CONVERGENCIA DE VARIABLES ALEATORIAS

JULIETA BAUTISTA DE LA CRUZ

TESIS

Presentada como Requisito Parcial para Obtener el Grado de:

MAESTRA EN ESTADÍSTICA APLICADA

UNIVERSIDAD AUTÓNOMA AGRARIA
"ANTONIO NARRO"

Buenavista, Saltillo, Coahuila, México
Mayo de 2012

Universidad Autónoma Agraria Antonio Narro
Subdirección de Postgrado

CONVERGENCIA DE VARIABLES ALEATORIAS

TESIS

Por:

JULIETA BAUTISTA DE LA CRUZ

Elaborada bajo la supervisión del comité particular de asesoría y aprobada como requisito parcial para optar al grado de

MAESTRA EN ESTADÍSTICA APLICADA

Comité Particular

Asesor principial: ______________________
Dr. Rolando Cavazos Cadena

Asesor: ______________________
M. C. Félix de Jesús Sánchez Pérez

Asesor: ______________________
M. C. Luis Rodríguez Gutiérrez

____
Dr. Fernando Ruiz Zárate
Subdirector de Postgrado

Buenavista, Saltillo, Coahuila, Mayo de 2012

Agradecimiento

Dedicatoria

COMPENDIO

CONVERGENCIA DE VARIABLES ALEATORIAS

Por

JULIETA BAUTISTA DE LA CRUZ

MAESTRÍA EN

ESTADÍSTICA APLICADA

UNIVERSIDAD AUTÓNOMA AGRARIA
ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA, Mayo de 2012

Dr. Rolando Cavazos Cadena –Asesor–

Palabras clave: Teorema central de límite, Convergencia en distribución, Convergencia en probabilidad, Función característica, Teorema de Continuidad, Ley de los grandes números.

Este trabajo trata sobre dos nociones de convergencia para variables aleatorias: convergencia en probabilidad y convergencia en distribución. El principal objetivo de la exposición es presentar las propiedades básicas de estas ideas, como las propiedades de invarianza, e ilustrar los conceptos y resultados por medio de ejemplos analizados en detalle. Se incia con la introducción de los dos modos de convergencia y posteriormente se estudia la idea de función característica, estableciendo su relación con los conceptos de convergencia. Usando el teorema de continuidad se demuestra el teroema central de límite, se muestran aplicaciones adiconales sobre aproximación de distribuciones, y se concluye estableciendo una versión de la ley de los grandes números para variables aleatorias independientes con distribución común que no necesariamente tiene varianza finita.

ABSTRACT

CONVERGENCE OF RANDOM VARIABLES

BY

JULIETA BAUTISTA DE LA CRUZ

MASTER IN

APPLIED STATISTICS

UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA, May, 2012

Dr. Rolando Cavazos Cadena –Advisor–

Key Words: Central limit theorem, Convergence in distribution, Convergence in probability, Characteristic function, Continuity theorem, Law of large numbers.

This work is concerned with two notions of covergence of random variables: convergence in distribution and convergence in probability. The main objective is to analyze the basic properties of those ideas, like the invariance property, and illustrate the concepts and results by using fully analyzed examples. The exposition begins by defining the two modes of convergence, and continues with the idea of characteristic function, establishing the connection with the convergence concepts. Using the continuity theorem, the central limit theorem is proved, and additional applications are shown about approximation of distributions. The presentation concludes with a version of the law of large numbers for independent random variables whose common distribution may have infinite variance.

ÍNDICE DE CONTENIDO

Capítulo 1

Presentación

En este capítulo se proporciona una visión general de este trabajo, y de la motivación que lo origina. Se establecen los objetivos que se pretenden alcanzar, y se describe la organización del material subsecuente.

1.1. Introducción

Este trabajo trata sobre nociones de *convergencia de variables aleatorias*, las cuales están presentes desde la interpretación misma de probabilidad en términos frecuenciales, hasta las aplicaciones más comunes en las que se utilizan aproximaciones (usualmente normales) para determinar tamaños de muestra, construir intervalos de confianza o probar hipótesis. El teorema central de límite, que se encuentra en el núcleo de la metodología estadística, se expresa mediante la *convergencia en distribución* de variables aleatorias, y la propiedad de consistencia, la cual es la propiedad más básica que se puede requerir de un método de estimación, se expresa por medio del concepto de *convergencia en probabilidad*. Así, los diversos modos de convergencia de variables aleatorias están presentes en la disciplina estadística, desde los aspectos más teóricos hasta los más aplicados , y su comprenisión es, sin duda alguna, relevante, como lo muestra el hecho de que el tema es abordado tanto en textos teóricos como en aquellos que enfatizan las aplicaciones; vea, por ejemplo, Dudewicz y Mishra (1998), Wackerly *et al.* (2009), Lehmann y Casella (1999), Graybill (2000), o Montgomery (2011).

La importancia que práctica y teóricamente tienen las ideas de convergencia de variables aleatorias es *la motivación* fundamental para tratar de entender de forma rigurosa estos conceptos, así como los métodos de estudio de sus propiedades. Como producto de este esfuerzo, al final de este trabajo se establecerá una versión de la ley de los grandes números que es más general que la que usualmente se encuentra en los textos de nivel intermedio (Mood *et al.* 1984, Dudewicz y Mishra 1998, Wackerly *et al.* 2009).

1.2. Objetivos y la Principal Contribución

Las principales metas de este trabajo se describen a continuación:

(i) Presentar una estudio formal y riguroso de las nociones de *convergencia en probabilidad* y *convergencia en distribución*;

(ii) Formular las *propiedades de invarianza de ambos modos de convergencia* e ilustrar su aplicación para establecer propiedades como la consistencia de estimadores, o para aproximar distribuciones.

(iii) Entender los instrumentos y resultados analíticos que se emplean para establecer la convergencia en distribución y probabilidad.

(iv) Ilustrar la aplicación de los instrumentos analíticos para analizar las nociones de convergencia de variables aleatorias mediante ejemplos desarrollados, y proporcionar una demostración del teorema central de límite.

(v) Establecer la ley de los grandes números para muestras de una distribución con esperanza finita, pero cuya varianza puede ser infinita.

El trabajo desarrollado en las siguientes secciones para alcanzar esas metas se basa en la función característica de una distribución de probabilidad, y utiliza intensivamente dos de sus propiedades:

(a) El principio de unicidad, el cual establece que si dos distribuciones de probabilidad tienen la misma distribución característica, entonces las distribuciones coinciden; en otras palabras, una distribución de probabilidad está *caracterizada* por su función característica.

(b) La propiedad de continuidad, la cual establece que la convergencia en ditribución de una sucesión de variables aleatorias $\{W_n\}$ hacia una variable aleatoria W equivale a que las funciones características de las variables W_n converjan a la función característica de la variable W.

Esta última propiedad permite trasladar un problema de naturaleza estadística (la convergencia en distribución) al terreno del análisis matemático (la convergencia de funciones).

Usualmente, los temas tratados en el desarrollo subsecuente se estudian en textos de nivel intermedio usando de la función generadora de momentos de una dsitribución; sin embargo, ésta no necesariamente está definida para una distribución arbitraria pues, por lo menos, se requiere que todos los momentos de la distribución de interés sean finitos, condición que no es universalmente válida; vea, por ejemplo Wackerly *et al.* (2009), o Mood *et al.* (1984).

La principal contribución de este trabajo es abordar los temas por medio de la función característica, e ilustrar de manera detallada los conceptos estudiados. Como ejemplo de la potencia del empleo de la función característica como instrumento de análisis, es oportuno señalar que la versión de la ley de los grandes números que se establece al final de este trabajo, no se encuentra en los dos textos anteriromente mencionados.

1.3. El Origen de Este Trabajo

Este trabajo se originó como producto de las actividades desarrolladas en el proyecto *Mathematical Statistics: Elements of Theory and Examples*, el cual inició en el mes de julio de 2011, dentro del Programa de Graduados en Estadística en la Universidad Autónoma Agraria Antonio Narro, con la participación, por parte de los estudiantes del programa, de Mary Carmen Ruiz Moreno y Alfonso Soto Almaguer.

Los objetivos básicos del proyecto son:

(i) Ser un entorno de trabajo *voluntario* donde los problemas estadísticos puedan ser discutidos de forma abierta y fructífera;

(ii) Promover el *entendimiento* de ideas estadísticas y matemáticas a través del análisis y solución detallada de ejercicios;

(iii) Desarrollar las *habilidades de escritura* de los participantes, generando un conjunto de ejemplos claramente resueltos, los cuales puedan ser utiliza-dos por otros participantes del proyecto, así como por otras comunidades interesadas en la estadística, tanto nacionales como extrajeras;

(iv) Propiciar el desarrollo de las *habilidades de comunicación* de los participantes del proyecto por medio de discusiones y exposición regular de temas en seminarios.

Actualmente, las actividades del proyecto giran alrededor de problemas fundamentales en le teoría estadística a un nivel intermedio—de manera que el enfoque no utiliza teoría de la medida de forma intensiva—al nivel del texto de Dudewicz y Mishra (1998). Cuando es necesario obtener información a un nivel más profundo, se ha recurrido a referencias como Lehmann y Casella (1998), Borobkov (1999) and Shao (2002), mientras que aspectos probabilísticos más profundos se han consultado en Loève (1984), Billinsley (1998), o Ash (2002). Por otro lado, el análisis estadístico requiere herramientas algebráicas y analíticas, y en esta dirección se han utilizado referencias como Graybill (2001), Harville (2008), Apostol (1980), Fulks (1980), Rudin (1984) Khuri (2002) and Royden (2003). whereasmientras que los aspectos algebráicos necesarios se presentan, por ejemplo, en Graybill (2001) y Harville (2008).

El enfoque y algunos de los ejemplos presentados más adelante reflejan el trabajo desarrollado en el proyecto.

1.4. La Organización.

El material subsecuente ha sido organizado de la siguiente manera:

En el Capítulo 2 se discute el comportamiento de una sucesión $\{W_n\}$ de variables aleatorias conforme n crece; en las aplicaciones, n representa el tamaño de la muestra, y W_n es un estadístico basado en la primeras n observaciones muestrales. El propósito principal es introducir la idea de *convergencia en probabilidad*, que se refiere al significado de que, conforme n aumenta, la variable W_n se aproxime a una variable aleatoria fija W. Después de analizar este concepto se presenta la propiedad de invarianza, la cual establece que, si una sucesión convergente en probabilidad se transforma mediante la aplicación de una función continua, la sucesión transformada también converge en probabilidad y, más aún, el nuevo límite se obtiene transformando el límite original.

A continuación, en el Capítulo 3 se estudia la noción de *convergencia en distribución* de una sucesión de variables aleatorias. En términos de esta idea se expresa el teorema central de límte, el cual es el resultado medular en la teoría estadística. La exposición analiza la relación entre los conceptos de convergencia en distribución y en probabilidad, y la invarianza de la convergencia en distribución bajo perturbaciones pequeñas.

La presentación continua en el Capítulo 4 estudiando la idea de función característica de una variable aleatoria (o de su distribución), el cual es el instrumento analítico esencial para estudiar la noción convergencia en distribución; se presentan las propiedades fundamentales de las funciones características, incluyendo los principios de unicidad y de continuidad.

Finalmente, la exposición concluye en el Capítulo 5 presentando aplicaciones del principio de continuidad; la primera de ellas es la demostración del teorema central de límite, y posteriormente se obtienen aproximaciones para una distribución geométrica en términos de variables exponenciales, y para una distribución binomial usando una distribución de Poisson. También se establece la aproximación clásica para un distribución binomial por medio de una distribución normal, y se concluye con una versión general de la ley de los grandes números obtenida por medio del teorema de continuidad.

Capítulo 2

Convergencia de Variables Aleatorias

En este capítulo se discute el comportamiento de una sucesión $\{W_n\}$ de variables aleatorias conforme n crece; en las aplicaciones, n representa el tamaño de la muestra, y W_n es un estadístico basado en la primeras n observaciones muestrales. El propósito principal es introducir la idea de *convergencia en probabilidad*, que se refiere al significado de que, conforme n crece, la variable W_n se aproxime a una variable aleatoria fija W. Después de analizar este concepto se presenta la propiedad de invarianza, la cual establece que, si una sucesión convergente en probabilidad se transforma mediante la aplicación de una función continua, la sucesión transformada también converge en probabilidad y, más aún, el nuevo límite se obtiene transformando el límite original.

2.1. Convergencia en Probabilidad

Considere una sucesión $\{W_n\}$ de variables aleatorias definidas en un mismo espacio de probabilidad. De manera intuitiva, la sucesión converge en probabilidad a una variable aleatoria fija W si, a medida que n crece, la probabilidad de que W_n se acerque a W tanto como se quiera se acerca a 1. Formalmente, esta idea se introduce a continuación:

Definición 2.1.1. Sea $\{W_n\}$ una sucesión de variables aleatorias definidas en un mismo espacio de probabilidad $(\Omega, \mathcal{F}, P)$. En este caso $\{W_n\}$ converge en probabilidad a la variable aleatoria $W\colon \Omega \to \mathbb{R}$ si, y sólo si, para cada $\varepsilon > 0$

$$\lim_{n\to\infty} P[|W_n - W| > \varepsilon] = 0.$$

Se utiliza la notación

$$W_n \xrightarrow{\text{P}} W$$

para indicar que $\{W_n\}$ converge en probabilidad a W.

Como punto de partida para analizar la definición precedente es conveniente analizar el siguiente caso sencillo.

Ejemplo 2.1.1. Sea W una variable aleatoria fija, y defina $W_n = W + 1/n$, En este caso, se verá que $W_n \xrightarrow{\text{P}} W$. Para verificar esta afirmación, sea $\varepsilon > 0$ y note que $|W_n - W| = 1/n$, así que el evento $[|W_n - W| > \varepsilon]$ está dado por

$$[|W_n - W| > \varepsilon] = \begin{cases} \emptyset, & \text{si } 1/n < \varepsilon \\ \Omega, & \text{si } 1/n \geq \varepsilon \end{cases}$$

Note ahora que $1/n < \varepsilon$ significa que $n > 1/\varepsilon$, así que cuando n excede a $1/\varepsilon$ se tiene que el evento $[|W_n - W| > \varepsilon]$ es vacío, y entonces $P[|W_n - W| > \varepsilon] = 0$ si $n > 1/\varepsilon$, lo cual muestra que

$$\lim_{n\to\infty} P[|W_n - W| > \varepsilon] = 0,$$

El significado intuitivo de que la sucesión $\{W_n\}$ converja en probabilidad a W es que, si n es suficientemente grande, entonces W_n y W diferirán por menos de ε, donde $\varepsilon > 0$ es la precisión deseada de aproximación. Note que, en el caso presente, si $\varepsilon = .001$, entonces $P[|W_n - W| > .001] = 1$ si $n = 1, 2, 3, \ldots, 999$, esto es, $P[|W_n - W| > .001]$ será 'grande' mientras n no exceda a 999. En general, que tan grande debe ser n para que $P[|W_n - W| > \varepsilon]$ sea 'pequeña', digamos menor a un número δ, depende tanto de ε como del número δ. □

El siguiente teorema se refiere a la preservación de la convergencia en probabilidad bajo transformaciones de las variables aleatorias involucradas.

Teorema 2.1.1. Suponga que la sucesión $\{W_n\}$ y W son tales que $W_n \xrightarrow{\text{P}} W$, y sea g una función continua en un intervalo abierto I que contiene a todos los posibles valores de W, esto es, $P[W \in I] = 1$. En este caso,

$$g(W_n) \xrightarrow{\text{P}} g(W).$$

Antes de proceder a demostrar este resultado, es conveniente entender su significado. El teorema establece que si $W_n \xrightarrow{\text{P}} W$, entonces $W_n^2 \xrightarrow{\text{P}} W^2$, pues $g(w) = w^2$ es una función continua; similarmente $\cos(W_n) \xrightarrow{\text{P}} \cos(W)$, ya que $g(x) = \cos(x)$ es continua en $\mathbb{R}$. Suponga que $P[W > 0] = 1$. En este caso $\log(W_n) \xrightarrow{\text{P}} \log(W)$ and $1/W_n \xrightarrow{\text{P}} 1/W$, como se desprende del teorema aplicado a las funciones $g(x) = \log(x)$ y $g(x) = 1/x$, respectivamente, las cuales son continuas en el intervalo $(0, \infty)$, un intervalo abierto en el que, con probabilidad 1, se ubican los valores de W. En esta discusión se ha enfatizado la continuidad de la función g en un conjunto que contenga los valores de W con probabilidad 1. El siguiente ejemplo muestra que el requerimiento de continuidad de $g(x)$ es esencial, esto es, que no puede garantizarse la convergencia $g(W_n) \xrightarrow{\text{P}} g(W)$ cuando la continuidad de g no se tiene.

