

*SIGNIFICADO DEL TEOREMA CENTRAL DEL LIMITE EN TEXTOS  
UNIVERSITARIOS DE PROBABILIDAD Y ESTADISTICA*

Meaning of Central Limit Theorem in University Statistics and Probability Textbooks

*Hugo Alvarado<sup>1</sup>, Carmen Batanero<sup>2</sup>*

<sup>1</sup>Universidad Católica de la Santísima Concepción, alvaradomartinez@ucsc.cl

<sup>2</sup>Universidad de Granada, batanero@ugr.es

**Resumen**

En este trabajo se analiza la presentación del teorema central del límite en una muestra de libros de textos de estadística destinados a la formación de ingenieros. Siguiendo un modelo teórico sobre el significado de un objeto matemático, los textos más utilizados en la enseñanza a ingenieros en Chile muestran una gran riqueza de lenguaje, herramientas de resolución de problemas, conceptos asociados, propiedades y tipos de argumentos que le confieren una gran complejidad y posibilita presentaciones muy diferentes a los alumnos. Como consecuencia se deducen algunos criterios para el diseño de futuras propuestas de enseñanza, como la implementación de diferentes formulaciones y representaciones de la suma de variables aleatorias según su naturaleza, en orden creciente de complejidad.

*Palabras clave:* teorema central del límite, significado de objetos matemáticos, análisis de libros de texto.

**Abstract**

In this paper we analyse the presentation of the central limit theorem in a sample of statistics textbooks for engineers. Our work is based on a theoretical model on the meaning of mathematical objects. The more employed textbooks in Chile present a great wealth of language, tools for solving problems, associated representations, problems fields and arguments that confer a great complexity and allow several quite different presentations for the students. As a consequence of such an analysis we infer some criteria for the future designing of teaching proposals, like the implementation of different formulations and representations of the sum of random variables according to its nature and in an increasing order of complexity.

*Key words:* central limit theorem, meaning of mathematical objects, analysing textbooks.

## INTRODUCCION

El teorema central del límite, uno de los fundamentales en estadística, estudia el comportamiento de la suma de variables aleatorias, cuando crece el número de sumandos, asegurando su convergencia hacia una distribución normal en condiciones muy generales. Este teorema, del cual existen diferentes versiones que se han ido desarrollando a lo largo de la historia, tiene una gran aplicación en inferencia estadística, pues muchos parámetros de diferentes distribuciones de probabilidad, como la media, pueden expresarse en función de una suma de variables. Permite también aproximar muchas distribuciones de uso frecuente: binomial, Poisson, chi cuadrado, t-student, gamma, etc., cuando sus parámetros crecen y el cálculo se hace difícil (Devore 2001: 232). Por otro lado, la suma de variables aleatorias aparece en forma natural en muchas aplicaciones de la ingeniería: determinación de masa forestal, carga soportada por una estructura, tiempo de espera de servicios, etc.

Todo ello explica por qué muchos métodos estadísticos requieren la condición de normalidad para su correcta aplicación y, en consecuencia, este teorema es un componente importante de la formación estadística de los ingenieros, ya que, por otro lado, su enseñanza plantea interrogantes importantes al profesor. El teorema se apoya y relaciona entre sí con otros conceptos y procedimientos básicos en estadística, como los de variable aleatoria y sus transformaciones, distribución muestral, convergencia, tipificación, cálculo de probabilidades, etc., algunos de los cuales podrían plantear problemas de aprendizaje.

En este trabajo se analiza el significado del teorema central del límite en los libros de texto de probabilidad y estadística dirigidos a ingenieros, siguiendo la misma metodología en análisis del significado de otros conceptos estadísticos en los libros de texto (Tauber 2001; Cobo y Batanero 2004). En el modelo de la teoría de los significados institucionales y personales de los objetos matemáticos (Godino y Batanero 2003), las matemáticas se asumen como una actividad humana implicada en la solución de cierta clase de situaciones problemáticas de la cual emergen y evolucionan progresivamente los objetos matemáticos.

El análisis de los textos, aunque no sustituye la observación de la enseñanza en el aula, puede proporcionar información para la construcción de instrumentos de evaluación o para mejorar la enseñanza. Se han encontrado significados sesgados en la presentación de la probabilidad en algunos libros (Ortiz, Serrano y Batanero 2002). Considerando las investigaciones relacionadas con la enseñanza del teorema se estudia la siguiente hipótesis: *es complejo el significado del teorema central del límite presentado en los textos de estadística aplicada a la ingeniería y se encontrarán una variedad de enfoques y aproximaciones*. Asimismo, puede ser un recurso para los profesores, contribuyendo a tomar decisiones sobre cuáles pueden recomendar a sus alumnos, proporcionándoles ideas para enriquecer su actividad docente en el aula y mostrando algunas dificultades que podrían tener los estudiantes al estudiar estos textos.

## FUNDAMENTOS

*Marco teórico.* La perspectiva didáctica que se empleará está basada en el modelo teórico denominado “enfoque ontosemiótico” propuesto por Godino y sus colaboradores (Godino

2002; Godino y Batanero 2003; Godino, Batanero y Font 2007; Godino, Contreras y Font 2006). Este enfoque teórico proporciona una perspectiva pragmático-antropológica sobre el conocimiento matemático y propone tres dimensiones en el análisis de la enseñanza y el aprendizaje de las matemáticas: epistemológica, cognitiva e instruccional. Cada una de ellas se aborda con herramientas agrupadas en tres modelos teóricos: teoría de los significados institucionales y personales de los objetos matemáticos, teoría de las funciones semióticas y teoría de las configuraciones didácticas. Se pretende elaborar un modelo de los procesos de comprensión de las matemáticas que tenga en cuenta los factores institucionales y socioculturales implicados en los mismos. Se considerará para el análisis del teorema central del límite la siguiente tipología de objetos matemáticos primarios, denominada “elementos del significado” y que a su vez se organizan en sistemas conceptuales, teorías, etc.

