LEY DE LOS GRANDES NÚMEROS Y TEOREMA CENTRAL DEL LÍMITE CON STATA

María Delfina Ramírez Cruz[1]

Ley de los Grandes Números. *Bajo este término se engloban varios teoremas que describen el comportamiento del promedio de una secuencia de variables aleatorias conforme aumenta el tamaño de la muestra de manera que este promedio tenderá a estar cerca de la media de la población completa.*

Ley de los Grandes Números de Kolmogorov*: Sea $\{x_i\}$ una secuencia de variables aleatorias independiente e idénticamente distribuidas donde cada una tiene $E(x_i) = \mu$. Si la media de la muestra es igual a: $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, entonces: $\bar{x}_n \xrightarrow{a.s.} \mu$*

Donde $\xrightarrow{a.s.}$ significa que la media de una muestra de tamaño n 'converge casi seguramente' a una constante μ que es la media de $\{x_i\}$ y esto se expresa como sigue:

$$Prob\left(\lim_{n\to\infty} \bar{x}_n = \mu\right) = 1$$

El anterior concepto de convergencia es más fuerte que la convergencia en probabilidad según la cual una secuencia de variables aleatorias $\{\bar{x}_n\}$ converge en probabilidad a una constante μ si para cualquier constante $\varepsilon > 0$:

$$\lim_{n\to\infty} Prob(\,|\,\bar{x}_n - \mu\,|\, > \varepsilon) = 0$$

Observe que en la convergencia en probabilidad se describe el valor límite de una secuencia de probabilidades. No se describe el valor límite de $\bar{x}_n$. Esto significa que si se calculara la diferencia $|\bar{x}_n - \mu|$ para valores crecientes de n, **la probabilidad** *de que esta diferencia exceda cualquier ε dada se vuelve más y más pequeña y converge a 0. Esto no implica que para una secuencia determinada $\{\bar{x}_n\}$ la diferencia observada $|\bar{x}_n - \mu|$ converge a 0.*

---

[1] *Profesora del Posgrado de la Facultad de Economía de la UNAM. Agradezco la colaboración de Francisco Javier Martínez Hernández, estudiante del último semestre de la licenciatura de Economía en la Facultad de Economía de la UNAM como asistente de investigación y en el procesamiento de datos con Stata. Este artículo forma parte de los trabajos realizados dentro del proyecto PAPIME PE300911 financiado por la DGAPA de la UNAM.*

El resultado anterior describe a la media muestral como un estimador consistente de la media poblacional y se conoce como Ley de los Grandes Números. Este corolario se puede probar para variables independiente e idénticamente distribuidas (iid) con varianza finita y también se puede probar para cualquier secuencia de variables iid con media finita μ[2].

Teorema Central del Límite (TCL)*. Una versión simple del Teorema Central del Límite puede expresarse de la manera siguiente. Si se toma una muestra aleatoria de tamaño n de cualquier distribución con media* μ *y varianza* σ^2*, o alternativamente si* $\{x_i\}$ *es una secuencia de variables aleatorias independientes e idénticamente distribuidas cada una con media* μ *y varianza* σ^2*, entonces la función de probabilidad para la media de la muestra converge a una normal con media* μ *y varianza* σ^2/n*.*

Este resultado se estableció para una muestra aleatoria de una distribución Bernoulli por A. de Moivre a principios del siglo dieciocho. La prueba para una muestra aleatoria grande de una distribución arbitraria, sin importar si la distribución es discreta o continua, fue establecida independientemente por J.W. Lindeberg y P. Levy a principios de los 1920s. El enunciado preciso del Teorema Central del Límite de Lindeberg y Levy para la media muestral es:

Si las variables aleatorias $x_1, \ldots, x_n$ *forman una muestra aleatoria de tamaño n de una distribución dada con media* μ *y varianza* σ^2*, con* $0 < \sigma^2 < \infty$*, entonces para cualquier número fijo x:*

$$\lim_{n\to\infty} \boldsymbol{Prob}\ \left(\frac{\sqrt{n}\,(\bar{x}_n - \mu)}{\sigma} \leq x\right) = \Phi(x)$$

Donde $\boldsymbol{\Phi}$ *denota la función de distribución de una normal estándar.*

Por tanto la distribución de $\bar{\mathbf{x}}_n$ *será aproximadamente Normal con media* μ *y varianza* σ^2/n*, o equivalentemente, la distribución de la suma* $\sum_{i=1}^{n} x_i$ *será aproximadamente una Normal con media* $\boldsymbol{n\mu}$ *y varianza*[3] *igual a* $\boldsymbol{n\sigma^2}$*.*

