|^{\infty}$$

applying H'opital

$$\begin{aligned} -\frac{\lambda^{\alpha-1}}{\Gamma(\alpha)}\frac{x^{\alpha}}{e^{\lambda x}}|^{\infty} &= \lim_{x\to\infty} -\frac{\lambda^{\alpha-1}}{\Gamma(\alpha)}\frac{x^{\alpha}}{e^{\lambda x}} = \lim_{x\to\infty} -\frac{\lambda^{\alpha-1}}{\Gamma(\alpha)}\frac{\alpha x^{\alpha-1}}{\lambda e^{\lambda x}} \\ &= \lim_{x\to\infty} -\frac{\lambda^{\alpha-1}}{\Gamma(\alpha)}\frac{\alpha(\alpha-1)x^{\alpha-2}}{\lambda e^{\lambda x}} \end{aligned}$$

replacing the value of parameters we have

$$\begin{aligned} -\frac{\lambda^{\alpha-1}}{\Gamma(\alpha)}\frac{x^{\alpha}}{e^{\lambda x}}|^{\infty} &= \lim_{x\to\infty} -\frac{3^{2-1}}{\Gamma(2)}\frac{2(2-1)x^{2-2}}{3e^{3x}} \\ &= 0 \end{aligned}$$

thus we have

$$\int_0^{\infty}\frac{\lambda^{\alpha}x^{\alpha}e^{-\lambda x}}{\Gamma(\alpha)}dx = \frac{\alpha}{\lambda}$$

In the case of the (raw) second moment

$$E\left(x^2\right) = \int_0^{\infty} x^2\frac{\lambda^{\alpha}x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}dx$$

using the following rule $uv|_0^{\infty} = \int_0^{\infty} udv + \int_0^{\infty} vdu$, being in our case $u = \frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha+1}$ and $dv = e^{-\lambda x}$

$$\int_0^{\infty}\frac{\lambda^{\alpha}x^{\alpha+1}e^{-\lambda x}}{\Gamma(\alpha)}dx = \frac{\lambda^{\alpha}}{\Gamma(\alpha)}\frac{x^{\alpha+1}e^{-\lambda x}}{-\lambda}|_0^{\infty} + \frac{(\alpha+1)}{\lambda}\int_0^{\infty}\frac{\lambda^{\alpha}}{\Gamma(\alpha)}e^{-\lambda x}x^{\alpha}dx$$

Using the result in (4) and evaluating the first right-side expression at limit equal to infinity

$$\begin{aligned} \int_0^{\infty}\frac{\lambda^{\alpha}x^{\alpha+1}e^{-\lambda x}}{\Gamma(\alpha)}dx &= \frac{\lambda^{\alpha}}{\Gamma(\alpha)}\frac{x^{\alpha+1}e^{-\lambda x}}{-\lambda}|_0^{\infty} + \frac{(\alpha+1)}{\lambda}\frac{\alpha}{\lambda} \\ &= \frac{(\alpha+1)}{\lambda}\frac{\alpha}{\lambda} \end{aligned}$$

the variance is $E\left(x^2\right) - [E(x)]^2$, then:

$$\begin{aligned} V(x) &= E\left(x^2\right) - [E(x)]^2 \\ &= \frac{(\alpha+1)}{\lambda}\frac{\alpha}{\lambda} - \left(\frac{\alpha}{\lambda}\right)^2 \\ &= \frac{\alpha}{\lambda^2} \end{aligned}$$

The exact mean and variance are $\frac{2}{3}$ and $\frac{2}{9}$ respectively.

(iii). The sample mean and standard deviation of data are $0.646 \approx \frac{2}{3}$ and $0.46 \approx \sqrt{\frac{2}{9}}$

(iv). The estimates and asymptotic standard errors of α and λ are

Table 4.1

	α	λ
point estimates	2.0373	3.1556
standard errors	0.0847	0.1487

(v). The algorithm I chose was the Newton-Raphson and I found solution in just 7 iterations (see the **GAUSS** code in the appendix).