- *Situaciones-problemas*: Situaciones fenomenológicas que originan actividades matemáticas (situaciones-problemas, aplicaciones) de donde surge el objeto; a veces las podemos categorizar en “tipos” o “campos” de problema. Por ejemplo, para el teorema central del límite serían situaciones problemas la “búsqueda de una aproximación a la distribución binomial cuando  $n$  es grande” o “determinación de la distribución de la suma o la media de variables aleatorias”.
- *Lenguaje*: Representaciones materiales utilizadas en la actividad matemática. Las notaciones, gráficos, palabras y otras representaciones del objeto que se pueden usar para referirnos a él. El lenguaje es esencial en la teoría del aprendizaje debido a su función comunicativa e instrumental, que modifica el propio sujeto que los utiliza como mediadores.
- *Procedimientos*: Modos de actuar ante situaciones o tareas (algoritmos, operaciones, reglas de cálculo). Cuando un sujeto se enfrenta a un problema y trata de resolverlo o comunicar la solución a otras personas, validar y generalizar la solución a otros contextos y problemas, etc., realiza distintos tipos de acciones que se llegan a algoritmizar.
- *Conceptos-definición*: (introducidos mediante definiciones o descripciones, por ejemplo: media, distribución muestral, ...). El caso que aborda este trabajo, y puesto que el objeto a estudiar es un teorema considerará sus diversos enunciados en esta categoría, ya que se dan las descripciones del objeto. Se podrían también incluir en esta categoría las definiciones de objetos ligados al teorema, como “distribución muestral”, “muestra”, etc. Pero, por limitar la investigación, ésta no se centra específicamente en ellas, aunque es posible en un momento dado referirse a alguno.
- *Proposiciones*: Se tratarán específicamente en este trabajo las propiedades asociadas al teorema central del límite y objetos relacionados, que no se limitan a descripciones de dichos objetos sino los ponen en relación. Aparte del propio teorema (que se ha incluido en la categoría anterior), aparecerán propiedades tales como las referidas a la media o varianza de la suma de variables aleatorias o la corrección de continuidad.
- *Argumentos*: Finalmente, todas estas acciones y objetos se ligan entre sí mediante argumentos o razonamientos que se usan para comprobar las soluciones de los problemas o explicar a otro la solución. La forma usual de demostración en matemáticas es la deductiva, que es la más extendida en los libros universitarios.

Este tipo de argumentación se completa o sustituye por otras como la búsqueda de contraejemplos, generalización, análisis y síntesis, simulaciones con ordenador, demostraciones, etc.

*Investigaciones sobre el teorema central del límite.* Las investigaciones psicológicas y didácticas han llamado la atención sobre los errores frecuentes en la realización e interpretación de la estadística, en relación al incumplimiento de los supuestos para aplicar la inferencia, intervalos de confianza y con los contrastes de hipótesis (Vallecillos 1996; Batanero 2000). Muchos de estos errores tienen su origen en una comprensión deficiente del concepto de distribución muestral y del teorema central del límite, pero la investigación didáctica sobre la enseñanza y aprendizaje de conceptos avanzados de estadística es todavía muy escasa.

Por otro lado, se han comparado las creencias sobre el teorema en estudiantes de doctorado y alumnos novales, representando el conjunto de conocimientos implícito en el teorema por medio de un mapa conceptual, estableciendo cuatro propiedades básicas que deben entenderse para poder lograr una comprensión sólida del teorema (Méndez 1991): a) La media de la distribución muestral es igual a la media de la población, e igual a la media de una muestra cuando el tamaño de la muestra tiende al infinito; b) La varianza de la distribución muestral es menor que la de la población (cuando  $n > 1$ ); c) La forma de la distribución muestral tiende a ser acampanada a medida que se incrementa el tamaño muestral, y aproximadamente normal, independientemente de la forma de la distribución en la población; d) La forma de la distribución muestral crece en altura y decrece en dispersión a medida que el tamaño muestral crece. La enseñanza basada en la simulación, utilizando un programa de elaboración propia, sugiere que la tecnología por sí sola no es suficiente para la comprensión del teorema, sino que las actividades y la enseñanza de tipo constructiva juega un papel relevante (delMas, Garfield y Chance 1999). También se han descrito los múltiples elementos de significado relacionados con la distribución normal, que son necesarios para la comprensión del teorema, puesto que en éste es necesario el uso de esta distribución (Tauber 2001).

Los anteriores trabajos, aunque proporcionan un primer análisis, no completan el estudio del significado del teorema central del límite, que es el objeto de esta investigación. El estudio de la evolución histórica del teorema (Alvarado 2004) revela diferentes versiones del mismo desarrolladas, entre otros, por de Moivre, Laplace, Poisson, Dirichlet, Bessel, Chebyshev, Markov, Liapounov y Féller para resolver diferentes campos de problemas. Su utilidad principal es proporcionar esta información a los profesores, tanto para planificar la enseñanza y evaluación del tema como para fundamentar la enseñanza posterior de la estimación por intervalos y los contrastes de hipótesis.

## METODOLOGIA

Se eligieron 16 textos que se listan en el Apéndice, de manera que se contemplasen distintas categorías y teniendo en cuenta los que aparecen en las bibliografías recomendadas para estudiantes de ingeniería, principalmente de Chile. Para elegir esta muestra se tomaron diferentes tipos de libros: a) textos de estadística aplicada a la ingeniería;

b) textos de probabilidad y estadística matemática; c) textos clásicos de autores de prestigio; d) textos recientes de enfoque novedoso, y e) textos con énfasis en los ejercicios y problemas.

Sobre ellos se llevó a cabo un análisis de contenido, que sirve para efectuar inferencias mediante la identificación sistemática y objetiva de las características específicas de un texto buscando el significado del discurso. Es un proceso complejo, seguramente el que más esfuerzo intelectual requiere de entre todas las técnicas de análisis de datos y es uno de los pocos campos de los comprendidos en las etapas finales del proceso de investigación en la que el investigador desempeña un papel individual y creativo (Ghiglione y Matalón 1991). En la línea de didáctica de la probabilidad y estadística un libro de texto se considera como un segundo nivel de transposición didáctica, después del primer nivel que lo constituirán los currículos y programas oficiales. Si en un texto aparece un significado sesgado, éste puede llegar a transmitirse a los alumnos, debiendo el profesor que los usa mantener una permanente vigilancia epistemológica sobre el contenido de los libros de texto (Ortiz 1999). Los pasos a seguir en el análisis de contenido son los siguientes:

- Determinar los textos de estadística universitaria a analizar;
- Seleccionar los capítulos, mediante una lectura detallada de los que tratan el tema, clasificando y agrupando las diferentes definiciones, propiedades, representaciones y justificaciones prototípicas;
- Determinar los elementos de significados, a partir del análisis epistémico, como guía para establecer el significado del teorema que se da en la institución universitaria;
- Elaborar tablas comparativas que recogen los elementos de significados en los distintos textos seleccionados;
- Análisis comparativo de contenido, entre lo que los autores de los libros seleccionados consideran más adecuado para este nivel educativo y el análisis del significado de referencia;
- Presentación de conclusiones, mediante el análisis descriptivo de la información obtenida.

A continuación se presentan los resultados y conclusiones de este análisis, proporcionando ejemplos que ayuden a clarificar las diferentes categorías encontradas.

## CAMPOS DE PROBLEMAS

En primer lugar se analizarán los campos diferenciados de problemas cuya resolución hace surgir la idea del teorema central del límite y su presencia en los libros de texto seleccionados. Se referirá indistintamente a la distribución de la suma o la media de variables aleatorias, pues son equivalentes salvo una constante.