Otra versión del TCL (Liapounov) se establece para una secuencia de variables aleatorias independientes pero no necesariamente idénticamente distribuidas[4]*. El TCL se refiere a la*

[2] La Ley de los Grandes Números para Procesos Estacionarios en Covarianza se puede ver en Hamilton (1994).
[3] Var $\sum x_i$ /n = σ^2/n, entonces Var $\sum x_i$ = Var(n$\bar{x}$) = $n^2 * \sigma^2/n = n\sigma^2$.
[4] Ver, por ejemplo, Aris Spanos (1986).

distribución de la suma de un número creciente de variables aleatorias que individualmente no ejercen ningún efecto significante sobre el comportamiento de la suma.

Ley de los Grandes Números con Stata (La media muestral es un estimador consistente de la media poblacional). *Para ilustrar que la media muestral es un estimador consistente de la media poblacional se creó una variable χ^2 con 1 grado de libertad. Esta variable tiene una distribución con marcada asimetría positiva. El valor medio de la χ^2 es igual a sus grados de libertad y su varianza es 2 veces sus grados de libertad, de modo que se generarán con números aleatorios tres tamaños de muestra (n=5, n=50, n=250) de una χ^2 con media 1 y varianza 2.*

```
. program onesample, rclass
  1. drop _all
  2. quietly set obs 5
  3. generate xchi = rchi2(1)
  4. summarize x
  5. return scalar meanx = r(mean)
  6. end

. onesample
```

Variable	Obs	Mean	Std. Dev.	Min	Max
xchi	5	.6349084	.5041356	.0478238	1.192997

```
. program secsample, rclass
  1. drop _all
  2. quietly set obs 50
  3. generate xxchi = rchi2(1)
  4. summarize xx
  5. return scalar meanxx = r(mean)
  6. end

. secsample
```

Variable	Obs	Mean	Std. Dev.	Min	Max
xxchi	50	.9306599	1.384531	.0001955	6.371782

```
. program thirdsample, rclass
  1. drop _all
  2. quietly set obs 250
  3. generate xxxchi = rchi2(1)
  4. summarize xxx
  5. return scalar meanxxx = r(mean)
  6. end

. thirdsample
```

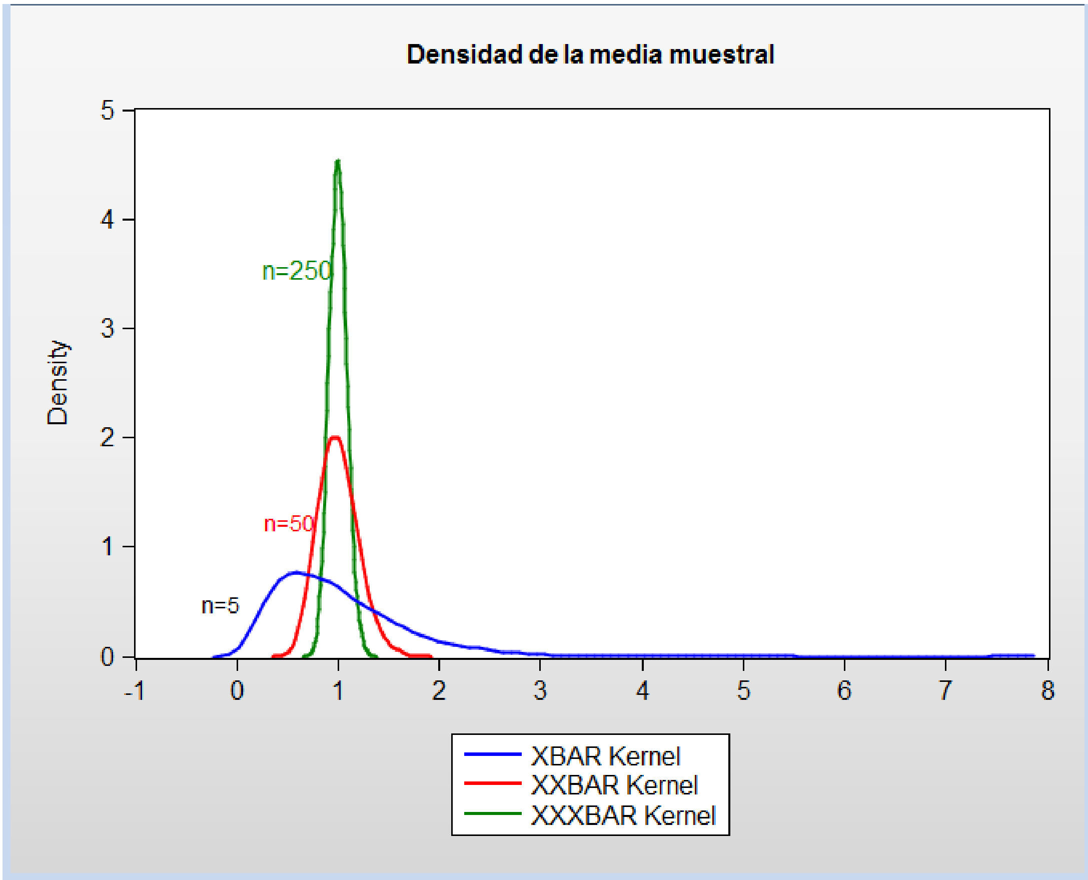
Variable	Obs	Mean	Std. Dev.	Min	Max
xxxchi	250	1.12057	1.35567	1.28e-06	8.855204