References

[1] Mendenhall and Scheaffer. 1973. Mathematical Statistics with Applications. Duxbury Press. North Scituate, Massachusetts.

[2] Casella, G and R. Berger. 2002. Statistical Inference. Second Edition, Duxbury Advanced Studies.

[3] `Mathworld website. http://mathworld.wolfram.com/`

[4] `GAUSS kernel density library. GAUSS.`

[5] M.P Wand & M.C Jones. 1995. Kernel Smoothing. Monographs on Statistics and Applied Probability. Chapman & Hall, 1995.

[6] Mittelhamer, R., G. Judge, and D. Miller (2000). Econometric Foundations. Cambridge Unviersity Press.

[7] Chumacero, R. 2003. Maximum LIkelihood Handout. Universidad de Chile.

[8] Amemiya, T., 1985, Advanced Econometrics, Cambridge: Harvard University Press.

1 Appendix

1.1 Question 2.(a): An elegant way

We have the following

$$\widehat{\lambda} = \frac{X'i}{n}$$

the square of the estimator is

$$\widehat{\lambda}^2 = \frac{X'i}{n}\frac{X'i}{n}$$

we can express the above identity as

$$\begin{aligned} \widehat{\lambda}\widehat{\lambda}' &= \frac{X'i}{n}\frac{i'X}{n} \\ \widehat{\lambda}\widehat{\lambda}' &= \frac{X'JX}{n^2} \end{aligned}$$

where $J = ii'$. Then, if we take expectations

$$E\left(\widehat{\lambda}\widehat{\lambda}'\right) = E\left(X'JX\right)\frac{1}{n^2}$$

Using trace properties and doing some algebra

$$\begin{aligned} \mathtt{tr}E\left(\widehat{\lambda}\widehat{\lambda}'\right) &= \mathtt{tr}E\left(\frac{X'JX}{n}\right)\frac{1}{n^2} \\ E\left(\widehat{\lambda}\widehat{\lambda}'\right) &= E\left(\mathtt{tr}\left(X'JX\right)\right)\frac{1}{n^2} \\ &= E\left(\mathtt{tr}\left(XX'J\right)\right)\frac{1}{n^2} \end{aligned}$$

Using the interchangeable property between trace and expectation

$$\begin{aligned} E\left(\widehat{\lambda}\widehat{\lambda}'\right) &= \mathtt{tr}\left(E\left(XX'J\right)\right)\frac{1}{n^2} \qquad (5) \\ &= \mathtt{tr}\left(J\cdot E\left(XX'\right)\right)\frac{1}{n^2} \end{aligned}$$

In this case $E(XX')$ is

$$E\left(XX'\right) = \begin{bmatrix} E\left(x_1^2\right) & E\left(x_1x_2\right) & E\left(x_1x_3\right) & \ldots & E\left(x_1x_n\right) \\ E\left(x_2x_1\right) & E\left(x_2^2\right) & E\left(x_2x_3\right) & \ldots & E\left(x_2x_n\right) \\ E\left(x_3x_1\right) & E\left(x_3x_2\right) & E\left(x_3^2\right) & \ldots & E\left(x_3x_n\right) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ E\left(x_nx_1\right) & E\left(x_nx_2\right) & E\left(x_nx_3\right) & \ldots & E\left(x_n^2\right) \end{bmatrix}$$

replacing by the known values

$$E\left(XX'\right) = \begin{bmatrix} \lambda^2+\lambda & \lambda^2 & \lambda^2 & \ldots & \lambda^2 \\ \lambda^2 & \lambda^2+\lambda & \lambda^2 & \ldots & \lambda^2 \\ \lambda^2 & \lambda^2 & \lambda^2+\lambda & \ldots & \lambda^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda^2 & \lambda^2 & \lambda^2 & \ldots & \lambda^2+\lambda \end{bmatrix} \qquad (6)$$