CP1: *Obtener una aproximación de la distribución binomial para valores grandes de  $n$ .* En el cálculo de la distribución binomial  $bin(n,p)$  y otras distribuciones discretas intervienen los términos factoriales, que crecen muy rápidamente al aumentar el valor

de  $n$ . Mucho antes de la aparición de los ordenadores, este problema llevó a distintos matemáticos a tratar de encontrar valores aproximados de estas probabilidades para valores de  $n$  grandes. Este desarrollo es debido a Abraham De Moivre (1667-1754) y constituye el primer paso en el desarrollo histórico del teorema (Xiuyu 2003). Se ha encontrado este tipo de problemas en los libros, como en el siguiente ejemplo situado en el área de la ingeniería:

Las compañías eléctricas podan los árboles que crecen cerca de sus líneas para evitar cortes eléctricos debidos a la caída de árboles durante las tormentas. La aplicación de un producto químico para retrasar el crecimiento de los árboles es más barato que podar los árboles, pero estos productos matan algunos de los árboles. Suponga que un producto químico de este tipo matará el 20% de los arces. La compañía prueba este producto con una muestra aleatoria de 250 arces. ¿Cuál es la probabilidad de que mueran al menos 60 árboles (el 24% de la muestra)? (Moore 1995: 327)

CP2: *Determinar la distribución de la suma de variables aleatorias discretas independientes e idénticamente distribuidas.* La distribución binomial  $bin(n,p)$  puede considerarse como la suma de  $n$  variables aleatorias independientes e idénticamente distribuidas  $bin(1,p)$  o variables de Bernoulli. La demostración que la distribución binomial se aproximaba a la distribución normal llevó a tratar de generalizar este resultado a la suma de otras variables discretas. Se distinguen tres subcampos:

- CP2.1. *Variables con distribución uniforme discreta.* Este primer caso fue estudiado por Laplace, quien se interesó por la distribución de la suma de errores, hacia 1810. Afrontó el problema de la desviación entre la media aritmética de los datos (diferentes medidas con errores observacionales) y los valores teóricos, suponiendo que todos éstos están distribuidos aleatoriamente. El teorema proporciona la distribución de la media muestral, incluso si la población muestreada no es normal, como vemos en el siguiente ejemplo en que la población tiene una distribución uniforme.

Considérese una población que consta de cinco enteros igualmente probables: 1, 2, 3, 4 y 5. Comprobar la veracidad de los tres hechos específicos sobre la distribución muestral de medias muestrales: a) la media de la distribución muestral es igual a la media de la población, b) el error estándar de la media para la distribución muestral es igual a la desviación estándar de la población dividida entre la raíz cuadrada del tamaño de la muestra, c) la distribución de la media muestral será aproximadamente distribuida normalmente. (Johnson y Kuby 2004: 278).

- CP2.2. *Variables discretas acotadas.* Laplace generalizó en forma posterior, en el año 1809, el trabajo de De Moivre a la suma de variables aleatorias discretas idénticamente distribuidas con media y varianza finita. Este campo es más general que el anterior; un caso particular podría ser considerar la distribución binomial como población inicial (Wisniewski y Velasco 2001: 228). Se presenta el siguiente ejercicio:

La variable aleatoria  $X$ , que representa el número de cerezas en una empanada, tiene la siguiente distribución de probabilidad:

x	4	5	6	7
P (X = x)	0,2	0,4	0,3	0,1

Encuentre la probabilidad de que el número promedio de cerezas en 36 empanadas sea menor que 5,5. (Walpole, Myers y Myers 1999: 223).

- CP2.3. *Variables discretas no acotadas.* Debido a que la distribución límite obtenida por Laplace sólo depende de la media y la varianza, el resultado es válido también para una *distribución discreta con rango infinito*, en el caso de que existan los momentos finitos. El siguiente ejemplo se refiere a variables con distribución Poisson, que, aunque en la práctica está acotada, no lo está teóricamente:

Unos biólogos aseguran que el número de semillas por limón, en cierta variedad de limones sigue una distribución de Poisson con parámetro  $\mu = 5$ . Determinar la probabilidad de que el número promedio de semillas por limón sea menor a 5,5 en una muestra aleatoria de  $n = 125$  limones. (Wisniewski y Velasco 2001: 231).

CP3: *Establecer la distribución de la suma de variables aleatorias discretas no idénticamente distribuidas.* Poisson generalizó la demostración de Laplace a la *suma de variables aleatorias discretas perteneciente a una misma familia de distribuciones aunque con parámetros diferentes*. Un ejemplo es el siguiente, donde las distribuciones son todas uniformes, pero el valor de  $n$  varía para cada variable:

Sean  $X_1, X_2, \dots$  variables aleatorias independientes,  $X_n \sim U[-n, n]$  y  $S_n = \sum_{i=1}^n X_i$ . Mostrar que  $[S_n - E(S_n)]/\sigma_n$  converge a la distribución normal estándar, verificando las condiciones de Linderberg y Liapounov. (James 1981: 267).

CP4: *Determinar la distribución de la suma de variables aleatorias continuas.* Posteriormente Poisson da una demostración más rigurosa para una *variable continua* (de este modo sembró la semilla para el concepto de variables aleatorias). Se presenta un ejemplo en este campo que tiene gran frecuencia de uso e introduce el concepto de variable continua.

Se tienen  $n$  voltajes  $V_1, V_2, \dots, V_n$  que se reciben en una especie de concentrador (sumador), de tal suerte que  $V = \sum V_i$  es la suma de los voltajes recibidos en ese punto. Cada voltaje es  $V_i$  una variable aleatoria distribuida uniforme en el intervalo  $[0, 10]$ . Luego,  $E(V_i) = 5$  volts y  $Var(V_i) = 100/12$ . Si  $n = 20$ , podemos calcular la probabilidad de que el voltaje total de entrada sobrepase los 105 volts. (Meyer 1992: 263).

CP5: *Estimar el error de aproximación en el teorema central del límite.* Dirichlet fue el primero en intentar estimar el error de aproximación, aunque no tuvo éxito. Fue Cauchy, mediante la función característica, quien dio una demostración del teorema siguiendo la demostración de Poisson y estableció una cota superior para la diferencia entre el valor exacto y la aproximación. Este estudio del error de estimación tuvo como precedente la ley de los grandes números, que establece que la media observada  $\bar{X}$  de un número grande de observaciones tiene que estar cerca de la media  $\mu$  de la población. Se presenta el siguiente ejemplo:

El decano de una universidad desea estimar cuántos puntos se puede esperar que obtengan los aspirantes que se someten a un examen de ingreso. Si utiliza una muestra aleatoria de 100 aspirantes y supone que la desviación estándar es 20 puntos, ¿qué puede aseverar acerca de la probabilidad de que su error sea menor que tres puntos si emplea a) el teorema de Chebyshev; b) el teorema central del límite? (Freund y Smith 1989: 312).

