TEOREMA DEL LIMITE CENTRAL:
UNA DEMOSTRACION ALTERNATIVA

GABRIELA GRIJALVA MONTEVERDE

T E S I S

PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE
MAESTRO EN CIENCIAS
EN ESTADISTICA EXPERIMENTAL

Universidad Autonoma Agraria
Antonio Narro
PROGRAMA DE GRADUADOS
Buenavista, Saltillo, Coah.
JUNIO DE 1991

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

MAESTRO EN CIENCIAS EN
ESTADISTICA EXPERIMENTAL

COMITE PARTICULAR

Asesor principal : ______________________
Dr. Rolando Cavazos Cadena

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

Asesor : ______________________
M.C. José Antonio Díaz García

____
Dr. José Manuel Fernández Brondo
Subdirector de Asuntos de Postgrado

Buenavista, Saltillo, Coah. Junio de 1991.

AGRADECIMIENTOS

Agradezco sinceramente al Dr. Rolando Cavazos Cadena por el apoyo que me brindó a lo largo de mis estudios y especialmente durante el desarrollo de este trabajo.

A mis hijos,

Dina y Miguel,

a pesar de los cuales me fue posible realizar este trabajo.

A mi compañero,

Miguel Angel Valdés Covarrubias.

COMPENDIO

Teorema del límite central: una demostración alternativa

por

GABRIELA GRIJALVA MONTEVERDE

MAESTRIA

ESTADISTICA EXPERIMENTAL

UNIVERSIDAD AUTONOMA AGRARIA ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA. JUNIO DE 1991

Dr. Rolando Cavazos Cadena - Asesor -

Palabras clave : vectores multinomiales, sucesión de variables aleatorias, sucesión de vectores aleatorios, distribución asintótica.

Este trabajo trata sobre el Teorema del Límite Central, resultado de indudable importancia en todas las áreas de la teoría y práctica estadísticas. El objetivo principal que se persigue es obtener una demostración elemental de la versión clásica de este teorema, tanto para el caso unidimensional como para el equivalente multivariado.

ABSTRACT

Central limit theorem: an alternative proof.

BY

GABRIELA GRIJALVA-MONTEVERDE

MASTER OF SCIENCE

EXPERIMENTAL STATISTICS

UNIVERSIDAD AUTONOMA AGRARIA ANTONIO NARRO

BUENAVISTA, SALTILLO, COAHUILA. MEXICO. JUNE 1991

Dr. Rolando Cavazos-Cadena - Advisor -

Key words : multinomial vectors, sequence of random variables, sequence of random vectors, asymptotic distribution.

We are concerned in this work with the Central Limit Theorem, a fundamental result in most theoretical and practical fields of Statistics. The main objective is to propose an elementary proof of the classical version of this theorem in both situations, the unidimensional and the multivariate cases.

INDICE DE CONTENIDO

INTRODUCCION

El Teorema del Límite Central (TLC) constituye, sin duda alguna, la herramienta más utilizada en la teoría y práctica de la Estadística. El presente trabajo trata sobre el TLC, proponiéndose como objetivo principal la derivación de una demostración *elemental* para este resultado.

Antes de continuar, debe mencionarse que bajo el nombre de teorema del límite central se incluye a una gran variedad de resultados, razón por la cual es necesario describir la versión a la que nos referiremos en adelante. Con este fin, suponga que $X_1, X_2, \ldots$ es una sucesión de variables aleatorias con media y varianza finitas y, para cada $n = 1,2,\ldots$, defina $\overline{X}_n$ y $\overline{\mu}_n$ como sigue :

$$\overline{X}_n := \sum_{i=1}^{n} X_i/n \ ;$$

$$\overline{\mu}_n := E(\overline{X}_n).$$

Entonces, un TLC es un resultado que establece condiciones sobre la sucesión $X_1, X_2, \ldots$, bajo las cuales la distribución asintótica de $(\overline{X}_n - \overline{\mu}_n)n^{1/2}$ es normal con media cero y cierta varianza finita y positiva σ^2. En términos más precisos, el requerimento sobre la distribución asintótica de $(\overline{X}_n - \overline{\mu}_n)n^{1/2}$ se refiere a si, para todos los números reales a y b, se tiene la siguiente convergencia

$$\lim_{n\to\infty} P[a < (\overline{X}_n - \overline{\mu}_n)n^{1/2} < b] = \Phi(b/\sigma) - \Phi(a/\sigma), \qquad (1)$$

donde, como es usual, $\Phi(\cdot)$ es la función de distribución normal estándar. Detalles adicionales acerca de ésta y otras formas del TLC se encuentran, por ejemplo, en Loéve (1977).

La forma más conocida del TLC establece que (1) ocurre si las variables aleatorias $X_1, X_2, \ldots$ son *independientes e idénticamente distribuidas*. El interés de este trabajo se centra precisamente en esta formulación, a la que nos referiremos como la *versión clásica* del TLC. Además del caso unidimensional recién descrito, también se considerará la extensión al caso multidimensional, en el cual las variables aleatorias X_i se reemplazan por vectores aleatorios $\mathbf{X}_i$.

A partir de lo anterior, es conveniente notar que la importancia de la versión clásica del TLC radica en que permite obtener la normalidad asintótica de muchas sucesiones interesantes de variables aleatorias. Por ejemplo, suponga que $X_1, X_2, \ldots$ es una muestra de una población con densidad $f(x)$ y que la mediana poblacional, digamos m, es tal que $f(m) > 0$. Además, sea W_n la mediana muestral basada en n observaciones. En este caso, aunque W_n no se expresa directamente como un múltiplo de una suma de variables aleatorias, se tiene que la distribución asintótica de $(W_n-m)/n^{1/2}$ es normal, hecho que puede establecerse a partir de la versión clásica del TLC. Aplicaciones del TLC en procesos de decisión Markovianos pueden encontrarse, por ejemplo, en Mandl (1974), así como en Ross (1980). Por otra parte, en Araujo y Giné (1980) pueden encontrarse algunas extensiones del TLC a el caso de vectores aleatorios con valores en espacios abstractos.

Indudablemente, el tema que nos ocupa representa un clásico en la literatura, así que es natural preguntarse por la motivación que dio origen al desarrollo de esta presentación. Para abordar esta cuestión, es útil comentar primero, brevemente, la forma en que el TLC ha evolucionado. De acuerdo a Feller (1986) y Maistrov (1974), la primera forma del TLC fue obtenida por DeMoivre en 1730, quien consideró variables aleatorias independientes $X_1, X_2, \ldots$ para las cuales $P[X_i = 1] = 1/2 = P[X_i = 0]$ y mostró que, en estas circunstancias, (1) ocurre con $\sigma = 1/2$. Posteriormente, en 1810, Laplace generalizó el resultado al caso en que las variables aleatorias X_i son independientes y tienen una distribución (arbitraria) del tipo de Bernoulli, es decir, con $P[X_i = 1] = p = P[X_i = 0]$, donde $p \in (0,1)$; en este caso, (1) se satisface con $\sigma = [p(1-p)]^{1/2}$. Las extensiones del

TLC a distribuciones menos restringidas tuvieron que esperar el advenimiento de herramientas matemáticas sofisticadas, como la teoría de las transformadas de Fourier y el correspondiente teorema de inversión, la teoría de la medida, la noción de convergencia débil de medidas y el teorema de continuidad de Lèvy. La forma en que estos conceptos y resultados se amalgaman para obtener versiones del TLC más generales que la versión clásica descrita anteriormente, puede encontrarse, por ejemplo, en Ash (1982), o en Dudley (1989). En este trabajo se pretende obtener una demostración de la versión clásica del TLC utilizando técnicas y conceptos probabilísticas elementales, de modo que ésta sea accesible a un público más amplio y de esta manera contribuir a un mejor entendimiento del TLC. Este objetivo fue la principal *motivación* para emprender esta investigación.

El material se ha organizado de la siguiente manera :

En el Capítulo I se introduce la distribución multinomial y se establecen los resultados preliminares para el desarrollo subsecuente del trabajo; entre los resultados más importantes, se obtiene una estimación del término máximo de la función de probabilidad multinomial para valores grandes de n. El Capítulo II se aboca a la extensión del teorema de DeMoivre Laplace al caso de vectores multinomiales, resultado sobre el cual se apoya el resto del trabajo, e incluye también la generalización respectiva a vectores aleatorios (arbitrarios) con soporte finito. La demostración del teorema del límite central para el caso unidimensional se proporciona en el Capítulo III. Esta demostración se divide en tres casos, a saber : variables aleatorias con soporte finito, variables acotadas y variables no acotadas. Por último, en el Capítulo IV se obtiene la versión clásica del TLC multivariado para vectores aleatorios acotados y no acotados.

I | LA DISTRIBUCION MULTINOMIAL

1 INTRODUCCION

El propósito de este capítulo es introducir la función de probabilidad multinomial, así como establecer algunos resultados acerca de ésta que nos serán de utilidad en el Capítulo II. El material se ha organizado de la siguiente manera : la sección de preliminares presenta una demostración de la fórmula de Stirling y un resumen de las principales propiedades algebraicas del valor esperado de una matriz aleatoria; en la sección 3 se introduce la función de probabilidad multinomial, en tanto que en la sección 4 se obtiene una estimación del término máximo de esta función para valores grandes de n.

2 PRELIMINARES

La fórmula de Stirling. En este capítulo estudiaremos algunas propiedades de la distribución multinomial; entre ellas, se establecerá una estimación del término máximo de esta distribución para valores grandes de n (Teorema 1.19). En la obtención de esta estimación utilizaremos un resultado clásico conocido como la *fórmula de Stirling*[1], el cual proporciona información acerca del comportamiento asintótico de n!. La demostración que aquí se presenta puede encontrarse en Fulks (1967).

Para empezar, necesitamos introducir la idea de igualdad asintótica de dos sucesiones, la cual desempeña un papel relevante en el desarrollo subsecuente.

Definición 1.1. Sean $\{a_n\}$ y $\{b_n\}$ dos sucesiones de números reales no nulos. Se dice que $\{a_n\}$ y $\{b_n\}$ son *asintóticamente iguales*, en cuyo caso escribimos $a_n \sim b_n$, si y sólo si

$$\lim_{n\to\infty} (a_n/b_n) = 1. \tag{1.1}$$

En otras palabras, $a_n \sim b_n$ si la *diferencia relativa* entre a_n y b_n tiende a cero cuando n se incrementa. Note que (1.1) no implica nada acerca de la *diferencia absoluta* $a_n - b_n$. Para ver esto, considere los siguientes casos :

(1) $a_n = n$, $b_n = n + 1/n$;

(2) $a_n = n$, $b_n = n + n^{1/2}$;

(3) $a_n = n$, $b_n = n + (-1)^n$;

(4) $a_n = n$, $b_n = n + (-1)^n n^{1/2}$.

En cada uno de los casos (1)-(4) se tiene que $a_n \sim b_n$ pero el comportamiento de $a_n - b_n$ conforme n tiende a ∞ es muy diferente : bajo (1) se tiene que $a_n - b_n \to 0$, en el caso

---

[1] Según Feller (1986), este resultado fue presentado por James Stirling en 1730.

(2) $a_n - b_n \to -\infty$, en tanto que $a_n - b_n$ oscila en (3) y (4); además, note que la sucesión $\{a_n - b_n\}$ es acotada en el caso (3), pero no en la situación (4). En resumen, esta discusión muestra que la igualdad asintótica de dos sucesiones no implica cosa alguna acerca del comportamiento de la sucesión formada por las diferencias de los términos correspondientes.

Como mencionamos anteriormente, nuestro primer objetivo en este capítulo es estudiar el comportamiento de la sucesión $a_n := n!$ conforme n tiende a ∞. Es bien sabido que n! se incrementa "rápidamente" a medida que n crece; por ejemplo, 10! = 3'628 800 , 11! = 39'916 800. Por esta razón, para cualquier fin práctico resulta conveniente disponer de una "estimación" de n! aûn para valores moderados de n. Tal estimación se proporciona en el siguiente teorema, en cuya demostración hacemos uso de dos resultados que pueden encontarse en Atkinson (1978) y Spivak (1978) y que presentamos en A1 y A2 del Apéndice al final de este trabajo.

Teorema 1.2. Las sucesiones $\{n!\}$ y $\{(2\pi n)^{1/2} n^n \exp(-n)\}$ son asintóticamente iguales, esto es,

$$n! \sim (2\pi n)^{1/2} n^n \exp(-n). \tag{1.2}$$

Demostración. Primeramente, observe que para todo entero positivo n se tiene que

$$\ln(n!) = \sum_{k=1}^{n} \ln(k) \quad .$$

Por otro lado, la función f(x) = ln(x) es dos veces diferenciable en $[1,\infty)$ y su segunda dereivada está dada por

$$f''(x) = - 1/x^2.$$

Luego, la regla del trapecio (ver A1 en Apéndice), implica

$$\int_k^{k+1} \ln(x)\, dx$$

$$= [\ln(k) + \ln(k+1)]/2 + 1/(12\xi_k^2) \quad , \ k = 1,\ldots,n-1,$$

donde ξ_k es un punto en el intervalo [k,k+1]. Sumando sobre k se obtiene

$$\int_1^n \ln(x)\, dx = \sum_{k=1}^{n} \ln(k) - (1/2)\ln(n) + (1/12)\sum_{k=1}^{n-1} \xi_k^{-2} \quad . \tag{1}$$

Por otra parte, un sencillo argumento de integración por partes muestra que

$$\int_1^n \ln(x)\, dx = n \ln(n) - n + 1. \tag{2}$$

A partir de (1) y (2) se desprende inmediatamente que

$$\ln n! = (n + 1/2) \ln(n) - n + \gamma_n , \tag{3}$$

donde

$$\gamma_n := 1 - (1/12)\sum_{k=1}^{n-1} \xi_k^{-2} \quad .$$

Luego, tomando la función exponencial en ambos miembros de (3) se obtiene que

$$n! = n^{n+1/2}\exp(-n)c_n , \quad c_n := \exp(\gamma_n) . \tag{4}$$

Dado que la serie $\sum_{k=1}^{\infty} k^{-2}$ es convergente y $0 < \xi_k^{-2} \le k^{-2}$, vemos que existe el límite

$$\gamma := \lim_{n\to\infty} \gamma_n = 1 - (1/12)\sum_{k=1}^{\infty} \xi_k^{-2}.$$

Defina entonces

$$c := \lim_{n\to\infty} c_n = \lim_{n\to\infty} \exp(\gamma_n),$$

y observe que

$$c = \lim_{n\to\infty} [n!/(n^{n+1/2}\exp(-n))]. \tag{5}$$

Así, para concluir la demostración basta establecer que $c = (2\pi)^{1/2}$. Para ver esto, primero observe que $\lim_{n\to\infty} c_{2n} = c$ y $\lim_{n\to\infty} c_n^2 = c^2$. Luego

$$c = \lim_{n\to\infty} (c_n^2/c_{2n}), \tag{6}$$

y usando (4) se sigue que

$$c_n^2/c_{2n} = n!^2(2n)^{2n+1/2}/(n^{2n+1}(2n)!),$$

esto es

$$c_n^2/c_{2n} = 2^{1/2}2^{2n}\ n!^2/(n^{1/2}(2n)!). \tag{7}$$

Para evaluar explícitamente el límite del lado derecho de (7), recordemos la representación de π conocida como la *fórmula del producto de Wallis* (ver A2 en el Apéndice) :

$$\pi = 2\prod_{k=1}^{\infty}(4k^2)/(4k^2-1)$$

$$= 2 \lim_{n\to\infty} \prod_{k=1}^{n}(4k^2)/(4k^2-1)$$

$$= 2 \lim_{n\to\infty} \frac{2\cdot 2\cdot 4\cdot 4\cdot \ldots \cdot 2n\cdot 2n}{1\cdot 3\cdot 3\cdot 5\cdot \ldots \cdot (2n-1)(2n+1)}. \tag{8}$$

Observe ahora que para los productos parciales se tiene

$$[2\prod_{k=1}^{n}(4k^2)/(4k^2-1)]^{1/2} = (2)^{1/2}\frac{2\cdot 4\cdot\ldots\cdot 2n}{3\cdot 5\cdot\ldots\cdot(2n-1)(2n+1)^{1/2}}.$$

Multiplicando numerador y denominador por $2\cdot 4\cdot\ldots\cdot 2n$, se desprende que

$$[2\prod_{k=1}^{n}(4k^2)/(4k^2-1)]^{1/2} = \frac{1}{(n+1/2)^{1/2}}\frac{2^2\cdot 4^2\cdot\ldots\cdot(2n)^2}{2\cdot 3\cdot 4\cdot 5\cdot\ldots\cdot(2n-1)\cdot 2n},$$

de donde se obtiene la expresión

$$[2\prod_{k=1}^{n}(4k^2)/(4k^2-1)]^{1/2} = 2^{2n}(n!)^2/[(n+1/2)^{1/2}(2n)!]. \qquad (9)$$

Combinando (6)-(9) se sigue que

$$(\pi)^{1/2} = \lim_{n\to\infty}\ [2\prod_{k=1}^{n}(4k^2)/(4k^2-1)]^{1/2}$$

$$= \lim_{n\to\infty}\ (2^{2n}(n!)^2/[(n+1/2)^{1/2}(2n)!])$$

$$= \lim_{n\to\infty}\ (2^{2n}(n!)^2/[(n)^{1/2}(2n)!])$$

$$= c/(2)^{1/2}.$$

Finalmente, esta última convergencia junto con (5) implican que

$$\lim_{n\to\infty}\ [c_n/(2\pi)^{1/2}] = \lim_{n\to\infty}\ n!/(2\pi n)^{1/2}n^n\exp(-n) = 1,$$

esto es,

$$\lim_{n\to\infty} c_n = (2\pi)^{1/2},$$

y como se mecionó anteriormente, esto concluye la demostración. ∎

La demostración del teorema 1.2 permite establecer inmediatamente una cota del error de la estimación en (1.2). En efecto, usando la notación empleada en la demostración del teorema anterior se tiene que, para todo n mayor o igual a 2, se satisface la desigualdad

$$0 < \gamma_n - \gamma = (1/12)\sum_{k=n}^{\infty} \xi_k^{-2}$$

$$\leq (1/12)\sum_{k=n}^{\infty} k^{-2}$$

$$< (1/12)\int_{n-1}^{\infty} x^{-2}dx = 1/[12(n-1)],$$

y entonces,

$$1 < \exp(\gamma_n - \gamma) < \exp(1/[12(n-1)]).$$

Usando (4) se obtiene

$$\exp(\gamma_n - \gamma) = n!/[cn^{n+1/2}\exp(-n)]$$

$$= n!/[(2\pi n)^{1/2}n^n\exp(-n)],$$

de donde se desprende que

$$(2\pi n)^{1/2}n^n\exp(-n) < n!$$

$$< (2\pi n)^{1/2}n^n\exp(-n)\exp(1/[12(n-1)]).$$

Luego, la aproximación $(2\pi n)^{1/2}n^n\exp(-n)$ *subestima* a n! con un error relativo menor a $\exp(1/[12(n-1)]) - 1$.

Esperanza de una matriz aleatoria. En este apartado se presentan las propiedades principales de la esperanza y matriz de covarianza de un vector aleatorio que utilizaremos en el presente trabajo.

Definición 1.3. Sea $W = [w_{ij}]$, $i = 1,\ldots,k_1$, $j = 1,\ldots,k_2$ una matriz de orden $k_1\times k_2$, donde w_{ij} son variables aleatorias definidas en un mismo espacio muestral Ω. En este caso, W es una *matriz aleatoria*. Si $E(w_{ij})$ existe para toda (i,j), se define la *esperanza de la matriz* W, denotada por $E(W)$, como la matriz $A = [a_{ij}]$ de orden $k_1\times k_2$ tal que

$$a_{ij} = E(w_{ij}).$$

Como caso particular de la definición anterior tenemos que si $\mathbf{X} = (X_1,\ldots,X_k)'$ es un vector aleatorio, entonces

$$E(\mathbf{X}) := (E(X_1),\ldots,E(X_k))', \tag{1.3}$$

siempre y cuando $E(X_i)$ esté definida para toda $i = 1,\ldots,k$.

Las propiedades del valor esperado de una variable aleatoria y un cálculo algebraico directo conducen en forma inmediata al siguiente resultado (Graybill, 1976).

Teorema 1.4. Sean V y W dos matrices aleatorias cuya esperanza esté definida y sean A_1, A_2 matrices de constantes. Suponga que las dimensiones de V, W, A_1, A_2 son tales que las operaciones matriciales que aparecen a continuación tienen sentido. Entonces

i) $E(A_1) = A_1$,

ii) $E(A_1VA_2) = A_1E(V)A_2$ (1.4)

iii) $E(V + W) = E(V) + E(W)$.

Si X y Y son dos variables aleatorias definidas en un mismo espacio muestral, recuérdese que la *covarianza* entre X y Y se define por

$$\mathrm{cov}(X,Y) := E[(X - E(X))(Y - E(Y))]; \tag{1.5}$$

en particular, la varianza de una variable aleatoria X – a saber, Var(X)– está dada por

$$\mathrm{Var}(X) := \mathrm{cov}(X,X) = E([X - E(X)]^2). \tag{1.6}$$

Estas nociones desempeñan un importante papel en el análisis de variables aleatorias. Para vectores aleatorios se tienen conceptos análogos, los cuales se introducen a continuación.

Definición 1.5. Si $\mathbf{X} = (X_1,\ldots,X_k)'$ es un vector aleatorio, la *matriz de covarianza* de **X** se denota COV(**X**) y se define por

$$\mathrm{COV}(\mathbf{X}) := E[(\mathbf{X} - E(\mathbf{X}))(\mathbf{X} - E(\mathbf{X}))'], \tag{1.7}$$

siempre y cuando las esperanzas del lado derecho existan. Más generalmente, si **X** y **Y** son vectores aleatorios definidos en un mismo espacio, la *matriz de covarianza entre* **X** *y* **Y** se denota por COV(**X**,**Y**) y está dada por

$$\mathrm{COV}(\mathbf{X},\mathbf{Y}) := E[(\mathbf{X} - E(\mathbf{X}))(\mathbf{Y} - E(\mathbf{Y}))'],$$

siempre y cuando las esperanzas del lado derecho existan.

Observación 1.6. Note que el elemento ij de la matriz COV(**X**) está dado por

$$E[(X_i - E(X_i))(X_j - E(X_j))] = \mathrm{cov}(X_i,X_j);$$

vea (1.5). Esta igualdad da origen a la denominación de la matriz COV(**X**). Además, debido a que

$$\mathrm{cov}(X_i,X_j) = \mathrm{cov}(X_j,X_i) \ , \ i,j = 1,\ldots,k \ ,$$

se tiene que COV(**X**) en (1.7) es una matriz simétrica, esto es,

$$\mathrm{COV}(\mathbf{X}) = (\mathrm{COV}(\mathbf{X}))'. \tag{1.8}$$

Con frecuencia resulta útil representar a un conjunto de variables aleatorias $Y_1,\ldots,Y_p$ como combinaciones lineales (no homogéneas) de otro conjunto de variables aleatorias $X_1,\ldots,X_k$. El siguiente resultado, consecuencia inmediata del Teorema 1.4, relaciona la esperanza y matriz de covarianza de los vectores aleatorios **Y** y **X**.

Teorema 1.7. Sea **X** un vector aleatorio de dimensión k, A una matriz constante de orden $p \times k$, y $\mathbf{a} \in \mathbb{E}_p$ un vector arbitrario. Entonces, para el vector aleatorio $\mathbf{Y} = A\mathbf{X} + \mathbf{a}$ se tiene que

i) $E(\mathbf{Y}) = A\,E(\mathbf{X}) + \mathbf{a}$,

ii) $\mathrm{COV}(\mathbf{Y}) = A\,\mathrm{COV}(\mathbf{X})\,A'$. (1.9)

3 LA DISTRIBUCION MULTINOMIAL

Uno de los modelos probabilísticos más sencillos, pero que ha tenido una gran importancia teórica y práctica en el desarrollo de la Estadística, es el conocido como *ensayos o experimentos de Bernoulli*; este término se utiliza para referirse a una serie de repeticiones *independientes* de un experimento aleatorio en el que sólo se distinguen dos resultados mutuamente excluyentes y que conservan probabilidades de ocurrencia constantes en cada repetición. Usualmente los dos resultados posibles se denominan éxito y falla.