Then we replace matrix in (6) into $\operatorname{tr}\left(J \cdot E\left(XX'\right)\right)\frac{1}{n^2}$ and perform the product $J \cdot E\left(XX'\right)$

$$J \cdot E\left(XX'\right) = \left(n\lambda^2 + \lambda\right) \cdot I_{n\times n}$$

thus

$$\begin{aligned} E\left(\widehat{\lambda}\widehat{\lambda}'\right) &= \operatorname{tr}\left(\left(n\lambda^2+\lambda\right)\cdot I_{n\times n}\right)\frac{1}{n^2} \\ &= n\left(n\lambda^2+\lambda\right)\frac{1}{n^2} \\ E\left(\widehat{\lambda}^2\right) &= \lambda^2 + \frac{\lambda}{n} \end{aligned}$$

1.2 Question 3: GAUSS code

```
/*******************************************************************
   Assignment 8
   MLE of Logit model (Tanner p.14) also we add
    probit model---
   Program:  c:\ec506\logitprob.pro@
   code made by Professor Tsurumi but modified by Freddy Rojas Cama
   *******************************************************************/
   new;
   library maxlik, pgraph;
   #include maxlik.ext;
   cls;

   // Loading data
   load z[24,2]=C:\Users\Maro\Desktop\assignment_8\logit.dat;
   // path where I save results
   chDir C:\Users\Maro\Desktop\assignment_8;
   path="C:\Users\Maro\Desktop\assignment_8" ;
   // In this question we need to
   // Setting up the algorithm for seeking a local maximum
   // Algorithm:  =1 (steep) =2 (BFGS) =3 (DFP) =4 (Newton-Raphson)
   // =5 (BHHH) =6 (PRCG) =7 (BFGS-S) =8 (DFP-SC)
   maxset;
   _max_Algorithm=4;
   //_max_Algorithm=7;
   __output=1; /*===If 0, print out of each iteration is supressed;
    if 1, print out of each iteration is given.  =======*/

   /*====MLE of logit and probit models using MAXLIK======*/
   y=z[.,2];
   n=rows(y);
   x=ones(n,1)~z[.,1];
   /* logit & probit */
```

```
yx=y~x;
k=cols(x);
start=y/x; /*===using OLS as the initial valaues ====*/
start_set=seqa(start[1],0.2,10)~seqa(start[2],0.2,10);
print;
print "Logit Model ";
j=1;
do while j <= 6;
if j==5;
j=j+1;
endif;
maxset;
_max_Algorithm=j;
_max_GradTol=1e-5;
_max_MaxIters=3000;
//_max_LineSearch=5; //BHHH step
__output=0; /*===If 0, print out of each iteration is supressed;
 if 1, print out of each iteration is given.  =======*/
field = 1;
prec = 0;
fmat = "%*.*lf";
zz = ftos(j,fmat,field,prec);
max_al="Logit_algorithm_" $+ zz;
output file=^max_al reset;
__title="MLE Logit using " $+ zz $+ " Algorithm";
{coeff,f,g,c,ret}=maxlik(yx,0,&logit,start);
call maxprt(coeff,f,g,c,ret);
output off;
j=j+1;
endo;
//calculating marginal effects for logit model
xb_bar=meanc(x.*coeff');
f_xb_bar=exp(-sumc(xb_bar))/(1+exp(-sumc(xb_bar)))^2;
me_logit_coeff1=f_xb_bar*coeff[1];
me_logit_coeff2=f_xb_bar*coeff[2];
xb_bar_i=x.*coeff';
f_xb_bar_i=exp(-sumc(xb_bar_i')).(1+exp(-sumc(xb_bar_i'))).^2;
