Artículo Original
Revista Varianza - Año 2022, Nro. 20
ISSN 2789-3510 versión impresa
DOI: https://doi.org/10.53287/uufr9587pp15b
Páginas: 35-42

EL TEOREMA CENTRAL DEL LÍMITE: UNA VISIÓN ILUSTRATIVA

THE CENTRAL LIMIT THEOREM: AN ILLUSTRATIVE VIEW

Raúl León Delgado Álvarez [1]
Carrera de Estadística, Universidad Mayor de San Andrés, La Paz-Bolivia
✉ rldelgado3@umsa.bo

Artículo recibido: 17-08-2022
Artículo aceptado: 06-09-2022

RESUMEN

El Teorema Central del Límite es de gran importancia, en el estudio de la Estadística, para su compresión además de la variedad de problemas prácticos, está la certeza de su convergencia.

Objetivo: Comprender el Teorema Central del Límite, y su convergencia.

Método: Se utilizarán cuadros y gráficos de la distribución del promedio de variables aleatorias independientes cada una con distribución Uniforme en el intervalo (0,1) y comparándola con la distribución Normal standard.

Datos: Para este fin se utilizará una partición del intervalo comprendido entre 0 y 1, en distancias de un décimo.

Resultados: Se mostrará la rápida convergencia del promedio de las variables citadas con distribución Uniforme hacia una distribución Normal estándar; para esta comparación se consideran dos cuadros que mostrarán esta afinidad de forma numérica y posteriormente (para cuando n aumente su valor) se expondrá el gráfico de la distribución del promedio de variables que completará el panorama.

Palabras clave: Convergencia, distribución Irwin Hall, ilustración, Teorema Central del Límite.

ABSTRACT

The Central Limit Theorem is of great importance, in the study of Statistics, for its understanding, in addition to the variety of practical problems, there is the certainty of its convergence.

Objective: To understand the Central Limit Theorem, and its convergence.

Method: Tables and graphs of the distribution of the average of independent random variables will be used, each with a Uniform distribution in the interval (0.1) and comparing it with the standard Normal distribution.

Data: For this purpose, a partition of the interval between 0 and 1, in distances of one tenth, will be used.

Results: The rapid convergence of the average of the variables mentioned with Uniform distribution towards a standard Normal distribution will be shown; For this comparison, two tables are considered that will show this affinity numerically and later (for when n increases its value) the graph of the distribution of the average of variables will be exposed that will complete the panorama.

Key words: Convergence, Irwin Hall distribution, illustration, Central Limit Theorem.

[1] Docente emérito, Carrera de Estadística Universidad Mayor de San Andrés, La Paz Bolivia
Orcid: 0000-0001-7886-5264

I. INTRODUCCIÓN

El proceso de enseñanza aprendizaje de la Estadística, amerita reflexionar sobre, cuan útiles son las segmentaciones de esta disciplina, traducidas en asignaturas que marcan contenidos secuenciales y cuáles son las mejores estrategias que deberían utilizar los docentes para conseguir un aprendizaje significativo en el colectivo estudiantil.

Una parte sin duda importante, es la comprensión del enunciado teórico de un tema, que se realiza a la luz del método científico, utilizando axiomas, lemas, teoremas, y corolarios cuyas demostraciones se desarrollan en clase.

Otra parte es el argumento ilustrativo en la enseñanza para fijar los conocimientos, cómo son los cuadros y gráficos que sin duda son elementos que ayudan a este cometido.

El Teorema Central del límite, se desarrolla haciendo énfasis en una de sus principales características como es su convergencia, la que muestra a través del comportamiento de la distribución del promedio aritmético de variables aleatorias. El teorema es muy fuerte en sus aplicaciones dado que ni siquiera es necesario conocer la distribución de las variables. La distribución Uniforme es un ejemplo ideal de una distribución simple y se puede considerar genérica pues, cada una de las variables aleatorias con distribución de densidad de probabilidades se encuentra en el intervalo (0,1) a diferencia de la distribución Beta que cuya gráfica ya muestra una comparación con la Normal.