Ejemplo 2.1.2. Suponga que

$$W_n = 1/n$$

para todo n, esto es, W_n es la variable aleatoria que toma el único valor $1/n$ con probabilidad 1. Si W es la variable aleatoria nula, esto es,

$$P[W = 0] = 1,$$

el Ejemplo 2.1.1 muestra que

$$W_n \xrightarrow{\text{P}} W.$$

Ahora considere la función $g(x)$ definida mediante

$$g(x) = \begin{cases} 1, & \text{si } x \leq 0 \\ 0, & \text{si } x > 0. \end{cases}$$

En este caso, ya que $W_n > 0$ y $W = 0$ con probabilidad 1, se tiene que

$$g(W_n) = 0, \quad \text{y} \quad g(W) = 1.$$

Por lo tanto, $|g(W_n) - g(W)| = |0 - 1| = 1$, de donde se desprende que para cada $\varepsilon \in (0, 1)$ la desigualdad $|g(W_n) - g(W)| > \varepsilon$ es siempre válida, esto es, $P[|g(W_n) - g(W)| > \varepsilon] = 1$ para todo n, lo cual muestra que $\{g(W_n)\}$ *no converge en probabilidad a* $g(W)$, a pesar de que $W_n \xrightarrow{\mathrm{P}} W$. Desde luego, esto se debe a que la función $g(\cdot)$ no es continua en cero, el único valor que la variable aleatoria W asume. Este análisis muestra que la condición de continuidad sobre la función g requerida en el Teorema 2.1.1 es esencial para preservar la convergencia en probabilidad. $\square$

2.2. Demostración del Teorema 2.1.1

En esta sección se establece el teorema enunciado anteriormente sobre la preservación de la convergencia en probabilidad bajo la aplicación de funciones continuas. El argumento depende de dos propiedades fundamentales de los números reales, las cuales pueden consultarse, por ejemplo, en Apostol (1980), Fulks (1980), o en Khuri (2002):

(i) Cada intervalo abierto es la unión creciente de una cantidad numerable de intervalos cerrados y acotados; por ejemplo, el intervalo abierto $(0, 1)$ es la unión de todos los intervalos $[1/n, 1 - 1/n]$ donde $n = 1, 2, 3, 4, \ldots$, y

(ii) Si una función g es continua en un intervalo cerrado y acotado $J = [a, b]$, entonces, la función g es uniformemente continua en J, esto es, dado $\varepsilon > 0$, existe $\delta > 0$ tal que si x y x_1 son dos puntos arbitrarios del intervalo J tales que $|x - x_1| < \delta$, entonces $|g(x) - g(x_1)| < \varepsilon$.

Demostración del Teorema 2.1.1. Sea I un conjunto abierto tal que $P[W \in I] = 1$ y g es continua en I. Se verificará que si $W_n \xrightarrow{\mathrm{P}} W$, entonces

$$\lim_{n\to\infty} P[|g(W_n) - g(W)| > \varepsilon] = 0,$$

esto es, que dado $\delta > 0$ existe un entero positivo N tal que

$$P[|g(W_n) - g(W)| > \varepsilon] < \delta \quad \text{si } n > N. \tag{2.2.1}$$

Para alcanzar este objetivo, se seguirán los siguientes tres pasos:

(i) El intervalo I es la unión de una sucesión creciente de conjuntos cerrados y acotados (comúnmente denominados compactos) K_n, esto es, $I = \bigcup_{n=1}^{\infty} K_n$, donde $K_n \subset K_{n+1}$ (Khuri, 2002). Por lo tanto, la propiedad de continuidad de una medida de probabilidad implica que $\lim_{n\to\infty} P[W \in K_n] = P[W \in I] = 1$. Luego, existe un entero positivo N_1 tal que $P[W \in K_{N_1}] > 1 - \delta/2$, esto es,

$$P[W \in I \setminus K_{N_1}] < \delta/2. \tag{2.2.2}$$

(ii) Como g es continua en I y $K_{N_1} \subset I$ es un conjunto cerrado y acotado, la función g es uniformemente continua en K_{N_1}, lo cual implica que existe un número $\Delta > 0$ tal que

$$|g(x) - g(W)| \leq \varepsilon \quad \text{si } W \in K_{N_1} \quad \text{y} \quad |x - W| \leq \Delta. \tag{2.2.3}$$

(iii) Utilizando ahora que $W_n \xrightarrow{\text{P}} W$, se desprende que existe un entero positivo N tal que

$$P[|W_n - W| > \Delta] < \delta/2, \qquad \text{si } n > N. \tag{2.2.4}$$

Para concluir, observe que (2.2.3) permite establecer que

$$[W \in K_{N_1}] \cap [|W_n - W| \leq \Delta] \subset [|g(W_n) - g(W)| \leq \varepsilon],$$

y por lo tanto, tomando el complemento en ambos lados de esta inclusión, se desprende que

$$[W \in K_{N_1}]^c \cup [|W_n - W| \leq \Delta]^c \supset [|g(W_n) - g(W)| \leq \varepsilon]^c,$$

es decir,

$$[W \in I \setminus K_{N_1}] \cup [|W_n - W| > \Delta] \supset [|g(W_n) - g(W)| > \varepsilon],$$

lo cual equivale a

$$[|g(W_n) - g(W)| > \varepsilon] \subset [W \in I \setminus K_{N_1}] \cup [|W_n - W| > \Delta].$$

Tomando la probabilidad en ambos lados de esta inclusión se desprende que

$$\begin{aligned} P[|g(W_n) - g(W)| > \varepsilon] &\leq P[W \in I \setminus K_{N_1}] + P[|W_n - W| > \Delta] \\ &< \delta/2 + P[|W_n - W| > \Delta], \end{aligned}$$

donde se utilizó (2.2.2) para establecer la segunda desigualdad. Combinando esta última relación con (2.2.4) se desprende que $P[|g(W_n) - g(W)| > \varepsilon] \leq \delta/2 + \delta/2 = \delta$ cuando $n > N$, concluyendo la demostración. $\square$

Mediante un argumento similar al utilizado para demostrar el Teorema 2.1.1, es posible establecer el siguiente resultado.

Teorema 2.2.1. Suponga que las sucesiones $\{W_{n,1}\}, \{W_{n,2}\}, \ldots, \{W_{n,k}\}$ son tales que

$$W_{n,1} \xrightarrow{\mathrm{P}} W_1,\ W_{n,2} \xrightarrow{\mathrm{P}} W_2,\ \ldots,\ W_{n,k} \xrightarrow{\mathrm{P}} W_k$$

y sea $G(x_1, \ldots, x_k)$ una función continua en una región abierta $\mathcal{R}$ del espacio $\mathbb{R}^k$ tal que

$$P[(W_1, W_2, \ldots, W_k) \in \mathcal{R}] = 1.$$

En este caso,

$$G(W_{n,1}, W_{n,2}, \ldots, W_{n,k}) \xrightarrow{\mathrm{P}} G(W_1, W_2, \ldots, W_k).$$

2.3. Convergencia Hacia una Constante

En esta sección se analiza un caso importante de la idea de convergencia en probabilidad, a saber, aquél en que la variable aleatoria límite W es constante, esto es,

$$P[W = c] = 1,$$

para alguna constante c. En estas circunstancias, si $\{W_n\}$ converge en probabilidad a W se tiene que

$$\lim_{n\to\infty} P[|W_n - c| > \varepsilon] = 0. \tag{2.3.1}$$

Una razón por la cual este tipo de convergencia es importante es que un experimento estadístico se desarrolla con la finalidad de conocer algunos aspectos desconocidos del fenómeno que genera los datos, y que dichos aspectos están formalmente representados en el modelo mediante números cuyo valor se desconoce. Si c es uno de esos números, entonces una tarea básica del analista consiste en utilizar los datos disponibles $X_1, X_2, \ldots, X_n$ para construir una nueva variable aleatoria W_n cuyo valor se utilizarán como una aproximación a la cantidad desconocida c, y entonces se desea que, a medida que la muestra crezca, la aproximación W_n se ubique cada vez más cerca del número c con una alta probabilidad, esto es, que la convergencia (2.3.1) ocurra; la variable aleatoria W_n, construida a partir de $X_1, \ldots, X_n$, se denomina un *estimador* de c basado en los datos X_i, $1 \leq i \leq n$, y cuando (2.3.1) ocurre, la sucesión de estimadores $\{W_n\}$ se denomina consistente. En general, determinar si una sucesión $\{W_n\}$ converge a una constante requiere analizar la distribución de las variables W_n.

Ejemplo 2.3.1. Considere la distribución con densidad

$$f(x) = \frac{2x}{\beta^2 - \alpha^2} I_{(\alpha,\beta)}(x).$$

donde $\alpha < \beta$ y suponga que $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes e idénticamente distribuidas con esta densidad.

(a) Defina $W_n = \max\{X_1, X_2, \ldots, X_n\}$. Se verificará que

$$W_n \xrightarrow{\text{P}} \beta$$

Para alcanzar este objetivo, note que $[X_i \in (\alpha, \beta)] = 1$, de manera que $P[W_n \in (\alpha, \beta)] = 1$ para todo n. Luego, si $0 < \varepsilon < \beta - \alpha$, se tiene que

$$\begin{aligned} P[|W_n - \beta| < \varepsilon] &= P[|W_n - \beta| < \varepsilon,\, W_n \in (\alpha, \beta)] \\ &= P[\beta - \varepsilon < W_n < \beta + \varepsilon,\, W_n \in (\alpha, \beta)] \\ &= P[\beta - \varepsilon < W_n < \beta] \\ &= 1 - P[W_n \leq \beta - \varepsilon] \end{aligned}$$

Debido a que la relación $W_n \leq x$ ocurre si y sólo si $X_i \leq x$ para todo $i = 1, 2, \ldots, n$, se desprende que

$$\begin{aligned} P[W_n \leq \beta - \varepsilon] &= P[X_i \leq \beta - \varepsilon,\ i = 1, 2, \ldots, n] \\ &= \prod_{i=1}^{n} P[X_i \leq \beta - \varepsilon] \\ &= \left[\int_{\alpha}^{\beta-\varepsilon} \frac{2x}{\beta^2 - \alpha^2}\, dx\right]^n \\ &= \left[\frac{(\beta - \varepsilon)^2 - \alpha^2}{\beta^2 - \alpha^2}\right]^n \to 0 \quad \text{conforme } n \to \infty, \end{aligned}$$

donde la convergencia se debe al hecho de que $r^n \to 0$ cuando $|r| < 1$. Combinando las dos últimas relaciones desplegadas se concluye que $P[|W_n - \beta| < \varepsilon] \to 1$ cuando $n \to \infty$, lo cual equivale a $\lim_{n\to\infty} P[|W_n - \beta| > \varepsilon] = 0$, y entonces $W_n \xrightarrow{\mathrm{P}} \beta$.

(b) Sea $V_n = \min\{X_1, X_2, \ldots, X_n\}$. A continuación se comprobará que

$$V_n \xrightarrow{\mathrm{P}} \alpha$$

Como punto de partida observe que $[X_i \in (\alpha, \beta)] = 1$, implica que $P[V_n \in (\alpha, \beta)] = 1$ para todo n. Por lo tanto, para cada ε que satisface $0 < \varepsilon < \beta - \alpha$ se tiene que

$$\begin{aligned} P[|V_n - \alpha| > \varepsilon] &= P[|V_n - \alpha| > \varepsilon,\, V_n \in (\alpha, \beta)] \\ &= P[V > \alpha + \varepsilon,\, V_n \in (\alpha, \beta)] \\ &= P[\alpha + \varepsilon < V_n] \end{aligned}$$

Observando que la relación $V_n > x$ ocurre si y sólo si $X_i > x$ para todo $i = 1, 2, \ldots, n$, se tiene que que

$$\begin{aligned} P[\alpha + \varepsilon < V_n] &= P[\alpha + \varepsilon < X_i,\ i = 1, 2, \ldots, n] \\ &= \prod_{i=1}^{n} P[\alpha + \varepsilon < X_i] \\ &= \left[\int_{\alpha+\varepsilon}^{\beta} \frac{2x}{\beta^2 - \alpha^2}\, dx \right]^n \\ &= \left[\frac{\beta^2 - (\alpha + \varepsilon)^2}{\beta^2 - \alpha^2} \right]^n \to 0 \quad \text{cuando } n \to \infty. \end{aligned}$$

Combinando las dos últimas relaciones desplegadas se obtiene que $P[|V_n - \alpha| > \varepsilon] \to 0$ cuando $n \to \infty$, de manera que $V_n \xrightarrow{\mathrm{P}} \alpha$. □

En el siguiente resultado se establecen propiedades algebráicas básicas de la convergencia en propabilidad hacia una constante.

Teorema 2.3.1. Suponga que $\{W_n\}$ y $\{V_n\}$ son dos sucesiones de variables aleatorias tales que $W_n \xrightarrow{\mathrm{P}} \beta$ y $V_n \xrightarrow{\mathrm{P}} \alpha$.

(i) Las siguientes convergencias en probabilidad ocurren:

$$W_n + V_n \xrightarrow{\mathrm{P}} \beta + \alpha, \qquad W_n - V_n \xrightarrow{\mathrm{P}} \beta - \alpha,$$

$$\frac{W_n}{V_n} \xrightarrow{\mathrm{P}} \frac{\beta}{\alpha} \quad \text{si} \quad \alpha \neq 0, \qquad W_n V_n \xrightarrow{\mathrm{P}} \beta\alpha.$$

(ii) Más generalmente, si $g(x, y)$ es una función de dos variables que es continua en la pareja (α, β), entonces

$$g(V_n, W_n) \xrightarrow{\mathrm{P}} g(\alpha, \beta).$$

(iii) Suponga que las sucesiones $\{W_{n,1}\}, \{W_{n,2}\}, \ldots, \{W_{n,k}\}$ son tales que

$$W_{n,1} \xrightarrow{\mathrm{P}} c_1,\ W_{n,2} \xrightarrow{\mathrm{P}} c_2,\ \ldots,\ W_{n,k} \xrightarrow{\mathrm{P}} c_k$$

y sea $G(x_1, \ldots, x_k)$ una función continua en el punto $(c_1, c_2, \ldots, c_k) \in \mathbb{R}^k$. En este caso

$$G(W_{n,1}, W_{n,2}, \ldots, W_{n,k}) \xrightarrow{\mathrm{P}} G(c_1, c_2, \ldots, c_k).$$

Demostración. Note que la parte (i) se desprende de la parte (ii). En efecto, aplicando la segunda parte con la función $g(x, y) = x + y$ se obtiene que

$$W_n + V_n = g(V_n, W_n) \xrightarrow{\mathrm{P}} g(\alpha, \beta) = \alpha + \beta,$$

mientras que considerando $g(x, y) = y/x$, la cual es continua en (α, β) si $\alpha \neq 0$, se desprende que $W_n/V_n = g(V_n, W_n) \xrightarrow{\mathrm{P}} g(\alpha, \beta) = \beta/\alpha$; las otras convergencias en la parte (i) se obtienen similarmente. A continuación se establecerá la segunda afirmación en el teorema. Sea $\varepsilon > 0$ arbitrario. Usando que $g(x, y)$ es continua en (α, β), seleccione $\Delta > 0$ de manera que

$$|g(x, y) - g(\alpha, \beta)| \leq \varepsilon \quad \text{cuando} \quad |x - \alpha| \leq \Delta \quad \text{y} \quad |y - \beta| \leq \Delta.$$

Esta relación implica que si la desigualdad $|g(V_n, W_n) - g(\alpha, \beta)| > \varepsilon$ ocurre, entonces debe tenerse que $|V_n - \alpha| > \Delta$ o $|W_n - \beta| > \Delta$, esto es

$$[|g(V_n, W_n) - g(\alpha, \beta)| > \varepsilon] \subset [|V_n - \alpha| > \Delta] \cup [|W_n - \beta| > \Delta],$$

Tomando la probabilidad de ambos lados de esta inclusión se obtiene que

$$\begin{aligned}
&P[|g(V_n, W_n) - g(\alpha, \beta)| > \varepsilon] \\
&\qquad \leq P\left[|V_n - \alpha| > \Delta] \cup [|W_n - \beta| > \Delta\right] \\
&\qquad \leq P[|V_n - \alpha| > \Delta] + P[|W_n - \beta| > \Delta] \to 0 \quad \text{conforme } n \to \infty,
\end{aligned}$$

lo cual muestra que $g(V_n, W_n) \xrightarrow{\text{P}} g(\alpha, \beta)$, estableciendo la parte (ii), mientras que la tercera parte se puede demostrar mediante argumentos similares. $\square$

2.4. La Ley de los Grandes Números

En esta sección se presenta el resultado general más importante respecto a la convergencia en probabilidad de una sucesión de variables aleatorias hacia una constante. Dicho resultado, el cual está íntimamente relacionado con la interpretación de una probabilidad en términos frecuenciales, se conoce como 'ley de los grandes números', y establece que al tomar una muestra de una distribución con varianza finita, el promedio de los primeros n datos observados converge en probabilidad hacia el valor esperado de la población.