Con cada experimento de Bernoulli podemos asociar una variable aleatoria, digamos X, definida de la siguiente manera :

$$X := 1 \text{ si el resultado es éxito;}$$
$$X := 0 \text{ si el resultado es falla.}$$

De este modo, X proporciona una codificación *numérica* del resultado *(cualitativo)* de un experimento de Bernoulli. La función de probabilidad de X es

$$\begin{aligned} f_X(1) &= P[X = 1] = p \equiv \text{probabilidad de éxito;} \\ f_X(0) &= P[X = 0] \\ &= 1 - p := q \equiv \text{probabilidad de falla;} \qquad (1.10) \\ f_X(x) &= 0 \text{ si } x \neq 0,1. \end{aligned}$$

En este caso, la variable X se denomina *variable de Bernoulli con parámetro* p = P[X = 1]; expresamos este hecho escribiendo $X \simeq Ber(p)$. Usando (1.10) se tiene que

si $X \simeq Ber(p)$, entonces

$$E(X) = p \ , \ y \quad \mathrm{Var}(p) = p(1 - p) \equiv pq. \tag{1.11}$$

En general, con cualquier experimento aleatorio pueden asociarse variables de Bernoulli como se describe a continuación. Suponga que Ω es el espacio muestral para un experimento aleatorio y sea $A \subset \Omega$ un evento arbitrario. La *función indicadora* de A, denotada por $I_A\colon \Omega \to \mathbb{R}$, se define por

$$\begin{aligned} I_A(w) &:= 1 \quad \text{si } \omega \in A, \\ &:= 0 \quad \text{si } \omega \notin A. \end{aligned} \tag{1.12}$$

En palabras, $I_A = 1$ si y sólo si el resultado del experimento pertenece a A, mientras que $I_A = 0$ de otra forma. En este caso, I_A tiene distribución de Bernoulli con parámetro $p = P[I_A = 1] = P[A]$.

Para un evento A contenido en el espacio muestral Ω defina

$$Y_1 := I_A \quad , \ Y_2 := I_{A^c} \ .$$

En este caso, $Y_1 \simeq \mathrm{Ber}(p_1)$ y $Y_2 \simeq \mathrm{Ber}(p_2)$, donde $p_1 = P[A] = p$ y $p_2 = P[A^c]$, $p_1 + p_2 = 1$. Note además que $Y_1 + Y_2 = 1$, de modo que Y_1 y Y_2 son *dependientes*.

La distribución del vector $\mathbf{Y} = (Y_1, Y_2)'$ está determinada por

$$\begin{aligned} f_{\mathbf{Y}}(1,0) &= P[Y_1 = 1, Y_2 = 0] = P[A] = p_1; \\ f_{\mathbf{Y}}(0,1) &= P[Y_1 = 0,\ Y_2 = 1] = P[A^c] = p_2, \end{aligned} \tag{1.13}$$

y, desde luego,

$$f_{\mathbf{Y}}(\mathbf{Y}) = f_{\mathbf{Y}}(y_1,y_2)$$
$$= P[Y_1 = y_1,\ Y_2 = y_2] = 0,\ (y_1,y_2) \neq (1,0),(1).$$

Este es un ejemplo de un vector aleatorio con distribución multinomial, la cual se introduce formalmente más adelante. Note que $Y_1 = 1$ (0) implica $Y_2 = 0$ (1), y entonces $Y_1Y_2 = 0$. Usando (1.13) se obtiene

$$E(\mathbf{Y}) = (p_1,p_2)';$$

$$\mathrm{Var}(Y_i) = p_i(1 - p_i),\ i = 1,2;$$

$$\mathrm{cov}(Y_1,Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2) = -p_1p_2.$$

Las últimas dos ecuaciones pueden resumirse como sigue :

$$\mathrm{COV}(\mathbf{Y}) = \begin{bmatrix} p_1(1-p_1) & -p_1p_2 \\ -p_1p_2 & p_2(1-p_2) \end{bmatrix}$$
$$= \mathrm{Diag}(p_1,p_2) - (p_1,p_2)'(p_1,p_2). \qquad (1.14)$$

Esta discusión se extiende inmediatamente al caso de r resultados mutuamente excluyentes. Con este fin, sea Ω el espacio muestral de un experimento aleatorio y suponga que los eventos $A_1,\ldots,A_r \subset \Omega$ satisfacen las siguientes condiciones *(i)* y *(ii)*:

(i) $A_i \cap A_j = \emptyset$, $i < j$, y

(ii) $\bigcup_{i=1}^{r} A_i = \Omega$.

Defina $Y_i := I_{A_i}$, $i = 1,2,\ldots,r$. En este caso $Y_i \simeq \mathrm{Ber}(p_i)$, donde $p_i = P[A_i]$, $i = 1,2,\ldots,r$. Además, observe que $Y_i = 1$ si y sólo si $Y_j = 0$ para todo $j \neq i$, lo cual es equivalente a la siguiente afirmación : el vector aleatorio $\mathbf{Y} = (Y_1,\ldots,Y_r)'$ toma valores en el conjunto $\{\mathbf{e}_1,\ldots,\mathbf{e}_r\}$, donde $\mathbf{e}_i$ es el i-ésimo elemento de la base canónica en $\mathbb{E}_r$. Un vector aleatorio con estas características tiene una distribución que es un caso especial de la distribución multinomial.

Definición 1.8. Sea $\mathbf{Y} = (Y_1,\ldots,Y_r)'$ un vector r-dimensional, con $r \geq 2$. Suponga que

$$P[\mathbf{Y} = \mathbf{e}_i] = p_i > 0 \;,\; i = 1,\ldots,r, \tag{1.15}$$

donde $\{\mathbf{e}_i\}_{i=1}^{r}$ es la base canónica de $\mathbb{E}_r$, y $\sum_{i=1}^{r} p_i = 1$. En este caso, se dice que $\mathbf{Y}$ tiene *distribución multinomial r-dimensional con parámetros* 1 y $p_1,\ldots,p_r$, y escribimos

$$\mathbf{Y} \simeq \mathcal{M}_r(1;p_1,\ldots,p_r). \tag{1.15*}$$

Notación. Acerca de la notación establecida en (1.15*) se puede comentar lo siguiente : el subíndice r indica la dimensión del vector $\mathbf{Y}$, el parámetro 1 indica que se ha efectuado *una* sola repetición del experimento, mientras que $p_1,\ldots,p_r$ son las probabilidades de obtener los diversos resultados del experimento.

Como caso particular, podemos observar que

$\mathcal{M}_2(1;p_1,p_2)$ define la distribución (1.13); en este sentido, podemos decir que la distribución (1.15) extiende la noción de ensayo de Bernoulli a experimentos en los que los resultados se dividen no en dos, sino en r categorías mutuamente excluyentes. El siguiente lema proporciona la generalización de la relación (1.14).

Lema 1.9. Suponga que $\mathbf{Y} \simeq \mathcal{M}_r(1;p_1,\ldots,p_r)$. Entonces,

$$E(\mathbf{Y}) = (p_1,\ldots,p_r)' := \mathbf{p}, \tag{1.16}$$

y

$$\mathrm{COV}(\mathbf{Y}) = \mathrm{Diag}(p_1,\ldots,p_r) - \mathbf{p}\cdot\mathbf{p}'. \tag{1.17}$$

Demostración. Basta observar que cada componente del vector $\mathbf{Y}$ tiene distribución Bernoulli con parámetro p_i, por lo que (1.11) implica

$$E(Y_i) = p_i \ , \ \mathrm{COV}(Y_i,Y_i) = p_i(1-p_i) \ , \ i = 1,\ldots r.$$

Además, para cualesquier $i,j = 1,\ldots,r$, $i\neq j$, se tiene

$$\mathrm{COV}(Y_i,Y_j) = E(Y_iY_j) - E(Y_i)E(Y_j) = -\ p_ip_j$$

pues $P[Y_iY_j = 0] = 1$. ∎

Observación 1.10. Dada la relación existente entre $Y_1,\ldots,Y_r$ es fácil verificar que la matriz de covarianza (1.17) es singular. Para ver esto observe que, de acuerdo a la definición 1.8, $P[Y_1+\ldots+ Y_r = 1] = 1$, y por lo tanto $\mathrm{Var}(Y_1+\ldots+Y_r) = 0$. Esta última igualdad es equivalente a

$\mathrm{Var}(\mathbf{1}'\mathbf{Y}) = 0$, donde $\mathbf{1} \in \mathbb{E}_r$ está dado por $\mathbf{1} := (1,\ldots,1)'$, y entonces, utilizando (1.9), se obtiene que

$$0 = \mathrm{Var}(\mathbf{1}'\mathbf{Y}) = \mathbf{1}'\mathrm{COV}(\mathbf{Y})\mathbf{1},$$

de donde se desprende que $\mathbf{1}'\mathrm{COV}(\mathbf{Y}) = 0$ y de aquí que $\mathrm{COV}(\mathbf{Y})$ no sea inversible.

Tomando en cuenta la observación anterior y debido a que es más cómodo trabajar con vectores aleatorios cuya matriz de covarianza sea no singular, preferiremos en adelante trabajar con la proyección de $\mathbf{Y}$ sobre $\mathbb{E}_{r-1}$. Además, ésta es una elección lógica si consideramos que al conocer el resultado de $Y_1,\ldots,Y_{r-1}$ automáticamente se conoce $Y_r = 1 - \sum_{i=1}^{r-1} Y_i$. En el resto del trabajo emplearemos la siguiente convención notacional.

Notación. Si $\mathbf{Y} = (Y_1,\ldots,Y_r)' \simeq \mathcal{M}_r(1;p_1,\ldots,p_r)$, $\tilde{\mathbf{Y}}$ denotará al vector aleatorio de dimensión $r - 1$ que consiste de las primeras $r - 1$ componentes de $\mathbf{Y}$, esto es

$$\tilde{\mathbf{Y}} := (Y_1,\ldots,Y_{r-1})'. \tag{1.18*}$$

Similarmente, si $\mathbf{Y}_1,\mathbf{Y}_2,\ldots$ es una sucesión de vectores aleatorios independientes e idénticamente distribuidos, $\mathbf{Y}_k \simeq \mathcal{M}_r(1;p_1,\ldots,p_r)$, para todo entero positivo n defina el vector aleatorio

$$\tilde{\mathbf{S}}_n := \sum_{k=1}^{n} \tilde{\mathbf{Y}}_k. \tag{1.18}$$

Con esta notación (1.16) y (1.17) pueden expresarse en términos de $\tilde{Y}$ como se indica en el siguiente corolario.

Corolario 1.11. Si $Y \simeq \mathcal{M}_r(1;p_1,\ldots,p_r)$ entonces

$$E(\tilde{Y}) = (p_1,\ldots,p_{r-1})' := \tilde{\mathbf{p}} \,, \tag{1.19}$$

$$\mathrm{COV}(\tilde{Y}) = \mathrm{Diag}(p_1,\ldots,p_{r-1}) - \tilde{\mathbf{p}}\cdot\tilde{\mathbf{p}}' := \tilde{V}. \tag{1.20}$$

A continuación veremos que la matriz $\tilde{V}$ definida en (1.20) es inversible y encontraremos explícitamente la matriz $\tilde{V}^{-1}$ así como det(V). Empezamos la discusión con el siguiente resultado preliminar.

Lema 1.12. Sea $\mathbf{z} \in \mathbb{E}_k - \{0\}$ un vector arbitrario y defina la matriz W de orden $k \times k$ mediante

$$W := I + a\mathbf{z}\mathbf{z}',$$

donde I es la matriz identidad de orden k y $a \in \mathbb{R}$ es una constante. Entonces,

(*i*) $\det(W) = 1 + a\mathbf{z}'\mathbf{z} = 1 + a\|\mathbf{z}\|^2$.

(*ii*) Si $1 + a\|\mathbf{z}\|^2 \neq 0$, se tiene que

$$W^{-1} = I - a(1 + a\|\mathbf{z}\|^2)^{-1}\mathbf{z}\mathbf{z}'.$$

Demostración. (*i*) Sean $\mathbf{v}_1,\ldots,\mathbf{v}_{k-1}$ vectores en $\mathbb{E}_k$ tales que $\mathcal{B} := \{\mathbf{v}_1,\ldots,\mathbf{v}_{k-1},\mathbf{z}\}$ es una base *ortogonal* de $\mathbb{E}_k$; note que en este caso se tiene que $\mathbf{z}'\mathbf{v}_i = 0$, $i = 1,2,\ldots,k-1$.

Entonces,

$$Wv_i = Iv_i + azz'v_i = v_i, \quad i = 1,\ldots,k - 1,$$

y también,

$$Wz = Iz + azz'z = (1 + az'z)z.$$

Estas igualdades establecen que cada vector en la base $\mathcal{B}$ es un vector propio de W; el valor propio correspondiente a v_i es 1, $i = 1,\ldots,k-1$, mientras que a **z** le corresponde el valor propio $1 + az'z$. Luego (Hoffman y Kunze, 1973),

$$\det(W) = 1\cdot\ldots\cdot1\cdot(1 + az'z) = 1 + a\|z\|^2.$$

(*ii*) Suponga que $1 + a\|z\|^2 \neq 0$ y defina

$$b := -a/(1 + a\|z\|^2).$$

En este caso,

$$\begin{aligned} abz'z &= ab\|z\|^2 \\ &= -a^2\|z\|^2/(1 + a\|z\|^2) \\ &= -(a^2\|z\|^2 + a)/(1 + a\|z\|^2) + a/(1 + a\|z\|^2) \\ &= - a - b, \end{aligned}$$

esto es,

$$ab\|z\|^2 + a + b = 0. \tag{1}$$

Luego,

$$\begin{aligned} [I - a(1 + a\|z\|^2)^{-1}zz']W &= (I + bzz')(I + azz') \\ &= I + azz' + bzz' + (abz'z)zz' \\ &= I + (a + b + ab\|z\|^2)zz' \\ &= I, \end{aligned}$$

donde la última igualdad se desprende de (1). ∎

Teorema 1.13. Sean $\tilde{p}$ y $\tilde{V}$ como en (1.19) y (1.20), respectivamente, y suponga que

(*i*) $p_i > 0$ para $i = 1,2,\cdots,r-1$, y

(*ii*) $p_r := 1 - \sum_{i=1}^{r-1} p_i > 0$.

Entonces,

(*a*) $\tilde{V}$ es no singular y se tiene

$$\det(\tilde{V}) = p_1 \cdot p_2 \cdot \ldots \cdot p_r. \tag{1.21}$$

Mas aún,

(b) $$\tilde{V}^{-1} = p_r^{-1}[\mathrm{Diag}(p_r/p_1,\ldots,p_r/p_{r-1}) + J_{r-1}], \tag{1.22}$$

donde J_{r-1} es la matriz cuadrada de orden r-1 con todas sus componentes iguales a 1.

Demostración. Defina los vectores z_1 y z_2 mediante

$$z_1 := (p_1^{1/2},\ldots,p_{r-1}^{1/2})',$$

$$z_2 := (p_1^{-1/2},\ldots,p_{r-1}^{-1/2})'.$$

Usando la definición de $\tilde{V}$ no es difícil verificar que

$$W := \mathrm{Diag}(z_2')\ \tilde{V}\ \mathrm{Diag}(z_2')$$

$$= I - z_1 z_1'.$$

Observe ahora que $\|z_1\|^2 = p_1+\ldots+p_{r-1}$. Entonces, usando el lema 1.12 (i) con a = -1, vemos que

$$\det(W) = 1 - \|z_1\|^2 = p_r > 0,$$

de donde se desprende que W es no singular, lo cual a su vez conduce a la no singularidad de $\tilde{V}$. Además, la parte (ii) del lema 1.12 establece

$$W^{-1} = I + p_r^{-1} z_1 z_1'.$$

Por otra parte, note que $[\mathrm{Diag}(z_2')]^{-1} = \mathrm{Diag}(z_1')$ y así

$$\tilde{V} = \mathrm{Diag}(z_1')\ W\ \mathrm{Diag}(z_1'),$$

de donde se concluye que

i) $\det(\tilde{V}) = [\det(\mathrm{Diag}(z_1'))]^2 \cdot \det(W) = p_1 \cdot p_2 \cdot \ldots \cdot p_r$.

ii)
$$\begin{aligned}
\tilde{V}^{-1} &= \mathrm{Diag}(z_2')\ W^{-1}\mathrm{Diag}(z_2') \\
&= \mathrm{Diag}(z_2')\ (I + p_r^{-1} z_1 z_1')\ \mathrm{Diag}(z_2') \\
&= \mathrm{Diag}(z_2')\ \mathrm{Diag}(z_2') + p_r^{-1}\mathrm{Diag}(z_2')\ z_1 z_1'\mathrm{Diag}(z_2') \\
&= \mathrm{Diag}(1/p_1, \ldots, 1/p_{r-1}) + p_r^{-1}\mathbf{1}\cdot\mathbf{1}' \\
&= p_r^{-1}[\mathrm{Diag}(p_r/p_1, \ldots, p_r/p_{r-1}) + J_{r-1}].
\end{aligned}$$ ■

Suponga ahora que un experimento de Bernoulli se repite n veces, y sea $X_i = 1$ si el i-ésimo ensayo resulta éxito, $X_i = 0$ de otro modo. Como mencionamos antes, cada X_i tiene distribución Ber(p), donde p = probabilidad de éxito en cada repetición, y si los ensayos son independientes entonces $T := X_1 + \ldots + X_n$ tiene distribución binomial con parámetros n y p; en este caso, escribimos $T \simeq B(n,p)$. Note que T es el número total de éxitos en las n repeticiones.

En general, suponga que $Y_1, \ldots, Y_n$ es una sucesión de vectores aleatorios independientes con distribución común $\mathcal{M}_r(1;p_1,\ldots,p_r)$. Nuestro interés ahora radica en determinar la distribución de

$$\mathbf{S}_n := \mathbf{Y}_1+\ldots+ \mathbf{Y}_n.$$

Para determinar la función de probabilidad de $\mathbf{S}_n$ observe que (1.15) implica que, para $\mathbf{m} = (m_1,\ldots,m_r) \in \mathbb{E}_r$ el evento $[\mathbf{S}_n = \mathbf{m}]$ tiene probabilidad positiva si y sólo si : (a) cada m_i es un entero no negativo, y (b) $m_1+\ldots+m_r= n$. En estas circunstancias, a partir de la independencia de los vectores $\mathbf{Y}_k$ se desprende que cualquier resultado para el que $\mathbf{S}_n = \mathbf{m}$ ocurre tiene probabilidad

$$p_1^{m_1}\ldots p_r^{m_r} .$$

Por otro lado, el número de maneras en que $[\mathbf{S}_n = \mathbf{m}]$ puede ocurrir es el mismo que el número de formas en que es posible agrupar n objetos en r subgrupos de tamaño $m_1,m_2,\ldots,m_r$, el cual está dado por

$$\binom{n}{m_1\ldots m_r} := n!/(m_1!\ldots m_r!).$$

De esta manera, la distribución de $\mathbf{S}_n$ está determinada por la siguiente expresión :

$$P[\mathbf{S}_n= \mathbf{m}] = \binom{n}{m_1\ldots m_r} p_1^{m_1}\ldots p_r^{m_r} , \qquad (1.23)$$

si $m_1,\ldots,m_r$ son enteros no negativos cuya suma es n, mientras que

$$P[\mathbf{S}_n= \mathbf{m}] = 0 \text{ en otro caso.}$$

Definición 1.14. Un vector aleatorio $\mathbf{X} = (X_1,...,X_r)'$ tiene *distribución multinomial* con parámetros n y $p_1,...,p_r$ si la función de probabilidad de $\mathbf{X}$ está dada por

$$f_{\mathbf{X}}(\mathbf{x}) = \binom{n}{x_1...x_r} p_1^{x_1}...p_r^{x_r} \text{ , si } x_1,...,x_r \text{ son enteros } \geq 0 \text{ cuya suma es } n; \qquad (1.24)$$

$f_{\mathbf{X}}(\mathbf{x}) = 0$ de otra forma.

Como mencionamos anteriormente, la función de probabilidad (1.15) es un caso particular de una distribución multinomial. En efecto, observe que (1.15) se obtiene de (1.24) si $n = 1$.

Notación. Si $\mathbf{X}$ es r-dimensional y se distribuye según (1.24) escribimos

$$\mathbf{X} \simeq \mathcal{M}_r(n;p_1,...,p_r). \qquad (1.25)$$

La discusión anterior puede resumirse como sigue : si los vectores aleatorios $\mathbf{Y}_1,...,\mathbf{Y}_n$ son independientes con distribución común $\mathcal{M}_r(1;p_1,...,p_r)$, entonces el vector aleatorio $\mathbf{S}_n = \mathbf{Y}_1+...+ \mathbf{Y}_n \simeq \mathcal{M}_r(n;p_1,...,p_r)$. Más generalmente se tiene el siguiente resultado.

Lema 1.15. Si $\mathbf{Y}_1,...,\mathbf{Y}_k$ son vectores aleatorios independientes tales que

$$\mathbf{Y}_i \simeq \mathcal{M}_r(n_i;p_1,...,p_r), \; i = 1,...,k,$$

entonces

$$\mathbf{Y}_1+...+ \mathbf{Y}_n \simeq \mathcal{M}_r(n_1+...+ n_k;p_1,...,p_r).$$

Suponga ahora que $Y_1,\dots,Y_n$ son vectores aleatorios independientes con distribución $\mathcal{M}_r(1;p_1,\dots,p_r)$, y $\tilde{Y}_1,\dots,\tilde{Y}_n$ son las proyecciones respectivas sobre $\mathbb{E}_{r-1}$ como en (1.18*). Considere el vector aleatorio $\tilde{S}_n = \tilde{Y}_1+\dots+\tilde{Y}_n$. Para determinar la distribución de $\tilde{S}_n$ observe que el evento $[\tilde{S}_n = \tilde{m}]$, $\tilde{m} = (m_1,\dots,m_{r-1})' \in \mathbb{E}_{r-1}$ tiene probabilidad positiva si y sólo si : (a) cada m_i es un entero no negativo, y (b) $m_1+\dots+m_{r-1} \le n$; en este caso, es claro que

$$P[\tilde{S}_n = \tilde{m}] = P[S_n = m],$$

donde $m = (\tilde{m},m_r)' \in \mathbb{E}_r$, $m_r = n - (m_1+\dots+m_{r-1}) \ge 0$.

De esta manera, la función de probabilidad de $\tilde{S}_n$ puede resumirse como sigue :

$$\begin{aligned} P[\tilde{S}_n = \tilde{m}] &= \binom{n}{m_1\dots m_r} p_1^{m_1}\dots p_r^{m_r}, \quad m_1,\dots,m_r \ge 0 \text{ enteros,} \\ P[\tilde{S}_n = \tilde{m}] &= 0 \qquad \text{de otra forma.} \end{aligned} \tag{1.26}$$

Observación 1.16. Observe que los términos de la función de probabilidad (1.26) coinciden numéricamente con los de la distribución multinomial con parámetros n, $p_1,\dots,p_r$ en (1.24).

Notación. En lo que sigue tendremos que considerar frecuentemente los términos de una distribución multinomial. Para simplificar la notación procederemos como sigue: si $m_1,\cdots,$ m_r son enteros no negativos cuya suma es menor o igual a n, entonces escribiremos

$$a(\tilde{m}) := P[\tilde{S}_n = \tilde{m}]$$

$$= \binom{n}{m_1 \ldots m_r} p_1^{m_1} \ldots p_r^{m_r},$$

donde

(i) $\tilde{m} = (m_1, \ldots, m_{r-1})' \in \mathbb{N}^{r-1}$;

(ii) $m_r = n - (m_1 + \ldots + m_{r-1})$, y

(iii) los valores de r, n, así como los de $p_1, \ldots p_r$ se *sobreentenderán de acuerdo al contexto.*

4 MAXIMO DE UNA DISTRIBUCION MULTINOMIAL

La distribución multinomial en (1.24) es una distribución discreta cuya función de probabilidad se anula fuera de un conjunto finito y, por lo tanto, esta función de probabilidad alcanza su máximo. El objetivo de esta sección es encontrar una estimación de dicho valor máximo para valores grandes de n.

Con este fin, empezaremos determinando una condición necesaria y suficiente para los puntos en que se alcanza el máximo de la función de probabilidad multinomial. El resultado preciso se establece a continuación. (Vea la convención notacional acordada al final de la sección precedente).