CP6: *Encontrar condiciones necesarias y suficientes para el teorema.* La demostración de Cauchy, limitada sólo al caso de distribuciones de rango finito, no explicita las condiciones de validez ni estudia la razón de convergencia. Estos problemas fueron resueltos por Chebychev y sus discípulos Markov y Liapunov. Linderberg consideró que los momentos de tercer orden han de estar acotados para validar la convergencia, como condición suficiente del teorema. Feller y Levy probaron la condición necesaria del teorema. Un ejemplo es el siguiente:

Sea  $X_n$  una sucesión de variables aleatorias independientes, cada una distribuida uniformemente en el intervalo de 0 a  $\pi$ . Sea  $\{A_n\}$  una sucesión de constantes positivas. Establezca condiciones bajo las cuales la sucesión  $X_n = A_n \cos U_n$  obedezca el teorema central del límite (Parzen 1987: 477).

CP7: *Obtener la distribución de la suma de variables aleatorias dependientes.* El supuesto de dependencia es menos frecuente y presenta una dificultad de especificar modelos en más de dos variables. Se ilustra un ejemplo puramente matemático, referido a procesos autorregresivos de primer orden y cadenas de Markov en dos estados.

Proceso autorregresivo de primer orden estacionario. Se definen sucesivamente las variables  $X_{i+1} = \theta + \beta(X_i - \theta) + U_{i+1}$ ,  $|\beta| < 1$ ,  $i = 1, 2, \dots$  donde las variables independientes  $U_i \approx N(0, \sigma_0^2)$ . Vemos que  $X_{i+k} - \theta = \beta^k(X_i - \theta) + \beta^{k-1}U_{i+1} + \dots + \beta U_{i+k-1} + U_{i+k}$ . Además, la varianza de las variables satisface lo siguiente  $Var(X_{i+1}) = \beta^2 Var(X_i) + \sigma_0^2$ . Si la sucesión de las variables es estacionaria, debemos tener  $Var(X_{i+1}) = Var(X_i) \quad \forall i = 1, 2, \dots$  y por tanto  $\sigma^2 = \sigma_0^2 / (1 - \beta^2)$ , donde  $\sigma^2$  es la varianza común de las  $X_i$ . Suponga que la sucesión comienza con un  $X_1$  el cual es normal  $N(\theta, \sigma^2)$ . Entonces, esta es la distribución para todos los  $X_i$ , y por tanto la sucesión es estacionaria. (Lehmann 1999: 109).



Los campos de problemas mencionados están directamente relacionados con la evolución histórica del teorema central del límite (Alvarado y Batanero 2006). También se han encontrado otros campos de problemas indirectos, cuyas soluciones no llevan directamente al teorema, sino a uno de los problemas anteriores, es decir, a la búsqueda de una distribución que aproxime la suma de variables aleatorias. Son los siguientes:

CP8: *Obtener el tamaño adecuado de una muestra de poblaciones desconocidas*

CP9: *Obtener la distribución de la suma de los logaritmos de variables aleatorias independientes*

CP10: *Obtener la distribución de diferencias de medias muestrales en dos poblaciones*

CP11: *Determinar la distribución de funciones de variables aleatorias*

CP12: *Estimar por intervalos de confianza la media y otros parámetros para muestras grandes*

CP13: *Establecer pruebas de hipótesis de la media y otros parámetros para muestras grandes.*

En la tabla 1 se observa que los campos de problemas comunes en todos los libros son el de aproximación de la distribución binomial (CP1) y la distribución de la suma de variables aleatorias continuas (CP4). En cambio, los campos CP7 y CP3, carácter teórico, son inexistentes en la mayoría de los libros. Los ocho textos específicos de estadística aplicada a la ingeniería, ordenados por año, carecen de la mayoría de estos campos de problemas; sólo uno de ellos muestra aplicaciones basadas en la distribución de Poisson; modelo que es de importancia para los ingenieros.

Entre los campos de problemas indirectos (CP8 a CP13), los campos de intervalos de confianza (CP12) y pruebas de hipótesis (CP13) son los más frecuentes y están presentes hoy en día en la mayoría de los textos de estadística para ingenieros.

En general, los libros de textos muestran muchos ejercicios de aplicación de esta proposición para variables con distribución continua, en particular a la distribución exponencial y uniforme, y, por otro lado, carecen de problemas elementales, basados en extracción a partir de urnas, lanzamientos de dados o monedas, que permitan un primer acercamiento tangible del teorema. Además, falta rigurosidad en los ejemplos resueltos en algunos libros, pues no mencionan explícitamente que se está utilizando el teorema en el desarrollo del ejercicio, ni tampoco indican que el resultado es sólo aproximado. Estos serían puntos a mejorar en la presentación de los problemas. Asimismo, si se quiere que los alumnos aprecien la utilidad y alcance de este teorema, sería necesario mostrar diferentes aplicaciones de los campos de problemas a problemas específicos en su profesión.

Tabla 1

Campos de problemas que presentan los libros seleccionados

Situaciones problemas	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
CP1: Aproximación de la distribución binomial	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
CP2.1: Distribución suma v.a. discretas uniforme						x				x					x	x
CP2.2: Distribución suma v.a. discretas acotadas		x				x		x		x	x					x
CP2.3: Distribución suma v.a. discretas no acotadas								x	x	x	x	x		x		
CP3: Distribución suma v.a. no idénticamente distribuidas									x	x						
CP4: Distribución de la suma de variables aleatorias continuas		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
CP5: Estimación del error de aproximación	x	x	x	x		x			x	x		x	x			
CP6: Condiciones necesarias y suficientes									x		x	x				
CP7: Distribución de la suma de v.a. aleatorias dependientes									x							
CP8: Tamaño adecuado de una muestra aleatoria de poblaciones desconocidas	x		x					x	x	x	x	x				
CP9: Distribución de la suma de logaritmos	x							x								
CP10: Distribución de diferencias de medias muestrales			x	x	x	x		x				x				
CP11: Distribuciones de funciones de variables aleatorias			x		x				x		x	x				
CP12: Estimación por intervalos de confianza	x	x	x	x	x	x	x	x			x	x	x	x		x
CP13: Pruebas de hipótesis para muestras grandes	x	x	x	x	x	x	x	x				x	x			x

## LENGUAJE

El lenguaje matemático utilizado es el segundo elemento de significado, considerando las palabras, notaciones y todas las representaciones materiales del objeto abstracto y sus propiedades, así como los utilizados para describir los problemas y operar con sus datos (Godino 2002). Se diferencian tres tipos (Ortiz 1999):

*Términos y expresiones verbales.* Un primer tipo son las palabras y frases que se usan para describir los conceptos, sus operaciones y transformaciones. Se diferencian tres categorías de palabras usadas en la enseñanza de las matemáticas:

1. Palabras o expresiones matemáticas específicas que, normalmente, no forman parte del lenguaje cotidiano. Por ejemplo: Distribución límite, teorema, distribución asintótica en el muestreo, convergencia en ley y de sucesiones de variables aleatorias, error de estimación, variable transformada, función característica.
2. Palabras que aparecen en las matemáticas y en el lenguaje ordinario, aunque *no siempre con el mismo significado* en los dos contextos. Por ejemplo: Suficientemente grande, simulación manipulable, contraejemplo, muestra grande, generalización, bajo ciertas condiciones.
3. Palabras que tienen *significados iguales* o muy próximos en ambos contextos. Por ejemplo: Central, muestra, caso particular.