Teorema 2.4.1. Suponga que $X_i, X_2, \ldots, X_n$ son variables aleatorias independientes con media y varianza comunes μ y $\sigma^2 < \infty$, respectivamente, y defina la media muestral de los primeros n datos mediante.

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Con esta notación, $\overline{X}_n \xrightarrow{\text{P}} \mu$, esto es, para cada $\varepsilon > 0$

$$\lim_{n\to\infty} P[|\overline{X}_n - \mu| > \varepsilon] = 0.$$

El argumento para establecer este teorema depende del siguiente resultado preliminar.

Lema 2.4.1. (i) [Desigualdad de Markov.] Para cualquier variable aleatoria X y todos los números $a, r > 0$,

$$P[|X| > a] \leq \frac{E[|X|^r]}{a^r}.$$

En consecuencia,

(ii) [Desigualdad de Chebichev.] Si X tiene esperanza finita μ, entonces

$$P[|X - \mu| > a] \leq \frac{E[(X - \mu)^2]}{a^2} = \frac{\text{Var}\,[X]}{a^2}.$$

Demostración. (i) Observe que

$$|X|^r \geq |X|^r I[|X| \geq a] \geq a^r I[|X| \geq a]$$

donde $I[A]$ denota la función indicadora del evento A. Tomando el valor esperado en los lados extremos de esta desigualdad se desprende que

$$E[|X|^r] \geq E[a^r I[|X| \geq a]] = a^r E[I[|X| \geq a]] = a^r P[|X| \geq a],$$

de donde se sigue que $E[|X|^r]/a^r \geq P[|X| \geq a]$.

(ii) Sustituyendo X por $X - \mu$ en la parte (i) y poniendo $r = 2$ se llega a

$$\frac{E[(X-\mu)^2]}{a^2} \geq P[|X - \mu| \geq a],$$

la cual es la conclusión deseada. □

Demostración del Teorema 2.4.1. Como las variables $X_1, X_2, \ldots, X_n$ son independientes con media y varianza comunes μ y σ^2, respectivamente, se tiene que

$$E[\overline{X}_n] = \mu, \qquad \text{y} \quad \text{Var}\left[\overline{X}_n\right] = \frac{\text{Var}\,[X_1]}{n} = \frac{\sigma^2}{n}.$$

Por lo tanto, aplicando la desigualdad de Chebichev en el Lema 2.4.1(ii) con $\overline{X}_n$ en lugar de X se desprende que

$$P[|\overline{X}_n - \mu| > a] \leq \frac{\text{Var}\left[\overline{X}_n\right]}{a^2} = \frac{\sigma^2}{na^2}, \quad a > 0.$$

Tomando el límite cuando n tiende a infinito se obtiene que, para cada $a > 0$,

$$\lim_{n\to\infty} P[|\overline{X}_n - \mu| > a] \leq \lim_{n\to\infty} \frac{\sigma^2}{na^2} = 0$$

y, debido a que el número $a > 0$ en esta relación es arbitrario, se concluye que $\overline{X}_n \xrightarrow{\text{P}} \mu$; vea la Definición 2.1.1. □

Combinando los Teoremas 2.4.1 y 2.3.1(iii) se obtiene la siguiente conclusión.

Corolario 2.4.1. Suponga que las sucesiones $\{X_n\}, \{Y_n\}, \dots \{W_n\}$ consisten de variables aleatorias independientes con media y varianza finitas, donde

$$\begin{aligned}
\mu_X &= E[X_1], \quad \sigma_X^2 = \mathrm{Var}\,[X_1] \\
\mu_Y &= E[Y_1], \quad \sigma_Y^2 = \mathrm{Var}\,[Y_1] \\
&\vdots \\
\mu_W &= E[W_1], \quad \sigma_W^2 = \mathrm{Var}\,[W_1]\,.
\end{aligned}$$

En este caso, si $G(x, y, \dots, w)$ es una función continua en $(\mu_X, \mu_y, \dots, \mu_W)$, se tiene que

$$G(\overline{X}_n, \overline{Y}_n, \dots, \overline{W}_n) \xrightarrow{\mathrm{P}} G(\mu_X, \mu_Y, \dots, \mu_W).$$

Ejemplo 2.4.1. Suponga que $X_1, X_2, \dots, X_n$ es una muestra de la densidad gamma con parámetros α y λ, la cual está dada por

$$f(x; \alpha, \lambda) = \begin{cases} \dfrac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x}, & \text{si } x > 0 \\ 0, & \text{si } x \leq 0, \end{cases}$$

donde α y λ son números positivos. El objetivo de este ejemplo es construir estimadores Y_n y W_n basados en $X_1, \dots, X_n$ para los cuales se tiene que $Y_n \xrightarrow{\mathrm{P}} \alpha$ y $W_n \xrightarrow{\mathrm{P}} \lambda$. Como punto de partida, note que

$$\mu = E[X_1] = \frac{\alpha}{\lambda}, \quad E[X_1^2] = \frac{\alpha(\alpha+1)}{\lambda^2}.$$

Luego, del Teorema 2.4.1 aplicado a la sucesión $X_1, X_2, X_3, \dots$ se obtiene que

$$\overline{X}_n \xrightarrow{\mathrm{P}} \mu = \frac{\alpha}{\lambda} \tag{2.4.1}$$

Por otro lado, definiendo las variables aleatorias W_i mediante $W_i = X_i^2$, $i = 1, 2, 3, \dots$ se tiene que $W_1, W_2, W_3, \dots$ son independientes con la misma distribución, y que su media común es

$$E[W_i] = \mu_W = \frac{\alpha(\alpha+1)}{\lambda^2},$$

de manera que

$$\overline{W}_n \xrightarrow{\mathrm{P}} \mu_W = \frac{\alpha(\alpha+1)}{\lambda^2}. \tag{2.4.2}$$

Note ahora que

$$\mu_W = \frac{\alpha(\alpha+1)}{\lambda^2} = \left(\frac{\alpha}{\lambda}\right)^2 \left[1 + \frac{1}{\alpha}\right] = \mu^2 \left[1 + \frac{1}{\alpha}\right]$$

de donde se desprende que $\dfrac{\mu_W}{\mu^2} = 1 + \dfrac{1}{\alpha}$ y entonces

$$\alpha = \frac{\mu^2}{\mu_W - \mu^2}. \tag{2.4.3}$$

Observando que $\lambda = \alpha/\mu$ (vea 2.4.1), se concluye que

$$\lambda = \frac{\mu}{\mu_W - \mu^2}, \tag{2.4.4}$$

y utilizando que $\mu_W > \mu^2$, se desprende que tanto α como λ son funciones continuas de μ y μ_W:

$$\alpha = G(\mu, \mu_W) = \frac{\mu^2}{\mu_W - \mu^2}, \qquad \lambda = G_1(\mu, \mu_W) = \frac{\mu}{\mu_W - \mu^2}.$$

Combinando esta expresión con las convergencias (2.4.1) y (2.4.2), el Corolario 2.4.1 permite establecer que

$$G(\overline{X}_n, \overline{W}_n) \xrightarrow{\text{P}} G(\mu, \mu_W), \quad \text{y} \quad G_1(\overline{X}_n, \overline{W}_n) \xrightarrow{\text{P}} G_1(\mu, \mu_W),$$

esto es, $\dfrac{\overline{X}_n^2}{\overline{W}_n - \overline{X}_n^2} \xrightarrow{\text{P}} \alpha, \quad \text{y} \quad \dfrac{\overline{X}_n}{\overline{W}_n - \overline{X}_n^2} \xrightarrow{\text{P}} \lambda.$ □

2.5. Consistencia del Coeficiente de Correlación Muestral

En esta sección se presenta una ilustración adicional de la ley de los grandes números enunciada en el Teorema 2.4.1. Considere un vector aleatorio (X, Y) donde las componentes tienen segundo momento finito. En este caso, el coeficiente de correlación ρ entre X y Y está dado por

$$\rho = \frac{\text{Cov}\,(X, Y)}{\sqrt{\text{Var}\,[X]\, Var[Y]}},$$

donde, a partir de este punto, se supone que la varianza de las variables X y Y es finita y positiva. El coeficiente de correlación es una característica importante de la distribución conjunta del vector (X, Y), pues es una medida de la asociación lineal entre las componentes X y Y. Suponga ahora que el valor de ρ no se conoce, de manera que será necesario estimarlo. Con esa finalidad se obtiene una muestra (X_i, Y_i), $i = 1, 2, \ldots, n$ de la distribución del vector (X, Y), y se calcula el coeficiente de correlación muestral basado en las n parejas (X_i, Y_i), , el cual está dado por

$$R_n = \frac{\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\sqrt{\sum_{i=1}^n (X_i - \overline{X}_n)^2 \sum_{i=1}^n (Y_i - \overline{Y}_n)^2}} \tag{2.5.1}$$

A continuación se verá como el Corolario 2.4.1 permite concluir que

$$R_n \xrightarrow{\text{P}} \rho. \tag{2.5.2}$$

El argumento que sigue supondrá que X y Y tienen cuarto momento finito, y en la observación al final de esta sección se indicará como evitar este supuesto adicional. El desarrollo subsecuente consiste de tres fases. Primero se analizarán las varianzas muestrales de $X_1, X_2, \ldots, X_n$ y $Y_1, Y_2, \ldots, Y_n$, y posteriormente se estudiará la covarianza muestral entre los datos X_i y Y_i.

Etapa 1: Considere la varianza muestral $S^2_{X,n}$ de los datos $X_1, X_2, \ldots, X_n$, la cual está dada por

$$S^2_{X,n} = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}_n^2.$$

Note ahora que las variables $X_1^2, X_2^2, X_3^2, \ldots, X_n^2$ son independientes con la misma distribución y que $E[(X_i^2)^2] = E[X_i^4] < \infty$; luego, $\text{Var}\left[X_i^2\right] < \infty$. Aplicando la ley de los grandes números se desprede que

$$\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{\text{P}} E[X^2].$$

Por otro lado, debido a que $\overline{X}_n \xrightarrow{\text{P}} E[X]$, el Corolario 2.4.1 aplicado a la función $G(x, y) = x - y^2$ implica que

$$G \left. \quad \frac{1}{n}\sum_{i=1}^n X_i^2,\ \overline{X}_n \right) \xrightarrow{\text{P}} G(E[X^2],\ E[X])$$

esto es,

$$S_{X,n}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2 \xrightarrow{\text{P}} E[X^2] - E[X]^2 = \text{Var}\,[X]$$

Etapa 2: Argumentando de forma similar, puede demostrarse que la varianza muestral de $Y_1, Y_2, \ldots, Y_n$, la cual está dada por

$$S_{Y,n}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}_n^2,$$

satisface

$$S_{Y,n}^2 \xrightarrow{\text{P}} \text{Var}\,[Y]$$

Etapa 3: Se analizará ahora la covarianza muestral entre los dos vectores de datos $X_1, X_2, \ldots, X_n$ y $Y_1, Y_2, \ldots, Y_n$, la cual está dada por

$$S_{XY,n}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)(Y_i - \overline{Y}_n) = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \overline{X}_n \overline{Y}_n.$$

Observe que los productos $X_1Y_1, X_2Y_2, X_3Y_3, \ldots, X_nY_n$ son independientes con la misma distribución y que

$$E[(X_iY_i)^2] = E[X_i^2 Y_i^2] \leq \left(E[X_i^4]\right)^{1/2} \left(E[Y_i^4]\right)^{1/2},$$

(por la desigualdad de Cauchy-Schwarz), y entonces el supuesto de que las variables X_i y Y_i tienen cuarto momento finito implica que $E[(X_iY_i)^2] < \infty$, y por lo tanto $\text{Var}\,[X_iY_i] < \infty$. En este caso, la ley de los grandes números en el Teorema 2.4.1 permite concluir que

$$\frac{1}{n}\sum_{i=1}^{n} X_i Y_i \xrightarrow{\text{P}} E[XY]$$

Combinando esta convergencia con $\overline{X}_n \xrightarrow{\text{P}} E[X]$ y $\overline{Y}_n \xrightarrow{\text{P}} E[Y]$, el Corolario 2.4.1 aplicado a la función $G(x, y, w) = x - yw$ implica que

$$G \left.\quad \frac{1}{n}\sum_{i=1}^{n} X_i^2,\ \overline{X}_n,\ \overline{Y}_n \right) \xrightarrow{\text{P}} G(E[XY],\ E[X],\ E[Y])$$

esto es,

$$S_{XY,n}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \overline{X}_n \overline{Y}_n \xrightarrow{\text{P}} E[XY] - E[X]E[Y] = \text{Cov}\,(X, Y)$$

Para concluir, observe que el coeficiente de correlación muestral R_n se expresa en términos de $S_{XY,n}$, $S^2_{X,n}$ y $S^2_{Y,n}$ como

$$R_n = \frac{S_{XY,n}}{\sqrt{S^2_{X,n} S^2_{Y,n}}} = G(S_{XY,n},\ S^2_{X,n},\ S^2_{Y,n})$$

donde

$$G(x,y,z) = \frac{x}{\sqrt{yz}}$$

es una función continua en la región $y > 0,\ z > 0$, y por lo tanto es continua en el punto $(\mathrm{Cov}\,(X,Y)\,,\mathrm{Var}\,[X]\,,\mathrm{Var}\,[Y)]$. Luego, una aplicación adicional del Corolario 2.4.1 conduce a

$$R_n = G(S_{XY,n},\ S^2_{X,n},\ S^2_{Y,n}) \xrightarrow{\mathrm{P}} G(\mathrm{Cov}\,(X,Y)\,,\mathrm{Var}\,[X)]\,,\mathrm{Var}\,[Y]) = \rho,$$

alcanzando el objetivo deseado □

Observación 2.5.1. Las condicones bajo las cuales se obtiene que $\overline{X}_n \xrightarrow{\mathrm{P}} \mu$ en el Teorema 2.4.1 requieren que las variables X_i sean independientes y con distribución común de media μ *y varianza* σ^2 finitas. En este punto es oportuno mencionar que la condición de que la varianza común sea finita puede omitirse, y que, bajo el supuesto de que las variables sean independientes con la misma distribución, la convergencia $\overline{X}_n \xrightarrow{\mathrm{P}} \mu$ ocurre cuando la esperanza (o media) común de las variables X_i está bien definida. Esta afirmación se establecerá en el último teorema de este trabajo; vea también la denominada ley *fuerte* de los grandes números, la cual puede encontartse demostrada en Loève (1984), Billingsley (1998), o Ash (2002). □

Capítulo 3

El Teorema Central de Límite

Este capítulo trata sobre la noción de *convergencia en distribución* de una sucesión de variables aleatorias. En términos de esta idea se expresa el teorema central de límte, el cual es el resultado medular en la teoría estadística. La exposición analiza la relación entre los conceptos de convergencia en distribución y en probabilidad, y la invarianza de la convergencia en distribución bajo perturbaciones pequeñas.