mei_logit_coeff1=f_xb_bar_i*coeff[1];
mei_logit_coeff2=f_xb_bar_i*coeff[2];
mean_nu=meanc(mei_logit_coeff2);
median_nu=median(mei_logit_coeff2);
max_nu=maxc(mei_logit_coeff2);
min_nu=minc(mei_logit_coeff2);
output file=Logit_q3_marginal_effects reset;
" ";
```

```
" ";
" --------------- Question 3 -------------------------------------";
" ";
" Results in question 3.i, Logit ";
" AT the sample mean ";; me_logit_coeff2;
" The mean of the marginal effects of xij's ";; mean_nu;
" The median of the marginal effects of xij's ";; median_nu;
" The max of the marginal effects of xij's ";; max_nu;
" The min of the marginal effects of xij's ";; min_nu;
" ";
output off;
print;
print "Probit Model ";
j=1;
do while j <= 6;
if j==5 or j==3;
j=j+1;
endif;
maxset;
_max_Algorithm=j;
_max_GradTol=1e-5;
_max_MaxIters=3000;
//_max_LineSearch=5; //BHHH step
__output=0; /*===If 0, print out of each iteration is supressed;
 if 1, print out of each iteration is given.  =======*/
field = 1;
prec = 0;
fmat = "%*.*lf";
zz = ftos(j,fmat,field,prec);
max_al="Probit_algorithm_" $+ zz;
output file=^max_al reset;
__title="MLE Probit using " $+ zz $+ " Algorithm";
{coeff,f,g,c,ret}=maxlik(yx,0,&probit,start);
call maxprt(coeff,f,g,c,ret);
output off;
j=j+1;
endo;
//calculating marginal effects for logit model
xb_bar=meanc(x.*coeff');
f_xb_bar=pdfn(sumc(xb_bar));
me_probit_coeff1=f_xb_bar*coeff[1];
me_probit_coeff2=f_xb_bar*coeff[2];
xb_bar_i=x.*coeff';
f_xb_bar_i=pdfn(sumc(xb_bar_i'));
mei_probit_coeff1=f_xb_bar_i*coeff[1];
```

```
mei_probit_coeff2=f_xb_bar_i*coeff[2];
mean_nu=meanc(mei_probit_coeff2);
median_nu=median(mei_probit_coeff2);
max_nu=maxc(mei_probit_coeff2);
min_nu=minc(mei_probit_coeff2);
output file=Probit_q3_marginal_effects reset;
" ";
" ";
" --------------- Question 3 -------------------------------------";
" ";
" Results in question 3.ii, Probit";
" At the sample mean ";; me_probit_coeff2;
" The mean of the marginal effects of xij's ";; mean_nu;
" The median of the marginal effects of xij's ";; median_nu;
" The max of the marginal effects of xij's ";; max_nu;
" The min of the marginal effects of xij's ";; min_nu;
" ";
output off;
// Procedures
proc logit(beta,dat);
local y,x,xb,p0,p1,logl;
y=dat[.,1];
x=dat[.,2:k+1];
xb=x*beta;
p1=exp(xb)./(1+exp(xb));
p0=1-p1;
logl=y.*ln(p1)+(1-y).*ln(p0);
logl=sumc(logl); /*====logl is nx1 vector =====*/
retp(logl); /*=== returning logl as a vector rather than a scalar ====*/
endp;
proc probit(beta,dat);
local y,x,xb,p0,p1,logl;
y=dat[.,1];
x=dat[.,2:k+1];
xb=x*beta;
p1=cdfn(xb);
p0=1-p1;
logl=y.*ln(p1)+(1-y).*ln(p0);
logl=sumc(logl); /*===logl is nx1 vector ====*/
retp(logl); /*===returning logl as a vector rather than a scalar====*/
endp;
```