En estas categorías podríamos incluir un gran número de expresiones asociadas al teorema, debido a los diversos elementos relacionados con el tema. La categoría con mayor número de palabras y frases es la de palabras específicas ya que el teorema es un objeto puramente matemático, que ha sido analizado a lo largo de la historia con un nivel avanzado de conocimiento matemático y últimamente por simulación.

*Notaciones y símbolos.* Las notaciones simbólicas permiten realizar operaciones con los conceptos, trabajando a un alto nivel de complejidad. Se han encontrado las siguientes expresiones simbólicas del teorema tales como  $X_1, X_2, \dots, X_n$  para referirse a una muestra aleatoria,  $S_n = \sum_{i=1}^n X_i$  al enunciar la suma de  $n$  variables aleatorias,  $\bar{X} \approx N(\mu, \sigma^2/n)$  al describir la distribución muestral de la media de la muestra para  $n$  suficientemente grande, la fórmula de estandarización de la media muestral  $\frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ , etc.

*Representaciones gráficas.* Uno de los elementos característicos de la estadística son los gráficos. En la mayoría de los textos es común ver el histograma como primer acercamiento al teorema. Se han encontrado *histogramas* para representar la distribución empírica de las medias muestrales y *gráficos de control de medias* para representar que la distribución de las medias muestrales estarán más próximas a una distribución normal que las mediciones individuales. Al hablar de conceptos relacionados aparecen *gráficos de barras* para ilustrar la distribución de probabilidad teórica de la población, distribución de frecuencias relativas y distribuciones poblacionales.

*Simulaciones.* La simulación es una representación en el sentido que sustituye un experimento estocástico por otro y su empleo es especialmente útil en la enseñanza de conceptos en el campo de las distribuciones en el muestreo. Los textos de Mendenhall y Sincich (1997) y Johnson y Kuby (2004) proponen ejercicios para generar muestras aleatorias de observaciones de diversas distribuciones de probabilidad, usando los paquetes de software estadístico MINITAB, SAS o planilla electrónica EXCEL.

En la tabla 2 se aprecia que los términos específicos del teorema están presentes en los textos de estadística matemática [N° 9,10,11,12], además del texto para ingenieros Scheaffer y McClave (1993). Lo más común para referirse al teorema es a través de los símbolos, presentes en todos los textos a excepción de Johnson y Kuby (2004). Es débil la presencia de representaciones gráficas y la simulación; de los ocho textos de estadística para ingenieros sólo tres la presentan.

Tabla 2

Lenguaje y representaciones del teorema presente en los textos

Expresión	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Términos			x						x	x	x	x	x		x	
Símbolos	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Gráficos	x	x	x	x	x									x		x
Simulación			x		x			x			x			x		x

## PROCEDIMIENTOS

Las situaciones-problemas relacionadas con el teorema se pueden resolver mediante las siguientes técnicas:

AP1: *Cálculo algebraico y transformación de variables aleatorias.* El cálculo algebraico de la media muestral de variables aleatorias es el más típico y presente en todos los libros de textos analizados. Las situaciones planteadas no son explícitas en la aplicación del teorema, en cuanto a separar el procedimiento de cálculo, primero para la suma de variables aleatorias y luego para la media aritmética, lo que puede ser una dificultad de aprendizaje del tema en los alumnos.

AP2: *Tipificación/destipificación del cálculo de probabilidades para obtener una muestra aleatoria de tamaño adecuado.* El siguiente ejemplo donde se desarrolla la expresión  $P(|\bar{X} - \mu| \leq 0.3) = 0.95$  conduce a encontrar un tamaño de muestra adecuado.

Una máquina para llenar botellas tiene una varianza en las cantidades de llenado aproximadamente de  $\sigma^2 = 1$  onza. Sin embargo el promedio de las onzas de llenado  $\mu$  depende de un ajuste que puede cambiar de día a día. ¿Cuántas observaciones se deben efectuar en la muestra para que  $\bar{X}$  quede a menos de 0,3 onzas de  $\mu$  con una probabilidad de 0,95? (Scheaffer y Mc Clave 1993: 236).

AP3: *Cálculo de probabilidades referidas al teorema con calculadora, tablas de distribución o programa de ordenador.* Para muestras de tamaño grande es útil el cálculo de probabilidades mediante simulación con el ordenador, en la cual la obtención de muestras de tamaño  $n$  en una población se sustituye por obtención de números aleatorios y a partir de ellos se calcula el estadístico. De los textos analizados sólo dos libros plantean problemas con ordenador, a pesar de los enfoques actuales de la enseñanza de la estadística basados en los paquetes estadísticos, datos reales y situaciones de interés del educando.

AP4: *Cálculo de probabilidades referidas al teorema a partir de simulación con materiales manipulables.* Johnson y Kuby (2004: 284) presentan una población para la que puede elaborarse la distribución muestral teórica de todas las muestras posibles de tamaño 2 y a partir del procedimiento de resolución establece diferentes propiedades del teorema, incluyendo una aproximación empírica a la distribución normal.

Se observa en la tabla 3 que la estrategia más común para resolver problemas es el cálculo algebraico de la suma o media de variables aleatorias (AP1) presente en todos los textos analizados. Por el contrario, la técnica menos empleada es la operación inversa a la transformación estándar (AP2), que le exige al alumno habilidad en la operatoria algebraica y dominio del teorema.

Tabla 3

Algoritmos y procedimientos que presentan los libros seleccionados

Procedimientos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
AP1: Cálculo algebraico y transformación de variables aleatorias	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
AP2: Tipificación/destipificación	x		x						x							
AP3: Cálculo de probabilidades con calculadora, tablas estadísticas u ordenador				x		x			x						x	
AP4: Cálculo de probabilidades a partir de simulación					x			x	x		x	x				x

En los textos nuevos, la simulación, que no aparece en los textos de orientación matemática, se hace presente cada vez más como técnica específica de presentar el teorema, aunque sólo en tres textos para ingenieros se apoyan con el ordenador. Se destacan los textos para ingenieros, Scheaffer y Mc Clave (1993) y Devore (2001) que presentan tres de los cuatro algoritmos de aplicación del teorema, así también el texto moderno de Johnson y Kuby (2004).

