

Estudiando el libro “Probability, random variables, and stochastic processes” de Athanasios Papoulis y S. Unnikrishna Pillai

Ernesto López

21 de enero de 2022

Índice general

1. Conceptos básicos	2
1.1. El espacio muestral	2
1.2. Eventos	2
1.2.1. Secuencias de eventos	2
1.2.2. Álgebra de eventos	7
1.2.3. σ -álgebra de eventos	8
1.3. Espacios de probabilidad	9
1.3.1. Espacio medible	9
1.3.2. Espacio de probabilidad	9
1.3.3. Continuidad de la medida de probabilidad	15
1.3.4. Probabilidad condicional	23
1.3.5. Eventos independientes	23
2. Ensayos repetidos	25
2.1. Experimentos combinados	25
2.1.1. Producto cartesiano de conjuntos	26
2.1.2. Producto cartesiano de dos experimentos	26
2.2. Ensayos de Bernoulli	27
2.2.1. Generalización de ensayos de Bernoulli	29
2.3. Teorema 3.1: Teorema de Bernoulli	29
2.4. Ejemplo 3.17: Estrategia para un juego desfavorable	32
2.4.1. Solución	32
3. El concepto de variable aleatoria	38
3.1. Variable aleatoria	38
3.1.1. Eventos generados por variables aleatorias	38
3.1.2. Definición de variable aleatoria	38
3.2. Distribución y densidad de probabilidad	39
3.2.1. La función distribución de probabilidad	39
3.2.2. La función densidad de probabilidad	40
3.3. Variables aleatorias específicas	41
3.3.1. Variables aleatorias continuas	41
3.3.2. Variables aleatorias discretas	47
3.4. Distribuciones condicionales	47
3.4.1. Probabilidad total y el teorema de Bayes	49
3.4.2. Ejemplo 4-19	51
3.5. Aproximaciones asintóticas de una variable aleatoria binomial	52
3.5.1. Aproximación normal de una variable aleatoria binomial	53
4. Funciones de una variable aleatoria	55
4.1. Distribución de $y = g(x)$	55
4.1.1. Ejemplo 5-1: Distribución de $y = ax + b$	55
4.1.2. Ejemplo 5-2: Distribución de $y = x^2$	56
4.2. Densidad de $y = g(x)$	57

4.2.1. Teorema fundamental	57
4.2.2. Ejemplo	57
4.3. Media y varianza	60
4.3.1. Definiciones	60
4.3.2. Media en función de la distribución	61
4.3.3. Ejemplo 5-18	62
4.3.4. Media y varianza de una variable aleatoria normal	63
4.3.5. Media y varianza de una variable aleatoria Bernoulli	64
4.3.6. Media y varianza de una variable aleatoria Poisson	64
4.3.7. Ejemplo: media y mediana	65
4.4. Momentos	67
4.4.1. Definiciones	67
4.4.2. Momentos de $\mathbf{y} = \mathbf{x} + b$	68
4.4.3. Momentos de una variable aleatoria normal	68
4.4.4. Momentos de una variable aleatoria beta	71
4.4.5. Desigualdad de Chebyshev	72
4.4.6. Desigualdad de Markov	72
4.5. Funciones características	73
4.5.1. Función característica de una variable aleatoria normal	74
4.5.2. Momentos de una variable aleatoria normal	78
4.5.3. Momentos de una variable aleatoria gamma	78
4.5.4. Momentos de una variable aleatoria chi-cuadrado	79
4.5.5. Momentos de una variable aleatoria uniforme	79
4.5.6. Momentos de una variable aleatoria binomial	82
4.5.7. Momentos de una variable aleatoria de Poisson	82
4.5.8. Teorema de DeMoivre-Laplace	83
4.6. Ejemplo 5-32: Probabilidad de sucesión de éxitos	84
4.7. Ejemplo 5-33: El problema de las parejas	87
4.7.1. Solución del problema	87
4.7.2. Otras consideraciones	91
5. Dos variables aleatorias	93
5.1. Algunas definiciones y propiedades	93
5.1.1. Distribución de probabilidad conjunta	93
5.1.2. Densidad de probabilidad conjunta	95
5.1.3. Estadísticas conjuntas	95
5.1.4. Estadísticas marginales	95
5.1.5. Independencia	96
5.1.6. Variables aleatorias conjuntamente normales	99
5.2. Función de dos variables aleatorias	101
5.2.1. Ejemplo: Densidad de probabilidad de $\mathbf{z} = \mathbf{x} + \mathbf{y}$	101
5.2.2. Ejemplo: Densidad de probabilidad de $\mathbf{z} = \mathbf{x}/\mathbf{y}$	103
5.2.3. Estadísticos de orden	107
5.3. Dos funciones de dos variables aleatorias	110
5.3.1. Densidad de probabilidad conjunta	114
5.4. Momentos conjuntos	123
5.4.1. Covarianza	124
5.4.2. No correlación y ortogonalidad	125
5.5. Funciones características conjuntas	131
5.5.1. Independencia y convolución	132
5.5.2. Teorema de los momentos	138
5.6. Distribuciones condicionales	141
5.6.1. Teorema de Bayes y probabilidad total	142
5.7. Valores esperados condicionales	143
5.7.1. Valores esperados condicionados a variables aleatorias	145

6. Secuencias de variables aleatorias	148
6.1. Conceptos generales	148
6.1.1. Independencia	149
6.1.2. Experimentos independientes y ensayos repetidos	150
6.1.3. Media y covarianza	151
6.2. Densidades condicionales, funciones características y normalidad	161
6.2.1. Valores esperados condicionales	162
6.2.2. Funciones características	163
6.2.3. Vectores aleatorios normales	164
6.3. Estimación cuadrática media	177
6.3.1. Estimación cuadrática media lineal	180
6.3.2. Principio de ortogonalidad	181
6.3.3. Estimación cuadrática media lineal: caso general	183
6.3.4. Estimación cuadrática media no lineal: caso general	185
6.4. Convergencia estocástica y teoremas de límite	187
6.4.1. Modos de convergencia de secuencias aleatorias	188
6.4.2. La ley de los grandes números	197
6.4.3. Teorema central del límite	201
7. Estadística	210
7.1. Introducción	210
7.2. Estimación	213
7.2.1. Estimación de intervalos	214
7.3. Estimación de parámetros	219
7.3.1. Método de máxima verosimilitud	219
7.3.2. Estadísticos suficientes	220
7.3.3. Estimadores insesgados	224
7.3.4. Estimador insesgado de varianza mínima	228
7.3.5. El problema de los taxis	232
7.3.6. Cota inferior de Crámer-Rao para el caso de múltiples parámetros	237
7.3.7. Mejoras a la cota de Crámer-Rao: la cota de Bhattacharya	240

Prefacio

Cuando se estudian tópicos de ingeniería como procesamiento estadístico de señales, reconocimiento de patrones, aprendizaje automático o cualquier disciplina que involucre el análisis cuantitativo de datos, es imprescindible el dominio de la teoría de la probabilidad. El presente documento consiste en apuntes sobre el libro “Probability, random variables, and stochastic processes” [1]. Se seleccionó este texto debido a que desarrolla la teoría de forma profunda pero a su vez desde un enfoque práctico, evitando las formalidades matemáticas de la teoría de la probabilidad moderna basada en la teoría de la medida. El objetivo de este documento es explicar mas detalladamente algunos razonamientos que el autor omite posiblemente por asumir triviales.

Capítulo 1

Conceptos básicos

En este capítulo inicial, se incluyen algunas definiciones básicas de la teoría de la probabilidad. A diferencia del resto del documento, este capítulo está basado en el texto [2], aunque también se empleó como soporte el libro [3], mas estricto matemáticamente. Se hará hincapié en las definiciones y teoremas vinculados a secuencias de eventos, con el objetivo de facilitar el estudio de los modos de convergencia de secuencias de variables aleatorias, que se explica en el capítulo 7 de [1] pero con poca profundidad.

1.1. El espacio muestral

Si bien la teoría de la probabilidad está basada en un conjunto de reglas formales abstractas, también posee un significado intuitivo que está relacionado siempre a la realización de un *experimento*. Este experimento puede ser real o conceptual. En la teoría de la probabilidad se asignan probabilidades a todos los resultados posibles del experimento. En el modelo abstracto de un experimento, el conjunto de todos los resultados posibles del experimento se llama *espacio muestral*. Debe tenerse en cuenta que este conjunto es una entidad matemática independientemente del significado intuitivo. El espacio muestral se denota como Ω o S y sus elementos como ω o ζ^1 .

1.2. Eventos

Considérese el experimento de lanzar un dado. Los resultados posibles del experimento son los números naturales del 1 al 6, o en otras palabras, el espacio muestral es $\Omega = \{1, \dots, 6\}$. Sin embargo, podría interesar el resultado “salió un número par”. Este resultado no es un elemento de Ω , si no el *subconjunto* $\{2, 4, 6\}$ de Ω . El resultado “salió un número par” no es un *resultado elemental* del experimento. Cuando el resultado es un subconjunto del espacio muestral en lugar de un resultado elemental, se denomina *evento*. La realización de un experimento se llama *ensayo*. En cada ensayo, ocurre un único resultado ζ_i . Se dice que el evento A ocurrió en el ensayo si contiene el elemento ζ_i , es decir, $\zeta_i \in A$.

Los conceptos intuitivos de resultado y evento incorporan dentro un contexto matemático abstracto como un conjunto y sus subconjuntos. De esta forma, es posible aplicar las operaciones unión, intersección y complementación de conjuntos sobre eventos. Ocurre el evento $A \cup B$ cuando ocurre el evento A o el evento B . Ocurre el evento $A \cap B$ cuando ocurre el evento A y el evento B . Si los eventos A y B son mutuamente excluyentes, si ocurre A no ocurre B y viceversa. Si $A \subset B$ y ocurre A , entonces también ocurre B . El complemento \bar{A} del evento A es el evento que cumple que $A \cup \bar{A} = \Omega$. En cada ensayo, ocurre el evento A o el evento \bar{A} .

En la bibliografía, los eventos se denotan típicamente con letras mayúsculas del comienzo del alfabeto latino: A , B , etcétera. Como se mencionó, un evento es un subconjunto de Ω .

1.2.1. Secuencias de eventos

Sea la secuencia de eventos $\{A_n\}_{n=1}^{\infty} \subset \Omega$. Se define:

¹En este capítulo se usa la notación del texto de referencia ([2]). En el resto del documento se emplea la notación de [1], en donde el espacio muestral se denota como S y sus elementos como ζ .

- *Intersección de la secuencia:*

$$\bigcap_{n=1}^{\infty} A_n \triangleq \{\omega \in \Omega : \omega \in A_n \text{ para todo } n\}.$$

- *Unión de la secuencia:*

$$\bigcup_{n=1}^{\infty} A_n \triangleq \{\omega \in \Omega : \omega \in A_n \text{ para al menos algún } n\}.$$

Límites de secuencias de eventos

Las definiciones de los distintos modos de convergencia de secuencias de variables aleatorias requiere el concepto de límite de secuencias de conjuntos. Sea $\{A_n\}_{n=1}^{\infty} \subset \Omega$. Se define

- *Ínfimo de una secuencia de eventos:*

$$\inf_{k \geq n} A_k \triangleq \bigcap_{k=n}^{\infty} A_k = \{\omega \in \Omega : \omega \in A_k \text{ para todo } k \geq n\}$$

- *Supremo de una secuencia de eventos:*

$$\sup_{k \geq n} A_k \triangleq \bigcup_{k=n}^{\infty} A_k = \{\omega \in \Omega : \omega \in A_k \text{ para al menos un } k \geq n\}$$

Notar que el ínfimo es una secuencia de conjuntos monótonamente creciente con n y el supremo es una secuencia monótonamente decreciente con n (ver sección 1.2.1 sobre secuencias monótonas).

- *Límite inferior de una secuencia de eventos:*

$$\liminf_{n \rightarrow \infty} A_n \triangleq \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

- *Límite superior de una secuencia de eventos:*

$$\limsup_{n \rightarrow \infty} A_n \triangleq \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

Para interpretar estos conceptos, se requiere un análisis profundo de las definiciones, y es lo que se hará a continuación. Sea la secuencia de eventos $\{A_n\}_{n=1}^{\infty} \subset \Omega$,

- (a) la interpretación de \liminf es la siguiente:

$$\begin{aligned} \liminf_{n \rightarrow \infty} A_n &\triangleq \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \{\omega \in \Omega : \omega \in A_n \ \forall n \geq n_0(\omega) \text{ para algún } n_0(\omega) \in \mathbb{N}\} \\ &= \{\omega \in \Omega : \omega \in A_n \text{ para todo } n \text{ excepto para un número finito}\} \\ &= \{\omega \in \Omega : \omega \in A_n \text{ “eventualmente siempre”}\} \end{aligned}$$

- (b) la interpretación de \limsup es la siguiente:

$$\begin{aligned} \limsup_{n \rightarrow \infty} A_n &\triangleq \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \{\omega \in \Omega : \omega \in A_k \text{ para al menos un } k \geq n, \ \forall n\} \\ &= \{\omega \in \Omega : \omega \in A_{n_k}, k = 1, 2, \dots, \text{ para alguna subsecuencia } n_k(\omega)\} \\ &= \{\omega \in \Omega : \omega \in A_n \text{ “infinitamente a menudo”}\} \end{aligned}$$

En la bibliografía, se emplea la siguiente notación:

$$\liminf_{n \rightarrow \infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ e.a.}\}, \quad \limsup_{n \rightarrow \infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ i.o.}\},$$

donde la sigla *e.a.* proviene de la sigla de “eventualmente siempre” en inglés (*eventually always*) y la sigla *i.o.* proviene de la sigla “infinitamente a menudo” en inglés (*infinitely often*).

Dos conexiones entre los conceptos de \liminf y \limsup son las siguientes:

1. Se cumple la relación de inclusión

$$\liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n.$$

Esta relación proviene de que

$$\{\omega \in \Omega : \omega \in A_n \forall n \geq n_0(\omega)\} \subseteq \{\omega \in \Omega : \omega \in A_n \text{ “infinitamente a menudo”}\}$$

2. Además, se cumple que

$$\left(\liminf_{n \rightarrow \infty} A_n\right)^c = \limsup_{n \rightarrow \infty} A_n^c.$$

Esta identidad se obtiene aplicando las leyes de Morgan de la siguiente forma:

$$\left(\liminf_{n \rightarrow \infty} A_n\right)^c = \left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k\right)^c = \bigcap_{n=1}^{\infty} \left(\bigcap_{k=n}^{\infty} A_k\right)^c = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k^c = \limsup_{n \rightarrow \infty} A_n^c$$

De forma análoga, puede probarse que

$$\left(\limsup_{n \rightarrow \infty} A_n\right)^c = \liminf_{n \rightarrow \infty} A_n^c.$$

- *Límite de una secuencia de eventos:* si la secuencia de eventos $\{A_n\}_{n=1}^{\infty}$ es tal que $\liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n$, se dice que la secuencia tiene límite, el cual se define como

$$\lim_{n \rightarrow \infty} A_n = \liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n.$$

Ejemplos

- Sea Ω el conjunto \mathbb{N} de los números naturales y sea

$$A_n = \{\text{todos los números naturales pares/impares si } n \text{ es par/ímpar}\}.$$

Considerando que

$$\bigcup_{k=n}^{\infty} A_k = \mathbb{N}, \quad \bigcap_{k=n}^{\infty} A_k = \emptyset, \quad \forall n,$$

se tiene que

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} \mathbb{N} = \mathbb{N} = \Omega,$$

y

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} \emptyset = \emptyset.$$

Se concluye la secuencia no tiene límite.

(b) Sea $A_n = [0, n/(n+1))$. Como

$$\bigcup_{k=n}^{\infty} A_k = \bigcup_{k=n}^{\infty} \left[0, \frac{k}{k+1}\right) = \left[0, \lim_{k \rightarrow \infty} \frac{k}{k+1}\right) = [0, 1), \quad \forall n,$$

se tiene que

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} [0, 1) = [0, 1).$$

Por otro lado, se cumple que

$$\bigcap_{k=n}^{\infty} A_k = \bigcap_{k=n}^{\infty} \left[0, \frac{k}{k+1}\right) = \left[0, \frac{n}{n+1}\right),$$

donde en la última igualdad se empleó que $[0, n/(n+1)) \subset [0, m/(m+1))$ si $n < m$. Por lo tanto,

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} \left[0, \frac{n}{n+1}\right) = [0, 1).$$

Se concluye que existe el límite de A_n y es $\lim_{n \rightarrow \infty} A_n = [0, 1)$.

(c) Sea $A_n = \{0, 1/n, 2/n, \dots, (n-1)/n, 1\}$. De esta forma,

$$A_1 = \{0, 1\}, \quad A_2 = \left\{0, \frac{1}{2}, 1\right\}, \quad A_3 = \left\{0, \frac{1}{3}, \frac{2}{3}, 1\right\}, \quad A_4 = \left\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\right\}, \quad \dots$$

Notando que

$$\bigcup_{k=n}^{\infty} A_k = \mathbb{Q} \cap [0, 1] \quad \forall n,$$

esto es, todos los números racionales entre 0 y 1 inclusive, se tiene que

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} \mathbb{Q} \cap [0, 1] = \mathbb{Q} \cap [0, 1].$$

Por otro lado,

$$\bigcap_{k=n}^{\infty} A_k = \{0, 1\},$$

así que

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} \{0, 1\} = \{0, 1\}.$$

En este caso, la secuencia A_n no tiene límite.

Secuencias monótonas

Una secuencia de conjuntos $\{A_n\}_{n=1}^{\infty}$ se dice *monótona creciente* si $A_1 \subset A_2 \subset A_3 \subset \dots$ y *monótona decreciente* si $A_1 \supset A_2 \supset A_3 \supset \dots$. Se emplea la notación $A_n \nearrow$ o $A_n \uparrow$ para indicar que una secuencia es creciente y $A_n \searrow$ o $A_n \downarrow$ para indicar que una secuencia es decreciente. El límite de secuencias monótonas de conjuntos siempre existe, como se demostrará a continuación.

Teorema Si la secuencia de conjuntos $\{A_n\}_{n=1}^{\infty}$ es creciente, el límite existe y está dado por

$$\lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n. \quad (1.1)$$

Para demostrarlo, se observa que por tratarse de una secuencia creciente se cumple que

$$\bigcup_{k=n}^{\infty} A_k = \bigcup_{k=1}^{\infty} A_k, \quad \bigcap_{k=n}^{\infty} A_k = A_n,$$

ya que $A_1 \subset A_2 \subset A_3 \subset \dots$. Por lo tanto, el límite superior es

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} \bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} A_k,$$

y el límite inferior es

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} A_n.$$

Se concluye que el límite existe y es

$$\lim_{n \rightarrow \infty} A_n = \liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n.$$

Observar que el resultado del límite inferior es intuitivo, ya que si $\omega \in A_n$ para algún n , debido a que la secuencia es creciente, se cumple que $\omega \in A_{n+1}, A_{n+2}, \dots$. Esto significa que ω ocurre eventualmente siempre para todo ω en cualquier evento A_n de la secuencia, por lo que el límite es $\bigcup_{n=1}^{\infty} A_n$.

Teorema Si la secuencia de conjuntos $\{A_n\}_{n=1}^{\infty}$ es decreciente, el límite existe y está dado por

$$\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n. \quad (1.2)$$

Para demostrarlo, se parte de que por tratarse de una secuencia decreciente se cumple que

$$\bigcup_{k=n}^{\infty} A_k = A_n, \quad \bigcap_{k=n}^{\infty} A_k = \bigcap_{k=1}^{\infty} A_k,$$

ya que $A_1 \supset A_2 \supset A_3 \supset \dots$. Por lo tanto, el límite superior es

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} A_n,$$

y el límite inferior es

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} \bigcap_{k=1}^{\infty} A_k = \bigcap_{k=1}^{\infty} A_k.$$

Se concluye que el límite existe y es

$$\lim_{n \rightarrow \infty} A_n = \liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n.$$

1.2.2. Álgebra de eventos

Sea Ω el espacio muestral de cierto experimento. La colección de todos los posibles eventos del experimento es el conjunto de todos los subconjuntos de Ω . Este conjunto se denomina *conjunto potencia* y se denota como $\mathcal{P}(\Omega)$ o 2^Ω . Si un conjunto tiene n elementos, la cantidad total de subconjuntos es 2^n . Esto proviene de la identidad

$$\sum_{k=0}^n \binom{n}{k} = 2^n,$$

ya que la cantidad de subconjuntos distintos de k elementos que se pueden formar tomados de n elementos es $\binom{n}{k}$, por lo que el total de subconjuntos es la suma de estos subconjuntos para $k = 0, \dots, n$. El lado derecho de la igualdad se deduce del teorema del binomio, dado por la ecuación 2.9, con $x = y = 1$. Otra forma sencilla de deducir que el total de subconjuntos es 2^n es considerar un número binario de n dígitos, donde un 0 o un 1 en la posición i -ésima indica la ausencia o presencia respectivamente del elemento i -ésimo en el subconjunto. El total de números binarios distintos de n dígitos es 2^n . Por ejemplo, en el experimento de arrojar un dado, el espacio muestral contiene $n = 6$ elementos y el total de subconjuntos es $2^6 = 64$, como se muestra en la siguiente tabla

Subconjunto	Representación binaria
$\{\} = \emptyset$	000000
$\{1\}$	000001
$\{2\}$	000010
$\{1, 2\}$	000011
\vdots	\vdots
$\{2, 3, 4, 5, 6\}$	111110
$\{1, 2, 3, 4, 5, 6\} = \Omega$	111111

Los eventos de un experimento son subconjuntos del espacio muestral Ω a los cuales se le asignan probabilidades, como se verá mas adelante. Sin embargo, no necesariamente se considerarán como eventos todos los subconjuntos de Ω . Un motivo de esto es la naturaleza del experimento. Por ejemplo, si en el experimento de arrojar un dado interesa si el resultado es par o impar, alcanza con considerar como eventos únicamente los cuatro conjuntos \emptyset , {par}, {impar} y Ω . Pero la razón principal de no considerar todos los subconjuntos de Ω es de naturaleza matemática: en algunos casos en donde el espacio muestral contiene infinitos elementos, no es posible asignar probabilidades a todos sus subconjuntos satisfaciendo los axiomas de probabilidad, dados en la sección 1.3.2. Por lo tanto, la colección de eventos a considerar es en realidad un subconjunto de 2^Ω . Este subconjunto debe cumplir los requerimientos de un *álgebra de eventos*. En otras palabras, se considerarán como eventos los subconjuntos de 2^Ω que forman un álgebra de eventos.

Un conjunto de eventos se dice **álgebra de eventos** sobre Ω si cumple las siguientes tres condiciones:

- (I) Ω es un evento.
- (II) Si $A \subseteq \Omega$ es un evento, también lo es A^c . Es decir, un álgebra de eventos es cerrado bajo la operación de complementación.
- (III) Si $A, B \subseteq \Omega$ son eventos, también lo es $A \cap B$. Es decir, un álgebra de eventos es cerrado bajo la intersección finita.

Un sinónimo de álgebra es *cuerpo*. También ocasionalmente se emplea el término *campo* por la traducción de término usado en inglés (*field*).

Teorema Sea Ω un espacio muestral y \mathcal{C} un álgebra de eventos sobre Ω . Entonces, se cumple que

1. $\emptyset \in \mathcal{C}$.

Demostración: como \mathcal{C} es un álgebra de eventos, $\Omega \in \mathcal{C}$ y además $\Omega^c = \emptyset \in \mathcal{C}$.

2. Si $A_1, \dots, A_n \in \mathcal{C}$, se cumple que $\bigcup_{i=1}^n A_i \in \mathcal{C}$.

Demostración para $n = 2$: como \mathcal{C} es un álgebra de eventos y $A_1, A_2 \in \mathcal{C}$, se cumple que $A_1^c, A_2^c \in \mathcal{C}$. Además, como \mathcal{C} es cerrado bajo intersecciones finitas, se cumple que $A_1^c \cap A_2^c \in \mathcal{C}$ y además por ser cerrado a la complementación, también se cumple que

$$(A_1^c \cap A_2^c)^c = A_1 \cup A_2 \in \mathcal{C},$$

donde en la igualdad se emplearon las leyes de Morgan. La generalización para n arbitrario se puede demostrar por inducción.

3. Si $A_1, \dots, A_n \in \mathcal{C}$, se cumple que $\bigcap_{i=1}^n A_i \in \mathcal{C}$.

1.2.3. σ -álgebra de eventos

Una familia de eventos \mathcal{F} de un espacio muestral Ω es un σ -álgebra sobre Ω si es un álgebra sobre Ω y además cumple que

(III') Si $\{A_n\}_{n=1}^{\infty}$ es una sucesión infinita de eventos en \mathcal{F} , se cumple que

$$\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}.$$

Una familia de eventos que cumple esta propiedad se dice que es cerrada bajo intersecciones numerables. Es fácil ver que un conjunto de eventos que cumple esta propiedad también es cerrado bajo intersecciones finitas. Además, análogamente a un conjunto álgebra, las condiciones implican que un conjunto σ -álgebra es cerrado a uniones numerables, es decir, también se cumple que

■ Si $\{A_n\}_{n=1}^{\infty}$ es una sucesión infinita de eventos en \mathcal{F} , se cumple que

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}.$$

En la teoría de la probabilidad, el espacio (conjunto) de eventos es un σ -álgebra. De esta forma, gracias a la clausura de las operaciones en el espacio, se logra la suficiente flexibilidad para construir eventos a partir de otros eventos, pero no se brinda flexibilidad completa para evitar problemas para asignar probabilidades a los elementos del espacio.

Ejemplos de σ -álgebra

A continuación se dan algunos ejemplos de conjuntos σ -álgebra sencillos.

(a) *El conjunto potencia*: como el conjunto potencia $\mathcal{F} = \mathcal{P}(\Omega)$ es el conjunto de todos los posibles subconjuntos de Ω , satisface todos los postulados de clausura de un σ -álgebra.

(b) *El σ -álgebra mínimo*: considérese el conjunto $\mathcal{F} = \{\emptyset, \Omega\}$. Cumple los postulados de clausura, ya que,

$$\emptyset^c = \Omega \in \mathcal{F}, \quad \Omega^c = \emptyset \in \mathcal{F}, \quad \emptyset \cap \Omega = \emptyset \in \mathcal{F}.$$

(c) Para cada evento $A \subseteq \Omega$, el conjunto $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ es un σ -álgebra. Es fácil ver que este conjunto cumple con todos los postulados de un σ -álgebra, como por ejemplo

$$(A^c)^c = A \in \mathcal{F}, \quad A \cap A^c = \emptyset \in \mathcal{F}, \quad A \cap \Omega = A \in \mathcal{F}, \quad A \cap A^c \cap \Omega = \emptyset \in \mathcal{F},$$

y así sucesivamente.

(d) *El sigma-álgebra contable/co-contable.* Se entiende por conjunto *contable* aquel de cardinalidad finita o cardinalidad infinita pero numerable. Un ejemplo es el conjunto \mathbb{N} de los números naturales. El conjunto \mathbb{Q} de los números racionales también es contable, aunque esta afirmación es menos intuitiva.

Sea Ω un espacio muestral no contable, y se considera la familia

$$\mathcal{F} = \{A \subseteq \Omega : A \text{ es contable}\} \cup \{A \subseteq \Omega : A^c \text{ es contable}\}.$$

Por ejemplo, si $\Omega = \mathbb{R}$, \mathcal{F} consiste en los subconjuntos de \mathbb{R} que son contables o tienen complemento contable. \mathcal{F} es un σ -álgebra, ya que

- (I) $\Omega \in \mathcal{F}$, ya que $\Omega^c = \emptyset$ es contable por tener cero elementos.
- (II) $A \in \mathcal{F}$ implica que $A^c \in \mathcal{F}$. Efectivamente, si $A \in \mathcal{F}$ significa que A o A^c es contable. Por lo tanto, $A^c \in \mathcal{F}$ ya que A^c o $(A^c)^c = A$ es contable.
- (III) Si $\{A_n\}_{n=1}^{\infty} \in \mathcal{F}$, se cumple que $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$. Para verificar esto, hay que considerar dos casos
 - (a) si al menos un A_n es contable, $\bigcap_{n=1}^{\infty} A_n$ es contable y por lo tanto está en \mathcal{F} .
 - (b) si ningún A_n es contable, A_n^c es contable para todo n , y por lo tanto $\bigcup_{n=1}^{\infty} A_n^c$ es contable.

Como

$$\left(\bigcap_{n=1}^{\infty} A_n \right)^c = \bigcup_{n=1}^{\infty} A_n^c \text{ es contable,}$$

se concluye que $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$, ya que su complemento es contable.

1.3. Espacios de probabilidad

1.3.1. Espacio medible

Se llama *espacio medible* al par (Ω, \mathcal{F}) que consiste en un conjunto Ω y en un σ -álgebra \mathcal{F} de subconjuntos (eventos) de Ω . El nombre proviene de que se trata de un espacio que está listo para asignarle una medida. Cuando a cada elemento de \mathcal{F} se le asigna una medida de probabilidad P , se construye un *espacio de probabilidad*.

1.3.2. Espacio de probabilidad

Un espacio de probabilidad es el triplete (Ω, \mathcal{F}, P) , donde Ω es el espacio muestral de cierto experimento, \mathcal{F} es un σ -álgebra de subconjuntos de Ω y P es una medida de probabilidad con dominio \mathcal{F} y rango $[0, 1]$, es decir,

$$P : \mathcal{F} \rightarrow \mathbb{R},$$

que cumple las siguientes propiedades:

- (I) Para cada evento $A \in \mathcal{F}$, $P(A) > 0$.
- (II) P es *contable aditiva*: si $\{A_n\}_{n=1}^{\infty}$ es una secuencia de eventos disjuntos en \mathcal{F} , se cumple que

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n). \quad (1.3)$$

Esta propiedad también se llama σ -aditiva.

- (III) $P(\Omega) = 1$.

Algunas consecuencias de la definición de medida de probabilidad P son las siguientes:

1. Se cumple que

$$P(A^c) = 1 - P(A), \quad (1.4)$$

ya que,

$$1 \stackrel{(III)}{=} P(\Omega) = P(A \cup A^c) \stackrel{(II)}{=} P(A) + P(A^c).$$

2. Se cumple que

$$P(\emptyset) = 0,$$

ya que $P(\emptyset) = P(\Omega^c) = 1 - P(\Omega) = 1 - 1 = 0$.

3. Para dos eventos A, B cualesquiera, se cumple que²

$$P(A \cup B) = P(A) + P(B) - P(AB). \quad (1.5)$$

Para demostrarlo, se observa que

$$\begin{aligned} P(A \cup B) &\stackrel{(a)}{=} P(AB^c \cup BA^c \cup AB) \\ &\stackrel{(b)}{=} P(AB^c) + P(BA^c) + P(AB) \\ &\stackrel{(c)}{=} P(A) - P(AB) + P(B) - P(AB) + P(AB) \\ &= P(A) + P(B) - P(AB), \end{aligned}$$

donde la identidad (a) se observa en el diagrama de Venn de la figura 1.1, en (b) se usó que AB^c , BA^c y AB son disjuntos, como se aprecia también en la figura 1.1 y en (c) se sustituyó $P(AB^c)$ y $P(BA^c)$ a partir de las siguientes identidades,

$$\begin{aligned} P(A) &= P(AB^c) + P(AB) \\ P(B) &= P(BA^c) + P(AB), \end{aligned} \quad (1.6)$$

que provienen de notar por ejemplo que $A = (AB^c) \cup (AB)$ y demás que los conjuntos AB^c y AB son disjuntos.

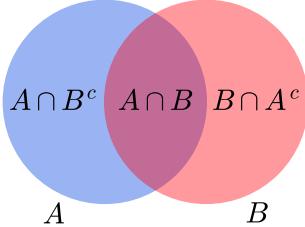


Figura 1.1: Diagrama de Venn con los eventos A y B . Se observa que se cumple que $A \cup B = AB^c \cup BA^c \cup AB$, donde los eventos AB^c , BA^c y AB son disjuntos.

4. *Principio de inclusión-exclusión*: si A_1, \dots, A_n son eventos, se cumple que

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} A_{i_2}) + \sum_{1 \leq i_1 < i_2 < i_3 \leq n} P(A_{i_1} A_{i_2} A_{i_3}) + \dots \\ &\quad + (-1)^{n-1} \sum_{1 \leq i_1 < \dots < i_{n-1} \leq n} P(A_{i_1} \dots A_{i_{n-1}}) + (-1)^n P(A_1 \dots A_n). \end{aligned} \quad (1.7)$$

El principio puede demostrarse por inducción. Ya se demostró que se cumple para $n = 2$, dado por la ecuación 1.5. Asumiendo que se cumple para n , como en la ecuación 1.7, hay que demostrar que se cumple para $n + 1$. Desarrollando el lado izquierdo para $n + 1$, se tiene que

$$\begin{aligned} P(A_1 \cup \dots \cup A_n \cup A_{n+1}) &\stackrel{(a)}{=} P[(A_1 \cup \dots \cup A_n) \cup A_{n+1}] \\ &\stackrel{(b)}{=} P(A_1 \cup \dots \cup A_n) + P(A_{n+1}) - P[(A_1 \cup \dots \cup A_n) \cap A_{n+1}] \\ &\stackrel{(c)}{=} P(A_1 \cup \dots \cup A_n) + P(A_{n+1}) - P[(A_1 \cap A_{n+1}) \cup \dots \cup (A_n \cap A_{n+1})], \end{aligned} \quad (1.8)$$

²De aquí en mas, se empleará la notación AB para indicar $A \cap B$.

donde en (a) se empleó la propiedad asociativa de la operación de unión, en (b) se empleó el principio de inclusión-exclusión para $n = 2$ y en (c) la propiedad distributiva de la operación de intersección. El primer y tercer sumando consisten en probabilidades de n uniones, y por hipótesis se cumple el principio de inclusión-exclusión, que para el primer sumando está dado por la ecuación 1.7 y para el tercer sumando es

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i A_{n+1}\right) &= \sum_{i=1}^n P(A_i A_{n+1}) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} A_{i_2} A_{n+1}) + \cdots \\ &\quad + (-1)^{n-1} \sum_{1 \leq i_1 < \dots < i_{n-1} \leq n} P(A_{i_1} \cdots A_{i_{n-1}} A_{n+1}) + (-1)^n P(A_1 \cdots A_n A_{n+1}). \end{aligned}$$

Sustituyendo esos resultados en la ecuación 1.8, se obtiene que

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= \left[\sum_{i=1}^n P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} A_{i_2}) + \sum_{1 \leq i_1 < i_2 < i_3 \leq n} P(A_{i_1} A_{i_2} A_{i_3}) + \cdots \right. \\ &\quad \left. + (-1)^n P(A_1 \cdots A_n) \right] \\ &\quad + P(A_{n+1}) \\ &\quad - \left[\sum_{i=1}^n P(A_i A_{n+1}) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} A_{i_2} A_{n+1}) + \cdots \right. \\ &\quad \left. + (-1)^{n-1} \sum_{1 \leq i_1 < \dots < i_{n-1} \leq n} P(A_{i_1} \cdots A_{i_{n-1}} A_{n+1}) + (-1)^n P(A_1 \cdots A_n A_{n+1}) \right]. \end{aligned}$$

y reordenando los sumandos, la expresión puede escribirse como

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= \left[\sum_{i=1}^n P(A_i) + P(A_{n+1}) \right] - \left[\sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} A_{i_2}) + \sum_{i=1}^n P(A_i A_{n+1}) \right] \\ &\quad + \left[\sum_{1 \leq i_1 < i_2 < i_3 \leq n} P(A_{i_1} A_{i_2} A_{i_3}) + \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} A_{i_2} A_{n+1}) \right] + \cdots \quad (1.9) \\ &\quad + (-1)^n \left[P(A_1 \cdots A_n) + \sum_{1 \leq i_1 < \dots < i_{n-1} \leq n} P(A_{i_1} \cdots A_{i_{n-1}} A_{n+1}) \right] \\ &\quad + (-1)^{n+1} P(A_1 \cdots A_n A_{n+1}). \end{aligned}$$

El paso siguiente, es notar que cada sumando entre paréntesis rectos puede agruparse en una sola sumatoria. Por ejemplo, los sumandos de la sumatoria en la penúltima expresión entre paréntesis

rectos son las $\binom{n}{n-1} = n$ probabilidades de los productos de los siguientes n factores

$$\begin{array}{cccccccccc}
 A_1 & A_2 & A_3 & \cdots & A_{n-2} & A_{n-1} & \bigcirc & A_{n+1} \\
 A_1 & A_2 & A_3 & \cdots & A_{n-2} & \bigcirc & A_n & A_{n+1} \\
 A_1 & A_2 & A_3 & \cdots & \bigcirc & A_{n-1} & A_n & A_{n+1} \\
 \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
 A_1 & A_2 & \bigcirc & \cdots & A_{n-2} & A_{n-1} & A_n & A_{n+1} \\
 A_1 & \bigcirc & A_3 & \cdots & A_{n-2} & A_{n-1} & A_n & A_{n+1} \\
 \bigcirc & A_2 & A_3 & \cdots & A_{n-2} & A_{n-1} & A_n & A_{n+1},
 \end{array}$$

y al sumar el término $P(A_1 \cdots A_n)$, se obtiene la suma de las $n+1$ probabilidades de los productos de n factores

$$\begin{array}{cccccccccc}
 A_1 & A_2 & A_3 & \cdots & A_{n-2} & A_{n-1} & A_n & \bigcirc \\
 A_1 & A_2 & A_3 & \cdots & A_{n-2} & A_{n-1} & \bigcirc & A_{n+1} \\
 A_1 & A_2 & A_3 & \cdots & A_{n-2} & \bigcirc & A_n & A_{n+1} \\
 A_1 & A_2 & A_3 & \cdots & \bigcirc & A_{n-1} & A_n & A_{n+1} \\
 \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
 A_1 & A_2 & \bigcirc & \cdots & A_{n-2} & A_{n-1} & A_n & A_{n+1} \\
 A_1 & \bigcirc & A_3 & \cdots & A_{n-2} & A_{n-1} & A_n & A_{n+1} \\
 \bigcirc & A_2 & A_3 & \cdots & A_{n-2} & A_{n-1} & A_n & A_{n+1},
 \end{array}$$

expresión que se puede escribir como

$$\sum_{1 \leq i_1 < \dots < i_n \leq n+1} P(A_{i_1} \cdots A_{i_n}).$$

Generalizando este razonamiento para todos los sumandos entre paréntesis rectos de la ecuación 1.9, se obtiene que

$$\begin{aligned}
 P\left(\bigcup_{i=1}^{n+1} A_i\right) &= \sum_{i=1}^{n+1} P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n+1} P(A_{i_1} A_{i_2}) + \sum_{1 \leq i_1 < i_2 < i_3 \leq n+1} P(A_{i_1} A_{i_2} A_{i_3}) + \dots \\
 &\quad + (-1)^n \sum_{1 \leq i_1 < \dots < i_n \leq n+1} P(A_{i_1} \cdots A_{i_n}) \\
 &\quad + (-1)^{n+1} P(A_1 \cdots A_{n+1}),
 \end{aligned}$$

con lo que queda demostrado el paso inductivo.

5. *Propiedad de monotonía.* La medida de probabilidad P es no decreciente: dados dos eventos A y B , se cumple que

$$\text{si } A \subset B \quad \Rightarrow \quad P(A) \leq P(B). \quad (1.10)$$

Efectivamente, como $B = A \cup (B \setminus A)$ y los conjuntos A y $B \setminus A$ son disjuntos, se cumple que

$$P(B) = P(A) + P(B \setminus A) \geq P(A).$$

6. *Desigualdad de Boole.* La medida de probabilidad P es contable sub-aditiva: para eventos cualesquiera A_n , $n \geq 1$, se cumple que

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n). \quad (1.11)$$

Para demostrarlo, considérese la secuencia de eventos

$$B_1 = A_1, \quad B_2 = A_2 \setminus A_1, \quad B_3 = A_3 \setminus (A_1 \cup A_2), \dots, \quad B_n = A_n \setminus \left(\bigcup_{k=1}^{n-1} A_k \right), \dots$$

Los eventos B_n son disjuntos, su unión es igual a la unión de los eventos A_n y además $B_n \subseteq A_n$. Por lo tanto, se cumple que

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n) \leq \sum_{n=1}^{\infty} P(A_n).$$

Espacios contables

Si el espacio muestral Ω tiene un número finito N de resultados posibles, la probabilidad de todos los eventos puede ser expresada en términos de las probabilidades

$$P\{\zeta_i\} = p_i$$

de los resultados elementales $\{\zeta_i\}$. Debido a las propiedades de la medida de probabilidad, los valores p_i deben ser no-negativos y su suma debe valer 1,

$$p_i \geq 0, \quad p_1 + \dots + p_N = 1.$$

Considérese el caso en que el evento A consiste en r resultados elementales ζ_{k_i} , $i = 1, \dots, r$. Como A puede expresarse como la unión de los resultados elementales $\{\zeta_{k_i}\}$, se cumple que

$$P(A) = P\{\zeta_{k_1}\} + \dots + P\{\zeta_{k_r}\} = p_{k_1} + \dots + p_{k_r}.$$

Por la propiedad contable aditiva de la medida de probabilidad (ecuación 1.3), esto también es cierto en el caso en que Ω consiste en una cantidad infinita pero contable de resultados elementales.

Ejemplo El siguiente ejemplo (ejemplo 2.8 de [1]) muestra como es posible determinar las probabilidades de eventos complejos expresándolos como la unión de eventos más simples mutuamente excluyentes.

Una urna contiene m bolas blancas y n bolas negras. Se extrae aleatoriamente una bola por vez sin reposición. Se quiere calcular la probabilidad de extraer alguna bola blanca antes de la k -ésima extracción (inclusive).

Denótese como W_k al evento

$$W_k = \{\text{una bola blanca es extraída antes de la } k\text{-ésima extracción}\}.$$

Notar que el complemento del evento W_k es la extracción de k bolas negras en las k extracciones. El evento W_k puede ocurrir de las siguientes formas mutuamente excluyentes: se extre una bola blanca en la primera extracción, o se extrae una bola negra seguida de una bola blanca, o se extraen dos bolas negras seguidas de una bola blanca, y así sucesivamente hasta la extracción de $k-1$ bolas negras seguida de una bola blanca. Sea el evento

$$X_i = \{\text{se extraen } i \text{ bolas negras seguidas de una bola blanca}\}, \quad i = 0, 1, \dots, n.$$

De esta forma,

$$W_k = X_0 \cup X_1 \cup \dots \cup X_{k-1},$$

y como los eventos X_i son mutuamente excluyentes, se cumple que

$$P(W_k) = \sum_{i=0}^{k-1} P(X_i).$$

Observando que

$$\begin{aligned}
P(X_0) &= \frac{m}{m+n} \\
P(X_1) &= \frac{n}{m+n} \cdot \frac{m}{m+n-1} \\
P(X_2) &= \frac{n}{m+n} \cdot \frac{n-1}{m+n-1} \cdot \frac{m}{m+n-2} \\
&\vdots \\
P(X_i) &= \frac{n}{m+n} \cdot \frac{n-1}{m+n-1} \cdots \frac{n-(i-1)}{m+n-(i-1)} \cdot \frac{m}{m+n-i} \\
&\vdots \\
P(X_{k-1}) &= \frac{n}{m+n} \cdot \frac{n-1}{m+n-1} \cdots \frac{n-(k-2)}{m+n-(k-2)} \cdot \frac{m}{m+n-(k-1)},
\end{aligned}$$

y sumando se obtiene que

$$\begin{aligned}
P(W_k) &= \frac{m}{m+n} \left\{ 1 + \frac{n}{m+n-1} + \frac{n[n-1]}{[m+n-1][m+n-2]} + \cdots \right. \\
&\quad \left. + \frac{n[n-1] \cdots [n-(k-2)]}{[m+n-1][m+n-2] \cdots [m+n-(k-1)]} \right\}, \tag{1.12}
\end{aligned}$$

que es el resultado que se buscaba.

Es interesante notar que en la extracción $n+1$, se debe haber extraído al menos una bola blanca, por lo que

$$P(W_{n+1}) = 1.$$

Sustituyendo este resultado en la ecuación 1.12, es decir, evaluando en $k = n+1$ e igualando a 1, se obtiene la siguiente interesante identidad:

$$1 + \frac{n}{m+n-1} + \frac{n(n-1)}{(m+n-1)(m+n-2)} + \cdots + \frac{n(n-1) \cdots 2 \cdot 1}{(m+n-1)(m+n-2) \cdots (m+1)m} = \frac{m+n}{m}.$$

También vale la pena observar que, como se mencionó previamente, el evento

$$Y_k = \{\text{se extraen } k \text{ bolas negras en } k \text{ extracciones}\}$$

es el complemento del evento W_k , y por lo tanto,

$$P(W_k) = 1 - P(Y_k).$$

Como Y_k ocurre cuando se extrae una bola negra en la primer extracción, y se extrae una bola negra en la segunda extracción y así sucesivamente hasta la k -ésima extracción, se tiene que

$$P(Y_k) = \frac{n}{m+n} \cdot \frac{n-1}{m+n-1} \cdots \frac{n-(k-1)}{m+n-(k-1)},$$

y por lo tanto,

$$P(W_k) = 1 - \frac{n(n-1) \cdots (n-k+1)}{(n+m)(n+m-1) \cdots (n+m-k+1)}.$$

Esta expresión es equivalente a la de la ecuación 1.12.

Espacios no contables. El eje real.

Si el espacio muestral Ω consiste en una cantidad infinita no contable de elementos, la probabilidad de los eventos no puede expresarse en función de la probabilidad de los resultados elementales. Considérese que Ω es el conjunto de todos los números reales. Sus subconjuntos son conjuntos arbitrarios de puntos

en el eje real. Puede demostrarse que no es posible asignar probabilidades a todos los subconjuntos de Ω de forma de satisfacer las propiedades de una medida de probabilidad.

Para construir una medida de probabilidad sobre el eje real, se considerarán como eventos todos los intervalos $x_1 \leq x \leq x_2$, y sus uniones e intersecciones contables. Puede demostrarse que estos eventos forman un σ -álgebra de eventos. Este σ -álgebra incluye todas las semirrectas $x \leq x_i$, donde x_i es cualquier número real. También contiene todos los intervalos abiertos y cerrados, todos los puntos individuales del eje real, y en general, todos los conjuntos de puntos de interés práctico en las aplicaciones. Sin embargo, es posible demostrar que existen conjuntos de puntos en el eje real que no son uniones e intersecciones contables de intervalos, y por lo tanto, el σ -álgebra generado por intervalos no contiene todos los subconjuntos de Ω .

Para obtener una especificación completa de un espacio de probabilidad sobre el eje real, alcanza con asignar probabilidades a los eventos $\{x \leq x_i\}$. El resto de las probabilidades surgen de las propiedades de la medida de probabilidad. Supóngase que $\alpha(x)$ es una función que cumple que

$$\int_{-\infty}^{\infty} \alpha(x)dx = 1, \quad \alpha(x) \geq 0.$$

Se define la probabilidad del evento $\{x \leq x_i\}$ como

$$P\{x \leq x_i\} = \int_{-\infty}^{x_i} \alpha(x)dx.$$

De esta forma, quedan especificadas todas las probabilidades del espacio de probabilidad. Por ejemplo, para obtener la probabilidad del evento $\{x_1 < x \leq x_2\}$, se observa que

$$\{x \leq x_1\} \cup \{x_1 < x \leq x_2\} = \{x \leq x_2\},$$

donde los eventos $\{x \leq x_1\}$ y $\{x_1 < x \leq x_2\}$ son mutuamente excluyentes. Por lo tanto, se cumple que

$$P\{x \leq x_1\} + P\{x_1 < x \leq x_2\} = P\{x \leq x_2\}.$$

De esta forma,

$$\begin{aligned} P\{x_1 < x \leq x_2\} &= P\{x \leq x_2\} - P\{x \leq x_1\} \\ &= \int_{-\infty}^{x_2} \alpha(x)dx - \int_{-\infty}^{x_1} \alpha(x)dx \\ &= \int_{x_1}^{x_2} \alpha(x)dx, \end{aligned}$$

quedando definida la probabilidad del evento $\{x_1 < x \leq x_2\}$ como

$$P\{x_1 < x \leq x_2\} = \int_{x_1}^{x_2} \alpha(x)dx. \quad (1.13)$$

Se observa además que si la función $\alpha(x)$ es acotada, la integral 1.13 tiende a 0 cuando $x_1 \rightarrow x_2$. Esto conduce a la conclusión de que la probabilidad del resultado elemental $\{x_2\}$ es nula para todo x_2 en el eje real. La probabilidad de cada resultado elemental de Ω es 0, a pesar de que la probabilidad de su unión es 1. Esto no entra en conflicto con la propiedad contable aditiva de la medida de probabilidad dada por la ecuación 1.3, ya que la cantidad de elementos de Ω no es contable.

1.3.3. Continuidad de la medida de probabilidad

Se presentan a continuación los teoremas vinculados a la *continuidad* de la función de probabilidad. La medida de probabilidad es una función continua en el sentido de que si una secuencia de eventos $\{A_n\}_{n=1}^{\infty}$ tiene límite, se cumple que $P(\lim A_n) = \lim P(A_n)$. Se comenzará demostrando una versión más liviana de este teorema para el caso particular de secuencias crecientes y decrecientes.

Teorema *Continuidad para secuencias crecientes:* sea $\{A_n\}_{n=1}^{\infty}$ una secuencia creciente de eventos, entonces se cumple que

$$P\left(\lim_n A_n\right) = \lim_n P(A_n).$$

Además, como $P(A_n)$ es una función creciente, se cumple que

$$P(A_n) \nearrow \lim_n P(A_n).$$

Demostración Como se vio en la sección 1.2.1, el límite de una secuencia creciente existe y está dado por

$$\lim_n A_n = \bigcup_{k=1}^{\infty} A_k.$$

Notar que por ser una secuencia creciente, para cualquier n se cumple que

$$\bigcup_{k=1}^n A_k = A_n \quad \Rightarrow \quad P\left(\bigcup_{k=1}^n A_k\right) = P(A_n).$$

Para demostrar el teorema, se define la secuencia de eventos disjuntos

$$B_1 = A_1, \quad B_2 = A_2 \setminus A_1, \dots, \quad B_n = A_n \setminus A_{n-1}, \dots$$

Notar que se cumple que

$$\bigcup_{k=1}^n B_k = \bigcup_{k=1}^n A_k = A_n,$$

y por lo tanto,

$$\bigcup_{k=1}^{\infty} B_k = \bigcup_{k=1}^{\infty} A_k.$$

De esta forma,

$$\begin{aligned} P\left(\lim_n A_n\right) &= P\left(\bigcup_{k=1}^{\infty} A_k\right) \\ &= P\left(\bigcup_{k=1}^{\infty} B_k\right) \\ &\stackrel{(a)}{=} \sum_{k=1}^{\infty} P(B_k) \\ &= \lim_n \sum_{k=1}^n P(B_k) \\ &\stackrel{(b)}{=} \lim_n P\left(\bigcup_{k=1}^n B_k\right) \\ &= \lim_n P(A_n), \end{aligned}$$

donde en (a) y (b) se empleó la propiedad contable aditiva de P para eventos disjuntos.

Teorema *Continuidad para secuencias decrecientes:* sea $\{A_n\}_{n=1}^{\infty}$ una secuencia decreciente de eventos, entonces se cumple que

$$P\left(\lim_n A_n\right) = \lim_n P(A_n).$$

Además, como $P(A_n)$ es una función decreciente, se cumple que

$$P(A_n) \searrow \lim_n P(A_n).$$

Demostración Como la sucesión $\{A_n\}_{n=1}^{\infty}$ es decreciente, la sucesión $\{A_n^c\}_{n=1}^{\infty}$ es creciente. Por lo tanto,

$$\begin{aligned}
P\left(\lim_n A_n\right) &\stackrel{(a)}{=} P\left(\bigcap_{k=1}^{\infty} A_k\right) \\
&\stackrel{(b)}{=} 1 - P\left[\left(\bigcap_{k=1}^{\infty} A_k\right)^c\right] \\
&\stackrel{(c)}{=} 1 - P\left(\bigcup_{k=1}^{\infty} A_k^c\right) \\
&\stackrel{(d)}{=} 1 - P\left(\lim_n A_n^c\right) \\
&\stackrel{(e)}{=} 1 - \lim_n P(A_n^c) \\
&\stackrel{(f)}{=} 1 - \lim_n [1 - P(A_n)] \\
&= \lim_n P(A_n),
\end{aligned}$$

donde en (a) se consideró que A_n es decreciente por lo que el límite existe y está dado por la ecuación 1.2, en (b) y en (f) se empleó la propiedad 1.4 de la medida de probabilidad, en (c) se aplicaron leyes de Morgan, en (d) se empleó que A_n^c es creciente y el límite está dado por la ecuación 1.1, y en (e) se aplicó el resultado del teorema anterior de la continuidad de la función de probabilidad para secuencias crecientes.

Lema *Lema de Fatou: sea $\{A_n\}_{n=1}^{\infty}$ una secuencia arbitraria de eventos, entonces se cumple que*

$$P\left(\liminf_n A_n\right) \leq \liminf_n P(A_n). \quad (1.14)$$

Demostración Recordar que

$$\liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} G_n,$$

donde se definió la secuencia creciente $G_n = \bigcap_{k=n}^{\infty} A_k$, es el conjunto de resultados que ocurren “eventualmente siempre”. Por lo tanto,

$$\begin{aligned}
\lim_n P(G_n) &\stackrel{(a)}{=} P\left(\lim_n G_n\right) \\
&\stackrel{(b)}{=} P\left(\bigcup_{n=1}^{\infty} G_n\right) \\
&= P\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k\right) \\
&= P\left(\liminf_n A_n\right),
\end{aligned} \quad (1.15)$$

donde en (a) se empleó la continuidad de la función de probabilidad para secuencias crecientes y en (b) se empleó que el límite de secuencias crecientes existe y está dado por la ecuación 1.1. Por otro lado, se tiene que como $G_n = \bigcap_{k=n}^{\infty} A_k$, se cumple que

$$P(G_n) \leq \inf_{k \geq n} P(A_k),$$

y tomando el límite cuando $n \rightarrow \infty$, se tiene que

$$\lim_n P(G_n) \leq \lim_n \left[\inf_{k \geq n} P(A_k) \right].$$

Como se vio previamente en la ecuación 1.15, el lado izquierdo de la desigualdad converge a $P(\liminf_n A_n)$, y el lado derecho converge a $\liminf_n P(A_n)$, resultando en 1.14, que es lo que se quería demostrar.

Lema *Lema de Fatou Reverso:* sea $\{A_n\}_{n=1}^{\infty}$ una secuencia arbitraria de eventos, entonces se cumple que

$$\limsup_n P(A_n) \leq P\left(\limsup_n A_n\right) \quad (1.16)$$

Demostración La demostración es análoga a la del lema de Fatou. Recordar que

$$\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} G_n,$$

donde se definió la secuencia decreciente $G_n = \bigcup_{k=n}^{\infty} A_k$, es el conjunto de resultados que ocurren “infinitamente a menudo”. Por un lado, se tiene que

$$\begin{aligned} \lim_n P(G_n) &\stackrel{(a)}{=} P\left(\lim_n G_n\right) \\ &\stackrel{(b)}{=} P\left(\bigcap_{n=1}^{\infty} G_n\right) \\ &= P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) \\ &= P\left(\limsup_n A_n\right), \end{aligned} \quad (1.17)$$

donde en (a) se empleó la continuidad de la función de probabilidad para secuencias decrecientes y en (b) se empleó que el límite de secuencias decrecientes existe y está dado por la ecuación 1.2. Por otro lado, se tiene que como $G_n = \bigcup_{k=n}^{\infty} A_k$, se cumple que

$$P(G_n) \geq \sup_{k \geq n} P(A_k),$$

y tomando el límite cuando $n \rightarrow \infty$, se tiene que

$$\lim_n P(G_n) \geq \lim_n \left[\sup_{k \geq n} P(A_k) \right].$$

Como se vio previamente en la ecuación 1.17, el lado izquierdo de la desigualdad converge a $P(\limsup_n A_n)$, y el lado derecho converge a $\limsup_n P(A_n)$, resultando en 1.16, que es lo que se quería demostrar.

Es necesario realizar una consideración sobre los lemas de Fatou. En la ecuación 1.14 por ejemplo, el \liminf_n del lado izquierdo de la desigualdad se aplica sobre la secuencia de eventos A_n , mientras que el \liminf_n del lado derecho se aplica sobre la sucesión de números $P(A_n)$. Tanto el límite inferior \liminf_n como el límite superior \limsup_n aplicados a secuencias de eventos, fueron definidos en la sección 1.2.1. La definición del límite inferior y el límite superior aplicados a la sucesión de números x_n son las siguientes:

$$\liminf_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \left(\inf_{k \geq n} x_k \right), \quad \limsup_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \left(\sup_{k \geq n} x_k \right),$$

como ya se sugirió en la demostración de los lemas de Fatou. La interpretación de el límite inferior y el límite superior aplicado a sucesiones numéricas, se ilustra en la figura 1.2³.

Combinando el lema de Fatou y el lema de Fatou reverso, se obtienen las siguientes desigualdades:

$$P\left(\liminf_n A_n\right) \leq \liminf_n P(A_n) \leq \limsup_n P(A_n) \leq P\left(\limsup_n A_n\right), \quad (1.18)$$

donde la primera y ultima desigualdad son el lema de Fatou y el lema de Fatou reverso respectivamente, y la desigualdad del centro proviene de que para cualquier secuencia numérica x_n , se cumple que

$$\inf_{k \geq n} x_k \leq \sup_{k \geq n} x_k, \quad \text{para todo } n.$$

³Fuente: https://en.wikipedia.org/wiki/Limit_superior_and_limit_inferior

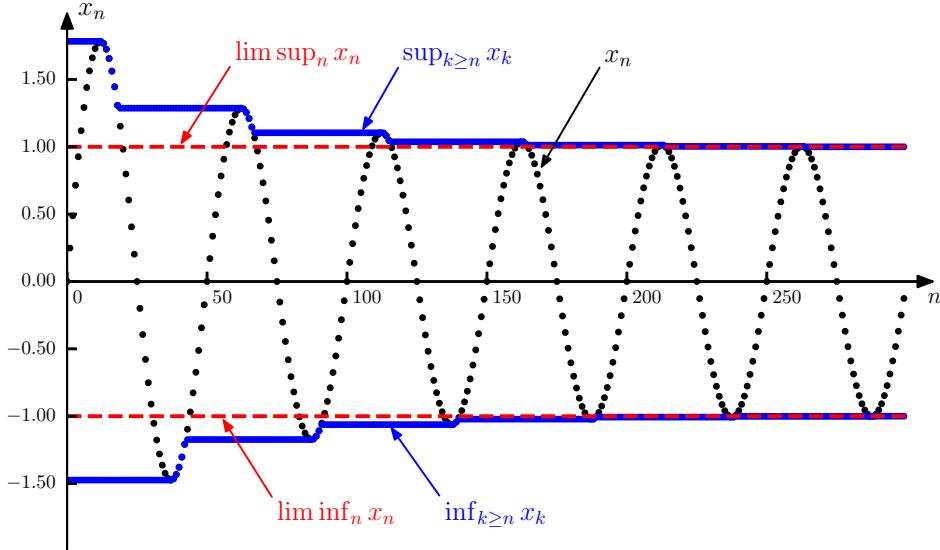


Figura 1.2: Ilustración del límite inferior y el límite superior aplicado a la secuencia de números x_n . La secuencia x_n se grafica en negro. La secuencia $\sup_{k \geq n} x_k$ es una secuencia en n , donde el valor que toma en n se obtiene como el máximo valor que toma x_k en $k \geq n$. El límite superior $\limsup_n x_n$ es el límite de la secuencia $\sup_{k \geq n} x_k$ cuando $n \rightarrow \infty$. Análogamente, $\inf_{k \geq n} x_k$ es la secuencia que consiste en el mínimo de x_k para $k \geq n$ para cada valor de n , y el límite inferior $\liminf_n x_n$ es el límite de dicha secuencia cuando $n \rightarrow \infty$. El límite inferior y el límite superior coinciden si y solo si la secuencia converge.

Ejemplo Sea $\Omega = \{0, 1\}$, con medida de probabilidad $P(\{0\}) = 1/3$ y $P(\{1\}) = 2/3$. El σ -álgebra de eventos es en este caso $\mathcal{F} = \{\emptyset, \{0\}, \{1\}, \Omega\}$ con probabilidades 0, 1/3, 2/3 y 1 respectivamente. Sea la secuencia de eventos

$$A_n = \begin{cases} \{0\}, & \text{si } n \text{ es par} \\ \{1\}, & \text{si } n \text{ es impar} \end{cases}.$$

Se calculará $P(\liminf_n A_n)$, $P(\limsup_n A_n)$, $\liminf_n P(A_n)$ y $\limsup_n P(A_n)$. Considerando que

$$\bigcup_{k=n}^{\infty} A_k = \{0, 1\}, \quad \bigcap_{k=n}^{\infty} A_k = \emptyset, \quad \forall n,$$

se tiene que

$$\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} \{0, 1\} = \{0, 1\} = \Omega,$$

y

$$\liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} \emptyset = \emptyset.$$

Por lo tanto,

$$P\left(\limsup_{n \rightarrow \infty} A_n\right) = 1, \quad P\left(\liminf_{n \rightarrow \infty} A_n\right) = 0.$$

Además, como

$$\sup_{k \geq n} P(A_k) = \frac{2}{3}, \quad \inf_{k \geq n} P(A_k) = \frac{1}{3}, \quad \forall n$$

se tiene que

$$\limsup_n P(A_n) = \frac{2}{3}, \quad \liminf_n P(A_n) = \frac{1}{3}.$$

Notar que se cumplen las desigualdades de los lemas de Fatou dadas por la ecuación 1.18.

Teorema *Continuidad de la medida de probabilidad:* si una secuencia de eventos $\{A_n\}_{n=1}^{\infty}$ tiene límite, entonces se cumple que

$$P\left(\lim_n A_n\right) = \lim_n P(A_n). \quad (1.19)$$

Demostración Como la secuencia tiene límite, se cumple que (ver sección 1.2.1)

$$\lim_n A_n = \limsup_n A_n = \liminf_n A_n$$

y por lo tanto,

$$P\left(\lim_n A_n\right) = P\left(\limsup_n A_n\right) = P\left(\liminf_n A_n\right).$$

Esto indica que las desigualdades de los lemas de Fatou de la ecuación 1.18 deben ser igualdades, y por lo tanto

$$\limsup_n P(A_n) = \liminf_n P(A_n) = \lim_n P(A_n),$$

concluyendo la prueba.

Se presenta a continuación el lema de Borel-Cantelli, que es la herramienta básica para demostrar convergencia casi segura.

Lema *Lema de Borel-Cantelli:* sea $\{A_n\}_{n=1}^{\infty}$ una secuencia de eventos tal que

$$\sum_n P(A_n) < \infty.$$

Entonces, se cumple que

$$P\left(\limsup_n A_n\right) = P(\{A_n \text{ i.o.}\}) = 0.$$

Esto significa que el conjunto de elementos ω del espacio muestral Ω que pertenecen a una cantidad infinita de eventos A_1, A_2, \dots tiene probabilidad 0 en Ω .

Demostración Se define la secuencia decreciente

$$G_n = \bigcup_{k=n}^{\infty} A_k.$$

Se cumple que

$$P(G_n) \searrow P\left(\limsup_n A_n\right),$$

ya que

$$\lim_n P(G_n) \stackrel{(a)}{=} P\left(\lim_n G_n\right) \stackrel{(b)}{=} P\left(\bigcap_{n=1}^{\infty} G_n\right) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = P\left(\limsup_n A_n\right),$$

donde en (a) se empleó la continuidad de la medida de probabilidad para secuencias decrecientes y en (b) que el límite de una secuencia decreciente existe y está dado por la ecuación 1.2. Por lo tanto, para todo m se cumple que

$$P\left(\limsup_n A_n\right) \leq P(G_m) = P\left(\bigcup_{k=m}^{\infty} A_k\right) \leq \sum_{k=m}^{\infty} P(A_k)$$

donde la última desigualdad proviene de la propiedad de sub-aditividad de la medida de probabilidad dada por la ecuación 1.11. Tomando $m \rightarrow \infty$ y considerando que el lado derecho es la cola de una serie convergente y por lo tanto

$$\lim_{m \rightarrow \infty} \sum_{k=m}^{\infty} P(A_k) = 0,$$

se concluye la prueba.

Ejemplo *Paradoja de Ross–Littlewood* [4]: supongase que se dispone de una urna infinitamente grande y una colección de infinitas bolillas etiquetadas con los números 1, 2, 3, etcétera. Considérese que se realiza el siguiente experimento: al minuto 1 antes del mediodía, las bolillas numeradas del 1 al 10 se introducen en la urna, y la bolilla de número 10 se extrae (asúmase que la extracción consume tiempo 0). Luego, a $1/2$ minuto antes del mediodía, se introducen las bolillas numeradas del 11 al 20 y se extrae la bolilla 20. A $1/4$ de minuto antes del mediodía, se introducen las bolillas numeradas del 21 al 30 y se extrae la bolilla 30, y así sucesivamente. Los tiempos en donde se ejecuta cada paso son así definidos para asegurar que una cantidad contable infinita de pasos se ejecutaron al mediodía. Una secuencia contable infinita de operaciones que se realizan en un tiempo finito se denomina *supertarea*. La pregunta de interés es cuantas bolillas hay en la urna al mediodía.

La respuesta a la pregunta claramente es que hay infinitas bolillas en la urna al mediodía, ya que toda bolilla cuyo número no sea de la forma $10n$, para todo $n \geq 1$, permanece en la urna.

Supóngase que ahora se modifica el experimento de la siguiente forma: al minuto 1 antes del mediodía, las bolillas numeradas del 1 al 10 se introducen en la urna, y la bolilla de número 1 se extrae. Luego, a $1/2$ minuto antes del mediodía, se introducen las bolillas numeradas del 11 al 20 y se extrae la bolilla 2. A $1/4$ de minuto antes del mediodía, se introducen las bolillas numeradas del 21 al 30 y se extrae la bolilla 3, y así sucesivamente. En este nuevo experimento, ¿cuantas bolillas hay en la urna al mediodía?

Tal vez sorprendentemente, la respuesta ahora es que la urna está vacía al mediodía. Efectivamente, considérese cualquier bolilla, la n -ésima por ejemplo. En algún momento previo al mediodía, esta bolilla será extraída. Concretamente, la bolilla n será extraída $(1/2)^{n-1}$ minutos antes del mediodía. Se concluye que para cada n , la bolilla n no estará en la urna al mediodía, y por lo tanto, la urna estará vacía a esa hora.

Como se ve, la diferencia de los resultados de los experimentos proviene de la forma en que se extraen las bolillas. Se considera a continuación el caso en que cuando se extrae una bolilla, se hace de forma aleatoria entre las bolillas presentes en la urna. Es decir, 1 minuto antes del mediodía, se introducen las bolillas numeradas del 1 al 10 y se extrae un bolilla al azar, y así sucesivamente. En este caso, ¿cuantas bolillas hay en la urna al mediodía?

Considérese la bolilla etiquetada con el número 1, y defínase E_n al evento que consiste en que la bolilla 1 está aún en la urna luego de n extracciones. Se comenzará calculando $P(E_n)$. Para esto, se define A_n como el evento que consiste que no se extrae la bolilla 1 en el paso n . De esta forma,

$$E_n = A_1 \cap A_2 \cap \cdots \cap A_n,$$

y como los eventos A_k son independientes (este concepto se verá mas adelante en el capítulo 2), se cumple que

$$P(E_n) = P(A_1) \times P(A_2) \times \cdots \times P(A_n) = \prod_{k=1}^n P(A_k).$$

Para calcular $P(A_k)$, se observa que en el paso k -ésimo luego de insertar las 10 bolillas y previo a la extracción hay $10k - (k - 1) = 9k + 1$ bolillas en la urna, ya que se insertaron 10 bolillas en cada paso hasta el paso k -ésimo y se sacó una en los $k - 1$ pasos previos. De esas $9k + 1$ bolillas, $9k$ no son la bolilla 1. Por lo tanto, la probabilidad de no extraer la bolilla en el paso k es

$$P(A_k) = \frac{9k}{9k + 1},$$

y por lo tanto,

$$P(E_n) = \prod_{k=1}^n \frac{9k}{9k + 1}.$$

Si bien los conceptos para llegar a este resultado no se han desarrollado aún en estos apuntes, el mismo puede deducirse con técnicas simple de conteo y la noción frecuentista de la probabilidad.

Ahora, la probabilidad de que la bolilla 1 está aún en la urna al mediodía es

$$P\{\text{bolilla 1 está en la urna al mediodía}\} = P\left(\lim_{n \rightarrow \infty} E_n\right).$$

Para continuar con el razonamiento es necesario notar que la secuencia de eventos E_n es decreciente. Efectivamente, que la bolilla 1 esté en la urna en el paso $n + 1$ implica que esté en la urna en el paso n ,

y por lo tanto,

$$E_{n+1} \subset E_n, \quad \forall n \geq 1.$$

De esta forma,

$$\begin{aligned} P\{\text{bolilla 1 está en la urna al mediodía}\} &= P\left(\lim_{n \rightarrow \infty} E_n\right) \\ &\stackrel{(a)}{=} \lim_{n \rightarrow \infty} P(E_n) \\ &= \lim_{n \rightarrow \infty} \prod_{k=1}^n \frac{9k}{9k+1} \\ &= \lim_{n \rightarrow \infty} \left[\prod_{k=1}^n \frac{9k+1}{9k} \right]^{-1} \\ &= \lim_{n \rightarrow \infty} \left[\prod_{k=1}^n \left(1 + \frac{1}{9k}\right) \right]^{-1}, \end{aligned}$$

donde en (a) se empleó la continuidad de la medida de probabilidad para secuencias de eventos decrecientes. Notando que para todo $n \geq 1$ se cumple que

$$\prod_{k=1}^n \left(1 + \frac{1}{9k}\right) = \left(1 + \frac{1}{9}\right) \left(1 + \frac{1}{18}\right) \cdots \left(1 + \frac{1}{9n}\right) > \frac{1}{9} + \frac{1}{18} + \cdots + \frac{1}{9n} = \sum_{k=1}^n \frac{1}{9k}$$

se obtiene que

$$\left[\prod_{k=1}^n \left(1 + \frac{1}{9k}\right) \right]^{-1} < \left[\frac{1}{9} \sum_{k=1}^n \frac{1}{k} \right]^{-1}.$$

Tomando el límite con $n \rightarrow \infty$, se tiene que

$$\lim_{n \rightarrow \infty} \left[\prod_{k=1}^n \left(1 + \frac{1}{9k}\right) \right]^{-1} < \lim_{n \rightarrow \infty} \left[\frac{1}{9} \sum_{k=1}^n \frac{1}{k} \right]^{-1} = 0$$

ya que $\sum_{k=1}^{\infty} 1/k$ es la serie armónica y por lo tanto, diverge. Se concluye que

$$P\{\text{bolilla 1 está en la urna al mediodía}\} = 0,$$

es decir, la probabilidad de la bolilla 1 esté en la urna luego de infinitos pasos es 0.

Defínase a F_i como el evento de que la bolilla número i esté en la urna al mediodía. Se acaba de demostrar que $P(F_1) = 0$. De forma análoga se puede demostrar que $P(F_i) = 0$ para todo i . Por ejemplo, puede verse que

$$P(F_i) = \lim_{n \rightarrow \infty} \prod_{k=2}^n \frac{9k}{9k+1}, \quad \text{para } i = 11, 12, \dots, 20,$$

y mas en general, se tiene que

$$P(F_i) = \lim_{n \rightarrow \infty} \prod_{k=m}^n \frac{9k}{9k+1}, \quad \text{para } i = 10m-9, \dots, 10m-1, 10m.$$

El rango de i corresponde a la numeración de las bolillas introducidas en el paso m . A partir de esta expresión y siguiendo el mismo razonamiento que para F_1 , se obtiene que $P(F_i) = 0$ para todo i .

Finalmente, la probabilidad que la urna no esté vacía al mediodía es

$$P\left(\bigcup_{i=1}^{\infty} F_i\right) \leq \sum_{i=1}^{\infty} P(F_i) = 0,$$

donde se empleó la desigualdad de Boole (ecuación 1.11). Por lo tanto, la urna estará vacía al mediodía con probabilidad 1.

1.3.4. Probabilidad condicional

La probabilidad condicional de un evento A asumiendo la ocurrencia de otro evento M se define como

$$P(A|M) = \frac{P(AM)}{P(M)}, \quad (1.20)$$

donde se asume que $P(M) \neq 0$. Es referida como “la probabilidad de A dado M ”.

Se puede demostrar que para un evento M específico, la probabilidad condicional $P(A|M)$ es efectivamente una probabilidad, ya que cumple todas las propiedades de una medida de probabilidad. De esta forma, todos los resultados que involucran probabilidades se cumplen también para probabilidades condicionales.

Probabilidad total y el teorema de Bayes En el capítulo 2 de [1] se demuestra el teorema de la *probabilidad total*, que indica que si $U = [A_1, A_2, \dots, A_n]$ es una partición de S y B un evento arbitrario, se cumple que

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \dots + P(B|A_n)P(A_n). \quad (1.21)$$

Además, de 1.20 se tiene que

$$P(BA_i) = P(A_i|B)P(B) = P(B|A_i)P(A_i)$$

y por lo tanto

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}. \quad (1.22)$$

Sustituyendo la ecuación 1.21 en este último resultado, se obtiene el *teorema de Bayes*,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + \dots + P(B|A_n)P(A_n)}. \quad (1.23)$$

Usualmente se emplean los términos *a priori* y *a posteriori* para referirse a las probabilidades $P(A_i)$ y $P(A_i|B)$ respectivamente.

1.3.5. Eventos independientes

Se dice que los eventos A y B son *independientes* si se cumple que

$$P(AB) = P(A)P(B). \quad (1.24)$$

Se puede demostrar que si los eventos A y B son independientes, los eventos \bar{A} y B y los eventos \bar{A} y \bar{B} también son independientes.

Lema Segundo lema de Borel-Cantelli: sea $\{A_n\}_{n=1}^{\infty}$ una secuencia de eventos *mutuamente independientes* tal que

$$\sum_n P(A_n) = \infty.$$

Entonces, se cumple que

$$P\left(\limsup_n A_n\right) = P(\{A_n \text{ i.o.}\}) = 1.$$

Esto significa que si la suma de las probabilidades de una secuencia infinita de eventos independientes diverge, el conjunto de elementos ω del espacio muestral Ω que pertenece a infinitos eventos A_1, A_2, \dots tiene probabilidad 1 en Ω , o equivalentemente, el evento $\{\omega \in \Omega : \omega \in A_n \text{ i.o.}\}$ ocurre *casi seguramente* (ver sección 6.4.1).

Demostración Se parte observando que

$$\left(\limsup_{n \rightarrow \infty} A_n \right)^c = \left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \right)^c = \bigcup_{n=1}^{\infty} \left(\bigcup_{k=n}^{\infty} A_k \right)^c = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c,$$

donde las identidades provienen de aplicar las leyes de Morgan. Por lo tanto, se cumple que

$$\begin{aligned} P \left[\left(\limsup_{n \rightarrow \infty} A_n \right)^c \right] &= P \left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c \right) \\ &\leq \sum_{n=1}^{\infty} P \left(\bigcap_{k=n}^{\infty} A_k^c \right), \end{aligned} \quad (1.25)$$

donde se empleó la desigualdad de Boole, dada por la ecuación 1.11. Además, como los eventos de la secuencia $\{A_n\}_{n=1}^{\infty}$ son mutuamente independientes, los eventos de la secuencia complementaria $\{A_n^c\}_{n=1}^{\infty}$ también son mutuamente independientes, y por lo tanto

$$\begin{aligned} P \left(\bigcap_{k=n}^{\infty} A_k^c \right) &= \prod_{k=n}^{\infty} P(A_k^c) \\ &= \prod_{k=n}^{\infty} [1 - P(A_k)] \\ &\stackrel{(a)}{\leq} \prod_{k=n}^{\infty} e^{-P(A_k)} \\ &= \exp \left[- \sum_{k=n}^{\infty} P(A_k) \right] \\ &\stackrel{(b)}{=} 0, \end{aligned}$$

donde en (a) se empleó que $1 - x \leq e^{-x}$ para todo x y en (b) la hipótesis $\sum_n P(A_n) = \infty$. Combinando este resultado con el de la ecuación 1.25, se obtiene que

$$P \left[\left(\limsup_{n \rightarrow \infty} A_n \right)^c \right] \leq \sum_{n=1}^{\infty} P \left(\bigcap_{k=n}^{\infty} A_k^c \right) = 0.$$

y por lo tanto,

$$P \left(\limsup_{n \rightarrow \infty} A_n \right) = 1.$$

Ejemplo Un ensayo de Bernoulli consiste en la repetición de un experimento que tiene los dos resultados posibles h y t , con $P\{h\} = p$ y $P\{t\} = 1 - p$, como se explica mas detalladamente en la sección 2.2. El resultado de cada experimento es independiente del resultado de los demás experimentos. Un ejemplo de ensayo de Bernoulli, es lanzar al aire una moneda repetidamente.

Considérese el evento “ $hh \dots h$ ” en un ensayo de Bernoulli. Se quiere determinar la probabilidad de que dicha secuencia de n resultados h sucesivos ocurra infinitamente a menudo. Para hacerlo, se considera el evento $A_k = \{\text{el resultado en el } k\text{-ésimo experimento es } h\}$, y se define el evento B_i como

$$B_i = A_i \cap A_{i+1} \cap \dots \cap A_{i+n-1}, \quad i \geq 1,$$

que es la ocurrencia de n resultados h sucesivos entre el experimento i -ésimo y el experimento $i+n-1$. Puede demostrarse (ver sección 2.2) que $P(B_i) = p^n$. Los eventos B_i no son independientes, pero los eventos $B_1, B_{n+1}, B_{2n+1}, \dots$ son independientes, y además

$$\sum_{k=0}^{\infty} P(B_{kn+1}) = \sum_{k=0}^{\infty} p^n = \infty.$$

Por lo tanto, por el segundo lema de Borel-Cantelli, el patrón “ $hh \dots h$ ”, así como cualquier otro patrón arbitrario, ocurrirá infinitamente a menudo con probabilidad 1.

Capítulo 2

Ensayos repetidos

Se incluyen en este capítulo algunos conceptos desarrollados en el capítulo 3 de [1].