1.3 Question 4: GAUSS code

```
/*****************************************************************
   Assignment 8
   MLE of Logit model (Tanner p.14) also we add
    probit model---
   Program:  c:\ec506\logitprob.pro@
   code made by Professor Tsurumi but modified by Freddy Rojas Cama
   *****************************************************************/
   new;
   library maxlik, pgraph;
   cls;

   // path where I save results
   chDir C:\Users\Maro\Desktop\assignment_8;

   maxset;
   _max_Algorithm=4;
   _max_GradTol=1e-5;
   _max_MaxIters=1000;
   //_max_LineSearch=5; //BHHH step
   __output=1; /*===If 0, print out of each iteration is supressed;
    if 1, print out of each iteration is given.  =======*/
   /*
   field = 1;
   prec = 0;
   fmat = "%*.*lf";
   zz = ftos(j,fmat,field,prec);
   max_al="_" $+ zz;
   */
   // generating data from Gamma distribution alpha=2, theta=3^-1
   x=rndgam(1000,1,2)*3^-1;
   //describing the data (mean and standard error);
   mean_x=meanc(x);
   std_x=stdc(x);
    _pcolor = { 5}; /* Colors for series */
    _pmcolor = { 1, 8, 2, 8, 8, 8, 8, 8, 15 };
    /*Colors for axes, title, x and y labels, date, box, and background */
    _plwidth={12}; /*Controls line thickness for main curves*/
    // _paxht=0.20; /*Controls size of axes labels*/
    _ptitlht = 0.18; /*Controls main title size */
   _pltype={6 3};
   //_plegctl = { 1 7 0.04 13.5};
   //_plegstr="Data Coming from Gamma distribution";
   xlabel("Bins");
   title("Histogram of Gamma(2,3^-1) distribution");
```

```
{ b,m,freq }=hist(x,100);
output file=q4 reset;
start=0.2|0.3;
__title="MLE of Gamma Distribution";
{coeff,f,g,c,ret}=maxlik(x,0,&gam,start);
call maxprt(coeff,f,g,c,ret);
output off;
// Other Stuff
// calculation of asymptotic standard errors
proc gam(beta,dat);
local x,L,logL;
x=dat;
L=beta[1]^beta[2]*x.^(beta[2]-1).*exp(-beta[1].*x)./(gamma(beta[2]));
logL=sumc(ln(L));
retp(logL);
endp;
```