3.1. Convergencia en Distribución

En esta sección se analiza otra idea de convergencia de variables aleatorias, a saber, la noción de convergencia en distribución.

Definición 3.1.1. Sea $\{V_n\}$ una sucesión de variables aleatorias y sea F una función de distribución. En este caso, la sucesión $\{V_n\}$ converge en distribución a F si, para cada intervalo $I = (a, b] \neq \emptyset$ para el cual a y b son puntos de continuidad de F, se tiene que

$$\lim_{n\to\infty} P[V_n \in I] = \lim_{n\to\infty} P[a < V_n \leq b] = F(b) - F(a);$$

se utiliza la notación

$$V_n \xrightarrow{\mathrm{d}} F$$

para indicar que $\{V_n\}$ converge en distribución a F. Más aún, si X es una variable aleatoria cuya función de distribución es F, entonces también se escribe

$$V_n \xrightarrow{\text{d}} X$$

en lugar de $V_n \xrightarrow{\text{d}} F$.

Ejemplo 3.1.1. Considere el contexto del Ejemplo 2.3.1, en el cual la sucesión $X_1, X_2, X_3, \ldots$ consta de variables independientes con densidad común

$$f(x) = \frac{2x}{\beta^2 - \alpha^2} I_{(\alpha,\beta)}(x),$$

donde $\alpha < \beta$. En dicho ejemplo se demostró que si la variable aleatoria W_n está dada por $W_n = \max\{X_1, X_2, \ldots, X_n\}$, entonces $W_n \xrightarrow{\text{P}} \beta$. Ahora se estudiará la diferencia entre W_n y β. Con este fin, defina

$$V_n = n(\beta - W_n), \quad n = 1, 2, 3, \ldots,$$

A continuación se verificará que la sucesión $\{V_n\}$ converge en distribución hacia una variable exponencial. Para corroborar esta afirmación, note que V_n asume solamente valores positivos, y que para cada $x > 0$

$$\begin{aligned} P[V_n \leq x] &= P[n(\beta - W_n) \leq x] \\ &= P[(\beta - W_n) \leq x/n] \\ &= P[\beta - x/n \leq W_n] \\ &= 1 - P[W_n < \beta - x/n] \\ &= 1 - P[W_n \leq \beta - x/n] \end{aligned} \tag{3.1.1}$$

donde la última igualdad se debe a la continudad de la función de distribución de W_n. Note ahora que la desigualdad $W_n \leq \beta - x/n$ equivale a

$$X_i \leq \beta - x/n, \quad i = 1, 2, \ldots, n$$

y entonces

$$
\begin{aligned}
&P[W_n \le \beta - x/n] \\
&\quad = P[X_1 \le \beta - x/n,\, X_2 \le \beta - x/n,\, \ldots, X_n \le \beta - x/n] \\
&\quad = P[X_1 \le \beta - x/n]P[X_2 \le \beta - x/n]\cdots P[X_n \le \beta - x/n] \\
&\quad = \left(\frac{(\beta - x/n)^2 - \alpha^2}{\beta^2 - \alpha^2}\right)^n \\
&\quad = \left(\frac{\beta^2 - 2\beta x/n + x^2/n^2 - \alpha^2}{\beta^2 - \alpha^2}\right)^n \\
&\quad = \left(1 - \frac{2\beta x/n - x^2/n^2}{\beta^2 - \alpha^2}\right)^n \\
&\quad = (1 - A_n/n)^n
\end{aligned}
\tag{3.1.2}
$$

donde

$$A_n = \frac{2\beta x - x^2/n}{\beta^2 - \alpha^2}.$$

Por lo tanto, $\lim_{n\to\infty} A_n = 2\beta x/(\beta^2 - \alpha^2)$, de donde se desprende que

$$\lim_{n\to\infty} (1 - A_n/n)^n = e^{-2\beta x/(\beta^2 - \alpha^2)},$$

(vea Khuri 2002, Rudin 1984, o Royden 2003). Esta convergencia ocurre frecuentemente en el estudio de la estadística, y será demostrada en la Sección 3 del siguiente capítulo. Combinando la relación anterior con (3.1.1) y (3.1.2) se sigue que

$$\lim_{n\to\infty} P[V_n \le x] = 1 - e^{-2\beta x/(\beta^2 - \alpha^2)}, \quad x > 0.$$

Defina ahora la función de distribución $F(x)$ mediante

$$F(x) = \begin{cases} 1 - e^{-2\beta x/(\beta^2 - \alpha^2)}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

y note que la convergencia precedente muestra que

$$\lim_{n\to\infty} P[V_n \le x] = F(x), \quad x > 0;$$

como ambos lados de esta igualdad son nulos cuando $x \le 0$, se desprende que

$$V_n \xrightarrow{\mathrm{d}} F.$$

Note que F corresponde a la densidad exponencial con parámetro $\lambda = 2\beta/(\beta^2 - \alpha^2)$, así que si Y es una variable con distribución exponencial con dicho parámetro, la convergencia anterior puede escribirse como $V_n \xrightarrow{\mathrm{d}} Y$.□

3.2. Relación entre Modos de Convergencia

En el siguiente resultado se explora la relación entre las nociones de convergencia en probabilidad y convergencia en distribución. La principal conclusión es que la idea de convergencia en probabailidad es más fuerte que la noción de convergencia en distribución, en el sentido de que si se tiene convergencia en probabilidad, entonces, forzosamente, se tiene convergencia en distribución.

Teorema 3.2.1. Suponga que $\{W_n\}$ y W son variables aleatorias que satisfacen

$$W_n \xrightarrow{\mathrm{P}} W.$$

En este contexto,

$$W_n \xrightarrow{\mathrm{d}} W.$$

Demostración. Sea $F(x)$ la función de distribución de W, esto es,

$$F(x) = P[W \leq x], \quad x \in \mathbb{R}.$$

A continuación se verificará que, si x es un punto de continuidad de $F(\cdot)$, entonces

$$\lim_{n\to\infty} P[W_n \leq x] = F(x).$$

Para alcanzar este objetivo, sea $\varepsilon > 0$ un número arbitrario, y note que las siguiente inclusión es válida:

$$\begin{aligned}[W_n \leq x] &= [W_n \leq x, |W_n - W| \leq \varepsilon] \cup [W_n \leq x, |W_n - W| > \varepsilon] \\ &\subset [W \leq x + \varepsilon] \cup [|W_n - W| > \varepsilon]\end{aligned}$$

y por lo tanto,

$$\begin{aligned}P[W_n \leq x] &\leq P[W \leq x + \varepsilon] + P[|W_n - W| > \varepsilon] \\ &= F(x + \varepsilon) + P[|W_n - W| > \varepsilon]\end{aligned} \tag{3.2.1}$$

Similarmente,

$$\begin{aligned}[W_n > x] &= [W_n > x, |W_n - W| \le \varepsilon] \cup [W_n > x, |W_n - W| > \varepsilon] \\ &\subset [W > x - \varepsilon] \cup [|W_n - W| > \varepsilon]\end{aligned}$$

y entonces

$$P[W_n > x] \le P[W > x - \varepsilon] + P[|W_n - W| > \varepsilon]$$

de donde se desprende que

$$\begin{aligned}P[W_n \le x] &= 1 - P[W > x] \\ &\ge 1 - [P[W > x - \varepsilon] - P[|W_n - W| > \varepsilon]] \\ &= P[W \le x - \varepsilon] - P[|W_n - W| > \varepsilon],\end{aligned}$$

lo cual conduce a

$$P[W_n \le x] \ge F(x - \varepsilon) - P[|W_n - W| > \varepsilon]. \tag{3.2.2}$$

Para concluir, sea $\delta > 0$ un número arbitrario y note que, debido a que x es un punto de continuidad para $F(x)$, existe $\varepsilon^* > 0$ tal que

$$|F(x - \varepsilon) - F(x)| + |F(x) - F(x + \varepsilon)| \le \delta/2 \quad \text{si} \;\; 0 \le \varepsilon \le \varepsilon^*.$$

Combinando esta relación con (3.2.1) y (3.2.2) se desprende que

$$F(x) + \delta/2 + P[|W_n - W| > \varepsilon^*] \ge P[W_n \le x] \ge F(x) - \delta/2 - P[|W_n - W| > \varepsilon^*].$$

Finalmente, debido a que $W_n \xrightarrow{\text{P}} W$, se tiene que existe un entero N^* tal que

$$P[|W_n - W| > \varepsilon^*] < \delta/2 \quad \text{si } n > N^*,$$

y entonces la anterior relación desplegada implica que

$$F(x) + \delta/2 + \delta/2 \ge P[W_n \le x] \ge F(x) - \delta/2 - \delta/2 \quad \text{cuando} \;\; n > N^*.$$

esto es,

$$|P[W_n \le x] - F(x)| \le \delta \quad \text{si} \;\; n > N^*,$$

lo cual muestra que

$$\lim_{n\to\infty} P[W_n \le x] = F(x) \quad \text{si } x \text{ es punto de continuidad de } F.$$

Para concluir, sean a y b dos puntos de continuidad de $F(x)$. Observando que $P[a < W_n \leq b] = P[W_n \leq b] - P[W_n \leq a]$, se desprende que

$$\lim_{n\to\infty} P[a < W_n \leq b] = F(b) - F(a),$$

de donde se sigue que $W_n \xrightarrow{\mathrm{d}} F$, es decir, $W_n \xrightarrow{\mathrm{d}} W$; vea la Definición 3.1.1. $\square$

En general, si $\{W_n\}$ converge en distribución a W, no es necesario que la sucesión $\{W_n\}$ converja a W en probabilidad. Sin embargo, en un caso especial esto si es cierto, como se muestra a continuación.

Teorema 3.2.2. Suponga que W es una variable aleatoria constante e igual a c, esto es,

$$P[W = c] = 1.$$

En este caso, si $\{W_n\}$ satisface

$$W_n \xrightarrow{\mathrm{d}} c,$$

entonces se tiene que

$$W_n \xrightarrow{\mathrm{P}} c.$$

Demostración. Suponga que $W_n \xrightarrow{\mathrm{d}} c$ y observe que la función de distribución de la variable W que es constante e igual a c está dada por

$$F(x) = \begin{cases} 1, & \text{si } x \geq c, \\ 0, & \text{si } x < c, \end{cases}$$

la cual es continua en todos los puntos, excepto en $x = c$. Por lo tanto, aplicando la Definición 3.1.1, se tiene que

$$\lim_{n\to\infty} P[W_n \leq x] = \begin{cases} 1, & \text{si } x > c \\ 0, & \text{si } x < c. \end{cases}$$

Luego, para cada $\varepsilon > 0$,

$$\begin{aligned} &\lim_{n\to\infty} P[W_n < c - \varepsilon] = 0 \\ &\lim_{n\to\infty} P[W_n > c + \varepsilon] = \lim_{n\to\infty} [1 - P[W_n \leq c + \varepsilon]] = 1 - 1 = 0 \end{aligned}.$$

Para concluir, note que $[|W_n - c| > \varepsilon] = [W_n < c - \varepsilon] \cup [W_n > c + \varepsilon]$, de manera que

$$P[|W_n - c| > \varepsilon] \leq P[W_n < c - \varepsilon] + P[W_n > c + \varepsilon] \to 0 \quad \text{cuando } n \to \infty,$$

mostrando que $W_n \xrightarrow{\mathrm{P}} c$. □

3.3. Perturbación de una Sucesión Convergente

A continuación se analizará el efecto de someter a una sucesión $\{W_n\}$ que converge en distribución a perturbaciones 'pequeñas'. Diversas versiones del siguiente resultado se conocen en la literaura como *Teorema de Slutsky*; ve por ejemplo, Borovkov (1999), Dudewicz y Mishra (1988), Lehmann y Casell1 (1998), o Mood *et al.* (1984).

Teorema 3.3.1. Suponga que la sucesión $\{W_n\}$ de variables aleatorias satisface la relación

$$W_n \xrightarrow{\mathrm{d}} W.$$

(a) Si las variables aleatorias V_n son tales que $V_n \xrightarrow{\mathrm{P}} 0$, entonces

$$W_n + V_n \xrightarrow{\mathrm{d}} W.$$

(b) Si $V_n \xrightarrow{\mathrm{P}} 0$, entonces

$$V_n W_n \xrightarrow{\mathrm{d}} 0.$$

(c) Si $V_n \xrightarrow{\mathrm{P}} 1$ ocurre, entonces

$$V_n W_n \xrightarrow{\mathrm{d}} 0.$$

Demostración. Sea $F(x) = P[W \leq x]$ la función de distribución de W, de manera que el supuesto de que W_n converge en distribución a W equivale a

$$\lim_{n\to\infty} P[W_n \leq x] = F(x) \quad \text{si } F(\cdot) \text{ es continua en el punto } x. \tag{3.3.1}$$

(a) Sea $x \in \mathbb{R}$ un punto de continuidad de x y $\delta > 0$ un número arbitrario. Seleccione $\varepsilon^* > 0$ tal que $x \pm \varepsilon^*$ son puntos de continuidad de F; en este punto es oportuno observar que, debido a que F es discontinua en a lo más una cantidad numerable de puntos, siempre es posible seleccionar ε^* tan pequeño como se quiera con la propiedad de que F es continua en $x \pm \varepsilon^*$. Note también que, debido a que $V_n \xrightarrow{\text{P}} 0$,

$$P[|V_n| > \varepsilon^*] \to 0 \quad \text{si } n \to \infty. \tag{3.3.2}$$

A continuación observe que

$$[W_n + V_n \leq x] = [W_n + V_n \leq x, \, |V_n| > \varepsilon^*] \cup [W_n + V_n \leq x, \, |V_n| \leq \varepsilon].$$

Note ahora que $W_n + V_n \leq x$ y $|V_n| \leq \varepsilon$ implican que $W_n \leq x + \varepsilon^*$, de manera que la anterior relación desplegada implica que

$$[W_n + V_n \leq x] \subset [|V_n| > \varepsilon^*] \cup [W_n \leq x + \varepsilon^*],$$

lo cual conduce a

$$P[W_n + V_n \leq x] \leq P[|V_n| > \varepsilon^*] + P[W_n \leq x + \varepsilon^*],$$

Tomando el límite conforme n tiende a ∞, (3.3.1) y (3.3.2) implican que

$$\lim_n P[W_n + V_n \leq x] \leq \lim_n P[W_n \leq x + \varepsilon^*] = F(x + \varepsilon^*);$$

y dejando que $\varepsilon^* > 0$ converja a cero a través de puntos en los que $x \pm \varepsilon^*$ son puntos de continuidad de F, se desprende que

$$\lim_n P[W_n + V_n \leq x] \leq F(x); \tag{3.3.3}$$

Por otro lado note que

$$[W_n + V_n \leq x] \supset [W_n \leq x - \varepsilon^*, \, |V_n| \leq \varepsilon^*]$$

de donde se sigue que

$$[|V_n| > \varepsilon^*] \cup [W_n + V_n \leq x] \supset [W_n \leq x - \varepsilon^*],$$

y entonces

$$P[|V_n| > \varepsilon^*] + P[W_n + V_n \leq x] \geq P[W_n \leq x - \varepsilon^*].$$

Después de tomar el límite cuando n tiende a ∞, esta desigualdad implica, via (3.3.1) y (3.3.2), que

$$\lim_n P[W_n + V_n \leq x] \geq \lim_{n\to\infty} P[W_n \leq x - \varepsilon^*] = F(x - \varepsilon^*);$$

dejando que ε^* disminuya hacia cero y recordando que $F(\cdot)$ es continua en x se desprende que

$$\lim_n P[W_n + V_n \leq x] \geq F(x),$$

relación que combinada con (3.3.3) implica que $\lim_n P[W_n + V_n \leq x] = F(x)$; como x es un punto de continuidad arbitrario de $F(x)$ se tiene que $W_n + V_n \xrightarrow{\text{d}} F$, es decir, $W_n + V_n \xrightarrow{\text{d}} W$.