2.1. Experimentos combinados

Se consideran dos experimentos. El primero consiste en tirar un dado,

$$S_1 = \{f_1, \dots, f_6\} \quad P_1\{f_i\} = \frac{1}{6},$$

y el segundo en lanzar una moneda,

$$S_2 = \{h, t\} \quad P_2\{h\} = P_2\{t\} = \frac{1}{2}.$$

Se realizan los dos experimentos y se quiere encontrar la probabilidad de obtener “dos” (f_2) en el dado y “cara” (h) en la moneda. Si se asume la hipótesis razonable de que el resultado del primer experimento es independiente del segundo experimento, según la ecuación 1.24 esta probabilidad es $1/6 \times 1/2$. Sin embargo, en la definición de la ecuación 1.24 se asume que A y B son subconjuntos del mismo espacio muestral. Por lo tanto, para adaptar esta conclusión al caso del ejemplo considerado, hay que construir un espacio muestral S que contenga como subconjuntos los eventos “dos” y “cara”. Esto se explica a continuación.

Ambos experimentos son vistos como un único experimento cuyas salidas son pares $\zeta_1\zeta_2$, donde ζ_1 es una de las seis caras del dado y ζ_2 es uno de los dos resultados de la moneda. De esta forma, el espacio muestral S del experimento contiene 12 elementos,

$$S = \{f_1h, \dots, f_6h, f_1t, \dots, f_6t\}.$$

En este espacio, $\{\text{dos}\}$ no es un evento elemental, si no un subconjunto que contiene dos elementos,

$$\{\text{dos}\} = \{f_2h, f_2t\}.$$

Análogamente, $\{\text{cara}\}$ es un evento con seis elementos,

$$\{\text{cara}\} = \{f_1h, \dots, f_6h\}.$$

Para especificar completamente al experimento, hay que asignar probabilidades a todos los subconjuntos de S . Como el evento $\{\text{dos}\}$ ocurre si en el dado sale “dos” independientemente del resultado de la moneda,

$$P\{\text{dos}\} = \frac{1}{6}.$$

De igual forma, el evento $\{\text{cara}\}$ ocurre si sale “cara” en la moneda sin importar el resultado del dado, y por lo tanto

$$P\{\text{cara}\} = \frac{1}{2}.$$

Además, el evento {dos} es un resultado del primer experimento y el evento {cara} es resultado del segundo experimento, y por lo tanto es razonable asumirlos eventos independientes en el experimento combinado. La intersección de los eventos {dos} y {cara} es el evento elemental $\{f_2h\}$ de S . Por lo tanto, se aplica la ecuación 1.24 y se concluye que

$$P\{f_2h\} = P\{\text{dos}\} \times P\{\text{cara}\} = \frac{1}{6} \times \frac{1}{2} = \frac{1}{12}.$$

2.1.1. Producto cartesiano de conjuntos

Sean dos conjuntos S_1 y S_2 con elementos ζ_1 y ζ_2 respectivamente. El *producto cartesiano* de los conjuntos S_1 y S_2 es un conjunto S cuyos elementos son todos los pares ordenados $\zeta_1\zeta_2$, donde ζ_1 es cualquier elemento de S_1 y ζ_2 es cualquier elemento de S_2 . Se denota como

$$S = S_1 \times S_2.$$

Si A es un subconjunto de S_1 y B es un subconjunto de S_2 , el conjunto

$$C = A \times B,$$

que consiste en todos los pares $\zeta_1\zeta_2$, con $\zeta_1 \in A$ y $\zeta_2 \in B$, es un subconjunto de S . Considérense los conjuntos $A \times S_2$ y $S_1 \times B$. Claramente su intersección es el conjunto $A \times B$,

$$A \times B = (A \times S_2) \cap (S_1 \times B). \quad (2.1)$$

Ejemplo Supóngase que S_1 es el eje x , S_2 es el eje y , y A y B son los intervalos

$$A = \{x_1 \leq x \leq x_2\}, \quad b = \{y_1 \leq y \leq y_2\},$$

como se muestra en la figura 2.1. En ese caso, $A \times S_2$ es una banda vertical, $S_1 \times B$ es una banda horizontal, y $A \times B$ es un rectángulo.

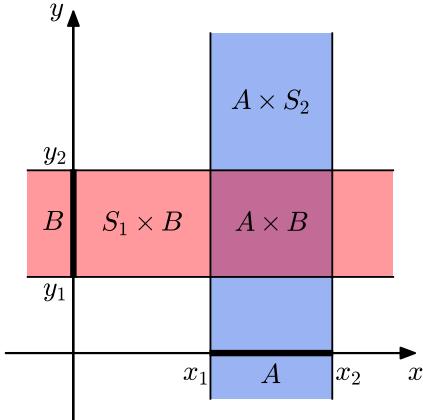


Figura 2.1: Conjuntos $A \times S_2$, $S_1 \times B$ y $A \times B$ cuando S_1 es el eje x , S_2 es el eje y y A y B son intervalos en el eje x e y respectivamente.

2.1.2. Producto cartesiano de dos experimentos

El producto cartesiano de dos experimentos S_1 y S_2 es un nuevo experimento $S = S_1 \times S_2$ cuyos eventos son todos los productos cartesianos de la forma

$$A \times B,$$

donde A es un evento de S_1 y B es un evento de S_2 , y sus uniones y sus intersecciones. En este experimento las probabilidades de los eventos $A \times S_2$ y $S_1 \times B$ son

$$P(A \times S_2) = P_1(A), \quad P(S_1 \times B) = P_2(B), \quad (2.2)$$

donde $P_1(A)$ es la probabilidad del evento A en el experimento S_1 y $P_2(B)$ es la probabilidad del evento B en el experimento S_2 . Efectivamente, el evento $A \times S_1$ del experimento S ocurre si el evento A del experimento S_1 ocurre sin importar el resultado del experimento S_2 . De la misma forma, el evento $S_1 \times B$ del experimento S ocurre si el evento B del experimento S_2 ocurre sin importar el resultado del experimento S_1 .

La ecuación 2.2 determina únicamente la probabilidad de los eventos $A \times S_2$ y $S_1 \times B$. La probabilidad de un evento arbitrario de la forma $A \times B$ y la de sus uniones e intersecciones no pueden en general expresarse en función de P_1 y P_2 . Para determinar estas probabilidades, se necesita más información de los experimentos S_1 y S_2 .

Experimentos independientes

En muchas aplicaciones, los eventos $A \times S_2$ y $S_1 \times B$ del experimento combinado son independientes para cualquier A y B . Como la intersección de esos eventos es $A \times B$, se tiene que

$$P(A \times B) \stackrel{(a)}{=} P((A \times S_2) \cap (B \times S_1)) \stackrel{(b)}{=} P(A \times S_2)P(S_1 \times B) \stackrel{(c)}{=} P_1(A)P_2(B),$$

donde en (a) se empleó la ecuación 2.1, en (b) se empleó la ecuación 1.24 y en (c) la ecuación 2.2. Esto completa la especificación del experimento S en el caso en que los experimentos S_1 y S_2 son independientes, ya que todos los eventos de S son uniones e intersecciones de eventos de la forma $A \times B$.

Notar que en particular el evento elemental $\{\zeta_1 \zeta_2\}$ puede expresarse como el producto cartesiano $\{\zeta_1\} \times \{\zeta_2\}$ de los eventos elementales $\{\zeta_1\}$ y $\{\zeta_2\}$ de S_1 y S_2 . Por lo tanto,

$$P\{\zeta_1 \zeta_2\} = P\{\zeta_1\}P\{\zeta_2\}.$$

2.2. Ensayos de Bernoulli

Un conjunto de n objetos puede ser ordenado de varias formas distintas obteniendo *permutaciones*. Por ejemplo, las posibles permutaciones de los tres objetos a, b, c son

$$abc \quad bac \quad bca \quad acb \quad cab \quad cba.$$

Como se ve, hay 6 permutaciones distintas de 3 objetos. En general, dado un conjunto de n objetos, el de la primera posición se puede elegir de n formas posibles, el de la segunda posición de $n - 1$ formas posibles, ya que se sacó uno previamente y quedan $n - 1$ objetos, y así sucesivamente. De esta forma, el número de permutaciones de n objetos es $n(n - 1)(n - 2) \cdots 3 \cdot 2 \cdot 1 = n!$.

Se consideran ahora permutaciones de $k < n$ objetos tomados de un conjunto de n objetos, prestando atención al orden de los objetos en cada subconjunto. Igual que antes, el objeto de la primera posición puede elegirse entre n objetos, el de la segunda posición, entre los $n - 1$ objetos restantes, y así sucesivamente hasta el objeto de la posición k -ésima, que puede seleccionarse de entre $n - k + 1$ objetos. Por lo tanto, el número total de subconjuntos de n objetos tomando k es

$$n(n - 1)(n - 2) \cdots (n - k + 1) = \frac{n!}{(n - k)!}.$$

Por ejemplo, tomando dos objetos del conjunto de tres objetos a, b, c , se tienen las permutaciones

$$ab \quad ba \quad ac \quad ca \quad bc \quad cb,$$

El total de permutaciones es $3!/(3 - 2)! = 6$. Notar que se consideró que por ejemplo, los subconjuntos ab y ba son distintos, es decir, se tuvo en cuenta el orden de los objetos en los subconjuntos.

Supóngase ahora que los k objetos son tomados de un conjunto de n objetos pero sin tener en cuenta el orden, es decir, los subconjuntos ab y ba se consideran el mismo. Estos subconjuntos son las *combinaciones* de n objetos tomando k objetos. En este caso, las $k!$ permutaciones generadas por cada grupo de k objetos contribuyen únicamente en una combinación. Por lo tanto, las combinaciones de n objetos tomando k objetos es

$$\frac{n(n - 1)(n - 2) \cdots (n - k + 1)}{k!} = \frac{n!}{(n - k)!k!} = \binom{n}{k}.$$

El numerador en la primera igualdad es la cantidad de subconjuntos de k objetos teniendo en cuenta el orden, y el denominador se incluye para considerar que las $k!$ permutaciones de los objetos de cada subconjunto contribuye solo en una combinación. Por ejemplo, si un conjunto tiene 4 elementos, la cantidad de subconjuntos de dos elementos que se pueden formar es

$$\binom{4}{2} = \frac{4!}{(4-2)!2!} = \frac{4 \cdot 3}{2} = 6.$$

Efectivamente, los subconjuntos de dos elementos formados del conjunto de cuatro elementos $abcd$ son

$$ab \quad ac \quad ad \quad bc \quad bd \quad cd.$$

Este resultado será empleado para obtener la probabilidad de que un evento ocurra k veces en n ensayos independientes de un experimento S . Como se ilustra en el siguiente ejemplo, el problema es esencialmente el mismo que calcular la probabilidad de obtener k caras al tirar una moneda n veces.

Ejemplo (3-6) La probabilidad de obtener cara al tirar una moneda es $P\{h\} = p$. Se probará que la probabilidad de obtener cara k veces al tirar la moneda n veces es

$$p_n(k) = \binom{n}{k} p^k q^{n-k}, \quad q = 1 - p. \quad (2.3)$$

El experimento en consideración es tirar una moneda n veces. El resultado del experimento es una secuencia particular de caras (h) y números (t). El evento $\{k$ caras en cualquier orden $\}$ consiste en todas las secuencias que contienen k caras y $n - k$ números. Para obtener la cantidad de secuencias distintas de n objetos que contienen k caras y $n - k$ números, hay que notar que si los objetos fueran todos distintos, habría $n!$ secuencias distintas. Sin embargo, como las k caras y los $n - k$ números son idénticos entre sí, las correspondientes $k!$ permutaciones entre las caras y las $(n - k)!$ entre los números de una secuencia específica contribuyen solo a una secuencia. Por lo tanto, el número total de secuencias distintas (combinaciones) con k caras y $n - k$ números es

$$\frac{n!}{k!(n-k)!} = \binom{n}{k}, \quad (2.4)$$

es decir, el evento $\{k$ caras en cualquier orden $\}$ consiste en $\binom{n}{k}$ eventos elementales conteniendo k caras y $n - k$ números en un orden específico. Como la probabilidad de cada uno de esos eventos elementales es $p^k q^{n-k}$, se concluye que

$$P\{k \text{ caras en cualquier orden}\} = \binom{n}{k} p^k q^{n-k}.$$

Como ejemplo, considérese el caso en que $n = 3$ y $k = 2$, y $p = q = 0.5$, es decir, los eventos cara y numero son equiprobables. Considerando el orden, hay $2^n = 2^3 = 8$ secuencias posibles (permutaciones con repetición), que son

$$hhh \quad hht \quad hth \quad thh \quad tth \quad tht \quad htt \quad ttt$$

y tres formas de obtener $k = 2$ caras, que son hht , hth , thh . Por lo tanto, como los ocho elementos del espacio muestral son equiprobables, la probabilidad de obtener dos caras es

$$P\{2 \text{ caras en cualquier orden}\} = \frac{\text{casos favorables}}{\text{casos posibles}} = \frac{3}{8}.$$

Realizando el mismo cálculo usando la ecuación 2.3 se tiene que

$$p_3(2) = \binom{3}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right) = \frac{3!}{2!1!} \left(\frac{1}{2}\right)^3 = \frac{3}{8}.$$

Finalmente, se calculará la probabilidad

$$P\{k_1 \leq k \leq k_2\}$$

de que el número de ocurrencias del evento A (obtención de cara) esté entre k_1 y k_2 . Para hacerlo, se observa que los eventos $\{A \text{ ocurre } k \text{ veces}\}$ con k tomando cualquier valor entre k_1 y k_2 , son mutuamente excluyentes y su unión es el evento $\{k_1 \leq k \leq k_2\}$. Por lo tanto, de la ecuación 2.3, se obtiene que

$$P\{k_1 \leq k \leq k_2\} = \sum_{k=k_1}^{k_2} p_n(k) = \sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k}. \quad (2.5)$$

2.2.1. Generalización de ensayos de Bernoulli

La generalización del teorema de la ecuación 2.3 cuando cada ensayo tiene mas de dos resultados posibles puede encontrarse en la sección 4-5 de [1] y se plantea como sigue.

Teorema Supóngase que

$$U = [A_1, A_2, \dots, A_r]$$

es una partición de S que consiste en r eventos A_i con

$$P(A_i) = p_i, \quad p_1 + p_2 + \dots + p_r = 1.$$

Se repite el experimento n veces y se denota la probabilidad del evento $\{A_1 \text{ ocurre } k_1 \text{ veces}, A_2 \text{ ocurre } k_2 \text{ veces}, \dots, A_r \text{ ocurre } k_r \text{ veces en cualquier orden}\}$ como $p_n(k_1, k_2, \dots, k_r)$, donde $k_1 + k_2 + \dots + k_r = n$. Se cumple que

$$p_n(k_1, k_2, \dots, k_r) = \frac{n!}{k_1! k_2! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r} \quad (2.6)$$

Demostración Para demostrarlo, se observa que usando el mismo argumento que en caso de ensayos con dos resultados posibles (ecuación 2.4), la cantidad de eventos de la forma $\{A_1 \text{ ocurre } k_1 \text{ veces}, \dots, A_r \text{ ocurre } k_r \text{ veces en un orden específico}\}$ es

$$\frac{n!}{k_1! k_2! \dots k_r!}.$$

Además, como cada ensayo es independiente, la probabilidad de cada uno de esos eventos es

$$p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}.$$

2.3. Teorema 3.1: Teorema de Bernoulli

El teorema de Bernoulli establece que la frecuencia relativa de sucesos en una secuencia de ensayos Bernoulli se acerca indefinidamente a la probabilidad de suceso del ensayo a medida que el número de ensayos crece. Es una versión simplificada de la *ley de los grandes números* y también se conoce como la *ley débil de los grandes números*. El teorema fue planteado y demostrado por Jacob Bernoulli en 1713 [1], pero la demostración que se incluye aquí fue propuesta por Chebyshev.

Teorema Sea A un evento cuya probabilidad de ocurrencia en un único ensayo es p . Si k es el número de ocurrencias en n ensayos independientes, se cumple que

$$P\left(\left|\frac{k}{n} - p\right| > \epsilon\right) < \frac{pq}{n\epsilon^2}. \quad (2.7)$$

Demostración La probabilidad $p_n(k)$ de k ocurrencias del evento A en n ensayos está dada por la ecuación 2.3 y es

$$p_n(k) = \binom{n}{k} p^k q^{n-k},$$

donde $q = 1 - p$ es la probabilidad de no ocurrencia del evento en un único ensayo. El evento A es una variable aleatoria Binomial. Se hará uso de las siguientes identidades,

$$\sum_{k=0}^n kp_n(k) = np \quad \text{y} \quad \sum_{k=0}^n k^2 p_n(k) = n^2 p^2 + npq. \quad (2.8)$$

que se demuestran a continuación. En cuanto a la primera identidad, se tiene que

$$\begin{aligned} \sum_{k=0}^n kp_n(k) &\stackrel{(a)}{=} \sum_{k=1}^n k \binom{n}{k} p^k q^{n-k} \\ &\stackrel{(b)}{=} \sum_{k=1}^n k \frac{n!}{(n-k)!k!} p^k q^{n-k} \\ &\stackrel{(c)}{=} \sum_{k=1}^n \frac{n!}{(n-k)!(k-1)!} p^k q^{n-k} \\ &\stackrel{(d)}{=} \sum_{i=0}^{n-1} \frac{n!}{(n-i-1)!i!} p^{i+1} q^{n-i-1} \\ &= np \sum_{i=0}^{n-1} \frac{(n-1)!}{(n-i-1)!i!} p^i q^{n-i-1} \\ &= np \sum_{i=0}^{n-1} \binom{n-1}{i} p^i q^{n-1-i} \\ &\stackrel{(e)}{=} np(p+q)^{n-1} \\ &\stackrel{(f)}{=} np, \end{aligned}$$

donde en (a) se usó la ecuación 2.3, en (b) se usó la definición de combinaciones, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, en (c) se usó que $k! = k(k-1)!$, en (d) se realizó el cambio de variable $i = k-1$, en (e) se usó el teorema del binomio¹, que establece que

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k, \quad (2.9)$$

y en (f) se usó que $p+q=1$. Procediendo de forma similar con la segunda igualdad, se ve que

$$\sum_{k=0}^n k^2 p_n(k) = \sum_{k=1}^n k \frac{n!}{(n-k)!(k-1)!} p^k q^{n-k}$$

Observando que

$$\frac{1}{(k-2)!} + \frac{1}{(k-1)!} = \frac{k-1}{(k-1)!} + \frac{1}{(k-1)!} = \frac{k}{(k-1)!},$$

se tiene que el factor $\frac{k}{(k-1)!}$ de la sumatoria se puede expresar como

$$\frac{k}{(k-1)!} = \begin{cases} \frac{1}{(k-1)!}, & \text{si } k=1 \\ \frac{1}{(k-2)!} + \frac{1}{(k-1)!}, & \text{si } k \geq 2 \end{cases},$$

y por lo tanto

$$\sum_{k=0}^n k^2 p_n(k) = \sum_{k=2}^n \frac{n!}{(n-k)!(k-2)!} p^k q^{n-k} + \sum_{k=1}^n \frac{n!}{(n-k)!(k-1)!} p^k q^{n-k}.$$

¹ver por ejemplo: https://en.wikipedia.org/wiki/Binomial_theorem

El segundo término de la suma a la derecha de la igualdad, coincide con la identidad demostrada previamente y por lo tanto vale np . Desarrollando el primer término de la suma, se tiene que

$$\begin{aligned}
\sum_{k=2}^n \frac{n!}{(n-k)!(k-2)!} p^k q^{n-k} &\stackrel{(a)}{=} \sum_{i=0}^{n-2} \frac{n!}{(n-i-2)!i!} p^{i+2} q^{n-i-2} \\
&= n(n-1)p^2 \sum_{i=0}^{n-2} \frac{(n-2)!}{(n-i-2)!i!} p^i q^{n-i-2} \\
&\stackrel{(b)}{=} n(n-1)p^2 \sum_{i=0}^{n-2} \binom{n-2}{i} p^i q^{n-i-2} \\
&\stackrel{(c)}{=} n(n-1)p^2(p+q)^{n-2} \\
&= n(n-1)p^2 \\
&= n^2 p^2 - np^2 \\
&= n^2 p^2 - np(1-q) \\
&= n^2 p^2 - np + npq,
\end{aligned}$$

donde en (a) se realizó el cambio de variable $i = k - 2$, en (b) se usó la definición de combinaciones y en (c) se usó el teorema del binomio. Finalmente,

$$\begin{aligned}
\sum_{k=0}^n k^2 p_n(k) &= \sum_{k=2}^n \frac{n!}{(n-k)!(k-2)!} p^k q^{n-k} + \sum_{k=1}^n \frac{n!}{(n-k)!(k-1)!} p^k q^{n-k} \\
&= n^2 p^2 - np + npq + np \\
&= n^2 p^2 + npq,
\end{aligned}$$

con lo que queda demostrada la segunda identidad. Habiendo demostrado ambas identidades, a continuación se procede a demostrar el teorema de Bernoulli dado por la ecuación 2.7. Partiendo de la hipótesis,

$$\left| \frac{k}{n} - p \right| > \epsilon \iff (k - np)^2 > n^2 \epsilon^2 \quad (2.10)$$

Multiplicando ambos lados de la desigualdad por $p_n(k)$ y sumando para todo k , se tiene que

$$\sum_{k=0}^n (k - np)^2 p_n(k) > \sum_{k=0}^n n^2 \epsilon^2 p_n(k) = n^2 \epsilon^2 \sum_{k=0}^n p_n(k) = n^2 \epsilon^2, \quad (2.11)$$

donde se usó que

$$\sum_{k=0}^n p_n(k) = 1, \quad (2.12)$$

ya que se trata de la suma de probabilidades de eventos mutuamente excluyentes que abarcan todo el espacio muestral. Desarrollando el lado izquierdo de la desigualdad, se tiene que

$$\begin{aligned}
\sum_{k=0}^n (k - np)^2 p_n(k) &= \sum_{k=0}^n (k^2 - 2knp + n^2 p^2) p_n(k) \\
&= \sum_{k=0}^n k^2 p_n(k) - 2np \sum_{k=0}^n kp_n(k) + n^2 p^2 \sum_{k=0}^n p_n(k) \\
&\stackrel{(a)}{=} n^2 p^2 + npq - 2np \cdot np + n^2 p^2 \\
&= npq,
\end{aligned} \quad (2.13)$$

donde en (a) se usaron las identidades de la ecuación 2.8 y la ecuación 2.12. De forma alternativa, el lado izquierdo de la desigualdad puede expresarse como

$$\begin{aligned}
\sum_{k=0}^n (k - np)^2 p_n(k) &= \sum_{|k-np| \leq n\epsilon} (k - np)^2 p_n(k) + \sum_{|k-np| > n\epsilon} (k - np)^2 p_n(k) \\
&\stackrel{(a)}{\geq} \sum_{|k-np| > n\epsilon} (k - np)^2 p_n(k) \\
&\stackrel{(b)}{>} n^2 \epsilon^2 \sum_{|k-np| > n\epsilon} p_n(k) \\
&= n^2 \epsilon^2 P(|k - np| > n\epsilon),
\end{aligned} \tag{2.14}$$

donde en (a) se tuvo en cuenta que todos los sumandos de la sumatoria son positivos y en (b) se usó la hipótesis, dada por la ecuación 2.10. Combinando las ecuaciones 2.13 y 2.14, se llega a que

$$n^2 \epsilon^2 P(|k - np| > n\epsilon) < npq,$$

es decir,

$$P\left(\left|\frac{k}{n} - p\right| > \epsilon\right) < \frac{pq}{n\epsilon^2}$$

2.4. Ejemplo 3.17: Estrategia para un juego desfavorable

Se considera el caso en que dos jugadores A y B juegan en una serie de partidas en donde en cada partida la probabilidad de ganar es desfavorable para A , es decir, menor que $1/2$. Sin embargo, A puede elegir la cantidad de partidas a jugar. Gana el juego el jugador que gana mas partidas. Si la cantidad de partidas de un juego es par, cual es el número de partidas que debe elegir A ?

2.4.1. Solución

En cada partida, A gana con probabilidad p y B gana con probabilidad $q = 1 - p > p$. En principio se podría pensar que la mejor estrategia para A es retirarse lo antes posible. Si A tiene que jugar un número par de partidas, tal vez lo mejor sería retirarse luego de dos partidas. Esto es cierto si p es extremadamente pequeño. Sin embargo, si p es solo un poco menor a $1/2$, la estrategia óptima de A puede ser jugar cierto número de partidas mayor a dos, como se verá a continuación.

Sea X_k el evento “ A gana k partidas en un juego de $2n$ partidas”. Como X_k es una variable aleatoria binomial, $X_k \sim B(k, 2n)$, se cumple que

$$P(X_k) = \binom{2n}{k} p^k q^{2n-k}, \quad k = 0, 1, 2, \dots, 2n.$$

Para que A gane un juego de $2n$ partidas, tiene que ganar al menos $n + 1$ partidas. Por lo tanto, si se denota P_{2n} a la probabilidad de que A gane un juego de $2n$ partidas, se tiene que

$$\begin{aligned}
P_{2n} &= P\left(\bigcup_{k=n+1}^{2n} X_k\right) \\
&\stackrel{(a)}{=} \sum_{k=n+1}^{2n} P(X_k) \\
&= \sum_{k=n+1}^{2n} \binom{2n}{k} p^k q^{2n-k},
\end{aligned} \tag{2.15}$$

donde en (a) se consideró que los eventos X_i son mutuamente excluyentes.

Si $2n$ fuera efectivamente el número óptimo de partidas, se debería cumplir que

$$P_{2n-2} \leq P_{2n} \geq P_{2n+2},$$

donde P_{2n-2} y P_{2n+2} es la probabilidad de que A gane en un juego de $2n-2$ y de $2n+2$ partidas respectivamente. Análogamente, para que A gane un juego de $2n+2$ partidas, debe ganar al menos $\frac{2n+2}{2} + 1 = n+2$ partidas, y por lo tanto,

$$P_{2n+2} = \sum_{k=n+2}^{2n+2} \binom{2n-2}{k} p^k q^{2n+2-k}. \quad (2.16)$$

Para intentar establecer una relación entre las probabilidades de las ecuaciones 2.15 y 2.16, se comienza observando que

$$\begin{aligned} \sum_{k=0}^{2n+2} \binom{2n+2}{k} p^k q^{2n+2-k} &\stackrel{(a)}{=} (p+q)^{2n+2} \\ &= (p+q)^{2n} (p+q)^2 \\ &\stackrel{(b)}{=} \left\{ \sum_{k=0}^{2n} \binom{2n}{k} p^k q^{2n-k} \right\} (p^2 + 2pq + q^2), \end{aligned} \quad (2.17)$$

donde en (a) y en (b) se usó el teorema del binomio. Notar que si se consideran solo los sumandos desde $n+2$ en la sumatoria de la izquierda de la igualdad, el resultado es P_{2n+2} y si se consideran solo los sumandos desde $n+1$ en la sumatoria de la última igualdad, el resultado es P_{2n} . Se considera por un lado P_{2n+2} , y expandiendo la ecuación 2.16 se tiene que

$$\begin{aligned} P_{2n+2} &= \sum_{k=n+2}^{2n+2} \binom{2n+2}{k} p^k q^{2n+2-k} \\ &= \binom{2n+2}{n+2} p^{n+2} q^n + \binom{2n+2}{n+3} p^{n+3} q^{n-1} + \cdots + \binom{2n+2}{2n+1} p^{2n+1} q + \binom{2n+2}{2n+2} p^{2n+2}. \end{aligned} \quad (2.18)$$

Haciendo lo mismo con P_{2n} dado por la ecuación 2.15, se ve que

$$\begin{aligned} P_{2n} &= \sum_{k=n+1}^{2n} \binom{2n}{k} p^k q^{2n-k} \\ &= \binom{2n}{n+1} p^{n+1} q^{n-1} + \binom{2n}{n+2} p^{n+2} q^{n-2} + \cdots + \binom{2n}{2n-1} p^{2n-1} q + \binom{2n}{2n} p^{2n}, \end{aligned}$$

y por lo tanto, al multiplicar por $(p^2 + 2pq + q^2)$ como establece la ecuación 2.17, se obtiene que

$$\begin{aligned} P_{2n}(p^2 + 2pq + q^2) &= \binom{2n}{n+1} p^{n+3} q^{n-1} + \binom{2n}{n+2} p^{n+4} q^{n-2} + \cdots + \binom{2n}{2n-1} p^{2n+1} q + \binom{2n}{2n} p^{2n+2} \\ &\quad + 2 \left[\binom{2n}{n+1} p^{n+2} q^n + \binom{2n}{n+2} p^{n+3} q^{n-1} + \cdots + \binom{2n}{2n-1} p^{2n} q^2 + \binom{2n}{2n} p^{2n+1} q \right] \\ &\quad + \binom{2n}{n+1} p^{n+1} q^{n+1} + \binom{2n}{n+2} p^{n+2} q^n + \cdots + \binom{2n}{2n-1} p^{2n-1} q^3 + \binom{2n}{2n} p^{2n} q^2, \end{aligned}$$

donde la primer fila es $p^2 P_{2n}$, la segunda es $2pq P_{2n}$ y la tercera es $q^2 P_{2n}$. Notar que en cada fila hay un sumando con factor p y q elevados a la misma potencia. Por ejemplo, el primer sumando de la primera fila, el segundo de la segunda fila y el tercero de la tercera fila tienen el factor común $p^{n+3} q^{n-1}$. Estos factores de potencias en p y q coinciden además con los factores de P_{2n+2} . Por ejemplo, el segundo sumando de P_{2n+2} tiene como factor $p^{n+3} q^{n-1}$, como se ve en la ecuación 2.18. Continuando con la

ecuación anterior, al sacar estos términos de factor común, se tiene que

$$\begin{aligned}
P_{2n}(p^2 + 2pq + q^2) &= \sum_{k=n+1}^{2n-2} p^{k+2} q^{2n-k} \left[\binom{2n}{k+2} + \binom{2n}{k+1} + \binom{2n}{k} \right] \\
&\quad + \binom{2n}{2n-1} p^{2n+1} q + \binom{2n}{2n} p^{2n+2} \\
&\quad + 2 \left[\binom{2n}{n+1} p^{n+2} q^n + \binom{2n}{2n} p^{2n+1} q \right] \\
&\quad + \binom{2n}{n+1} p^{n+1} q^{n+1} + \binom{2n}{n+2} p^{n+2} q^n,
\end{aligned} \tag{2.19}$$

donde los términos que quedan afuera de la sumatoria son los de los bordes, es decir, los dos últimos de la primer fila, el primero y el último de la segunda y los dos primeros de la tercera. Para continuar con el desarrollo de la ecuación 2.19, se considera primero la siguiente identidad de combinatoria,

$$\begin{aligned}
\binom{n}{k} &= \binom{n-1}{k} + \binom{n-1}{k-1} \\
&= \binom{n-2}{k} + \binom{n-2}{k-1} + \binom{n-2}{k-2} + \binom{n-2}{k-3} \\
&= \binom{n-2}{k} + 2 \binom{n-2}{k-1} + \binom{n-2}{k-2},
\end{aligned}$$

que se demuestran fácilmente a partir de la definición de combinaciones. Por lo tanto, el término entre paréntesis rectos en la sumatoria de la ecuación 2.19 se puede expresar como

$$\binom{2n}{k+2} + 2 \binom{2n}{k+1} + \binom{2n}{k} = \binom{2n+2}{k+2}, \tag{2.20}$$

y por lo tanto,

$$\begin{aligned}
P_{2n}(p^2 + 2pq + q^2) &= \sum_{k=n+1}^{2n-2} \binom{2n+2}{k+2} p^{k+2} q^{2n-k} \\
&\quad + p^{2n+1} q \left[\binom{2n}{2n-1} + 2 \binom{2n}{2n} \right] \\
&\quad + p^{n+2} q^n \left[2 \binom{2n}{n+1} + \binom{2n}{n+2} \right] \\
&\quad + \binom{2n}{2n} p^{2n+2} + \binom{2n}{n+1} p^{n+1} q^{n+1},
\end{aligned}$$

donde además se agruparon los términos con factores comunes. Luego,

$$\begin{aligned}
P_{2n} &= \sum_{i=n+3}^{2n} \binom{2n+2}{i} p^i q^{2n+2-i} \\
&\quad + p^{2n+1} q (2n+2) \\
&\quad + p^{n+2} q^n \left[\binom{2n+2}{n+2} - \binom{2n}{n} \right] \\
&\quad + p^{2n+2} + \binom{2n}{n+1} p^{n+1} q^{n+1},
\end{aligned}$$

donde se realizó el cambio de variable $i = k + 2$ en la sumatoria, se usó la ecuación 2.20 con $k = n$ en el tercer sumando, y se consideró que

$$\binom{n}{n} = 1, \quad \binom{n}{n-1} = n$$

en el segundo y cuarto sumando. También se tuvo en cuenta que el factor a la izquierda de la igualdad es $(p^2 + 2pq + q^2) = (p + q)^2 = 1$. Recordar que el objetivo de este razonamiento es relacionar P_{2n} con P_{2n+2} , para lo cual es necesario notar que, excepto por los límites, la sumatoria en P_{2n} es igual a la sumatoria en P_{2n+2} (ver la ecuación 2.16). Los términos que faltan en la sumatoria en P_{2n} para igualar P_{2n+2} son $i = n + 2, 2n + 1, 2n + 2$, y valen

$$\begin{aligned} i = n + 2 &: \binom{2n+2}{n+2} p^{n+2} q^n \\ i = 2n + 1 &: \binom{2n+2}{2n+1} p^{2n+1} q = (2n+2) p^{2n+1} q, \\ i = 2n + 2 &: \binom{2n+2}{2n+2} p^{2n+2} = p^{2n+2}, \end{aligned}$$

todos los cuales son sumandos en la ecuación de P_{2n} . Por lo tanto,

$$P_{2n} = \sum_{i=n+2}^{2n+2} \binom{2n+2}{i} p^i q^{2n+2-i} - \binom{2n}{n} p^{n+2} q^n + \binom{2n}{n+1} p^{n+1} q^{n+1},$$

y como la sumatoria de la derecha de la igualdad es P_{2n+2} , se concluye que

$$P_{2n+2} = P_{2n} + \binom{2n}{n} p^{n+2} q^n - \binom{2n}{n+1} p^{n+1} q^{n+1}. \quad (2.21)$$

Esta ecuación tiene una interpretación interesante. Un juego de $2n+2$ partidas puede considerarse como un juego de $2n$ partidas al que se le agregan 2 partidas. Notar que para ganar un juego de $2n$ partidas hay que ganar $n+1$ partidas y para ganar un juego de $2n+2$ partidas hay que ganar $n+2$ partidas. Esto implica que a menos que el jugador A haya ganado n o $n+1$ partidas en el juego de $2n$ partidas, su condición como ganador o perdedor en el juego de $2n+2$ partidas no cambia. Por lo tanto, excepto por esas dos posibilidades, P_{2n+2} es idéntico a P_{2n} . Una de las excepciones es que habiendo ganado n partidas en el juego de $2n$ partidas, gana las dos partidas siguientes, transformándose en ganador del juego de $2n+2$ partidas. Esto ocurre con probabilidad

$$p^2 \binom{2n}{n} p^n q^n = \binom{2n}{n} p^{n+2} q^n,$$

por lo que la probabilidad de P_{2n+2} respecto a P_{2n} se incrementa esa cantidad. La otra excepción es que habiendo ganado $n+1$ partidas en el juego de $2n$ partidas, pierde las dos partidas siguientes, pasando a ser perdedor del juego de $2n+2$ partidas. Esto ocurre con probabilidad

$$q^2 \binom{2n}{n+1} p^{n+1} q^{n-1} = q^2 \binom{2n}{n+1} p^{n+1} q^{n+1}$$

por lo que la probabilidad de P_{2n+2} respecto a P_{2n} se reduce esa cantidad [5].

Si $2n$ es el número óptimo de partidas, se debe de cumplir que $P_{2n} \geq P_{2n+2}$, y de la ecuación 2.21 se desprende que

$$\begin{aligned} \binom{2n}{n+1} p^{n+1} q^{n+1} &\geq \binom{2n}{n} p^{n+2} q^n \quad \Leftrightarrow \quad \frac{2n!}{(n+1)!(n-1)!} q \geq \frac{2n!}{n!n!} p \\ \Leftrightarrow \quad \frac{1}{(n+1)} q &\geq \frac{1}{n} p \quad \Leftrightarrow \quad nq \geq (n+1)p \quad \Leftrightarrow \quad n(q-p) \geq p, \end{aligned}$$

resultando en

$$n \geq \frac{p}{1-2p}. \quad (2.22)$$

De forma similar, si $2n$ es óptimo, se tiene que cumplir que $P_{2n} \geq P_{2n-2}$. La relación entre P_{2n} y P_{2n-2}

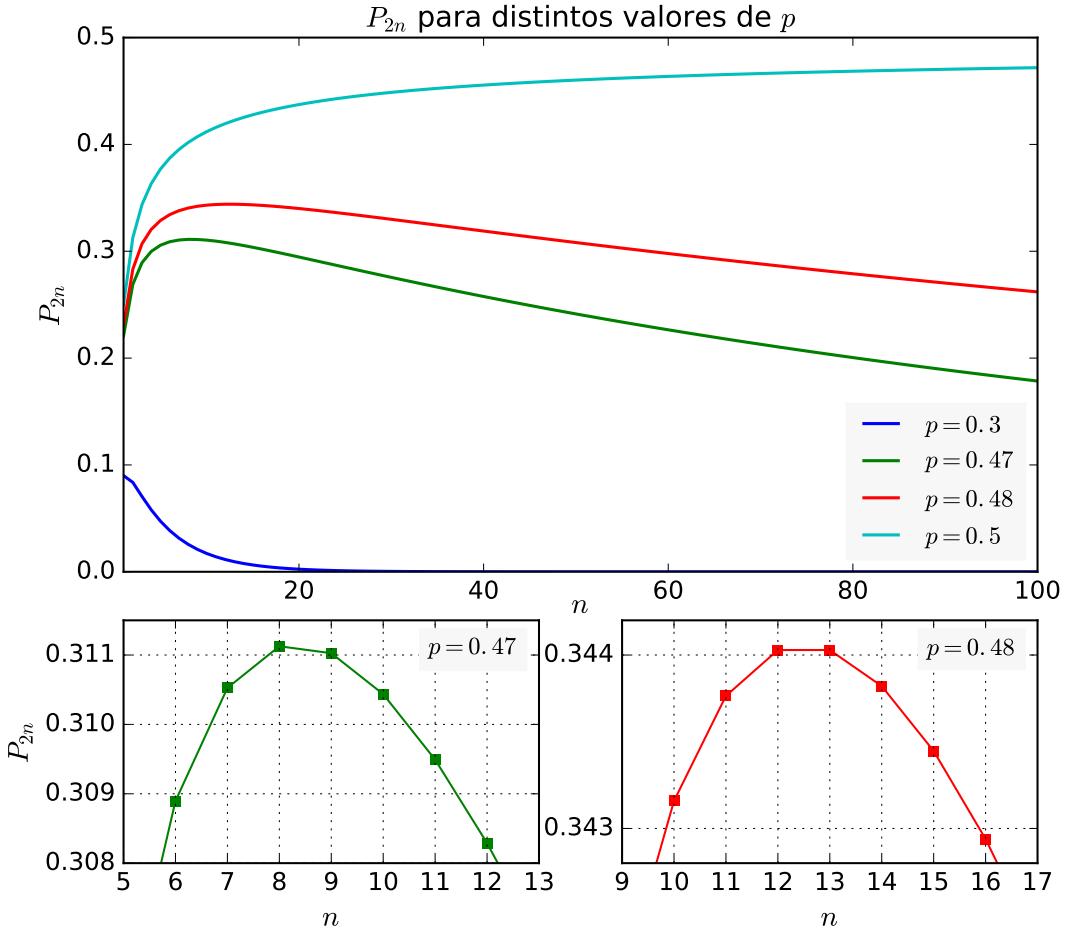


Figura 2.2: Probabilidad P_{2n} de ganar un juego de $2n$ partidas para valores de $p = 0.3, 0.47, 0.48$ y 0.5 . Como se observa en la gráfica de arriba, en el caso de que el juego es equitativo, es decir, cuando $p = 0.5$, la mejor estrategia es jugar la mayor cantidad de partidas posibles, ya que la probabilidad P_{2n} de ganar crece al valor asintótico de 0.5. En el caso en que el juego es desfavorable, esto es, cuando $p < 0.5$, hay dos posibilidades: (1) si p no es cercano a 0.5, P_{2n} es decreciente, y la mejor estrategia es jugar la menor cantidad posible de partidas, que es $2n = 2$. Este es el caso en la gráfica con $p = 0.3$. (2) si p es cercano a 0.5, P_{2n} tiene un máximo, lo que implica que hay un número óptimo de partidas. Si se continuara jugando indefinidamente, se perdería con certeza, ya que P_{2n} tiende a 0 con n , como se ve en las gráficas con $p = 0.47$ y $p = 0.48$. En las gráficas de abajo, se muestra P_{2n} para $p = 0.47$ y $p = 0.48$ en torno al máximo. En el caso en que $p = 0.47$, el máximo está en $n = 8$, por lo que el número óptimo de partidas es $2n = 16$. En el caso en que $p = 0.48$, el valor máximo de P_{2n} se da en dos valores de n , $n = 12$ y $n = 13$, por lo que conviene jugar $2n = 24$ o $2n = 26$ partidas indistintamente. Esto ocurre porque $1/(1-2p) = 25$ es un entero impar, y los dos enteros pares adyacentes producen la misma probabilidad óptima de ganar.

se obtiene de la ecuación 2.21 reemplazando n por $n - 1$, lo que resulta en

$$P_{2n} = P_{2n-2} + \binom{2n-2}{n-1} p^{n+1} q^{n-1} - \binom{2n-2}{n} p^n q^n,$$

y por lo tanto, se debe cumplir que

$$\binom{2n-2}{n-1} p^{n+1} q^{n-1} \geq \binom{2n-2}{n} p^n q^n.$$

Despejando n igual que antes, se llega a que

$$n \leq \frac{q}{1-2p}, \quad (2.23)$$

y combinando las ecuaciones 2.22 y 2.23 se obtiene que

$$\frac{p}{1-2p} \leq n \leq \frac{q}{1-2p},$$

o equivalentemente,

$$\frac{1}{1-2p} - 1 \leq 2n \leq \frac{1}{1-2p} + 1.$$

De esta forma, $2n$ queda determinado únicamente como el entero par más cercano a $1/(1-2p)$. Por ejemplo, si $p = 0.47$, $1/(1-2p) = 16.666\dots$, por lo que el número óptimo de partidas es $2n = 16$. Sin embargo, si $1/(1-2p)$ es un entero impar, los dos enteros pares adyacentes $1/(1-2p) - 1$ y $1/(1-2p) + 1$ producen la misma probabilidad óptima. Por ejemplo, si $p = 0.48$, $1/(1-2p) = 25$ y el número óptimo de partidas es $2n = 24$ o $2n = 26$, ambos con la misma probabilidad máxima de ganar. En la figura 2.2 se muestra la probabilidad P_{2n} de ganar el juego para algunos valores de p .

Capítulo 3

El concepto de variable aleatoria

Este capítulo contiene algunos conceptos desarrollados en el capítulo 4 de [1].

3.1. Variable aleatoria

Una variable aleatoria es una *función* cuyo dominio es el conjunto S de todos los resultados posibles de un experimento. Concretamente, un experimento está definido por el espacio muestral S (o Ω), el σ -álgebra de subconjuntos de S , denominados eventos, y una medida de probabilidad asignada a esos eventos. A cada resultado ζ del experimento, se asigna un número $\mathbf{x}(\zeta)$. De esta forma, se creó una función \mathbf{x} cuyo dominio es el conjunto S y el rango es un conjunto de números.

Respecto a la notación, se empleará el símbolo $\mathbf{x}(\zeta)$ para referirse al número asignado al resultado ζ específico, y el símbolo \mathbf{x} para indicar la regla de correspondencia entre los elementos de S y el número asignado a éstos.

3.1.1. Eventos generados por variables aleatorias

En el estudio de variables aleatorias, surgen preguntas del tipo “¿Cuál es la probabilidad de que la variable aleatoria \mathbf{x} sea menor que cierto valor x ?", o “¿Cuál es la probabilidad de que la variable aleatoria \mathbf{x} esté entre los números x_1 y x_2 ?" Debido a que las probabilidades se asignan a eventos, para poder responder esas preguntas, hay que expresar las condiciones impuestas a la variable aleatoria \mathbf{x} como eventos. En este sentido, considérese la expresión

$$\{\mathbf{x} \leq x\}.$$

Este evento es el subconjunto de S que consiste en los elementos ζ que cumplen que $\mathbf{x}(\zeta) \leq x$. De esta forma, $\{\mathbf{x} \leq x\}$ no es un conjunto de números, sino un conjunto de resultados elementales del experimento. Análogamente, la expresión

$$\{x_1 \leq \mathbf{x} \leq x_2\},$$

es el subconjunto de S que contiene los elementos ζ tales que $x_1 \leq \mathbf{x}(\zeta) \leq x_2$. De la misma forma, si R es un conjunto de números del eje x ,

$$\{x \in R\}$$

representa el subconjunto de S que contiene los elementos ζ que cumplen que $\mathbf{x}(\zeta) \in R$. Un comentario adicional de este caso: $\{x \in R\}$ es un evento solo si R es la unión e intersección contable de intervalos del eje x . Esta restricción tiene mas interés matemático que práctico.

3.1.2. Definición de variable aleatoria

Formalmente, una variable aleatoria \mathbf{x} es el proceso de asignación de un número $\mathbf{x}(\zeta)$ a cada elemento ζ del espacio muestral. Esta función de asignación es mas bien arbitraria, siempre que satisfaga las siguientes condiciones:

- I. El conjunto $\{\mathbf{x} \leq x\}$ es un evento para todo x .
- II. Las probabilidades de los eventos $\{x = -\infty\}$ y $\{x = \infty\}$ es nula:

$$P\{x = -\infty\} = 0 \quad P\{x = \infty\} = 0.$$

Esta condición indica que si bien se permite que $\mathbf{x}(\zeta) = -\infty$ o $\mathbf{x}(\zeta) = \infty$ para algunos elementos ζ del espacio muestral, se exige que el conjunto de estos elementos tenga probabilidad nula.

3.2. Distribución y densidad de probabilidad

3.2.1. La función distribución de probabilidad

Los elementos del espacio muestral S contenidos en el evento $\{\mathbf{x} \leq x\}$ cambian al cambiar el número x . Por lo tanto, la probabilidad $P\{\mathbf{x} \leq x\}$ del evento $\{\mathbf{x} \leq x\}$ es un número que depende de x . Este número se denota como $F_x(x)$ y se llama *función de distribución (acumulada)* de la variable aleatoria \mathbf{x} . Es decir, la función de distribución de una variable aleatoria \mathbf{x} se define como

$$F_x(x) = P\{\mathbf{x} \leq x\},$$

para todo x desde $-\infty$ a ∞ .

Variables aleatorias continuas y discretas Se dice que una variable aleatoria \mathbf{x} es *continua* si su distribución de probabilidad $F_x(x)$ es continua. En este caso, se puede demostrar que se cumple que $P\{\mathbf{x} = x\} = 0$ para todo x .

Si $F_x(x)$ es constante excepto por un número finito de discontinuidades de salto (constante a trozos), se dice que \mathbf{x} es una variable aleatoria *discreta*. Si x_i es uno de los puntos de discontinuidad, se puede demostrar que se cumple que

$$P\{\mathbf{x} = x_i\} = F_x(x_i) - F_x(x_i^-) = p_i.$$

Percentiles Se define el percentil u de la variable aleatoria \mathbf{x} como el menor valor x_u tal que

$$u = P\{\mathbf{x} \leq x_u\} = F_x(x_u).$$

De esta forma, $x_u = F^{-1}(u)$, por lo que su dominio es el intervalo $0 \leq u \leq 1$ y su rango es el eje x . La *mediana* de \mathbf{x} es el menor valor m tal que $F_x(m) = 0.5$. Por lo tanto, la mediana m es el percentil 0.5 de \mathbf{x} .

Supóngase que se dispone de una serie de observaciones de una variable aleatoria \mathbf{x} y se desea estimar cierto percentil. Esto se denomina *percentil empírico*. Por ejemplo, supóngase que hay 100 observaciones y se quiere calcular el percentil 0.6. Esto implica encontrar el valor $x_{0.6}$ de forma que $P\{\mathbf{x} \leq x_{0.6}\} = 0.6$. Desde el punto de vista de la interpretación en frecuencia de la probabilidad, una probabilidad de 0.6 significa que hay 60 casos favorables entre los 100 casos posibles. Por lo tanto, estimar el percentil 0.6 significa encontrar la observación $x_{0.6}$ tal que haya 60 observaciones de valor menor o igual a $x_{0.6}$. Entonces, para encontrar este valor, hay que ordenar las 100 observaciones de menor a mayor, y la observación en la posición 60 es el percentil 0.6 buscado $x_{0.6}$. Asimismo, la mediana de la variable aleatoria \mathbf{x} , que por definición es el percentil 0.5, es el valor de centro de las observaciones ordenadas de menor a mayor.

Distribución y percentil empíricos Supóngase que se realiza un experimento n veces y se obtienen los valores x_1, \dots, x_n de la variable aleatoria \mathbf{x} . Estos valores se ubican en el eje x y se construye una función $F_n(x)$ escalonada como se muestra en la figura 3.1. Los escalones se ubican en los valores x_i y tienen altura $1/n$. El primer escalón se ubica en el valor más pequeño x_{\min} . De esta forma $F_n(x) < 0$ para $x < x_{\min}$. Esta función es la *distribución de probabilidad empírica* de la variable aleatoria \mathbf{x} , ya que para un valor de x específico, el número de escalones n_x de $F_n(x)$ es la cantidad de valores x_i más pequeños que x , por lo que $F_n(x) = n_x/n$. Como $n_x/n \approx P\{\mathbf{x} \leq x\}$ si n es grande, se concluye que

$$F_n(n) = \frac{n_x}{n} \rightarrow P\{\mathbf{x} \leq x\} = F(x) \quad \text{con } n \rightarrow \infty.$$

El percentil empírico se construye de la siguiente forma: se forman n segmentos de longitud x_i y se ubican verticalmente en orden creciente separados una distancia $1/n$, y se construye una función escalonada como se muestra en la figura 3.1. Esta función es igual a la distribución empírica $F_n(x)$ si se intercambian los ejes.

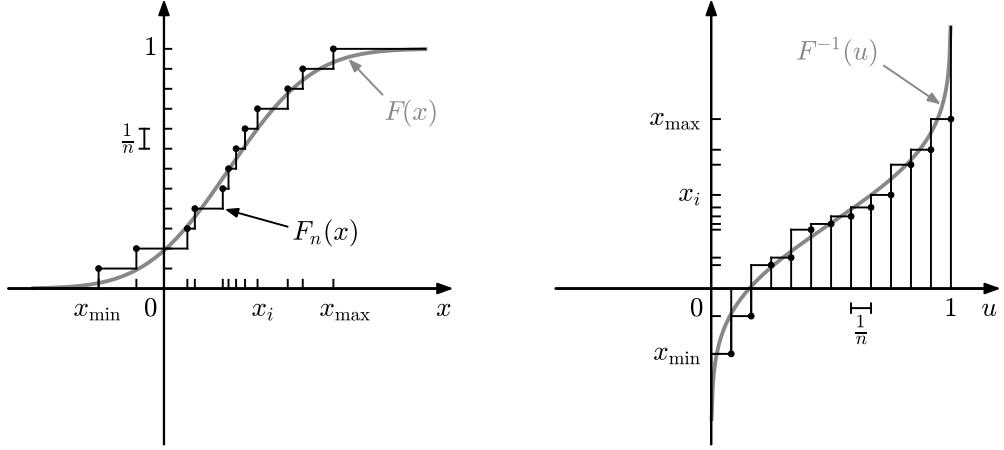


Figura 3.1: Distribución y percentil empíricos. En la gráfica de la izquierda se muestra la distribución de probabilidad $F(x)$ y la distribución empírica $F_n(x)$ con $n = 12$ para una variable aleatoria normal. A la izquierda se muestra la función inversa $x_u = F^{-1}(y)$ y el percentil empírico.

3.2.2. La función densidad de probabilidad

La derivada de la función distribución de probabilidad $F_x(x)$ se llama función densidad de probabilidad (p.d.f, *probability density function*) de la variable aleatoria \mathbf{x} y se denota como $f_x(x)$. De esta forma

$$f_x(x) \triangleq \frac{dF_x(x)}{dx}. \quad (3.1)$$

Si la variable aleatoria \mathbf{x} es continua, $f_x(x)$ es una función continua. Si la variable aleatoria \mathbf{x} es discreta, la densidad de probabilidad tiene la forma

$$f_x(x) = \sum_i p_i \delta(x - x_i),$$

donde x_i son los puntos de discontinuidad de salto en $F_x(x)$. En este caso, la densidad es conocida con el nombre de *función de masa de probabilidad* (p.m.f, *probability mass function*).

De la ecuación 3.1, se deduce que

$$F_x(x) = \int_{-\infty}^x f_x(u) du.$$

Además, puede probarse que

$$P\{x_1 < \mathbf{x} \leq x_2\} = F_x(x_2) - F_x(x_1) = \int_{x_1}^{x_2} f_x(x) dx, \quad (3.2)$$

es decir, el área bajo la densidad de probabilidad $f_x(x)$ en el intervalo (x_1, x_2) es la probabilidad de que la variable aleatoria \mathbf{x} pertenezca al intervalo (x_1, x_2) .

Si la variable aleatoria \mathbf{x} es continua, el intervalo en la probabilidad en la ecuación 3.2 puede cambiarse por $\{x_1 \leq \mathbf{x} \leq x_2\}$. Tomando $x_1 = x$ y $x_2 = x + \Delta x$ en la ecuación 3.2, se obtiene que

$$P\{x \leq \mathbf{x} \leq x + \Delta x\} \approx f_x(x)\Delta x \quad (3.3)$$

para Δx suficientemente pequeño. De aquí se obtiene que $f_x(x)$ puede ser directamente definida como

$$f_x(x) = \lim_{\Delta x \rightarrow 0} \frac{P\{x \leq \mathbf{x} \leq x + \Delta x\}}{\Delta x}. \quad (3.4)$$

3.3. Variables aleatorias específicas

En esta sección se describen algunas distribuciones de probabilidad usadas comúnmente. Todo lo aquí mencionado está basado en la sección 4.3 de [1]. Se hace hincapié en el desarrollo de algunos resultados que no se detallan en la referencia.

3.3.1. Variables aleatorias continuas

Distribución Normal Se dice que \mathbf{x} es una variable aleatoria normal o gaussiana de parámetros μ y σ^2 si su función densidad de probabilidad es

$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}. \quad (3.5)$$

Se denota $\mathbf{x} \sim N(\mu, \sigma^2)$.

La constante $\sqrt{2\pi\sigma^2}$ en la densidad de probabilidad es una constante de normalización para que el área bajo $f_x(x)$ sea uno. Para ver esto, se demostrará que

$$Q \triangleq \int_{-\infty}^{\infty} e^{-x^2/2\sigma^2} dx = \sqrt{2\pi\sigma^2}.$$

Es claro que el área bajo $e^{-x^2/2\sigma^2}$ coincide con el área bajo $e^{-(x-\mu)^2/2\sigma^2}$ de la ecuación 3.5, ya que el cambio de variable $x - \mu$ es solo en un desplazamiento de la curva en x una cantidad μ . Considérese

$$\begin{aligned} Q^2 &= \left(\int_{-\infty}^{\infty} e^{-x^2/2\sigma^2} dx \right)^2 \\ &= \left(\int_{-\infty}^{\infty} e^{-x^2/2\sigma^2} dx \right) \left(\int_{-\infty}^{\infty} e^{-y^2/2\sigma^2} dy \right) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2\sigma^2} dx dy \end{aligned} \quad (3.6)$$

La integral doble es una integral en todo el plano (x, y) . Se realizará el cambio de variable

$$x = r \cos \theta, \quad y = r \sin \theta, \quad (3.7)$$

que consiste en transformar la integral doble en coordenadas cartesianas a coordenadas polares. Para hacerlo, recordar que para integrar $f(x, y)$ sobre la región R del plano, bajo la transformación $x = g(u, v)$, $y = h(u, v)$, la región de integración se transforma en S y la integral es

$$\iint_R f(x, y) dA = \iint_S f(g(u, v), h(u, v)) |J(u, v)| du dv, \quad (3.8)$$

donde $|J(u, v)|$ es el determinante de la matriz Jacobiana. En el caso de la transformación a coordenadas polares dada por la ecuación 3.7, la matriz Jacobiana es

$$J(r, \theta) = \begin{pmatrix} \partial x / \partial r & \partial x / \partial \theta \\ \partial y / \partial r & \partial y / \partial \theta \end{pmatrix} = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix},$$

y el determinante es $|J(r, \theta)| = r$, por lo que al aplicar la transformación de la ecuación 3.8 a la ecuación

3.6 notando que $x^2 + y^2 = r^2$ se obtiene que

$$\begin{aligned} Q^2 &= \int_0^{2\pi} \int_0^\infty e^{-r^2/2\sigma^2} r dr d\theta \\ &= 2\pi \int_0^\infty e^{-r^2/2\sigma^2} r dr \\ &\stackrel{(a)}{=} 2\pi\sigma^2 \int_0^\infty e^{-u} du \\ &= 2\pi\sigma^2, \end{aligned}$$

donde en (a) se realizó el cambio de variable $u = r^2/2\sigma^2$, $r dr = \sigma^2 du$. Se concluye que

$$\int_{-\infty}^\infty e^{-x^2/2\sigma^2} dx = \sqrt{2\pi\sigma^2} \quad (3.9)$$

En el caso en que $\sigma^2 = 1/2$, la ecuación anterior queda

$$\int_{-\infty}^\infty e^{-x^2} dx = \sqrt{\pi}, \quad (3.10)$$

resultado que se conoce como integral gaussiana¹.

Distribución Exponencial La variable aleatoria \mathbf{x} es exponencial de parámetro λ si su densidad de probabilidad es

$$f_x(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{en otro caso} \end{cases}. \quad (3.11)$$

Si la ocurrencia de eventos en intervalos de tiempo que no se solapan son independientes, como los instantes de ocurrencia de llamadas telefónicas o los instantes de arribo de ómnibus a una parada, puede demostrarse que el tiempo de espera de esos eventos es una variable aleatoria exponencial.

La distribución de probabilidad de una variable aleatoria exponencial es

$$F_x(x) = 1 - e^{-\lambda x}.$$

Para verlo, se parte de la definición de densidad de probabilidad y se opera de la siguiente forma,

$$\begin{aligned} F_x(x) &= \int_{-\infty}^x f_x(u) du \\ &= \int_0^x \lambda e^{-\lambda u} du \\ &= -e^{-\lambda u} \Big|_0^x \\ &= - (e^{-\lambda x} - 1) \\ &= 1 - e^{-\lambda x}. \end{aligned}$$

De forma similar, puede probarse que la media de una variable aleatoria exponencial \mathbf{x} es $E\{\mathbf{x}\} = 1/\lambda$. El concepto de media y varianza se desarrolla mas adelante en [1], pero se incluye aquí por conveniencia.

¹ver https://en.wikipedia.org/wiki/Gaussian_integral

Usando la definición de la media de una variable aleatoria y operando, se tiene que

$$\begin{aligned}
E\{\mathbf{x}\} &= \int_{-\infty}^{\infty} tf_x(t)dt \\
&= \int_0^{\infty} t\lambda e^{-\lambda t}dt \\
&\stackrel{(a)}{=} -te^{-\lambda t}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda t}dt \\
&= -\frac{1}{\lambda}e^{-\lambda t}\Big|_0^{\infty} \\
&= -\frac{1}{\lambda}(0 - 1) \\
&= \frac{1}{\lambda},
\end{aligned}$$

donde en (a) se aplicó integración por partes, $\int u dv = uv - \int v du$, con $u = t$ y $dv = \lambda e^{-\lambda t}$, y entonces $du = 1$ y $v = -e^{-\lambda t}$. Esto implica que si la variable aleatoria exponencial \mathbf{x} es por ejemplo cierto tiempo de espera de ocurrencia de un evento, su media es $1/\lambda$ unidades de tiempo y λ es la tasa de ocurrencia del evento. En ocasiones, en lugar de la parametrización con λ se emplea el parámetro $\beta = 1/\lambda$. En este caso,

$$f_x(x) = \begin{cases} \frac{1}{\beta}e^{-x/\beta}, & x \geq 0 \\ 0, & \text{en otro caso} \end{cases},$$

y $E\{\mathbf{x}\} = \beta$.

Distribución Gamma Se dice que \mathbf{x} es una variable aleatoria gamma de parámetros α y β , $\alpha \geq 0$, $\beta \geq 0$, si su densidad de probabilidad es

$$f_x(x) = \begin{cases} \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha}e^{-x/\beta}, & x \geq 0 \\ 0, & \text{en otro caso} \end{cases}, \quad (3.12)$$

donde $\Gamma(\alpha)$ es la función gamma, definida como

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}e^{-x} dx. \quad (3.13)$$

A continuación se mencionarán algunas propiedades de la función gamma. La función gamma cumple la ecuación funcional

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1). \quad (3.14)$$

Para verlo, se parte de la definición de la función gamma y se opera de la siguiente forma,

$$\begin{aligned}
\Gamma(\alpha) &= \int_0^{\infty} x^{\alpha-1}e^{-x} dx \\
&\stackrel{(a)}{=} -x^{\alpha-1}e^{-x}\Big|_0^{\infty} + (\alpha - 1) \int_0^{\infty} x^{\alpha-2}e^{-x} dx \\
&= (\alpha - 1)\Gamma(\alpha - 1),
\end{aligned}$$

donde en (a) se aplicó integración por partes con $u = x^{\alpha-1}$ y $dv = e^{-x}$, y entonces $du = (\alpha - 1)x^{\alpha-2}$ y $v = -e^{-x}$. Si $\alpha = n$ es un número entero, se cumple que

$$\Gamma(n) = (n - 1)\Gamma(n - 1) = (n - 1)!, \quad (3.15)$$

ya que

$$\Gamma(1) = \int_0^\infty e^{-x} dx = -e^{-x} \Big|_0^\infty = -(0 - 1) = 1.$$

y por lo tanto

$$\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)(n-2) \cdots 2 \cdot 1 \cdot \Gamma(1) = (n-1)!.$$

Se cumple además que

$$\Gamma(1/2) = \sqrt{\pi}. \quad (3.16)$$

Efectivamente, evaluando la función gamma en $\alpha = 1/2$ y operando, se tiene que

$$\begin{aligned} \Gamma(1/2) &= \int_0^\infty x^{1/2-1} e^{-x} dx \\ &= \int_0^\infty x^{-1/2} e^{-x} dx \\ &\stackrel{(a)}{=} -x^{1/2} e^{-x} \Big|_0^\infty + 2 \int_0^\infty e^{-u^2} du \\ &= \int_{-\infty}^\infty e^{-x^2} dx \\ &\stackrel{(b)}{=} \sqrt{\pi}, \end{aligned}$$

donde en (a) se aplicó integración por partes con $u = x^{1/2}$ y $dv = e^{-x} dx$, y entonces $du = (1/2)x^{-1/2} dx$ y $v = -e^{-x}$, y en (b) se notó que la integral es gaussiana y se aplicó el resultado de la ecuación 3.10. Empleando la ecuación 3.14, se deduce que

$$\begin{aligned} \Gamma\left(\frac{3}{2}\right) &= \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2} \\ \Gamma\left(\frac{5}{2}\right) &= \frac{3}{2} \Gamma\left(\frac{3}{2}\right) = \frac{3\sqrt{\pi}}{4}, \end{aligned}$$

y así sucesivamente, resultando en el caso general, en

$$\begin{aligned} \Gamma\left(\frac{2n+1}{2}\right) &= \frac{(2n-1)(2n-3) \cdots 5 \cdot 3 \cdot 1}{2^n} \sqrt{\pi} \\ &\stackrel{(a)}{=} \frac{(2n-1)!!}{2^n} \sqrt{\pi} \\ &\stackrel{(b)}{=} \frac{(2n)!}{4^n n!} \sqrt{\pi} \end{aligned}$$

donde en (a) se expresó la ecuación empleando la operación de doble factorial² y en (b) se tuvieron las dos siguientes consideraciones:

1. Para un número par $2n$, se cumple que

$$(2n)!! = (2n)(2n-2)(2n-4) \cdots 4 \cdot 2 \stackrel{(a)}{=} 2^n n(n-1)(n-2) \cdots 2 \cdot 1 = 2^n n!,$$

donde en (a) se sacó 2 de factor común en cada uno de los n factores del producto.

2. Se cumple que

$$n! = n!!(n-1)!! \quad \Rightarrow \quad n!! = \frac{n!}{(n-1)!!} = \frac{n!(n+1)}{(n-1)!!(n+1)} = \frac{(n+1)!}{(n+1)!!}.$$

²Ver https://en.wikipedia.org/wiki/Double_factorial.

Empleando estos dos resultados, se observa que para un número impar $2n - 1$, se cumple que

$$(2n - 1)!! = \frac{(2n)!}{(2n)!!} = \frac{(2n)!}{2^n n!},$$

que es lo que se quería mostrar.

Una variable aleatoria \mathbf{x} gamma de parámetros α y β se denota como $\mathbf{x} \sim G(\alpha, \beta)$. La función de densidad gamma tiene un conjunto variado de formas en función del valor de los parámetros α y β . Una variable aleatoria exponencial es un caso particular de una variable aleatoria gamma con $\alpha = 1$. Con $\alpha = n/2$ y $\beta = 2$ se obtiene una variable aleatoria $\chi^2(n)$ (chi-cuadrado con n grados de libertad).

Con $\alpha = n$ entero en la ecuación 3.12 y definiendo $\beta = 1/\lambda$, la función densidad de probabilidad gamma es

$$f_x(x) = \begin{cases} \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}, & x \geq 0 \\ 0, & \text{en otro caso.} \end{cases} \quad (3.17)$$

Se demostrará a continuación que la distribución de probabilidad es

$$F_x(t) = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

Para hacerlo, se parte de la definición de distribución de probabilidad y se aplica integración por partes,

$$\begin{aligned} F_x(t) &= \int_0^t f_x(x) dx \\ &= \frac{\lambda^n}{(n-1)!} \int_0^t x^{n-1} e^{-\lambda x} dx \\ &= \frac{\lambda^{n-1}}{(n-2)!} \int_0^t \frac{x^{n-1}}{n-1} \lambda e^{-\lambda x} dx \\ &\stackrel{(a)}{=} \frac{\lambda^{n-1}}{(n-2)!} \left(-\frac{x^{n-1} e^{-\lambda x}}{n-1} \Big|_0^t + \int_0^t x^{n-2} e^{-\lambda x} dx \right) \\ &= \frac{\lambda^{n-1}}{(n-2)!} \left(-\frac{t^{n-1} e^{-\lambda t}}{n-1} + \int_0^t x^{n-2} e^{-\lambda x} dx \right) \\ &= -\frac{(\lambda t)^{n-1} e^{-\lambda t}}{(n-1)!} + \frac{\lambda^{n-1}}{(n-2)!} \int_0^t x^{n-2} e^{-\lambda x} dx, \end{aligned}$$

donde en (a) se aplicó integración por partes con $u = x^{n-1}/(n-1)$ y $dv = \lambda e^{-\lambda x}$, y entonces $du = x^{n-2}$ y $v = -e^{-\lambda x}$. Notar que el segundo sumando en la última ecuación es la distribución de probabilidad gamma con parámetro $\alpha = n - 1$. Aplicando integración por partes $n - 2$ veces mas, es decir, $n - 1$ veces en total, se obtiene que

$$\begin{aligned} F_x(t) &= -\frac{(\lambda t)^{n-1} e^{-\lambda t}}{(n-1)!} - \frac{(\lambda t)^{n-2} e^{-\lambda t}}{(n-2)!} - \cdots - \frac{(\lambda t) e^{-\lambda t}}{1!} + \frac{\lambda}{0!} \int_0^t e^{-\lambda x} dx \\ &= -\frac{(\lambda t)^{n-1} e^{-\lambda t}}{(n-1)!} - \frac{(\lambda t)^{n-2} e^{-\lambda t}}{(n-2)!} - \cdots - \frac{(\lambda t) e^{-\lambda t}}{1!} - e^{-\lambda t} + 1 \\ &= 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}. \end{aligned}$$

Distribución chi-cuadrado Se dice que una variable aleatoria \mathbf{x} es $\chi^2(n)$ (chi-cuadrado con n grados de libertad) si su densidad de probabilidad es

$$f_x(x) = \begin{cases} \frac{x^{n/2-1}}{2^{n/2} \Gamma(n/2)} e^{-x/2}, & x \geq 0 \\ 0, & \text{en otro caso} \end{cases}, \quad (3.18)$$

donde $\Gamma(\alpha)$ es la función gamma definida en la ecuación 3.13.

Distribución Beta Una variable aleatoria x tiene distribución beta con parámetros α y β no negativos si

$$f_x(x) = \begin{cases} \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, & 0 < x < 1 \\ 0, & \text{en otro caso,} \end{cases} \quad (3.19)$$

donde la función beta $B(\alpha, \beta)$ se define como

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = 2 \int_0^{\pi/2} (\sin \theta)^{2\alpha-1} (\cos \theta)^{2\beta-1} d\theta. \quad (3.20)$$

La forma trigonométrica en la ecuación 3.20 se obtiene mediante el cambio de variable $x = \sin^2 \theta$. Efectivamente, realizando dicho cambio de variable y teniendo en cuenta que entonces $dx = 2 \sin \theta \cos \theta d\theta$, se tiene que

$$\begin{aligned} B(\alpha, \beta) &= 2 \int_0^{\pi/2} (\sin^2 \theta)^{\alpha-1} (1 - \sin^2 \theta)^{\beta-1} \sin \theta \cos \theta d\theta \\ &= 2 \int_0^{\pi/2} (\sin \theta)^{2\alpha-2} (\cos \theta)^{2\beta-2} \sin \theta \cos \theta d\theta \\ &= 2 \int_0^{\pi/2} (\sin \theta)^{2\alpha-1} (\cos \theta)^{2\beta-1} d\theta. \end{aligned}$$

Es posible expresar la función beta en términos de la función gamma definida en la ecuación 3.13. Para hacerlo, primero se considera que

$$\begin{aligned} \Gamma(\alpha) &= \int_0^\infty x^{\alpha-1} e^{-x} dx \\ &\stackrel{(a)}{=} 2 \int_0^\infty y^{2\alpha-2} e^{-y^2} y dy, \end{aligned}$$

donde en (a) se realizó el cambio de variable $x = y^2$. Esto resulta en que la función gamma se puede expresar como

$$\Gamma(\alpha) = 2 \int_0^\infty y^{2\alpha-1} e^{-y^2} dy.$$

De esta forma,

$$\begin{aligned} \Gamma(\alpha)\Gamma(\beta) &= \left(2 \int_0^\infty x^{2\alpha-1} e^{-x^2} dx \right) \left(2 \int_0^\infty y^{2\beta-1} e^{-y^2} dy \right) \\ &= 4 \int_0^\infty \int_0^\infty x^{2\alpha-1} y^{2\beta-1} e^{-(x^2+y^2)} dx dy. \end{aligned}$$

Realizando el cambio a coordenadas polares $x = r \sin \theta$ y $y = r \cos \theta$, donde el determinante de la matriz Jacobiana de la transformación es $|J(r, \theta)| = r$, como se vió previamente, y $0 \leq r < \infty$ y $0 \leq \theta \leq \pi/2$, se tiene que

$$\begin{aligned} \Gamma(\alpha)\Gamma(\beta) &= 4 \int_0^{\pi/2} \int_0^\infty (r \sin \theta)^{2\alpha-1} (r \cos \theta)^{2\beta-1} e^{-r^2} r dr d\theta \\ &= 4 \int_0^{\pi/2} \int_0^\infty r^{2(\alpha+\beta)-1} e^{-r^2} (\sin \theta)^{2\alpha-1} (\cos \theta)^{2\beta-1} dr d\theta \\ &= \left(2 \int_0^\infty r^{2(\alpha+\beta)-1} e^{-r^2} dr \right) \left(2 \int_0^{\pi/2} (\sin \theta)^{2\alpha-1} (\cos \theta)^{2\beta-1} d\theta \right) \\ &= \Gamma(\alpha + \beta) B(\alpha, \beta), \end{aligned}$$

y por lo tanto,

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}. \quad (3.21)$$

Una variable aleatoria con distribución beta con parámetros α y β se denota como $\mathbf{x} \sim \text{Beta}(\alpha, \beta)$. La distribución beta tiene una gran flexibilidad, tomando diversas formas en función de los parámetros α y β . Con $\alpha = \beta = 1$ se obtiene la distribución uniforme en $(0, 1)$. Si $\alpha > 1$ y $\beta > 1$, $f_x(x) \rightarrow 0$ en $x = 0$ y $x = 1$ y tiene un único máximo. Si $0 < \alpha < 1$, $f_x(x) \rightarrow \infty$ con $x \rightarrow 0$, y si $0 < \beta < 1$, $f_x(x) \rightarrow \infty$ con $x \rightarrow 1$. Si $\alpha < 1$ y $\beta < 1$, $f_x(x)$ es convexa con un único mínimo. Si $\alpha = \beta$, la densidad es simétrica en torno a $x = 1/2$.

Otras distribuciones En [1] se describe con cierto detalle otros tipos de variables aleatorias continuas. Se mencionarán a continuación algunas de ellas de forma breve.

La *variable aleatoria de Erlang* es un caso particular de la variable aleatoria gamma con $G(n, 1/n\mu)$. Cuando el parámetro n entero de la distribución de Erlang es 1, se obtiene una distribución exponencial, y cuando $n \rightarrow \infty$ se obtiene una distribución constante $F_x(t) = 1$ para $t > 1/\mu$ y cero en otro caso. Por lo tanto, la distribución de Erlang abarca de aleatoriedad a certeza cuando n varía 1 a ∞ .

Otras distribuciones mencionadas en [1] son: Rayleigh, Nakagami, uniforme, Cauchy, Laplace y Maxwell.

3.3.2. Variables aleatorias discretas

En el caso de variables aleatorias discretas, la probabilidad de que la variable aleatoria tome un valor específico k es una función de k que se suele denominar *función de masa de probabilidad* (p.m.f, *probability mass function*). Es el concepto análogo a la función densidad de probabilidad de una variable aleatoria continua.

Distribución de Bernoulli La variable aleatoria \mathbf{x} es Bernoulli si toma valores 0 o 1 con probabilidad

$$P\{\mathbf{x} = 1\} = p, \quad P\{\mathbf{x} = 0\} = q = 1 - p.$$

Distribución Binomial En n ensayos Bernoulli independientes donde p es la probabilidad de éxito de cada ensayo, si \mathbf{x} es la cantidad de éxitos en los n experimentos, \mathbf{x} es una variable aleatoria Binomial (ver la sección 2.2). Por lo tanto, se dice que \mathbf{x} es una variable aleatoria Binomial de parámetros n y p si \mathbf{x} toma valores $0, 1, 2, \dots, n$ con probabilidad (ver la ecuación 2.3)

$$P\{\mathbf{x} = k\} = \binom{n}{k} p^k q^{n-k}, \quad p + q = 1, \quad k = 0, 1, 2, \dots, n. \quad (3.22)$$

Distribución de Poisson La distribución de Poisson representa el número de ocurrencias de un evento poco frecuente en un número grande de ensayos. La variable aleatoria \mathbf{x} se dice Poisson con parámetro λ si toma valores $0, 1, 2, \dots, \infty$ con probabilidad

$$P\{\mathbf{x} = k\} = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots, \infty. \quad (3.23)$$

3.4. Distribuciones condicionales

Todo lo incluido en esta sección está basado en la sección 4.4 de [1].

En la sección 1.3.4 se definió la probabilidad condicional de eventos. La *distribución condicional* $F(x|M)$ de la variable aleatoria \mathbf{x} dado el evento M se define como la probabilidad condicional del evento $\{\mathbf{x} \leq x\}$,

$$F(x|M) = P\{\mathbf{x} \leq x|M\} = \frac{P\{\mathbf{x} \leq x, M\}}{P(M)}, \quad (3.24)$$

donde $\{\mathbf{x} \leq x, M\}$ es la intersección de los eventos $\{\mathbf{x} \leq x\}$ y M , es decir, el evento que consiste en todos los resultados ζ tal que $\mathbf{x}(\zeta) \leq x$ y $\zeta \in M$. $F(x|M)$ tiene todas las propiedades de una distribución de probabilidad. En particular, se cumple que

$$F(\infty|M) = 1 \quad F(-\infty|M) = 0$$

$$P\{x_1 < \mathbf{x} \leq x_2|M\} = F(x_2|M) - F(x_1|M) = \frac{P\{x_1 < \mathbf{x} \leq x_2, M\}}{P(M)}.$$

La *densidad condicional* $f(x|M)$ es la derivada de $F(x|M)$,

$$f(x|M) = \frac{dF(x|M)}{dx} = \lim_{\Delta x \rightarrow 0} \frac{P\{x < \mathbf{x} \leq x + \Delta x|M\}}{\Delta x}.$$

Por ser una densidad de probabilidad es una función nonnegativa de área 1.

En general, para determinar $F(x|M)$ es necesario conocer el experimento subyacente. Sin embargo, si M puede expresarse en función de la variable aleatoria \mathbf{x} , alcanza con conocer $F(x)$ para determinar $F(x|M)$. A continuación se presentan dos casos como ejemplo.