### ENUNCIADOS DIFERENCIADOS DEL TEOREMA

Se encuentran diversos tipos de presentaciones para el teorema, según el grado de formalización, y que enfatizan diferentes aspectos del significado de los conceptos o se remiten a diferentes formas de aplicación:

E1: Enunciado *del teorema mediante la convergencia de sucesiones de variables aleatorias*, de manera formal y rigurosa con un nivel de matemática avanzado, como el dado por Cuadras (1999: 189):

Si es una sucesión de v.a. independientes de medias  $E(X_i) = \mu_i$  y varianzas  $Var(X_i) = \sigma_i^2$ , entonces, en ciertas condiciones generales, la variable suma, reducida:

$$(S_n - \sum \mu_i) / \sqrt{\sum \sigma_i^2} \xrightarrow{\text{CONVERGE EN LEY}} N(0,1).$$

E2: Enunciado *del teorema como límite ordinario de una sucesión de funciones*. En este caso la convergencia tiene un matiz determinista, mientras que en el anterior es aleatoria (en probabilidad). Se reproduce la formulación del teorema por Kalbfleisch (1984: 189):

Denotemos por  $f_n$  la función densidad de probabilidad de la suma  $S_n$ , o la altura del histograma de  $S_n$  en el caso discreto. El teorema central del límite afirma que, para todo número real  $z$ ,  $\lim_{n \rightarrow \infty} f_n(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2 / 2)$ .

E3: *Enunciado del teorema para la suma de variables independientes no idénticamente distribuidas*.

Sea  $X_1, X_2, \dots, X_n \dots$  una sucesión de variables aleatorias independientes con  $E(X_i) = \mu_i$  y  $Var(X_i) = \sigma_i^2$ ,  $i = 1, 2, \dots$ . Sea  $S_n = X_1 + X_2 + \dots + X_n$ . Luego bajo ciertas condiciones generales  $Z_n = (S_n - \sum_{i=1}^n \mu_i) / \sqrt{\sum_{i=1}^n \sigma_i^2}$  tiene aproximadamente la distribución  $N(0, 1)$ . (Meyer, 1973: 259).

E4: Enunciado *del teorema para la suma de variables independientes idénticamente distribuidas*. Esta presentación más restringida del teorema es la más común en los libros, pero en la mayoría se presenta para la media muestral. Se presenta un ejemplo para el caso de la suma:

Si se extrae una muestra aleatoria de  $n$  observaciones de una población con una media finita  $\mu$  y una varianza finita  $\sigma^2$ , entonces, si  $n$  es lo bastante grande, la distribución de muestreo de la suma  $S_n = \sum X_i$  se puede aproximar con una función de densidad normal cuya media es  $n\mu$  y varianza  $n\sigma^2$ . (Mendenhall y Sincich 1997: 315).

E5: *Enunciado del teorema de forma general*. Son varios los textos aplicados a la ingeniería que introducen el tema sin formulación matemática. Hoy en día es más conocido el teorema de manera general para el estimador de medias muestrales.

Obtén una muestra aleatoria simple de tamaño  $n$  de cualquier población de media  $\mu$  y desviación típica finita  $\sigma$ . Cuando  $n$  es grande, la distribución de la media muestral  $\bar{X}$  se aproxima mucho a la distribución normal  $N(\mu, \sigma/\sqrt{n})$  con media  $\mu$  y desviación típica  $\sigma/\sqrt{n}$ . Moore (1995: 304).

E6: *Enunciado intuitivo del teorema*. Otra forma de presentar el teorema es a través de la manipulación con objetos didácticos concretos de un experimento. Montgomery y Runger (1996: 302) introducen el teorema mostrando gráficamente que la aproximación normal para  $\bar{X}$  depende del tamaño  $n$  de la muestra, mediante la distribución de lanzamientos de varios dados legal de seis caras.

Se observa en la tabla 4 que los enunciados presentes en la mayoría de los textos es el correspondiente a la suma de variables aleatorias independientes idénticamente distribuidas (E4). Los menos encontrados en los libros son los presentados con rigor matemático (E1 y E2), seguido del enunciado del teorema para variables no idénticamente distribuidas. La tendencia de los textos actuales es hacia la presentación intuitiva del teorema, contrario a los libros más antiguos que introducen de manera formal el teorema. Por último, los enunciados E4 a E6 no son dados en forma precisa y clara, lo hacen por medio de un ejercicio planteado.

Tabla 4

Enunciados que presentan los libros seleccionados

Enunciados del teorema	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
E1: Mediante convergencia de sucesiones de v.a.									x	x						
E2: Como límite ordinario de una sucesión de funciones									x	x	x	x				x
E3: Para la suma de variables independientes no idénticamente distribuidas									x	x	x			x	x	
E4: Para la suma de variables independientes idénticamente distribuidas	x	x	x	x	x	x	x	x	x	x	x	x	x			x
E5: Enunciado de forma general	x				x				x	x	x		x			x
E6: Enunciado intuitivo del teorema			x	x	x			x				x	x			x

## PROPIEDADES

Tauber (2001: 139 a 144) encontró nueve propiedades de la distribución normal que clasificó en geométricas, estadísticas y algebraicas. Entre ellas, las siguientes ponen en correspondencia diferentes elementos de definición, lenguaje, representación y procedimiento del teorema central del límite:

P1: *La media de una suma de variables aleatorias es siempre la suma de las medias, sea aproximada o exacta la distribución de dicha suma;*

P2: *La varianza de la distribución de la suma de variables aleatorias independientes es la suma de las varianzas;*

P3: *La media aritmética de una muestra aleatoria de tamaño suficientemente grande sigue una distribución aproximadamente normal;*

P4: *La aproximación mejora con el número de sumandos;*

P5: *Las transformaciones lineales de variables aleatorias también siguen una distribución asintótica normal.*



Además se han encontrado las siguientes propiedades que tratan aplicaciones del teorema:

P6: *Las medias muestrales en dos poblaciones siguen una distribución aproximadamente normal;*

P7: *Aproximación de una distribución discreta por una continua;*

P8: *Aproximación de algunas distribuciones clásicas a la distribución normal;*

P9: *Los errores aleatorios siguen una distribución normal;*

P10: *Los estimadores de máxima verosimilitud tienen distribución asintótica normal;*