Lema 1.17. El término $a(\tilde{\mathbf{k}})$ es el valor máximo de la función de probabilidad multinomial con parámetros $n,p_1,...,p_r$ en (1.24) si y sólo si

$$p_i k_j \le p_j(k_i + 1) \quad \text{para todo } i,j = 1,...,r. \tag{1.27}$$

Demostración. Demostraremos por separado que la condición

$$p_i k_j \le p_j(k_i + 1) \ , \ i,j = 1,...,r, \tag{1}$$

es necesaria y suficiente para que el término $a(\tilde{\mathbf{k}})$ respectivo sea máximo de la función de probabilidad en (1.24).

$\Rightarrow$) Supongamos que $a_0 := a(\tilde{\mathbf{k}})$ es el valor máximo de la función de probabilidad multinomial con parámetros $n,p_1,...,p_r$, esto es

$$\begin{aligned} a_0 &= \binom{n}{k_1 \ldots k_r} p_1^{k_1} \ldots p_r^{k_r} \\ &\ge a(\tilde{\mathbf{m}}) \\ &= \binom{n}{m_1 \ldots m_r} p_1^{m_1} \ldots p_r^{m_r}, \end{aligned} \tag{2}$$

para cualesquier $\tilde{\mathbf{m}} \in \mathbb{E}_{r-1}$, $\tilde{\mathbf{m}} = (m_1,...,m_{r-1})'$; desde luego

$$k_r = n - \sum_{i=1}^{r-1} k_i, \ m_r = n - \sum_{i=1}^{r-1} m_i.$$

Si $k_i = n$ para alguna i, entonces $k_j = 0$ para toda $j \ne i$. Así, a partir de (2) se tiene, para cualesquier $j = 1,..,r$, $j \ne i$,

$$a_0 \ge a(k_1,...,k_j+1,...,k_i-1,...,k_{r-1});$$

esto es,

$$p_i^{\,n} = n!\; p_j p_i^{\,n-1}/(n-1)!,$$

de donde se sigue

$$p_j k_i = np_j \le p_i = p_i(k_j + 1).$$

Para las posibles combinaciones restantes se verifica inmediatamente

$$p_i k_j = 0 < (n + 1)p_j = p_j(k_i + 1),\ j \ne i\ ;$$

$$p_j k_{j'} < p_{j'}(k_j + 1),\quad k_j = k_{j'} = 0.$$

Si no estamos en el caso anterior, podemos distinguir tres combinaciones a investigar: $k_i, k_j \ne 0$; $k_i \ne 0$, $k_j = 0$; y $k_i, k_j = 0$. Claramente esta última combinación satisface (1) y para las dos primeras también puede verificarse a partir de la desigualdad

$$\alpha_0 \ge \alpha(k_1,\ldots,k_j+1,\ldots,k_i-1,\ldots,k_{r-1}).$$

Por lo tanto, si $\alpha(\tilde{\mathbf{k}})$ es el valor máximo se cumple

$$p_i k_j \le p_j(k_i + 1),\ i,j = 1,\ldots,r.$$

$\Leftarrow$) Supongamos ahora que $k_1,\ldots,k_r$ son enteros no negativos tales que $\sum_{i=1}^{r} k_i = n$ y satisfacen la condición (1). Sean $\alpha(\tilde{\mathbf{k}})$ el término correspondiente de la función de probabilidad (1.24) y $\alpha(\tilde{\mathbf{m}})$ cualquier otro valor no nulo de ésta. Entonces, se requiere demostrar que

$$\alpha(\tilde{\mathbf{m}}) \le \alpha(\tilde{\mathbf{k}}).$$

Expresemos cada m_i en la forma

$$m_i = k_i + d_i, \; i = 1,\ldots,r, \; \sum_{i=1}^{r} d_i = 0. \tag{3}$$

De esta manera

$$a(\tilde{m}) = [n!/(k_1+d_1)!\cdots(k_r+d_r)!]\; p_1^{k_1+d_1}\ldots p_r^{k_r+d_r},$$

y por lo tanto

$$a(\tilde{m})/a(\tilde{k}) = \prod_{i=1}^{r}[k_i!/(k_i+d_i)!]p_i^{d_i}. \tag{4}$$

Para comparar el cociente del miembro izquierdo de (4) con la unidad, necesitamos investigar la forma de los factores $[k_i!/(k_i+d_i)!]p_i^{d_i}$, $i = 1,2,\ldots,r$. Aquí podemos distinguir tres situaciones.

Caso 1. $d_i = 0$.

En este caso, observe que

$$[k_i!/(k_i+d_i)!]p_i^{d_i} = 1. \tag{5}$$

Caso 2. $d_i < 0$.

Para esta situación se tiene

$$[k_i!/(k_i+d_i)!]p_i^{d_i} = k_i(k_i-1)\cdots(k_i+d_i+1)p_i^{d_i},$$

y por lo tanto

$$[k_i!/(k_i+d_i)!]p_i^{d_i} \le (k_i/p_i)^{-d_i}. \tag{6}$$

Caso 3. $d_i > 0$.

Note que

$$[k_i!/(k_i+d_i)!]p_i^{d_i} = \left[1/[(k_i+1)\cdots(k_i+d_i)]\right]p_i^{d_i},$$

y dado que $d_i > 0$ se sigue que

$$[k_i!/(k_i+d_i)!]p_i^{d_i} \le [p_i/(k_i+1)]^{d_i}. \tag{7}$$

Con estas consideraciones, a partir de (4)-(7) se desprende que

$$\alpha(\tilde{m})/\alpha(\tilde{k}) \le \prod_{d_i<0}(k_i/p_i)^{-d_i} \prod_{d_i>0}[p_i/(k_i+1)]^{d_i}. \tag{8}$$

Defina

$$k/p := \text{máx}\{k_i/p_i : i = 1,\ldots,r\},$$

y, utilizando (3), sea

$$\begin{aligned} b &:= \Sigma\{-d_i : d_i < 0\} \\ &= \Sigma\{d_i : d_i > 0\}. \end{aligned} \tag{9}$$

Luego, de la desigualdad (8) obtenemos

$$\alpha(\tilde{m})/\alpha(\tilde{k}) \le (k/p)^b \prod_{d_i>0}[p_i/(k_i+1)]^{d_i} . \tag{10}$$

Por otra parte, si se satisface el conjunto de desigualdades (1), entonces en particular

$$p_i/(k_i+1) \le p/k \text{ , para toda } i$$

y por lo tanto (9) y la desigualdad (10) conducen al resultado :

$$\alpha(\tilde{m})/\alpha(\tilde{k}) \le (k/p)^b(p/k)^b = 1. \qquad \blacksquare$$

El lema anterior conduce en forma inmediata al siguiente resultado.

Lema 1.18. El punto $(k_1,\ldots,k_r)$ que maximiza la función de probabilidad en (1.24) satisface las desigualdades

$$np_i - 1 < k_i \le (n + r - 1)p_i \quad , \ i = 1,\ldots,r. \tag{1.28}$$

Demostración. Si $a(\tilde{\mathbf{k}})$ es el valor máximo, el lema 1.17 mostrado anteriormente establece que

$$p_i k_j \le p_j(k_i + 1) \quad \text{para todo } i,j = 1,\ldots,r.$$

Sumando estas desigualdades para i fijo y para todo j :

$$np_i < k_i + 1 \text{ , para cualquier } i = 1,\ldots,r. \tag{1}$$

Sumando ahora sobre i, i ≠ j, j arbitraria pero fija :

$$k_j \le (n + r - 1)p_j \text{ , } j = 1,\ldots,r \tag{2}$$

Conjuntando (1) y (2) se sigue el resultado. ■

Aquí cabe hacer notar que el conjunto de desigualdades (1.28) establecen un máximo *no necesariamente único* y, lo que es más importante, establecen la igualdad asintótica de k_i y np_i (ver la definición 1.1). Es precisamente esta implicación de (1.28) la que permite obtener el comportamiento asintótico del máximo de la función de probabilidad en (1.24), cuestión que pasamos a resolver a continuación y que jugará un importante papel para obtener el resultado del próximo capítulo.

Teorema 1.19. Si a_0 es máximo de la distribución (1.24), entonces

$$a_0 \sim (2\pi n)^{-(r-1)/2}(p_1 \cdots p_r)^{-1/2} \tag{1.29}$$

Demostración. Las desigualdades (1.29) establecen que si $a_0 = a(\tilde{\mathbf{k}})$ es máximo de la distribución (1.24), entonces los enteros k_i, $i = 1,\dots,r$, pueden expresarse como

$$k_i = np_i + \delta_i \quad , \ -1 < \delta_i \le (r-1)p_i \ ,$$

con $\sum_{i=1}^{r}\delta_i = 0$; esto es

$$np_i/k_i = 1 - (\delta_i/k_i) \ , \quad i = 1,\dots,r. \tag{1}$$

Defina

$$c_n := n!/[(2\pi n)^{1/2}n^n \exp(-n)].$$

Con esta notación, observe primeramente que

$$\begin{aligned}
a_0 &= n! \prod_{i=1}^{r}(p_i^{k_i}/k_i!) \\
&= c_n\cdot(2\pi n)^{1/2}n^n\exp(-n)\cdot\prod_{i=1}^{r}(p_i^{k_i}/k_i!) \\
&= c_n\cdot(2\pi n)^{1/2}\cdot\prod_{i=1}^{r}(np_i^{k_i})\prod_{i=1}^{r}[\exp(-k_i)/k_i!] \\
&= c_n\cdot(2\pi n)^{1/2}\cdot\prod_{i=1}^{r}(np_i/k_i)^{k_i}\prod_{i=1}^{r}[k_i^{k_i}\cdot\exp(-k_i)/k_i!] \qquad (2) \\
&= c_n\cdot(2\pi n)^{1/2}\cdot d_n,
\end{aligned}$$

donde d_n es la sucesión formada por los dos productos del lado derecho de (2).

Por otra parte,

$$d_n = \prod_{i=1}^{r}(np_i)^{-1/2}\prod_{i=1}^{r}(np_i/k_i)^{k_i+1/2}\prod_{i=1}^{r}[k_i^{k_i+1/2}\exp(-k_i)/k_i!],$$

y multiplicando por $(2\pi)^{r/2}$ en ambos miembros de la igualdad anterior y reordenando se obtiene

$$\begin{aligned}
&d_n(2\pi n)^{r/2}(p_1\cdots p_r)^{1/2} \\
&\qquad = \prod_{i=1}^{r}(np_i/k_i)^{k_i+1/2}\prod_{i=1}^{r}[(2\pi)^{1/2}k_i^{k_i+1/2}\exp(-k_i)/k_i!]. \qquad (3)
\end{aligned}$$

Combinando (2) y (3) se sigue que

$$a_0 = b_n \cdot c_n \cdot \gamma_n, \tag{4}$$

donde

$$b_n := (2\pi n)^{-(r-1)/2}(p_1 \cdots p_r)^{-1/2},$$

$$\gamma_n := \prod_{i=1}^{r}[(2\pi)^{1/2}k_i^{k_i+1/2}\exp(-k_i)/k_i!]\prod_{i=1}^{r}(np_i/k_i)^{k_i+1/2}.$$

De acuerdo a (4), para concluir el resultado (1.29) es suficiente mostrar que $c_n \cdot \gamma_n$ converge a 1, para lo cual considere los siguientes argumentos.

Para $\varepsilon > 0$ arbitrario, elija

$$\alpha < \text{mín}\{1 - (1-\varepsilon)^{1/2}, (1+\varepsilon)^{1/2}-1\}.$$

Note que, en este caso,

$$1 - \varepsilon < (1 - \alpha)^2 \text{ y } (1 + \alpha)^2 < 1 + \varepsilon. \tag{5}$$

Recuerde ahora que, para todo $x \in \mathbb{R}$,

$$(\text{i})\ (1 + x/m)^m \longrightarrow e^x \text{ y } (\text{ii})\ (1 + x/m)^{1/2} \longrightarrow 1,$$

conforme $m \to \infty$. Tomando en cuenta que las cantidades δ_i en (1) permanecen acotadas y suman cero para cualquier n, entonces de (1), (i), (ii) se desprende que, si $n \to \infty$,

$$\prod_{i=1}^{r}[np_i/k_i]^{k_i} \cdot \prod_{i=1}^{r}[np_i/k_i]^{1/2} \longrightarrow e^0 \cdot 1 = 1$$

y por lo tanto existe un entero positivo N_1 tal que

$$1 - \alpha < \prod_{i=1}^{r}(np_i/k_i)^{k_i+1/2} < 1 + \alpha \text{ , si } n > N_1. \tag{6}$$

Por otra parte, a partir de la fórmula de Stirling (1.2) se tiene que

$$(2\pi k_i)^{1/2} k_i^{k_i} \exp(-k_i)/k_i! \longrightarrow 1 \ , \ i = 1,\dots,r$$

$$n!/[(2\pi n)^{1/2} n^n \exp(-n)] \longrightarrow 1 \ ,$$

cuando $n \to \infty$, lo cual implica que el producto de estas sucesiones también converge a 1. Así, existe $N_2 \in \mathbb{N}$ tal que, para $n > N_2$,

$$1 - \alpha < c_n \cdot \prod_{i=1}^{r} [(2\pi)^{1/2} k_i^{k_i+1/2} \exp(-k_i)/k_i!] < 1 + \alpha. \qquad (7)$$

Considerando (5)-(7) se sigue que para $\varepsilon > 0$ arbitrario, si $n > N \equiv \max\{N_1, N_2\}$, entonces

$$1 - \varepsilon < c_n \cdot \gamma_n < 1 + \varepsilon. \qquad (8)$$

A partir de la definición de b_n, (4) y (8) se concluye que

$$\alpha_0/(2\pi n)^{-(r-1)/2}(p_1 \cdots p_r)^{-1/2} \longrightarrow 1, \ n \to \infty. \qquad \blacksquare$$

II TEOREMA DE DE MOIVRE-LAPLACE PARA VECTORES MULTINOMIALES

1 INTRODUCCION

En este capítulo se extiende el teorema de DeMoivre-Laplace a vectores aleatorios con distribución multinomial. Este resultado constituirá el pilar fundamental de este trabajo, y a partir de él nos será posible deducir la versión general del teorema del límite central. Asimismo, como una primera aplicación de esta extensión, obtendremos un teorema de límite central para vectores aleatorios con soporte finito.

El capítulo se ha organizado en la siguiente forma: en la sección 2 se presenta un resumen de las características algebraicas de las matrices no negativas y se introduce la distribución normal multivariada; en la sección 3 se realiza la extensión del teorema de De Moivre-Laplace a la distribución multinomial (teorema 2.13); y, por último, la sección 4 se ocupa de generalizar el teorema 2.13 a distribuciones arbitrarias con soporte finito.

2 PRELIMINARES

Matrices no negativas. En esta parte se presentan los resultados fundamentales del álgebra de matrices que se requieren para introducir la distribución normal multidimensional.

Si A es una matriz real de orden $m \times n$ hablaremos de *los vectores* de A para referirnos al conjunto de vectores formados por sus columnas. Es decir, si $A = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$, con $\mathbf{a}_i \in \mathbb{E}_m$, $i = 1,2,\ldots,n$, diremos que $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ son los vectores de A. El conjunto de combinaciones lineales de estos vectores, a saber

$$\mathcal{L}_A = \{\mathbf{v} \in \mathbb{E}_m : \mathbf{v} = \sum_{i=1}^{n} c_i \mathbf{a}_i,\ c_i \in \mathbb{R}\},$$

constituye un espacio vectorial y recibe el nombre de *espacio generado por los vectores de* A o *espacio columna de* A. Claramente $\dim(\mathcal{L}_A) \le n$, y $\dim(\mathcal{L}_A) > 0$ a menos que A sea una matriz nula.

Definición 2.1.a. Si A es una matriz $m \times n$ el *rango* de A se denota por $r(A)$ y se define como $r(A) := \dim(\mathcal{L}_A)$.

En palabras, $r(A)$ es el número máximo de vectores linealmente independientes del conjunto $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$. Si $r(A) = n$ diremos que A es de *rango completo*.

Por otra parte, la relación entre el determinante de una matriz y la dependencia o independencia lineal de sus vectores muestra la equivalencia de la anterior y la siguiente definición.

Definición 2.1.b. Si A es una matriz no nula de orden $m \times n$, entonces $r(A)$ es el mayor entero k tal que A posee una submatriz (cuadrada) no singular de orden $k \times k$.

A partir de la formulación anterior, no es difícil establecer las siguientes afirmaciones acerca de $r(A)$:

(i) $r(A) \leq \text{mín}\{m,n\}$.

(ii) Para cualesquier matriz A

$$r(A) = r(A'). \tag{2.1}$$

(iii) Si A es una matriz $n \times n$, entonces

$$\det(A) = 0 \text{ si y sólo si } r(A) < n. \tag{2.2}$$

Para el producto de matrices, el siguiente teorema relaciona los rangos del producto y los factores.

Teorema 2.2. Sean A,B,C tres matrices de órdenes adecuados para que exista el producto CAB. Entonces

(i) $r(AB) \leq \text{mín}\{r(A),r(B)\}$. (2.3)

(ii) Si B y C son no singulares

$$r(CAB) = r(CA) = r(AB) = r(A). \tag{2.4}$$

Demostración. (i) Como las columnas de AB son combinaciones lineales de las columnas de A, entonces $\mathcal{L}_{AB} \subseteq \mathcal{L}_{A}$ y por lo tanto $r(AB) \leq r(A)$. Por otra parte, de (2.1) se tiene que

$r(AB) = r(B'A') \leq r(B') = r(B)$.

(ii) Del inciso anterior se concluye que

$$r(CAB) \leq r(AB) = r(C^{-1}CAB) \leq r(CAB),$$

y entonces $r(AB) = r(CAB)$.

Análogamente

$$r(AB) \leq r(A) = r(ABB^{-1}) \leq r(AB),$$

de donde se desprende $r(A) = r(AB)$. ∎

Pasemos ahora a definir a las matrices no negativas. Para esto, recuerde primeramente que una matriz cuadrada A es simétrica si satisface la relación $A = A'$.

Definición 2.3. Sea A una matriz simétrica de orden $n \times n$. A es una matriz

(a) *positiva semidefinida* si y sólo si

$$\mathbf{y}'A\mathbf{y} \geq 0 \text{ para todo } \mathbf{y} \in \mathbb{E}_n,$$

y además (2.5)

$$\mathbf{y}'A\mathbf{y} = 0 \text{ para } \textit{algún} \text{ vector no nulo } \mathbf{y} \in \mathbb{E}_n.$$

(b) *positiva definida* si y sólo si

$$\mathbf{y}'A\mathbf{y} > 0 \quad \text{para cualquier vector } \mathbf{y} \in \mathbb{E}_n,\ \mathbf{y} \neq \mathbf{0}. \qquad (2.6)$$

(c) *no negativa* si A es positiva semidefinida o positiva definida.

De la definición anterior se siguen inmediatamente las siguientes propiedades de las matrices no negativas.

Lema 2.4. Sea $A = [a_{ij}]$, $i,j = 1,\ldots,n$, una matriz simétrica. Las condiciones (1.a),(2.a),(3.a) se satisfacen si A es positiva semidefinida, en tanto que (1.b),(2.b),(3.b) se siguen si a es positiva definida.

(1.a) $r(A) < n$.

(2.a) $a_{ii} \geq 0$, $i = 1,\ldots,n$.

(3.a) Para toda matriz P de orden $n \times n$, $P'AP$ es positiva semidefinida.

(1.b) $r(A) = n$.

(2.b) $a_{ii} > 0$, $i = 1,\ldots,n$.

(3.b) $P'AP$ es positiva definida, para toda matriz P no singular. En particular A^{-1} es una matriz positiva.

No obstante que el lema anterior proporciona cierta información acerca de las matrices no negativas, ésta no es suficiente en el sentido de que sólo establece condiciones necesarias. Para determinar con más precisión las propiedades de estas matrices se requiere estudiar el problema de los valores propios de una matriz.

Para una matriz A de orden $n \times n$, este problema consiste en encontrar los valores λ tales que

$$Ay = \lambda \mathbf{y}, \text{ para algún vector } \mathbf{y} \text{ no nulo.} \tag{2.7}$$

En caso de que esto ocurra, se dice que λ es un *valor propio* de A, y que el vector no nulo $\mathbf{y}$ es un *vector propio* correspondiente a λ. La condición (2.7) es equivalente a decir que

$$(A - \lambda I_n)\mathbf{y} = \mathbf{0}, \text{ para algún } \mathbf{y} \neq \mathbf{0},$$

lo cual puede ocurrir si y sólo si la matriz $A - \lambda I_n$ es singular, es decir, si y sólo si

$$\det(A - \lambda I_n) = 0. \qquad (2.8)$$

La ecuación (2.8), llamada *ecuación característica* de A, representa un polinomio de grado n en λ y por lo tanto tiene exactamente n soluciones, no necesariamente diferentes ni reales.

De esta manera, el problema (2.7) conduce en general a valores y vectores propios complejos. Sin embargo, si A es simétrica es conocido que *todos* sus valores propios (y por lo tanto también los vectores propios asociados) son necesariamente reales. Además, para este tipo de matrices se tiene un importante resultado, conocido como *teorema de los ejes principales*.

Teorema 2.5. Si A de orden $n \times n$ es simétrica existe una *base ortonormal* de $\mathbb{E}_n$ formada por vectores propios de A.

Se puede encontrar una demostración de este teorema en Máltsev (1978).

El teorema anterior es equivalente a afirmar que para toda matriz simétrica A existe una matriz ortogonal P (formada por la base ortonormal mencionada en el teorema 2.5) tal que

$$AP = P\Delta = P \text{ Diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}, \tag{2.9}$$

donde las λ_i son los valores propios de A. Como P es inversible, de (2.3) y (2.9) se sigue que $r(A) = r(\Delta)$; de esta manera, si A es simétrica, el número de valores propios diferentes de cero es igual al rango de A. Igualmente, la formulación (2.9) del teorema de los ejes principales y el lema 2.4 permiten establecer las condiciones que caracterizan a las matrices no negativas, las cuales se resumen en el siguiente teorema.

Teorema 2.6. Sea A una matriz simétrica de orden $n \times n$. Las condiciones (1.a) y (2.a) son suficientes y necesarias para que A sea positiva semidefinida, en tanto que (1.b) y (2.b) ocurren si y sólo si A es positiva definida.

(1.a) Los valores propios deA son no negativos y al menos uno es cero.

(2.a) Existe una matriz Γ de orden $k \times n$ tal que

i) $r(\Gamma) = r(A) = k$, y *ii*) $A = \Gamma'\Gamma$.

(1.b) Todos los valores propios de A son positivos.

(2.b) Existe una matriz inversible Γ de orden $n \times n$ tal que

$A = \Gamma'\Gamma$.

Concluimos esta sección aplicando estos resultados a la matriz de covarianza de un vector aleatorio $\mathbf{X}$ de dimensión p. Según la definición (1.7) y la propiedad (1.4) se concluye que $COV(\mathbf{X})$ es no negativa pues

$$\begin{aligned} \mathbf{y}'COV(\mathbf{X})\mathbf{y} &= E[\mathbf{y}'(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])'\mathbf{y}] \\ &= E[(\mathbf{y}'(\mathbf{X} - E[\mathbf{X}]))^2] \\ &\geq 0, \end{aligned}$$

para cualquier vector no nulo $\mathbf{y} \in \mathbb{E}_p$. Entonces, de acuerdo al Teorema 2.6, siempre podemos encontrar una factorización de rango completo de la forma $COV(\mathbf{X}) = \Gamma'\Gamma$, con Γ una matriz de orden $k \times p$, $k = r(COV(\mathbf{X}))$; además, $COV(\mathbf{X})$ será positiva definida si y sólo si $k = p$.

Distribución normal multivariada. A continuación se establecen las definiciones y algunos resultados importantes acerca de la distribución normal multidimensional. Las demostraciones y comentarios al respecto pueden encontrarse en Graybill (1976) y Scheffé (1959).

Definición 2.7. El vector aleatorio $\mathbf{Z}$ de dimensión k tiene *distribución normal estándar k-variada* si y sólo si $\mathbf{Z}$ tiene función de densidad dada por

$$\varphi_k(\mathbf{z};\mathbf{0},\mathbf{I}) := (2\pi)^{-k/2}\exp(-\tfrac{1}{2}\,\mathbf{z}'\mathbf{z}) \;,\; \forall \mathbf{z} \in \mathbb{E}_k; \tag{2.10}$$

en este caso escribimos $\mathbf{Z} \simeq \mathcal{N}_k(\mathbf{0},\mathrm{I})$.