2. PROBLEMA

En estudios realizados por (Johnson, Kotz, & Balakrishnan, 1995, pág. 296), en el libro *Continuous Univariate Distributions*, los autores citan a la distribución de Irwin Hall, para explicar la suma de variables aleatorias independientes, cada una con distribución Uniforme en el intervalo (0,1), que se constituye en un aporte para la comprensión del Teorema Central del Límite. Otro aspecto didáctico, se puede observar en el texto denominado Estadística Informatizada, (Barrera Ojeda, 2019), que hace referencia al aprendizaje utilizando programas informáticos.

3. OBJETIVO

Comprender el Teorema Central del Límite, y su convergencia, a través de cuadros y gráficos de la distribución del promedio de variables aleatorias independientes cada una con distribución Uniforme en el intervalo (0,1) y comparándola con la distribución Normal standard.

4. DESARROLLO

A diferencia del concepto de convergencia en sucesiones de números reales, en las sucesiones de variables aleatorias se debe considerar que las imágenes son números reales, pero siendo las variables aleatorias funciones de un espacio muestral hacia un conjunto numérico, la diferencia de una variable genérica X_n de la sucesión menos una variable aleatoria X, tanto por exceso como por defecto que sea menor que un ϵ, es un suceso que tendría cierta probabilidad de no ocurrir (Obregón Sanin, 1977, pág. 247).

Se denomina convergencia en probabilidad si:

Dado un $\epsilon > 0$ $\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$ (1)

Se denomina convergencia casi segura si:

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1 \qquad (2)$$

La convergencia casi segura, implica

convergencia en probabilidad, pero el reciproco no es cierto. Es decir la convergencia casi segura es más fuerte que la convergencia en probabilidad. La Ley de los grandes números afirma que si X es una variable aleatoria cuya media existe, y si X_1, $X_2,\ldots, X_n$ son observaciones independientes de ella, entonces $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ converge casi seguro a $E(X)$, esta versión se conoce como la Ley fuerte de los grandes números, si la convergencia es en probabilidad se denomina Ley débil de los grandes números.

El Teorema Central del Límite

El Teorema Central del Límite se enunciará en su forma general sin la demostración, dado que se pretende buscar su comprensión.

Sean X_1, $X_2,\ldots, X_n$, variables aleatorias no necesariamente independientes en el supuesto que las varianzas de las X_i sean finitas y que un número finito de ellas no predomine sobre las demás.

Definida por

$$Y_n = \frac{\sum_{i=1}^{n} X_i - E\left(\sum_{i=1}^{n} X_i\right)}{D\left(\sum_{i=1}^{n} X_i\right)} \qquad (3)$$

Donde E (.), es el valor esperado y D (.), la desviación estándar de la suma de variables aleatorias.

Y_n converge en probabilidad a una variable aleatoria con distribución Normal estándar.

Expresiones más restringidas son aquellas en que la independencia de las variables aleatorias, y su distribución común con media μ y varianza σ^2 nos permiten establecer que la sucesión:

$$Y_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \qquad (4)$$

Y la más conocida, de aplicaciones más variadas es:

$$Y_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \qquad (5)$$

Se interpreta que Y_n converge en probabilidad a una variable Normal estándar, esta última afirmación podría interpretarse como sigue:

Si $F_n(z)$, es la función de distribución acumulada del término genérico de la sucesión de variables y $\Phi(z)$ es la función de distribución acumulada de la densidad Normal estándar, entonces:

$$\lim_{n\to\infty} \frac{F_n(z)}{\Phi(z)} = 1 \qquad (6)$$

O equivalentemente:

$$\begin{aligned} &\lim_{n\to\infty} F_n(z) = \Phi(z) \\ &\lim_{n\to\infty} F_n(z) - \Phi(z) = 0 \end{aligned} \qquad (7)$$

Es decir, que cuando n crezca más, las funciones acumuladas tenderán a ser iguales, este hecho puede observarse en el comportamiento del promedio de distribuciones uniformes, pero ya desde el promedio de dos variables, se observa esta convergencia y más aún con el promedio de tres o más variables.

La manera de la aproximación para una n finita, depende de las características de las distribuciones originales, pero en general es sorprendente la aproximación para una n pequeña en esta distribución.

Para mostrar este hecho, hacemos el análisis del promedio aritmético o media aritmética de variables aleatorias independientes con distribución común Uniforme en el intervalo (0,1).