P11: *Los estimadores de los momentos tienen distribución asintótica normal;*

P12: *Corrección de continuidad.*

Tabla 5

Propiedades que presentan los libros seleccionados

Proposiciones	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
P1: Media de la distribución de la suma de variables aleatorias			x		x			x	x		x	x	x	x		
P2: Varianza de la distribución de la suma de variables aleatorias			x		x			x	x		x	x	x	x		
P3: Media aritmética sigue una distribución aproximadamente normal			x		x	x	x	x	x	x	x	x	x			
P4: La aproximación mejora con el número de sumandos			x		x			x						x		x
P5: Transformaciones lineales de variables aleatorias	x	x	x	x		x	x		x	x	x	x	x			x
P6: Medias muestrales en dos poblaciones siguen aproximación normal	x	x		x		x		x				x				x
P7: Aproximación de una distribución discreta por una continua		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P8: Aproximación de distribuciones clásicas a la distribución normal	x							x	x	x	x	x		x		
P9: Errores aleatorios con aprox. normal									x		x					
P10: Estimadores de máxima verosimilitud con distribución asintótica normal							x			x	x			x		
P11: Estimadores de momentos con distribución asintótica normal							x									
P12: Corrección de continuidad	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Las propiedades más frecuentes en los textos son las relacionadas con la corrección de continuidad (P12) presente en todos los textos pero con diferente intensidad y la convergencia a la normal (P5 y P3), siendo escasa la presencia de las propiedades referida a los errores (P9) y estimadores de máxima verosimilitud (P10), a pesar de que este método de estimación, según Peña (1995), es el más utilizado y produce estimadores con buenas propiedades estadísticas. Sólo la mitad de los textos consultados inicia el teorema con el cálculo algebraico de las propiedades de la suma y varianza de variables aleatorias (P1 y P2), lo que implica que en los otros textos el estudiante debe conocer esto anteriormente.

Los textos de estadística matemática son los que más propiedades tienen, siendo Meyer (1992) el mayor con nueve propiedades. Lo siguen los textos clásicos que también trabajan las propiedades algebraicas, a diferencia de los libros para ingenieros.

Las propiedades P6 y P7 no son definidas en forma precisa y clara; por ejemplo, Scheaffer y Mc Clave (1993) presentan la propiedad P6 sólo como un ejercicio planteado. Finalmente, de los 16 textos analizados, la mitad presenta el tema específico del teorema y a continuación como caso especial la aproximación de la binomial por la normal. Didácticamente sería preferible iniciar la enseñanza del teorema partiendo con el caso particular del estudio de la distribución binomial para muestras grandes, seguido del caso general para cualquier distribución, como lo enfoca Meyer (1992).

## ARGUMENTOS

Todos los enunciados, propiedades, problemas y algoritmos anteriores se ligan entre sí mediante argumentos o razonamientos que se usan para comprobar las soluciones de los problemas o demostrar las propiedades y relaciones que se pueden clasificar en la forma siguiente:

A1: *Demostraciones formales algebraicas y/o deductivas.* Un primer argumento para validar el teorema es la demostración matemática de forma lógico-deductiva. La más común es usando la función generadora de momentos que se puede ver en Mendenhall, Wackerly y Scheaffer (1994), Meyer (1973) y Canavos (1988) y usando la función característica en Parzen (1987).

A2: *Presentación del teorema como caso especial de un resultado general.* Meyer (1992) lo presenta mediante un ejemplo y Parzen (1987) lo argumenta mediante la función característica de la normalidad asintótica de variables aleatorias binomiales.

A3: *Simulaciones manipulables de distribuciones en el muestreo.* Aunque una demostración matemática establece la validez del teorema, puede que no contribuya mucho a la idea intuitiva del resultado. Algunos libros como Meyer (1992), incluyen otra forma de validar el teorema. A partir de una población sencilla se realiza una simulación con lápiz y papel de la elección de la muestra aleatoria de distintos tamaños de esta población.

A4: *Simulación gráfica con ordenador del teorema.* Una forma moderna de ilustrar y argumentar esta proposición consiste en realizar simulaciones con el ordenador aumentando

progresivamente el tamaño muestral. Johnson y Kuby (2004) muestran la aproximación de la distribución normal a la binomial para diferentes valores de los parámetros  $n$  y  $p$ .

*A5: Comprobación de ejemplos y contraejemplos, sin pretensión de generalizar.* Algunos textos utilizan distribuciones clásicas, la binomial o Poisson, para verificar la exactitud de la aproximación (Wisniewski y Velasco 2001; Peña 1995).

En la tabla 6, se observa que la forma de demostrar más reiterativa en los textos es mediante la comprobación de ejemplos, principalmente la distribución binomial. Parzen (1987) es el único que propone como contraejemplo la distribución de Cauchy.

*Tabla 6*

Argumentos que presentan los libros seleccionados

Argumentos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A1: Demostraciones formales algebraicas y/o deductivas									x	x	x	x	x			
A2: Caso especial de un resultado general			x						x	x	x					
A3: Simulaciones manipulables				x				x			x					x
A4: Simulación gráfica con ordenador			x		x			x						x		x
A5: Comprobar ejemplos y contraejemplos	x	x	x	x	x	x	x		x	x	x	x	x	x	X	

Son pocos los textos que utilizan la simulación con ordenador, o bien dispositivos de dados y monedas como un tipo de argumento que, aunque no proporciona una demostración, puede ayudar a la comprensión intuitiva del teorema. Las demostraciones algebraicas del teorema tienden a desaparecer en los textos más recientes. Los libros de estadística matemática son los más completos en las demostraciones aunque no trabajan la simulación como elemento validativo. En general, las demostraciones formales están ausentes en los textos para ingenieros; al parecer están más allá del propósito de los autores, dejándolo para cursos más avanzados.

## CONCLUSIONES

El análisis de contenido ha mostrado la complejidad del teorema y la riqueza de sus variados campos de problemas y representaciones. Se han identificado los elementos de significado más importantes del teorema, aunque muestra que no todos se utilizan de

manera significativa, debido a las orientaciones particulares de cada texto. Es preocupante que no todos los libros destinados a estudiantes de ingeniería consideren las representaciones y simulaciones, medios didácticos relevantes en el aprendizaje del teorema. En cuanto a los procedimientos de resolución de problemas, los textos son pobres al mostrar un algoritmo único y se inclinan por el cálculo algebraico. Respecto al enunciado del teorema, se ha pasado en los últimos años del rigor matemático actualmente a la presentación intuitiva. De las propiedades con significado asociado por la mayoría, la de mayor presencia es la aproximación de la media aritmética a la distribución normal. Percibimos una tendencia de la demostración deductiva a desaparecer en los textos, y obtuvimos el contraejemplo como elemento validativo más recurrente por los autores. Los ejemplos y situaciones no mencionan claramente que se está aplicando el teorema. Además no se diferencia la aplicación del teorema para la suma de variables aleatorias y la media aritmética, no se enfatizan problemas reales de la ingeniería con salida de software estadístico, ni tratan las propiedades matemáticas del teorema en tópicos tan importantes de la inferencia estadística como los métodos de estimación, errores de estimación y aplicaciones a las distribuciones de probabilidades clásicas.

