Economics 671
Problem Set #3:
Thinking about Asymptotic Approximations

The purpose of this question is to help solidify your understanding of sampling distributions and asymptotic approximations to them. We do so via a very simple case where exact, analytic results can be obtained to characterize the sampling distribution of the OLS estimator.

Consider a regression model with just an intercept parameter:

$$y_i = \mu + \epsilon_i,$$

where it is known that ϵ_i are iid draws from an Exponential distribution with mean λ, which are then standardized to have mean zero by subtracting off λ. That is, if $x_i \sim Exponential(\lambda)$ (an Exponential distribution with mean parameter λ and variance λ^2) then

$$p(x_i) = \lambda^{-1} \exp(-\lambda^{-1} x), \quad x, \lambda > 0, \tag{1}$$

and the errors are then assumed to be generated as follows:

$$\epsilon_i = x_i - \lambda, \quad i = 1, 2, \ldots, n. \tag{2}$$

The OLS estimator of μ is, obviously,

$$\hat{\mu}_n = \overline{y}_n,$$

the sample average of the n observations on y. In the remainder of this question, we seek to *analytically* characterize the sampling distribution of this estimator, and then compare the finite-sample distribution of our statistic to the asymptotic normal approximation.

Here are some helpful steps along the way:

(**a**) Derive the moment generating function (MGF) of x, $M(t) \equiv E(\exp[tx])$, when $x \sim Exponential(\lambda)$. (**Note:** A brief review of MGFs is given in Appendix B.6 of your book).

(**b**) Now, consider the distribution of $Y \equiv \sum_i x_i$, where $x_i \overset{iid}{\sim} Exponential(\lambda)$. Use the moment generating function technique to show that $Y \sim Gamma(n, \lambda)$. To do this, note that the MGF of the $Gamma(n, \lambda)$ distribution is:

$$M(t) \equiv E(\exp(tY)) = (1 - \lambda t)^{-n}, \quad t < \lambda^{-1}.$$

(**c**) Use the result above to derive the (exact) sampling distribution of the estimator $\hat{\mu}_n = \overline{y}_n$. Note that the gamma density is parameterized as follows:

$$Y \sim Gamma(n, \lambda) \Rightarrow p(y) = [\Gamma(n)\lambda^n]^{-1} y^{n-1} \exp(-y/\lambda), \quad n, \lambda, y > 0.$$

(**Note:** A solution to part (c) as well as part (d) below will likely make use of the change-of-variable technique. A brief review of the change of variable or transformation technique can be found in A.B.4 and B.5 of your book. In addition, be careful with the supports of the densities when you perform your changes of variable).

(**d**) Given what you established in part (c), obtain the *exact* sampling distribution of the statistic:

$$\sqrt{n}(\hat{\mu}_n - \mu).$$

(**e**) What is the *asymptotic distribution* associated with the statistic

$$\sqrt{n}(\hat{\mu}_n - \mu)?$$

(**f**) Now, bring all of this together using MATLAB. To do this, perform the following:

- First, fix $n = 2$ and set $\mu = 3$, $\lambda = 1$ throughout.

Begin by using the results above to determine the exact distribution of the statistic when $n = 2$ and, using MATLAB, plot $\sqrt{n}(\hat{\mu}_n - \mu)$ when $n = 2$.

Now, compare this analytic result with two alternatives:

First, the normal approximation of the sampling distribution in part (e).

Second, a numerical, computer-generated approximation of the sampling distribution of the statistic when $n = 2$. To do this, proceed as follows:

1. Begin a loop in MATLAB, from, say, 1 to 5,000.
2. Within an iteration of that loop, draw error vector from the de-meaned exponential distribution. The syntax you will need to use for this in matlab is `exprnd(lambda,n,1)`. This will produce n draws from the exponential distribution with mean λ.
3. Given the errors in the previous step, calculate the y vector and then calculate $\hat{\mu}_n = \overline{y_n}$ and $w_n \equiv \sqrt{n}(\hat{\mu}_n - \mu)$.

4. Store this value of w_n and close the loop.

When complete, you will have $5,000$ values of the statistic w_n. This collection of values is an approximation of the sampling distribution of the estimator. To convert this collection of w_n values into a density estimate, use the m-file I have supplied, epanech2. The syntax is

$$[dom\ ran] = epanech2(draws)$$

Here, *dom* is a set of gridpoints over which the density is calculated, *ran* are the ordinates of the density at those gridpoints and *draws* are the data values that you pass to the m-file (in this case, it will be your collection of $5,000$ values of w_n).

You can then plot the density by typing plot(dom,ran).

Plot all three densities (i.e., the analytic density, the asymptotic approximation and the numerical computer-generated approximation above) on the same graph and comment on the results.

Finally, repeat this process, this time setting $n = 50$. Once again, plot each of the three densities on the same graph for the $n = 50$ case. What do you find? How do your results compare to the $n = 1$ case?