Si se consideran dos variables aleatorias X_1, X_2 con distribución Uniforme en el intervalo

(0,1) e independientes, a través del método de transformación u otro método, se puede probar que su promedio $z = \frac{1}{2}(X_1 + X_2)$ sigue la densidad $f(z)$ y su acumulada es $F_2(z)$

$$f(z) = \begin{cases} 4z, & 0 \le z < \frac{1}{2} \\ 4(1-z), & \frac{1}{2} \le z < 1 \\ 0, & en\ otra\ parte \end{cases},$$

Con

$$F_2(z) = \begin{cases} 0, & z < 0 \\ 2z^2, & 0 \le z < \frac{1}{2} \\ 4z - 1 - 2z^2, & \frac{1}{2} \le z < 1 \\ 1, & z \ge 1 \end{cases} \quad (8)$$

Los valores $E(z) = \frac{1}{2}$ y $V(z) = \frac{1}{24}$, se utilizarán en la estandarización de z.

Se recordará que $F_2(z)$ es la función de distribución acumulada exacta del promedio de dos variables aleatorias independientes con distribución uniforme en el intervalo (0,1).

Cuadro No. 1

Comparación entre distribuciones acumuladas (*n*=2)

z	F_2 (z)	Φ(z,0,5;0,2041)	Error
0,0000	0,0000	0,0071	-0,0071
0,1000	0,0200	0,0250	-0,0050
0,2000	0,0800	0,0708	0,0092
0,3000	0,1800	0,1635	0,0165
0,4000	0,3200	0,3121	0,0079
0,5000	0,5000	0,5000	0,0000
0,6000	0,6800	0,6879	-0,0079
0,7000	0,8200	0,8365	-0,0165
0,8000	0,9200	0,9292	-0,0092
0,9000	0,9800	0,9750	0.0050
1.0000	1,0000	0,9929	0,0071

Fuente: Elaboración propia

Claramente el mayor error entre las distribuciones exacta y normal acumuladas, apenas excede el orden de los centésimos, este error es ínfimo sí se considerarán tres variables con distribución Uniforme en el intervalo (0,1) y siendo independientes, entre sí se puede hallar de forma equivalente la función de densidad del promedio y su distribución acumulada como:

$$f(z) = \begin{cases} \frac{27z^2}{2}, & 0 \le z < \frac{1}{3} \\ -27z^2 + 27z - \frac{9}{2}, & \frac{1}{3} \le z < \frac{2}{3} \\ \frac{27}{2}z^2 + \frac{27}{2} - 27z, & \frac{2}{3} \le z < 1 \\ 0, & en\ otra\ parte \end{cases}$$

$$F_3(z) = \begin{cases} 0, & z < 0 \\ \frac{9}{2}z^2, & 0 \le z < \frac{1}{3} \\ -9z^3 + \frac{27}{2}z^2 - \frac{9}{2}z + \frac{1}{2}, & \frac{1}{3} \le z < \frac{2}{3} \\ 1 - \frac{9}{2}(1-z)^3 & \frac{2}{3} \le z < 1 \\ 1, & z \ge 1 \end{cases} \quad (9)$$

Considerando los valores de $E(z) = \frac{1}{2}$, $V(Z) = \frac{1}{36}$, $D(Z) = \frac{1}{6}$, se realiza:

Cuadro No. 2

Comparación entre distribuciones acumuladas (*n*=3)

z	F_3 (z)	Φ(z,0,5;0,1667)	Error
0,0000	0,0000	0,0013	-0,0013
0,1000	0,0045	0,0082	-0,0037
0,2000	0,0360	0,0359	0,0001
0,3000	0,1215	0,1151	0,0064
0,4000	0,2840	0,2742	0,0098
0,5000	0,5000	0,5000	0,0000
0,6000	0,7160	0,7258	-0,0098
0,7000	0,8785	0,8849	-0,0064
0,8000	0,9640	0,9641	-0,0001
0,9000	0,9955	0,9918	0,0037
1.0000	1,0000	0,9987	0,0013

Fuente: Elaboración propia

Del Cuadro No. 2 se obtiene que el máximo error cometido al tomar la función de distribución de Z_3 como normal, es menor que 0,01, es decir, un error del 1%.

El cálculo numérico del error de estimación máximo no es el objetivo de este artículo, pero es un trabajo que puede lograr para el lector que quiere profundizar en este tema.