(b) Se tiene que demostrar que, si las convergencias $W_n \xrightarrow{\text{d}} W$ y $V_n \xrightarrow{\text{P}} 0$ se verifican, entonces $V_n W_n \xrightarrow{\text{d}} 0$, lo cual equivale a $V_n W_n \xrightarrow{\text{P}} 0$, por el Teorema 3.2.2. Para alcanzar este objetivo, sea $\delta > 0$ un número arbitrario, y seleccione $\Delta > 0$ tal que

$$P[|W| \leq \Delta] = F(\Delta) - F(-\Delta) > \delta/2$$

Debido a que $W_n \xrightarrow{\text{d}} W$ se tiene que, conforme n tiende a ∞,

$$P[|W_n| \leq \Delta] \to P[|W| \leq \Delta] = F(\Delta) - F(-\Delta) > \delta/2$$

y entonces $P[|W_n| > \Delta] < \delta/2$ si n es suficientemente grande, digamos $n > N^*$:

$$P[|W_n| > \Delta] < \delta/2, \quad n > N^*. \tag{3.3.4}$$

Sea ahora $\varepsilon > 0$ un número arbitrario, y observe que

$$[|V_n W_n| > \varepsilon] = [|V_n W_n| > \varepsilon,\, |W_n| \leq \Delta] \cup [|V_n W_n| > \varepsilon,\, |W_n| > \Delta];$$

usando que $|V_n W_n| > \varepsilon$ y $|W_n| \leq \Delta$ implican que $|V_n| > \Delta/\varepsilon$, la anterior igualdad desplegada permite concluir que

$$[|V_n W_n| > \varepsilon] = [|V_n| > \Delta/\varepsilon] \cup [|W_n| > \Delta].$$

Por lo tanto, (3.3.4) implica que

$$\begin{aligned} P[|V_n W_n| > \varepsilon] &= P[|V_n| > \Delta/\varepsilon] + P[|W_n| > \Delta] \\ &\leq P[|V_n| > \Delta/\varepsilon] + \delta/2, \quad n > N^*. \end{aligned}$$

Como $V_n \xrightarrow{\text{P}} 0$, existe un entero N^{**} tal que $P[|V_n| > \Delta/\varepsilon] < \delta/2$ si $n > N^{**}$, lo cual, al combinarse con la anterior relación permite establecer que

$$P[|V_n W_n| > \varepsilon] < \delta, \qquad \text{si } n > N,$$

donde $N = \max\{N^*, N^{**}\}$, mostrando que $P[|V_n W_n| > \varepsilon] \to 0$; puesto que $\varepsilon > 0$ es arbitrario, se sigue que $V_n W_n \xrightarrow{\text{P}} 0$.

(c) Suponga ahora que $V_n \xrightarrow{\text{P}} 1$ y $W_n \xrightarrow{\text{d}} W$. En este caso, escriba

$$V_n W_n = W_n + (V_n - 1) W_n$$

y note que $(V_n - 1) \xrightarrow{\text{P}} 0$, de manera que la parte (b) arroja que $(V_n - 1)W_n \xrightarrow{\text{P}} 0$, y combinando este hecho con la parte (a) y la convergencia $W_n \xrightarrow{\text{d}} W$, se desprende que $V_n W_n = W_n + (V_n - 1)W_n \xrightarrow{\text{d}} W$, concluyendo la demostración. □

3.4. Un Ejemplo

Los teoremas de la sección precedente son fundamentales en estudio de las distribuciones límite de una gran cantidad de estimadores en estadística. El punto de partida para aprovechar toda la potencia de los Teoremas 3.2.1–3.3.1, es el teorema central de límite que se estudiará en el siguiente capítulo, pero por el momento se ilustrará la aplicación de los resultados precendentes para establecer un hecho clásico, a saber, la convergencia de la distribución t hacia la distribución normal conforme el número de grados de libertad aumenta.

Ejemplo 3.4.1. Suponga que $X_1, X_2, X_3, \ldots$ son variables aleatorias independientes que tienen como distribución común a la normal con media μ y varianza $\sigma^2 > 0$:

$$X_1, X_2, X_3, \ldots \text{son i.i.d. } \mathcal{N}\left(\mu, \sigma^2\right)$$

En este caso, se sabe que $\overline{X}_n \sim \mathcal{N}\left(\mu, \sigma^2\right)$, y entonces

$$\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \sim \mathcal{N}\left(0, 1\right)$$

vea, por ejemplo, Dudewicz y Mishra (1998). La variable aleatoria en el lado izquierdo de esta expresión es la media estandarizada y, verbalmente, el anterior enunciado establece que al tomar una muestra de tamaño n de una distrinución normal, la media estandarizada tiene distribución normal estándar sin importar el valor de n. Observe que la anterior relación desplegada implica que

$$\sqrt{n}\frac{\overline{X}_n-\mu}{\sigma}\xrightarrow{\mathrm{d}} Z, \quad \text{donde } Z\sim\mathcal{N}(0,1) \tag{3.4.1}$$

Cuando se busca construir un intervalo de confianza para μ, es conveniente sustituir σ por su estimador S_n basado en $X_1, X_2, \ldots, X_n$:

$$S_n^2=\frac{1}{n-1}\sum_{k=1}^{n}(X_i-\overline{X}_n)^2$$

Como se mostró en el Ejemplo 2.4.1(ii), la convergencia

$$\sum_{k=1}^{n}(X_i-\overline{X}_n)^2/n\xrightarrow{\mathrm{P}}\sigma^2$$

ocurre, y entonces

$$S_n=\frac{n}{n-1}\left[\frac{1}{n}\sum_{k=1}^{n}(X_i-\overline{X}_n)^2\right]\xrightarrow{\mathrm{P}}\sigma^2,$$

por el Teorema 3.3.1. Por lo tanto, como la función raíz cuadrada es continua en $[0,\infty)$, se tiene que

$$S_n=\sqrt{S_n^2}\xrightarrow{\mathrm{P}}\sqrt{\sigma^2}=\sigma.$$

por el Teorema 2.1.1. Aplicando de nueva cuenta ese resultado con la función $g(x)=\sigma/x$, la cual es continua en $(0,\infty)$, se desprende que

$$\frac{\sigma}{S_n}=g(S_n)\xrightarrow{\mathrm{P}}g(\sigma)=1,$$

y entonces, a partir de (3.4.1), una aplicación del Teorema 3.3.1(iii) con $V_n=\dfrac{\sigma}{S_n}$ conduce a

$$\sqrt{n}\frac{\overline{X}_n-\mu}{S_n}=\frac{\sigma}{S_n}\sqrt{n}\frac{\overline{X}_n-\mu}{\sigma}\xrightarrow{\mathrm{d}} Z$$

La distribución de $\sqrt{n}(\overline{X}_n - \mu)/S_n$ es la distribución t con $n-1$ grados de libertad, de manera que la relación precedente puede expresarse verbalmente diciendo que, 'la distribución t se aproxima a la distribución normal estándar conforme el número de grados de libertad aumenta'. $\square$

3.5. Teorema Central del Límite

En esta sección se estudia el caso más importante de la noción de convergencia en distribución, a saber, el teorema central del límite, resultado que es la piedra angular de la teoría estadística. Su importancia se debe a que permite obtener aproximaciones a la distribución de un promedio de variables independientes e idénticamente distribuidas en términos de su media y varianza comunes, sin conocer ningún otro detalle de la distribución de la que provienen los datos, por lo cual es posible utilizarlo en una amplia gama de situaciones. El enunciado del teorema es el siguiente.

Teorema 3.5.1. [Teorema central de límite.] Suponga que $X_1, X_2, \ldots, X_n, \ldots$ es una sucesión de variables aleatorias independientes con la misma distribución, la cual tiene varianza σ^2 finita, y por lo tanto la media μ de las variables aleatorias también es finita. Considere la media muestral

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

y define la media estandarizada $\overline{X}_n^*$ mediante

$$\overline{X}_n^* = \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma}. \tag{3.5.1}$$

Con esta notación

$$\overline{X}_n^* \xrightarrow{\text{d}} \Phi \qquad (3.5.2),$$

donde, siguiendo la convención usual, Φ es la función de distribución de la medida de probabilidad normal estándar en la la recta, es decir,

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz. \tag{3.5.3}$$

Observación 3.5.1. (i) En forma más explícita, la convergencia (3.5.2) establece que

$$\lim_{n\to\infty} P\left[\overline{X}_n^* \leq x\right] = \Phi(x), \quad x \in \mathbb{R},$$

lo cual equivale a

$$\begin{aligned}\lim_{n\to\infty} P\left[a \leq \overline{X}_n^* \leq b\right] &= \Phi(b) - \Phi(a)\\ &= \int_a^b \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz, \quad a \leq b, \quad a, b \in \mathbb{R};\end{aligned}$$

esta convergencia significa que, cuando el tamaño de la muestra n es 'grande', la probabilidad $P[a \leq \overline{X}_n^* \leq b]$ puede aproximarse mediante $\Phi(b) - \Phi(a)$. En general, la aproximación será mejor si la diferencia de a y b respecto a μ medida en términos de la desviación estándar de la distribución es 'moderada', y el tamaño de la muestra es tan 'grande' como sea posible. Si Z es una variable aleatoria con distribución normal estándar, la relación anteriormente desplegada puede escribirse alternativamente como $\lim_{n\to\infty} P\left[a \leq \overline{X}_n^* \leq b\right] = P[a \leq Z \leq b]$, de manera que,

$$P\left[a \leq \overline{X}_n^* \leq b\right] \approx P[a \leq Z \leq b].$$

(ii) El Teorema 3.5.1 permite obtener aproximaciones para eventos expresados en términos de $\overline{X}_n$ en lugar de la media estandarizada $\overline{X}_n^*$. Note que $a \leq \overline{X}_n \leq b$ equivale a $\sqrt{n}(a-\mu)/\sigma \leq \overline{X}_n^* \leq \sqrt{n}(b-\mu)/\sigma$, de tal manera que

$$\begin{aligned}P[a \leq \overline{X}_n \leq b] &= P\left[\sqrt{n}\frac{a-\mu}{\sigma} \leq \overline{X}_n^* \leq \sqrt{n}\frac{b-\mu}{\sigma}\right]\\ &\approx \Phi\left(\sqrt{n}\frac{b-\mu}{\sigma}\right) - \Phi\left(\sqrt{n}\frac{a-\mu}{\sigma}\right) \qquad (3.5.4)\\ &= P\left[\sqrt{n}\frac{a-\mu}{\sigma} \leq Z \leq \sqrt{n}\frac{b-\mu}{\sigma}\right].\end{aligned}$$

de nueva cuenta, es conveniente enfatizar que la anterior aproximación no involucra el conocimiento de la distribución de $\overline{X}_n$, solo es necesario conocer la media y varianza de la distribución de la cual provienen los datos; ésta puede ser continua, discreta, o mixta, sin que sea necesario conocer su tipo o cualquier otra de sus características. □

La aproximación (3.5.4) se ilustra en los siguientes dos ejemplos.

Ejemplo 3.5.1. La resistencia a la ruptura del vidrio templado promedia 14 (medida en miles de libras por pulgada cuadrada) y tiene una desviación estándar de 2.

(a) ¿Cuál es la probabilidad de que el promedio de resistencia a la ruptura de 100 piezas seleccionadas aleatoriamente de este vidrio execeda de 14.5?

(b) Encuentre un intervalo que incluya, con probabilidad 0.95, el promedio de resistencia a la ruptura de 100 piezas de este vidrio seleccionadas aleatoriamente.

Solución. (a) Si $X_1, X_2, \ldots, X_{100}$ son las mediciones de la resistencia de las 100 piezas de vidrio seleccionadas, estos datos pueden pensarse con variables aleatorias independientes con la misma distribución. De acuerdo a la información proporcionada, la esperanza de las variables X_i es $\mu = 14$, mientras que la desviación estándar es $\sigma = 2$. En este contexto, se pide encontrar la probabilidad de que el promedio de las 100 mediciones, esto es $\overline{X}_{100}$, exceda de 14.5; en símbolos, se busca $P[\overline{X}_{100} > 14.5]$, y es conveniente notar que el valor exacto de esta probabilidad depende de la distribución de $\overline{X}_{100}$, y que no se tiene manera de determinar a partir de la información del problema. Sin embargo, el teorema central de límite permite establecer que

$$
\begin{aligned}
P[\overline{X}_{100} > 14.5] &= P\left[\sqrt{n}\frac{\overline{X}_{100} - \mu}{\sigma} > \sqrt{n}\frac{14.5 - \mu}{\sigma}\right] \\
&= P\left[\overline{X}^*_{100} > \sqrt{100}\frac{14.5 - 14}{2}\right] \\
&= P\left[\overline{X}^*_{100} > 2.5\right] \\
&\approx P[Z > 2.5] \ = 1 - \Phi(2.5) = 1 - 0.9938 = 0.0062;
\end{aligned}
$$

así, la probabilidad de que $\overline{X}_{100}$ exceda a 14.5 es, *aproximadamente*, de solo 6 milésimos, conclusión que se obtiene sin conocer la distribución exacta de $\overline{X}_{100}$.

(b) En este punto se pide encontrar números a y b tales que

$$P[a \le \overline{X}_{100} \le b] = .95$$

los cuales no pueden determinarse de exactamente, pues no se conoce la distribución de $\overline{X}_{100}$. Sin embargo, el teorema central de límite permite obtener una solución aproximada al problema. Observe que

$$
\begin{aligned}
P[a \le \overline{X}_{100} \le b] &= P\left[\sqrt{100}\frac{a - 14}{2} \le \sqrt{100}\frac{\overline{X}_{100} - 14}{2} \le \sqrt{100}\frac{b - 14}{2}\right] \\
&= P\left[5(a - 14) \le \overline{X}^*_{100} \le 5(b - 14)\right] \\
&\approx P[5(a - 14) \le Z \le 5(b - 14)]
\end{aligned}
$$

donde Z tiene la distribucón normal estándar, y la aproximación se debe al teorema central de límite. Recordando que

$$P[-1.96 \leq Z \leq 1.96] = 0.95$$

las tres últimas relaciones desplegadas permiten afirmar que

$$\text{Si } 5(a-14) = -1.96 \text{ y } 5(b-14) = 1.96 \text{ entonces} \quad P[a \leq \overline{X}_{100} \leq b] \approx 0.95$$

Resolviendo las ecuaciones para a y b se desprende que $b = 14.392$ y $a = 13.608$. Por lo tanto,

$$P[13.608 \leq \overline{X}_{100} \leq 14.392] \approx 0.95.$$

Luego, aunque no es posible determinar un intervalo cuya probabilidad de contener a $\overline{X}_{100}$ sea exactamente de 0.95, el teorema central de límite permite determinar un intervalo para el que la probabilidad de contener a $\overline{X}_{100}$ es *aproximadamente* de 0.95. Este análisis pone de manifiesto la potencia práctica del teorema central de límite; su empleo permite obtener aproximaciones a probabilidades interesantes que involucran una media muestral *sin conocer la distribución exacta de los datos involucrados.* □

Ejemplo 3.5.2. Una antropóloga desea calcular el promedio de estatura de los hombres de cierta raza. Si se supone que la desviación estándar poblacional es de 2.5 pulgadas y si ella toma una muestra aleatoria de 100 hombres, encuentre la probabilidad de que la diferencia entre la media muestral y la media poblacional no exceda de .5 pulgadas. □

Solución. Las 100 mediciones de estatura $X_1, X_2, \ldots, X_{100}$ conforman una muestra de una distribución con media desconocida y desviación estándar 2.5, y se requiere encontrar $P[|\overline{X}_{100} - \mu| \leq 0.5]$. Como en el ejemplo anterior, no es posible encontrar exactamente esta probabilidad, pero el teorema central de límite permite encontrar una aproximación. Como en este caso el tamaño de la muestra es $n = 100$ y $\sigma = 2$, la media estandarizada es $\overline{X}^*_{100} = \sqrt{100}[\overline{X}_{100} - \mu]/2 = 4[\overline{X}_{100} - \mu]$, y

$$P[|\overline{X}_{100} - \mu| \leq 0.5] = P\left[|\overline{X}^*_{100}| \leq 2\right]] \approx P[|Z| \leq 2] = .9544,$$

donde el último número se obtuvo a partir de la tabla normal. □

Capítulo 4

Funciones Características

En este capítulo se estudia concepto de función característica de una variable aleatoria (o de su distribución), el cual es el instrumento analítico esencial para estudiar la noción convergencia en distribución. Se presentan las propiedades fundamentales de la idea de función característica, incluyendo los principios de unicidad y de continuidad, los cuales se utilizarán posteriormente para establecer el teorema central de límite. La exposición concluye con una expansión básica de la función característica correspondiente a una variable aleatoria con media nula y varianza unitaria.

