

03061  
2ej.  
2

# Universidad Nacional Autónoma de México



INSTITUTO DE INVESTIGACIONES EN MATEMATICAS  
APLICADAS Y SISTEMAS.  
UNIDAD ACADÉMICA DE LOS CICLOS PROFESIONAL  
Y DE POSGRADO DEL C. C. H.

## TEOREMA CENTRAL DEL LIMITE: EVIDENCIA EMPÍRICA MEDIANTE SIMULACION

**T E S I S**

QUE PARA OBTENER EL GRADO DE:  
MAESTRO EN ESTADÍSTICA E  
INVESTIGACION DE OPERACIONES  
P R E S E N T A E L :  
ACT. VICTOR MANUEL ESPINOSA BALDERAS

**TESIS CON  
FALLA EN EL ORIGEN**

MEXICO, D. F.

1989



Universidad Nacional  
Autónoma de México



## **UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso**

### **DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

# INDICE

## Introducción.

I

### I. Teorema Central del Limite para variables aleatorias independientes con varianza finita.

1. Introducción.	1
2. Teorema de deMoivre-Laplace.	5
2.1 Teorema Local de deMoivre-Laplace.	5
2.2 Teorema Integral de deMoivre-Laplace.	12
3. Teorema Central del Limite para varianza finita.	21
4. Rapidez de convergencia.	23
4.1 Cota de Berry-Esseen.	23
4.2 Aplicación de la cota de Berry-Esseen.	28
5. Otra demostración del Teorema Central del Limite.	31
6. Extensiones.	33
7. Bibliografía comentada.	34

### II. Convergencia de sumas de variables aleatorias independientes: Leyes Estables e Infinitamente Divisibles.

1. Introducción.	37
2. Convergencia de sumas de variables aleatorias independientes idénticamente distribuidas: caracterización de Leyes Estables.	38
2.1 Aplicación y particularidades.	38
2.2 Ejemplos de leyes estables.	41
2.3 Dominio de Atracción Estándar y Condiciones de Atracción.	42
2.4 Consecuencias de las Condiciones de Atracción.	45
3. Convergencia de sumas de variables aleatorias independientes: Leyes Infinitamente Divisibles.	46
3.1 Definición y caracterización.	46
3.2 Ejemplos de Leyes Infinitamente Divisibles.	50
3.3 Convergencia a la distribución Poisson.	51
4. Convergencia a la Distribución Normal.	52
4.1 La distribución Normal como Ley Estable.	53
4.2 La distribución Normal como Ley Infinitamente Divisible.	55
5. Bibliografía Comentadas.	57

### III. Convergencia de sumas de variables aleatorias dependientes: Condiciones Mezclantes y Martingalas.

1. Introducción.	59
2. Condiciones Mezclantes.	63
2.1 Aplicaciones y definición.	63
2.2 Relación entre las condiciones	68

	mezclantes.	
2.3	Ejemplos de sucesiones mezclantes.	69
2.4	Comportamiento de condiciones mezclantes.	74
2.5	Teorema Central del Límite para sucesiones mezclantes.	75
2.6	Convergencia de sucesiones $\phi$ -mezclantes y $p$ -mezclantes.	76
2.7	Convergencia de sucesiones fuertemente mezclantes.	79
3.	Martingalas.	82
3.1	Origen y definición de martingalas.	82
3.2	Ejemplos de martingala.	85
3.3	Arreglo de martingalas.	86
3.4	Definiciones y propiedades.	88
3.5	Convergencia a mezcla de distribuciones normales.	92
3.6	Condiciones generales para el Teorema Central del Límite.	96
3.7	Refinamiento de las condiciones de convergencia.	98
3.8	Análisis de la estandarización aleatoria.	102
3.9	Aproximación de sucesiones mezclantes.	103
3.10	Algunas extensiones.	107
4.	Bibliografía comentada.	107

#### IV. Convergencia de sumas de variables aleatorias: Evidencia empírica mediante simulación.

1.	Introducción	109
2.	Conceptos básicos manejados por el paquete de simulación.	110
2.1	Distribución Normal Estándar.	110
2.2	Sumas de Variables Aleatorias.	111
2.3	Función de Distribución Empírica.	111
2.4	Convergencia en Distribución.	112
2.5	Cota de Berry-Esseen.	112
2.6	Prueba de Bondad de Ajuste de Anderson-Darling.	113
3.	Prácticas sugeridas para la explotación del paquete.	114
3.1	Práctica Número 1. Teorema de deMoivre-Laplace y Rapidez de Convergencia.	116
3.2	Práctica Número 2. Teorema Central del Límite para variables aleatorias independientes con Varianza Finita.	118
3.3	Práctica Número 3. Sumas de variables aleatorias independientes con Colas Pesadas. Leyes Estables.	120
3.4	Práctica Número 4. Sumas Aleatorias de Variables Aleatorias.	122
3.5	Práctica Número 5. Cadenas de Markov.	123
3.6	Práctica Número 6. Procesos Estacionarios: Medias Móviles.	124
3.7	Práctica Número 7. Dependencia en el	125

valor de la suma de los  $m(n)$  primeros términos.

4. Evidencia Empírica.	128
5. Variables aleatorias simuladas por el paquete de cómputo.	131
5.1 Distribución Bernoulli.	131
5.2 Distribución Uniforme Discreta.	132
5.3 Distribución Geométrica.	132
5.4 Distribución Binomial.	133
5.5 Distribución Poisson.	134
5.6 Distribución Exponencial.	134
5.7 Distribución Uniforme Continua.	135
5.8 Distribución Normal.	136
5.9 Distribución Ji-Cuadrada.	136
5.10 Distribución T de Student.	137
5.11 Distribución F.	137
5.12 Distribución Rayleigh.	138
5.13 Distribución Cauchy.	139
5.14 Cadena de Markov.	140

CONCLUSIONES 141

Apéndice. 145

Bibliografía. 151

## INTRODUCCION.

El importante papel que ha jugado el Teorema Central del Límite en el desarrollo de la Probabilidad y la Estadística está fuera de discusión. Por más de dos siglos y medio, generaciones de investigadores han dedicado su esfuerzo a profundizar en su estudio -baste mencionar que aproximadamente desde 1900 se han estado reuniendo semana a semana estadísticos soviéticos para discutir el tema-. Los alcances conseguidos se han vertido en una extensísima bibliografía, donde además de registrar los resultados logrados, se ha plaseado la forma en que se ha extendido su dominio de aplicación.

Como concepto, el Teorema Central del Límite representa más bien toda una temática que un sólo resultado, y para efectos de este trabajo se considera como la convergencia en distribución a la normal estándar de la suma estandarizada de variables aleatorias de una sucesión (o arreglo de sucesiones). También conviene establecer desde el principio una nota sobre el nombre, pues en nuestro idioma se ha generalizado el que se presenta en el título del trabajo, aunque el que le correspondería originalmente es el de Teorema del Límite Central, pues el adjetivo "central" califica al límite, no al teorema (en inglés: Central Limit Theorem). Hecha esta advertencia, se conviene que en todo el trabajo se mantendrá el nombre con el que mejor se conoce en español.

En el presente trabajo se incorpora directamente el uso de la computadora para introducir este concepto, aprovechando su capacidad de representación gráfica y de velocidad de cálculo. Es entonces factible sustituir fórmulas por gráficas que resultan de un proceso de simulación relativamente rápido, con la ventaja de explotar la intuición para comprender su significado, contando así con un recurso didáctico de importancia y una herramienta en la investigación y aplicación de resultados estadísticos. Es por ello que se ha desarrollado un paquete computacional en el que la representación central consiste en la gráfica de la función de distribución normal estándar superpuesta a la función de distribución empírica correspondiente a la suma estandarizada de las variables aleatorias simuladas. Se indican también los parámetros involucrados en la simulación y luego de evaluarse, el resultado de la prueba de Anderson-Darling de bondad de ajuste; lo cual da una idea global del concepto de convergencia y del papel que desempeñan los elementos estadísticos involucrados.

Para aprovechar al máximo este recurso se requieren dos

Características: diversidad de opciones y facilidad de manejo. La primera se pretende lograr permitiendo escoger las variables aleatorias involucradas, sus parámetros o el número considerado en la sucesión, las relaciones existentes entre ellas, e inclusive el número de simulaciones. El paquete de cómputo elaborado para manejar estas opciones está estructurado de forma que se elige la opción a simular por medio de menús sucesivos, pero no excesivos. Se han fijado alternativas en que varios factores están determinados previamente, y se incorporan además, pantallas informativas sobre la selección hecha.

Con estas particularidades, es posible simular las condiciones de una gran variedad de resultados relativos al Teorema Central del Límite, explotando su ventaja intuitiva sobre las demostraciones, que sin menosprecio de su imprescindible papel, resultan generalmente demasiado técnicas. Así, es posible realizar un recorrido histórico en que se va acumulando evidencia empírica sobre la validez de las proposiciones; desde el resultado pionero de deMoivre-Laplace que aplica a un particularísimo caso, pero que contiene la esencia del problema, para avanzar a sumas estandarizadas de cualquier tipo de variables aleatorias con la misma distribución, aunque con varianzas finitas, pues resulta ser requisito indispensable, como puede observarse con el paquete mediante contraejemplos. Más aún, es posible desechar la condición de idéntica distribución para encontrarse con que se puede ensayar en el paquete con las condiciones necesarias y suficientes en que se verifica la convergencia a la distribución normal, ya establecidas en el teorema de Lindeberg-Feller. Siguiendo el orden cronológico de las investigaciones, y del nivel de generalidad, se aborda la situación de dependencia dentro de la sucesión, en la que también se pueden experimentar casos en que ocurre la convergencia en distribución señalada.

Sin embargo, la aportación de este trabajo no se refiere únicamente al paquete de cómputo porque resultaría incompleto y sin orientación. Incluye material escrito con la intención de que guíe la utilización del programa computacional y proporcione además documentación sobre los teoremas que justifican los resultados que deben alcanzarse.

Por esa razón se presentan al final del trabajo sugerencias de prácticas que dosifican el avance de las simulaciones, invitan a reflexionar acerca del papel que juegan las condiciones involucradas, apuntan hacia las características relevantes en cada simulación y presentan las conclusiones a las que se supone debió conducir la experimentación. El tipo de resultados que se abordan en las prácticas va haciéndose más complejo, lo cual permite un avance gradual de todo el tema, o bien ubicarse en el estudio de algún teorema específico.

Respecto a la documentación, se han recopilado resultados que abarcan toda la variedad de situaciones descritas más arriba, muy bien tratados en distintos textos y artículos, pero que por la forzosa especialidad con que se orientan, se encuentran dispersos dificultando su consulta. Por evidentes razones de extensión, el acervo no pretende ser exhaustivo, pero se intenta proporcionar una selección de lo más importante y representativo sobre el tema; otra característica es que no se han incorporado, salvo contadas

excepciones que justificaremos más adelante, las demostraciones de los resultados establecidos. Ello obedece a que se pretende resaltar el enfoque intuitivo que impone el uso de esta herramienta empírica, prefiriéndose hacer énfasis en la interpretación de las condiciones impuestas en cada teorema y en ejemplificar con casos en que se aplican estos resultados. Otros elementos que enriquecen el material es una amplia referencia bibliográfica donde pueden consultarse las demostraciones -inclusive alguna es comentada en el trabajo-, y un apéndice que pretende hacer al material lo más autosuficiente posible.

El Capítulo 1 presenta el teorema de deMoivre-Laplace, donde se planteó por primera vez el problema de convergencia de sumas estandarizadas de variables aleatorias, y donde se mencionó también por primera vez la distribución normal. Se presenta en el mismo capítulo la extensión del Teorema Central del Límite a sucesiones de variables aleatorias independientes, idénticamente distribuidas y con varianzas finitas, para el que se presenta la demostración que aparece usualmente en la literatura, así como otra más reciente en que se utiliza el recurso de la regla de L'Hospital. De notable importancia es el resultado que igualmente se incluye sobre rapidez de convergencia debido a Berry y Esseen. Con esto están incorporados los ingredientes básicos, a saber, la convergencia en distribución que será interés en todo el trabajo, la posibilidad de generalizarla, y al que la velocidad con que se cumple puede variar de un caso a otro. Cabe señalar que es en este capítulo donde únicamente se incluyen todas las demostraciones, a fin de que el lector pueda comparar la conveniencia que sobre ellas tiene un argumento empírico, pues las primeras pretenden validar el resultado más que interpretarlo. Adicionalmente, se considera conveniente que cualquier lector debe conocer las demostraciones aquí incluidas por el conocimiento que brindan los métodos que éstas utilizan: al menos el teorema de deMoivre-Laplace, el Clásico y el de Cadenas de Markov de dos estados.

El contenido de este capítulo es el objeto de las primeras prácticas sugeridas a realizar con el paquete en las que se sugiere simular las condiciones del teorema de deMoivre-Laplace para confirmar su validez; también se recomienda observar el comportamiento de la distribución empírica cuando el parámetro de la distribución bernoulli tiende a los valores extremos cero o uno, en preparación a un resultado que compete al siguiente capítulo. Están diseñadas otras prácticas en que se varía la distribución a simular, tanto continuas como discretas, a fin de confirmar la amplitud del resultado y experimentar los efectos de la rapidez de convergencia.

El Capítulo 2 se ocupa de generalizar el estudio de la convergencia de sumas normadas a sucesiones de variables aleatorias independientes con la misma distribución, pero eliminando la restricción sobre la varianzas, lo que implica la inclusión de la familia de distribuciones estables, que han resultado adecuadas para modelar fenómenos en distintos campos tales como cotizaciones de acciones en la Bolsa de Valores, la distribución de las altas temperaturas que se registran en un reactor nuclear, en el cálculo de probabilidades en caminatas aleatorias y otros. De ellas se presenta una caracterización así como algunos ejemplos, y se muestra que son precisamente las únicas distribuciones límite en



esta situación. Se presenta además el concepto de dominio de atracción y las condiciones para pertenecer a él. Semejante papel y tratamiento se aplica a las distribuciones infinitamente divisibles cuando se descarta el supuesto de idéntica distribución, en cuyo contexto se han logrado establecer las condiciones necesarias y suficientes que posibilitan el cumplimiento del Teorema Central del Límite. Por otro lado, se incluye el interesante y útil teorema que muestra que bajo ciertas circunstancias, la distribución Poisson es también límite de sumas de variables aleatorias independientes.

Se concluye el capítulo al considerar a la distribución normal como un caso particular de ley estable e infinitamente divisible, anotándose teoremas en que se plantean distintas condiciones de suficiencia para alcanzar la convergencia a la distribución normal, y más aún, también condiciones necesarias y suficientes mediante el Teorema Central del Límite Clásico en el contexto de las leyes estables, y el conocido teorema de Lindeberg-Feller dentro del ámbito de las distribuciones infinitamente divisibles.

Las prácticas elaboradas para este capítulo tienden a presentar contraejemplos para la situación de idéntica distribución de tal manera que se resalte la necesidad de las condiciones. Por otra parte, se recomienda ensayar con casos en que las variables ya no compartan la misma distribución, para estudiar la validez del Teorema Central del Límite. Adicionalmente, se induce a considerar como una aproximación adecuada a la distribución Poisson para las sucesiones que representan los llamados eventos raros.

Una situación que puede resultar más real es la que corresponde al supuesto de dependencia entre las variables aleatorias de la sucesión y es el que se analiza en el Capítulo 3, que implica el estudio de las condiciones mezclantes para medir el grado de asociación existente dentro de la sucesión, y que aparte de definir las, relacionarlas y ejemplificarlas, se establecen resultados que garantizan para sucesiones que cumplen ese tipo de condiciones la convergencia a la distribución normal estándar. Un tratamiento más general y moderno lo representa el enfoque de martingalas que constituye la segunda parte del capítulo, y en el que la metodología seguida es relativamente la misma: definición, ilustración y aplicación al Teorema Central del Límite. Son estos tipos de dependencia, condiciones mezclantes y martingalas, las que más comúnmente se utilizan en desarrollos de Probabilidad y Estadística.

Como único ejemplo del método de demostración en este tipo de sucesiones, se desarrolla la correspondiente a una cadena de Markov de dos estados, en la que se aproxima la sucesión con una de comportamiento semejante, pero de elementos independientes.

En las prácticas correspondientes a este capítulo se aprovecha la capacidad del paquete para simular procesos incluidos como ejemplos en el texto para confirmar la validez de la convergencia a la distribución normal, ya sea en algún tipo de sucesión mezclante o porque se cumplen los requisitos para formar una martingala, combinados con la satisfacción de las condiciones impuestas en los teoremas respectivos. Asimismo, existe la sugerencia de simular sucesiones en que las circunstancias

impiden el logro del Teorema Central del Límite para evidenciar la imposibilidad de generalizarlo a cualquier situación.

Todas las prácticas a las que se ha hecho referencia han sido recopiladas en el Capítulo 4 para facilitar su localización y por tanto su aprovechamiento como recurso didáctico en el estudio del Teorema Central del Límite y en su aplicación. El formato empleado en las prácticas pretende ser accesible y de lenguaje sencillo, identificando el objetivo de cada práctica, planteando una serie de interrogantes que el usuario del paquete debe contestar conforme va observando la simulación, y la presentación sucinta de conclusiones al respecto. Sobra decir que debido a la versatilidad del paquete, el número de prácticas que se pueden implementar es mucho mayor que el de las presentadas, posibilitando elaborar otras con fines particulares. Contiene también las instrucciones necesarias para acceder al paquete, la definición de los elementos estadísticos que se hallan presentes y un resumen sobre la caracterización de las variables aleatorias que se simulan.

Como síntesis del trabajo en la que confluyen todos los factores: sustento teórico, paquete de cómputo y guía de uso, se realizaron repetidamente simulaciones bajo diversas condiciones con el objeto de reunir amplias y variadas pruebas experimentales sobre la veracidad de los enunciados expuestos a lo largo del trabajo. La presentación utiliza resúmenes, cuadros y comparaciones para facilitar su análisis y explotación.

Se considera que esta aportación puede facilitar el proceso de enseñanza-aprendizaje del tema, e inclusive es útil para reafirmar el significado de los conceptos involucrados. Está dirigido a un amplio público: estudiantes de licenciatura del Área de matemáticas que por primera vez abordan el estudio de la Probabilidad y que por tanto, el Teorema Central del Límite es tema obligado en sus versiones más típicas; para ellos se sugiere cubrir las prácticas y resultados correspondientes al primer capítulo y una introducción al contenido del segundo. Para el nivel de maestría la sugerencia no puede ser tan específica, pero puede servir de repaso el material del primer capítulo y estudiarse partes de los dos siguientes en distintos cursos, siempre agilizándolo con la realización de las prácticas. Para el estudioso en la materia, le apoya como documentación y amplia referencia bibliográfica, le auxilia en la aclaración de conceptos pues contiene definiciones y ejemplos y, por medio de las simulaciones le ayuda a estudiar algún teorema en particular, conocer si es válida la convergencia a la distribución normal en alguna aplicación y a reconocer el efecto que tienen las condiciones impuestas.

En fin, las referencias en las prácticas al texto, la intención de ser autosuficiente, la inclusión de datos históricos, el manejo del paquete por menús y sus cuadros de presentación así como el conjunto de simulaciones realizadas con ayuda del paquete y sus conclusiones, son algunas de las características del trabajo desarrollado que pretenden conferirle un carácter didáctico y de fácil consulta.

# CAPITULO I.

## TEOREMA CENTRAL DEL LIMITE PARA VARIABLES ALEATORIAS INDEPENDIENTES, IDENTICAMENTE DISTRIBUIDAS CON VARIANZA FINITA.

### 1. INTRODUCCION.

Con un pequeño bulto bajo el brazo, Abraham caminaba presuroso de regreso a casa, atravesó sin mucho cuidado la calle empedrada entre dos carruajes tirados por un par de caballos, de esos que estaban muy de moda en Londres; era evidente que sus pensamientos estaban en otro lado.

Al llegar, el ama de llaves, que ya lo esperaba, le ayudó silenciosamente a quitarse el pesado abrigo, la bufanda, el austero sombrero de copa y los guantes. Ahora ya podía ocuparse del legajo que recién había dejado en su escritorio.

Estaba consciente de la importancia del trabajo que había desarrollado en los papeles que tenía ante sí, aunque ese campo fuera prácticamente nuevo y estuviera muy restringido en su área de aplicación. Habían pasado 79 años desde que los respetuosos Monsieur Pascal y Monsieur Fermat se habían ocupado de contestar seriamente inquietudes surgidas a la clase aristócrata sobre resultados que se presentaban en juegos de azar, muy en voga en aquél entonces. Lo acertado de sus respuestas popularizó grandemente sus hallazgos.

Ahora corría el año 1733 y Abraham deMoivre, infundido por el espíritu del cálculo de probabilidades en los juegos de azar, pero no movido por una vana intención de ociosidad del juego, sino por el deseo de conocer las leyes más generales que rigen sus resultados, se dedicaba con ahínco a sus investigaciones, mismas que esperaba dar a conocer a varios de sus amigos ese mismo año; y en realidad darían fruto: significaban el comienzo en el trabajo que ocuparía a destacados matemáticos por espacio de más de dos siglos: el Teorema Central del Límite.

Las características del problema consistían en suponer, primero, que el juego se podría repetir indefinidamente para así estudiar su comportamiento a largo plazo, en el límite; segundo, que las reglas del juego no varían en el tiempo, manteniendo fija la probabilidad del jugador de ganar; y tercero, que no determinan de ninguna manera el resultado siguiente los anteriores, implicando que el jugador no adquiere experiencia a medida que se va repitiendo el juego. Este esquema se ajusta perfectamente, por ejemplo, al lanzamiento de una moneda en que el jugador gana si cae hacia arriba una cara de la moneda, y pierde si cae la otra.

Las interrogantes que se planteó eran las siguientes: si

supuestamente esa persona juega un gran número de partidas, ¿ cómo se podría calcular la probabilidad de que al menos un número determinado de ellas sea favorable? o ¿ cuál es la probabilidad de que gane a lo más un número específico?, en general, ¿qué probabilidad existe de ganar cierto número de partidas que esté entre unos límites previamente determinados?. Más aún, puesto que los cálculos directos resultan ser sumamente engorrosos e involucran cifras muy grandes -aún para las computadoras, que obviamente no existían en ese entonces-, ¿ existe alguna forma apropiada para encontrar dicha probabilidad con una precisión aceptable?

El planteamiento y solución a estas cuestiones los plasmaba finalmente A. deMoivre en una obra que tanto le había ocupado y que intituló *Doctrine of Chances* (exactamente en la segunda edición de 1738).

Pero hubieron de correr 74 años, hasta 1812, para que Pierre Simon de Laplace, aquél excelente y controvertido matemático expusiera con mayor abstracción el teorema de deMoivre, y aunque buscó generalizar más el enunciado, no lo logró plenamente pues su demostración carecía desafortunadamente de la formalidad requerida: era válida solamente para un conjunto restringido de funciones de distribución (el tipo de las llamadas latices). Sin embargo, en esa medio frustrada demostración, Laplace introdujo además un potente método conocido actualmente como función característica, que ha resultado de gran utilidad en el Cálculo de Probabilidades.

El problema de deMoivre y Laplace queda completamente establecido si consideramos la sucesión de variables aleatorias  $(X_n)$  con distribución  $b(p)$   $i=1,2,..$  (Bernoulli con parámetro  $p$ ) y deseamos encontrar la convergencia en distribución de la suma estandarizada de las variables aleatorias:

$$P \left\{ z_1 < \frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} < z_2 \right\}$$

Vale la pena señalar que si bien las restricciones sobre la sucesión de variables aleatorias son demasiadas inicialmente, cayendo aparentemente en un caso muy particular, los elementos primordiales de un concepto más amplio denominado Teorema Central del Límite ya están presentes y el enfoque de deMoivre apunta a la esencia del problema. Más aún, el citado caso particular tiene gran relevancia: es de vida o muerte.

Existe en efecto tal distribución y se trata de la muy conocida (¡originada por este resultado!) función de distribución normal o de Gauss, por lo que podemos afirmar que

$$P \left\{ z_1 < \frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} < z_2 \right\} \xrightarrow{d} \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-z^2/2} dz$$

Es conveniente comentar acerca de la expresión  $(\sum X_i - np) / \sqrt{np(1-p)}$ , que es una estandarización de la variable aleatoria "número de ocurrencias en n ensayos" -la cual sigue una distribución binomial con parámetros n y p-, que hace que la esperanza sea cero y la varianza igual a la unidad.

Para ejemplificar los conceptos expresados hasta aquí, supongamos que efectuamos n observaciones de una sucesión de variables aleatorias con distribución Bernoulli (que puede concebirse como un conjunto de personas con una probabilidad fija p de sobrevivir cada una cierto periodo de tiempo) y registramos el total de sobrevivientes al término del periodo -la suma de las  $X_i$  observadas-, haciéndolo no solo para un conjunto, sino para m grupos con el mismo número n de individuos. Entonces, construyamos la función de distribución empírica de

$$S_j = \frac{\sum_{i=1}^n X_{ji} - np}{\sqrt{np(1-p)}} \quad j = 1, 2, \dots, m$$

o sea

$$F_n(x) = \frac{1}{m} \sum_{j=1}^m I_{\{S_j \leq x\}}$$

donde  $I_{\{S_j \leq x\}}$  es la función indicadora, que vale 1 si  $S_j$  es menor o igual que el valor escogido x, y cero en caso contrario (se llama así pues 'indica' si algún elemento pertenece o no al conjunto expresado dentro de las llaves).

La función resultante será de tipo escalonado pues se incrementará un m-ésimo cada vez que se tenga un valor para el total de ocurrencias (estandarizadas) en alguno de los m grupos, y si el tamaño n en cada grupo es 'grande', los incrementos ocurrirán de tal manera que la función se aproxime al comportamiento de la distribución de Gauss. Esto es precisamente lo que se ha hecho por medio de simulaciones en computadora, fijando previamente n, p, m y generando aleatoriamente las series de observaciones lo cual muestra, de manera empírica, que la función de distribución  $F_n(x)$  se aproxima a la de la distribución normal estándar, como lo afirma el Teorema de deMoivre-Laplace.

Haciendo una consideración adicional, si bien podemos afirmar que al final de cuentas la serie converge a la distribución de Gauss, no podemos esperar hasta el infinito para ver que ello ocurre, por lo que requeriremos una medida que nos indique cómo se realiza esta aproximación y que sea capaz de determinar para cualquier número de elementos en la sucesión que tan 'cerca' se halla de la distribución límite. A esta medida se le conoce como 'rapidez de convergencia' y le dedicaremos nuestra atención más

adelante.

En nuestros días, al resultado conjunto de las investigaciones planteadas por estos dos protagonistas se le conoce con justicia como teorema de deMoivre-Laplace, e involucra formalmente varios conceptos ahora comunes en el estudio de la probabilidad, como son la sucesión de variables aleatorias, independencia a pares, compartición de la misma distribución, y convergencia en distribución a otra variable aleatoria. En este trabajo detallaremos cada uno de estos conceptos para posteriormente abrir un poco la discusión del tema al relajar la severidad de las condiciones y ampliar los resultados.

Con este preámbulo, se puede decir que aparte de su importancia, el Teorema Central del Límite se escribe con mayúsculas debido a que se refiere más bien a una propiedad que posee una familia de variables aleatorias que a un solo resultado: establece la convergencia en distribución de la suma estandarizada de las variables en la familia a la distribución normal estándar. Es aplicable a situaciones diversas que presentan las familias; en el sentido clásico la atención se centra en sumas de sucesiones de variables independientes e idénticamente distribuidas, y en este capítulo impondremos una restricción adicional, la de tener varianzas finitas para las cuales, adelantamos, siempre se cumple el Teorema Central del Límite.

La razón de dedicar un capítulo aparte a este caso específico reside no solo en cuestiones históricas, pues fueron las primeras en ser abordadas, sino en primer lugar, en que el papel que juegan en Probabilidad y Estadística variables aleatorias comprendidas en este caso es sumamente relevante, ya que éstas se encuentran en una amplia gama de aplicaciones (e incluso podríamos decir, casi la totalidad de distribuciones que tienen 'nombre' poseen varianzas finitas). En segundo lugar, las herramientas y conceptos empleados aquí serán utilizados de nueva cuenta más adelante, por lo que su introducción en un esquema más simple en este momento facilitará su comprensión así como el avance posterior.

Aunque lo dicho en el párrafo anterior pretende destacar la importancia de este capítulo, dista mucho de significar que no existen otros casos de interés en que la aproximación a la distribución normal se sigue cumpliendo, por lo que varianzas finitas y Teorema Central del Límite no son sinónimos; en el caso contrario, también es cierto que existen sucesiones con elementos independientes e idénticamente distribuidos cuyas sumas no convergen en distribución a la normal, contradiciendo lo que en ocasiones se supone un poco a la ligera, acerca de la infalible validez de la aproximación de la distribución de estas sumas estandarizadas por la distribución normal.

En la sección 2 trataremos el ya mencionado Teorema de DeMoivre-Laplace, primeramente en su forma local, como aproximación a probabilidades puntuales mediante la distribución normal, lo que servirá como plataforma para demostrar la forma integral que corresponde a probabilidades de intervalo.

Posteriormente, en la sección 3, presentaremos el resultado central del capítulo que asegura la propiedad del Teorema Central del Límite al caso en que existe varianza finita, con una demostración más general basada en la función característica. Se incluye también en la sección 4 un resultado debido a Berry y Esseen que permite conocer en algunos casos particulares la rapidez con que se converge a la distribución normal. Aparece finalmente en la sección 5 una variante de la demostración del Teorema Central del Límite para varianza finita, que recurre a otras herramientas como son la función generadora de momentos y la regla de L'Hospital.

No debe perderse de vista que los resultados aquí alcanzados están implementados en la computadora en el paquete anexo, con el objeto de ejemplificarlos y discutir algunas características y diferencias.

## 2. TEOREMA DE DeMOIVRE-LAPLACE.

### 2.1 Teorema Local de DeMoivre-Laplace.

Para calcular  $P_n(k)$ , la probabilidad de  $k$  ocurrencias del evento  $\xi$  en  $n$  ensayos independientes de Bernoulli donde  $\xi$  tiene probabilidad fija igual a  $p$ , desarrollaremos una fórmula asintótica debida a DeMoivre y Laplace.

Centrando el valor de  $k$

$$x = \frac{k - np}{\sqrt{npq}}$$

artificio con el cual se tiene una nueva variable cuya media es igual a cero y la varianza igual a uno.

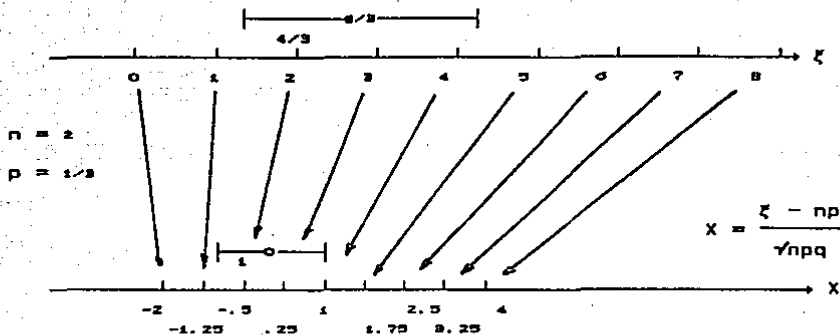


Figura 1.1  
Transformación de  $\zeta$  en  $X$ .

Queremos calcular

$$P_n(k) = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

con  $k = 0, 1, \dots, n$      $0 < p < 1$ ,     $q = 1-p$

que es igual a

$$\sqrt{\frac{n}{2\pi k(n-k)}} \frac{n^n p^k q^{n-k}}{k^k (n-k)^{(n-k)}} e^{(\theta_n - \theta_k - \theta_{n-k})} \quad (1)$$

en donde hemos utilizado la fórmula de Stirling

$$s! = \sqrt{2\pi s} s^s e^{-s} e^{\theta_s},$$

siendo  $\theta_s$  el residuo que satisface  $|\theta_s| \leq 1/12s$  (anecdóticamente, esta fórmula se debe también al mismo DeMoivre, excepto por la constante  $\sqrt{2\pi}$ ).



Cuadro 1.1  
Aproximación de Stirling para los 10 primeros valores de s.

s	s!	T=Stirling	s!/T	$\theta_s = \ln(s!/T)$	1/12s
1	1	.922	1.084	.0811	-.0833
2	2	1.919	1.042	.0413	-.0417
3	6	5.866	1.028	.0277	-.0278
4	24	25.506	1.021	.0208	-.0208
5	120	118.019	1.017	.0166	-.0167
6	720	710.078	1.014	.0139	-.0139
7	5 040	4 980.396	1.012	.0119	-.0119
8	40 320	39 902.395	1.010	.0104	-.0104
9	362 880	359 536.873	1.009	.0093	-.0093
10	3 628 800	3 598 695.619	1.008	.0083	-.0083

Considerando que a x la podemos mantener acotada por dos valores finitos a y b ( $a \leq x \leq b$ ) encontraremos por separado al limite de cada uno de los tres factores establecidos en (1), para lo cual utilizaremos las expresiones

$$k = np + x \sqrt{npq}$$

y

$$n-k = nq - x \sqrt{npq}$$

derivada la primera de la definición de x, y de restar ésta a n para hallar la segunda.

Comenzaremos por el ultimo factor. Por la acotación de x, se cumple que

$$k \geq np + a\sqrt{npq} \quad \text{y} \quad n - k \geq nq - b\sqrt{npq}$$

y puesto que

$$|\theta_n - \theta_k - \theta_{n-k}| \leq |\theta_n| + |\theta_k| + |\theta_{n-k}| \leq \frac{1}{12} \left( \frac{1}{n} + \frac{1}{k} + \frac{1}{n-k} \right)$$

tenemos

$$|\theta_n - \theta_k - \theta_{n-k}| \leq \frac{1}{12n} \left[ 1 + \frac{1}{p + a\sqrt{pq/n}} + \frac{1}{q - b\sqrt{pq/n}} \right]$$

que es ya una forma apropiada para mostrar que si n tiende a infinito, sin importar de que intervalo [a,b] se trate, esta expresion tiende a cero uniformemente en x, por lo que

$$\exp(\theta_n - \theta_k - \theta_{n-k}) \rightarrow 1 \quad \text{si } n \rightarrow \infty \quad (2)$$

Tratando ahora el primer factor de (1)

$$\begin{aligned} \sqrt{\frac{n}{2\pi k(n-k)}} &= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \\ &= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{(np+x\sqrt{npq})(nq-x\sqrt{npq})}} \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} \sqrt{\frac{1}{(1+x\sqrt{q/np})(1-x\sqrt{p/nq})}} \end{aligned}$$

y como el último radical tiende a uno si  $n$  tiende a infinito, podemos concluir que

$$\sqrt{\frac{n}{2\pi k(n-k)}} \sim \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} \quad (3)$$

para toda  $k$ , donde el símbolo  $\sim$  significa que ambas expresiones son asintóticas al crecer ilimitadamente  $n$ .

Cuadro 1.2  
Razón de la aproximación del primer factor en (1).

$n$	$k$	$\sqrt{n/(2\pi k(n-k))}$	$1/\sqrt{2\pi npq}$	Razón
50	23	.11320	.11968	.9458
100	42	.08083	.08428	.9551
500	187	.03687	.03785	.9743
1 000	363	.02624	.02676	.9803
2 000	708	.01865	.01892	.9857
5 000	1 733	.01186	.01197	.9906
10 000	3 427	.00841	.00846	.9932
15 000	5 115	.00687	.00691	.9944

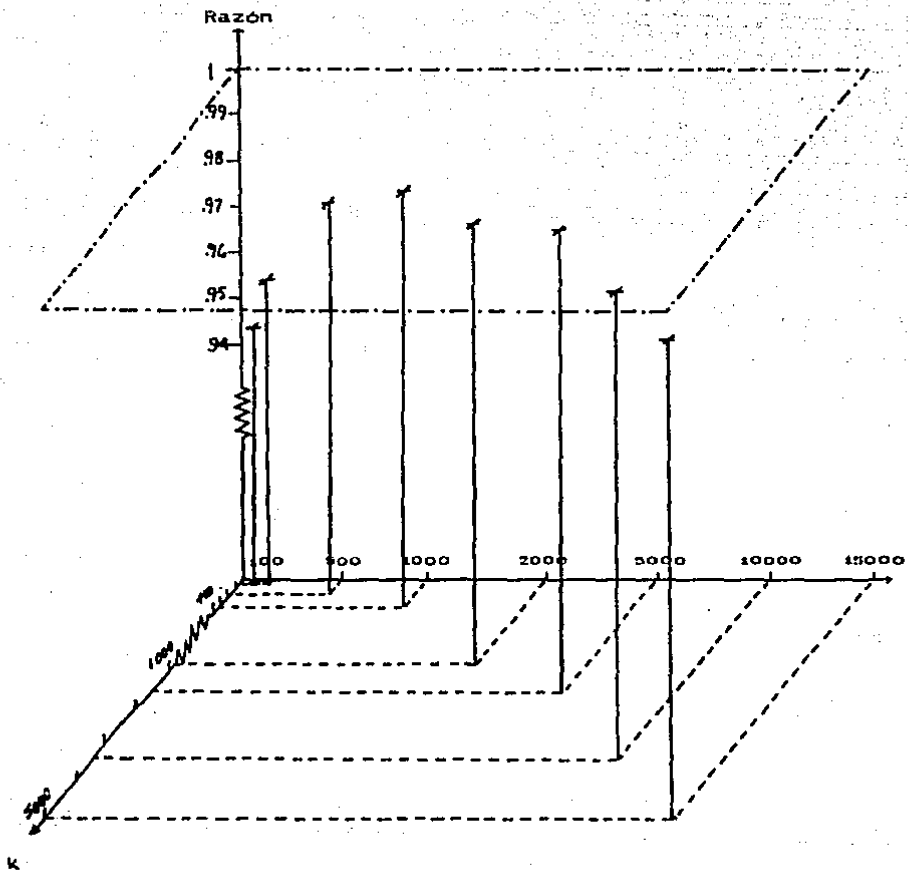


Figura 1.2  
Razón de la aproximación del primer factor en (1).

Por último, aprovechemos las propiedades de la función logaritmo:

$$\ln \left[ \frac{n^n p^k q^{n-k}}{k^k (n-k)^{(n-k)}} \right] = \ln \left[ \frac{np}{k} \right]^k + \ln \left[ \frac{nq}{n-k} \right]^{(n-k)}$$

$$= -k \ln \left[ \frac{k}{np} \right] - (n-k) \ln \left[ \frac{n-k}{nq} \right]$$

y usando nuevamente las expresiones alternativas para  $n$  y  $n-k$

$$\ln \left[ \frac{n^n p^k q^{n-k}}{k^k (n-k)^{(n-k)}} \right] = -(np + x\sqrt{npq}) \ln(1 + x\sqrt{q/np}) -$$

$$-(nq - x\sqrt{npq}) \ln(1 - x\sqrt{p/nq})$$

Puesto que  $x$  está acotada, podemos hacer  $x\sqrt{p/nq}$  y  $x\sqrt{q/np}$  tan pequeños que podemos suponerlos menores a uno y entonces expandir los logaritmos en series de potencia (explicitando únicamente los dos primeros términos):

$$\ln \left[ \frac{n^n p^k q^{n-k}}{k^k (n-k)^{(n-k)}} \right] = -(np + x\sqrt{npq}) [x\sqrt{q/np} - x^2 q/2np + O_1(1/n^{3/2})]$$

$$- (nq - x\sqrt{npq}) [-x\sqrt{p/nq} - x^2 p/2nq + O_2(1/n^{3/2})]$$

$$= -\sqrt{x^2 npq} -x^2 q + x^2 q/2 - \frac{x^3}{2} \sqrt{q^3/np} + O_3(1/n^{1/2})]$$

$$+ \sqrt{x^2 npq} -x^2 p + x^2 p/2 - \frac{x^3}{2} \sqrt{p^3/nq} + O_4(1/n^{1/2})]$$

$$= -\frac{x^2}{2} + O(1/n^{1/2})$$

(ver definición de  $O(n^k)$  en A1.14) por lo que también podemos concluir

$$\frac{n^n p^k q^{n-k}}{k^k (n-k)^{(n-k)}} \sim e^{-x^2/2} \quad (4)$$

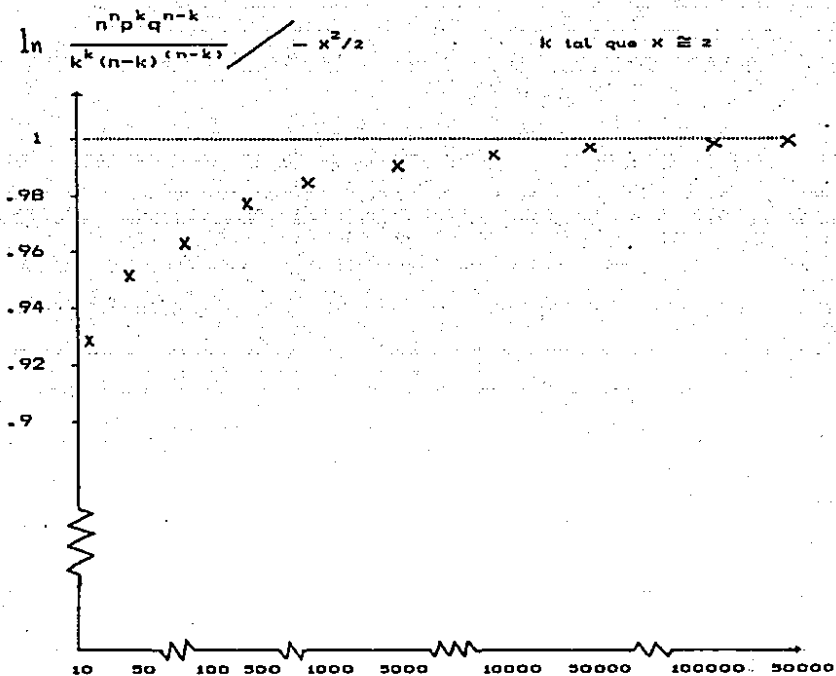


Figura 1.3  
Razón de aproximación del segundo factor de (1).

Ensamblando nuevamente las partes (2), (3) y (4), tenemos

$$P_n(k) \sim \frac{1}{\sqrt{2\pi} \sqrt{npq}} e^{-x^2/2}$$

que resuelve el problema planteado inicialmente: encontrar una estimación para la probabilidad

$$P_n(k) \cong \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \frac{p(x)}{\sqrt{npq}} \quad (5)$$

donde  $p(x)$  es precisamente la función de densidad de una normal con  $\mu = 0$ ,  $\sigma^2 = 1$ , y que se conoce tradicionalmente como el teorema local de DeMoivre-Laplace (Teorema 1.1).

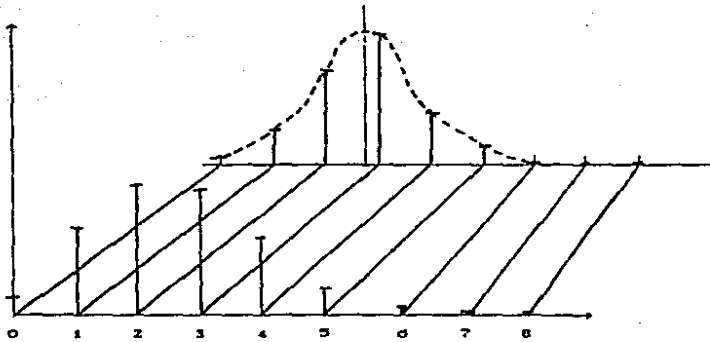


Figura 1.4 .  
Teorema Local de deMoivre-Laplace.