Lo observado anteriormente, conduce a la afirmación de que cuanto más tiende a crecer

n, la aproximación a la distribución Normal de la distribución del promedio de variables aleatorias independientes con distribución Uniforme en el intervalo (0,1) es más visible, por la interpretación de los Cuadros No. 1 y 2, sin embargo, para interpretar que n crece grandemente, se hará uso de la distribución de Irwin Hall para el promedio de variables independientes con distribución Uniforme, entonces la distribución del promedio indicada, se obtiene como:

Siendo: $z = \frac{1}{n}(X_1 + X_2 + \cdots \quad X_n)$, su función de densidad de probabilidad es:

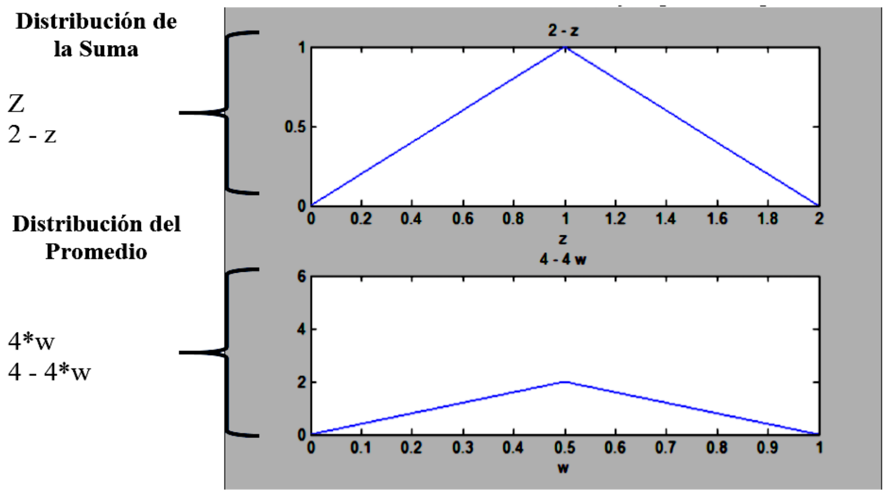
$$f(z) = \frac{n}{(n-1)!}\sum_{j=0}^{k}\binom{n}{j}(-1)^j(nz-j)^{n-1} \quad (10)$$

Donde $\frac{k}{n} \leq z < \frac{k+1}{n}$, $\quad k = 0, 1, 2, \ldots, (n-1)$

En el entendido de proseguir con la visión ilustrativa, se ha desarrollado un programa en MATLAB que muestra la expresión de la distribución exacta y la gráfica de la misma, para su comparación..

Gráfico No. 1

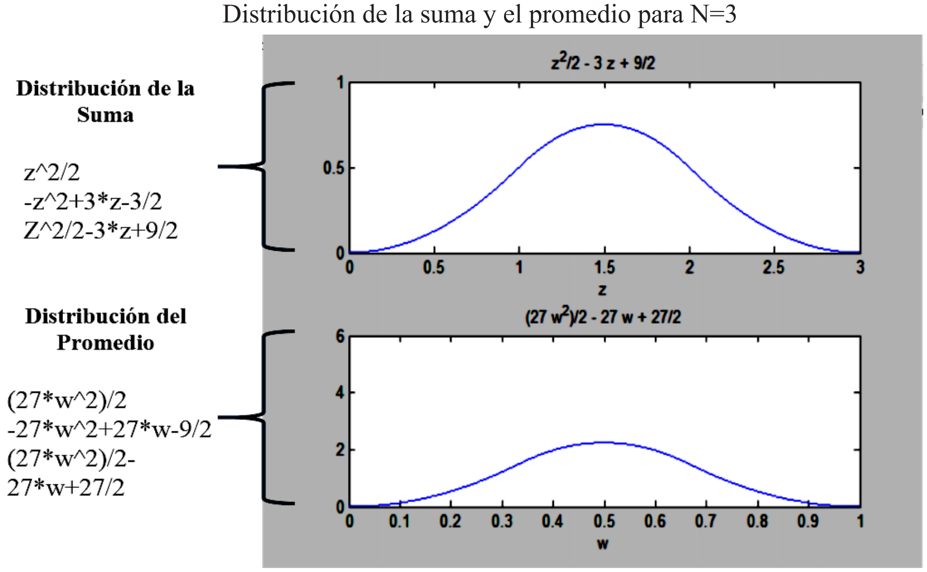
Distribución de la suma y el promedio para N=2



Fuente: Elaboración propia.