La función de densidad (2.10) puede considerarse una extensión de la distribución normal estándar unidimensional, en tanto representa la distribución conjunta de k variables $Z_1,\ldots,Z_k$ mutuamente independientes, cada una con distribución marginal normal estándar. A partir de esta observación, es fácil ver que, si $\mathbf{Z} \simeq \mathcal{N}_k(\mathbf{0},\mathrm{I})$, entonces

$$E[\mathbf{Z}] = \mathbf{0}, \ \mathrm{COV}(\mathbf{Z}) = \mathrm{I}_k. \tag{2.11}$$

Análogamente al caso unidimensional, la distribución normal estándar multidimensional genera otras distribuciones para permitir momentos diferentes a (2.11).

Definición 2.8. Sea $\mathbf{X}$ un vector aleatorio p-dimensional con $E[\mathbf{X}] = \mathbf{b}$, $\mathrm{COV}(\mathbf{X}) = \Sigma$, $r(\Sigma) = k > 0$. Se dice que $\mathbf{X}$ tiene *distribución normal* p-*variada de rango* k *con parámetros* $\mathbf{b}$ *y* Σ, y se denota por $\mathbf{X} \simeq \mathcal{N}_p(\mathbf{b},\Sigma)$, si y sólo si $\mathbf{X}$ tiene la misma distribución que el vector aleatorio $\Gamma'\mathbf{Z} + \mathbf{b}$, donde $\Sigma = \Gamma'\Gamma$ es una factorización de rango completo para Σ y $\mathbf{Z} \simeq \mathcal{N}_k(\mathbf{0},\mathrm{I})$.

La definición anterior establece que $\mathbf{X}$ tiene distribución normal p-variada si su distribución es la misma que la distribución conjunta de p funciones lineales (no homogéneas) de k variables mutuamente independientes, cada una con distribución normal estándar. En este sentido, observemos que si $\mathbf{X} \simeq \mathcal{N}_p(\mathbf{b},\Sigma)$, $r(\Sigma) = k$, el soporte de la distribución de $\mathbf{X}$ puede reducirse a un subespacio de dimensión $k \le p$.

Si $\mathbf{X} \simeq \mathcal{N}_p(\mathbf{b},\Sigma)$, se pueden establecer los siguientes resultados fundamentales para esta distribución :

(1) A pesar de que la factorización de Σ que aparece en la definición no es única, la distribución de $\mathbf{X}$ queda perfectamente determinada por $\mathbf{b}$ y Σ.

(2) $\mathbf{X} \simeq \mathcal{N}_p(\mathbf{b},\Sigma)$ tiene función de densidad si y sólo si $k = p$. En este caso la función de densidad está dada por

$$\varphi_p(\mathbf{x};\mathbf{b},\Sigma) = [(2\pi)^p \det(\Sigma)]^{-1/2} \exp\{-\tfrac{1}{2}(\mathbf{x}-\mathbf{b})'\Sigma^{-1}(\mathbf{x}-\mathbf{b})\}. \quad (2.12)$$

Por último, dado que se ha definido la distribución normal multivariada con parámetros $\mathbf{b}$ y Σ utilizando transformaciones lineales de $\mathbf{Z} \simeq \mathcal{N}_k(\mathbf{0},I)$, no es difícil ver que las transformaciones lineales de vectores con distribución normal $\mathbf{b}$, Σ también tienen distribuión normal. Precisando, se tiene el siguiente resultado.

Teorema 2.9. Sean B una matriz constante de orden $q \times p$, $\mathbf{a} \in \mathbb{E}_q$ un vector arbitrario. Si $\mathbf{X} \simeq \mathcal{N}_p(\mathbf{b},\Sigma)$, entonces

$$Y = BX + \mathbf{a} \simeq \mathcal{N}_q(B\mathbf{b} + \mathbf{a}, B\Sigma B'). \quad (2.13)$$

3 EL TEOREMA DE DE MOIVRE-LAPLACE Y LA DIST. MULTINOMIAL

El Teorema de De Moivre-Laplace, formulación del

Finalmente, note que $\mathbf{1'X} = \sum_{i=1}^{r-1} x_i = -x_r$, $\mathbf{1} = (1,1,\ldots,1)' \in \mathbb{E}_{r-1}$. Luego

$$\begin{aligned}\mathbf{X'}J_{r-1}\mathbf{X} &= \mathbf{X'(1\cdot 1')X} \\ &= \mathbf{(1'X)'1'X} \\ &= x_r^{\,2},\end{aligned}$$

y combinando esta igualdad con (1) se concluye que

$$\begin{aligned}\mathbf{X'}\tilde{V}^{-1}\mathbf{X} &= p_r^{-1}x_r^{\,2} + \sum_{i=1}^{r-1} x_i^{\,2}/p_i \\ &= \sum_{i=1}^{r} x_i^{\,2}/p_i. \qquad \blacksquare\end{aligned}$$

Lema 2.11. **(i)** Suponga que $t \in \mathbb{R}$ es tal que $|t| \le 1/2$. Entonces

$$1 + t = \exp(t)\exp(\Delta(t)),$$

donde $|\Delta(t)| \le 2t^2$.

(ii) Si $t_1,\ldots,t_l$ satisfacen $|t_i| \le 1/2$, entonces

$$\prod_{i=1}^{l}(1 + t_i) = \exp\left(\sum_{i=1}^{l} t_i\right)\exp(\Delta_l),$$

donde $\Delta_l \equiv \Delta_l(t_1,\ldots,t_l)$ es tal que $|\Delta_l| \le 2\sum_{i=1}^{l} t_i^{\,2}$.

Demostración. (i) Sea $f(t) := \ln(1 + t)$, $t \in [-1/2,1/2]$. Usando la fórmula de Taylor con residuo vemos que

$$f(t) = f(0) + f'(0)\cdot t + f''(\tilde{t})\cdot t^2/2,$$

donde $\tilde{t}$ está entre 0 y t; en particular $\tilde{t} \in [-1/2,1/2]$. Ahora, note que

$$f(0) = 0,\ f'(0) = 1,\ \text{y}\ f''(\tilde{t}) = -1/(1 + \tilde{t})^2,$$

de donde se desprende que

$$\ln(1 + t) = t + \Delta(t), \tag{1}$$

donde

$$\Delta(t) := [- 1/(1 + \tilde{t})^2](t^2/2). \tag{2}$$

Por otra parte, como $\tilde{t} \geq - 1/2$, se tiene

$$1/(1 + \tilde{t})^2 \leq 4,$$

y entonces de (2) se sigue que

$$|\Delta(t)| = [1/(1 + \tilde{t})^2](t^2/2) \leq 2t^2. \tag{3}$$

Finalmente, tomando la función exponencial en ambos miembros de (1) se concluye que

$$1 + t = \exp(t)\exp(\Delta(t)),$$

y el resultado se sigue usando (3).

(ii) Esta afirmación es consecuencia inmediata de (i). ■

Teorema 2.12. Sea $Y_1,Y_2,\ldots$ una sucesión de vectores aleatorios independientes con distribución común $\mathcal{M}_r(1;p_1,\ldots,p_r)$. Si K y ε son números positivos arbitrarios, entonces existe un entero N > 0 con la siguiente propiedad :

Para todo vector $\tilde{m} = (m_1,\ldots,m_{r-1})'$ de enteros no negativos tal que

$$|m_i - np_i| \leq K\, n^{1/2},\ i = 1,2,\ldots,r-1 \tag{*}$$

se tiene que, para todo $n \geq N$,

$$1 - \varepsilon < P[\tilde{S}_n = \tilde{m}]/[h^{r-1}\varphi_{r-1}(h\mathbf{x};0,\tilde{V})]$$

$$< 1 + \varepsilon \,, \tag{2.14}$$

donde

$$h := 1/n^{1/2}, \quad \mathbf{x} := \tilde{m} - n\tilde{p},$$

y φ_{r-1} es la función de densidad normal multivariada definida en (2.12).

Demostración. Por claridad, dividiremos la demostración en varias etapas.

Etapa 1. Comparación de $a(\tilde{m}) = P[\tilde{S}_n = \tilde{m}]$ con el término máximo de la distribución $\mathcal{M}_r(n;p_1,\ldots,p_r)$. Recuerde que para $\mathbf{m} = (m_1,\ldots,m_{r-1})' \in \mathbb{N}^{r-1}$:

$$a(\tilde{m}) = \frac{n!}{m_1!\ldots m_r!}\, p_1^{m_1}\ldots p_r^{m_r} \,, \tag{1}$$

donde

$$m_r := n - (m_1+\ldots+m_{r-1}). \tag{2}$$

El punto importante ahora es comparar $a(\tilde{m})$ con $a_0 \equiv$ término máximo de la función de probabilidad multinomial. Observe que

$$a_0 = \frac{n!}{k_1!\ldots k_r!}\, p_1^{k_1}\ldots p_r^{k_r} \,, \tag{3}$$

donde los enteros no negativos k_i satisfacen

$$k_i = np_i + \delta_i \,, \; -1 < \delta_i \le (r-1)p_i \,, \; i = 1,\ldots,r, \tag{4}$$

y, desde luego, $n = k_1+\ldots+k_r$; vea lema 1.18 en el cap. I. Defina

$$d_i := m_i - k_i \ , \ i = 1,\ldots,r. \tag{5}$$

Usando (4), note que

$$|d_i| = |m_i - np_i - \delta_i|$$

$$\leq |m_i - np_i| + r,$$

y, a partir de la condición (*) se sigue

$$|d_i| \leq Kn^{1/2} + r \ . \tag{6}$$

Combinando (1) y (3) vemos que

$$a(\tilde{m})/a_0 = \prod_{i=1}^{r}(k_i!/m_i!)p_i^{m_i - k_i}$$

y usando (5) obtenemos

$$a(\tilde{m})/a_0 = \prod_{i=1}^{r}[k_i!/(k_i + d_i)!]p_i^{d_i} \ . \tag{7}$$

Etapa 2. Estimación de los factores en (7).

A continuación obtendremos una estimación para cada uno de los factores $[k_i!/(k_i + d_i)!]p_i^{d_i}$, $i = 1,\ldots,r$.

Caso 1. Suponga que $d_i < 0$. En esta situación se tiene que

$$[k_i!/(k_i + d_i)!]p_i^{d_i}$$

$$= k_i(k_i - 1)\cdots(k_i + d_i + 1)p_i^{d_i}$$

$$= (np_i + \delta_i)(np_i + \delta_i - 1)\cdots(np_i + \delta_i + d_i + 1)p_i^{d_i},$$

donde se ha utilizado (4) para la última igualdad.

Luego

$$[k_i!/(k_i + d_i)!]p_i^{d_i} = [\prod_{u=0}^{-d_i - 1}(np_i + \delta_i - u)] \ p_i^{d_i}$$

$$= p_i^{d_i} \prod_{u=0}^{-d_i - 1} np_i[1 + (\delta_i - u)/np_i]$$

$$= p_i^{d_i} \prod_{u=0}^{-d_i-1} np_i \prod_{u=0}^{-d_i-1} [1 + (\delta_i - u)/np_i]. \qquad (8)$$

Recuerde ahora que $d_i < 0$. Entonces,

$$\prod_{u=0}^{-d_i-1} np_i = n^{-d_i} p_i^{-d_i},$$

y combinando esta igualdad con (8) vemos que

$$[k_i!/(k_i + d_i)!]p_i^{d_i} = n^{-d_i} \prod_{u=0}^{-d_i-1} [1 + (\delta_i - u)/np_i]. \qquad (9)$$

Por otro lado, utilizando (4) y (6), observe que para $u = 0,1,\ldots,-d_i - 1$:

$$\begin{aligned} |(\delta_i - u)/np_i| &\leq (|\delta_i| + |d_i| + 1)/np_i \\ &\leq (r - 1 + |d_i| + 1)/np_i \\ &\leq (2r + Kn^{1/2})/np_i, \end{aligned}$$

y en consecuencia

$$\begin{aligned} |(\delta_i - u)/np_i| &\leq (2r + Kn^{1/2})/np_i \\ &\leq (2r + K)n^{1/2}/np_i \\ &\leq [(2r + K)/p]/n^{1/2} \equiv M/n^{1/2} \qquad (10) \end{aligned}$$

donde

$$\begin{aligned} p &:= \min\{p_i : i = 1,2,\ldots,r\}, \\ M &:= (2r + K)/p. \end{aligned} \qquad (10')$$

A partir de (10) se sigue que para $n > (2M)^2$:

$$|(\delta_i - u)/np_i| \leq 1/2, \text{ para todo } u = 0,1,\ldots,-d_i - 1.$$

Finalmente, a partir de (9) y usando (ii) del lema 2.11 con $t_u = (\delta_i - u)/np_i$, $u = 0,1,\ldots,-d_i-1$ se tiene, para $n > (2M)^2$,

$$[k_i!/(k_i+d_i)!]p_i^{d_i} = n^{-d_i}\exp\Big(\sum_{u=0}^{-d_i-1}(\delta_i-u)/np_i\Big)\exp(\Delta_i), \quad (11)$$

donde se satisface

$$\begin{aligned} |\Delta_i| &\le 2\sum_{u=0}^{-d_i-1}[(\delta_i-u)/np_i]^2 \\ &\le 2\sum_{u=0}^{-d_i-1}(M^2/n) \\ &= 2(-d_i)M^2/n \\ &\le 2(K+r)n^{1/2}M^2/n\ ; \end{aligned}$$

vea (10) y (6).

Equivalentemente,

$$|\Delta_i| \le 2(K+r)M^2/n^{1/2}. \quad (12)$$

Finalmente, observe que

$$\sum_{u=0}^{-d_i-1}(\delta_i-u) = -\delta_i d_i - d_i(d_i+1)/2 \quad (13)$$

y así,

$$\begin{aligned} &\exp\Big(\sum_{u=0}^{-d_i-1}(\delta_i-u)/np_i\Big) \\ &\quad = \exp[-\delta_i d_i/(np_i) - d_i^2/(2np_i)]\cdot\exp[-d_i/(2np_i)]. \end{aligned}$$

Usando (11) concluimos que, si $n > (2M)^2$, entonces

$$\begin{aligned} &[k_i!/(k_i+d_i)!]p_i^{d_i} \\ &\quad = n^{-d_i}\exp[-\delta_i d_i/(np_i) - d_i^2/(2np_i)]\cdot\exp(\overline{\Delta}_i), \end{aligned} \quad (14)$$

donde

$$\overline{\Delta}_i := \Delta_i - d_i/(2np_i),$$

cantidad para la cual se tiene, a partir de (6) y (10') :

$$|\overline{\Delta}_i| \le |\Delta_i| + |d_i|/(2np_i)$$

$$\leq 2(K + r)M^2/n^{1/2} + (K + r)n^{1/2}/(2np)$$

$$\leq C/n^{1/2}, \tag{15}$$

con $C := 2(K + r)M^2 + M$.

En este punto se debe señalar que la parte importante de esta discusión ha sido obtener las expresiones (14) y (15).

Caso 2. Suponga que $d_i > 0$. En este caso, argumentos similares a los utilizados en el caso 1 permiten concluir que, si $n > (2M)^2$, entonces

$$[k_i!/(k_i + d_i)!]p_i^{d_i}$$

$$= n^{-d_i}\exp[-\delta_i d_i/(np_i) - d_i^2/(2np_i)]\cdot\exp(\overline{\Delta}_i), \tag{16}$$

donde

$$|\overline{\Delta}_i| \leq C/n^{1/2}. \tag{17}$$

Etapa 3. Estimación de la razón $a(\tilde{m})/a_0$.

Primero note que $[k_i!/(k_i + d_i)!]p_i^{d_i} = 1$ cuando $d_i = 0$. Entonces, usando (7),(14) y (16) concluimos que

$$a(\tilde{m})/a_0$$

$$= \prod_{i=1}^{r} n^{-d_i} \prod_{i=1}^{r} \exp[-\delta_i d_i/(np_i) - d_i^2/(2np_i)] \prod_{i=1}^{r} \exp(\overline{\Delta}_i). \tag{18}$$

para todo $n > (2M)^{1/2}$.

A continuación observe que

$$\prod_{i=1}^{r} n^{-d_i} = n^{-(d_1+d_2+\ldots+d_r)} = n^0 = 1,$$

donde hemos usado (5) y el hecho de que $\Sigma m_i = \Sigma k_i = n$. Luego, (18) es equivalente a la siguiente afirmación :

Si $n > (2M)^2$ tenemos

$a(\tilde{m})/a_0$

$$= \exp[-\sum_{i=1}^{r}(\delta_i d_i/(np_i) + d_i^2/(2np_i))]\cdot\exp[\sum_{i=1}^{r}\overline{\Delta}_i]$$

$$= \exp[-\sum_{i=1}^{r}(\delta_i + d_i)^2/(2np_i)]\cdot\exp[\sum_{i=1}^{r}\overline{\Delta}_i + \sum_{i=1}^{r}\delta_i^2/(2np_i)] \quad (19)$$

Por otra parte, usando (4) y (5) vemos que para $i = 1,..,r$

$$d_i + \delta_i = m_i - k_i + \delta_i$$

$$= m_i - (np_i + \delta_i) + \delta_i \equiv x_i. \quad (20)$$

Además, si

$$\overline{\Delta} := \sum_{i=1}^{r}\overline{\Delta}_i + \sum_{i=1}^{r}\delta_i^2/(2np_i) \quad , \quad (21)$$

entonces (15),(17),(4) y (10′) implican que

$$|\overline{\Delta}| \le \sum_{i=1}^{r} C/n^{1/2} + \sum_{i=1}^{r} r^2/(2np)$$

$$\le (rC + r^3 p)/n^{1/2} \equiv \overline{C}/n^{1/2} \quad (22)$$

Combinando (19)-(22) y el resultado del lema 2.10 se obtiene

$$a(\tilde{m})/a_0 = \exp[-\sum_{i=1}^{r} x_i^2/(2np_i)]\cdot\exp(\overline{\Delta})$$

$$= \exp[-(\hbar X)'\tilde{V}^{-1}(\hbar X)]\cdot\exp(\overline{\Delta}) \quad , \ n > (2M)^2, \quad (23)$$

donde $\hbar := 1/n^{1/2}$ y $|\overline{\Delta}| \le \overline{C}/n^{1/2}$.

Para concluir, recuerde el resultado del teorema 1.19, según el cual

$$a_0 = (2\pi n)^{-(r-1)/2}(p_1 p_2 \cdots p_r)^{-1/2}\exp(\rho(n))$$

$$= (2\pi n)^{-(r-1)/2}\,(\det(\tilde{V}))^{-1/2}\exp(\rho(n)), \quad (24)$$

donde

$$\rho(n) \to 0 \text{ si } n \to \infty \tag{25}$$

y se ha utilizado (1.21) para la última igualdad en (24). Luego, (23) y (24), junto con la definición (2.12), implican que si $n > (2M)^2$

$a(\tilde{m})$

$$= h^{r-1}(2\pi)^{-(r-1)/2}|\tilde{V}|^{-1/2}\exp[-(hX)'\tilde{V}^{-1}(hX)]\exp(\bar{\Lambda} + \rho(n))$$

$$= h^{r-1}\varphi_{r-1}(hX;0,\tilde{V})\cdot\exp(\bar{\Lambda} + \rho(n)). \tag{26}$$

Finalmente, usando (22),(25) y la continuidad de la función exponencial se tiene que, dado $\varepsilon > 0$ existe $N_1 \in \mathbb{N}$ tal que

$$1 - \varepsilon < \exp(\bar{\Lambda} + \rho(n)) < 1 + \varepsilon \quad , \text{ si } n > N_1. \tag{27}$$

Luego, si $n > N \equiv \max\{(2M)^2, N_1\}$, (26) y (27) implican

$$1 - \varepsilon < a(\tilde{m})/[h^{r-1}\varphi_{r-1}(hX;0,\tilde{V})] < 1 + \varepsilon, \text{ para todo } n > N.\blacksquare$$

La estimación en el teorema 2.12 permite extender el teorema de De Moivre-Laplace al caso de vectores aleatorios con distribución multinomial. Para esto, considere primeramente la siguiente convención notacional.

Notación. Si $Y_1, Y_2, \ldots$ son vectores aleatorios independientes con distribución común $\mathcal{M}_r(1;p_1,\ldots,p_r)$, para todo $n \in \mathbb{N}$ defina

$$\tilde{S}_n^* := (\tilde{S}_n - n\tilde{p})/n^{1/2}, \tag{2.15}$$

donde $\tilde{p} = E(\tilde{Y}_1)$ es como en (1.19).

Teorema 2.13. Sean $Y_1, Y_2, \ldots$ vectores aleatorios independientes con distribución común $\mathcal{M}_r(1; p_1, \ldots, p_r)$, donde $p_1, \ldots, p_r$ son números positivos. Sea $\mathcal{A} \subset \mathbb{E}_{r-1}$ una región como las que se estudian en cálculo[1]. Entonces

$$\lim_{n\to\infty} P[\tilde{S}_n^{\,*} \in \mathcal{A}] = \int_{\mathcal{A}} \varphi_{r-1}(\mathbf{x}; 0, \tilde{V})\, d\mathbf{x}\ , \tag{2.16}$$

donde $\tilde{S}_n^{\,*}$, $\tilde{V} = \mathrm{COV}(\tilde{Y})$ son como en (2.15), (1.20), respectivamente.

Demostración. Dividiremos la discusión en dos partes.

Caso 1. $\mathcal{A}$ es una región *acotada*, digamos

$$\mathcal{A} \subset \{\mathbf{x} = (x_1, \ldots, x_{r-1})': |x_i| < K,\ i = 1, \ldots, r-1\}, \tag{1}$$

donde K es una constante positiva.

Para empezar observe que

$$P[\tilde{S}_n^{\,*} \in \mathcal{A}] = P[(\tilde{S}_n - n\tilde{p})/n^{1/2} \in \mathcal{A}]$$

$$= \Sigma\{P[\tilde{S}_n = \tilde{m}] : (\tilde{m} - n\tilde{p})/n^{1/2} \in \mathcal{A}\} \tag{2}$$

Considere ahora el término típico de la sumatoria en (2), esto es, $P[\tilde{S}_n = \tilde{m}] \equiv a(\tilde{m})$ donde $(\tilde{m} - n\tilde{p})/n^{1/2} \in \mathcal{A}$. Usando (1) vemos que $\mathbf{x} \equiv \mathbf{x}(\tilde{m}) := (\tilde{m} - n\tilde{p})/n^{1/2}$ satisface $|x_i| < K$, para $i = 1, \ldots, r-1$. Luego, usando el teorema 2.10 vemos que, dado $\varepsilon > 0$, existe $N \in \mathbb{N}$ tal que para $n > N$ se tiene

---

[1]La frase "una región como las que se estudian en cálculo" se refiere a que $\mathcal{A}$ sea un conjunto *medible en el sentido de Jordan* (Spivak, 1972).

$$(1 - \varepsilon)h_n^{r-1}\varphi_{r-1}(\mathbf{x}(\tilde{\mathrm{m}});\mathbf{0},\tilde{\mathrm{V}}) < a(\tilde{\mathrm{m}})$$

$$< (1+\varepsilon)h_n^{r-1}\varphi_{r-1}(\mathbf{x}(\tilde{\mathrm{m}});\mathbf{0},\tilde{\mathrm{V}}), \quad (3)$$

donde $h_n := 1/n^{1/2}$. Luego, sumando sobre todos los $\mathbf{m} \in \mathbb{N}^{r-1}$ para los que $\mathbf{x}(\tilde{\mathrm{m}}) \in \mathcal{A}$ y usando (2) se obtiene que, si $n > N$

$$(1 - \varepsilon)\Sigma\{h_n^{r-1}\varphi_{r-1}(\mathbf{x}(\tilde{\mathrm{m}});\mathbf{0},\tilde{\mathrm{V}}) : \mathbf{x}(\tilde{\mathrm{m}}) \in \mathcal{A}\}$$

$$< \mathrm{P}[\tilde{\mathrm{S}}_n^* \in \mathcal{A}]$$

$$< (1 + \varepsilon)\Sigma\{h_n^{r-1}\varphi_{r-1}(\mathbf{x}(\tilde{\mathrm{m}});\mathbf{0},\tilde{\mathrm{V}}) : \mathbf{x}(\tilde{\mathrm{m}}) \in \mathcal{A}\}. \quad (4)$$

Por otro lado, para n suficientemente grande, la sumatoria en (4) es una suma de Riemann para la integral de $\varphi_{r-1}(\cdot;\mathbf{0},\tilde{\mathrm{V}})$ sobre la región $\mathcal{A}$; esta suma corresponde a una partición determinada por cubos de dimensión $r - 1$ con lados de longitud h_n. Usando el hecho de que $h_n = 1/n^{1/2} \to 0$, cuando $n \to \infty$, vemos que

$$\lim_{n\to\infty} \Sigma\{h_n^{r-1}\varphi_{r-1}(\mathbf{x}(\tilde{\mathrm{m}});\mathbf{0},\tilde{\mathrm{V}}):\mathbf{x}(\tilde{\mathrm{m}}) \in \mathcal{A}\} = \int_{\mathcal{A}} \varphi_{r-1}(\mathbf{x};\mathbf{0},\tilde{\mathrm{V}})d\mathbf{x}. \quad (5)$$

Finalmente, tome límite cuando $n \to \infty$ en (4). Usando (5) así como el hecho de que $\varepsilon > 0$, concluimos que

$$\lim_{n\to\infty} \mathrm{P}[\tilde{\mathrm{S}}_n^* \in \mathcal{A}] = \int_{\mathcal{A}} \varphi_{r-1}(\mathbf{x};\mathbf{0},\tilde{\mathrm{V}})\ d\mathbf{x}.$$

Caso 2. $\mathcal{A} \subset \mathbb{E}_{r-1}$ es una región (conjunto medible en el sentido de Jordan) no acotada.