## 2.2 Teorema Integral de DeMoivre-Laplace.

Con este resultado buscaremos encontrar una expresión para la probabilidad del intervalo

$$P_n(a,b) = P \left\{ a \leq X = \frac{\xi - np}{\sqrt{npq}} < b \right\}$$

que por definición es igual a

$$\sum_{a \leq x_k < b} P_n(k) \quad \text{con } x_k = \frac{k - np}{\sqrt{npq}}$$

Mostraremos que una expresión adecuada para aproximar esta probabilidad está dada en términos de la distribución normal, por lo que un primer paso será aproximarla a una función con dominio continuo, a saber

$$H_n(x) = \begin{cases} \sqrt{npq} P_n(k) & x_k \leq x < x_{k+1} \quad k = 0, 1, \dots, n \\ 0 & \text{otro caso} \end{cases}$$

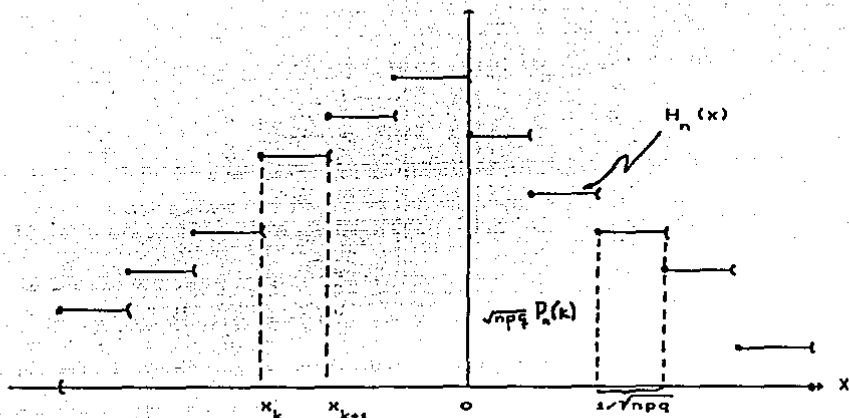


Figura 1.5  
Función escalonada  $H_n(x)$ .

hallando que una aproximación para  $P_n(a,b)$  está dada por  $\int_a^b H_n(x) dx$  pues aunque lo que realmente se cumple es

$$P_n(a,b) = \int_{x_l}^{x_u} H_n(x) dx$$

donde  $x_l$  y  $x_u$  son los valores transformados inmediatamente superiores a  $a$  y  $b$  respectivamente ( $a \leq x_l < a + 1/\sqrt{npq}$ ,  $b \leq x_u < b + 1/\sqrt{npq}$ ). La diferencia debida a cambiar los límites en la integral es despreciable, como veremos en seguida.

Localizando el punto donde se hace máximo  $P_n(k)$ ,  $m_0 = [(n+1)p]$  (donde los corchetes representan la función máximo entero contenido), el máximo de  $H_n(x)$  se halla a su vez en el intervalo

$$a_0 \leq \frac{-1}{\sqrt{npq}} \leq \frac{m_0 - np}{\sqrt{npq}} \leq x < \frac{m_0 + 1 - np}{\sqrt{npq}} \leq \frac{2}{\sqrt{npq}} \leq b_0$$

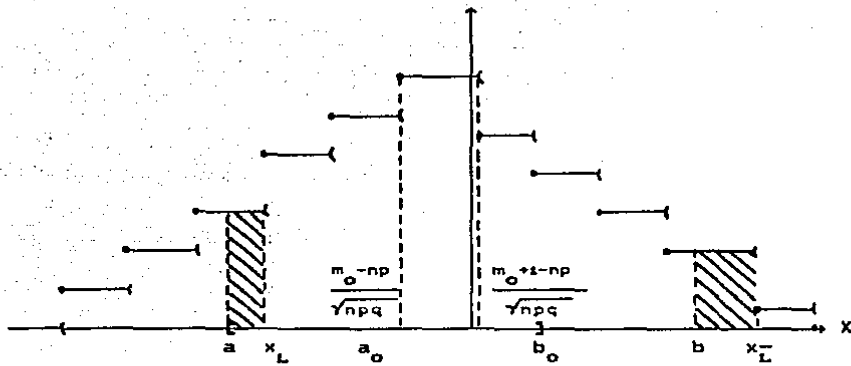


Figura 1.6  
Máximo de  $H_n(x)$ .

con  $a_0, b_0$  constantes, por lo que se trata de un intervalo finito y por lo tanto podemos aplicar (5) para encontrar una cota para  $H_n(x)$ :

$$\max_{-\infty < x < \infty} H_n(x) = \max_{a_0 \leq x \leq b_0} H_n(x) \cong \sqrt{npq} \max_{a_0 \leq x \leq b_0} \frac{e^{-x^2/2}}{\sqrt{2\pi} \sqrt{npq}} \leq \frac{1}{\sqrt{2\pi}}$$

pues el cero pertenece a  $[a_0, b_0]$ . Entonces

$$P_n(a, b) = \int_{x_L}^{x_u} H_n(x) dx = \int_a^b H_n(x) dx + \int_b^{x_u} H_n(x) dx - \int_a^{x_L} H_n(x) dx$$

pero

$$\begin{aligned} \left| \int_b^{x_u} H_n(x) dx - \int_a^{x_L} H_n(x) dx \right| &\leq \int_b^{x_u} \max H_n(x) dx + \int_a^{x_L} \max H_n(x) dx < \\ &< \frac{1}{\sqrt{2\pi}} [ (x_u - b) + (x_L - a) ] \leq \\ &\leq \frac{1}{\sqrt{2\pi}} \frac{2}{\sqrt{npq}} \rightarrow 0 \quad \text{si } n \rightarrow \infty \end{aligned}$$



por lo que concluimos que en efecto

$$P_n(a,b) \xrightarrow{b} \int_a^b H_n(x) dx \quad \text{si } n \rightarrow \infty$$

Ahora probaremos que  $H_n(x)$  puede aproximarse por la distribución normal, considerando primero que  $a$  y  $b$  son finitos, en cuyo caso podemos aplicar nuevamente (5), haciendo explícito un residuo

$$H_n(x_k) = \frac{e^{-x_k^2/2}}{\sqrt{2\pi}} [1 + R_n(x_k)]$$

donde  $R_n(x_k) \rightarrow 0$  si  $n \rightarrow \infty$ .

Para extender al intervalo  $x_k \leq x < x_{k+1}$

$$\begin{aligned} H_n(x) = H_n(x_k) &= \frac{e^{[-x_k^2/2 + x^2/2 - x^2/2]}}{\sqrt{2\pi}} [1 + R_n(x_k)] \\ &= \frac{e^{-x^2/2}}{\sqrt{2\pi}} [1 - 1 + e^{(x^2 - x_k^2)/2} [1 + R_n(x_k)]] \\ &= \frac{e^{-x^2/2}}{\sqrt{2\pi}} [1 + R_n(x)] \end{aligned}$$

donde  $R_n(x) = -1 + e^{(x^2 - x_k^2)/2} [1 + R_n(x_k)]$  tiende a cero pues  $R_n(x_k) \rightarrow 0$  y porque

$$\left| \frac{x^2 - x_k^2}{2} \right| = \frac{|x - x_k| |x + x_k|}{2} \leq \frac{|x|}{\sqrt{npq}} \leq \frac{\max(|a|, |b|)}{\sqrt{npq}} \rightarrow 0$$

cuando  $n \rightarrow \infty$ .

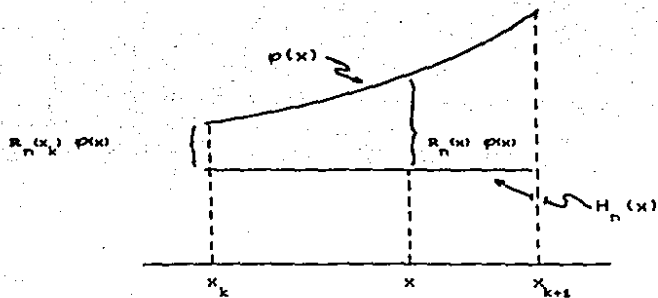


Figura 1.7  
Aproximación de  $H_n(x)$  por la distribución normal.

Entonces

$$\int_a^b H_n(x) dx = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx + \frac{1}{\sqrt{2\pi}} \int_a^b R_n(x) e^{-x^2/2} dx$$

y

$$\frac{1}{\sqrt{2\pi}} \int_a^b R_n(x) e^{-x^2/2} dx \leq \max_{a \leq x \leq b} \frac{R_n(x)}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx \rightarrow 0$$

Por lo tanto

$$P_n(a, b) \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx \quad \text{si } a, b \in \mathbb{R}$$

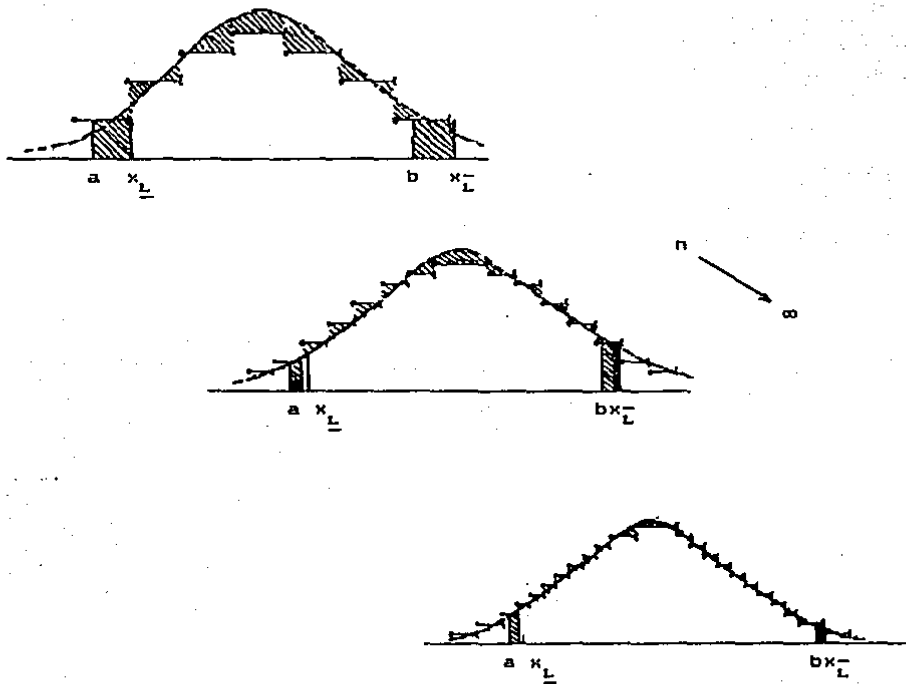


Figura 1.B  
Refinamiento de la aproximación cuando  $n$  tiende a infinito.

Para el caso en que  $a$  o  $b$  sean infinitos, hagamos las siguientes consideraciones.

Puesto que  $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1$ , podemos elegir para toda  $\epsilon > 0$  una constante  $L$  suficientemente grande que cumpla

$$\frac{1}{\sqrt{2\pi}} \int_{-L}^L e^{-x^2/2} dx > 1 - \epsilon/4 \quad (6)$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-L} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_L^{\infty} e^{-x^2/2} dx < \epsilon/4 \quad (7)$$

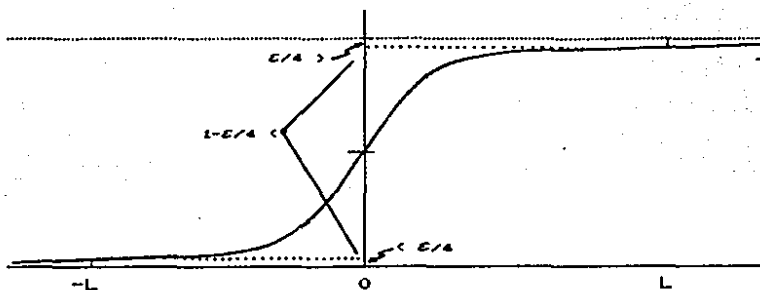


Figura 1.9  
Acotamiento de los extremos de la distribución normal.

Y aprovechando el caso ya probado de  $a, b$  finitos, se cumple para toda  $a_0, b_0$  en  $-L \leq a_0 \leq b_0 \leq L$  que

$$\left| P_n(a_0, b_0) - \frac{1}{\sqrt{2\pi}} \int_{a_0}^{b_0} e^{-x^2/2} dx \right| < \epsilon/4 \quad (8)$$

en particular

$$\left| P_n(-L, L) - \frac{1}{\sqrt{2\pi}} \int_{-L}^L e^{-x^2/2} dx \right| < \epsilon/4$$

de donde establecemos

$$\frac{1}{\sqrt{2\pi}} \int_{-L}^L e^{-x^2/2} dx < \epsilon/4 + P_n(-L, L)$$

y por (6)

$$P_n(-L, L) > 1 - \epsilon/2$$

cumpliéndose además que

$$P_n(-\infty, L) + P_n(L, \infty) = 1 - P_n(-L, L) < \epsilon/2 \quad (9)$$

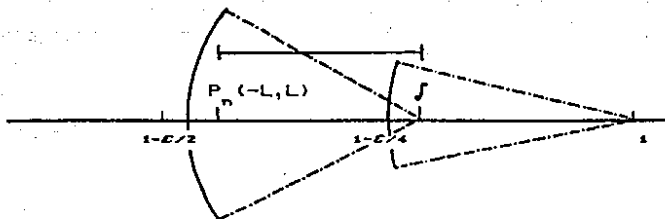


Figura 1.10  
Acotación del valor de  $P_n(-L, L)$ .

Estas acotaciones pueden ser utilizadas para terminar la demostración, cualquiera que sean los valores que tomen  $a$  y  $b$  respecto al intervalo ya fijado por  $\epsilon$ ,  $(-L, L)$ . Como ilustración, y puesto que los demás casos son equivalentes, consideremos solamente  $-L \leq a < L$  y  $b = \infty$ . Así

$$\begin{aligned} & \left| P_n(a, b) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx \right| \\ &= \left| P_n(a, L) - \frac{1}{\sqrt{2\pi}} \int_a^L e^{-x^2/2} dx + P_n(L, \infty) - \frac{1}{\sqrt{2\pi}} \int_L^\infty e^{-x^2/2} dx \right| \\ &\leq \left| P_n(a, L) - \frac{1}{\sqrt{2\pi}} \int_a^L e^{-x^2/2} dx \right| + P_n(L, \infty) + \frac{1}{\sqrt{2\pi}} \int_L^\infty e^{-x^2/2} dx \end{aligned}$$

$$\leq \frac{\epsilon}{4} + \frac{\epsilon}{2} + \frac{\epsilon}{4} = \epsilon$$

por (8), (9) y (7), respectivamente.

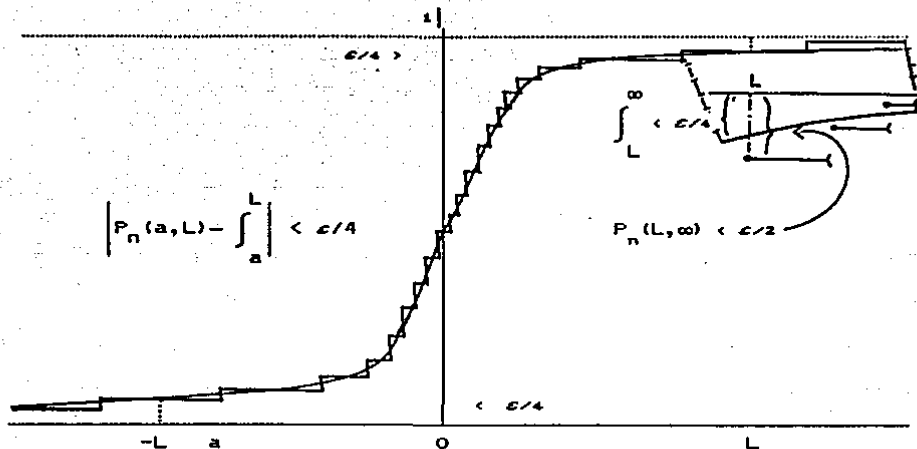


Figura 1.11  
Aproximación de  $P_n(a, b)$  cuando  $b$  es infinito.

El desarrollo anterior constituye la demostración del siguiente enunciado:

**Teorema 1.2.** (DeMoivre, 1738, LaPlace, 1812) Sean  $\zeta_1, \dots, \zeta_n$  una sucesión de variables aleatorias que se distribuyen independientemente e idénticamente con una distribución Bernoulli con parámetro  $p$ , y defínase

$$X = \frac{\sum \zeta_i - np}{\sqrt{npq}}$$

y

$$P_n(a, b) = P(a \leq X < b)$$

entonces

$$\lim_{n \rightarrow \infty} P_n(a, b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

Antes de dejar este resultado, puede observarse que en la demostración se dedicó gran esfuerzo a manejar los residuales de la aproximación, y a justificar el uso de una función continua como es la de la distribución normal para evaluar una discreta. Esto no ayuda demasiado a comprender la naturaleza del resultado, por lo que se sugiere utilizar como complemento argumentos empíricos obtenidos por medio de simulación.

### 3. TEOREMA CENTRAL DEL LIMITE PARA VARIANZA FINITA.

Hemos probado que una sucesión de variables aleatorias independientes que se distribuyen idénticamente como una Bernoulli con parámetro  $p$  cumplen la condición del Teorema Central del Limite, a saber: "Las sumas parciales de las variables aleatorias en la sucesión, ajustadas mediante una transformación lineal, convergen en distribución a la normal con media cero y varianza la unidad". Sin embargo, se trata solamente de una distribución en particular por lo que resulta conveniente ampliar el resultado, lo que haremos en seguida para todas aquellas sucesiones de variables independientes e idénticamente distribuidas que posean varianza finita.

Para ello necesitaremos algunas herramientas y remitimos al lector al apéndice para mayores referencias, en particular sobre la función característica.

Así, procedamos a enunciar y demostrar el Teorema Central del Limite para variables aleatorias independientes idénticamente distribuidas con varianza finita, propuesto por Lindeberg en 1922.

**Teorema 1.3** Sea  $X_1, X_2, \dots$  una sucesión de variables aleatorias idénticamente distribuidas con  $E(X_i) = \mu$ ,  $E(X_i^2) < \infty$ . Sean  $\text{Var}(X_i) = \sigma^2$  y  $S_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Entonces

$$P \left[ \frac{S_n - \mu}{\sigma/\sqrt{n}} \leq x \right] \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy \quad \text{si } n \rightarrow \infty \quad \triangleright$$

**Demostración.** Utilizando la función característica  $\phi_x(t) = E(e^{itx})$  y sus propiedades (A1.1, A1.3, A1.4), obtenemos

$$\begin{aligned} \frac{\phi_{S_n - \mu}}{\sigma/\sqrt{n}}(t) &= \phi_{S_n - \mu} \left( \frac{\sqrt{nt}}{\sigma} \right) \\ &= \phi_{\sum_{i=1}^n X_i - n\mu} (t/\sqrt{no}) \\ &= \prod_{i=1}^n \phi_{X_i - \mu} (t/\sqrt{no}) \\ &= \left[ \phi_{X_1 - \mu} (t/\sqrt{no}) \right]^n \end{aligned}$$

y puesto que  $E(X_i - \mu) = 0$ ,  $\text{Var}(X_i - \mu) = \sigma^2$ , podemos aplicar la expansión de la función característica en términos de los momentos

no centrales (A1.6),

$$\phi_{x-\mu}(t) = 1 - \sigma^2 t^2 / 2 + o(t^2) \quad (10)$$

con  $\lim_{t \rightarrow 0} \frac{o(t^2)}{t^2} = 0$ , como se especifica en general en A1.14,

para tener la expresión

$$\phi_{\frac{s_{n-\mu}}{\sigma\sqrt{n}}}(t) = \left[ 1 - \frac{t^2}{2n} + o(t^2/\sqrt{no}) \right]^n$$

o

$$\ln \phi_{\frac{s_{n-\mu}}{\sigma\sqrt{n}}}(t) = n \ln \left[ 1 - \frac{t^2}{2n} + o(t^2/\sqrt{no}) \right]$$

y tomando la expansión de  $\ln(1+x)$  en serie de MacLaurin, considerando solamente el primer término y acotando el resto (ver A1.15 a A1.17)

$$\ln(1+x) = x + \rho(x) \quad -1 < x \leq 1 \quad (11)$$

donde  $|\rho(x)| \leq C|x|^2$

escribimos

$$\ln \phi_{\frac{s_{n-\mu}}{\sigma\sqrt{n}}}(t) = n g(t,n) + n \rho(g(t,n))$$

con  $g(t,n) = -\frac{t^2}{2n} + o(t^2/\sqrt{no})$

entonces

$$\left| -\frac{t^2}{2n} - \ln \phi_{\frac{s_{n-\mu}}{\sigma\sqrt{n}}}(t) \right| \leq n \left| o(t^2/\sqrt{no}) \right| + n \left| \rho(g(t,n)) \right| \quad (12)$$

pero por (10)

$$\lim_{n \rightarrow \infty} \frac{no^2}{t^2} o(t^2/\sqrt{no}) = 0 \quad \forall t \text{ fija}$$

que implica que  $n |o(t^2/\sqrt{no})| \rightarrow 0$  si  $n \rightarrow \infty$ . Por argumentos similares, junto con (11),



$$\lim_{n \rightarrow \infty} n | \rho(g(t, n)) | = 0$$

$\forall t$  fija

por lo tanto, tenemos debido a (12)

$$\lim_{n \rightarrow \infty} \ln \frac{\phi_{S_n - \mu}}{\sigma \sqrt{n}}(t) = -\frac{t^2}{2}$$

$$\lim_{n \rightarrow \infty} \frac{\phi_{S_n - \mu}}{\sigma \sqrt{n}}(t) = e^{-t^2/2}$$

Por el teorema de continuidad (A1.9), el cual establece que dada una sucesión de funciones de distribución  $F_n$  siendo  $\phi_n$  sus correspondientes funciones características, y dada también  $F$  una función de distribución con  $\phi$  su función característica, entonces  $F_n \xrightarrow{d} F \Leftrightarrow \phi_n(t) \rightarrow \phi(t)$ , por lo que la función característica de  $(S_n - \mu)/\sigma\sqrt{n}$  es en el límite  $e^{-t^2/2}$ , la cual por A1.4 —que establece la relación biunívoca entre las funciones características y de distribución—, la reconocemos como la función característica de la distribución normal estandarizada (media cero y varianza la unidad) (A1.2), con lo que concluimos la demostración. ■

#### 4. RAPIDEZ DE CONVERGENCIA.

##### 4.1 Cota de Berry-Esseen.

Por el resultado anterior estamos justificados a utilizar la distribución normal como una aproximación para el cálculo de probabilidades de la suma estandarizada de variables idénticamente distribuidas con varianza finita; sin embargo, esta aproximación, ¿qué tan cerca se encuentra del valor correcto? Una manera de cuantificar esa "cercanía" consiste en encontrar una cota para el supremo de las distancias entre la distribución exacta y la aproximada, donde la distancia se mide como la diferencia en valor absoluto entre ambas funciones.

En este sentido contamos con una cota que depende de los primeros momentos de la distribución, en el caso de que  $E(|X_k|^3)$  exista y es el siguiente resultado debido a Berry (1941) y a Esseen (1945).

**Teorema 1.4.** Sea  $X_1, X_2, X_3, \dots$  una sucesión de variables aleatorias independientes e idénticamente distribuidas con  $E(X_k) = 0$ ,  $\text{Var}(X_k) = \sigma^2$  y  $E(|X_k|^3) < \infty$ . Sea  $F_n(x)$  la distribución de

$S_n/\sigma\sqrt{n}$ ,  $S_n = \sum_{k=1}^n X_k$  y  $\Phi(x)$  la función de distribución normal estándar. Entonces

$$\sup_x |F_n(x) - \Phi(x)| \leq \frac{C E(|X_k|^3)}{\sigma^3 \sqrt{n}}$$

donde C es una constante tal que  $1/\sqrt{2\pi} \leq C < .8$ .

Demostración. Sin pérdida de generalidad y para simplificar las expresiones, hagamos  $\sigma^2 = 1$  y  $\beta_3 = E(|X_k|^3)$ . Partimos de la desigualdad de Esseen (A1.19) para encontrar la siguiente cota:

$$\sup_x |F_n(x) - \Phi(x)| \leq \frac{2}{\pi} \int_0^T \left| \frac{\phi_n(t) - \theta(t)}{t} \right| dt + \frac{24}{\pi T} \frac{1}{\sqrt{2\pi}} \quad (13)$$

donde T es un número real arbitrario,  $\phi_n(t)$  y  $\theta(t) = e^{-t^2/2}$  son las funciones características de  $F_n(x)$  y de la distribución normal estándar, respectivamente; y donde además se ha utilizado

$$\sup_x \left| \theta'(x) \right| = \sup_x \left| \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right| = \frac{1}{\sqrt{2\pi}}$$

Además, especificando a  $\phi_n(t)$  tenemos

$$\phi_n(t) = \phi_{S_n}(t) = \phi_{\sum_{k=1}^n X_k}(t) = \phi(t/\sqrt{n}) = [\phi(t/\sqrt{n})]^n \quad (14)$$

donde  $\phi(t) = E(e^{itX})$

Como en (13) podemos tomar arbitrariamente a T, hagámosla igual a

$$T = \sqrt{n}/(5\beta_3)$$

Con esta T acotaremos a  $|\phi_n(t) - \theta(t)|$ .

Por ello, con la fórmula de Taylor (A1.15), considerando el residuo integral (A1.18) tenemos

$$\phi(t) = \phi(0) + \frac{t}{1!} \phi'(0) + \frac{t^2}{2!} \phi''(0) + \frac{t^3}{2} \int_0^1 (1-y)^2 \phi'''(yt) dy \quad (15)$$

y con  $\phi^{(k)}(0) = i^k E(X^k)$

$$\phi(t) = 1 - \frac{t^2}{2!} + \frac{t^3}{2} \int_0^1 (1-y)^2 \phi'''(yt) dy$$

entonces

$$\phi(t/\sqrt{n}) = 1 - \frac{t^2}{2n} + \frac{t^3 \beta_3 \theta}{6n^{3/2}} \quad (16)$$

donde  $|\theta| \leq 1$  pues  $|\phi'''(yt)| \leq |\phi'''(0)| = |i^3 E(X^3)| \leq E(|X|^3) = \beta_3$  y por lo tanto  $\phi'''(yt) = \theta \beta_3$ .

Ahora, y por la desigualdad de Lyapunov (A1.20) tenemos que  $\beta_3 = E(|X|^3) \geq E(|X|^2)^{3/2} \geq E(X^2)^{3/2} = \sigma^3 = 1$  y si  $|t| \leq T$  podemos afirmar que

$$t^2 \leq \frac{n}{25} \quad \text{y} \quad |t^3| \leq \frac{n^{3/2}}{125}$$

con lo cual obtenemos

$$\frac{t^2}{2n} + \frac{|t^3| \beta_3 \theta}{6n^{3/2}} \leq \frac{1}{50} + \frac{1}{750} \leq \frac{1}{25}$$

y en consecuencia acotamos (16)

$$\phi(t/\sqrt{n}) \geq 1 - \frac{t^2}{2n} - \frac{|t^3| \beta_3 \theta}{6n^{3/2}} \geq \frac{24}{25} \quad (17)$$

Por otro lado, apliquemos (15) a  $\ln \phi(t/\sqrt{n})$ , considerando por supuesto sus derivadas, las cuales quedan nuevamente en términos de las de  $\phi(t/\sqrt{n})$ , las evaluamos en cero y obtenemos

$$\begin{aligned} \ln \phi(t/\sqrt{n}) &= -\frac{t^2}{2n} + \frac{t^3}{2n^{3/2}} \int_0^1 (1-y)^2 \ln \phi'''(yt/\sqrt{n}) dy \\ &= -\frac{t^2}{2n} + \frac{t^3}{6n^{3/2}} \ln \phi'''(\theta_1 t/\sqrt{n}) \end{aligned}$$

la segunda igualdad debida a la continuidad de  $\phi(t)$  y al teorema del valor intermedio generalizado (A1.23) con  $|\theta_1| \leq 1$ . Asimismo, podemos expresar la tercera derivada del logaritmo en términos de las de la función característica  $\phi$ , acotarlas por los momentos de los valores absolutos, y conseguir

$$|\ln \phi''(t)| =$$

$$\begin{aligned} &= | [\phi''(t) (\phi(t))^2 - 3\phi'(t)\phi'(t)\phi(t) + 2(\phi'(t))^2] \phi(t)^{-3} | \leq \\ &\leq | [\phi''(0) (\phi(0))^2 - 3\phi'(0)\phi'(0)\phi(0) + 2(\phi'(0))^2] \phi(t)^{-3} | = \\ &= | [E(x^2)(E(x^0))^2 - 3E(x^2)E(x)E(x^0) + 2(E(x))^2] (24/25)^{-3} | \leq \\ &= [ |E(x^2)| + 3 |E(x)|^2 + 2 |E(x)|^2 ] (25/24)^3 \leq \\ &= 6\beta_2 (25/24)^3 \leq 7\beta_2 \end{aligned}$$

apoyados en A1.7, A1.8, A1.20 y (17).

Entonces tenemos

$$\ln \phi(t/\sqrt{n}) \leq -\frac{t^2}{2n} + \frac{t^3}{6n^{3/2}} \frac{7\beta_2}{n} \quad (18)$$

Estamos ahora en posición de acotar la desigualdad de interés:

$$|\phi_n(t) - \theta(t)| = |e^{n \ln \phi(t/\sqrt{n})} - e^{-t^2/2}| \leq$$

(por (14))

$$\leq \left| \exp n \left[ -\frac{t^2}{2n} + \frac{t^3}{6n^{3/2}} \frac{7\beta_2}{n} \right] - e^{-t^2/2} \right|$$

(por (16))

$$\leq \frac{7\beta_2 |t^3|}{6\sqrt{n}} \exp \left[ -\frac{t^2}{2} + \frac{7}{6} |t|^3 \frac{\beta_2}{\sqrt{n}} \right]$$

(por A1.21)

$$\leq \frac{7\beta_2 |t|^3}{6\sqrt{n}} \exp \left[ -\frac{t^2}{2} + \frac{7t^2}{30} \right]$$

$$\text{si } |t| \leq T = \frac{\sqrt{n}}{5\beta_2}$$

O sea,

$$|\phi_n(t) - \theta(t)| \leq \frac{7\beta_2 |t|^3}{6\sqrt{n}} \exp \left[ -\frac{t^2}{4} \right]$$

Utilizando esta desigualdad en (13) podemos concluir que

$$\begin{aligned}
\sup_x |F_n(x) - \Phi(x)| &\leq \frac{2}{\pi} \int_0^{\pi} \left| \frac{\phi_n(t) - \theta(t)}{t} \right| dt + \frac{24}{\pi!} \frac{1}{\sqrt{2\pi}} \omega \\
&\leq \frac{2}{\pi} \int_0^{\sqrt{n}/\sigma\beta_2} \frac{7\beta_2 t^2}{6\sqrt{n}} e^{-t^2/4} dt + \frac{120 \beta_2}{\pi \sqrt{2\pi n}} \\
&\leq \frac{\beta_2}{\sqrt{n}} \left[ \frac{56 \Gamma(3/2)}{6\pi} + \frac{120}{\pi\sqrt{2\pi}} \right] \\
&= \frac{C\beta_2}{\sqrt{n}}
\end{aligned}$$

(un análisis más detallado muestra que  $C < .8$ )

□

#### 4.2 Aplicación de la cota de Berry-Esseen.

En el siguiente cuadro aparecen distribuciones "estandarizadas" (media cero y varianza la unidad) que presentan distintas características para las cuales se ha calculado la cota de Berry y Esseen con el objeto de observar cómo varía la rapidez con que se cumple la convergencia a la distribución normal.

Rapidez de Convergencia.

$$\mu = 0$$

$$\sigma^2 = 1$$

$$\text{cota: } \frac{C E(|x|^3)}{\sigma^3 \sqrt{n}}$$

$$1/\sqrt{2\pi} \leq C < .6$$

- 1 -

DISTRIBUCION	$f(x)$	DISC	SIMET	DOM AC	$E( x ^3)$	COTA
Bernoulli	$1/2$ $(-1, 1)$	si	si	si	1	$\frac{C}{\sqrt{n}}$
Uniforme	$1/2\sqrt{3}$ $(-\sqrt{3}, \sqrt{3})$	no	si	si	$\frac{3\sqrt{3}}{4}$	$\frac{1.299C}{\sqrt{n}}$
Bernoulli	$1/4$ $x = -\sqrt{3}$ $3/4$ $x = 1/\sqrt{3}$	si	no	si	$\frac{5\sqrt{3}}{6}$	$\frac{1.443C}{\sqrt{n}}$
Triangular	$(x+2\sqrt{2})/9$ $(-2\sqrt{2}, \sqrt{2})$	no	no	si	$\frac{46\sqrt{2}}{45}$	$\frac{1.446C}{\sqrt{n}}$

Rapidez de Convergencia.

$\mu = 0$

$\sigma^2 = 1$

cota:  $\frac{E(|x|^3)}{\sigma^3 \sqrt{n}}$

$1/\sqrt{2\pi} \leq C < .8$

- 2 -

DISTRIBUCION	f(x)	DISC	SIMET	DOM AC	E( x ^3)	COTA
t de Student k = 4	$\frac{\Gamma(5/2)}{\Gamma(2)\sqrt{2\pi}(1+x^2)^{5/2}}$ $(-\infty, \infty)$	no	si	no	$2\sqrt{2}$	$\frac{2.028C}{\sqrt{n}}$
t de Student k = 3	$\frac{\Gamma(2)}{\Gamma(\frac{3}{2})\sqrt{n}(1+x^2)^2}$ $(-\infty, \infty)$	no	si	no	$\infty$	No existe
Geométrica p = 3/4	$\frac{3}{4} \left[ \frac{1}{4} \right]^{\frac{2}{3}} x + \frac{1}{3}$ $-\frac{1}{2}, 1, \frac{5}{2}, \dots$	si	no	no	$\frac{83}{32}$	$\frac{2.594C}{\sqrt{n}}$
Geométrica p = 1/4	$\frac{1}{4} \left[ \frac{3}{4} \right] 2\sqrt{x} + 3$ $-\frac{\sqrt{3}}{2}, -\frac{\sqrt{3}}{3}, -\frac{\sqrt{3}}{6}, 0$	si	no	no	$\frac{1291 \sqrt{3}}{256}$	$\frac{8.735C}{\sqrt{n}}$
Geométrica p = 1/2	$\left[ \frac{1}{2} \right] \sqrt{x} + 2$ $-\frac{\sqrt{2}}{2}, 0, \frac{\sqrt{2}}{2}, \dots$	si	no	no	$\frac{53 \sqrt{2}}{8}$	$\frac{9.369C}{\sqrt{n}}$

Rapidez de Convergencia.

$$\mu = 0$$

$$\sigma^2 = 1$$

$$\text{cota: } \frac{C E(|x|^q)}{\sigma^q \sqrt{n}}$$

$$1/\sqrt{2\pi} \leq C < .8$$

- 3 -

DISTRIBUCION	f(x)	DISC	SIMET	DOM AC	E( x ^q)	COTA
Exponencial	$e^{-(x+1)}$ $(-1, \infty)$	no	no	no	$\frac{12}{e} - 2$	$\frac{2.415C}{\sqrt{n}}$
Pareto	$81(x+2\sqrt{2})^{-5}$ $(-1/\sqrt{2}, \infty)$	no	no	no	$\frac{41\sqrt{2}}{8}$	$\frac{7.248C}{\sqrt{n}}$
Pareto	$\frac{8\sqrt{2}}{9}(x+\sqrt{3})^{-4}$ $(-1/\sqrt{3}, \infty)$	no	no	no	$\infty$	No existe



## 5. OTRA DEMOSTRACION DEL TEOREMA CENTRAL DEL LIMITE.

En la literatura sobre el tema se puede encontrar una demostración alternativa basada en la función generadora de momentos (R. Tardiff, 1981), la cual es relativamente más fácil de manipular pues a diferencia de la función característica, se trata de una función real; sin embargo es esta una restricción adicional ya que no necesariamente una variable aleatoria posee función generadora.

En seguida se presenta esta prueba, donde se utiliza la regla de L'Hospital para encontrar el valor del límite que juega un papel central en la demostración, en vez del desarrollo de Taylor que es el usualmente aplicado.

**Teorema 1.5.** Sea  $\{X_n\}$  una sucesión de variables aleatorias independientes e idénticamente distribuidas con media  $\mu$ , varianza  $\sigma^2$  y función generadora  $\psi_X(t)$ , que existe en  $-h < t < h$ . Entonces

$$P\left\{ z_1 < \frac{\sum X_i - n\mu}{\sqrt{n}\sigma} < z_2 \right\} \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-z^2/2} dz$$

Demostración. Sea

$$\zeta(t) = E\left( e^{t(X-\mu)/\sigma} \right) = e^{-\mu t/\sigma} \psi_X(t/\sigma)$$

por lo que, considerando A1.13,  $\zeta(0) = 1$ ,  $\zeta'(0) = 0$ , y  $\zeta''(0) = 1$ ; y desarrollando

$$\begin{aligned} \psi_n(t) &= E\left[ \exp\left( t \frac{\sum X_i - n\mu}{\sqrt{n}\sigma} \right) \right] \\ &= E\left[ \prod_{i=1}^n \exp\left( t \frac{X_i - \mu}{\sqrt{n}\sigma} \right) \right] \\ &= \prod_{i=1}^n E\left[ \exp\left( t \frac{X_i - \mu}{\sqrt{n}\sigma} \right) \right] \\ &= \left\{ E\left[ \exp\left( t \frac{X - \mu}{\sqrt{n}\sigma} \right) \right] \right\}^n \\ &= \left[ \zeta\left( \frac{t}{\sqrt{n}} \right) \right]^n \end{aligned}$$

para  $-h < \frac{t}{\sqrt{n}} < h$ .

Encontremos ahora el límite de la expresión que hemos hallado:

$$\begin{aligned} L &= \lim_{n \rightarrow \infty} \psi_n(t) \\ &= \lim_{n \rightarrow \infty} \left[ \zeta \left( \frac{t}{\sqrt{n}} \right) \right]^n \end{aligned}$$

y tomando logaritmos

$$\ln(L) = \lim_{n \rightarrow \infty} \left[ \frac{\theta \left( t/\sqrt{n} \right)}{1/\sqrt{n}} \right]$$

donde  $\theta(t) = \ln \zeta(t)$

Puesto que  $\theta(0) = \ln \zeta(0) = \ln 1 = 0$ , tanto el límite del numerador como del denominador tienden a cero cuando  $n$  tiende a infinito, y por lo tanto podemos aplicar la regla de L'Hospital (A1.22) para evaluarlo, usando la regla de la cadena para hallar la derivada del denominador:

$$\ln(L) = \frac{1}{2} \lim_{n \rightarrow \infty} \frac{t \theta' \left( t/\sqrt{n} \right)}{1/\sqrt{n}}$$

Observando ahora que  $\theta'(t) = (\ln \zeta)'(t) = \zeta'(t) / \zeta(t)$ , y por lo tanto  $\theta'(0) = 0$ , puede aplicarse nuevamente la regla de L'Hospital y obtener

$$\ln(L) = \frac{1}{2} \lim_{n \rightarrow \infty} t^2 \theta'' \left( t/\sqrt{n} \right)$$

para que finalmente, por  $\theta''(t) = [\zeta(t)\zeta''(t) - (\zeta'(t))^2] / \zeta^2(t)$  tengamos  $\theta''(0) = 1$  y entonces

$$\ln(L) = \frac{1}{2} t^2$$

o

$$L = \exp(t^2/2)$$

que reconocemos como la función generadora de la distribución normal (A1.11), y por A1.12, aseguramos que la convergencia de la probabilidad del enunciado del teorema se cumple. ■

## 6. EXTENSIONES.

De acuerdo a las cotas obtenidas para las distintas distribuciones en los cuadros de la sección 4 sobre rapidez de convergencia, puede apreciarse que el factor más relevante para lograr una aproximación más rápida es que la distribución sea de dominio acotado, sin importar demasiado si la función es continua o simétrica.

En caso contrario, si el dominio de la función de distribución no está acotado, la velocidad con que se cumple la convergencia es menor, pues crece la cota de Berry-Esseen, dándose incluso ejemplos en que ésta no existe debido a que el "peso" de las colas es muy alto respecto al "centro" de la distribución. Precisamente son estas situaciones las que nos hacen reflexionar acerca del papel que juega la existencia de una varianza finita para cada variable aleatoria en la sucesión (aunque en nuestros cuadros todas las distribuciones presentadas tienen varianza igual a uno para fines de comparación).

Aún más, lo anterior invita a reflexionar sobre lo que ocurriría si no contáramos con esta condición. ¿Cabría esperar que se siguiera cumpliendo la convergencia de la suma de variables aleatorias a la distribución Normal?, o, ante una respuesta negativa, ¿Seguiría convergiendo a otra u otras distribuciones?, O, ¿ni siquiera el límite es una función de distribución, o bien no siempre existe?. Estas interrogantes constituyen precisamente parte del tema de estudio del siguiente capítulo donde trataremos con la familia de distribuciones conocidas como "Leyes Estables".

Las direcciones en que se puede explorar la validez del Teorema Central del Límite son variadas. Como ejemplo de esta afirmación supongamos que mantenemos las condiciones del teorema 1.3, excepto que no se fija de antemano el número de sumandos sino que se considera aleatorio bajo ciertas condiciones muy generales. Esta situación se presenta en el contexto del Muestreo cuando el número de individuos a analizar se determina por el propio proceso por el cumplimiento de cierto requisito. Cabe preguntarse si la distribución normal sigue siendo útil como aproximación. El siguiente resultado contesta afirmativamente tal cuestión.

Teorema 1.6. Sea  $X_1, X_2, \dots$  una sucesión de variables aleatorias independientes idénticamente distribuidas con  $E(X_i) = \mu$ ,  $E(X_i^2) < \infty$ , y denotemos  $\text{Var}(X_i) = \sigma^2$ . Además sea  $V_1, V_2, \dots$  una sucesión de variables aleatorias que sólo toma valores enteros positivos y que cumplen con que  $V_n \rightarrow \infty$  si  $n \rightarrow \infty$ . Entonces si

denotamos  $S_{V_n} = \sum_{i=1}^{V_n} X_i$  se cumple que

$$P \left[ \frac{S_n - \mu}{\sigma/\sqrt{n}} \leq x \right] \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy \quad \text{si } n \rightarrow \infty$$

Este es un caso particular que representa un elemento motivador de las extensiones que se establecerán en los capítulos subsiguientes.