4.1. Función Característica

En esta sección se presenta el instrumento analítico esencial que se utilizará posteriormente para establecer el teorema central de límite. El eje del argumento es un resultado que se denomina la 'propiedad de continuidad', y que involucra la idea de función característica introducida a continuación.

Definición 4.1.1. La función característica de una variablea aleatoria X se denota mediante $\varphi_X(t)$ y se define mediante

$$\varphi_X(t) := E[e^{itX}], \quad t \in \mathbb{R}. \tag{4.1.1}$$

En el lado derecho de (4.1.1), la variable aleatoria bajo el signo del valor esperado es $e^{itX} = \cos(tX) + i\sin(tX)$, la cual toma valores complejos. Debido a que tanto las funciones seno y coseno toman valores entre -1 y 1, el valor esparado siempre existe, y φ_X está definida y toma un valor finito para todo número real t; note que siempre se tiene que

$$\varphi_X(t) \text{ es continua en todo punto } t \in \mathbb{R}, \text{ y } \varphi_X(0) = 1.$$

Como se establece a continuación, a partir de la función característica de una variable aleatoria X es posible obtener los momentos correspondientes cuando éstos son finitos.

Lema 4.1.1. Suponga que una variable aleatoria X tiene momento finito de orden k para algún entero $k \geq 1$, esto es, que $E[|X|^k] < \infty$. En este caso, las siguientes afirmaciones (i) y (ii) son válidas.

(i) $\varphi_X(t)$ tiene derivadas continuas de ordenes $1, 2, 3, \ldots, k$ en cada punto t y, más aún,

$$\frac{d^r}{dt^r}\varphi(t) = i^r E[X^r e^{itX}], \quad r = 1, 2, \ldots, k, \quad t \in \mathbb{R}.$$

(ii) En particular, las derivadas de $\varphi_X(t)$ de ordenes $1, 2, 3, \ldots, k$ en el punto $t = 0$ están dadas por

$$\left.\frac{d^r}{dt^r}\varphi(t)\right|_{t=0} = \varphi_X^{(r)}(0) = i^r E[X^r] \quad r = 1, 2, \ldots, k$$

Este resultado se desprende de un teorema de la teoría general de la integración, denominado el teorema de convergencia dominada (Fulks 1980, Khuri 2002, Royden 2002). Una demostración detallada del lema precedente puede encontrarse, por ejemplo, en Loève (1984), Billingley (1998), Asf (2002). A continuación se evalúa la función característica de tres distribuciones familiares.

Ejemplo 4.1.1. (i) Suponga que la variable aleatoria X es tal que

$$X \sim \mathit{Binomial}(n, p)\,.$$

En este caso, la función característica está dada por

$$\begin{aligned}
\varphi_X(t) &= E[e^{itX}] \\
&= \sum_{k=1}^{n} e^{itk} \binom{n}{k} p^k (q-p)^{n-k} \\
&= \sum_{k=1}^{n} \binom{n}{k} (e^{it}p)^k (q-p)^{n-k} \\
&= (1-p+pe^{it})^n
\end{aligned}$$

donde la última igualdad se debe a la fórmula para el desarrollo de una potencia positiva de un binomio. Las primeras dos derivadas de φ_X son

$$\begin{aligned}
\frac{d}{dt}\varphi(t) &= n((1-p+pe^{it})^{n-1}(ipe^{it}) \\
\frac{d^2}{dt^2}\varphi(t) &= n(n-1)((1-p+pe^{it})^{n-2}(ipe^{it})^2 + n((1-p+pe^{it})^{n-1}(i^2pe^{it})
\end{aligned}$$

Evaluando en $t = 0$ se desprende que

$$\begin{aligned}
\varphi^{(1)}(0) &= n((1-p+p)^{n-1}(ip) = inp \\
\varphi^{(2)}(0) &= n(n-1)((1-p+p)^{n-2}(ip)^2 + n((1-p+p)^{n-1}(i^2p) \\
&= i^2 n(n-1)p^2 + i^2 np
\end{aligned}$$

y el lema precedente permite concluir que

$$iE[X] = inp \quad \text{y} \quad i^2E[X^2] = i^2[n(n-1)p^2 + np],$$

es decir, $E[X] = np$ y $E[X^2] = n(n-1)p^2 + np = n^2p^2 - np^2 + np = np(1-p) + n^2p^2$, de donde se desprende que $\mathrm{Var}\,[X] = E[X^2] - (E[X])^2 = np(1-p) + n^2p^2 - (np)^2 = np(1-p)$, expresiones que coinciden con las obtenidas directamente por medio de métodos más elementales.

(ii) Suponga que X tiene la distribución *Poisson* (λ). En este caso,

$$\begin{aligned}
\varphi_X(t) &= E[e^{itX}] \\
&= \sum_{k=1}^{\infty} e^{itk} \frac{\lambda^k}{k!} e^{-\lambda k} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{(\lambda e^{it})^k}{k!} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda[e^{it}-1]}
\end{aligned}$$

donde, para establecer la cuarta igualdad, se utilizó, la expansión $e^a = \sum_{k=0}^{\infty} a^k/k!$ con $a = \lambda e^{it}$. A partir de esta fórmula se desprende que

$$\frac{d}{dt}\varphi_X(t) = \varphi_X(t) i\lambda e^{it}$$

y por lo tanto

$$\frac{d^2}{dt^2}\varphi_X(t) = \frac{d}{dt}\varphi_X(t) i\lambda e^{it} + \varphi_X(t) i^2 \lambda e^{it\lambda} = \varphi_X(t)(i\lambda e^{it})^2 + \varphi_X(t) i^2 \lambda e^{it}.$$

Usando que $\varphi_X(0) = 1$, las expresiones anteriores implican que

$$\begin{aligned}\varphi_X^{(1)}(0) &= \frac{d}{dt}\varphi_X(t)\bigg|_{t=0} = i\lambda \\ \varphi_X^{(2)}(0) &= \frac{d^2}{dt^2}\varphi_X(t)\bigg|_{t=0} = i^2[\lambda^2 + \lambda],\end{aligned}$$

y aplicando el Lema 4.1.1 se obtienen las igualdades $iE[X] = i\lambda$, así como $i^2 E[X^2] = i^2[\lambda^2 + \lambda]$, de manera que $E[X] = \lambda$ y $E[X^2] = \lambda^2 + \lambda$; luego, $\mathrm{Var}\,[X] = E[X^2] - (E[X])^2 = [\lambda^2 + \lambda] - \lambda^2 = \lambda$, en concordancia con los fórmulas conocidas. □

A continuación se presenta la función característica de una distribución normal estándar.

Ejemplo 4.1.2. Suponga que Z tiene distribución $\mathcal{N}(0, 1)$. En este caso, $\varphi_Z(t)$ está dada por

$$\varphi_Z(t) = E[e^{itZ}] = \int_{\mathbb{R}} e^{itz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-[z^2 - 2itz]/2}\, dz.$$

Para evaluar esta integral, note que

$$z^2 - 2itz = (z - it)^2 - (it)^2 = (z - it)^2 + t^2$$

donde se ha utilizado que $i^2 = -1$. Por lo tanto,

$$\begin{aligned}\varphi_Z(t) &= E[e^{itZ}] \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-[(z-it)^2 + t^2]/2}\, dz = e^{-t^2/2} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-(z-it)^2/2}\, dz = e^{-t^2/2}.\end{aligned}$$

donde la última igualdad se debe a la relación $\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-(z-it)^2/2}\, dz = 1$. A partir de esta igualdad se desprende que

$$\frac{d}{dt}\varphi_Z(t) = -t\varphi_Z(t)$$

y por lo tanto

$$\frac{d^2}{dt^2}\varphi_Z(t) = -t\frac{d}{dt}\varphi_Z(t) - \varphi_Z(t)$$

Aplicando la igualdad $\varphi_X(0) = 1$, después de evaluar las relaciones anteriores en $t = 0$ se desprende que

$$\varphi_Z^{(1)}(0) = \frac{d}{dt}\varphi_Z(t)\bigg|_{t=0} = 0,$$

y

$$\varphi_Z^{(2)}(0) = \frac{d^2}{dt^2}\varphi_Z(t)\bigg|_{t=0} = -1$$

Estas expresiones permiten concluir, por medio del Lema 4.1.1, que $0 = iE[Z]$ y que $i^2E[Z^2] = -1$. Luego, $E[Z] = 0$ y $E[Z^2] = 1$, de manera que $\mathrm{Var}\,[Z] = 1$, como corresponde a una variable aleatoria con distribución normal estándar. $\square$

Los ejemplos anteriores han mostrado como a partir de la función característica de una variable aleatoria X es posible determinar sus momentos. Sin embargo, la aplicación más importante de la idea de función característica no tiene que ver con esos cálculos, sino con propiedades más profundas, que se presentan a continuación.

4.2. El Principio de Unicidad

En la sección se analizan una importante propiedad de una función característica, la cual se describe verbalmente como sigue:

La función característica de una variable aleatoria *caracteriza* su distribución. De manera más explícita, si dos variables aleatorias X y Y tienen la misma función característica, entonces sus distribuciones coinciden.

El enunciado formal de esta propiedad es el siguiente:

Teorema 4.2.1. . Sean X y Y dos variables aleatorias con funciones características φ_X y φ_Y, respectivamente. En este caso, si

$$\varphi_X(t) = \varphi_Y(t) \quad \text{para todo } t \in \mathbb{R},$$

entonces X y Y tienen la misma distribución, lo cual se expresa escribiendo

$$X \stackrel{\mathrm{d}}{=} Y,$$

y significa que

$$P[X \in J] = P[Y \in J], \qquad \text{para todo intervalo } J \subset \mathbb{R}.$$

Ejemplo 4.2.1. (i) Suponga que por algún medio se ha obtenido que una variable aleatoria Y tiene la función característica

$$\varphi_Y(t) = (.7 + .3e^{it})^9.$$

Recordando el Ejemplo 4.1.1, se sabe que si $X \sim \mathit{Binomial}(9, 0, 3)$, entonces su función característica es $(.7 + .3e^{it})^9$, de manera que $\varphi_Y(t) = \varphi_X(t)$. A partir de este punto, el Teorema 4.2.1 permite concluir que Y y X tienen la misma distribución, es decir, que $Y \sim \mathit{Binomial}(9, 0.3)$.

(ii) Suponga ahora que una variable aleatoria Y satisface que

$$\varphi_Y(t) = e^{7[e^{it}-1]}$$

Como se estableció en el Ejemplo 4.1.1(ii) si $X \sim \mathit{Poisson}(7)$, entonces su función característica está dada por $\varphi_X(t) = e^{7[e^{7it}-1]} = \varphi_Y(t)$, y el Teorema 4.2.1 permite concluir que X y Y tienen la misma distribución, esto es, $Y \sim \mathit{Poisson}(7)$. □

Además de que será útil más adelante, el siguiente lema permite presentar aplicaciones interesantes del teorema de unicidad.

Lema 4.2.1. (i) Para cualquier variable aleatoria X, y cada número real a, las funciones características de X y aX están relacionadas por

$$\varphi_{aX}(t) = \varphi_X(at), \quad t \in \mathbb{R}.$$

(ii) Si X y Y son dos variables aleatorias independientes, entonces la función característica de la suma $X+Y$ es el producto de las funciones características de X y Y, esto es,

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t), \quad t \in \mathbb{R}.$$

Más generalmente,

(iii) Si $Y_1, Y_2, \dots, Y_n$ son variables aleatorias independientes y $a_1, a_2, \dots, a_n$ son números reales, entonces la función característica de la combinación lineal $a_1Y_1 + a_2Y_2 + \cdots + a_nY_n$ está dada por

$$\varphi_{a_1Y_1+a_2Y_2+\cdots+a_nY_n}(t) = \varphi_{Y_1}(a_1t)\varphi_{Y_2}(a_2t)\cdots\varphi_{Y_n}(a_nt), \quad t \in \mathbb{R}.$$

(iv) Si $Y_1, Y_2, \dots, Y_n$ es una muestra aleatoria de una distribución con función característica $\varphi(t)$, entonces, para todos los números reales $a_1, a_2, \dots, a_n$, la combinación lineal $a_1Y_1 + a_2Y_2 + \cdots + a_nY_n$ tiene función característica

$$\varphi_{a_1Y_1+a_2Y_2+\cdots+a_nY_n}(t) = \varphi(a_1t)\varphi(a_2t)\cdots\varphi(a_nt) = \prod_{i=1}^{n}\varphi(a_it), \quad t \in \mathbb{R}.$$

Demostración. (i) Sólo note que $\varphi_{aX}(t) = E[e^{it(aX)}] = E[e^{i(at)X}] = \varphi_X(at)$.

(ii) Observe que

$$\varphi_{X+Y}(t) = E[e^{it(X+Y)}] = E[e^{itX}e^{itY}] = E[e^{itX}]E[e^{itY}] = \varphi_X(t)\varphi_Y(t),$$

donde la tercera igualdad se debe a la independencia de X y Y.

(iii) La función característica de la combinación lineal $a_1Y_1+a_2Y_2+\cdots+a_nY_n$ está dada por

$$\begin{aligned}
\varphi_{a_1Y_1+a_2Y_2+\cdots+a_nY_n}(t) &= E[e^{t(a_1Y_1+a_2Y_2+\cdots+a_nY_n)}] \\
&= E[e^{(a_1t)Y_1}e^{(a_2t)Y_2}\cdots e^{(a_nt)Y_n}] \\
&= E[e^{(a_1t)Y_1}]E[e^{(a_2t)Y_2}]\cdots E[e^{(a_nt)Y_n}] \\
&= \varphi_{Y_1}(a_1t)\varphi_{Y_2}(a_2t)\cdots\varphi_{Y_2}(a_nt)
\end{aligned}$$

donde la independencia de las variables Y_i se utilizó en la tercera etapa.

(iv) Cuando las variables Y_i provienen de una misma distribución con función característica $\varphi(t)$, se tiene que $\varphi_{Y_i}(t) = \varphi(t)$, de manera que la conclusión se desprende directamente de la parte (iii). □

Ejemplo 4.2.2. (i) Suponga que Y tiene función característica

$$\varphi_Y(t) = (.7 + .3e^{i4t})^9.$$

Con esta información, el problema es describir de forma precisa la distribución de Y. Para alcanzar este objetivo, note que si

$$X \sim \mathit{Binomial}(9, 0.3),$$

el Ejemplo 4.2.1(i) permite concluir que $\varphi_X(t) = (.7 + .3e^{it})^9$, y entonces

$$\varphi_{4X}(t) = (.7 + .3e^{i4t})^9 = \varphi_Y(t).$$

A partir del Teorema 4.2.1 se concluye que $4X$ y Y tienen la misma distribución. Así, la información proporcionada permite concluir que Y tiene la misma distribución que una variable $\mathit{Binomial}(9, 0.3)$ multiplicada por 4.

(ii) Ahora suponga que Y tiene función característica

$$\varphi_Y(t) = e^{7[e^{i5t}-1]}$$

Para describir de forma precisa la distribución de Y, note que si $X \sim \mathit{Poisson}(7)$, entonces $\varphi_X(t) = e^{7[e^{it}-1]}$, por el Ejemplo 4.2.1(ii). Por lo tanto, $\varphi_{5X}(t) = e^{7[e^{i5t}-1]} = \varphi_Y(t)$, de tal manera que el Teorema 4.2.1 permite concluir que $5X$ y Y tienen la misma distribución. Luego, a partir de la información proporcionada se concluye que Y tiene la misma distribución que una variable con distribución $\mathit{Poisson}(7)$ multiplicada por 5.