Sea $\varepsilon > 0$ arbitrario. Para cada $K > 0$ defina

$$\mathcal{A}(K) := \mathcal{A} \cap \{\mathbf{x} : |x_i| < K\};$$

note que $\mathcal{A}(K)$ es una región acotada. Además, dado $\varepsilon > 0$ existe $K_0 > 0$ tal que

$$\left| \int_{\mathcal{A}} \varphi_{r-1}(\mathbf{x};0,\tilde{V})\, d\mathbf{x} - \int_{\mathcal{A}(K)} \varphi_{r-1}(\mathbf{x};0,\tilde{V})\, d\mathbf{x} \right| < \varepsilon, \text{ para } K > K_0. \qquad (6)$$

Ahora, escriba $\tilde{S}_n^* = (S_n^1, \ldots, S_n^{r-1})$ y observe que

(a) $E(S_n^i) = 0$, y

(b) $\mathrm{Var}(S_n^i) = p_i(1-p_i) \leq 1/4$, $i = 1,2,\ldots,r-1$.

Luego, de la desigualdad de Chebyschev se sigue que para todo $i = 1,\ldots,r-1$

$$P[|S_n^i| \geq K] \leq \mathrm{Var}(S_n^i)/K^2 \leq 1/(4K^2). \qquad (7)$$

Por otra parte, note que

$$\begin{aligned} |P[\tilde{S}_n^* \in \mathcal{A}] - P[\tilde{S}_n^* \in \mathcal{A}(K)]| &= P[\tilde{S}_n^* \in \mathcal{A} - \mathcal{A}(K)] \\ &\leq P[|S_n^i| \geq K \text{ para algún } i = 1,\ldots,r-1] \\ &\leq \sum_{i=1}^{r-1} P[|S_n^i| \geq K], \end{aligned}$$

y conjuntando con (7) se obtiene

$$|P[\tilde{S}_n^* \in \mathcal{A}] - P[\tilde{S}_n^* \in \mathcal{A}(K)]| \leq (r-1)/4K^2,$$

de donde se desprende que existe $K_1 > 0$ tal que

$$|P[\tilde{S}_n^* \in \mathcal{A}] - P[\tilde{S}_n^* \in \mathcal{A}(K)]| < \varepsilon \text{ , para todo } K > K_1. \qquad (8)$$

Finalmente, seleccione $K > \text{máx}\{K_0, K_1\}$, con lo cual (6) y (8) ocurren. Además, debido a que $\mathcal{A}(K)$ es acotada, el caso

1 discutido anteriormente implica que existe $N(\varepsilon) \in \mathbb{N}$ tal que

$$|P[\tilde{S}_n^* \in \mathcal{A}(K)] - \int_{\mathcal{A}(K)} \varphi_{r-1}(\mathbf{x};\mathbf{0},\tilde{V})\, d\mathbf{x}\ | < \varepsilon \quad \text{si } n > N. \qquad (9)$$

Combinando (6),(8) y (9) y usando la desigualdad del triángulo se tiene

$$|P[\tilde{S}_n^* \in \mathcal{A}] - \int_{\mathcal{A}} \varphi_{r-1}(\mathbf{x};\mathbf{0},\tilde{V})\, d\mathbf{x}\ | < 3\varepsilon \quad \text{si } n > N,$$

lo cual demuestra (2.16) pues $\varepsilon > 0$ es arbitrario. ∎

Para concluir, observe que el teorema (2.15) es un teorema de límite central para los vectores $\tilde{\mathbf{Y}}$. Además, debemos mencionar que la convergencia en (2.16) es el resultado alrededor del cual gira el resto de este trabajo.

4 VECTORES ALEATORIOS CON SOPORTE FINITO

Como se mencionó anteriormente, el Teorema 2.13 puede considerarse como un teorema de límite central genaralizado. A continuación, utilizando argumentos que involucran transformaciones lineales sencillas, extenderemos este resultado al caso de una distribución arbitraria con soporte finito. Esta extensión se presenta en el teorema 2.15.

Sea $\mathbf{X}$ un vector aleatorio de dimensión p y suponga que $\mathbf{X}$ tiene soporte finito, esto es, que existe un conjunto $\mathcal{X} = \{\mathbf{x}_1,\ldots,\mathbf{x}_r\} \subset \mathbb{E}_p$ tal que

$$P[X = \mathbf{x}_i] = p_i > 0,\ i = 1,2,\ldots,r,\ y\ \sum_{i=1}^{r} p_i = 1. \qquad (2.17)$$

Supongamos ahora que $\mathcal{X}$ contiene al vector nulo $\mathbf{0}$. En este caso, sin pérdida de generalidad se puede suponer que $\mathbf{x}_r = \mathbf{0}$. Ahora, defina la matriz B de orden $p \times (r-1)$ mediante

$$B = [\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{r-1}], \qquad (2.18)$$

esto es, la i-ésima columna de B es el vector no nulo $\mathbf{x}_i$. Por otro lado, sea $\mathbf{Y}$ un vector aleatorio con distribución $\mathcal{M}_1(1;p_1,\ldots,p_r)$ y $\tilde{\mathbf{Y}}$ la respectiva proyección sobre $\mathbb{E}_{r-1}$. Note que (1.15), (2.17) y (2.18) conducen a que el vector aleatorio $\tilde{\mathbf{X}}$

$$\tilde{\mathbf{X}} := B\tilde{\mathbf{Y}}, \qquad (2.19)$$

tenga la misma distribución que $\mathbf{X}$. En efecto, si $\tilde{\mathbf{e}}_1,\ldots,\tilde{\mathbf{e}}_{r-1}$ es la base canónica de $\mathbb{E}_{r-1}$,

$$\begin{aligned} P[\tilde{\mathbf{X}} = \mathbf{x}_i] &= P[\tilde{\mathbf{Y}} = \mathbf{e}_i] \\ &= p_i \\ &= P[\mathbf{X} = \mathbf{x}_i], \end{aligned}$$

$i = 1,\ldots,r-1$; además,

$$\begin{aligned} P[\tilde{\mathbf{X}} = \mathbf{0}] &= P[\tilde{\mathbf{Y}} = \mathbf{0}] \\ &= p_r \\ &= P[\mathbf{X} = \mathbf{0}]. \end{aligned}$$

Luego, las propiedades (1.9), (1.19) y (1.20) facilitan la labor de encontrar una expresión para los primeros momentos del vector $\mathbf{X}$, pues éstos coinciden con los de $\tilde{\mathbf{X}} = B\tilde{\mathbf{Y}}$:

independientes con distribución común $\mathcal{M}_r(1;p_1,\ldots,p_r)$. De la discusión realizada al inicio de esta sección se desprende que los vectores aleatorios $\tilde{\mathbf{X}}_i = \mathrm{B}\tilde{\mathbf{Y}}_i$, con B definida en (2.18), son independientes y tienen la misma distribución que $\mathbf{X}$; vea (2.19). Luego, el vector aleatorio

$$\mathbf{X}_n^* = (\sum_{i=1}^{n}\mathbf{X}_i - n\mu)/n^{1/2}$$

tiene la misma distribución que

$$\tilde{\mathbf{X}}_n^* := (\sum_{i=1}^{n}\mathrm{B}\tilde{\mathbf{Y}}_i - n\mathrm{B}\tilde{\mathbf{p}})/n^{1/2}$$

$$= \mathrm{B}(\sum_{i=1}^{n}\tilde{\mathbf{Y}}_i - n\tilde{\mathbf{p}})/n^{1/2}$$

$$= \mathrm{B}\tilde{\mathbf{S}}_n^* ;$$

vea (2.20) y (2.15).

Así, para toda región $\mathcal{A} \subset \mathbb{E}_p$, se tiene

$$\mathrm{P}[\mathbf{X}_n^* \in \mathcal{A}] = \mathrm{P}[\mathrm{B}\tilde{\mathbf{S}}_n^* \in \mathcal{A}]$$

$$= \mathrm{P}[\tilde{\mathbf{S}}_n^* \in \mathcal{A}_1],$$

donde

$$\mathcal{A}_1 := \{\mathrm{Z} : \mathrm{BZ} \in \mathcal{A}\} \subset \mathbb{E}_{r-1}. \tag{1}$$

Luego, el teorema 2.13 implica que

$$\lim_{n\to\infty} \mathrm{P}[\mathbf{X}_n^* \in \mathcal{A}] = \lim_{n\to\infty} \mathrm{P}[\tilde{\mathbf{S}}_n^* \in \mathcal{A}_1]$$

$$= \int_{\mathcal{A}_1} \varphi_{r-1}(\mathbf{x};\mathbf{0},\tilde{\mathrm{V}})\ d\mathbf{x}$$

$$= \mathrm{P}[\mathrm{Z} \in \mathcal{A}_1], \tag{2}$$

con Z tal que

$$\mathrm{Z} \simeq \mathcal{N}_{r-1}(\mathbf{0},\tilde{\mathrm{V}}). \tag{3}$$

Ahora, note que de la definición (1) de $\mathcal{A}_1$ se desprende que $Z \in \mathcal{A}_1$ es equivalente a $BZ \in \mathcal{A}$. Usando esta observación en (2) se sigue que

$$\lim_{n\to\infty} P[X^* \in \mathcal{A}] = P[BZ \in \mathcal{A}]$$

$$= P[\mathbf{W} \in \mathcal{A}], \tag{4}$$

donde

$$\mathbf{W} := BZ. \tag{5}$$

Conjuntando (4),(5),(2.20) y (2.21) se sigue el resultado.

Caso 2. $\mathbf{0} \notin \mathcal{X}$.

En este caso, defina el vector aleatorio

$$\mathbf{T} := \mathbf{X} - \mathbf{x}_r, \tag{6}$$

y observe entonces que el soporte de $\mathbf{T}$ está dado por

$$\mathcal{T} := \{\mathbf{x}_1 - \mathbf{x}_r, \ldots, \mathbf{x}_{r-1} - \mathbf{x}_r, \mathbf{0}\}. \tag{7}$$

De esta manera, si $\mathbf{T}_1, \mathbf{T}_2, \ldots$ es una sucesión de vectores aleatorios independientes con la misma distribución de $\mathbf{T}$, el resultado demostrado anteriormente en el caso 1 implican que, para toda región $\mathcal{A} \subset \mathbb{E}_p$,

$$\lim_{n\to\infty} P[\mathbf{T}_n^* \in \mathcal{A}] = P[\mathbf{W} \in \mathcal{A}], \tag{8}$$

donde

$$\mathbf{W} \simeq \mathcal{N}_p(\mathbf{0}, \mathrm{COV}(\mathbf{T})), \tag{9}$$

$$\mathbf{T}_n^* := [\sum_{i=1}^{n} \mathbf{T}_i - nE(\mathbf{T})]/n^{1/2}. \tag{10}$$

Pero a partir de (6) y (1.9) se sigue que

$$E(\mathbf{T}) = E(\mathbf{X}) - \mathbf{x}_r, \text{ y } \mathrm{COV}(\mathbf{T}) = \mathrm{COV}(\mathbf{X}), \tag{11}$$

y por lo tanto

$$\begin{aligned} \mathbf{T}_n^* &= [\sum_{i=1}^{n}(\mathbf{X}_i - \mathbf{x}_r) - n(\mu - \mathbf{x}_r)]/n^{1/2} \\ &= [\sum_{i=1}^{n}\mathbf{X}_i - n\mathbf{x}_r - n\mu + n\mathbf{x}_r]/n^{1/2} \\ &= \mathbf{X}_n^*. \end{aligned} \tag{12}$$

Conjuntando (8),(9),(11),(12) se concluye que para toda región $\mathcal{A} \subset \mathbb{E}_p$,

$$\lim_{n\to\infty} \mathrm{P}[\mathbf{X}_n^* \in \mathcal{A}] = \mathrm{P}[\mathbf{W} \in \mathcal{A}],$$

donde $\mathbf{W}$ es un vector aleatorio tal que

$$\mathbf{W} \simeq \mathcal{N}_p(\mathbf{0},\mathrm{M}). \qquad \blacksquare$$

III TEOREMA DEL LIMITE CENTRAL UNIDIMENSIONAL

1 INTRODUCCION

En los capítulos anteriores se han establecido los resultados fundamentales que permitirán ahora demostrar el Teorema del límite central (TLC) para variables aleatorias independientes e idénticamente distribuidas con segundo momento finito. El esquema general de esta demostración consiste en establecer primeramente, a partir del teorema 2.15,el teorema para variables aleatorias discretas con soporte finito. Posteriormente -aproximando por variables con soporte finito- se demostrará el teorema para variables acotadas, y, por último, se abordará el caso de una variable no acotada.

De acuerdo al comentario anterior, el material se ha estructurado en las secciones correspondientes a cada una de las etapas mencionadas.

A lo largo del presente capítulo haremos uso de la siguiente convención.

Notación. Si $X_1, X_2, \ldots$ es una sucesión de variables aleatorias independientes con $E(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2 < \infty$, $i =$

1,2,..., escribiremos

$$X_n^* := n^{1/2}(\overline{X}_n - \mu)$$
$$= (\sum_{i=1}^{n} X_i - n\mu)/n^{1/2}. \qquad (3.1)$$

2 EL CASO DE VARIABLES DISCRETAS CON SOPORTE FINITO

Como primer paso hacia la versión general del teorema del límite central, enunciaremos el siguiente resultado, el cual se desprende del teorema 2.15.

Teorema 3.1. Sea X una variable aleatoria con soporte finito y segundo momento finito. Sea $\mu := E(X)$ y suponga que $0 < \mathrm{Var}(X) := \sigma^2 < \infty$. Entonces, si $X_1, X_2, \ldots$ es una sucesión de variables aleatorias independientes con la misma distribución de X, se tiene que

la distribución asintótica de X_n^* es $\mathcal{N}(0,\sigma^2)$.

Más explícitamente, para todo intervalo $A \subset \mathbb{R}$,

$$\lim_{n\to\infty} P[X_n^* \in A] = \int_A \varphi_1(z;0,\sigma^2)\, dz \; ; \qquad (3.2)$$

equivalentemente,

$$\lim_{n\to\infty} P[X_n^*/\sigma \in A] = \int_A \varphi_1(z;0,1)\, dz. \qquad (3.3)$$

Observe que este resultado no es otra cosa que un caso particular del teorema 2.15, correspondiente a p = 1; se ha enunciado separadamente porque es el punto de partida de los argumentos de este capítulo.

3 TLC PARA VARIABLES ALEATORIAS ACOTADAS

Abordaremos ahora el problema cuando X es una variable aleatoria acotada cualesquiera, independientemente de su carácter discreto o continuo. Para esta situación, el teorema del límite central se establecerá en el teorema 3.7, en cuya demostración utilizaremos el hecho de que este tipo de variables pueden aproximarse "suficientemente bien" mediante variables discretas con soporte finito.

Definición 3.2. Sea $A \subseteq \mathbb{R}$ un conjunto arbitrario. La *función indicadora* de A se denota por I_A: $\mathbb{R} \to \mathbb{R}$ y se define por

$$I_A(x) := 1 \text{ si } x \in A, \qquad := 0 \text{ se } x \notin A. \tag{3.4}$$

Observación 3.3. Sea X una variable aleatoria y $A \subset \mathbb{R}$ un intervalo arbitrario.

(i) Note que de (3.4) se sigue que

$$P[I_A(X) = 1] = P[X \in A] \text{ y } P[I_A(X) = 0] = P[A^c].$$

Luego,

$$E(I_A(X)) = P[X \in A]. \tag{3.5}$$

(ii) Suponga que $A_1,\ldots,A_r \subset \mathbb{R}$ son conjuntos disjuntos y que $\bigcup_{i=1}^{r} A_i \supset \mathcal{X} \equiv$ soporte de X. Si G(X) es una función arbitraria de X con esperanza finita, entonces

$$E(G(X)) = \sum_{i=1}^{r} E[G(X) I_{A_r}(X)]. \tag{3.6}$$

Definición 3.4. Sea X una variable aleatoria acotada con media μ y varianza $\sigma^2 > 0$. Sea $M > 0$ tal que soporte $X = \mathcal{X} \subset (-M,M)$. Para cada $\delta \in (0,M)$ sea $a_0, a_1, \ldots, a_r$ la sucesión que satisface

(i) $a_o = -M < a_1 < \ldots < a_{r-1} < M = a_r$, y

(ii) $a_i - a_{i-1} = \delta$, $i = 1,\ldots,r$.

En este caso escribiremos

$$A_j := [a_{j-1}, a_j), \; j = 1,2,\ldots,r.$$

Defina $x_1, x_2, \ldots, x_r$ mediante

$$\begin{aligned} x_j &:= E(X \cdot I_{A_j}(X))/P[X \in A_j] \text{ si } P[X \in A_j] > 0, \\ &:= 0 \quad \text{de otro modo.} \end{aligned} \tag{3.7}$$

Finalmente, la variable aleatoria X^δ se define por

$$X^\delta := \sum_{j=1}^{r} [x_j I_{A_j}(X)] \; ; \tag{3.8}$$

note que esto es equivalente a $X^\delta = x_j$ si $X \in A_j$.
Nos referiremos a X^δ como la *δ-discretizacioń de* X.

En los siguientes lemas se muestra que X^δ puede aproximar tan bien como se quiera a X eligiendo adecuadamente a δ.

Lema 3.5. Si X y X^δ son dos variables aleatorias como en la definición 3.4, entonces

(i) $P[|X - X^{\delta}| \le \delta] = 1$.

(ii) $E(X^{\delta}) = E(X)$.

(iii) $0 \le \mathrm{Var}(X) - \mathrm{Var}(X^{\delta}) \le \delta^2$.

Demostración. *i)* Observe primeramente que

$$X - X^{\delta} = X - x_j \ , \ \text{si } X \in A_j, \tag{1}$$

donde x_j está definida en (3.7) para todo $j = 1,2,\ldots,r$. Mostraremos que si $P[X \in A_j] > 0$ entonces $x_j \in A_j$.

Suponga que A_j es un intervalo, $A_j = [a_{j-1},a_j)$. Observe que si $X \in A_j$, esto es

$$a_{j-1} \le X < a_j,$$

entonces

$$a_{j-1}I_{A_j}(X) \le X\, I_{A_j}(X) < a_jI_{A_j}(X). \tag{2}$$

Utilizando (3.5), (2) y las propiedades del valor esperado se sigue que

$$a_{j-1}P[X \in A_j] \le E(X\, I_{A_j}(X)) \le a_jP[X \in A_j].$$

De esta manera, si $P[X \in A_j] > 0$ se tiene

$$a_{j-1} \le E(X\, I_{A_j}(X))/P[X \in A_j] \le a_j, \tag{3}$$

esto es, si $P[X \in A_j] > 0$ entonces

$$x_j \in A_j \cup \{a_j\}. \tag{4}$$

Por otra parte, se puede verificar que, en este caso, suponer $x_j = a_j$ conduce a una contradicción, y por lo tanto

$$P[X \in A_j] > 0 \text{ implica } x_j \in A_j. \tag{5}$$

A partir de (1) y (5) se sigue

$$P[|X - X^\delta| \leq \delta] = 1.$$

ii) Observe primeramente que si A_j es tal que $P[X \in A_j] = 0$ entonces

$$E(X\, I_{A_j}(X)) = 0. \tag{6}$$

A partir de (3.7), (3.8) y (4) se desprende que

$$\begin{aligned} E(X^\delta) &= \sum_{j=1}^{r} x_j\ P[X^\delta = x_j] \\ &= \sum_{j=1}^{r} x_j\ P[X \in A_j] \\ &= \sum_{j=1}^{r} E[X\, I_{A_j}(X)\ ], \\ &= E(X). \end{aligned}$$

iii) Note que el evento $[(X - X^\delta)^2 \leq \delta^2]$ es equivalente a $[|X - X^\delta| \leq \delta]$. A partir de i) se tiene entonces

$$P[(X - X^\delta)^2 \leq \delta^2] = 1,$$

de donde se deprende que

$$0 \leq E[(X - X^\delta)^2] \leq \delta^2. \tag{7}$$

Por otra parte,

$$\begin{aligned} E[(X - X^\delta)^2] &= \sum_{j=1}^{r} E[(X - X^\delta)^2 I_{A_j}(X)] \\ &= \sum_{j=1}^{r} E[(X^2 - 2XX^\delta + (X^\delta)^2) I_{A_j}(X)] \end{aligned}$$

$$= \sum_{j=1}^{r} E(X^2 I_{A_j}(X)) - 2\sum_{j=1}^{r} E(XX^{\delta} I_{A_j}(X)) + \sum_{j=1}^{r} E[(X^{\delta})^2 I_{A_j}(X)]. \qquad (8)$$

Observe ahora que para todo j = 1,2,...,r se tiene :

$$\begin{aligned} E(XX^{\delta} I_{A_j}(X)) &= x_j E(XI_{A_j}(X)) \\ &= x_j^2 P[X \in A_j] \\ &= x_j^2 E(I_{A_j}(X)) \\ &= E((X^{\delta})^2 I_{A_j}(X)], \end{aligned}$$

por lo que de (8) y el resultado de (ii) se sigue que

$$\begin{aligned} E[(X - X^{\delta})^2] &= \sum_{j=1}^{r} E(X^2 I_{A_j}(X)) - \sum_{j=1}^{r} E[(X^{\delta})^2 I_{A_j}(X)] \\ &= \sum_{j=1}^{r} E(X^2 I_{A_j}(X)) - \mu^2 - \sum_{j=1}^{r} E[(X^{\delta})^2 I_{A_j}(X)] + \mu^2 \\ &= \mathrm{Var}(X) - \mathrm{Var}(X^{\delta}). \qquad (9) \end{aligned}$$

Combinando (7) y (9) se obtiene el resultado. ■

En el siguiente lema se establece que la δ-discretización de X definida en (3.7) y (3.8) es tal que la probabilidad de que $X^{\delta}{}_n{}^*$ esté muy cercana a X_n^* se puede controlar con una elección adecuada de δ. Este resultado es el que permitirá desprender el TLC para variables acotadas a partir del caso de variables con soporte finito.

Lema 3.6. Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias acotadas independientes e idénticamente distribuidas, con media μ y varianza $\sigma^2 > 0$. Para $\delta > 0$, $\delta \ll 1$ sea X_i^δ la δ-discretrización de X_i. Entonces, para todo $n > 0$

$$P[|\ \overline{X}_n - \overline{X_n^\delta}|n^{1/2}/\sigma \geq \delta^{1/2}\ ] \leq \delta/\sigma^2.$$

Demostración. Primero observe que las diferencias $X_i - X_i^\delta$ son independientes e idénticamente distribuidas. Luego, usando la desigualdad de Chebyshev, (i) y (ii) del lema 3.5 se tiene

$$\begin{aligned} P[|\overline{X}_n - \overline{X_n^\delta}|n^{1/2} \geq \sigma\delta^{1/2}\ ] &\leq Var[n^{1/2}(\overline{X}_n - \overline{X_n^\delta})]/\sigma^2\delta \\ &= Var(X - X^\delta)/\sigma^2\delta \\ &= E[(X - X^\delta)^2]/\sigma^2\delta \\ &\leq \delta/\sigma^2. \end{aligned}$$

■

Teorema 3.7. Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias *acotadas* independientes e idénticamente distribuidas, con media μ y varianza $\sigma^2 > 0$. Entonces, para todo intervalo $A \subset \mathbb{R}$,

$$\lim_{n\to\infty} P[X_n^*/\sigma \in A] = \int_A \varphi_1(z;0,1)\ dz.$$

Demostración. La idea de la demostración consiste en construir X^δ "tan parecida a X" de tal forma que lo que se pueda afirmar acerca de $X_n^{\delta *}$ sea también válido para X_n^*. Con el objeto de exponer con mayor claridad los argumentos, dividiremos la demostración en varias etapas.