## 7. BIBLIOGRAFIA COMENTADA.

- Gnedenko, 1949. Este es un libro clásico de Probabilidad en el que se aborda el Teorema Central del Límite. Dedicó un capítulo a la discusión de los ensayos de Bernoulli donde se presentan y demuestran los teoremas 1.1 y 1.2 correspondientes al caso local y de intervalos, respectivamente, del resultado debido a de Moivre y Laplace. Se demuestra el teorema 1.3 como corolario de uno más amplio que corresponde al contenido del siguiente capítulo de este trabajo.
- Rényi, 1976. Este libro trata de manera extensa el tema, presenta resultados interesantes un tanto particulares, pero con relativa sencillez. Presenta con detalle las demostraciones de los teoremas 1.1 y 1.2; incluye también la del 1.3 donde retoma el 1.2 para señalarlo como un caso particular. La demostración del teorema 1.6 aparece en esta obra.
- Shiryayev, 1984. El Teorema Central del Límite está muy bien tratado en este libro, incluso más allá de lo contemplado en el presente capítulo. Demuestra con cuidado los teoremas 1.1, 1.2, 1.3 y 1.4 (acerca de este último enfatiza el concepto de rapidez de convergencia).
- Feller, 1950. Es otra obra que se anota frecuentemente como referencia en la literatura. Aquí primero se aborda el teorema 1.1 para  $p = 1/2$  por razones de sencillez y para distinguir el caso simétrico, para después extenderlo a  $p \neq 1/2$ . También se presenta el teorema 1.2 con una demostración muy simplificada. Demuestra el teorema 1.3 por dos métodos; justifica el 1.4 incluyendo discusiones en torno a él, así como la del 1.6. Desafortunadamente, el material se presenta muy disperso en el texto, lo que dificulta su consulta.
- Gnedenko y Kolmogorov, 1954. Es una obra clásica dedicada por entero al estudio de la convergencia de sumas de variables aleatorias que sean independientes que avanza de lo general a lo particular. En una brillante introducción aborda el enunciado del teorema 1.2, aunque sin demostrarlo. En cambio, aparecen las demostraciones del teorema 1.3 y 1.4.
- Harris, 1966. Hasta el último capítulo de este libro se han

concentrado los resultados acerca de teoremas límite, y en él aparece la demostración del teorema 1.3 en la línea aquí seguida. El teorema 1.2 es presentado como caso particular del anterior y se utiliza para ilustrar el procedimiento de corrección por discontinuidad.

Tucker, 1967. En un tratamiento similar a Gnedenko y Kolmogorov (1954), presenta un enfoque más general en que el teorema 1.3 se reporta como corolario.

Hoel, Port y Stone, 1971. El teorema 1.3 es demostrado en este texto utilizando también el método de funciones características.

Todhunter, 1865. Se trata de una extensa obra que hace una revisión histórica que abarca desde los orígenes de la Probabilidad hasta los trabajos de Laplace. A ella se recurre como fuente fidedigna para establecer créditos y fechas de publicación de aportaciones tales como las de deMoivre y Laplace sobre el Teorema Central del Límite.

## CAPITULO II

### CONVERGENCIA DE SUMAS DE VARIABLES ALEATORIAS INDEPENDIENTES: LEYES ESTABLES E INFINITAMENTE DIVISIBLES.

#### 1. INTRODUCCION.

El objetivo de este capítulo consiste en generalizar los resultados alcanzados al atenuar las condiciones impuestas a distribuciones de las variables aleatorias y encontrar alcances y limitaciones de la convergencia a la distribución normal. Primero se elimina la restricción de que la varianza de las variables aleatorias sea finita, lo cual abre paso a los casos donde existe una mayor dispersión en la distribución. En segundo lugar, se quita la petición de idéntica distribución, que en la práctica significa no circunscribirse a situaciones donde las condiciones experimentales se preservan totalmente.

En el primer caso conviene dar un esbozo de las distribuciones o leyes estables que son la familia de variables aleatorias con la propiedad de que la convergencia de sumas estandarizadas de variables aleatorias independientes e idénticamente distribuidas se realiza hacia alguna ley estable, y solamente a alguna de ellas. De esto se ocupa la sección 2. Asimismo se mencionan algunas particularidades, como la forma de su función característica o que son cerradas respecto a la convolución; también se pretende destacar algunos miembros importantes de esta familia. Para una distribución estable dada, al conjunto de variables aleatorias cuya suma estandarizada converge a ella se conoce como el dominio de atracción, lo cual, junto con las condiciones que se requieren para determinar si alguna variable aleatoria pertenece a este dominio, es tratado también en la sección 2.

La familia de distribuciones a las que converge una suma cuando los componentes no son variables idénticamente distribuidas se amplía enormemente y es la constituida por las llamadas infinitamente divisibles, aquellas que se pueden representar como la suma de cualquier número finito de variables aleatorias independientes e idénticamente distribuidas. Por ello, se presenta en la sección 3 una caracterización y propiedades de las distribuciones infinitamente divisibles, así como el resultado que establece la convergencia de sumas de variables aleatorias independientes solo y únicamente a distribuciones de esta familia; en este contexto, aparte de la convergencia a la distribución normal, se menciona como ejemplo la aproximación de Poisson a la distribución binomial.

Como conclusión del capítulo, se presenta en la sección 4 a la distribución normal como un caso particular de las distribuciones estables y de las infinitamente divisibles para

delimitar los casos en que se cumple la convergencia a esta distribución y ubicarla en un contexto más amplio donde la distribución límite puede ser alguna otra. Se presenta el teorema que establece que la suma de variables aleatorias independientes idénticamente distribuidas converge a la distribución normal solamente si poseen varianzas finitas, así como el famoso teorema de Lindeberg-Feller que explicita las condiciones necesarias y suficientes para la convergencia a la normal en el caso de que a los sumandos sólo se les requiera que sean independientes.

## 2. CONVERGENCIA DE SUMAS DE VARIABLES ALEATORIAS INDEPENDIENTES IDÉNTICAMENTE DISTRIBUIDAS: CARACTERIZACIÓN DE LEYES ESTABLES.

### 2.1 Aplicación y particularidades.

Ante la interrogante de reconocer la familia de distribuciones a las que puede converger la suma estandarizada de variables aleatorias independientes e idénticamente distribuidas, en la década de los 30's, matemáticos de la talla de Khintchine, Lévy y Gnedenko, identificaron esa familia como las leyes estables y se avocaron a caracterizar propiedades y condiciones requeridas para la convergencia. Los resultados establecieron en definitiva la respuesta a la convergencia y mostraron una familia de distribuciones muy interesante pues presentan colas en la distribución cuyo peso probabilístico es no despreciable, son flexibles y capaces de reflejar las propiedades de estructuras de probabilidad en las que los momentos muestrales tienen un comportamiento errático.

Como un ejemplo simple de la presencia de distribuciones estables podemos considerar una caminata aleatoria formada de la siguiente manera: se observa la diferencia  $D_n$  entre el número de "soles" y "águilas" que ocurren al lanzar una moneda justa  $n$  lanzamientos ( $D_0 = 0$ ), para la que interesa conocer el evento de que  $D_n$  alcance cierto valor fijo  $r$  en los primeros  $N$  lanzamientos. Si este número de lanzamientos se pone en función cuadrática de la diferencia deseada  $r$  (específicamente  $N = 2 \lfloor kr^2 \rfloor$ ,  $k$  constante mayor que cero) entonces la distribución límite de esta variable cuando  $r \rightarrow \infty$  es la ley estable conocida con el nombre de distribución de Lévy.

Las características de las leyes estables han encontrado aplicación en distintos campos, pudiéndose destacar el estudio sobre cotizaciones de acciones (common stock returns), donde son de utilidad propiedades consistentes con la teoría económica de precios, tales como la convergencia de las leyes estables y la peculiaridad -a la que deben su nombre- de converger su suma estandarizada a sí mismas. Estudios de Fama (1965) y Fama y Roll (1971) modelan satisfactoriamente el comportamiento de precios de

acciones con distribuciones estables simétricas; posteriormente Rozelle y Fielitz (1980), Simkowitz y Beedles (1980) y Fielitz y Rozelle (1983), han empleado leyes estables asimétricas con el objeto de explicar más adecuadamente la distribución muestral sesgada de las cotizaciones.

Aparecen también en procesos de ramificación como el siguiente. Dada una partícula de cierto tipo, después de un intervalo de tiempo igual a la unidad, ésta puede desaparecer (muerte) con probabilidad  $p > 0$ , dividirse en dos partículas idénticas con la misma probabilidad  $p$ , o permanecer sin cambio con probabilidad  $1 - 2p$ . Entonces puede modelarse el comportamiento del número total de partículas que desaparecen cuando en el proceso se considera un número infinito de generaciones mediante una distribución estable con parámetro de escala  $\alpha = 1/2$  (más adelante se especificará quién es este parámetro). Con procesos de ramificación semejantes al descrito pueden resolverse problemas relacionados con la teoría de la evolución de las especies, tales como la determinación del número de especies con que puede contar un género dado, cuya probabilidad decrece a razón  $n^{-1/2}$ , por lo que ha sido posible modelarla con la ley debida a Willis (1922) y Yule (1924) que involucra la distribución estable de Lévy mencionada anteriormente.

Por último, para dar mejor cuenta de la amplia gama de aplicaciones de las leyes estables, se menciona una en el campo de la radiación, donde se desarrolla un experimento consistente en colocar una fuente radioactiva frente a una pantalla plana en la que se registran los haces luminosos producidos al chocar con ella las partículas radioactivas. La distribución de los haces luminosos en el plano es circular y simétrica por condiciones de regularidad supuestas en la fuente emisora, con lo que su posición aleatoria (U,V) sigue la misma distribución en ambos componentes; dicha distribución sigue la ley de Cauchy, que resulta ser una ley estable. Otros fenómenos físicos obedecen la distribución Cauchy, tales como la distribución de la energía en estados inestables en reacciones de productos radioactivos llamada distribución de Lorentz; y la distribución de Breit-Wigner de la masa de partículas en el mismo tipo de reacciones de materiales radioactivos. Las dos distribuciones son casos particulares de la distribución de Cauchy.

Definición 2.1. La función de distribución  $F(x)$  es una ley estable si existe una sucesión  $X_1, X_2, \dots, X_n, \dots$  de variables aleatorias independientes idénticamente distribuidas tal que  $F(x)$  es el límite en distribución de la suma estandarizada de la forma

$$Y_n = \frac{1}{B_n} \sum_{k=1}^n X_k - A_n$$

También se dice que la variable aleatoria con función de distribución estable, es estable. ▶

Resulta relevante conocer cuándo se da la convergencia a alguna ley estable en particular, máxime que la distribución normal representa uno de estos casos. Ello nos conduce a la siguiente:



**Definición 2.2.** Sean  $X_1, \dots, X_n, \dots$  variables aleatorias independientes con idéntica distribución  $F(x)$ . Si para unas constantes  $A_n$  y  $B_n > 0$  las funciones de distribución de las sumas

$$Y_n = \frac{1}{B_n} \sum_{k=1}^n X_k - A_n$$

convergen en distribución a una función de distribución  $V(x)$  cuando  $n$  tiende a infinito, se dice que  $F(x)$  es atraída por  $V(x)$ . Al conjunto de funciones atraídas por  $V(x)$  se le llama dominio de atracción de  $V(x)$ .

De esta definición y de la de leyes estables se desprende que las únicas distribuciones con dominio de atracción no vacío son precisamente las estables.

Alternativamente, se puede mostrar que la función de distribución  $F(x)$  es estable si para cada  $a_1 > 0, a_2 > 0, b_1, b_2$  corresponden constantes  $a > 0$  y  $b$  tales que se cumple la ecuación

$$F(a_1 x + b_1) * F(a_2 x + b_2) = F(ax + b)$$

donde el operador  $*$  significa convolución de funciones. Esta equivalencia puede ser interpretada diciendo que una ley es estable si todas las convoluciones que se formen con ella tienen una distribución de la misma familia.

Si se generaliza a  $n$  términos, la variable aleatoria  $X$  y su función de distribución  $F(x)$  son estables si para toda  $n \geq 1$  y  $X_1, \dots, X_n$  idénticamente distribuidas como  $X$ , existen constantes  $a_n > 0$  y  $b_n$  tales que

$$a_n X + b_n \stackrel{d}{=} X_1 + \dots + X_n$$

$$F[(x-b_n)/a_n] = F * \dots * F(x) \quad (n \text{ veces})$$

$$[\phi(t)]^n = [\phi(a_n t)] e^{i b_n t}$$

siendo  $\phi(t)$  la función característica de  $X$ .

De aquí se puede desprender una manera alternativa de reconocer una distribución estable.

**Teorema 2.1.** (A. Ya. Khintchine y P. Lévy, 1936). Para que la función de distribución  $F(x)$  sea estable es necesario y suficiente que su función característica tenga la forma

$$\phi(t) = \exp [i\gamma t - c|t|^\alpha [1 + i\beta \frac{t}{|t|} \omega(t, \alpha)]]$$

donde  $0 < \alpha \leq 2$ ,  $-1 \leq \beta \leq 1$ ,  $\gamma \in \mathbb{R}$ ,  $c \geq 0$  y

$$\omega(t, \alpha) = \begin{cases} \tan(\pi\alpha/2) & \text{si } \alpha \neq 1 \\ 2/\pi \ln(|t|) & \text{si } \alpha = 1 \end{cases}$$

El parámetro  $\alpha$ , conocido como exponente característico o índice de estabilidad, está asociado con el nivel de dispersión de los valores que toma la variable aleatoria, y está restringido al intervalo  $(0, 2]$  para que se trate de una función característica;  $\gamma$  es un parámetro de localización,  $\beta$  es el parámetro relacionado con el sesgo, siendo la distribución simétrica respecto a  $\gamma$  si  $\beta = 0$ , sesgada a la derecha si es positiva y a la izquierda si es negativa. Además, la variable aleatoria toma solo valores positivos si y sólo si  $\beta = -1$ ,  $\alpha < 1$  y  $\gamma > 0$  y, caso contrario, toma valores negativos si y sólo si  $\beta = 1$ ,  $\alpha < 1$ , y  $\gamma > 0$ . La constante  $c$  es un parámetro de escala.

También se puede mostrar que todas las distribuciones estables propias (aquellas cuya función de distribución no está concentrada en un solo punto) son unimodales (I. A. Ibragimov y Chernin, 1959), continuas y tienen derivadas de todos los órdenes en cada punto (A Ya. Khintchine, 1938). Así, podemos eliminar un sinnúmero de distribuciones que no son estables por no tener su función característica la forma adecuada, por ser discretas, multimodales o no tener en todo punto derivadas de cualquier orden.

## 2.2 Ejemplos de leyes estables.

Podemos ejemplificar como miembros de la familia de las leyes estables a la distribución normal tomando  $\alpha = 2$ ; a la distribución Cauchy si hacemos  $\alpha$  igual a 1 y  $\beta$  igual a cero; a la distribución con densidad

$$f(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{1}{\sqrt{2\pi}} e^{-1/2x} x^{-3/2} & \text{si } x > 0 \end{cases}$$

que corresponde al caso en que  $\alpha = 1/2$ ,  $\beta = 1$ ,  $\gamma = 0$  y  $c = 1$ , la cual se conoce como distribución de Lévy. La distribución unitaria es otro ejemplo de distribución estable; y aquellas que tienen como función característica a

$$\phi(t) = e^{-d|t|^\alpha}$$

con  $d \geq 0$  y  $0 < \alpha \leq 2$ , además de ser estables, son simétricas.

Desafortunadamente, y a pesar de que la existencia de derivadas de todos los órdenes en cada punto de las distribuciones estables asegura la existencia de la función de densidad, no es posible hallar una expresión cerrada para ésta excepto para los casos de las distribuciones Normal, Cauchy y de Lévy, arriba

señalados.

### 2.3 Dominio de Atracción Estándar y Condiciones de Atracción.

Se anotan en el siguiente resultado las condiciones mínimas suficientes para determinar la pertenencia de una distribución al dominio de atracción de una ley estable, excluyendo el caso de la distribución normal, que dejaremos para la sección número 4.

Teorema 2.2. (B. V. Gnedenko 1939, y W. Doeblin 1940).  $F(x)$  pertenece al dominio de atracción de una ley estable con exponente  $\alpha$  ( $0 < \alpha < 2$ ) si y sólo si

$$\lim_{x \rightarrow \infty} \frac{F(-x)}{1 - F(x)} = \frac{c_1}{c_2}$$

y si para toda constante  $k > 0$

$$\lim_{x \rightarrow \infty} \frac{1 - F(x) + F(-x)}{1 - F(kx) + F(-kx)} = k^\alpha$$

donde  $c_1$  y  $c_2$  cumplen con  $c_1 \geq 0$ ,  $c_2 > 0$  aparte de estar involucradas en la representación de Lévy para la función característica de la ley estable, considerada como distribución infinitamente divisible, tal como se verá en la sección 3.  $\uparrow$

La primera condición significa que las colas izquierda y derecha de la distribución están "balanceadas", en el sentido de que en el límite son proporcionales entre sí, con proporcionalidad igual a  $c_1/c_2$ . Por el segundo límite se dice que las leyes estables (excepto la distribución normal) tienen "colas pesadas" y que decaen regularmente a infinito, puesto que la razón entre el "peso" de las colas truncadas hasta un punto cualquiera en una cola y hasta su inverso aditivo en la otra, y el correspondiente a las colas truncadas hasta un múltiplo de los mismos puntos -que se encuentran  $k$  veces más alejados hacia los extremos- tiende a mantenerse igual a una constante, determinada solamente por el múltiplo considerado y por el índice de estabilidad  $\alpha$ .

En este estudio resulta de singular interés escoger los coeficientes  $B_n$  proporcionales a  $n^{1/\alpha}$ , ya que precisamente son de esa forma los que se toman en la teoría clásica de convergencia a la normal, y lo que pretendemos es generalizar esos resultados. También resulta claro que al agregar esta restricción, el dominio de atracción de una distribución estable se reducirá, o a lo más se mantendrá inalterado. Tales ideas serán abordadas en seguida.

Definición 2.3. Se dice que la distribución  $F(x)$  pertenece al dominio de atracción estándar de la ley estable  $V(x)$  si para una sucesión  $(X_n)$  de variables aleatorias independientes con idéntica

distribución  $F(x)$ , alguna  $a > 0$  y algunas  $A_n$  se cumple que las variables aleatorias

$$Y_n = \frac{1}{an^{1/\alpha}} \sum_{k=1}^n X_k - A_n$$

convergen en distribución a  $V(x)$ .

**Teorema 2.3.** (B. V. Gnedenko 1939). La distribución  $F(x)$  pertenece al dominio de atracción estándar de la distribución estable  $V(x)$  con parámetro  $\alpha$ ,  $0 < \alpha < 2$  si y sólo si

$$F(x) = (c_1 a^\alpha + r_1(x)) / |x|^\alpha \quad \text{para } x < 0$$

$$F(x) = 1 - (c_2 a^\alpha + r_2(x)) / |x|^\alpha \quad \text{para } x > 0$$

donde  $r_1$  y  $r_2$  cumplen con

$$\lim_{x \rightarrow -\infty} r_1(x) = \lim_{x \rightarrow \infty} r_2(x) = 0$$

Las constantes  $c_1$  y  $c_2$  son tales que  $c_1 \geq 0$ ,  $c_2 \geq 0$  y  $c_1 + c_2 > 0$  (las mismas que aparecen en el teorema anterior, excepto que se permite a  $c_2$  tomar el valor de cero). ▶

Además, se puede afirmar que toda distribución estable pertenece a su propio dominio de atracción estándar de acuerdo a la siguiente

**Proposición 2.1.** Sea  $\{X_k\}$  una sucesión de variables aleatorias independientes idénticamente distribuidas con  $X_1$ , cuya distribución  $V(x)$  es estable con exponente  $\alpha$ . Entonces para unas constantes  $a$  y  $\{A_n\}$  la nueva sucesión

$$Y_n = \frac{1}{an^{1/\alpha}} \sum_{k=1}^n X_k - A_n$$

converge en distribución a la misma  $V(x)$ .

**Demostración.** El resultado buscado equivale a

$$\phi_{Y_n}(t) \rightarrow \phi_X(t)$$

que es lo que mostraremos directamente.

Consideremos en primera instancia  $\alpha \neq 1$  y escojamos  $a = 1$ ,

$A_n = \gamma (n^{(\alpha-1)/\alpha} - 1)$ ; entonces

$$\begin{aligned}
\phi_{Y_n}(t) &= \exp[-it\gamma (n^{(\alpha-1)/\alpha} - 1)] \phi_{\sum X_i}(t/n^{1/\alpha}) \\
&= \exp[-it\gamma (n^{(\alpha-1)/\alpha} - 1)] [\phi_X(t/n^{1/\alpha})]^n \\
&= \exp \left\{ -it\gamma (n^{(\alpha-1)/\alpha} - 1) + \right. \\
&\quad \left. + n \left[ \frac{it\gamma}{n^{1/\alpha}} - c \left| \frac{t}{n^{1/\alpha}} \right| \left\{ 1 + i\beta \frac{t/n^{1/\alpha}}{|t/n^{1/\alpha}|} \omega\left(\frac{t}{n^{1/\alpha}}, \alpha\right) \right\} \right] \right\} \\
&= \exp \left\{ -it\gamma (n^{(\alpha-1)/\alpha} - 1) + \right. \\
&\quad \left. + it\gamma n^{1-1/\alpha} - c|t|^\alpha \left\{ 1 + i\beta \frac{t}{|t|} \omega\left(\frac{t}{n^{1/\alpha}}, \alpha\right) \right\} \right\} \\
&= \exp \left\{ it\gamma - c|t|^\alpha \left\{ 1 + i\beta \frac{t}{|t|} \omega(t, \alpha) \right\} \right\}
\end{aligned}$$

por la definición de  $\omega(t, \alpha)$ , con lo que identificamos nuevamente a la función característica de X.

Ahora, si  $\alpha = 1$  se escoge  $a = 1$ ,  $A_n = \frac{2}{\pi} \beta c \ln n$ . Entonces

$$\begin{aligned}
\phi_{Y_n}(t) &= \exp[-it \frac{2}{\pi} \beta c \ln n] \phi_{\sum X_i}(t/n) \\
&= \exp[-it \frac{2}{\pi} \beta c \ln n] [\phi_X(t/n)]^n \\
&= \exp \left\{ -it \frac{2}{\pi} \beta c \ln n + \right. \\
&\quad \left. + n \left[ \frac{it\gamma}{n} - c \left| \frac{t}{n} \right| \left\{ 1 + i\beta \frac{t/n}{|t/n|} \frac{2}{\pi} \ln \left| \frac{t}{n} \right| \right\} \right] \right\} \\
&= \exp \left\{ -it \frac{2}{\pi} \beta c \ln n + \right. \\
&\quad \left. + it\gamma - c|t| \left\{ 1 + i\beta \frac{t}{|t|} \frac{2}{\pi} [\ln|t| - \ln n] \right\} \right\}
\end{aligned}$$

$$= \exp \left\{ it\gamma - c|t| \left\{ 1 + i\beta \frac{t}{|t|} \omega(t,1) \right\} \right\}$$

Es interesante observar que la estandarización requerida es distinta si  $\alpha = 1$  a los demás casos; en estos últimos la  $A_n$  elegida resulta ser lógica en algún modo, pues está centrando respecto al parámetro de localización  $\gamma$ , ponderado por el tamaño  $n$  y el índice de estabilidad  $\alpha$ . En cambio, cuando  $\alpha = 1$ , la corrección será mayor a medida que aumenta el sesgo de la distribución sin que intervenga el parámetro de localización.

## 2.4 Consecuencias de las Condiciones de Atracción.

Los resultados alcanzados sobre dominios de atracción permiten establecer otras características concernientes a sus momentos y a cualidades analíticas, que darán una idea más completa acerca del comportamiento en general de las leyes estables.

Teorema 2.4. (P. Lévy 1937). Para toda distribución estable, excepto la distribuciones normal y unitaria, existen constantes  $\alpha$ ,  $0 < \alpha < 2$  y  $c_\alpha > 0$  tales que se verifica el siguiente límite

$$\lim_{x \rightarrow \infty} x^\alpha [ 1 - F(x) + F(-x) ] = c_\alpha$$

Este resultado da una interpretación más precisa de lo que significa que las distribuciones estables tengan "colas pesadas" pues da una medida de la rapidez de convergencia de las colas de la distribución al cero. En efecto, para que el límite sea finito y mayor que cero, se requiere que la velocidad con que el peso de las colas tienda a cero sea del mismo orden con el cual  $x^\alpha$  tiende a infinito; de ahí se desprende que cuanto más cercano es  $\alpha$  a cero, más lento decaen las colas, y cuanto más se aproxima  $\alpha$  a 2, el decaimiento será aproximadamente de segundo orden, que corresponde al de la distribución normal, como se verá en la sección 4.

Un resultado estrechamente relacionado con el anterior asevera que para una distribución estable ( $0 < \alpha < 2$ ) los momentos absolutos menores a  $\alpha$  son finitos, mientras que los de orden mayor o igual son infinitos, lo cual da una medida de la dispersión de la distribución. Concretando el resultado en términos de los dos momentos más usuales, la media y la varianza se concluye: las leyes estables con el exponente característico  $\alpha$  mayor que 1 y menor que 2, solamente tienen esperanza matemática y no varianza, y aquellas con  $0 < \alpha \leq 1$ , no tienen ni esperanza ni varianza.

De aquí concluimos que, excepto la distribución normal, ninguna ley estable posee varianza finita, lo cual conduce en primer término a reconsiderar el significado de "varianza" y en segundo, a destacar que las distribuciones estables tienen en

efecto colas "pesadas" pues presentan tan alta dispersión que la varianza no la puede medir; más aún, un factor determinante para que una función de distribución sea atraída o no por la normal es el comportamiento de las colas de la distribución bajo estudio, tal como se afirmará en la sección 4, donde se establecerá que la condición de convergencia es que el peso de segundo orden de las colas resulte despreciable respecto de la parte central de la distribución.

### 3. CONVERGENCIA DE SUMAS DE VARIABLES ALEATORIAS INDEPENDIENTES: LEYES INFINITAMENTE DIVISIBLES.

#### 3.1 Definición y caracterización.

Se ha resuelto la convergencia de sumas de variables aleatorias independientes e idénticamente distribuidas y se ha visto el tipo de distribuciones límite, las leyes estables. Es ahora posible desechar un supuesto más, y será el que las variables que conforman la suma compartan la misma distribución.

Directamente asociado con el intento de extender la aplicación del Teorema Central del Límite es el descubrimiento de las distribuciones infinitamente divisibles en 1937 por el francés Paul Lévy quien contribuyó ampliamente en el tema, pues fue él mismo quien replanteó las transformadas de Fourier o funciones características para reconocer a las leyes estables, además de cierto número de teoremas en que se establecen condiciones de suficiencia para asegurar el cumplimiento de la convergencia a la distribución normal. Existe además controversia sobre el crédito que corresponde a Lévy en las condiciones estrictamente necesarias para el caso de sumas de variables aleatorias independientes con distribución arbitraria, generalmente atribuidas a William Feller y que se han plasmado y difundido por medio del teorema de Lindeberg-Feller con el que se cerró en definitiva un capítulo en el estudio del Teorema Central del Límite. En este contexto también resultó significativa la incorporación del análisis sobre los dominios de atracción hecha por W. Doblin en 1939 en que se incorporaba el concepto de variación regular, que a la postre representó un enfoque más moderno pues unificó criterios y permitió reducciones significativas en los desarrollos.

Como ilustración de la aplicación de las distribuciones infinitamente divisibles, puede anotarse que en un proceso estocástico  $X(t)$  con incrementos independientes  $X(t+s) - X(s)$  y además estacionario -porque sólo depende de la longitud  $t$ , mas no de  $s$ -, precisamente los incrementos tienen una distribución que es infinitamente divisible.

Comencemos por proporcionar la definición de un variable aleatoria infinitamente divisible.

**Definición 2.4.** La variable aleatoria  $X$  es infinitamente divisible si para toda  $n \in \mathbb{R}$ , existen variables aleatorias  $X_{n1}, \dots, X_{nn}$  independientes e idénticamente distribuidas tales que

$$X = \sum_{l=1}^n X_{nl}$$

o equivalentemente, si  $\phi(t)$ , la función característica de  $X$  cumple con

$$\phi(t) = [\phi_n(t)]^n$$

para alguna función característica  $\phi_n(t)$ .

Si una variable aleatoria es infinitamente divisible, se dice también que su función de distribución es infinitamente divisible.

Tal parece que en esta ocasión el nombre de las distribuciones si resulta ser adecuado, pues su definición indica que deben tener la propiedad de poder descomponerse como la suma de  $n$  "partes iguales", sin importar que tan grande sea  $n$ ; esto es, la variable puede hacerse "infinitamente divisible".

Se anota aquí el teorema que justifica la inclusión de las distribuciones infinitamente divisibles en el trabajo, a saber, son las que resuelven la convergencia de sumas de variables aleatorias independientes (bajo algunas restricciones que den sentido al resultado).

**Teorema 2.5.** (A. Ya. Khintchine, 1937). Para cada  $n \geq 1$ , sea  $(Y_{nk})$  una sucesión de variables aleatorias independientes e idénticamente distribuidas, y sean

$$X_n = \sum_{k=1}^n Y_{nk}$$

entonces la variable aleatoria  $X$  puede cumplir con

$$X_n \xrightarrow{d} X$$

si y sólo si  $X$  es infinitamente divisible. ▶

Observese que por el enunciado se está tratando con un arreglo triangular donde el número de variables aleatorias aumenta en uno renglón a renglón, y donde las colocadas en uno mismo comparten la distribución, mientras que para las situadas en diferentes renglones no existe restricción alguna.

De ello resulta que los elementos de la propia sucesión  $(X_n)$  no tienen por que estar idénticamente distribuidos, pero si ser



independientes entre sí.

La razón de establecer la restricción sobre las variables aleatorias de un mismo renglón de compartir la misma distribución es que de otro modo la distribución límite podría ser totalmente arbitraria; por ejemplo, se puede hacer una construcción muy artificial, como escoger una variable aleatoria  $X$  cualquiera, y tomar  $Y_{nk} = X \forall n$ , y  $Y_{nk} = 0 \forall n$  y  $k > 1$ , con lo cual todas las  $X_n$  coincidirían con  $X$ , y por lo tanto evidentemente el límite sería  $X$  misma, es decir, el límite podría ser cualquier distribución.

Vale la pena hacer notar que la bien conocida aproximación de Poisson a la distribución binomial es un caso particular del teorema anterior y por lo tanto la distribución Poisson es infinitamente divisible. En las últimas líneas de esta sección se retomará esta afirmación.

Una propiedad de las distribuciones infinitamente divisibles es que  $\phi(t)$ , la función característica, cumple con que  $\phi(t) > 0 \forall t \in \mathbb{R}$ . El que la función característica de una variable aleatoria infinitamente divisible nunca se anule permite invertir la relación entre las funciones características anotadas en la definición, pudiendo expresar de manera unívoca

$$\phi_n(t) = [\phi(t)]^{1/n}$$

tomando en consideración los siguientes atributos propios de toda función característica:

- i)  $\phi_n(0) = 1$
- ii)  $\phi_n(t)$  es continua

Por lo tanto,  $\phi(t)$  o  $F(x)$  determinan sin ambigüedades a  $\phi_n(t)$  o a  $F_n(x)$ .

También se puede probar que si  $\phi(t)$  es la función característica de una distribución infinitamente divisible, entonces  $\forall c > 0$   $[\phi(t)]^c$  también es una función característica.

A continuación se especificará la forma que toma la función característica de una distribución infinitamente divisible, iniciando con un caso específico al considerar la condición de varianza finita.

Teorema 2.6. Sea  $X$  una variable aleatoria con función de distribución  $F(x)$ , función característica  $\phi(t)$ , y varianza finita. Entonces,  $F(x)$  es infinitamente divisible si y sólo si

$$\phi(t) = \exp \left[ i \gamma t + \int_{-\infty}^{\infty} [e^{itu} - 1 - i t u] \frac{1}{u^2} dK(u) \right]$$

donde  $\gamma$  es una constante, y  $K(u)$  es una función no decreciente de variación acotada que cumple con  $K(-\infty) = 0$  y  $K(\infty) < \infty$ . ▶

Esta expresión para la función característica es única y se conoce como representación de Kolmogorov, quien la desarrolló en 1932.

A  $\gamma$  se le identifica como el parámetro de localización y a  $K(u)$  se le asocia con la dispersión de la distribución pues se puede mostrar por medio de esta fórmula de Kolmogorov que  $E(x) = \gamma$  y  $\text{Var}(x) = K(\omega)$ .

Para el caso en que la varianza no sea finita, se tiene el siguiente resultado.

**Teorema 2.7.** La variable aleatoria  $X$  es infinitamente divisible si y sólo si su función característica  $\phi(t)$  tiene la forma

$$\phi(t) = \exp \left[ i c t + \int_{-\infty}^{\infty} \left[ e^{itu} - 1 - \frac{itu}{1+u^2} \right] \frac{1+u^2}{u^2} dG(u) \right]$$

donde  $c > 0$ ,  $G(u)$  es una función no decreciente de variación acotada y el integrando es igual a  $-t^2/2$  cuando  $u = 0$ .

La expresión anterior es única y se conoce como representación de Lévy-Khintchine, expuesta por ambos en 1935 y 1937, respectivamente.

Otra representación que sirve para expresar una distribución infinitamente divisible está dada por el siguiente resultado:

**Teorema 2.8.** (P. Lévy 1935). La función característica de una distribución infinitamente divisible puede expresarse como

$$\phi(t) = \exp \left[ i \gamma t - \frac{\sigma^2}{2} t^2 + \int_{-\infty}^0 \left[ e^{itu} - 1 - \frac{itu}{1+u^2} \right] dM(u) + \int_0^{\infty} \left[ e^{itu} - 1 - \frac{itu}{1+u^2} \right] dN(u) \right]$$

donde  $M(u)$ ,  $N(u)$  y  $\sigma^2$  están dadas por

$$M(u) = \int_{-\infty}^u \frac{1+z^2}{z^2} dG(z) \quad \text{para } u < 0 \quad (1)$$

$$N(u) = - \int_u^{\infty} \frac{1+z^2}{z^2} dG(z) \quad \text{para } u > 0 \quad (2)$$

$$\sigma^2 = G(+0) - G(-0)$$

Bajo esta definición,  $M(u)$  y  $N(u)$  cumplen las siguientes condiciones:

1) son no decrecientes en los dominios en que están definidas.

ii) sus puntos de continuidad coinciden exactamente con los de  $B(u)$ .

iii)  $M(-\infty) = N(\infty) = 0$ .

iv)  $\int_{-\epsilon}^0 u^2 dM(u) + \int_0^{\epsilon} u^2 dN(u) < \infty \quad \forall \epsilon > 0$ .

Recíprocamente, funciones  $M(u)$  y  $N(u)$  que satisfagan las condiciones (i), (iii) y (iv) y alguna constante  $\sigma > 0$  determinan la función característica de alguna ley infinitamente divisible mediante las mismas relaciones dadas en (1) y (2). ▶

A esta representación de la función característica de una distribución infinitamente divisible se le conoce como fórmula de Lévy.

### 3.2 Ejemplos de Leyes Infinitamente Divisibles.

Se ha pasado revista a algunas expresiones que permiten identificar una distribución infinitamente divisible mediante la forma de su función característica. Para continuar este apartado se anotarán algunos ejemplos de distribuciones pertenecientes a esta familia.

En primer lugar se señala como ley infinitamente divisible a la distribución normal, aunque se dejarán los puntos relacionados con ella para la siguiente sección, como ya se había indicado. Otros ejemplos lo constituyen la ley de Poisson y la distribución de Cauchy pues sus funciones características se pueden expresar como

$$\phi(t) = \exp [iat + \lambda(e^{it} - 1)] = \{\exp [iat/n + \lambda(e^{it} - 1)/n]\}^n$$

y

$$\phi(t) = \exp [ibt - a|t|] = \{\exp [ibt/n - a|t|/n]\}^n$$

respectivamente, donde las expresiones entre corchetes del lado derecho son a su vez funciones características de distribuciones del mismo tipo, con lo cual se confirma la aseveración.

La familia de distribuciones estables resulta ser un caso destacado y de carácter general de las infinitamente divisibles ya que la función característica de cualquier ley estable se puede expresar mediante la representación de Lévy:

$$M(u) = \frac{c_1}{|u|^\alpha} \quad N(u) = -\frac{c_2}{u^\alpha} \quad \sigma = 0 \quad \text{si } 0 < \alpha < 2$$

y

$$M(u) = N(u) = 0 \quad \sigma \geq 0 \quad \text{si } \alpha = 2$$

donde  $c_1$  y  $c_2$  son las constantes del teorema 2.3.

También son leyes infinitamente divisibles las distribuciones Gama cuya función de densidad es de la forma

$$f(x) = \begin{cases} \beta^\alpha x^{\alpha-1} e^{-\beta x} / \Gamma(\alpha) & x > 0 \\ 0 & x \leq 0 \end{cases}$$

con  $\alpha > 0$ ,  $\beta > 0$  y  $\Gamma(\alpha)$  es la función analítica Gama (ver en el apéndice A2.1), pues su función característica toma la expresión

$$\phi(t) = (1 - it/\beta)^{-\alpha}$$

que puede expresarse de acuerdo a la fórmula de Kolmogorov con  $\gamma = \alpha/\beta$  y  $K(u) = [1 - e^{-\beta u}(u\beta + 1)]/\beta^2$  ( $dK(u) = \alpha u e^{-\beta u}$ ). A estas distribuciones pertenece la Ji Cuadrada si  $\beta = 1$  y si  $2\alpha$  es un entero.

### 3.3 Convergencia a la distribución Poisson.

Separándose un poco de la línea, pero reconociendo que se trata de un interesante ejemplo de convergencia a una distribución distinta de la normal entre la familia de las infinitamente divisibles, se presenta el siguiente teorema.

Teorema 2.9. (B. V. Gnedenko 1938 y J. Marcinkiewicz 1938).

Sean  $X_{n1}, \dots, X_{nk}$   $n = 1, 2, \dots$  sucesiones de variables aleatorias infinitesimales (ver A2.2) e independientes con distribuciones  $F_{n1}, F_{n2}, \dots, F_{nk}$  respectivamente, y sea

$$X_n = \sum_{i=1}^k X_{ni}$$

La sucesión  $\{X_n\}$  converge a la distribución Poisson

$$F(x) = \sum_{k=0}^{[x]} \frac{e^{-\lambda} \lambda^k}{k!} \quad \lambda > 0$$

(donde  $[x]$  es la función máximo entero) si y sólo si se cumplen las siguientes condiciones para toda  $\epsilon$ ,  $0 < \epsilon < 1$ :

$$(i) \quad \lim_{n \rightarrow \infty} \sum_{i=1}^k \int_{\mathcal{R}_\epsilon} dF_{ni}(x) = 0$$

$$(ii) \quad \lim_{n \rightarrow \infty} \sum_{l=1}^k \int_{|x-l| < \epsilon} dF_{nl}(x) = \lambda$$

$$(iii) \quad \lim_{n \rightarrow \infty} \sum_{l=1}^k \int_{|x| < \epsilon} x dF_{nl}(x) = 0$$

$$(iv) \quad \lim_{n \rightarrow \infty} \sum_{l=1}^k \left\{ \int_{|x| < \epsilon} x^2 dF_{nl}(x) - \left[ \int_{|x| < \epsilon} x dF_{nl}(x) \right]^2 \right\} = 0$$

donde  $\mathcal{X}_\epsilon$  denota el dominio de  $X$  en  $\mathbb{R}$  quitando los intervalos  $|x| < \epsilon$  y  $|x-1| < \epsilon$ . ▶

Es decir, son condiciones necesarias y suficientes que la densidad de las distribuciones debe estar concentrada con probabilidad uno alrededor del origen y la unidad; que la densidad acumulada alrededor del uno debe tender a  $\lambda$ , el parámetro de la distribución Poisson; y que los pesos correspondientes a una vecindad alrededor del origen de la suma de los valores esperados y las varianzas, sea igual a cero.

De particular interés resulta el caso en que se aplica este teorema cuando las variables aleatorias  $X_{ni}$ ,  $i = 1, 2, \dots, k_n = n$  se distribuyen de acuerdo a una variable aleatoria Bernoulli con parámetro  $p_n$  de tal forma que  $np_n \rightarrow \lambda$  cuando  $n \rightarrow \infty$ . Fácilmente se puede mostrar que son satisfechas las condiciones del teorema. En la práctica, la aplicación de este resultado representa una alternativa para aproximar la distribución de una variable aleatoria binomial con un parámetro  $n$  pequeño o incluso moderado, y un valor del parámetro  $p$  muy pequeño, pues la suma converge más rápidamente a la distribución Poisson que la suma estandarizada a la distribución normal. Este esquema corresponde a la observación de eventos "raros" en una sucesión de ensayos realizados de manera independiente en que se conservan las mismas condiciones.

#### 4. CONVERGENCIA A LA DISTRIBUCION NORMAL.

En este punto se concentran los resultados relativos a la convergencia a la distribución normal, caso particular de las leyes estables y de las infinitamente divisibles, con lo que se destaca su importancia y las condiciones en que esta convergencia se efectúa.

Desde el primer capítulo son conocidas algunas condiciones requeridas para que se cumpla dicha convergencia, pero aquí serán relajadas hasta alcanzar las condiciones mínimas en que se sigue cumpliendo para el caso de variables aleatorias independientes.

#### 4.1 La distribución Normal como Ley Estable.

En efecto, la distribución normal pertenece a la familia de distribuciones estables, como a continuación se mostrará; y por eso solo hecho a ella convergen sumas estandarizadas de variables aleatorias independientes idénticamente distribuidas.

Puede verse directamente por definición que la distribución normal es estable, pues la convolución de variables aleatorias normales es normal, o bien, por la expresión de su función característica

$$\phi(t) = \exp [ i\mu t - \sigma^2 t^2 / 2 ]$$

que se ajusta a la forma general de las funciones características de leyes estables tomando  $\alpha = 2$ ,  $\beta = 0$ ,  $\gamma = \mu$  y  $c = \sigma^2/2$  (lo cual refleja que el orden de dispersión es igual a 2, es insesgada, está localizada alrededor de  $\mu$  y su parámetro de escala es igual a  $\sigma^2/2$ ).

Más aún, de la forma general, puede observarse que si  $\alpha = 2$ ,  $\beta$  no aparece en la expresión pues el término donde aparece se anula por el valor que toma  $\omega(t, \alpha)$ . Esto significa que es la distribución normal la única ley estable con índice de estabilidad igual a 2.

Con apoyo de resultados alcanzados en el apartado dedicado a las distribuciones estables, se obtienen las siguientes conclusiones: la distribución normal es unimodal, continua y tiene derivadas de todos los ordenes en cada punto y el dominio de atracción de la distribución normal es no vacío. Por ser de particular interés, se anotan los requisitos que debe cumplir una variable aleatoria para ser atraída por la distribución normal, establecidos por A. Ya. Khintchine (1935), W. Feller (1937) y P. Lévy (1925).

Teorema 2.10. Una condición necesaria y suficiente para que la función de distribución  $F(x)$  pertenezca al dominio de atracción de la distribución normal (no degenerada) es que

$$\lim_{x \rightarrow \infty} \frac{x^2 \int_{|t| > x} dF(t)}{\int_{|t| < x} t^2 dF(t)} = 0$$

Que este límite sea igual a cero significa que el "peso" de segundo orden que tienen las colas comparado con la parte central es despreciable; esto es, para que una distribución sea atraída por la normal, es necesario y suficiente que no tenga "colas pesadas". Y si adicionalmente se toma en consideración que toda ley estable pertenece a su propio dominio de atracción, y en especial la distribución normal, se concluye que de las leyes estables, solo ésta tiene "colas no pesadas".