(iii) Sea W una variable aleatoria con función característica dada por

$$\varphi_W(t) = (.7 + .3e^{i4t})^9 e^{7[e^{i5t}-1]}.$$

La distribución de W puede describirse como sigue: De acuerdo a las partes (i) y (ii), $(.7 + .3e^{i4t})^9$ es la función característica de $4X$, donde $X \sim \mathit{Binomial}(9, 0.3)$, mientras que $e^{7[e^{i5t}-1]}$ es la función característica de $5Y$, donde $Y \sim \mathit{Poisson}(7)$. Cuando X y Y son independientes, $4X + 5Y$ tiene función característica $\varphi_{4X+5Y}(t) = (.7 + .3e^{i4t})^9 e^{7[e^{i5t}-1]}$, por el lema anterior, y se concluye que

$$\varphi_W(t) = \varphi_{4X+5Y}(t),$$

y el Teorema 4.2.1 permite concluir que

$$W \stackrel{\mathrm{d}}{=} 4X + 5Y.$$

$\square$

4.3. La Propiedad de Continuidad

En esta sección se presenta otra propiedad crucial relacionada con las funciones características, se denomina 'principio de continuidad', la cual permite caraterizar la convergencia en distribución de una sucesión de variables aleatorias $\{W_n\}$ hacia la variable aleatoria W por medio de la convergencia de las funciones características. Esta propiedad se aplicará posteriormente para demostrar el teorema central de límite, y para demostrar una versión general de la ley de los grandes números.

Teorema 4.3.1. Sea $\{W_n\}$ un sucesión de variables aleatorias y considere una variable aleatoria adicional W. En este caso, la sucesión $\{W_n\}$ converge en distribución a W si, y sólo si, la sucesión de funciones características $\{\varphi_{W_n}(t)\}$ converge a $\varphi_W(t)$ para todo número real t:

$$W_n \xrightarrow{\mathrm{d}} W \iff \lim_{n\to\infty} \varphi_{W_n}(t) = \varphi(t) \quad t \in \mathbb{R}. \tag{4.3.1}$$

Conjuntamente con el Teorema 4.2.1, este resultado es el centro de la teoría asintótica de variables aleatorias y su demostración puede verse, por ejemplo, en Billingley (1998), Borovkov (1999), o Loève (1984). Note que el criterio (4.3.1) permite estudiar la convergencia en distribución de variables aleatorias analizando un problema más familiar, la convergencia de las funciones características involucradas. Como se menciónо anteriormente, el Teorema 4.3.1 se utilizará en el siguiente capítulo para establecer el teorema central de límite. Por el momento, es conveniente estudiar un caso particular, a saber, el de convergencia de variables aleatorias hacia una constante.

Corolario 4.3.1. Una sucesión de variables aleatorias $\{W_n\}$ converge en probabilidad hacia una constante c si, y sólo si la sucesión $\{\varphi_{W_n}(t)\}$ converge a e^{ict} para todo $t \in \mathbb{R}$:

$$W_n \xrightarrow{\mathrm{P}} c \iff \lim_{n\to\infty} \varphi_{W_n}(t) = e^{ict}, \quad t \in R.$$

Demostración. El argumento depende de dos hechos. Primero, note que los Teoremas 3.2.1 y 3.2.2 establecen que

$$W_n \xrightarrow{\mathrm{P}} c \iff W_n \xrightarrow{\mathrm{d}} c.$$

En segundo lugar, note que si X es la variable aleatoria constante c—es decir, $P[X = c] = 1$—entonces $\varphi_c(t) = \varphi_X(t) = E[e^{itX}] = e^{ict}$. Combinando esta igualdad y la anterior relación desplegada, el criterio (4.3.1) permite concluir que $\{W_n\}$ converge en probabilidad a la constante c si, y sólo si, $\varphi_{W_n}(t) \to e^{ict}$ para todo $t \in \mathbb{R}$, estableciendo la conclusión deseada. □

En las aplicaciones de este resultado, con frecuencia es útil tener en mente el siguiente hecho auxiliar.

Lema 4.3.1. Suponga que $\{A_n\}$ es una sucesión de números (posiblemente) complejos tales que

$$A_n \to A \quad \text{si } n \to \infty.$$

En este caso,

$$\lim_{n\to\infty} \left(1 + \frac{A_n}{n}\right)^n = e^A$$

Demostración. Tomando logaritmos, la convergencia que se desa establecer equivale a

$$\lim_{n\to\infty} n \log\left(1 + \frac{A_n}{n}\right) = A. \tag{4.3.2}$$

Para verificar esta igualdad, recuerde que

$$\frac{d}{dz}\log(z) = \frac{1}{z}$$

de manera que

$$\frac{\log(z+h) - \log(z)}{h} \to \frac{1}{z}, \quad \text{cuando } h \to 0.$$

Usando esta relación con $z = 1$ y recordando que $\log(1) = 0$ se obtiene que

$$\frac{\log(1+h)}{h} = \frac{\log(1+h) - \log(1)}{h} \to \frac{1}{1} = 1, \quad \text{cuando } h \to 0;$$

por lo tanto

$$\lim_{n\to\infty} \frac{1}{A_n}\left[n \log\left(1 + A_n/n\right)\right] = \lim_{n\to\infty} \frac{\log(1 + A_n/n)}{A_n/n} = 1,$$

relación que equivale a (4.3.2), pues $A_n \to A$. □

El propósito del siguiente ejemplo es ilustrar la aplicación del Corolario 4.3.1y del Lema 4.3.1.

Ejemplo 4.3.1. Para cada entero positivo n, considere una variable aleatoria Y_n con distribución binomial negativa, de tal manera que

$$P[Y_n = k] = \binom{n+k-1}{k-1} p^{n-1}(1-p)^{k-n}, \quad k = n, n+1.n+2, \ldots;$$

note que se tiene la igualdad

$$\sum_{k=1}^{\infty} \binom{n+k-1}{k-1} p^{n-1}(1-p)^{k-n} = 1, \quad |p| < 1$$

la cual equivale a

$$\sum_{k=1}^{\infty} \binom{n+k-1}{k-1} q^k = \frac{q^n}{(1-q)^{n-1}} = (1-q)\left(\frac{q}{1-q}\right)^n, \quad \text{donde } q = 1-p. \tag{4.3.3}$$

A continuación se analizará la convergencia de la sucesión $\{Y_n/n\}$ hacia una constante usando el Corolario 4.3.1. Como punto de partida, se calculará la función característica de Y_n. Note que

$$\begin{aligned}
\varphi_{Y_n}(t) &= E[e^{itY_n}] \\
&= \sum_{k=n}^{\infty} e^{itk} \binom{n+k-1}{k-1} p^{n-1}(1-p)^{k-n} \\
&= \frac{p^{n-1}}{(1-p)^n} \sum_{k=n}^{\infty} e^{itk} \binom{n+k-1}{k-1} (1-p)^k \\
&= \frac{1}{p}\left(\frac{p}{1-p}\right)^n \sum_{k=n}^{\infty} \binom{n+k-1}{k-1} [e^{it}q]^k \\
&= \frac{1}{p}\left(\frac{1-q}{q}\right)^n \sum_{k=n}^{\infty} \binom{n+k-1}{k-1} [e^{it}q]^k
\end{aligned}$$

Para evaluar la última sumatoria, utilice (4.3.3) con $e^{it}q$ en lugar de q para obtener que

$$\sum_{k=n}^{\infty} \binom{n+k-1}{k-1} [e^{it}q]^k = (1-e^{it}q)\left(\frac{e^{it}q}{1-e^{it}q}\right)^n$$

de manera que

$$\begin{aligned}
\varphi_{Y_n}(t) &= \frac{1-e^{it}q}{p}\left(\frac{1-q}{q}\right)^n\left(\frac{e^{it}q}{1-e^{it}q}\right)^n \\
&= \frac{1-e^{it}q}{p}\left(\frac{(1-q)e^{it}}{1-e^{it}q}\right)^n \\
&= \frac{1-e^{it}q}{p}\left(1+f(t)\right)^n
\end{aligned} \tag{4.3.4}$$

donde

$$f(t) = \frac{(1-q)e^{it}}{1-e^{it}q} - 1 = \frac{e^{it}-1}{1-e^{it}q};$$

observe que $f(0) = 0$ y que

$$f'(0) = \lim_{h\to 0}\frac{f(h)-f(0)}{h} = \lim_{h\to 0}\frac{f(h)}{h} = \lim_{h\to 0}\frac{(e^{ih}-1)/h}{1-e^{ih}q} = \frac{i}{1-q} = \frac{i}{p}. \tag{4.3.5}$$

Ahora considere la variable aleatoria $W_n = Y_n/n$ y note que su función característica está dada por

$$\varphi_{W_n}(t) = \varphi_{Y_n/n}(t) = \varphi_{Y_n}(t/n) = \frac{1-e^{it/n}q}{p}\left(1+f(t/n)\right)^n,$$

donde la segunda igualdad se desprende del Lema 4.2.1(i), y la relación (4.3.4) se utilizó en la última etapa. Note ahora que

$$f(t/n) = A_n/n, \quad \text{donde } A_n = nf(t/n) = t\frac{f(t/n)}{t/n} \to t\frac{i}{p} = \frac{it}{p}, \tag{4.3.6}$$

Con esta especificación de A_n, la anterior expresión para $\varphi_{W_n}(t)$ equivale a

$$\varphi_{W_n}(t) = \frac{1-e^{it/n}q}{p}\left(1+\frac{A_n}{n}\right)^n.$$

Por otro lado, (4.3.6) y el Lema 4.3.1 conducen a

$$\lim_{n\to\infty}\left(1+\frac{A_n}{n}\right)^n = e^{it/p},$$

mientras que la continuidad de la función exponencial implica que

$$\lim_{n\to\infty}\frac{1-e^{it/n}q}{p} = \frac{1-e^{i0}q}{p} = \frac{1-q}{p} = 1.$$

Combinando las tres últimas relaciones desplegadas se obtiene

$$\lim_{n\to\infty} \varphi_{Y_n/n}(t) = \lim_{n\to\infty} \varphi_{W_n}(t) = e^{it/p},$$

y el Corolario 4.3.1 permite concluir que $Y_n/n \xrightarrow{\text{P}} 1/p$. □

4.4. Variables Estandarizadas

En el siguiente capítulo el Teorema 4.3.1 será usado para establecer el teorema central de límite. El argumento involucra el análisis de la función característica de una variable con media cero y varianza 1, la cual se denomina *estandarizda*. A continuacún se enuncia el resultado técnico auxiliar que se empleará en la demostración.

Lema 4.4.1. Suponga que Y es una variable aleatoria con $E[Y] = 0$ y $E[Y^2] = \text{Var}\,[Y] = 1$. En este caso, la función característica de Y puede escribirse como

$$\varphi_Y(t) = 1 - \frac{t^2}{2}(1 + R_n(t)), \tag{4.4.1}$$

donde $R_n(t)$ satisface

$$\lim_{t\to 0} R_n(t) = 0. \tag{4.4.2}$$

Demostración. Defina $R_n(t)$ mediante

$$R_n(t) := -\frac{\varphi_Y(t) - 1 + t^2/2}{t^2/2},$$

y note que (4.4.1) se satisface. Ahora se verificará que (4.4.2) también ocurre. Observe que $\lim_{t\to 0}[\varphi_Y(t) - 1 + t^2/2] = \varphi_Y(0) - 1 + 0 = 1$, pues $\varphi(0) = 1$, mientras que es claro que $\lim_{t\to 0} t^2/2 = 0$. Luego, al tratar de calcular el límite conforme $t \to 0$ del cociente que define $R_n(t)$ como 'el cociente de los límites', se llega a la forma indeterminada $0/0$, de manera que, usando la regla de L'hopital, el límite del cociente coincide con el límite del cociente de las derivadas:

$$\lim_{t\to 0} R_n(t) = -\lim_{t\to 0} \frac{\varphi_Y(t) - 1 + t^2/2}{t^2/2} = -\lim_{t\to 0} \frac{\varphi_Y'(t) + t}{t}.$$

De nueva cuenta, al analizar el último cociente en esta expresión, se tiene que el numerador satisface $\lim_{t\to 0} \varphi_Y'(t) + t = \varphi_Y'(0) + 0 = 0$, pues $0 = iE[Y] =$

$\varphi'_Y(0)$, por el Lema 4.1.1, mientras que es claro que el denominador satisface $\lim_{t\to 0} t = 0$. Por lo tanto, al intentar calcular el límite del último cociente en la anterior relación desplegada, se llega a la forma indeterminada $0/0$ y, de nuevo, la regla de L'hopital conduce a que el límite del cociente coincide con el límite del cociente de las derivadas:

$$\lim_{t\to 0} \frac{\varphi'_Y(t) + t}{t} = \lim_{t\to 0} \frac{\varphi''_Y(t) + 1}{1} = \lim_{t\to 0}[\varphi''_Y(t) + 1] = \varphi''_Y(0) + 1.$$

Recuerde ahora que $i^2 E[Y^2] = \varphi''_Y(0)$, por el Lema 4.1.1. Como $i^2 = -1$ y $E[Y^2] = 1$, se concluye que $\varphi''_Y(0) = -1$, de manera que

$$\varphi''_Y(0) + 1 = 0.$$

Combinando las tres relaciones desplegadas anteriormente se desprende que

$$\lim_{t\to 0} R_n(t) = 0,$$

estableciendo (4.4.2). □

Capítulo 5

Aplicaciones de la Propiedad de Continuidad

En este capítulo se analizan apliaciones del principio de continuidad. La primera de ellas es la demostración del teorema central de límite, y posteriormente se obtienen aproximaciones para una distribución geométrica en términos de variables exponenciales, y para una distribución binomial usando una distribución de Poisson. También se establece la aproximación clásica para un distribución binomial por medio de una distribución normal, y la exposición concluye con una versión general de la ley de los grandes números obtenida por medio del teorema de continuidad.

5.1. Demostración del Teorema Central de Límite

En esta sección se establece el Teorema Central de límite enunciado en la Sección 3.5. Recuerde que dicho resultado establece que, si $X_1, X_2, X_3, \ldots$ son variables aleatorias independientes con la misma distribución con media μ y varianza σ^2 finita y positiva, entonces la media estandarizada $\overline{X}_n^*$ converge en distribución hacia la meida de probabilidad normal estándar en $\mathbb{R}$, esto es,

$$\overline{X}_n^* \xrightarrow{\mathrm{d}} \Phi;$$

vea (3.5.2) y (3.5.3). Como la función característica de la distribución normal estándar es $e^{-t^2/2}$, el Teorema 4.3.1 muestra que la anterior convergencia en distribución equivale a

$$\lim_{n\to\infty} \varphi_{\overline{X}_n^*}(t) = e^{-t^2/2}. \tag{5.1.1}$$

Demostración del Teorema 3.5.1. Se verificará (5.1.1) en dos fases. Primero, se expresará $\overline{X}_n^*$ en términos de variables aleatorias de media cero y varianza 1. Con ese fin, defina

$$Y_i = \frac{X_i - \mu}{\sigma}, \quad i = 1, 2, 3, \ldots,$$

y note que las variables Y_i son independientes con la misma distribución, lo cual se debe a que las X_js tienen esta propiedad. Más aún,

$$E[Y_i] = E\left[\frac{X_i - \mu}{\sigma}\right] = \frac{E\left[X_i\right] - \mu}{\sigma} = 0,$$

y

$$\mathrm{Var}\left[Y_i\right] = \mathrm{Var}\left[\frac{X_i - \mu}{\sigma}\right] = \frac{\mathrm{Var}\left[X_i\right]}{\sigma^2} = 1.$$

Note ahora que la especificación de las variables Y_i implica que

$$\frac{Y_1 + Y_2 + \cdots + Y_n}{n} = \overline{Y}_n = \frac{\overline{X}_n - \mu}{\sigma} \tag{5.1.2}$$

de manera que

$$\frac{Y_1 + Y_2 + \cdots + Y_n}{\sqrt{n}} = \sqrt{n}\overline{Y}_n = \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} = \overline{X}_n^*.$$

A continuación, denote mediante $\varphi(t)$ a la función característica de la distribución común de las variables Y_i, de manera que $\varphi_{Y_i}(t) = \varphi(t)$ para todo $i = 1, 2, 3, \ldots$ y $t \in \mathbb{R}$. Combinado la expresión (5.1.2) con el Lema 4.2.1(iii), se desprende que

$$\varphi_{\overline{X}_n^*}(t) = \varphi_{\frac{Y_1+Y_2+\cdots+Y_n}{\sqrt{n}}}(t) = \prod_{i=1}^{n} \varphi_{Y_i}(t/\sqrt{n}) = \varphi(t/\sqrt{n})^n \tag{5.1.3}$$

Puesto que $\varphi(t)$ corresponde a una distribución de media nula y varianza unitaria, el Lema 4.4.1 permite establecer que

$$\varphi(h) = 1 - \frac{t^2}{2}(1 + R_n(h)), \quad \text{donde } \lim_{h\to 0} R_n(h) = 0,$$

de manera que

$$\varphi(t/\sqrt{n}) = 1 - \frac{t^2}{2n}(1 + R_n(t/\sqrt{n})), \quad \text{donde } \lim_{n\to 0} R_n(t/\sqrt{n}) = 0.$$

Considerando a $t \in \mathbb{R}$ fijo, por el momento, defina

$$A_n := -\frac{t^2}{2}(1 + R_n(t/\sqrt{n})$$

y note que

$$\varphi(t/\sqrt{n}) = 1 + \frac{A_n}{n},$$

de tal manera que (5.1.3) puede escribirse como

$$\varphi_{\overline{X}^*_n}(t) = \left(1 + \frac{A_n}{n}\right)^n \quad \text{donde} \quad \lim_{n\to\infty} A_n = -\frac{t^2}{2},$$

y entonces, una aplicación del Lema 4.3.1 con $A = -t^2/2$ permite concluir que

$$\lim_{n\to\infty} \varphi_{\overline{X}^*_n}(t) = e^{-t^2/2},$$

estableciendo (5.1.1), lo cual concluye la demostración. □

5.2. Aproximación Exponencial a una Distribución Geométrica

En esta sección se muestra otra aplicación de la propiedad de continuidad en el Teorema 4.3.1, a saber, que una medida de probabilidad exponencial en $\mathbb{R}$ puede obtenerse como el límite en distribución de variables aleatorias con distribución gemoétrica. Primero, se encontrarán las funciones características correspondientes.