La información proporcionada por los libros, aunque valiosa es limitada, ya que sólo proporciona una primera aproximación a la enseñanza del teorema. Sin embargo, se ha establecido un significado de referencia para diseñar una propuesta didáctica específica de enseñanza. Su adecuación se basará en el análisis de textos realizado y de las investigaciones previas, el uso de la tecnología como recurso didáctico y el marco teórico mencionado. El contenido de la lección será en orden creciente de complejidad, implementando diferentes formulaciones y representaciones de la suma de variables aleatorias discretas y continuas. Un primer acercamiento del alumno hacia la construcción del significado del teorema en el aula deberá considerar la simulación manipulable con lápiz y papel para muestras pequeñas. Luego continuar la introducción para muestras grandes de forma intuitiva por medio de la simulación gráfica con apoyo del ordenador y posteriormente analizar de forma algebraica con distintos campos de problemas que requieren de la proposición como herramienta de análisis de datos en ingeniería.

## BIBLIOGRAFIA

- Alvarado, H. (2004). *Significados del teorema central del límite y sus campos de problemas en los textos de estadística para ingenieros*. Memoria de Tercer Ciclo, Universidad de Granada, España.
- Alvarado, H. y C. Batanero (2006). El significado del teorema central del límite: Evolución histórica a partir de sus campos de problemas. En A. Contreras, L. Ordóñez y C. Batanero (eds.), *Investigación en Didáctica de las Matemáticas/ Congreso Internacional sobre Aplicaciones y Desarrollos de la Teoría de las Funciones Semióticas* (pp. 257-277). Jaén: Universidad de Jaén, España.
- Batanero, C. (2000). Controversies around significance tests, *Mathematical Thinking and Learning* 2 (1-2): 75-98.
- Cobo, B. y C. Batanero (2004). Significado de la media en los libros de texto de secundaria. *Enseñanza de las ciencias* 22 (1), 5-18.
- Cuadras, C. (1999). *Problemas de probabilidades y estadística*. Barcelona: EUB.

- delMas, R. C., J. B. Garfield y B. L. Chance (1999). A model of classroom research in action: developing simulation activities to improve students' statistical reasoning. *Journal of Statistic Education*, 7, 3. On line: <http://www.amstat.org/publications/jse>.
- Ghiglione, R. y B. Matalón (1991). *Les enquêtes sociologiques. Théories et pratique*. París: Armand Colin.
- Godino, J. D. (2002). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactique des Mathématiques* 22 (2 y 3).
- Godino, J. D. y C. Batanero (2003). Semiotic functions in teaching and learning mathematics. En M. Anderson, A. Sáenz-Ludlow, S. Zellweger y V.V. Cifarelli (eds.), *Educational perspectives on mathematics as semiosis: From thinking to interpreting to knowing* (pp. 149-168). New York: LEGAS.
- Godino, J. D., C. Batanero y V. Fonts (2007). The onto-semiotic approach to research in mathematics education. *Zentralblatt für Didaktik der Mathematik* 39 (1-2): 127-135.
- Godino, J. D., A. Contreras y V. Font (2006). Análisis de procesos de instrucción basado en el enfoque ontológico-semiótico de la cognición matemática. *Recherches en Didactiques des Mathématiques* 26 (1): 39-88.
- Kalbfleisch, J. (1984). *Probabilidad e inferencia estadística*. Madrid: AC.
- Lehmann, E. L. (1999). *Elements of large-sample theory*. New York: Springer.
- Méndez, H. (1991). *Understanding the central limit theorem*. Tesis Doctoral. Universidad de California. UMI 6369.
- Ortiz, J. J. (1999). *Significado de los conceptos probabilísticos elementales en los textos de Bachillerato*. Tesis Doctoral. Universidad de Granada.
- Ortiz, J. J., L. Serrano y C. Batanero (2002). El lenguaje probabilístico en los libros de texto. *SUMA* 36.
- Tauber, L. (2001). *Significado y comprensión de la distribución normal a partir de actividades de análisis de datos*. Tesis Doctoral. Universidad de Sevilla.
- Valleccillos, A. (1996). *Inferencia estadística y enseñanza: Un análisis didáctico del contraste de hipótesis estadísticas*. Madrid: Comares.
- Xiuyu, J. (2003). Historical development of central limit theorem. On line: <http://www.stat.rice.edu/~blairc/seminar/Files/julieTalk.pdf>.

## APENDICE

### *Libros de textos incluidos en el análisis:*

- Canavos, G. (1992). *Probabilidad y estadística*. México: McGraw-Hill.
- De Groot, M. (1988). *Probabilidad y Estadística*. Wilmington: Addison-Wesley.
- Devore, J. (2001). *Probabilidad y Estadística para Ingeniería y Ciencias*. Quinta edición. México: Thompson.
- Johnson, R. y P. Kuby (2004). *Estadística Elemental*. Tercera edición. México: Thompson.
- Kennedy, J. Y A. Neville (1982). *Estadística para Ciencias e Ingeniería*. Segunda edición. México: Harla.
- Mendenhall, W. y T. Sincich (1997). *Probabilidad y Estadística para Ingeniería y Ciencias*. Cuarta edición. México: Prentice Hall.
- Mendenhall, W., D. Wackerly y R. Scheaffer (1994). *Estadística Matemática con aplicaciones*. Segunda edición. México: Grupo Editorial Iberoamericana.
- Meyer, P. (1992). *Probabilidad y Aplicaciones Estadística*. Segunda edición. México: Addison-Wesley.
- Miller, I., J. Freund y R. Johnson (1992). *Probabilidad y Estadística para Ingenieros*. Cuarta edición. México: Prentice Hall.

- Montgomery, D. y G. Runger (1996). *Probabilidad y Estadística aplicadas a la Ingeniería*. México: McGraw-Hill.
- Parzen, E. (1987). *Teoría Moderna de Probabilidades y sus Aplicaciones*. Quinta edición. México: Limusa.
- Peña, D. (1995). *Estadística. Modelos y Métodos*. Octava edición. Madrid: Alianza Universitaria.
- Scheaffer, R. y J. McClave (1993). *Probabilidad y Estadística para Ingenieros*. Tercera edición. México: Grupo Editorial Iberoamericana.
- Velasco, G. y P. Wisniewski (2001). *Probabilidad y Estadística para Ingeniería y Ciencias*. México: Thompson.
- Walpole, R., R. Myers y S. Myers (1999). *Probabilidad y Estadística para Ingenieros*. Sexta edición. México: Prentice Hall, Pearson.
- Wisniewski, P. y G. Velasco (2001). *Problemario de Probabilidad*. México: Thompson.