Del siguiente resultado se desprende que para el caso de

variables aleatorias independientes idénticamente distribuidas, no sólo es suficiente — como se vió en el teorema central del Capítulo I — sino necesario, que la distribución tenga varianza finita para que su suma estandarizada converja a la normal.

Teorema 2.11. Una condición necesaria y suficiente para que la función de distribución  $F(x)$  pertenezca al dominio de atracción estándar de la distribución normal (no degenerada) es que tenga varianza finita  $\sigma^2$ . En ese caso, en la expresión (1) se escogen

$$a = \sigma \quad \text{y} \quad A_n = n \int_{-\infty}^{\infty} x \, dF(x). \quad \triangleright$$

Se puede dar otra particularidad de la distribución normal entre las leyes estables, debida a la existencia de momentos iguales al exponente característico  $\alpha$ : de las leyes estables, solo la distribución normal tiene varianza finita, y más aún, es la única cuyos momentos de cualquier orden son finitos.

Pasando al entorno de las leyes infinitamente divisibles, se puede mostrar que la distribución normal con media  $\mu$  y varianza  $\sigma^2$  pertenece a esta familia, pues es posible expresar su función característica como la  $n$ ésima potencia de la función característica de una distribución también normal con media  $\mu/n$  y varianza  $\sigma^2/n$ :

$$\phi(t) = \exp [ i\mu t - \sigma^2 t^2 / 2 ] = \exp [ i\mu t/n - \sigma^2 t^2 / 2n ]^n$$

que asimismo refleja el hecho de que la distribución normal es reproductiva respecto a la suma.

Además, es posible expresar esta función característica en las formas de Kolmogorov, Lévy-Kolmogorov y de Lévy considerando:

$$\gamma = \mu \quad K(u) = \begin{cases} 0 & \text{si } u \leq 0 \\ \sigma^2 & \text{si } u > 0, \end{cases}$$

$$\gamma = \mu \quad B(u) = \begin{cases} 0 & \text{si } u \leq 0 \\ \sigma^2 & \text{si } u > 0, \end{cases}$$

y

$$\gamma = \mu \quad M(u) = 0 \quad N(u) = 0 \quad \sigma = \sigma$$

respectivamente.

Conviene señalar que por lo tanto cualquier potencia positiva de la función característica,  $[\phi(t)]^c$  también es una función característica, y corresponde a la de otra distribución normal, con la media recorrida  $c$  veces y la varianza expandida (o contraída si  $c < 1$ ) por el mismo factor; lo que viene a significar una generalización de la reproductividad de la suma al no considerar un número entero de sumandos, sino cualquier fracción  $c$ .

También se puede afirmar (H. Cramér 1936) que si la convolución de dos leyes infinitamente divisibles es una distribución normal, entonces las dos son distribuciones normales, lo cual resulta ser una manera exclusiva de la distribución normal, junto con la de Poisson, de caracterizarlas.

#### 4.2 La distribución Normal como Ley Infinitamente Divisible.

Continuaremos con el establecimiento de algunas condiciones requeridas para la convergencia a la distribución normal, dentro del ámbito de las leyes infinitamente divisibles con otro resultado de A. Ya. Khintchine (1938).

Teorema 2.12. Sean  $X_{n1}, \dots, X_{nk}$   $n = 1, 2, \dots$  sucesiones de variables aleatorias infinitesimales e independientes en cada renglón, y sea

$$X_n = \sum_{i=1}^k X_{ni}$$

Si la sucesión  $\{X_n\}$  converge a algún límite, entonces la relación

$$\lim_{n \rightarrow \infty} \sum_{i=1}^k \int_{|x| \geq c} dF_{ni}(x) = 0$$

se cumple  $\forall c > 0$  si y sólo si el límite de la sucesión es la distribución normal.  $\blacktriangleright$

Esta condición establecida por A. Ya. Khintchine establece que sólo que se trate de la convergencia a la distribución normal, la suma de las probabilidades con que las componentes infinitesimales toman valores "grandes" en valor absoluto es despreciable en el límite.

Teorema 2.13. Sean  $X_{n1}, \dots, X_{nk}$   $n = 1, 2, \dots$  variables aleatorias independientes y sea

$$X_n = \sum_{i=1}^k X_{ni} - A_n$$

con  $A_n$  constantes elegidas adecuadamente. La sucesión  $\{X_n\}$  converge a la distribución normal con media cero y varianza unitaria, y los componentes  $X_{ni}$   $i = 1, \dots, k$  son infinitesimales si y sólo si  $\forall c > 0$  se satisfacen las siguientes condiciones:



$$(i) \quad \lim_{n \rightarrow \infty} \sum_{l=1}^k \int_{|x| \geq c} dF_{nl}(x) = 0$$

$$(ii) \quad \lim_{n \rightarrow \infty} \sum_{l=1}^k \left\{ \int_{|x| < c} x^2 dF_{nl}(x) - \left[ \int_{|x| < c} x dF_{nl}(x) \right]^2 \right\} = 1$$

Teorema 2.14. Sean  $X_{n1}, \dots, X_{nk}$ ,  $n = 1, 2, \dots$  variables aleatorias infinitesimales independientes y sea

$$X_n = \sum_{l=1}^k X_{nl} - A_n$$

con  $A_n$  una constante elegida adecuadamente. La sucesión  $(X_n)$  converge a la distribución normal si y sólo si  $\forall c > 0$  se satisfacen las siguientes condiciones:

$$(i) \quad \lim_{n \rightarrow \infty} \sum_{l=1}^k \int_{|x| > c} dF_{nl}(x + \alpha_{nl}) = 0$$

$$(ii) \quad \lim_{n \rightarrow \infty} \sum_{l=1}^k \int_{|x| < c} x^2 dF_{nl}(x + \alpha_{nl}) = 1$$

donde

$$\alpha_{nl} = \int_{|x| < \tau} x dF_{nl}(x) \quad \text{y } \tau > 0.$$

Teorema 2.15. (S. N. Bernstein 1926 y W. Feller 1935). Sean  $X_1, \dots, X_n$  variables aleatorias independientes. La sucesión

$$\frac{1}{B_n} \sum_{l=1}^n X_l - A_n$$

con las constantes  $A_n$  y  $B_n > 0$  elegidas adecuadamente, converge a la distribución normal, y los componentes  $X_{ni} = X_i/B_n$ ,  $i = 1, \dots, n$  son infinitesimales si y sólo si existe una sucesión de constantes  $C_n$  ( $C_n \rightarrow \infty$ ) tal que se satisfacen las siguientes condiciones:

$$(i) \quad \lim_{n \rightarrow \infty} \sum_{l=1}^n \int_{|x| > C_n} dF_l(x) = 0$$

$$(i) \quad \lim_{n \rightarrow \infty} \frac{1}{C_n^2} \sum_{i=1}^n \left\{ \int_{|x| < C_n} x^2 dF_i(x) - \left[ \int_{|x| < C_n} x dF_i(x) \right]^2 \right\} = 1$$

Finalmente, se presenta el famoso teorema de Lindeberg-Feller que marca el punto final en las condiciones necesarias (establecidas por W. Feller, 1935) y suficientes (demostradas por Y. W. Lindeberg, 1922) para determinar la convergencia de sumas de variables aleatorias independientes a la distribución normal, en la que destaca nuevamente la importancia de que el peso de las "colas" de las distribuciones  $F_k(x)$  sea despreciable respecto a la dispersión total.

Teorema 2.16. (Y. W. Lindeberg, 1922 y W. Feller). Sea  $(X_k)$  una sucesión de variables aleatorias independientes con función de distribución  $F_k(x)$  y  $E(X_k) = 0$ . La sucesión normalizada

$$\frac{1}{B_n} \left[ \sum_{i=1}^n X_i \right]$$

con  $B_n^2 = \sum_{k=1}^n \text{Var}(X_k)$ , converge a la distribución normal y los sumandos son infinitesimales si y sólo si

$$\lim_{n \rightarrow \infty} \frac{1}{B_n^2} \sum_{k=1}^n \int_{|x| > c B_n} x^2 dF_k(x) = 0 \quad \forall c > 0. \quad \triangleright$$

## 5. BIBLIOGRAFIA COMENTADA.

- B. V. Gnedenko y A. N. Kolmogorov, 1954. Es un libro fundamental para el tema, circunscrito al caso de variables aleatorias independientes. Da cuenta de los resultados alcanzados hasta la fecha de publicación, con un tratamiento formal y ordenado, por lo que se logra dar una visión general y a la vez detallada. Parte de lo general a lo particular, presentando primero en el capítulo 3 la convergencia de sumas de variables aleatorias independientes a las distribuciones infinitamente divisibles, centrándose posteriormente, en el capítulo 5 en la convergencia a la distribución normal y a la Poisson. En el capítulo 7 estudia los sumandos idénticamente distribuidos y la convergencia a las leyes estables. Incluye la totalidad de las demostraciones enunciadas en el capítulo, por lo que se remite al lector interesado en profundizar en el tema a este texto.
- V. M. Zolotarev, 1986. Es un trabajo dedicado por completo al estudio de las leyes estables, donde aborda con detalle y de excelente forma las aplicaciones y propiedades de estas

distribuciones, así como problemas de estimación de sus parámetros. En particular, las aplicaciones aquí presentadas pueden consultarse con mayor detalle en este texto, al igual que otras en distintas áreas.

- A. N. Shiryayev, 1984. Inserto en un contexto de probabilidad más amplio, aborda mediante un enfoque formal los resultados más importantes y de carácter general sobre convergencia de sumas de variables aleatorias, aunque sólo incluye las demostraciones de los teoremas 2.5 y 2.16.
- B. Harris, 1966. Contiene un capítulo bien desarrollado acerca de la convergencia a la distribución normal para sumas de variables aleatorias independientes, sin ocuparse de las generalizaciones dentro de las leyes infinitamente divisibles o las estables.
- B. V. Gnedenko, 1969. En su capítulo 2 da una introducción al problema de convergencia particularizando al caso de variables aleatorias Bernoulli. El capítulo 8 lo dedica al estudio del Teorema Central del Límite Clásico, y el capítulo 9 a las Leyes Infinitamente Divisibles. Tiene buenos comentarios, un enfoque intuitivo y presenta los principales resultados, demostrando solamente los teoremas 2.5 y 2.6 (aparte de otros más generales aquí no tratados).

## CAPITULO III.

### CONVERGENCIA DE SUMAS DE VARIABLES ALEATORIAS DEPENDIENTES: CONDICIONES MEZCLANTES Y MARTINGALAS.

#### 1. INTRODUCCION.

En los capítulos precedentes se ha avanzado en el estudio del Teorema Central del Límite hasta conservar únicamente la característica de independencia entre los elementos de las sucesiones consideradas. Entonces resulta razonable cuestionarse si aún para situaciones de dependencia en la sucesión, existen condiciones en que se verifica la convergencia de sumas estandarizadas de variables aleatorias a la distribución normal. Y así es en efecto, pues en trabajos más recientes que los antes establecidos, se ha llegado a tal conclusión. Por lo tanto, es la presentación y discusión de estos resultados el objeto del presente capítulo. La atención se centrará en el estudio de las condiciones mezclantes y martingalas, por ser éstas las de mayor desarrollo y difusión en el tratamiento de sucesiones dependientes. Las condiciones mezclantes representan una manera directa de definir el grado de dependencia y su conceptualización ocurre en la década de los 50's de nuestro siglo, manteniéndose como un área de investigación activa hasta nuestros días, encontrando aplicación en las sucesiones  $m$ -dependientes, algunas cadenas de Markov y procesos gaussianos, entre otros. El otro enfoque, el de martingalas, si bien data de épocas anteriores con los trabajos de Bernstein y Lévy, su definición formal dada por Ville y los avances logrados por Dobb acaecidos entre 1927 y 1953, ha tomado un nuevo impulso recientemente generalizando el tratamiento y mejorando resultados conseguidos con anterioridad, en procesos estacionarios por ejemplo, y encontrando una amplia aplicación en campos tales como procesos de ramificación, modelos ARMA, etc.

Como ilustración, abordemos una situación de dependencia de las más simples y conocidas, la de una cadena de Markov de dos estados, en que para no estropear la presentación dejaremos de lado la definición de los conceptos empleados, remitiendo al lector que así lo requiera a la siguiente sección en la cual se detallan algunos ejemplos. En contrapartida, se incluye la demostración de este resultado con el objeto de mostrar el tipo de razonamiento que se sigue para su solución, primordialmente en el manejo que hace para tratar la dependencia entre las variables.

Teorema 3.1. Considere que la sucesión  $X_1, X_2, \dots$  forma una cadena de Markov homogénea con dos estados. Sea  $X_n = 0$ ,  $X_n = 1$  según el sistema se encuentre en el instante  $n$  en el estado  $A_0$  o en el estado  $A_1$ . Sea la matriz de transición

$$(p_{ij}) = \begin{bmatrix} 1 - \lambda & \lambda \\ \mu & 1 - \mu \end{bmatrix}$$

con  $0 < \lambda < 1$  y  $0 < \mu < 1$ . Si denotamos  $\sum X_i$  como  $S_n$ , entonces

$$P \left[ \frac{S_n - \frac{\lambda}{\lambda + \mu} n}{\sqrt{\frac{n\lambda\mu(2-\lambda-\mu)}{(\lambda+\mu)^3}}} \leq x \right] \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^x e^{-y^2/2} dy \quad \text{si } n \rightarrow \infty$$

Demostración. Construyamos la sucesión de variables aleatorias  $(T_n)$ ,  $n = 1, 2, \dots$  que representa el instante en que el sistema regresa por  $n$ -ésima ocasión al estado  $A_1$ . Entonces la sucesión es creciente:  $T_1 < T_2 < \dots < T_n < \dots$  y

$$X_{T_n} = \begin{cases} 1 & \text{si } k = T_n, & n \in \mathbb{N} \\ 0 & \text{si } k < T_1 \text{ o } T_n < k < T_{n+1}, & n \in \mathbb{N} \end{cases}$$

Además, las variables

$$D_n = \begin{cases} T_1 & \text{si } n = 1 \\ T_n - T_{n-1} & \text{si } n > 1 \end{cases}$$

representan los lapsos entre ocurrencias sucesivas, y puesto que  $D_n$  no depende de los estados del sistema en los instantes  $t < T_{n-1}$  (por las propiedades de la cadena de Markov), se trata de una sucesión de variables aleatorias independientes. Más aún, la homogeneidad de la cadena de Markov asegura que las variables  $D_n$ , excepto quizá  $D_1$ , comparten la misma distribución y para  $n > 1$

$$P(D_n = 1) = 1 - \mu$$

y

$$P(D_n = k) = \mu\lambda(1-\lambda)^{k-2} \quad k \geq 2$$

con lo que calculamos su media y varianza:

$$E(D_n) = \frac{\lambda + \mu}{\lambda}$$

y

$$\text{Var}(D_n) = \frac{\mu(2-\lambda-\mu)}{\lambda^2}$$

$D_1$  tiene la misma distribución si  $X_0 = 1$  (pues entonces comienza ya igual que las otras), y si  $X_0 = 0$ , entonces  $P(D_1 = 1) = \lambda$ ,  $P(D_1 = k) = \lambda(1-\lambda)^{k-1}$  para  $k > 1$ , de donde

$$E(D_1) = \frac{1}{\lambda} \quad \text{y} \quad \text{Var}(D_1) = \frac{1-\lambda}{\lambda^2}$$

La sucesión  $\{D_n\}$ , relacionada con  $\{S_n\}$  a través de las  $T_n$ , es de gran utilidad por estar compuesta de variables aleatorias independientes e idénticamente distribuidas a las que se puede aplicar el Teorema Central del Límite en la versión del teorema 1.3, cuando se cuenta con varianzas finitas.

En efecto, se cumple que  $T_k = \sum_{i=1}^k D_i$ , el tiempo transcurrido para regresar por  $n$ -ésima vez a  $A_1$  es igual a la suma de los tiempos transcurridos entre las anteriores ocurrencias. Y además, se tiene que

$$P(S_n < k) = P(T_k > n)$$

pues el evento de ocurrir el evento  $A_1$  menos de  $k$  veces en  $n$  periodos es equivalente a que el tiempo de ocurrencia del  $k$ -ésimo retorno al mismo estado sea mayor que los  $n$  periodos. Esta desigualdad resulta determinante en la demostración.

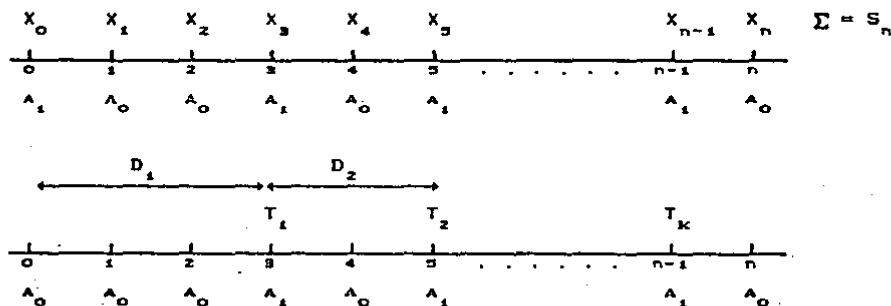


Figura 3.1  
Variables  $X_n$ ,  $S_n$ ,  $T_k$  y  $D_k$ .

Entonces, al estandarizar  $S_n$  y relacionarla con  $T_n$  tenemos

$$P \left[ \frac{S_n - \frac{\lambda}{\lambda + \mu} n}{\sqrt{\frac{n\lambda\mu(2-\lambda-\mu)}{(\lambda+\mu)^2}}} < x \right] = P \left[ S_n < x \sqrt{\frac{n\lambda\mu(2-\lambda-\mu)}{(\lambda+\mu)^2}} + \frac{\lambda}{\lambda + \mu} n \right]$$

$$= P \left[ T_k > -x \frac{\sqrt{k\mu(2-\lambda-\mu)}}{\lambda} + \frac{\lambda + \mu}{\lambda} k + O(1) \right]$$

en donde se ha considerado que  $n - k$  está acotado en un intervalo finito al despejar  $n$ , dando lugar adicionalmente al residuo  $O(1)$ . Entonces es equivalente a

$$P \left[ \frac{T_n - \frac{\lambda + \mu}{\lambda} k}{\frac{\sqrt{k\mu(2-\lambda-\mu)}}{\lambda}} > -x + O(k^{-1/2}) \right]$$

por lo que tomando el límite cuando  $n$  tiende a infinito se tiene

$$\lim_{n \rightarrow \infty} P \left[ \frac{S_n - \frac{\lambda + \mu}{\lambda} n}{\frac{\sqrt{n\lambda\mu(2-\lambda-\mu)}}{(\lambda + \mu)^2}} < x \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$$

(si  $X_0 = 0$ , el teorema 1.3 no es estrictamente aplicable, pero por tener varianza finita la conclusión sigue siendo válida con ligeras modificaciones en la estandarización, como lo afirma el teorema 2.16 de Lindeberg-Feller).

Lo ilustrativo de la demostración consiste en que se ha construido una sucesión de variables aleatorias independientes relacionadas con la sucesión original de variables dependientes, para así aplicar el Teorema Central del Límite con resultados ya conocidos y trasladarlos al caso que nos ocupa. Esta técnica es utilizada con frecuencia en demostraciones de la convergencia a la distribución normal para variables aleatorias dependientes.

Este resultado da una prueba de existencia, de que en efecto el Teorema Central del Límite es todavía válido para situaciones de dependencia. Mas aún, y antes de continuar, nótese que si  $\lambda + \mu = 1$ , la cadena se reduce a una sucesión de variables aleatorias independientes pues  $p_{i1} = p_{i2}$  para  $i = 1, 2$  que se distribuyen como una bernoulli de parámetro  $\lambda$  debido a que  $p_{12} = p_{22} = \lambda$ , y puesto que la estandarización de la expresión anterior se reduce a

$$\frac{S_n - \lambda n}{\sqrt{n\lambda(1-\lambda)}}$$

podemos concluir de manera interesante que el resultado expuesto para cadenas de Markov de dos estados es una generalización del teorema de deMoivre-Laplace en el caso de dependencia entre las variables aleatorias.

Satisfecha la inquietud, nos preocuparemos por determinar las condiciones requeridas para que sucesiones de variables aleatorias dependientes converjan a la distribución normal, utilizando para ello tratamientos más generales en lugar de buscar resultados a casos particulares. Los enfoques que se abordarán son los de condiciones mezclantes y de martingalas, a los cuales se dedican las secciones 2 y 3, respectivamente.

En todo el capítulo, como es común suponer, se considerarán condiciones mínimas que den sentido al problema abordado, y se refieren a tratar con sucesiones estacionarias definidas por

Definición 3.1. Una sucesión  $\{X_t\}$  de variables aleatorias es estrictamente estacionaria si la distribución del vector

$$(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_s+h})$$

no depende de  $h$ . Y es estacionaria en sentido débil si

$$E[X_t^2] < \infty \quad \forall t$$

y

$$E[X_s] \quad \text{y} \quad E[X_s X_{s+t}] \quad \text{no dependen de } s.$$

Sin pérdida de generalidad, supondremos que  $E[X_s] = 0$ , salvo que se indique lo contrario.

## 2. CONDICIONES MEZCLANTES.

A continuación se definirán distintas condiciones mezclantes, se darán propiedades y relaciones entre ellas para después presentar varios ejemplos y enunciar algunos resultados existentes acerca del cumplimiento del Teorema Central del Límite en este tipo de sucesiones.

### 2.1 Aplicaciones y definición.

Para abordar el problema de convergencia a la distribución normal de expresiones del tipo

$$\frac{1}{A_n} \sum_{k=1}^n X_k - B_n$$

donde la sucesión  $\{X_k\}$  de variables aleatorias no son ya independientes, se requirió en principio, como en otros contextos, establecer un mínimo de condiciones que dieran sentido al estudio, porque si se planteara un marco muy general, tal como la sola



suposición de estacionariedad, podría establecerse como límite cualquier distribución, únicamente con tomar la misma variable para formar la sucesión. Sin embargo, en casos como ese, sucede que la dependencia entre dos miembros de la sucesión puede ser muy alta, por mucho que estén separados en el tiempo esos elementos en la sucesión. Para obviar esa situación, cabe imponer patrones de comportamiento al grado de dependencia, y uno natural sería exigir que disminuyera a medida que se escogieran dos elementos cualesquiera, pero que se encontraran en el tiempo cada vez más separados entre sí.

De esa manera surgen las sucesiones mezclantes para las cuales se establecen condiciones en que el pasado remoto y el futuro distante son asintóticamente independientes, dando lugar a distintos tipos mezclantes, de acuerdo a la manera en que se mida la relación entre las variables de la sucesión.

Un área de gran aplicación para las condiciones mezclantes ha sido la Econometría, donde se ha demostrado que son las condiciones adecuadas para resolver problemas de estimación y de propiedades asintóticas en cierto tipo de modelos. Específicamente, en White y Domowitz (1981) se trata con modelos no lineales con observaciones dependientes, y los mismos autores, en Domowitz y White (1982) abordan el caso de modelos no lineales con problemas de especificación en presencia de heteroscedasticidad y dependencia en las observaciones; y el tratamiento de modelos dinámicos con condiciones mezclantes es introducido por Gallant (1986). Para estos casos los datos pueden provenir de series de tiempo (modelos ARIMA no lineales, por ejemplo), cross-section o en panel, y en todos ellos juegan un papel importante las condiciones uniforme y fuertemente mezclantes -que se verán enseguida- para establecer propiedades asintóticas de los estimadores.

Se establecen, entonces las definiciones básicas para el enfoque de las condiciones mezclantes.

Sea  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad formado con el conjunto  $\Omega$ , la  $\sigma$ -álgebra  $\mathcal{F}$  contenida en  $\Omega$  y  $P$  una medida de probabilidad definida en  $\mathcal{F}$ , y sean  $A$  y  $B$  dos  $\sigma$ -álgebras contenidas en  $\mathcal{F}$  para las que se definen las siguientes medidas de dependencia:

$$\alpha(A, B) = \sup |P(A \cap B) - P(A)P(B)|, \quad A \in \mathcal{A}, \quad B \in \mathcal{B}$$

$$\phi(A, B) = \sup |P(B|A) - P(B)| = \sup \frac{|P(A \cap B) - P(A)P(B)|}{P(A)},$$

$$A \in \mathcal{A}, \quad B \in \mathcal{B}, \quad P(A) > 0$$

$$\phi_{rev}(A, B) = \phi(B, A)$$

$$\psi(A, B) = \sup \frac{|P(A \cap B) - P(A)P(B)|}{P(A)P(B)}, \quad A \in \mathcal{A}, \quad B \in \mathcal{B}, \quad P(A) > 0, \quad P(B) > 0$$

$$\rho(A, B) = \sup |\text{Corr}(X, Y)|, \quad X \in L_2(\mathcal{A}), \quad Y \in L_2(\mathcal{B});$$

X, Y variables aleatorias reales  
(ver definición de  $L_2(\mathcal{B})$  en A3.1)

$$\beta(A, B) = \sup \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J |P(A_i \cap B_j) - P(A_i)P(B_j)|,$$

donde el supremo se toma sobre todos los pares de particiones  $\{A_1, A_2, \dots, A_I\}$  y  $\{B_1, B_2, \dots, B_J\}$  de  $\Omega$  tales que  $A_i \in \mathcal{A} \forall i$ , y  $B_j \in \mathcal{B} \forall j$ .

Como se puede observar, dadas dos  $\sigma$ -álgebras contenidas en un mismo espacio, puede medirse su grado de asociación de distintas maneras. Las dos primeras y más simples,  $\alpha(A, B)$  y  $\phi(A, B)$  lo hacen evaluando la discrepancia entre las probabilidades de la intersección de dos conjuntos cualesquiera, uno de cada  $\sigma$ -álgebra, considerando y no el supuesto de independencia, para luego tomar el supremo sobre todos los conjuntos. Para el caso de  $\phi(A, B)$ , puesto que la condicional se fija sobre una de las dos  $\sigma$ -álgebras, no es simétrica y tiene sentido definir  $\phi_{\text{ov}}(A, B)$ , que junto con  $\psi(A, B)$  representan estandarizaciones de  $\alpha(A, B)$ , mientras que  $\rho(A, B)$  es un caso particular cuando las variables aleatorias en las  $\sigma$ -álgebras son reales y tienen segundo momento finito, midiendo, como es sabido, el grado de dependencia lineal. Para  $\beta(A, B)$  la característica estriba en que se toman todas las particiones posibles de los conjuntos A y B para calcular la medida de no independencia.

Con estas definiciones estableceremos cinco condiciones mezclantes fuertes:

Sea  $\{X_k\}$ ,  $k \in \mathbb{Z}$  una sucesión de variables aleatorias,  $\mathcal{F}_J^L = \sigma(X_k, J \leq k \leq L)$ , la sigma-álgebra generada, con  $-\infty \leq J \leq L \leq \infty$ . Para  $n \in \mathbb{N}$  sean

$$\alpha(n) = \sup_{j \in \mathbb{Z}} \alpha(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^{\infty})$$

$$\phi(n) = \sup_{j \in \mathbb{Z}} \phi(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^{\infty})$$

$$\psi(n) = \sup_{j \in \mathbb{Z}} \psi(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^{\infty})$$

$$\rho(n) = \sup_{j \in \mathbb{Z}} \rho(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^{\infty})$$

$$\beta(n) = \sup_{j \in \mathbb{Z}} \beta(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^{\infty})$$

Bajo estas condiciones, la sucesión se denomina fuertemente mezclante (Rosenblatt, 1956) si  $\lim_{n \rightarrow \infty} \alpha(n) = 0$

$\phi$ -mezclante o uniformemente mezclante (Ibragimov, 1959) si

$$\lim_{n \rightarrow \infty} \phi(n) = 0$$

$\psi$ -mezclante (Blum, Hanson y Koopmans, 1963) si  $\lim_{n \rightarrow \infty} \psi(n) = 0$

$\rho$ -mezclante (Kolmogorov y Rosanov, 1960) si  $\lim_{n \rightarrow \infty} \rho(n) = 0$

regularidad absoluta (Volkonskii y Rozanov, 1959) si  $\lim_{n \rightarrow \infty} \beta(n) = 0$

La condición  $\psi$ -mezclante implica la conocida como  $\ast$ -mezclante:  $\lim_{n \rightarrow \infty} \sup_{j \in \mathbb{Z}} \psi(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^{\ast}) = 0$  introducida en Blum, Hanson y Koopmans, 1963. La condición uniformemente mezclante no es simétrica respecto al pasado y al futuro.

Entonces, las condiciones mezclantes resultan de aplicar las medidas de dependencia a una sucesión de variables aleatorias, considerando las  $\sigma$ -álgebras generadas por los elementos de la sucesión que dejen en medio  $n-1$  variables aleatorias. Para su interpretación es útil concebir las  $\sigma$ -álgebras como conjuntos de información respecto a un proceso en que se consideran dos momentos separados entre sí  $n$  periodos de tiempo; entonces se toma la mayor medida de las obtenidas al recorrer todo el proceso dejando siempre los  $n$  periodos separando los conjuntos de información; finalmente, si al aumentar al infinito la separación entre los periodos de observación, resulta que esa medida de dependencia se desvaneca, se establece la condición mezclante respectiva y puede considerarse que el pasado remoto y el futuro lejano del proceso son asintóticamente independientes entre sí.

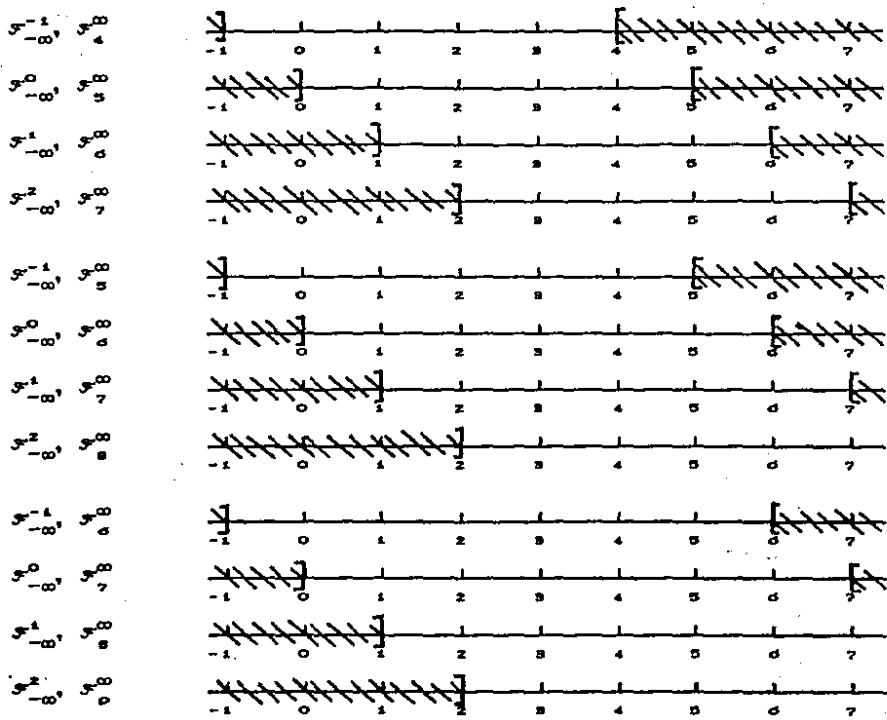


Figura 3.2  
 Recorrido para las condiciones mezclantes.

## 2.2 Relación entre las condiciones mezclantes.

Entre estas medidas de dependencia se cumplen las siguientes implicaciones:

$\psi$ -mezclante  $\Rightarrow$

$\phi$ -mezclante  $\Rightarrow$

$$\left\{ \begin{array}{l} \rho\text{-mezclante} \\ \text{regularidad absoluta} \end{array} \right\} \Rightarrow$$

fuertemente mezclante

Para extender estas condiciones a sucesiones estrictamente estacionarias  $\{X_k\}$ ,  $k \in \mathbb{Z}$  se redefine, por ejemplo

$$\alpha(n) = \alpha(\mathcal{F}_{-\infty}^0, \mathcal{F}_n^\infty)$$

y de manera análoga las demás medidas de dependencia.

Conviene incluir en nuestro estudio a las sucesiones que cumplen otras características de dependencia, como es la de regularidad, cuya definición para sucesiones estacionarias es la siguiente:

Definición 3.2. Se dice que una sucesión  $\{X_k\}$  es regular si la  $\sigma$ -álgebra generada con la "cola" izquierda

$$\bigcap_{j=1}^{\infty} \mathcal{F}_{-j}^\infty$$

es trivial en el sentido de que contiene sólo eventos con probabilidad cero o uno.  $\blacktriangleright$

Se puede mostrar (Ibragimov-Linnik, 1971) que se cumple la siguiente implicación:

fuertemente mezclante  $\Rightarrow$  regular

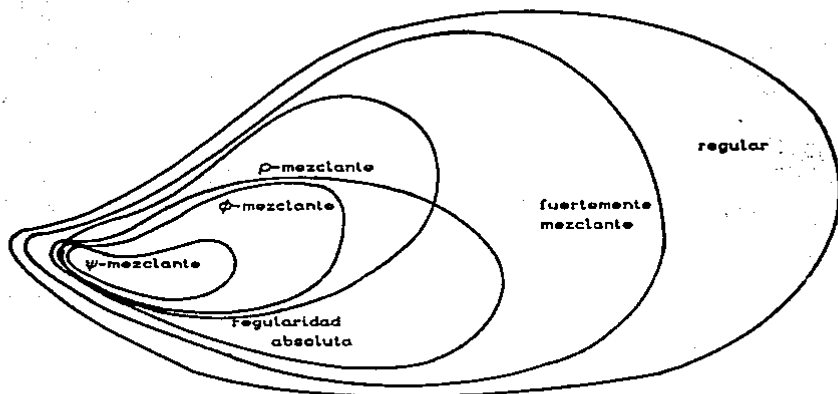


Figura 3.3  
Relación entre condiciones mezclantes.

### 2.3 Ejemplos de sucesiones mezclantes.

A continuación, se mencionarán algunos ejemplos en que se verifican una ó varias de estas condiciones.

Son satisfechas las cinco condiciones mezclantes por sucesiones de variables aleatorias independientes, pues los coeficientes mezclantes son iguales a cero para toda  $n \geq 1$ .

También cumplen todas las condiciones mezclantes las llamadas sucesiones  $m$ -dependientes  $(X_k)$ ,  $k = \dots, -2, -1, 0, 1, 2, \dots$  que para  $(X_{a-p}, X_{a-p+1}, \dots, X_a)$  y  $(X_b, X_{b+1}, \dots, X_{b+q})$ , con  $b-a > m$ ,  $p$  y  $q \in \mathbb{N}^*$  son independientes (cualquier par de vectores aleatorios de longitud arbitraria -pero fija- y separados entre sí por más de  $m$  elementos son independientes). La razón es que todos los coeficientes mezclantes se anulan para  $n > m$ .

Otro ejemplo lo constituyen las sucesiones Gaussianas que son aquellas en las que para cualquier  $t_1, t_2, \dots, t_n$  la función característica del vector  $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$  tiene la forma

$$\phi(\theta_{t_1}, \theta_{t_2}, \dots, \theta_{t_n}) = \exp \left[ i \sum_{j=1}^n a_{t_j} \theta_{t_j} - \frac{1}{2} \sum_{k=1}^n \sum_{j=1}^n R_{t_k t_j} \theta_{t_k} \theta_{t_j} \right]$$

donde  $a_{t_j}$  son números reales cualesquiera y la matriz  $R = R_{t_k t_j}$  es

semipositiva definida, siendo posible mostrar que  $a_{t_j} = E [X_{t_j}]$ ,  $R_{t_k t_j} = E [(X_{t_k} - a_{t_k}) (X_{t_j} - a_{t_j})]$  y que si  $R$  es no singular entonces  $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$  tiene una distribución conjunta normal  $n$ -variante.

Debido a que variables aleatorias normalmente distribuidas son independientes si son ortogonales, entonces una sucesión gaussiana es  $m$ -dependiente si y sólo si su función de autocovarianza  $R_{t_k t_j} = 0$  para  $|t_k - t_j| > m$ , y por lo tanto cumple todas las condiciones mezclantes.

En general, se cuenta con el siguiente resultado debido a I.A. Ibragimov y Y.A. Rozanov (1970 y 1978) en que se establecen las condiciones mezclantes satisfechas por una sucesión gaussiana de acuerdo a la forma que adopte la función de densidad espectral (ver A3.2):

**Teorema 3.2.** Sea  $\{X_k\}$ ,  $k \in \mathbb{Z}$  una sucesión Gaussiana estacionaria de variable real. Entonces se verifican los siguientes cuatro enunciados:

a) Las dos siguientes condiciones son equivalentes:

- i)  $\{X_k\}$  es regular.
- ii)  $\{X_k\}$  tiene una función de distribución espectral absolutamente continua y su densidad espectral  $f$  (definida en  $[-\pi, \pi]$ ) cumple con

$$\int_{-\pi}^{\pi} \ln f(\lambda) d\lambda > -\infty$$

b) Las siguientes tres condiciones son equivalentes:

- i)  $\{X_k\}$  es fuertemente mezclante.
- ii)  $\{X_k\}$  es  $\rho$ -mezclante.
- iii) La función espectral  $f$  de  $\{X_k\}$  se puede expresar en la forma

$$f(\lambda) = |P(e^{i\lambda})|^2 \exp \{ u(e^{i\lambda}) + \bar{z}(e^{i\lambda}) \}$$

donde  $P$  es un polinomio,  $u$  y  $a$  son funciones reales continuas definidas en el círculo unitario complejo y  $\bar{z}$  es el conjugado de  $a$ .

c) Son equivalentes las siguientes dos condiciones:

- i)  $\{X_k\}$  es absolutamente regular.
- ii) La función espectral  $f$  de  $\{X_k\}$  se puede expresar en la forma

$$f(\lambda) = |P(e^{i\lambda})|^2 \exp \left( \sum_{j=-\infty}^{\infty} a_j e^{ij\lambda} \right)$$

donde  $P$  es un polinomio cuyas raíces, si existen, caen en el círculo unitario y  $\sum_{j=-\infty}^{\infty} |j| |a_j|^2 < \infty$ .

d) Son equivalentes las siguientes cuatro condiciones:

- i)  $\{X_k\}$  es  $\phi$ -mezclante.
- ii)  $\{X_k\}$  es  $\psi$ -mezclante.
- iii)  $\{X_k\}$  es  $m$ -dependiente.
- iv) La función espectral  $f$  de  $\{X_k\}$  se puede expresar en la forma

$$f(\lambda) = |P(e^{i\lambda})|^2$$

donde  $P$  es un polinomio.

Consideremos ahora las Cadenas de Markov. Sean  $A_1, A_2, \dots, A_k, \dots$  los posibles resultados o estados que pueden ocurrir en los momentos  $t_1, t_2, \dots, t_r, \dots$  para una sucesión de variables aleatorias  $\{X_t\}$  que forman un sistema. Si la probabilidad de que el sistema esté en el estado  $A_k$  en el tiempo  $t_n$  depende solamente del estado en que se encontraba en el tiempo  $t_{n-1}$  (propiedad markoviana), se trata de una cadena de Markov y es finita o infinita de acuerdo a la cardinalidad de los estados. Definamos  $p_{ij} = P(X_{t_n} = A_j | X_{t_{n-1}} = A_i)$  como la probabilidad de transición a un paso del estado  $A_i$  al estado  $A_j$ , y si no depende del tiempo  $t_n$  que se trate, la cadena es homogénea. La matriz  $P = (p_{ij})$  es la matriz de transición a un paso y  $P^n = (p_{ij}^{(n)})$  resulta ser la matriz de transición a  $n$  pasos. Esa matriz junto con las probabilidades  $\pi_i = P(X_{t_0} = A_i)$  definen por completo a la cadena de Markov. Además, si  $p_{ij}^{(n)} > 0 \forall i, j$  y alguna  $n \in \mathbb{N}$  (el sistema puede pasar de cualquier estado a otro en un número finito de pasos) la cadena es irreducible, y es aperiódica si  $p_{ii}^{(n)} > 0 \forall i, n$  (el sistema puede regresar a un estado dado en cualquier número de pasos).

Con estos conceptos podemos formular las siguientes condiciones mezclantes para cadenas de Markov: cualquier cadena de Markov finita, estacionaria, irreducible y aperiódica es  $\psi$ -mezclante con decaimiento exponencial de los coeficientes (esto es, para algún  $r > 0$ ,  $\psi(n) = O(e^{-rn})$  si  $n$  tiende a infinito —recordar la definición de  $O()$  en A1.14—), y por lo tanto también



es mezclante en cualquier otro sentido. El resultado presentado en la introducción del capítulo que generaliza el teorema de deMoivre-Laplace es un caso particular de este ejemplo por tratarse de una cadena de Markov de dos estados.

Si la cadena es estrictamente estacionaria con un conjunto de estados numerable, irreducible y aperiódica es absolutamente regular, lo cual es una consecuencia del teorema de convergencia de las matrices de transición  $P^n$ ; en este caso, el decaimiento exponencial de  $\alpha(n)$  y  $\beta(n)$  es menor que exponencial. Sin embargo, estas cadenas ya no son  $\rho$ -mezclantes.

Inversamente, existen cadenas que son  $\rho$ -mezclantes, pero no absolutamente regulares, como el ejemplo que a continuación se construirá. En principio, considérese la cadena de Markov estacionaria  $\{Y_k\}$  de dos estados  $(A_1, A_2) = \{0, 1\}$  con probabilidades  $\pi_i = 1/2$ ,  $i = 1, 2$  y matriz de transición

$$P = \begin{bmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{bmatrix}$$

de la que se puede mostrar por inducción que la matriz de transición a  $n$  pasos está dada por

$$P^n = \begin{bmatrix} (1+2^{-n})/2 & (1-2^{-n})/2 \\ (1-2^{-n})/2 & (1+2^{-n})/2 \end{bmatrix}$$

con la cual puede mostrarse que  $\rho(n) = 2^{-n}$ . Ahora, considérense  $\{Y_k^{(i)}\}$ ,  $i = 1, 2, \dots$  cadenas de Markov independientes con la misma distribución de la cadena  $\{Y_k\}$  y sea  $\{X_k\}$  la sucesión definida por

$$X_k = \sum_{l=1}^{\infty} 2^{-l} Y_k^{(l)}.$$

Excepto por conjuntos de medida cero, se cumple que  $\sigma(X_k) = \sigma(Y_k^{(i)})$   $i = 1, 2, \dots$  (lo que significa que a partir de  $X_k$  se pueden recuperar los valores de  $Y_k^{(i)}$   $\forall i$ ), de donde se sigue que  $X_k$  es una cadena de Markov, además es estrictamente estacionaria y  $\rho(n) = 2^{-n}$  y por lo tanto  $\rho$ -mezclante. Sin embargo  $\beta(n) = 1 \forall n \geq 1$  no siendo entonces absolutamente regular.

De estos dos últimos ejemplos se destaca que si bien la condición  $\rho$ -mezclante y la de regularidad absoluta son consecuencia de la  $\phi$ -mezclante y a su vez implican la condición fuertemente mezclante, no son equivalentes entre sí.