Etapa 1. Para $\delta > 0$, $\delta \ll 1$ sea X^δ la δ-discretización de X. Sea $X_1^\delta, X_2^\delta, \ldots$ una sucesión de variables aleatorias independientes y con la misma distribución que X^δ.

Defina

$$D_n := (\bar{X}_n - \overline{X_n^\delta}\,)n^{1/2},$$

y note que para cualquier intervalo $A \subset \mathbb{R}$ se tiene

$$\begin{aligned} P[\, X_n^*/\sigma \in A\,] &= P[(X_n^{\delta *}/\sigma) + D_n/\sigma \in A] \\ &= P[(X_n^{\delta *}/\sigma) + D_n/\sigma \in A,\ |D_n/\sigma| \ge \delta^{1/2}] \\ &\quad + P[X_n^{\delta *}/\sigma + D_n/\sigma \in A,\ |D_n/\sigma| < \delta^{1/2}]; \qquad (1) \end{aligned}$$

vea la convención notacional (3.1).

Etapa 2. Estimación de los términos del lado derecho de (1).

Primeramente observe que del resultado del lema 3.6 se sigue inmediatamente que

$$\begin{aligned} 0 &\le P[(X_n^{\delta *}/\sigma) + D_n/\sigma \in A,\ |D_n/\sigma| \ge \delta^{1/2}] \\ &\le P[\,|D_n/\sigma| \ge \delta^{1/2}] \\ &\le \delta/\sigma^2. \qquad (2) \end{aligned}$$

Ahora, defina

$$A_{-\delta} := \{x \in \mathbb{R} : (x - \delta^{1/2},\ x + \delta^{1/2}) \subset A\},$$

$$A_\delta := \{x \in \mathbb{R} : (x - \delta^{1/2},\ x + \delta^{1/2}) \cap A \ne \emptyset\}.$$

Con esta notación, note que

$$(*)\ P[(X_n^{\delta *}/\sigma) + D_n/\sigma \in A,\ |D_n/\sigma| < \delta^{1/2}] \le P[X_n^{\delta *}/\sigma \in A_\delta].$$

(**) Note que el evento $[X_n^{\delta *}/\sigma \in A_{-\delta}]$ se puede "descomponer" como

$$[X_n^{\delta *}/\sigma \in A_{-\delta},\ |D_n/\sigma| < \delta^{1/2}] \cup [X_n^{\delta *}/\sigma \in A_{-\delta}, |D_n/\sigma| \geq \delta^{1/2}].$$

En este caso, es claro que, si $X_n^{\delta *}/\sigma \in A_{-\delta}$, entonces

$$(X_n^{\delta *}/\sigma) + D_n/\sigma \in A,\ |D_n/\sigma| < \delta^{1/2};$$

así, $[X_n^{\delta *}/\sigma \in A_{-\delta}]$ está contenido en

$$[(X_n^{\delta *}/\sigma) + D_n/\sigma \in A,\ |D_n/\sigma| < \delta^{1/2}] \cup [|D_n/\sigma| \geq \delta^{1/2}],$$

de donde se desprende que

$$P[X_n^{\delta *}/\sigma \in A_{-\delta}]$$

$$\leq P[(X_n^{\delta *}/\sigma) + D_n/\sigma \in A,\ |D_n|/\sigma < \delta^{1/2}] + \delta/\sigma^2.$$

De la discusión en (*) y (**) se sigue que

$$P[X_n^{\delta *}/\sigma \in A_{-\delta}] - \delta/\sigma^2$$

$$\leq P[(X_n^{\delta *}/\sigma) + D_n/\sigma \in A,\ |D_n|/\sigma < \delta^{1/2}]$$

$$\leq P[X_n^{\delta *}/\sigma \in A_{\delta}]. \tag{3}$$

Combinando (1)-(3) se obtiene

$$P[X_n^{\delta *}/\sigma \in A_{-\delta}] - \delta/\sigma^2 \leq P[\ X_n^{*}/\sigma \in A\ ]$$

$$\leq P[X_n^{\delta *}/\sigma \in A_{\delta}] + \delta/\sigma^2. \tag{4}$$

Defina ahora

$$A_{-\delta\sigma} := \{x \in \mathbb{R} : [\sigma(\delta)/\sigma]x \in A_{-\delta}\}\ ,$$

$$A_{\delta\sigma} = \{x \in \mathbb{R} : [\sigma(\delta)/\sigma]x \in A_{\delta}\}\ ,$$

donde

$$\sigma(\delta) := \mathrm{Var}(X^{\delta}).$$

Observe que

$$[X_n^{\delta *}/\sigma \in A_{-\delta}] = [X_n^{\delta *}/\sigma(\delta) \in A_{-\delta\sigma}],$$

$$[X_n^{\delta *}/\sigma \in A_{-\delta}] = [X_n^{\delta *}/\sigma_{\delta} \in A_{\delta\sigma}],$$

y por lo tanto, usando (4) se sigue que

$$P[X_n^{\delta *}/\sigma(\delta) \in A_{-\delta\sigma}] - \delta/\sigma^2 \leq P[\ X_n^{*}/\sigma \in A\ ]$$

$$\leq P[X_n^{\delta *}/\sigma(\delta) \in A_{-\delta\sigma}] + \delta/\sigma^2. \qquad (5)$$

Etapa 3. Derivación del resultado a partir del teorema 3.1.

Sea $\varepsilon > 0$ arbitrario.

Por la forma de los conjuntos $A_{-\delta\sigma}, A_{\delta\sigma}$[2] y el resultado (iii) del lema 3.5 se puede elegir $\delta_0 > 0$ lo suficientemente pequeño tal que

$$\left| \int_{A_{-\delta\sigma}} \varphi(z;0,1)\ dz - \int_{A} \varphi(z;0,1)\ dz \right| < \varepsilon/3\ , \qquad (6)$$

$$\left| \int_{A_{\delta\sigma}} \varphi(z;0,1)\ dz - \int_{A} \varphi(z;0,1)\ dz \right| < \varepsilon/3\ , \qquad (7)$$

para todo $\delta < \delta_0$.

Por otra parte, como la variable X^{δ} tiene soporte finito, el teorema 3.1 garantiza la existencia de $N(\varepsilon) > 0$ tal que

---

[2]Por ejemplo, observe que si A es un intervalo de la forma (a,∞), entonces $A_{-\delta} = (a+\delta^{1/2},\infty)$ y por lo tanto

$$A_{-\delta\sigma} = (\ [\sigma/\sigma(\delta)](a+\delta^{1/2}),\infty).$$

$$|P[X_n^{\delta *}/\sigma(\delta) \in A_{-\delta\sigma}] - \int_{A_{-\delta\sigma}} \varphi(z;0,1)\, dz| < \varepsilon/3 \;, \tag{8}$$

$$|P[X_n^{\delta *}/\sigma(\delta) \in A_{\delta\sigma}] - \int_{A_{\delta\sigma}} \varphi(z;0,1)\, dz| < \varepsilon/3 \;, \tag{9}$$

para todo $n > N(\varepsilon)$.

Elija ahora $\delta_1 := \text{mín}\{\delta_0, \varepsilon\sigma^2/3\}$. Entonces, a partir de (6), (8) y la desigualdad del triángulo se sigue que si $\delta < \delta_1$

$$|P[X_n^{\delta *}/\sigma(\delta) \in A_{-\delta\sigma}] - \delta/\sigma^2 - \int_A \varphi(z;0,1)\, dz \;| < \varepsilon, \tag{10}$$

para todo $n > N(\varepsilon)$.

Análogamente, a partir de (7),(9) y la definición de δ_1 se obtiene para $\delta < \delta_1$

$$|P[X_n^{\delta *}/\sigma(\delta) \in A_{\delta\sigma}] + \delta/\sigma^2 - \int_A \varphi(z;0,1)\, dz \;| < \varepsilon, \tag{11}$$

si $n > N(\varepsilon)$.

Combinando (5), (10) y (11) se sigue que para $\varepsilon > 0$ arbitrario, si $n > N(\varepsilon)$, entonces

$$\begin{aligned}
-\varepsilon & \\
&< P[X_n^{\delta *}/\sigma_\delta \in A_{-\delta\sigma}] - \int_A \varphi(z;0,1)\, dz - \delta/\sigma^2 \\
&\le P[\, X_n^{*}/\sigma \in A \,] - \int_A \varphi(z;0,1)\, dz \\
&\le P[X_n^{\delta *}/\sigma_\delta \in A_{-\delta\sigma}] - \int_A \varphi(z;0,1)\, dz + \delta/\sigma^2 \\
&< \varepsilon,
\end{aligned}$$

y por lo tanto

$$\lim_{n\to\infty} P[X_n^{*}/\sigma \in A] = \int_A \varphi_1(z;0,1)\, dz. \qquad \blacksquare$$

4 TLC PARA VARIABLES NO ACOTADAS

Supongamos ahora que X es una variable aleatoria no acotada con $E(X) = \mu$, $\mathrm{Var}(X) = \sigma^2 < \infty$. Mostaremos cómo X puede aproximarse por una variable acotada, a partir de lo cual se podrá establecer el teorema del límite central para variables no acotadas.

Observación 3.8. La existencia del segundo momento de X implica que para todo $\delta > 0$ existe $M(\delta) > 0$ tal que

$$E(X^2) - E(X^2 I_{B(\delta)}(X)) \leq \delta, \tag{3.9}$$

donde

$$B(\delta) := \{x \in \mathbb{R} : |x| \leq M(\delta)\}.$$

La condición (3.9) es equivalente a afirmar que para todo δ positivo existe $B(\delta) \subset \mathbb{R}$ acotado tal que

$$E(X^2 I_{B(\delta)^c}(X)) \leq \delta. \tag{3.10}$$

Por otra parte, para $\delta \ll 1$ tal que $M(\delta) > 1$ se tiene

(1) $P[X \in B(\delta)^c] = E(I_{B(\delta)^c}(X)) < E(X^2 I_{B(\delta)^c}(X)) \leq \delta$;

(2) $|E(X I_{B(\delta)^c}(X))| \leq E(|X| I_{B(\delta)^c}(X)] < E(X^2 I_{B(\delta)^c}(X)) \leq \delta.$

Así, si $\delta \ll 1$, $M(\delta) > 1$ la condición (3.10) implica

$$P[X \in B(\delta)^c] < \delta, \tag{3.11}$$

$$|E(X I_{B(\delta)^c}(X))| < \delta. \tag{3.12}$$

La observación 3.8 sugiere la definición de una variable acotada que aproxime a X. En efecto, usando la notación anterior, observe que si se define una nueva variable soportada únicamente en $B(\delta)$, la expresión (3.11) garantiza que la "pérdida de información" se encuentra en un conjunto de probabilidad menor a δ. Precisando, considere la siguiente definición.

Definición 3.9. Sea X una varible aleatoria no acotada , con $E(X) = \mu$, $0 < \mathrm{Var}(X) = \sigma^2 < \infty$. Para $\delta > 0$, $\delta \ll 1$, sea $M(\delta) > 0$ tal que

$$E(X^2 I_{B(\delta)^c}(X)) \leq \delta,$$

donde $B(\delta) := \{x \in \mathbb{R} : |x| \leq M(\delta)\}$.

Defina la variable aleatoria X^δ como

$$\begin{aligned} X^\delta &:= X + E(XI_{B(\delta)^c}(X))/P[X \in B(\delta)] \ , \text{ si } X \in B(\delta); \\ &:= 0 \ , \text{ si } X \in B(\delta)^c. \end{aligned} \tag{3.13}$$

Los siguientes lemas, equivalentes para este caso a los lemas 3.5 y 3.6 en la sección 3, muestran la bondad de la aproximación de X por X^δ.

Lema 3.10. Sean X y X^δ como en la Def. 3.9. Entonces

(*i*) $E[X^\delta] = E[X]$,

(*ii*) $|\sigma(\delta)^2 - \sigma^2| \leq C_1(\delta)$,

donde $\sigma(\delta)^2 := \mathrm{Var}(X^\delta)$ y $C_1(\delta)$ es cierta función tal que $C_1(\delta) \to 0$ cuando $\delta \to 0$.

Demostración. *i*) A partir de la definición de X^δ y las propiedades (3.5) y (3.6) se tiene

$$E(X^\delta) = E[(X + E(XI_{B(\delta)^c}(X))/P[X \in B(\delta)])I_{B(\delta)}(X)]$$

$$= E(XI_{B(\delta)}(X)) + E(XI_{B(\delta)^c}(X))$$

$$= E(X).$$

(*ii*) Note que

$$E[(X^\delta)^2] = E([(X + E(XI_{B(\delta)^c}(X))/P[X \in B(\delta)]]^2 I_{B(\delta)}(X)) ,$$

de donde se obtiene

$$E[(X^\delta)^2] = c_1 + c_2 + c_3, \tag{1}$$

con

$$c_1 := E(X^2 I_{B(\delta)}(X)),$$

$$c_2 := 2 \cdot E[XI_{B(\delta)^c}(X)] \cdot E[XI_{B(\delta)}(X)]/P[X \in B(\delta)],$$

$$c_3 := (E[XI_{B(\delta)^c}(X)])^2/P[X \in B(\delta)].$$

Observe ahora que

$$E(X^2) = c_1 + E(X^2 I_{B(\delta)^c}(X));$$

luego, combinando (1) con el resultado en (i) se sigue

$$\sigma(\delta)^2 - \sigma^2 = c_2 + c_3 + c_4, \tag{2}$$

donde

$$c_4 := -E(X^2 I_{B(\delta)^c}(X)).$$

Por otra parte, de la definición de X^δ, (3.11) y (3.12) se sigue que

$$|c_2| \leq 2\ \delta(|\mu| + \delta)/(1 - \delta), \tag{3}$$

$$|c_3| \le \delta^2/(1 - \delta), \tag{4}$$

$$|c_4| \le \delta. \tag{5}$$

Conjuntando (2)-(5) y utilizando la desigualdad del triángulo se sigue que

$$|\sigma(\delta)^2 - \sigma^2| \le C_1(\delta),$$

donde $C_1(\delta) := 2\,\delta(|\mu| + \delta)/(1 - \delta) + \delta^2/(1 - \delta) + \delta.$ ■

Lema 3.11. Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias no acotadas independientes e idénticamente distribuidas, con media μ, y varianza $\sigma^2 > 0$. Para $\delta > 0$, $\delta \ll 1$ defina la sucesión $X_1^\delta, X_2^\delta, \ldots$ de variables aleatorias independientes con la misma distribución que X^δ, usando la definición 3.9. Entonces

$$P[|\,\overline{X}_n - \overline{X_n^\delta}|n^{1/2}/\sigma \ge \delta^{1/4}\,] \le C_2(\delta),$$

donde $C_2(\delta)$ es cierta función tal que $C_2(\delta) \to 0$ cuando δ tiende a cero.

Demostración. Observe que las variables $X_i - X_i^\delta$ son independientes y tienen la misma distribución, para cada $i = 1,2,\ldots$. Así, a partir de la desigualdad de Chebyshev y el resultado (i) del lema 3.10 se sigue que

$$\begin{aligned} P[|\,\overline{X}_n - \overline{X_n^\delta}|n^{1/2} \ge \sigma\delta^{1/4}\,] &\le \mathrm{Var}[n^{1/2}(\,\overline{X}_n - \overline{X_n^\delta})]/(\sigma^2\delta^{1/2}) \\ &= \mathrm{Var}(X - X^\delta)/(\sigma^2\delta^{1/2}) \\ &= E[(X - X^\delta)^2]/(\sigma^2\delta^{1/2}). \qquad (1) \end{aligned}$$

Ahora, observe que

$$X - X^{\delta} := - E[XI_{B(\delta)^c}(X)]/P[X \in B(\delta)] \ , \ \text{si } X \in B(\delta),$$

$$:= X \ , \ \text{si } X \in B(\delta)^c,$$

de donde se obtiene

$$E[(X - X^{\delta})^2] = (E[XI_{B(\delta)^c}(X)])^2/P[X \in B(\delta)]+E[X^2I_{B(\delta)^c}(X)]$$

$$\leq \delta^2/(1-\delta) + \delta. \qquad (2)$$

Conjuntando (1) y (2) se sigue el resultado, donde

$$C_2(\delta) := [\delta^{1/2} + \delta^{3/2}/(1-\delta)]/\sigma^2. \qquad \blacksquare$$

Teorema 3.12. Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias *no acotadas*, independientes e idénticamente distribuidas, con media μ y varianza $\sigma^2 > 0$. Entonces, para todo intervalo $A \subset \mathbb{R}$,

$$\lim_{n\to\infty} P[X_n^*/\sigma \in A] = \int_A \varphi_1(z;0,1)\ dz.$$

Demostración. De acuerdo a la discusión realizada entre la observación 3.8 y el lema 3.11, no es difícil ver que la demostración de este teorema puede verificarse en forma análoga a la del teorema 3.7. En efecto, considere la siguiente argumentación :

Para alguna $\delta > 0$, $\delta \ll 1$ se construye la sucesión $\{X_i^{\delta}\}_{i=1}^{\infty}$ de variables acotadas independientes e idénticamente distribuidas que X^{δ}, utilizando la definición 3.9. Defina

$$D_n := |\overline{X}_n - \overline{X_n^{\delta}}| n^{1/2},$$

y observe entonces que, para todo intervalo $A \subset \mathbb{R}$,

$$P[\, X_n^*/\sigma \in A \,] = P[(X_n^{\delta *}/\sigma) + D_n/\sigma \in A]$$

$$= P[(X_n^{\delta *}/\sigma) + D_n/\sigma \in A,\ |D_n/\sigma| \geq \delta^{1/4}]$$

$$+ P[(X_n^{\delta *}/\sigma) + D_n/\sigma \in A,\ |D_n/\sigma| < \delta^{1/4}]. \quad (1)$$

A partir de (1), los resultados de los lemas 3.10, 3.11 y el Teorema 3.7 permiten realizar la analogía con las etapas 2 y 3 de la demostración de ese teorema. ■

5 TEOREMA DEL LIMITE CENTRAL UNIDIMENSIONAL

Para concluir este capítulo, y como una forma de resumir la discusión hasta aquí realizada, a continuación se enuncia la versión general del TLC para variables aleatorias con segundo momento finito.

Teorema del Límite Central. Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias independientes e idénticamente distribuidas, con media μ y varianza $\sigma^2 > 0$. Entonces

la distribución asintótica de X_n^*/σ *es* $\mathcal{N}(0,1)$,

esto es, para todo intervalo $A \subset \mathbb{R}$,

$$\lim_{n\to\infty} P[X_n^*/\sigma \in A] = \int_A \varphi_1(z;0,1)\, dz.$$

IV | TEOREMA DEL LIMITE CENTRAL MULTIDIMENSIONAL

1 INTRODUCCION

En la sección 4 del capítulo II se estableció un teorema del límite central para vectores aleatorios con soporte finito (Teorema 2.15), el cual sirvió de punto de partida para la discusión del TLC unidimensional. En el presente capítulo se extienden las ideas del capítulo III al caso multivariado, estableciéndose una versión general del teorema del límite central para vectores aleatorios independientes con matriz de covarianza.

El material se ha dividido en forma tal de considerar las dos posibilidades restantes de investigar : en la sección 2 se aborda el caso de vectores aleatorios con soporte acotado y la sección 3 se ocupa de vectores no acotados.

Notación. Sea $\mathbf{X}$ un vector aleatorio k-dimensional, para el cual existen $\mu := E(\mathbf{X})$, $M := \mathrm{COV}(\mathbf{X})$. Si $\mathbf{X}_1, \mathbf{X}_2, \ldots$ es una sucesión de vectores independientes con la misma distribución de $\mathbf{X}$, defina

$$\mathbf{X}_n^* := \left(\sum_{i=1}^{n} \mathbf{X}_i - n\mu\right)/n^{1/2}. \tag{4.1}$$

2 VECTORES ALEATORIOS ACOTADOS

Definición 4.1. Un conjunto $\mathcal{S} \subset \mathbb{R}^k$ es un *rectangulo* si $\mathcal{S}$ es de la forma

$$A_1 \times A_2 \times \ldots \times A_k,$$

donde A_i es un intervalo acotado en $\mathbb{R}$, $i = 1,\ldots,k$. Esto es, $\mathcal{S}$ es un rectángulo en $\mathbb{R}^k$ si

$$\mathcal{S} = \{\mathbf{x} = (x_1,\ldots,x_k)': x_i \in A_i,\ A_i \subset \mathbb{R} \text{ un intervalo acotado}\}.$$

Como ejemplos de la definición anterior considere los siguientes conjuntos.

(1) $\mathcal{S} := [0,5) \subset \mathbb{R}$.

(2) $\mathcal{S} := \{\mathbf{x} = (x_1,x_2)': -1 < x_1 < 1,\ -1/2 < x_2 < 1\} \subset \mathbb{R}^2$.

(3) $\mathcal{S} := \{\mathbf{x} = (x_1,x_2,x_3)': (-1/2) \le x_1,x_2,x_3 \le 1/2\} \subset \mathbb{R}^3$.

(4) $\mathcal{S} := [a_1,b_1) \times [a_2,b_2) \times \ldots \times [a_k,b_k)$.

Por lo que respecta al caso en que $k = 1$, observe que de la definición 4.1 se sigue que los rectángulos en $\mathbb{R}$ son los intervalos de la forma $[a,b],(a,b),(a,b]$ o $[a,b)$, como es el caso del ejemplo (1). En (2) note que $\mathcal{S}$ es el rectángulo en el plano comprendido entre las rectas $x_1 = -1$, $x_1 = 1$, $x_2 = -1/2$, $x_2 = 1$; $\mathcal{S}$ en (3) es el cubo con centro en el origen y lado 1, incluyendo sus "caras".

Definición. 4.2. Sean $\mathbf{x}_0 \in \mathbb{R}^k$ un punto arbitrario y r un número positivo. *La vecindad abierta de* $\mathbf{x}$ *con radio* r se denota por $V(\mathbf{x},r)$ y se define como

$$\mathcal{V}(\mathbf{x},r) := \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}_0\| < r\}.$$

En particular, para todo $r > 0$,

$$\mathcal{V}(\mathbf{0},r) = \{\mathbf{x} : \|\mathbf{x}\| < r\}.$$

Un conjunto $\mathcal{A} \subset \mathbb{R}^k$ es *acotado* si existe $R > 0$ tal que $\mathcal{A} \subset \mathcal{V}(\mathbf{0},R)$.

En este caso, existe un rectángulo $\mathcal{S} \subset \mathbb{R}^k$,

$$\mathcal{S} = [a_1,b_1) \times [a_2,b_2) \times \ldots \times [a_k,b_k), \tag{4.2}$$

tal que $\mathcal{A} \subset \mathcal{S}$.

Definición 4.3. Sea $\mathcal{A} \subset \mathbb{R}^k$ un conjunto arbitrario. La *función indicadora* de $\mathcal{A}$ se denota por $I_{\mathcal{A}}: \mathbb{R}^k \to \mathbb{R}$ y se define por

$$\begin{aligned} I_{\mathcal{A}}(\mathbf{x}) &:= 1 \text{ si } \mathbf{x} \in \mathcal{A}, \\ &:= 0 \text{ se } \mathbf{x} \notin \mathcal{A}. \end{aligned}$$

Observación 4.4. Sea $\mathbf{X}$ un vector aleatorio de dimensión k y $\mathcal{A} \subset \mathbb{R}^k$ un rectángulo arbitrario. Entonces

(i) $P[I_{\mathcal{A}}(\mathbf{X}) = 1] = P[\mathbf{X} \in \mathcal{A}]$ y $P[I_{\mathcal{A}}(\mathbf{X}) = 0] = P[\mathcal{A}^c]$.