- I. Se quiere encontrar la distribución condicional de la variable aleatoria \mathbf{x} dado $\mathbf{x} \leq a$, asumiendo que $F(a) \neq 0$. Esto es un caso especial de la ecuación 3.24 con $M = \{\mathbf{x} \leq a\}$. Por lo tanto, como indica la ecuación, el problema consiste en encontrar la función

$$F(x|\mathbf{x} \leq a) = P\{\mathbf{x} \leq x|\mathbf{x} \leq a\} = \frac{P\{\mathbf{x} \leq x, \mathbf{x} \leq a\}}{P\{\mathbf{x} \leq a\}}.$$

- Si $x \geq a$, se cumple que $\{\mathbf{x} \leq x, \mathbf{x} \leq a\} = \{\mathbf{x} \leq a\}$, ya que $\{\mathbf{x} \leq a\} \subset \{\mathbf{x} \leq x\}$. Por lo tanto,

$$F(x|\mathbf{x} \leq a) = \frac{P\{\mathbf{x} \leq a\}}{P\{\mathbf{x} \leq a\}} = 1, \quad x \geq a.$$

- Si $x < a$, se cumple que $\{\mathbf{x} \leq x, \mathbf{x} \leq a\} = \{\mathbf{x} \leq x\}$ y por lo tanto,

$$F(x|\mathbf{x} \leq a) = \frac{P\{\mathbf{x} \leq x\}}{P\{\mathbf{x} \leq a\}} = \frac{F(x)}{F(a)}, \quad x < a.$$

Es decir

$$F(x|\mathbf{x} \leq a) = \begin{cases} \frac{F(x)}{F(a)}, & x < a. \\ 1, & x \geq a \end{cases} \quad (3.25)$$

Para obtener la densidad condicional, hay que derivar $F(x|\mathbf{x} \leq a)$ respecto a x . Como $F'(x) = f(x)$, se obtiene que

$$f(x|\mathbf{x} \leq a) = \begin{cases} \frac{f(x)}{F(a)} = \frac{f(x)}{\int_{-\infty}^a f(x)dx}, & x < a. \\ 0, & x \geq a \end{cases} \quad (3.26)$$

- II. Se supone ahora que $M = \{b < \mathbf{x} \leq a\}$, por lo que la ecuación 3.24 queda

$$F(x|b < \mathbf{x} \leq a) = \frac{P\{\mathbf{x} \leq x, b < \mathbf{x} \leq a\}}{P\{b < \mathbf{x} \leq a\}}.$$

- Si $x \geq a$, se cumple que $\{\mathbf{x} \leq x, b < \mathbf{x} \leq a\} = \{b < \mathbf{x} \leq a\}$. Por lo tanto,

$$F(x|b < \mathbf{x} \leq a) = \frac{P\{b < \mathbf{x} \leq a\}}{P\{b < \mathbf{x} \leq a\}} = 1, \quad x \geq a.$$

- Si $b \leq x < a$, se cumple que $\{\mathbf{x} \leq x, b < \mathbf{x} \leq a\} = \{b < \mathbf{x} \leq x\}$ y Por lo tanto,

$$F(x|b < \mathbf{x} \leq a) = \frac{P\{b < \mathbf{x} \leq x\}}{P\{b < \mathbf{x} \leq a\}} = \frac{F(x) - F(b)}{F(a) - F(b)}, \quad b \leq x < a.$$

- Si $x < b$, se cumple que $\{\mathbf{x} \leq x, b < \mathbf{x} \leq a\} = \{\emptyset\}$. Por lo tanto,

$$F(x|b < \mathbf{x} \leq a) = 0, \quad x < b.$$

Por lo tanto

$$F(x|b < \mathbf{x} \leq a) = \begin{cases} 0, & x < b \\ \frac{F(x) - F(a)}{F(a) - F(b)}, & b \leq x < a. \\ 1, & x \geq a \end{cases} \quad (3.27)$$

Derivando $F(x|\mathbf{x} \leq a)$ respecto a \mathbf{x} , se obtiene la densidad condicional, que es

$$f(x|b < \mathbf{x} \leq a) = \begin{cases} \frac{f(x)}{F(a) - F(b)} = \frac{f(x)}{\int_b^a f(x)dx}, & b \leq x < a. \\ 0, & \text{en otro caso} \end{cases} \quad (3.28)$$

En la figura 3.2 se ilustran los resultados cuando la variable aleatoria \mathbf{x} es normal.

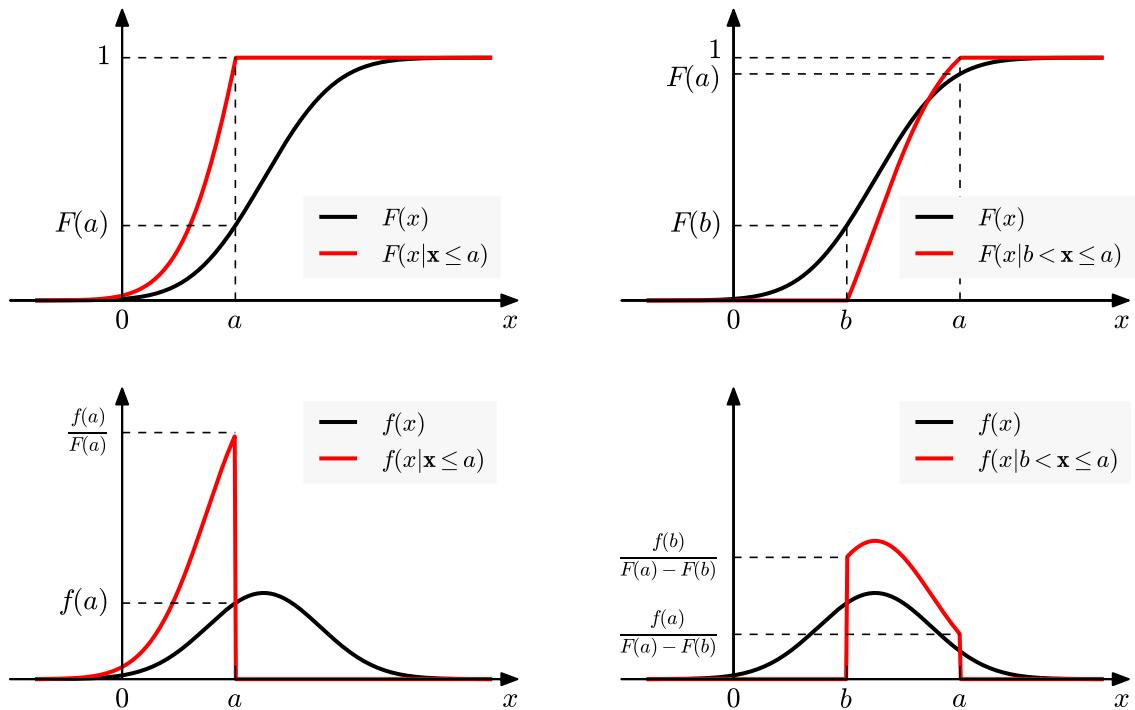


Figura 3.2: Distribución $F(x|M)$ y densidad $f(x|M)$ condicional cuando M puede expresarse en términos de \mathbf{x} . En este ejemplo, \mathbf{x} es una variable aleatoria normal, $\mathbf{x} \sim N(\mu, \sigma^2)$. En las gráficas de la izquierda se muestra la distribución y la densidad en el caso en que $M = \{\mathbf{x} \leq a\}$, dadas por las ecuaciones 3.25 y 3.26 respectivamente. Se usó $a = \mu - \sigma/2$ en el ejemplo. En las gráficas de la derecha se muestra la distribución y la densidad en el caso en que $M = \{b < \mathbf{x} \leq a\}$ dadas por las ecuaciones 3.27 y 3.28 respectivamente. Se usó $b = \mu - \sigma/2$ y $a = \mu + 3\sigma/2$ en el ejemplo.

3.4.1. Probabilidad total y el teorema de Bayes

Para extender estos resultados a variables aleatorias, se establece $B = \{\mathbf{x} \leq x\}$ en la ecuación 1.21

$$P(\{\mathbf{x} \leq x\}) = P(\mathbf{x} \leq x|A_1)P(A_1) + \cdots + P(\mathbf{x} \leq x|A_n)P(A_n)$$

y usando las definiciones de distribución y densidad condicional, se obtiene que

$$\begin{aligned} F(x) &= F(x|A_1)P(A_1) + \cdots + F(x|A_n)P(A_n) \\ f(x) &= f(x|A_1)P(A_1) + \cdots + f(x|A_n)P(A_n), \end{aligned}$$

donde los eventos A_1, \dots, A_n forman una partición del espacio muestral S . Procediendo de igual forma con la ecuación 1.22 se obtiene que

$$P\{A|\mathbf{x} \leq x\} = \frac{P\{\mathbf{x} \leq x|A\}}{P\{\mathbf{x} \leq x\}} P(A) = \frac{F(x|A)}{F(x)} P(A).$$

Estableciendo B como $\{x_1 < \mathbf{x} \leq x_2\}$ en la ecuación 1.22 se llega a que

$$\begin{aligned} P\{A|x_1 < \mathbf{x} \leq x_2\} &= \frac{P\{x_1 < \mathbf{x} \leq x_2|A\}}{P\{x_1 < \mathbf{x} \leq x_2\}} P(A) \\ &= \frac{F(x_2|A) - F(x_1|A)}{F(x_2) - F(x_1)} P(A). \end{aligned}$$

La probabilidad condicional $P\{A|\mathbf{x} = x\}$ no se puede definir a partir de la ecuación 1.22 debido a que en general $P\{\mathbf{x} = x\} = 0$, obteniendo un resultado indeterminado, pero si puede hacerse como el siguiente límite

$$\begin{aligned} P\{A|\mathbf{x} = x\} &= \lim_{\Delta x \rightarrow 0} P\{A|x < \mathbf{x} \leq x + \Delta x\} \\ &\stackrel{(a)}{=} \lim_{\Delta x \rightarrow 0} \frac{P\{x < \mathbf{x} \leq x + \Delta x|A\}}{P\{x < \mathbf{x} \leq x + \Delta x\}} P(A) \\ &\stackrel{(b)}{=} \lim_{\Delta x \rightarrow 0} \frac{\frac{P\{x < \mathbf{x} \leq x + \Delta x|A\}}{\Delta x}}{\frac{P\{x < \mathbf{x} \leq x + \Delta x\}}{\Delta x}} P(A), \end{aligned}$$

donde en (a) se usó el resultado anterior y en (b) se dividió el numerador y el denominador entre Δx y se aplicó una propiedad de límites. Finalmente, a partir de la definición de densidad y densidad condicional se obtiene que

$$P\{A|\mathbf{x} = x\} = \frac{f(x|A)}{f(x)} P(A). \quad (3.29)$$

Teorema de la probabilidad total Como se mencionó previamente,

$$F(\infty|A) = \int_{-\infty}^{\infty} f(x|A) dx = 1.$$

Por lo tanto, multiplicando la ecuación 3.29 por $f(x)$ e integrando, se obtiene que

$$\int_{-\infty}^{\infty} P\{A|\mathbf{x} = x\} f(x) dx = P(A), \quad (3.30)$$

que es la versión continua del teorema de la probabilidad total dada por la ecuación 1.21.

Teorema de Bayes Despejando $f(x|A)$ en la ecuación 3.29, se obtiene que

$$f(x|A) = \frac{P\{A|\mathbf{x} = x\}}{P(A)} f(x) = \frac{P\{A|\mathbf{x} = x\} f(x)}{\int_{-\infty}^{\infty} P\{A|\mathbf{x} = x\} f(x) dx} \quad (3.31)$$

donde en la segunda igualdad se sustituyó $P(A)$ por el resultado de la ecuación 3.30. Esta es la versión continua del teorema de Bayes dado por la ecuación 1.23.

3.4.2. Ejemplo 4-19

Sea $p = P(H)$ la probabilidad de obtener cara al tirar una moneda. Para determinada moneda, la probabilidad a priori p es desconocida y puede ser cualquier valor en el intervalo $(0, 1)$. Ante la ausencia de mas información, se debería asumir que la densidad de probabilidad $f_p(p)$ a priori es uniforme en ese intervalo. De esta forma, no se le da preferencia a algún valor de p en particular. Supóngase que se realiza el experimento de lanzar la moneda n veces y se obtiene cara k veces. Con esta nueva información, ¿cómo se actualiza $f_p(p)$?

Sea el evento $A = "k$ caras en n tiradas". Como en esas n tiradas se obtiene una secuencia específica de k caras, se cumple que

$$P(A|\mathbf{p} = p) = p^k q^{n-k} = p^k (1-p)^{n-k}.$$

Usando la ecuación 3.30, se obtiene que

$$\begin{aligned} P(A) &= \int_{-\infty}^{\infty} P\{A|\mathbf{p} = p\} f_p(p) dp \\ &\stackrel{(a)}{=} \int_0^1 p^k (1-p)^{n-k} dp \end{aligned}$$

en donde en (a) se usó el resultado anterior y que $f_p(p) = 1$ si $p \in (0, 1)$ y $f_p(p) = 0$ en otro caso. La última integral puede resolverse aplicando integración por partes k veces. Al aplicar partes una vez con $u = p^k$ y $dv = (1-p)^{n-k}$, por lo que $du = kp^{k-1}$ y $v = -(1-p)^{n-k+1}/(n-k+1)$, se obtiene que

$$\begin{aligned} \int_0^1 p^k (1-p)^{n-k} dp &= -\frac{p^k (1-p)^{n-k+1}}{n-k+1} \Big|_0^1 + \int_0^1 \frac{kp^{k-1}(1-p)^{n-k+1}}{n-k+1} dp \\ &= \frac{k}{n-(k-1)} \int_0^1 p^{k-1}(1-p)^{n-(k-1)} dp. \end{aligned}$$

Luego de aplicar partes k veces, se obtiene que

$$\begin{aligned} \int_0^1 p^k (1-p)^{n-k} dp &= \frac{k(k-1)\dots 1}{[n-(k-1)][n-(k-2)]\dots n} \int_0^1 p^0 (1-p)^n dp \\ &\stackrel{(a)}{=} \frac{(n-k)!k!}{n!} \left[-\frac{(1-p)^{n+1}}{n+1} \Big|_0^1 \right] \\ &= \frac{(n-k)!k!}{n!} \left(\frac{1}{n+1} \right) \\ &= \frac{(n-k)!k!}{(n+1)!}, \end{aligned}$$

donde en (a) se observó que $[n-(k-1)][n-(k-2)]\dots n = n!/(n-k)!$. En conclusión, se obtuvo que

$$P(A) = \frac{(n-k)!k!}{(n+1)!}.$$

La densidad de probabilidad a posteriori $f_{p|A}(p|A)$ es la información actualizada sobre p dada la ocurrencia del evento A , y puede obtenerse de la ecuación 3.31,

$$\begin{aligned} f_{p|A}(p|A) &= \frac{P\{A|\mathbf{p} = p\}}{P(A)} f_p(p) \\ &= \frac{(n+1)!}{(n-k)!k!} p^k q^{n-k}, \quad 0 < p < 1. \end{aligned} \tag{3.32}$$

Comparando el resultado con la ecuación 3.19, se deduce que $p|A \sim \text{Beta}(k+1, n-k+1)$. Observar que con estos parámetros, la función beta, dada por la ecuación 3.21, en la ecuación 3.19 es

$$B(k+1, n-k+1) = \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} = \frac{k!(n-k)!}{(n+1)!},$$

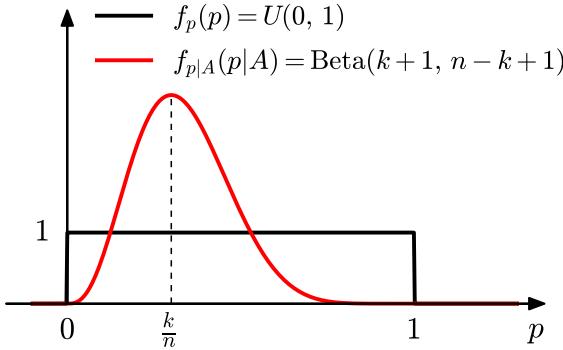


Figura 3.3: Densidad $f_p(p)$ a priori y densidad $f_{p|A}(p|A)$ a posteriori. Luego de haber realizado el experimento n veces, los resultados se reflejan en la densidad a posteriori. En el ejemplo de la figura, $n = 10$ y $k = 3$. La densidad a posteriori es $\text{Beta}(k + 1, n - k + 1)$, cuyo máximo se da en k/n .

acorde con la ecuación 3.32. Además, de la ecuación 4.21, el máximo de la densidad se da en k/n . Se obtuvo que la densidad a posteriori ya no es uniforme sino beta, es decir, $p \sim U(0, 1)$ pero $p|A \sim \text{Beta}(k + 1, n - k + 1)$. El resultado se muestra en la figura 3.3.

La densidad de probabilidad a posteriori puede ser empleada para realizar predicciones en nuevos experimentos. Por ejemplo, teniendo en cuenta el resultado del experimento anterior, puede calcularse la probabilidad de que salga cara en la tirada $n + 1$. Sea el evento B = “sale cara en la tirada $n + 1$ dado que k caras salieron en las n tiradas anteriores”. Del teorema de la probabilidad total dado por la ecuación 3.30 se tiene que

$$P(B) = \int_{-\infty}^{\infty} P\{B|\mathbf{p} = p\} f_{p|A}(p|A) dp.$$

Notar que a diferencia de la ecuación 3.30 se usó la densidad de probabilidad a posteriori de forma de reflejar el conocimiento adquirido en el experimento realizado previamente. Teniendo en cuenta que $P(B|\mathbf{p} = p) = p$, y usando de valor calculado antes de la densidad condicional, se llega a que

$$\begin{aligned} P(B) &= \int_{-\infty}^{\infty} p \frac{(n+1)!}{(n-k)!k!} p^k (1-p)^{n-k} dp \\ &= \frac{(n+1)!}{(n-k)!k!} \int_{-\infty}^{\infty} p^{k+1} (1-p)^{n-k} dp. \end{aligned}$$

La última integral puede calcularse de forma análoga a la calculada previamente. Aplicando integración por partes $k + 1$ veces, puede verse que

$$\int_{-\infty}^{\infty} p^{k+1} (1-p)^{n-k} dp = \frac{(n-k)!(k+1)!}{(n+2)!},$$

y por lo tanto, al sustituir el resultado en la ecuación anterior, se llega a que

$$P(B) = \frac{k+1}{n+2}.$$

Por ejemplo, si en $n = 10$ tiradas salieron $k = 5$ caras, la probabilidad que en la onceava tirada salga cara es $P(B) = 6/12 = 0.5$.

3.5. Aproximaciones asintóticas de una variable aleatoria binomial

Sea \mathbf{x} una variable aleatoria binomial, caracterizada por la ecuación 3.22. La probabilidad de que \mathbf{x} se encuentre entre los valores k_1 y k_2 está dada por la ecuación 2.5 y es

$$P\{k_1 \leq \mathbf{x} \leq k_2\} = \sum_{k=k_1}^{k_2} p_n(k) = \sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k}. \quad (3.33)$$

Como el coeficiente binomial

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

crece rápidamente con n , el cálculo de 3.33 puede ser difícil para valores grandes de n . Por este motivo son útiles dos aproximaciones, la aproximación normal y la aproximación de Poisson. Se verá a continuación la aproximación normal.

3.5.1. Aproximación normal de una variable aleatoria binomial

Supóngase que $n \rightarrow \infty$ y p es fijo. Como se verá en la sección 4.5.8 mediante funciones características, para k en la vecindad \sqrt{npq} de np , es válida la aproximación

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} e^{-(k-np)^2/2npq}, \quad p+q=1, \quad (3.34)$$

conocida como el teorema de DeMoivre-Laplace. Empleando este resultado se mostrará que una aproximación de $P\{k_1 \leq x \leq k_2\}$ está dada por

$$P\{k_1 \leq x \leq k_2\} = \sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k} \approx G\left(\frac{k_2 - np}{\sqrt{npq}}\right) - G\left(\frac{k_1 - np}{\sqrt{npq}}\right), \quad (3.35)$$

donde

$$G(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

es la distribución de probabilidad normal estándar. Esto indica que para calcular la probabilidad de que el número de ocurrencias del evento A en n ensayos esté entre k_1 y k_2 alcanza con evaluar la distribución normal estándar $G(x)$. La aproximación es satisfactoria si $\sigma^2 = npq \gg 1$ y k_1 y k_2 están dentro o cerca del intervalo $(np - \sqrt{npq}, np + \sqrt{npq})$. Para la demostración, se parte sustituyendo la ecuación 3.34 en la ecuación 3.33, obteniendo

$$\sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sigma \sqrt{2\pi}} \sum_{k=k_1}^{k_2} e^{-(k-np)^2/2\sigma^2}. \quad (3.36)$$

La curva normal es aproximadamente constante en cualquier intervalo de largo 1 debido a que $\sigma^2 = npq \gg 1$ por hipótesis, y por lo tanto, el área en un intervalo de largo 1 iguala aproximadamente a su ordenada. Esto implica que el lado derecho de la ecuación anterior puede aproximarse como

$$\frac{1}{\sigma \sqrt{2\pi}} \sum_{k=k_1}^{k_2} e^{-(k-np)^2/2\sigma^2} \approx \frac{1}{\sigma \sqrt{2\pi}} \int_{k_1}^{k_2} e^{-(x-np)^2/2\sigma^2} dx. \quad (3.37)$$

De esta forma

$$P\{k_1 \leq x \leq k_2\} \approx \frac{1}{\sqrt{2\pi}} \int_{k_1}^{k_2} e^{-(x-np)^2/2\sigma^2} dx = \frac{1}{\sqrt{2\pi}} \int_{y_1}^{y_2} e^{-y^2/2} dy = G(y_2) - G(y_1)$$

donde

$$y_1 = \frac{k_1 - np}{\sqrt{npq}} \quad \text{y} \quad y_2 = \frac{k_2 - np}{\sqrt{npq}},$$

concluyendo la prueba.

Corrección del error La suma en la izquierda de la ecuación 3.36 consiste en $k_2 - k_1 + 1$ términos, y la aproximación de la derecha consiste en la suma de $k_2 - k_1 + 1$ ordenadas de la densidad normal. La integral en la ecuación 3.37 es una aproximación del área sombreada entre k_1 y k_2 en la figura 3.4, que consiste en $k_2 - k_1$ rectángulos cuya área es la ordenada de la densidad normal. Si $k_2 - k_1 \gg 1$, el error es despreciable. Para valores moderados de $k_2 - k_1$ el error no es despreciable. Para reducirlo,

se reemplazan los límites de integración k_1 y k_2 por $k_1 - 1/2$ y $k_2 + 1/2$ en la ecuación 3.37, como se muestra en la figura. Esto conduce a la mejor aproximación

$$P\{k_1 \leq \mathbf{x} \leq k_2\} = \sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k} \approx G\left(\frac{k_2 + 0.5 - np}{\sqrt{npq}}\right) - G\left(\frac{k_1 - 0.5 - np}{\sqrt{npq}}\right). \quad (3.38)$$

La figura 3.4 fue generada con los parámetros $n = 30$, $p = 0.6$ y $k_1 = 15$ y $k_2 = 23$. Con esos valores, se

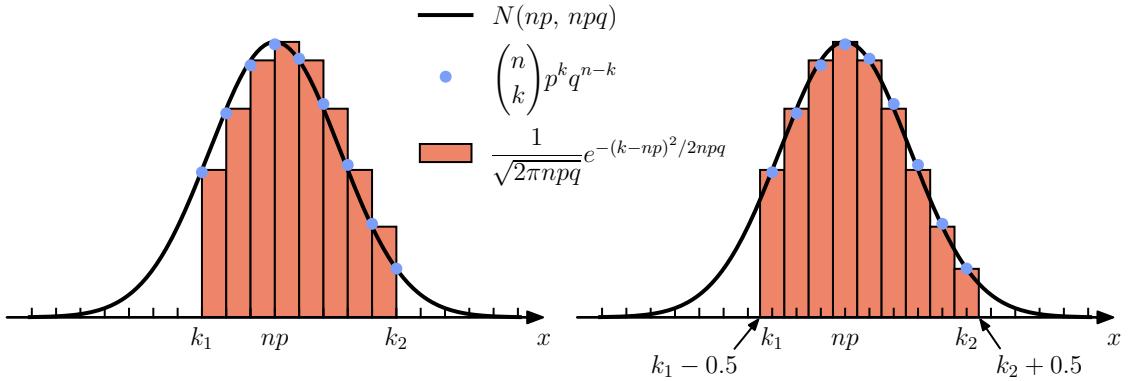


Figura 3.4: Aproximación normal de la probabilidad de un intervalo de una variable aleatoria binomial. Se muestran los valores $P\{\mathbf{x} = k\} = \binom{n}{k} p^k q^{n-k}$ para cada valor de k en el intervalo entre k_1 y k_2 de una variable aleatoria $\mathbf{x} \sim B(n, p)$. La suma de dichos valores es $P\{k_1 \leq \mathbf{x} \leq k_2\}$. Se grafica además la densidad de probabilidad gaussiana con la misma media y varianza, $N(np, npq)$. El área sombreada en la figura de la derecha es exactamente el lado derecho de la ecuación 3.36, y se puede aproximar por el área de la normal entre $k_1 - 0.5$ y $k_2 + 0.5$. Los valores empleados para generar la figura son $n = 30$, $p = 0.6$ y $k_1 = 15$ y $k_2 = 23$. De esta forma, $\eta = np = 18$ y $\sigma^2 = npq = 7.2$.

obtiene que la probabilidad exacta del intervalo es

$$P\{15 \leq \mathbf{x} \leq 23\} = \sum_{k=15}^{23} \binom{30}{k} 0.6^k 0.4^{30-k} = 0.88576,$$

mientras que la aproximación normal mediante la ecuación 3.35 es 0.83702 y la aproximación normal corregida dada por la ecuación 3.38 es 0.88375.

Capítulo 4

Funciones de una variable aleatoria

Esta sección sigue lo explicado en el capítulo 5 de [1].

4.1. Distribución de $\mathbf{y} = g(\mathbf{x})$

Sean \mathbf{x} e \mathbf{y} dos variables aleatorias tal que $\mathbf{y} = g(\mathbf{x})$, con $g(x)$ una función de variable real x . Es de interés expresar la distribución de probabilidad $F_y(y)$ de la variable aleatoria \mathbf{y} en términos de la distribución de probabilidad $F_x(x)$ de la variable aleatoria \mathbf{x} y la función $g(x)$. Para hacerlo, hay que determinar el conjunto R_y de valores de x tal que $g(x) < y$, y la probabilidad de que \mathbf{x} esté en ese conjunto.

4.1.1. Ejemplo 5-1: Distribución de $\mathbf{y} = ax + b$

Sea

$$\mathbf{y} = a\mathbf{x} + b.$$

Para determinar $F_y(y)$ hay que encontrar los valores de x de forma que $ax + b \leq y$. Se consideran dos casos en función de si a es mayor o menor que cero.

- $a > 0$: $ax + b \leq y$ si $x \leq (y - b)/a$ y por lo tanto

$$F_y(y) = P \left\{ x \leq \frac{y - b}{a} \right\} = F_x \left(\frac{y - b}{a} \right), \quad a > 0.$$

- $a < 0$: $ax + b \leq y$ si $x \geq (y - b)/a$ y por lo tanto

$$F_y(y) = P \left\{ x \geq \frac{y - b}{a} \right\} = 1 - F_x \left(\frac{y - b}{a} \right), \quad a < 0.$$

Para obtener la densidad de probabilidad $f_y(y)$, se diferencia $F_y(y)$, por lo que

- $a > 0$:

$$f_y(y) = \frac{1}{a} f_x \left(\frac{y - b}{a} \right), \quad a > 0.$$

- $a < 0$:

$$f_y(y) = -\frac{1}{a} f_x \left(\frac{y - b}{a} \right), \quad a < 0,$$

que se puede escribir en una ecuación para todo a como

$$f_y(y) = \frac{1}{|a|} f_x \left(\frac{y - b}{a} \right). \quad (4.1)$$

4.1.2. Ejemplo 5-2: Distribución de $\mathbf{y} = \mathbf{x}^2$

Considérese el caso en que

$$\mathbf{y} = \mathbf{x}^2$$

y se quiere determinar $F_y(y) = P\{\mathbf{y} \leq y\} = P\{\mathbf{x}^2 \leq y\}$

- Si $y \geq 0$, se cumple que $x^2 \leq y$ si $-\sqrt{y} \leq x \leq \sqrt{y}$ y por lo tanto

$$F_y(y) = P\{-\sqrt{y} \leq \mathbf{x} \leq \sqrt{y}\} = F_x(\sqrt{y}) - F_x(-\sqrt{y}), \quad y \geq 0.$$

- Si $y < 0$, no hay valores de x tal que $x^2 \leq y$ y por lo tanto

$$F_y(y) = P\{\emptyset\} = 0, \quad y < 0.$$

La densidad de probabilidad $f_y(y)$ puede obtenerse diferenciando $F_y(y)$, y queda

$$f_y(y) = \begin{cases} \frac{1}{2\sqrt{y}} (f_x(\sqrt{y}) + f_x(-\sqrt{y})), & y \geq 0 \\ 0, & \text{en otro caso} \end{cases},$$

En el caso en que $f_x(x)$ es una función par, es decir, $f_x(x) = f_x(-x)$, la ecuación anterior se reduce a

$$f_y(y) = \frac{1}{\sqrt{y}} f_x(\sqrt{y}) U(y), \quad (4.2)$$

donde $U(x)$ es la función escalón de Heaviside. En particular, si $\mathbf{x} \sim N(0, 1)$ y por lo tanto

$$f_x(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

sustituyendo en la ecuación 4.2, se obtiene que la distribución de la variable aleatoria $\mathbf{y} = \mathbf{x}^2$ es

$$f_y(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2} U(y). \quad (4.3)$$

Por otro lado, la densidad de probabilidad de una variable aleatoria chi-cuadrado con $n = 1$ grado de libertad es (ecuación 3.18)

$$f_z(z) = \frac{z^{-1/2}}{2^{1/2} \Gamma(1/2)} e^{z/2} U(z) = \frac{1}{\sqrt{2z} \Gamma(1/2)} e^{z/2} U(z)$$

y que con la función $\Gamma(\alpha)$ definida por la ecuación 3.13, $\Gamma(1/2)$ es

$$\begin{aligned} \Gamma(1/2) &= \int_0^{-\infty} z^{-1/2} e^{-z} dz \\ &= \int_0^{\infty} \frac{e^{-z}}{\sqrt{z}} dz \\ &\stackrel{(a)}{=} 2 \int_0^{\infty} e^{-u^2} du \\ &= \int_{-\infty}^{\infty} e^{-u^2} du \\ &\stackrel{(b)}{=} \sqrt{\pi}, \end{aligned}$$

donde en (a) se realizó el cambio de variable $u = \sqrt{z}$, por lo que $du = dz/(2\sqrt{z})$, y en (b) se notó que la integral es gaussiana (ecuación 3.10). Por lo tanto, si $\mathbf{z} \sim \chi^2(1)$, entonces

$$f_z(z) = \frac{1}{\sqrt{2\pi z}} e^{z/2} U(z).$$

Como esta distribución coincide con la ecuación 4.3, se concluye que si $\mathbf{x} \sim N(0, 1)$, entonces $\mathbf{y} = \mathbf{x}^2 \sim \chi^2(1)$.

4.2. Densidad de $\mathbf{y} = g(\mathbf{x})$

La determinación de la densidad de probabilidad de $\mathbf{y} = g(\mathbf{x})$ en función de $f_x(x)$ y $g(x)$ está dada por el *teorema fundamental*, que se plantea como sigue.

4.2.1. Teorema fundamental

Para calcular $f_y(y)$ para un valor específico de y , se resuelve la ecuación $y = g(x)$. Sean x_n las raíces reales de dicha ecuación, es decir,

$$y = g(x_1) = \dots = g(x_n) = \dots$$

El teorema fundamental indica que la densidad de probabilidad $f_y(y)$ para ese valor de y es

$$f_y(y) = \frac{f_x(x_1)}{|g'(x_1)|} + \dots + \frac{f_x(x_n)}{|g'(x_n)|} + \dots \quad (4.4)$$

Para demostrarlo, se considera que la ecuación $y = g(x)$ tiene tres raíces sin perder generalidad, como se muestra en la figura 4.1. Como indica la ecuación 3.3, se cumple que

$$f_y(y) = P\{y < \mathbf{y} \leq y + dy\}.$$

Como $\mathbf{y} = g(\mathbf{x})$, hay que encontrar el conjunto de valores de x que cumplen que $y < g(x) \leq y + dy$ y calcular la probabilidad de que \mathbf{x} esté en dicho conjunto. Observando la figura 4.1, ese conjunto consiste en los tres intervalos

$$x_1 < x \leq x_1 + dx_1, \quad x_2 + dx_2 \leq x < x_2, \quad x_3 < x \leq x_3 + dx_3,$$

con $dx_1 > 0$ y $dx_3 > 0$ pero $dx_2 < 0$. Por lo tanto, se cumple que

$$\begin{aligned} P\{y < \mathbf{y} \leq y + dy\} &= P\{x_1 < x \leq x_1 + dx_1\} \\ &\quad + P\{x_2 + dx_2 \leq x < x_2\} + P\{x_3 < x \leq x_3 + dx_3\}. \end{aligned}$$

El lado derecho de la igualdad es el área sombreada en la figura 4.1. Como

$$\begin{aligned} P\{x_1 < x \leq x_1 + dx_1\} &= f_x(x_1)dx_1 & dx_1 &= dy/g'(x_1) \\ P\{x_2 + dx_2 \leq x < x_2\} &= f_x(x_2)|dx_2| & dx_2 &= dy/g'(x_2) \\ P\{x_3 < x \leq x_3 + dx_3\} &= f_x(x_3)dx_3 & dx_3 &= dy/g'(x_3), \end{aligned}$$

sustituyendo estos resultados en la ecuación anterior, se concluye que

$$f_y(y)dy = \frac{f_x(x_1)}{|g'(x_1)|}dy + \frac{f_x(x_2)}{|g'(x_2)|}dy + \frac{f_x(x_3)}{|g'(x_3)|}dy,$$

lo que resulta en la ecuación 4.4.

A continuación se considera una aplicación que tiene cierta dificultad.

4.2.2. Ejemplo

Se quiere determinar $f_y(y)$ en el caso en que

$$\mathbf{y} = a \sin(\mathbf{x} + \theta), \quad a > 0.$$

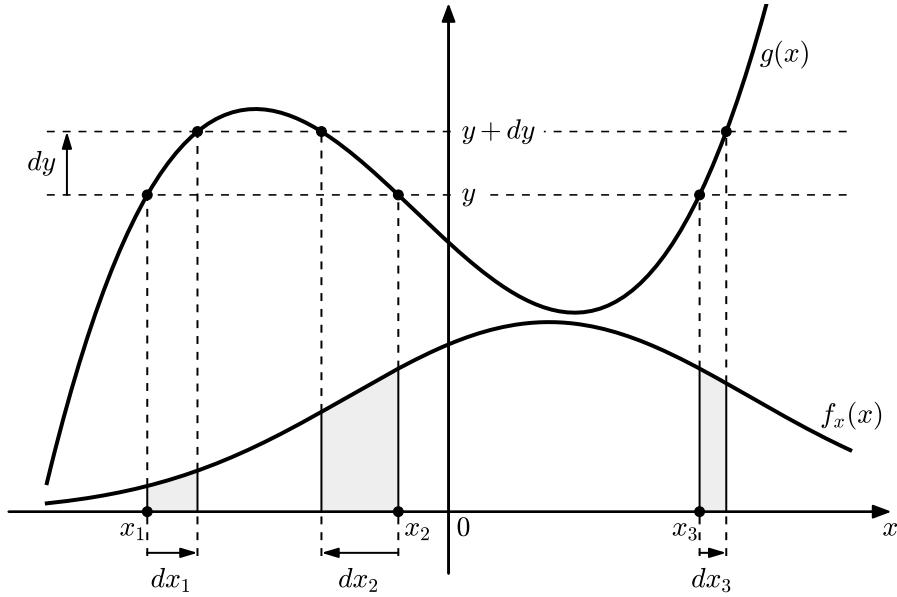


Figura 4.1: Teorema fundamental. Se grafica la función $g(x)$ y la densidad de probabilidad $f_x(x)$. La ecuación $g(x) = y$ tiene raíces x_1, x_2 y x_3 y la ecuación $g(x) = y + dy$ tiene raíces $x_1 + dx_1, x_2 + dx_2$ y $x_3 + dx_3$. Por lo tanto, $P\{y < y \leq y + dy\}$ corresponde a la probabilidad de que x esté en los intervalos $(x_1, x_1 + dx_1)$, $(x_2 + dx_2, x_2)$ y $(x_3, x_3 + dx_3)$ y es el área sombreada.

Si $|y| > a$, la ecuación $y = a \sin(x + \theta)$ no tiene soluciones, y por lo tanto $f_y(y) = 0$. Si $|y| < a$, resolviendo la ecuación $y = a \sin(x + \theta)$ se tiene que

$$\begin{aligned}
 y = a \sin(x + \theta) &\Leftrightarrow \begin{cases} y = a \sin(x + \theta + 2k\pi) \\ y = a \sin(\pi - x - \theta + 2k\pi) \end{cases} \text{ o} \\
 &\Leftrightarrow \begin{cases} x + \theta + 2k\pi = \arcsin(y/a) \\ \pi - x - \theta + 2k\pi = \arcsin(y/a) \end{cases} \text{ o} \\
 &\Leftrightarrow \begin{cases} x = \arcsin(y/a) - \theta + 2k\pi \\ x = \pi - \arcsin(y/a) - \theta + 2k\pi \end{cases} \text{ o} \\
 &\Leftrightarrow \begin{cases} x = \arcsin(y/a) - \theta + 2k\pi \\ x = -\arcsin(y/a) - \theta + (2k + 1)\pi \end{cases} \text{ o}
 \end{aligned}$$

Este resultado puede expresarse en una única ecuación como

$$x_n = (-1)^n \arcsin\left(\frac{y}{a}\right) - \theta + n\pi, \quad n = \dots, -2, -1, 0, 1, 2, \dots \quad (4.5)$$

es decir, si $|y| < a$ hay infinitas soluciones $y = g(x_n) = a \sin(x_n + \theta)$. En la figura 4.2 se ilustra la deducción de las soluciones.

Por otro lado, se tiene que

$$\begin{aligned}
 g'(x_n) &= a \cos(x_n + \theta) \\
 &\stackrel{(a)}{=} \sqrt{a^2 - a^2 \sin^2(x_n + \theta)} \\
 &= \sqrt{a^2 - y^2},
 \end{aligned}$$

donde en (a) se empleó la propiedad trigonométrica $\sin^2 x + \cos^2 x = 1$. Sustituyendo estos resultados en la ecuación 4.4, se llega a que

$$f_y(y) = \frac{1}{\sqrt{a^2 - y^2}} \sum_{n=-\infty}^{\infty} f_x(x_n), \quad |y| < a, \quad (4.6)$$

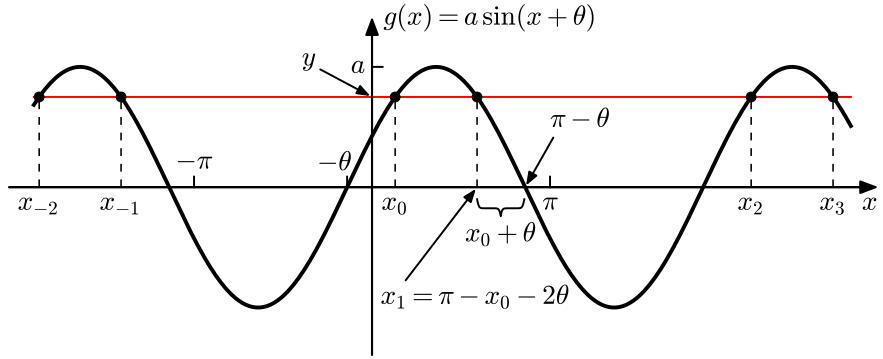


Figura 4.2: Soluciones de $y = a \sin(x + \theta)$ cuando $|y| < a$. Si x_0 es solución, también son raíces los valores $x_0 + 2k\pi$. Estas raíces son de la forma $x = \arcsin(y/a) - \theta + 2k\pi$. Además, también es solución el valor $x_1 = (\pi - \theta) - (x_0 + \theta) = \pi - x_0 - 2\theta$ y todos los valores que distan 2π de este, $x_1 + 2k\pi$. Estas raíces son de la forma $x = -\arcsin(y/a) - \theta + (2k + 1)\pi$.

con x_n dado por la ecuación 4.5.

Caso particular: considérese el caso en que \mathbf{x} es uniforme en el intervalo $(-\pi, \pi)$. La ecuación $y = a \sin(x + \theta)$ tiene únicamente dos raíces el intervalo $(-\pi, \pi)$ para cualquier valor de θ . En la figura 4.2 estas raíces son x_0 y x_1 . La función $f_x(x)$ vale $1/2\pi$ para esas dos raíces y 0 para las raíces fuera del intervalo $(-\pi, \pi)$. Por lo tanto, la sumatoria de la ecuación 4.6 tiene solo dos sumandos no nulos de valor $1/2\pi$ resultando en

$$f_y(y) = \frac{1}{\pi\sqrt{a^2 - y^2}}, \quad |y| < a.$$

Para calcular la distribución de probabilidad $F_y(y)$, se observa que $\mathbf{y} < y$ si \mathbf{x} está en el intervalo $(-\pi, x_0)$ o (x_1, π) , como se ve en la figura 4.2. El primer intervalo tiene una longitud de $x_0 + \pi$ y el segundo intervalo tiene una longitud de

$$\pi - x_1 = \pi - (\pi - x_0 - 2\theta) = x_0 + 2\theta,$$

por lo que la suma de los largos de ambos intervalos es

$$\begin{aligned} (x_0 + \pi) + (x_0 + 2\theta) &= \pi + 2x_0 + 2\theta \\ &\stackrel{(a)}{=} \pi + 2 \left[\arcsin\left(\frac{y}{a}\right) - \theta \right] + 2\theta \\ &= \pi + 2 \arcsin\left(\frac{y}{a}\right), \end{aligned}$$

donde en (a) se sustituyó x_0 usando la ecuación 4.5 con $n = 0$. Este resultado es el intervalo correspondiente al evento $\{\mathbf{y} < y\}$. Dividiendo entre el largo total del intervalo $(-\pi, \pi)$, que es 2π , se obtiene la distribución de probabilidad (casos favorables sobre casos posibles),

$$F_y(y) = P\{\mathbf{y} < y\} = \frac{1}{2} + \frac{1}{\pi} \arcsin\left(\frac{y}{a}\right).$$

4.3. Media y varianza

4.3.1. Definiciones

Media (η)	
\mathbf{x} variable aleatoria continua	\mathbf{x} variable aleatoria discreta
$E\{\mathbf{x}\} = \int_{-\infty}^{\infty} x f(x) dx$	$E\{\mathbf{x}\} = \sum_i x_i P\{\mathbf{x} = x_i\}$
Media condicional	
\mathbf{x} variable aleatoria continua	\mathbf{x} variable aleatoria discreta
$E\{\mathbf{x} M\} = \int_{-\infty}^{\infty} x f(x M) dx$	$E\{\mathbf{x} M\} = \sum_i x_i P\{\mathbf{x} = x_i M\}$
Media de $g(\mathbf{x})$	
\mathbf{x} variable aleatoria continua	\mathbf{x} variable aleatoria discreta
$E\{g(\mathbf{x})\} = \int_{-\infty}^{\infty} g(x) f_x(x) dx$	$E\{g(\mathbf{x})\} = \sum_i g(x_i) P\{\mathbf{x} = x_i\}$
Varianza ($\sigma^2 \triangleq E\{(\mathbf{x} - \eta)^2\}$)	
\mathbf{x} variable aleatoria continua	\mathbf{x} variable aleatoria discreta
$\sigma^2 = \int_{-\infty}^{\infty} (x - \eta)^2 f(x) dx$	$\sigma^2 = \sum_i (x_i - \eta)^2 P\{\mathbf{x} = x_i\}$

Algunas observaciones son las siguientes:

Media de $g(x)$ Arriba se indicó que

$$E\{g(\mathbf{x})\} = \int_{-\infty}^{\infty} g(x) f_x(x) dx. \quad (4.7)$$

Esto en realidad no es una definición, sino que debe ser demostrado. Sea la variable aleatoria $\mathbf{y} = g(\mathbf{x})$ construida a partir de una variable aleatoria \mathbf{x} y una función $g(x)$. Por definición, la media de \mathbf{y} es

$$E\{\mathbf{y}\} = \int_{-\infty}^{\infty} y f_y(y) dy. \quad (4.8)$$

Supóngase que para cierto valor de y , la función $g(x)$ tiene tres raíces x_1 , x_2 y x_3 , es decir, $y = g(x_1) = g(x_2) = g(x_3)$, como en la figura 4.1. Como se vio en la sección 4.2.1, se cumple que¹

$$f_y(y) dy = f_x(x_1) dx_1 + f_x(x_2) dx_2 + f_x(x_3) dx_3.$$

Multiplicando por y se tiene que

$$y f_y(y) dy = g(x_1) f_x(x_1) dx_1 + g(x_2) f_x(x_2) dx_2 + g(x_3) f_x(x_3) dx_3.$$

Por lo tanto, a cada diferencial en la ecuación 4.8, corresponde uno o mas diferenciales en la ecuación 4.7. Como dy abarca todo el eje y , los dx correspondientes abarcan todo el eje x y no se solapan, las ecuaciones 4.7 y 4.8 son equivalentes. Notar que entonces, para calcular la media de $\mathbf{y} = g(\mathbf{x})$ no es necesario conocer la densidad de probabilidad $f_y(y)$ de \mathbf{y} .

¹En esta demostración, a diferencia de la demostración de la sección 4.2.1, se consideran todos los dx_i positivos

Varianza Debido a que la esperanza es lineal (ecuación 5.48 mas adelante), la varianza puede expresarse como

$$\begin{aligned}
 \sigma^2 &\triangleq E\{(\mathbf{x} - \eta)^2\} \\
 &= E\{\mathbf{x}^2 - 2\mathbf{x}\eta + \eta^2\} \\
 &= E\{\mathbf{x}^2\} - 2\eta E\{\mathbf{x}\} + \eta^2 \\
 &= E\{\mathbf{x}^2\} - \eta^2,
 \end{aligned} \tag{4.9}$$

es decir,

$$\sigma^2 = E\{\mathbf{x}^2\} - E^2\{\mathbf{x}\}. \tag{4.10}$$

Transformación lineal: la varianza de la variable aleatoria $\mathbf{y} = a\mathbf{x} + b$, donde \mathbf{x} es una variable aleatoria de media η_x y varianza σ_x^2 es

$$\begin{aligned}
 \sigma_y^2 &\triangleq E\{(\mathbf{y} - \eta_y)^2\} \\
 &\stackrel{(a)}{=} E\{[(a\mathbf{x} + b) - (a\eta_x + b)]^2\} \\
 &= E\{[a(\mathbf{x} - \eta_x)]^2\} \\
 &= E\{a^2(\mathbf{x} - \eta_x)^2\} \\
 &\stackrel{(b)}{=} a^2 E\{(\mathbf{x} - \eta_x)^2\} \\
 &= a^2 \sigma_x^2,
 \end{aligned} \tag{4.11}$$

donde en (a) se empleó que $\eta_y = a\eta_x + b$ y en (b) se empleó la propiedad de linealidad de la esperanza.

4.3.2. Media en función de la distribución

Sea una variable aleatoria \mathbf{x} con distribución de probabilidad $F(x)$. Se divide el eje x en intervalos (x_k, x_{k+1}) de largo Δx como se muestra en la figura 4.3. En la figura, se observa que el área de cada región diferencial rectangular es $x_k[F(x_{k+1}) - F(x_k)] = x_k[F(x_k + \Delta x) - F(x_k)]$. Además, se cumple que

$$F(x_k + \Delta x) - F(x_k) = P\{x_k < \mathbf{x} < x_k + \Delta x\} \approx f(x_k)\Delta x.$$

Por lo tanto, la suma de todas las regiones diferenciales rectangulares es

$$\sum_k x_k [F(x_k + \Delta x) - F(x_k)] \approx \sum_k x_k f(x_k) \Delta x \rightarrow \int_{-\infty}^{\infty} x f(x) dx = E\{\mathbf{x}\}$$

cuando $\Delta x \rightarrow 0$. Notando que las regiones rectangulares correspondientes a x_k negativos contribuyen con áreas negativas, se concluye que

$$E\{\mathbf{x}\} = (1BD) - (0AB),$$

donde $(1BD)$ y $(0AB)$ son las áreas coloreadas en la figura 4.3. Por lo tanto,

$$E\{\mathbf{x}\} = \int_0^{\infty} [1 - F(x)] dx - \int_{-\infty}^0 F(x) dx.$$

Como caso particular, si una variable aleatoria toma solo valores positivos, se cumple que

$$E\{\mathbf{x}\} = \int_0^{\infty} [1 - F(x)] dx.$$

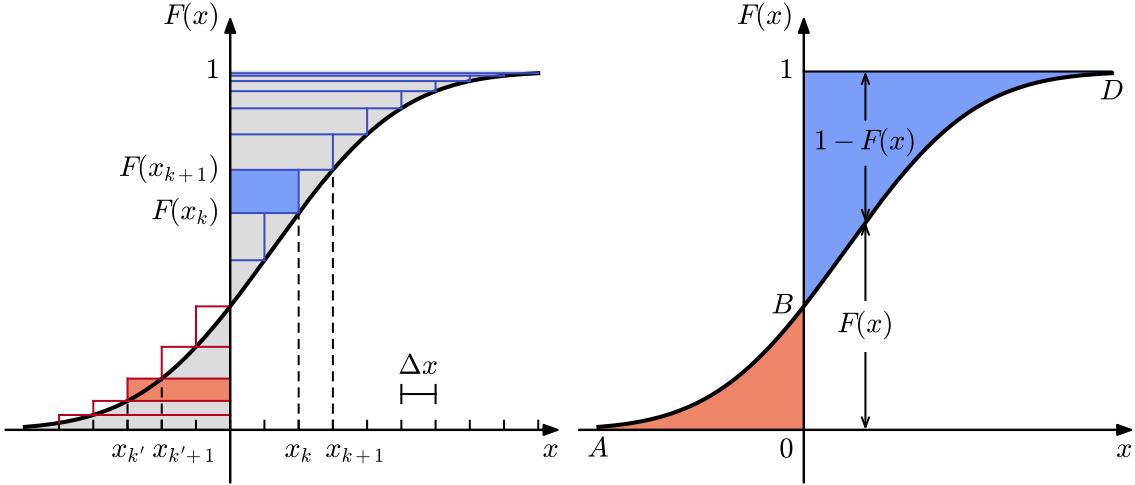


Figura 4.3: Deducción de la media de \mathbf{x} en función de la distribución de probabilidad. La media de \mathbf{x} esta dada por $(1BD) - (0AB)$, donde $(1BD)$ y $(0AB)$ son las áreas coloreadas en la figura de la derecha.

4.3.3. Ejemplo 5-18

Se quiere calcular $E\{\mathbf{x}|M\}$ con $M = \{\mathbf{x} \geq a\}$. A partir de la definición, se tiene que

$$E\{\mathbf{x}|\mathbf{x} \geq a\} = \int_{-\infty}^{\infty} x f(x|\mathbf{x} \geq a) dx.$$

Para determinar $f(x|\mathbf{x} \geq a)$ hay que determinar primero $F(x|\mathbf{x} \geq a)$ de forma similar a como se realizó en la sección 3.4 usando la definición dada por la ecuación 3.24,

$$F(x|\mathbf{x} \geq a) = P\{\mathbf{x} \leq x | \mathbf{x} \geq a\} = \frac{P\{\mathbf{x} \leq x, \mathbf{x} \geq a\}}{P\{\mathbf{x} \geq a\}},$$

- Si $x < a$, se cumple que $\{\mathbf{x} \leq x, \mathbf{x} \geq a\} = \{\emptyset\}$ y como $P\{\emptyset\} = 0$ se tiene que,

$$F(x|\mathbf{x} \geq a) = 0, \quad x < a.$$

- Si $x \geq a$, se cumple que $\{\mathbf{x} \leq x, \mathbf{x} \geq a\} = \{a \leq \mathbf{x} \leq x\}$ y por lo tanto,

$$F(x|\mathbf{x} \geq a) = \frac{P\{a \leq \mathbf{x} \leq x\}}{P\{\mathbf{x} \geq a\}} = \frac{F(x) - F(a)}{1 - F(a)}, \quad x \geq a.$$

Derivando respecto a x , se obtiene la densidad condicional, que es

$$f(x|\mathbf{x} \geq a) = \begin{cases} 0, & x < a \\ \frac{f(x)}{1 - F(a)} = \frac{f(x)}{\int_a^{\infty} f(x) dx}, & x \geq a. \end{cases}$$

Por lo tanto, se concluye que

$$E\{\mathbf{x}|\mathbf{x} \geq a\} = \frac{\int_a^{\infty} x f(x) dx}{\int_a^{\infty} f(x) dx}.$$

4.3.4. Media y varianza de una variable aleatoria normal

Una variable aleatoria normal tiene densidad de probabilidad (ecuación 3.5)

$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\eta)^2/2\sigma^2}.$$

Se verá a continuación que las constantes η y σ^2 son la media y la varianza respectivamente. Para esto, hay que demostrar que

$$\begin{aligned} E\{\mathbf{x}\} &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-(x-\eta)^2/2\sigma^2} dx = \eta \\ E\{(\mathbf{x} - \eta)^2\} &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x - \eta)^2 e^{-(x-\eta)^2/2\sigma^2} dx = \sigma^2. \end{aligned}$$

Hay varias formas de hacerlo. Una de ellas es resolver las integrales usando métodos de integración convencionales. Otras formas explotan las propiedades de la densidad de probabilidad y de la esperanza.

Se comenzará realizando la demostración resolviendo las integrales de forma convencional. En el caso de la media, se tiene que

$$\begin{aligned} E\{\mathbf{x}\} &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-(x-\eta)^2/2\sigma^2} dx \\ &\stackrel{(a)}{=} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (\sigma z + \eta) e^{-z^2/2\sigma^2} \sigma dz \\ &= \frac{1}{\sqrt{2\pi}} \left(\sigma \int_{-\infty}^{\infty} z e^{-z^2/2\sigma^2} dz + \eta \int_{-\infty}^{\infty} e^{-z^2/2\sigma^2} dz \right) \\ &\stackrel{(b)}{=} \frac{1}{\sqrt{2\pi}} \left(\sigma \cdot 0 + \eta \cdot \sqrt{2\pi} \right) \\ &= \eta, \end{aligned}$$

donde en (a) se realizó el cambio de variable $z = (x - \eta)/\sigma$, por lo que $\sigma dz = dx$, y en (b) se observó que el integrando de la primera integral es una función impar por lo que integra a cero, y la segunda integral es gaussiana como la de la ecuación 3.9 con $\sigma^2 = 1$, por lo que integra a $\sqrt{2\pi}$. Para el cálculo de la varianza, se ve que

$$\begin{aligned} E\{(\mathbf{x} - \eta)^2\} &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x - \eta)^2 e^{-(x-\eta)^2/2\sigma^2} dx \\ &\stackrel{(a)}{=} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \sigma^3 z^2 e^{-z^2/2\sigma^2} dz \\ &\stackrel{(b)}{=} \frac{\sigma^2}{\sqrt{2\pi}} \left(-ze^{-z^2/2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-z^2/2\sigma^2} dz \right) \\ &\stackrel{(c)}{=} \frac{\sigma^2}{\sqrt{2\pi}} \left(0 + \sqrt{2\pi} \right) \\ &= \sigma^2, \end{aligned}$$

donde en (a) se realizó el cambio de variable $z = (x - \eta)/\sigma$ igual que antes, en (b) se aplicó integración por partes, $\int u dv = uv - \int v du$, con $u = z$ y $dv = ze^{-z^2/2\sigma^2}$ por lo que $du = 1$ y $v = -e^{-z^2/2\sigma^2}$, y en (c) se observó que el primer término es nulo debido a que la velocidad de convergencia de $e^{-z^2/2\sigma^2} \rightarrow 0$ es mayor que la de $z \rightarrow \infty$, y la segunda integral coincide con la de la ecuación 3.9 con $\sigma^2 = 1$, por lo que integra a $\sqrt{2\pi}$.

Otra forma de realizar el mismo cálculo es la siguiente. Debido a que $f(x)$ es una densidad de probabilidad, se cumple que

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-(x-\eta)^2/2\sigma^2} dx = 1 \Leftrightarrow \int_{-\infty}^{\infty} e^{-(x-\eta)^2/2\sigma^2} dx = \sigma\sqrt{2\pi}. \quad (4.12)$$

Esto además, para el caso particular de una distribución gaussiana, se demostró en la sección 3.3.1. Para calcular la media, se diferencia la ecuación anterior respecto a η . Esto resulta en

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{(x-\eta)}{\sigma^2} e^{-(x-\eta)^2/2\sigma^2} dx &= 0 & \Rightarrow \\ \int_{-\infty}^{\infty} x e^{-(x-\eta)^2/2\sigma^2} dx &= \eta \int_{-\infty}^{\infty} e^{-(x-\eta)^2/2\sigma^2} dx & \Rightarrow \\ \int_{-\infty}^{\infty} x e^{-(x-\eta)^2/2\sigma^2} dx &= \eta \sqrt{2\pi\sigma^2} & \Rightarrow \\ \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-(x-\eta)^2/2\sigma^2} dx &= \eta, \end{aligned}$$

que es lo que se quería demostrar. Para calcular la varianza, se diferencia la ecuación 4.12 respecto a σ . De esta forma, se tiene que

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{(x-\eta)^2}{\sigma^3} e^{-(x-\eta)^2/2\sigma^2} dx &= \sqrt{2\pi} & \Rightarrow \\ \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-\eta)^2 e^{-(x-\eta)^2/2\sigma^2} dx &= \sigma^2, \end{aligned}$$

que es lo que se quería demostrar.

4.3.5. Media y varianza de una variable aleatoria Bernoulli

Como se definió en la sección 3.3.2, una variable aleatoria Bernoulli \mathbf{x} toma valores 1 y 0 con probabilidades p y $q = 1 - p$ respectivamente, es decir,

$$P\{\mathbf{x} = 1\} = p, \quad P\{\mathbf{x} = 0\} = q = 1 - p.$$

Por lo tanto,

$$E\{\mathbf{x}\} = \sum_i x_i P\{\mathbf{x} = x_i\} = 1 \times p + 0 \times q = p$$

$$E\{\mathbf{x}^2\} = \sum_i x_i^2 P\{\mathbf{x} = x_i\} = 1^2 \times p + 0^2 \times q = p.$$

Además,

$$\sigma^2 = E\{\mathbf{x}^2\} - E^2\{\mathbf{x}\} = p - p^2 = p(1 - p) = pq.$$

4.3.6. Media y varianza de una variable aleatoria Poisson

Una variable aleatoria con distribución de Poisson con parámetro λ toma valores 0, 1, ... con probabilidades

$$P\{\mathbf{x} = k\} = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Se mostrará que la media y la varianza son iguales y toman el valor λ , es decir,

$$\eta = E\{\mathbf{x}\} = \lambda, \quad E\{\mathbf{x}^2\} = \lambda^2 + \lambda, \quad \sigma^2 = E\{\mathbf{x}^2\} - E^2\{\mathbf{x}\} = \lambda.$$

Para calcular la media, se observa que

$$\begin{aligned}
\eta = E\{\mathbf{x}^2\} &= e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \\
&\stackrel{(a)}{=} e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^{n+1}}{n!} \\
&= \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \\
&\stackrel{(b)}{=} \lambda,
\end{aligned}$$

donde en (a) se realizó el cambio de variable $k = n + 1$ y en (b) consideró que la sumatoria es la expansión de Taylor de e^λ ,

$$e^\lambda = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}.$$

De forma similar, el segundo momento es

$$\begin{aligned}
E\{\mathbf{x}^2\} &= e^{-\lambda} \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} \\
&= e^{-\lambda} \sum_{n=0}^{\infty} (n+1) \frac{\lambda^{n+1}}{n!} \\
&= e^{-\lambda} \left(\lambda \sum_{n=1}^{\infty} n \frac{\lambda^n}{n!} + \lambda \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \right) \\
&= e^{-\lambda} \left(\lambda \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} + \lambda e^\lambda \right) \\
&= e^{-\lambda} \left(\lambda \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{k!} + \lambda e^\lambda \right) \\
&= e^{-\lambda} (\lambda^2 e^\lambda + \lambda e^\lambda) \\
&= \lambda^2 + \lambda,
\end{aligned}$$

culminando la prueba.

4.3.7. Ejemplo: media y mediana

Mediante el siguiente ejemplo, se pretende clarificar los conceptos de media y mediana. Se considera la variable aleatoria \mathbf{x} con densidad de probabilidad

$$f_x(x) = \begin{cases} \frac{2x}{a^2}, & 0 \leq x \leq a \\ 0, & \text{en otro caso,} \end{cases}$$

donde a es una constante positiva. Claramente, por ser una densidad de probabilidad, el área bajo su gráfica tiene que ser uno. Observando en la figura 4.4 que la gráfica es triangular con base de longitud a y altura $f_x(a) = 2/a$, se confirma que el área es uno. La media de \mathbf{x} es

$$\eta_x = E\{\mathbf{x}\} = \int_{-\infty}^{\infty} x f_x(x) dx = \frac{2}{a^2} \int_0^a x^2 dx = \frac{2}{a^2} \left(\frac{x^3}{3} \Big|_0^a \right) = \frac{2a}{3}.$$

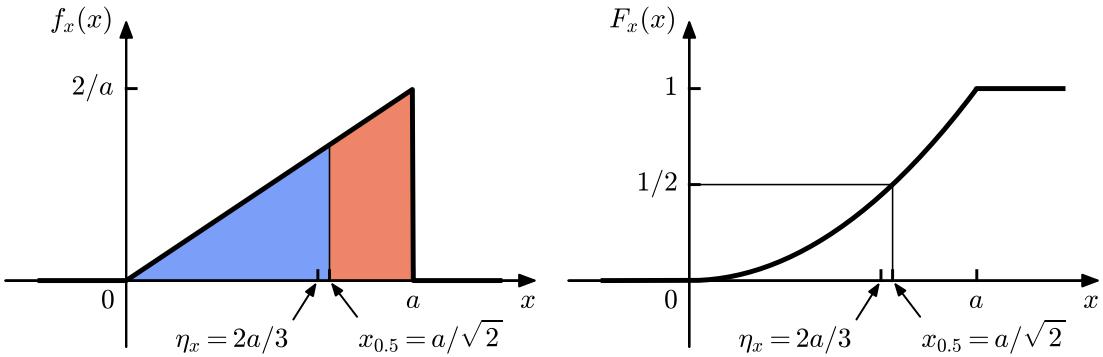


Figura 4.4: Funciones densidad y distribución de probabilidad de la variable aleatoria \mathbf{x} . Por ser una densidad no simétrica, la media y la mediana no coinciden. La mediana es el valor de x que divide a la densidad de probabilidad en áreas iguales de valor $1/2$. Asimismo, la mediana es el valor de x donde la distribución de probabilidad vale $1/2$.

Se calculará a continuación la mediana m de la variable aleatoria \mathbf{x} , que por definición es $F_x(m) = 1/2$, o equivalentemente, el percentil 0.5 de \mathbf{x} , $m = x_{0.5}$. Para hacerlo, se comienza calculando la distribución de probabilidad $F_x(x)$,

$$F_x(x) = \int_{-\infty}^x f_x(u) du = \frac{2}{a^2} \int_0^x u du = \frac{2}{a^2} \left(\frac{u^2}{2} \Big|_0^x \right) = \frac{x^2}{a^2}, \quad 0 \leq x \leq a,$$

es decir,

$$F_x(x) = \begin{cases} 0, & x < 0 \\ \left(\frac{x}{a}\right)^2, & 0 \leq x \leq a \\ 1, & x > a \end{cases}$$

como se muestra en la figura 4.4. La mediana de \mathbf{x} es $x_{0.5}$ tal que

$$F_x(x_{0.5}) = \left(\frac{x_{0.5}}{a}\right)^2 = \frac{1}{2},$$

por lo que despejando se obtiene que $x_{0.5} = a/\sqrt{2}$. Se concluye que la media y la mediana de la variable aleatoria \mathbf{x} son respectivamente

$$\eta_x = \frac{2a}{3}, \quad x_{0.5} = \frac{a}{\sqrt{2}}.$$

Notar que por definición, la mediana es el valor de x que divide la densidad de probabilidad en áreas iguales, como se muestra en la figura 4.4, o equivalentemente, el valor de x tal que la distribución de probabilidad valga $1/2$. Efectivamente, el área a la izquierda de la mediana en la densidad de probabilidad es

$$\frac{1}{2}x_{0.5}f_x(x_{0.5}) = \frac{1}{2}x_{0.5} \left(\frac{2}{a^2}x_{0.5} \right) = \left(\frac{x_{0.5}}{a} \right)^2 = \frac{1}{2},$$

donde se consideró que la superficie es triangular. Como el área bajo la densidad de probabilidad es uno, el área a la derecha de la mediana también es $1/2$.

Observar que en general, la media es distinta a la mediana. La situación particular en que son iguales es cuando la variable aleatoria \mathbf{x} tiene densidad de probabilidad simétrica respecto a algún valor de x , en cuyo caso, la media y la mediana son ese valor de simetría.

4.4. Momentos

4.4.1. Definiciones

Momentos:

$$m_n = E\{\mathbf{x}^n\} = \int_{-\infty}^{\infty} x^n f(x) dx$$

Momentos centrales:

$$\mu_n = E\{(\mathbf{x} - \eta)^n\} = \int_{-\infty}^{\infty} (x - \eta)^n f(x) dx$$

Momentos absolutos:

$$E\{|\mathbf{x}|^n\}, \quad E\{|\mathbf{x} - \eta|^n\}$$

Momentos generalizados:

$$E\{(\mathbf{x} - a)^n\}, \quad E\{|\mathbf{x} - a|^n\}$$

Consideraciones Los momentos de menor orden son

Momentos $m_0 = E\{\mathbf{x}^0\} = E\{1\} = 1$ $m_1 = E\{\mathbf{x}\} \triangleq \eta$	Momentos centrales $\mu_0 = E\{(\mathbf{x} - \eta)^0\} = E\{1\} = 1$ $\mu_1 = E\{(\mathbf{x} - \eta)\} = E\{\mathbf{x}\} - \eta = 0$ $\mu_2 = E\{(\mathbf{x} - \eta)^2\} \triangleq \sigma^2$
--	---

Los momentos centrales pueden expresarse en función de los momentos, ya que

$$\begin{aligned} \mu_n &= E\{(\mathbf{x} - \eta)^n\} \\ &\stackrel{(a)}{=} E\left\{ \sum_{k=0}^n \binom{n}{k} \mathbf{x}^k (-\eta)^{n-k} \right\} \\ &\stackrel{(b)}{=} \sum_{k=0}^n \binom{n}{k} E\{\mathbf{x}^k\} (-\eta)^{n-k}, \end{aligned}$$

donde en (a) se empleó el teorema del binomio, dado por la ecuación 2.9, y en (b) la propiedad de linealidad de la esperanza. Esto resulta en

$$\mu_n = \sum_{k=0}^n \binom{n}{k} m_k (-\eta)^{n-k}. \quad (4.13)$$

Análogamente, los momentos pueden expresarse en función de los momentos centrales,

$$\begin{aligned} m_n &= E\{\mathbf{x}^n\} \\ &= E\{[(\mathbf{x} - \eta) + \eta]^n\} \\ &= E\left\{ \sum_{k=0}^n \binom{n}{k} (\mathbf{x} - \eta)^k \eta^{n-k} \right\} \\ &= \sum_{k=0}^n \binom{n}{k} E\{(\mathbf{x} - \eta)^k\} \eta^{n-k}, \end{aligned}$$

resultando en

$$m_n = \sum_{k=0}^n \binom{n}{k} \mu_k \eta^{n-k}. \quad (4.14)$$

Por ejemplo, el momento cuarto $m_4 = E\{\mathbf{x}^4\}$ en función de los momentos centrales es

$$\begin{aligned} m_4 &= \binom{4}{0} \mu_0 \eta^4 + \binom{4}{1} \mu_1 \eta^3 + \binom{4}{2} \mu_2 \eta^2 + \binom{4}{3} \mu_3 \eta + \binom{4}{4} \mu_4 \\ &= 1 \cdot 1 \cdot \eta^4 + 4 \cdot 0 \cdot \eta^3 + 6\sigma^2 \eta^2 + 4\mu_3 \eta + \mu_4, \\ &= \eta^4 + 6\sigma^2 \eta^2 + 4\mu_3 \eta + \mu_4. \end{aligned}$$

A modo de resumen, se cumple que los primeros momentos en función de los momentos centrales son

$$\begin{aligned} m_2 &= \eta^2 + \sigma^2 \\ m_3 &= \eta^3 + 3\sigma^2 \eta + \mu_3 \\ m_4 &= \eta^4 + 6\sigma^2 \eta^2 + 4\mu_3 \eta + \mu_4, \end{aligned} \tag{4.15}$$

los cuales se deducen de la ecuación 4.14. Notar que la primera ecuación es la misma que la ecuación 4.10.

4.4.2. Momentos de $\mathbf{y} = \mathbf{x} + b$

Media En la sección 4.1.1 se calculó la densidad de probabilidad de $\mathbf{y} = a\mathbf{x} + b$ obteniendo la ecuación 4.1. En el caso en que $a = 1$ en la ecuación 4.1, se tiene que

$$\mathbf{y} = \mathbf{x} + b \quad \Rightarrow \quad f_y(y) = f_x(x - b),$$

es decir, la densidad de probabilidad de \mathbf{y} es la densidad de probabilidad de \mathbf{x} desplazada horizontalmente una cantidad b . Si $b > 0$, el desplazamiento es hacia la derecha. Evidentemente, la media de \mathbf{y} va a estar desplazada la misma cantidad. Si la media de \mathbf{x} es η_x , entonces la media η_y de \mathbf{y} es

$$\eta_y = E\{\mathbf{y}\} = E\{\mathbf{x} + b\} = E\{\mathbf{x}\} + b = \eta_x + b.$$

Momentos centrales Los momentos centrales de las variables aleatorias \mathbf{x} e $\mathbf{y} = \mathbf{x} + b$ son iguales. Efectivamente, los momentos centrales $\mu_n(y)$ de \mathbf{y} son

$$\begin{aligned} \mu_n(y) &\triangleq \int_{-\infty}^{\infty} (y - \eta_y)^n f_y(y) dy \\ &\stackrel{(a)}{=} \int_{-\infty}^{\infty} [y - (\eta_x + b)]^n f_x(y - b) dy \\ &\stackrel{(b)}{=} \int_{-\infty}^{\infty} (x - \eta_x)^n f_x(x) dx \\ &= \mu_n(x), \end{aligned}$$

donde en (a) se usó que $f_y(y) = f_x(y - b)$ y $\eta_y = \eta_x + b$ como se mencionó arriba, y en (b) se realizó el cambio de variable $x = y - b$.

4.4.3. Momentos de una variable aleatoria normal

Se demostrará a continuación que si

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$$

entonces

$$E\{\mathbf{x}^n\} = \begin{cases} 1 \cdot 3 \dots (n-1)\sigma^n, & \text{si } n = 2k \\ 0, & \text{si } n = 2k+1 \end{cases} \tag{4.16}$$

$$E\{|\mathbf{x}|^n\} = \begin{cases} 1 \cdot 3 \dots (n-1)\sigma^n, & \text{si } n = 2k \\ 2^k k! \sigma^{2k+1} \sqrt{2/\pi}, & \text{si } n = 2k+1. \end{cases} \tag{4.17}$$

Se comenzará demostrando la ecuación 4.16. Los momentos impares son nulos debido a que si n es impar, la función $x^n f(x)$ es impar por ser el producto de una función impar con una función par, por lo que su integral es nula. Esto se cumple para cualquier variable aleatoria de densidad de probabilidad par. Para demostrar el resultado con n par, se considera la siguiente identidad,

$$\int_{-\infty}^{\infty} e^{-\alpha x^2} dx = \sqrt{\frac{\pi}{\alpha}}, \quad (4.18)$$

que se verifica notando que se trata de una integral gaussiana de la forma de la ecuación 3.9 con $2\sigma^2 = 1/\alpha$. Se diferenciará la ecuación 4.18 respecto a α k veces. La derivada del integrando es

$$\begin{aligned} \frac{\partial}{\partial \alpha} (e^{-\alpha x^2}) &= -x^2 e^{-\alpha x^2} \\ \frac{\partial^2}{\partial \alpha^2} (e^{-\alpha x^2}) &= x^4 e^{-\alpha x^2} \quad \Rightarrow \quad \frac{\partial^k}{\partial \alpha^k} (e^{-\alpha x^2}) = (-1)^k x^{2k} e^{-\alpha x^2} \\ \frac{\partial^3}{\partial \alpha^3} (e^{-\alpha x^2}) &= -x^6 e^{-\alpha x^2} \end{aligned}$$

El lado derecho de la ecuación 4.18, es una constante multiplicada por $\alpha^{-1/2}$. Derivando sucesivamente $\alpha^{-1/2}$ se ve que

$$\begin{aligned} \frac{\partial}{\partial \alpha} (\alpha^{-1/2}) &= -\frac{1}{2} \alpha^{-3/2} = -\frac{1}{2} \cdot \frac{1}{\sqrt{\alpha^3}} \\ \frac{\partial^2}{\partial \alpha^2} (\alpha^{-1/2}) &= \frac{1 \cdot 3}{2^2} \alpha^{-5/2} = \frac{1 \cdot 3}{2^2} \cdot \frac{1}{\sqrt{\alpha^5}} \\ \frac{\partial^3}{\partial \alpha^3} (\alpha^{-1/2}) &= -\frac{1 \cdot 3 \cdot 5}{2^3} \alpha^{-7/2} = -\frac{1 \cdot 3 \cdot 5}{2^3} \cdot \frac{1}{\sqrt{\alpha^7}} \end{aligned}$$

y luego de derivar k veces, se obtiene que

$$\frac{\partial^k}{\partial \alpha^k} (\alpha^{-1/2}) = (-1)^k \frac{1 \cdot 3 \cdot 5 \cdots (2k-1)}{2^k} \cdot \frac{1}{\sqrt{\alpha^{2k+1}}}.$$

A partir de estos resultados, se obtiene que la derivada k -ésima de la ecuación 4.18 es

$$\int_{-\infty}^{\infty} x^{2k} e^{-\alpha x^2} dx = \frac{1 \cdot 3 \cdot 5 \cdots (2k-1)}{2^k} \sqrt{\frac{\pi}{\alpha^{2k+1}}}.$$

Sustituyendo $\alpha = 1/2\sigma^2$, se obtiene que

$$\int_{-\infty}^{\infty} x^{2k} e^{-x^2/2\sigma^2} dx = \frac{1 \cdot 3 \cdot 5 \cdots (2k-1)}{2^k} \sqrt{2^{2k+1} \sigma^{2(2k+1)} \pi}.$$

Operando el término de la raíz cuadrada, se ve que

$$\sqrt{2^{2k+1} \sigma^{2(2k+1)} \pi} = \sqrt{2 \cdot 2^{2k} \sigma^2 \sigma^{4k} \pi} = \sqrt{2\pi\sigma^2} \sqrt{2^{2k} \sigma^{4k}} = \sqrt{2\pi\sigma^2} 2^k \sigma^{2k},$$

y por lo tanto

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x^{2k} e^{-x^2/2\sigma^2} dx = 1 \cdot 3 \cdot 5 \cdots (2k-1) \sigma^{2k},$$

y con $n = 2k$ se obtiene la ecuación 4.16.

Para demostrar la ecuación 4.17, se observa que $E\{|\mathbf{x}|^{2k}\} = E\{\mathbf{x}^{2k}\}$, y por lo tanto coincide con el

resultado de la ecuación 4.16 si n es par. Si $n = 2k + 1$ impar, se ve que

$$\begin{aligned}
E\{|x|^{2k+1}\} &\stackrel{(a)}{=} \frac{2}{\sqrt{2\pi\sigma^2}} \int_0^\infty x^{2k+1} e^{-x^2/2\sigma^2} dx \\
&= \frac{1}{\sigma} \sqrt{\frac{2}{\pi}} \int_0^\infty (x^2)^k e^{-x^2/2\sigma^2} x dx \\
&\stackrel{(b)}{=} \frac{1}{\sigma} \sqrt{\frac{2}{\pi}} \int_0^\infty (2\sigma^2 y)^k e^{-y} \sigma^2 dy \\
&= 2^k \sigma^{2k+1} \sqrt{\frac{2}{\pi}} \int_0^\infty y^k e^{-y} dy \\
&\stackrel{(c)}{=} 2^k \sigma^{2k+1} \sqrt{\frac{2}{\pi}} \Gamma(k+1) \\
&\stackrel{(d)}{=} 2^k \sigma^{2k+1} \sqrt{\frac{2}{\pi}} k!,
\end{aligned}$$

donde en (a) se usó el hecho de que $|x|^{2k+1} f(x)$ es una función par, por lo que la integral entre $-\infty$ a ∞ es el doble que la integral entre 0 a ∞ , en (b) se realizó el cambio de variable $y = x^2/2\sigma^2$ por lo que $xdx = \sigma^2 dy$, en (c) se notó que la integral es la función gamma definida en la ecuación 3.13 evaluada en $k+1$ y en (d) se usó el resultado de la ecuación 3.15 que indica que $\Gamma(k+1) = k!$.

Notar además que como en este caso la variable aleatoria es de media nula, los momentos y los momentos centrales coinciden. Por lo tanto, los momentos centrales de $\mathbf{x} \sim N(0, \sigma^2)$ están dados por la ecuación 4.16.

A modo ilustrativo, se calculará el momento cuarto de una variable aleatoria $\mathbf{x} \sim N(0, \sigma^2)$ resolviendo la integral de forma similar a la realizada en la sección 4.3.4. Para hacerlo, se puede aplicar la técnica de integración por partes como sigue,

$$\begin{aligned}
E\{\mathbf{x}^4\} &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^\infty x^4 e^{-x^2/2\sigma^2} dx \\
&\stackrel{(a)}{=} \frac{1}{\sqrt{2\pi\sigma^2}} \left(-\sigma^2 x^3 e^{-x^2/2\sigma^2} \Big|_{-\infty}^\infty + 3\sigma^2 \int_{-\infty}^\infty e^{-x^2/2\sigma^2} dx \right) \\
&= 3\sigma^2 \left(\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^\infty e^{-x^2/2\sigma^2} dx \right) \\
&\stackrel{(b)}{=} 3\sigma^4,
\end{aligned}$$

donde en (a) se aplicó integración por partes, $\int u dv = uv - \int v du$, con $u = x^3$ y $dv = xe^{-x^2/2\sigma^2} dx$ por lo que $du = 3x^2 dx$ y $v = -\sigma^2 e^{-x^2/2\sigma^2}$, y en (b) se observó que el factor entre paréntesis es σ^2 , como se calculó en la sección 4.3.4.

