Department of Economics **University of Manitoba**

ECON 7010: Econometrics I
Fall 2014

Assignment #4

Question 1:
Background -
The **Maximum Likelihood** estimation principle gives us a method of obtaining estimators which are generally very well-behaved, if the sample size is large enough, in a very wide range of situations. To illustrate this technique, consider a random sample of *n* observations on a variable *X*, where $x_i \sim N(\mu, \sigma^2)$ for all *i*. This implies that the density function for a single x_i value is the familiar "bell-shaped" curve, the formula for which is:

$$f(x_i \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\tfrac{1}{2\sigma^2}(x_i - \mu)^2\} .$$

Now consider the *joint* density function of the *n* sample values, given the parameters. Using the *independence* of these values:

$$f(x_1, x_2, \ldots, x_n \mid \mu, \sigma^2) = f(x_1 \mid \mu, \sigma^2) \,.\, f(x_2 \mid \mu, \sigma^2) \ldots\ldots f(x_n \mid \mu, \sigma^2)$$
$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp(-\tfrac{1}{2\sigma^2} \sum_i (x_i - \mu)^2) .$$

Here, the joint density function is viewed as a ***function of the x data***, given the values of the two parameters, μ and σ^2. Without changing the form of this expression, we could also view it from a different perspective – namely, as a ***function of the parameters***, given the values of the data. When we view it in this alternative way we give the function a different name, even though the algebraic expression is the same – we call it the "**Likelihood Function**". That is:

$$L(\mu, \sigma^2 \mid x_1, x_2, \ldots x_n) = f(x_1, x_2, \ldots, x_n \mid \mu, \sigma^2).$$

One general estimation principle that we could adopt is to try and choose the values of the parameters that are "***most likely***" to have generated the observed sample of data. Note that all of the (data-related) information that we have about the parameters is summarized in the joint density function for the sample observations (*i.e.*, the likelihood function). In other words, we could choose a formula for the estimator which maximizes the value of the likelihood function (or the value of the logarithm of the likelihood function, as this will yield the same result, because the logarithmic transformation is strictly monotonic increasing). The estimator that we obtain by following this logic is called the Maximum Likelihood Estimator (MLE), and is perhaps the most widely used estimation principle in statistics (and econometrics).

To get the MLE's of μ and σ^2 in our example, we will have to take the partial derivatives of the likelihood function, or its logarithm (denoted $lnL(\mu, \sigma^2)$) for convenience, with respect to μ and σ^2, and set these derivatives equal to zero to obtain the first-order conditions We then solve this pair of simultaneous equations μ and σ^2:

$$\ln L(\mu, \sigma^2) = -(n/2)\ln(2\pi\sigma^2) - (1/2\sigma^2)\sum_{i=1}^{n}(x_i - \mu)^2 \ ,$$

so,

$$\frac{\partial \ln L}{\partial \mu} = 0 - \frac{1}{2\sigma^2}.2(-1)\sum_{i=1}^{n}(x_i - \mu) \quad , \qquad \text{(i)}$$

and

$$\frac{\partial \ln L}{\partial \sigma^2} = -(n/2\sigma^2) + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 \quad . \qquad \text{(ii)}$$

[Note: To get (ii) we differentiated with respect to σ^2, ***not*** with respect to σ.]

So, setting (i) equal to zero:

$$\sum_{i=1}^{n}(x_1 - \tilde{\mu}) = 0 \ ,$$

or, $$n\tilde{\mu} = \sum_{i=1}^{n} x_i \ ,$$

or, $$\tilde{\mu} = \frac{1}{n}\sum_{i=-1}^{n} x_i = \bar{x} \ . \qquad \text{(iii)}$$

Similarly, setting (ii) equal to zero:

$$\sum_{i=1}^{n}(x_1 - \tilde{\mu})^2 / \tilde{\sigma}^2 = n \ ,$$

so, $$\tilde{\sigma}^2 = (1/n)\sum_{i=1}^{n}(x_1 - \tilde{\mu})^2 \ . \qquad \text{(iv)}$$

Our two MLE's are given by the expressions in (iii) and (iv). In this case, the MLE for the population mean is just the sample mean, and the MLE for the population variance is the sample mean squared deviation (not the sample *variance*, as the divisor in (iv) is n, not (n-1)).

Now, here is the exercise for you to work through:

Consider our usual linear multiple regression model:

$$y = X\beta + \varepsilon \quad ; \quad \varepsilon \sim \text{N}(0, \sigma^2 \text{I})$$

(a) Write down the likelihood function, $L(\beta, \sigma^2 \mid y_1, y_2, \ldots, y_n)$. (Note that if the ε_i's are uncorrelated and normally distributed, then they are independent.)

(b) Derive the maximum likelihood estimator for β in this model. Is this estimator biased or unbiased?

(c) Compare the assumptions that we used in class to derive the OLS estimator of β with the assumptions used here to obtain the MLE.

(d) Derive the MLE for σ^2 in this model. Is this estimator biased or unbiased?

(e) One of the two estimators you have obtained is biased. Show that this bias vanishes if the sample size is sufficiently large.