Luego,

$$E(I_{\mathcal{A}}(\mathbf{X})) = P[\mathbf{X} \in \mathcal{A}]. \tag{4.3}$$

(ii) Suponga que $\mathcal{A}_1,\ldots,\mathcal{A}_r \subset \mathbb{R}^k$ son conjuntos disjuntos y que $\bigcup_{i=1}^{r} \mathcal{A}_i \supset \mathcal{X} \equiv$ soporte de $\mathbf{X}$. Si $G : \mathbb{R}^k \to \mathbb{R}$ es una función arbitraria de $\mathbf{X}$ con esperanza finita, entonces

$$E(G(\mathbf{X})) = \sum_{i=1}^{r} E[G(\mathbf{X})I_{\mathcal{A}_r}(\mathbf{X})]. \tag{4.4}$$

Sea **X** un vector aleatorio k-dimensional con soporte $\mathcal{X} \subset \mathbb{R}^k$ acotado y tal que existe M := COV(**X**). Análogamente a como se hizo en el caso unidimensional, se definirá un vector con soporte finito que aproxime el comportamiento de **X**; la idea consiste en dividir el soporte de **X** en "partes muy pequeñas" y elegir dentro de ellas un "buen representante". Precisando, tenemos la siguiente definición.

Definición 4.5. Sean $\mathbf{X} := (X_1,\ldots,X_k)'$ un vector aleatorio con soporte $\mathcal{X} \subset \mathbb{R}^k$ acotado, $\mu := E(\mathbf{X})$, $M := \mathrm{COV}(\mathbf{X})$. Sea $\mathcal{S} \subset \mathbb{R}^k$ un rectángulo de la forma (4.2) tal que $\mathcal{X} \subset \mathcal{S}$. Si δ es un número positivo, $\delta < \min\{ b_i - a_i : i = 1,\ldots,k\}$, considere una partición de $\mathcal{S}$ en rectángulos $\mathcal{S}_1,\mathcal{S}_2,\ldots,\mathcal{S}_r$ tales que

(i) $\mathcal{S}_j := A_{1j}\times A_{2j}\times\ldots\times A_{kj}$, $A_{ij} := [a_j^i, b_j^i)$, $i = 1,\ldots,k;$
$j = 1,\ldots,r.$

(ii) $b_j^i - a_j^i = \delta$, $i = 1,\ldots,k;$ $j = 1,\ldots,r.$

(iii) $\bigcup_{j=1}^{r} \mathcal{S}_j = \mathcal{S}.$

Defina los vectores $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_r$ mediante

$$\begin{aligned} \mathbf{x}_j &:= E(I_{\mathcal{S}_j}(\mathbf{X})\cdot\mathbf{X})/P[\mathbf{X} \in \mathcal{S}_j] \text{ si } P[\mathbf{X} \in \mathcal{S}_j] > 0, \\ &:= \mathbf{0} \text{ de otro modo.} \end{aligned} \tag{4.5}$$

Finalmente, defina el vector aleatorio $\mathbf{X}^\delta$ de dimensión k mediante la expresión

$$\mathbf{X}^{\delta} := \sum_{j=1}^{r}[I_{\mathscr{S}_j}(\mathbf{X})\mathbf{x}_j], \tag{4.6}$$

esto es

$$\mathbf{X}^{\delta} = \mathbf{x}_j \text{ si } \mathbf{X} \in \mathscr{S}_j.$$

Llamaremos a $\mathbf{X}^{\delta}$ la *δ-discretización de* **X**.

A partir de la definición anterior es claro que, si $\mathbf{x}_j \neq \mathbf{0}$,

$$P[\mathbf{X}^{\delta} = \mathbf{x}_j] = P[\mathbf{X} \in \mathscr{S}_j].$$

Mostraremos ahora que, en este caso, $\mathbf{x}_j \in \mathscr{S}_j$. En este sentido, podemos decir que $\mathbf{x}_j$ es un buen representante de **X** en $\mathscr{S}_j$.

Lema 4.6. Si $P[\mathbf{X} \in \mathscr{S}_j] > 0$, entonces $\mathbf{x}_j$ definida en (4.6) es tal que

$$\mathbf{x}_j \in \mathscr{S}_j.$$

Demostración. Sea

$$\mathscr{S}_j := A_{1j} \times A_{2j} \times \ldots \times A_{kj},$$

donde $A_{ij} := [a_i^{\,j}, b_i^{\,j})$, $i = 1,\ldots,k$.

Si $P[\mathbf{X} \in \mathscr{S}_j] > 0$, de la definición (4.6) se sigue que

$$x_j^{\,i} := E(X_i I_{\mathscr{S}_j}(\mathbf{X}))/P[\mathbf{X} \in \mathscr{S}_j], \; i = 1,\ldots,k.$$

Utilizando (4.3) se tiene

$$E[I_{\mathscr{S}_j}(\mathbf{X})]x_j^{\,i} = E[X_i I_{\mathscr{S}_j}(\mathbf{X})] \;, \; i = 1,\ldots,k,$$

de donde se deprende que, para toda $i = 1,\dots,k$

$$E[(X_i - x_j^{\,i}) I_{\mathscr{S}_j}(\mathbf{X})] = E[(x_j^{\,i} - X_i) I_{\mathscr{S}_j}(\mathbf{X})] = 0. \tag{2}$$

Como se ha supuesto que $E[I_{\mathscr{S}_j}(\mathbf{X})] > 0$, las igualdades (2) implican (Rudin, 1977) que no es posible tener $x_j^{\,i} - X_i > 0$ ni $x_j^{\,i} - X_i < 0$ en todo $\mathscr{S}_j$, de donde se sigue que $x_j^{\,i} \in A_{ij}$. Por lo tanto

$$\mathbf{x}_j \in \mathscr{S}_j. \qquad \blacksquare$$

Observación 4.7. Por la construcción de los subrectángulos $\mathscr{S}_j$, del lema 4.6 se desprenden las siguientes afirmaciones

(i) $P[|X_i - X_i^{\delta}| \le \delta] = 1,\ i = 1,2,\dots,k.$

(ii) $P[\|\mathbf{X} - \mathbf{X}^{\delta}\| \le k^{1/2}\delta] = 1.$

A continuación, los lemas 4.8 y 4.10 establecen la relación existente entre los momentos de $\mathbf{X}$ y los de la discretización $\mathbf{X}^{\delta}$.

Lema 4.8. Sean $\mathbf{X}$ y $\mathbf{X}^{\delta}$ vectores aleatorios como en la definición 4.5. Entonces

$$E(\mathbf{X}^{\delta}) = E(\mathbf{X}).$$

Demostración. Primeramente, observe que si $P[\mathbf{X} \in \mathscr{S}_j] = E(I_{\mathscr{S}_j}(\mathbf{X})) = 0$, entonces $E(I_{\mathscr{S}_j}(\mathbf{X})\cdot\mathbf{X}) = \mathbf{0}$.

Con esta consideración, (4.3)-(4.6) implican

$$E(\mathbf{X}^{\delta}) = \sum_{j=1}^{r} \mathbf{x}_j E(I_{\mathcal{S}_j}(\mathbf{X}))$$

$$= \sum_{j=1}^{r} E(I_{\mathcal{S}_j}(\mathbf{X})\cdot\mathbf{X})$$

$$= E(\mathbf{X}). \qquad \blacksquare$$

Recordemos ahora la definición 1.5, cap. I, según la cual,para dos vectores **Y**,**Z** definidos en un mismo espacio, la matriz de covarianza entre **Y** y **Z** está dada por

$$\mathrm{COV}(\mathbf{Y},\mathbf{Z}) = E[(\mathbf{Y} - E(\mathbf{Y}))(\mathbf{Z} - E(\mathbf{Z})'],$$

siempre y cuando las esperanzas del lado derecho existan. Un cálculo algebraico directo a partir de esta definición proporciona el siguiente resultado, el cual será utilizado en el lema 4.10.

Lema 4.9. Sean **Y**,**Z** vectores aleatorios de la misma dimensión y definidos en el mismo espacio. Si COV(**Y**),COV(**Z**) existen, entonces COV(**Y** - **Z**) también existe y

$$\mathrm{COV}(\mathbf{Y} - \mathbf{Z}) = \mathrm{COV}(\mathbf{Y}) + \mathrm{COV}(\mathbf{Z}) - \mathrm{COV}(\mathbf{Y},\mathbf{Z}) - \mathrm{COV}(\mathbf{Z},\mathbf{Y});$$

recuerde que COV(**Y**) = COV(**Y**,**Y**).

Lema 4.10. Sea **X** un vector aleatorio acotado y $\mathbf{X}^{\delta}$ la δ-discretización de **X** en (4.6). Entonces

$$\mathrm{COV}(\mathbf{X}^{\delta}) = \mathrm{COV}(\mathbf{X}) + A,$$

donde $A = [a_{ii'}]$ es una matriz de orden $k \times k$ tal que

$$|a_{ii'}| \le \delta^2, \text{ para todo } i,i' = 1,2,\ldots,k.$$

Demostración. Utilizando el lema 4.9 se tiene

$$\mathrm{cov}(\mathbf{X} - \mathbf{X}^\delta) = \mathrm{cov}(\mathbf{X}) + \mathrm{cov}(\mathbf{X}^\delta) - \mathrm{cov}(\mathbf{X},\mathbf{X}^\delta) - \mathrm{cov}(\mathbf{X}^\delta,\mathbf{X}). \quad (1)$$

i) Mostraremos primeramente que

$$\mathrm{cov}(\mathbf{X}^\delta) = \mathrm{cov}(\mathbf{X},\mathbf{X}^\delta) = \mathrm{cov}(\mathbf{X}^\delta,\mathbf{X}). \qquad (**)$$

Para todo $i,i' = 1,\ldots,k$ se tiene

$$\begin{aligned} E(X_i^\delta X_{i'}^\delta) &= \sum_{j=1}^r E(X_i^\delta X_{i'}^\delta \cdot I_{\mathscr{S}_j}(\mathbf{X})) \\ &= \sum_{j=1}^r x_j^i x_j^{i'} E(I_{\mathscr{S}_j}(\mathbf{X})) \\ &= \Sigma\{E(X_i I_{\mathscr{S}_j}(\mathbf{X}))\cdot E(X_{i'} I_{\mathscr{S}_j}(\mathbf{X}))/P[\mathbf{X} \in \mathscr{S}_j]\} \ , \end{aligned}$$

donde la última sumatoria se extiende sólo para los j tales que $P[\mathbf{X} \in \mathscr{S}_j] > 0$. Con esta notación observe entonces que

$$\begin{aligned} E(X_i^\delta X_{i'}^\delta) &= \Sigma\{E(X_i I_{\mathscr{S}_j}(\mathbf{X}))\cdot x_j^{i'}\} \\ &= \Sigma\{E(x_j^{i'} X_i I_{\mathscr{S}_j}(\mathbf{X}))\} \\ &= \sum_{j=1}^r E(X_{i'}^\delta X_i I_{\mathscr{S}_j}(\mathbf{X})) \\ &= E(X_{i'}^\delta X_i), \end{aligned} \qquad (1)$$

y procediendo en forma similar se obtiene

$$E(X_i^\delta X_{i'}^\delta) = E(X_i^\delta X_{i'}). \qquad (2)$$

Conjuntando (1),(2) y el resultado del lema 4.8 se obtiene

$$\mathrm{cov}(X_i^\delta, X_{i'}^\delta) = \mathrm{cov}(X_i, X_{i'}^\delta) = \mathrm{cov}(X_i^\delta, X_{i'}),$$

para todo $i,i' = 1,\ldots,k$, y, por lo tanto, (**) se satisface.

ii) Combinando (*) y (**) se tiene

$$\mathrm{COV}(\mathbf{X}^{\delta}) = \mathrm{COV}(\mathbf{X}) - \mathrm{COV}(\mathbf{X} - \mathbf{X}^{\delta}).$$

De esta manera, definiendo

$$A \equiv [a_{ii'}] := - \mathrm{COV}(\mathbf{X} - \mathbf{X}^{\delta}),$$

restaría demostrar que $|a_{ii'}| \le \delta^2$. Pero

$$\begin{aligned} |a_{ii'}| &= |\mathrm{cov}(X_i - X_i^{\delta}, X_{i'} - X_{i'}^{\delta})| \\ &= |E[(X_i - X_i^{\delta})(X_{i'} - X_{i'}^{\delta})]| \\ &\le E[|X_i - X_i^{\delta}||X_{i'} - X_{i'}^{\delta}|] \\ &\le \delta^2, \end{aligned}$$

donde la última desigualdad se sigue de (i) en la observación 4.7. ■

Si $\mathbf{X}_1, \mathbf{X}_2, \ldots$ es una sucesión de vectores aleatorios acotados independientes e idénticamente distribuidos, recuerde que nuestro interés radica en obtener la distribución asintótica del vector aleatorio $\mathbf{X}_n^*$; vea la convención notacional en la sección 1. El siguiente lema garantiza la proximidad -en términos de probabilidad- de $\mathbf{X}_n^*$ y $\mathbf{X}^{\delta}{}_n^*$, cuando se elige adecuadamente a δ. Este resultado conduce a la formulación del TLC para vectores acotados en el teorema 4.12, en la medida en que permite extender a $\mathbf{X}_n^*$ la afirmación que se hace sobre la distribución asintótica de $\mathbf{X}^{\delta}{}_n^*$ en el teorema 2.13.

Lema 4.11. Sea $\mathbf{X}_1, \mathbf{X}_2, \ldots$ una sucesión de vectores aleatorios de dimensión k acotados, independientes e idénticamente distribuidos, con media μ y matriz de covarianza M. Para $\delta > 0$, $\delta \ll 1/k$, sea $\mathbf{X}_i^{\delta}$ la δ-discretrización de $\mathbf{X}_i$. Entonces, para todo $n > 0$

$$P[\|\mathbf{X}_n^* - \mathbf{X}^{\delta}{}_n^*\| \geq \delta^{1/2}] \leq k\delta.$$

Demostración. Defina

$$\mathbf{Z}_n := \mathbf{X}_n^* - \mathbf{X}^{\delta}{}_n^*$$

$$= [\sum_{i=1}^{n} (\mathbf{X}_i - \mathbf{X}_i^{\delta})]/n^{1/2}$$

y observe que el lema 4.8 y la independencia de los vectores $\mathbf{X} - \mathbf{X}^{\delta}$ implican

$$E(\mathbf{Z}_n) = \mathbf{0}, \tag{1}$$

$$\mathrm{COV}(\mathbf{Z}_n) = \mathrm{COV}(\mathbf{X} - \mathbf{X}^{\delta}). \tag{2}$$

De esta manera, utilizando la desigualdad de Chebyshev para vectores aleatorios (ver A3), se tiene

$$P[\|\mathbf{Z}_n\| \geq \delta^{1/2}] \leq \mathrm{tr}[\mathrm{COV}(\mathbf{X} - \mathbf{X}^{\delta})]/\delta. \tag{3}$$

Por otra parte, utilizando (i) de la observación 4.7 se sigue que

$$\mathrm{tr}[\mathrm{COV}(\mathbf{X} - \mathbf{X}^{\delta})] = \sum_{i=1}^{k} \mathrm{Var}(X_i - X^{\delta}{}_i)$$

$$= \sum_{m=1}^{k} E[(X_i - X^{\delta}{}_i)^2]$$

$$\leq k\delta^2. \tag{4}$$

Combinando (3),(4) y la definición de $\mathbf{Z}_n$ se obtiene el resultado. ■

Teorema 4.12. Sea **X** un vector aleatorio *acotado* de dimensión k, con esperanza $\mu \in \mathbb{E}_k$ y matriz de covarianza M. Si $\mathbf{X}_1, \mathbf{X}_2, \ldots$ es una sucesión de vectores aleatorios independientes con la misma distribución de **X**, entonces *la distribución asintótica de* $\mathbf{X}_n^*$ es $\mathcal{N}_k(\mathbf{0}, M)$; *esto es, si* $\mathcal{A}$ es una región (conjunto Jordán-medible) en $\mathbb{E}_k$, entonces

$$\lim_{n\to\infty} P[\mathbf{X}_n^* \in \mathcal{A}] = P[\mathbf{W} \in \mathcal{A}],$$

donde $\mathbf{W} \simeq \mathcal{N}_k(\mathbf{0}, M)$.

Demostración. Por claridad, dividiremos la demostración en tres etapas.

Etapa 1. Para $\delta > 0$, $\delta \ll 1$ sea $\mathbf{X}^\delta$ la δ-discretización de **X**. Sea $\mathbf{X}_1^\delta, \mathbf{X}_2^\delta, \ldots$ una sucesión de variables aleatorias independientes y con la misma distribución que $\mathbf{X}^\delta$.
Defina

$$\mathbf{Z}_n := (\mathbf{X}_n^* - \mathbf{X}^{\delta}{}_n^*),$$

y note que para cualquier región $\mathcal{A} \subset \mathbb{E}_k$ se tiene

$$\begin{aligned} P[\, \mathbf{X}_n^* \in \mathcal{A}\,] &= P[(\mathbf{X}_n^{\delta *} + \mathbf{Z}_n) \in \mathcal{A}] \\ &= P[(\mathbf{X}_n^{\delta *} + \mathbf{Z}_n) \in \mathcal{A},\ \|\mathbf{Z}_n\| \geq \delta^{1/2}] \\ &\quad + P[(\mathbf{X}_n^{\delta *} + \mathbf{Z}_n) \in \mathcal{A},\ \|\mathbf{Z}_n\| < \delta^{1/2}]; \qquad (1) \end{aligned}$$

Etapa 2. Empleando el resultado del lema 4.11 se tiene

$$0 \le P[(X_n^{\delta *} + Z_n) \in \mathscr{A},\ \|Z_n\| \ge \delta^{1/2}]$$

$$\le P[\|Z_n\| \ge \delta^{1/2}]$$

$$\le k\delta \tag{2}$$

Ahora, defina

$$\mathscr{A}_{-\delta} := \{\mathbf{x} : V(\mathbf{x}, \delta^{1/2}) \subset \mathscr{A}\},$$

$$\mathscr{A}_{\delta} := \{\mathbf{x} : V(\mathbf{x}, \delta^{1/2}) \cap \mathscr{A} \neq \emptyset\};$$

con esta notación, considere las siguientes observaciones :

(*) $P[(X_n^{\delta *} + Z_n) \in \mathscr{A},\ \|Z_n\| < \delta^{1/2}] \le P[X_n^{\delta *} \in \mathscr{A}_{\delta}]$.

(**) Note que

$$[X_n^{\delta *} \in \mathscr{A}_{-\delta}] = [X_n^{\delta *} \in \mathscr{A}_{-\delta},\ \|Z_n\| < \delta^{1/2}]$$

$$\cup\ [X_n^{\delta *} \in \mathscr{A}_{-\delta},\ \|Z_n\| \ge \delta^{1/2}].$$

En este caso, es claro que si $X_n^{\delta *} \in \mathscr{A}_{-\delta}$ y Z_n es tal que $\|Z_n\| < \delta^{1/2}$, entonces $X_n^{\delta *} + Z_n \in \mathscr{A}$. Así,

$$[X_n^{\delta *} \in \mathscr{A}_{-\delta}] \subset [X_n^{\delta *} + Z_n \in \mathscr{A},\ \|Z_n\| < \delta^{1/2}] \cup [\|Z_n\| \ge \delta^{1/2}].$$

Utilizando (2) se desprende entonces que

$$P[X_n^{\delta *} \in \mathscr{A}_{-\delta}] \le P[X_n^{\delta *} + Z_n \in \mathscr{A},\ \|Z_n\| < \delta^{1/2}] + k\delta.$$

Resumiendo la discusión en (*) y (**) se tiene :

$$P[X_n^{\delta *} \in \mathscr{A}_{-\delta}] - k\delta \le P[X_n^{\delta *} + Z_n \in \mathscr{A},\ \|Z_n\| < \delta^{1/2}]$$

$$\le P[X_n^{\delta *} \in \mathscr{A}_{\delta}]. \tag{3}$$

A partir de (1)-(3) se obtiene

$$P[X_n^{\delta *} \in \mathcal{A}_{-\delta}] - k\delta \le P[\ X_n^{\,*} \in \mathcal{A}\ ] \le P[X_n^{\delta *} \in \mathcal{A}_{\delta}] + k\delta. \qquad (4)$$

Etapa 3. Sea $\varepsilon > 0$ arbitrario.

Por la forma de los conjuntos $\mathcal{A}_{-\delta}, \mathcal{A}_{\delta}{}^{1}$, se puede elegir $\delta_0 > 0$ lo suficientemente pequeño tal que

$$|P[X_n^{\delta *} \in \mathcal{A}_{-\delta}] - P[X_n^{\delta *} \in \mathcal{A}]| < \varepsilon/4\ , \qquad (5)$$

$$|P[X_n^{\delta *} \in \mathcal{A}_{\delta}] - P[X_n^{\delta *} \in \mathcal{A}]| < \varepsilon/4\ , \qquad (6)$$

para todo $\delta < \delta_0$.

Por otra parte, como $\mathbf{X}^{\delta}$ tiene soporte finito, del teorema 2.15 se desprende que existe $N \equiv N(\varepsilon) > 0$ tal que, para todo $n > N$,

$$|P[X_n^{\delta *} \in \mathcal{A}] - P[\mathbf{W}^{\delta} \in \mathcal{A}]| < \varepsilon/4\ , \qquad (7)$$

donde $\mathbf{W}^{\delta} \simeq \mathcal{N}_k(\mathbf{0}, \mathrm{COV}(\mathbf{X}^{\delta}))$.

Recuerde ahora que, para todo $\delta > 0$,

$$\mathrm{COV}(\mathbf{X}^{\delta}) = M + A,$$

donde A es una matriz cuyos elementos son, en valor absoluto, menores a δ^2; ver lema 4.10. De esta manera, se puede encontrar δ_1 positiva tal que, para todo $\mathcal{A} \subset \mathbb{E}_k$,

$$|P[\mathbf{W}^{\delta} \in \mathcal{A}] - P[\mathbf{W} \in \mathcal{A}]| < \varepsilon/4,\ \mathbf{W} \simeq \mathcal{N}_k(\mathbf{0}, M), \qquad (8)$$

para todo $\delta < \delta_1$.

A partir de (5)-(8) y la desigualdad del triángulo se desprende que, para $\delta < \mathrm{mín}\{\delta_0, \delta_1, \varepsilon/4k\}$,

$$|P[X_n^{\delta *} \in \mathcal{A}_{-\delta}] - k\delta - P[\mathbf{W} \in \mathcal{A}]| < \varepsilon, \qquad (9)$$

$$|P[X_n^{\delta *} \in \mathcal{A}_\delta] + k\delta - P[W \in \mathcal{A}]| < \varepsilon, \tag{10}$$

para todo n > N, $W \simeq \mathcal{N}_k(0,M)$.

Por lo tanto, combinando (4),(9) y (10) se sigue que

$$\lim_{n\to\infty} P[X_n^{\,*} \in \mathcal{A}] = P[W \in \mathcal{A}].$$ ■

3 VECTORES ALEATORIOS NO ACOTADOS

Supongamos ahora que $\mathbf{X} = (X_1,\ldots,X_k)'$ es un vector aleatorio con soporte $\mathcal{X} \subset \mathbb{E}_k$ no acotado, $\mu := E(\mathbf{X})$, M := COV($\mathbf{X}$). Análogamente al caso unidimensional, se buscará la forma de "truncar" el soporte de $\mathbf{X}$ en forma conveniente para extender a este caso el resultado del teorema 4.12.

Primeramente, observe que la existencia de la matriz de covarianza M significa que

$$|\mathrm{cov}(X_i,X_j)| < \infty \text{ , para todo } i,j = 1,\ldots,k.$$

De esta manera, para todo $\delta > 0$ existe $\mathcal{B}(\delta) \subset \mathbb{E}_k$ *acotado* tal que

$$|E(X_iX_jI_{\mathcal{B}(\delta)^c}(\mathbf{X}))| \le \delta, \tag{4.7}$$

para todo i,j = 1,2,...,k.