Caso con media no nula Se calcularán a continuación los momentos de una variable aleatoria $\mathbf{x} \sim N(\eta, \sigma^2)$. Debido al argumento dado en la sección 4.4.2, los momentos centrales son iguales a los de la variable aleatoria $\mathbf{x}' \sim N(0, \sigma^2)$, ya que $\mathbf{x} = \mathbf{x}' + \eta$, y por lo tanto, están dados por la ecuación 4.16. Los primeros momentos centrales son

$$\mu_2 = \sigma^2, \quad \mu_4 = 3\sigma^4, \quad \mu_6 = 15\sigma^6 \quad (4.19)$$

Los momentos $m_n = E\{\mathbf{x}^n\}$ pueden calcularse a partir de los momentos centrales empleando la ecuación 4.14. En la ecuación 4.15 se indican los cuatro primeros momentos, y sustituyendo los momentos centrales, se obtiene que

$$\begin{aligned}
m_2 &= \eta^2 + \sigma^2 \\
m_3 &= \eta(\eta^2 + 3\sigma^2) \\
m_4 &= \eta^4 + 6\sigma^2\eta^2 + 3\sigma^4.
\end{aligned} \quad (4.20)$$

4.4.4. Momentos de una variable aleatoria beta

Se demostrará a continuación que si $\mathbf{x} \sim \text{Beta}(\alpha, \beta)$ (ver la ecuación 3.19), los momentos son [6]

$$m_n = \frac{(\alpha + n - 1)(\alpha + n - 2) \dots \alpha}{(\alpha + \beta + n - 1)(\alpha + \beta + n - 2) \dots (\alpha + \beta)}.$$

Partiendo de la definición del momento n -ésimo, se ve que

$$\begin{aligned} m_n &= E\{\mathbf{x}^n\} = \int_{-\infty}^{\infty} x^n f(x) dx \\ &= \frac{1}{B(\alpha, \beta)} \int_0^1 x^n x^{\alpha-1} (1-x)^{\beta-1} dx \\ &= \frac{1}{B(\alpha, \beta)} \int_0^1 x^{\alpha+n-1} (1-x)^{\beta-1} dx \\ &= \frac{B(\alpha + n, \beta)}{B(\alpha, \beta)} \left[\frac{1}{B(\alpha + n, \beta)} \int_0^1 x^{\alpha+n-1} (1-x)^{\beta-1} dx \right]. \end{aligned}$$

Considerando que el término entre paréntesis rectos es la integral de la densidad de una variable aleatoria $\text{Beta}(\alpha + n, \beta)$ en todo su soporte, y por lo tanto, vale 1, se obtiene que

$$m_n = E\{\mathbf{x}^n\} = \frac{B(\alpha + n, \beta)}{B(\alpha, \beta)}.$$

Empleando la ecuación 3.21, este resultado puede expresarse en términos de la función gamma como

$$m_n = \frac{\Gamma(\alpha + n)\Gamma(\beta)}{\Gamma(\alpha + \beta + n)} \times \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)},$$

resultando en que

$$m_n = E\{\mathbf{x}^n\} = \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + n)}{\Gamma(\alpha)\Gamma(\alpha + \beta + n)}$$

Además, a partir de la propiedad de la función gamma dada por la ecuación 3.14, se tiene que

$$m_n = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)} \times \frac{(\alpha + n - 1)(\alpha + n - 2) \dots \alpha \Gamma(\alpha)}{(\alpha + \beta + n)(\alpha + \beta + n - 1) \dots (\alpha + \beta) \Gamma(\alpha + \beta)},$$

resultando en que

$$m_n = E\{\mathbf{x}^n\} = \frac{(\alpha + n - 1)(\alpha + n - 2) \dots \alpha}{(\alpha + \beta + n)(\alpha + \beta + n - 1) \dots (\alpha + \beta)}.$$

Por lo tanto, la media y el segundo momento son respectivamente

$$m_1 = E\{\mathbf{x}\} = \frac{\alpha}{\alpha + \beta} \quad \text{y} \quad m_2 = E\{\mathbf{x}^2\} = \frac{(\alpha + 1)\alpha}{(\alpha + \beta + 1)(\alpha + \beta)},$$

y la varianza es

$$\begin{aligned} \sigma^2 &= E\{\mathbf{x}^2\} - E^2\{\mathbf{x}\} \\ &= \frac{(\alpha + 1)\alpha}{(\alpha + \beta + 1)(\alpha + \beta)} - \frac{\alpha^2}{(\alpha + \beta)^2} \\ &= \frac{(\alpha + 1)\alpha(\alpha + \beta) - \alpha^2(\alpha + \beta + 1)}{(\alpha + \beta + 1)(\alpha + \beta)^2}, \end{aligned}$$

resultando en

$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}.$$

Como resultado adicional, se calculará la moda. Para eso, se observa que la derivada de la densidad de probabilidad es

$$\begin{aligned} f'(x) &= \frac{1}{B(\alpha, \beta)} [(\alpha - 1)x^{\alpha-2}(1-x)^{\beta-1} - x^{\alpha-1}(\beta - 1)(1-x)^{\beta-2}] \\ &= \frac{1}{B(\alpha, \beta)} x^{\alpha-2}(1-x)^{\beta-2} [(\alpha - 1)(1-x) - x(\beta - 1)], \end{aligned}$$

e igualando a 0, se obtiene que la moda cumple que

$$(\alpha - 1)(1-x) - x(\beta - 1) = 0,$$

resultando en

$$x = \frac{\alpha - 1}{\alpha + \beta - 2}. \quad (4.21)$$

Este resultado es válido en el caso en que $\alpha > 1$ y $\beta > 1$. Si $\alpha < 1$ y $\beta < 1$, se trata de la antimoda, es decir, de la ubicación del mínimo. Se observa además que si $\alpha = \beta$, la moda es $1/2$.

4.4.5. Desigualdad de Chebyshev

Una medida de la concentración de una variable aleatoria en torno a su media η es su varianza σ^2 . El siguiente teorema muestra que la probabilidad de que la variable aleatoria \mathbf{x} tome un valor fuera del intervalo $(\eta - \epsilon, \eta + \epsilon)$ es despreciable si la razón σ/ϵ es suficientemente pequeña.

Teorema: desigualdad de Chebyshev Para cualquier $\epsilon > 0$, se cumple que

$$P\{|\mathbf{x} - \eta| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2} \quad (4.22)$$

Demostración Se parte notando que

$$\{|\mathbf{x} - \eta| \geq \epsilon\} \Leftrightarrow \{\mathbf{x} - \eta \geq \epsilon\} \cup \{-\mathbf{x} + \eta \geq \epsilon\} \Leftrightarrow \{\mathbf{x} \geq \eta + \epsilon\} \cup \{\mathbf{x} \leq \eta - \epsilon\},$$

y por lo tanto

$$P\{|\mathbf{x} - \eta| \geq \epsilon\} = \int_{-\infty}^{\eta-\epsilon} f(x) dx + \int_{\eta+\epsilon}^{\infty} f(x) dx = \int_{|\mathbf{x}-\eta| \geq \epsilon} f(x) dx.$$

Por otro lado, se tiene que

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} (x - \eta)^2 f(x) dx \\ &\stackrel{(a)}{\geq} \int_{|\mathbf{x}-\eta| \geq \epsilon} (x - \eta)^2 f(x) dx \\ &\stackrel{(b)}{\geq} \epsilon^2 \int_{|\mathbf{x}-\eta| \geq \epsilon} f(x) dx \\ &\stackrel{(c)}{=} \epsilon^2 P\{|\mathbf{x} - \eta| \geq \epsilon\}, \end{aligned}$$

donde la desigualdad en (a) proviene de que el integrando es positivo y se redujo el intervalo de integración, en (b) se notó que $(x - \eta)^2 \geq \epsilon^2$ para todo x del intervalo de integración y en (c) que la integral es $P\{|\mathbf{x} - \eta| \geq \epsilon\}$.

La importancia de la desigualdad de Chebyshev es que es válida para variables aleatorias de cualquier distribución, y por lo tanto puede ser usada incluso con variables aleatorias de distribución desconocida. Sin embargo, la cota establecida por la desigualdad, es en general demasiado alta.

4.4.6. Desigualdad de Markov

Teorema: desigualdad de Markov Si la variable aleatoria \mathbf{x} toma solo valores no-negativos, es decir, $f(x) = 0$ si $x < 0$, se cumple para todo $\alpha > 0$ que

$$P\{\mathbf{x} \geq \alpha\} \leq \frac{\eta}{\alpha}. \quad (4.23)$$

Demostración Partiendo de la definición de la media $\eta = E\{\mathbf{x}\}$ de una variable aleatoria \mathbf{x} , se tiene que

$$\begin{aligned} E\{\mathbf{x}\} &\stackrel{(a)}{=} \int_0^\infty xf(x) dx \\ &\stackrel{(b)}{\geq} \int_\alpha^\infty xf(x) dx \\ &\stackrel{(c)}{=} \alpha \int_\alpha^\infty f(x) dx, \end{aligned}$$

donde en (a) se empleó el hecho de que la variable aleatoria es no-negativa, la desigualdad en (b) proviene de que el integrando es no-negativo y se redujo el intervalo de integración y en (c) se observó que $x \geq \alpha$ para todo x en el intervalo de integración.

La desigualdad de Chebyshev, dada por la ecuación 4.22 es un caso particular de la desigualdad de Markov. Efectivamente, considerando la variable aleatoria no-negativa $|\mathbf{x} - \eta|^2$ y $\alpha = \epsilon^2$ en la desigualdad de Markov, se tiene que

$$P\{|\mathbf{x} - \eta|^2 \geq \epsilon^2\} \leq \frac{E\{|\mathbf{x} - \eta|^2\}}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2},$$

y considerando que

$$P\{|\mathbf{x} - \eta|^2 \geq \epsilon^2\} = P\{|\mathbf{x} - \eta| \geq \epsilon\},$$

se obtiene la ecuación de Chebyshev.

4.5. Funciones características

La *función característica* de una variable aleatoria \mathbf{x} con densidad de probabilidad $f(x)$ se define como

$$\Phi_x(\omega) = \int_{-\infty}^\infty f(x)e^{j\omega x} dx. \quad (4.24)$$

Como $f(x) > 0$, esta función es máxima en el origen,

$$|\Phi_x(\omega)| \leq \Phi_x(0) = 1, \quad (4.25)$$

ya que

$$|\Phi_x(\omega)| = \left| \int_{-\infty}^\infty f(x)e^{j\omega x} dx \right| \leq \int_{-\infty}^\infty |f(x)e^{j\omega x}| dx = \int_{-\infty}^\infty f(x) dx = \Phi_x(0) = 1,$$

donde se empleó la desigualdad triangular y el hecho de que la densidad de probabilidad intergrala a 1.

Si se sustituye $j\omega$ por s se obtiene la *función generadora de momentos*,

$$\Phi(s) = \int_{-\infty}^\infty f(x)e^{sx} dx. \quad (4.26)$$

La función

$$\Psi(\omega) = \ln \Phi_x(\omega) = \Psi(j\omega)$$

es la *segunda función característica* de \mathbf{x} .

Notar que se cumple que

$$\Phi(\omega) = E\{e^{j\omega \mathbf{x}}\}, \quad \Phi(s) = E\{e^{s\mathbf{x}}\}.$$

Esto conduce al hecho de que si

$$\mathbf{y} = a\mathbf{x} + b \quad \Rightarrow \quad \Phi_y(\omega) = e^{jb\omega} \Phi_x(a\omega) \quad (4.27)$$

ya que

$$\Phi_y(\omega) = E\{e^{j\omega \mathbf{y}}\} = E\{e^{j\omega(a\mathbf{x}+b)}\} = e^{jb\omega} E\{e^{ja\omega \mathbf{x}}\} = e^{jb\omega} \Phi_x(a\omega).$$

Análogamente, puede verse que la función generadora de momentos cumple que

$$\Phi_y(s) = e^{bs} \Phi_x(as). \quad (4.28)$$

4.5.1. Función característica de una variable aleatoria normal

Se demostrará que la función característica de $\mathbf{x} \sim N(\eta, \sigma^2)$ es

$$\Phi_x(\omega) = e^{j\eta\omega - \frac{1}{2}\sigma^2\omega^2}. \quad (4.29)$$

Para ver esto, se considera la variable aleatoria normalizada $\mathbf{z} = (\mathbf{x} - \eta)/\sigma$. De esta forma, $\mathbf{z} \sim N(0, 1)$, y su función generadora de momentos es

$$\Phi_z(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{sz} e^{-z^2/2} dz.$$

Observando que el exponente de la exponencial se puede escribir como

$$sz - \frac{z^2}{2} = -\frac{1}{2}(z - s)^2 + \frac{s^2}{2},$$

se concluye que

$$\Phi_z(s) = e^{s^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z-s)^2/2} dz = e^{s^2/2},$$

donde en la segunda igualdad se notó que el integrando es una densidad de probabilidad $N(s, 1)$ y por lo tanto tiene área 1. Ahora, como $\mathbf{x} = \sigma\mathbf{z} + \eta$, usando la ecuación 4.28 se llega a que

$$\Phi_x(s) = e^{\eta s} e^{\sigma^2 s^2/2} = e^{\eta s + \frac{1}{2}\sigma^2 s^2}, \quad (4.30)$$

y sustituyendo s por $j\omega$ se obtiene la función característica dada por la ecuación 4.29. Además, la segunda función característica de \mathbf{x} es

$$\Psi_x(\omega) = \ln \Phi_x(\omega) = j\eta\omega - \frac{1}{2}\sigma^2\omega^2. \quad (4.31)$$

Fórmula de inversión De la definición de la función característica dada por la ecuación 4.24, se observa que $\Phi_x(\omega)$ es la transformada de Fourier de $f(x)$. Por lo tanto, $f(x)$ puede expresarse en términos de $\Phi_x(\omega)$ a partir de la transformada inversa,

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_x(\omega) e^{-j\omega x} d\omega. \quad (4.32)$$

Inversión de la función característica de una variable aleatoria normal Como ejemplo, se invertirá la función característica de una variable aleatoria normal. Aplicando la fórmula de inversión, dada por la ecuación 4.32, a la ecuación 4.29, se tiene que

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\eta\omega - \frac{1}{2}\sigma^2\omega^2} e^{-j\omega x} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\sigma^2\omega^2 - j\omega(x-\eta)} d\omega. \quad (4.33)$$

Se expresará el exponente de la exponencial del integrando usando la técnica de “completar el cuadrado”², que consiste en observar que un polinomio de segundo grado puede expresarse como

$$\frac{1}{2}az^2 + bz + c = \frac{1}{2}a \left(z + \frac{b}{a} \right)^2 + c - \frac{b^2}{2a}. \quad (4.34)$$

Tomando

$$z = \omega, \quad a = \sigma^2, \quad b = j(x - \eta), \quad c = 0,$$

el exponente del integrando puede escribirse como

$$-\frac{1}{2}\sigma^2\omega^2 - j\omega(x - \eta) = -\frac{1}{2}\sigma^2 \left[\omega + \frac{j(x - \eta)}{\sigma^2} \right]^2 - \frac{(x - \eta)^2}{2\sigma^2}.$$

²ver https://en.wikipedia.org/wiki/Completing_the_square

De esta forma,

$$f(x) = \frac{1}{2\pi} \exp\left[-\frac{(x-\eta)^2}{2\sigma^2}\right] \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\sigma^2\left[\omega + \frac{j(x-\eta)}{\sigma^2}\right]^2\right\} d\omega. \quad (4.35)$$

Notar que el beneficio de expresar el exponente de esta forma es que un término del producto de exponentiales no depende de la variable de integración y por lo tanto puede sacarse de la integral, y el otro término es una exponencial con exponente al cuadrado. Se obtiene entonces una integral gaussiana similar a la de la ecuación 4.18, excepto que en este caso, la variable de integración está afectada por el desplazamiento complejo $j(x-\eta)/\sigma^2$. Para resolver la integral, se demostrará que [7]

$$\int_{-\infty}^{\infty} e^{-p(t+c)^2} dt = \int_{-\infty}^{\infty} e^{-pt^2} dt, \quad \text{con } p, c \in \mathbb{C}, \operatorname{Re}\{p\} > 0.$$

Considérese la curva rectangular $\Gamma_c(T)$ que se ilustra en la figura 4.5, y está formada por las rectas

$$\begin{aligned} \Gamma_{c_1}(T) : z &= x, & -T < x < T \\ \Gamma_{c_2}(T) : z &= T + iy, & 0 < y < b \\ \Gamma_{c_3}(T) : z &= x + ib, & -T < x < T \\ \Gamma_{c_4}(T) : z &= -T + iy, & 0 < y < b. \end{aligned}$$

La función $f(z) \triangleq e^{-pz^2}$ es analítica dentro de la región limitada por $\Gamma_c(T)$ y por lo tanto el teorema integral de Cauchy³ indica que

$$\oint_{\Gamma_c} f(z) dz = 0. \quad (4.36)$$

Considerando la integral en cada uno de los caminos rectos se tiene que

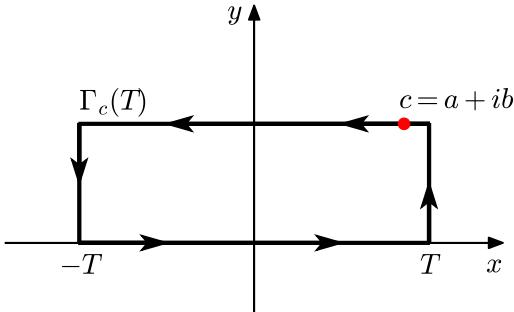


Figura 4.5: Camino de integración $\Gamma_c(T)$ de la ecuación 4.36. $\Gamma_c(T)$ es un rectángulo con un lado que coincide con el eje x y el lado opuesto pasa por el desplazamiento complejo c .

$$\begin{aligned} \oint_{\Gamma_c(T)} f(z) dz &= \oint_{\Gamma_{c_1}(T)} f(z) dz + \oint_{\Gamma_{c_2}(T)} f(z) dz + \oint_{\Gamma_{c_3}(T)} f(z) dz + \oint_{\Gamma_{c_4}(T)} f(z) dz \\ &\quad + \oint_{\Gamma_{c_4}(T)} f(z) dz \\ &= \int_{-T}^T f(x) dx + \int_0^b f(T + iy) i dy + \int_T^{-T} f(x + ib) dx \\ &\quad + \int_b^0 f(-T + iy) i dy, \end{aligned}$$

donde $x, y \in \mathbb{R}$. Tomando $|T| \rightarrow \infty$, el integrando del segundo sumando $e^{-p(T+iy)^2} \rightarrow 0$ y lo mismo ocurre con el cuarto sumando. Por lo tanto,

$$\int_{-\infty}^{\infty} e^{-p(x+ib)^2} dx = \int_{-\infty}^{\infty} e^{-px^2} dx.$$

³**Enunciado:** Sea D un subconjunto abierto de \mathbb{C} , sea $f : D \rightarrow \mathbb{C}$ una función analítica y sea Γ un contorno cerrado simple contenido en D , entonces $\oint_{\Gamma} f(z) dz = 0$.

Haciendo el cambio de variable $x = t + a = t + c - ib$ en la integral de la izquierda de la igualdad, se concluye que

$$\int_{-\infty}^{\infty} e^{-p(t+c)^2} dt = \int_{-\infty}^{\infty} e^{-pt^2} dt,$$

que es lo que se quería demostrar. Este resultado indica que la integral de la ecuación 4.35 es

$$f(x) = \frac{1}{2\pi} \exp \left[-\frac{(x-\eta)^2}{2\sigma^2} \right] \int_{-\infty}^{\infty} e^{-\sigma^2\omega^2/2} d\omega.$$

Considerando que la integral es gaussiana, usando el resultado de la ecuación 4.18, se tiene que

$$\int_{-\infty}^{\infty} e^{-\sigma^2\omega^2/2} d\omega = \sqrt{\frac{2\pi}{\sigma^2}},$$

por lo que se concluye que

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x-\eta)^2}{2\sigma^2} \right],$$

que es la densidad de probabilidad de una variable aleatoria normal de media η y varianza σ^2 .

Se explica a continuación una deducción alternativa de la inversión de la función característica de una variable aleatoria normal⁴. Partiendo de la ecuación 4.33, se ve que

$$\begin{aligned} f(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\sigma^2\omega^2} e^{-j\omega(x-\eta)} d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^0 e^{-\frac{1}{2}\sigma^2\omega^2} e^{-j\omega(x-\eta)} d\omega + \frac{1}{2\pi} \int_0^{\infty} e^{-\frac{1}{2}\sigma^2\omega^2} e^{-j\omega(x-\eta)} d\omega \\ &= -\frac{1}{2\pi} \int_{\infty}^0 e^{-\frac{1}{2}\sigma^2u^2} e^{ju(x-\eta)} du + \frac{1}{2\pi} \int_0^{\infty} e^{-\frac{1}{2}\sigma^2\omega^2} e^{-j\omega(x-\eta)} d\omega, \end{aligned}$$

donde en la última igualdad se realizó el cambio de variable $u = -\omega$ en la primer integral. Intercambiando los límites de integración y cambiando el signo en la primer integral, así como renombrando la variable de integración a ω , se tiene que

$$\begin{aligned} f(x) &= \frac{1}{2\pi} \int_0^{\infty} e^{-\frac{1}{2}\sigma^2\omega^2} e^{j\omega(x-\eta)} d\omega + \frac{1}{2\pi} \int_0^{\infty} e^{-\frac{1}{2}\sigma^2\omega^2} e^{-j\omega(x-\eta)} d\omega \\ &= \frac{1}{\pi} \int_0^{\infty} e^{-\frac{1}{2}\sigma^2\omega^2} \frac{e^{j\omega(x-\eta)} + e^{-j\omega(x-\eta)}}{2} d\omega \end{aligned}$$

Considerando la identidad $\cos x = (e^{jx} + e^{-jx})/2$, se concluye que

$$f(x) = \frac{1}{\pi} \int_0^{\infty} e^{-\frac{1}{2}\sigma^2\omega^2} \cos \omega(x-\eta) d\omega. \quad (4.37)$$

Derivando respecto a x , se tiene que

$$f'(x) = -\frac{1}{\pi} \int_0^{\infty} \omega e^{-\frac{1}{2}\sigma^2\omega^2} \sin \omega(x-\eta) d\omega,$$

y aplicando integración por partes, $\int u dv = uv - \int v du$, con $u = \sin \omega(x-\eta)$ y $dv = \omega e^{-\frac{1}{2}\sigma^2\omega^2}$ y por lo tanto $du = (x-\eta) \cos \omega(x-\eta)$ y $v = -e^{-\frac{1}{2}\sigma^2\omega^2}/\sigma^2$, se tiene que

$$\begin{aligned} f'(x) &= -\frac{1}{\pi} \left[-\frac{e^{-\frac{1}{2}\sigma^2\omega^2}}{\sigma^2} \sin \omega(x-\eta) \Big|_0^{\infty} + \int_0^{\infty} \frac{e^{-\frac{1}{2}\sigma^2\omega^2}}{\sigma^2} (x-\eta) \cos \omega(x-\eta) d\omega \right] \\ &= -\frac{(x-\eta)}{\sigma^2} \left[\frac{1}{\pi} \int_0^{\infty} e^{-\frac{1}{2}\sigma^2\omega^2} \cos \omega(x-\eta) d\omega \right], \end{aligned}$$

⁴ Esta deducción está basada en: Sasha, Characteristic function of a standard normal random variable, URL (version: 2011-12-10): <https://math.stackexchange.com/q/86048>.

donde se consideró que el primer sumando a la derecha de la primer igualdad se anula, ya que la exponencial se anula con $\omega \rightarrow \infty$ y $\sin 0 = 0$. Por lo tanto, se obtuvo que

$$f'(x) = -\frac{(x - \eta)}{\sigma^2} f(x).$$

Esta ecuación diferencial puede resolverse dividiendo la igualdad entre $f(x)$ y considerando que

$$\frac{f'(x)}{f(x)} = \frac{d}{dx} \ln f(x).$$

De esta forma, se obtiene que

$$\frac{d}{dx} \ln f(x) = -\frac{(x - \eta)}{\sigma^2}$$

e integrando en ambos lados de la igualdad se llega a que

$$\ln f(x) = -\frac{(x - \eta)^2}{2\sigma^2} + C',$$

donde C' es una constante a determinar. Despejando $f(x)$ se obtiene que

$$f(x) = C \exp \left[-\frac{(x - \eta)^2}{2\sigma^2} \right].$$

Para determinar la constante C se evalúa esta última ecuación y la ecuación 4.37 en $x = \eta$ y se igualan los resultados. De esta forma

$$\begin{aligned} f(\eta) &= C = \frac{1}{\pi} \int_0^\infty e^{-\frac{1}{2}\sigma^2\omega^2} d\omega \\ &\stackrel{(a)}{=} \frac{1}{2\pi} \int_{-\infty}^\infty e^{-\frac{1}{2}\sigma^2\omega^2} d\omega \\ &\stackrel{(b)}{=} \frac{1}{2\pi} \sqrt{\frac{2\pi}{\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}, \end{aligned}$$

donde en (a) se consideró que el integrando es una función par y por lo tanto, la integral entre 0 y ∞ es el doble que la integral entre $-\infty$ y ∞ y en (b) que es una integral gaussiana y por lo tanto se aplica el resultado de la ecuación 4.18. De esta forma, se concluye que

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x - \eta)^2}{2\sigma^2} \right].$$

Teorema de los momentos Derivando la ecuación 4.26 n veces se obtiene que

$$\Phi^{(n)}(s) = E \{ \mathbf{x}^n e^{s\mathbf{x}} \},$$

y evaluando en $s = 0$ se concluye que

$$\Phi^{(n)}(0) = E \{ \mathbf{x}^n \} = m_n, \quad (4.38)$$

es decir, las derivadas de $\Phi^{(n)}(s)$ en el origen son los momentos de la variable aleatoria \mathbf{x} . De aquí el nombre de función generadora de momentos. En particular,

$$\Phi'(0) = m_1 = \eta, \quad \Phi''(0) = m_2 = \eta^2 + \sigma^2.$$

Observación Expandiendo $\Phi(s)$ en una serie de potencias en torno al origen, se tiene que

$$\Phi(s) = \sum_{n=0}^{\infty} \frac{m_n}{n!} s^n.$$

Esta expresión es válida si todos los momentos son finitos y la serie converge absolutamente en torno al origen. Debido a que $f(x)$ puede obtenerse a partir de $\Phi(s)$, se concluye que en las condiciones indicadas, la densidad de probabilidad de una variable aleatoria queda únicamente determinada si se conocen todos sus momentos.

4.5.2. Momentos de una variable aleatoria normal

Como se vió previamente, la función generadora de momentos de una variable aleatoria normal está dada por la ecuación 4.30,

$$\Phi(s) = e^{\eta s + \frac{1}{2} \sigma^2 s^2}.$$

Por lo tanto,

$$\begin{aligned}\Phi'(s) &= (\eta + \sigma^2 s) e^{\eta s + \frac{1}{2} \sigma^2 s^2} \\ \Phi''(s) &= \sigma^2 e^{\eta s + \frac{1}{2} \sigma^2 s^2} + (\eta + \sigma^2 s)^2 e^{\eta s + \frac{1}{2} \sigma^2 s^2} \\ \Phi'''(s) &= 3\sigma^2(\eta + \sigma^2 s) e^{\eta s + \frac{1}{2} \sigma^2 s^2} + (\eta + \sigma^2 s)^3 e^{\eta s + \frac{1}{2} \sigma^2 s^2}\end{aligned}$$

y evaluando en $s = 0$ se obtiene que

$$\begin{aligned}m_1 &= \Phi'(0) = \eta \\ m_2 &= \Phi''(0) = \sigma^2 + \eta^2 \\ m_3 &= \Phi'''(0) = 3\sigma^2\eta + \eta^3\end{aligned}$$

coincidiendo con los resultados obtenidos previamente en la ecuación 4.20.

4.5.3. Momentos de una variable aleatoria gamma

Se calcularán los momentos de una variable aleatoria gamma, cuya densidad de probabilidad es

$$f_x(x) = \gamma x^{b-1} e^{-\lambda x} U(x), \quad \gamma = \frac{\lambda^b}{\Gamma(b)}, \quad (4.39)$$

donde se redefinieron los parámetros de la densidad dada en la ecuación 3.12 como $\alpha = b$ y $\beta = 1/\lambda$. A partir de la ecuación 4.26 se obtiene que la función generadora de momentos es

$$\begin{aligned}\Phi(s) &= \gamma \int_0^\infty x^{b-1} e^{-(\lambda-s)x} dx \\ &\stackrel{(a)}{=} \gamma \int_0^\infty \frac{u^{b-1}}{(\lambda-s)^{b-1}} e^{-u} \frac{du}{\lambda-s} \\ &= \frac{\gamma}{(\lambda-s)^b} \int_0^\infty u^{b-1} e^{-u} du \\ &\stackrel{(b)}{=} \frac{\gamma}{(\lambda-s)^b} \Gamma(b),\end{aligned}$$

donde en (a) se realizó el cambio de variable $u = (c-s)x$, por lo que $du = (c-s)dx$ y en (b) se notó que la integral es la función gamma, definida en la ecuación 3.13. Finalmente, se obtiene que

$$\Phi(s) = \frac{\lambda^b}{(\lambda-s)^b} = \lambda^b (\lambda-s)^{-b} \quad (4.40)$$

y derivando sucesivamente respecto a s , se ve que

$$\begin{aligned}\Phi'(s) &= \lambda^b b (\lambda-s)^{-(b+1)} \\ \Phi''(s) &= \lambda^b b (b+1) (\lambda-s)^{-(b+2)}\end{aligned}$$

Luego de derivar n veces, se obtiene que

$$\Phi^{(n)}(s) = \frac{\lambda^b b(b+1) \cdots (b+n-1)}{(\lambda-s)^{(b+n)}}, \quad (4.41)$$

y evaluando en $s = 0$, se concluye que

$$\Phi^{(n)}(0) = \frac{b(b+1) \cdots (b+n-1)}{\lambda^n} = E\{\mathbf{x}^n\}.$$

Los dos primeros momentos son

$$E\{\mathbf{x}\} = \frac{b}{\lambda}, \quad E\{\mathbf{x}^2\} = \frac{b(b+1)}{\lambda^2}, \quad \sigma^2 = E\{\mathbf{x}^2\} - E^2\{\mathbf{x}\} = \frac{b}{\lambda^2}$$

4.5.4. Momentos de una variable aleatoria chi-cuadrado

Una variable aleatoria chi-cuadrado con m grados de libertad, cuya densidad de probabilidad está dada por la ecuación 3.18, es un caso particular de una variable aleatoria gamma con parámetros $b = m/2$ y $\lambda = 1/2$ en la densidad de probabilidad de la ecuación 4.39. De esta forma, sustituyendo $b = m/2$ y $\lambda = 1/2$ en la función generadora de momentos (ecuación 4.40) se obtiene que

$$\Phi(s) = \frac{\left(\frac{1}{2}\right)^{m/2}}{\left(\frac{1}{2} - s\right)^{m/2}} = \frac{\left(\frac{1}{2}\right)^{m/2}}{\left(\frac{1}{2}\right)^{m/2} (1 - 2s)^{m/2}} = \frac{1}{(1 - 2s)^{m/2}},$$

concluyendo que la función generadora de momentos de una variable aleatoria chi-cuadrado con m grados de libertad es

$$\Phi(s) = \frac{1}{\sqrt{(1 - 2s)^m}}. \quad (4.42)$$

De la misma forma, sustituyendo $b = m/2$ y $\lambda = 1/2$ en la derivada n -ésima de la función generadora de momentos de la variable aleatoria gamma (ecuación 4.41) se obtiene que

$$\begin{aligned} \Phi^{(n)}(s) &= \frac{\left(\frac{1}{2}\right)^{m/2} \left[\frac{m}{2} \left(\frac{m}{2} + 1 \right) \cdots \left(\frac{m}{2} + n - 1 \right) \right]}{\left(\frac{1}{2} - s\right)^{m/2+n}} \\ &\stackrel{(a)}{=} \frac{\left(\frac{1}{2}\right)^{m/2+n} [m(m+2) \cdots (m+2n-2)]}{\left(\frac{1}{2}\right)^{m/2+n} (1 - 2s)^{m/2+n}} \\ &= \frac{m(m+2) \cdots (m+2n-2)}{(1 - 2s)^{(m+2n)/2}}, \end{aligned}$$

donde en (a) se sacó denominador común 2 en los n términos en el producto entre paréntesis rectos y se sacó el factor $1/2^n$ resultante fuera de los paréntesis, y se hizo lo análogo en el denominador. Se concluye que la derivada n -ésima de la función generadora de momentos de una variable aleatoria chi-cuadrado con m grados de libertad es

$$\Phi^{(n)}(s) = \frac{m(m+2) \cdots (m+2n-2)}{\sqrt{(1 - 2s)^{m+2n}}}.$$

Evaluando en $s = 0$, se llega a que

$$\Phi^{(n)}(0) = m(m+2) \cdots (m+2n-2) = E\{\mathbf{x}^n\}.$$

Por lo tanto, los dos primeros momentos son

$$E\{\mathbf{x}\} = m, \quad E\{\mathbf{x}^2\} = m(m+2),$$

y el segundo momento central es

$$\sigma^2 = E\{\mathbf{x}^2\} - E^2\{\mathbf{x}\} = m(m+2) - m^2 = 2m.$$

4.5.5. Momentos de una variable aleatoria uniforme

Sea una variable aleatoria con distribución uniforme en el intervalo (a, b) , cuya densidad de probabilidad esta dada por

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{en otro caso.} \end{cases}$$

La función generadora de momentos es

$$\Phi(s) = \frac{1}{b-a} \int_a^b e^{sx} dx = \frac{1}{b-a} \frac{e^{sx}}{s} \bigg|_a^b = \frac{e^{sb} - e^{sa}}{s(b-a)}.$$

Considérese el caso particular en que $a = -T/2$ y $b = T/2$, por lo que la variable aleatoria tiene distribución uniforme en el intervalo $(-T/2, T/2)$. En ese caso, la función generadora de momentos queda

$$\begin{aligned} \Phi(s) &= \frac{e^{sT/2} - e^{-sT/2}}{sT} \\ &\stackrel{(a)}{=} \frac{2 \sinh(sT/2)}{sT}, \end{aligned}$$

donde en (a) se escribió la expresión en términos de la función seno hiperbólico, que junto a la función coseno hiperbólico se definen respectivamente como

$$\sinh x = \frac{e^x - e^{-x}}{2} \quad \text{y} \quad \cosh x = \frac{e^x + e^{-x}}{2}.$$

Para calcular los momentos, resulta útil realizar una expansión de Taylor en torno a $s = 0$ de la función generadora de momentos. La expansión de Taylor de una función $f(x)$ en torno a $x = 0$ es

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \dots$$

De las definiciones de seno y coseno hiperbólico es fácil ver que

$$\sinh 0 = 0, \quad \cosh 0 = 1, \quad \frac{d}{dx} \sinh x = \cosh x, \quad \frac{d}{dx} \cosh x = \sinh x,$$

por lo que el desarrollo de Taylor de la función seno hiperbólico es

$$\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \dots = \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!}.$$

Aplicando este resultado, el desarrollo de Taylor de la función generadora de momentos es

$$\Phi(s) = \frac{2}{sT} \sum_{k=0}^{\infty} \frac{\left(\frac{sT}{2}\right)^{2k+1}}{(2k+1)!} = \sum_{k=0}^{\infty} \frac{\left(\frac{sT}{2}\right)^{2k}}{(2k+1)!},$$

es decir,

$$\Phi(s) = \sum_{k=0}^{\infty} \frac{T^{2k}}{2^{2k}(2k+1)!} s^{2k}. \quad (4.43)$$

Como indica la ecuación 4.38, la derivada n -ésima evaluada en cero de la función generadora de momentos, es el momento n -ésimo de la variable aleatoria. Al diferenciar $2n$ veces y evaluar en $s = 0$, el único término que sobrevive es el que corresponde al coeficiente de grado $2n$, ya que los de grado menor se anulan al diferenciar y los de grado mayor se anulan al evaluar en $s = 0$, resultando en

$$\Phi^{(2n)}(0) = \frac{T^{2n}(2n)!}{2^{2n}(2n+1)!} = \frac{T^{2n}}{2^{2n}(2n+1)},$$

donde se tuvo en cuenta además que $d^n x^n / dx^n = n!$. Se concluye que los momentos de una variable aleatoria uniforme en el intervalo $(-T/2, T/2)$ son

$$m_{2n} = E\{\mathbf{x}^{2n}\} = \frac{T^{2n}}{2^{2n}(2n+1)},$$

y los momentos impares son nulos debido a que la densidad de probabilidad es una función simétrica. Por lo tanto, los primeros momentos son

$$E\{\mathbf{x}^2\} = \frac{T^2}{2^2 \times 3} = \frac{T^2}{12}, \quad E\{\mathbf{x}^4\} = \frac{T^4}{2^3 \times 5} = \frac{T^4}{80}.$$

Cumulantes Los cumulantes λ_n de una variable aleatoria \mathbf{x} se definen como las derivadas de la función generadora de momentos segunda (también llamada función generadora de cumulantes),

$$\frac{d\Psi(0)}{ds^n} = \lambda_n,$$

y se cumple que

$$\lambda_1 = \eta, \quad \lambda_2 = \sigma^2. \quad (4.44)$$

Para mostrarlo, se observa que por definición, $\Phi = e^{\Psi}$ y por lo tanto,

$$\Phi' = \Psi' e^{\Psi}, \quad \Phi'' = [\Psi'' + (\Psi')^2] e^{\Psi}$$

Evaluando en $s = 0$ y considerando que $\Psi(0) = \lambda_0 = 0$ ya que

$$\Psi(0) = \ln \Phi(0) = \ln \left[\int_{-\infty}^{\infty} f(x) dx \right] = \ln 1 = 0$$

y teniendo en cuenta que $\Phi^{(n)}(0) = E\{\mathbf{x}^n\} = m_n$, se tiene que

$$\Phi'(0) = \Psi'(0) = m_1 = \eta$$

y

$$\Phi''(0) = \Psi''(0) + [\Psi'(0)]^2$$

por lo que

$$\Psi''(0) = \Phi''(0) - [\Psi'(0)]^2 = m_2 - m_1^2 = E\{\mathbf{x}^2\} - E^2\{\mathbf{x}\} = \sigma^2.$$

Caso discreto En el caso en que \mathbf{x} es una variable aleatoria discreta con $p_i = P\{\mathbf{x} = x_i\}$, la *función característica* se define como

$$\Phi_x(\omega) = \sum_i p_i e^{j\omega x_i}$$

La *función generadora de momentos* puede definirse igual que en el caso continuo (ecuación 4.26), pero si la variable aleatoria toma solo valores enteros, es más apropiado definirla mediante la transformada z como

$$\Gamma(z) = E\{z^{\mathbf{x}}\} = \sum_{n=-\infty}^{\infty} P\{\mathbf{x} = n\} z^n = \sum_{n=-\infty}^{\infty} p_n z^n. \quad (4.45)$$

De esta forma, $\Gamma(1/z)$ es la transformada z de la secuencia $p_n = P\{\mathbf{x} = n\}$. La función característica en este caso es

$$\Phi_x(\omega) = \sum_{n=-\infty}^{\infty} p_n e^{j\omega n},$$

y se trata de la transformada discreta de Fourier de la secuencia $\{p_n\}$. La *segunda función de momentos*, también llamada *función generadora de cumulantes*, se define como

$$\Psi(s) = \ln \Gamma(e^s). \quad (4.46)$$

El *teorema de los momentos* se deriva diferenciando la función generadora de momentos. Luego de diferenciar k veces la ecuación 4.45 se obtiene que

$$\Gamma^{(k)}(z) = E\{\mathbf{x}(\mathbf{x} - 1) \cdots (\mathbf{x} - k + 1) z^{\mathbf{x}-k}\}.$$

Evaluando en $z = 1$, se obtiene que

$$\Gamma^{(k)}(1) = E\{\mathbf{x}(\mathbf{x} - 1) \cdots (\mathbf{x} - k + 1)\},$$

y en particular,

$$\Gamma'(1) = E\{\mathbf{x}\}, \quad \Gamma''(1) = E\{\mathbf{x}^2\} - E\{\mathbf{x}\}.$$

4.5.6. Momentos de una variable aleatoria binomial

Sea $\mathbf{x} \sim B(m, p)$. La densidad de probabilidad está dada por la ecuación 2.3,

$$p_n = P\{\mathbf{x} = n\} = \binom{m}{n} p^n q^{m-n}, \quad 0 \leq n \leq m,$$

y la función generadora de momentos es

$$\Gamma(z) = \sum_{n=0}^m \binom{m}{n} p^n q^{m-n} z^n = (pz + q)^m, \quad (4.47)$$

donde en la segunda igualdad se usó el teorema del binomio dado por la ecuación 2.9. Las derivadas primeras y segunda son

$$\Gamma'(z) = mp(pz + q)^{m-1}, \quad \Gamma''(z) = m(m-1)p^2(pz + q)^{m-2}$$

y evaluando en $z = 1$ se obtiene que

$$\Gamma'(1) = mp, \quad \Gamma''(1) = m(m-1)p^2,$$

donde se usó que $p + q = 1$. Se tiene entonces que

$$E\{\mathbf{x}\} = mp$$

Además,

$$E\{\mathbf{x}^2\} = \Gamma''(1) + \Gamma'(1) = m^2 p^2 - mp^2 + mp = m^2 p^2 + mpq,$$

y

$$\sigma^2 = E\{\mathbf{x}^2\} - E^2\{\mathbf{x}\} = mpq.$$

4.5.7. Momentos de una variable aleatoria de Poisson

Si \mathbf{x} una variable aleatoria de Poisson con parámetro λ (ver la ecuación 3.23),

$$P\{\mathbf{x} = n\} = e^{-\lambda} \frac{\lambda^n}{n!}, \quad n = 0, 1, 2, \dots,$$

y la función generadora de momentos es

$$\Gamma(z) = e^{-\lambda} \sum_{n=0}^m \lambda^n \frac{z^n}{n!} = e^{-\lambda} e^{\lambda z} = e^{\lambda(z-1)}. \quad (4.48)$$

La función de momentos segunda, dada por la ecuación 4.46, es

$$\Psi(s) = \lambda(e^s - 1). \quad (4.49)$$

Observando que

$$\Psi'(0) = \lambda \quad \text{y} \quad \Psi''(0) = \lambda,$$

de la ecuación 4.44 se concluye que

$$E\{\mathbf{x}\} = \lambda \quad \text{y} \quad \sigma^2 = \lambda,$$

resultado acorde al obtenido en la sección 4.3.6.

4.5.8. Teorema de DeMoivre-Laplace

Sea $\mathbf{x} \sim B(n, p)$. Se demostrará que cuando $n \rightarrow \infty$, \mathbf{x} tiende a una distribución normal, $N(np, npq)$. La función característica de \mathbf{x} se obtiene evaluando la ecuación 4.47 en $z = e^{j\omega}$, y es

$$\Phi_x(\omega) = (pe^{j\omega} + q)^n.$$

Se considera la variable aleatoria

$$\mathbf{y} = \frac{\mathbf{x} - np}{\sqrt{npq}}$$

La función característica de \mathbf{y} se obtiene a partir de $\Phi_x(\omega)$ mediante la ecuación 4.27,

$$\begin{aligned}\Phi_y(\omega) &= e^{-jnp\omega/\sqrt{npq}} \Phi_x\left(\frac{\omega}{\sqrt{npq}}\right) \\ &= e^{-jnp\omega/\sqrt{npq}} \left(pe^{j\omega/\sqrt{npq}} + q\right)^n \\ &= \left(pe^{j\omega/\sqrt{npq}} e^{-j\omega/\sqrt{npq}} + q e^{-j\omega/\sqrt{npq}}\right)^n \\ &= \left(pe^{jq\omega/\sqrt{npq}} + q e^{-jp\omega/\sqrt{npq}}\right)^n\end{aligned}$$

Expresando las exponentiales como una serie de potencias,

$$e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!} = 1 + z + \frac{z^2}{2} + \sum_{k=3}^{\infty} \frac{z^k}{k!},$$

se tiene que

$$\begin{aligned}\Phi_y(\omega) &= \left\{ p \left[1 + \frac{jq\omega}{\sqrt{npq}} - \frac{q^2\omega^2}{2npq} + \sum_{k=3}^{\infty} \frac{1}{k!} \left(\frac{jq\omega}{\sqrt{npq}} \right)^k \right] \right. \\ &\quad \left. + q \left[1 - \frac{jp\omega}{\sqrt{npq}} - \frac{p^2\omega^2}{2npq} + \sum_{k=3}^{\infty} \frac{1}{k!} \left(\frac{-jp\omega}{\sqrt{npq}} \right)^k \right] \right\}^n \\ &= \left\{ \left[p + \frac{jpq\omega}{\sqrt{npq}} - \frac{pq^2\omega^2}{2npq} + \sum_{k=3}^{\infty} \frac{1}{k!} \left(\frac{j\omega}{\sqrt{n}} \right)^k \frac{pq^k}{(\sqrt{pq})^k} \right] \right. \\ &\quad \left. + \left[q - \frac{jpq\omega}{\sqrt{npq}} - \frac{p^2q\omega^2}{2npq} + \sum_{k=3}^{\infty} \frac{1}{k!} \left(\frac{j\omega}{\sqrt{n}} \right)^k \frac{(-p)^k q}{(\sqrt{pq})^k} \right] \right\}^n \\ &= \left\{ p + q - \frac{pq^2\omega^2}{2npq} - \frac{p^2q\omega^2}{2npq} + \sum_{k=3}^{\infty} \frac{1}{k!} \left(\frac{j\omega}{\sqrt{n}} \right)^k \frac{pq^k + (-p)^k q}{(\sqrt{pq})^k} \right\}^n.\end{aligned}$$

Observando que

- la suma de los dos primeros sumandos es $p + q = 1$.
- la suma del tercer y cuarto sumando es

$$-\frac{pq^2\omega^2}{2npq} - \frac{p^2q\omega^2}{2npq} = -\frac{q\omega^2}{2n} - \frac{p\omega^2}{2n} = -\frac{\omega^2}{2n}$$

- El último sumando se puede escribir como

$$\sum_{k=3}^{\infty} \frac{1}{k!} \left(\frac{j\omega}{\sqrt{n}} \right)^k \frac{pq^k + (-p)^k q}{(\sqrt{pq})^k} = -\frac{\omega^2}{2n} \left[2 \sum_{k=3}^{\infty} \frac{1}{k!} \left(\frac{j\omega}{\sqrt{n}} \right)^{k-2} \frac{pq^k + (-p)^k q}{(\sqrt{pq})^k} \right]$$

y definiendo

$$\phi(n) \triangleq 2 \sum_{k=3}^{\infty} \frac{1}{k!} \left(\frac{j\omega}{\sqrt{n}} \right)^{k-2} \frac{pq^k + (-p)^k q}{(\sqrt{pq})^k}$$

se llega a que

$$\Phi_y(\omega) = \left\{ 1 - \frac{\omega^2}{2n} [1 + \phi(n)] \right\}^n \rightarrow e^{-\omega^2/2} \quad \text{con } n \rightarrow \infty$$

ya que $\phi(n) \rightarrow 0$ con $n \rightarrow \infty$ y por definición,

$$e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n} \right)^n.$$

Comparando la ecuación característica de \mathbf{y} con la ecuación 4.29 se concluye que con $n \rightarrow \infty$ \mathbf{y} tiende a la distribución normal estándar y por lo tanto, \mathbf{x} tiende a $N(np, npq)$.

4.6. Ejemplo 5-32: Probabilidad de sucesión de éxitos

Cierto evento A ocurre con probabilidad p constante en ensayos independientes. Si A ocurre al menos r veces de forma sucesiva, se dice que se produjo una racha de largo r . Se quiere calcular la probabilidad de obtener una racha de largo r en n ensayos. Este problema es de interés histórico, ya que fue propuesto y resuelto por Abraham de Moivre [1]. La solución está basada en la inversión de la función generadora de momentos.

Sea p_n la probabilidad del evento $X_n = \{\text{racha de largo } r \text{ de } A \text{ en } n \text{ ensayos}\}$. Una racha de largo r en $n+1$ ensayos puede ocurrir de dos formas mutuamente excluyentes: o ocurre una racha de largo r en los n primeros ensayos (X_n) o la racha ocurre en los últimos r de los $n+1$ ensayos y no antes (B_{n+1}). De esta forma

$$X_{n+1} = X_n \cup B_{n+1} \quad (4.50)$$

donde

$$\begin{aligned} B_{n+1} &= \{\text{No ocurre racha de largo } r \text{ de } A \text{ en los primeros } n-r \text{ ensayos}\} \\ &\cap \{A \text{ no ocurre en el } (n-r+1)\text{-ésimo ensayo}\} \\ &\cap \{\text{Corrida de largo } r \text{ de } A \text{ en los últimos } r \text{ ensayos}\} \\ &= \overline{X}_{n-r} \cap \overline{A} \cap \underbrace{A \cap A \cap \cdots \cap A}_r \end{aligned}$$

y por la independencia de estos eventos, se tiene que

$$P\{B_{n+1}\} = (1 - p_{n-r})qp^r.$$

y de la ecuación 4.50, se concluye que

$$p_{n+1} = P\{X_{n+1}\} = P\{X_n\} + P\{B_{n+1}\} = p_n + (1 - p_{n-r})qp^r.$$

Esta ecuación se trata de una ecuación en diferencias, y las condiciones iniciales son

$$p_0 = p_1 = \cdots = p_{r-1} = 0 \quad \text{y} \quad p_r = p^r,$$

ya que no puede haber una racha de largo r en $n < r$ ensayos y la probabilidad de una racha de largo r en r ensayos es p^r . La ecuación en diferencias puede resolverse recursivamente obteniendo

$$p_{r+1} = p^r(1+q), p_{r+2} = p^r(1+2q), \dots, p_{r+m} = p^r(1+mq)$$

para $m \leq r$, pero la recursión se hace complicada para valores grandes de n . Se empleará el método de las funciones generadoras de momentos para obtener una expresión analítica general para p_n . La ecuación en diferencias, reemplazando n por $n+r$ puede escribirse como

$$p_{n+r+1} = p_{n+r} + (1 - p_n)qp^r.$$

y definiendo

$$q_n \triangleq 1 - p_n,$$

que es la probabilidad de que no ocurra una racha de largo r en n ensayos, la ecuación en diferencias queda

$$1 - q_{n+r+1} = 1 - q_{n+r} + q_n q p^r$$

y reordenando términos se llega a la ecuación en diferencias

$$q_{n+r+1} = q_{n+r} - q p^r q_n \quad n \geq 0 \quad (4.51)$$

con las nuevas condiciones iniciales

$$q_0 = q_1 = \cdots = q_{r-1} = 1 \quad \text{y} \quad q_r = (1 - p^r). \quad (4.52)$$

A partir de la ecuación 4.45 se define la función generadora de momentos,

$$\phi(z) = \sum_{n=0}^{\infty} q_n z^n.$$

Despejando q_n de la ecuación 4.51 se tiene que,

$$q_n = \frac{1}{q p^r} (q_{n+r} - q_{n+r+1})$$

y sustituyendo en la función generadora de momentos, se obtiene que

$$\phi(z) = \frac{1}{q p^r} \sum_{n=0}^{\infty} (q_{n+r} - q_{n+r+1}) z^n,$$

es decir,

$$q p^r \phi(z) = \sum_{n=0}^{\infty} q_{n+r} z^n + \sum_{n=0}^{\infty} q_{n+r+1} z^n. \quad (4.53)$$

Operando con la primer sumatoria del lado derecho de la igualdad, se ve que

$$\sum_{n=0}^{\infty} q_{n+r} z^n \stackrel{(a)}{=} \sum_{k=r}^{\infty} q_k z^{k-r} = \frac{1}{z^r} \sum_{k=r}^{\infty} q_k z^k \stackrel{(b)}{=} \frac{1}{z^r} \left[\phi(z) - \sum_{k=0}^{r-1} q_k z^k \right],$$

donde en (a) se realizó el cambio de variable $k = n + r$ y en (b) se notó que la sumatoria es igual a la de la función generadora de momentos pero comenzando en r . Haciendo las operaciones análogas con la segunda sumatoria de la ecuación 4.53, se llega a que

$$\sum_{n=0}^{\infty} q_{n+r+1} z^n = \frac{1}{z^{r+1}} \left[\phi(z) - \sum_{k=0}^r q_k z^k \right].$$

Sustituyendo estos resultados en la ecuación 4.53, se obtiene que

$$\begin{aligned} q p^r \phi(z) &= \frac{\phi(z) - \sum_{k=0}^{r-1} q_k z^k}{z^r} - \frac{\phi(z) - \sum_{k=0}^r q_k z^k}{z^{r+1}} \\ &\stackrel{(a)}{=} \frac{\phi(z) - \sum_{k=0}^{r-1} q_k z^k}{z^r} - \frac{\phi(z) - \left[\sum_{k=0}^{r-1} q_k z^k + q_r z^r \right]}{z^{r+1}} \\ &\stackrel{(b)}{=} \frac{\phi(z) - \sum_{k=0}^{r-1} z^k}{z^r} - \frac{\phi(z) - \left[\sum_{k=0}^{r-1} z^k + (1 - p^r) z^r \right]}{z^{r+1}} \\ &\stackrel{(c)}{=} \frac{z \phi(z) - \sum_{k=0}^{r-1} z^{k+1} - \phi(z) + \sum_{k=0}^{r-1} z^k + (1 - p^r) z^r}{z^{r+1}} \\ &\stackrel{(d)}{=} \frac{(z - 1) \phi(z) - \sum_{n=1}^r z^n + \sum_{k=0}^{r-1} z^k + (1 - p^r) z^r}{z^{r+1}}, \end{aligned}$$

donde en (a) se separó la sumatoria que va de 0 a r en una sumatoria que va de 0 a $r - 1$ mas el término r , en (b) se sustituyó q_k por su valores correspondientes dados por las condiciones iniciales de la ecuación 4.52, en (c) se sacó denominador común y en (d) se reagruparon términos y se realizó el cambio de variable $n = k + 1$ en la primer sumatoria. Observando que entre las sumatorias se cancelan todos los términos desde 1 a $r - 1$ se obtiene que

$$\begin{aligned} qp^r \phi(z) &= \frac{(z - 1)\phi(z) + 1 - z^r + (1 - p^r)z^r}{z^{r+1}} \\ &= \frac{(z - 1)\phi(z) + 1 - p^r z^r}{z^{r+1}}. \end{aligned}$$

Finalmente, despejando $\phi(z)$ se obtiene que la función generadora de momentos es

$$\phi(z) = \frac{1 - p^r z^r}{1 - z + qp^r z^{r+1}}.$$

$\phi(z)$ es una función racional en z , y el coeficiente de z^n en su serie de potencias es q_n . Con el fin de expresar $\phi(z)$ como una serie de potencias, se escribe como

$$\phi(z) = (1 - p^r z^r) \left[\frac{1}{1 - z(1 - qp^r z^r)} \right].$$

Considerando la igualdad⁵

$$\frac{1}{1 - z} = \sum_{m=0}^{\infty} z^m, \quad |z| < 1, \quad (4.54)$$

el término entre paréntesis rectos se puede expandir como

$$\begin{aligned} \frac{1}{1 - z(1 - qp^r z^r)} &= \sum_{m=0}^{\infty} z^m (1 - qp^r z^r)^m \\ &\stackrel{(a)}{=} \sum_{m=0}^{\infty} z^m \sum_{k=0}^m \binom{m}{k} 1^{n-k} (-qp^r z^r)^k \\ &= \sum_{m=0}^{\infty} z^m \sum_{k=0}^m \binom{m}{k} (-1)^k (qp^r)^k z^{kr} \\ &= \sum_{m=0}^{\infty} \sum_{k=0}^m \binom{m}{k} (-1)^k (qp^r)^k z^{m+kr}, \end{aligned}$$

donde en (a) se aplicó el teorema del binomio (ecuación 2.9). Realizando el cambio de variable $n = m + kr$, se llega a que

$$\frac{1}{1 - z(1 - qp^r z^r)} = \sum_{n=0}^{\infty} \sum_{k=0}^{\lfloor n/(r+1) \rfloor} \binom{n-k}{k} (-1)^k (qp^r)^k z^n.$$

El límite inferior de la sumatoria en n es 0 ya que $n = 0$ cuando $m = 0$ y $k = 0$. El límite superior de la sumatoria en k es $m = n - kr$ evaluado en $k = m$, y el resultado es la parte entera de $n/(r - 1)$. Este límite es consistente con la condición $n - kr \geq k$ para que $\binom{n-k}{k}$ esté bien definido. Definiendo

$$\alpha_{n,r} = \sum_{k=0}^{\lfloor n/(r+1) \rfloor} \binom{n-k}{k} (-1)^k (qp^r)^k \quad (4.55)$$

la ecuación anterior se puede expresar como

$$\frac{1}{1 - z(1 - qp^r z^r)} = \sum_{n=0}^{\infty} \alpha_{n,r} z^n,$$

⁵ver https://en.wikipedia.org/wiki/Geometric_series

y la función generadora de momentos es

$$\begin{aligned}\phi(z) &= (1 - p^r z^r) \sum_{n=0}^{\infty} \alpha_{n,r} z^n \\ &= \sum_{n=0}^{\infty} \alpha_{n,r} z^n - \sum_{n=0}^{\infty} p^r \alpha_{n,r} z^{n+r}.\end{aligned}$$

Operando con la segunda sumatoria, se tiene que

$$\sum_{n=0}^{\infty} p^r \alpha_{n,r} z^{n+r} \stackrel{(a)}{=} \sum_{k=r}^{\infty} p^r \alpha_{k-r,r} z^k \stackrel{(b)}{=} \sum_{k=0}^{\infty} p^r \alpha_{k-r,r} z^k,$$

donde en (a) se realizó el cambio de variable $k = n + r$ y en (b) se consideró que $\alpha_{n,r} = 0$ si $n < 0$. Por lo tanto,

$$\begin{aligned}\phi(z) &= \sum_{n=0}^{\infty} \alpha_{n,r} z^n - \sum_{n=0}^{\infty} p^r \alpha_{n-r,r} z^{n+r} \\ &= \sum_{n=0}^{\infty} (\alpha_{n,r} - p^r \alpha_{n-r,r}) z^n,\end{aligned}$$

concluyendo que

$$q_n = \alpha_{n,r} - p^r \alpha_{n-r,r},$$

por lo que

$$p_n = 1 - q_n = 1 - \alpha_{n,r} + p^r \alpha_{n-r,r},$$

con $\alpha_{n,r}$ definido por la ecuación 4.55.

4.7. Ejemplo 5-33: El problema de las parejas

El problema de las parejas se formula como sigue. Una persona escribe n cartas y las correspondientes direcciones en n sobres. Luego, introduce cada carta en un sobre de forma aleatoria. ¿Cuál es la probabilidad de que al menos una carta llegue al destino correcto? ¿Qué ocurre cuando $n \rightarrow \infty$?

Otra formulación clásica del problema es con personas con sombreros. n personas con sombrero van a una fiesta. Durante la fiesta ingieren alcohol en grandes cantidades, y al retirarse, cada una toma un sombrero de forma aleatoria. ¿Cuál es la probabilidad de que al menos uno tenga el sombrero correcto?

Este problema fue propuesto por Pierre Raymond de Montmort en 1708, y resuelto por el mismo Montmort en 1710 en una carta dirigida a Johann Bernoulli [8]. En la literatura, es usual referirse al problema como *Montmort's matching problem*. El problema se reduce en calcular *desarreglos* (derangement⁶), esto es, el número de permutaciones de n objetos en donde ningún objeto está en su posición original. El número de desarreglos se suele denotar con el símbolo *subfactorial*, que consiste en el signo de exclamación del factorial pero a la izquierda del número, $!n$. A continuación se desarrolla la solución usando la formulación de las cartas y los sobres tal como aparece en el libro, pero se agrega la deducción de las ecuaciones (5-140) y (5-141), que el libro omite y no me parecen nada triviales.

4.7.1. Solución del problema

Cuando una carta es introducida en el sobre con la dirección correcta, se dirá que es una *coincidencia*. Sea X_k el evento “hay exactamente k coincidencias en los n sobres”. Los eventos X_0, X_1, \dots, X_n forman una partición del espacio muestral, ya que son eventos mutuamente excluyentes y alguno debe ocurrir. Por lo tanto,

$$p_n(0) + p_n(1) + p_n(2) + \dots + p_n(n) = 1, \quad (4.56)$$

⁶ver <https://en.wikipedia.org/wiki/Derangement>

donde $p_n(k) \triangleq P\{X_k\}$. Para determinar $p_n(k)$, se observa que hay $\binom{n}{k}$ formas posibles de que haya k coincidencias, cada una con probabilidad

$$\frac{1}{n} \cdot \frac{1}{n-1} \cdots \frac{1}{n-k+1},$$

y además, en los $n - k$ sobres restantes no debe haber ninguna coincidencia, evento de probabilidad $p_{n-k}(0)$. Por lo tanto,

$$p_n(k) = P\{X_k\} = \binom{n}{k} \frac{1}{n(n-1)\cdots(n-k+1)} p_{n-k}(0),$$

y usando que $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, se obtiene que

$$p_n(k) = \frac{p_{n-k}(0)}{k!}. \quad (4.57)$$

Por lo tanto, para calcular $p_n(k)$ hay que calcular $p_{n-k}(0)$, para lo cual se necesita contar la cantidad de permutaciones de $n - k$ objetos sin ninguna coincidencia de posición. Observar que en el caso en que todas las cartas coincidan con su sobre, es decir, cuando $k = n$, se obtiene que $p_n(n) = p_0(0)/n!$, y como $p_n(n) = 1/n!$, se concluye que $p_0(0) = 1$.

Sustituyendo la ecuación 4.57 en 4.56 término a término, se obtiene que

$$p_n(0) + \frac{p_{n-1}(0)}{1!} + \frac{p_{n-2}(0)}{2!} + \cdots + \frac{p_1(0)}{(n-1)!} + \frac{p_0(0)}{n!} = 1, \quad (4.58)$$

a partir de lo cual es posible calcular los valores de $p_n(0)$ recursivamente,

$$\begin{aligned} p_1(0) &= 1 - \frac{p_0(0)}{1!} = 1 - 1 = 0, \\ p_2(0) &= 1 - \frac{p_1(0)}{1!} - \frac{p_0(0)}{2!} = 1 - 0 - \frac{1}{2} = \frac{1}{2}, \\ p_3(0) &= 1 - \frac{p_2(0)}{1!} - \frac{p_1(0)}{2!} - \frac{p_0(0)}{3!} = 1 - \frac{1}{2} - 0 - \frac{1}{6} = \frac{1}{3}, \end{aligned}$$

y así sucesivamente. A partir de $p_n(0)$ es posible calcular $p_n(k)$ usando la ecuación 4.57. Se obtendrá a continuación una expresión explícita para $p_n(0)$ a partir de la función generadora de momentos $\phi(z)$, definida como

$$\phi(z) = \sum_{n=0}^{\infty} p_n(0) z^n. \quad (4.59)$$

El desarrollo de esta parte se omite en el libro, pero puede encontrarse en [9] bajo el nombre de *Inversión Binomial*. No he encontrado mas fuentes en la bibliografía que enfrenten el problema usando este enfoque. En [8] también se resuelve el problema mediante la función generadora de momentos pero tampoco se detalla el desarrollo.

La ecuación 4.58 puede escribirse como

$$\sum_{k=0}^n \frac{p_k(0)}{(n-k)!} = 1.$$

Multiplicando por z^n y sumando para todo $n \geq 0$, se obtiene que

$$\sum_{n=0}^{\infty} \left(\sum_{k=0}^n \frac{p_k(0)}{(n-k)!} \right) z^n = \sum_{n=0}^{\infty} z^n. \quad (4.60)$$

El lado derecho de la igualdad es una serie geométrica de converge a $1/(1-z)$ si $|z| < 1$. Operando con

el lado izquierdo, se tiene que

$$\begin{aligned}
\sum_{n=0}^{\infty} \left(\sum_{k=0}^n \frac{p_k(0)}{(n-k)!} \right) z^n &= \sum_{n=0}^{\infty} \sum_{k=0}^n p_k(0) z^k \frac{z^{n-k}}{(n-k)!} \\
&\stackrel{(a)}{=} \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} p_k(0) z^k \frac{z^{n-k}}{(n-k)!} \\
&= \sum_{k=0}^{\infty} p_k(0) z^k \sum_{n=k}^{\infty} \frac{z^{n-k}}{(n-k)!} \\
&\stackrel{(b)}{=} \sum_{k=0}^{\infty} p_k(0) z^k \sum_{i=0}^{\infty} \frac{z^i}{i!} \\
&\stackrel{(c)}{=} \phi(z) e^z,
\end{aligned} \tag{4.61}$$

donde en (a) se intercambió el orden de las sumatorias manteniendo la suma sobre los mismos pares de valores (n, k) en la mitad del cuadrante de la sumatoria original, como se ilustra en la figura 4.6, en (b) se realizó el cambio de variable $i = n - k$ y en (c) se notó que el primer factor es la función generadora de momentos definida en la ecuación 4.59 y el segundo factor es la representación en serie de potencias de la función exponencial. De las ecuaciones 4.60 y 4.61, se deduce que la función generadora de momentos

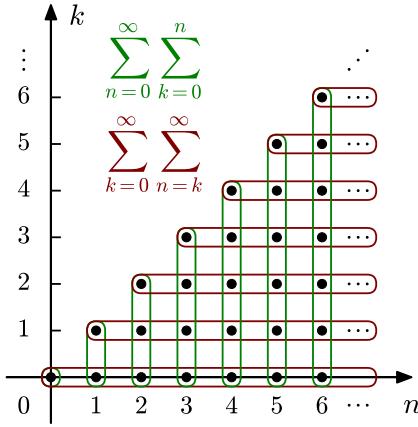


Figura 4.6: Equivalencia de dobles sumatorias cuando el par de valores (n, k) consisten en los números naturales de la mitad del cuadrante.

es

$$\phi(z) = \frac{e^{-z}}{1-z}.$$

Además, puede demostrarse que

$$\frac{e^{-z}}{1-z} = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n \frac{(-1)^k}{k!} \right) z^n. \tag{4.62}$$

Para hacerlo, se procede de forma análoga a como se realizó con la deducción de la función generadora

de momentos. Comenzando con el lado derecho de la igualdad y operando, se tiene que,

$$\begin{aligned}
\sum_{n=0}^{\infty} \left(\sum_{k=0}^n \frac{(-1)^k}{k!} \right) z^n &= \sum_{n=0}^{\infty} \sum_{k=0}^n \frac{(-1)^k z^k}{k!} z^{n-k} \\
&= \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} \frac{(-1)^k z^k}{k!} z^{n-k} \\
&= \sum_{k=0}^{\infty} \frac{(-1)^k z^k}{k!} \sum_{n=k}^{\infty} z^{n-k} \\
&= \sum_{k=0}^{\infty} \frac{(-1)^k z^k}{k!} \sum_{i=0}^{\infty} z^i \\
&= e^{-z} \frac{1}{1-z},
\end{aligned}$$

en donde en la última igualdad se tomó en cuenta que el primer factor del producto de sumatorias es la serie de potencias de e^{-z} . Por lo tanto, se concluye que

$$\phi(z) = \frac{e^{-z}}{1-z} = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n \frac{(-1)^k}{k!} \right) z^n.$$

Comparando esta ecuación con la ecuación 4.59, se obtiene que la expresión de $p_n(0)$ que se buscaba,

$$p_n(0) = \sum_{k=0}^n \frac{(-1)^k}{k!}. \quad (4.63)$$

Notar que

$$e^{-z} = \sum_{n=0}^{\infty} \frac{(-1)^n z^n}{n!} \quad \Rightarrow \quad e^{-1} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!}$$

y por lo tanto,

$$\lim_{n \rightarrow \infty} p_n(0) = \frac{1}{e}.$$

Finalmente, sustituyendo la ecuación 4.63 en la ecuación 4.57 se obtiene la probabilidad de que exactamente k cartas de las n lleguen al destino correcto, y es

$$p_n(k) = \frac{1}{k!} \sum_{m=0}^{n-k} \frac{(-1)^m}{m!}.$$

Con esto, queda completamente caracterizado el evento X_k . La respuesta a la pregunta realizada inicialmente de cual es la probabilidad de que al menos una de las n cartas llegue al destino correcto es

$$\begin{aligned}
P\{\text{Al menos una carta llegue al destino correcto}\} &= 1 - p_n(0) \\
&= 1 - \sum_{k=0}^n \frac{(-1)^k}{k!} \\
&\rightarrow 1 - \frac{1}{e} \approx 0.63212056
\end{aligned}$$

Además, como $p_n(0)$ es la probabilidad de que no haya ninguna coincidencia en los n sobres, se calcula como el cociente entre la cantidad de permutaciones donde no hay ninguna coincidencia (desarreglos, $!n$), y la cantidad de permutaciones posibles,

$$p_n(0) = \frac{!n}{n!}$$

y por lo tanto, la cantidad de desarreglos de n objetos es

$$!n = n! \sum_{k=0}^n \frac{(-1)^k}{k!}. \quad (4.64)$$

4.7.2. Otras consideraciones

El resultado de la ecuación 4.62 puede obtenerse usando la siguiente propiedad de las funciones generadoras de momentos. Sean $\phi_a(z)$ y $\phi_b(z)$ las funciones generadoras de las variables aleatorias discretas X e Y respectivamente, con $P\{X = n\} = a_n$ y $P\{Y = n\} = b_n$ para $-\infty < n < \infty$. La función generadora de $c_n = (a * b)(n)$ es

$$\phi_c(z) = \phi_a(z)\phi_b(z) = \sum_{n=-\infty}^{\infty} c_n z^n,$$

donde

$$c_n = (a * b)(n) = \sum_{k=-\infty}^{\infty} a_k b_{n-k},$$

es la convolución entre las secuencias a_n y b_n [10]. Esto significa que la función generadora de la convolución de secuencias es el producto de las funciones generadoras de las secuencias. Tener en cuenta que en este contexto, una secuencia es la densidad de probabilidad de una variable aleatoria discreta. Este resultado proviene de la multiplicación de polinomios:

$$\begin{aligned} f(x) &= \sum_{n=-\infty}^{\infty} a_n x^n & h(x) &= f(x)g(x) = \sum_{n=-\infty}^{\infty} c_n x^n, \\ \text{si} & & \Rightarrow & \\ g(x) &= \sum_{m=-\infty}^{\infty} b_m x^m & \text{con} & c_n = (a * b)(n). \end{aligned}$$

En el caso en que el soporte de la secuencias sean los naturales, es decir, $n \geq 0$, se cumple que

$$\phi_c(z) = \phi_a(z)\phi_b(z) = \sum_{n=0}^{\infty} c_n z^n, \quad \text{con} \quad c_n = (a * b)(n) = \sum_{k=0}^n a_k b_{n-k}.$$

Teniendo esto en cuenta y volviendo a la ecuación 4.62, sean

$$\phi_a(z) = e^{-z} = \sum_{n=0}^{\infty} \frac{(-1)^n z^n}{n!} \quad \text{y} \quad \phi_b(z) = \frac{1}{1-z} = \sum_{n=0}^{\infty} z^n,$$

Las secuencias de estas funciones generadoras son

$$a_n = \begin{cases} \frac{(-1)^n}{n!}, & \text{si } n \geq 0 \\ 0, & \text{si } n < 0 \end{cases} \quad \text{y} \quad b_n = \begin{cases} 1, & \text{si } n \geq 0 \\ 0, & \text{si } n < 0 \end{cases},$$

respectivamente. Por lo tanto, la secuencia de la función generadora

$$\phi_c(z) = \phi_a(z)\phi_b(z) = \frac{e^{-z}}{1-z}$$

es

$$c_n = (a * b)(n) = \sum_{k=0}^n \frac{(-1)^k}{k!},$$

ya que b_n es el escalón de Heaviside, y la convolución de una secuencia por un escalón de Heaviside es la suma acumulada de la secuencia. Se concluye que

$$\phi_c(z) = \frac{e^{-z}}{1-z} = \sum_{n=0}^{\infty} \sum_{k=0}^n \frac{(-1)^k}{k!} z^n,$$

de forma acorde con la ecuación 4.62

El problema de las parejas también puede resolverse recursivamente de la siguiente forma [9]. Sea D_n la cantidad de desarreglos de las n cartas ($D_n \equiv !n$), es decir, la cantidad de formas posibles que se pueden poner las cartas en los sobres de forma de que no haya ninguna coincidencia. Llámese carta 1 a la primera carta. Hay $n - 1$ formas posibles de poner la carta en un sobre sin que haya coincidencia. Supóngase que la carta 1 se pone en el sobre i , con $i \neq 1$. Ahora se consideran dos casos: a) la carta correspondiente al sobre i se guarda en el sobre 1, y b) la carta del sobre i se guarda en otro sobre distinto del 1. Entonces,

- a) el resto de las $n - 2$ cartas se guarda en un sobre incorrecto, y la cantidad de formas posibles de hacerlo es D_{n-2} .
- b) cada una de las $n - 1$ cartas tiene una opción prohibida. La opción prohibida de la carta i es el sobre 1, y la opción prohibida de las otras cartas es su propio sobre. Por lo tanto, la cantidad de formas posibles de guardar las $n - 1$ cartas sin que haya coincidencia en estas condiciones es D_{n-1} .

Como los eventos a) y b) son mutuamente excluyentes y abarcan todas las posibilidades, se cumple que

$$D_n = (n - 1)(D_{n-1} + D_{n-2}).$$

Además, se tiene como condiciones iniciales que $D_0 = 1$ y $D_1 = 0$. Ahora, la probabilidad de que no haya ninguna coincidencia es

$$\begin{aligned} P_n &= \frac{D_n}{n!} \\ &= \frac{n-1}{n!} (D_{n-1} + D_{n-2}) \\ &= (n-1) \left[\frac{D_{n-1}}{n(n-1)!} + \frac{D_{n-2}}{n(n-1)(n-2)!} \right] \\ &\stackrel{(a)}{=} (n-1) \left[\frac{P_{n-1}}{n} + \frac{P_{n-2}}{n(n-1)} \right] \\ &= \left(1 - \frac{1}{n} \right) P_{n-1} + \frac{1}{n} P_{n-2}, \end{aligned}$$

donde en (a) se usó que $P_k = D_k/k!$. Finalmente, se obtiene que

$$P_n = P_{n-1} - \frac{1}{n} (P_{n-1} - P_{n-2})$$

con las condiciones iniciales $P_0 = 1$ y $P_1 = 0$. Resolviendo recursivamente, se ve que

$$\begin{aligned} P_2 &= P_1 - \frac{1}{2} (P_1 - P_0) = 0 + \frac{1}{2} = \frac{1}{2!} \\ P_3 &= P_2 - \frac{1}{3} (P_2 - P_1) = \frac{1}{2!} - \frac{1}{3} \times \frac{1}{2} = \frac{1}{2!} - \frac{1}{3!} \\ P_4 &= P_3 - \frac{1}{4} (P_3 - P_2) = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4} \times \frac{1}{3!} = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} \\ &\vdots \\ P_n &= \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \cdots + \frac{(-1)^n}{n!}, \end{aligned}$$

es decir,

$$P_n = \sum_{k=2}^n \frac{(-1)^k}{k!} = \sum_{k=0}^n \frac{(-1)^k}{k!},$$

en concordancia con el resultado obtenido mediante la función generadora de momentos de la ecuación 4.63, y

$$D_n = n! P_n = n! \sum_{k=0}^n \frac{(-1)^k}{k!}.$$

Capítulo 5

Dos variables aleatorias

Continuando con el estudio del libro [1], en lo que sigue se incluye lo desarrollado en el capítulo 6.

5.1. Algunas definiciones y propiedades

Se incluyen a continuación algunas definiciones y propiedades debido a que serán mencionadas más adelante.

5.1.1. Distribución de probabilidad conjunta

La distribución de probabilidad conjunta $F_{xy}(x, y)$, o simplemente $F(x, y)$, de dos variables aleatorias \mathbf{x} e \mathbf{y} es la probabilidad del evento

$$\{\mathbf{x} \leq x, \mathbf{y} \leq y\} = \{(\mathbf{x}, \mathbf{y}) \in D_1\}$$

donde x e y son dos números reales arbitrarios y D_1 es el cuadrante mostrado en la figura 5.1. Entonces,

$$F(x, y) = P\{\mathbf{x} \leq x, \mathbf{y} \leq y\}.$$

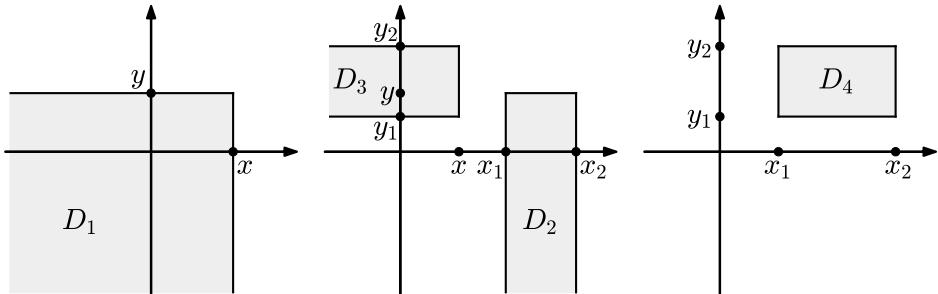


Figura 5.1: Regiones del plano xy correspondientes a distintos eventos. La gráfica de la izquierda resalta los puntos del plano del evento $\{\mathbf{x} \leq x, \mathbf{y} \leq y\} = \{(\mathbf{x}, \mathbf{y}) \in D_1\}$. En la gráfica del centro, la semi-banda vertical D_2 son los puntos del plano del evento $\{x_1 < \mathbf{x} \leq x_2, \mathbf{y} \leq y\}$ y la semi-banda horizontal D_3 son los puntos del plano del evento $\{\mathbf{x} \leq x, y_1 < \mathbf{y} \leq y_2\}$. La gráfica de la derecha consiste en los puntos de la región D_4 , correspondiente al evento $\{x_1 < \mathbf{x} \leq x_2, y_1 < \mathbf{y} \leq y_2\}$.