Otro ejemplo importante en el que se satisfacen las condiciones mezclantes es el siguiente. Sea  $\{X_k\}$  la sucesión de variables aleatorias dada por

$$X_k = \sum_{j=0}^{\infty} \beta_j \epsilon_{k-j}$$

con  $\epsilon_k$  variables aleatorias independientes idénticamente distribuidas con función de distribución continua. Entonces  $X_k$  está bien definida, es estrictamente estacionaria y cumple la condición de regularidad absoluta si las constantes  $\beta_j$  satisfacen  $|\beta_j| \rightarrow 0$  exponencialmente o al menos a una razón polinomial rápida, mostrando los coeficientes mezclantes un comportamiento asintótico semejante. En esta clase pueden mencionarse los modelos Autorregresivos y de Medias móviles (modelos ARMA) débilmente estacionarios.

Sin embargo, se presenta a continuación un contraejemplo dentro de este tipo de sucesiones al considerar los  $\epsilon_k$  con distribución discreta: supóngase que  $(Y_k)$  es una sucesión de variables aleatorias independientes con idéntica distribución bernoulli  $b(1/2)$  y con la cual se define la sucesión

$$X_k = \sum_{j=0}^{\infty} \frac{1}{2^{j+1}} Y_{k-j}$$

que se puede representar también por

$$X_k = \frac{1}{2} X_{k-1} + \frac{1}{2} Y_k$$

por lo que se trata de un proceso AR(1) que es estrictamente estacionario. Puede mostrarse que  $X_k$  se distribuye uniformemente en  $[0,1]$  (los dígitos de  $X_k$  en base 2 son precisamente  $Y_k, Y_{k-1}, \dots$  debido a que sus coeficientes son las potencias negativas de 2). También, para toda  $k$ ,  $X_k$  se puede expresar como función Borel-medible de  $X_{k+1}$ :  $X_k = 2X_{k+1} - [2X_{k+1}]$  (la parte fraccional de  $2X_{k+1}$ ) casi seguro, por lo que -por inducción-  $X_0$  es una función Borel-medible de  $X_n$ ,  $\forall n \geq 1$  y por lo tanto se justifica

$$\begin{aligned} \alpha(n) &\geq \alpha(\sigma(X_0), \sigma(X_n)) \geq \alpha(\sigma(X_0), \sigma(X_0)) \geq \\ &\geq P(X_0 < 1/2) = [P(X_0 < 1/2)]^2 = 1/4 \end{aligned}$$

por lo que  $(X_k)$  no es ni siquiera fuertemente mezclante (de hecho  $\alpha(n) = 1/4 \forall n \geq 1$ ); más aún, aunque se puede mostrar que es regular, la  $\sigma$ -álgebra generada con la "cola" derecha  $\bigcap_{j=1}^{\infty} \mathcal{F}_j^{\infty}$  es no trivial pues coincide, excepto por conjuntos de medida "cero", con  $\mathcal{F}_{-\infty}^{\infty}$ . Como conclusión, la clave del contraejemplo fue considerar una variable aleatoria discreta, por lo que no es posible generalizar a este tipo de distribuciones el cumplimiento de las condiciones mezclantes.

Corresponde el turno a un ejemplo de sucesión regular. Sea  $Y_k$  una sucesión de variables aleatorias definida como

$$Y_k = \sum_{j=1}^{\infty} j^{-r} \epsilon_{k-j}$$

con  $\epsilon_k$  variables aleatorias independientes idénticamente distribuidas como una normal con media cero y varianza la unidad, y  $1/2 \leq r \leq 3/4$ ; por representarse  $Y_k$  de esta manera, es una sucesión gaussiana regular. Ahora, sea

$$X_k = Y_k^2 - E[Y_k^2]$$

que por lo tanto también es regular. Este ejemplo se retomará más adelante.

#### 2.4 Comportamiento de condiciones mezclantes.

Como complemento, se enuncia el siguiente resultado acerca del comportamiento de las condiciones mezclantes (en plural):

Teorema 3.3. (Bradley, 1980 y 1983) Sea  $\{X_k\}$ ,  $k \in \mathbb{Z}$  una sucesión de variables aleatorias estrictamente estacionaria y mezclante en el sentido ergódico (ver definición A3.3), entonces se cumple que:

$$i) \quad \begin{cases} \lim_{n \rightarrow \infty} \beta(n) = 0 & \text{ó} \\ \beta(n) = 1 & \forall n \geq 1 \end{cases}$$

$$ii) \quad \begin{cases} \lim_{n \rightarrow \infty} \phi(n) = 0 & \text{ó} \\ \phi(n) = 1 & \forall n \geq 1 \end{cases}$$

$$iii) \quad \begin{cases} \lim_{n \rightarrow \infty} \psi(n) = 0 & \text{ó} \\ \lim_{n \rightarrow \infty} \psi(n) = 1 & \text{ó} \\ \psi(n) = \infty & \forall n \geq 1 \end{cases}$$

Para establecer el comportamiento de los coeficientes  $\alpha(n)$  y  $\rho(n)$  se tiene el siguiente resultado debido a Bradley (1981):

Teorema 3.4. Sea  $\{X_k\}$ ,  $k \in \mathbb{Z}$  una sucesión de variables aleatorias regular y estacionaria, entonces se cumple que:

$$i) 0 \leq \lim_{n \rightarrow \infty} \alpha(n) \leq 1/4$$

$$ii) 0 \leq \lim_{n \rightarrow \infty} \rho(n) \leq 1$$

Si no se impone la condición mezclante a la sucesión, se puede establecer el siguiente comportamiento establecido por Berbee (1981) y Bradley (1986):

**Teorema 3.5.** Sea  $\{X_k\}$ ,  $k \in \mathbb{Z}$  una sucesión de variables aleatorias estrictamente estacionaria y ergódica, entonces se tiene

$$i) \lim_{n \rightarrow \infty} \beta(n) = 1 - 1/p \quad \text{para algún } p \in \{1, 2, 3, \dots\} \cup \{\infty\}$$

$$ii) \lim_{n \rightarrow \infty} \phi(n) = 1 - 1/p \quad \text{para algún } p \in \{1, 2, 3, \dots\} \cup \{\infty\}$$

$$iii) \lim_{n \rightarrow \infty} \psi(n) = p - 1 \quad \text{para algún } p \in \{1, 2, 3, \dots\} \cup \{\infty\}$$

Sin embargo, no es posible establecer un teorema análogo a los anteriores para sucesiones no estacionarias.

## 2.5 Teorema Central del Límite para sucesiones mezclantes.

Trataremos ahora las condiciones relacionadas con el establecimiento del Teorema Central del Límite para sucesiones de variables aleatorias dependientes, que es la razón principal por la cual se introdujeron las condiciones mezclantes. En lo que sigue, se tratará con sucesiones  $\{X_k\}$  de variables aleatorias, y se establece como notación general

$$S_n = \sum_{k=1}^n X_k$$

y

$$\text{Var}[S_n] = \sigma_n^2$$

como las sumas parciales y sus varianzas. En general, se pedirá que  $\sigma_n^2$  tienda a infinito cuando  $n$  también tienda a infinito para evitar casos triviales de convergencia en que  $S_n$  no crezca (como en el caso de tomar a  $X_k$  como la diferencia de variables aleatorias independientes e idénticamente distribuidas).

El comportamiento de las sucesiones para la condición fuertemente mezclante es el mejor conocido de todos, donde se cuenta con un resultado general sobre leyes de atracción equivalente a los del capítulo anterior: dada la sucesión  $\{X_k\}$

estacionaria y fuertemente mezclante con coeficientes  $\alpha(n)$ , si las nuevas variables

$$\frac{1}{A_n} \sum_{k=1}^n X_k - B_n$$

convergen en distribución, entonces lo hacen necesariamente a una ley estable con parámetro  $\alpha$ , y las constantes  $A_n$  se escogen como

$$A_n = n^{1/\alpha} h(n)$$

donde  $h(n)$  es una función lentamente variante (ver A3.2) conforme  $n \rightarrow \infty$ .

Con este resultado como marco de referencia, nos concentraremos en el estudio en que  $\alpha = 2$ , es decir, en el caso de la convergencia a la distribución normal, que es la de nuestro interés. En esa situación, es la medida de dependencia fuertemente mezclante el único caso en que se han podido establecer exactamente condiciones necesarias y suficientes, las cuales están relacionadas con la existencia y comportamiento de momentos de las sumas parciales  $S_n$ . Sin embargo, en las otras condiciones mezclantes menos restrictivas los requisitos de suficiencia, asociados al comportamiento de los momentos de  $X_k$  y de la razón de decaimiento a cero de los coeficientes mezclantes, están próximos a ser necesarios, como lo indican algunos trabajos al respecto, pues al relajar un poco esos requerimientos, no es posible asegurar ya la convergencia a la distribución normal.

Iniciaremos con el estudio del Teorema Central del Límite para la condición  $\phi$ -mezclante, luego la  $\rho$ -mezclante y finalmente la fuertemente mezclante.

## 2.6 Convergencia de sucesiones $\phi$ -mezclantes y $\rho$ -mezclantes.

Respecto a las sucesiones  $\phi$ -mezclantes se desconoce en general el comportamiento de  $\sigma_n^2$ , pudiendo decirse solamente que es asintóticamente lineal con respecto a  $n$ ; por ello la condición que aquí se establece, debida a Peligrad (1986), es únicamente de suficiencia, y está dada como una exigencia algo más débil que pedir la existencia de momentos de orden  $2+\delta$ , con  $\delta > 0$  y sin imponer ningún requisito a los coeficientes  $\phi(n)$ :

Teorema 3.6. Sea  $\{X_k\}$  una sucesión estrictamente estacionaria  $\phi$ -mezclante que satisface  $E[X_0] = 0$  y  $E[g|X_1|] < \infty$  para una función  $g$  tal que  $x^2 < g(x) < x^{2+\delta}$  para alguna  $\delta > 0$ . Si además

$$g\left(\ln \exp \left[-d \sum_{k=1}^n k^{-1} \phi(k)\right]^{1/2}\right) \gg n \quad \text{si } n \rightarrow \infty \quad \forall d > 0$$

entonces

$$\frac{S_n}{\sigma_n} \xrightarrow{d} N(0,1)$$

**Teorema 3.7.** Sea  $\{X_k\}$  una sucesión uniformemente mezclante que satisface  $E[X_0] = 0$ ,  $E[|X_j|^{2+\delta}] < \infty$  para alguna  $\delta > 0$ . Si  $\sigma_n^2 \rightarrow \infty$  si  $n \rightarrow \infty$ , entonces

$$\frac{S_n}{\sigma \sqrt{n}} \xrightarrow{d} N(0,1)$$

**Teorema 3.8.** Sea  $\{X_k\}$  una sucesión uniformemente mezclante que satisface  $E[X_0] = 0$ , y los coeficientes  $\phi(n)$  cumplen con

$$\sum_{k=1}^{\infty} [\phi(k)]^{1/2} < \infty$$

Entonces

$$\sigma^2 = E[X_0^2] + 2 \sum_{k=1}^{\infty} E[X_0 X_k] < \infty$$

y si  $\sigma^2 > 0$ , también se tiene que

$$\frac{S_n}{\sqrt{n} \sigma} \xrightarrow{d} N(0,1)$$

En los dos últimos teoremas se han impuesto condiciones menos restrictivas a las sucesiones uniformemente mezclantes que para las que se establecerán al caso fuertemente mezclante más adelante. En el primer teorema, la condicionante se ha hecho sobre los momentos de  $X_k$ , y en el otro sobre la razón de decaimiento de los coeficientes  $\phi(n)$  pero, cabe resaltar, basta con uno de los requisitos para asegurar la convergencia a la distribución normal.

Como  $\phi(n) = 0$  si  $n > m$  en una sucesión  $m$ -dependiente, entonces se puede aplicar el último resultado para establecer la convergencia a la distribución normal: si  $\{X_k\}$  es una sucesión estacionaria  $m$ -dependiente con  $E[X_0^2] < \infty$  entonces  $\sigma^2$  está dada por

$$\sigma^2 = E[X_0^2] + 2 \sum_{k=1}^{\infty} E[X_0 X_k] < \infty$$

y si es mayor que cero, se cumple la convergencia a la distribución

normal de  $S_n/\sqrt{n}$ .

Si bien las cadenas de Markov homogéneas con un número finito de estados cumplen todas las condiciones mezclantes, aprovecharemos nuevamente el último teorema relativo a sucesiones uniformemente mezclantes para explicitar el logro del Teorema Central del Límite. Si para tales cadenas existe alguna  $s > 0$  tal que  $p_{ij}^{(n)} > 0 \forall i, j$  entonces los límites

$$p_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$$

existen  $\forall i, j$  y no dependen de  $i$ , además de que para algunas constantes  $C$  y  $r$ ,  $0 < r < 1$ , se verifica

$$\max_{i, j} | p_{ij}^{(n)} - p_j | \leq C r^n$$

que es una desigualdad crucial para establecer la velocidad de decaimiento de los coeficientes mezclantes, como a continuación observaremos.

Se supondrá que las probabilidades iniciales están dadas por  $P(X_1 = j) = p_j \forall j$ , por lo que la sucesión  $(X_n)$  es estacionaria.

Ahora sean

$$A = \{X_{t_1} = i_1, X_{t_2} = i_2, \dots, X_{t_u} = i_u\}$$

y

$$B = \{X_{t_{n+u}} = i_{n+u}, \dots, X_{t_v} = i_v\}$$

y calculemos

$$\begin{aligned} P(AB) &= p_{i_1} p_{i_2} \dots p_{i_u} p_{i_u}^{(n)} p_{i_{n+u}} \dots p_{i_v} = \\ &= P(A) p_{i_u}^{(n)} p_{i_{n+u}} \dots p_{i_v} \end{aligned}$$

$$P(A) P(B) = P(A) p_{i_{n+u}} p_{i_{n+u+1}} \dots p_{i_v}$$

por lo que

$$| P(AB) - P(A) P(B) | \leq P(A) | p_{i_u}^{(n)} - p_{i_{n+u}} | \leq P(A) C r^n$$

o sea

$$\phi(n) = \frac{| P(AB) - P(A) P(B) |}{P(A)} \leq C r^n$$

P(A)

y así

$$\sum_{n=1}^{\infty} [\rho(n)]^{1/2} \leq C r^{1/2} \sum_{n=1}^{\infty} r^n < \infty \quad \text{pues } 0 < r < 1$$

y entonces concluimos que

$$\frac{1}{\sqrt{n\sigma}} \sum_{k=1}^n (X_k - E[X_k]) \xrightarrow{d} N(0,1).$$

$$\text{si } 0 < \sigma^2 = E[X_0^2] + 2 \sum_{k=1}^{\infty} E[X_0 X_k].$$

Es posible generalizar a cualquier distribución inicial  $\pi$  con un poco de manipulación de las probabilidades y esperanzas condicionales en  $\pi$ , que aunque no desarrollaremos, consideraremos como cierto.

El resultado que sigue debido a Ibraгимov (1975), es relativo a sucesiones  $\rho$ -mezclantes, donde las condiciones impuestas a la velocidad de decaimiento de los coeficientes  $\rho(n)$  son tan débiles casi como es posible, tal como lo muestra Bradley (1980 y 1984b) con ejemplos donde otras restricciones más relajadas fallan en el cumplimiento de convergencia (en ellos se cambia la restricción sobre  $\rho(n)$  señalada en el teorema por la de  $\rho(n) = O((\ln n)^{-1})$  y se muestra que ya no se logra la convergencia a la distribución normal).

Teorema 3.9. Sea  $\{X_k\}$  una sucesión estacionaria de segundo orden, con  $E[X_0] = 0$  y  $\rho$ -mezclante que satisfice  $\sum_{k=1}^n \rho(2^k) < \infty$ ; y tal que  $\sigma_n^2 \rightarrow \infty$  si  $n \rightarrow \infty$ . Entonces

$$\frac{S_n}{\sigma_n} \xrightarrow{d} N(0,1)$$

## 2.7 Convergencia de sucesiones fuertemente mezclantes.

Los siguientes teoremas corresponden a sucesiones fuertemente mezclantes y difieren del caso  $\rho$ -mezclante por el comportamiento de  $\sigma_n^2$  en esas sucesiones, que está mejor determinado, por lo que aquí si se podrán establecer condiciones de necesidad y suficiencia.

Teorema 3.10. (Ibraгимov-Linnik, 1971) Sea  $\{X_k\}$  un proceso



fuertemente mezclante con coeficientes  $\alpha(n)$ . Para que se cumpla la condición del Teorema Central del Límite

$$\frac{S_n}{\sigma_n} \xrightarrow{d} N(0,1)$$

es necesario que se cumplan las siguientes dos condiciones:

a) que  $\sigma_n^2$  tenga la forma  $\sigma_n^2 = n h(n)$ , donde  $h(x)$  es una función lentamente variante en el dominio continuo  $x > 0$ .

b) que la función de distribución de  $S_n$ ,  $F_{S_n}(z)$  cumpla con

$$\lim_{n \rightarrow \infty} \frac{n}{p_n \sigma_n^2} \int_{|z| < \epsilon \sigma_n} z^2 dF_{S_n}(z) = 0$$

para cualquier  $\epsilon > 0$  y para cualquier par de sucesiones  $p_n$ ,  $q_n$  tales que

$$(i) \lim_{n \rightarrow \infty} p_n = \lim_{n \rightarrow \infty} q_n = \infty, \quad q_n = O(p_n), \quad p_n = O(n)$$

$$(ii) \lim_{n \rightarrow \infty} p_n^{-2} q_n^{1+\beta} n^{1-\beta} = 0 \quad \forall \beta > 0$$

$$(iii) \lim_{n \rightarrow \infty} p_n^{-1} \alpha(q_n) n = 0$$

Recíprocamente, es suficiente para la convergencia a la distribución normal que se satisfagan los requisitos establecidos en la primera condición, y en la segunda para algún par de sucesiones  $p_n$  y  $q_n$ .

La primera condición establece el comportamiento de la varianza  $\sigma_n^2$ , y la segunda limita el peso de los valores en las colas de la distribución de  $S_n$ , en función de un par de sucesiones relacionadas con el método utilizado en la demostración, debido a S. N. Bernstein, en que se descompone a  $S_n$  en bloques de tamaño  $p_n$  de variables aleatorias independientes, alternando con otros más pequeños de tamaño  $q_n$  cuya suma es despreciable en  $L_2$ . La manera de escoger  $p_n$  y  $q_n$  está dada en los incisos (i), (ii) y (iii).

Tal como lo mostraron Volkonskii y Rozanov (1959), el teorema anterior se simplifica si se supone que  $\sigma_n^2$  es asintóticamente lineal, requiriéndose en esencia que  $S_n^2/\sigma_n^2$  sea uniformemente integrable:

**Corolario 3.1.** Sea  $(X_k)$  una sucesión fuertemente mezclante que cumple  $\sigma_n^2 = n \sigma^2 (1 + O(1/n))$  cuando  $n \rightarrow \infty$ , y  $\sigma^2 > 0$ , entonces se verifica la condición

$$\frac{S_n}{\sigma_n} \xrightarrow{d} N(0,1)$$

si y sólo si

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} \int_{|z| > N} z^2 dF_{z_n}(z) = 0$$

donde  $F_{z_n}(z)$  es la función de distribución de  $S_n/\sigma_n$ . ▶

Otra versión, mejorada, del Teorema Central del Límite debida a Denker (1984), descarta la exigencia asintótica dejando, básicamente como único requisito necesario y suficiente que  $S_n^2/\sigma_n^2$  sea uniformemente integrable. Veamos:

**Teorema 3.11.** Sea  $(X_k)$  una sucesión estrictamente estacionaria, con  $E[X_0] = 0$  y segundos momentos finitos, fuertemente mezclante que cumple con  $\sigma_n^2 \rightarrow \infty$  cuando  $n \rightarrow \infty$ ; entonces

$$\frac{S_n}{\sigma_n} \xrightarrow{d} N(0,1)$$

si y sólo si  $S_n^2/\sigma_n^2$  es uniformemente integrable. ▶

**Teorema 3.12.** Sea  $(X_k)$  una sucesión fuertemente mezclante con coeficientes  $\alpha(n)$ , que cumple alguna de las dos siguientes condiciones:

i) para alguna  $0 \leq \delta < \infty$ ,

$$E[|X_1|^{2+\delta}] < \infty \quad \text{y} \quad \sum_{k=1}^{\infty} [\alpha(n)]^{\delta/(2+\delta)} < \infty$$

ii) para alguna  $c < \infty$ ,

$$P(|X_j| < c) = 1 \quad \text{y} \quad \sum_{k=1}^{\infty} \alpha(n) < \infty$$

Entonces se verifica

$$\sigma^2 = E[X_0^2] + 2 \sum_{k=1}^{\infty} E[X_0 X_k] < \infty$$

y si  $\sigma^2 > 0$ , entonces

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} N(0,1)$$

Esta es la clásica versión del Teorema Central del Límite desarrollada por Ibragimov (1962) en que los requisitos están planteados ya sea en términos de la existencia de momentos de orden superior a 2 para los elementos  $X_k$  de la sucesión, o que estén acotados en forma casi segura, además de determinar el orden de convergencia de los coeficientes mezclantes.

Se han podido construir contraejemplos de cadenas de Markov con un conjunto numerable de estados, estrictamente estacionarias, a las que se les ha impuesto alguna de las siguientes condiciones:

i) para alguna  $0 < \delta < \infty$ ,

$$E[|X_k|^{2+\delta}] < \infty \quad \text{y} \quad \alpha(n) = O(n^{-(1-\delta)(n+2/\delta)}), \quad \text{si } n \rightarrow \infty$$

o

ii) para alguna  $0 < c < \infty$ ,

$$P(|X_k| < c) = 1 \quad \text{y} \quad \alpha(n) = O(n^{-(1-\epsilon)}), \quad \text{si } n \rightarrow \infty$$

y escogiendo  $\sigma_n^2 = n^r$  para alguna  $r$  tal que  $1 < r < 2$ , entonces  $S_n$  no converge en distribución a la normal, sino a una ley estable simétrica con parámetro  $\alpha = 2/r$  que en consecuencia satisface  $1 < \alpha < 2$ .

### 3. MARTINGALAS.

#### 3.1 Origen y definición de martingalas.

El nombre de martingala tiene sus orígenes en los juegos de azar correspondiendo a una estrategia seguida por el jugador en sus apuestas, y lo toma de una localidad francesa cuyo nombre provenzal es *Martigues*. Sin embargo el concepto moderno que tiene en Probabilidad aparece con los trabajos de S. N. Bernstein (1927) y P. Lévy (1935) desarrollados desde fines de los años veinte a principios de los cuarenta de este siglo, para generalizar el Teorema Central del Límite al considerar sumas consecutivas de variables aleatorias. Sin embargo, ellos no emplearon específicamente ese término, mas bien fue el francés J. Ville en su trabajo de 1939 quien lo mencionó por primera vez, para que posteriormente le diera un fuerte impulso J.L. Dobb con sus investigaciones, entre las que resalta el Teorema de Convergencia en Martingalas, que representa un papel análogo a la ley fuerte de los grandes números en sucesiones de variables aleatorias

independientes. En las últimas décadas ha recibido nuevamente atención en un intento por generalizar las condiciones de convergencia a la distribución normal, que es el interés de este trabajo.

Como punto de partida se anotará la definición de martingala:

Definición 3.3. Sea  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad y  $\{\mathcal{F}_n\}$  una sucesión creciente de  $\sigma$ -álgebras:  $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots \subseteq \mathcal{F}_n \subseteq \dots \subseteq \mathcal{F}$ ,  $n \geq 1$ . Sea  $\{X_n\}$  una sucesión de variables aleatorias definidas en  $\Omega$  que satisfacen:

- i)  $X_n$  es medible respecto a  $\mathcal{F}_n$ .
- ii)  $E[|X_n|] < \infty$ .
- iii)  $E[X_n | \mathcal{F}_m] = X_m$  casi seguro,  $\forall m < n$ ,  $m, n \geq 1$ .

Entonces  $\{X_n\}$  es una martingala respecto a  $\{\mathcal{F}_n\}$  o simplemente  $\{X_n, \mathcal{F}_n\}$ ,  $n \geq 1$  es una martingala.

Si el signo de igualdad en la condición (iii) es sustituido por el de  $\geq$  ó  $\leq$ , entonces se denomina a  $\{X_n, \mathcal{F}_n\}$  una submartingala o supermartingala, respectivamente. ▶

Por una razón que se hará evidente más adelante, consideremos el siguiente ejemplo: sea  $\{Y_k\}$ ,  $k \geq 1$  una sucesión de variables aleatorias independientes con la misma distribución:  $P\{Y_k = 1\} = p$  y  $P\{Y_k = -1\} = q$ ,  $p + q = 1$  que podemos interpretar -por conveniencia- como el evento de ganar o perder con probabilidades  $p$  y  $q$  respectivamente en el  $k$ -ésimo turno de cierto juego. Si se considera  $A_k$  el monto de apuesta en ese turno, la cantidad ganada (o perdida) hasta entonces será

$$X_k = \sum_{j=1}^k A_j Y_j = X_{k-1} + A_k Y_k, \quad X_0 = 0$$

También resulta razonable suponer, como aparentemente ocurre en la práctica, que la cantidad apostada en un turno depende de la experiencia del jugador sobre sus apuestas anteriores y los resultados del juego. Esto es, podemos considerar  $\mathcal{F}_0 = (\emptyset, \Omega), \dots, \mathcal{F}_k = \sigma(Y_1, Y_2, \dots, Y_k)$  como los conjuntos de información disponibles por el jugador en cada turno, y por lo tanto considerar que  $A_k$  es  $\mathcal{F}_{k-1}$ -medible, es decir, la sucesión  $\{A_k, \mathcal{F}_{k-1}\}$  que determina la estrategia seguida por el jugador, es previsible.

El juego resulta ser justo, favorable o desfavorable al jugador hipotético de acuerdo a que

$$E [X_k - X_{k-1} | \mathcal{F}_{k-1}] = 0, \quad \geq 0, \quad \leq 0$$

lo que ocurre según

$$p = q = 1/2, \quad p > q, \quad p < q$$

respectivamente. Más aún, por la igualdad

$$E [X_k - X_{k-1} | \mathcal{F}_{k-1}] = E [X_k | \mathcal{F}_{k-1}] - E [X_{k-1} | \mathcal{F}_{k-1}] = E [X_k | \mathcal{F}_{k-1}] - X_{k-1}$$

los mismos casos corresponden a que  $(X_k, \mathcal{F}_k)$  sea una martingala, submartingala o supermartingala.

Particularizemos un poco más el ejemplo considerando que el jugador inicia sus apuestas con un peso (o un millón, si usted quiere) y dobla su apuesta cada turno hasta que logre un éxito, momento en el cual se retirará; es decir, supóngase que  $A_1 = 1$  y

$$A_k = \begin{cases} 2^{k-1} & \text{si } Y_j = -1 \quad \forall j < k \\ 0 & \text{otro caso} \end{cases}$$

Por lo tanto, la ganancia total del jugador si el primer éxito ocurre en el  $k$ -ésimo turno será

$$X_k = X_{k-1} + Y_k A_k = \sum_{j=1}^{k-1} (-1) 2^{j-1} + (1) 2^{k-1} = -(2^{k-1} - 1) + 2^{k-1} = 1$$

y si se supone que  $p = q = 1/2$ , entonces se puede demostrar que la estrategia finalizará en un número finito de turnos casi seguramente y que el valor esperado de  $X_k$  es igual a 1, o sea que aunque se trate de un juego justo con igual probabilidad de ganar o perder en cada turno y que se inicia con cero pérdidas o ganancias, se puede concluir que el jugador tendrá la expectativa de incrementar su capital en una unidad.

A esta estrategia se la conoce en juegos de azar precisamente como "martingala" y es a quien debe su nombre el concepto matemático definido al inicio de la sección. Más en general, la formalización matemática de cualquier estrategia de un jugador que se base en la experiencia de los resultados previos en el juego donde el monto esperado de su fortuna en el  $n$ -ésimo juego sea igual a su fortuna en el  $(n-1)$ -ésimo juego, es decir, que su ganancia esperada sea cero, es precisamente el concepto de martingala.

### 3.2 Ejemplos de martingala.

Considere ahora  $(Y_k)$  una sucesión de variables aleatorias independientes con la misma distribución:

$$P\{Y_k = 1\} = p, \quad P\{Y_k = -1\} = q$$

y hágase

$$S_k = \sum_{j=1}^k Y_j$$

y

$$X_k = \left[ \frac{q}{p} \right]^{S_k}$$

entonces  $(X_k, \mathcal{F}_k)$ ,  $\mathcal{F}_k = \sigma(S_1, S_2, \dots, S_k)$  es otro ejemplo de martingala puesto que al ser  $X_k$  una función de  $S_k$  y ser ésta  $\mathcal{F}_k$ -medible, entonces  $X_k$  también lo es, y porque se verifica

$$\begin{aligned} E[X_{k+1} | \mathcal{F}_k] &= E\left[ \left[ \frac{q}{p} \right]^{S_k} \left[ \frac{q}{p} \right]^{Y_{k+1}} | \mathcal{F}_k \right] = \\ &= \left[ \frac{q}{p} \right]^{S_k} E\left[ \left[ \frac{q}{p} \right]^{Y_{k+1}} | \mathcal{F}_k \right] = \\ &= \left[ \frac{q}{p} \right]^{S_k} E\left[ \left[ \frac{q}{p} \right]^{Y_{k+1}} \right] \\ &= \left[ \frac{q}{p} \right]^{S_k} = X_k \end{aligned}$$

$$\text{(pues } E\left[ \left[ \frac{q}{p} \right]^{Y_{k+1}} \right] = \left[ \frac{q}{p} \right]^1 p + \left[ \frac{q}{p} \right]^{-1} q = 1).$$

Otro caso de martingala lo constituyen las sumas de variables aleatorias independientes centradas en su valor esperado: sea  $(X_k)$  una sucesión de variables aleatorias tal que, por

simplicidad,  $E[X_k] = 0 \forall k$ , y hágase  $S_k = \sum_{j=1}^k X_j$ . Entonces, puesto

que  $\sigma(S_1, S_2, \dots, S_k) = \sigma(X_1, X_2, \dots, X_k)$  se tiene que

$$\begin{aligned} E[S_{k+n} | S_1, S_2, \dots, S_k] &= E[S_{k+n} | X_1, X_2, \dots, X_k] = \\ &= E[X_{k+1} + X_{k+2} + \dots + X_{k+n} | X_1, X_2, \dots, X_k] + E[S_k | X_1, X_2, \dots, X_k] = \end{aligned}$$

$$= 0 + S_k,$$

por lo que  $\{S_k\}$  cumple con las condiciones de martingala.

Siguiendo con los ejemplos, se construirá uno en que los elementos de la suma  $S_n$  son dependientes entre sí. Para ello, considérese en principio a  $\{Y_i\}$ , una sucesión de variables aleatorias independientes con idéntica distribución simétrica respecto al origen

$$P(Y_i = \pm 1) = 1/2 \quad i = 1, 2, \dots$$

y  $\{m_i\}$  una sucesión de enteros definida por la regla  $m_1 = 1$ ,

$$m_{i+1} = m_i + 2m_i^2$$

(obsérvese que los primeros valores son  $m_1 = 1$ ,  $m_2 = 3$ ,  $m_3 = 21$ ,  $m_4 = 903$ ,  $m_5 = 1\ 631\ 721$ ) con lo cual se establece

$$X_k = \begin{cases} Y_k & \text{si } m_i < k \leq (m_i + m_{i+1})/2 \\ Y_k \cdot I [Y_{m_i} > 0] & \text{si } (m_i + m_{i+1})/2 < k \leq m_{i+1} \end{cases}$$

la sucesión  $\{X_k\}$  coincide con  $\{Y_k\}$ , excepto que a cada intervalo definido por las  $m_i$ , la mitad se puede anular dependiendo del signo de una sola observación, la primera del intervalo correspondiente -y vale la pena observar que cada vez los intervalos se hacen más grandes, con crecimiento cuadrático, por lo que son más los términos que dependen del comportamiento de aquel elemento-.

Entonces, si  $S_n = \sum_{j=1}^n X_k$  y  $\mathcal{F}_n = \sigma(Y_1, Y_2, \dots, Y_n)$ ,  $(S_n, \mathcal{F}_n)$  es una martingala pues  $S_n$  es medible respecto a  $\mathcal{F}_n$ ,  $E[|S_n|]$  es finita y  $E[S_{n+1} | \mathcal{F}_n] = S_n$  casi seguro (para la última aseveración hay que proceder en forma parecida a los ejemplos anteriores, distinguiendo entre las dos posibles de definir  $X_k$ ).

### 3.3 Arreglo de martingalas.

La siguiente definición establece una extensión del concepto de martingalas y que resulta una manera análoga en que se definieron los arreglos infinitesimales en el capítulo anterior para estudiar la convergencia dentro de las leyes infinitamente divisibles.

Definición 3.4. Sea  $(S_{n_i}, \mathcal{F}_{n_i})$ ,  $1 \leq i \leq k_n$  una martingala

$\forall n \geq 1$  y si  $k_n \rightarrow \infty$  cuando  $n \rightarrow \infty$ , entonces a la doble sucesión así formada se le denomina arreglo de martingalas.  $\triangleright$

Con esta definición se considerará otro caso de martingala que también ilustra la condición de dependencia. Retomemos  $\{Y_i\}$ , la sucesión de variables aleatorias independientes con idéntica distribución simétrica respecto al origen

$$P(Y_i = \pm 1) = 1/2 \quad i = 1, 2, \dots$$

y constrúyase el arreglo  $X_{ni}$  con  $k_n = n$ ,  $X_{ni} = \frac{1}{\sqrt{n}} Y_i$  y

$$X_{ni} = \frac{1}{\sqrt{n}} Y_i \left[ \sum_{j=1}^{i-1} \frac{1}{j} Y_j \right]$$

entonces, cada  $X_{ni}$  depende de  $n, Y_1, Y_2, \dots, Y_i$  por lo que resultan ser dependientes entre sí; y para una  $i$  fija,  $X_{ni}$  y  $X_{ni}$  sólo difieren en el factor raíz cuadrado de estandarización, que las hace más pequeñas a medida que  $n$  crece. Si tomamos  $\mathcal{F}_{ni} = \sigma(Y_1, Y_2, \dots, Y_i) \forall n$ , entonces  $\{S_{ni}, \mathcal{F}_{ni}\}$  forman otro caso de martingala.

Demos otro caso de arreglo de martingala. Sea  $\{Y_i\}$  como la del ejemplo anterior, es decir,

$$P(Y_i = \pm 1) = 1/2 \quad i = 1, \dots, 2n.$$

Considérese  $k_n = 2n$  y  $m(n)$  una constante entre 1 y  $n$ ; hágase

$$X_{ni} = \begin{cases} n^{-1/2} Y_i & \text{si } i = 1, \dots, m(n) \\ n^{-1/2} Y_i I \left[ \frac{1}{\sqrt{n}} \sum_{l=1}^{m(n)} Y_l > 0 \right] & \text{si } i = m(n)+1, \dots, 2n \end{cases}$$

donde  $I[\cdot]$  es la función indicadora, y

$$S_{n,k_n} = \sum_{i=1}^{k_n} X_{ni} \\ = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i + \frac{1}{\sqrt{n}} \sum_{i=n+1}^{2n} Y_i I \left[ \frac{1}{\sqrt{n}} \sum_{l=1}^{m(n)} Y_l > 0 \right]$$

Se trata de una suma de  $2n$  términos donde los últimos  $n$  sólo aparecen si la suma de los primeros  $m(n)$  es positiva, que es la condición que establece la dependencia entre las  $X_{ni}$ . Si  $\mathcal{F}_{ni} = \sigma(Y_1, Y_2, \dots, Y_i)$ , se puede mostrar de manera análoga a los ejemplos



anteriores que  $(S_{n_i}, \mathcal{F}_{n_i})$  es una martingala para cada  $n$ , y por lo tanto constituyen un arreglo de martingalas.

Consideremos una variante del anterior ejemplo, que retomaremos más adelante. Sea  $(c_n)$  una sucesión de constantes tal que  $c_n \leq 1$  y  $c_n \rightarrow 0$  si  $n \rightarrow \infty$ , y  $m_n = [n(1-c_n)]$  (la parte entera de  $n(1-c_n)$ ). Considérese  $(Y_i)$  una sucesión de variables aleatorias independientes distribuidas como una normal estándar y

$$X_{n_i} = \begin{cases} n^{-1/2} Y_i & \text{si } i = 1, \dots, m_n \\ \sqrt{2} n^{-1/2} Y_i I \left[ \sum_{i=1}^{m_n} Y_i > 0 \right] & \text{si } i = m_n + 1, \dots, n \end{cases}$$

donde  $I[\cdot]$  es la función indicadora, y si

$$S_{n_i} = \sum_{j=1}^i X_{n_j}, \quad \mathcal{F}_{n_i} = \sigma(Y_1, Y_2, \dots, Y_i)$$

entonces  $(S_{n_i}, \mathcal{F}_{n_i})$  forman un arreglo de martingalas.

### 3.4 Definiciones y propiedades.

Abordemos algunas definiciones complementarias. En primer lugar, debe hacerse notar que la exigencia de que  $E[|X_n|]$  sea finita no es estrictamente necesaria para que  $E[X_n | \mathcal{F}_m]$  exista, por lo que se puede sustituir por una condición más amplia, que es definir

$$E[X_{n+1} | \mathcal{F}_m] = E[X_{n+1}^+ | \mathcal{F}_m] - E[X_{n+1}^- | \mathcal{F}_m]$$

siempre que

$$(\omega | E[X_n^+ | \mathcal{F}_m] < \infty) \cup (\omega | E[X_n^- | \mathcal{F}_m] < \infty) = \Omega$$

con lo que se tiene el concepto de Martingala Generalizada, (o submartingala o supermartingala según el caso) (Definición 3.5).

También conviene definir lo que se entiende por diferencia de martingala, concepto que de hecho ha intervenido ya en algunos de los ejemplos.

Definición 3.6. Sea  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad y  $(\mathcal{F}_n)$  una sucesión creciente de  $\sigma$ -álgebras:  $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots \subseteq \mathcal{F}_n \subseteq \dots \subseteq \mathcal{F}$ ,  $n \geq 1$ . Sea  $(X_n)$  una sucesión de variables aleatorias definidas en  $\Omega$  que satisfacen:

i)  $X_n$  es medible respecto a  $\mathcal{F}_n$ .

ii)  $E[|X_n|] < \infty$ .

iii)  $E[X_{n+1} | \mathcal{F}_n] = 0$  casi seguro,  $\forall m < n, m, n \geq 1$ .

Entonces  $\{X_n\}$  es una diferencia de martingala respecto a  $\{\mathcal{F}_n\}$  o simplemente  $(X_n, \mathcal{F}_n)$  es una diferencia de martingala.  $\triangleright$

La relación existente entre martingalas y diferencia de martingalas la sugiere el mismo nombre: si  $(S_n, \mathcal{F}_n)$  es una martingala, entonces  $(X_n, \mathcal{F}_n)$  con  $X_0 = S_0$  y  $X_n = S_n - S_{n-1}$ , es una diferencia de martingala; inversamente, si  $(X_n, \mathcal{F}_n)$  es una diferencia de martingala, entonces  $(S_n, \mathcal{F}_n)$  es una martingala, con

$$S_n = \sum_{j=1}^n X_j.$$

Si  $(X_n, \mathcal{F}_n)$  es una martingala que cumple con  $E[X_n^2] < \infty$  se dice que es de cuadrado integrable, característica que será muy recurrida en los resultados posteriores. También se requerirá en ocasiones que sea uniformemente integrable, lo que significa que  $\forall \epsilon > 0$  existe una constante  $K_\epsilon > 0$  tal que  $E[|X_n| I(|X_n| \geq K_\epsilon)] < \epsilon$ ,  $\forall n$ .

Resulta conveniente anotar en este momento un resultado teórico muy importante en el estudio de martingalas, y aunque no será utilizado directamente en este trabajo, interviene en el logro de los resultados que aquí se anotarán. Es el conocido como Teorema de Convergencia de Martingalas y resuelve un problema del tipo de existencia.

Teorema 3.13. Sea  $(X_n, \mathcal{F}_n)$   $n \geq 1$ , una submartingala acotada en  $L^1$ . Entonces existe una variable aleatoria  $X$  tal que  $\lim_{n \rightarrow \infty} X_n = X$  casi seguro y cumple con  $E[|X|] \leq \liminf_{n \rightarrow \infty} E[|X_n|] < \infty$ . Además, si la submartingala es uniformemente integrable, entonces  $X_n$  converge a  $X$  en  $L^1$  y si  $(X_n, \mathcal{F}_n)$   $n \geq 1$ , es una martingala acotada en  $L^2$ , entonces  $X_n$  converge a  $X$  en  $L^2$ .  $\triangleright$

Para poder plantear algunos resultados que confirmen el Teorema Central del Límite, objeto último de nuestro estudio, se introducirán y comentarán brevemente otros elementos que servirán como condiciones en tales teoremas. Si no se indica lo contrario, en la descripción se trabajará con arreglos de martingalas.

Se requerirá que las  $X_{n_i}$  en la diferencia de cada martingala del arreglo sean despreciables, de tal manera que no tengan un peso dominante en la suma, peculiaridad solicitada desde los teoremas anotados en el capítulo anterior, como el de Lindeberg-Feller. La

consideración de ser despreciable puede ser evaluada de distintas maneras, algunas de las cuales mencionaremos aquí, así como la relación que guardan entre sí.

La primera condición es la que se utiliza usualmente, la de ser asintóticamente uniformemente despreciable:

$$\lim_{n \rightarrow \infty} \max_l P(|X_{nl}| > \epsilon) = 0 \quad \forall \epsilon > 0$$

que es ligeramente más débil que la condición sobre la suma:

$$\lim_{n \rightarrow \infty} \sum_{l=1}^{k_n} P(|X_{nl}| > \epsilon) = 0 \quad \forall \epsilon > 0$$

aunque generalmente resultan ser equivalentes si  $S_{nk}$  converge en distribución. Y si  $X_{nl}$  son independientes, la anterior condición equivale a pedir que

$$\max_l |X_{nl}| \xrightarrow{P} 0$$

la cual a su vez es ligeramente más débil que la condición de Lindeberg:

$$\lim_{n \rightarrow \infty} \sum_{l=1}^{k_n} E[X_{nl}^2 I(|X_{nl}| > \epsilon)] = 0 \quad \forall \epsilon > 0$$

aunque en el caso de que  $\sum_{l=1}^{k_n} X_{nl}^2$  sea uniformemente integrable, resultan ser equivalentes. Por último, suele presentarse la misma condición de Lindeberg en versión condicional (a pesar de que en la mayoría de los casos resultan ser equivalentes):

$$\lim_{n \rightarrow \infty} \sum_{l=1}^{k_n} E[X_{nl}^2 I(|X_{nl}| > \epsilon) | \mathcal{F}_{n,l-1}] = 0 \quad \forall \epsilon > 0$$

Ahora considérese por simplicidad de notación una sola martingala  $(S_n, \mathcal{F}_n)$ , donde  $S_n = \sum_{k=1}^n X_k$ . Para estimar la variancia

$E[S_n^2]$  usualmente se considera  $V_n^2 = \sum_{k=1}^n E[X_k^2 | \mathcal{F}_{k-1}]$  que puede

interpretarse como la cantidad de información contenida en el pasado de la sucesión, explicación que resulta conveniente en la teoría inferencial sobre procesos estocásticos, ya que es posible mostrar que si las  $X_k$  son variables aleatorias independientes,  $V_n^2$  resulta ser la información de Fisher condicional, por lo que viene a ser una generalización de dicho concepto.