Por otra parte, si $\delta \ll 1$ es de esperar que $\mathcal{B}(\delta)$ sea tal que, para todo $\mathbf{X} \in \mathcal{B}(\delta)^c$, se tenga $|X_i| > 1$, i = 1,..,k. Si éste es el caso, note que :

(i) $P[\mathbf{X} \in \mathcal{B}(\delta)^c] = E(I_{\mathcal{B}(\delta)^c}(\mathbf{X})) < E(X_i^{\,2}I_{\mathcal{B}(\delta)^c}(\mathbf{X}) \le \delta$;

(ii) $|E(X_i I_{\mathcal{B}(\delta)^c}(\mathbf{X}))| \le E(|X_i| I_{\mathcal{B}(\delta)^c}(\mathbf{X})] < E(X_i^2 I_{\mathcal{B}(\delta)^c}(\mathbf{X}))$

$\le \delta.$

Así, si $\delta \ll 1$ se elige adecuadamente la condición (4.7) implica

$$P[\mathbf{X} \in \mathcal{B}(\delta)^c] < \delta, \tag{4.8}$$

$$|E(X_i I_{\mathcal{B}(\delta)^c}(\mathbf{X}))| < \delta. \tag{4.9}$$

A partir de esta discusión, se define el vector aleatorio $\mathbf{X}^\delta$ con soporte acotado que aproxime a $\mathbf{X}$ de la siguiente manera.

Definicioń 4.13. Sea $\mathbf{X} = (X_1,\ldots,X_k)'$ un vector aleatorio no acotado , con esperanza μ, y matriz de covarianza M. Para $\delta > 0$, $\delta \ll 1$, sea $\mathcal{B}(\delta)$ un subconjunto acotado de $\mathbb{E}_k$ tal que

$$|E(X_i X_j I_{\mathcal{B}(\delta)^c}(\mathbf{X}))| \le \delta, \text{ para toda } i = 1,2,\ldots,k.$$

Defina entonces el vector aleatorio $\mathbf{X}^\delta$ como

$$\begin{aligned} \mathbf{X}^\delta &:= \mathbf{X} + E(\mathbf{X} I_{\mathcal{B}(\delta)^c}(\mathbf{X}))/P[\mathbf{X} \in \mathcal{B}(\delta)] \text{ , si } \mathbf{X} \in \mathcal{B}(\delta); \\ &:= \mathbf{0} \text{ , si } \mathbf{X} \in \mathcal{B}(\delta)^c. \end{aligned} \tag{4.10}$$

Los lemas 4.14-4.16 que se muestran a continuación constituyen los resultados equivalentes -para el caso que nos ocupa- a los lemas 4.8,4.10,4.11 de la sección 2.

Lema 4.14. Sean $\mathbf{X}$ y $\mathbf{X}^\delta$ vectores aleatorios como en la definición 4.13. Entonces

$$E(\mathbf{X}^\delta) = E(\mathbf{X}).$$

Demostración. Usando (4.10) se tiene que

$$E(\mathbf{X}^\delta) = E(\mathbf{X}^\delta I_{\mathcal{B}(\delta)}(\mathbf{X}))$$

$$= E[(\mathbf{X} + E(I_{\mathcal{B}(\delta)^c}(\mathbf{X})\cdot\mathbf{X})/P[\mathbf{X} \in \mathcal{B}(\delta)])I_{\mathcal{B}(\delta)}(\mathbf{X})]$$

$$= E(\mathbf{X}).$$ ■

Lema 4.15. Sea $\mathbf{X}$ un vector aleatorio no acotado y $\mathbf{X}^\delta$ como en (4.10). Entonces

$$\mathrm{COV}(\mathbf{X}^\delta) = \mathrm{COV}(\mathbf{X}) + A,$$

donde $A = [a_{ij}]$ es una matriz de orden $k \times k$ tal que, para todo $i,j = 1,2,\ldots,k$,

$$|a_{ij}| \le C_{ij}(\delta),$$

$C_{ij}(\delta) \to 0$ conforme $\delta \to 0$.

Demostración. Defina $p := P[\mathbf{X} \in \mathcal{B}(\delta)]$ y observe que para $i,j = 1,\ldots,k$ se tiene

$$E(X_i^\delta X_j^\delta) = E(X_i^\delta X_j^\delta I_{\mathcal{B}(\delta)}(\mathbf{X}))$$

$$= E[(X_i + E(X_i I_{\mathcal{B}(\delta)^c}(\mathbf{X}))/p)(X_j + E(X_j I_{\mathcal{B}(\delta)^c}(\mathbf{X}))/p)I_{\mathcal{B}(\delta)}(\mathbf{X})]$$

$$= E(X_i X_j I_{\mathcal{B}(\delta)}(\mathbf{X})) + c_1 + c_2 + c_3, \tag{1}$$

donde

$$c_1 := E[X_i I_{\mathcal{B}(\delta)}(\mathbf{X})]E[X_j I_{\mathcal{B}(\delta)^c}(\mathbf{X})]/p,$$

$$c_2 := E[X_i I_{\mathcal{B}(\delta)^c}(\mathbf{X})]E[X_j I_{\mathcal{B}(\delta)}(\mathbf{X})]/p,$$

$$c_3 := E[X_i I_{\mathcal{B}(\delta)^c}(\mathbf{X})]E[X_j I_{\mathcal{B}(\delta)^c}(\mathbf{X})]/p.$$

Utilizando la notación

$$\mathrm{COV}(\mathbf{X}^\delta) := [\tau_{ij}] \ , \ \mathrm{COV}(\mathbf{X}) := [\sigma_{ij}],$$

se tiene, a partir de (1) y el lema (4.9), que para cualquier $i,j = 1,\ldots,k$,

$$\tau_{ij} = \sigma_{ij} + c_1 + c_2 + c_3 + c_4, \tag{2}$$

con $c_4 := -E(X_i X_j I_{\mathcal{B}(\delta)^c}(\mathbf{X}))$.

De esta manera, definiendo

$$A := [a_{ij}],$$

donde

$$a_{ij} := \tau_{ij} - \sigma_{ij},$$

se tiene

$$\mathrm{COV}(\mathbf{X}^\delta) = \mathrm{COV}(\mathbf{X}) + A.$$

Por otra parte, utilizando (2),(4.7)-(4.9) se puede verificar que

$$|a_{ij}| \leq \delta + [\delta^2 + (|\mu_i| + |\mu_j|)\delta]/(1-\delta),$$

donde $\mu_i := E(X_i)$. ■

Lema 4.16. Sea $\mathbf{X}_1,\mathbf{X}_2,\ldots$ una sucesión de vectores aleatorios de dimensión k no acotados, independientes e idénticamente distribuidos, con media μ y matriz de covarianza M. Para $\delta > 0$, $\delta \ll 1$, sean $\mathbf{X}^\delta{}_1,\mathbf{X}^\delta{}_2,\ldots$ vectores independientes con la misma distribución que $\mathbf{X}^{\delta}$ en (4.10). Entonces, para todo n > 0

$$P[\|\mathbf{X}_n^* - \mathbf{X}^{\delta}{}_n^*\| \geq \delta^{1/4}] \leq C(\delta),$$

donde $C(\delta) \to 0$ conforme $\delta \to 0$.

Demostración. Defina

$$\begin{aligned} \mathbf{Z}_n &:= \mathbf{X}_n^* - \mathbf{X}^{\delta}{}_n^* \\ &= [\sum_{i=1}^{n}(\mathbf{X}_i - \mathbf{X}^{\delta}{}_i)]/n^{1/2} \end{aligned}$$

y observe que el lema 4.14 y la independencia de los vectores $\mathbf{X}_i - \mathbf{X}^{\delta}{}_i$ implican

$$E(\mathbf{Z}_n) = \mathbf{0}, \tag{1}$$

$$\mathrm{COV}(\mathbf{Z}_n) = \mathrm{COV}(\mathbf{X} - \mathbf{X}^{\delta}). \tag{2}$$

De esta manera, utilizando la desigualdad de Chebyshev se tiene

$$P[\|\mathbf{Z}_n\| \geq \delta^{1/4}] \leq \mathrm{tr}[\mathrm{COV}(\mathbf{X} - \mathbf{X}^{\delta})]/\delta^{1/2}. \tag{3}$$

Por otra parte, a partir de la definición de $\mathbf{X}^{\delta}$ sigue que

$$\begin{aligned} \mathbf{X} - \mathbf{X}^{\delta} &:= -E(I_{\mathcal{B}(\delta)^c}(\mathbf{X})\cdot\mathbf{X})/P[\mathbf{X} \in \mathcal{B}(\delta)], \text{ si } \mathbf{X} \in \mathcal{B}(\delta) \\ &:= \mathbf{X}, \text{ si } \mathbf{X} \in \mathcal{B}(\delta)^c, \end{aligned}$$

y por lo tanto, utilizando (4.7)-(4.9) se sigue que

$$\begin{aligned} \mathrm{tr}[\mathrm{COV}(\mathbf{X} - \mathbf{X}^{\delta})] &= \sum_{i=1}^{k}\mathrm{Var}(X_i - X^{\delta}{}_i) \\ &= \sum_{i=1}^{k}E[(X_i - X^{\delta}{}_i)^2] \\ &= \sum_{i=1}^{k}\{(E(X_i I_{\mathcal{B}(\delta)^c}(\mathbf{X}))^2/P[\mathbf{X} \in \mathcal{B}(\delta)] + E(X_i^2 I_{\mathcal{B}(\delta)^c}(\mathbf{X})\} \end{aligned}$$

$$\leq k[\delta^2/(1 - \delta) + \delta]. \tag{4}$$

Combinando (3) y (4) se sigue el resultado, con

$$C(\delta) := k\delta^{1/2}/(1-\delta).$$ ■

Concluimos enunciando el teorema del límite central para vectores aleatorios no acotados, cuya demostración se sigue a partir de los lemas 4.14 - 4.16 en forma totalmente análoga a la argumentación propuesta para el teorema 4.12.

Teorema 4.17. Sea $\mathbf{X}$ un vector aleatorio *no acotado* de dimensión k, con esperanza $\mu \in \mathbb{E}_k$ y matriz de covarianza M. Si $\mathbf{X}_1, \mathbf{X}_2, \ldots$ es una sucesión de vectores aleatorios independientes con la misma distribución de $\mathbf{X}$, entonces, si $\mathcal{A}$ es una región (Jordán-medible) en $\mathbb{E}_k$,

$$\lim_{n\to\infty} P[\mathbf{X}_n^* \in \mathcal{A}] = P[\mathbf{W} \in \mathcal{A}],$$

donde $\mathbf{W} \simeq \mathcal{N}_k(0,M)$.

LITERATURA CITADA

Araujo, A. y E. Giné. 1980. *The central limit theorem for real Banach valued random variables*, Wiley, New York, 152 p.

Ash, R.B. 1974. *Real analysis and probability*, Academic Press, new York, 476 p.

Atkinson, K.E. 1978. *An introduction to numerical analysis*, Wiley, New York, 245 p.

Brunk, H.D. 1979. *Introducción a la Estadística matemática*, Trillas, México D.F., 597 p.

Dudley, R.M. 1989. *Real analysis and probability*, Wadsworth Cole, Belmont CA, 328 p.

Feller, W. 1986. *Introducción a la teoría de probabilidades y sus aplicaciones*, Vol. 1, Limusa, México D.F., 504 p.

Fulks, W. 1967. *Advanced calculus*, Wiley, New York, 551 p.

Graybill, F.A. 1976. *Theory and applications of the linear model*, Wadsworth, Belmont CA., 704 p.

Hoffman, K. y R. Kunze. 1973. *Algebra lineal*, Prentice Hall Hispanoamericana, México D.F., 400 p.

Loève, M. 1977. *Probability theory I*, Springer-Verlag, New York, 425 p.

Maistrov, L.E. 1974. *Probability theory, an historical sketch*, Academic Press, New York, 425 p.

Máltsev, A.I. 1978. *Fundamentos de álgebra lineal*, Mir, Moscú, 400 p.

Mandl, P. 1974. Estimation and control in Markov chains, *Advances in applied probability*, 6, 40-60.

Ross, S.M. 1980. *An introduction to stochastic processes*, Springer-Verlag, New York, 347 p.

Rudin, W. 1977. *Real and complex analysis*, Wiley, New York, 397 p.

Scheffé, H. 1959. *The analysis of variance*, Wiley, New York, 477 p.

Spivak, M. 1972. *Cálculo en variedades*, Reverté, Barcelona, 123 p.

Spivak, M. 1978. *Cálculo infinitesimal*, Reverté, Bogotá, 843 p.

APENDICE

A1 LA REGLA DEL TRAPECIO

Antes de mostrar este resultado analítico empleado en la demostración de la fórmula de Stirling (teorema 1.2) recordemos la formulación general del teorema del valor medio para integrales, también conocido como segundo teorema del valor medio (para integrales). (Spivak, 1978).

Teorema A.1. Sean $f,g:[a,b] \to \mathbb{R}$ funciones continuas, $g \geq 0$. Entonces, existe $\xi \in [a,b]$ tal que

$$\int_a^b f(x)g(x)\ dx = f(\xi) \int_a^b g(x)\ dx\ .$$

Demostración. Como f es acotada en $[a,b]$, sean

$$m := \inf\{f(x) : x \in [a,b]\},$$

$$M := \sup\{f(x) : x \in [a,b]\}.$$

Defina $\mu \in \mathbb{R}$ tal que

$$\int_a^b f(x)g(x)\ dx = \mu \int_a^b g(x)\ dx\ . \qquad (*)$$

Mostraremos primeramente que $\mu \in [m,M]$. A partir de la definición de m y M, y dado que $g \geq 0$ en $[a,b]$, se cumple

$$m\ g(x) \leq f(x)g(x) \leq M\ g(x), \text{ para todo } x \in [a,b],$$

de donde se sigue

$$m \int_a^b g(x)\ dx \le \int_a^b f(x)g(x)\ dx \le M \int_a^b g(x)\ dx\ . \qquad (**)$$

Conjuntando (*) y (**) obtenemos $m \le \mu \le M$. De la continuidad de f se sigue entonces que existe $\xi \in [a,b]$ tal que

$$\mu = f(\xi). \qquad \blacksquare$$

Observación. Como caso particular, si $g \equiv 1$ obtenemos la formulación del teorema A.1 que aparece comúnmente como teorema del valor medio para integrales en los textos de cálculo, a saber

$$\int_a^b f(x)g(x)\ dx = f(\xi)\ (b - a), \text{ para algún } \xi \in [a,b].$$

Teorema A.2 (Regla del trapecio). Sea $f:[a,b] \to \mathbb{R}$ una función con segunda derivada continua. Entonces, existe $\xi \in [a,b]$ tal que

$$\int_a^b f(x)\ dx = (1/2)(b-a)[f(a) + f(b)] - (1/12)(b-a)^3 f''(\xi).$$

Demostración. Defina $g(x) := (1/2)(x - a)(b - x)$. Claramente, g es continua y además note que

i) $g \ge 0$ en $[a,b]$; en especial $g(a) = g(b) = 0$.

ii) $g'(x) = -(1/2)(2x - a - b)$; en especial

$$g'(a) = -g'(b) = (1/2)(b - a).$$

iii) $g''(x) \equiv -1$.

A partir de *iii)* se tiene entonces

$$\int_a^b f(x)\ dx = -\int_a^b f(x)g''(x)\ dx.$$

Realizando una integración por partes en el miembro derecho de la igualdad anterior y empleando *ii)* :

$$\int_a^b f(x)\ dx = -f(x)g'(x)\Big|_a^b + \int_a^b f'(x)g'(x)\ dx$$

$$= (1/2)(b-a)[f(a) + f(b)] + \int_a^b f'(x)g'(x)\ dx\ . \quad (*)$$

Mediante un nuevo argumento de integración por partes para el miembro derecho de (*) :

$$\int_a^b f'(x)g'(x)\ dx = f'(x)g(x)\Big|_a^b - \int_a^b f''(x)g(x)\ dx$$

$$= -\int_a^b f''(x)g(x)\ dx\ , \quad (**)$$

donde la última igualdad se obtiene usando *i)*.

Ahora, dado que se ha supuesto la continuidad de f'', aplicando el resultado del teorema A.1 se tiene que existe $\xi \in [a,b]$ tal que

$$\int_a^b f''(x)g(x)\ dx = f''(\xi)\int_a^b g(x)\ dx. \quad (***)$$

Por último, se puede verificar directamente que

$$\int_a^b g(x)\ dx = (1/12)(b-a)^3.$$

Así, a partir de (*)-(***) se sigue el resultado. ∎

La denominación "regla del trapecio" del teorema A.2 viene de la observación de que, si $f > 0$, la expresión

$$(1/2)(b-a)[f(a) + f(b)]$$

puede interpretarse geométricamente como el área del trapecio formado por los puntos $(a,0),(b,0),(a,f(a)),(b,f(b))$.

A2 LA FORMULA DEL PRODUCTO DE WALLIS

En esta sección se proporciona una demostración de la fórmula de Wallis, utilizada en la demostración del teorema 1.2, la cual establece una representación de π como un producto infinito.

Para empezar, considere la sucesión $I_0, I_1, \ldots, I_m, \ldots$ definida por

$$I_m := \int \operatorname{sen}^m x \, dx \ , \ m = 0,1,\ldots$$

Encontremos una fórmula recursiva para esta sucesión. Observe primeramente que

$$I_0 = \int \operatorname{sen}^0 x \, dx = x,$$

$$I_1 = \int \operatorname{sen} x \, dx = -\cos x \ .$$

Para $m \geq 2$, la técnica de integración por partes proporciona el resultado

$$I_m = \int \operatorname{sen}^m x \, dx$$

$$= -\int \frac{d}{dx}(\cos x) \operatorname{sen}^{m-1} x \, dx$$

$$= -\cos x \operatorname{sen}^{m-1} x + (m-1)\int \cos^2 x \operatorname{sen}^{m-2} x \, dx \ ;$$

Utilizando la identidad trigonométrica $\operatorname{sen}^2 x + \cos^2 x = 1$ se tiene:

$$I_m = - \cos x \, \mathrm{sen}^{m-1}x + (m-1)\int (1 - \mathrm{sen}^2 x)\mathrm{sen}^{m-2}x \, dx$$

$$= - \cos x \, \mathrm{sen}^{m-1}x + (m-1)\int \mathrm{sen}^{m-2}x \, dx - (m-1)\int \mathrm{sen}^m x \, dx$$

$$= - \cos x \, \mathrm{sen}^{m-1}x + (m-1) I_{m-2} - (m-1) I_m ,$$

de donde se sigue

$$I_m = -(1/m)\cos x \, \mathrm{sen}^{m-1}x + (m-1/m) I_{m-2}, \quad m \geq 2.$$

Resumiendo, para $I_0, I_1, \ldots, I_m, \ldots$ se tiene la fórmula recursiva

$$\begin{aligned} &I_0 = x \ , \ I_1 = -\cos x \\ &I_m = -(1/m)\cos x \, \mathrm{sen}^{m-1}x + (m-1/m) I_{m-2}, \quad m \geq 2. \end{aligned} \tag{a}$$

Defina ahora la sucesión $a_0, a_1, \ldots, a_m, \ldots$ mediante

$$a_m := \int_0^{\pi/2} \mathrm{sen}^m x \, dx \ , \quad m = 0, 1, \ldots \tag{b}$$

Dado que $g(x) = \mathrm{sen}\, x \cos x$ se anula en 0 y $\pi/2$, de (a) se desprende que $\{a_m\}$ puede definirse recursivamente por

$$\begin{aligned} &a_0 = \pi/2 \ , \ a_1 = 1, \\ &a_m = (m-1/m) \, a_{m-2} \ , \ m \geq 2, \end{aligned} \tag{c}$$

con lo cual se puede calcular a_m en forma recursiva para cualquier m natural. Por ejemplo, observe que

$$a_{10} = \frac{9}{10} a_8 = \frac{9}{10}\frac{7}{8} a_6 = \ldots = \frac{9\cdot7\cdot5\cdot3\cdot1}{10\cdot8\cdot6\cdot4\cdot2}\frac{\pi}{2} .$$

Teorema A.3 (Fórmula del producto de Wallis). Para toda n = 1,2,..defina

$$\pi_n := \prod_{k=1}^{n} 4k^2/(4k^2-1).$$

Entonces la sucesión $\{\pi_n\}_{n=1}^{\infty}$ es convergente y

$$\prod_{k=1}^{\infty} 4k^2/(4k^2-1) := \lim_{n\to\infty} \prod_{k=1}^{n} 4k^2/(4k^2-1) = \pi/2.$$

Demostración. Defina la sucesión $a_0, a_1, \ldots, a_m, \ldots$ como en (b). A partir de la fórmula de recursión (c) observe que las subsucesiones $\{a_2, a_4, \ldots, a_{2n}, \ldots\}, \{a_3, a_5, \ldots, a_{2n+1}, \ldots\}$, $\{a_4, a_6, \ldots, a_{2n+2}, \ldots\}$ pueden definirse por

$$a_{2n} = \frac{(2n-1)(2n-3)\cdots 3\cdot 1}{2n\ (2n-2)\cdots 4\cdot 2}\ \frac{\pi}{2},$$

$$a_{2n+1} = \frac{2n\ (2n-2)\cdots 4\cdot 2}{(2n+1)(2n-1)\cdots 5\cdot 3}, \tag{*}$$

$$a_{2n+2} = \frac{2n+1}{2n+2}\ a_{2n},$$

$n = 1,2,\ldots$, respectivamente. De (*) y la definición de π_n se desprende que para cualesquier $n \in \mathbb{N}$

$$a_{2n+1}/a_{2n} = \frac{2n\cdot 2n\cdots 4\cdot 2\cdot 2}{(2n+1)(2n-1)\cdots 3\cdot 3\cdot 1}\ \frac{2}{\pi}$$

$$= (2/\pi)\prod_{k=1}^{n}(2k)^2/[(2k+1)(2k-1)]$$

$$= (2/\pi)\pi_n . \tag{**}$$

Mostraremos entonces que a_{2n+1}/a_{2n} es convergente.

Primeramente, observe que $0 \le \operatorname{sen} x \le 1$ si $x \in [0,\pi/2]$, de donde se sigue que para todo $n \in \mathbb{N}$

$$\operatorname{sen}^{2n+2} x \le \operatorname{sen}^{2n+1} x \le \operatorname{sen}^{2n} x;$$

entonces, por la definición (b) de a_m, se tiene

$$a_{2n+2} \le a_{2n+1} \le a_{2n}, \quad n = 1,2,\dots ,$$

o lo que es lo mismo

$$a_{2n+2}/a_{2n} \le a_{2n+1}/a_{2n} \le 1 \quad , \quad n = 1,2,\dots \qquad (***)$$

pues $a_{2n} > 0$ para todo n.

Observe ahora de (*) que

$$a_{2n+2}/a_{2n} = \frac{2n+1}{2n+2} ,$$

de donde obtenemos

$$\lim_{n\to\infty} [a_{2n+2}/a_{2n}] = 1.$$

A partir de este último resultado y las desigualdades (***) se sigue que a_{2n+1}/a_{2n} converge, y

$$\lim_{n\to\infty} [a_{2n+1}/a_{2n}] = 1,$$

por lo cual (vea (**)) π_n es convergente y

$$\lim_{n\to\infty} \pi_n = \pi/2. \qquad \blacksquare$$

A3 LA DESIGUALDAD DE CHEBYSHEV

En esta sección presentamos una demostración sencilla de la formulación general de la desigualdad de Chebyshev para vectores aleatorios con matriz de covarianza.

Lema A.4. Sea **Y** un vector aleatorio k-dimensional con media μ y matriz de covarianza M. Entonces, para todo $\varepsilon > 0$,

$$P[\|\mathbf{Y} - \mu\| \ge \varepsilon] \le tr(M)/\varepsilon^2.$$

Demostración. Defina la variable aletoria

$$Z := \|Y - \mu\|$$

$$= [\sum_{i=1}^{k} (Y_i - \mu_i)^2]^{1/2} \tag{1}$$

y para $\varepsilon > 0$ arbitrario sea A el conjunto

$$A := [\varepsilon, \infty) \subset \mathbb{R}. \tag{2}$$

Utilizando la notación $M \equiv [m_{ij}]$, $i,j = 1,\ldots,k$, observe primeramente que $E(Z^2)$ existe y está dada por

$$E[Z^2] = \sum_{i=1}^{k} m_{ii}$$

$$= tr(M). \tag{3}$$

Ahora, a partir de la definición del conjunto A se tiene

$$E(Z^2) \geq E[Z^2 I_A(Z)] \geq \varepsilon^2 E[I_A(Z)],$$

de donde se sigue que

$$E[Z^2]/\varepsilon^2 \geq P[Z \in A]. \tag{4}$$

Combinando (1)-(4) se concluye que, para todo número positivo ε,

$$P[\|Y - \mu\| \geq \varepsilon] \leq tr(M)/\varepsilon^2. \qquad \blacksquare$$