La distribución de probabilidad cumple las siguientes propiedades:

1. La función $F(x, y)$ es tal que

$$F(-\infty, y) = 0, \quad F(x, -\infty) = 0, \quad F(\infty, \infty) = 1.$$

Demostración: las dos primeras igualdades indican que

$$\begin{aligned} F(-\infty, y) &= P\{\mathbf{x} = -\infty, \mathbf{y} \leq y\} = 0 \\ F(x, -\infty) &= P\{\mathbf{x} \leq x, \mathbf{y} = -\infty\} = 0. \end{aligned} \quad (5.1)$$

Como se cumple que

$$P\{\mathbf{x} = -\infty\} = P\{\mathbf{y} = -\infty\} = 0$$

y además

$$\{\mathbf{x} = -\infty, \mathbf{y} \leq y\} \subset \{\mathbf{x} = -\infty\}, \quad \{\mathbf{x} \leq x, \mathbf{y} = -\infty\} \subset \{\mathbf{y} = -\infty\},$$

queda demostrada la ecuación 5.1. La última igualdad surge de que

$$F(\infty, \infty) = P\{\mathbf{x} \leq \infty, \mathbf{y} \leq \infty\},$$

y de observar que

$$\{\mathbf{x} \leq \infty, \mathbf{y} \leq \infty\} = S, \quad P(S) = 1,$$

es decir, el evento $\{\mathbf{x} \leq \infty, \mathbf{y} \leq \infty\}$ es todo el espacio muestral, y por lo tanto, su probabilidad de ocurrencia es 1.

2. El evento $\{x_1 < \mathbf{x} \leq x_2, \mathbf{y} \leq y\}$ consiste en todos los puntos (\mathbf{x}, \mathbf{y}) en la semi-banda vertical D_2 y el evento $\{\mathbf{x} \leq x, y_1 < \mathbf{y} \leq y_2\}$ consiste en todos los puntos (\mathbf{x}, \mathbf{y}) en la semi-banda horizontal D_3 en la figura 5.1. La probabilidad de estos eventos es

$$\begin{aligned} P\{x_1 < \mathbf{x} \leq x_2, \mathbf{y} \leq y\} &= F(x_2, y) - F(x_1, y) \\ P\{\mathbf{x} \leq x, y_1 < \mathbf{y} \leq y_2\} &= F(x, y_2) - F(x, y_1). \end{aligned} \quad (5.2)$$

Demostración: observando que

$$\{\mathbf{x} \leq x_2, \mathbf{y} \leq y\} = \{\mathbf{x} \leq x_1, \mathbf{y} \leq y\} \cup \{x_1 < \mathbf{x} \leq x_2, \mathbf{y} \leq y\},$$

y teniendo en cuenta que los eventos a la derecha de la igualdad son mutuamente excluyentes, se cumple que

$$P\{\mathbf{x} \leq x_2, \mathbf{y} \leq y\} = P\{\mathbf{x} \leq x_1, \mathbf{y} \leq y\} + P\{x_1 < \mathbf{x} \leq x_2, \mathbf{y} \leq y\},$$

y por lo tanto

$$\begin{aligned} P\{x_1 < \mathbf{x} \leq x_2, \mathbf{y} \leq y\} &= P\{\mathbf{x} \leq x_2, \mathbf{y} \leq y\} - P\{\mathbf{x} \leq x_1, \mathbf{y} \leq y\} \\ &= F(x_2, y) - F(x_1, y), \end{aligned}$$

que es lo que se quería demostrar. La segunda igualdad en la ecuación 5.2 se demuestra de forma análoga.

3. Se cumple que

$$P\{x_1 < \mathbf{x} \leq x_2, y_1 < \mathbf{y} \leq y_2\} = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1), \quad (5.3)$$

que es la probabilidad de que (\mathbf{x}, \mathbf{y}) esté en el rectángulo D_4 de la figura 5.1. Demostración: Como

$$\{x_1 < \mathbf{x} \leq x_2, \mathbf{y} \leq y_2\} = \{x_1 < \mathbf{x} \leq x_2, \mathbf{y} \leq y_1\} \cup \{x_1 < \mathbf{x} \leq x_2, y_1 < \mathbf{y} \leq y_2\}$$

y los eventos a la derecha de la igualdad son mutuamente excluyentes, se cumple que

$$P\{x_1 < \mathbf{x} \leq x_2, \mathbf{y} \leq y_2\} = P\{x_1 < \mathbf{x} \leq x_2, \mathbf{y} \leq y_1\} + P\{x_1 < \mathbf{x} \leq x_2, y_1 < \mathbf{y} \leq y_2\}.$$

Por lo tanto,

$$\begin{aligned} P\{x_1 < \mathbf{x} \leq x_2, y_1 < \mathbf{y} \leq y_2\} &= P\{x_1 < \mathbf{x} \leq x_2, \mathbf{y} \leq y_2\} - P\{x_1 < \mathbf{x} \leq x_2, \mathbf{y} \leq y_1\} \\ &\stackrel{(a)}{=} [F(x_2, y_2) - F(x_1, y_2)] - [F(x_2, y_1) - F(x_1, y_1)] \\ &= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1), \end{aligned}$$

donde en (a) se usó la primera ecuación en 5.2.

5.1.2. Densidad de probabilidad conjunta

La densidad de probabilidad conjunta de \mathbf{x} e \mathbf{y} es por definición la función

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}. \quad (5.4)$$

Se cumple por lo tanto que

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(\alpha, \beta) d\alpha d\beta.$$

De esta forma,

$$F(-\infty, y) = 0 \quad \text{y} \quad F(x, -\infty) = 0,$$

acorde con una de las propiedades de la distribución de probabilidad conjunta.

5.1.3. Estadísticas conjuntas

Se probará que la probabilidad de que el punto (\mathbf{x}, \mathbf{y}) esté en una región D del plano xy es la integral de $f(x, y)$ en D , es decir,

$$P\{(\mathbf{x}, \mathbf{y}) \in D\} = \iint_D f(x, y) dx dy, \quad (5.5)$$

donde $\{(\mathbf{x}, \mathbf{y}) \in D\}$ es el evento que consiste en todos los resultados ζ en los cuales el punto $[\mathbf{x}(\zeta), \mathbf{y}(\zeta)]$ está en D . Para mostrarlo, se observa que por la definición de derivada, se cumple que

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = \lim_{\Delta x \rightarrow 0} \lim_{\Delta y \rightarrow 0} \frac{F(x + \Delta x, y + \Delta y) - F(x, y + \Delta y) - F(x + \Delta x, y) + F(x, y)}{\Delta x \Delta y}. \quad (5.6)$$

Efectivamente,

$$\begin{aligned} \frac{\partial^2 F(x, y)}{\partial x \partial y} &= \frac{\partial}{\partial x} \left[\frac{\partial F(x, y)}{\partial y} \right] \\ &= \lim_{\Delta x \rightarrow 0} \frac{\frac{\partial F}{\partial y}(x + \Delta x, y) - \frac{\partial F}{\partial y}(x, y)}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \frac{\left[\lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x, y + \Delta y) - F(x + \Delta x, y)}{\Delta y} \right] - \left[\lim_{\Delta x \rightarrow 0} \frac{F(x, y + \Delta y) - F(x, y)}{\Delta y} \right]}{\Delta x}, \end{aligned}$$

obteniendo la ecuación 5.6. Empleando los resultados de las ecuaciones 5.3 y 5.4 en la ecuación 5.6, se llega a que

$$P\{x < \mathbf{x} \leq x + \Delta x, y < \mathbf{y} \leq y + \Delta y\} \approx f(x, y) \Delta x \Delta y. \quad (5.7)$$

Esto indica que la probabilidad de que el punto (\mathbf{x}, \mathbf{y}) esté en un rectángulo diferencial es el producto entre $f(x, y)$ y el área $\Delta x \Delta y$ del rectángulo, lo que prueba la ecuación 5.5, ya que la región D se puede expresar como el límite de la unión de dichos rectángulos.

5.1.4. Estadísticas marginales

En el análisis de varias variables aleatorias, la estadística de cada una por separado se llama *marginal*. De esta forma, $F_x(x)$ es la *distribución marginal* y $f_x(x)$ es la *densidad marginal* de \mathbf{x} . Se puede demostrar que la estadística marginal de \mathbf{x} e \mathbf{y} en función de su estadística conjunta $F(x, y)$ y $f(x, y)$ es

$$F_x(x) = F(x, \infty), \quad F_y(y) = F(\infty, y)$$

$$f_x(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad f_y(y) = \int_{-\infty}^{\infty} f(x, y) dx. \quad (5.8)$$

5.1.5. Independencia

En la sección 1.3.5 se definió independencia de eventos. Se dice que dos variables aleatorias son *independientes* si los eventos $\{\mathbf{x} \in A\}$ y $\{\mathbf{y} \in B\}$ son independientes, es decir,

$$P\{\mathbf{x} \in A, \mathbf{y} \in B\} = P\{\mathbf{x} \in A\}P\{\mathbf{y} \in B\},$$

donde A y B son dos conjuntos arbitrarios en los ejes x e y respectivamente. Aplicando esto a los eventos $\{\mathbf{x} \leq x\}$ y $\{\mathbf{y} \leq y\}$, se concluye que si \mathbf{x} e \mathbf{y} son independientes, se cumple que

$$F(x, y) = F_x(x)F_y(y), \quad (5.9)$$

y por lo tanto, también se cumple que

$$f(x, y) = f_x(x)f_y(y). \quad (5.10)$$

Puede demostrarse a partir de la ecuación 5.5 que lo recíproco también es cierto: si se cumple la ecuación 5.9 o 5.10, las variables aleatorias \mathbf{x} e \mathbf{y} son independientes.

Ejemplo: la aguja de Buffon. Un plano tiene dibujadas rectas paralelas separadas una distancia d , y se tira de forma aleatoria una aguja de largo l sobre el. Se quiere averiguar la probabilidad de que la aguja intersecte alguna recta. Este problema fue propuesto por el naturalista francés Georges-Louis Leclerc, conde de Buffon, en 1733 y publicado por él en 1777 con la solución¹.

Solución: sea z la distancia entre el centro de la aguja y la recta más cercana y sea θ el ángulo entre la aguja y la recta. El hecho de que la aguja sea tirada de forma aleatoria permite asumir que:

- z es una variable aleatoria uniformemente distribuida en $[0, d/2]$, por lo que su densidad de probabilidad es

$$f_z(z) = \begin{cases} 2/d, & \text{si } 0 \leq z \leq d/2 \\ 0, & \text{en otro caso.} \end{cases}$$

- θ es una variable aleatoria uniformemente distribuida en $[0, \pi/2]$, por lo que su densidad de probabilidad es

$$f_\theta(\theta) = \begin{cases} 2/\pi, & \text{si } 0 \leq \theta \leq \pi/2 \\ 0, & \text{en otro caso.} \end{cases}$$

De forma equivalente, se podría haber considerado que $0 \leq \theta \leq \pi$ con densidad de probabilidad $f_\theta = 1/\pi$, pero las probabilidades de θ en los intervalos $[0, \pi/2]$ y $[\pi/2, \pi]$ son simétricas, por lo que alcanza considerar solo el intervalo $[0, \pi/2]$ con el doble de probabilidad.

- z y θ son variables aleatorias independientes, y por lo tanto se cumple que

$$f_{z\theta}(z, \theta) = f_z(z)f_\theta(\theta) = \begin{cases} 4/\pi d, & \text{si } 0 \leq z \leq d/2 \text{ y } 0 \leq \theta \leq \pi/2 \\ 0, & \text{en otro caso.} \end{cases}$$

Según estas especificaciones, el espacio muestral del experimento es

$$\Omega = \left\{ (z, \theta) : 0 \leq z \leq \frac{d}{2}, 0 \leq \theta \leq \frac{\pi}{2} \right\}.$$

Para encontrar la probabilidad de intersección hay que separar los casos $l \leq d$ y $l > d$. Se comenzará resolviendo el primero.

Como se observa en el caso límite de la figura 5.2, si $l \leq d$, la aguja intersecta una recta si se cumple que

$$z \leq \frac{l}{2} \sin \theta.$$

Llámese D a esta región del plano (z, θ) , es decir,

$$D = \left\{ (z, \theta) : 0 \leq z \leq \frac{l}{2} \sin \theta, 0 \leq \theta \leq \frac{\pi}{2} \right\}.$$

¹Fuente: <http://mathworld.wolfram.com/BuffonsNeedleProblem.html>

La probabilidad de intersección es

$$\begin{aligned}
 P\{\text{intersección}\} &= \iint_D f_{z\theta}(z, \theta) dz d\theta \\
 &= \int_0^{\pi/2} \int_0^{(l/2)\sin\theta} \frac{4}{\pi d} dz d\theta \\
 &= \frac{4}{\pi d} \int_0^{\pi/2} \left(z \Big|_0^{(l/2)\sin\theta} \right) d\theta \\
 &= \frac{4}{\pi d} \int_0^{\pi/2} \frac{l}{2} \sin\theta d\theta \\
 &= \frac{2l}{\pi d} \left(-\cos\theta \Big|_0^{\pi/2} \right) \\
 &= \frac{2l}{\pi d} \left(-\cos\frac{\pi}{2} + 1 \right),
 \end{aligned}$$

concluyendo que

$$P\{\text{intersección}\} = \frac{2l}{\pi d}.$$

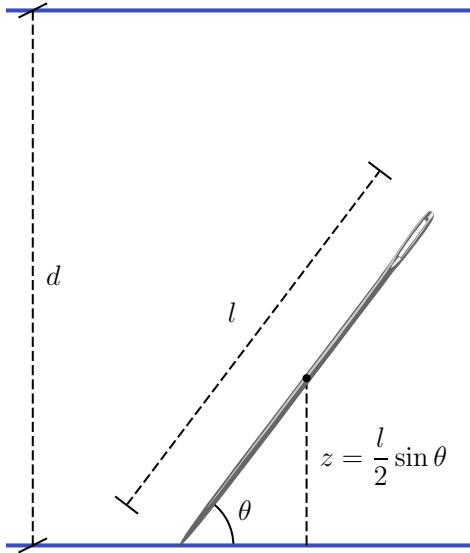


Figura 5.2: Experimento de la aguja de Buffon para el caso en que el largo l de la aguja es menor que la distancia d entre las líneas del plano. La condición para que la aguja intersecte una recta del plano, es que la distancia z entre el centro de la aguja y la recta mas cercana sea menor a $(l/2)\sin\theta$, donde θ es el ángulo entre la aguja y las rectas del plano. En la figura se muestra el caso especial en que $z = (l/2)\sin\theta$.

Una aplicación interesante de este experimento es en la estimación del valor de π . Supóngase que se arroja la aguja n veces, y de esas n veces h veces intersecta una recta del plano. Empleando la interpretación frecuentista del concepto de probabilidad, se cumple que

$$P\{\text{intersección}\} = \frac{2l}{\pi d} \approx \frac{h}{n},$$

por lo que el valor de π puede estimarse como

$$\pi \approx \frac{2ln}{dh}.$$

En 1901, el matemático italiano Mario Lazzarini realizó el experimento de la aguja de Buffon. Lazzarini arrojó una aguja $n = 3408$ veces y obtuvo la estimación de π de $355/113 = 3,14159292$, correcta hasta 6 dígitos después de la coma. Esta estimación es demasiado precisa para la baja cantidad n de veces que

repitió el experimento debido a que conocía el valor de π y realizó el experimento de forma sesgada para obtener un buen resultado². De todas formas, el experimento realizado por Lazzarini es considerado la primer aplicación de la técnica de integración de Montecarlo.

Se calculará a continuación la probabilidad de intersección para el caso en que la aguja es mas larga que la separación de los renglones en el plano, es decir, $l > d$. Para hacerlo, se comienza considerando el caso límite en que los extremos de la aguja tocan sin cortar dos rectas sucesivas, como se muestra en la figura 5.3. En ese caso, $z = d/2$ y

$$\sin \theta = \frac{d}{l} \quad \Rightarrow \quad \theta = \arcsin \frac{d}{l}.$$

Se deduce entonces que si $\theta \geq \arcsin d/l$, la aguja va a intersectar al menos una recta, independientemente del valor de z . La región del plano (z, θ) de esta situación es

$$D_1 = \left\{ (z, \theta) : 0 \leq z \leq \frac{d}{2}, \arcsin \frac{d}{l} \leq \theta \leq \frac{\pi}{2} \right\}.$$

En el caso en que $\theta < \arcsin d/l$, la aguja intersecta una recta si $z \leq (l/2) \sin \theta$, como se muestra también en la figura 5.3, y la región del plano (z, θ) de esta situación es

$$D_2 = \left\{ (z, \theta) : 0 \leq z \leq \frac{l}{2} \sin \theta, 0 \leq \theta < \arcsin \frac{d}{l} \right\}.$$

Teniendo en cuenta que las regiones D_1 y D_2 son disjuntas (ver figura 5.3), la probabilidad de intersección

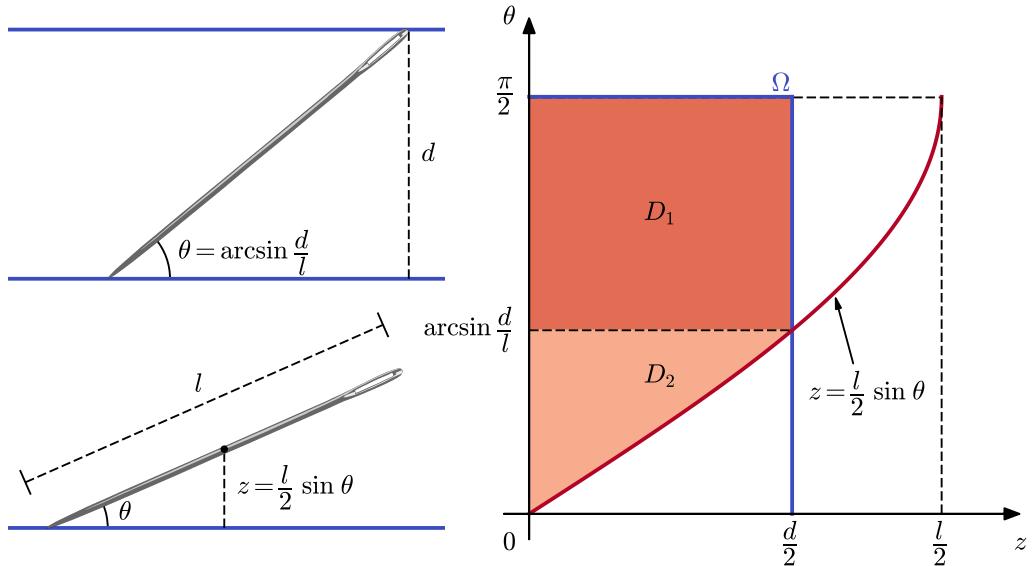


Figura 5.3: Experimento de la aguja de Buffon para el caso en que el largo l de la aguja es mayor que la distancia d entre las líneas del plano. Si el ángulo de la aguja cumple que $\theta \geq \arcsin d/l$, la aguja intersecta al menos una recta del plano. Esa situación consiste en los puntos (z, θ) en la región D_1 del espacio muestral Ω . En el caso en que $\theta < \arcsin d/l$, la aguja intersecta una recta si $z \leq (l/2) \sin \theta$. Esta situación consiste en los puntos de la región D_2 .

es

$$P\{\text{intersección}\} = \iint_{D_1 \cup D_2} f_{z\theta}(z, \theta) dz d\theta = \iint_{D_1} f_{z\theta}(z, \theta) dz d\theta + \iint_{D_2} f_{z\theta}(z, \theta) dz d\theta.$$

²Ver https://en.wikipedia.org/wiki/Buffon%27s_needle_problem#Estimating_%CF%80

La primera integral de la suma es

$$\begin{aligned}
\iint_{D_1} f_{z\theta}(z, \theta) dz d\theta &= \int_{\arcsin d/l}^{\pi/2} \int_0^{d/2} \frac{4}{\pi d} dz d\theta \\
&= \frac{4}{\pi d} \int_{\arcsin d/l}^{\pi/2} \left(z \Big|_0^{d/2} \right) d\theta \\
&= \frac{4}{\pi d} \int_{\arcsin d/l}^{\pi/2} \frac{d}{2} d\theta \\
&= \frac{2}{\pi} \left(\theta \Big|_{\arcsin d/l}^{\pi/2} \right) \\
&= \frac{2}{\pi} \left(\frac{\pi}{2} - \arcsin \frac{d}{l} \right),
\end{aligned}$$

y la segunda es

$$\begin{aligned}
\iint_{D_2} f_{z\theta}(z, \theta) dz d\theta &= \int_0^{\arcsin d/l} \int_0^{(l/2) \sin \theta} \frac{4}{\pi d} dz d\theta \\
&= \frac{4}{\pi d} \int_0^{\arcsin d/l} \left(z \Big|_0^{(l/2) \sin \theta} \right) d\theta \\
&= \frac{4}{\pi d} \int_0^{\arcsin d/l} \frac{l}{2} \sin \theta d\theta \\
&= \frac{2l}{\pi d} \left(-\cos \theta \Big|_0^{\arcsin d/l} \right) \\
&= \frac{2l}{\pi d} \left[1 - \cos \left(\arcsin \frac{d}{l} \right) \right] \\
&\stackrel{(a)}{=} \frac{2l}{\pi d} \left[1 - \sqrt{1 - \sin^2 \left(\arcsin \frac{d}{l} \right)} \right] \\
&= \frac{2l}{\pi d} \left(1 - \sqrt{1 - \frac{d^2}{l^2}} \right) \\
&= \frac{2}{\pi d} \left(l - \sqrt{l^2 - d^2} \right),
\end{aligned}$$

donde en (a) se empleó la identidad $\sin^2 \theta + \cos^2 \theta = 1$. Combinando ambos resultados se obtiene que

$$\begin{aligned}
P\{\text{intersección}\} &= \frac{2}{\pi} \left(\frac{\pi}{2} - \arcsin \frac{d}{l} \right) + \frac{2}{\pi d} \left(l - \sqrt{l^2 - d^2} \right) \\
&= \frac{2}{\pi d} \left(\frac{\pi d}{2} - d \arcsin \frac{d}{l} + l - \sqrt{l^2 - d^2} \right).
\end{aligned}$$

5.1.6. Variables aleatorias conjuntamente normales

Se dice que dos variables aleatorias \mathbf{x} e \mathbf{y} son conjuntamente normales si su densidad de probabilidad conjunta es

$$f(x, y) = A \exp \left\{ -\frac{1}{2(1-r^2)} \left[\frac{(x-\eta_1)^2}{\sigma_1^2} - 2r \frac{(x-\eta_1)(y-\eta_2)}{\sigma_1 \sigma_2} + \frac{(y-\eta_2)^2}{\sigma_2^2} \right] \right\}, \quad (5.11)$$

con

$$A = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}, \quad |r| < 1. \quad (5.12)$$

La función $f(x, y)$ es una exponencial de exponente negativo si $|r| < 1$, es positiva para todo (x, y) , e integra a 1, como se mostrará a continuación. Se denota como

$$N(\eta_1, \eta_2, \sigma_1^2, \sigma_2^2, r).$$

Como se verá, η_1 y η_2 son los valores esperados de \mathbf{x} e \mathbf{y} respectivamente, y σ_1^2 y σ_2^2 sus varianzas. El parámetro r se llama coeficiente de correlación, y su significado será explicado mas adelante.

Se demostrará que las densidades marginales de \mathbf{x} e \mathbf{y} son

$$f_x(x) = \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-(x-\eta_1)^2/2\sigma_1^2} \quad f_y(y) = \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-(y-\eta_2)^2/2\sigma_2^2},$$

es decir, \mathbf{x} e \mathbf{y} son variables aleatorias normales con $\mathbf{x} \sim N(\eta_1, \sigma_1^2)$ y $\mathbf{y} \sim N(\eta_2, \sigma_2^2)$. Para esto, hay que probar que si se integra la ecuación 5.11 respecto a x se obtiene $f_y(y)$, como indica la ecuación 5.8, y si se integra respecto a y se obtiene $f_x(x)$. Se parte observando que el término entre paréntesis rectos del exponente en 5.11 se puede escribir como

$$\frac{(x-\eta_1)^2}{\sigma_1^2} - 2r \frac{(x-\eta_1)(y-\eta_2)}{\sigma_1 \sigma_2} + \frac{(y-\eta_2)^2}{\sigma_2^2} = \left(\frac{x-\eta_1}{\sigma_1} - r \frac{y-\eta_2}{\sigma_2} \right)^2 + (1-r^2) \frac{(y-\eta_2)^2}{\sigma_2^2}$$

es decir, como un cuadrado perfecto y un segundo sumando para mantener la igualdad. Observando el lado derecho de la igualdad, se ve que si $|r| < 1$, la expresión es positiva, y por lo tanto, el exponente de la ecuación 5.11 es negativo, como se mencionó previamente. Sacando denominador común σ_1 en el primer sumando a la derecha de la igualdad, se llega a que el término entre paréntesis rectos del exponente en 5.11 queda

$$\left\{ \frac{x - \left[\eta_1 + r \frac{\sigma_1}{\sigma_2} (y - \eta_2) \right]}{\sigma_1} \right\}^2 + (1-r^2) \frac{(y-\eta_2)^2}{\sigma_2^2} \quad (5.13)$$

y definiendo

$$\eta \triangleq \eta_1 + r \frac{\sigma_1}{\sigma_2} (y - \eta_2),$$

se tiene que la densidad marginal de \mathbf{y} es

$$\begin{aligned} f_y(y) &= \int_{-\infty}^{\infty} f(x, y) dx \\ &= \int_{-\infty}^{\infty} A \exp \left[-\frac{(x-\eta)^2}{2(1-r^2)\sigma_1^2} - \frac{(y-\eta_2)^2}{2\sigma_2^2} \right] dx \\ &= A e^{-(y-\eta_2)^2/2\sigma_2^2} \int_{-\infty}^{\infty} e^{-(x-\eta)^2/2(1-r^2)\sigma_1^2} dx. \end{aligned}$$

La integral en la última igualdad representa una variable aleatoria normal de media η y varianza $2(1-r^2)\sigma_1^2$ y por lo tanto, integra a $\sqrt{2\pi(1-r^2)\sigma_1^2}$. Se llega al mismo resultado considerando que se trata de una integral gaussiana y aplicando la ecuación 3.9. Teniendo en cuenta este resultado y sustituyendo el valor de A dado por la ecuación 5.12, se tiene que

$$f_y(y) = \frac{\sigma_1 \sqrt{2\pi(1-r^2)}}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} e^{-(y-\eta_2)^2/2\sigma_2^2},$$

concluyendo que

$$f_y(y) = \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-(y-\eta_2)^2/2\sigma_2^2},$$

que era lo que se quería demostrar. La densidad marginal de \mathbf{x} se obtiene de forma análoga. Como $f_y(y)$ es una densidad de probabilidad normal, integra a 1. Por lo tanto, la integral doble de $f(x, y)$ es 1 con el valor de A indicado en la ecuación 5.12, como se mencionó antes.

Vale la pena tener en cuenta las siguientes consideraciones:

- Si dos variables aleatorias son conjuntamente normales, sus marginales son normales, como se acaba de mostrar. Sin embargo, lo recíproco no es cierto: dos variables aleatorias que tienen marginales normales, no necesariamente son conjuntamente normales.
- La definición de variables aleatorias conjuntamente normales puede plantearse también de la siguiente forma: dos variables aleatorias \mathbf{x} e \mathbf{y} son conjuntamente normales si la suma $a\mathbf{x} + b\mathbf{y}$ es normal para todo a y b .

5.2. Función de dos variables aleatorias

Dadas dos variables aleatorias \mathbf{x} e \mathbf{y} y una función $g(x, y)$, se construye una nueva variable aleatoria \mathbf{z} como

$$\mathbf{z} = g(\mathbf{x}, \mathbf{y}).$$

A continuación se verá como calcular la densidad de probabilidad $f_z(z)$ de \mathbf{z} a partir de la densidad de probabilidad conjunta $f_{xy}(x, y)$ de \mathbf{x} e \mathbf{y} . De forma general, se tiene que

$$\begin{aligned} F_z(z) &= P\{\mathbf{z} \leq z\} = P\{g(\mathbf{x}, \mathbf{y}) \leq z\} = P\{(\mathbf{x}, \mathbf{y}) \in D_z\} \\ &= \iint_{(x, y) \in D_z} f_{xy}(x, y) dx dy, \end{aligned} \quad (5.14)$$

donde D_z es la región del plano xy donde se cumple que $g(x, y) \leq z$.

5.2.1. Ejemplo: Densidad de probabilidad de $\mathbf{z} = \mathbf{x} + \mathbf{y}$

Sea $\mathbf{z} = \mathbf{x} + \mathbf{y}$. Se quiere determinar $f_z(z)$. Se parte determinando la región D_z del plano xy en la ecuación 5.14, que en este caso es $x + y \leq z$ y corresponde a la región del plano a la izquierda de la recta $x + y = z$, como se muestra en la figura 5.4. Integrando sobre una banda a lo largo del eje x desde $-\infty$ a $x = z - y$ y luego desplazando la banda a lo largo del eje y desde $-\infty$ a ∞ , se cubre toda la región. Por lo tanto,

$$F_z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_{xy}(x, y) dx dy.$$

Para calcular la densidad de probabilidad $f_z(z)$ hay que diferenciar $F_z(z)$ respecto a z . Para esto, hay que aplicar la regla de integración de Leibnitz, que indica que si

$$F_z(z) = \int_{a(z)}^{b(z)} f(x, z) dx, \quad (5.15)$$

la derivada es

$$\frac{dF_z(z)}{dz} = \frac{db(z)}{dz} f(b(z), z) - \frac{da(z)}{dz} f(a(z), z) + \int_{a(z)}^{b(z)} \frac{\partial f(x, z)}{\partial z} dx. \quad (5.16)$$

Por lo tanto

$$\begin{aligned} f_z(z) &= \frac{dF_z(z)}{dz} = \frac{d}{dz} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_{xy}(x, y) dx dy \right] \\ &= \int_{-\infty}^{\infty} \frac{\partial}{\partial z} \left[\int_{-\infty}^{z-y} f_{xy}(x, y) dx \right] dy \\ &\stackrel{(a)}{=} \int_{-\infty}^{\infty} \left[\frac{\partial(z-y)}{\partial z} f_{xy}(z-y, y) - 0 + \int_{-\infty}^{z-y} \frac{\partial f_{xy}(x, y)}{\partial z} dx \right] dy \\ &\stackrel{(b)}{=} \int_{-\infty}^{\infty} f_{xy}(z-y, y) dy. \end{aligned}$$

donde en (a) se aplicó la regla de integración de Leibnitz, donde el sumando 0 se debe a que el límite inferior de la integral es independiente de z y en (b) se tuvo en cuenta que $\partial(z-y)/\partial z = 1$ y $\partial f_{xy}(x, y)/\partial z = 0$ ya que $f_{xy}(x, y)$ es independiente de z .

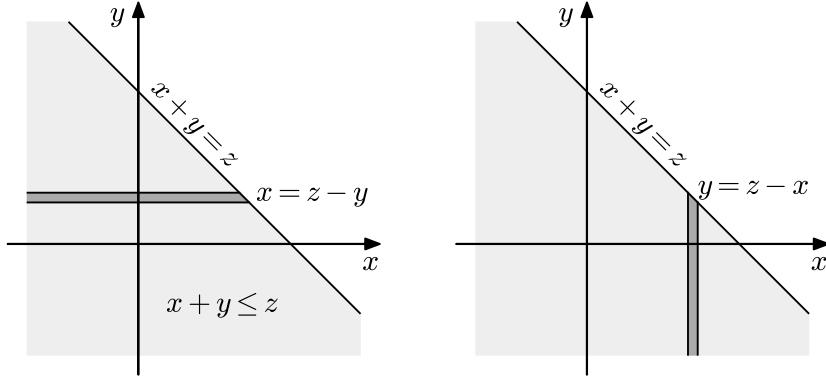


Figura 5.4: Región de integración para determinar la densidad de probabilidad de $\mathbf{z} = \mathbf{x} + \mathbf{y}$. La región de integración puede cubrirse de dos formas: *i*) integrando sobre una banda a lo largo del eje x desde $-\infty$ a $x = z - y$ y luego desplazando la banda a lo largo del eje y desde $-\infty$ a ∞ , como se muestra en la gráfica de la izquierda, o *ii*) integrando sobre una banda a lo largo del eje y desde $-\infty$ a $y = z - x$ y luego desplazando la banda a lo largo del eje x desde $-\infty$ a ∞ , como se muestra en la gráfica de la derecha.

Como se ilustra en la figura 5.4, si en lugar de integrar en una banda horizontal que se desplaza a lo largo del eje y , se integra en una banda vertical que se desplaza a lo largo del eje x se obtiene un resultado equivalente y realizando un razonamiento análogo se llega a que

$$f_z(z) = \int_{-\infty}^{\infty} f_{xy}(z - y, y) dy = \int_{-\infty}^{\infty} f_{xy}(x, z - x) dx. \quad (5.17)$$

Si las variables aleatorias \mathbf{x} e \mathbf{y} son independientes, se cumple que

$$f_{xy}(x, y) = f_x(x) f_y(y),$$

y la ecuación 5.17 queda

$$f_z(z) = \int_{-\infty}^{\infty} f_x(z - y) f_y(y) dy = \int_{-\infty}^{\infty} f_x(x) f_y(z - x) dx. \quad (5.18)$$

Esta integral es la convolución entre $f_x(z)$ y $f_y(z)$. Se concluye que si dos variables aleatorias son independientes, la densidad de probabilidad de la suma es la convolución de las densidades de probabilidad.

Caso especial Se considera el caso especial en que $f_x(x) = 0$ si $x < 0$ y $f_y(y) = 0$ si $y < 0$. La región de integración se muestra en la figura 5.5, y se puede cubrir considerando una banda horizontal desde 0 a $x = z - y$ que se desplaza a lo largo del eje y desde 0 a $y = z$. Por lo tanto, la distribución de probabilidad es

$$F_z(z) = \int_0^z \int_0^{z-y} f_{xy}(x, y) dx dy.$$

y la densidad de probabilidad se obtiene como

$$\begin{aligned} f_z(z) &= \frac{dF_z(z)}{dz} = \frac{d}{dz} \left[\int_0^z \int_0^{z-y} f_{xy}(x, y) dx dy \right] \\ &= \frac{d}{dz} \left[\int_0^z g(z, y) dy \right], \end{aligned}$$

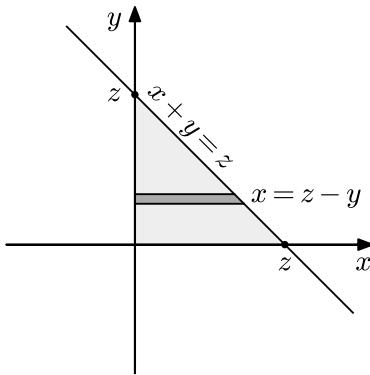


Figura 5.5: Región en el plano xy correspondiente al evento $\{\mathbf{x} + \mathbf{y} \leq z\}$ cuando $f_x(x) = 0$ si $x < 0$ y $f_y(y) = 0$ si $y < 0$.

donde se definió

$$g(z, y) \triangleq \int_0^{z-y} f_{xy}(x, y) dx.$$

Aplicando la regla de integración de Leibnitz se tiene que

$$\frac{d}{dz} \left[\int_0^z g(z, y) dy \right] = g(z, z) + \int_0^z \frac{\partial g(z, x)}{\partial z} dx.$$

Como

$$g(z, z) = \int_0^{z-z} f_{xy}(x, y) dx = \int_0^0 f_{xy}(x, y) dx = 0,$$

se obtiene que

$$\begin{aligned} f_z(z) &= \frac{d}{dz} \left[\int_0^z \int_0^{z-y} f_{xy}(x, y) dx dy \right] \\ &= \int_0^z \left[\frac{\partial}{\partial z} \int_0^{z-y} f_{xy}(x, y) dx \right] dy \\ &\stackrel{(a)}{=} \begin{cases} \int_0^z f_{xy}(z-y, y) dy, & z > 0 \\ 0, & z \leq 0 \end{cases} \end{aligned}$$

donde en (a) se diferenció aplicando la regla de Leibnitz y se tomó en cuenta que $\mathbf{z} = \mathbf{x} + \mathbf{y} \geq 0$ ya que $\mathbf{x} \geq 0$ y $\mathbf{y} \geq 0$.

5.2.2. Ejemplo: Densidad de probabilidad de $\mathbf{z} = \mathbf{x}/\mathbf{y}$

Sea $\mathbf{z} = \mathbf{x}/\mathbf{y}$. Se quiere determinar $f_z(z)$ en términos de $f_{xy}(x, y)$. Se comenzará calculando

$$F_z(z) = P\{\mathbf{x}/\mathbf{y} \leq z\}.$$

La inecuación $x/y \leq z$ puede escribirse como $\mathbf{x} \leq \mathbf{y}z$ si $\mathbf{y} > 0$ y como $\mathbf{x} \geq \mathbf{y}z$ si $\mathbf{y} < 0$. Por lo tanto, el evento $\{\mathbf{x}/\mathbf{y}\}$ tiene que ser condicionado al evento $A = \{\mathbf{y} > 0\}$ y su complemento \bar{A} . Como $A \cup \bar{A} = S$ y A y \bar{A} son mutuamente excluyentes, se cumple que

$$\begin{aligned} P\{\mathbf{x}/\mathbf{y} \leq z\} &= P\{\mathbf{x}/\mathbf{y} \leq z \cap (A \cup \bar{A})\} \\ &= P\{(\mathbf{x}/\mathbf{y} \leq z \cap A) \cup (\mathbf{x}/\mathbf{y} \leq z \cap \bar{A})\} \\ &= P\{\mathbf{x}/\mathbf{y} \leq z, \mathbf{y} > 0\} + P\{\mathbf{x}/\mathbf{y} \leq z, \mathbf{y} < 0\} \\ &= P\{\mathbf{x} \leq \mathbf{y}z, \mathbf{y} > 0\} + P\{\mathbf{x} \leq \mathbf{y}z, \mathbf{y} < 0\} \end{aligned}$$

En la figura 5.6 se muestra el área de integración correspondiente a cada sumando. Integrando sobre

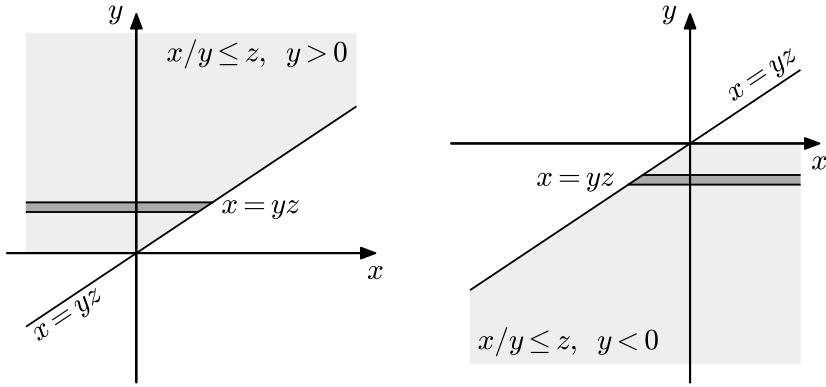


Figura 5.6: Región de integración para determinar la densidad de probabilidad de $\mathbf{z} = \mathbf{x}/\mathbf{y}$. A la derecha se muestra la región de integración de $f_{xy}(x, y)$ para obtener $P\{\mathbf{x} \leq \mathbf{yz}, \mathbf{y} > 0\}$. La región se puede cubrir considerando una banda a lo largo del eje x desde $x = -\infty$ a $x = yz$ y desplazando la banda desde $y = 0$ hasta $y = \infty$. A la izquierda se muestra la región de integración para obtener $P\{\mathbf{x} \leq \mathbf{yz}, \mathbf{y} < 0\}$. La región se puede cubrir considerando una banda a lo largo del eje x desde $x = yz$ a $x = \infty$ y desplazando la banda desde $y = -\infty$ hasta $y = 0$.

esas regiones, se obtiene que

$$F_z(z) = \int_0^\infty \int_{-\infty}^{yz} f_{xy}(x, y) dx dy + \int_{-\infty}^0 \int_{yz}^\infty f_{xy}(x, y) dx dy.$$

y diferenciando respecto a z empleando la regla de integración de Leibnitz, se obtiene la densidad de probabilidad,

$$f_z(z) = \int_0^\infty y f_{xy}(yz, y) dy + \int_{-\infty}^0 -y f_{xy}(yz, y) dy,$$

es decir,

$$f_z(z) = \int_{-\infty}^\infty |y| f_{xy}(yz, y) dy. \quad (5.19)$$

En el caso en que \mathbf{x} e \mathbf{y} son variables aleatorias no negativas, la distribución de probabilidad queda

$$F_z(z) = \int_0^\infty \int_0^{yz} f_{xy}(x, y) dx dy$$

y la densidad de probabilidad es

$$f_z(z) = \int_0^\infty y f_{xy}(yz, y) dy.$$

Caso particular (Ejemplo 6-11) Se considera el caso en que las variables aleatorias \mathbf{x} e \mathbf{y} son conjuntamente normales de media nula y densidad de probabilidad

$$f_{xy}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left[-\frac{1}{2(1-r^2)}\left(\frac{x^2}{\sigma_1^2} - \frac{2rxy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right)\right],$$

y se quiere calcular la distribución y densidad de probabilidad de $\mathbf{z} = \mathbf{x}/\mathbf{y}$.

Como en este caso la densidad de probabilidad cumple que $f_{xy}(-x, -y) = f_{xy}(x, y)$, la ecuación 5.19 queda

$$f_z(z) = \int_{-\infty}^\infty |y| f_{xy}(yz, y) dy = f_z(z) = 2 \int_0^\infty y f_{xy}(yz, y) dy. \quad (5.20)$$

El integrando contiene el término $f_{xy}(yz, y)$, es decir, la densidad de probabilidad evaluada en $x = yz$, y es

$$\begin{aligned} f_{xy}(yz, y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left[-\frac{1}{2(1-r^2)}\left(\frac{y^2z^2}{\sigma_1^2} - \frac{2ry^2z}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right)\right] \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left[-\frac{y^2}{2(1-r^2)}\left(\frac{z^2}{\sigma_1^2} - \frac{2rz}{\sigma_1\sigma_2} + \frac{1}{\sigma_2^2}\right)\right] \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left(-\frac{y^2}{2\sigma_0^2}\right), \end{aligned} \quad (5.21)$$

donde se definió

$$\sigma_0^2 \triangleq \frac{1-r^2}{\frac{z^2}{\sigma_1^2} - \frac{2rz}{\sigma_1\sigma_2} + \frac{1}{\sigma_2^2}}.$$

Notando que

$$\left(\frac{z}{\sigma_1} - \frac{r}{\sigma_2}\right)^2 = \frac{z^2}{\sigma_1^2} - \frac{2rz}{\sigma_1\sigma_2} + \frac{r^2}{\sigma_2^2},$$

el denominador de σ_0^2 se puede expresar como

$$\frac{z^2}{\sigma_1^2} - \frac{2rz}{\sigma_1\sigma_2} + \frac{1}{\sigma_2^2} = \left(\frac{z}{\sigma_1} - \frac{r}{\sigma_2}\right)^2 + \frac{1}{\sigma_2^2} - \frac{r^2}{\sigma_2^2},$$

por lo que

$$\sigma_0^2 = \frac{1-r^2}{\left(\frac{z}{\sigma_1} - \frac{r}{\sigma_2}\right)^2 + \frac{1-r^2}{\sigma_2^2}} = \frac{\sigma_1^2\sigma_2^2(1-r^2)}{\sigma_2^2\left(z - \frac{\sigma_1}{\sigma_2}r\right)^2 + \sigma_1^2(1-r^2)}, \quad (5.22)$$

donde en la segunda igualdad se multiplicó el numerador y el denominador entre $\sigma_1^2\sigma_2^2$. A partir de las ecuaciones 5.20 y 5.21, la densidad de probabilidad queda,

$$\begin{aligned} f_z(z) &= 2 \int_0^\infty y f_{xy}(yz, y) dy \\ &= \frac{1}{\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int_0^\infty y e^{-y^2/2\sigma_0^2} dy \\ &= \frac{1}{\pi\sigma_1\sigma_2\sqrt{1-r^2}} \left[-\sigma_0^2 e^{-y^2/2\sigma_0^2} \right]_0^\infty \\ &= \frac{1}{\pi\sigma_1\sigma_2\sqrt{1-r^2}} [-\sigma_0^2 (0-1)] \\ &= \frac{\sigma_0^2}{\pi\sigma_1\sigma_2\sqrt{1-r^2}}, \end{aligned}$$

y sustituyendo σ_0^2 dado por la ecuación 5.22 se llega a que

$$f_z(z) = \frac{\sigma_1\sigma_2\sqrt{1-r^2}}{\pi \left[\sigma_2^2 \left(z - \frac{\sigma_1}{\sigma_2}r \right)^2 + \sigma_1^2(1-r^2) \right]}.$$

Operando, se ve que $f_z(z)$ se puede escribir como

$$\begin{aligned} f_z(z) &\stackrel{(a)}{=} \frac{\sigma_1\sigma_2\sqrt{1-r^2}}{\pi\sigma_1^2(1-r^2) \left[\frac{\sigma_2^2(z-r\sigma_1/\sigma_2)^2}{\sigma_1^2(1-r^2)} + 1 \right]} \\ &= \frac{1}{\pi \frac{\sigma_1}{\sigma_2} \sqrt{1-r^2} \left[\left(\frac{z-r\sigma_1/\sigma_2}{\frac{\sigma_1}{\sigma_2}\sqrt{1-r^2}} \right)^2 + 1 \right]}, \end{aligned}$$

donde en (a) se multiplicó y dividió el denominador entre $\sigma_1^2(1 - r^2)$. Definiendo

$$\gamma \triangleq \frac{\sigma_1}{\sigma_2} \sqrt{1 - r^2}, \quad z_0 \triangleq r\sigma_1/\sigma_2,$$

la densidad de probabilidad se puede expresar como

$$f_z(z) = \frac{1}{\pi\gamma \left[\left(\frac{z - z_0}{\gamma} \right)^2 + 1 \right]}, \quad (5.23)$$

por lo que se concluye que $\mathbf{z} = \mathbf{x}/\mathbf{y}$ es una variable aleatoria de Cauchy con centro $z_0 = r\sigma_1/\sigma_2$.

Integrando la ecuación 5.23 entre $-\infty$ a z se obtiene la distribución de probabilidad. Para hacerlo, se toma en cuenta que

$$\frac{d}{dx}(\arctan x) = \frac{1}{x^2 + 1}. \quad (5.24)$$

Esta igualdad puede deducirse mediante diferenciación implícita de la siguiente forma. Se quiere calcular la derivada de la función

$$y(x) = \arctan x, \quad -\frac{\pi}{2} < y < \frac{\pi}{2}.$$

Se comienza despejando x

$$\tan y = x \quad (5.25)$$

y luego se diferencia respecto a x ,

$$\frac{d}{dx}(\tan y) = \frac{d}{dx}x = 1$$

Desarrollando el lado izquierdo de la igualdad, se ve que

$$\frac{d}{dx}(\tan y) = \frac{d}{dx} \left(\frac{\sin y}{\cos y} \right) = \frac{\frac{dy}{dx} \cos^2 y + \frac{dy}{dx} \sin^2 y}{\cos^2 y} = \frac{dy}{dx} (1 + \tan^2 y)$$

y por lo tanto,

$$\frac{dy}{dx} (1 + \tan^2 y) = 1.$$

Despejando dy/dx se llega a que

$$\frac{dy}{dx} = \frac{1}{1 + \tan^2 y} = \frac{1}{1 + x^2},$$

donde en la última igualdad se usó la ecuación 5.25. Se concluye que se verifica la ecuación 5.24, a partir de la cual es fácil ver que

$$\frac{d}{dx} \left(\arctan \frac{x}{a} \right) = \frac{1}{a \left[\left(\frac{x}{a} \right)^2 + 1 \right]}.$$

De aquí se deduce que la primitiva de la densidad de probabilidad de la ecuación 5.23 es

$$\frac{d}{dz} \left(\frac{1}{\pi} \arctan \frac{z - z_0}{\gamma} \right) = -\frac{1}{\pi\gamma \left[\left(\frac{z - z_0}{\gamma} \right)^2 + 1 \right]} = f_z(z)$$

Por lo tanto, la distribución de probabilidad es

$$\begin{aligned} F_z(z) &= \int_{-\infty}^z f_z(u) du \\ &= \frac{1}{\pi} \left(\arctan \frac{u - z_0}{\gamma} \Big|_{-\infty}^z \right) \\ &\stackrel{(a)}{=} \frac{1}{\pi} \left(\arctan \frac{z - z_0}{\gamma} + \frac{\pi}{2} \right) \\ &= \frac{1}{\pi} \arctan \frac{z - z_0}{\gamma} + \frac{1}{2}, \end{aligned}$$

donde en (a) se usó que $\arctan u \rightarrow -\pi/2$ con $u \rightarrow -\infty$. Sustituyendo z_0 y γ se obtiene que

$$\begin{aligned} F_z(z) &= \frac{1}{\pi} \arctan \frac{z - r\sigma_1/\sigma_2}{\frac{\sigma_1}{\sigma_2} \sqrt{1 - r^2}} + \frac{1}{2} \\ &= \frac{1}{\pi} \arctan \frac{z\sigma_2 - r\sigma_1}{\sigma_1 \sqrt{1 - r^2}} + \frac{1}{2}. \end{aligned}$$

5.2.3. Estadísticos de orden

Considérese una n -tupla de valores $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$. El subíndice k en \mathbf{x}_k puede indicar un orden temporal, pero usualmente no es significativo. Un ejemplo en donde el subíndice es significativo es en el caso en que la n -tupla forma parte de una serie temporal. La n -tupla puede ordenarse de forma creciente como

$$\mathbf{x}_{(1)} \leq \mathbf{x}_{(2)} \leq \dots \leq \mathbf{x}_{(n)},$$

donde $\mathbf{x}_{(1)} = \min(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, $\mathbf{x}_{(2)}$ es el segundo valor más pequeño y así sucesivamente hasta $\mathbf{x}_{(n)} = \max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$. En el caso en que $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ son variables aleatorias, la función $\mathbf{x}_{(k)}$ que toma el valor $x_{(k)}$ en cada posible secuencia (x_1, x_2, \dots, x_n) se llama estadístico de orden k -ésimo. El conjunto $\{\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \dots, \mathbf{x}_{(n)}\}$ representa los estadísticos de orden entre n variables aleatorias. En este contexto,

$$\mathbf{R} = \mathbf{x}_{(n)} - \mathbf{x}_{(1)}$$

representa el estadístico *rango*. En el caso en que $n = 2$, se tienen los estadísticos *máximo* y *mínimo*.

Ejemplo: Densidad de probabilidad de $\mathbf{z} = \max(\mathbf{x}, \mathbf{y})$ Sea $\mathbf{z} = \max(\mathbf{x}, \mathbf{y})$. Se quiere determinar $f_z(z)$. Como

$$\mathbf{z} = \max(\mathbf{x}, \mathbf{y}) = \begin{cases} \mathbf{x}, & \mathbf{x} > \mathbf{y} \\ \mathbf{y}, & \mathbf{x} \leq \mathbf{y} \end{cases},$$

la distribución de probabilidad es

$$\begin{aligned} F_z(z) &= P\{\max(\mathbf{x}, \mathbf{y}) \leq z\} \\ &= P\{(\max(\mathbf{x}, \mathbf{y}) \leq z) \cap (\mathbf{x} > \mathbf{y} \cup \mathbf{x} \leq \mathbf{y})\} \\ &= P\{(\max(\mathbf{x}, \mathbf{y}) \leq z \cap \mathbf{x} > \mathbf{y}) \cup (\max(\mathbf{x}, \mathbf{y}) \leq z \cap \mathbf{x} \leq \mathbf{y})\} \\ &= P\{\mathbf{x} \leq z, \mathbf{x} > \mathbf{y}\} + P\{\mathbf{y} \leq z, \mathbf{x} \leq \mathbf{y}\}, \end{aligned} \quad (5.26)$$

ya que $\{\mathbf{x} > \mathbf{y}\}$ y $\{\mathbf{x} \leq \mathbf{y}\}$ son eventos mutuamente excluyentes y forman una partición del espacio muestral. En la figura 5.7 se muestran las regiones correspondientes a cada sumando de la ecuación 5.26, así como la unión de las regiones. De la figura, se obtiene que

$$F_z(z) = P\{\mathbf{x} \leq z, \mathbf{y} \leq z\} = F_{xy}(z, z). \quad (5.27)$$

En el caso en que \mathbf{x} e \mathbf{y} son variables aleatorias independientes, la distribución queda

$$F_z(z) = F_x(z)F_y(z),$$

y diferenciando respecto a z se obtiene la densidad de probabilidad, que es

$$f_z(z) = F_x(z)f_y(z) + f_x(z)F_y(z).$$

Ejemplo: Densidad de probabilidad de $\mathbf{w} = \min(\mathbf{x}, \mathbf{y})$ Considérese ahora el caso en que $\mathbf{z} = \min(\mathbf{x}, \mathbf{y})$. Se quiere determinar $f_z(z)$. Como

$$\mathbf{w} = \min(\mathbf{x}, \mathbf{y}) = \begin{cases} \mathbf{y}, & \mathbf{x} > \mathbf{y} \\ \mathbf{x}, & \mathbf{x} \leq \mathbf{y} \end{cases},$$

realizando un razonamiento análogo al caso anterior, se tiene que

$$\begin{aligned} F_w(w) &= P\{\min(\mathbf{x}, \mathbf{y}) \leq w\} \\ &= P\{\mathbf{y} \leq w, \mathbf{x} > \mathbf{y}\} + P\{\mathbf{x} \leq w, \mathbf{x} \leq \mathbf{y}\}. \end{aligned} \quad (5.28)$$

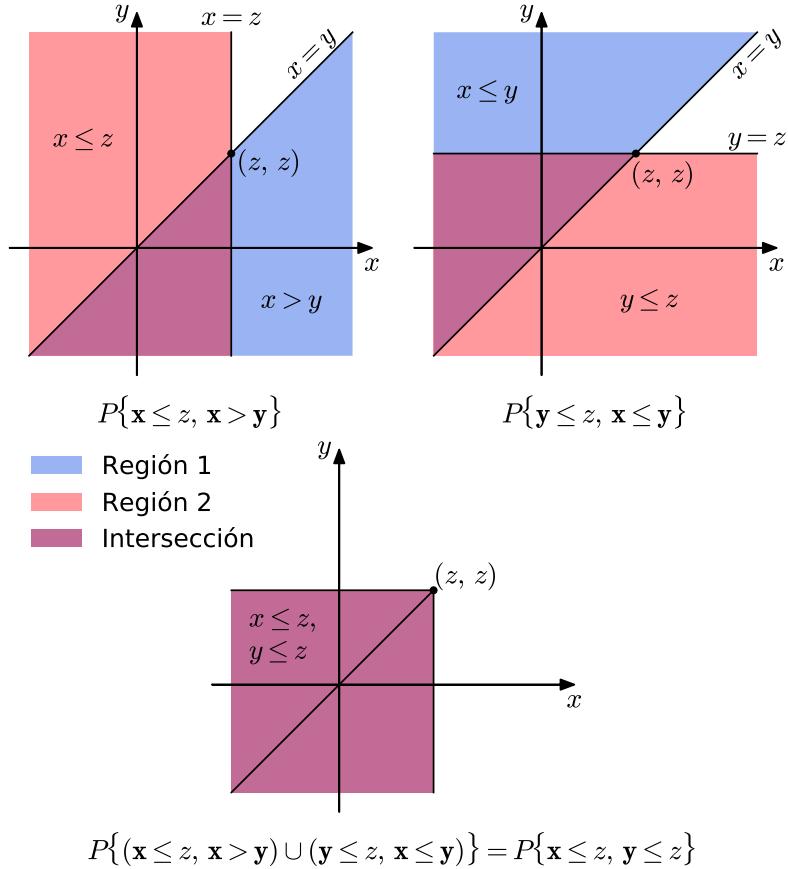


Figura 5.7: Región del plano correspondiente a $F_z(z) = P\{\max(\mathbf{x}, \mathbf{y}) \leq z\}$.

En la figura 5.8 se muestran las regiones del plano correspondientes a las desigualdades de la ecuación 5.28, y también la unión de dichas regiones. A partir de la figura, se tiene que

$$F_w(w) = 1 - P\{\mathbf{x} > w, \mathbf{y} > w\},$$

ya que la probabilidad conjunta del todo el plano es $F(\infty, \infty) = P\{x \leq \infty, y \leq \infty\} = 1$. Además, se ve que

$$\begin{aligned} P\{\mathbf{x} > w, \mathbf{y} > w\} &= P\{w < \mathbf{x} \leq \infty, w < \mathbf{y} \leq \infty\} \\ &\stackrel{(a)}{=} F(\infty, \infty) - F(w, \infty) - F(\infty, w) + F(w, w) \\ &\stackrel{(b)}{=} 1 - F_x(w) - F_y(w) + F_{xy}(w, w), \end{aligned}$$

donde en (a) se usó la ecuación 5.3 y en (b) las estadísticas marginales. Sustituyendo este resultado en la ecuación anterior, se obtiene que

$$F_w(w) = 1 - [1 - F_x(w) - F_y(w) + F_{xy}(w, w)],$$

concluyendo que

$$F_w(w) = F_x(w) + F_y(w) - F_{xy}(w, w). \quad (5.29)$$

Ejemplo: caso discreto Sean \mathbf{x} y \mathbf{y} variables aleatorias de Poisson con parámetros λ_1 y λ_2 respectivamente. Si $\mathbf{z} = \mathbf{x} + \mathbf{y}$, se quiere determinar la p.m.f de \mathbf{z} .

Como \mathbf{x} y \mathbf{y} toman los valores $\{0, 1, 2, \dots\}$, lo mismo ocurre con \mathbf{z} . Para cada $n = 0, 1, 2, \dots$, $\{\mathbf{x} + \mathbf{y} = n\}$ implica un número finito de valores posibles para \mathbf{x} y \mathbf{y} . De hecho, si $\mathbf{x} = 0$, \mathbf{y} debe ser n ,

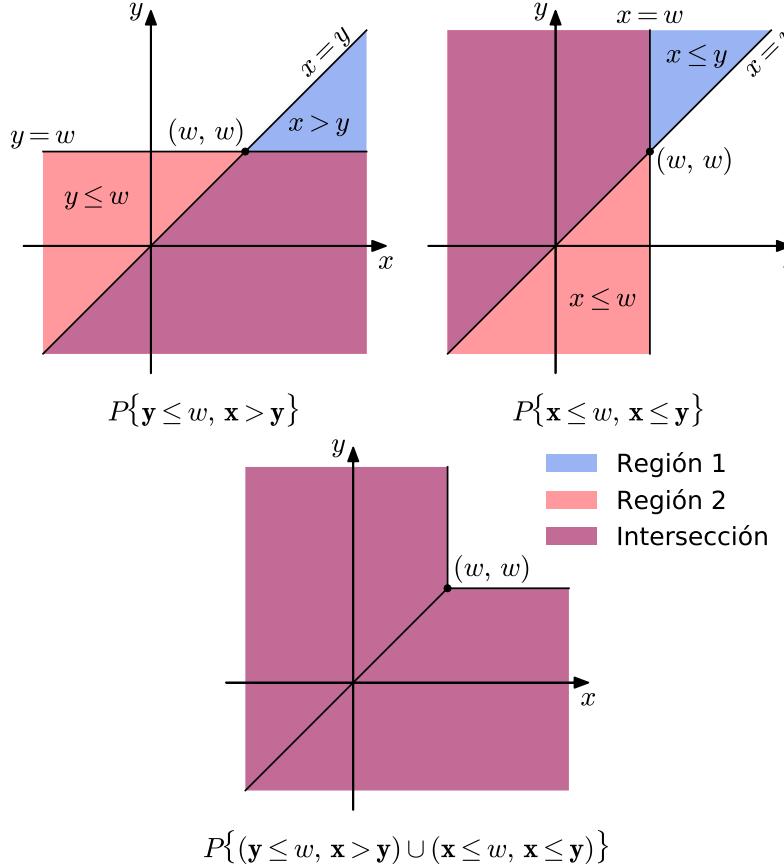


Figura 5.8: Región del plano correspondiente a $F_z(z) = P\{\min(\mathbf{x}, \mathbf{y}) \leq z\}$.

si $\mathbf{x} = 1$, \mathbf{y} debe ser $n - 1$, y así sucesivamente. Por lo tanto, el evento $\{\mathbf{x} + \mathbf{y} = n\}$ es la unión de los eventos mutuamente excluyentes $A_k = \{\mathbf{x} = k, \mathbf{y} = n - k\}$, con $k = 0, 1, \dots, n$. Por lo tanto,

$$P\{\mathbf{z} = n\} = P\{\mathbf{x} + \mathbf{y} = n\} = P\left\{\bigcup_{k=0}^n \{\mathbf{x} = k, \mathbf{y} = n - k\}\right\} = \sum_{k=0}^n P\{\mathbf{x} = k, \mathbf{y} = n - k\}.$$

Además, como \mathbf{x} y \mathbf{y} son independientes, se cumple que

$$P\{\mathbf{x} = k, \mathbf{y} = n - k\} = P\{\mathbf{x} = k\}P\{\mathbf{y} = n - k\}.$$

De esta forma,

$$\begin{aligned} P\{\mathbf{z} = n\} &= \sum_{k=0}^n P\{\mathbf{x} = k\}P\{\mathbf{y} = n - k\} \\ &\stackrel{(a)}{=} \sum_{k=0}^n e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!} \\ &\stackrel{(b)}{=} \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \sum_{k=0}^n \frac{n!}{k!(n-k)!} \lambda_1^k \lambda_2^{n-k}, \end{aligned}$$

donde en (a) se sustituyó la p.m.f. de Poisson, dada por la ecuación 3.23 y en (b) se multiplicó y dividió entre $n!$. Notando que la sumatoria es $(\lambda_1 + \lambda_2)^n$, como indica la ecuación 2.9, se obtiene que

$$P\{\mathbf{z} = n\} = e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1 + \lambda_2)^n}{n!}, \quad n = 0, 1, \dots \quad (5.30)$$

Se concluye que \mathbf{z} es una variable aleatoria de Poisson con parámetro $\lambda_1 + \lambda_2$, indicando que la suma de variables aleatorias de Poisson independientes es una variable aleatoria de Poisson con parámetro la suma de los parámetros de las variables aleatorias originales.

Como muestra este ejemplo simple, el procedimiento es en general tedioso para el caso de variables aleatorias discretas. En la sección 5.5 se verá que problemas de este tipo pueden resolverse de forma mucho mas fácil mediante funciones características o funciones generadoras de momentos conjuntas.

5.3. Dos funciones de dos variables aleatorias

Se considera ahora una generalización del caso de la sección anterior. Sean \mathbf{x} e \mathbf{y} variables aleatorias con densidad de probabilidad conjunta $f_{xy}(x, y)$. Dadas dos funciones $g(x, y)$ y $h(x, y)$, se definen dos nuevas variables aleatorias

$$\begin{aligned}\mathbf{z} &= g(\mathbf{x}, \mathbf{y}) \\ \mathbf{w} &= h(\mathbf{x}, \mathbf{y}),\end{aligned}$$

y se quiere determinar la densidad de probabilidad conjunta $f_{zw}(z, w)$. Luego, los marginales $f_z(z)$ y $f_w(w)$ pueden obtenerse integrando $f_{zw}(z, w)$.

De forma análoga al caso de una función de dos variables aleatorias dado en la ecuación 5.14, la distribución de probabilidad es

$$\begin{aligned}F_{zw}(z, w) &= P\{\mathbf{z} \leq z, \mathbf{w} \leq w\} \\ &= P\{g(\mathbf{x}, \mathbf{y}) \leq z, h(\mathbf{x}, \mathbf{y}) \leq w\} \\ &= P\{(\mathbf{x}, \mathbf{y}) \in D_{z, w}\} \\ &= \iint_{(x, y) \in D_{z, w}} f_{xy}(x, y) dx dy,\end{aligned}\tag{5.31}$$

donde $D_{z, w}$ es la región en el plano xy donde las inecuaciones $g(x, y) \leq z$ y $h(x, y) \leq w$ se cumplen simultáneamente. La densidad de probabilidad conjunta $f_{zw}(z, w)$ puede obtenerse a partir de las derivadas de la distribución de probabilidad $F_{zw}(z, w)$ (ecuación 5.4), pero como se verá mas adelante, también puede obtenerse directamente de $f_{xy}(x, y)$.

Ejemplo Sean \mathbf{x} e \mathbf{y} variables aleatorias independientes de distribución uniforme en el intervalo $(0, \theta)$. Se definen las variables aleatorias $\mathbf{z} = \max(\mathbf{x}, \mathbf{y})$ y $\mathbf{w} = \min(\mathbf{x}, \mathbf{y})$. Se quiere determinar $f_{zw}(z, w)$.

Se comenzará calculando la distribución de probabilidad $F_{zw}(z, w)$. La densidad de probabilidad $f_{zw}(z, w)$ se obtiene diferenciando la distribución. Por definición,

$$F_{zw}(z, w) = P\{\mathbf{z} \leq z, \mathbf{w} \leq w\} = P\{\max(\mathbf{x}, \mathbf{y}) \leq z, \min(\mathbf{x}, \mathbf{y}) \leq w\}.$$

Como \mathbf{x} e \mathbf{y} varían en el intervalo $(0, \theta)$, $\mathbf{z} = \max(\mathbf{x}, \mathbf{y})$ y $\mathbf{w} = \min(\mathbf{x}, \mathbf{y})$ varían en el mismo intervalo. Por lo tanto,

$$F_{zw}(z, w) = 0, \quad \text{si } z < 0 \text{ o } w < 0.$$

La región de integración $D_{z, w}$ en la ecuación 5.31 se trata de la intersección de las regiones de integración de las figuras 5.7 y 5.8, por lo que hay que diferenciar los casos $w \geq z$ y $w < z$, ya que dan lugar a regiones distintas, como se muestra en la figura 5.9.

En el caso en que $w < z$, observando la región intersección en la figura 5.9, se deduce que

$$F_{zw}(z, w) = F_{xy}(z, w) + F_{xy}(w, z) - F_{xy}(w, w), \quad 0 \leq w < z < \theta,$$

ya que $F_{xy}(z, w)$ es el rectángulo horizontal de vértices (z, w) y $(0, 0)$ en las diagonales, $F_{xy}(z, w)$ es el rectángulo vertical de vértices (w, z) y $(0, 0)$ en las diagonales, y como estas regiones se solapan, se está considerando dos veces el cuadrado de vértices (w, w) y $(0, 0)$ en las diagonales, por lo que hay que restar $F_{xy}(w, w)$. En el caso en que $w \geq z$, de la figura 5.9 se obtiene que

$$F_{zw}(z, w) = F_{xy}(z, z), \quad 0 \leq z \leq w < \theta.$$

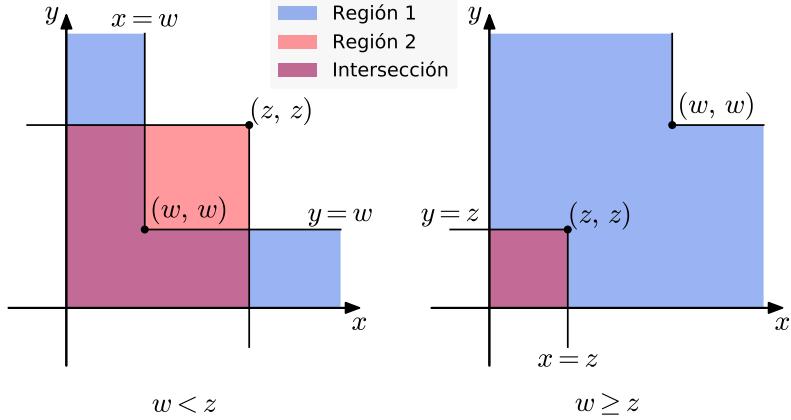


Figura 5.9: Regiones del plano correspondientes a $F_z(z, w) = P\{\max(\mathbf{x}, \mathbf{y}) \leq z, \min(\mathbf{x}, \mathbf{y}) \leq w\}$.

Además, en el caso en que $w > \theta$ se cumple que

$$\begin{aligned} F_{zw}(z, w) &= P\{\mathbf{z} \leq z, \mathbf{w} \leq w\} \\ &= P\{(\mathbf{z} \leq z) \cap (\mathbf{w} \leq w)\} \\ &\stackrel{(a)}{=} P\{(\mathbf{z} \leq z) \cap S\} \\ &\stackrel{(b)}{=} P\{\mathbf{z} \leq z\} \\ &= F_z(z), \end{aligned}$$

donde en (a) se tuvo en cuenta que como $w > \theta$ y \mathbf{w} varía en el intervalo $(0, \theta)$, se cumple que

$$\{\mathbf{w} \leq w\} = \{(\mathbf{w} < \theta) \cup (\theta \leq \mathbf{w} \leq w)\} = \{(\mathbf{w} < \theta) \cup \emptyset\} = \{\mathbf{w} < \theta\} = S$$

es decir, el evento $\{\mathbf{w} \leq w\}$ es todo el espacio muestral S , y en (b) se consideró que si A es un evento arbitrario, se cumple que $A \cap S = A$. De este razonamiento se concluye que

$$\begin{aligned} F_{zw}(z, w) &= F_z(z) \stackrel{(a)}{=} F_{xy}(z, z), & 0 \leq z < \theta \leq w. \\ F_{zw}(z, w) &= F_w(w) \stackrel{(b)}{=} F_x(w) + F_y(w) - F_{xy}(w, w), & 0 \leq w < \theta \leq z \\ F_{zw}(z, w) &= 1, & z \geq \theta \quad y \quad w \geq \theta, \end{aligned}$$

donde en (a) y en (b) se usaron los resultados de las ecuaciones 5.27 y 5.29 respectivamente. Juntando estos resultados se llega a que

$$F_{zw}(z, w) = \begin{cases} 0, & z < 0 \quad o \quad w < 0 \\ F_{xy}(z, w) + F_{xy}(w, z) - F_{xy}(w, w), & 0 \leq w < z < \theta \\ F_{xy}(z, z), & 0 \leq z \leq w < \theta \\ F_{xy}(z, z), & 0 \leq z < \theta \leq w \\ F_x(w) + F_y(w) - F_{xy}(w, w), & 0 \leq w < \theta \leq z \\ 1, & z \geq \theta \quad y \quad z \geq \theta. \end{cases} \quad (5.32)$$

Por otro lado, como $\mathbf{x} \sim U(0, \theta)$ e $\mathbf{y} \sim U(0, \theta)$, su distribución de probabilidad es

$$F_x(x) = F_y(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{\theta}, & 0 \leq x < \theta \\ 1, & x \geq \theta \end{cases}$$

y como además son independientes,

$$F_{xy}(x, y) = F_x(x)F_y(y) = \begin{cases} 0, & x < 0 \text{ o } y < 0 \\ \frac{x}{\theta} \cdot \frac{y}{\theta} = \frac{xy}{\theta^2}, & 0 \leq x < \theta \text{ y } 0 \leq y < \theta \\ \frac{x}{\theta}, & 0 \leq x < \theta \leq y \\ \frac{y}{\theta}, & 0 \leq y < \theta \leq x \\ 1, & x \geq \theta \text{ y } y \geq \theta. \end{cases}$$

Finalmente, sustituyendo estos resultados en la ecuación 5.32 se llega a que

$$F_{zw}(z, w) = \begin{cases} 0, & z < 0 \text{ o } w < 0 \\ \frac{zw}{\theta^2} + \frac{zw}{\theta^2} - \frac{w^2}{\theta^2} = \frac{2zw - w^2}{\theta^2}, & 0 \leq w < z < \theta \\ \frac{z^2}{\theta^2}, & 0 \leq z \leq w < \theta \\ \frac{z^2}{\theta^2}, & 0 \leq z < \theta \leq w \\ \frac{w}{\theta} + \frac{w}{\theta} - \frac{w^2}{\theta^2} = \frac{2\theta w - w^2}{\theta^2}, & 0 \leq w < \theta \leq z \\ 1, & z \geq \theta \text{ y } z \geq \theta. \end{cases} \quad (5.33)$$

En la figura 5.10 se muestra la distribución de probabilidad $F_{zw}(z, w)$.

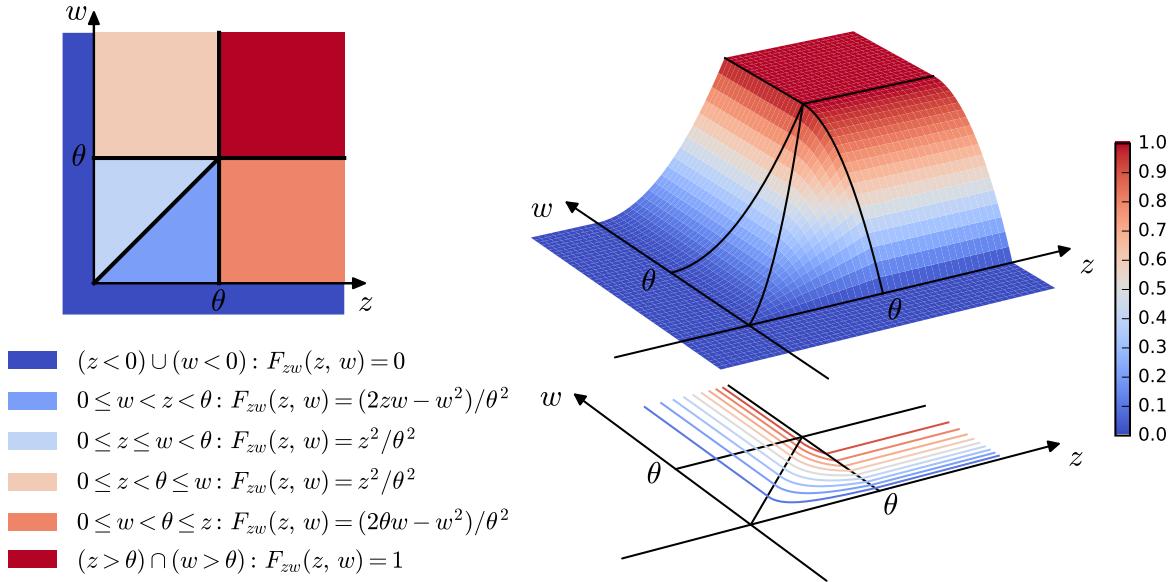


Figura 5.10: Distribución de probabilidad $F_z(z, w) = P\{\max(\mathbf{x}, \mathbf{y}) \leq z, \min(\mathbf{x}, \mathbf{y}) \leq w\}$. A la izquierda se muestra el valor que toma la función en las distintas regiones del plano. A la derecha se muestra la superficie $F_{zw}(z, w)$ y las curvas de nivel en función de z y w .

La densidad de probabilidad puede calcularse como

$$f_{zw}(z, w) = \frac{\partial^2 F_{zw}(z, w)}{\partial z \partial w},$$

lo que resulta en

$$f_{zw}(z, w) = \begin{cases} \frac{2}{\theta^2}, & 0 \leq w < z < \theta \\ 0, & \text{en otro caso.} \end{cases}$$

la cual se muestra en la figura 5.11.

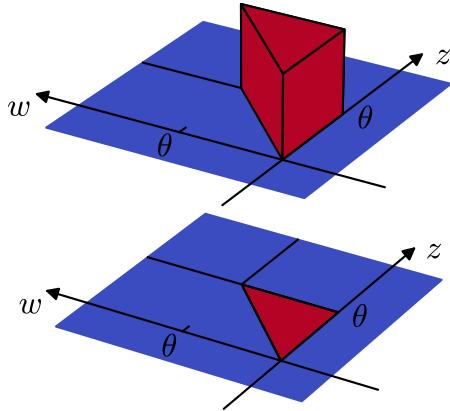


Figura 5.11: Densidad de probabilidad $f_{zw}(z, w)$ de las variables aleatorias $\mathbf{z} = \max(\mathbf{x}, \mathbf{y})$ y $\mathbf{w} = \min(\mathbf{x}, \mathbf{y})$. $f_{zw}(z, w)$ toma el valor constante $2/\theta^2$ en la región triangular $0 \leq w < z < \theta$ y 0 en el resto del plano zw .

Integrando la densidad de probabilidad $f_{zw}(z, w)$ como indica la ecuación 5.31, puede obtenerse la distribución de probabilidad $F_{zw}(z, w)$,

$$F_{zw}(z, w) = \int_{-\infty}^z \int_{-\infty}^w f_{zw}(z, w) dw dz,$$

Esto consiste en calcular el volumen de $f_{zw}(z, w)$ en el cuadrante determinado por las semirrectas $(-\infty, z)$ y $(-\infty, w)$ para todo valor de (z, w) . Como en este ejemplo $f_{zw}(z, w)$ es una función sencilla, constante en una región triangular, el cálculo puede realizarse de forma geométrica, teniendo en cuenta que el volumen cambia según la región del plano donde se encuentren z y w , como se muestra en la figura 5.12. Para hacerlo, se observa que:

- si $0 \leq w < z < \theta$ como z_1 y w_1 en la figura 5.12, el área de la base es

$$\text{Área} = \frac{w^2}{2} + (z-w)w = zw - \frac{w^2}{2},$$

y la altura es $f_{zw}(z, w) = 2/\theta^2$, por lo que el volumen es

$$F_{zw}(z, w) = \left(zw - \frac{w^2}{2} \right) \frac{2}{\theta^2} = \frac{2zw - w^2}{\theta^2}.$$

- si $0 \leq z < \theta$ y $z < w$ como z_2 y w_2 en la figura 5.12, se tiene que

$$F_{zw}(z, w) = \frac{z^2}{2} \cdot \frac{2}{\theta^2} = \frac{z^2}{\theta^2}.$$

- si $0 \leq w < \theta \leq z$ como z_3 y w_3 en la figura 5.12, se tiene que

$$F_{zw}(z, w) = \left((\theta - w)w + \frac{w^2}{2} \right) \frac{2}{\theta^2} = \frac{2\theta w - w^2}{\theta^2}.$$

- si $z > \theta$ y $w > \theta$ como z_4 y w_4 en la figura 5.12, se tiene que

$$F_{zw}(z, w) = \frac{\theta^2}{2} \cdot \frac{2}{\theta^2} = 1.$$

Este resultado coincide con el de la ecuación 5.33.