$V_n^2$  también sirve para relacionar martingalas con submartingalas, pues dada  $(Y_n, \mathcal{F}_n)$  una submartingala, existe una

martingala  $\{S_n, \mathcal{F}_n\}$  y una sucesión creciente de variables aleatorias no negativas  $A_n$  tal que

$$V_n^2 = S_n + A_n$$

y si  $A_n$  es  $\mathcal{F}_{n-1}$ -medible, es decir previsible, entonces está unívocamente determinada y es igual a  $V_n^2$ , resultado que es conocido como descomposición de Dobb.

La importancia de  $V_n^2$  radica en que su comportamiento está estrechamente ligado al de  $S_n$ , por lo que si este comportamiento es muy errático, también lo será el de  $S_n$  y por lo tanto no se garantizaría el cumplimiento del Teorema Central del Límite. Entonces, aparece en cierta forma lógico pedir como condición que

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} V_n^2 = 1 \quad s_n^2 = E[V_n^2]$$

Aunque en las primeras versiones del Teorema Central del Límite para martingalas se utilizó  $V_{ni}^2$  para estandarizar la suma, en la práctica resulta más conveniente tratar en su lugar con  $U_{ni}^2 = \sum_{j=1}^i X_{nj}^2$  pues así se evita trabajar con esperanzas condicionales y porque en el caso de arreglos de martingalas, bajo ciertas restricciones comúnmente impuestas a  $V_{ni}^2$  se verifica

$$\max_i |V_{ni}^2 - U_{ni}^2| \xrightarrow{P} 0$$

es decir, es virtualmente equivalente estandarizar con  $U_{ni}^2$  que con  $V_{ni}^2$ .

Para preparar otros resultados, introduciremos tres definiciones, la de convergencia en distribución establemente y mezclante, y la de convergencia débil en  $L^1$ , las cuales están relacionadas, además de que la primera permite cambiar el tipo de normas requeridas en las condiciones del Teorema Central del Límite.

Definición 3.7. Sea  $\{X_n\}$  una sucesión de variables aleatorias definida en  $(\Omega, \mathcal{F}, P)$  y que convergen en distribución a la variable aleatoria  $X$ ; entonces la convergencia es estable si para todos los puntos de continuidad  $x$  de la distribución de  $X$  y para todo evento  $E$  contenido en  $\mathcal{F}$  existe el límite

$$\lim_{n \rightarrow \infty} P(\{X_n \leq x\} \cap E) = Q_x(E)$$

y si  $Q_n(E) \rightarrow P(E)$  cuando  $n \rightarrow \infty$ . Esta convergencia se denota por

$$X_n \xrightarrow{d} X \quad (\text{establemente}) \quad \triangleright$$

Como se verá, la convergencia en distribución mezclante es un caso particular de la anterior.

Definición 3.8. Sea  $\{X_n\}$  una sucesión de variables aleatorias definida en  $(\Omega, \mathcal{F}, P)$  y que convergen en distribución a la variable aleatoria  $X$ ; entonces la convergencia es mezclante (en el sentido de Rényi) si y sólo si para todos los puntos de continuidad  $x$  de  $X$  y para todo evento  $E$  contenido en  $\mathcal{F}$  se cumple

$$\lim_{n \rightarrow \infty} P(X_n \leq x) \cap E \rightarrow P(X \leq x) P(E) \quad \text{cuando } x \rightarrow \infty$$

Esta convergencia se denota por

$$X_n \xrightarrow{d} X \quad (\text{mezclante}) \quad \triangleright$$

Definición 3.9. una sucesión  $\{X_n\}$  de variables aleatorias integrables en  $(\Omega, \mathcal{F}, P)$  converge debilmente en  $L^1$  a la variable aleatoria integrable  $X$  si  $\forall E \in \mathcal{F}$  se cumple

$$\lim_{n \rightarrow \infty} E [X_n I [E]] = E [X I [E]]$$

y se denota por

$$X_n \xrightarrow{L^1} X \quad (\text{debilmente en } L^1) \quad \triangleright$$

### 3.5 Convergencia a mezcla de distribuciones normales.

Apliquemos las definiciones vistas hasta aquí a arreglos de martingalas para obtener el siguiente teorema.

Teorema 3.14. Sea  $\{S_{n_i}, \mathcal{F}_{n_i}\}$ ,  $1 \leq i \leq k_n$ ,  $k_n \rightarrow \infty$  un arreglo de martingala con media cero, de cuadrado integrable y con  $X_{n_i}$  sus correspondientes diferencias de martingala. Sea  $X^2$  una variable aleatoria finita casi seguramente. Si

$$\max_i |X_{n_i}| \xrightarrow{P} 0,$$

$$\sum_{i=1}^{k_n} X_{n_i}^2 \xrightarrow{P} X^2,$$

$E \left[ \max_l X_{nl}^2 \right]$  está acotado en  $n$

y las  $\sigma$ -álgebras están anidadas:

$$\mathcal{F}_{nl} \subseteq \mathcal{F}_{n+1,l} \quad l = 1, \dots, k_n, \quad \forall n \geq 1,$$

entonces

$$S_{n,k_n} = \sum_{l=1}^{k_n} X_{nl} \xrightarrow{d} Z \quad (\text{establemente})$$

donde  $Z$  es una variable aleatoria con función característica

$$\phi(t) = E \left[ \exp\left(-\frac{1}{2} Z^2 t^2\right) \right]$$

El siguiente corolario establece el mismo resultado con diferentes restricciones.

Corolario 3.2. En el teorema anterior supóngase que se cumple la condición de Lindeberg en su modalidad condicional

$$\lim_{n \rightarrow \infty} \sum_{l=1}^{k_n} E \left[ X_{nl}^2 \mathbb{1} \{ |X_{nl}| > \epsilon \} \mid \mathcal{F}_{n,l-1} \right] = 0 \quad \forall \epsilon > 0,$$

en lugar de que  $\max_l |X_{nl}| \xrightarrow{P} 0$  y  $E \left[ \max_l X_{nl}^2 \right]$  esté acotado en  $n$ ,

y que  $\sum_{l=1}^{k_n} X_{nl}^2 \xrightarrow{P} X^2$  se sustituye por

$$V_{n,k_n}^2 = \sum_{l=1}^{k_n} E \left[ X_{nl}^2 \mid \mathcal{F}_{n,l-1} \right] \xrightarrow{P} X^2,$$

entonces la conclusión del teorema se sigue cumpliendo.

Este teorema y su corolario significan en cierta forma una generalización del Teorema Central del Límite pues en el caso particular en que  $X^2 = 1$  casi seguramente,  $Z$  se distribuye como una normal estándar. En general, si  $N$  es una variable aleatoria normal estándar independiente de  $X$ , la raíz de  $X^2$ , entonces  $XN$  tiene precisamente como función característica a  $\phi(t) = E \left[ \exp\left(-\frac{1}{2} Z^2 t^2\right) \right]$ , por lo que se dice que  $XN$  es una mezcla de distribuciones normales, o una normal con varianza aleatoria.

Podemos ilustrar este teorema retomando la martingala ejemplificada más arriba donde  $\{Y_i\}$  es una sucesión de variables aleatorias independientes con distribución  $P\{Y_i = \pm 1\} = 1/2$ ,

$i=1, \dots, k_n = n$ ,  $X_{nl} = \frac{1}{\sqrt{n}} Y_l$  y  $X_{nl} = \frac{1}{\sqrt{n}} Y_l \left[ \sum_{j=1}^{l-1} \frac{1}{j} Y_j \right]$ . Puesto que  $\mathcal{F}_{nl}$

$= \sigma(Y_1, Y_2, \dots, Y_l) \forall n$ , por supuesto que se cumple  $\mathcal{F}_{n,l} \subseteq \mathcal{F}_{n+1,l}$ .  
Además, se tiene que

$$\sum_{j=1}^{l-1} \frac{1}{J} Y_j \xrightarrow{c.s.} Y = \sum_{j=1}^{\infty} \frac{1}{J} Y_j$$

y como es finito casi seguramente, se puede afirmar que

$$U_{nn}^2 = \sum_{l=1}^n X_{nl}^2 = \frac{1}{n} \sum_{l=1}^n \left[ \sum_{j=1}^{l-1} \frac{1}{J} Y_j \right]^2 \xrightarrow{c.s.} Y^2$$

además

$$\max_l |X_{nl}| = \frac{1}{\sqrt{n}} \max_l \left| \sum_{j=1}^{l-1} \frac{1}{J} Y_j \right| \xrightarrow{c.s.} 0,$$

y por

$$\max_l X_{nl}^2 \leq U_{nn}^2 \xrightarrow{c.s.} Y^2,$$

junto con

$$E [U_{nn}^2] = \frac{1}{n} \sum_{l=1}^n E \left[ \sum_{j=1}^{l-1} \frac{1}{J} Y_j \right]^2 \rightarrow \sum_{j=1}^{\infty} \frac{1}{J^2} = E [Y^2]$$

se tiene que

$$E \left[ \max_l X_{nl}^2 \right] \text{ está acotado en } n$$

por lo tanto se cumplen las condiciones estipuladas en el teorema tomando  $X^2 = Y^2$ , por lo que concluimos que

$$S_{n,n} = \sum_{l=1}^n X_{nl} \xrightarrow{d} Z \quad (\text{establemente})$$

donde la función característica de  $Z$  es  $\phi(t) = E [\exp (-\frac{1}{2} X^2 t^2)]$ .  
Es decir, nuestro ejemplo cumple el teorema establecido siendo  $X^2$  una variable aleatoria que no es constante casi seguramente.

Mantengamos un momento más las condiciones del teorema anterior y supongamos además que  $X^2 > 0$  casi seguro y es  $\mathcal{F}_{n,l}$ -medible  $\forall n, l$ . Entonces, para  $n = 2, \dots, k_n$

$$E [X_{nl}^{-1} S_{nl} | \mathcal{F}_{n,l-1}] = X_{nl}^{-1} E [S_{nl} | \mathcal{F}_{n,l-1}] = X_{nl}^{-1} S_{n,l-1}$$

por lo cual  $(X_{nl}^{-1} S_{nl}, \mathcal{F}_{n,l})$ ,  $n \geq 2, l=2, \dots, k_n$  también es un arreglo de martingala y las condiciones especificadas en el teorema siguen siendo válidas, condicionadas en  $X$ , por ejemplo

$$E \left[ \max_l X_{nl}^{-2} X_{nl}^2 | X \right] = X^{-2} E \left[ \max_l X_{nl}^2 | X \right] \text{ y está acotado en } n$$

y

$$\sum_{l=2}^{k_n} X^{-2} X_{nl}^2 = X^{-2} \sum_{l=1}^{k_n} X_{nl}^2 - X^{-2} X_{n1}^2 \xrightarrow{P} 1$$

y como este último límite es constante, podemos validar la conclusión del teorema, condicionando en  $X$ , quedando de la forma

$$\sum_{l=2}^{k_n} X^{-1} X_{nl} \xrightarrow{d} N(0,1) \quad (\text{establemente})$$

o

$$X^{-1} S_{nk_n} = \sum_{l=1}^{k_n} X^{-1} X_{nl} \xrightarrow{d} N(0,1) \quad (\text{establemente})$$

pues  $X^{-1} X_{nl} \xrightarrow{P} 0$ , y además la convergencia es incondicional porque el límite no depende de  $X$ . Finalmente,

$$U_{nk_n}^{-1} S_{nk_n} = [U_{nk_n}^{-1} X] [X^{-1} S_{nk_n}] \xrightarrow{d} N(0,1)$$

pues se cumple también  $U_{nk_n}^{-1} X \xrightarrow{P} 1$ .

Por este razonamiento se ha alcanzado el primer resultado sobre el Teorema Central del Límite para martingalas donde estandarizar con  $U_{nk_n}$  ha tenido el efecto de cancelar las fluctuaciones de  $S_{nk_n}$  para permitir la convergencia a la normal estándar. Sin embargo, la solicitud de que  $X$  sea medible en todas las  $\sigma$ -álgebras es demasiado exigente y puede reemplazarse por  $P(X^2 > 0) = 1$ . Estamos ahora en posibilidad de formalizar el resultado, y de especificarlo un poco más, pues el tipo de convergencia que se verifica es el caso particular del estable, llamado mezclante (recuerde su definición dada en 3.6).

**Teorema 3.15.** Supóngase que las condiciones del teorema 3.14 se mantienen y que  $P(X^2 > 0) = 1$ . Entonces

$$U_{nk_n}^{-1} S_{nk_n} \xrightarrow{d} N(0,1) \quad (\text{mezclante}) \quad \blacktriangleright$$

**Corolario 3.3.** Supóngase que las condiciones del corolario 3.2 se mantienen y que  $P(X^2 > 0) = 1$ . Entonces

$$U_{nk_n}^{-1} S_{nk_n} \xrightarrow{d} N(0,1) \quad (\text{mezclante}) \quad \blacktriangleright$$



### 3.6 Condiciones generales para el Teorema Central del Limite.

A continuación se presenta otro resultado de tipo general, que como veremos, considera como caso particular a sucesiones que forman martingalas.

**Teorema 3.16.** Sea  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad y  $(\mathcal{F}_{nk})$ ,  $n = 0, 1, \dots$ ,  $k = 0, 1, \dots, n$  una doble sucesión de  $\sigma$ -álgebras que cumple con  $\mathcal{F}_{n0} = \{\emptyset, \Omega\} \subseteq \mathcal{F}_{n1} \subseteq \mathcal{F}_{n2} \subseteq \dots \subseteq \mathcal{F}_{nn} \subseteq \mathcal{F} \forall n$ , y sea  $(X_{nk})$  una doble sucesión de variables aleatorias definidas en  $\Omega$  tal que  $X_{nk}$  es medible respecto a  $\mathcal{F}_{nk}$ . Para alguna  $t$ ,  $0 < t \leq 1$ , hágase  $S_{nt} = \sum_{k=0}^{[tn]} X_{nk}$ . Si  $\forall \epsilon > 0$  se cumplen las tres siguientes condiciones :

i)

$$\max_{1 \leq i \leq [tn]} |X_{ni}| \xrightarrow{P} 0$$

o

$$\sum_{i=1}^{[tn]} P(|X_{ni}| > \epsilon \mid \mathcal{F}_{n,i-1}) \xrightarrow{P} 0$$

ii)

$$\sum_{i=1}^{[tn]} E[X_{ni} \mid \mathcal{F}_{n,i-1}] \xrightarrow{P} 0$$

iii)

$$\sum_{i=0}^{[tn]} \{X_{ni} - E[X_{ni} \mid \mathcal{F}_{n,i-1}]\}^2 \xrightarrow{P} \sigma_t^2$$

o

$$\sum_{i=1}^{[tn]} V[X_{ni} \mid \mathcal{F}_{n,i-1}] \xrightarrow{P} \sigma_t^2,$$

donde  $\sigma_t^2 \geq 0$ .

Entonces

$$S_{nt} \xrightarrow{d} N(0, \sigma_t^2)$$

Resulta interesante destacar que en el caso de que las variables fueran independientes, escogiendo  $t = 1$  y  $\sigma_t^2 = \sigma^2$ , las condiciones coinciden con las establecidas en el teorema 2.12 del capítulo anterior, por lo que este resultado representa una generalización de aquél en el dominio de las variables aleatorias dependientes.

La condición (iii) representa una estimación de la varianza condicional, y el supuesto (i) impone a cada componente la característica de ser despreciable respecto de la suma, lo que

Junto con la condición (ii) asegura que es posible expresar a  $S_{nt}$  como

$$S_{nt} = Z_{nt} + W_{nt}$$

donde  $W_{nt} \xrightarrow{P} 0$  y  $Z_{nt} = \sum_{k=0}^{[tn]} Y_{nk}$  es una martingala con diferencia de martingala  $Y_{nk}$  que cumple con

$$E [Y_{nk} | \mathcal{F}_{n,i-1}] = 0 \quad \text{y} \quad |Y_{nk}| \leq c$$

uniformemente  $\forall i, n$ . Por lo tanto, puede establecerse como resultado particular para martingalas el siguiente

**Teorema 3.17.** Sea  $(X_{nk}, \mathcal{F}_{nk})$  una diferencia de martingala de cuadrado integrable que satisface para una  $t$   $0 < t \leq 1$ , la condición de Lindeberg condicional:

$$\sum_{l=0}^{[tn]} E [X_{nl}^2 I [ |X_{nl}| > \epsilon ] | \mathcal{F}_{n,i-1}] \xrightarrow{P} 0 \quad \forall \epsilon > 0$$

Y si

$$\sum_{l=0}^{[tn]} E [X_{nl}^2 | \mathcal{F}_{n,i-1}] \xrightarrow{P} \sigma_t^2$$

o

$$\sum_{l=0}^{[tn]} X_{nl}^2 \xrightarrow{P} \sigma_t^2$$

entonces, para  $S_{nt} = \sum_{k=0}^{[tn]} X_{nk}$

$$S_{nt} \xrightarrow{d} N(0, \sigma_t^2)$$

En el caso de variables aleatorias independientes, es posible intercambiar el supuesto de ser asintóticamente infinitesimales de manera uniforme, utilizado también en los teoremas anteriores, por condiciones sobre los momentos, como el de ser finitos los de primer o segundo orden. Un resultado análogo para el caso dependiente está dado por el siguiente teorema.

**Teorema 3.18.** Sea  $(X_{nk}, \mathcal{F}_{nk})$  una diferencia de martingala de cuadrado integrable,  $X_{n0} = 0$ ; sea además  $S_{nt} = \sum_{k=0}^{[tn]} X_{nk}$  y  $F_{nk}(x) = P(X_{nk} \leq x | \mathcal{F}_{nk-1})$ . Si se cumple

$$\sum_{l=0}^{(l)n} E [X_{nl}^2 | X_{n,l-1}] = \sum_{l=0}^{(l)n} \sigma_{nl}^2 = V_{nl} \xrightarrow{P} \sigma_l^2$$

con  $0 \leq \sigma_l^2 < \infty$ , y

$$\sum_{l=0}^{(l)n} \int_{|x|>c} \left\{ |x| \left| F_{nk}(x) - \int_{-\infty}^{x/\sigma_{nl}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \right| \right\} dx \xrightarrow{P} 0 \quad \forall c > 0$$

entonces

$$S_{nl} \xrightarrow{d} N(0, \sigma_l^2)$$

### 3.7 Refinamiento de las condiciones de convergencia.

Para mostrar que las condiciones del teorema 3.14 no son precisamente las indicadas, retomaremos un ejemplo presentado al principio de la sección que no satisface las hipótesis planteadas, pero que para cierta variante si se cumple el Teorema Central del Límite mientras que para otra no.

Específicamente, sea  $(Y_i)$  tal que

$$P(Y_i = \pm 1) = 1/2 \quad i = 1, \dots, 2n.$$

y con  $k_n = 2n$

$$X_{nl} = \begin{cases} n^{-1/2} Y_l & \text{si } l = 1, \dots, m(n) \\ n^{-1/2} Y_l I \left[ \frac{1}{\sqrt{n}} \sum_{l=1}^{m(n)} Y_l > 0 \right] & \text{si } l = m(n)+1, \dots, 2n \end{cases}$$

$$S_{n,k_n} = \sum_{l=1}^{k_n} X_{nl} = \frac{1}{\sqrt{n}} \sum_{l=1}^n Y_l + \frac{1}{\sqrt{n}} \sum_{l=n+1}^{2n} Y_l I \left[ \frac{1}{\sqrt{n}} \sum_{l=1}^{m(n)} Y_l > 0 \right]$$

Escogeremos  $m(n)$  de dos maneras,  $m(n) = n$  y  $m(n) = \lfloor \sqrt{n} \rfloor$ . En el primer caso

$$S_{n,2n} = \sum_{l=1}^{k_n} X_{nl} = \frac{1}{\sqrt{n}} \sum_{l=1}^n Y_l + \frac{1}{\sqrt{n}} \sum_{l=n+1}^{2n} Y_l I \left[ \frac{1}{\sqrt{n}} \sum_{l=1}^n Y_l > 0 \right]$$

$$\xrightarrow{d} N_1 + N_2 I [N_1 > 0],$$

donde  $N_1$  y  $N_2$  son variables aleatorias independientes con distribución normal estándar, aunque no se trata de una mezcla de normales. Aún más,

$$U_{n,2n}^2 = \sum_{j=1}^{2n} X_{nj}^2 = 1 + I \left[ \frac{1}{\sqrt{n}} \sum_{l=1}^n Y_l > 0 \right] \xrightarrow{d} X^2,$$

donde  $X^2$  es una variable aleatoria con probabilidades  $P(X^2 = 1) = P(X^2 = 2) = 1/2$ . Con ello establecemos

$$\frac{S_{n,2n}}{U_{n,2n}} \xrightarrow{d} N_1 I [N_1 \leq 0] + \frac{1}{\sqrt{2}} (N_1 + N_2) I [N_1 > 0]$$

que no tiene una distribución normal estándar.

Ahora, para el caso en que  $m(n) = \lfloor \sqrt{n} \rfloor$  tenemos

$$\begin{aligned} S_{n,2n} &= \frac{1}{\sqrt{n}} \sum_{l=1}^n Y_l + \frac{1}{\sqrt{n}} \sum_{l=n+1}^{2n} Y_l I \left[ \frac{1}{\sqrt{n}} \sum_{l=1}^n Y_l > 0 \right] \\ &\xrightarrow{d} N_1 + N_2 I [N_3 > 0], \end{aligned}$$

que son también variables independientes con distribución  $N(0,1)$ , y en este caso  $U_{n,2n}^2$  sigue convergiendo en distribución a  $X^2$  (pues la función indicadora sigue teniendo igual probabilidad  $1/2$  de valer cero o uno aunque el número de sumandos haya cambiado). Combinando ambos resultados, obtenemos

$$\frac{S_{n,2n}}{U_{n,2n}} \xrightarrow{d} N_1 I [N_3 \leq 0] + \frac{1}{\sqrt{2}} (N_1 + N_2) I [N_3 > 0]$$

que sí se distribuye como una normal estándar.

En conclusión, tenemos dos ejemplos muy similares en que sólo para uno de ellos se verifica la convergencia a la distribución normal. Revisando el teorema 3.14, puede observarse que para ambos se cumplen todas las condiciones, excepto que

$$U_{nk_n}^2 = \sum_{l=1}^{k_n} X_{nl}^2 \xrightarrow{p} X^2$$

pues la convergencia en distribución de  $U_{n,2n}^2$  no puede ampliarse a convergencia en probabilidad. Por lo tanto, se hace evidente que esta restricción no resulta ser la más apropiada, como se pretendía mostrar.

Entonces, para reconocer si es válido el Teorema Central del Límite es preciso observar la diferencia existente entre una y otra variante del ejemplo, y se nota que si bien en ambos casos la sucesión de constantes  $m(n)$  es divergente, la relación existente con  $n$  misma es distinta, pues únicamente cuando se cumple la convergencia a la normal el cociente  $m(n)/2n$  tiende a cero. En consecuencia, parece ser que la restricción adecuada establece que

la información contenida en  $U_{nk}^2$  debe corresponder a una parte cada vez más irrelevante de la sucesión  $(X_{ni})$  a medida que  $n$  aumenta. Esta discusión nos conduce al siguiente teorema.

**Teorema 3.19.** Para todo  $n \geq 1$ , sean  $S_{n1}, S_{n2}, \dots, S_{nk_n}$  variables aleatorias definidas en el espacio de probabilidad  $(\Omega_n, \mathcal{F}_n, P_n)$  y de cuadrado integrable. Sean  $\mathcal{F}_{ni}$   $\sigma$ -álgebras anidadas:  $\mathcal{F}_{n1} \subseteq \mathcal{F}_{n2} \subseteq \dots \subseteq \mathcal{F}_{nk} \subseteq \mathcal{F}_n$  y tal que  $S_{ni}$  es  $\mathcal{F}_{ni}$ -medible. Sean  $X_{ni} = S_{ni} - S_{ni-1}$  ( $S_{n0} = 0$ ) y  $U_{ni}^2 = \sum_{j=1}^i X_{nj}^2$ . Si  $G_n$  es una sub  $\sigma$ -álgebra de  $\mathcal{F}_n$ , sea  $G_{ni} = \sigma(\mathcal{F}_{ni} \cup G_n)$ , con  $G_{n0} = (\Omega_n, \phi)$  la  $\sigma$ -álgebra trivial.

Si se cumple que para cada  $n$

$$\max_i |X_{ni}| \xrightarrow{P} 0, \quad E \left[ \max_i X_{ni}^2 \right] \text{ está acotado en } n,$$

existen variables aleatorias  $u_n^2$   $G_n$ -medibles tales que

$$U_{nk_n}^2 - u_n^2 \xrightarrow{P} 0, \\ \sum E(X_{ni} | G_{ni-1}) \xrightarrow{P} 0 \quad \sum |E(X_{ni} | G_{ni-1})|^2 \xrightarrow{P} 0$$

y

$$\lim_{\delta \rightarrow 0} \liminf_{n \rightarrow \infty} P(U_{nk_n}^2 > \delta) = 1$$

entonces

$$\frac{S_{nk_n}}{U_{nk_n}} \xrightarrow{d} N(0, 1)$$

Si en lugar de la última condición se pide que  $U_{nk_n}^2 \xrightarrow{d} \chi^2$ , entonces

$$S_{n,k_n} \xrightarrow{d} Z$$

donde  $Z$  es una variable aleatoria con función característica

$$\phi(t) = E \left[ \exp\left(-\frac{1}{2} \chi^2 t^2\right) \right]$$

En este resultado se condiciona a que ningún elemento de la diferencia de martingalas debe ser demasiado grande con respecto a la suma en primero y segundo momentos de la misma manera que en el teorema 3.14; la restricción sobre  $U_{nk_n}^2$  significa que sólo contiene información sobre la  $\sigma$ -álgebra  $G_n$ . Si agregamos el supuesto

$E(X_{n_i} | \mathcal{F}_{n,i-1}) = 0$  la sucesión se convierte en una diferencia de martingalas, y que en el teorema se establezca como condición que la suma de los valores esperados condicionados por la dependencia converja a cero y que su dispersión medida por el cuadrado de los valores absolutos tienda a anularse, significa que realmente no se altera esa propiedad de martingala por haber extendido la  $\sigma$ -álgebra  $\mathcal{F}_{n_i}^{n_i}$  a  $G_{n_i}$ . Finalmente, la última condición equivale a pedir que  $U_{nk_n}^2$  sea positiva casi seguramente, que si se reemplaza porque  $U_{nk_n}^2$  converja en distribución a la variable aleatoria  $\chi^2$  entonces la propia sucesión  $S_{nk_n}$  converge en distribución a una variable aleatoria que es una mezcla de normales.

Apliquemos el teorema a las dos variantes del ejemplo para resaltar que ahora las condiciones establecidas se ajustan mejor a los requisitos verdaderamente necesarios.

Se había comentado que se satisfacía la condición de despreciabilidad de los elementos  $(X_{n_i})$ , por lo que podemos pasar a la siguiente. Para ello, escójase

$$U_n^2 = U_{nk_n}^2 = 1 + I \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{m(n)} Y_i > 0 \right]$$

por lo que evidentemente  $U_{nk_n}^2 \xrightarrow{p} 0$ .

Ahora, por la definición de  $G_{n_i}$  se tiene

$$G_{n_i} = \begin{cases} \mathcal{F}_{n, m(n)} & i = 1, \dots, m(n) \\ \mathcal{F}_{n_i} & i = m(n)+1, \dots, 2n \end{cases}$$

con lo cual se tiene

$$\begin{aligned} E[X_{n_i} | G_{n,i-1}] &= \begin{cases} E[X_{n_i} | \mathcal{F}_{n, m(n)}] & i = 1, \dots, m(n) \\ E[X_{n_i} | \mathcal{F}_{n, i-1}] & i = m(n)+1, \dots, 2n \end{cases} \\ &= \begin{cases} X_{n_i} & i = 1, \dots, m(n) \\ 0 & i = m(n)+1, \dots, 2n \end{cases} \end{aligned}$$

y por lo tanto

$$\sum_{i=1}^{2n} E[X_{n_i} | G_{n,i-1}] = \sum_{i=1}^{m(n)} X_{n_i} = \frac{1}{\sqrt{n}} \sum_{i=1}^{m(n)} Y_i$$

que nos sirve para concluir que

$$P \left[ \left| \frac{1}{\sqrt{n}} \sum_{l=1}^{m(n)} Y_l \right| \geq \epsilon \right] = P \left[ \left| \frac{1}{\sqrt{m(n)}} \sum_{l=1}^{m(n)} Y_l \right| \geq 2 \sqrt{\frac{n}{m(n)}} \epsilon \right] \approx \\ \approx 1 - \phi \left[ 2 \sqrt{\frac{n}{m(n)}} \epsilon \right] \xrightarrow{p} 0$$

(por el teorema de deMoivre-Laplace visto en el primer capítulo)

y

$$\sum_{l=1}^{2n} |E[X_{nl} | \mathcal{G}_{n,l-1}]|^2 = \sum_{l=1}^{m(n)} |X_{nl}|^2 = \frac{1}{n} \sum_{l=1}^{m(n)} |Y_l|^2 = \frac{m(n)}{n} \xrightarrow{p} 0$$

si y sólo si  $\frac{m(n)}{n} \rightarrow 0$ .

Por último, como

$U_{n,2n}^2 \xrightarrow{d} \eta^2 > 0$  casi seguro,  $P(\eta=1) = P(\eta=2) = 1/2$ ; entonces

$$\lim_{\delta \rightarrow 0} \liminf_{n \rightarrow \infty} P\{U > \delta\} = 1$$

con lo que, juntando todas las condiciones que se satisfacen y que son las exigidas en el teorema 3.19, podemos afirmar que

$$\frac{S_{n,2n}}{U_{n,2n}} \xrightarrow{d} N(0,1)$$

si y sólo si  $\frac{m(n)}{n} \rightarrow 0$ .

### 3.8 Análisis de la estandarización aleatoria.

Por otro lado, para destacar el efecto que resulta algo sorpresivo de estandarizar aleatoriamente para conseguir la convergencia a la distribución normal, utilizemos otro de los ejemplos de martingala que se presentaron al inicio. En él, se había dado  $\{Y_i\}$  tal que  $P(Y_i = \pm 1) = 1/2$ ,  $i \geq 1$  y  $\{m_i\}$  definida por  $m_1 = 1$ ,  $m_{i+1} = m_i + 2m_i^2$ ,  $i > 1$ , para formar

$$X_k = \begin{cases} Y_k & \text{si } m_i < k \leq (m_i + m_{i+1})/2 \\ Y_k I [Y_{m_i} > 0] & \text{si } (m_i + m_{i+1})/2 < k \leq m_{i+1} \end{cases}$$

Entonces, si  $S_n = \sum_{k=1}^n X_k$  y  $U_n^2 = \sum_{k=1}^n X_k^2$ , se mostrará que ninguna de las dos converge a alguna distribución, exhibiendo dos subsucesiones que tengan diferente límite; y sin embargo, el cociente  $S_n/U_n$  tiene un límite bien determinado.

En efecto,

$$m_k^{-1} U_{m_k}^2 = m_k^{-1} [U_{m_{k-1}}^2 + m_{k-1}^2 + m_{k-1}^2 I [Y_{m_{k-1}} > 0]] \\ \xrightarrow{d} \frac{1}{2} [1 + I [Y_1 > 0]] \quad \text{si } k \rightarrow \infty$$

mientras que

$$(m_k + m_k^2)^{-1} U_{m_k + m_k^2}^2 = (m_k + m_k^2)^{-1} [U_{m_k}^2 + m_k^2] \xrightarrow{p} 1$$

entonces, no existe sucesión de constantes  $(c_n)$  tal que  $U_n^2/c_n$  converja débilmente a alguna distribución propia. De la misma manera, no existe  $(c_n)$  tal que  $S_n/c_n$  pueda converger a alguna distribución. Pero,  $(S_n, \mathcal{F}_n)$ , con  $\mathcal{F}_n = \sigma(Y_1, Y_2, \dots, Y_n)$ , forman una martingala que satisface los requisitos del teorema 3.19, por lo que

$$\frac{S_n}{U_n} \xrightarrow{d} N(0, 1)$$

es decir, se cumple el Teorema Central del Límite.

### 3.9 Aproximación de sucesiones mezclantes.

La estructura de martingala es bastante general, y puede abarcar casos como los tratados en la sección anterior, esto es, pueden aproximarse mediante martingalas sucesiones estacionarias de variables aleatorias que cumplen condiciones mezclantes. En lo que resta de la sección, se presentarán algunos resultados, varios de los cuales amplían los ya anotados, y que se han logrado por esta vía. M. I. Gordin presentó en 1969 resultados importantes en ese sentido, y a él se debe un resultado que, aunque de carácter general, involucra condiciones difíciles de comprobar. Sin embargo, a partir de él se pueden derivar otros teoremas que en seguida se presentan, los identificados como 3.20 y 3.21, y que aunque más particulares, resultan útiles en la práctica.

Teorema 3.20. Sea  $\{X_k\}$  una sucesión estacionaria y ergódica con  $E[X_0] = 0$ ,  $E[X_0^2] < \infty$ , y  $S_n = \sum_{k=1}^n X_k$ . Si  $G_0$  es una sub  $\sigma$ -álgebra de  $\mathcal{F}$  tal que  $X_0$  es  $G_0$ -medible. Si  $\forall M \geq 0$



$$\sum_{k=1}^{\infty} E [X_k E (X_k | G_0)]$$

converge, y

$$\lim_{M \rightarrow \infty} \sum_{k=N}^{\infty} E [X_k E (X_k | G_0)] = 0$$

uniformemente para toda  $N \geq 1$ , entonces

$$\frac{1}{n} E [S_n^2] \rightarrow \sigma^2 \quad \text{si } n \rightarrow \infty$$

$0 \leq \sigma^2 < \infty$ ; y si  $\sigma^2 > 0$ ,

$$\frac{S_n}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$$

Si se satisface la condición fuertemente mezclante, este resultado involucra como caso particular al teorema 3.10 de Ibragimov. La demostración usando martingalas, tanto del teorema como de la implicación se debe a C.C. Heyde (1974a).

A continuación se anota una variación del teorema anterior, cuya demostración también proporcionó C.C. Heyde (1974a), y en que no se requiere que  $X_0$  sea  $G_0$ -medible, a cambio de imponerle otras restricciones menos exigentes (de hecho, las condiciones del teorema 3.20 implican las del 3.21).

**Teorema 3.21.** Sea  $\{X_k\}$  una sucesión estacionaria y ergódica con transformación  $T$  uno a uno que preserve la medida, y con  $E[X_0] = 0$ ,  $E[X_0^2] < \infty$ , y  $S_n = \sum_{k=1}^n X_k$ . Si  $G_0$  es una sub  $\sigma$ -álgebra de  $\mathcal{F}$  tal que  $G_0 \subseteq T^{-1}(G_0)$ ,  $G_k \subseteq T^{-k}(G_0)$  y sea

$$Y_k = E(X_k | G_0) - E(X_k | G_{-1}) \quad -\infty < k < \infty.$$

Si se cumple que

$$\sum_{k=-\infty}^{\infty} Y_k = Y_0 \in L^2, \quad E[Y_0^2] = \sigma^2 > 0$$

y

$$\frac{1}{n} E [S_n^2] \rightarrow \sigma^2 \quad \text{si } n \rightarrow \infty$$

entonces

$$\frac{S_n}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$$

Este teorema se puede aplicar a procesos de medias móviles para afinar los requisitos estipulados en la sección anterior para garantizar la convergencia a la distribución normal, pues ya no

exige que la serie de los valores absolutos de los coeficientes sea convergente, y sobre la densidad espectral, remueve la petición de que sea uniformemente acotada (véase el ejemplo de las páginas 72-73 y consúltese Anderson (1970) y Hannan (1970)).

**Corolario 3.4.** Sea el proceso estacionario lineal ergódico  $\{X_n\}$  dado por

$$X_n - \mu = \sum_{j=-\infty}^{\infty} \beta_j \varepsilon_{n-j} \quad \sum_{j=-\infty}^{\infty} \beta_j^2 < \infty$$

con  $\{\varepsilon_n\}$  variables aleatorias independientes idénticamente distribuidas con  $E[\varepsilon_n] = 0$  y  $\text{Var}[\varepsilon_n] = \sigma^2$ . Sea

$$f(\lambda) = \frac{\sigma^2}{\sqrt{2\pi}} \left[ \sum_{j=-\infty}^{\infty} \beta_j e^{i\lambda j} \right]^2$$

la densidad espectral de  $\{X_n\}$ . Si  $f(\lambda)$  es continua en  $\lambda = 0$  y  $f(0) > 0$  entonces

$$\sum_{j=1}^n [X_j - \mu] \xrightarrow{d} N(0, 2\pi f(0))$$

si además  $\{X_n\}$  es un proceso estacionario puramente no determinístico, se puede expresar como

$$X_n - \mu = \sum_{j=0}^{\infty} \beta_j \varepsilon_{n-j} \quad \sum_{j=0}^{\infty} \beta_j^2 < \infty$$

$$\beta_0 = 1 \quad E[\varepsilon_n] = 0 \quad E[\varepsilon_n \varepsilon_m] = 0 \quad m \neq n$$

y el corolario se sigue cumpliendo. >

Es posible remover la restricción en el segundo momento  $E[X_0^2] < \infty$ , imponiendo condiciones solamente sobre  $E[X_0]$ , como lo demuestra el siguiente resultado propuesto por Gordin en 1975, y cuya demostración fue presentada por P. Hall y C.C. Heyde en 1980.

**Teorema 3.22.** Sea  $\{X_k\}$  una sucesión estacionaria y ergódica con transformación  $T$  uno a uno que preserve la medida, y con  $E[X_0] = 0$ ,  $E[|X_0|] < \infty$ ,  $X_k(\omega) = X_0(T^k\omega)$  y sea  $S_n = \sum_{k=1}^n X_k$ . Si  $G_0$  es una sub  $\sigma$ -álgebra de  $\mathcal{F}$  tal que  $G_0 \subseteq T^{-1}(G_0)$ ,  $G_k = T^{-k}(G_0)$ . Si se satisfacen las condiciones

$$\sum_{k=0}^{\infty} \left\{ E[|E(X_0 | G_{-k})|] + E[|X_0 - E(X_0 | G_k)|] \right\} < \infty$$

y

$$\limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} E [|S_n|] < \infty$$

entonces

$$\lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} E [|S_n|] = \sigma^2$$

con  $0 \leq \sigma^2 < \infty$ ; y si  $\sigma^2 > 0$

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2)$$

Al aplicar este resultado a sucesiones uniforme o fuertemente mezclantes obtenemos el siguiente

Corolario 3.5. Sea  $\{X_k\}$  un proceso estacionario que satisface  $E[X_0] = 0$ ,  $E[|X_0|^{2+\delta}] < \infty$  para alguna  $0 \leq \delta < \infty$  y que

$$\frac{1}{n} E[S_n^2] \rightarrow \sigma^2 > 0 \quad \text{si } n \rightarrow \infty$$

Si se cumple alguna de las condiciones siguientes:

i) El proceso invertido en el tiempo  $\{X_{-k}\}$ , resultante simplemente de cambiar el orden en el índice, cumple la condición fuertemente mezclante con  $\sum_{k=1}^{\infty} \gamma(k) < \infty$  (tiene sentido considerar el proceso invertido pues  $\gamma(n)$  no es simétrico).

ii) El proceso  $\{X_k\}$  cumple con la condición fuertemente mezclante con  $\sum_{k=1}^{\infty} [\alpha(k)]^{(k+\delta)/(2+\delta)} < \infty$ .

Entonces

$$\frac{S_n}{\sigma\sqrt{n}} \xrightarrow{d} N(0, 1)$$

El inciso (i) de este corolario requiere una razón de decaimiento de los coeficientes mezclantes más lenta que la establecida en el teorema 3.12 pero aumenta el requisito de que  $\frac{1}{n} E[S_n^2] \rightarrow \sigma^2$ . En cambio, el inciso (ii) resulta más exigente que las condiciones propuestas en el correspondiente teorema 3.10, inciso (i) porque solicita una mayor razón de decaimiento de los coeficientes mezclantes y agrega la misma necesidad de convergencia sobre  $E[S_n^2]$ .

### 3.10 Algunas extensiones.

Como comentarios finales, se puede apuntar que existen resultados sobre convergencia de los momentos de la suma estandarizada de la sucesión a los respectivos de la distribución normal, que para variables aleatorias independientes datan desde los trabajos de S.N. Bernstein en 1939, hasta otros más recientes de B.M. Brown (1969 y 1970), y en particular para martingalas destacan los de P. Hall (1978b). Para la discusión de algunos resultados precisamente sobre martingalas, el lector puede consultar a P. Hall y C.C. Heyde (1980).

Análogos a la cota de Berry-Esseen para variables aleatorias independientes, existen resultados sobre rapidez de convergencia para martingalas, tanto del tipo uniforme como puntual. Sin embargo, la rapidez depende en gran parte de la de  $V_{nk}^2 = \sum_{i=1}^n E [X_{ni}^2 | \mathcal{F}_{n,i-1}] \xrightarrow{P} 1$ , la cual imposibilita que sea del orden de  $n^{-1/2}$ , la cual corresponde al caso independiente. Por esta razón deben incorporarse algunas restricciones adicionales que sean viables de imponer en el aspecto práctico para garantizar una rapidez de convergencia del orden de al menos, digamos,  $O(n^{-1/4})$ . Los resultados más importantes se encuentran en Ibragimov (1963), Heyde y Brown (1970), Grams (1972), Basu (1976), Nakata (1976), Kato (1978) y Erickson (1978); nuevamente se remite al trabajo de P. Hall y C.C. Heyde (1980) para el estudio del tema.

### 4. BIBLIOGRAFIA COMENTADA.

- I.A. Ibragimov y Yu. V. Linnik, 1971. Un libro ya clásico en el tratamiento de sucesiones de variables aleatorias, en que se exponen con notable claridad los resultados sobre el Teorema Central del Límite para sucesiones mezclantes; a pesar del tiempo transcurrido desde su publicación sigue siendo una consulta de actualidad. En él aparecen las definiciones de condiciones mezclantes y las demostraciones de los teoremas 3.8, 3.10, 3.11, 3.12 y el corolario 3.1.
- R.C. Bradley, 1986. Se trata de un artículo de reciente aparición que establece, con ayuda de los resultados disponibles hasta la fecha, una versión unificada de las condiciones mezclantes, sin abordar la teoría asintótica. Aunque no incluye las demostraciones de los teoremas considerados (3.1 a 3.5) debido a que se trata más bien de una visión global sobre el tema, su consulta resulta de mucha utilidad por los comentarios y la amplia referencia bibliográfica que contiene.
- M. Poligrad, 1986. También se trata de una referencia de reciente aparición, en la que se resumen los resultados alcanzados en materia del Teorema Central del Límite para sucesiones mezclantes, aunque sin demasiada claridad. Presenta los

teoremas 3.6 al 3.10, el corolario 3.1 y bibliografía actualizada.

P. Hall y C.C. Heyde, 1980. Un libro de alto contenido sobre martingalas, donde se expresan con detalle y profundidad las condiciones de convergencia a la distribución normal para sucesiones de este tipo. El conjunto de resultados presentados es bastante amplio e incluye los obtenidos por los autores. Contiene las demostraciones de la segunda parte de este capítulo: teoremas 3.13 a 3.22 y corolarios 3.2 a 3.5. Se recomienda su lectura para quien desee profundizar en el tema.