Gráfico No. 2

Distribución de la suma y el promedio para N=3



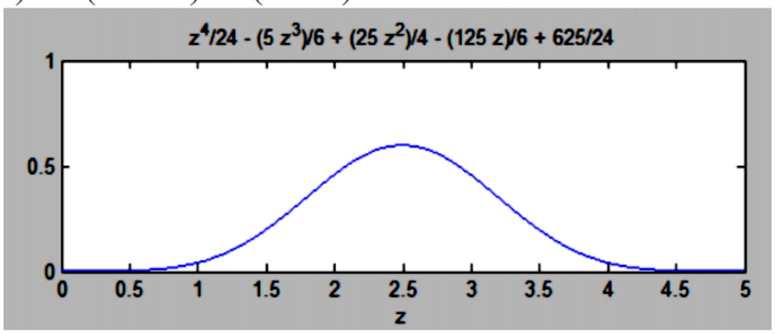
Fuente: Elaboración propia.

Gráfico No. 3

Distribución de la suma y el promedio para N=5

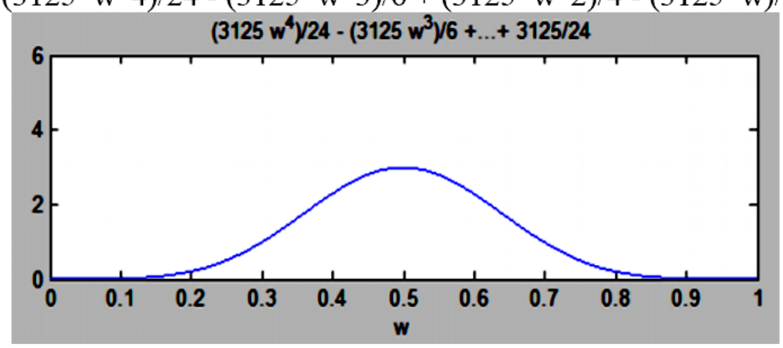
Distribución de la Suma

z^4/24
- z^4/6 + (5*z^3)/6 - (5*z^2)/4 + (5*z)/6 - 5/24
z^4/4 - (5*z^3)/2 + (35*z^2)/4 - (25*z)/2 + 155/24
- z^4/6 + (5*z/^3)/2 - (55*z^2)/4 + (65*z)/2 - 655/24
z^4/24 - (5*z^3)/6 + (25*z^2)/4 - (125*z)/6 + 625/24



Distribución del Promedio

(3125*w^4)/24
- (3125*w^4)/6 + (3125*w^3)/6 - (625*w^2)/4 + (125*w)/6 - 25/24
(3125*w^4)/4 - (3125*w^3)/2 + (4375*w^2)/4 - (625*w)/2 + 775/24
- (3125*w^4)/6 + (3125*w^3)/2 - (6875*w^2)/4 + (1625*w)/2 - 3275/24
(3125*w^4)/24 - (3125*w^3)/6 + (3125*w^2)/4 - (3125*w)/6 + 3125/24



Fuente: Elaboración propia.

5. RESULTADOS.

En las Cuadros No. 1 y 2 se puede observar una columna de errores, que muestran una diferencia muy pequeña entre las funciones acumuladas exactas y la distribución Normal, este análisis implica que el error absoluto, para n=2, es casi igual a 0,01, la convergencia se nota más cuando para n=3, hacemos la comparación y el error es menor que 0,01, para n mayor se realiza la descripción de la función de densidad del promedio de n variables aleatorias, a través de la fórmula de Irwin Hall y haciendo una transformación que permite mostrar, la cercanía entre la densidad del promedio con la distribución Normal estándar Gráficas No. 1, 2 y 3.

6. DISCUSIÓN

El proceso de enseñanza y aprendizaje en la universidad tiene una vivencia día a día en el aula que es función del contenido científico que el docente pretende que se apropie el estudiante, en esta tendencia la estrategia para este fin cambia con mucha frecuencia de clase en clase, es así que esta ilustración pretende afirmar el conocimiento que debe tener el estudiante acerca del Teorema Central del Límite.