Variable	Núm. obs	Media	$\lvert\bar{x}-\mu\rvert$	Varianza est	|Varestim-2|
xchi	5	.634908	0.365	.2541	1.746
xxchi	50	.949422	0.051	1.905	0.095
xxxchi	250	1.10299	0.103	1.910	0.090

Teorema Central del Límite: *Si $\{x_i\}$ es una secuencia de variables aleatorias independientes e idénticamente distribuidas cada una con media μ y varianza σ^2, entonces la función de probabilidad para la media muestral converge a una normal con media μ y varianza σ^2/n. Para mostrar la convergencia en distribución de la media muestral, se generaron para cada tamaño de muestra 10,000 repeticiones, a fin de obtener una secuencia de la media muestral: $\{\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_{10000}\}$, así como la media y la varianza de la media muestral:*

	Obs	Media	$\bar{x}$ - μ	Desv.Estándar.	Varianza	Asimetría	Curtosis
xbar n=5	10,000	**.9975947**	-0.00241	.620956	**.3855864**	1.287337	6.27142
xxbar n=50	10,000	**.9985424**	-0.00146	.2001771	**.0400709**	.3919977	3.227154
xxxbar n=250	10,000	**1.000265**	0.000265	.0892111	**.0079586**	.1483689	3.05947

Del resultado se observa que la diferencia entre la media muestral y la media poblacional μ se va acercando a 0 y que la varianza de la media muestral $\boldsymbol{\sigma^2/n}$ *se va reduciendo con el tamaño de la muestra: 0.3855864 ≈ 2/5; 0.0400709 ≈ 2/50; 0.0079586 ≈ 2/250. Esto es, la media muestral converge a una distribución Normal con media μ y varianza* $\sigma^2/_n$.
Para ilustrar el Teorema Central del Límite se grafican las densidades de la media muestral generadas con 10,000 muestras aleatorias para los tres tamaños de muestra:



Inicialmente la distribución con n=5 exhibe la asimetría positiva de la χ^2 *con 1 grado de libertad. A medida que crece el tamaño de la muestra la distribución de la media muestral se va concentrando simétricamente alrededor de la media poblacional* $\mu=1$.

Comandos en Stata para generar los resultados presentados:

(Las líneas con asteriscos identifican los comentarios. Los comandos de Stata aparecen en negritas).

*Simulación de 10000 muestras para una variable Chi cuadrado con 1 grado de libertad

(A) n=5

Inicio de comandos

**Programa para obtener una muestra de tamaño 5 de una Chi cuadrada y devolver la media muestral*

. program onesample, rclass

**En la pantalla se despliega automáticamente un menú con números donde señalaremos los pasos del programa.*

1. drop _all
2. quietly set obs 5
3. generate xchi = rchi2(1)
4. summarize x
5. return scalar meanx = r(mean)
6. end

**Corremos el programa onesample para corroborar los resultados.*

. onesample

```
Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
xchi |         5    .6349084    .5041356   .0478238   1.192997
```

**Ejecución del programa onesample 10000 veces para obtener 10000 medias muestrales*

. simulate xbar = r(meanx), reps(10000) nodots: onesample

```
    command:  onesample
       xbar:  r(meanx)
```

**Pedimos un resumen de las 10000 medias muestrales*

. summarize xbar, detail

**Salvar el resultado ahora. Aparecerá en la ventana la ubicación donde queda guardado el archivo:*

file "C:\Users\Maria\Documents\Econometría II \Teorema CL1.dta" saved

(B) n=50

A continuación se repite el procedimiento pero con el tamaño de muestra n=50. Para no borrar las variables se abre Stata nuevamente, no se cierra la ventana donde tenemos el archivo del procedimiento con muestra n=5, para navegar fácilmente entre múltiples ventanas abiertas. Al abrir otra ventana de Stata se usan los comandos establecidos con anterioridad:

Inicio de comandos

**Programa para obtener una muestra de tamaño 50 de una Chi cuadrada y devolver la media muestral*

. program secsample, rclass

**En la pantalla se despliega automáticamente un menú con números donde señalaremos los pasos del programa.*

1. drop _all
2. quietly set obs 50
3. generate xxchi = rchi2(1)
4. summarize xx
5. return scalar meanxx = r(mean)
6. end

**Corremos el programa secsample para corroborar los resultados.*

. secsample

```
Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
   xxchi |        50    .9494227    1.380405   .0034754   6.371782
```

**Ejecución del programa secsample 10000 veces para obtener 10000 medias muestrales*

. simulate xxbar = r(meanxx), reps(10000) nodots: secsample

```
      command:  secsample
        xxbar:  r(meanxx)
```

**Pedimos un resumen de las 10000 medias muestrales*

. summarize xxbar, detail

**Salvar el resultado ahora. Aparecerá en la ventana la ubicación donde queda guardado el archivo:*

file "C:\Users\Maria\Documents\EconometríaII\TeoremaCL2.dta" saved

(C) n=250

A continuación se repite el procedimiento para el tamaño de muestra n=250. Abrimos Stata nuevamente, no se cierra la ventana donde tenemos el archivo del procedimiento con muestra n=50 si queremos navegar fácilmente entre múltiples ventanas. Al abrir otra ventana de Stata se usan los comandos establecidos con anterioridad:

Inicio de comandos

**Programa para obtener una muestra de tamaño 250 de una Chi cuadrada y devolver la media muestral*

. program thirdsample, rclass

**En la pantalla se despliega automáticamente un menú con números donde señalaremos los pasos del programa.*

1. drop _all
2. quietly set obs 250
3. generate xxxchi = rchi2(1)
4. summarize xxx
5. return scalar meanxxx = r(mean)
6. end

**Corremos el programa thirdsample para corroborar los resultados.*

. thirdsample

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      xxxchi |       250    1.102998    1.382137   1.28e-06   8.855204
```

** Ejecución del programa thirdsample 10000 veces para obtener 10000 medias muestrales*

. simulate xxxbar = r(meanxxx), reps(10000) nodots: thirdsample

```
      command:  thirdsample
       xxxbar:  r(meanxxx)
```

**Pedimos un resumen de las 10000 medias muestrales*

. summarize xxxbar, detail

**Salvar el resultado ahora. Aparecerá en la ventana la ubicación del archivo guardado:*

save "C:\Users\Maria\Documents\Econometría II \Teorema CL3.dta"
file C:\Users\Maria\Documents\Econometría II\Teorema CL3.dta saved

*Una vez que los tres archivos están salvados podemos unirlos. En este caso desde la ventana que queramos, por ejemplo desde la ventana con Teorema CL3.dta. Para unirlos usamos el comando: *merge using ubicación del archivo*

En la instrucción merge es importante incluir las comillas al dar la ruta para el archivo que se va a unir.

. merge using "C:\Users\Maria\Documents\Econometría II\CL1.dta"

*Una vez unidos dos archivos, salvamos el nuevo archivo

. save "C:\Users\Maria\Documents\Econometría II\CL3.dta", replace
file C:\Users\Maria\Documents\Econometría II\CL3.dta saved

**Al unir los archivos se crea una variable _merge, la eliminamos:*

. drop _merge

**Unimos el archivo CL2:*

. merge using "C:\Users\Maria\Documents\Econometría II\CL2.dta"

**Una vez unidas las 3 bases de datos se puede proseguir para obtener la gráfica conjunta:*

graph twoway (histogram xbar) (histogram xxbar) (histogram xxxbar) (kdensity xbar) (kdensity xxbar) (kdensity xxxbar), title(Densidad de la media muestral).

Referencias:

Hamilton, J.D. (1994), <u>Times Series Analysis</u>, Princeton University Press.

Hayashi, Fumio (2000), <u>Econometrics</u>, Princeton University Press.

Larson, H.J. (1982), <u>Introduction to Probability Theory and Statistical Inference</u>, Wiley Series in Probability and Mathematical Statistics.

Spanos, Aris. (1986), <u>Statistical Foundations of Econometric Modeling</u>, Cambridge University Press.

Cameron A. Colin & Trivedi Pravin K. (2010), Microeconometrics Using Stata, Stata Press Publication.

Febrero/2016