## CAPITULO IV.

### CONVERGENCIA DE SUMAS DE VARIABLES ALEATORIAS: EVIDENCIA EMPIRICA MEDIANTE SIMULACION.

#### 1. INTRODUCCION.

En este material se presenta el apoyo necesario para utilizar, explorar y aprovechar el paquete de cómputo que pretende dar evidencia empírica de la convergencia de sumas estandarizadas a la función de distribución normal estándar y de resultados teóricos relacionados con esta convergencia.

En la sección 2, con el objeto de brindar información sobre los elementos involucrados en la simulación, se presentan brevemente los conceptos de sumas de variables aleatorias, función de distribución normal estándar, función de distribución empírica, convergencia en distribución, cota de Berry-Esseen y la prueba de bondad de ajuste de Anderson-Darling.

En la sección 3, que constituye la parte central de este documento, aparecen algunos ejercicios sugeridos para utilizar el paquete de simulación, resaltando en cada uno de ellos algún resultado específico de la correspondiente parte teórica. El objetivo de esta sección es optimizar el uso de esta herramienta de cómputo, servir de apoyo al docente para relacionar el contenido de su curso con prácticas que reafirmen los conceptos adquiridos por el estudiante, y al estudioso del tema para confirmar los resultados manejados en su trabajo personal.

Con el fin de facilitar la realización de las prácticas, en una última sección se incluyen las referencias resumidas de las variables aleatorias que son simuladas por el paquete: función de densidad, dominio de la variable, utilidad, gráfica y primeros momentos.

Finalmente, se presentan resultados de simulación realizados con el paquete que siguen aproximadamente la secuencia de las prácticas propuestas, y que resumen un gran número de corridas por medio de tablas y gráficas; se indica el grado de aproximación a la distribución normal bajo el criterio de la prueba Anderson-Darling, la cantidad de simulaciones realizadas, y los valores de los parámetros considerados; todo esto significa una valiosa experiencia que ha sido posible conseguir por este medio acerca del comportamiento del Teorema Central del Límite bajo distintas circunstancias.

## 2. CONCEPTOS BASICOS MANEJADOS POR EL PAQUETE DE SIMULACION.

El paquete se maneja mediante sencillos menús interactivos que permiten seleccionar la o las distribuciones a simular y los valores de los parámetros involucrados.

Inicialmente aparece un cuadro que identifica la distribución a simular, la gráfica de su función de densidad o de probabilidades, y remarca el hecho de que las variables simuladas son idénticamente distribuidas o no.

El cuadro central donde se presentan los resultados de la simulación se compone de la graficación de ejes estandarizados, la impresión de títulos que identifican el caso que se está tratando, el dibujo de la distribución normal estandar, la simulación y graficación de la función de distribución empírica, y el cálculo y presentación de la conclusión de la prueba de bondad de ajuste de Anderson-Darling.

Para una mejor comprensión de estos elementos, se presenta a continuación una descripción de cada uno de los conceptos involucrados.

### 2.1 Distribución Normal Estándar.

Se dice que una variable aleatoria  $X$  se distribuye como una normal estandar si su función de densidad está dada por

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad -\infty < x < \infty$$

mientras que la función de distribución por no tener una expresión elemental, adquiere la forma

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

La función de densidad tiene la particularidad de ser simétrica respecto al eje de las ordenadas, por lo que su media es precisamente igual a 0, mientras que su varianza es igual a 1. Esto se denota escribiendo

$$X \sim N(0,1)$$

Precisamente esta distribución juega un papel preponderante en la Teoría de la Probabilidad, debido en mucho a los resultados que aquí se abordarán.

## 2.2 Sumas de Variables Aleatorias.

Puesto que los resultados que se obtendrán serán referidos a sumas de variables aleatorias, es conveniente definir las. Sea una sucesión  $X_1, \dots, X_n, \dots$  de variables aleatorias; entonces la suma estandarizada de la sucesión está dada por

$$\frac{1}{B_n} \left[ \sum_{l=1}^n X_l - A_n \right]$$

donde  $A_n$  y  $B_n$  son constantes que sólo dependen de  $n$ . Usualmente, y cuando existen, estas constantes se escogen como la media y la desviación estándar de la suma de las variables. La forma de calcularlas, bajo condiciones de independencia está dada por la siguiente afirmación

Sea una sucesión  $X_1, \dots, X_n, \dots$  de variables aleatorias independientes. Entonces

$$E \left[ \sum_{l=1}^n X_l \right] = \sum_{l=1}^n E(X_l)$$

y

$$\text{Var} \left[ \sum_{l=1}^n X_l \right] = \sum_{l=1}^n \text{Var}(X_l)$$

## 2.3 Función de Distribución Empírica.

En el proceso de simulación la función de distribución empírica juega un papel importante por lo que se anota aquí lo que se entiende por este concepto. Sea  $X_1, \dots, X_n$  una muestra aleatoria que proviene de una población con función de distribución  $F(x)$ . La función de distribución empírica está dada por

$$F_n(x) = \frac{1}{n} \sum_{l=1}^n I_{(-\infty, x]}(X_l)$$

donde  $I_{(a,b)}(x)$  es la función indicadora definida por



$$I_{(a,b)}(x) = \begin{cases} 1 & \text{si } x \in (a,b] \\ 0 & \text{si } x \notin (a,b] \end{cases}$$

entonces, la función empírica es monótona no decreciente, de forma escalonada y cumple con que  $F_n(-\infty) = 0$  y  $F_n(\infty) = 1$ . Un resultado importante es el teorema de Glivenko-Cantelli que asegura que  $F_n(x)$  converge con probabilidad uno a la función de distribución de las variables  $X_1, \dots, X_n$ .

## 2.4 Convergencia en Distribución.

La aproximación de la distribución normal a la distribución de sumas estandarizadas se da en el sentido de la convergencia en distribución, para lo cual supongamos que  $X_1, \dots, X_n$  es una sucesión de variables aleatorias con funciones de distribución  $F_n(x)$  y  $X$  otra variable aleatoria cualquiera con función de distribución  $F(x)$ . Se dice que la sucesión converge en distribución a  $X$  si

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad \forall x \text{ punto de continuidad de } F(x)$$

Por comodidad, se referirá la convergencia en distribución simplemente como la convergencia de la sucesión a la variable  $X$ .

## 2.5 Cota de Berry-Esseen.

Asociada con el concepto de rapidez de convergencia para variables aleatorias independientes e idénticamente distribuidas, la cota de Berry-Esseen establece la máxima diferencia que puede presentarse entre la función de distribución empírica y la distribución normal con parámetros  $\mu = 0$  y  $\sigma^2 = 1$ , en todo el recorrido de la variable aleatoria  $X$ , quedando definida la cota en función del tercer momento absoluto de la variable aleatoria, de acuerdo al resultado de la sección 4 del capítulo 1. Por lo tanto, para casos específicos de variables aleatorias, a medida que la cota es menor, se dice que su suma estandarizada converge más rápidamente a la distribución normal estándar. Al final de la misma sección aparecen cuadros con las cotas correspondientes a algunas distribuciones.

## 2.6 Prueba de Bondad de Ajuste de Anderson-Darling.

Para evaluar si la hipótesis de convergencia de una suma estandarizada a la distribución normal se cumple, se utiliza la prueba de Anderson-Darling que está dada por

$$A^2 = n \int_{-\infty}^{\infty} \{F_n(x) - F(x)\}^2 / \{F(x)\}\{1-F(x)\} dF(x)$$

donde  $F_n(x)$  es la función de distribución empírica, y  $F(x)$  es la distribución bajo la hipótesis.

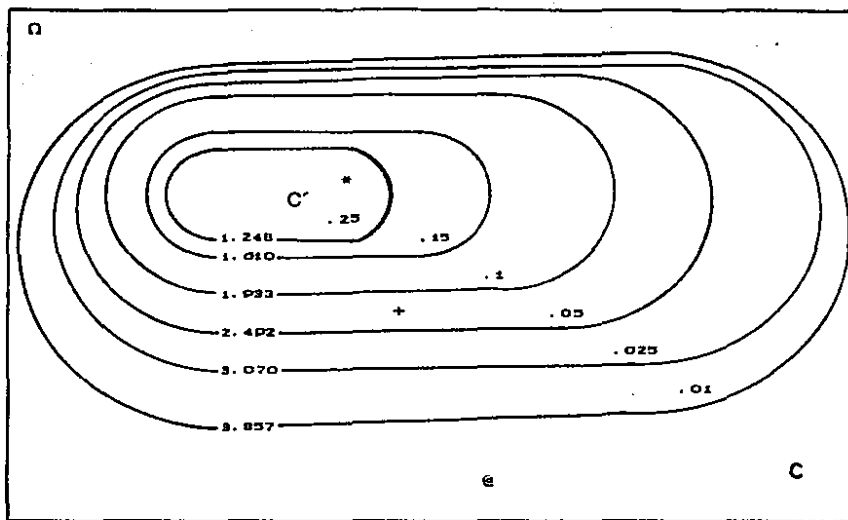
Dada una muestra aleatoria,  $X_1, \dots, X_n$  una forma alternativa para  $A^2$  que se utiliza para el cálculo es la siguiente:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n [ (2i-1) \ln Z_{(i)} + (2n+1-2i) \ln (1-Z_{(i)}) ]$$

donde  $Z_i = F(x_i)$  y  $Z_{(i)}$  son las estadísticas de orden de las  $Z_i$ .

La hipótesis de nulidad se rechaza si  $A^2 > A_{(1-\alpha)}$ , el  $\alpha$ -ésimo percentil superior de la distribución. En el siguiente cuadro se presentan los valores críticos  $A_{(1-\alpha)}$  correspondientes a algunos  $\alpha$  de interés.

$\alpha$	.25	.15	.1	.05	.025	.01
$A_{(1-\alpha)}$	1.248	1.610	1.933	2.492	3.070	3.857



- \* muy chica, se acepta con  $\alpha \leq .25$
- @ muy grande, se rechaza con  $\alpha \geq .01$
- + se acepta con  $\alpha \leq .05$  y se rechaza con  $\alpha \geq .1$

Figura 4.1  
Crecimiento de la Región Crítica cuando  $\alpha$  tiende a 1.

### 3. PRACTICAS SUGERIDAS PARA LA EXPLOTACION DEL PAQUETE.

Existen diferentes aspectos sobre la convergencia de sumas de variables aleatorias que pueden ser estudiados mediante la simulación tales como la rapidez de convergencia, afectación por la variación de parámetros, diferencias causadas por los distintos tipos de variables aleatorias involucradas, contraejemplos, y otros. Aspectos todos analizados y sustentados teóricamente, pero que pueden ser confirmados heurísticamente con la ventaja de recurrir a elementos visuales y experimentales que coadyuvan en la comprensión del concepto.

A continuación se presenta un conjunto de prácticas conducentes a experimentar mediante simulación los tópicos arriba mencionados, abordándolos en una secuencia que coincide con el

avance teórico que usualmente se da al tema y que es el seguido en los capítulos anteriores. Ello permite que este material pueda alternar con sesiones en que se presenten y demuestren formalmente los resultados con las prácticas que reafirmen los conocimientos adquiridos y familiaricen al lector con su significado; también facilita la localización de condiciones específicas para el Teorema Central del Límite.

Se recomienda seleccionar la práctica que mejor corresponda al avance del estudio, de acuerdo al objetivo que se establece al principio de cada una. Para su realización se proponen una o varias actividades que implican la ejecución de algunas simulaciones, la observación de las condiciones en que se efectúan, recabación de características acerca de las distribuciones simuladas, y la contestación a interrogantes que conducen al interesado a la confirmación del objetivo planteado. Se incluyen además conclusiones generales que señalan los posibles resultados e interpretaciones que el usuario debió obtener, así como referencias a los teoremas que apoyan tales conclusiones.

Resultaría provechoso que el usuario del paquete leyera todas las instrucciones de la actividad que realizará antes de comenzar, pues con ello podría planear mejor su trabajo, estaría atento a los detalles que debe observar en cada simulación y podría recabar en el momento oportuno toda la información que se le requiere.

Para utilizar el paquete en una microcomputadora personal deberá encenderse la máquina con el sistema operativo MS-DOS instalado. Inmediatamente después de dar la fecha y hora es necesario seguir los siguientes pasos:

1) Teclar

>a:  (cambia a la unidad de diskette)

2) Instalar el diskette del paquete en la ranura de la unidad de disco flexible.

3) Teclar

>CONVERGE TCL   
(lee el archivo y corre el programa)

En este momento pueden realizarse ya las prácticas que se deseen mediante la selección apropiada en los menús que aparecen.

→ Práctica Número 1. Teorema de deMoivre-Laplace y Rapidez de Convergencia.

**Objetivo:**

Experimentar la convergencia a la distribución normal de la suma estandarizada de ensayos Bernoulli. Observar el efecto que causa la variación de la probabilidad de éxito.

\* Actividad (a):

Simular:

"S.Control Completo de Parámetros.", "1.bernoulli"  
"Probabilidad de éxito? " .5  
"Valor de N (5-1000)? " 50  
"Simulaciones (5-1000)? " 50

- ▶ ¿Qué tipo de distribución es la Bernoulli? ¿Es continua o discreta? con  $p=1/2$ , ¿Es simétrica o no? ¿tiene dominio acotado?
- ▶ ¿Podría decirse que guarda algún parecido con la distribución normal?
- ▶ ¿Parecería entonces 'lógico' aproximar las probabilidades de la suma de variables aleatorias con distribución Bernoulli y  $p = 1/2$  por medio de la distribución normal?
- ▶ Reconociendo la curva suave al ejecutar la simulación propuesta como la función de distribución normal y la función escalonada como la distribución empírica de la suma de las variables Bernoulli, ¿Se puede concluir que por este argumento gráfico que en efecto se puede aproximar la distribución de la suma por medio de la normal?
- ▶ Una vez graficada la distribución teórica de la normal y la función de distribución empírica, y antes de que aparezca el resultado de la prueba de Anderson-Darling, pruebe visualmente: ¿Se puede considerar que se aproxima a la distribución normal? ¿Es muy alta la aproximación?, ¿satisfactoria o decididamente deficiente?
- ▶ Acerca de las preguntas anteriores, y habiendo anotado ya el resultado de la prueba de bondad de ajuste, ¿Cuál fue la conclusión proporcionada? ¿A que nivel de significancia? ¿Coincide con la apreciación hecha a priori? En caso negativo, ¿A que puede atribuirse?

**Conclusión:**

Desde 1733, Abraham deMoivre estableció que cuando el número de sumandos en una suma de variables aleatorias Bernoulli con parámetro  $p = 1/2$  -modelo asociado desde los inicios de la probabilidad con el fenómeno del número de 'águilas' observadas en N lanzamientos de una moneda legal- su función de distribución tendía a la distribución normal. En la práctica significa que es posible calcular de manera bastante aproximada la probabilidad de que el número de éxitos sea igual o inferior a un número dado

haciendo uso de la función de distribución normal. La simulación efectuada debió confirmar de manera empírica tan histórico resultado, presentado y demostrado en el teorema 1.2 de la sección 2, capítulo 1.

**\* Actividad (b):**

Simular: "1.Automática: Variación de N", "1.bernoulli"

- ▶ ¿Cuál es el valor de  $p$ ?, ¿Cuál es el número de simulaciones?
- ▶ ¿Qué valores va tomando  $N$ ?,  
▶ Conforme aumenta  $N$ , ¿Cuántos "escalones" de la función de distribución empírica aparecen?, ¿A que se debe?
- ▶ ¿Son equidistantes estos "escalones"?, ¿Por qué?
- ▶ Nuevamente, antes de observar el resultado de la prueba de Anderson-Darling, ¿Cuál sería la apreciación visual acerca de la convergencia a la distribución normal?
- ▶ Ahora, ¿Coincide la apreciación subjetiva con la prueba estadística? ¿Debió corregirse?
- ▶ ¿Cómo se va dando la aproximación a la normal al variar  $N$ ?

**Conclusión:**

Generalizando, de acuerdo al Teorema de deMoivre-Laplace, para una suma estandarizada de variables Bernoulli con parámetro  $p$  fijo, al tender el número de sumandos  $N$  a infinito, la probabilidad de que la suma sea menor o igual a cierto valor puede evaluarse con la función de distribución normal calculada en ese mismo punto; esto debe reflejarse en las simulaciones observando que la distribución empírica y la distribución normal acumulativa estándar tienden a aproximarse a medida que crece  $N$ , ¿Ha ocurrido eso?. Nuevamente, la respuesta la da el teorema 1.2.

**\* Actividad (c):**

Simular: "3.Automática: Varian Parámetros", "1.bernoulli"

- ▶ ¿Cuál es el valor de  $N$ ?, ¿Cuántas simulaciones se hacen para construir la distribución empírica?
- ▶ ¿Qué valores va tomando el parámetro  $p$ ?, ¿Hacia que valor tiende?
- ▶ A medida que cambia el valor de  $p$ , ¿Varia la conclusión acerca de la convergencia a la distribución normal? Reflexionando un poco, ¿Qué fenómeno podría ejemplificar este caso en que el parámetro  $p$  tiende a cero?
- ▶ Recuerde: trate de establecer su propia conclusión antes de conocer el resultado de la prueba de bondad de ajuste. Si en algún caso no se ha dado la aproximación a la normal, ¿Cómo puede explicarse la diferencia? ¿Está presente una discrepancia en los extremos? ¿Puede considerarse determinante una diferencia en la parte central?
- ▶ ¿Qué comportamiento general podría deducirse?

### Conclusión:

Si bien el teorema de deMoivre-Laplace asegura la convergencia a la distribución normal de sumas estandarizadas de variables aleatorias Bernoulli con  $p$  fijo, la manera en que se da la aproximación no es uniforme, como lo asegura el Teorema de Berry-Esseen, por lo que con una  $N$  en particular, esta aproximación puede ser satisfactoria para algunos valores de  $p$  y no serlo para otros. Como comportamiento general, para valores de  $p$  cercanos a cero (o a uno, por simetría) la convergencia a la distribución normal es cada vez más lenta. Por ello, para el caso en que la máxima discrepancia esté dentro de límites aceptables signifique un tamaño de  $N$  muy grande, se cuenta como alternativa con la aproximación de la distribución Poisson a la binomial que pide que  $np \rightarrow \lambda$  si  $n$  tiende a infinito, por lo que en la práctica, si se tiene una  $p$  pequeña, con una  $N$  grande, o incluso moderada, se puede aproximar la distribución de la suma (no estandarizada) por medio de esa ley. Es este otro ejemplo de convergencia en un ámbito más amplio, que corresponde a las leyes infinitamente divisibles y que puede consultarse en la sección 3 del capítulo 2; en particular puede verse el teorema 2.9 para este resultado, que es una situación que corresponde a fenómenos en que, por ser  $p$  cercano a cero, el interés se centra en observar resultados difíciles de ocurrir, por lo que suelen denominarse de "eventos raros".

→ Práctica Número 2. Teorema Central del Límite para variables aleatorias independientes con Varianza Finita.

#### Objetivos:

Reconocer que por el Teorema Central del Límite Clásico la convergencia de sumas de variables aleatorias a la distribución normal se aplica a sucesiones de variables aleatorias independientes idénticamente distribuidas con varianza finita.

Reafirmar el concepto de rapidez de convergencia y constatar que varía de acuerdo al tipo de variable que se trata.

#### \* Actividad (a):

Simular: "2. Automática: Varía Distribución",  
"1. Discretas con Varianza Finita"

- ▶ ¿Cuáles son las variables aleatorias consideradas en esta opción?
- ▶ ¿Qué características distinguen a las distribuciones simuladas en este grupo?
- ▶ ¿Se puede considerar que la función de distribución empírica se aproxima a la distribución normal estándar? ¿Visual y estadísticamente? ¿En todos los casos?
- ▶ ¿Es esta actividad una generalización de la Práctica No. 1? ¿En qué sentido?
- ▶ ¿Qué se puede concluir?

### Conclusión:

Todas las variables aleatorias consideradas en esta actividad son discretas con varianza finita y las simulaciones realizadas debieron dar como resultado una aproximación satisfactoria a la normal estándar, por lo que la convergencia a esta distribución puede extenderse, hasta ahora, de la distribución Bernoulli a la familia de las distribuciones discretas con segundo momento central finito.

### \* Actividad (b):

Similar: "2. Automática: Varía Distribución",  
"2. Continuas con Varianza Finita"

- ▶ ¿Qué variables fueron simuladas?
- ▶ Respecto a la actividad anterior, ¿Qué característica de las variables aleatorias cambió?
- ▶ ¿Cuáles son las conclusiones acerca de la aproximación a la distribución normal estándar? ¿Se sigue cumpliendo la convergencia? ¿Coinciden dichas conclusiones según los criterios visual y estadístico? ¿Se ha dado el peso correcto a las discrepancias en la parte central de las distribuciones y en los extremos?
- ▶ ¿Siguen siendo equidistantes los escalones de la distribución empírica? ¿Cuál es la explicación?
- ▶ ¿Cuanto vale la varianza de las distribuciones simuladas? ¿Es alguna de ellas infinita?

### Conclusión:

La experimentación sugerida en esta actividad conduce a extender la convergencia de sumas de variables aleatorias independientes idénticamente distribuidas a la normal estándar, al caso en que las variables involucradas sean continuas. Por ello debe esperarse que en todas las simulaciones realizadas, la aproximación de la función de distribución empírica a la distribución normal se halla cumplido.

Las conclusiones alcanzadas en las actividades (a) y (b) se deben a que la versión del Teorema Central del Límite asentada en el teorema 1.3 asegura que cuando se toman sumas estandarizadas de variables aleatorias independientes e idénticamente distribuidas, basta con que tengan varianza finita para que su distribución converja a la normal estándar, sin importar que otra característica compartan con la variable aleatoria normalmente distribuida, como puede ser la continuidad, la simetría o el dominio no acotado.

### \* Actividad (c):

Elaborar un cuadro comparativo de las simulaciones efectuadas en las actividades (a) y (b) para contestar:

- ▶ ¿Que distribuciones se simularon?
- ▶ Tipo de variables de que se trata, ¿Continuas o discretas?, ¿Simétricas o asimétricas?, ¿De dominio acotado o infinito?



- ▶ ¿Cuánto vale la varianza de las variables consideradas en la sucesión?
- ▶ Para cada una, ¿Cuánto vale el tercer momento absoluto?
- ▶ ¿Que valor tomó  $N$ ?, ¿Es el mismo para cada simulación?
- ▶ ¿Se aproxima la función de distribución empírica a la distribución normal en todos los casos?
- ▶ Considerando el nivel mínimo de aceptación para  $\alpha$ , la probabilidad de error tipo I, ¿Cuál se aproximó mejor y cual peor?. Ordena las variables simuladas de acuerdo a este criterio.

**Conclusión:**

Se ha constatado que para una gran variedad de variables aleatorias su suma estandarizada converge a la distribución normal; sin embargo, la velocidad con que se cumple esta convergencia depende del tercer momento absoluto de la variable simulada, tal como lo afirma la cota de Berry-Esseen, establecida en el teorema 1.4.

De acuerdo a las cotas anotadas para distintas distribuciones en los cuadros que aparecen en la sección 4 del primer capítulo, puede apreciarse que el factor más relevante para lograr una rápida convergencia es que la distribución sea de dominio acotado, sin importar demasiado si la función es o no continua, simétrica o asimétrica. Por el contrario, si el dominio de la función de distribución no está acotado, la velocidad con que se cumple la convergencia es menor, pues crece la cota de Berry-Esseen. ¿Han sido en ese sentido sus conclusiones en las simulaciones?.

+ Práctica Número 3. Sumas de variables aleatorias independientes con Colas Pesadas. Leyes Estables.

**Objetivo:**

Verificar que existen casos en que la distribución de sumas estandarizadas de variables aleatorias independientes e idénticamente distribuidas no se aproximan a la distribución normal y resaltar que se debe a la carencia de varianza finita.

Mostrar que la distribución normal pertenece a su propio dominio de atracción, considerada como una ley estable.

**\* Actividad (a):**

Simular: "2. Automáticas: Varía Distribución";  
"3. Con Colas Pesadas".

- ▶ En las dos distribuciones simuladas, ¿Se aproxima la función de distribución empírica a la distribución normal?
- ▶ ¿En que falla?, ¿Qué comportamiento presentan las simulaciones que sea distinto al esperado?
- ▶ ¿Que características tienen en común con la

distribución normal?, ¿Son continuas, simétricas o de dominio infinito?

- ▶ ¿Cuanto vale la varianza en cada caso?
- ▶ ¿A que se puede atribuir la no convergencia a la distribución normal?

#### Conclusión:

En efecto, existen sumas estandarizadas de variables aleatorias independientes e idénticamente distribuidas que no convergen a la distribución normal. Ello se debe a que la existencia de una varianza finita no sólo es una condición suficiente, sino que también es necesaria, como se afirma en el teorema 2.11. Cuando este atributo no está presente, la función de distribución empírica muestra valores más grandes en términos absolutos que los correspondientes a la distribución normal, esto es, los "escalones" de la distribución empírica, corregida por factores de localización y escala, empiezan antes y terminan después que el intervalo seleccionado de  $\pm 4$  desviaciones estándar (existe carencia de ajuste en las colas). Este fenómeno se presenta cuando se trata de variables aleatorias con "colas pesadas" y de lugar, como generalización, al estudio de las distribuciones estables, que es la única familia a donde pueden converger sumas estandarizadas de este tipo de variables, que no consideran necesariamente la condición de varianza finita (vea la definición 2.1).

\* Actividad (b):

Simular: "1. Automática: Variación de N", "B. normal"

- ▶ ¿Que valores de N se consideran? ¿Influyen en la convergencia?
- ▶ ¿Se cumple la convergencia a la distribución normal? Visualmente, ¿Fue la misma apreciación?
- ▶ ¿Cómo explica que no coincida la distribución empírica exactamente con la curva suave si se trata de la distribución normal en ambos casos?
- ▶ ¿Que tan buena es la aproximación observada?
- ▶ Intuitivamente, ¿Es lógico este resultado?

#### Conclusión:

Es sabido que la suma finita de variables aleatorias normalmente distribuidas es también normal, por lo que, con la adecuada estandarización, su distribución es la normal con media cero y varianza uno, lo cual es mucho más que sólo afirmar que converge a ella. Sin embargo, existen otros argumentos, pues cabe señalar que por contar la misma distribución normal con varianza finita, el Teorema Central del Límite en la versión de la sección 3 del capítulo 1, asegura que la suma estandarizada de una sucesión de variables aleatorias independientes distribuidas de acuerdo a la ley normal, convergen asimismo a la normal estándar. Refleja además, el hecho de que toda distribución estable pertenece a su propio dominio de atracción, como lo afirma la proposición 2.1, puesto que la misma distribución normal es un ejemplo de ley estable, y por lo tanto, se asegura de nueva cuenta que a ella

converge una suma estandarizada de variables aleatorias normales.

+ Práctica Número 4. Sumas Aleatorias de Variables Aleatorias

Objetivo:

Verificar que la convergencia a la distribución normal de sumas estandarizadas se sigue cumpliendo aún cuando el número de sumandos también sea aleatorio.

\* Actividad (a):

Simular:

"4. Automática: Sumas Aleatorias", "3. geométrica"  
"4. Automática: Sumas Aleatorias", "6. exponencial"

- ▶ Observe que aparece un número de variable aleatoria;
- ▶ ¿Coincide ese número con el valor de  $N$ ?
- ▶ ¿Cómo es el comportamiento de  $N$ ? ¿Estrictamente creciente?
- ▶ ¿Termina cumpliéndose la convergencia a la distribución normal estándar? (Continúe ejercitando su capacidad de observación: ¿Llega por su cuenta a la misma conclusión que la prueba de Anderson-Darling? ¿Coincide el grado de su conclusión con el nivel de significancia expuesto por la prueba?
- ▶ ¿Podría decirse que ante la presencia de un número aleatorio de sumandos persiste la convergencia a la normal estándar? ¿Debería imponerse alguna restricción en esa aleatoriedad?

Conclusión:

Considerando que se toman sumas estandarizadas de variables aleatorias independientes idénticamente distribuidas con varianzas finitas, donde el número de sumandos es escogido aleatoriamente de acuerdo a la distribución de una sucesión de variables que sólo toman valores positivos, y que la probabilidad de tomar valores mayores a un número fijo, tiende a uno si el número de variables consideradas tiende a infinito. Entonces, bajo estas condiciones se puede asegurar (véase el teorema 1.6) que la convergencia de las sumas así formadas se realiza a la distribución normal estándar. Esto puede explicarse porque si bien el comportamiento del número de componentes en la suma es errático, finalmente termina por tender a infinito para lograr la aproximación deseada. ¿Las simulaciones presentaron evidencia en ese sentido? ¿Fueron sus conclusiones correctas? ¿Intuyó la necesidad de la condición impuesta al patrón aleatorio del número de sumandos (las variables aleatorias con valores positivos en el paquete son uniformes discretas con valores de 0 a  $T$ , con  $T$  aumentando en cada simulación)?.

+ Práctica Número 5. Cadenas de Markov.

**Objetivo:**

Mostrar que en algunos casos, la convergencia a la distribución normal sigue cumpliéndose cuando las variables aleatorias no son ya independientes. Ejemplificar cuando se trata de Cadenas de Markov de dos Estados.

\* Actividad (a):

Simular: "1. Automática: Variación de N",  
"15. cadenas de Markov"

- ▶ En una cadena de Markov de dos estados, ¿Cuántos valores puede tomar la variable aleatoria? ¿Qué probabilidades define la matriz de transición?
- ▶ Para cada valor de N, según su apreciación visual y la prueba estadística, ¿Se cumple la convergencia a la distribución normal? En términos de rapidez. ¿Podría considerarse satisfactoria?
- ▶ ¿Puede considerarse cierto el postulado de que la distribución normal puede aproximar a la distribución de la suma estandarizada de variables aleatorias, aunque sean dependientes?

**Conclusión:**

La convergencia de sumas estandarizadas de variables aleatorias que forman una cadena de Markov a la distribución normal se cumple en el caso general, cuando se tienen n estados. En el caso particular de dos estados, por supuesto se sigue verificando la convergencia, como lo asegura el teorema 3.1, incluso con una alta rapidez. Si las probabilidades de transición de éxito a fracaso y viceversa suman uno, se rompe la dependencia y se reduce al caso de las sucesiones de ensayos Bernoulli. Por ello, se puede considerar de alguna manera este ejemplo como una generalización del teorema de deMoivre-Laplace, que ilustra además la convergencia a la distribución normal en el caso de variables aleatorias dependientes.

→ Práctica Número 6. Procesos Estacionarios: Medias Móviles.

**Objetivo:**

Mostrar que para los procesos estacionarios de medias móviles los cuales representan una aplicación muy extendida en el estudio de series aleatorias dependientes se verifica la convergencia a la distribución normal.

\* Actividad (a):

Simular:

"1. Automática: Variación de N".  
"15. medias móviles"

- ▶ ¿Cuál es la distribución de las variables aleatorias que forman el ruido blanco según se especifica en la presentación de la simulación? ¿Cómo puede denotarse esta sucesión de ruido blanco con un índice que las ordene (con frecuencia es un índice asociado con el tiempo)?
- ▶ Si se forma una nueva sucesión con cada elemento expresado como una combinación lineal de observaciones sucesivas del ruido blanco desde aquella que le corresponde el mismo índice y hacia atrás hasta considerar un número de variables igual a un valor denotado como orden + 1, ¿Cómo se puede expresar mediante símbolos matemáticos?
- ▶ Considerando que la sucesión de ruido blanco está formada por variables aleatorias independientes, ¿La nueva sucesión sigue siendo independiente?
- ▶ ¿Reconoce el término de 'medias móviles'? ¿Podría asociarse con la sucesión formada a partir del ruido blanco?
- ▶ ¿En qué área de la estadística se utiliza? ¿Qué aplicaciones tiene?
- ▶ ¿Qué significa que un proceso sea estacionario? ¿Y que sea ergódico?
- ▶ ¿Es un proceso de medias móviles estacionario y ergódico?
- ▶ Al realizar cada simulación anote los valores de M, N, el orden del proceso y el valor de los coeficientes ¿Cuáles cambian y cuáles se mantienen fijos?
- ▶ ¿Cuál es la conclusión en cada simulación? ¿Se mejora la aproximación a la distribución normal si N aumenta? ¿Puede decirse que converge a ella si  $n \rightarrow \infty$ ?
- ▶ ¿Qué aplicación puede tener este resultado?

**Conclusión:**

El proceso simulado se conoce como de medias móviles y puede expresarse de la forma siguiente:

$$x(n) - \mu = \sum_{j=0}^{\infty} \beta(j) \varepsilon(n-j) \quad \sum_{j=0}^{\infty} \beta^2(j) < \infty$$

y cumple con

$$\beta(0) = 1 \quad E [z(n)z(m)] = 0 \quad m \neq n$$

$$E [z(n)] = 0 \quad y \quad \text{Var}[z(n)] = \sigma^2$$

por lo que es débilmente estacionario.

Es un proceso que encuentra su utilidad en el área de las Series de Tiempo y cuya aplicación está muy difundida, en Economía y Meteorología, por ejemplo.

La simulación realizada considera en particular el caso más común de suponer normalidad sobre el ruido blanco, aunque no es esencial para establecer la convergencia a la distribución normal:

$$\sum_{j=1}^n x(j) \xrightarrow{d} N(0, 2nf(0))$$

con

$$f(\lambda) = \frac{\sigma^2}{\sqrt{2\pi}} \left[ \sum_{j=-\infty}^{\infty} \beta(j) e^{i\lambda j} \right]^2$$

la densidad espectral de  $\{x(n)\}$ , que debe ser continua en  $\lambda = 0$  y  $f(0) > 0$ . Consúltese el corolario 3.4.

+ Práctica Número 7. Dependencia en el valor de la suma de los  $m(n)$  primeros términos.

Objetivo:

Observar que en casos de dependencia entre las variables aleatorias existen condiciones en que se cumple la convergencia a la distribución normal, así como otras en que esto no ocurre.

\* Actividad (a):

Simular: "5.Control Completo de Parámetros.", "17.dependencia de los primeros  $\sqrt{N}$  terminos"  
"Valor de  $N$  (5-1000)? " 50  
"Simulaciones, (5-1000)? " 50

- ▶ Por la gráfica de presentación, ¿Qué tipo de distribución tienen las variables aleatorias simuladas? ¿Son continuas o discretas? ¿Qué valores toman?, ¿Son simétricas respecto al origen?
- ▶ En la misma presentación se anota un criterio de decisión sobre el número de componentes en la suma: Si  $X_1, X_2, \dots, X_{2N}$  son las variables aleatorias simuladas, ¿Cómo podría expresarse esta condición con ayuda de la función indicadora?

- ▶ Después de haber impuesto la condición establecida, ¿Siguen siendo independientes las variables aleatorias?. En caso negativo, ¿Cómo se da la dependencia? ¿Qué variables dependen de cuáles?
- ▶ ¿Es entonces constante el número de sumandos?
- ▶ Tomando en cuenta la pregunta anterior, para estandarizar la suma de las variables aleatorias, ¿Podría suponerse constante la forma de estandarizar? ¿Porqué?
- ▶ Una vez graficada la distribución teórica de la normal y la función de distribución empírica, ¿Se puede considerar que se aproxima a la distribución normal?
- ▶ ¿Coincide la apreciación hecha a priori con la conclusión de la prueba de Anderson-Darling?
- ▶ Este caso de convergencia, ¿Qué tipo de condiciones ejemplifica?

**Conclusión:**

Este es un caso particular de sumas de variables aleatorias dependientes en que se sigue cumpliendo la convergencia a la distribución normal pues cumple los requisitos del teorema 3.19. La simulación puede sintetizarse de la manera siguiente: Se consideran variables aleatorias idénticamente distribuidas que toman los valores  $\pm 1$  con probabilidad  $1/2$  cada uno, y se suman  $\lfloor \sqrt{N} \rfloor$  (los corchetes significan parte entera) de ellas; si la suma es mayor que cero, se continúa la suma hasta  $2N$  términos, pero si es menor o igual a cero, sólo se suman  $N$  términos. Representa un ejemplo de Martingala en que la relación de dependencia no muy fuerte, podría decirse- resultante de considerar la suma de  $N$  términos adicionales de acuerdo al signo de la suma de solamente los primeros  $\lfloor \sqrt{N} \rfloor$  sumandos, permite asegurar esta convergencia; además se requiere de una estandarización que depende del número de sumandos que hayan resultado de la simulación, siendo igual a la raíz del número de sumandos ( $\sqrt{N}$  o  $\sqrt{2N}$ ). Esto puede expresarse de la siguiente manera:

$$\frac{S_{N,2N}}{U_{N,2N}} \xrightarrow{d} N(0,1)$$

donde

$$S_{N,2N} = \frac{1}{\sqrt{N}} \sum_{i=1}^N Y_i + \frac{1}{\sqrt{N}} \sum_{i=N+1}^{2N} Y_i I \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^{\lfloor \sqrt{N} \rfloor} Y_i > 0 \right],$$

$$U_{N,2N} = 1 + I \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^{\lfloor \sqrt{N} \rfloor} Y_i > 0 \right]$$

$P(Y_i = \pm 1) = 1/2$        $I[\ ]$  es la función indicadora

**\* Actividad (b):**

Similar: "5.Control Completo de Parámetros" "18.dependencia de los primeros  $N/2$  términos"

"Valor de  $N$  (5-1000)? " 50

"Simulaciones (5-1000)? " 50

- ▶ ¿Cuál es la distribución de las variables aleatorias? ¿Es la misma que en la simulación de la actividad anterior?
- ▶ ¿Existe también en este caso una restricción al número de sumandos? ¿Cómo se modifica respecto al caso anterior?
- ▶ Si  $X_1, X_2, \dots, X_{2N}$  son nuevamente las variables aleatorias simuladas, ¿Cómo podría establecerse esta condición con la función indicadora?  
¿Se trata de un caso de variables aleatorias dependientes o independientes?
- ▶ Cuáles son ahora las conclusiones sobre la convergencia de la simulación a la distribución normal, visualmente y de acuerdo a la prueba de Anderson-Darling?
- ▶ Si los dos tipos de simulación son parecidos pero la conclusión cambia, ¿Cómo puede explicarse la importancia de la forma sobre la condición de dependencia?

**Conclusión:**

Se ha presentado un caso donde la convergencia a la distribución normal ya no se cumple, por lo que se evidencia que no es posible extenderla a cualquier situación de dependencia. En este ejemplo, las condición de dependencia es semejante a la de la actividad (a), aunque el factor determinante resulta ser que la restricción para extender la suma a  $2N$  componentes consiste en que la adición de los primeros  $N/2$  términos haya sido mayor que cero, mientras que en la actividad (a) donde sí se cumple la convergencia, la condición se establece en función de la suma de los primeros  $\sqrt{N}$  sumandos (más exactamente, de su parte entera). Entonces, cuando no se cumple la convergencia, el orden de  $N/2$  y de  $N$  es el mismo, mientras que cuando sí se satisface, la potencia de  $\sqrt{N}$  y  $N$  es  $1/2$  y  $1$ , respectivamente, lo que se traduce en que a medida que  $N$  crece la condición de dependencia va perdiendo importancia relativa. En general, si la condición depende de los primeros  $m(N)$  términos, se verifica la convergencia a la distribución normal solamente si  $m(N)/N$  tiende a cero, como se establece en el ejemplo que sigue al teorema 3.19 en el capítulo 3 de este trabajo. Este contraejemplo también está relacionado con las condiciones impuestas en el teorema 3.14.



#### 4.EVIDENCIA EMPIRICA.

En esta sección se resumen un conjunto de simulaciones que se realizaron con el paquete con el objeto de brindar elementos empíricos que confirmen las aseveraciones de los teoremas vistos en la teoría, den información adicional sobre la manera en que éstos se cumplen, y aporten indicios sobre otros posibles resultados que no se incluyen en este trabajo.

Se trata de simulaciones de sumas estandarizadas de variables aleatorias que cumple alguna condición que es especificada en el encabezado de cada uno de los cuadros, pudiendo cambiar los parámetros, las distribuciones o el número de sumandos, por ejemplo; sin embargo el número de repeticiones involucradas para construir la función de distribución empírica,  $M$ , siempre se mantiene igual a 50 por razones de comparación. A su vez, para observar el comportamiento al que tienden las simulaciones se realizaron 50 repeticiones bajo las mismas circunstancias, y se reporta el porcentaje observado en cada categoría en que es clasificada la simulación por la prueba de Anderson-Darling.

El orden de presentación sigue en líneas generales el de los resultados teóricos comprendidos en los capítulos anteriores y el de las prácticas de la sección precedente.

Bernoulli,  $p = 1/2$ .

N	$\alpha$ :	<.25	<.15	<.1	<.05	<.025	<.01
10		28	28	18	16	8	2
50		78	8	6	4	0	4
75		72	12	8	4	2	2
100		70	8	10	4	4	4
500		74	10	8	6	2	0
1000		88	8	4	0	0	0

Bernoulli.

p	N	$\alpha$ :	<.25	<.15	<.1	<.05	<.025	<.01
.2	10		0	12	6	38	32	12
	50		50	18	14	6	6	6
	100		64	14	4	12	2	4
.1	10		0	0	0	0	20	80
	50		56	16	18	6	0	4
	100		60	10	16	6	4	4
.05	10		0	0	0	0	0	100
	50		12	30	24	14	10	10
	100		50	20	14	10	6	0
.01	10		0	0	0	0	0	100
	50		0	0	0	0	0	100
	100		0	0	0	0	26	74

**VARIABLES ALEATORIAS CON VARIANZA FINITA.**

distribución	N	$\alpha$ : <.25	<.15	<.1	<.05	<.025	<.01
Uniforme	10	68	8	12	10	2	0
Discreta	50	78	6	4	4	4	4
rango 0,10	100	80	4	10	2	2	2
Geométrica	10	26	24	16	16	8	10
$p=1/4$	50	64	14	8	6	6	2
	100	60	12	6	14	4	4
Geométrica	10	74	8	6	6	2	4
$p=1/2$	50	76	18	0	4	2	0
	100	66	18	4	4	4	4
Geométrica	10	74	16	4	0	4	2
$p=3/4$	50	72	10	8	4	2	4
	100	64	18	6	6	2	4
Poisson	10	46	10	12	10	12	10
$\lambda=3$	50	64	8	2	12	8	6
	100	60	16	4	12	2	6
Uniforme	10	78	10	4	4	2	2
Continua	50	78	10	0	6	2	4
	100	70	12	8	10	0	0
Exponencial	10	70	18	2	6	2	2
media=1	50	74	10	12	2	0	2
	100	70	12	6	2	4	6
t	10	62	16	6	4	6	6
$g.l.= 4$	50	74	8	4	6	2	6
	100	68	8	6	10	2	6
t	10	70	8	6	8	6	2
$g.l.= 3$	50	68	16	4	4	6	2
	100	76	6	4	6	6	2
$\chi^2$	10	14	12	32	16	8	18
$g.l.= 2$	50	22	16	12	24	16	10
	100	22	10	28	16	18	6

**VARIABLES ALEATORIAS CON ALTA DISPERSION.**

distribución	N	$\alpha$ : <.25	<.15	<.1	<.05	<.025	<.01
t	10	0	0	0	0	0	100
$g.l.= 2$	50	0	0	0	0	0	100
	100	0	0	0	0	0	100
Cauchy	10	0	0	0	0	0	100
	50	0	0	0	0	0	100
	100	0	0	0	0	0	100

Cadenas de Markov.  
 Porcentajes de aceptación para  $\alpha \geq .05$ .  
 N = 10

P(F→E) \ P(E→F)	.1	.4	.5	.6	.9
.1	32				
.4	30	86			
.5	48	90	80		
.6	26	96	84	78	
.9	0	44	18	0	0

N = 50

P(F→E) \ P(E→F)	.1	.4	.5	.6	.9
.1	88				
.4	88	96			
.5	92	96	100		
.6	90	90	98	98	
.9	90	90	90	94	60

N = 100

P(F→E) \ P(E→F)	.1	.4	.5	.6	.9
.1	88				
.4	94	98			
.5	90	96	96		
.6	94	90	94	88	
.9	92	90	92	94	82

Medias Móviles.

distribución	N	$\alpha: <.25$	$<.15$	$<.1$	$<.05$	$<.025$	$<.01$
MA(2)	10	56	16	4	14	4	6
$\theta = .5$	50	70	10	12	6	0	2
$\theta = -.1$	100	80	14	4	0	2	0

**Martingalas.**  
**Dependencia de los primeros sumandos.**

distribución	N	$\alpha: <.25$	$<.15$	$<.1$	$<.05$	$<.025$	$<.01$
$\sqrt{N}$ Primeros	10	52	18	4	12	2	12
	50	42	12	8	6	6	6
	100	74	8	10	6	0	2
	300	78	6	8	4	2	2
	500	70	14	4	4	4	2
N Primeros	10	18	24	14	22	10	12
	50	40	14	14	16	10	6
	100	52	14	4	12	6	12
	300	38	20	8	14	8	12
	500	38	18	4	18	2	20

**5. VARIABLES ALEATORIAS SIMULADAS  
 POR EL PAQUETE DE COMPUTO.**

Como rápida referencia al manejar el paquete, se enlistan las características sobresalientes de las variables simuladas. Se presentan en el orden en que aparecen en los menús.