1.4 Question 3: GAUSS code by using MAXLIK library with 7 algorithms

```
/*****************************************************************
   Assignment 8
   MLE of Logit model (Tanner p.14) also we add
    probit model---
   Program:  c:\ec506\logitprob.pro@
   code made by Professor Tsurumi but modified by Freddy Rojas Cama
   *****************************************************************/
   new;
   library maxlik, pgraph;
   #include maxlik.ext;
   cls;

   // Loading data
   load z[24,2]=C:\Users\Maro\Desktop\assignment_8\logit.dat;
   // path where I save results
   chDir C:\Users\Maro\Desktop\assignment_8;
   path="C:\Users\Maro\Desktop\assignment_8" ;
   // In this question we need to
   // Setting up the algorithm for seeking a local maximum
   // Algorithm:  =1 (steep) =2 (BFGS) =3 (DFP) =4 (Newton-Raphson)
   // =5 (BHHH) =6 (PRCG) =7 (BFGS-S) =8 (DFP-SC)

   /*====MLE of logit and probit models using MAXLIK======*/
   y=z[.,2];
   n=rows(y);
   x=ones(n,1)~z[.,1];
   /* logit & probit */
```

```
yx=y~x;
k=cols(x);
start=y/x; /*===using OLS as the initial valaues ====*/
start_set=seqa(start[1],0.2,10)~seqa(start[2],0.2,10);
start=0.05*(3.8192|-0.0895);
print;
print "Logit Model ";
j=3;
do while j == 3;
if j==3;
//j=j+1;
endif;
maxset;
_mlalgr=j;
 _mlgtol=1e-5;
_mlmiter=3000;
__output=1;
 //_mlstep=3;
/*===If 0, print out of each iteration is supressed;
 if 1, print out of each iteration is given.  =======*/
field = 1;
prec = 0;
fmat = "%*.*lf";
zz = ftos(j,fmat,field,prec);
max_al="Logit_algorithm_" $+ zz;
output file=^max_al reset;
__title="MLE Logit using " $+ zz $+ " Algorithm";
{coeff,f,g,c,ret}=maxlik(yx,0,&logit,start);
call maxprt(coeff,f,g,c,ret);
output off;
j=j+1;
endo;
//calculating marginal effects for logit model
xb_bar=meanc(x.*coeff');
f_xb_bar=exp(-sumc(xb_bar))/(1+exp(-sumc(xb_bar)))^2;
me_logit_coeff1=f_xb_bar*coeff[1];
me_logit_coeff2=f_xb_bar*coeff[2];
xb_bar_i=x.*coeff';
f_xb_bar_i=exp(-sumc(xb_bar_i')))./(1+exp(-sumc(xb_bar_i'))).^2;
mei_logit_coeff1=f_xb_bar_i*coeff[1];
mei_logit_coeff2=f_xb_bar_i*coeff[2];
mean_nu=meanc(mei_logit_coeff2);
median_nu=median(mei_logit_coeff2);
max_nu=maxc(mei_logit_coeff2);
min_nu=minc(mei_logit_coeff2);
```

```
output file=Logit_q3_marginal_effects reset;
" ";
" ";
" --------------- Question 3 -------------------------------------";
" ";
" Results in question 3.i, Logit ";
" AT the sample mean ";; me_logit_coeff2;
" The mean of the marginal effects of xij's ";; mean_nu;
" The median of the marginal effects of xij's ";; median_nu;
" The max of the marginal effects of xij's ";; max_nu;
" The min of the marginal effects of xij's ";; min_nu;
" ";
output off;
print;
print "Probit Model ";
j=3;
do while j == 3;
if j==3;
//j=j+1;
endif;
maxset;
_mlalgr=j;
 _mlgtol=1e-5;
_mlmiter=3000;
__output=0;
field = 1;
prec = 0;
fmat = "%*.*lf";
zz = ftos(j,fmat,field,prec);
max_al="Probit_algorithm_" $+ zz;
output file=^max_al reset;
__title="MLE Probit using " $+ zz $+ " Algorithm";
{coeff,f,g,c,ret}=maxlik(yx,0,&probit,start);
call maxprt(coeff,f,g,c,ret);
output off;
j=j+1;
endo;
//calculating marginal effects for logit model
xb_bar=meanc(x.*coeff');
f_xb_bar=pdfn(sumc(xb_bar));
me_probit_coeff1=f_xb_bar*coeff[1];
me_probit_coeff2=f_xb_bar*coeff[2];
xb_bar_i=x.*coeff';
f_xb_bar_i=pdfn(sumc(xb_bar_i'));
mei_probit_coeff1=f_xb_bar_i*coeff[1];
```

```
mei_probit_coeff2=f_xb_bar_i*coeff[2];
mean_nu=meanc(mei_probit_coeff2);
median_nu=median(mei_probit_coeff2);
max_nu=maxc(mei_probit_coeff2);
min_nu=minc(mei_probit_coeff2);
output file=Probit_q3_marginal_effects reset;
" ";
" ";
" ---------------- Question 3 --------------------------------------";
" ";
" Results in question 3.ii, Probit";
" At the sample mean ";; me_probit_coeff2;
" The mean of the marginal effects of xij's ";; mean_nu;
" The median of the marginal effects of xij's ";; median_nu;
" The max of the marginal effects of xij's ";; max_nu;
" The min of the marginal effects of xij's ";; min_nu;
" ";
output off;
// Procedures
proc logit(beta,dat);
local y,x,xb,p0,p1,logl;
y=dat[.,1];
x=dat[.,2:k+1];
xb=x*beta;
p1=exp(xb)./(1+exp(xb));
p0=1-p1;
logl=y.*ln(p1)+(1-y).*ln(p0);
logl=sumc(logl); /*====logl is nx1 vector =====*/
retp(logl); /*=== returning logl as a vector rather than a scalar ====*/
endp;
proc probit(beta,dat);
local y,x,xb,p0,p1,logl;
y=dat[.,1];
x=dat[.,2:k+1];
xb=x*beta;
p1=cdfn(xb);
p0=1-p1;
logl=y.*ln(p1)+(1-y).*ln(p0);
logl=sumc(logl); /*===logl is nx1 vector ====*/
retp(logl); /*===returning logl as a vector rather than a scalar====*/
endp;
```