Seguramente hay otras estrategias para la enseñanza de este teorema y uso de nuevos recursos, la visión ilustrativa es una alternativa para la compresión de este teorema.

7. CONCLUSIÓN

Para la comprensión del Teorema Central del Límite, se mostró que la distribución exacta del promedio de variables aleatorias independientes con distribución Uniforme en el intervalo (0,1) tiende a la distribución Normal estándar, incluso con valores de n pequeños, como se vio en los Cuadros No. 1 y 2 y en las Gráficas No. 1, 2 y 3.

REFERENCIA BIBLIOGRÁFICA

Barrera Ojeda, D. (2019). *Estadística Informatizada*. Stigma.

Bernard, F. (1997). *A first course in Multivariate Statistic*. Springer.

Billingsley, P. (1986). *Probability and Measure*. John Wiley and Sons.

Casella George y Berger Roger. (2002). *Statistical Inference.* Duxbury Thomson.

Dela Cruz, R. (1981). *Razonamiento Matemático Superior.* Lima Perú.

Efimov A, Karakulin A., Pospelov P., Tereschenko A., Vukolov E., Zemskov V., Zolotarev Yu. (1986). *Problemas de las Matemáticas Superiores*. Mir Moscu.

Fernández Cuesta y Fuentes García, Felipe. (1995). *Curso de Estadística Descriptiva Teoría y Práctica*. Barcelona: Ariel.

Freund Jhon y Walpole, Ronald . (1990). *Estadística Matemática con Aplicaciones.* Prentice Hall hispanoamericana.

García Ore, C. (1997). *Distribuciones y Estadística Inferencial.* Lima Peru.

Grimett, G. R., & Stirzaker, D. (1993). *Probability and Random Processes Problems and Solutions*. Oxford University.

Gutierrez, B. (2004). *Respuestas a dudas típicas de Estadística*. Diaz de Santos.

Hogg, R. (2005). *Introducción to Mathematical Statistics.* Pearson Prentice Hall.

Johnson, N., Kotz, S., & Balakrishnan, N. (1995). *Continuous Univariate Distributions.* Wiley an Sons.

Kendall Maurice y Buckland William. (1980). *Diccionario de Estadística*. Pirámide.

Kreyszig, E. (1973). Introducción a la Estadística Matemática. Limusa Wiley.

Lazaro, M. (2007). *Inferencia Estadística*. Moshera.

Lijoletov I.I.y Matzkevich I.P. (1977). *Problemas de Matemática Superior Teoría de Probabilidad y de Estadística Matemática*. Paraninfo.

Meyer, P. L. (1998). *Probabilidad y Aplicaciones Estadísticas*. Paraninfo.

Mitacc Meza, M. (1998). *Tópicos de Estadística descriptiva y Probabilidad*. Lima Peru.

Mitacc, M. (1998). *Tópicos de Inferencia estadística.* Lima Perú: Thales.

Mood Alexander y Graybill Franklin. (1978). *Introducción a la teoría de la Estadística*. Aguilar.

Obregón Sanin, I. (1977). *Teoría de la Probabilidad.* p. 247: Limusa.

Parzen, E. (1979). *Teoría Moderna de Probabilidades y sus Aplicaciones*. Limusa.

Rincon, L. (2007). *Curso Intermedio de Probabilidad*. UNAM.

Ruiz Maya , L., & Martin Pliego , J. (2002). *Estadística II Inferencia*. Thomson.

Ruiz Maya, L. (1992). *Problemas de Estadística*. AC Madrid.

Sevastianov, B., Chistiakov , V., & Zubkov. (1985). *Problemas de Cálculo de Probabilidades.* Mir Moscu.

Tucker, H. G. (1966). *Introducción a la teoría matemática de las probabilidades y a la estadística.* Vinces vives.

Velez Barrola, R. (2012). *Principios de Inferencia Estadística*. Madrid: Uned.

Wisniewski Piotr Marian y Velasco Sotomayor Gabriel, Blanco Castaneda Liliana , Arunachalam Viswanathan y Dharmaraja Selvamuthu. (2012). *Introduction to Probability and Stochastic Processes*. Wiley and Sons.

Yohai, V. (2008). *Notas de Probabilidades y Estadística*.

Zacks, S. (2014). *Examples and Problems in Mathematical Statistics*. Wiley.