**5.1 Distribución Bernoulli.**

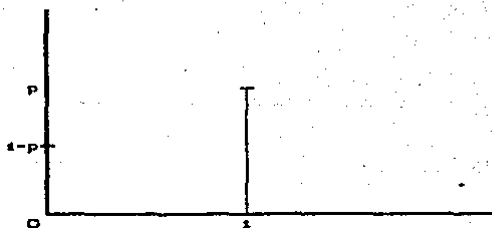
Función de probabilidad:

$$p(x) = \begin{cases} p^x(1-p)^{1-x} & x = 0, 1 \\ 0 & \text{otro caso} \end{cases} \quad 0 \leq p \leq 1$$

Se trata de una función discreta que sólo toma dos valores, por lo general es asimétrica (siempre que  $p \neq 1/2$ ). Corresponde a un experimento donde se distingue únicamente entre caso favorable o adverso, con probabilidades  $p$  y  $1-p$ , respectivamente.

$$\mu = p$$

$$\sigma^2 = p(1-p)$$



### 5.2 Distribución Uniforme Discreta.

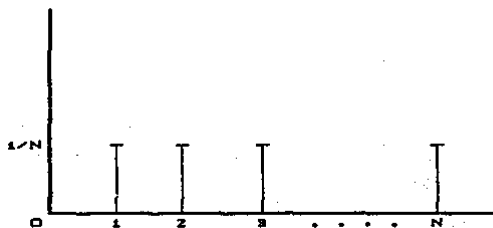
Función de probabilidad:

$$p(x) = \begin{cases} 1/N & x = 0, 1, \dots, N \\ 0 & \text{otro caso} \end{cases} \quad N \in \mathbb{Z}^+$$

Se trata de una función discreta que toma  $N$  valores con igual probabilidad, y es simétrica respecto a su media. Representa fenómenos donde se reconocen  $N$  clases, todas con igual probabilidad de ocurrir (completa incertidumbre).

$$\mu = \frac{N + 1}{2}$$

$$\sigma^2 = \frac{N^2 - 1}{12}$$



### 5.3 Distribución Geométrica.

Función de probabilidad:

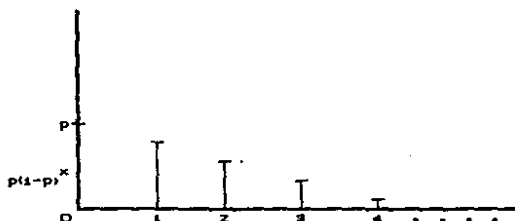
$$p(x) = \begin{cases} p(1-p)^x & x = 0, 1, \dots \\ 0 & \text{otro caso} \end{cases} \quad 0 \leq p \leq 1$$

Es una función discreta que toma valores en los números

enteros positivos con probabilidad decreciente que puede interpretarse como la variable aleatoria: número de éxitos que ocurren en una sucesión de ensayos Bernoulli antes de ocurrir el primer fracaso; es una función asimétrica.

$$\mu = \frac{1-p}{p}$$

$$\sigma^2 = \frac{1-p}{p^2}$$



#### 5.4 Distribución Binomial.

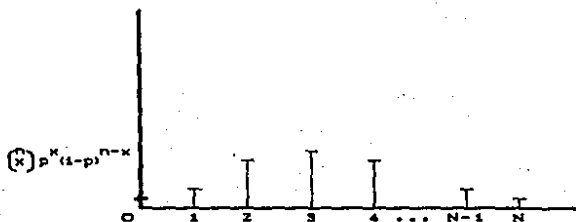
Función de probabilidad:

$$p(x) = \begin{cases} \binom{N}{x} p^x (1-p)^{N-x} & x = 0, 1, \dots, N \\ 0 & \text{otro caso} \end{cases} \quad 0 \leq p \leq 1$$

También se trata de una función discreta, toma valores distintos de cero en los enteros entre cero y N, y representa a la variable aleatoria número de éxitos en N ensayos independientes de Bernoulli con probabilidad p fija. Excepto que p sea igual a 1/2, la función es asimétrica.

$$\mu = N p$$

$$\sigma^2 = N p (1-p)$$



### 5.5 Distribución Poisson.

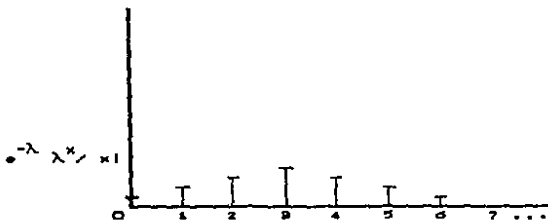
Función de probabilidad:

$$p(x) = \begin{cases} e^{-\lambda} \lambda^x / x! & x = 0, 1, 2, \dots \\ 0 & \text{otro caso} \end{cases} \quad \lambda > 0$$

Se trata de la última variable discreta considerada en el paquete, que asume valores distintos de cero para los enteros positivos, por lo que resulta adecuada para modelar fenómenos de conteo, hallando gran aplicación en el área de los Procesos Estocásticos y en donde está sumamente ligada a la distribución exponencial. La función de probabilidad es asimétrica y, después de alcanzar su máximo, tiende a cero a medida que  $x$  tiende a infinito. Tiene la peculiaridad para todos los miembros de la familia de que los dos primeros momentos coinciden con el parámetro de la distribución.

$$\mu = \lambda$$

$$\sigma^2 = \lambda$$



### 5.6 Distribución Exponencial.

Función de densidad:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otro caso} \end{cases} \quad \lambda > 0$$

Esta variable aleatoria es la primera continua de las consideradas en el paquete, su recorrido se restringe a valores positivos y se le asocia a fenómenos de tiempos de vida, o también, al tiempo transcurrido entre la ocurrencia de dos eventos sucesivos en un fenómeno regido por la distribución Poisson. La densidad tiene un comportamiento exponencial negativo por lo que decrece desde el valor  $\lambda$  en  $x = 0$  hasta 0 cuando  $x$  tiende a infinito.

$$\mu = 1/\lambda$$

$$\sigma^2 = 1/\lambda^2$$



### 5.7 Distribución Uniforme Continua.

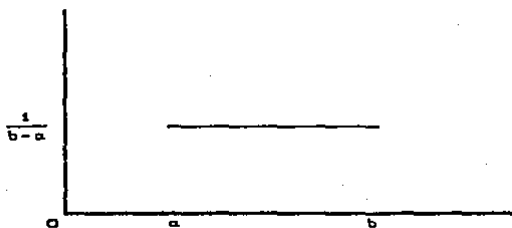
Función de densidad:

$$f(x) = \begin{cases} 1/(b-a) & -\infty < a < x < b < \infty \\ 0 & \text{otro caso} \end{cases}$$

Es una variable aleatoria continua muy sencilla, que toma un valor constante dentro de un intervalo finito, por lo que también es simétrica respecto a su media. En el caso de que el intervalo sea el unitario [0,1], sirve de base para la generación de números aleatorios de una diversidad de variables tanto de tipo continuo como discreto; se verifica además que la media y la varianza valen 1/2 y 1/12, respectivamente.

$$\mu = (a + b) / 2$$

$$\sigma^2 = (b - a)^2 / 12$$





### 5.8 Distribución Normal.

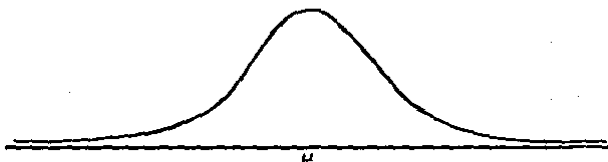
Función de densidad:

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-(x-\mu)^2 / 2\sigma^2} \quad -\infty < \mu < \infty, \quad \sigma^2 > 0$$

Esta variable aleatoria continua juega un papel central en la Estadística, pues en ella descansa la teoría clásica; ello se debe en buena medida a que a ella converge la distribución de sumas de variables aleatorias, resultado que aborda este paquete. Se trata de una variable de dominio no acotado, unimodal y simétrica respecto a su media.

$$\mu = \mu$$

$$\sigma^2 = \sigma^2$$



### 5.9 Distribución Ji-Cuadrada.

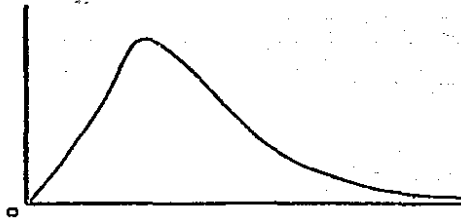
Función de densidad:

$$f(x) = \begin{cases} \frac{1}{2^{k/2} \Gamma(k/2)} e^{-x/2} x^{k/2-1} & x > 0 \\ 0 & \text{otro caso.} \end{cases} \quad k = 1, 2, \dots$$

Esta distribución, también conocida como  $\chi^2$ , donde el parámetro  $k$  se denomina grados de libertad, es una variable aleatoria continua que pertenece a la familia de distribuciones Gamma que está íntimamente ligada con la distribución normal pues la suma de cuadrados de variables aleatorias normales estandarizadas se distribuye como una  $\chi^2$ . Es una variable asimétrica, unimodal y definida exclusivamente para los valores positivos.

$$\mu = k$$

$$\sigma^2 = 2k$$



### 5.10 Distribución T de Student.

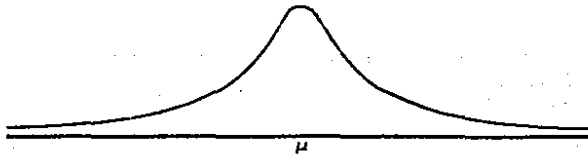
Función de densidad:

$$f(x) = \frac{\Gamma((k+1)/2)}{\Gamma(k/2) \sqrt{k\pi} (1+x^2/k)^{(k+1)/2}} \quad k=1,2,\dots$$

Dentro de esta distribución, a  $k$  también se le nombra 'grados de libertad'; es otra variable aleatoria continua relacionada con la distribución normal pues el cociente de una normal estándar y la raíz de una ji-cuadrada dividida por sus grados de libertad es una distribución t de Student con los mismos grados de libertad que la distribución ji-cuadrada involucrada. Tiene la misma figura que la distribución normal, sólo que presenta una mayor dispersión; es por tanto simétrica y de dominio no acotado.

$$\mu = 0 \quad \text{si } k > 1$$

$$\sigma^2 = k/(k-2) \quad \text{si } k > 2$$



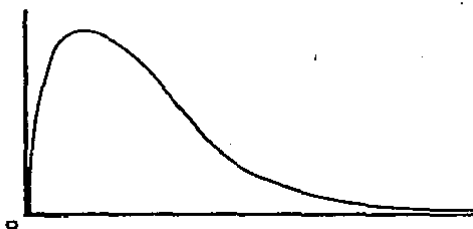
### 5.11 Distribución F.

Función de densidad:

$$f(x) = \begin{cases} \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \left[\frac{m}{n}\right]^{m/2} \frac{x^{(m-2)/2}}{[1+(m/n)x]^{(m+n)/2}} & x > 0 \\ 0 & \text{otro caso} \end{cases} \quad m, n = 1, 2, \dots$$

Para esta función, los parámetros  $m$  y  $n$  se conocen como grados de libertad en el numerador y denominador, respectivamente. Por estar relacionada también con la distribución normal via el cociente de dos distribuciones ji-cuadradas divididas por sus grados de libertad, es sumamente útil en Estadística. Su recorrido se restringe a los valores positivos y es asimétrica.  $\mu = n/(n-2)$  si  $n > 2$

$$\sigma^2 = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \quad \text{si } n > 4$$



### 5.12 Distribución Rayleigh.

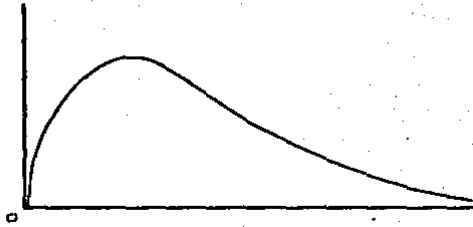
Función de densidad:

$$f(x) = \begin{cases} x/\beta^2 e^{-x^2/2\beta^2} & x > 0 \\ 0 & \text{otro caso} \end{cases} \quad \beta > 0$$

Es una variable aleatoria continua asimétrica y unimodal. Brinda un método para simular números aleatorios con distribución normal; en ingeniería eléctrica resulta de utilidad pues puede representar la distancia al origen de un punto en el plano cuyas coordenadas están determinadas por dos variables independientes cuya distribución es normal con media cero e igual varianza.

$$\mu = \beta \sqrt{\pi/2}$$

$$\sigma^2 = (4-\pi) \beta^2/2$$



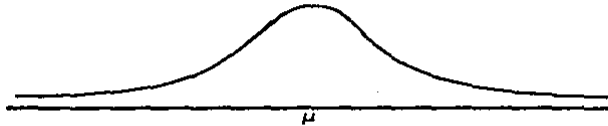
### 5.13 Distribución Cauchy.

Función de densidad:

$$f(x) = \frac{1}{\pi \beta (1+(x/\beta)^2)} \quad \beta > 0$$

Es una variable aleatoria continua, simétrica respecto al origen, de dominio no acotado, y tiene la peculiaridad de no tener media ni momentos de mayor orden; es un miembro de la familia de las leyes estables, por lo que se dice que tiene 'colas pesadas'.

$\mu$  y  $\sigma^2$  no existen



### 5.14 Cadena de Markov.

Una sucesión de variables aleatorias  $\xi_0, \xi_1, \xi_2, \dots$  se llama cadena de Markov de dos estados si para todo  $n$  y para  $x_i = 0, 1$   $i = 0, 1, 2, \dots, n$  se tiene

$$P(\xi_{n+1} = x_{n+1} | \xi_0 = x_0, \xi_1 = x_1, \dots, \xi_n = x_n) = P(\xi_{n+1} = x_{n+1} | \xi_n = x_n)$$

este es el ejemplo de variables aleatorias dependientes que maneja el paquete, tratándose de un caso discreto. Se puede interpretar como un sistema físico en que se distinguen dos estados posibles  $A_0$  y  $A_1$ , observándose el estado del sistema en el instante  $t=0, 1, 2, \dots$  donde los cambios de estado se registran aleatoriamente obedeciendo la siguiente regla: 'los estados pasados de la cadena influyen en los estados futuros únicamente a través del estado presente'.

A  $P(\xi_n = x_n | \xi_m = x_m)$  ( $m < n$ ) se le llama la probabilidad de transición en  $n-m$  etapas, y si no depende de  $n$  ni de  $m$ , sino sólo de  $n-m$  se dice que la cadena es homogénea. En ese caso basta definir las probabilidades iniciales  $P(\xi_0 = 0)$  y  $P(\xi_0 = 1)$  y las de transición en una etapa, que usualmente se denotan en forma matricial

$$\begin{bmatrix} 1-\lambda & \lambda \\ \mu & 1-\mu \end{bmatrix}$$

donde

$$\lambda = P(\xi_{n+1} = 1 | \xi_n = 0) \quad \text{y} \quad \mu = P(\xi_{n+1} = 0 | \xi_n = 1)$$

## CONCLUSIONES.

El trabajo de recopilación resulta ser laborioso pero no deja de ser importante. La colección de resultados aquí presentada es un conjunto valioso que tiene una gran amplitud, pues abarca desde los teoremas clásicos hasta versiones recientes de gran aplicabilidad en distintas ramas de la Estadística. La bibliografía consultada comprende incluso títulos de reciente aparición.

Este material es recomendable por lo tanto a estudiosos en la materia que deseen fundamentar su trabajo o analizar las condiciones que se están tratando.

Por el esfuerzo de plasmarlo en un lenguaje claro, que enfatiza en la interpretación de las condiciones establecidas en los teoremas, por incluir varios ejemplos de cada tipo de sucesión de variables aleatorias considerado, así como por incorporar algunos dibujos explicativos, no sólo es un material de consulta sino un apoyo directo en el aprendizaje de conceptos relacionados con el tema tales como leyes estables, infinitamente divisibles, condiciones mezclantes o martingalas, por mencionar algunos. Aunque, obviamente, se considera que este trabajo permite reconocer los supuestos requeridos para aplicar el Teorema Central del Límite, y distinguir las diferencias entre los distintos contextos en que es válido.

El paquete de cómputo merece una valoración aparte, pues se trata de un producto que explota adecuadamente un recurso tan importante como lo es la computadora, utilizando en provecho de una temática muy particular -la convergencia a la distribución normal- características propias de esta herramienta: velocidad de cálculo para ejecutar las simulaciones, y capacidad de graficación para presentar los resultados. Estas peculiaridades, junto con la estructura de manejo por menús, hacen del paquete un elemento de inmediata utilización que incrementa su explotación potencial.

Se tienen pruebas, aunque en pequeña escala, de que este material facilita la comprensión del Teorema Central del Límite y de la forma en que éste se verifica, acelerando además la etapa de asimilación.

En esta faceta, resulta ser de notable aportación la serie de prácticas que se sugieren para el uso del paquete, pues no únicamente orienta su explotación, sino que también la vincula con los resultados teóricos respectivos, racionalizando su utilización y graduando su aprovechamiento. Sobra decir que esta colección de prácticas no es exhaustiva y que pueden diseñarse específicamente algunas de acuerdo a la situación en que se maneja el paquete.

La posibilidad de concentrar gran cantidad de información mediante simulaciones convierte al paquete en una alternativa para contrastar los resultados teóricos manejados, e intuir otros. Se han presentado someramente los resultados de algunas simulaciones efectuadas con el paquete, donde se ha preferido no

incluir ahí conclusiones por no correr el riesgo de ser redundantes o pecar de parcos. Sin embargo, una revisión de los resultados anotados en los cuadros pone de manifiesto que la evidencia empírica viene a confirmar todas y cada una de las contrapartes teóricas, con sus matices específicos de cada caso.

Cabe anotar al respecto algunos comentarios de carácter general a manera de ilustración de las conclusiones que se pueden obtener de las simulaciones.

Los porcentajes presentados en la tabla correspondiente a la distribución Bernoulli con parámetro igual a  $1/2$ , confirman la validez del teorema de deMoivre-Laplace pues la convergencia a la distribución normal se acentúa en un mayor número de casos a medida que la cantidad de sumandos considerada aumenta; asimismo, se observa que el comportamiento no es completamente ascendente, lo cual además de reflejar el hecho de que se trata de una "muestra", señala que el tipo de límite involucrado es estocástico.

En el cuadro donde se varía el valor del parámetro para la misma distribución Bernoulli evidencia la generalización del teorema mencionado en el párrafo anterior, pero también resalta la dificultad de conseguir la convergencia cuando dicho parámetro es muy pequeño. Este análisis destaca la importancia de otro tipo de aproximaciones, como la que se consigue a través de la distribución Poisson. Este mismo comportamiento también sugiere el concepto de rapidez de convergencia.

El grupo de simulaciones en que se consideraron variables aleatorias independientes con varianza finita muestra que para todas ellas se verifica la convergencia a la distribución normal, lo que reafirma la necesidad y suficiencia de esta condición. Nuevamente, se intuye el concepto de rapidez de convergencia por las diferencias observadas entre las distintas simulaciones, y se recomienda compararlas con los cuadros presentados en la subsección 4.2 del primer capítulo correspondientes al cálculo de la cota de Berry-Esseen.

En oposición se presentan dos contraejemplos en que es irrefutable la no convergencia a la distribución normal: se trata de la distribución  $t$  de Student con dos grados de libertad, y la de Cauchy. La discrepancia se observa en los extremos de la distribución lo que resalta la importancia que tienen las "colas pesadas" para impedir el cumplimiento del Teorema Central del Límite.

Respecto a los cuadros sobre cadenas de Markov, aparte de observar el predecible efecto que tiene aumentar el número de sumandos, puede observarse que existe dificultad para conseguir la convergencia deseada en los casos localizados en las esquinas de las tablas, y que corresponden a situaciones en que las probabilidades de cambio de estado toman valores extremos, por lo que se puede sospechar una problemática análoga a la de la distribución Bernoulli cuando  $p$  tiende a cero o a uno.

Para el ejemplo presentado de medias móviles, se observa un buen comportamiento, incluso para valores pequeños de  $N$ , el número de sumandos.

Finalmente, sobre la tabla que trata ejemplos de martingalas, puede observarse que no hay dudas en que se verifica el Teorema Central del Límite cuando la dependencia está solamente en términos de los primeros  $\sqrt{N}$  sumandos, mientras que cuando está en función de los primeros  $N$  sumandos el comportamiento es sumamente errático, aunque no parece terminar por decidirse a converger a la distribución normal, o no hacerlo. Esta problemática es la que se aborda en la subsección 3.7 del tercer capítulo, y se sugiere su revisión para profundizar en la discusión.

Innegablemente, el tema no está agotado; existen varias direcciones donde el trabajo puede continuarse; por ejemplo, prestar mayor atención al concepto de rapidez de convergencia, de gran relevancia práctica, o estudiar otros casos de dependencia.

En conclusión, se espera que este trabajo sea de utilidad para el investigador en el área como material de consulta o de experimentación para analizar ciertas condiciones; o bien, como un instrumento que coadyuve en el mejoramiento del proceso enseñanza-aprendizaje.



## APENDICE.

### Función Característica.

A1.1 Definición. Sea  $X$  variable aleatoria y  $\varphi(x)$  una función cualquiera, se conoce como la función característica de  $\varphi(x)$

$$\phi_{\varphi(x)}(t) = E [ e^{it\varphi(x)} ]$$

donde  $E[*]$  es la función esperanza e  $i = \sqrt{-1}$ .

A1.2 Sea  $X \sim N(\mu, \sigma^2)$ , entonces

$$\phi_X(t) = \exp [ it\mu - t^2\sigma^2/2 ]$$

en particular, si  $X \sim N(0, 1)$

$$\phi_X(t) = e^{-t^2/2}$$

A1.3 Sea  $X$  variable aleatoria y  $a, b \in \mathbb{R}$ , entonces

$$\phi_{ax+b}(t) = e^{itb} \phi_X(at)$$

A1.4 Existe una relación uno a uno entre las funciones características y las funciones de distribución.

A1.5 Sea  $X$  variable aleatoria, entonces su función característica es uniformemente continua en  $t$ .

A1.6 Sea  $X$  variable aleatoria. Si  $E(X^k)$  existe, entonces  $\phi_X^{(j)}(t)$  existe y es continua para  $j=1,2,3,\dots,k$  y

$$\phi_X^{(j)}(t) = \sum_{j=0}^k \frac{i^j t^j E(X^j)}{j!} + h(t)$$

donde  $\lim_{t \rightarrow 0} \frac{h(t)}{t^k} = 0$

Además, si  $\phi_X^{(2k)}(0)$  existe, entonces  $E(X^j)$  existe para  $j=1,2,\dots,2k$ .

A1.7 Sea  $X$  variable aleatoria. Si  $E(X^k)$  existe, entonces

$$\phi_x^{(k)}(0) = i^k E(X^k)$$

A1.8 Sea  $X$  variable aleatoria, entonces  $|\phi_x(t)| \leq \phi_x(0) = 1$ .

A1.9 Teorema de Continuidad. Sea  $(F_n(x))$  una sucesión de funciones de distribución y  $(\phi_n(t))$  la correspondiente sucesión de funciones características. Entonces

1) Si  $F_n(x) \xrightarrow{d} F(x)$ , donde  $F(x)$  es una función de distribución,  $\Rightarrow \phi_n(t) \rightarrow \phi_x(t) \forall t \in \mathbb{R}$ , siendo  $\phi_x(t)$  la función característica de  $F(x)$ .

2) Si  $\lim_{n \rightarrow \infty} \phi_n(t)$  existe para cada  $t \in \mathbb{R}$  y es igual a  $\phi_x(t)$ , la cual es continua en  $t = 0$ , entonces  $\phi_x(t)$  es la función característica de una función de distribución  $F$  y  $F_n(x) \xrightarrow{d} F(x)$ .

Función Generadora.

A1.10 Definición. Sea  $X$  variable aleatoria y  $g(x)$  una función cualquiera, se conoce como la función generadora de  $g(x)$

$$\psi_{g(x)}(t) = E [ e^{tg(x)} ]$$

donde  $E[*]$  es la función esperanza.

A1.11 Sea  $X \sim N(\mu, \sigma^2)$ , entonces

$$\psi_x(t) = \exp [ t\mu + t^2\sigma^2/2 ]$$

en particular, si  $X \sim N(0, 1)$

$$\psi_x(t) = e^{t^2/2}$$

A1.12 Si existe la función generadora de una variable aleatoria, está unívocamente determinada.

A1.13 Sea  $X$  variable aleatoria. Si  $\psi_x(t)$  existe, entonces es continuamente diferenciable en alguna vecindad alrededor del origen y

$$\psi_x^{(k)}(0) = E(X^k) = \mu_k'$$

Orden de residuos.

A1.14 Definición. Una sucesión de variables reales  $a_n$  es un residuo de orden menor que  $b_n$ , y se denota por

$$a_n = o(b_n),$$

si

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0.$$

Y es de a lo más del mismo orden que  $b_n$  si existe un número real  $M$  tal que

$$\frac{|a_n|}{b_n} \leq M \quad \forall n$$

lo que se denota por

$$a_n = O(b_n)$$

Series de Taylor.

A1.15 Definición. Sea  $f(x)$  una función real de variable real, entonces

$$f(x) = \sum_{l=0}^{n-1} \frac{f^{(l)}(a) (x-a)^l}{l!} + r_n$$

donde  $f^{(i)}(\cdot)$  es la derivada  $i$ -ésima de  $f$ , y

$$r_n = \frac{f^{(n)}(\eta) (x-\eta)^n}{n!} \quad \text{con } \eta \in (\min [a, x], \max [a, x])$$

si  $\lim_{n \rightarrow \infty} r_n = 0$ , la expresión se conoce como Serie de Taylor de  $f$  en torno al valor  $a$ .

A1.16 Definición. En el punto anterior, si  $a = 0$ , la expansión se conoce como Serie de Maclaurin.

$$A1.17 \quad \ln(1+x) = \sum_{j=1}^{\infty} \frac{(-1)^{j+1} x^j}{j} \quad -1 < x \leq 1$$

A1.18 (Residuo integral) En el desarrollo de Taylor el resto

integral está dado por

$$r_n = \int_a^b \frac{(b-y)^n}{n!} f^{(n+1)}(y) dy$$

con  $f$  diferenciable en  $(a, b)$

Desigualdades.

A1.19 (Esseen). Sean  $F_x(x)$  y  $G_Y(x)$  funciones de distribución con funciones características  $\phi_x(t)$  y  $\phi_Y(t)$  respectivamente, tal que  $G_Y'(x)$  existe y  $\sup_x |G_Y'(x)| \leq c$ ,  $c \in \mathbb{R}$ . Entonces, para toda  $T > 0$

$$\sup_x |F_x(x) - G_Y(x)| \leq \frac{2}{\pi} \int_0^T \left| \frac{\phi_x(t) - \phi_Y(t)}{t} \right| dt + \frac{24}{\pi T} \sup_x |G_Y'(x)|$$

A1.20 (Lyapunov). Si  $0 < r < t$

$$[E(|X|^r)]^{1/r} \leq [E(|X|^t)]^{1/t}$$

A1.21  $|e^z - 1| \leq |z| e^{|z|}$ .

Regla de L'Hospital.

A1.22 Sea  $I$  un intervalo abierto (que puede ser semi-infinito), uno de cuyos extremos es  $a$ . Sean  $f$  y  $g$  funciones derivables en  $I$  para las cuales

1)  $g(x) \neq 0$ ,  $g'(x) \neq 0 \quad \forall x \in I$

$$2) \lim_{x \rightarrow a} \frac{f'(x)}{g'(x)} = L$$

3)  $f(x) \rightarrow 0$  y  $g(x) \rightarrow 0$  conforme  $x \rightarrow a$ , o bien

3')  $g(x) \rightarrow \pm \infty$  conforme  $x \rightarrow a$ .

$$\text{Entonces } \lim_{x \rightarrow a} \frac{f(x)}{g(x)} = L.$$

y donde todos los límites de este punto se deben considerar como límites laterales dentro de  $I$ .

A1.23 Si  $f(x)$  y  $p(x)$  son continuas en  $[a, b]$ , y además  $p(x)$  es positiva en ese intervalo, entonces existe  $c \in [a, b]$  tal que

$$\int_a^b f(x)p(x) dx = f(c) \int_a^b p(x) dx$$

**Función Analítica Gama.**

A2.1 La función Gama está definida por

$$\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt, \quad x > 0$$

teniendo, entre otras, las siguientes propiedades:

$$\Gamma(x) = (x-1) \Gamma(x-1) \quad \forall x > 1$$

$$\Gamma(x) = (x-1)! \quad x \in \mathbb{N}^+$$

$$\Gamma(1/2) = \sqrt{\pi}$$

**Variables Aleatorias Infinitesimales.**

A2.2 Las variables aleatorias en la doble sucesión  $(X_{nk})$   $n = 1, 2, \dots$   $k = 1, 2, \dots, k_n$  son infinitesimales si

$$\sup_{1 \leq i \leq k_n} P(X_{nk} \geq \epsilon) \rightarrow 0 \quad \text{si } n \rightarrow \infty \quad \forall \epsilon > 0$$

**Espacio  $L^2$ .**

A3.1 El espacio  $L^2 = L^2(\Omega, \mathcal{F}, P)$  es el espacio de clases equivalentes de variables aleatorias ( $X$  es equivalente con  $Y$  si  $X = Y$  casi seguramente) con segundos momentos finitos.

**Variación Lenta.**

A3.2 Una función  $h(x)$  positiva (no necesariamente monótona) definida en  $(0, \infty)$  varía lentamente en infinito si y sólo si

$$\lim_{t \rightarrow \infty} \frac{h(tx)}{h(x)} = 1 \quad \forall x > 0$$

**Mezclante en sentido ergódico.**

A3.3 La sucesión  $(X_k)$ ,  $k \in \mathbb{Z}$  es mezclante en el sentido ergódico si  $\forall A, B \in \mathcal{F}^{\infty}$   $\lim_{n \rightarrow \infty} P(A \cap T^n B) = P(A)P(B)$ .

## BIBLIOGRAFIA.

- Akgyray, V., Booth, G.G. (1988) *«The Stable-Law Model of Stock Returns»* Journal of Business & Economic Statistics Vol.6, No.1.
- Anderson, T.W. (1970) *«Time Series Analysis»* Wiley, New York.
- Berbee, H. (1981) *«Weak Bernoulli processes, periodicity and renewal theory on the integers»* Report no. 167, Mathematical Centre, Amsterdam.
- Bernstein, S.N., (1926) *«Sur l'extention du théorème limite du calcul des probabilités aux sommes de quantités dépendantes»* Math Ann 97.
- Bernstein, S.N., (1927) Math. Ann. 85, 1-57.
- Bernstein, S. (1939) *«Quelques remarques sur le théorème limite Liapounoff»*, Dokl. Akad. Nauk, USSR.
- Berry, A.C. (1941) *«The Accuracy of the Gaussian Aproximation to the sum of independent variates»* Trans. Amer. Math. Soc. 49.
- Blum, J.R., Hanson D.L y Koopmans, L.H. (1963) *«On the strong law of large numbers for a class of stochastic processes»* Z. Wahrsch. Verw. Gebiete 2, 1-11.
- Bradley R.C. (1980) *«On the  $\phi$ -mixing condition for stationary random sequences»* Duke Math. J. 47, 421-433.
- Bradley R.C. (1980b) *«A remark on the Central Limit Question for Dependent Random Variances»*, J. Appl. Prob. 17.
- Bradley R.C. (1981) *«Central limit theorems under weak dependence»* J. Multivariate Anal. 11, 1-16.
- Bradley R.C. (1982) *«Counterexamples to the Central Limit Theorem Under Strong Mixing Conditions»*, Colloquia Mathematica Soc. János Bolyai. 36 Limit Theorems en Probability and Statistics, Veszprém (Hungary).
- Bradley R.C. (1983) *«On the  $\psi$ -mixing condition for stationary random sequences»* Trans. Amer. Math. Soc. 276, 55-66.
- Bradley R.C. (1984) *«The Central Limit Question under  $\rho$ -mixing»* (por aparecer en Rocky Mtn. J. Math.).
- Bradley R. (1986) *«Basic properties of strong mixing conditions»* en E Eberlein and M.S. Taqqu Eds: *«Dependence in Probability and Statistics: A Survey of Recent Results»* Birkhauser, Boston.
- Brown, B.M. (1969) *«Moments of a stopping rule related to the Central Limit Theorem»*, Ann. Math. Statist. 41.
- Brown, B.M. (1970) *«Characteristic Functions, moments and the Central Limit Theorem»*, Ann. Math. Statist. 42.
- Cramér, H. (1936) *«Sur uner propriété de la loi de Gauss»* C.R. Acad. Sci., Paris.
- Collette, J.P. (1985) *«Historia de las Matemáticas »*. Vol II. Siglo XXI, México.
- DeMoivre, A. (1738) *«The Doctrine of Chances »*. 2nd. ed.
- Denker, M. (1986) *«Uniform integrability and the central limit theorem»* en E Eberlein and M.S. Taqqu Eds: *Dependence in Probability and Statistics: A Survey of Recent Results»* Birkhauser, Boston.
- Doebelin, W. (1940) *«Sur l'ensemble de puissances d'une loi de probabilité»* Studia Math 9.
- Domowitz, I. y White, H. (1982) *«Misspecified Models with Dependent Observations»*. Journal of Econometrics 20, 35-58.

- D'Agostino, Stephens (1986) « *Godness of fit Technics* ». M. Dekker, New York.
- Esseen, C.G. (1945) « *Fourier Analysis of Distribution Functions. A Mathematical Study of the Laplace Gaussian Law* ». Acta Math 77.
- Fama, E. (1965) « *The Behavior of Stock Market Prices* » Journal of Business 38.
- Fama, E. Roll, R. (1971) « *Parameter Estimates for Symmetric Stable Distributions* » Journal of the American Statistical Association, 66.
- Feller, W. (1950) « *An Introduction to Probability Theory and its Applications* », Wiley & Sons, New York.
- Feller, W. (1935, 1937) « *Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung* » Math Z 40.
- Fielitz, B., Rozelle, J. (1983) « *Stable Distributions and Mixtures of Distributions Hypotheses for Common Stock Returns* » Journal the American Statistical Association, 78.
- Fulks, W. (1973) « *Calculo Avanzado* ». Limusa, México.
- Gallant, A.R. (1986) « *Nonlinear Statistical Models* ». John Wiley & Sons, New York.
- Gnedenko, B.V. (1969) « *The Theory of Probability* ». MIR, Moscow.
- Gnedenko, B.V. (1939) « *On the Theory of Domains of attraction of Stable Laws* » Uchenye Zapiski Moskow Gos. Univ. Matematika 30.
- Gnedenko, B.V. (1938) « *Über die Konvergenz der Verteilungsgesetze von Summen voneinander unabhängiger Summanden* » C.r. (Doklady) Acad. Sci. URSS (N.S.).
- Gnedenko, B.V., y Kolmogorov, A.N. (1954) « *Limit Distributions for Sums of Independent Random Variables* ». Addison-Wesley, Massachusetts.
- Gordin, M.I. (1969) « *The Central Limit Theorem for Stationary Processes* », Soviet Math Dokl. 10.
- Hall, P. (1978) « *The convergence of moments in the Martingale Central Limit Theorem* » Wahrsch Verw. Gebiete 44.
- Hall, P. (1983) « *Martingales* ». en S. Kotz and N. Johnson Eds: Encyclopedia of Statistical Sciences 4, 278-285.
- Hall, P. y Heyde, C.C. (1980) « *Martingale Limit Theory and its Application* ». Academic Press, New York.
- Hannan, E.J. (1970) « *Multiple Time Series* », Wiley, New York.
- Hamming, R.W. (1962) « *Numerical Methods for Scientists and Engineers* ». McGraw-Hill, New York.
- Harris, B. (1966) « *Theory of Probability* ». Addison-Wesley, Massachusetts.
- Heyde C.C., (1974) « *On the Central Limit Theorem for Stationary Processes* » Wahrsch Verw. Gebiete 30.
- Hogg, R. y Craig, A. (1969) « *Introduction to Mathematical Statistics* ». MacMillan, New York.
- Ibragimov, I.A. (1959) « *Some limits theorems for strictly stationary stochastic processes* » Dokl. Akad. Nauk SSSR 125, 711-714.
- Ibragimov, I.A. (1962) « *Some Limit Theorem for Stationary Processes* » Theory Prob. Appl. 7, 349-382.
- Ibragimov, I.A. (1975) « *A Note on the Central Limit Theorem for Dependent random variables* ». Theory Prob. Appl. 20, 135-141.
- Ibragimov, I.A. y Linnik, Yu. V. (1971) « *Independent and Stationary Sequences of Random Variables* ». Walters-Noordhoff The Netherlands.

- Ibragimov, I.A. y Rozanov, Y.A. (1978) «Gaussian Random Processes» Springer-Verlag, Berlin.
- Kemeny, J.G. y Kurtz, T.E. (1985a) « True Basic Reference Manual ». Addison-Wesley, Massachusetts.
- Kemeny, J.G., y Kurtz, T.E. (1985b) « True Basic User's Guide ». Addison-Wesley, Massachusetts.
- Kennedy, W. y Gentle, J. (1980) « Statistical Computing ». M. Dekker, New York.
- Khintchine, A.Ya., Lévy, P. (1936) «Sur les lois stables» C.r. Acad. Sci., Paris.
- Khintchine, A.Ya. (1935) «Sul dominio di Attrazione della legge di Gauss» Giorn Ist. Ital. Attuari 6.
- Khintchine, A.Ya. (1938) «Limit Theorem of Sums of independent random variables» Gonti, Moscow-Leningrad.
- Khintchine, A.Ya. (1937a) «Theorie del ubners chranti teilbaren Verteilungsgesetze» Rec. Math (Mat. Sbornik) N.S. 2 (44).
- Khintchine, A.Ya. (1937b) «Deduction Nouvelle d'une formule de H Paul Lévy» Bull Univ. d'état Moscow. Sér Internat. Sect. A.1 no.1.
- Kolmogorov, A.N. (1932) «Sulla forma Generale di un processo Stocastico omogeneo» Atti Acad. Nat. Lincei Rend. Cl. Csi Fis. Mat. Nat. 6.
- Kolmogorov, A.N. y Rozanov, Y.A. (1960) «On strong mixing conditions for stationary Gaussian processes» Theory Probab. Appl. 5, 204-208.
- Laplace, R. (1820) «Theorie analytique des probabilités» Paris.
- Le Cam, L. (1986) « The Central Limit Theorem around 1935 ». Statistical Science, Vol 1, No. 1, pp.78-96.
- Lévy, P. (1925) «Calcul des probabilités» Gauthier-Villars, Paris.
- Lévy, P. (1935a) «Sur les intégrales dont les éléments sont des variables aléatoires indépendantes» Annali R. Scuola-Norm. Sup. Pisa (2).
- Lévy, P. (1935b) J. Math Pures Appl. 14, 347-402.
- Lévy, P. (1937) «Theorie de l'addition des variables aléatoires» Gauthier-Villars, Paris.
- Marankiewicz, J. (1938) «Sur les fonctions indépendantes II» Fund. Math. 30.
- Peligrad, M. (1986) «Recent Advances on the Central Limit Theorem and its Weak Invariance Principle for Mixing Sequences of Random Variables (A Survey)». en E Eberlein and M.S. Taqqu Eds: «Dependence in Probability and Statistics: A Survey of Recent Results» Birkhauser, Boston.
- Rényi, A. (1976) « Calculo de Probabilidades » Reverte, S.A., Barcelona.
- Rosenblatt, M. (1956) «A central limit theorem and a strong mixing condition» Proc. Nat. Acad. Sci. U.S.A. 42, 43-47.
- Rosenblatt (1971) «Markov Processes, structure and asymptotic behavior» Springer-Verlag, Berlin.
- Rozelle, J., Fielitz, B. (1980) «Skewness in Common Stock Returns» Financial Review, 15.
- Shiryayev, A.N. (1984) «Probability» Springer-Verlag, New York.
- Simkowitz, M., Beedles, W. (1980) «Asymmetric Stable Distributed Security Returns» Journal of the American Statistical Association, 75.
- Tardiff, R.M. (1981) «L'Hospital's Rule and the Central Limit Theorem». The American Statistician, Vol 35, No. 1, pp.43.
- Todhunter, M.A., F.R.S. (1865) « A History of the Mathematical



- Theory of Probability* ». Chelsea, New York.
- Tucker, H.G. (1967) « *A Graduate Course in Probability* ». Academic Press, New York.
- Ville, J. (1939) « *Etude Critique de la Notion de Collectif* ». Gauthiers-Villars, Paris.
- Volkanskii, V.A. y Rozanov, Y.A. (1959) « *Some limit theorems for random functions I* » *Theory Probab. Appl.* 4, 178-197.
- White, H. y Domowitz, I. (1981) « *Nonlinear Regression with Dependent Observations* ». Department of Economics discussion paper no. 81-32, University of California, San Diego.
- Willis, J.C. (1922) « *Age and Area* ». Cambridge University Press.
- Yule, G.U. (1925) « *A mathematical theory of Evolution, based on the conclusions of Dr. J.C. Willis, F.R.S.* ». *Philos. Trans. Roy. Soc. London Ser B* 213, 21-87.
- Zolotarev, V.M. (1986) « *One-Dimensional Stable Distributions* ». American Mathematical Society, Rhode Island.