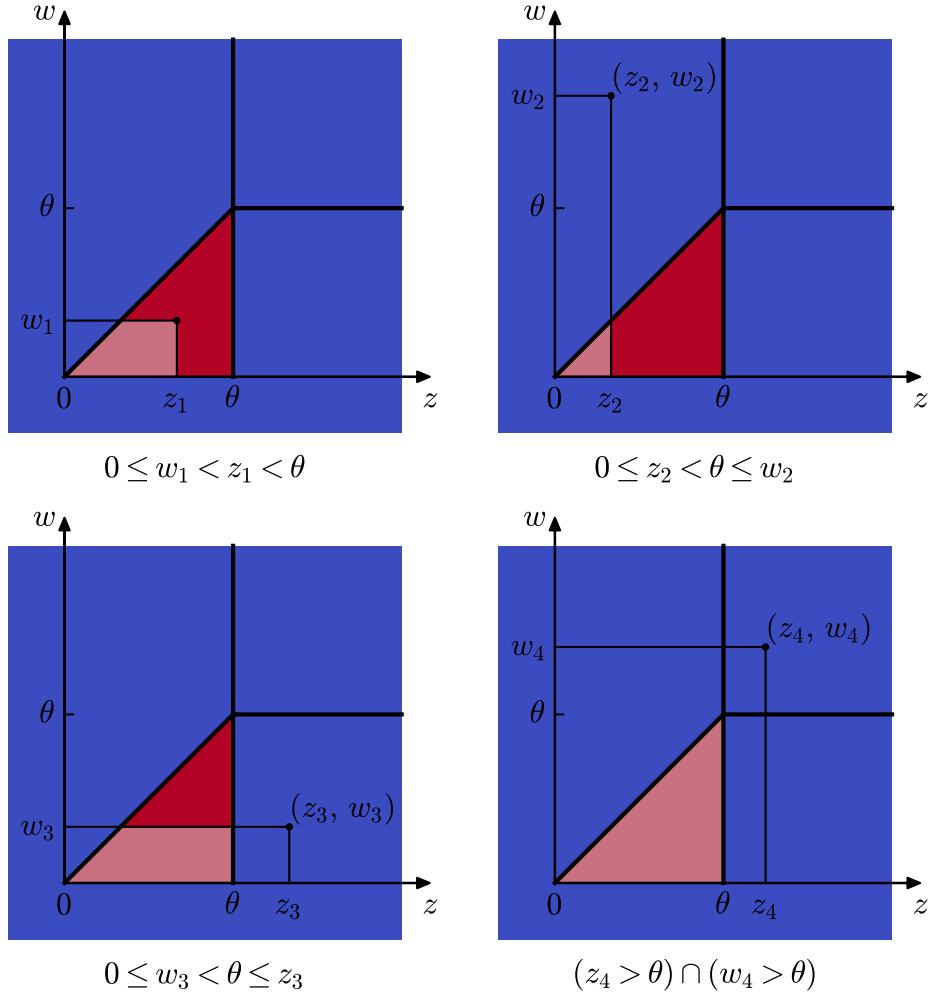


Figura 5.12: Cálculo de $F_{zw}(z, w)$ a partir de la integración de $f_{zw}(z, w)$. En la figura se muestra el área de la base para el cálculo del volumen abarcado por $f_{zw}(z, w)$ según la región del plano donde se encuentren z y w .

5.3.1. Densidad de probabilidad conjunta

Sean $g(x, y)$ y $h(x, y)$ dos funciones continuas y diferenciables y dos variables aleatorias \mathbf{x} e \mathbf{y} . Se definen las variables aleatorias \mathbf{z} y \mathbf{w} como

$$\begin{aligned}\mathbf{z} &= g(\mathbf{x}, \mathbf{y}) \\ \mathbf{w} &= h(\mathbf{x}, \mathbf{y}).\end{aligned}$$

Al igual que en el caso de una única variable aleatoria (ecuación 4.4), es posible desarrollar una fórmula para obtener la densidad de probabilidad conjunta $f_{zw}(z, w)$ directamente. Se probará que las densidades de probabilidad $f_{xy}(x, y)$ y $f_{zw}(z, w)$ se vinculan como

$$f_{zw}(z, w) = \sum_i |J(z, w)| f_{xy}(x_i, y_i) = \sum_i \frac{1}{|J(x_i, y_i)|} f_{xy}(x_i, y_i), \quad (5.34)$$

donde (x_i, y_i) son todas las soluciones del sistema de ecuaciones

$$g(x, y) = z, \quad h(x, y) = w \quad (5.35)$$

para un punto dado (z, w) , es decir, $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ son todos los puntos tal que

$$g(x_i, y_i) = z, \quad h(x_i, y_i) = w.$$

Equivalentemente, si g_1 y h_1 es la transformación inversa a 5.35, se cumple que

$$x_i = g_1(z, w), \quad y_i = h_1(z, w). \quad (5.36)$$

Además, $|J(z, w)|$ y $|J(x_i, y_i)|$ son el valor absoluto del determinante de la matriz Jacobiana de la transformación inversa y directa respectivamente, es decir,

$$J(z, w) = \begin{vmatrix} \frac{\partial g_1}{\partial z} & \frac{\partial g_1}{\partial w} \\ \frac{\partial h_1}{\partial z} & \frac{\partial h_1}{\partial w} \end{vmatrix} \quad J(x_i, y_i) = \left. \begin{vmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{vmatrix} \right|_{x=x_i, y=y_i}.$$

Como la matriz inversa de la matriz Jacobiana de una transformación invertible es la matriz Jacobiana de la transformación inversa, se cumple que

$$|J(z, w)| = \frac{1}{|J(x_i, y_i)|}.$$

Para la demostración, hay que evaluar la probabilidad

$$P\{z < \mathbf{z} \leq z + \Delta z, w < \mathbf{w} \leq w + \Delta w\} = P\{z < g(\mathbf{x}, \mathbf{y}) \leq z + \Delta z, w < h(\mathbf{x}, \mathbf{y}) \leq w + \Delta w\}. \quad (5.37)$$

Para un punto dado (z, w) , el sistema de ecuaciones 5.35 puede tener múltiples soluciones. Sean $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ dichas soluciones, las cuales se representan en la figura 5.13, y cumplen que

$$g(x_i, y_i) = z, \quad h(x_i, y_i) = w.$$

Empleando la ecuación 5.7, la probabilidad de la ecuación 5.37 puede escribirse como

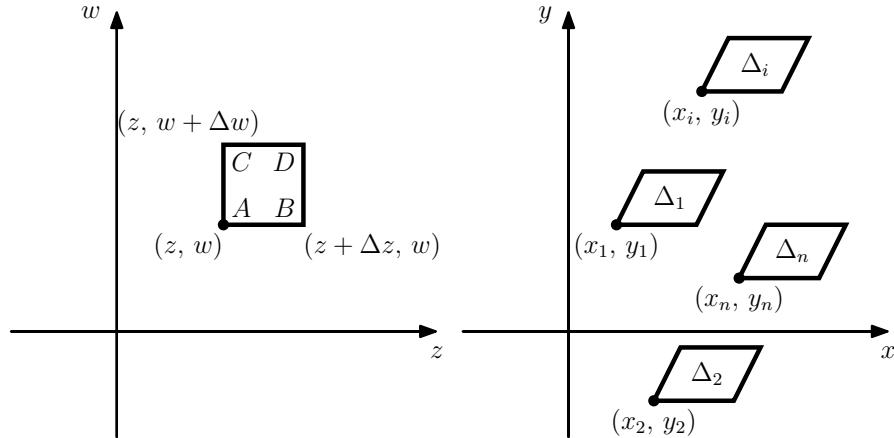


Figura 5.13: Mapeo del punto (z, w) y el rectángulo de área $\Delta z \Delta w$ en el plano xy mediante la transformación 5.35. Para un punto (z, w) dado, $(x_1, y_1), \dots, (x_n, y_n)$ son las soluciones del sistema.

$$P\{z < \mathbf{z} \leq z + \Delta z, w < \mathbf{w} \leq w + \Delta w\} = f_{zw}(z, w) \Delta z \Delta w. \quad (5.38)$$

Para expresar esta probabilidad en términos de $f_{xy}(x, y)$, es necesario evaluar la región en la que se mapea el rectángulo de área $\Delta z \Delta w$ del plano zw en el plano xy . Respecto a esto, como se muestra en la

figura 5.14, el punto A con coordenadas (z, w) se mapea en el punto A' con coordenadas (x_i, y_i) , además de en otros puntos, como se mencionó previamente y se muestra en la figura 5.13. Cuando z cambia a $z + \Delta z$ hacia el punto B , sea B' su imagen en el plano xy , y de forma similar, cuando w cambia a $w + \Delta w$ hacia el punto C , sea C' su imagen en el plano xy , como se muestra en la figura 5.14. Finalmente, D se mapea en el punto D' , y $A'B'C'D'$ representa el paralelogramo equivalente en el plano xy con área Δ_i . Debido a que estas regiones no se solapan, como en la figura 5.13, la probabilidad de la ecuación 5.7 puede expresarse en el plano xy como

$$\sum_i P\{(\mathbf{x}, \mathbf{y}) \in \Delta_i\} = \sum_i f_{xy}(x_i, y_i) \Delta_i. \quad (5.39)$$

Igualando las ecuaciones 5.38 y 5.39 se obtiene que

$$f_{zw}(z, w) = \sum_i f_{xy}(x_i, y_i) \frac{\Delta_i}{\Delta z \Delta w}. \quad (5.40)$$

Para desarrollar la ecuación 5.40, se necesita calcular el área del paralelogramo de la figura 5.14 en

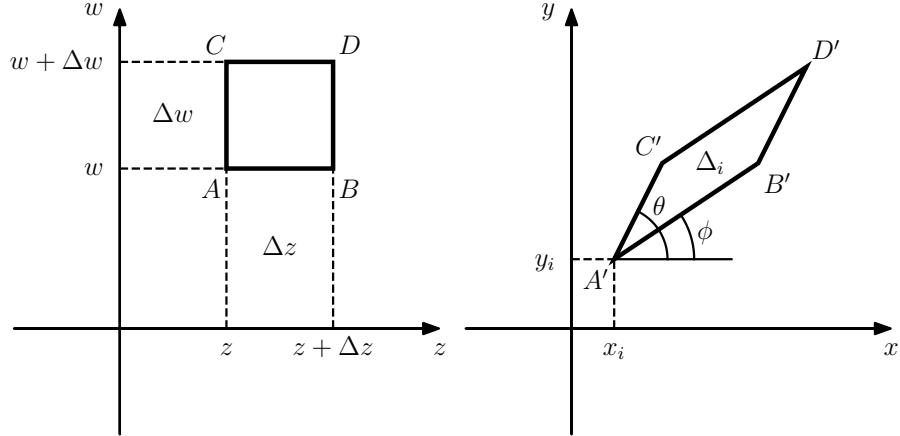


Figura 5.14: Imagen $A'B'C'D'$ en el plano xy del rectángulo $ABCD$ en el plano zw correspondiente a una solución (x_i, y_i) de la transformación 5.35. El rectángulo de área $\Delta z \Delta w$ se mapea en un paralelogramo de área Δ_i .

términos de $\Delta z \Delta w$. Con ese objetivo, sea g_1 y h_1 la transformación inversa a 5.35, de forma que

$$x_i = g_1(z, w), \quad y_i = h_1(z, w).$$

Teniendo esto en cuenta, las coordenadas de los puntos en los planos zw y xy son:

Plano zw		Plano xy	
Punto	Coordenadas	Punto	Coordenadas
A	(z, w)	A'	$[g_1(z, w), h_1(z, w)]$
B	$(z + \Delta z, w)$	B'	$[g_1(z + \Delta z, w), h_1(z + \Delta z, w)]$
C	$(z, w + \Delta w)$	C'	$[g_1(z, w + \Delta w), h_1(z, w + \Delta w)]$
D	$(z + \Delta z, w + \Delta w)$	D'	$[g_1(z + \Delta z, w + \Delta w), h_1(z + \Delta z, w + \Delta w)]$

Realizando una aproximación de Taylor con $\Delta z \rightarrow 0$ y $\Delta w \rightarrow 0$, se tiene que

$$\begin{aligned} g_1(z + \Delta z, w) &= g_1(z, w) + \frac{\partial g_1}{\partial z} \Delta z = x_i + \frac{\partial g_1}{\partial z} \Delta z \\ h_1(z + \Delta z, w) &= h_1(z, w) + \frac{\partial h_1}{\partial z} \Delta z = y_i + \frac{\partial h_1}{\partial z} \Delta z \\ g_1(z, w + \Delta w) &= g_1(z, w) + \frac{\partial g_1}{\partial w} \Delta w = x_i + \frac{\partial g_1}{\partial w} \Delta w \\ h_1(z, w + \Delta w) &= h_1(z, w) + \frac{\partial h_1}{\partial w} \Delta w = y_i + \frac{\partial h_1}{\partial w} \Delta w, \end{aligned}$$

por lo que las coordenadas de los puntos en el plano xy resultan en los valores mostrados en la siguiente tabla:

A'	B'	C'
(x_i, y_i)	$\left(x_i + \frac{\partial g_1}{\partial z} \Delta z, y_i + \frac{\partial h_1}{\partial z} \Delta z\right)$	$\left(x_i + \frac{\partial g_1}{\partial w} \Delta w, y_i + \frac{\partial h_1}{\partial w} \Delta w\right)$

El área del paralelogramo $A'B'C'D'$ en la figura 5.14 está dado por

$$\begin{aligned} \Delta_i &\stackrel{(a)}{=} (A'B')(A'C') \sin(\theta - \phi) \\ &\stackrel{(b)}{=} (A'B')(A'C') [\sin \theta \cos \phi - \cos \theta \sin \phi] \\ &= (A'B' \cos \phi)(A'C' \sin \theta) - (A'B' \sin \phi)(A'C' \cos \theta) \\ &\stackrel{(c)}{=} \frac{\partial g_1}{\partial z} \Delta z \frac{\partial h_1}{\partial w} \Delta w - \frac{\partial h_1}{\partial z} \Delta z \frac{\partial g_1}{\partial w} \Delta w, \end{aligned}$$

donde en (a) se empleó que el área del paralelogramo es $\Delta_i = bh$ (base \times altura), donde la base es $b = A'B'$ y la altura es $h = (A'C') \sin(\theta - \phi)$, en (b) se empleó la identidad trigonométrica $\sin(\theta - \phi) = \sin \theta \cos \phi - \cos \theta \sin \phi$ y en (c) se consideró que, como se observa en la figura 5.14, las coordenadas de los puntos B' y C' se pueden expresar como

$$B' : (x_i + A'B' \cos \phi, y_i + A'B' \sin \phi) \quad \text{y} \quad C' : (x_i + A'C' \cos \theta, y_i + A'C' \sin \theta),$$

por lo que comparando con la tabla, se deduce que

$$\begin{aligned} A'B' \cos \phi &= \frac{\partial g_1}{\partial z} \Delta z & A'C' \sin \theta &= \frac{\partial h_1}{\partial w} \Delta w \\ A'B' \sin \phi &= \frac{\partial h_1}{\partial z} \Delta z & A'C' \cos \theta &= \frac{\partial g_1}{\partial w} \Delta w. \end{aligned}$$

Se obtuvo que el área del paralelogramo es

$$\Delta_i = \left(\frac{\partial g_1}{\partial z} \frac{\partial h_1}{\partial w} - \frac{\partial g_1}{\partial w} \frac{\partial h_1}{\partial z} \right) \Delta z \Delta w$$

y por lo tanto

$$\frac{\Delta_i}{\Delta z \Delta w} = \left(\frac{\partial g_1}{\partial z} \frac{\partial h_1}{\partial w} - \frac{\partial g_1}{\partial w} \frac{\partial h_1}{\partial z} \right) = \begin{vmatrix} \frac{\partial g_1}{\partial z} & \frac{\partial g_1}{\partial w} \\ \frac{\partial h_1}{\partial z} & \frac{\partial h_1}{\partial w} \end{vmatrix}.$$

El determinante en el lado derecho de la igualdad representa el valor absoluto de la matriz Jacobiana $J(z, w)$ de la transformación inversa 5.36. Sustituyendo este resultado en la ecuación 5.40, se concluye que

$$f_{zw}(z, w) = \sum_i |J(z, w)| f_{xy}(x_i, y_i) = \sum_i \frac{1}{|J(x_i, y_i)|} f_{xy}(x_i, y_i),$$

donde $J(x_i, y_i)$ es la matriz Jacobiana de la transformación original dada por la ecuación 5.35, y por el teorema de la función inversa³, se cumple que $J(z, w) = J^{-1}(x_i, y_i)$ y por lo tanto, $\det[J(z, w)] = 1/\det[J(x_i, y_i)]$.

Transformación lineal Sean

$$\mathbf{z} = a\mathbf{x} + b\mathbf{y}, \quad \mathbf{w} = c\mathbf{x} + d\mathbf{y}.$$

Se quiere calcular $f_{zw}(z, w)$. Se considera el sistema de ecuaciones

$$\begin{cases} ax + by = z \\ cx + dy = w \end{cases}.$$

La solución del sistema es (regla de Cramer)

$$x = \frac{\Delta x}{\Delta}, \quad y = \frac{\Delta y}{\Delta}$$

con

$$\Delta = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc, \quad \Delta x = \begin{vmatrix} z & b \\ w & d \end{vmatrix} = zd - bw, \quad \Delta y = \begin{vmatrix} a & z \\ c & w \end{vmatrix} = aw - zc$$

y el sistema tiene una única solución si $\Delta = ad - bc \neq 0$. En ese caso, se obtiene que

$$x = \frac{dz - bw}{ad - bc} = Az + Bw, \quad y = \frac{-cz + aw}{ad - bc} = Cz + Dw.$$

Como

$$g(x, y) = ax + by, \quad h(x, y) = cx + dy,$$

$J(x, y) = ad - bc$, por lo que sustituyendo estos resultados en la ecuación 5.34 se llega a que

$$f_{zw}(z, w) = \frac{1}{|ad - bc|} f_{xy}(Az + Bw, Cz + Dw). \quad (5.41)$$

Ejemplo: variables aleatorias conjuntamente normales Sean las variables aleatorias \mathbf{x} e \mathbf{y} conjuntamente normales con densidad $N(\eta_x, \eta_y, \sigma_x^2, \sigma_y^2, \rho)$. Se definen las variables aleatorias \mathbf{z} y \mathbf{w} como la siguiente transformación lineal de \mathbf{x} e \mathbf{y} ,

$$\mathbf{z} = a\mathbf{x} + b\mathbf{y}, \quad \mathbf{w} = c\mathbf{x} + d\mathbf{y}.$$

La densidad de probabilidad conjunta $f_{zw}(z, w)$ se obtiene a partir de la ecuación 5.41 con $f_{xy}(x, y)$ dada por las ecuaciones 5.11 y 5.12. Operando, puede verse que $f_{zw}(z, w)$ es una exponencial con exponente cuadrático en z y w de la misma forma de $f_{xy}(x, y)$. Se concluye que \mathbf{z} y \mathbf{w} también son conjuntamente normales con densidad conjunta $N(\eta_z, \eta_w, \sigma_z^2, \sigma_w^2, \rho_{zw})$ con

$$\begin{aligned} \eta_z &= a\eta_x + b\eta_y \\ \eta_w &= c\eta_x + d\eta_y \\ \sigma_z^2 &= a^2\sigma_x^2 + 2ab\rho\sigma_x\sigma_y + b^2\sigma_y^2 \\ \sigma_w^2 &= c^2\sigma_x^2 + 2cd\rho\sigma_x\sigma_y + d^2\sigma_y^2 \\ \rho_{zw} &= \frac{ac\sigma_x^2 + (ad + bc)\rho\sigma_x\sigma_y + bd\sigma_y^2}{\sigma_z\sigma_w} \end{aligned} \quad (5.42)$$

En general, cualquier combinación lineal de dos variables aleatorias conjuntamente normales es normal. La deducción de estos resultados es tediosa, por lo que se omite aquí pero se hará mas adelante usando la función característica.

³Ver https://en.wikipedia.org/wiki/Inverse_function_theorem, por ejemplo.

Ejemplo 6-22 Sean \mathbf{x} e \mathbf{y} dos variables aleatorias independientes, de distribución gaussiana, con media nula y varianza común σ^2 . Se definen las variables aleatorias

$$\mathbf{r} = \sqrt{\mathbf{x}^2 + \mathbf{y}^2}, \quad \theta = \tan^{-1}(\mathbf{y}/\mathbf{x}), \quad (5.43)$$

con $|\theta| < \pi$, y se quiere calcular su densidad de probabilidad conjunta.

Como las variables aleatorias \mathbf{x} e \mathbf{y} son independientes, su densidad de probabilidad conjunta es

$$\begin{aligned} f_{xy}(x, y) &= f_x(x)f_y(y) \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} \right) \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-y^2/2\sigma^2} \right) \\ &= \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \end{aligned}$$

Teniendo en cuenta que $|\theta| < \pi$, el sistema de ecuaciones

$$\begin{cases} r = g(x, y) = \sqrt{x^2 + y^2} \\ \theta = h(x, y) = \tan^{-1}(y/x) \end{cases} \quad (5.44)$$

tiene como única solución

$$x_1 = r \cos \theta, \quad y_1 = r \sin \theta,$$

y la transformación inversa es

$$\begin{cases} g_1(r, \theta) = r \cos \theta \\ h_1(r, \theta) = r \sin \theta. \end{cases}$$

El determinante de la matriz jacobiana de la transformación inversa es

$$J(r, \theta) = \begin{vmatrix} \frac{\partial g_1}{\partial r} & \frac{\partial g_1}{\partial \theta} \\ \frac{\partial h_1}{\partial r} & \frac{\partial h_1}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r \cos^2 \theta + r \sin^2 \theta = r,$$

por lo que

$$|J(r, \theta)| = r.$$

Aplicando la ecuación 5.34, se llega a que

$$f_{r\theta}(r, \theta) = r f_{xy}(x_1, y_1) = \frac{r}{2\pi\sigma^2} e^{-r^2/2\sigma^2}, \quad 0 < r < \infty, \quad |\theta| < \pi.$$

Las distribuciones marginales son

$$\begin{aligned} f_r(r) &= \int_{-\pi}^{\pi} f_{r\theta}(r, \theta) d\theta \\ &= \frac{r}{2\pi\sigma^2} e^{-r^2/2\sigma^2} \int_{-\pi}^{\pi} 1 \cdot d\theta \\ &= \frac{r}{2\pi\sigma^2} e^{-r^2/2\sigma^2} \left(\theta \Big|_{-\pi}^{\pi} \right) \\ &= \frac{r}{2\pi\sigma^2} e^{-r^2/2\sigma^2} [\pi - (-\pi)] \\ &= \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} \end{aligned}$$

y

$$\begin{aligned}
f_\theta(\theta) &= \int_0^\infty f_{r\theta}(r, \theta) dr \\
&= \frac{1}{2\pi} \int_0^\infty \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} d\theta \\
&= \frac{1}{2\pi} \left(-e^{-r^2/2\sigma^2} \Big|_0^\infty \right) \\
&= \frac{1}{2\pi} [-(0-1)] \\
&= \frac{1}{2\pi}.
\end{aligned}$$

En resumen, la densidad de \mathbf{r} es

$$f_r(r) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2}, \quad 0 < r < \infty,$$

que representa una variable aleatoria Rayleigh con parámetro σ^2 , y la densidad de θ es

$$f_\theta(\theta) = \frac{1}{2\pi}, \quad |\theta| < \pi,$$

que representa una variable aleatoria uniforme en el intervalo $(-\pi, \pi)$. Notar que se cumple que

$$f_{r\theta}(r, \theta) = f_r(r)f_\theta(\theta),$$

lo que implica que las variables aleatorias \mathbf{r} y θ son independientes.

Notar que el mismo cálculo se podría haber realizado a partir de la matriz jacobiana de la transformación directa 5.44 en la ecuación 5.34. Efectivamente, la derivadas parciales en

$$J(x, y) = \begin{vmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{vmatrix}$$

son

$$\frac{\partial}{\partial x} \left[(x^2 + y^2)^{\frac{1}{2}} \right] = \frac{1}{2} (x^2 + y^2)^{-\frac{1}{2}} \cdot 2x = \frac{x}{\sqrt{x^2 + y^2}}$$

y con el mismo razonamiento se obtiene la derivada respecto a y . Además

$$\begin{aligned}
\frac{\partial}{\partial x} [\tan^{-1}(y/x)] &= \frac{1}{1 + y^2/x^2} \cdot \left(-\frac{y}{x^2} \right) = \frac{-y}{x^2 + y^2} \\
\frac{\partial}{\partial y} [\tan^{-1}(y/x)] &= \frac{1}{1 + y^2/x^2} \cdot \frac{1}{x} = \frac{x}{x^2 + y^2},
\end{aligned}$$

donde se usó que $(\tan^{-1})'(x) = 1/(1 + x^2)$ (ecuación 5.24) y la regla de la cadena. Por lo tanto

$$J(x, y) = \begin{vmatrix} \frac{x}{\sqrt{x^2 + y^2}} & \frac{y}{\sqrt{x^2 + y^2}} \\ \frac{-y}{x^2 + y^2} & \frac{x}{x^2 + y^2} \end{vmatrix} = \frac{x^2 + y^2}{(x^2 + y^2)\sqrt{x^2 + y^2}} = \frac{1}{\sqrt{x^2 + y^2}} = \frac{1}{r},$$

obteniéndose que

$$J(r, \theta) = \frac{1}{J(x, y)},$$

como se mencionó previamente.

Como corolario de este ejemplo, considérese la variable aleatoria compleja gaussiana $\mathbf{x} + j\mathbf{y}$, con \mathbf{x} e \mathbf{y} variables aleatorias independientes, gaussianas y de media nula. Como

$$\mathbf{x} + j\mathbf{y} = \mathbf{r}e^{j\theta},$$

con \mathbf{r} y θ definidas como en 5.43, se concluye que la magnitud y la fase de una variable aleatoria gaussiana son variables aleatorias independientes, con distribuciones de probabilidad Rayleigh y uniforme respectivamente.

Variables auxiliares Se considera a continuación el caso de una función de dos variables aleatorias. Sea

$$\mathbf{z} = g(\mathbf{x}, \mathbf{y}),$$

con \mathbf{x} e \mathbf{y} variables aleatorias. Es posible determinar la densidad de probabilidad $f_z(z)$ de \mathbf{z} mediante la ecuación 5.34. Para esto, se define una nueva variable aleatoria

$$\mathbf{w} = \mathbf{x}, \quad \text{o} \quad \mathbf{w} = \mathbf{y},$$

se determina $f_{zw}(z, w)$ y se integra.

Ejemplo 6-24 Como ejemplo, considérese la variable aleatoria $\mathbf{z} = \mathbf{x} + \mathbf{y}$, y se quiere determinar $f_z(z)$. Para hacerlo, se define la variable aleatoria $\mathbf{w} = \mathbf{y}$. El sistema de ecuaciones

$$\begin{cases} z = g(x, y) = x + y \\ w = h(x, y) = y \end{cases}$$

tiene como solución única

$$y_1 = w, \quad x_1 = z - w.$$

El determinante de la matriz jacobiana de la transformación es

$$J(x, y) = \begin{vmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1.$$

Sustituyendo estos resultados en la ecuación 5.34 se obtiene que

$$f_{zw}(z, w) = f_{xy}(x_1, y_1) = f_{xy}(z - w, w).$$

y la densidad de probabilidad de \mathbf{z} es

$$f_z(z) = \int_{-\infty}^{\infty} f_{zw}(z, w) dw = \int_{-\infty}^{\infty} f_{xy}(z - w, w) dw,$$

coincidiendo con el resultado de la ecuación 5.17.

Ejemplo 6-28: la distribución t Student Una variable aleatoria \mathbf{z} tiene distribución t Student $t(n)$ con n grados de libertad si para $-\infty < z < \infty$

$$f_z(z) = \frac{\gamma_1}{\sqrt{\left(1 + \frac{z^2}{n}\right)^{n+1}}}, \quad \text{con} \quad \gamma_1 = \frac{\Gamma((n+1)/2)}{\sqrt{\pi n} \Gamma(n/2)}. \quad (5.45)$$

Se demostrará que si \mathbf{x} e \mathbf{y} son dos variables aleatorias independientes, con $\mathbf{x} \sim N(0, 1)$ e $\mathbf{y} \sim \chi^2(n)$, la variable aleatoria

$$\mathbf{z} = \frac{\mathbf{x}}{\sqrt{\mathbf{y}/n}}$$

tiene distribución $t(n)$. Observar que la distribución t Student representa el cociente de una variable aleatoria normal y la raíz cuadrada de una variable aleatoria independiente χ^2 dividida entre sus grados de libertad. Para demostrarlo, se introduce la variable auxiliar $\mathbf{w} = \mathbf{y}$ y se emplea la ecuación 5.34, donde la transformación es

$$\mathbf{z} = \frac{\mathbf{x}}{\sqrt{\mathbf{y}/n}}, \quad \mathbf{w} = \mathbf{y}.$$

Fijando (z, w) , el sistema de ecuaciones

$$\frac{x}{\sqrt{y/n}} = z, \quad y = w$$

tiene una solución única (x_i, y_i) , y la transformación inversa es

$$x_i = g_1(z, w) = z\sqrt{\frac{w}{n}}, \quad y_i = h_1(z, w) = w.$$

Considerando que

$$\frac{\partial g_1}{\partial z} = \sqrt{\frac{w}{n}}, \quad \frac{\partial g_1}{\partial w} = \frac{z}{2\sqrt{nw}}, \quad \frac{\partial h_1}{\partial z} = 0, \quad \text{y} \quad \frac{\partial h_1}{\partial w} = 1,$$

la matriz Jacobiana de la transformación inversa es

$$J(z, w) = \begin{vmatrix} \frac{\partial g_1}{\partial z} & \frac{\partial g_1}{\partial w} \\ \frac{\partial h_1}{\partial z} & \frac{\partial h_1}{\partial w} \end{vmatrix} = \begin{vmatrix} \sqrt{\frac{w}{n}} & \frac{z}{2\sqrt{nw}} \\ 0 & 1 \end{vmatrix},$$

y el valor absoluto del determinante es

$$|J(z, w)| = \sqrt{\frac{w}{n}}.$$

Además, como \mathbf{x} e \mathbf{y} son independientes, se cumple que

$$f_{xy}(x, y) = f_x(x)f_y(y) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \frac{y^{n/2-1}}{2^{n/2}\Gamma(n/2)}e^{-y/2}U(y),$$

donde en la última igualdad se emplearon la ecuaciones 3.5 y 3.18 de las densidades de probabilidad de \mathbf{x} e \mathbf{y} respectivamente. Sustituyendo estos resultados en la ecuación 5.34, se tiene que

$$\begin{aligned} f_{zw}(z, w) &= |J(z, w)|f_{xy}(x_i, y_i) \\ &= \sqrt{\frac{w}{n}} \frac{1}{\sqrt{2\pi}}e^{-x_i^2/2} \frac{y_i^{n/2-1}}{2^{n/2}\Gamma(n/2)}e^{-y_i/2}U(y_i) \\ &= \sqrt{\frac{w}{n}} \frac{1}{\sqrt{2\pi}}e^{-z^2w/2n} \frac{w^{n/2-1}}{2^{n/2}\Gamma(n/2)}e^{-w/2}U(w) \\ &= \frac{w^{(n-1)/2}}{\sqrt{2\pi n}2^{n/2}\Gamma(n/2)}e^{-w(1+z^2/n)/2}U(w), \end{aligned}$$

e integrando respecto a w se obtiene que

$$\begin{aligned} f_z(z) &= \int_{-\infty}^{\infty} f_{zw}(z, w) dw \\ &= \frac{1}{\sqrt{2\pi n}2^{n/2}\Gamma(n/2)} \int_0^{\infty} w^{(n-1)/2}e^{-w(1+z^2/n)/2} dw \\ &\stackrel{(a)}{=} \frac{1}{\sqrt{2\pi n}2^{n/2}\Gamma(n/2)} \int_0^{\infty} \frac{2^{(n-1)/2}u^{(n-1)/2}}{(1+z^2/n)^{(n-1)/2}}e^{-u} \frac{du}{(1+z^2/n)/2} \\ &= \frac{1}{\sqrt{\pi n}\Gamma(n/2)(1+z^2/n)^{(n-1)/2}} \int_0^{\infty} u^{(n-1)/2}e^{-u} du \\ &\stackrel{(b)}{=} \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\Gamma(n/2)(1+z^2/n)^{(n-1)/2}}, \end{aligned}$$

donde en (a) se realizó el cambio de variable

$$u = w(1 + z^2/n)/2 \quad \Rightarrow \quad du = dw(1 + z^2/n)/2 \quad \text{y} \quad w = 2u/(1 + z^2/n),$$

y en (b) se notó que la integral es $\Gamma((n+1)/2)$, como indica la ecuación 3.13. Esto concluye la prueba. Observando además que

$$\frac{\sqrt{\pi} \Gamma(n/2)}{\Gamma((n+1)/2)} \stackrel{(a)}{=} \frac{\Gamma(1/2)\Gamma(n/2)}{\Gamma((n/2+1)/2)} \stackrel{(b)}{=} B(1/2, n/2),$$

donde en (a) se consideró que $\Gamma(n/2) = \sqrt{\pi}$, como indica la ecuación 3.16, y en (b) la definición de la función beta, dada por la ecuación 3.21. Mediante este resultado, la distribución de t Student se puede expresar en términos de la función beta como

$$f_z(z) = \frac{1}{\sqrt{n} B(1/2, n/2)(1 + z^2/n)^{(n-1)/2}}. \quad (5.46)$$

Cuando n crece, la distribución t tiene a la distribución normal. De hecho, en la ecuación 5.46 se cumple que

$$\left(1 + \frac{z^2}{n}\right)^{-(n+1)/2} \rightarrow e^{-z^2/2} \quad \text{con} \quad n \rightarrow \infty.$$

Esto se debe a que, tomando logaritmo

$$-\frac{n+1}{2} \ln \left(1 + \frac{z^2}{n}\right) = -\frac{\ln \left(1 + \frac{z^2}{n}\right)}{\frac{2}{n+1}} \rightarrow -\frac{\frac{z^2}{n}}{\frac{2}{n+1}} = -\frac{z^2}{2} \quad \text{con} \quad n \rightarrow \infty.$$

Para n pequeño, la distribución t tiene colas mas anchas que la distribución normal debido a su forma polinómica.

5.4. Momentos conjuntos

Dadas dos variables aleatorias \mathbf{x} e \mathbf{y} y una función $g(x, y)$, se construye una nueva variable aleatoria $\mathbf{z} = g(\mathbf{x}, \mathbf{y})$. Por definición, el valor esperado de \mathbf{z} es

$$E\{\mathbf{z}\} = \int_{-\infty}^{\infty} z f_z(z) dz.$$

De forma análoga al caso de una función de una variable aleatoria (ver ecuación 4.7), se puede demostrar que la media de \mathbf{z} se puede expresar directamente en términos de la función $g(x, y)$ y la densidad de probabilidad conjunta $f(x, y)$ como

$$E\{g(\mathbf{x}, \mathbf{y})\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy \quad (5.47)$$

Si las variables aleatorias \mathbf{x} e \mathbf{y} son discretas y toman los valores x_i y y_k con probabilidad p_{ik} , la esperanza de \mathbf{z} se calcula como

$$E\{g(\mathbf{x}, \mathbf{y})\} = \sum_i \sum_k g(x_i, y_k) p_{ik}$$

Linealidad de la esperanza A partir de la ecuación 5.47, teniendo en cuenta que la operación de integración es lineal, se deduce que

$$E \left\{ \sum_{k=1}^n a_k g_k(\mathbf{x}, \mathbf{y}) \right\} = \sum_{k=1}^n a_k E\{g_k(\mathbf{x}, \mathbf{y})\} \quad (5.48)$$

En particular, se cumple que

$$E\{\mathbf{x} + \mathbf{y}\} = E\{\mathbf{x}\} + E\{\mathbf{y}\}.$$

5.4.1. Covarianza

En el caso de una variable aleatoria, se definieron los parámetros media y varianza para describir el comportamiento promedio. El concepto de varianza se puede generalizar para el caso de dos variables aleatorias para medir el comportamiento entre ellas.

La covarianza C o C_{xy} de dos variables aleatorias \mathbf{x} e \mathbf{y} se define como el número

$$C_{xy} = E\{(\mathbf{x} - \eta_x)(\mathbf{y} - \eta_y)\}, \quad (5.49)$$

donde $\eta_x = E\{\mathbf{x}\}$ y $\eta_y = E\{\mathbf{y}\}$. A partir de la propiedad de linealidad de la esperanza (ecuación 5.48), se cumple que

$$C_{xy} = E\{\mathbf{x}\mathbf{y}\} - E\{\mathbf{x}\}E\{\mathbf{y}\}.$$

Notar que se cumple que

$$C_{xx} = E\{\mathbf{x}^2\} - E^2\{\mathbf{x}\} = \sigma_x^2.$$

Coeficiente de correlación El coeficiente de correlación ρ o ρ_{xy} de las variables aleatorias \mathbf{x} e \mathbf{y} se define como el cociente

$$\rho_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}. \quad (5.50)$$

Se demostrará que

$$|\rho_{xy}| \leq 1, \quad |C_{xy}| \leq \sigma_x \sigma_y.$$

Para hacerlo, considerar que

$$E\{[a(\mathbf{x} - \eta_x) + (\mathbf{y} - \eta_y)]^2\} \geq 0, \quad \forall a,$$

ya que la expresión dentro de la esperanza es no negativa. Operando, se ve que

$$\begin{aligned} E\{[a(\mathbf{x} - \eta_x) + (\mathbf{y} - \eta_y)]^2\} &= E\{a^2(\mathbf{x} - \eta_x)^2 + 2a(\mathbf{x} - \eta_x)(\mathbf{y} - \eta_y) + (\mathbf{y} - \eta_y)^2\} \\ &= a^2 E\{(\mathbf{x} - \eta_x)^2\} + 2a E\{(\mathbf{x} - \eta_x)(\mathbf{y} - \eta_y)\} \\ &\quad + E\{(\mathbf{y} - \eta_y)^2\} \\ &= a^2 \sigma_x^2 + 2a C_{xy} + \sigma_y^2 \geq 0, \end{aligned} \quad (5.51)$$

donde en la segunda igualdad se empleó la linealidad de la esperanza y en la última igualdad se aplicaron las definiciones de varianza y covarianza. Como el polinomio en a de la última igualdad es nonnegativo, no puede tener raíces, por lo que el discriminante debe ser negativo,

$$\Delta = 4C_{xy}^2 - 4\sigma_x^2\sigma_y^2 \leq 0,$$

es decir,

$$C_{xy}^2 \leq \sigma_x^2\sigma_y^2.$$

Como se mencionó, la covarianza es una medida de la variabilidad conjunta de las variables aleatorias. Si a valores grandes de una de ellas corresponden valores grandes de la otra, y a valores pequeños de una corresponden valores pequeños de la otra, es decir, ambas variables aleatorias tienden a comportarse de forma similar, su covarianza es positiva. Por el contrario, si a valores grandes de una corresponden valores pequeños de la otra y viceversa, la covarianza es negativa. Por lo tanto, el signo de la covarianza indica la tendencia en la relación lineal entre las variables aleatorias. La magnitud de la covarianza es difícil de interpretar porque al no ser una magnitud normalizada, depende de las magnitudes de las variables aleatorias. El coeficiente de correlación, al tratarse de una versión normalizada de la covarianza, indica mediante su magnitud cuan estrecho es el grado de relación lineal entre ellas⁴.

Observar que las variables aleatorias \mathbf{x} , \mathbf{y} y las variables aleatorias $\mathbf{x} - a$, $\mathbf{y} - b$, con a y b números arbitrarios, tienen la misma covarianza y coeficiente de correlación.

⁴Weisstein, Eric W. "Covariance." From MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/Covariance.html>

Variables aleatorias conjuntamente normales La densidad de probabilidad de dos variables aleatorias conjuntamente normales está dada por las ecuaciones 5.11 y 5.12. Se demostrará que su coeficiente de correlación es el parámetro r de la densidad de probabilidad.

Como el coeficiente de correlación no depende de la media de las variables aleatorias, se considerará el caso en que $\eta_x = \eta_y = 0$. Por definición del coeficiente de correlación (ecuación 5.50)

$$C_{xy} = \rho_{xy}\sigma_1\sigma_2$$

y como $\eta_x = \eta_y = 0$, de la ecuación 5.49 se ve que $C_{xy} = E\{\mathbf{xy}\}$. Por lo tanto, como se quiere demostrar que $\rho_{xy} = r$, alcanza con demostrar que

$$E\{\mathbf{xy}\} = r\sigma_1\sigma_2.$$

A partir de la ecuación 5.47, se cumple que

$$E\{\mathbf{xy}\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{xy}(x, y) dx dy,$$

donde la densidad de probabilidad conjunta $f_{xy}(x, y)$ está dada por las ecuaciones 5.11 y 5.12. Como se explicó en la sección 5.1.6, el exponente de la densidad conjunta, teniendo en cuenta que las medias de \mathbf{x} e \mathbf{y} son nulas, se puede escribir como

$$\begin{aligned} -\frac{1}{2(1-r^2)} \left[\frac{x^2}{\sigma_1^2} - 2r \frac{xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2} \right] &= -\frac{1}{2(1-r^2)} \left[\left(\frac{x}{\sigma_1} - r \frac{y}{\sigma_2} \right)^2 + (1-r^2) \frac{y^2}{\sigma_2^2} \right] \\ &= -\frac{1}{2(1-r^2)} \left[\frac{1}{\sigma_1^2} \left(x - ry \frac{\sigma_1}{\sigma_2} \right)^2 + (1-r^2) \frac{y^2}{\sigma_2^2} \right] \\ &= -\frac{(x - ry\sigma_1/\sigma_2)^2}{2\sigma_1^2(1-r^2)} - \frac{y^2}{2\sigma_2^2}, \end{aligned}$$

por lo tanto

$$\begin{aligned} E\{\mathbf{xy}\} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp \left[-\frac{(x - ry\sigma_1/\sigma_2)^2}{2\sigma_1^2(1-r^2)} - \frac{y^2}{2\sigma_2^2} \right] dx dy \\ &= \frac{1}{\sigma_2\sqrt{2\pi}} \int_{-\infty}^{\infty} ye^{-y^2/2\sigma_2^2} \int_{-\infty}^{\infty} \frac{x}{\sigma_1\sqrt{2\pi(1-r^2)}} \exp \left[-\frac{(x - ry\sigma_1/\sigma_2)^2}{2\sigma_1^2(1-r^2)} \right] dx dy. \end{aligned}$$

La integral interna es la media de una variable aleatoria de distribución normal de media $ry\sigma_1/\sigma_2$ y varianza $\sigma_1^2(1-r^2)$, y por lo tanto vale $ry\sigma_1/\sigma_2$. Esto lleva a que

$$E\{\mathbf{xy}\} = r \frac{\sigma_1}{\sigma_2} \int_{-\infty}^{\infty} \frac{1}{\sigma_2\sqrt{2\pi}} y^2 e^{-y^2/2\sigma_2^2} dy.$$

La integral es la varianza de una variable aleatoria de distribución normal de media nula y varianza σ_2^2 , por lo que el resultado es σ_2^2 . Se concluye que

$$E\{\mathbf{xy}\} = r\sigma_1\sigma_2,$$

que era lo que se quería demostrar.

5.4.2. No correlación y ortogonalidad

No correlación Se dice que dos variables aleatorias son no correlacionadas si su covarianza es 0. Matemáticamente, esto se puede expresar de las siguientes tres formas equivalentes

$$C_{xy} = 0, \quad \rho_{xy} = 0, \quad E\{\mathbf{xy}\} = E\{\mathbf{x}\}E\{\mathbf{y}\},$$

donde la última igualdad proviene de que $C_{xy} = E\{\mathbf{xy}\} - E\{\mathbf{x}\}E\{\mathbf{y}\} = 0$.

Ortogonalidad Se dice que dos variables aleatorias son ortogonales si cumplen que

$$E\{\mathbf{x}\mathbf{y}\} = 0,$$

y para indicarlo se usa la siguiente notación:

$$\mathbf{x} \perp \mathbf{y}.$$

Notar que

(a) Si \mathbf{x} e \mathbf{y} son no correlacionadas, entonces $\mathbf{x} - \eta_x \perp \mathbf{y} - \eta_y$. Efectivamente,

$$C_{xy} = E\{(\mathbf{x} - \eta_x)(\mathbf{y} - \eta_y)\} \stackrel{(a)}{=} 0,$$

donde en (a) se usó la hipótesis de que \mathbf{x} e \mathbf{y} son no correlacionadas y por lo tanto $C_{xy} = 0$.

(b) Si \mathbf{x} e \mathbf{y} son no correlacionadas y $\eta_x = 0$ o $\eta_y = 0$, se cumple que $\mathbf{x} \perp \mathbf{y}$. Para ver esto, por hipótesis se cumple que

$$E\{\mathbf{x}\mathbf{y}\} = E\{\mathbf{x}\}E\{\mathbf{y}\} = \eta_x\eta_y = 0$$

si $\eta_x = 0$ o $\eta_y = 0$.

Teorema: Independencia y no correlación Si dos variables aleatorias son independientes, entonces son no correlacionadas.

Por definición de variables aleatorias independientes, se cumple que (ecuación 5.10)

$$f(x, y) = f_x(x)f_y(y).$$

Alcanza con demostrar que

$$E\{\mathbf{x}\mathbf{y}\} = E\{\mathbf{x}\}E\{\mathbf{y}\}.$$

Para esto, se ve que

$$\begin{aligned} E\{\mathbf{x}\mathbf{y}\} &\stackrel{(a)}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y) dx dy \\ &\stackrel{(b)}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf_x(x)f_y(y) dx dy \\ &= \int_{-\infty}^{\infty} xf_x(x) dx \int_{-\infty}^{\infty} yf_y(y) dy \\ &= E\{\mathbf{x}\}E\{\mathbf{y}\}, \end{aligned}$$

donde en (a) se usó la ecuación 5.47 y en (b) la hipótesis.

Notar que

- si las variables aleatorias \mathbf{x} e \mathbf{y} son independientes, las variables aleatorias $g(\mathbf{x})$ y $h(\mathbf{y})$ también son independientes, y por lo tanto cumplen que

$$E\{g(\mathbf{x})h(\mathbf{y})\} = E\{g(\mathbf{x})\}E\{h(\mathbf{y})\}. \quad (5.52)$$

La demostración es similar a la realizada arriba. Esto no es cierto en general si \mathbf{x} e \mathbf{y} son solamente no correlacionadas.

- Si dos variables aleatorias son no correlacionadas, no necesariamente son independientes. Sin embargo, para variables aleatorias normales, no correlación es equivalente a independencia. Efectivamente, si las variables aleatorias \mathbf{x} e \mathbf{y} son conjuntamente normales y su coeficiente de correlación es $r = 0$, se cumple que $f_{xy}(x, y) = f_x(x)f_y(y)$ (ver la ecuación 5.11).

Ejemplo Se considera a continuación un ejemplo de dos variables aleatorias que no son independientes pero son no correlacionadas. Sean $\mathbf{x} \sim U(0, 1)$, $\mathbf{y} \sim U(0, 1)$ independientes. Se definen las variables aleatorias $\mathbf{z} = \mathbf{x} + \mathbf{y}$ y $\mathbf{w} = \mathbf{x} - \mathbf{y}$, y se quiere evaluar su correlación y dependencia.

Se comenzará determinando $f_{zw}(z, w)$, y luego $f_z(z)$ y $f_w(w)$. $f_{zw}(z, w)$ se puede calcular usando la ecuación 5.34, que indica que

$$f_{zw}(z, w) = \sum_i |J(z, w)| f_{xy}(x_i, y_i)$$

Es fácil ver que el sistema de ecuaciones

$$\begin{cases} z = x + y \\ w = x - y \end{cases}$$

tiene como única solución

$$x = g_1(z, w) = \frac{z + w}{2}, \quad y = h_1(z, w) = \frac{z - w}{2}.$$

Además,

$$J(z, w) = \begin{vmatrix} \frac{\partial g_1}{\partial z} & \frac{\partial g_1}{\partial w} \\ \frac{\partial h_1}{\partial z} & \frac{\partial h_1}{\partial w} \end{vmatrix} = \begin{vmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{vmatrix} = -\frac{1}{4} - \frac{1}{4} = -\frac{1}{2},$$

es decir, $|J(z, w)| = 1/2$. También se puede ver que como $0 < x < 1$ y $0 < y < 1$ se cumple que

$$\begin{aligned} z = x + y &\Rightarrow 0 < z < 2 \\ w = x - y &\Rightarrow -1 < w < 1 \\ (z + w) = 2x &\Rightarrow 0 < z + w < 2 \\ (z - w) = 2y &\Rightarrow 0 < z - w < 2 \end{aligned}$$

Estas condiciones determinan la región del plano (z, w) donde $f_{zw}(z, w)$ no es nula, y se muestra en la figura 5.15. Finalmente, se tiene que

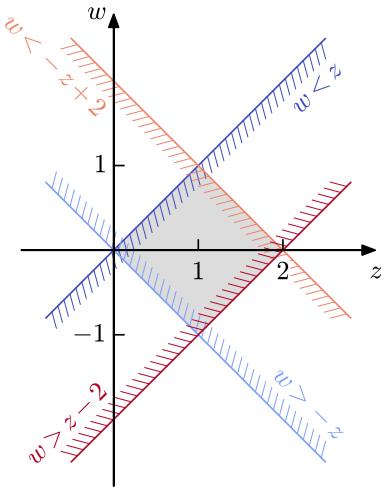


Figura 5.15: Región del plano (z, w) donde $f_{zw}(z, w)$ no es nula. Consiste en la intersección de los semiplanos determinados por las condiciones $0 < z + w < 2$ y $0 < z - w < 2$.

$$f_{xy}(x, y) = f_x(x)f_y(y) = 1, \quad 0 < x < 1, 0 < y < 1.$$

Sustituyendo estos resultados en la ecuación 5.34, se concluye que

$$f_{zw}(z, w) = \begin{cases} 1/2, & 0 < z + w < 2, 0 < z - w < 2 \\ 0, & \text{en otro caso.} \end{cases}$$

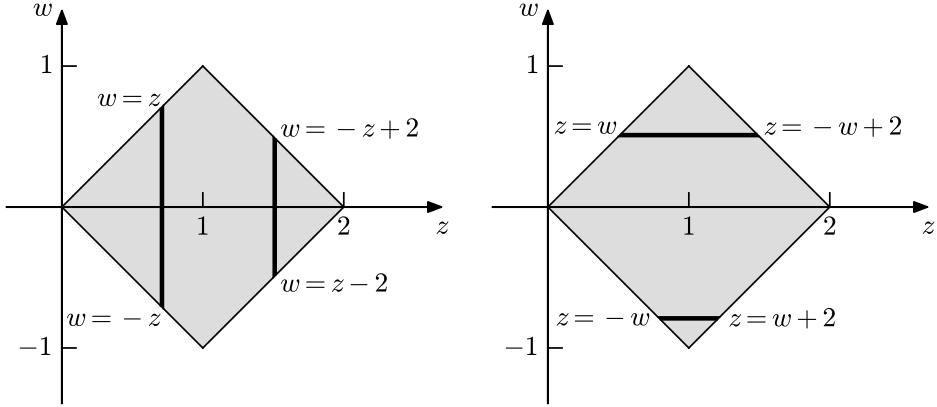


Figura 5.16: Límites de integración de $f_{zw}(z, w)$ para el cálculo de las densidades marginales. Se observa que los límites de integración dependen del valor de la variable de la que se calcula el marginal. A la izquierda se muestran los intervalos de integración para calcular la densidad marginal de z . Se distinguen dos casos: si $0 < z < 1$, los límites de integración son $-z < w < z$ y si $1 < z < 2$, los límites de integración son $z - 2 < w < -z + 2$. A la derecha, se muestran los intervalos de integración para calcular el marginal de w . Si $0 < w < 1$, los límites de integración son $w < z < -w + 2$, y si $-1 < w < 0$, los límites de integración son $-w < z < w + 2$. En este caso, esto se puede resumir como $|w| < z < 2 - |w|$ para $-1 < w < 1$.

La densidad marginal $f_z(z)$ se obtiene integrando $f_{zw}(z, w)$ respecto a w . Para esto, hay que tener en cuenta que los límites de integración dependen del valor de z , como se muestra en la figura 5.16. De esta forma,

$$f_z(z) = \int f_{zw}(z, w) dw = \begin{cases} \int_{-z}^z \frac{1}{2} dw = z, & 0 < z < 1 \\ \int_{z-2}^{-z+2} \frac{1}{2} dw = -z + 2, & 1 < z < 2 \\ 0 & \text{en otro caso.} \end{cases}, \quad (5.53)$$

y la densidad marginal de w es

$$f_w(w) = \int f_{zw}(z, w) dz = \int_{|w|}^{2-|w|} \frac{1}{2} dz = \frac{z}{2} \Big|_{|w|}^{2-|w|} = \frac{2-|w|}{2} - \frac{|w|}{2} = 1 - |w|,$$

si $-1 < w < 1$. Es decir

$$f_w(w) = \int f_{zw}(z, w) dz = \begin{cases} 1 - |w|, & -1 < w < 1 \\ 0 & \text{en otro caso.} \end{cases}. \quad (5.54)$$

De las ecuaciones 5.53 y 5.54 se deduce que

$$f_z(z) f_w(w) = \begin{cases} z(1 - |w|), & 0 < z < 1, -1 < w < 1 \\ (2-z)(1 - |w|), & 1 < z < 2, -1 < w < 1 \\ 0 & \text{en otro caso.} \end{cases},$$

y comparando este resultado con $f_{zw}(z, w)$, se ve que $f_{zw}(z, w) \neq f_z(z) f_w(w)$, lo que indica que las variables aleatorias \mathbf{z} y \mathbf{w} no son independientes.

Se calculará ahora la covarianza entre \mathbf{z} y \mathbf{w} para estudiar la correlación. Por un lado, se tiene que

$$\begin{aligned} E\{\mathbf{z}\mathbf{w}\} &= E\{(\mathbf{x} + \mathbf{y})(\mathbf{x} - \mathbf{y})\} \\ &= E\{\mathbf{x}^2 - \mathbf{x}\mathbf{y} + \mathbf{y}\mathbf{x} - \mathbf{y}^2\} \\ &= E\{\mathbf{x}^2 - \mathbf{y}^2\} \\ &\stackrel{(a)}{=} E\{\mathbf{x}^2\} - E\{\mathbf{y}^2\} \\ &\stackrel{(b)}{=} 0, \end{aligned}$$

donde en (a) se empleó la propiedad de linealidad de la esperanza y en (b) se usó que $E\{\mathbf{x}^2\} = E\{\mathbf{y}^2\}$ debido a que \mathbf{x} e \mathbf{y} tienen igual densidad de probabilidad. Esto indica que las variables aleatorias \mathbf{z} y \mathbf{w} son ortogonales. Por otro lado, se ve que

$$E\{\mathbf{w}\} = E\{\mathbf{x} - \mathbf{y}\} = E\{\mathbf{x}\} - E\{\mathbf{y}\} = 0.$$

A partir de estos dos resultados, se llega a que

$$C_{zw} = E\{\mathbf{z}\mathbf{w}\} - E\{\mathbf{z}\}E\{\mathbf{w}\} = 0,$$

implicando que las variables aleatorias \mathbf{z} y \mathbf{w} son no correlacionadas. Se concluye que \mathbf{z} y \mathbf{w} son variables aleatorias que no son independientes pero son no correlacionadas.

Las densidades $f_z(z)$ y $f_w(w)$ de las ecuaciones 5.53 y 5.54 se obtuvieron como los marginales de $f_{zw}(z, w)$. Como $\mathbf{z} = \mathbf{x} + \mathbf{y}$, $f_z(z)$ se podría haber calculado como en la sección 5.2.1, donde se concluyó que la densidad de la suma de dos variables aleatorias independientes es la convolución de las densidades. Así,

$$f_z(z) = \int_{-\infty}^{\infty} f_x(z - y) f_y(y) dy.$$

Como

$$f_x(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{en otro caso.} \end{cases},$$

se tiene que

$$f_x(z - y) = \begin{cases} 1, & 0 < z - y < 1 \\ 0, & \text{en otro caso.} \end{cases} = \begin{cases} 1, & z - 1 < y < z \\ 0, & \text{en otro caso.} \end{cases}.$$

Por lo tanto,

$$f_x(z - y) f_y(y) = \begin{cases} 0 & \text{en } -\infty < y < \infty \text{ si } z < 0 \text{ o } z > 2 \\ 1 & \text{en } 0 < y < z \text{ si } 0 < z < 1 \\ 1 & \text{en } z - 1 < y < 1 \text{ si } 1 < z < 2 \end{cases}, \quad (5.55)$$

como se muestra en la figura 5.17, e integrando se obtiene que

$$f_z(z) = \int_{-\infty}^{\infty} f_x(z - y) f_y(y) dy = \begin{cases} \int_0^z 1 dy = z, & 0 < z < 1 \\ \int_{z-1}^1 1 dy = 2 - z, & 1 < z < 2 \\ 0 & \text{en otro caso.} \end{cases},$$

resultado que coincide con la ecuación 5.53. Para calcular la densidad $f_w(w)$, procediendo como en la sección 5.2.1, se puede demostrar que si $\mathbf{w} = \mathbf{x} - \mathbf{y}$,

$$f_w(w) = \int_{-\infty}^{\infty} f_{xy}(y + w, y) dy = \int_{-\infty}^{\infty} f_{xy}(x, x - w) dx,$$

y si \mathbf{x} e \mathbf{y} son independientes, queda

$$f_w(w) = \int_{-\infty}^{\infty} f_x(y + w) f_y(y) dy = \int_{-\infty}^{\infty} f_x(x) f_y(x - w) dx, \quad (5.56)$$

que es la convolución entre las funciones $f_x(\tau)$ y $f_y(-\tau)$. Observando el que

$$f_y(x-w) = \begin{cases} 1, & 0 < x-w < 1 \\ 0, & \text{en otro caso.} \end{cases} = \begin{cases} 1, & w < x < 1+w \\ 0, & \text{en otro caso.} \end{cases},$$

el integrando de $f_w(w)$ es

$$f_x(x)f_y(x-w) = \begin{cases} 0 & \text{en } -\infty < x < \infty \text{ si } w < -1 \text{ o } w > 1 \\ 1 & \text{en } 0 < x < w+1 \text{ si } -1 < w < 0 \\ 1 & \text{en } w < x < 1 \text{ si } 0 < w < 1 \end{cases}, \quad (5.57)$$

como se muestra en la figura 5.17, por lo que

$$f_w(w) = \int_{-\infty}^{\infty} f_x(x)f_y(x-w) dx = \begin{cases} \int_0^{w+1} 1 dx = w+1, & -1 < w < 0 \\ \int_w^1 1 dx = 1-w, & 0 < w < 1 \\ 0 & \text{en otro caso} \end{cases},$$

coincidiendo con la ecuación 5.54. Las distribuciones marginales $f_z(z)$ y $f_w(w)$ se muestran en la figura 5.17.

Varianza de la suma de dos variables aleatorias Sea $\mathbf{z} = \mathbf{x} + \mathbf{y}$. Por la propiedad de linealidad de la esperanza, se cumple que $\eta_z = \eta_x + \eta_y$. Por lo tanto, la varianza de \mathbf{z} es

$$\begin{aligned} \sigma_z^2 &= E\{(\mathbf{z} - \eta_z)^2\} \\ &= E\{[(\mathbf{x} + \mathbf{y}) - (\eta_x + \eta_y)]^2\} \\ &\stackrel{(a)}{=} \sigma_x^2 + 2C_{xy} + \sigma_y^2 \\ &\stackrel{(b)}{=} \sigma_x^2 + 2\rho_{xy}\sigma_x\sigma_y + \sigma_y^2, \end{aligned} \quad (5.58)$$

donde en (a) se empleó la igualdad de la ecuación 5.51 con $a = 1$ y en (b) la definición del coeficiente de correlación, dado por la ecuación 5.50. Este resultado indica que si $\rho_{xy} = 0$, se cumple que

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2.$$

Se concluye que si dos variables aleatorias son no correlacionadas, la varianza de su suma es la suma de las varianzas. Como dos variables aleatorias independientes son no correlacionadas, esto también es cierto para el caso de variables aleatorias independientes.

Momentos El valor

$$m_{kr} = E\{\mathbf{x}^k \mathbf{y}^r\} = \int_{-\infty}^{\infty} x^k y^r f_{xy}(x, y) dx dy$$

es el momento conjunto de orden $n = k+r$ de las variables aleatorias \mathbf{x} e \mathbf{y} . De esta forma, los momentos de primer orden son $m_{10} = \eta_x$ y $m_{01} = \eta_y$ y los momentos de segundo orden son

$$m_{20} = E\{\mathbf{x}^2\}, \quad m_{11} = E\{\mathbf{x}\mathbf{y}\}, \quad m_{02} = E\{\mathbf{y}^2\}.$$

Los momentos centrales son los momentos de $\mathbf{x} - \eta_x$ y $\mathbf{y} - \eta_y$,

$$\mu_{kr} = E\{(\mathbf{x} - \eta_x)^k (\mathbf{y} - \eta_y)^r\} = \int_{-\infty}^{\infty} (\mathbf{x} - \eta_x)^k (\mathbf{y} - \eta_y)^r f_{xy}(x, y) dx dy.$$

Así, $\mu_{10} = \mu_{01} = 0$ y

$$\mu_{11} = C_{xy}, \quad \mu_{20} = \sigma_x^2, \quad \mu_{02} = \sigma_y^2.$$

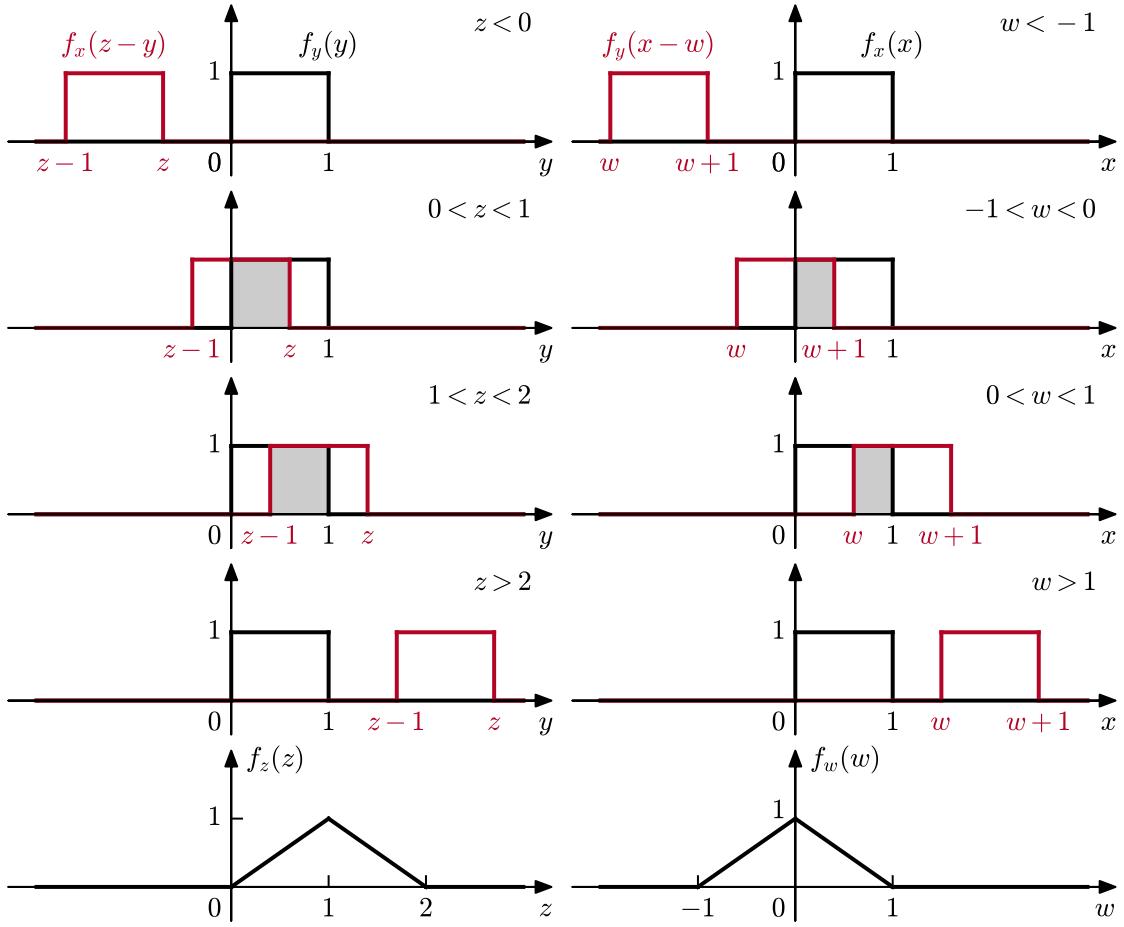


Figura 5.17: Deducción de las ecuaciones 5.55 y 5.57. A la izquierda se muestra la deducción de la densidad $f_z(z)$ de $\mathbf{z} = \mathbf{x} + \mathbf{y}$ a partir de la ecuación 5.18 para el caso en que $\mathbf{x} \sim U(0, 1)$ e $\mathbf{y} \sim U(0, 1)$. $f_z(z)$ es la convolución entre $f_x(x)$ y $f_y(y)$, es decir, la integral del producto de las funciones $f_y(y)$ y $f_x(z-y)$, las cuales se muestran para distintos valores de z y consiste en el área de la región sombreada. A la derecha se muestra la deducción de la densidad $f_w(w)$ de $\mathbf{w} = \mathbf{x} - \mathbf{y}$, que está dada por la ecuación 5.56.

5.5. Funciones características conjuntas

La función característica conjunta de las variables aleatorias \mathbf{x} e \mathbf{y} es por definición

$$\Phi(\omega_1, \omega_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{j(\omega_1 x + \omega_2 y)} dx dy. \quad (5.59)$$

Esto es la transformada de Fourier de dos dimensiones de $f(x, y)$. Aplicando la fórmula de inversión de la transformada de Fourier, se ve que se cumple que

$$f(x, y) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi(\omega_1, \omega_2) e^{-j(\omega_1 x + \omega_2 y)} d\omega_1 d\omega_2. \quad (5.60)$$

De esta forma,

$$\Phi(\omega_1, \omega_2) = E\{e^{j(\omega_1 \mathbf{x} + \omega_2 \mathbf{y})}\}.$$

La función característica conjunta logarítmica se define como

$$\Psi(\omega_1, \omega_2) = \ln \Phi(\omega_1, \omega_2).$$

Las funciones características marginales (ecuación 4.24)

$$\Phi_x(\omega) = E\{e^{j\omega x}\}, \quad \Phi_y(\omega) = E\{e^{j\omega y}\}$$

pueden expresarse en función de la función característica conjunta como

$$\Phi_x(\omega) = \Phi(\omega, 0), \quad \Phi_y(\omega) = \Phi(0, \omega),$$

ya que,

$$\begin{aligned} \Phi(\omega, 0) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{j\omega_1 x} dx dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x, y) dy \right) e^{j\omega_1 x} dx \\ &= \int_{-\infty}^{\infty} f_x(x) e^{j\omega_1 x} dx \\ &= \Phi_x(\omega). \end{aligned}$$

Notar que si $\mathbf{z} = a\mathbf{x} + b\mathbf{y}$,

$$\Phi_z(\omega) = E\{e^{j(\omega_1 x + \omega_2 y)\omega}\} = \Phi(a\omega, b\omega). \quad (5.61)$$

5.5.1. Independencia y convolución

Si las variables aleatorias \mathbf{x} e \mathbf{y} son independientes, se cumple que (ecuación 5.52)

$$E\{e^{j(\omega_1 x + \omega_2 y)}\} = E\{e^{j\omega_1 x}\} E\{e^{j\omega_2 y}\}$$

y por lo tanto, la función característica conjunta es

$$\Phi(\omega_1, \omega_2) = \Phi_x(\omega_1) \Phi_y(\omega_2). \quad (5.62)$$

Recíprocamente, si se cumple la ecuación 5.62, las variables aleatorias \mathbf{x} e \mathbf{y} son independientes. Para ver esto, se parte sustituyendo la ecuación 5.62 en la fórmula de inversión,

$$\begin{aligned} f(x, y) &= \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi_x(\omega_1) \Phi_y(\omega_2) e^{-j(\omega_1 x + \omega_2 y)} d\omega_1 d\omega_2 \\ &= \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_x(\omega_1) e^{-j\omega_1 x} d\omega_1 \right) \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_y(\omega_2) e^{-j\omega_2 y} d\omega_2 \right) \\ &= f_x(x) f_y(y), \end{aligned}$$

donde en la última igualdad se usó la fórmula de inversión de la función característica de una variable aleatoria (ecuación 4.32). Se concluye que como $f(x, y) = f_x(x) f_y(y)$, \mathbf{x} e \mathbf{y} son independientes.

Teorema de la convolución Si las variables aleatorias \mathbf{x} e \mathbf{y} son independientes y $\mathbf{z} = \mathbf{x} + \mathbf{y}$, se cumple que

$$E\{e^{j\omega z}\} = E\{e^{j\omega(x+y)}\} = E\{e^{j\omega x}\} E\{e^{j\omega y}\}$$

y por lo tanto

$$\Phi_z(\omega) = \Phi_x(\omega) \Phi_y(\omega), \quad \Psi_z(\omega) = \Psi_x(\omega) + \Psi_y(\omega) \quad (5.63)$$

Como se vio en la sección 5.2.1, la densidad $f_z(z)$ de \mathbf{z} es la convolución de las densidades $f_x(x)$ y $f_y(y)$. Se concluye que la función característica de la convolución de dos densidades es el producto de las funciones características.

Se demostrará nuevamente este resultado empleando propiedades de la transformada de Fourier. Sean $f_x(x)$ y $f_y(y)$ las densidades de probabilidad de las variables aleatorias \mathbf{x} e \mathbf{y} respectivamente, y se

considera una variable aleatoria \mathbf{z} con densidad de probabilidad la convolución entre $f_x(x)$ y $f_y(y)$. De esta forma, como indica la ecuación 5.18,

$$f_z(z) = \int_{-\infty}^{\infty} f_x(x) f_y(z-x) dx.$$

Se demostrará que la función característica de \mathbf{z} es $\Phi_z(\omega) = \Phi_x(\omega)\Phi_y(\omega)$. Partiendo de la definición de la función característica, se tiene que

$$\begin{aligned} \Phi_z(\omega) &= \int_{-\infty}^{\infty} f_z(z) e^{j\omega z} dz \\ &\stackrel{(a)}{=} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f_x(x) f_y(z-x) dx \right] e^{j\omega z} dz \\ &\stackrel{(b)}{=} \int_{-\infty}^{\infty} f_x(x) \left[\int_{-\infty}^{\infty} f_y(z-x) e^{j\omega z} dz \right] dx \\ &\stackrel{(c)}{=} \int_{-\infty}^{\infty} f_x(x) \left[\int_{-\infty}^{\infty} f_y(y) e^{j\omega(x+y)} dy \right] dx \\ &= \int_{-\infty}^{\infty} f_x(x) e^{j\omega x} dx \int_{-\infty}^{\infty} f_y(y) e^{j\omega y} dy \\ &= \Phi_x(\omega)\Phi_y(\omega), \end{aligned}$$

donde en (a) se sustituyó $f_z(z)$ por la convolución, en (b) se intercambió el orden de integración y en (c) se realizó el cambio de variable $z = x + y$. De forma análoga, puede demostrarse lo recíproco, esto es, si la función característica de \mathbf{z} es $\Phi_z(\omega) = \Phi_x(\omega)\Phi_y(\omega)$, la densidad de \mathbf{z} es la convolución entre las densidades de \mathbf{x} e \mathbf{y} . Para ver esto, se parte de la transformada inversa de la función característica y se opera de la siguiente forma,

$$\begin{aligned} f_z(z) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_z(\omega) e^{-j\omega z} d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_x(\omega)\Phi_y(\omega) e^{-j\omega z} d\omega \\ &\stackrel{(a)}{=} \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f_x(x) e^{j\omega x} dx \right] \Phi_y(\omega) e^{-j\omega z} d\omega \\ &\stackrel{(b)}{=} \int_{-\infty}^{\infty} f_x(x) \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_y(\omega) e^{-j\omega(z-x)} d\omega \right] dx \\ &\stackrel{(c)}{=} \int_{-\infty}^{\infty} f_x(x) f_y(z-x) dx, \end{aligned}$$

donde en (a) se sustituyó $\Phi_x(\omega)$ por su definición, en (b) se intercambió el orden de integración y en (c) se tuvo en cuenta que el término entre paréntesis rectos es la transformada inversa de la función característica $\Phi_y(\omega)$ evaluada en $z - x$, es decir, $f_y(z - x)$.

Ejemplo: suma de variables aleatorias de Poisson Se mostrará que si las variables aleatorias \mathbf{x} y \mathbf{y} son independientes y tienen distribución de Poisson con parámetro a y b respectivamente, su suma $\mathbf{z} = \mathbf{x} + \mathbf{y}$ también tiene distribución de Poisson con parámetro $a + b$.

De la ecuación 4.49 con $s = j\omega$ se deduce que las funciones características segundas son

$$\Psi_x(\omega) = a(e^{j\omega} - 1) \quad \text{y} \quad \Psi_y(\omega) = b(e^{j\omega} - 1),$$

por lo que de la ecuación 5.63 se concluye que

$$\Psi_z(\omega) = \Psi_x(\omega) + \Psi_y(\omega) = (a+b)(e^{j\omega} - 1), \quad (5.64)$$

que es la función característica segunda de una variable aleatoria de Poisson con parámetro $a + b$. Puede demostrarse que lo recíproco también es cierto: si las variables \mathbf{x} y \mathbf{y} son independientes y su suma es una variable aleatoria de Poisson, \mathbf{x} y \mathbf{y} son variables aleatorias de Poisson.

Ejemplo: variables aleatorias normales independientes Como se indicó previamente (ver la ecuación 5.42), si las variables aleatorias \mathbf{x} e \mathbf{y} son conjuntamente normales, cualquier combinación lineal $a\mathbf{x} + b\mathbf{y}$ es una variable aleatoria normal. Se demostrará el siguiente caso particular de este resultado: si las variables aleatorias \mathbf{x} e \mathbf{y} son normales e independientes, su suma $\mathbf{z} = \mathbf{x} + \mathbf{y}$ también es normal. Para hacerlo, se parte de las funciones características de \mathbf{x} e \mathbf{y} , que por ser normales son (ver la ecuación 4.29),

$$\Phi_x(\omega) = e^{j\eta_x\omega - \frac{1}{2}\sigma_x^2\omega^2}, \quad \Phi_y(\omega) = e^{j\eta_y\omega - \frac{1}{2}\sigma_y^2\omega^2}.$$

Por ser variables aleatorias independientes, por el teorema de la convolución dado por la ecuación 5.63, se cumple que

$$\Phi_z(\omega) = \Phi_x(\omega)\Phi_y(\omega) = \exp \left\{ j(\eta_x + \eta_y)\omega - \frac{1}{2}(\sigma_x^2 + \sigma_y^2)\omega^2 \right\},$$

que es la función característica de una variable aleatoria normal de media $\eta_x + \eta_y$ y varianza $\sigma_x^2 + \sigma_y^2$. Se concluye que $\mathbf{z} \sim N(\eta_x + \eta_y, \sigma_x^2 + \sigma_y^2)$.

Puede demostrarse que lo recíproco también es cierto (teorema de descomposición de Cramér⁵): si las variables aleatorias \mathbf{x} e \mathbf{y} son independientes y su suma es normal, entonces \mathbf{x} e \mathbf{y} son normales. Esto es equivalente a afirmar que la convolución de dos densidades es normal si y solo si ambas densidades son normales. También se puede demostrar que si \mathbf{x} e \mathbf{y} son variables aleatorias normales independientes, se cumple también que $\mathbf{x} + \mathbf{y}$ y $\mathbf{x} - \mathbf{y}$ también son independientes y normales. En este caso, el recíproco también es cierto (teorema de Bernstein⁶): si \mathbf{x} e \mathbf{y} son independientes e idénticamente distribuidas y si $\mathbf{x} + \mathbf{y}$ y $\mathbf{x} - \mathbf{y}$ también son independientes, necesariamente las variables aleatorias $(\mathbf{x}, \mathbf{y}, \mathbf{x} + \mathbf{y}, \mathbf{x} - \mathbf{y})$ son normales.

En el caso en que $\mathbf{z} = a\mathbf{x} + b\mathbf{y}$, se cumple que

$$\Phi_z(\omega) \stackrel{(a)}{=} \Phi_{xy}(a\omega, b\omega) \stackrel{(b)}{=} \Phi_x(a\omega)\Phi_y(b\omega),$$

donde en (a) se empleó el resultado de la ecuación 5.61 y en (b) se consideró que \mathbf{x} e \mathbf{y} son variables aleatorias independientes. De esta forma,

$$\Phi_z(\omega) = \exp \left\{ j(a\eta_x + b\eta_y)\omega - \frac{1}{2}(a^2\sigma_x^2 + b^2\sigma_y^2)\omega^2 \right\},$$

por lo que $\mathbf{z} \sim N(a\eta_x + b\eta_y, a^2\sigma_x^2 + b^2\sigma_y^2)$.

Variables aleatorias conjuntamente normales Sean \mathbf{x} e \mathbf{y} variables aleatorias conjuntamente normales con densidad $N(\eta_1, \eta_2, \sigma_1^2, \sigma_2^2, r)$. Su función característica conjunta está dada por

$$\Phi(\omega_1, \omega_2) = e^{j(\eta_1\omega_1 + \eta_2\omega_2)} e^{-(\omega_1^2\sigma_1^2 + 2r\sigma_1\sigma_2\omega_1\omega_2 + \omega_2^2\sigma_2^2)/2}. \quad (5.65)$$

Una forma de demostrarlo, es insertando la densidad de probabilidad conjunta, dada por las ecuaciones 5.11 y 5.12, en la definición de la función característica conjunta (ecuación 5.59) y calcular la integral. Esto se hará mas adelante. Otra forma mas simple, es la siguiente. Se considera la variable aleatoria $\mathbf{z} = \omega_1\mathbf{x} + \omega_2\mathbf{y}$, que por ser una combinación lineal de variables aleatorias normales, es normal, y por lo tanto, su función característica es (ecuación 4.29)

$$\Phi_z(\omega) = e^{j\eta_z\omega - \sigma_z^2\omega^2/2}.$$

La media de \mathbf{z} es

$$\eta_z = \omega_1\eta_1 + \omega_2\eta_2$$

⁵Ver https://en.wikipedia.org/wiki/Cramér's_theorem#Normal_random_variables

⁶Ver https://en.wikipedia.org/wiki/Normal_distribution#Bernstein's_theorem

y su varianza es

$$\begin{aligned}
\sigma_z^2 &= E\{[(\omega_1\mathbf{x} + \omega_2\mathbf{y}) - (\omega_1\eta_1 + \omega_2\eta_2)]^2\} \\
&= E\{[(\omega_1\mathbf{x} - \omega_1\eta_1) + (\omega_2\mathbf{y} - \omega_2\eta_2)]^2\} \\
&= E\{[\omega_1(\mathbf{x} - \eta_1) + \omega_2(\mathbf{y} - \eta_2)]^2\} \\
&= E\{\omega_1^2(\mathbf{x} - \eta_1)^2 + 2\omega_1\omega_2(\mathbf{x} - \eta_1)(\mathbf{y} - \eta_2) + \omega_2^2(\mathbf{y} - \eta_2)^2\} \\
&= \omega_1^2 E\{(\mathbf{x} - \eta_1)^2\} + 2\omega_1\omega_2 E\{(\mathbf{x} - \eta_1)(\mathbf{y} - \eta_2)\} + \omega_2^2 E\{(\mathbf{y} - \eta_2)^2\} \\
&= \omega_1^2\sigma_1^2 + 2\omega_1\omega_2 C_{xy} + \omega_2^2\sigma_2^2
\end{aligned}$$

concluyendo que

$$\sigma_z^2 = \omega_1^2\sigma_1^2 + 2r\omega_1\omega_2\sigma_1\sigma_2 + \omega_2^2\sigma_2^2.$$

Además, como indica la ecuación 5.61, se cumple que

$$\Phi_z(\omega) = \Phi(\omega_1\omega, \omega_2\omega),$$

por lo que con $\omega = 1$ se llega al resultado que se quería demostrar.