Lema 5.2.1. (i) Suponga que X tiene distribución *Geometrica*(p), esto es,

$$P[X = k] = (1-p)^{k-1}p, \quad x = 1, 2, 3, \ldots$$

En este caso,

$$\varphi_X(t) = \frac{pe^{it}}{1-(1-p)e^{it}}, \quad t \in \mathbb{R}.$$

(ii) Si X tiene distribución *Exponencial*(λ), la cual corresponde a la densidad

$$f(x;\lambda) = e^{-\lambda x} I_{(0,\infty)}(x),$$

entonces

$$\varphi_X(t) = \frac{\lambda}{\lambda - it}, \quad t \in \mathbb{R}.$$

Demostración. (i) Note que

$$\begin{aligned}\varphi_X(t) &= E[e^{itX}] \\ &= \sum_{k=1}^{\infty} e^{itk}(1-p)^{k-1}p = pe^{it}\sum_{k=1}^{\infty}[e^{it}(1-p)]^{k-1} = \frac{pe^{it}}{1-e^{it}(1-p)},\end{aligned}$$

donde la última igualdad se debe a la fórmula para la serie geométrica $\sum_{k=0}^{\infty} r^k = 1/(1-r)$, válida cuando $|r| < 1$.

(ii) Si $X \sim$ *Exponencial* (λ), entonces

$$\begin{aligned}\varphi_X(t) &= E[e^{itX}] \\ &= \int_0^{\infty} e^{itx}\lambda e^{-\lambda x}\,dx = \lambda\int_0^{\infty} e^{-[\lambda - it]x}\,dx = \frac{\lambda}{\lambda - it}\end{aligned}$$

El principal resultado de esta sección es el siguiente.

Teorema 5.2.1. Suponga que, para cada $n = 1, 2, 3, \ldots$,

$$Y_n \sim \textit{Geometrica}\,(p_n)\,, \quad \text{donde } p_n > 0 \text{ y } \quad \lim_{n\to\infty} p_n = 0,$$

y sea Y una variable aleatoria con distribución *Exponencial* (1). Con esta notación se tiene que

$$p_n Y_n \xrightarrow{\text{d}} Y.$$

Demostración. De acuerdo al Teorema 4.3.1, debe verificarse que

$$\lim_{n\to\infty} \varphi_{p_n Y_n}(t) = \varphi_Y(t), \quad t \in \mathbb{R}. \tag{5.2.1}$$

Para alcanzar este objetivo note que, combinando los Lemas 5.2.1(i) y 4.2.1(i) se tiene que

$$\begin{aligned}\varphi_{p_n Y_n}(t) &= \varphi_{Y_n}(p_n t) \\ &= \frac{p_n e^{ip_n t}}{1 - e^{ip_n t}(1-p_n)} \\ &= \frac{p_n e^{ip_n t}}{p_n e^{ip_n t} - [e^{ip_n t} - 1]} = \frac{e^{ip_n t}}{e^{ip_n t} - [e^{ip_n t} - 1]/p_n}\end{aligned}$$

.

A continuación, observe que

$$\lim_{n\to\infty} e^{ip_n t} = 1,$$

puesto que $p_n \to 0$. Por otro lado, observe que la derivada respecto a h de la función $g(h) = e^{iht}$ es $g'(h) = ite^{iht}$, de modo que $g'(0) = it$. Usando que $g(0) = 1$, recordando la definición de derivada se tiene que

$$\begin{aligned} 1 = g'(0) &= \lim_{h\to 0} \frac{g(h) - g(0)}{h} \\ &= \lim_{h\to 0} \frac{e^{ith} - 1}{h} = \lim_{n\to\infty} \frac{e^{itp_n} - 1}{tp_n} = \frac{1}{t} \lim_{n\to\infty} \frac{e^{itp_n} - 1}{p_n}, \end{aligned}$$

de manera que

$$\lim_{n\to\infty} \frac{e^{itp_n} - 1}{p_n} = t.$$

Combinando las cuatro últimas relaciones desplegadas se desprende que

$$\begin{aligned} \lim_{n\to\infty} \varphi_{p_n Y_n}(t) &= \lim_{n\to\infty} \frac{e^{ip_n t}}{e^{ip_n t} - [e^{ip_n t} - 1]/p_n} \\ &= \frac{\lim_{n\to\infty} e^{ip_n t}}{\lim_{n\to\infty} e^{ip_n t} - \lim_{n\to\infty} [e^{ip_n t} - 1]/p_n} \\ &= \frac{1}{1 - it} \end{aligned}$$

Para concluir, note que como Y tiene distribución *Exponencial*(1), se tiene que $\varphi_Y(t) = 1/(1 - it)$, así que la anterior convergencia equivale a (5.2.1), completando la demostración. □

5.3. Aproximación de Poisson a una Distribución Binomial

En esta seccíon se analiza el comportamiento asintótico de variables Y_n con distribución *Binomial*(n, p_n). Cuando el número de ensayos n y la probabilidad de éxito p_n son tales que el número de ensayos n crece hacia infinito, y producto np_n es 'moderado'. En este contexto, se demostrará que la distribución de Y_n puede aproximarse mediante una distribución de Poisson. El enunciado preciso de esta conclusión es como sigue:

Teorema 5.3.1. Suponga que, para cada $n = 1, 2, 3, \ldots$, la variable aleatoria Y_n tiene distribucíon *Binomial*(n, p_n), donde

$$\lim_{n\to\infty} np_n = \lambda \in (0, \infty), \tag{5.3.1}$$

y sea Y una variable aleatoria con distribución *Poisson*(λ). En este caso,

$$Y_n \xrightarrow{\text{d}} Y.$$

Demostración. De acuerdo al Ejemplo 4.1.1, se tiene que las funciones características de las variables Y y Y_n están dadas por

$$\begin{aligned} \varphi_Y(t) &= e^{\lambda[e^{it}-1]} \\ \varphi_{Y_n}(t) &= (1 - p_n + p_n e^{it})^n \end{aligned} \tag{5.3.2}$$

Ahora se estudiará el comportamiento de $\varphi_{Y_n}(t)$ conforme n crece. Con este fin, sea $t \in \mathbb{R}$ un número arbitrario pero fijo, y observe que

$$1 - p_n + p_n e^{it} = 1 + p_n[e^{it} - 1] = 1 + \frac{A_n}{n}, \quad \text{donde} \quad A_n := np_n[e^{it} - 1].$$

Con esta notación, la fórmula para $\varphi_{Y_n}(t)$ en (5.3.2) equivale a

$$\varphi_{Y_n}(t) = \left(1 + \frac{A_n}{n}\right)^n.$$

Observando que (5.3.1) y la especificación de A_n implican que $A_n \to \lambda[e^{it} - 1]$ cuando n tiende a ∞, el Lema 4.3.1 permite concluir que

$$\lim_{n\to\infty} \varphi_{Y_n}(t) = \lim_{n\to\infty} \left(1 + \frac{A_n}{n}\right)^n = e^{\lambda[e^{it}-1]} = \varphi_Y(t),$$

donde se utilizó la expresión para $\varphi_Y(t)$ en (5.3.2) para establecer la tercera igualdad. A partir de esta convergencia, el Teorema 4.3.1 implica que $Y_n \xrightarrow{\text{d}} Y$, concluyendo la demostración. □

En el Teorema anterior las variables binomiales Y_n son tales que el parámetro de éxito p_n asociado con Y_n se aproxima a cero conforme n crece, de tal manera que el producto np_n se aproxima a una constante positiva. En la siguiente sección se estudian variables binomiales, en las que el parámetro de éxito permanece constante.

5.4. Aproximación Normal a una Distribución Binomial

En esta sección se establece un resultado clásico que muestra bajo que condiciones es posible aproximar una distribución binomial por una distribución normal.

Teorema 5.4.1. Considere una sucesión $\{Y_n\}$ tales que

$$Y_n \sim \mathit{Binomial}\,(n, p)\,,$$

y sea Y_n^* la versión estandarizada de Y_n, esto es,

$$Y_n^* = \frac{Y_n - np}{\sqrt{np(1-p)}}.$$

Con esta notación

$$Y_n^* \xrightarrow{\mathrm{d}} Z,$$

donde

$$Z \sim \mathcal{N}\,(0, 1)\,.$$

Demostración. Aunque es posible establecer este resultado directamente de la propiedad de continuidad en el Teorema 4.3.1, es más simple utilizar el teorema central de límite. Con este fin, considere una sucesión $\{X_n\}$ de variables aleatorias independientes e idénticamente distribuidas con distribución de $\mathit{Bernoulli}\,(p)$, de manera que

$$X_i + X_2 + \cdots + X_n \sim \mathit{Binomial}\,(n, p)$$

Por lo tanto,

$$Y_n \overset{\mathrm{d}}{=} X_i + X_2 + \cdots + X_n,$$

de donde se desprende que

$$Y_n^* = \frac{Y_n - np}{\sqrt{n(1-p)}} \overset{\mathrm{d}}{=} \frac{X_i + X_2 + \cdots + X_n - np}{\sqrt{np(1-p)}} = \sqrt{n}\frac{\overline{X}_n - p}{\sqrt{p(1-p)}} = \overline{X}_n^*.$$

Por el teorema central de límite, $\overline{X}_n^* \xrightarrow{\text{d}} Z$, y puesto que Y_n^* tiene la misma distribución que $\overline{X}_n^*$, se concluye que

$$Y_n^* \xrightarrow{\text{d}} Z,$$

completando la demostración. □

5.5. Otra Versión de la Ley de los Grandes Números

Como última aplicación del Teorem 4.3.1, se establecerá ahora una versión refinada de la ley de los grandes números que se enunció en el Teorema 2.4.1.

Teorema 5.5.1. Suponga que $X_1, X_2, \ldots, X_n, \ldots$ son variables aleatorias independientes e idénticamente distribuidas y suponga que la distribución común tiene media finita μ. Sea $\overline{X}_n$ la media muestral de los primero n datos, esto es,

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Con esta notación,

$$\overline{X}_n \xrightarrow{\text{P}} \mu. \tag{5.5.1}$$

Note que, en contraste con el Teorema 2.4.1, en el presente resultado *no se supone que la varianza de la distribución común de las variables X_i sea finita.*

Demostración. Para obtener la convergencia (5.5.1), note que los Teoremas 3.2.1 y 3.2.2 implican que (5.5.1) es equivalente a

$$\overline{X}_n \xrightarrow{\text{d}} \mu, \tag{5.5.2}$$

y es esta convergencia la que se demostrará. Con este fin, note que el Lema 4.2.1(iii) permite concluir que

$$\varphi_{\overline{X}_n}(t) = \varphi_{X_/n+X_2/n+\cdots+X_n/n}(t) = \prod_{i=1}^{n} \varphi_{X_i}(t/n) = \varphi(t/n)^n$$

donde $\varphi(t)$ es la función característica de la distribución común de las variables X_i. Seleccione un número fijo $t \in \mathbb{R}$ y defina A_n mediante

$$A_n := n[\varphi(t/n) - 1], \tag{5.5.3}$$

de tal manera que

$$\varphi(t/n) = 1 + \frac{A_n}{n}.$$

Con esta notación,

$$\varphi_{\overline{X}_n}(t) = \left(1 + \frac{A_n}{n}\right)^n. \tag{5.5.4}$$

Observe ahora que, como las variables X_i tienen media finita, su función característica tiene derivada de primer orden, y $\varphi'(0) = i\mu$, por el Lema 4.1.1. Por lo tanto,

$$\begin{aligned}
i\mu = \varphi(0) &= \lim_{h\to 0} \frac{\varphi(h) - \varphi(0)}{h} \\
&= \lim_{h\to 0} \frac{\varphi(h) - 1}{h} \\
&= \lim_{n\to\infty} \frac{\varphi(t/n) - 1}{t/n} \\
&= \lim_{n\to\infty} n[\frac{\varphi(t/n) - 1}{t} \\
&= \lim_{n\to\infty} \frac{n[\varphi(t/n) - 1]}{t} = \frac{\lim_{n\to\infty} A_n}{t},
\end{aligned}$$

de donde se desprende que ,

$$\lim_{n\to\infty} A_n = it\mu.$$

Combinando esta convergencia con (5.5.4), el Lema 4.3.1 implica que

$$\lim_{n\to\infty} \varphi_{\overline{X}_n}(t) = \lim_{n\to\infty} \left(1 + \frac{A_n}{n}\right)^n = e^{it\mu},$$

lo cual, por el Corolario 4.3.1, equivale a (5.5.2), concluyendo el argumento.
□

Literatura Citada

.

[1].T. M. Apostol (1980), Mathematical Analysis, *Addison Wesley*, Reading, Massachusetts.

[2].A. A. Borovkov (1999), Mathematical Statistics, *Gordon and Breach*, New York

[3].E. Dudewicz y S. Mishra (1998). Mathematical Statistics, *Wiley*, New York.

[4].W. Fulks (1980), Cálculo Avanzado, *Limusa*, México, D. F.

[5].F. A. Graybill (2000), Theory and Application of the Linear Model, *Duxbury*, New York.

[6].F. A. Graybill (2001), Matrices with Applications in Statistics *Duxbury*, New York.

[7].D. A. Harville (2008), Matrix Algebra Form a Statistician's Perspective, *Sprin-ger-Verlag*, New York.

[8].A. I. Khuri (2002), Advanced Calculus with Applications in Statistics, *Wiley*, New York.

[9].E. L. Lehmann and G. B. Casella, (1998), Theory of Point Estimation, Springer, New York.

[10].M. Loève (1984), Probability Theory, I, Springer-Verlag, New York.

[11].D. C. Montgomery (2011), Introduction to Statistical Quality Control, 6th Edition, Wiley, New York.

[12].A. M. Mood, D. C. Boes and F. A. Graybill (1984), Introduction to the Theory of Statistics, McGraw-Hill, New York.

[13].W. Rudin (1984), Real and Complex Analysis, *McGraw-Hill*, New York.

[14].H. L. Royden (2003), Real Analysis, *MacMillan*, London.

[15].J. Shao (2010), Mathematical Statistics, *Springer*, New York.

[16].D. Wackerly, W. Mendenhall y R. L. Scheaffer (2009), Mathematical Statistics with Applications, *Prentice-Hall*, New York.