Se demostrará a continuación el mismo resultado resolviendo explícitamente la integral de la definición de la función característica. Si bien esta demostración es mas larga, tiene la ventaja de que no hay que emplear la hipótesis de que la combinación lineal de variables aleatorias normales es normal. La demostración aquí presentada está basada en [11]. Se parte sustituyendo la densidad de probabilidad conjunta $N(\eta_1, \eta_2, \sigma_1^2, \sigma_2^2, r)$, dada por las ecuaciones 5.11 y 5.12, en la función característica, dada por la ecuación 5.59,

$$\begin{aligned}
\Phi(\omega_1, \omega_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \exp\{j(\omega_1 x + \omega_2 y)\} dx dy \\
&= A \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\{j(\omega_1 x + \omega_2 y)\} \exp\left\{-\frac{1}{2(1-r^2)} \times \right. \\
&\quad \left. \left[\frac{(x - \eta_1)^2}{\sigma_1^2} - 2r \frac{(x - \eta_1)(y - \eta_2)}{\sigma_1\sigma_2} + \frac{(y - \eta_2)^2}{\sigma_2^2} \right] \right\} dx dy,
\end{aligned}$$

donde

$$A = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}.$$

Realizando el cambio de variable $u = x - \eta_1$ y $w = y - \eta_2$ se obtiene que

$$\begin{aligned}
\Phi(\omega_1, \omega_2) &= A \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\{j[\omega_1(u + \eta_1) + \omega_2(w + \eta_2)]\} \times \\
&\quad \exp\left\{-\frac{1}{2(1-r^2)} \left[\frac{u^2}{\sigma_1^2} - 2r \frac{uw}{\sigma_1\sigma_2} + \frac{w^2}{\sigma_2^2} \right] \right\} du dw \\
&= A \exp\{j(\omega_1\eta_1 + \omega_2\eta_2)\} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\{j(\omega_1 u + \omega_2 w)\} \times \\
&\quad \exp\left\{-\frac{1}{2(1-r^2)} \left[\frac{u^2}{\sigma_1^2} - 2r \frac{uw}{\sigma_1\sigma_2} + \frac{w^2}{\sigma_2^2} \right] \right\} du dw \\
&= A' \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\{j(\omega_1 u + \omega_2 w)\} \times \\
&\quad \exp\left\{-\frac{1}{2(1-r^2)} \left[\frac{u^2}{\sigma_1^2} - 2r \frac{uw}{\sigma_1\sigma_2} + \frac{w^2}{\sigma_2^2} \right] \right\} du dw,
\end{aligned}$$

donde en la última igualdad se definió

$$A' = A \exp\{j(\omega_1\eta_1 + \omega_2\eta_2)\} = \frac{\exp\{j(\omega_1\eta_1 + \omega_2\eta_2)\}}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}. \quad (5.66)$$

La integral doble puede separarse como

$$\Phi(\omega_1, \omega_2) = A' \int_{-\infty}^{\infty} e^{j\omega_2 w} \exp \left\{ -\frac{1}{2(1-r^2)} \frac{w^2}{\sigma_2^2} \right\} \times \left(\int_{-\infty}^{\infty} e^{j\omega_1 u} \exp \left\{ -\frac{1}{2(1-r^2)} \left[\frac{u^2}{\sigma_1^2} - 2r \frac{uw}{\sigma_1 \sigma_2} \right] \right\} du \right) dw. \quad (5.67)$$

Se operará a continuación con la integral interna entre paréntesis. La exponencial de la integral interna puede escribirse como

$$\exp \left\{ -\frac{1}{2(1-r^2)} \left[\frac{u^2}{\sigma_1^2} - 2r \frac{uw}{\sigma_1 \sigma_2} \right] \right\} = \exp \left\{ -\frac{1}{2\sigma_1^2(1-r^2)} \left[u^2 - \frac{2r\sigma_1 w}{\sigma_2} u \right] \right\}.$$

Empleando la técnica de completar el cuadrado con el término entre paréntesis rectos, éste puede escribirse como

$$u^2 - \frac{2r\sigma_1 w}{\sigma_2} u = \left[u - \frac{r\sigma_1 w}{\sigma_2} \right]^2 - \frac{r^2 \sigma_1^2 w^2}{\sigma_2^2}$$

ya que

$$\left[u - \frac{r\sigma_1 w}{\sigma_2} \right]^2 = u^2 - 2 \frac{r\sigma_1 w}{\sigma_2} u + \frac{r^2 \sigma_1^2 w^2}{\sigma_2^2}.$$

Por lo tanto, la integral interna puede escribirse como

$$\int_{-\infty}^{\infty} e^{j\omega_1 u} \exp \left\{ -\frac{1}{2\sigma_1^2(1-r^2)} \left[u - \frac{r\sigma_1 w}{\sigma_2} \right]^2 \right\} \exp \left\{ \frac{1}{2\sigma_1^2(1-r^2)} \frac{r^2 \sigma_1^2 w^2}{\sigma_2^2} \right\} du$$

y como la última exponencial no depende de la variable de integración u , puede sacarse fuera de la integral, resultando en

$$\exp \left\{ \frac{1}{2(1-r^2)} \frac{r^2 w^2}{\sigma_2^2} \right\} \int_{-\infty}^{\infty} e^{j\omega_1 u} \exp \left\{ -\frac{1}{2\sigma_1^2(1-r^2)} \left[u - \frac{r\sigma_1 w}{\sigma_2} \right]^2 \right\} du.$$

Al realizar el cambio de variable $v = u - r\sigma_1 w/\sigma_2$, se obtiene que

$$\exp \left\{ \frac{1}{2(1-r^2)} \frac{r^2 w^2}{\sigma_2^2} \right\} \int_{-\infty}^{\infty} \exp \left\{ j\omega_1 \left(v + \frac{r\sigma_1 w}{\sigma_2} \right) \right\} \exp \left\{ -\frac{v^2}{2\sigma_1^2(1-r^2)} \right\} dv,$$

es decir,

$$\exp \left\{ \frac{1}{2(1-r^2)} \frac{r^2 w^2}{\sigma_2^2} \right\} \exp \left\{ j\omega_1 \frac{r\sigma_1 w}{\sigma_2} \right\} \int_{-\infty}^{\infty} e^{j\omega_1 v} \exp \left\{ -\frac{v^2}{2\sigma_1^2(1-r^2)} \right\} dv,$$

donde se desarrolló la primer exponencial del integrando y se sacó afuera de la integral el término que no depende de la variable de integración. Sustituyendo este resultado por la integral interna en la ecuación 5.67, se obtiene que

$$\Phi(\omega_1, \omega_2) = A' \int_{-\infty}^{\infty} e^{j\omega_2 w} \exp \left\{ -\frac{1}{2(1-r^2)} \frac{w^2}{\sigma_2^2} \right\} \left(\exp \left\{ \frac{1}{2(1-r^2)} \frac{r^2 w^2}{\sigma_2^2} \right\} \times \exp \left\{ j\omega_1 \frac{r\sigma_1 w}{\sigma_2} \right\} \int_{-\infty}^{\infty} e^{j\omega_1 v} \exp \left\{ -\frac{v^2}{2\sigma_1^2(1-r^2)} \right\} dv \right) dw,$$

y operando, se llega a que

$$\Phi(\omega_1, \omega_2) = A' \int_{-\infty}^{\infty} \exp \left\{ jw \left(\omega_1 \frac{r\sigma_1}{\sigma_2} + \omega_2 \right) \right\} \exp \left\{ -\frac{w^2}{2\sigma_2^2} \right\} dw \times \int_{-\infty}^{\infty} \exp \{ jv\omega_1 \} \exp \left\{ -\frac{v^2}{2\sigma_1^2(1-r^2)} \right\} dv, \quad (5.68)$$

donde como se separó el integrando de la integral doble en $h(w, v) = h_1(w)h_2(v)$, la integral doble se separó en el producto de dos integrales simples, y además se sacaron factores comunes en el exponente de las exponenciales de la primer integral.

Ambas integrales de la ecuación 5.68 son de la forma

$$f(b) = \int_{-\infty}^{\infty} e^{-ax^2} e^{jbx} dx,$$

por lo que a continuación, se resolverá esta integral y se aplicará el resultado en la resolución de la ecuación 5.68. Teniendo en cuenta que $e^{jbx} = \cos(bx) + j \sin(bx)$, se ve que

$$\begin{aligned} f(b) &= \int_{-\infty}^{\infty} e^{-ax^2} e^{jbx} dx \\ &= \int_{-\infty}^{\infty} e^{-ax^2} \cos(bx) dx + j \int_{-\infty}^{\infty} e^{-ax^2} \sin(bx) dx \\ &\stackrel{(a)}{=} \int_{-\infty}^{\infty} e^{-ax^2} \cos(bx) dx, \end{aligned} \quad (5.69)$$

donde en (a) se consideró que la segunda integral es nula ya que integrando es una función impar por tratarse del producto de una función par con una función impar. Considérese ahora

$$\begin{aligned} f'(b) &= \frac{d}{db} \left(\int_{-\infty}^{\infty} e^{-ax^2} \cos(bx) dx \right) \\ &= \int_{-\infty}^{\infty} e^{-ax^2} \frac{d}{db} (\cos(bx)) dx \\ &= - \int_{-\infty}^{\infty} e^{-ax^2} x \sin(bx) dx \\ &\stackrel{(a)}{=} \frac{1}{2a} e^{-ax^2} \sin(bx) \Big|_{-\infty}^{\infty} - \frac{b}{2a} \int_{-\infty}^{\infty} e^{-ax^2} \cos(bx) dx \\ &\stackrel{(b)}{=} -\frac{b}{2a} \int_{-\infty}^{\infty} e^{-ax^2} \cos(bx) dx \end{aligned}$$

donde en (a) se aplicó integración por partes, $\int u dv = uv - \int v du$, con $u = \sin(bx)$ y $dv = -xe^{-ax^2}$ por lo que $du = b \cos(bx)$ y $v = e^{-ax^2}/2a$ y en (b) se tuvo en cuenta que el primer sumando se anula ya que $\lim_{x \rightarrow \pm\infty} e^{-ax^2} \sin(bx) = 0$. Comparando este resultado con la ecuación 5.69, se concluye que

$$f'(b) = -\frac{b}{2a} f(b).$$

Para resolver esta ecuación diferencial, se observa que

$$\frac{f'(b)}{f(b)} = \frac{d}{db} \ln f(b) = -\frac{b}{2a},$$

por lo que integrando respecto a b se obtiene que

$$\ln f(b) = -\frac{b^2}{4a} + C'$$

y despejando $f(b)$ se obtiene que

$$f(b) = Ce^{-b^2/4a}.$$

Para determinar la constante C , se evalúa en $b = 0$,

$$f(0) = C = \int_{-\infty}^{\infty} e^{-ax^2} dx \stackrel{(a)}{=} \sqrt{\frac{\pi}{a}},$$

donde en (a) se consideró que la integral es gaussiana por lo que se cumple el resultado de la ecuación 4.18. Se concluye que

$$\int_{-\infty}^{\infty} e^{-ax^2} e^{jbx} dx = \int_{-\infty}^{\infty} e^{-ax^2} \cos(bx) dx = \sqrt{\frac{\pi}{a}} e^{-b^2/4a}. \quad (5.70)$$

Aplicando el resultado de la ecuación 5.70 a las integrales de la ecuación 5.68 con $a = 1/2\sigma_2^2$ y $b = \omega_1 r \sigma_1 / \sigma_2 + \omega_2$ en la primer integral y $a = 1/[2\sigma_1^2(1-r^2)]$ y $b = \omega_1$ en la segunda, y reemplazando A' por su valor, dado por la ecuación 5.66, se obtiene que

$$\begin{aligned} \Phi(\omega_1, \omega_2) &= \frac{\exp\{j(\omega_1\eta_1 + \omega_2\eta_2)\}}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \sqrt{\pi 2\sigma_2^2} \exp\left\{-\left(\omega_1 \frac{r\sigma_1}{\sigma_2} + \omega_2\right)^2 \frac{2\sigma_2^2}{4}\right\} \times \\ &\quad \sqrt{\pi 2\sigma_1^2(1-r^2)} \exp\left\{-\omega_1^2 \frac{2\sigma_1^2(1-r^2)}{4}\right\} \\ &= \exp\{j(\omega_1\eta_1 + \omega_2\eta_2)\} \exp\left\{-\frac{1}{2}(\omega_1 r \sigma_1 + \omega_2 \sigma_2)^2 - \frac{1}{2}\omega_1^2 \sigma_1^2(1-r^2)\right\} \\ &= \exp\{j(\omega_1\eta_1 + \omega_2\eta_2)\} \exp\left\{-\frac{1}{2}(\omega_1^2 r^2 \sigma_1^2 + \omega_2^2 \sigma_2^2 + 2\omega_1 \omega_2 r \sigma_1 \sigma_2 + \omega_1^2 \sigma_1^2 - \omega_1^2 \sigma_1^2 r^2)\right\}. \end{aligned}$$

Se concluye que

$$\Phi(\omega_1, \omega_2) = \exp\{j(\omega_1\eta_1 + \omega_2\eta_2)\} \exp\left\{-\frac{1}{2}(\omega_1^2 \sigma_1^2 + 2\omega_1 \omega_2 r \sigma_1 \sigma_2 + \omega_2^2 \sigma_2^2)\right\},$$

que es lo que se quería demostrar.

5.5.2. Teorema de los momentos

La función generadora de momentos de \mathbf{x} e \mathbf{y} está dada por

$$\Phi(s_1, s_2) = E\{e^{s_1\mathbf{x} + s_2\mathbf{y}}\}.$$

Teniendo en cuenta la expansión de la función exponencial,

$$e^{-x} = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

y considerando la linealidad de la esperanza, se ve que

$$\Phi(s_1, s_2) = E\{e^{s_1\mathbf{x} + s_2\mathbf{y}}\} = \sum_{n=0}^{\infty} \frac{E\{(s_1\mathbf{x} + s_2\mathbf{y})^n\}}{n!},$$

y usando el teorema del binomio, que indica que

$$(x+y)^n = \sum_{k=0}^n x^k y^{n-k},$$

se obtiene que

$$\begin{aligned} \Phi(s_1, s_2) &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} E\{\mathbf{x}^k \mathbf{y}^{n-k}\} s_1^k s_2^{n-k} \\ &= 1 + m_{10}s_1 + m_{01}s_2 + \frac{1}{2}(m_{20}s_1^2 + 2m_{11}s_1s_2 + m_{02}s_2^2) \\ &\quad + \frac{1}{6}(m_{30}s_1^3 + 3m_{21}s_1^2s_2 + 3m_{12}s_1s_2^2 + m_{03}s_2^3) + \dots \end{aligned} \quad (5.71)$$

Es fácil ver que se cumple que

$$\left. \frac{\partial^k \partial^r}{\partial s_1^k \partial s_2^r} \Phi(s_1, s_2) \right|_{s_1=0, s_2=0} = m_{kr}.$$

Ejemplo Si \mathbf{x} e \mathbf{y} son variables aleatorias conjuntamente normales de media nula, se cumple que

$$E\{\mathbf{x}^2\mathbf{y}^2\} = E\{\mathbf{x}^2\}E\{\mathbf{y}^2\} + 2E^2\{\mathbf{xy}\}. \quad (5.72)$$

La función generadora de momentos de \mathbf{x} e \mathbf{y} se obtiene evaluando la ecuación 5.65 en $s_1 = j\omega_1$ y $s_2 = j\omega_2$, y es

$$\Phi(s_1, s_2) = e^{-A}, \quad A = \frac{1}{2} (\sigma_1^2 s_1^2 + 2Cs_1 s_2 + \sigma_2^2 s_2^2),$$

con $C = E\{\mathbf{xy}\} = r\sigma_1\sigma_2$. Notando que $E\{\mathbf{x}^2\mathbf{y}^2\}$ multiplica al término $s_1^2 s_2^2$ en la función generadora de momentos de la ecuación 5.71, la igualdad 5.72 puede demostrarse igualando los coeficientes que multiplican a $s_1^2 s_2^2$ en la función generadora de momentos y en la expansión de e^{-A} . En la función generadora de momentos 5.71, el coeficiente que multiplica a $s_1^2 s_2^2$ en el correspondiente a $n = 4$ y $k = 2$, y es

$$\frac{1}{4!} \binom{4}{2} E\{\mathbf{x}^2\mathbf{y}^2\} = \frac{1}{4} E\{\mathbf{x}^2\mathbf{y}^2\}.$$

Por otro lado, en

$$\Phi(s_1, s_2) = e^{-A} = \sum_{n=0}^{\infty} \frac{A^n}{n!} = 1 + A + \frac{A^2}{2} + \frac{A^3}{6} + \dots$$

el único término que contiene el factor $s_1^2 s_2^2$ es

$$\frac{A^2}{2} = \frac{1}{8} (\sigma_1^2 s_1^2 + 2Cs_1 s_2 + \sigma_2^2 s_2^2)^2,$$

y el coeficiente que lo multiplica es

$$\frac{1}{8} (2\sigma_1^2 \sigma_2^2 + 4C^2) = \frac{1}{4} (\sigma_1^2 \sigma_2^2 + 2C^2).$$

Igualando ambas ecuaciones, se tiene que

$$E\{\mathbf{x}^2\mathbf{y}^2\} = \sigma_1^2 \sigma_2^2 + 2C^2,$$

y teniendo en cuenta que \mathbf{x} e \mathbf{y} tienen media nula y por lo tanto $\sigma_1^2 = E\{\mathbf{x}^2\}$, $\sigma_2^2 = E\{\mathbf{y}^2\}$ y $C = E\{\mathbf{xy}\}$, se obtiene la ecuación 5.72.

Teorema de Price Dadas dos variables aleatorias \mathbf{x} e \mathbf{y} conjuntamente normales, se forma la media de una función $g(x, y)$,

$$I = E\{g(\mathbf{x}, \mathbf{y})\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy. \quad (5.73)$$

Esta integral es una función $I(\mu)$ de la covarianza μ de las variables aleatorias \mathbf{x} e \mathbf{y} , además de los otros cuatro parámetros que caracterizan la densidad normal conjunta $f(x, y)$ (medias y varianzas de \mathbf{x} e \mathbf{y}). Si $g(x, y)f(x, y) \rightarrow 0$ con $(x, y) \rightarrow \infty$, se cumple que

$$\frac{\partial^n I(\mu)}{\partial \mu^n} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial^{2n} g(x, y)}{\partial x^n \partial y^n} f(x, y) dx dy = E \left\{ \frac{\partial^{2n} g(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}^n \partial \mathbf{y}^n} \right\} \quad (5.74)$$

Para demostrarlo, se parte expresando la densidad conjunta $f(x, y)$ en términos de la función característica conjunta usando la fórmula de inversión de la transformada de Fourier de la ecuación 5.60, y se sustituye en la ecuación 5.73,

$$I(\mu) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi(\omega_1, \omega_2) e^{-j(\omega_1 x + \omega_2 y)} d\omega_1 d\omega_2 dx dy,$$

donde por ser $f(x, y)$ una densidad normal, $\Phi(\omega_1, \omega_2)$ está dada por la ecuación 5.65 con $\mu = r\sigma_1\sigma_2$,

$$\Phi(\omega_1, \omega_2) = e^{j(\eta_1 \omega_1 + \eta_2 \omega_2)} e^{-(\omega_1^2 \sigma_1^2 + 2\mu \omega_1 \omega_2 + \omega_2^2 \sigma_2^2)/2}.$$

Por lo tanto, diferenciando $I(\mu)$ n veces respecto a μ se obtiene que

$$\begin{aligned}\frac{\partial^n I(\mu)}{\partial \mu^n} &= \frac{(-1)^n}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \\ &\quad \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega_1^n \omega_2^n \Phi(\omega_1, \omega_2) e^{-j(\omega_1 x + \omega_2 y)} d\omega_1 d\omega_2 dx dy.\end{aligned}$$

Por otro lado, si diferencia la ecuación 5.60 n veces respecto a x y n veces respecto a y resulta en

$$\begin{aligned}\frac{\partial^{2n} f(x, y)}{\partial x^n \partial y^n} &= \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (-j\omega_1)^n (-j\omega_2)^n \Phi(\omega_1, \omega_2) e^{-j(\omega_1 x + \omega_2 y)} d\omega_1 d\omega_2 \\ &= \frac{(-1)^n}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega_1^n \omega_2^n \Phi(\omega_1, \omega_2) e^{-j(\omega_1 x + \omega_2 y)} d\omega_1 d\omega_2\end{aligned}$$

por lo que la ecuación anterior queda

$$\frac{\partial^n I(\mu)}{\partial \mu^n} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \frac{\partial^{2n} f(x, y)}{\partial x^n \partial y^n} dx dy.$$

Finalmente, aplicando integración por partes repetidamente y empleando la condición en ∞ se obtiene la ecuación 5.74, que es lo que se quería demostrar.

La integración por partes se aplica de la siguiente forma. Se considera primero la integral interna,

$$\frac{\partial^n I(\mu)}{\partial \mu^n} = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} g(x, y) \frac{\partial^{2n} f(x, y)}{\partial x^n \partial y^n} dx \right) dy,$$

en donde el integrando es solo función de la variable de integración x , e y puede considerarse una constante. En esta integral se aplica integración por partes, $\int u dv = uv - \int v du$, con

$$\begin{aligned}u &= g(x, y) & du &= \frac{\partial g(x, y)}{\partial x} \\ dv &= \frac{\partial^{2n} f(x, y)}{\partial x^n \partial y^n} & v &= \frac{\partial^{2n-1} f(x, y)}{\partial x^{n-1} \partial y^n},\end{aligned}$$

resultando en

$$\frac{\partial^n I(\mu)}{\partial \mu^n} = \int_{-\infty}^{\infty} \left(g(x, y) \frac{\partial^{2n-1} f(x, y)}{\partial x^{n-1} \partial y^n} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{\partial g(x, y)}{\partial x} \cdot \frac{\partial^{2n-1} f(x, y)}{\partial x^{n-1} \partial y^n} dx \right) dy.$$

El primer sumando en el integrando de la integral externa se anula debido a la hipótesis $g(x, y)f(x, y) \rightarrow 0$ con $(x, y) \rightarrow \infty$, ya que si $f(x) \rightarrow 0$ con $x \rightarrow \infty$, también se cumple que $f'(x) \rightarrow 0$ [12]. Por lo tanto, se obtiene que

$$\frac{\partial^n I(\mu)}{\partial \mu^n} = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial g(x, y)}{\partial x} \cdot \frac{\partial^{2n-1} f(x, y)}{\partial x^{n-1} \partial y^n} dx dy.$$

Si se cambia el orden de integración y se aplica integración por partes a la integral de variable de integración y , se obtiene que

$$\frac{\partial^n I(\mu)}{\partial \mu^n} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial^2 g(x, y)}{\partial x \partial y} \cdot \frac{\partial^{2n-2} f(x, y)}{\partial x^{n-1} \partial y^{n-1}} dx dy.$$

Realizando este procedimiento n veces en total, se obtiene el resultado de la ecuación 5.74.

Ejemplo Se empleará el teorema de Price para derivar el resultado de la ecuación 5.72. El teorema de Price con $g(x, y) = x^2 y^2$ y $n = 1$ indica que

$$\frac{\partial I(\mu)}{\partial \mu} = E \left\{ \frac{\partial^2 g(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x} \partial \mathbf{y}} \right\} = 4E\{\mathbf{xy}\} = 4\mu,$$

e integrando se obtiene que

$$I(\mu) = \frac{4\mu^2}{2} + I(0) \quad (5.75)$$

Si $\mu = 0$, las variables aleatorias son no correlacionadas, y por ser conjuntamente normales, son independientes. Por lo tanto,

$$I(0) = E\{\mathbf{x}^2\mathbf{y}^2\} = E\{\mathbf{x}^2\}E\{\mathbf{y}^2\}.$$

Con este resultado y $I(\mu) = E\{\mathbf{x}^2\mathbf{y}^2\}$ y $\mu = E\{\mathbf{xy}\}$ en la ecuación 5.75 se obtiene el resultado buscado.

5.6. Distribuciones condicionales

Extendiendo el resultado de la ecuación 3.24 en la sección 3.4 para el caso de dos variables, la *distribución condicional conjunta* puede expresarse como

$$F_{zw}(z, w | M) = P\{\mathbf{z} \leq z, \mathbf{w} \leq w | M\} = \frac{P\{\mathbf{z} \leq z, \mathbf{w} \leq w, M\}}{P(M)}. \quad (5.76)$$

La densidad de probabilidad conjunta y los marginales se obtienen mediante las derivadas apropiadas.

Ejemplo (6-39) Se determinará la distribución de probabilidad $F(x, y | M)$ con $M = \{x_1 < \mathbf{x} \leq x_2\}$. A partir de la ecuación 5.76 se tiene que

$$\begin{aligned} F(x, y | x_1 < \mathbf{x} \leq x_2) &= \frac{P\{\mathbf{x} < x, \mathbf{y} < y, x_1 < \mathbf{x} \leq x_2\}}{P\{x_1 < \mathbf{x} \leq x_2\}} \\ &= \begin{cases} \frac{P\{x_1 < \mathbf{x} \leq x_2, \mathbf{y} < y\}}{P\{x_1 < \mathbf{x} \leq x_2\}}, & x > x_2 \\ \frac{P\{x_1 < \mathbf{x} \leq x, \mathbf{y} < y\}}{P\{x_1 < \mathbf{x} \leq x_2\}}, & x_1 < x \leq x_2 \\ 0, & x < x_1 \end{cases} \\ &= \begin{cases} \frac{F(x_2, y) - F(x_1, y)}{F_x(x_2) - F_x(x_1)}, & x > x_2 \\ \frac{F(x, y) - F(x_1, y)}{F_x(x_2) - F_x(x_1)}, & x_1 < x \leq x_2 \\ 0, & x < x_1 \end{cases}. \end{aligned}$$

La densidad condicional conjunta se obtiene diferenciando la distribución condicional respecto a x y a y , y es

$$f(x, y | x_1 < \mathbf{x} \leq x_2) = \begin{cases} \frac{f(x, y)}{F_x(x_2) - F_x(x_1)}, & x_1 < x \leq x_2 \\ 0, & \text{en otro caso} \end{cases}. \quad (5.77)$$

Se determinará ahora la densidad condicional de \mathbf{y} asumiendo $\mathbf{x} = x$, $f_y(y | \mathbf{x} = x)$. Esto no puede deducirse directamente de la ecuación 5.76 debido a que en general el evento $\{\mathbf{x} = x\}$ tiene probabilidad 0. Sin embargo, puede ser definido como el límite. Para esto, se parte considerando el evento $M = \{x_1 < \mathbf{x} \leq x_2\}$. Integrando la ecuación 5.77 respecto a x se obtiene el marginal de y , que es

$$f_y(y | x_1 < \mathbf{x} \leq x_2) = \int_{-\infty}^{\infty} f(x, y | x_1 < \mathbf{x} \leq x_2) dx = \frac{\int_{x_1}^{x_2} f(\alpha, y) d\alpha}{F_x(x_2) - F_x(x_1)},$$

donde se tuvo en cuenta que el integrando en la primera igualdad es distinto de cero en $x_1 < \mathbf{x} \leq x_2$. Para determinar $f_y(y | \mathbf{x} = x)$, se establece $x_1 = x$ y $x_2 = x + \Delta x$ en la ecuación anterior, obteniéndose

$$f_y(y | x < \mathbf{x} \leq x + \Delta x) = \frac{\int_x^{x+\Delta x} f(\alpha, y) d\alpha}{F_x(x + \Delta x) - F_x(x)} \approx \frac{f(x, y) \Delta x}{f_x(x) \Delta x}.$$

Se concluye que

$$f_y(y | \mathbf{x} = x) = \lim_{\Delta x \rightarrow 0} f_y(y | x < \mathbf{x} \leq x + \Delta x) = \frac{f(x, y)}{f_x(x)}.$$

En los casos en que no haya ambigüedad la función $f_y(y | \mathbf{x} = x) = f_{y|x}(y|x)$ se denotará como $f(y|x)$. $f(x|y)$ se obtiene de forma análoga. En resumen,

$$f(y | x) = \frac{f(x, y)}{f(x)}, \quad f(x | y) = \frac{f(x, y)}{f(y)}. \quad (5.78)$$

Si las variables aleatorias \mathbf{x} e \mathbf{y} son independientes, se cumple que $f(x, y) = f(x)f(y)$ y por lo tanto

$$f(y | x) = f(y), \quad f(x | y) = f(x).$$

Ejemplo (6-41) Se calculará la densidad condicional $f(x | y)$ en el caso en que las variables aleatorias \mathbf{x} e \mathbf{y} son conjuntamente normales.

Para hacerlo, se puede emplear la ecuación 5.78. La densidad de probabilidad conjunta $f(x, y)$ está dada por las ecuaciones 5.11 y 5.12. Como se mostró en la sección 5.1.6, el exponente de $f(x, y)$ puede escribirse como (ecuación 5.13)

$$\frac{\left\{x - \left[\eta_1 + r\frac{\sigma_1}{\sigma_2}(y - \eta_2)\right]\right\}^2}{2\sigma_1^2(1 - r^2)} + \frac{(y - \eta_2)^2}{2\sigma_2^2},$$

y por lo tanto, al dividir entre $f(y)$, el término exponencial $(y - \eta_2)^2/2\sigma_2^2$ se elimina, así como un factor $\sigma_2\sqrt{2\pi}$ de la constante multiplicativa. Esto conduce a que

$$f(x | y) = \frac{1}{\sigma_1\sqrt{2\pi(1 - r^2)}} \exp\left(\frac{\left\{x - \left[\eta_1 + r\frac{\sigma_1}{\sigma_2}(y - \eta_2)\right]\right\}^2}{2\sigma_1^2(1 - r^2)}\right), \quad (5.79)$$

es decir, para un y dado, $f(x | y)$ es una densidad de probabilidad normal de media $\eta_1 + r\sigma_1(y - \eta_2)/\sigma_2$ y varianza $\sigma_1^2(1 - r^2)$.

5.6.1. Teorema de Bayes y probabilidad total

Despejando $f(x, y)$ en una de las ecuaciones 5.78 y sustituyendo el resultado en la otra, se obtiene que

$$f(x | y) = \frac{f(y | x)f(x)}{f(y)}. \quad (5.80)$$

Esta es la versión en densidades de probabilidad de la ecuación 1.22. Teniendo en cuenta que

$$f(y) = \int_{-\infty}^{\infty} f(x, y) dx \quad y \quad f(x, y) = f(y | x)f(x),$$

se concluye que

$$f(y) = \int_{-\infty}^{\infty} f(y | x)f(x) dx.$$

Esta es la ecuación análoga a la ecuación 3.30, y consiste en el teorema de probabilidad total para la densidad en lugar de la probabilidad. La ecuación indica que para eliminar la condición $\mathbf{x} = x$ de la densidad condicional $f(y | x)$, hay que multiplicar por la densidad de $f(x)$ de \mathbf{x} e integrar. Sustituyendo este resultado en la ecuación 5.80, se obtiene el teorema de Bayes para densidades de probabilidad, e indica que

$$f(x | y) = \frac{f(y | x)f(x)}{\int_{-\infty}^{\infty} f(y | x)f(x) dx}. \quad (5.81)$$

Esta ecuación es la análoga a la ecuación 3.31.

Ejemplo (6-42) El teorema de Bayes puede usarse para actualizar el conocimiento de algún parámetro desconocido a partir de observaciones, como se ilustra en este ejemplo.

Una fase θ desconocida se modela como una variable aleatoria con distribución uniforme en el intervalo $(0, 2\pi)$. Este modelo es consistente con el hecho de que no se tiene conocimiento a priori de la fase. Se dispone de una observación ruidosa $\mathbf{r} = \theta + \mathbf{n}$, con $\mathbf{n} \sim N(0, \sigma^2)$. Se quiere determinar $f(\theta | r)$.

Como \mathbf{n} modela el ruido en la observación \mathbf{r} , una hipótesis razonable es considerar que \mathbf{n} y θ son variables aleatorias independientes. Con esta hipótesis, se cumple que $f(r | \theta = \theta) \sim N(\theta, \sigma^2)$, ya que al considerar θ dado, se asume como constante, y en ese caso la variable aleatoria $\mathbf{r} = \theta + \mathbf{n}$ se comporta como \mathbf{n} , con la constante θ afectando únicamente la media. Por lo tanto,

$$f(r | \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(r-\theta)^2/2\sigma^2} \quad \text{y} \quad f_\theta(\theta) = \frac{1}{2\pi} \quad \text{en } 0 < \theta < 2\pi.$$

Aplicando la ecuación 5.81 se obtiene que la densidad de probabilidad a posteriori de θ dado \mathbf{r} es

$$f(\theta | r) = \frac{f(r | \theta) f_\theta(\theta)}{\int_{-\infty}^{\infty} f(r | \theta) f_\theta(\theta) d\theta} = \frac{e^{-(r-\theta)^2/2\sigma^2}}{\int_0^{2\pi} e^{-(r-\theta)^2/2\sigma^2} d\theta} \quad \text{en } 0 < \theta < 2\pi,$$

donde se tuvo en cuenta que el denominador en las expresiones del numerador y el denominador es el mismo y por lo tanto se anula, y también que $f_\theta(\theta) \neq 0$ solo en $0 < \theta < 2\pi$. Notar que el denominador es una constante para un r dado y además la expresión integra a 1 respecto a θ . El conocimiento adquirido a través la observación se refleja en la distribución a posteriori de θ , la cual se muestra en la figura 5.18. En la figura se aprecia que la fase ya no es uniforme, sino que tiene mayor probabilidad en torno a $\theta = r$.

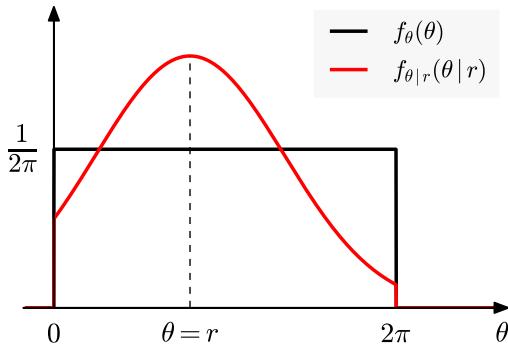


Figura 5.18: Densidad $f_\theta(\theta)$ a priori y densidad $f_{\theta|r}(\theta | r)$ a posteriori. La distribución a posteriori refleja el conocimiento brindado por la observación \mathbf{r} . Luego de conocer \mathbf{r} , la probabilidad de θ ya no es uniforme, sino que tiene valores más probables en torno a r .

5.7. Valores esperados condicionales

Aplicando la ecuación 4.7 a densidades condicionales, se observa que la esperanza de $g(\mathbf{y})$ es

$$E\{g(\mathbf{y}) | M\} = \int_{-\infty}^{\infty} g(y) f(y | M) dy. \quad (5.82)$$

Además, si $M = \{x < \mathbf{x} \leq x + \Delta x\}$ con $\Delta x \rightarrow 0$ se obtiene que

$$E\{g(\mathbf{y}) | x\} = \int_{-\infty}^{\infty} g(y) f(y | x) dy.$$

En particular, la *media condicional* de \mathbf{y} dado que $\mathbf{x} = x$ es

$$\eta_{y|x} = E\{\mathbf{y} | x\} = \int_{-\infty}^{\infty} y f(y | x) dy, \quad (5.83)$$

y la *varianza condicional* es

$$\sigma_{y|x}^2 = E\{(\mathbf{y} - \eta_{y|x})^2 | x\} = \int_{-\infty}^{\infty} (y - \eta_{y|x})^2 f(y | x) dy.$$

Ejemplo (6-48) Sea

$$f_{xy}(x, y) = \begin{cases} 1 & 0 < |y| < x < 1 \\ 0, & \text{en otro caso} \end{cases}.$$

Se quiere determinar $E\{\mathbf{x} | y\}$ y $E\{\mathbf{y} | x\}$.

Como se muestra en la figura 5.19, $f_{xy}(x, y) = 1$ en la región sombreada y 0 en el resto del plano. Por lo tanto, el marginal de \mathbf{x} es

$$f_x(x) = \int_{-\infty}^{\infty} f_{xy}(x, y) dy = \int_{-x}^x 1 \cdot dy = 2x, \quad 0 < x < 1$$

y la densidad marginal de \mathbf{y} es

$$f_y(y) = \int_{-\infty}^{\infty} f_{xy}(x, y) dx = \int_{|y|}^1 1 \cdot dx = 1 - |y|, \quad |y| < 1.$$

De esta forma, a partir de las ecuaciones 5.78, se tiene que

$$f_{x|y}(x | y) = \frac{f_{xy}(x, y)}{f_y(y)} = \frac{1}{1 - |y|}, \quad 0 < |y| < x < 1$$

y

$$f_{y|x}(y | x) = \frac{f_{xy}(x, y)}{f_x(x)} = \frac{1}{2x}, \quad 0 < |y| < x < 1.$$

Por lo tanto,

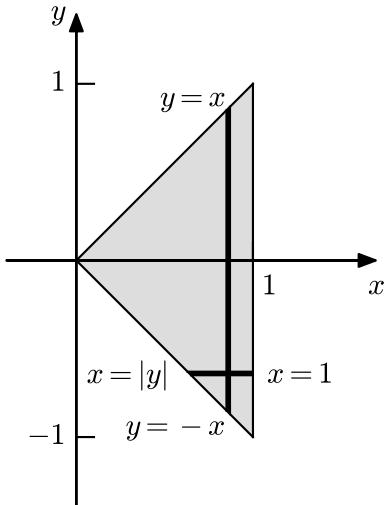


Figura 5.19: Densidad de probabilidad conjunta $f_{xy}(x, y)$.

$$\begin{aligned} E\{\mathbf{x} | y\} &= \int_{-\infty}^{\infty} x f_{x|y}(x | y) dx \\ &= \int_{|y|}^1 \frac{x}{1 - |y|} dx \\ &= \frac{1}{1 - |y|} \cdot \frac{x^2}{2} \Big|_{|y|}^1 \\ &= \frac{1 - |y|^2}{2(1 - |y|)} \\ &= \frac{(1 + |y|)(1 - |y|)}{2(1 - |y|)} \\ &= \frac{1 + |y|}{2}, \quad |y| < 1, \end{aligned}$$

y

$$\begin{aligned}
E\{\mathbf{y} \mid x\} &= \int_{-\infty}^{\infty} y f_{y|x}(y \mid x) dy \\
&= \int_{-x}^x \frac{y}{2x} dy \\
&= \frac{1}{2x} \cdot \frac{y^2}{2} \Big|_{-x}^x \\
&= \frac{x^2 - (-x)^2}{4x} \\
&= 0, \quad 0 < x < 1,
\end{aligned}$$

Para un x dado, la integral de la ecuación 5.83 es el centro de gravedad de la masa en la banda vertical $(x, x + dx)$. El lugar de esos puntos al variar x es la función

$$\varphi(x) = \int_{-\infty}^{\infty} y f(y \mid x) dy, \quad (5.84)$$

denominada *línea de regresión*.

A partir de los resultados de las ecuaciones 5.47 y 5.82 se deduce que

$$E\{g(\mathbf{x}, \mathbf{y}) \mid M\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y \mid M) dx dy. \quad (5.85)$$

Para definir $E\{g(\mathbf{x}, \mathbf{y}) \mid x\}$, hay que considerar límites. Estableciendo $x_1 = x$ y $x_2 = x + \Delta x$ en la ecuación 5.77, se tiene que

$$f(x, y \mid x < \mathbf{x} \leq x + \Delta x) = \begin{cases} \frac{f(x, y)}{F_x(x + \Delta x) - F_x(x)}, & x < x \leq x + \Delta x \\ 0, & \text{en otro caso} \end{cases}$$

y sustituyendo en la ecuación 5.85 con $M = \{x < \mathbf{x} \leq x + \Delta x\}$ se tiene que

$$E\{g(\mathbf{x}, \mathbf{y}) \mid x < \mathbf{x} \leq x + \Delta x\} = \int_{-\infty}^{\infty} \int_x^{x + \Delta x} \frac{g(\alpha, y) f(\alpha, y)}{F_x(\alpha + \Delta x) - F_x(\alpha)} d\alpha dy.$$

Con $\Delta x \rightarrow 0$, la integral interior tiende a $g(x, y) f(x, y) / f(x)$. Definiendo $E\{g(\mathbf{x}, \mathbf{y}) \mid x\}$ como el límite de la integral, se concluye que

$$E\{g(\mathbf{x}, \mathbf{y}) \mid x\} = \int_{-\infty}^{\infty} g(x, y) f(y \mid x) dy,$$

donde se usó que $f(y \mid x) = f(x, y) / f(x)$. Notar que también se cumple que

$$E\{g(x, \mathbf{y}) \mid x\} = \int_{-\infty}^{\infty} g(x, y) f(y \mid x) dy,$$

ya que $g(x, \mathbf{y})$ es una función de la variable aleatoria \mathbf{y} con un parámetro x y por lo tanto se aplica la ecuación 5.82. Por lo tanto,

$$E\{g(\mathbf{x}, \mathbf{y}) \mid x\} = E\{g(x, \mathbf{y}) \mid x\}.$$

5.7.1. Valores esperados condicionados a variables aleatorias

La media condicional de \mathbf{y} dado $\mathbf{x} = x$ es una función $\varphi(x) = E\{\mathbf{y} \mid x\}$ de x dada por la ecuación 5.84. Usando esta función, se puede construir la variable aleatoria $\varphi(\mathbf{x}) = E\{\mathbf{y} \mid \mathbf{x}\}$, y como indica la ecuación 4.7, la media es

$$E\{\varphi(\mathbf{x})\} = \int_{-\infty}^{\infty} f(x) \int_{-\infty}^{\infty} y f(y \mid x) dy dx.$$

Considerando que $f(x, y) = f(x)f(y | x)$, se obtiene que

$$\begin{aligned} E\{E\{\mathbf{y} | \mathbf{x}\}\} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x, y) dx dy \\ &= \int_{-\infty}^{\infty} y \left[\int_{-\infty}^{\infty} f(x, y) dx \right] dy \\ &= \int_{-\infty}^{\infty} yf_y(y) dy \\ &= E\{\mathbf{y}\}, \end{aligned}$$

es decir,

$$E\{E\{\mathbf{y} | \mathbf{x}\}\} = E\{\mathbf{y}\}. \quad (5.86)$$

Este resultado se denomina *ley de la esperanza total*⁷ y puede generalizarse. La media condicional $E\{g(\mathbf{x}, \mathbf{y}) | \mathbf{x}\}$ de $g(\mathbf{x}, \mathbf{y})$ dado $\mathbf{x} = x$ es una función de la variable real x . A partir de esta función, puede definirse una nueva variable aleatoria $E\{g(\mathbf{x}, \mathbf{y}) | \mathbf{x}\}$,

$$E\{g(\mathbf{x}, \mathbf{y}) | \mathbf{x}\} = \int_{-\infty}^{\infty} g(\mathbf{x}, y)f(y | \mathbf{x}) dy.$$

La media de esta variable aleatoria es

$$\begin{aligned} E\{E\{g(\mathbf{x}, \mathbf{y}) | \mathbf{x}\}\} &= \int_{-\infty}^{\infty} f(x) \int_{-\infty}^{\infty} g(x, y)f(y | x) dy dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y) dx dy \\ &= E\{g(\mathbf{x}, \mathbf{y})\}, \end{aligned}$$

es decir

$$E\{E\{g(\mathbf{x}, \mathbf{y}) | \mathbf{x}\}\} = E\{g(\mathbf{x}, \mathbf{y})\}.$$

Finalmente, se observa que

$$\begin{aligned} E\{g_1(\mathbf{x})g_2(\mathbf{y}) | x\} &= E\{g_1(x)g_2(\mathbf{y}) | x\} = g_1(x)E\{g_2(\mathbf{y}) | x\} \\ E\{g_1(\mathbf{x})g_2(\mathbf{y})\} &= E\{E\{g_1(\mathbf{x})g_2(\mathbf{y}) | \mathbf{x}\}\} = E\{g_1(\mathbf{x})E\{g_2(\mathbf{y}) | \mathbf{x}\}\} \end{aligned}$$

La última igualdad de la segunda ecuación se puede comprobar expresando la esperanza explícitamente como una integral,

$$\begin{aligned} E\{g_1(\mathbf{x})E\{g_2(\mathbf{y}) | \mathbf{x}\}\} &= \int_{-\infty}^{\infty} g_1(x) \left[\int_{-\infty}^{\infty} g_2(y)f(y | x) dy \right] f(x) dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x)g_2(y)f(x, y) dx dy \\ &= E\{g_1(\mathbf{x})g_2(\mathbf{y})\}. \end{aligned}$$

Ejemplo (6-50) Sean las variables aleatorias \mathbf{x} e \mathbf{y} conjuntamente normales de distribución $N(0, 0, \sigma_1^2, \sigma_2^2, r)$. Se quiere mostrar que se cumple que

$$E\{\mathbf{x}\mathbf{y}\} = r\sigma_1\sigma_2, \quad E\{\mathbf{x}^2\mathbf{y}^2\} = E\{\mathbf{x}^2\}E\{\mathbf{y}^2\} + 2E^2\{\mathbf{x}\mathbf{y}\}.$$

Se parte teniendo en cuenta que (ver ecuación 4.16)

$$E\{\mathbf{x}^2\} = \sigma_1^2, \quad E\{\mathbf{x}^4\} = 3\sigma_1^4.$$

⁷Ver https://en.wikipedia.org/wiki/Law_of_total_expectation, por ejemplo.

Además, como indica la ecuación 5.79, $f(y|x)$ es normal con media $r\sigma_2 x/\sigma_1$ y varianza $\sigma_2^2(1-r^2)$. Por lo tanto, de la ecuación 4.20 se tiene que

$$E\{\mathbf{y}^2|x\} = \eta_{y|x}^2 + \sigma_{y|x}^2 = \left(\frac{r\sigma_2 x}{\sigma_1}\right)^2 + \sigma_2^2(1-r^2).$$

Para probar la primera identidad, se observa que

$$\begin{aligned} E\{\mathbf{x}\mathbf{y}\} &= E\{E\{\mathbf{x}\mathbf{y}|\mathbf{x}\}\} = E\{\mathbf{x}E\{\mathbf{y}|\mathbf{x}\}\} = E\left\{\mathbf{x}\frac{r\sigma_2 \mathbf{x}}{\sigma_1}\right\} \\ &= \frac{r\sigma_2}{\sigma_1} E\{\mathbf{x}^2\} = \frac{r\sigma_2}{\sigma_1} \sigma_1^2 = r\sigma_1\sigma_2. \end{aligned}$$

Para probar la segunda identidad, se ve que

$$\begin{aligned} E\{\mathbf{x}^2\mathbf{y}^2\} &= E\{E\{\mathbf{x}^2\mathbf{y}^2|\mathbf{x}\}\} = E\{\mathbf{x}^2E\{\mathbf{y}^2|\mathbf{x}\}\} \\ &= E\left\{\mathbf{x}^2\left[\frac{r^2\sigma_2^2\mathbf{x}^2}{\sigma_1^2} + \sigma_2^2(1-r^2)\right]\right\} \\ &= \frac{r^2\sigma_2^2}{\sigma_1^2} E\{\mathbf{x}^4\} + \sigma_2^2(1-r^2)E\{\mathbf{x}^2\} \\ &= \frac{r^2\sigma_2^2}{\sigma_1^2} (3\sigma_1^4) + \sigma_2^2(1-r^2)(\sigma_1^2) \\ &= 3r^2\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_2^2 - r^2\sigma_1^2\sigma_2^2 \\ &= \sigma_1^2\sigma_2^2 + 2r^2\sigma_1^2\sigma_2^2 \\ &= E\{\mathbf{x}^2\}E\{\mathbf{y}^2\} + 2E^2\{\mathbf{x}\mathbf{y}\}. \end{aligned}$$

Este resultado ya se obtuvo previamente a partir de la función generadora de momentos (ecuación 5.72) y también mediante el teorema de Price.

Capítulo 6

Secuencias de variables aleatorias

6.1. Conceptos generales

Un vector aleatorio es un vector

$$\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$$

cuyos componentes \mathbf{x}_i son variables aleatorias. La probabilidad de que \mathbf{X} pertenezca a una región D del espacio n -dimensional es

$$P\{\mathbf{X} \in D\} = \int_D f(X) dX, \quad X = [x_1, \dots, x_n]$$

donde

$$f(X) = f(x_1, \dots, x_n) = \frac{\partial^n F(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n}$$

es la densidad de probabilidad conjunta (o multivariada) de las variables aleatorias \mathbf{x}_i y

$$F(X) = F(x_1, \dots, x_n) = P\{\mathbf{x}_1 \leq x_1, \dots, \mathbf{x}_n \leq x_n\}$$

es la distribución de probabilidad conjunta.

Generalizando los resultados de la sección 5.1.4, si en $F(x_1, \dots, x_n)$ se sustituyen ciertas variables por ∞ , se obtiene la distribución de las variables restantes. Si se integra $f(x_1, \dots, x_n)$ respecto a ciertas variables, se obtiene la densidad de las variables restantes. Como ejemplo, se cumple que

$$\begin{aligned} F(x_1, x_3) &= F(x_1, \infty, x_3, \infty) \\ f(x_1, x_3) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2, x_3, x_4) dx_2 dx_4 \end{aligned} \tag{6.1}$$

Transformaciones Dadas k funciones

$$g_1(X), \dots, g_k(X), \quad X = [x_1, \dots, x_n]$$

se construyen las k variables aleatorias

$$\mathbf{y}_1 = g_1(\mathbf{X}), \dots, \mathbf{y}_k = g_k(\mathbf{X}).$$

La estadística de estas variables aleatorias puede determinarse en términos de la estadística de \mathbf{X} , de forma análoga a lo explicado en la sección 5.3. Se considerará el caso en que $k = n$, ya que los casos en que $k < n$ o $k > n$ pueden reducirse a $k = n$, como se explica en el capítulo 7 de [1].

Para obtener la densidad $f_y(y_1, \dots, y_n)$ del vector aleatorio $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]$ para un conjunto específico de valores y_1, \dots, y_n , se resuelve el sistema

$$g_1(X) = y_1, \dots, g_n(X) = y_n. \tag{6.2}$$

Si el sistema no tiene soluciones, $f_y(y_1, \dots, y_n) = 0$. Si el sistema tiene una solución única $X = [x_1, \dots, x_n]$,

$$f_y(y_1, \dots, y_n) = \frac{f_x(x_1, \dots, x_n)}{|J(x_1, \dots, x_n)|} \quad (6.3)$$

donde

$$J(x_1, \dots, x_n) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \dots & \frac{\partial g_n}{\partial x_n} \end{vmatrix}$$

es la matriz jacobiana de la transformación 6.2. Si el sistema tiene múltiples soluciones, se suman los términos correspondientes como en la ecuación 5.34.

6.1.1. Independencia

Se dice que las variables aleatorias $\mathbf{x}_1, \dots, \mathbf{x}_n$ son mutuamente independientes si los eventos $\{\mathbf{x}_1 \leq x_1\}, \dots, \{\mathbf{x}_n \leq x_n\}$ son independientes. En ese caso, se cumple que

$$\begin{aligned} F_x(x_1, \dots, x_n) &= F(x_1) \dots F(x_n) \\ f_x(x_1, \dots, x_n) &= f(x_1) \dots f(x_n). \end{aligned}$$

Una observación respecto a la notación, es que las funciones se están identificando a través del nombre de la variable. Las funciones $f(x_i)$ en la ecuación anterior no necesariamente son iguales, sino que la densidad $f_i(x_i)$ de la variable aleatoria \mathbf{x}_i se denotó $f(x_i)$.

Ejemplo (7-1) Sean n variables aleatorias \mathbf{x}_i con respectivas densidades de probabilidad $f_i(x_i)$. Se construyen las variables aleatorias

$$\mathbf{y}_k = \mathbf{x}_1 + \dots + \mathbf{x}_k, \quad k = 1, \dots, n.$$

Se quiere determinar la densidad de probabilidad conjunta de \mathbf{y}_k .

La solución del sistema

$$\begin{aligned} x_1 &= y_1 = g_1(X) \\ x_1 + x_2 &= y_2 = g_2(X) \\ &\vdots \\ x_1 + \dots + x_n &= y_n = g_n(X) \end{aligned} \quad (6.4)$$

es

$$\begin{aligned} x_1 &= y_1 \\ x_2 &= y_2 - x_1 = y_2 - y_1 \\ x_3 &= y_3 - (x_1 + x_2) = y_3 - y_2 \\ &\vdots \\ x_n &= y_n - (x_1 + \dots + x_{n-1}) = y_n - y_{n-1}, \end{aligned}$$

es decir,

$$x_k = y_k - y_{k-1}, \quad 1 \leq k \leq n.$$

A partir de 6.4 se obtiene que el valor absoluto del determinante de la matriz jacobiana es

$$|J(x_1, \dots, x_n)| = \begin{vmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 & 0 \\ 1 & 1 & 1 & \dots & 1 & 1 \end{vmatrix} = 1.$$

Además, teniendo en cuenta que las variables aleatorias \mathbf{x}_i son independientes y por lo tanto se cumple que

$$f_x(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n),$$

aplicando la ecuación 6.3 se obtiene que

$$f_y(y_1, \dots, y_n) = f_1(y_1) f_2(y_2 - y_1) \dots f_n(y_n - y_{n-1}).$$

6.1.2. Experimentos independientes y ensayos repetidos

Considérese el experimento combinado

$$S^n = S_1 \times \dots \times S_n$$

donde las variables aleatorias \mathbf{x}_i dependen únicamente de las salidas ζ_i de S_i , es decir

$$\mathbf{x}_i(\zeta_1 \dots \zeta_i \dots \zeta_n) = \mathbf{x}_i(\zeta_i), \quad i = 1, \dots, n.$$

Si los experimentos S_i son independientes, las variables aleatorias \mathbf{x}_i son independientes.

Como caso particular de interés, supóngase que \mathbf{x} es una variable aleatoria definida en un experimento S , y el experimento se repite n veces generando el experimento $S^n = S \times \dots \times S$. En este experimento se definen las variables aleatorias \mathbf{x}_i tal que

$$\mathbf{x}_i(\zeta_1 \dots \zeta_i \dots \zeta_n) = \mathbf{x}(\zeta_i), \quad i = 1, \dots, n. \quad (6.5)$$

Por lo tanto, la distribución $F_i(x_i)$ de \mathbf{x}_i es $F(x)$, la distribución de la variable aleatoria \mathbf{x} . Como conclusión, si un experimento se repite n veces, las variables aleatorias \mathbf{x}_i definidas como en 6.5 son independientes y tienen la misma distribución $F_x(x)$. Estas variables aleatorias se denominan *i.i.d.* (independientes e idénticamente distribuidas).

Estadísticos de orden Los estadísticos de orden de las variables aleatorias \mathbf{x}_i son n variables aleatorias \mathbf{y}_k definidas de la siguiente forma: para determinada salida ζ , las variables aleatorias \mathbf{x}_i toman valores $\mathbf{x}_i(\zeta)$. Ordenando esos valores de forma creciente, se obtiene la secuencia

$$\mathbf{x}_{r_1}(\zeta) \leq \dots \leq \mathbf{x}_{r_k}(\zeta) \leq \dots \leq \mathbf{x}_{r_n}(\zeta)$$

y se define la variable aleatoria \mathbf{y}_k tal que

$$\mathbf{y}_1 = \mathbf{x}_{r_1}(\zeta) \leq \dots \leq \mathbf{y}_k = \mathbf{x}_{r_k}(\zeta) \leq \dots \leq \mathbf{y}_n = \mathbf{x}_{r_n}(\zeta). \quad (6.6)$$

Se demostrará que la densidad $f_k(y)$ del k -ésimo estadístico \mathbf{y}_k es

$$f_k(y) = \frac{n!}{(k-1)!(n-k)!} F_x^{k-1}(y) [1 - F_x(y)]^{n-k} f_x(y). \quad (6.7)$$

donde $F_x(x)$ y $f_x(x)$ son la distribución y la densidad de las variables aleatorias i.i.d. \mathbf{x}_i .

Para demostrarlo, se parte observando que el evento $B = \{y < \mathbf{y}_k \leq y + dy\}$ ocurre si exactamente $k-1$ variables aleatorias \mathbf{x}_i toman un valor menor a y y una toma un valor en el intervalo $(y, y+dy)$. Además, se cumple que

$$f_k(y)dy = P\{y < \mathbf{y}_k \leq y + dy\}.$$

En el experimento original S , los eventos

$$A_1 = \{\mathbf{x} \leq y\} \quad A_2 = \{y < \mathbf{x} \leq y + dy\} \quad A_3 = \{\mathbf{x} > y + dy\}$$

forman una partición y

$$P(A_1) = F_x(y) \quad P(A_2) = f_x(y)dy \quad P(A_3) = 1 - F_x(y).$$

En el experimento S^n , el evento B ocurre si A_1 ocurre $k-1$ veces, A_2 ocurre una vez y A_3 ocurre $n-k$ veces. Sustituyendo estos resultados en la ecuación 2.6 se llega al resultado que se quería demostrar.

Notar en particular que

$$f_1(y) = n[1 - F_x(y)]^{n-1} f_x(y), \quad f_n(y) = nF_x^{n-1}(y) f_x(y).$$

Como ejemplo, considérese el caso en que las variables aleatorias \mathbf{x}_i son uniformes en $(0, \theta)$. De esa forma,

$$f_x(x) = \begin{cases} 1/\theta, & 0 < x < \theta \\ 0, & \text{en otro caso} \end{cases} \quad \text{y} \quad F_x(x) = \begin{cases} 0, & x < 0 \\ x/\theta, & 0 < x < \theta \\ 1, & x > \theta \end{cases}.$$

En este caso, usando la ecuación 6.7 se concluye que

$$f_1(y) = \frac{n}{\theta} \left(1 - \frac{y}{\theta}\right)^{n-1}, \quad f_n(y) = \frac{ny^{n-1}}{\theta^n}. \quad (6.8)$$

Estos resultados son la densidad de probabilidad del mínimo y el máximo respectivamente de n variables aleatorias i.i.d. uniformes.

6.1.3. Media y covarianza

Extendiendo la ecuación 5.47 para n variables aleatorias, se concluye que la media de $g(\mathbf{x}_1, \dots, \mathbf{x}_n)$ es

$$E\{g(\mathbf{x}_1, \dots, \mathbf{x}_n)\} = \int_{-\infty}^{-\infty} \cdots \int_{-\infty}^{-\infty} g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \dots dx_n.$$

En el caso en que las variables aleatorias $\mathbf{z}_i = \mathbf{x}_i + j\mathbf{y}_i$ son complejas, la media de $g(\mathbf{z}_1, \dots, \mathbf{z}_n)$ es

$$\int_{-\infty}^{-\infty} \cdots \int_{-\infty}^{-\infty} g(z_1, \dots, z_n) f(x_1, \dots, x_n, y_1, \dots, y_n) dx_1 \dots dx_n dy_1 \dots dy_n.$$

A partir de estas consideraciones, se deduce que la media es lineal,

$$E\{a_1g_1(\mathbf{X}) + \dots + a_ng_n(\mathbf{X})\} = a_1E\{g_1(\mathbf{X})\} + \dots + a_nE\{g_n(\mathbf{X})\},$$

para cualquier vector \mathbf{X} real o complejo.

Correlación y covarianza La covarianza de dos variables aleatorias reales está definida por la ecuación 5.49. La covarianza C_{ij} entre dos variables aleatorias complejas \mathbf{x}_i y \mathbf{x}_j se define como

$$C_{ij} = E\{(\mathbf{x}_i - \eta_i)(\mathbf{x}_j^* - \eta_j^*)\} = E\{\mathbf{x}_i \mathbf{x}_j^*\} - E\{\mathbf{x}_i\} E\{\mathbf{x}_j^*\},$$

y la varianza de la variable aleatoria compleja \mathbf{x}_i se define como

$$\sigma_i^2 = C_{ii} = E\{(\mathbf{x}_i - \eta_i)(\mathbf{x}_i^* - \eta_i^*)\} = E\{|\mathbf{x}_i - \eta_i|^2\} = E\{|\mathbf{x}_i|^2\} - |E\{\mathbf{x}_i\}|^2.$$

Se dice que las variables aleatorias \mathbf{x}_i son *no correlacionadas* si $C_{ij} = 0$ para todo $i \neq j$. En ese caso se cumple que si

$$\mathbf{x} = \mathbf{x}_1 + \dots + \mathbf{x}_n, \quad \text{entonces} \quad \sigma_x^2 = \sigma_1^2 + \dots + \sigma_n^2, \quad (6.9)$$

como se mencionó en la sección 5.4.2.

Ejemplo 7-5: media y varianza muestral La *media muestral* y la *varianza muestral* de las variables aleatorias \mathbf{x}_i se definen como las variables aleatorias

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i, \quad \bar{\mathbf{v}} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2 \quad (6.10)$$

respectivamente. Se demostrará que si las variables aleatorias \mathbf{x}_i son no correlacionadas, todas con media $E\{\mathbf{x}_i\} = \eta$ y varianza $\sigma_i^2 = \sigma^2$, la media y la varianza de la media muestral es

$$E\{\bar{\mathbf{x}}\} = \eta, \quad \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \quad (6.11)$$

y la media y la varianza de la varianza muestral es

$$E\{\bar{\mathbf{v}}\} = \sigma^2, \quad \sigma_{\bar{v}}^2 = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right). \quad (6.12)$$

La media de la media muestral se obtiene directamente a partir de la linealidad de la esperanza,

$$E\{\bar{\mathbf{x}}\} = E\left\{ \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \right\} = \frac{1}{n} \sum_{i=1}^n E\{\mathbf{x}_i\} = \frac{1}{n} \sum_{i=1}^n \eta = \frac{1}{n} (n\eta) = \eta.$$

Para obtener la varianza, se observa que para cualquier constante a se cumple que

$$\begin{aligned} \sigma_{ax}^2 &\stackrel{(a)}{=} E\{(a\mathbf{x} - E\{a\mathbf{x}\})^2\} \\ &\stackrel{(b)}{=} E\{(a\mathbf{x} - aE\{\mathbf{x}\})^2\} \\ &= E\{a^2(\mathbf{x} - E\{\mathbf{x}\})^2\} \\ &\stackrel{(c)}{=} a^2 E\{(\mathbf{x} - E\{\mathbf{x}\})^2\} \\ &= a^2 \sigma_x^2, \end{aligned}$$

donde en (a) se usó la definición de varianza (ecuación 4.9) y en (b) y (c) la propiedad de linealidad de la esperanza. Combinando este resultado con el de la ecuación 6.9, se tiene que

$$\sigma_{\bar{x}}^2 = \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} (n\sigma^2) = \frac{\sigma^2}{n}.$$

A continuación se obtendrá el mismo resultado sin emplear las propiedades de la varianza. Partiendo de la definición de varianza de la ecuación 4.10, se observa que

$$\begin{aligned} \sigma_{\bar{x}}^2 &= E\{\bar{x}^2\} - E^2\{\bar{x}\} \\ &\stackrel{(a)}{=} E\left\{ \left(\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \right)^2 \right\} - \eta^2 \\ &= \frac{1}{n^2} E\left\{ \left(\sum_{i=1}^n \mathbf{x}_i \right)^2 \right\} - \eta^2, \end{aligned} \quad (6.13)$$

donde en (a) se empleó que $E\{\bar{\mathbf{x}}\} = \eta$, resultado obtenido antes. Observando que

$$\left(\sum_{i=1}^n \mathbf{x}_i \right)^2 = \sum_{i=1}^n \mathbf{x}_i^2 + \sum_{i=1}^n \sum_{j=1, j \neq i}^n \mathbf{x}_i \mathbf{x}_j$$

se tiene que

$$\begin{aligned} E\left\{ \left(\sum_{i=1}^n \mathbf{x}_i \right)^2 \right\} &\stackrel{(a)}{=} \sum_{i=1}^n E\{\mathbf{x}_i^2\} + \sum_{i=1}^n \sum_{j=1, j \neq i}^n E\{\mathbf{x}_i \mathbf{x}_j\} \\ &\stackrel{(b)}{=} \sum_{i=1}^n E\{\mathbf{x}_i^2\} + \sum_{i=1}^n \sum_{j=1, j \neq i}^n E\{\mathbf{x}_i\} E\{\mathbf{x}_j\} \\ &\stackrel{(c)}{=} \sum_{i=1}^n (\sigma^2 + \eta^2) + \sum_{i=1}^n \sum_{j=1, j \neq i}^n \eta^2 \\ &\stackrel{(d)}{=} n(\sigma^2 + \eta^2) + n(n-1)\eta^2 \\ &= n\sigma^2 + n^2\eta^2, \end{aligned}$$

donde en (a) se empleó la linealidad de la esperanza, en (b) que las variables \mathbf{x}_i son no correlacionadas y por lo tanto $E\{\mathbf{x}_i \mathbf{x}_j\} = E\{\mathbf{x}_i\}E\{\mathbf{x}_j\}$ si $i \neq j$, en (c) el resultado de la ecuación 4.15 y que todas las variables aleatorias \mathbf{x}_i tienen la misma media y varianza, y en (d) que la primer sumatoria tiene n sumandos y la segunda doble sumatoria tiene $n(n - 1)$ sumandos. Sustituyendo este resultado en la ecuación 6.13, se obtiene que

$$\sigma_{\bar{x}}^2 = \frac{n\sigma^2 + n^2\eta^2}{n^2} - \eta^2 = \frac{\sigma^2}{n},$$

que es lo que se quería demostrar.

Para calcular la media de la varianza muestral,

$$E\{\bar{v}\} = \frac{1}{n-1} \sum_{i=1}^n E\{(\mathbf{x}_i - \bar{\mathbf{x}})^2\}$$

se observa que

$$\begin{aligned} E\{(\mathbf{x}_i - \bar{\mathbf{x}})^2\} &= E\{[(\mathbf{x}_i - \eta) - (\bar{\mathbf{x}} - \eta)]^2\} \\ &= E\{(\mathbf{x}_i - \eta)^2 + (\bar{\mathbf{x}} - \eta)^2 - 2(\mathbf{x}_i - \eta)(\bar{\mathbf{x}} - \eta)\} \\ &= E\{(\mathbf{x}_i - \eta)^2\} + E\{(\bar{\mathbf{x}} - \eta)^2\} - 2E\{(\mathbf{x}_i - \eta)(\bar{\mathbf{x}} - \eta)\} \end{aligned}$$

El primer sumando en la última ecuación es σ^2 y el segundo es $\sigma_{\bar{x}}^2 = \sigma^2/n$, calculado antes. El tercer sumando es

$$\begin{aligned} E\{(\mathbf{x}_i - \eta)(\bar{\mathbf{x}} - \eta)\} &= E\left\{(\mathbf{x}_i - \eta)\left(\frac{1}{n} \sum_{j=1}^n \mathbf{x}_j - \eta\right)\right\} \\ &= \frac{1}{n} E\left\{(\mathbf{x}_i - \eta) \sum_{j=1}^n (\mathbf{x}_j - \eta)\right\} \\ &= \frac{1}{n} \sum_{j=1}^n E\{(\mathbf{x}_i - \eta)(\mathbf{x}_j - \eta)\} \\ &\stackrel{(a)}{=} \frac{1}{n} E\{(\mathbf{x}_i - \eta)(\mathbf{x}_i - \eta)\} \\ &= \frac{\sigma^2}{n}, \end{aligned}$$

donde en (a) se empleó que las variables \mathbf{x}_i son no correlacionadas por hipótesis y por lo tanto $E\{(\mathbf{x}_i - \eta)(\mathbf{x}_j - \eta)\} = 0$ si $i \neq j$. Por lo tanto,

$$E\{(\mathbf{x}_i - \bar{\mathbf{x}})^2\} = \sigma^2 + \frac{\sigma^2}{n} - 2\frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2.$$

Finalmente,

$$E\{\bar{v}\} = \frac{1}{n-1} \sum_{i=1}^n E\{(\mathbf{x}_i - \bar{\mathbf{x}})^2\} = \frac{1}{n-1} \sum_{i=1}^n \frac{n-1}{n} \sigma^2 = \frac{n}{n-1} \frac{n-1}{n} \sigma^2 = \sigma^2.$$

Si las variables aleatorias \mathbf{x}_i son independientes e idénticamente distribuidas con momento central cuarto $\mu_4 = E\{|\mathbf{x}_i - \eta|^4\}$, se cumple que la varianza de la varianza muestral es

$$\sigma_{\bar{v}}^2 = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right). \quad (6.14)$$

La demostración de esto es considerablemente tediosa y en [1] está como ejercicio. La demostración desarrollada aquí está basada en [13]¹. Se comenzará mostrando que la varianza muestral puede escribirse de las siguientes formas alternativas,

$$\bar{v} = \frac{1}{n-1} \left(\sum_{i=1}^n \mathbf{x}_i^2 - n\bar{\mathbf{x}}^2 \right), \quad \bar{v} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n \frac{(\mathbf{x}_i - \mathbf{x}_j)^2}{2}.$$

¹Otra demostración similar puede encontrarse en <http://mathworld.wolfram.com/SampleVarianceDistribution.html>,

Partiendo de la definición de la varianza muestral, para mostrar la primer identidad se ve que

$$\begin{aligned}
\bar{v} &= \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2 \\
&= \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i^2 + \bar{\mathbf{x}}^2 - 2\mathbf{x}_i \bar{\mathbf{x}}) \\
&= \frac{1}{n-1} \left(\sum_{i=1}^n \mathbf{x}_i^2 + \sum_{i=1}^n \bar{\mathbf{x}}^2 - 2\bar{\mathbf{x}} \sum_{i=1}^n \mathbf{x}_i \right) \\
&= \frac{1}{n-1} \left(\sum_{i=1}^n \mathbf{x}_i^2 + n\bar{\mathbf{x}}^2 - 2\bar{\mathbf{x}}(n\bar{\mathbf{x}}) \right) \\
&= \frac{1}{n-1} \left(\sum_{i=1}^n \mathbf{x}_i^2 - n\bar{\mathbf{x}}^2 \right).
\end{aligned}$$

Para demostrar la segunda identidad, se desarrolla el lado derecho de la siguiente forma,

$$\begin{aligned}
\frac{1}{n(n-1)} \sum_{i,j} \frac{(\mathbf{x}_i - \mathbf{x}_j)^2}{2} &= \frac{1}{2n(n-1)} \sum_{i,j} (\mathbf{x}_i^2 + \mathbf{x}_j^2 - 2\mathbf{x}_i \mathbf{x}_j) \\
&= \frac{1}{2n(n-1)} \left(\sum_{i,j} \mathbf{x}_i^2 + \sum_{i,j} \mathbf{x}_j^2 - 2 \sum_{i,j} \mathbf{x}_i \mathbf{x}_j \right) \\
&= \frac{1}{2n(n-1)} \left(n \sum_{i=1}^n \mathbf{x}_i^2 + n \sum_{j=1}^n \mathbf{x}_j^2 - 2 \sum_{i=1}^n \mathbf{x}_i \sum_{j=1}^n \mathbf{x}_j \right) \\
&= \frac{1}{2n(n-1)} \left[2n \sum_{i=1}^n \mathbf{x}_i^2 - 2 \left(\sum_{i=1}^n \mathbf{x}_i \right)^2 \right] \\
&= \frac{1}{(n-1)} \left[\sum_{i=1}^n \mathbf{x}_i^2 - \frac{1}{n} (n\bar{\mathbf{x}})^2 \right] \\
&= \frac{1}{(n-1)} \left(\sum_{i=1}^n \mathbf{x}_i^2 - n\bar{\mathbf{x}}^2 \right) = \bar{v}.
\end{aligned}$$

Para calcular la varianza $\sigma_{\bar{v}}^2$ de la varianza muestral, se utilizará esta expresión de la varianza muestral. De la definición de varianza, se tiene que

$$\sigma_{\bar{v}}^2 = E\{(\bar{v} - E\{\bar{v}\})^2\}.$$

Teniendo en cuenta que $E\{\bar{v}\} = \sigma^2$, como se demostró previamente, hay que calcular

$$\begin{aligned}
\sigma_{\bar{v}}^2 &= E \left\{ \left[\frac{1}{n(n-1)} \sum_{i,j} \frac{(\mathbf{x}_i - \mathbf{x}_j)^2}{2} - \sigma^2 \right] \left[\frac{1}{n(n-1)} \sum_{k,l} \frac{(\mathbf{x}_k - \mathbf{x}_l)^2}{2} - \sigma^2 \right] \right\} \\
&= E \left\{ \frac{1}{n^2(n-1)^2} \sum_{i,j} \frac{(\mathbf{x}_i - \mathbf{x}_j)^2}{2} \sum_{k,l} \frac{(\mathbf{x}_k - \mathbf{x}_l)^2}{2} - \frac{2\sigma^2}{n(n-1)} \sum_{i,j} \frac{(\mathbf{x}_i - \mathbf{x}_j)^2}{2} + \sigma^4 \right\} \\
&= \frac{1}{4n^2(n-1)^2} \sum_{i,j} \sum_{k,l} E\{(\mathbf{x}_i - \mathbf{x}_j)^2(\mathbf{x}_k - \mathbf{x}_l)^2\} - \frac{\sigma^2}{n(n-1)} \sum_{i,j} E\{(\mathbf{x}_i - \mathbf{x}_j)^2\} + \sigma^4. \quad (6.15)
\end{aligned}$$

Comenzando con el segundo sumando de esta expresión, se observa que

$$\begin{aligned}
E\{(\mathbf{x}_i - \mathbf{x}_j)^2\} &= E\{[(\mathbf{x}_i - \eta) - (\mathbf{x}_j - \eta)]^2\} \\
&= E\{(\mathbf{x}_i - \eta)^2\} + E\{(\mathbf{x}_j - \eta)^2\} - 2E\{(\mathbf{x}_i - \eta)(\mathbf{x}_j - \eta)\} \\
&\stackrel{(a)}{=} 2\sigma^2,
\end{aligned}$$

donde en (a) se consideró que \mathbf{x}_i y \mathbf{x}_j son independientes y por lo tanto $C_{ij} = E\{(\mathbf{x}_i - \eta)(\mathbf{x}_j - \eta)\} = 0$. Esta igualdad es válida si $i \neq j$. Si $i = j$, el resultado es 0. Además, la sumatoria en (i, j) con $i \neq j$ tiene $n(n - 1)$ sumandos. De esta forma, el segundo sumando de 6.15 es

$$\frac{\sigma^2}{n(n-1)} \sum_{i,j} E\{(\mathbf{x}_i - \mathbf{x}_j)^2\} = \frac{\sigma^2}{n(n-1)} 2\sigma^2 n(n-1) = 2\sigma^4. \quad (6.16)$$

Para calcular el primer sumando de 6.15, hay que calcular

$$E\{(\mathbf{x}_i - \mathbf{x}_j)^2(\mathbf{x}_k - \mathbf{x}_l)^2\}.$$

Claramente, esta expresión es nula si $i = j$ o $k = l$. Asimismo, toma valores distintos dependiendo de la cantidad de elementos de $\{i, j\} \cap \{k, l\}$, que puede ser 0, 1 o 2. Se discuten a continuación cada uno de estos casos.

1. $|\{i, j\} \cap \{k, l\}| = 0$, es decir, i, j, k, l son todos distintos. Como las variables aleatorias son independientes, se cumple que

$$E\{(\mathbf{x}_i - \mathbf{x}_j)^2(\mathbf{x}_k - \mathbf{x}_l)^2\} = E\{(\mathbf{x}_i - \mathbf{x}_j)^2\} E\{(\mathbf{x}_k - \mathbf{x}_l)^2\} = 4\sigma^4,$$

donde en la última igualdad se empleó el resultado obtenido antes.

2. $|\{i, j\} \cap \{k, l\}| = 1$, es decir, i o j es igual a k o l . En este caso, hay que hacer todas las cuentas. Expandiendo el producto dentro de la esperanza, se tiene que

$$\begin{aligned} (\mathbf{x}_i^2 + \mathbf{x}_j^2 - 2\mathbf{x}_i\mathbf{x}_j)(\mathbf{x}_k^2 + \mathbf{x}_l^2 - 2\mathbf{x}_k\mathbf{x}_l) &= \mathbf{x}_i^2\mathbf{x}_k^2 + \mathbf{x}_i^2\mathbf{x}_l^2 + \mathbf{x}_j^2\mathbf{x}_k^2 + \mathbf{x}_j^2\mathbf{x}_l^2 \\ &\quad - 2\mathbf{x}_i^2\mathbf{x}_k\mathbf{x}_l - 2\mathbf{x}_j^2\mathbf{x}_k\mathbf{x}_l \\ &\quad - 2\mathbf{x}_k^2\mathbf{x}_i\mathbf{x}_j - 2\mathbf{x}_l^2\mathbf{x}_i\mathbf{x}_j \\ &\quad + 4\mathbf{x}_i\mathbf{x}_j\mathbf{x}_k\mathbf{x}_l. \end{aligned}$$

Notar que esta expresión es simétrica en todas las variables y por lo tanto, en el caso en que $|\{i, j\} \cap \{k, l\}| = 1$, la expresión es equivalente sin importar el par de variables que sean iguales. Supóngase por ejemplo que $k = i$ y $j \neq l$. En ese caso,

$$\begin{aligned} (\mathbf{x}_i - \mathbf{x}_j)^2(\mathbf{x}_i - \mathbf{x}_l)^2 &= \mathbf{x}_i^4 + \mathbf{x}_i^2\mathbf{x}_l^2 + \mathbf{x}_j^2\mathbf{x}_i^2 + \mathbf{x}_j^2\mathbf{x}_l^2 \\ &\quad - 2\mathbf{x}_i^3\mathbf{x}_l - 2\mathbf{x}_j^2\mathbf{x}_i\mathbf{x}_l - 2\mathbf{x}_i^3\mathbf{x}_j - 2\mathbf{x}_l^2\mathbf{x}_i\mathbf{x}_j \\ &\quad + 4\mathbf{x}_i^2\mathbf{x}_j\mathbf{x}_l. \end{aligned}$$

Tomando esperanza y considerando que las variables aleatorias son independientes y que $E\{\mathbf{x}_i^n\} = m_n$ es el momento n -ésimo de \mathbf{x}_i , se tiene que

$$\begin{aligned} E\{(\mathbf{x}_i - \mathbf{x}_j)^2(\mathbf{x}_i - \mathbf{x}_l)^2\} &= m_4 + 3m_2m_2 - 4m_3m \\ &\stackrel{(a)}{=} (\eta^4 + 6\sigma^2\eta^2 + 4\mu_3\eta + \mu_4) + 3(\sigma^2 + \eta^2)^2 \\ &\quad - 4(\eta^3 + 3\sigma^2\eta + \mu_3)\eta \\ &= (\eta^4 + 6\sigma^2\eta^2 + 4\mu_3\eta + \mu_4) \\ &\quad + (3\sigma^4 + 3\eta^4 + 6\sigma^2\eta^2) \\ &\quad - (4\eta^4 + 12\sigma^2\eta^2 + 4\mu_3\eta) \\ &= \mu_4 + 3\sigma^4, \end{aligned}$$

donde en (a) se usó la ecuación 4.15, que expresa los momentos en función de los momentos centrales.

3. $|\{i, j\} \cap \{k, l\}| = 2$, es decir, $i = k$ y $j = l$ o $i = l$ y $j = k$. En este caso, se tiene que

$$\begin{aligned} E\{(\mathbf{x}_i - \mathbf{x}_j)^4\} &= E\{[(\mathbf{x}_i - \eta) + (\mathbf{x}_j - \eta)]^4\} \\ &= E\{(\mathbf{x}_i - \eta)^4\} + E\{(\mathbf{x}_j - \eta)^4\} \\ &\quad - 4E\{(\mathbf{x}_i - \eta)^3(\mathbf{x}_j - \eta)\} - 4E\{(\mathbf{x}_i - \eta)(\mathbf{x}_j - \eta)^3\} \\ &\quad + 6E\{(\mathbf{x}_i - \eta)^2(\mathbf{x}_j - \eta)^2\} \\ &\stackrel{(a)}{=} 2\mu_4 + 6\sigma^4, \end{aligned}$$

donde en (a) se consideró que \mathbf{x}_i y \mathbf{x}_j son independientes y por lo tanto, la esperanza del producto es el producto de las esperanzas, y también que $E\{\mathbf{x}_i - \eta\} = 0$.

Estos son los tres tipos de sumandos distintos en la sumatoria cuádruple de índices i, j, k, l , todos variando de 1 a n con $i \neq j$ y $k \neq l$. Por lo tanto, para obtener el resultado final, hay que contar la cantidad de sumandos que tienen $|\{i, j\} \cap \{k, l\}| = 0, 1$ y 2 en dicha sumatoria. Para hacerlo, notar que por ejemplo, si $n = 4$, la sumatoria se desarrolla como

$$\{[(\mathbf{x}_1 - \mathbf{x}_2)^2 + (\mathbf{x}_1 - \mathbf{x}_3)^2 + (\mathbf{x}_1 - \mathbf{x}_4)^2] + [(\mathbf{x}_2 - \mathbf{x}_1)^2 + (\mathbf{x}_2 - \mathbf{x}_3)^2 + (\mathbf{x}_2 - \mathbf{x}_4)^2] + \\ [(\mathbf{x}_3 - \mathbf{x}_1)^2 + (\mathbf{x}_3 - \mathbf{x}_2)^2 + (\mathbf{x}_3 - \mathbf{x}_4)^2] + [(\mathbf{x}_4 - \mathbf{x}_1)^2 + (\mathbf{x}_4 - \mathbf{x}_2)^2 + (\mathbf{x}_4 - \mathbf{x}_3)^2]\}^2.$$

Se observa que hay $n(n - 1)$ términos de la forma $(\mathbf{x}_i - \mathbf{x}_j)^2$ entre las llaves, ya que hay n términos agrupados por corchetes rectos, que se diferencian por el valor del primer índice, y $(n - 1)$ términos $(\mathbf{x}_i - \mathbf{x}_j)^2$ dentro de los corchetes rectos, diferenciados por el valor del segundo índice. Por lo tanto, al desarrollar el cuadrado se obtienen $n^2(n - 1)^2$ términos de la forma $(\mathbf{x}_i - \mathbf{x}_j)^2(\mathbf{x}_k - \mathbf{x}_l)^2$. A continuación se contarán la cantidad de términos con $|\{i, j\} \cap \{k, l\}| = 0, 1$ y 2.

1. $|\{i, j\} \cap \{k, l\}| = 0$, es decir, todos los subíndices distintos.

- La cantidad de términos de la forma $(\mathbf{x}_i - \mathbf{x}_j)^2$ que tienen cierto valor específico $(i, j) = (r, s)$ o $(i, j) = (s, r)$ en la expresión entre llaves es 2.
- La cantidad de términos con subíndices (k, l) distintos a r y a s entre llaves es $(n - 2)(n - 3)$. Esto es porque no hay que considerar 2 de los n términos agrupados entre corchetes (los que tienen primer índice r y s) y no hay que considerar 2 términos de la forma $(\mathbf{x}_i - \mathbf{x}_j)^2$ de los $n - 1$ dentro de los corchetes (los que tienen como segundo índice r y s).
- La cantidad de pares (r, s) posibles distintos es

$$\binom{n}{2} = \frac{n!}{2!(n-2)!} = \frac{n(n-1)}{2}.$$

Se concluye que la cantidad de términos de la forma $(\mathbf{x}_i - \mathbf{x}_j)^2(\mathbf{x}_k - \mathbf{x}_l)^2$ con todos los subíndices distintos es

$$2(n - 2)(n - 3) \frac{n(n - 1)}{2} = n(n - 1)(n - 2)(n - 3).$$

Este resultado podría haberse obtenido de forma mas directa observando que los términos con todos los subíndices distintos aparecen solo una vez y en todos los ordenes posibles vez para un grupo de cuatro valores de subíndices específico. Por lo tanto, la cantidad de términos son las permutaciones de 4 elementos tomados de un conjunto de n , es decir,

$$P(n, 4) = \frac{n!}{(n - 4)!} = n(n - 1)(n - 2)(n - 3).$$

2. $|\{i, j\} \cap \{k, l\}| = 1$

- La cantidad de términos de la forma $(\mathbf{x}_i - \mathbf{x}_j)^2$ que tienen cierto valor específico $(i, j) = (r, s)$ o $(i, j) = (s, r)$ en la expresión entre llaves es 2.

- Hay que calcular la cantidad de términos con uno de los subíndices (k, l) igual a r o s . Para esto se observa que la cantidad de términos con $k = r$ y $l \neq s$ o $l = r$ y $k \neq s$ es

$$(n-1) + (n-1) - 2 = 2(n-2),$$

ya que en uno de los factores agrupados entre corchetes, los $n-1$ términos entre paréntesis tienen como primer subíndice $k = r$, y en el resto de los $n-1$ factores agrupados entre corchetes hay 1 término entre paréntesis cuyo segundo subíndice es $l = r$. Además, entre todos éstos, hay 1 término con $(k, l) = (r, s)$ y 1 término con $(k, l) = (s, r)$ que no deben ser considerados. De forma análoga, también hay $2(n-2)$ términos con $k = s$ y $l \neq r$ o $l = s$ y $k \neq r$. Por lo tanto, el total es

$$4(n-2).$$

- La cantidad de pares (r, s) posibles distintos es $\binom{n}{2} = n(n-1)/2$.

Se concluye que la cantidad de términos de la forma $(\mathbf{x}_i - \mathbf{x}_j)^2(\mathbf{x}_k - \mathbf{x}_l)^2$ con $|\{i, j\} \cap \{k, l\}| = 1$ es

$$2 \cdot 4(n-2) \frac{n(n-1)}{2} = 4n(n-1)(n-2)$$

3. $|\{i, j\} \cap \{k, l\}| = 2$.

- Hay 4 términos de la forma

$$(\mathbf{x}_i - \mathbf{x}_j)^2(\mathbf{x}_k - \mathbf{x}_l)^2 = (\mathbf{x}_r - \mathbf{x}_s)^4$$

para un par de valores (r, s) específico. Estos son los siguientes:

(i, j)	(r, s)	(r, s)	(s, r)	(s, r)
(k, l)	(r, s)	(s, r)	(r, s)	(s, r)

- La cantidad de pares (r, s) posibles distintos es $\binom{n}{2} = n(n-1)/2$.

Por lo tanto, la cantidad total de términos es

$$4 \frac{n(n-1)}{2} = 2n(n-1).$$

Para verificar estos resultados, se puede comprobar que la suma de la cantidad de términos en cada situación debería coincidir con la cantidad total de términos de la sumatoria cuádruple, que como ya se mencionó es $n^2(n-1)^2$. Haciendo esto se ve que

$$\begin{aligned} n(n-1)(n-2)(n-3) + 4n(n-1)(n-2) + 2n(n-1) &= n(n-1) \{ (n-2) [(n-3) + 4] + 2 \} \\ &= n(n-1) [(n-2)(n+1) + 2] \\ &= n(n-1)(n^2 - n) \\ &= n^2(n-1)^2. \end{aligned}$$

Como resumen, en el término

$$\sum_{i, j} \sum_{k, l} E\{(\mathbf{x}_i - \mathbf{x}_j)^2(\mathbf{x}_k - \mathbf{x}_l)^2\}$$

de la ecuación 6.15 se obtuvieron los resultados que se muestran en la siguiente tabla:

$ \{i, j\} \cap \{k, l\} $	$E\{(\mathbf{x}_i - \mathbf{x}_j)^2(\mathbf{x}_k - \mathbf{x}_l)^2\}$	Número de sumandos
0	$4\sigma^4$	$n(n-1)(n-2)(n-3)$
1	$\mu_4 + 3\sigma^4$	$4n(n-1)(n-2)$
2	$2(\mu_4 + 3\sigma^4)$	$2n(n-1)$

Por lo tanto, se tiene que

$$\begin{aligned}
& \frac{1}{4n^2(n-1)^2} \sum_{i,j} \sum_{k,l} E\{(\mathbf{x}_i - \mathbf{x}_j)^2(\mathbf{x}_k - \mathbf{x}_l)^2\} \\
&= \frac{1}{4n^2(n-1)^2} [n(n-1)(n-2)(n-3)4\sigma^4 \\
&\quad + 4n(n-1)(n-2)(\mu_4 + 3\sigma^4) + 2n(n-1)2(\mu_4 + 3\sigma^4)] \\
&= \frac{1}{n(n-1)} [(n-2)(n-3)\sigma^4 + (n-2)(\mu_4 + 3\sigma^4) + (\mu_4 + 3\sigma^4)] \\
&= \frac{1}{n(n-1)} [(n-2)(n\sigma^4 + \mu_4) + (\mu_4 + 3\sigma^4)] \\
&= \frac{1}{n(n-1)} [(n^2 - 2n + 3)\sigma^4 + (n-1)\mu_4]. \tag{6.17}
\end{aligned}$$

Sustituyendo los resultados de las ecuaciones 6.16 y 6.17 en la ecuación 6.15 se obtiene que

$$\begin{aligned}
\sigma_{\bar{v}}^2 &= \frac{1}{n(n-1)} [(n^2 - 2n + 3)\sigma^4 + (n-1)\mu_4] - 2\sigma^4 + \sigma^4 \\
&= \frac{1}{n(n-1)} [(n^2 - 2n + 3)\sigma^4 + (n-1)\mu_4 - n(n-1)\sigma^4] \\
&= \frac{1}{n(n-1)} [(-n+3)\sigma^4 + (n-1)\mu_4] \\
&= \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right),
\end{aligned}$$

que es lo que se quería demostrar.

Matrices de correlación y covarianza Las matrices de correlación y covarianza del vector aleatorio $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ son respectivamente

$$R_n = \begin{bmatrix} R_{11} & \cdots & R_{1n} \\ \vdots & \ddots & \vdots \\ R_{n1} & \cdots & R_{nn} \end{bmatrix}, \quad C_n = \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \cdots & C_{nn} \end{bmatrix}$$

donde

$$R_{ij} = E\{\mathbf{x}_i \mathbf{x}_j^*\} = R_{ji}^*, \quad C_{ij} = R_{ij} - \eta_i \eta_j = C_{ji}^*.$$

Notar que

$$R_n = E\{\mathbf{X}^t \mathbf{X}^*\},$$

donde \mathbf{X}^t es el vector traspuesto de \mathbf{X} , resultando en un vector columna. Se verán algunas propiedades de la matriz de correlación R_n y su determinante Δ_n . Las propiedades de la matriz de covarianza C_n son similares por ser la matriz de correlación de las variables aleatorias centradas $\mathbf{x}_i - \eta_i$.

- La matriz R_n es *semidefinida positiva*. Esto significa que para cualquier vector $A = [a_1, \dots, a_n]$ se cumple que

$$Q = \sum_{i,j} a_i a_j^* R_{ij} = A R_n A^H \geq 0, \tag{6.18}$$

donde A^H es el vector complejo conjugado transpuesto de A . Para demostrarlo, se observa que

$$E\{|a_1 \mathbf{x}_1 + \cdots + a_n \mathbf{x}_n|^2\} \geq 0,$$

y además que

$$\begin{aligned}
E\{|a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n|^2\} &= E\{(a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n)(a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n)^*\} \\
&= E\{(a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n)(a_1^*\mathbf{x}_1^* + \cdots + a_n^*\mathbf{x}_n^*)\} \\
&= E\left\{\sum_{i,j} a_i a_j^* \mathbf{x}_i \mathbf{x}_j^*\right\} \\
&= \sum_{i,j} a_i a_j^* E\{\mathbf{x}_i \mathbf{x}_j^*\} \\
&= \sum_{i,j} a_i a_j^* R_{ij} = Q.
\end{aligned}$$

- Si se cumple que Q es estrictamente positivo, es decir $Q > 0$ para cualquier $A \neq 0$, entonces R_n es *definida positiva*.

Definición: independencia lineal

- Las variables aleatorias \mathbf{x}_i se dicen *linealmente independientes* si

$$E\{|a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n|^2\} > 0$$

para todo $A \neq 0$. En ese caso, su matriz de correlación R_n es definida positiva.

- Las variables aleatorias \mathbf{x}_i se dicen *linealmente dependientes* si

$$E\{|a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n|^2\} = 0$$

para algún $A \neq 0$. En ese caso, el valor Q correspondiente es 0 y su matriz R_n es singular, o equivalentemente, no invertible.

A partir de la definición se deduce que si las variables aleatorias \mathbf{x}_i son linealmente independientes, cualquier subconjunto también es linealmente independiente.

El determinante de la matriz de correlación

- El determinante Δ_n de la matriz de correlación es real debido a que $R_{ij} = R_{ji}^*$, es decir, debido a que la matriz R_n es hermítica. Para ver esto, hay que considerar las siguientes propiedades del determinante de una matriz,

$$\det(A) = \det(A^t), \quad \det(A^*) = \det(A)^*,$$

que implican que

$$\det(A^h) = \det(A)^*.$$

Por lo tanto, si A es hermítica, esto es, $A^h = A$, se cumple que

$$\det(A) = \det(A)^*,$$

es decir, el determinante es real.

- El determinante es no negativo,

$$\Delta_n \geq 0,$$

y la igualdad se da si las variables aleatorias \mathbf{x}_i son linealmente dependientes.

- Considérese el caso en que las variables aleatorias \mathbf{x}_i son linealmente independientes. Se cumple que el determinante Δ_n y el determinante de todos los menores principales son estrictamente positivos,

$$\Delta_k > 0, \quad k \leq n. \tag{6.19}$$

Para demostrarlo, se parte observando que la afirmación es cierta para $n = 1$, ya que

$$\Delta_1 = R_{11} = E\{\mathbf{x}_1^2\} > 0.$$

Como las variables aleatorias de cualquier subconjunto de \mathbf{x}_i son linealmente independientes, se puede asumir que la ecuación 6.19 se cumple para $k \leq n-1$, así que solo hay que demostrar que se cumple para $k = n$. Para esto, considérese el sistema de ecuaciones

$$\begin{aligned} R_{11}a_1 + \cdots + R_{1n}a_n &= 1 \\ R_{21}a_1 + \cdots + R_{2n}a_n &= 0 \\ \vdots &\quad \vdots \quad \vdots \\ R_{n1}a_1 + \cdots + R_{nn}a_n &= 0 \end{aligned} \tag{6.20}$$

Se resolverá el sistema para a_1 . Usando la regla de Cramer, se tiene que

$$a_1 = \frac{\Delta a_1}{\Delta_n},$$

donde

$$\begin{aligned} \Delta a_1 &= \begin{vmatrix} 1 & R_{12} & \cdots & R_{1n} \\ 0 & R_{22} & \cdots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & R_{n2} & \cdots & R_{nn} \end{vmatrix} \\ &= 1 \begin{vmatrix} R_{22} & \cdots & R_{2n} \\ \vdots & \ddots & \vdots \\ R_{n2} & \cdots & R_{nn} \end{vmatrix} - 0 \begin{vmatrix} \cdots \\ \vdots \\ \cdots \end{vmatrix} + 0 \begin{vmatrix} \cdots \\ \vdots \\ \cdots \end{vmatrix} + \cdots \pm 0 \begin{vmatrix} \cdots \\ \vdots \\ \cdots \end{vmatrix} \\ &= \Delta_{n-1}, \end{aligned}$$

con Δ_{n-1} el determinante de la matriz de correlación de las variables aleatorias $\mathbf{x}_2, \dots, \mathbf{x}_n$. De esta forma

$$a_1 = \frac{\Delta_{n-1}}{\Delta_n},$$

es un número real. Multiplicando la ecuación j -ésima del sistema de ecuaciones 6.20 por a_j^* se obtiene el sistema

$$\begin{aligned} R_{11}a_1a_1^* + \cdots + R_{1n}a_na_1^* &= a_1^* \\ R_{21}a_1a_2^* + \cdots + R_{2n}a_na_2^* &= 0 \\ \vdots &\quad \vdots \quad \vdots \\ R_{n1}a_1a_n^* + \cdots + R_{nn}a_na_n^* &= 0 \end{aligned}$$

Sumando las n ecuaciones y teniendo en cuenta que a_1 es real, se obtiene que

$$Q = \sum_{i,j} a_i a_j^* R_{ij} = a_1 = \frac{\Delta_{n-1}}{\Delta_n}.$$

Como por hipótesis las variables aleatorias \mathbf{x}_i son linealmente independientes, $Q > 0$. Además, por la hipótesis de inducción, $\Delta_{n-1} > 0$. Se concluye entonces que $\Delta_n > 0$.

- En el caso en que las variables aleatorias \mathbf{x}_i son linealmente dependientes, se cumple que

$$\Delta_n = 0.$$

Para demostrarlo, se observa que por definición de dependencia lineal, se cumple que existe un vector $A \neq 0$ tal que

$$a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n = 0.$$

Multiplicando por \mathbf{x}_i y tomando esperanza, se obtiene el sistema de ecuaciones

$$a_1R_{i1} + \cdots + a_nR_{in} = 0, \quad i = 1, \dots, n.$$

Este es un sistema de ecuaciones homogéneo que tiene además de la solución trivial, el vector $A \neq 0$. Por lo tanto, $\Delta_n = 0^2$.

Finalmente, se observa que

$$\Delta_n \leq R_{11}R_{22} \cdots R_{nn},$$

donde se cumple la igualdad si las variables aleatorias \mathbf{x}_i son mutuamente ortogonales, es decir, si la matriz R_n es diagonal. Esto es por una propiedad del determinante de una matriz que indica que

$$\det(A) \leq \prod_{i=1}^n a_{ii}.$$

6.2. Densidades condicionales, funciones características y normalidad

En el caso de múltiples variables aleatorias, las densidades condicionales conjuntas se definen como en la sección 5.6. La densidad condicional de las variables aleatorias $\mathbf{x}_n, \dots, \mathbf{x}_{k+1}$ dadas $\mathbf{x}_k, \dots, \mathbf{x}_1$ se obtiene de forma análoga a la ecuación 5.78, y es

$$f(x_n, \dots, x_{k+1} | x_k, \dots, x_1) = \frac{f(x_1, \dots, x_k, \dots, x_n)}{f(x_1, \dots, x_k)}. \quad (6.21)$$

La distribución de probabilidad correspondiente, al igual que antes, se obtiene mediante integración,

$$F(x_n, \dots, x_{k+1} | x_k, \dots, x_1) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_{k+1}} f(x_n, \dots, x_{k+1} | x_k, \dots, x_1) dx_{k+1} \cdots dx_n$$

Por ejemplo,

$$f(x_1 | x_2, x_3) = \frac{f(x_1, x_2, x_3)}{f(x_2, x_3)} = \frac{dF(x_1 | x_2, x_3)}{dx_1},$$

donde en la primera igualdad se empleó la ecuación 6.21.

Regla de la cadena La regla de la cadena indica que

$$f(x_1, \dots, x_n) = f(x_n | x_{n-1}, \dots, x_1) f(x_{n-1} | x_{n-2}, \dots, x_1) \cdots f(x_2 | x_1) f(x_1). \quad (6.22)$$

Se deduce a partir de la ecuación 6.21, ya que según dicha ecuación,

$$\begin{aligned} f(x_n, \dots, x_1) &= f(x_n | x_{n-1}, \dots, x_1) f(x_{n-1}, \dots, x_1) \\ f(x_{n-1}, \dots, x_1) &= f(x_{n-1} | x_{n-2}, \dots, x_1) f(x_{n-2}, \dots, x_1) \\ &\vdots \\ f(x_2, x_1) &= f(x_2 | x_1) f(x_1), \end{aligned}$$

y sustituyendo sucesivamente se obtiene la ecuación 6.22.

Para ver como eliminar variables del lado izquierdo o derecho de la barra del condicional, notar que por ejemplo, se cumplen las siguientes identidades

$$\begin{aligned} f(x_1 | x_3) &= \int_{-\infty}^{\infty} f(x_1, x_2 | x_3) dx_2 \\ f(x_1 | x_4) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1 | x_2, x_3, x_4) f(x_2, x_3 | x_4) dx_2 dx_3. \end{aligned}$$

La primera identidad se obtiene directamente de la ecuación 6.1, es decir, para eliminar variables del lado izquierdo alcanza con integrar respecto a las variables que se quiere eliminar, al igual que con una densidad no condicionada. Para que se cumpla la segunda igualdad, por la primera igualdad el integrando

²Ver https://en.wikipedia.org/wiki/System_of_linear_equations#Homogeneous_systems

debería ser $f(x_1, x_2, x_3 | x_4)$, ya que al integrar respecto a x_2 y x_3 , se eliminan esas variables. Para verlo, se observa que

$$\begin{aligned} f(x_1, x_2, x_3, x_4) &= f(x_1 | x_2, x_3, x_4) f(x_2, x_3, x_4) \\ &= f(x_1 | x_2, x_3, x_4) f(x_2, x_3 | x_4) f(x_4). \end{aligned}$$

Despejando $f(x_4)$, se tiene que

$$f(x_1 | x_2, x_3, x_4) f(x_2, x_3 | x_4) = \frac{f(x_1, x_2, x_3, x_4)}{f(x_4)} = f(x_1, x_2, x_3 | x_4),$$

que es lo que se quería demostrar. En todas las igualdades se empleó la ecuación 6.21. Otra forma de razonarlo, es ignorar las variables que están siempre del lado derecho del condicional, que en este caso es x_4 . A partir de la ecuación 6.21, se cumple que

$$f(x_1, x_2, x_3) = f(x_1 | x_2, x_3) f(x_2, x_3),$$

y condicionando respecto a x_4 se obtiene la igualdad

$$f(x_1, x_2, x_3 | x_4) = f(x_1 | x_2, x_3, x_4) f(x_2, x_3 | x_4).$$

Un caso particular importante, es la ecuación de Chapman-Kolmogoroff, que indica que

$$f(x_1 | x_3) = \int_{-\infty}^{\infty} f(x_1 | x_2, x_3) f(x_2 | x_3) dx_2.$$

La interpretación de esta ecuación es que la probabilidad de transición desde un estado 1 a un estado 3 puede obtenerse como la probabilidad de pasar del estado 1 a un estado intermedio 2 y luego pasar del estado 2 al 3 y sumar sobre todos los estados 2 intermedios posibles.

Resumiendo, las reglas para eliminar variables del lado izquierdo o del lado derecho del condicional, son las siguientes:

- para eliminar variables del lado izquierdo del condicional, se integra respecto a ellas.
- para eliminar variables del lado derecho del condicional, se multiplica por su densidad condicional respecto a las variables que se desea mantener y se integra el producto respecto a las variables que se quieren eliminar.

6.2.1. Valores esperados condicionales

La esperanza condicional de la variable aleatoria $g(\mathbf{x}_1, \dots, \mathbf{x}_n)$ dado el evento M es

$$E\{g(\mathbf{x}_1, \dots, \mathbf{x}_n) | M\} = \int_{-\infty}^{-\infty} \dots \int_{-\infty}^{-\infty} g(x_1, \dots, x_n) f(x_1, \dots, x_n | M) dx_1 \dots dx_n.$$

En particular,

$$E\{\mathbf{x}_1 | x_2, \dots, x_n\} = \int_{-\infty}^{-\infty} x_1 f(x_1 | x_2, \dots, x_n) dx_1.$$

Esta es una función de x_2, \dots, x_n , a partir de la cual puede definirse la variable aleatoria $E\{\mathbf{x}_1 | \mathbf{x}_2, \dots, \mathbf{x}_n\}$. De forma análoga a lo realizado en la sección 5.7.1, se cumple que

$$E\{E\{\mathbf{x}_1 | \mathbf{x}_2, \dots, \mathbf{x}_n\}\} = E\{\mathbf{x}_1\},$$

ya que

$$\begin{aligned} E\{E\{\mathbf{x}_1 | \mathbf{x}_2, \dots, \mathbf{x}_n\}\} &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} E\{\mathbf{x}_1 | x_2, \dots, x_n\} f(x_2, \dots, x_n) dx_2 \dots dx_n \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\int_{-\infty}^{-\infty} x_1 f(x_1 | x_2, \dots, x_n) dx_1 \right) \times \\ &\quad f(x_2, \dots, x_n) dx_2 \dots dx_n \end{aligned}$$

Intercambiando el orden de integración, se tiene que

$$\begin{aligned}
E\{E\{\mathbf{x}_1 | \mathbf{x}_2, \dots, \mathbf{x}_n\}\} &= \int_{-\infty}^{-\infty} x_1 \left(\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1 | x_2, \dots, x_n) \times \right. \\
&\quad \left. f(x_2, \dots, x_n) dx_2 \dots dx_n \right) dx_1 \\
&\stackrel{(a)}{=} \int_{-\infty}^{-\infty} x_1 \left(\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_2 \dots dx_n \right) dx_1 \\
&\stackrel{(b)}{=} \int_{-\infty}^{-\infty} x_1 f(x_1) dx_1 \\
&= E\{\mathbf{x}_1\},
\end{aligned}$$

donde en (a) se usó la ecuación 6.21 y en (b) la ecuación 6.1.

Razonando de forma similar, se observa que

$$\begin{aligned}
E\{\mathbf{x}_1 | x_2, x_3\} &= E\{E\{\mathbf{x}_1 | x_2, x_3, \mathbf{x}_4\}\} \\
&= \int_{-\infty}^{\infty} E\{\mathbf{x}_1 | x_2, x_3, x_4\} f(x_4 | x_2, x_3) dx_4.
\end{aligned}$$

Para confirmar este resultado, se tiene en cuenta que

$$E\{\mathbf{x}_1 | x_2, x_3, x_4\} = \int_{-\infty}^{-\infty} x_1 f(x_1 | x_2, x_3, x_4) dx_1.$$

y sustituyendo en la integral anterior, ésta queda

$$\int_{-\infty}^{\infty} x_1 \int_{-\infty}^{\infty} f(x_1 | x_2, x_3, x_4) f(x_4 | x_2, x_3) dx_4 dx_1.$$

Luego, como

$$f(x_1 | x_4) f(x_4) = f(x_1, x_4)$$

condicionando respecto a x_2 y x_3 , se tiene la igualdad

$$f(x_1 | x_2, x_3, x_4) f(x_4 | x_2, x_3) = f(x_1, x_4 | x_2, x_3)$$

por lo que la integral se puede expresar como

$$\begin{aligned}
\int_{-\infty}^{\infty} x_1 \int_{-\infty}^{\infty} f(x_1, x_4 | x_2, x_3) dx_4 dx_1 &= \int_{-\infty}^{\infty} x_1 f(x_1 | x_2, x_3) dx_1 \\
&= E\{\mathbf{x}_1 | x_2, x_3\}.
\end{aligned}$$

Como regla general para eliminar variables del lado derecho del valor medio condicional, se multiplica por la densidad de la variable que se desea eliminar condicionada al resto de las variables del lado derecho y se integra el producto.

6.2.2. Funciones características

La función característica de un vector aleatorio es por definición la función

$$\Phi(\Omega) = E\{e^{j\Omega\mathbf{X}^t}\} = E\{e^{j(\omega_1\mathbf{x}_1 + \dots + \omega_n\mathbf{x}_n)}\} = \Phi(j\Omega),$$

donde

$$\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \quad \Omega = [\omega_1, \dots, \omega_n].$$

Notar que los vectores son vectores fila, a diferencia de lo usual en la literatura. De esta forma, $\Omega\mathbf{X}^t$ es un escalar.

Como aplicación, se demostrará que si las variables aleatorias \mathbf{x}_i son independientes con densidades $f_i(x_i)$, la densidad $f_z(z)$ de su suma $\mathbf{z} = \mathbf{x}_1 + \dots + \mathbf{x}_n$ es la convolución de sus densidades,

$$f_z(z) = f_1(z) * \dots * f_n(z). \quad (6.23)$$

Para ver esto, se parte de la función característica de \mathbf{z} ,

$$\begin{aligned} \Phi_z(\omega) &= E\{e^{j\omega\mathbf{z}}\} \\ &= E\{e^{j\omega(\mathbf{x}_1+\dots+\mathbf{x}_n)}\} \\ &= E\{e^{j\omega\mathbf{x}_1} \dots e^{j\omega\mathbf{x}_n}\} \\ &\stackrel{(a)}{=} E\{e^{j\omega\mathbf{x}_1}\} \dots E\{e^{j\omega\mathbf{x}_n}\}, \end{aligned}$$

donde en (a) se usó la generalización de la ecuación 5.52. Por lo tanto,

$$\Phi_z(\omega) = \Phi_1(\omega) \dots \Phi_n(\omega). \quad (6.24)$$

Aplicando el teorema de la convolución de la transformada de Fourier (ver sección 5.5.1), se obtiene 6.23.

6.2.3. Vectores aleatorios normales

La normalidad conjunta de n variables aleatorias puede definirse como una generalización de las ecuaciones 5.11 y 5.12, esto es, su densidad de probabilidad conjunta es una exponencial cuyo exponente es una función cuadrática negativa. A continuación se dará una definición equivalente que expresa la normalidad conjunta de n variables aleatorias en función de la normalidad de una única variable aleatoria.

Definición: normalidad conjunta Las variables aleatorias \mathbf{x}_i son conjuntamente normales si y solo si la suma

$$a_1\mathbf{x}_1 + \dots + a_n\mathbf{x}_n = A\mathbf{X}^t$$

es una variable aleatoria normal para cualquier vector A .

Se demostrará que esta definición tiene las siguientes consecuencias: si las variables aleatorias \mathbf{x}_i tienen media η_i y matriz de covarianza C , la función característica conjunta es

$$\Phi(\Omega) = \exp \left\{ jH\Omega^t - \frac{1}{2}\Omega C \Omega^t \right\}, \quad (6.25)$$

donde $H = [\eta_1, \dots, \eta_n]$, es decir, $E\{\mathbf{X}\} = H$. Además, la densidad de probabilidad conjunta es

$$f(X) = \frac{1}{\sqrt{(2\pi)^n \Delta}} \exp \left\{ -\frac{1}{2}(X - H)C^{-1}(X - H)^t \right\}, \quad (6.26)$$

donde Δ es el determinante de C . Para demostrarlo, se parte construyendo la variable aleatoria

$$\mathbf{w} = \omega_1\mathbf{x}_1 + \dots + \omega_n\mathbf{x}_n,$$

que, por la definición dada, es normal si las variables aleatorias \mathbf{x}_i son conjuntamente normales. Además, la media η_w de \mathbf{w} es

$$\begin{aligned} \eta_w &= E\{\mathbf{w}\} = E\{\omega_1\mathbf{x}_1 + \dots + \omega_n\mathbf{x}_n\} \\ &= \omega_1 E\{\mathbf{x}_1\} + \dots + \omega_n E\{\mathbf{x}_n\} \\ &= \omega_1 \eta_1 + \dots + \omega_n \eta_n, \end{aligned}$$

es decir,

$$\eta_w = H\Omega^t.$$

La varianza σ_w^2 de \mathbf{w} es

$$\begin{aligned}
\sigma_w^2 &= E\{\mathbf{w}^2\} - E^2\{\mathbf{w}\} \\
&= E\left\{\sum_{i,j} \omega_i \omega_j \mathbf{w}_i \mathbf{w}_j\right\} - \sum_{i,j} \omega_i \omega_j \eta_i \eta_j \\
&= \sum_{i,j} \omega_i \omega_j E\{\mathbf{w}_i \mathbf{w}_j\} - \sum_{i,j} \omega_i \omega_j \eta_i \eta_j \\
&= \sum_{i,j} \omega_i \omega_j (E\{\mathbf{w}_i \mathbf{w}_j\} - \eta_i \eta_j) \\
&= \sum_{i,j} \omega_i \omega_j C_{ij},
\end{aligned}$$

es decir,

$$\sigma_w^2 = \Omega C \Omega^t.$$

Como la variable aleatoria \mathbf{z} es normal, de la ecuación 4.29, se sabe que su función característica es

$$\Phi_w(\omega) \triangleq E\{e^{j\omega \mathbf{w}}\} = e^{j\eta_w \omega - \frac{1}{2}\sigma_w^2 \omega^2}.$$

Evaluando la ecuación característica en $\omega = 1$, y sustituyendo η_w y σ_w^2 por los valores calculados previamente, se tiene que

$$\Phi_w(1) = E\{e^{j\mathbf{w}}\} = E\{e^{j\Omega \mathbf{X}^t}\} = \exp\left\{jH\Omega^t - \frac{1}{2}\Omega C \Omega^t\right\},$$

y como $\Phi(\Omega) \triangleq E\{e^{j\Omega \mathbf{X}^t}\}$, se concluye que

$$\Phi(\Omega) = \exp\left\{jH\Omega^t - \frac{1}{2}\Omega C \Omega^t\right\}.$$

Una forma de demostrar que la densidad de probabilidad está dada por la ecuación 6.26, es aplicar la transformada inversa de Fourier a la ecuación 6.25, como indica la ecuación 4.32. En el caso multivariado, la transformación inversa es

$$f(X) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Phi(\Omega) e^{-jX\Omega^t} d\omega_1 \cdots d\omega_n.$$

Para simplificar la notación, esta integral múltiple se expresará como

$$f(X) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \Phi(\Omega) e^{-jX\Omega^t} d\Omega.$$

Sustituyendo la función característica en la integral, se tiene que

$$\begin{aligned}
f(X) &= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{jH\Omega^t - \frac{1}{2}\Omega C \Omega^t} e^{-jX\Omega^t} d\Omega \\
&= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \exp\left\{-\frac{1}{2}\Omega C \Omega^t - j(X - H)\Omega^t\right\} d\Omega. \tag{6.27}
\end{aligned}$$

Una forma de resolver esta integral es usar la técnica de “completar el cuadrado”, como se hizo para resolver la integral de la ecuación 4.33. La ecuación equivalente a la ecuación 4.34 para completar un cuadrado perfecto en el caso multivariado es [14]

$$\frac{1}{2}zAz^t + bz^t + c = \frac{1}{2} (z^t + A^{-1}b^t)^t A (z^t + A^{-1}b^t) - \frac{1}{2}bA^{-1}b^t + c, \tag{6.28}$$

donde A es una matriz simétrica invertible y z , b y c son vectores fila. Para verificar la igualdad, se observa que al desarrollar el primer término a la derecha de la igualdad, se obtiene que

$$\begin{aligned} \frac{1}{2} (z^t + A^{-1}b^t)^t A (z^t + A^{-1}b^t) &= \frac{1}{2} \left[z + b (A^{-1})^t \right] A [z^t + A^{-1}b^t] \\ &\stackrel{(a)}{=} \frac{1}{2} (z + bA^{-1}) A (z^t + A^{-1}b^t) \\ &= \frac{1}{2} (zAz^t + zAA^{-1}b^t + bA^{-1}Az^t + bA^{-1}AA^{-1}b^t) \\ &= \frac{1}{2} (zAz^t + zb^t + bz^t + bA^{-1}b^t) \\ &\stackrel{(b)}{=} \frac{1}{2} zAz^t + bz^t + \frac{1}{2} bA^{-1}b^t, \end{aligned}$$

donde en (a) se usó que $(A^{-1})^t = (A^t)^{-1} = A^{-1}$. En esta ecuación, la primera igualdad surge de observar que

$$A^t (A^{-1})^t = (A^{-1}A)^t = I^t = I,$$

de donde al despejar $(A^{-1})^t$ se obtiene que $(A^{-1})^t = (A^t)^{-1}$, y la segunda igualdad se cumple por ser A simétrica y por lo tanto $A^t = A$. En (b) se usó que $zb^t = bz^t$. El exponente de la exponencial en la ecuación 6.27 puede expresarse completando el cuadrado con

$$z = \Omega, \quad A = C, \quad b = j(X - H), \quad c = 0,$$

en la ecuación 6.28. Así, dicho exponente puede escribirse como

$$\begin{aligned} \frac{1}{2} \Omega C \Omega^t + j(X - H) \Omega^t &= \frac{1}{2} [\Omega^t + jC^{-1}(X - H)^t]^t C [\Omega^t + jC^{-1}(X - H)^t] \\ &\quad + \frac{1}{2} (X - H) C^{-1} (X - H)^t. \end{aligned}$$

Ahora, como el segundo sumando no depende de la variable de integración Ω , puede sacarse afuera de la integral, y la ecuación 6.27 queda

$$\begin{aligned} f(X) &= \frac{1}{(2\pi)^n} \exp \left\{ -\frac{1}{2} (X - H) C^{-1} (X - H)^t \right\} \times \\ &\quad \int_{\mathbb{R}^n} \exp \left\{ -\frac{1}{2} [\Omega^t + jC^{-1}(X - H)^t]^t C [\Omega^t + jC^{-1}(X - H)^t] \right\} d\Omega \end{aligned} \quad (6.29)$$

La integral en esta ecuación es una *integral gaussiana multivariada*.

Una integral gaussiana multivariada tiene la forma

$$\int_{\mathbb{R}^n} \exp \left\{ -\frac{1}{2} X A X^t \right\} dX,$$

donde $X = [x_1, \dots, x_n]$ es un vector fila de largo n y A es una matriz simétrica e invertible $n \times n$. Se demostrará que³

$$\int_{\mathbb{R}^n} \exp \left\{ -\frac{1}{2} X A X^t \right\} dX = \sqrt{\frac{(2\pi)^n}{\det(A)}}. \quad (6.30)$$

Para hacerlo, se observa que por ser A simétrica, puede diagonalizarse mediante matrices ortogonales de la siguiente forma,

$$A = Q \Lambda Q^t,$$

donde Q es una matriz ortogonal, que por definición cumple que $QQ^t = I$, y Λ es una matriz diagonal con los valores propios de A en la diagonal, es decir, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, con λ_i los valores propios de A . Esta diagonalización motiva a realizar el cambio de variable $Z^t = Q^t X^t$ en la integral gaussiana. Un

³Esta deducción está basada en:
user26872, reference for multidimensional gaussian integral, URL (version: 2012-04-01): <https://math.stackexchange.com/q/126767>.

cambio de variable $(x_1, \dots, x_n) = g(z_1, \dots, z_n)$, con $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$, en una integral de una función de n variables se realiza como

$$\int_D f(x_1, \dots, x_n) dx_1 \cdots dx_n = \int_S (f \circ g)(z_1, \dots, z_n) |J_g(z_1, \dots, z_n)| dz_1 \cdots dz_n,$$

donde $|J_g(z_1, \dots, z_n)|$ es el valor absoluto del determinante de la matriz jacobiana de la función g , con $J_{g_{ij}} = \partial g_i(z_1, \dots, z_n) / \partial z_j$. Al realizar el cambio de variable

$$Z^t = Q^t X^t \quad \Rightarrow \quad QZ^t = X^t,$$

la transformación g es en este caso

$$g_i(z_1, \dots, z_n) = x_i = \sum_{j=1}^n Q_{ij} z_j,$$

y la matriz jacobiana queda

$$J_{g_{ij}}(z_1, \dots, z_n) = \frac{\partial x_i}{\partial z_j} = Q_{ij},$$

es decir, $J_g = Q$. Como la matriz Q es ortogonal, se cumple que $|\det(Q)| = 1$, ya que

$$1 = \det(I) = \det(QQ^t) = \det(Q) \det(Q^t) = \det(Q)^2.$$

Se concluye que $|\det(J_g)| = |\det(Q)| = 1$. Además, con el cambio de variable, el exponente de la exponencial en la integral queda

$$-\frac{1}{2} X A X^t = -\frac{1}{2} (Z Q^t) A (Q Z^t) = -\frac{1}{2} Z (Q^t A Q) Z^t = -\frac{1}{2} Z \Lambda Z^t,$$

en donde en la última igualdad se usó que como $A = Q \Lambda Q^t$, se cumple que $\Lambda = Q^t A Q$ por ser Q una matriz ortogonal. Teniendo en cuenta estas consideraciones, con el cambio de variable la integral gaussiana queda

$$\begin{aligned} \int_{\mathbb{R}^n} \exp \left\{ -\frac{1}{2} X A X^t \right\} dX &= \int_{\mathbb{R}^n} \exp \left\{ -\frac{1}{2} Z \Lambda Z^t \right\} dZ \\ &\stackrel{(a)}{=} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \lambda_i z_i^2 \right\} dz_1 \cdots dz_n \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^n e^{-\lambda_i z_i^2 / 2} dz_1 \cdots dz_n \\ &\stackrel{(b)}{=} \prod_{i=1}^n \int_{-\infty}^{\infty} e^{-\lambda_i z_i^2 / 2} dz_i \\ &\stackrel{(c)}{=} \prod_{i=1}^n \sqrt{\frac{2\pi}{\lambda_i}} \\ &= \sqrt{\frac{(2\pi)^n}{\prod_{i=1}^n \lambda_i}}, \end{aligned}$$

donde en (a) se tuvo en cuenta que Λ es una matriz diagonal, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, en (b) se observó que el integrando de la integral múltiple es de la forma $f(z_1) \times \cdots \times f(z_n)$ y por lo tanto se puede separar en el producto de integrales de una variable y en (c) se consideró que cada integral es una integral gaussiana de una variable, por que se aplica la ecuación 4.18. Finalmente, teniendo en cuenta que el producto de los valores propios de una matriz es igual al determinante, se llega a que

$$\int_{\mathbb{R}^n} \exp \left\{ -\frac{1}{2} X A X^t \right\} dX = \sqrt{\frac{(2\pi)^n}{\det(A)}},$$

que es lo que se quería demostrar.

Continuando con la inversión de la función característica de un vector aleatorio normal, hay que resolver la integral de la ecuación 6.29. Dicha integral puede llevarse a la forma de una integral gaussiana de la ecuación 6.30 realizando el cambio de variable $U^t = \Omega^t + jC^{-1}(X - H)^t$. Por lo tanto, teniendo en cuenta que la matriz C es simétrica por ser una matriz de covarianza, la ecuación 6.30 indica que se cumple que

$$\int_{\mathbb{R}^n} \exp \left\{ -\frac{1}{2} [\Omega^t + jC^{-1}(X - H)^t]^t C [\Omega^t + jC^{-1}(X - H)^t] \right\} d\Omega = \sqrt{\frac{(2\pi)^n}{\det(C)}}$$

y sustituyendo este resultado en la ecuación 6.29 se llega a que

$$\begin{aligned} f(X) &= \frac{1}{(2\pi)^n} \exp \left\{ -\frac{1}{2} (X - H) C^{-1} (X - H)^t \right\} \sqrt{\frac{(2\pi)^n}{\det(C)}} \\ &= \frac{1}{\sqrt{(2\pi)^n \det(C)}} \exp \left\{ -\frac{1}{2} (X - H) C^{-1} (X - H)^t \right\}, \end{aligned}$$

como indica la ecuación 6.26.

Se muestra a continuación una forma de realizar el cálculo de la función característica de un vector normal a partir de la densidad de probabilidad. La función característica se obtiene como la transformada de Fourier de la densidad de probabilidad. Por lo tanto,

$$\begin{aligned} \Phi(\Omega) &= \int_{\mathbb{R}^n} f(X) \exp\{j\Omega X^t\} dX \\ &= \frac{1}{\sqrt{(2\pi)^n \Delta}} \int_{\mathbb{R}^n} \exp \left\{ -\frac{1}{2} (X - H) C^{-1} (X - H)^t \right\} \exp\{j\Omega X^t\} dX, \end{aligned} \quad (6.31)$$

donde en la segunda igualdad se sustituyó $f(X)$ usando la ecuación 6.26. Operando con el exponente de la exponencial, se ve que

$$\begin{aligned} -\frac{1}{2} (X - H) C^{-1} (X - H)^t + j\Omega X^t &= -\frac{1}{2} X C^{-1} X^t + \frac{1}{2} X C^{-1} H^t + \frac{1}{2} H C^{-1} X^t - \frac{1}{2} H C^{-1} H^t + j\Omega X^t \\ &\stackrel{(a)}{=} -\frac{1}{2} X C^{-1} X^t + H C^{-1} X^t - \frac{1}{2} H C^{-1} H^t + j\Omega X^t \\ &= -\frac{1}{2} X C^{-1} X^t + (H C^{-1} + j\Omega) X^t - \frac{1}{2} H C^{-1} H^t, \end{aligned}$$

donde en (a) se usó que $X C^{-1} H^t = H C^{-1} X^t$. El exponente es una cuadrática en X , y se puede expresar como una cuadrado perfecto y completando términos como la igualdad de la ecuación 6.28, con

$$z = X, \quad A = -C^{-1}, \quad b = H C^{-1} + j\Omega, \quad c = -\frac{1}{2} H C^{-1} H^t.$$

Expresado de esta forma, el exponente queda

$$-\frac{1}{2} (X^t - Cb^t)^t C^{-1} (X^t - Cb^t) + \frac{1}{2} b C b^t - \frac{1}{2} H C^{-1} H^t,$$

donde por el momento no se sustituyó el vector b , que no depende de la variable de integración X . Los dos últimos sumandos del exponente no dependen de la variable de integración, por lo que pueden sacarse afuera de la integral en la ecuación 6.31, resultando en

$$\Phi(\Omega) = \frac{1}{\sqrt{(2\pi)^n \Delta}} \exp \left\{ \frac{1}{2} b C b^t - \frac{1}{2} H C^{-1} H^t \right\} \int_{\mathbb{R}^n} \exp \left\{ -\frac{1}{2} (X^t - Cb^t)^t C^{-1} (X^t - Cb^t) \right\} dX.$$

La integral es gaussiana, por lo que por la ecuación 6.30 integra a

$$\int_{\mathbb{R}^n} \exp \left\{ -\frac{1}{2} (X^t - Cb^t)^t C^{-1} (X^t - Cb^t) \right\} dX = \sqrt{\frac{(2\pi)^n}{\det(C^{-1})}} = \sqrt{(2\pi)^n \det(C)},$$

donde en la segunda igualdad se usó que $\det(C^{-1}) = 1/\det(C)$, ya que

$$1 = \det(I) = \det(CC^{-1}) = \det(C)\det(C^{-1}).$$

Sustituyendo este resultado en la función característica y considerando que previamente se definió $\Delta = \det(C)$ se tiene que

$$\begin{aligned}\Phi(\Omega) &= \frac{1}{\sqrt{(2\pi)^n \Delta}} \exp \left\{ \frac{1}{2} b C b^t - \frac{1}{2} H C^{-1} H^t \right\} \sqrt{(2\pi)^n \Delta} \\ &= \exp \left\{ \frac{1}{2} b C b^t - \frac{1}{2} H C^{-1} H^t \right\}.\end{aligned}$$

Sustituyendo $b = H C^{-1} + j\Omega$, el exponente queda

$$\begin{aligned}\frac{1}{2} b C b^t - \frac{1}{2} H C^{-1} H^t &= \frac{1}{2} (H C^{-1} + j\Omega) C (H C^{-1} + j\Omega)^t - \frac{1}{2} H C^{-1} H^t \\ &= \frac{1}{2} H C^{-1} C (C^{-1})^t H^t + \frac{1}{2} j H C^{-1} C \Omega^t + \frac{1}{2} j \Omega C (C^{-1})^t H^t - \frac{1}{2} \Omega C \Omega^t - \frac{1}{2} H C^{-1} H^t \\ &\stackrel{(a)}{=} \frac{1}{2} H C^{-1} H^t + \frac{1}{2} j H \Omega^t + \frac{1}{2} j \Omega H^t - \frac{1}{2} \Omega C \Omega^t - \frac{1}{2} H C^{-1} H^t \\ &\stackrel{(b)}{=} j H \Omega^t - \frac{1}{2} \Omega C \Omega^t,\end{aligned}$$

donde en (a) se usó que $CC^{-1} = C^{-1}C = I$ y $C(C^{-1})^t = I$ por ser C simétrica, y en (b) que $\Omega H^t = H \Omega^t$ y se anularon los términos opuestos. Se concluye que

$$\Phi(\Omega) = \exp \left\{ j H \Omega^t - \frac{1}{2} \Omega C \Omega^t \right\},$$

coincidiendo con el resultado obtenido antes en la ecuación 6.25.

En el caso particular en que el vector aleatorio normal tenga media nula, es decir, $H = [0, \dots, 0]$, la función característica y la densidad de probabilidad conjunta son respectivamente

$$\Phi(\Omega) = \exp \left\{ -\frac{1}{2} \Omega C \Omega^t \right\}, \quad (6.32)$$

$$f(X) = \frac{1}{\sqrt{(2\pi)^n \Delta}} \exp \left\{ -\frac{1}{2} X C^{-1} X^t \right\}. \quad (6.33)$$

Además, si las variables aleatorias \mathbf{x}_i son conjuntamente normales y no correlacionadas, son independientes (ver sección 5.4.2). Efectivamente, por ser no correlacionadas y tener media nula, la matriz de covarianza es diagonal con $C_{ii} = \sigma_i^2$. Por lo tanto, C^{-1} también es diagonal, con $C_{ii}^{-1} = 1/\sigma_i^2$. Considerando esto, la ecuación 6.33 queda

$$\begin{aligned}f(X) &= \frac{1}{\sigma_1 \cdots \sigma_n \sqrt{(2\pi)^n}} \exp \left\{ -\frac{1}{2} \left(\frac{x_1^2}{\sigma_1^2} + \cdots + \frac{x_n^2}{\sigma_n^2} \right) \right\} \\ &= \prod_{i=1}^n \frac{1}{\sigma_i \sqrt{2\pi}} e^{-x_i^2/2\sigma_i^2} \\ &= \prod_{i=1}^n f(x_i).\end{aligned}$$

Ejemplo (7-9) Se demostrará usando la función característica que si las variables aleatorias \mathbf{x}_i son conjuntamente normales de media nula y $E\{\mathbf{x}_i \mathbf{x}_j\} = C_{ij}$, se cumple que

$$E\{\mathbf{x}_1 \mathbf{x}_2 \mathbf{x}_3 \mathbf{x}_4\} = C_{12}C_{34} + C_{13}C_{24} + C_{14}C_{23}. \quad (6.34)$$

La función característica, que está dada por la ecuación 6.32, puede escribirse como

$$E\{e^{j(\omega_1\mathbf{x}_1+\dots+\omega_4\mathbf{x}_4)}\} = \exp\left\{-\frac{1}{2}\sum_{i=1}^4\sum_{j=1}^4\omega_i\omega_jC_{ij}\right\}, \quad (6.35)$$

donde el lado izquierdo de la igualdad es la definición de la función característica. Se expandirá en series de potencias las exponenciales a ambos lados de la igualdad⁴. La expansión del lado izquierdo de la igualdad es

$$\begin{aligned} E\{e^{j(\omega_1\mathbf{x}_1+\dots+\omega_4\mathbf{x}_4)}\} &= E\left\{\sum_{k=0}^{\infty}\frac{j^k(\omega_1\mathbf{x}_1+\dots+\omega_4\mathbf{x}_4)^k}{k!}\right\} \\ &\stackrel{(a)}{=} \sum_{k=0}^{\infty}\frac{j^k}{k!}E\{(\omega_1\mathbf{x}_1+\dots+\omega_4\mathbf{x}_4)^k\} \\ &\stackrel{(b)}{=} \dots + \frac{1}{4!}E\{(\omega_1\mathbf{x}_1+\dots+\omega_4\mathbf{x}_4)^4\} + \dots, \end{aligned}$$

donde en (a) se usó la propiedad de linealidad de la esperanza y en (b) se muestra explícitamente el único sumando de la sumatoria que contiene el factor $\omega_1\omega_2\omega_3\omega_4$, que es cuando $k=4$. Desarrollando el cuatrínomio a la cuarta, puede verse que el término $E\{\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3\mathbf{x}_4\}\omega_1\omega_2\omega_3\omega_4$ aparece en todos los órdenes posibles de los subíndices, es decir 4! veces, y por lo tanto

$$E\{(\omega_1\mathbf{x}_1+\dots+\omega_4\mathbf{x}_4)^4\} = (4!)E\{\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3\mathbf{x}_4\}\omega_1\omega_2\omega_3\omega_4 + \dots$$

De esta forma,

$$\begin{aligned} E\{e^{j(\omega_1\mathbf{x}_1+\dots+\omega_4\mathbf{x}_4)}\} &= \dots + \frac{4!}{4!}E\{\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3\mathbf{x}_4\}\omega_1\omega_2\omega_3\omega_4 + \dots \\ &= \dots + E\{\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3\mathbf{x}_4\}\omega_1\omega_2\omega_3\omega_4 + \dots \end{aligned} \quad (6.36)$$

Expandiendo el lado derecho de la igualdad de la ecuación 6.35, se tiene que

$$\begin{aligned} \exp\left\{-\frac{1}{2}\sum_{i=1}^4\sum_{j=1}^4\omega_i\omega_jC_{ij}\right\} &= \sum_{k=0}^{\infty}\frac{1}{k!}\left(-\frac{1}{2}\sum_{i=1}^4\sum_{j=1}^4\omega_i\omega_jC_{ij}\right)^k \\ &\stackrel{(a)}{=} \dots + \frac{1}{2}\left(-\frac{1}{2}\sum_{i=1}^4\sum_{j=1}^4\omega_i\omega_jC_{ij}\right)^2 + \dots \\ &= \dots + \frac{1}{8}\left(\sum_{i=1}^4\sum_{j=1}^4\omega_i\omega_jC_{ij}\right)^2 + \dots \end{aligned} \quad (6.37)$$

donde nuevamente en (a) se muestra el único sumando que contiene el factor $\omega_1\omega_2\omega_3\omega_4$. Para analizar los términos con dicho factor, se escribe explícitamente la doble sumatoria en la última ecuación,

$$\begin{aligned} \sum_{i=1}^4\sum_{j=1}^4\omega_i\omega_jC_{ij} &= \omega_1(\omega_1C_{11} + \omega_2C_{12} + \omega_3C_{13} + \omega_4C_{14}) \\ &\quad + \omega_2(\omega_1C_{21} + \omega_2C_{22} + \omega_3C_{23} + \omega_4C_{24}) \\ &\quad + \omega_3(\omega_1C_{31} + \omega_2C_{32} + \omega_3C_{33} + \omega_4C_{34}) \\ &\quad + \omega_4(\omega_1C_{41} + \omega_2C_{42} + \omega_3C_{43} + \omega_4C_{44}). \end{aligned}$$

En esta expresión, considérese por ejemplo el término $\omega_1\omega_2C_{12}$. Al elevar al cuadrado, se observa que al multiplicar ese término por $\omega_3\omega_4C_{34}$ y $\omega_4\omega_3C_{43}$ se obtienen los dos únicos términos con el factor

⁴La representación en serie de potencias de la exponencial es $e^x = \sum_{k=0}^{\infty}x^k/k!$.

$\omega_1\omega_2\omega_3\omega_4$ asociados a $\omega_1\omega_2C_{12}$. Lo mismo ocurre si se considera el término $\omega_2\omega_1C_{21}$. Análogamente, hay dos términos con el factor $\omega_1\omega_2\omega_3\omega_4$ asociados a $\omega_3\omega_4C_{34}$, que provienen de multiplicar por $\omega_1\omega_2C_{12}$ y $\omega_2\omega_1C_{21}$. Lo mismo ocurre si se considera el término $\omega_4\omega_3C_{43}$. En resumen, los términos con el factor $\omega_1\omega_2\omega_3\omega_4$ asociados a $\omega_1\omega_2C_{12}$, $\omega_2\omega_1C_{21}$, $\omega_3\omega_4C_{34}$ y $\omega_4\omega_3C_{43}$ son

$$\begin{aligned} & \omega_1\omega_2C_{12}(\omega_3\omega_4C_{34} + \omega_4\omega_3C_{34}) + \omega_2\omega_1C_{21}(\omega_3\omega_4C_{34} + \omega_4\omega_3C_{34}) \\ & + \omega_3\omega_4C_{34}(\omega_1\omega_2C_{12} + \omega_2\omega_1C_{21}) + \omega_4\omega_3C_{43}(\omega_1\omega_2C_{12} + \omega_2\omega_1C_{21}) \end{aligned}$$

y teniendo en cuenta que la matriz de covarianza es simétrica y por lo tanto se cumple que $C_{12} = C_{21}$ y $C_{34} = C_{43}$, la expresión anterior es

$$8C_{12}C_{34}\omega_1\omega_2\omega_3\omega_4.$$

Razonando de la misma forma puede verse también que los otros términos que contienen el factor $\omega_1\omega_2\omega_3\omega_4$ son

$$8C_{13}C_{24}\omega_1\omega_2\omega_3\omega_4, \quad 8C_{14}C_{23}\omega_1\omega_2\omega_3\omega_4.$$

Por lo tanto, sustituyendo estos resultados en la ecuación 6.37 se obtiene que

$$\exp \left\{ -\frac{1}{2} \sum_{i=1}^4 \sum_{j=1}^4 \omega_i \omega_j C_{ij} \right\} = \cdots + \frac{8}{8} (C_{12}C_{34} + C_{13}C_{24} + C_{14}C_{23})\omega_1\omega_2\omega_3\omega_4 + \cdots$$

e igualando coeficientes con la ecuación 6.36 se obtiene la ecuación 6.34.

Vectores aleatorios normales complejos Un vector aleatorio normal complejo es un vector $\mathbf{Z} = \mathbf{X} + j\mathbf{Y} = [\mathbf{z}_1, \dots, \mathbf{z}_n]$, cuyos componentes son las n variables aleatorias conjuntamente normales $\mathbf{z}_i = \mathbf{x}_i + j\mathbf{y}_i$. Se asumirá que $E\{\mathbf{z}_i\} = 0$. Las propiedades estadísticas del vector \mathbf{Z} están especificadas por la densidad de probabilidad conjunta

$$f_Z(Z) = f(x_1, \dots, x_n, y_1, \dots, y_n)$$

de las $2n$ variables aleatorias reales \mathbf{x}_i e \mathbf{y}_i . Esta función es una exponencial de la forma de la ecuación 6.33, donde la matriz de covarianza es la matriz real $2n \times 2n$ de covarianza del vector aleatorio $[\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{y}_1, \dots, \mathbf{y}_n]$

$$D = \begin{bmatrix} C_{XX} & C_{XY} \\ C_{YX} & C_{YY} \end{bmatrix}.$$

Como las matrices C_{XX} y C_{YY} son simétricas y $C_{YX} = C_{XY}^t$, la matriz D consiste en $2n^2 + n$ parámetros reales. Esto es porque que el número de entradas distintas $E\{\mathbf{x}_i\mathbf{x}_j\}$ y $E\{\mathbf{y}_i\mathbf{y}_j\}$ de las matrices C_{XX} y C_{YY} respectivamente, es

$$\binom{n}{2} + n = \frac{n!}{2!(n-2)!} + n = \frac{n(n-1)}{2} + n = \frac{n(n+1)}{2},$$

ya que hay combinaciones de 2 elementos tomados de n (las matrices son simétricas y por eso no son arreglos) mas los n casos con $i = j$. Además, el número de entradas $E\{\mathbf{x}_i\mathbf{y}_j\}$ de la matriz C_{XY} , que coinciden con las del la matriz C_{YX} , es n^2 . De esta forma, el total de parámetros es

$$2 \frac{n(n+1)}{2} + n^2 = 2n^2 + n.$$

La función característica correspondiente,

$$\Phi_Z(\Omega) = E\{\exp[j(u_1\mathbf{x}_1, \dots, u_n\mathbf{x}_n, v_1\mathbf{y}_1, \dots, v_n\mathbf{y}_n)]\},$$

es una exponencial como la ecuación 6.32,

$$\Phi_Z(\Omega) = \exp \left\{ -\frac{1}{2} Q \right\}, \quad Q = [U \quad V] \begin{bmatrix} C_{XX} & C_{XY} \\ C_{YX} & C_{YY} \end{bmatrix} \begin{bmatrix} U^t \\ V^t \end{bmatrix} \quad (6.38)$$

donde $U = [u_1, \dots, u_n]$ y $V = [v_1, \dots, v_n]$, y $\Omega = U + jV$. Notar que la definición de la función característica cuando la variable aleatoria es un vector complejo \mathbf{Z} se puede expresar como

$$\Phi_Z(\Omega) = E\{j\Re(\Omega^* \mathbf{Z}^t)\},$$

ya que

$$\begin{aligned}\Re\{\Omega^* \mathbf{Z}^t\} &= \Re\{(U + jV)^*(\mathbf{X} + j\mathbf{Y})^t\} \\ &= \Re\left\{\sum_{i=1}^n (u_i + jv_i)^*(\mathbf{x}_i + j\mathbf{y}_i)\right\} \\ &= \Re\left\{\sum_{i=1}^n (u_i - jv_i)(\mathbf{x}_i + j\mathbf{y}_i)\right\} \\ &= \Re\left\{\sum_{i=1}^n u_i \mathbf{x}_i + j u_i \mathbf{y}_i - j v_i \mathbf{x}_i + v_i \mathbf{y}_i\right\} \\ &= \sum_{i=1}^n u_i \mathbf{x}_i + v_i \mathbf{y}_i.\end{aligned}$$

La matriz de covarianza del vector complejo \mathbf{Z} es la matriz hermítica $n \times n$

$$C_{ZZ} = E\{\mathbf{Z}^t \mathbf{Z}^*\} = C_{XX} + C_{YY} - j(C_{XY} - C_{YX}), \quad (6.39)$$

donde la última igualdad proviene de observar que

$$\begin{aligned}E\{\mathbf{Z}^t \mathbf{Z}^*\} &= E\{(\mathbf{X} + j\mathbf{Y})^t (\mathbf{X} + j\mathbf{Y})^*\} \\ &= E\{(\mathbf{X}^t + j\mathbf{Y}^t)(\mathbf{X} - j\mathbf{Y})\} \\ &= E\{\mathbf{X}^t \mathbf{X} - j\mathbf{X}^t \mathbf{Y} + j\mathbf{Y}^t \mathbf{X} + \mathbf{Y}^t \mathbf{Y}\} \\ &= E\{\mathbf{X}^t \mathbf{X}\} + E\{\mathbf{Y}^t \mathbf{Y}\} - j(E\{\mathbf{X}^t \mathbf{Y}\} - E\{\mathbf{Y}^t \mathbf{X}\}) \\ &= C_{XX} + C_{YY} - j(C_{XY} - C_{YX}).\end{aligned}$$

La matriz C_{ZZ} está especificada mediante n^2 parámetros reales. Esto es porque la cantidad de elementos sobre la diagonal, sin incluirla, es

$$\sum_{i=1}^{n-1} i = \frac{n(n-1)}{2}.$$

Estos elementos son números complejos, por lo que determinan dos parámetros cada uno. Además, hay n elementos reales en la diagonal ($E\{\mathbf{z}_i \mathbf{z}_i^*\} = E\{|\mathbf{z}_i|^2\}$ real). El total de parámetros es por lo tanto,

$$2 \frac{n(n-1)}{2} + n = n^2 - n + n = n^2.$$

Como la matriz de covarianza C_{ZZ} del vector complejo \mathbf{Z} contiene n^2 parámetros y la densidad de probabilidad normal $f_Z(Z)$ está especificada por $2n^2 + n$ parámetros, en general la densidad $f_Z(Z)$ no queda determinada en términos de la matriz de covarianza C_{ZZ} , a diferencia del caso real. Por ejemplo, si $n = 1$, $\mathbf{Z} = \mathbf{z} = \mathbf{x} + j\mathbf{y}$, por lo que $C_{ZZ} = E\{\mathbf{z} \mathbf{z}^*\} = E\{|\mathbf{z}|^2\}$. Así, C_{ZZ} determina un único parámetro $\sigma_z^2 = E\{\mathbf{x}^2 + \mathbf{y}^2\}$. Sin embargo, la densidad de \mathbf{Z} $f_z(z) = f(x, y)$ es una densidad normal bivariada determinada por los tres parámetros σ_x^2 , σ_y^2 y $E\{\mathbf{x} \mathbf{y}\}$. A continuación se presenta una clase particular de vectores normales cuya estadística está determinada por su matriz de covarianza:

Teorema de Goodman: si los vectores \mathbf{X} e \mathbf{Y} son tales que

$$C_{XX} = C_{YY}, \quad C_{XY} = -C_{YX}$$

y $\mathbf{Z} = \mathbf{X} + j\mathbf{Y}$, se cumple que

$$C_{ZZ} = 2(C_{XX} - jC_{XY})$$

$$f_Z(Z) = \frac{1}{\pi^n |C_{ZZ}|} \exp \{-Z C_{ZZ}^{-1} Z^h\} \quad (6.40)$$

$$\Phi_Z(\Omega) = \exp \left\{ -\frac{1}{4} \Omega C_{ZZ} \Omega^h \right\}. \quad (6.41)$$

La ecuación de C_{ZZ} se obtiene inmediatamente de la ecuación 6.39 de la matriz de covarianza de un vector aleatorio complejo aplicando las hipótesis. Alcanza con demostrar 6.41 ya que 6.40 se obtiene a través de la transformada inversa de Fourier de 6.41. Por un lado, el exponente de la función característica es $-Q/2$ (ver ecuación 6.38) con

$$\begin{aligned} Q &= [U \ V] D \begin{bmatrix} U^t \\ V^t \end{bmatrix} \\ &\stackrel{(a)}{=} [U \ V] \begin{bmatrix} C_{XX} & C_{XY} \\ -C_{XY} & C_{XX} \end{bmatrix} \begin{bmatrix} U^t \\ V^t \end{bmatrix} \\ &= [U \ V] \begin{bmatrix} C_{XX}U^t + C_{XY}V^t \\ -C_{XY}U^t + C_{XX}V^t \end{bmatrix} \\ &= UC_{XX}U^t + UC_{XY}V^t - VC_{XY}U^t + VC_{XX}V^t, \end{aligned}$$

donde en (a) se consideró la hipótesis de que $C_{XX} = C_{YY}$ y $C_{YX} = -C_{XY}$. Considerando que

$$UC_{XY}V^t = (UC_{XY}V^t)^t = VC_{XY}^tU^t = -VC_{XY}U^t, \quad (6.42)$$

por lo que

$$Q = UC_{XX}U^t + 2UC_{XY}V^t + VC_{XX}V^t. \quad (6.43)$$

Por otro lado, en el exponente de la ecuación 6.41 aparece el término

$$\begin{aligned} \frac{1}{2} \Omega C_{ZZ} \Omega^h &= \frac{1}{2} (U + jV) 2(C_{XX} - jC_{XY})(U^t - jV^t) \\ &= (U + jV)(C_{XX}U^t - jC_{XX}V^t - jC_{XY}U^t - C_{XY}V^t) \\ &= (UC_{XX}U^t - jUC_{XX}V^t - jUC_{XY}U^t - UC_{XY}V^t) \\ &\quad + (jVC_{XX}U^t + VC_{XX}V^t + VC_{XY}U^t - jVC_{XY}V^t) \\ &= (UC_{XX}U^t - UC_{XY}V^t + VC_{XX}V^t + VC_{XY}U^t) \\ &\quad + j(-UC_{XX}V^t - UC_{XY}U^t + VC_{XX}U^t - VC_{XY}V^t). \end{aligned} \quad (6.44)$$

Notando que C_{XX} es simétrica y por lo tanto $C_{XX}^t = C_{XX}$, se cumple que

$$UC_{XX}V^t = VC_{XX}U^t.$$

Además, también se cumple que

$$UC_{XY}U^t = VC_{XY}V^t = 0.$$

Esto es porque $C_{XYii} = E\{x_i y_i\}$ y $C_{YXii} = E\{y_i x_i\}$, por lo que $C_{XYii} = C_{YXii}$. Pero como por hipótesis $C_{YX} = -C_{XY}$, se concluye que $C_{XYii} = C_{YXii} = 0$, es decir, la diagonal de las matrices C_{XY} y C_{YX} es nula. Teniendo esto en cuenta, se ve que

$$\begin{aligned} UC_{XY}U^t &= \sum_{i=1}^n \sum_{j=1}^n u_i u_j C_{XYij} \\ &\stackrel{(a)}{=} \sum_{i=1}^{n-1} \sum_{j=i+1}^n u_i u_j C_{XYij} + \sum_{m=1}^{n-1} \sum_{l=m+1}^n u_m u_l C_{XYlm} + \sum_{i=1}^n u_i^2 C_{XYii} \\ &\stackrel{(b)}{=} \sum_{i=1}^{n-1} \sum_{j=i+1}^n u_i u_j C_{XYij} + \sum_{i=1}^{n-1} \sum_{j=i+1}^n u_i u_j C_{XYji} \\ &= \sum_{i=1}^{n-1} \sum_{j=i+1}^n u_i u_j (C_{XYij} + C_{XYji}) \\ &\stackrel{(c)}{=} 0, \end{aligned}$$

donde en (a) se separó la sumatoria en todos los elementos de la matriz en la sumatoria de los elementos sobre la diagonal sumando a lo largo de las filas, la sumatoria bajo la diagonal sumando a lo largo de las columnas y la sumatoria en la diagonal, en (b) se consideró que los elementos de la diagonal son cero, por lo que la sumatoria simple se anula y se realizó el cambio de variable $m = i$ y $l = j$ en la segunda doble sumatoria y en (c) se consideró la hipótesis de que $C_{YX} = -C_{XY}$, que es equivalente a $C_{XY} = -C_{XY}^t$. Estas consideraciones permiten concluir que el parte imaginaria de la ecuación 6.44 se anula, ya que el primer y tercer sumando de la parte imaginaria se anulan entre si, y el segundo y cuarto sumando son nulos. De esta forma, la ecuación 6.44 queda

$$\begin{aligned} \frac{1}{2}\Omega C_{ZZ}\Omega^h &= UC_{XX}U^t - UC_{XY}V^t + VC_{XX}V^t + VC_{XY}U^t \\ &\stackrel{(a)}{=} UC_{XX}U^t + 2VC_{XY}U^t + VC_{XX}V^t \end{aligned}$$

donde en (a) se usó la ecuación 6.42. Comparando este resultado con la ecuación 6.43 se concluye que

$$\frac{1}{2}\Omega C_{ZZ}\Omega^h = Q,$$

por lo que la ecuación característica del vector aleatorio \mathbf{Z} está dada por la ecuación 6.41, que es lo que se quería demostrar.

Formas cuadráticas normales A partir de n variables aleatorias \mathbf{z}_i independientes con distribución $N(0, 1)$, se construye la variable aleatoria

$$\mathbf{x} = \mathbf{z}_1^2 + \cdots + \mathbf{z}_n^2.$$

Se demostrará usando funciones características que la variable aleatoria \mathbf{x} tiene distribución chi-cuadrado con n grados de libertad, es decir (ver la ecuación 3.18),

$$f_x(x) = \gamma x^{n/2-1} e^{-x/2} U(x), \quad \gamma = \frac{1}{2^{n/2}\Gamma(n/2)}.$$

Para hacerlo, se parte observando que como las variables \mathbf{z}_i tienen distribución $N(0, 1)$, las variables \mathbf{z}_i^2 tienen distribución $\chi^2(1)$, como se explicó en la sección 4.1.2. Por lo tanto, sus funciones características están dadas por la ecuación 4.42 con $m = 1$,

$$\Phi_i(s) = E\{e^{s\mathbf{z}_i}\} = \frac{1}{\sqrt{1-2s}}.$$

Como \mathbf{x} es la suma de las variables independientes \mathbf{z}_i^2 , la función característica de \mathbf{x} es el producto de las funciones características de \mathbf{z}_i , como indica la ecuación 6.24. De esta forma

$$\Phi_x(s) = \Phi_1(s) \cdots \Phi_n(s) = \frac{1}{\sqrt{(1-2s)^n}}.$$

Comparando esta ecuación con la ecuación 4.42, se ve que se trata de la función característica de una variable aleatoria chi-cuadrado con n grados de libertad, y por lo tanto, se concluye que $\mathbf{x} \sim \chi^2(n)$.

Notar que si las variables aleatorias $\mathbf{x} \sim \chi^2(m)$ e $\mathbf{y} \sim \chi^2(n)$ son independientes se cumple que $\mathbf{z} = \mathbf{x} + \mathbf{y} \sim \chi^2(m+n)$, ya que

$$\Phi_z(s) = \Phi_x(s)\Phi_y(s) = \frac{1}{\sqrt{(1-2s)^m}} \frac{1}{\sqrt{(1-2s)^n}} = \frac{1}{\sqrt{(1-2s)^{m+n}}}.$$

Recíprocamente, si $\mathbf{z} = \mathbf{x} + \mathbf{y} \sim \chi^2(m+n)$, \mathbf{x} e \mathbf{y} son variables aleatorias independientes y $\mathbf{x} \sim \chi^2(m)$, se cumple que $\mathbf{y} \sim \chi^2(n)$.

Varianza muestral Dadas n variables aleatorias \mathbf{x}_i i.i.d. con distribución $N(\eta, \sigma^2)$, se construye la varianza muestral (ver la ecuación 6.10),

$$\mathbf{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2, \quad \bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i.$$

Se demostrará que la variable aleatoria

$$\frac{(n-1)\mathbf{s}^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{\mathbf{x}_i - \bar{\mathbf{x}}}{\sigma} \right)^2 \quad \text{es} \quad \chi^2(n-1). \quad (6.45)$$

Como resultado previo, se demostrará que las variables aleatorias \mathbf{s}^2 y $\bar{\mathbf{x}}$ son independientes. Se considera primero el caso de n variables aleatorias \mathbf{y}_i i.i.d. con distribución $N(0, 1)$. Por lo tanto, $\bar{\mathbf{y}}$ y $\mathbf{y}_i - \bar{\mathbf{y}}$ son variables aleatorias normales por tratarse de una combinación lineal de variables aleatorias normales, como se mostró en la sección 5.5.1 para el caso de la combinación lineal de dos variables aleatorias. Además, como \mathbf{y}_i tiene media nula, $\bar{\mathbf{y}}$ e $\mathbf{y}_i - \bar{\mathbf{y}}$ también tienen media nula (ver la sección 6.1.3). La covarianza entre $\bar{\mathbf{y}}$ e $\mathbf{y}_i - \bar{\mathbf{y}}$ es

$$\text{Cov}\{\bar{\mathbf{y}}, \mathbf{y}_i - \bar{\mathbf{y}}\} \stackrel{(a)}{=} E\{\bar{\mathbf{y}}(\mathbf{y}_i - \bar{\mathbf{y}})\} = E\{\bar{\mathbf{y}}\mathbf{y}_i\} - E\{\bar{\mathbf{y}}^2\},$$

donde en (a) se consideró que las variables aleatorias tienen media nula. Además

$$E\{\bar{\mathbf{y}}\mathbf{y}_i\} = E\left\{ \frac{1}{n} \sum_{k=1}^n \mathbf{y}_k \mathbf{y}_i \right\} = \frac{1}{n} \sum_{k=1}^n E\{\mathbf{y}_k \mathbf{y}_i\} \stackrel{(a)}{=} \frac{1}{n} E\{\mathbf{y}_i^2\} \stackrel{(b)}{=} \frac{1}{n} \sigma_{y_i}^2 = \frac{1}{n},$$

donde en (a) se tuvo en cuenta que las variables \mathbf{y}_i son independientes y de media nula y por lo tanto $E\{\mathbf{y}_k \mathbf{y}_i\} = E\{\mathbf{y}_k\}E\{\mathbf{y}_i\} = 0$ si $i \neq k$ y en (b) nuevamente que \mathbf{y}_i es de media nula. También se observa que

$$E\{\bar{\mathbf{y}}^2\} = \sigma_{\bar{\mathbf{y}}}^2 = \frac{\sigma_{y_i}^2}{n} = \frac{1}{n},$$

como se calculó en la sección 6.1.3. Sustituyendo estos resultados en la ecuación de la covarianza, se obtiene que

$$\text{Cov}\{\bar{\mathbf{y}}, \mathbf{y}_i - \bar{\mathbf{y}}\} = \frac{1}{n} - \frac{1}{n} = 0,$$

lo que indica que las variables aleatorias $\bar{\mathbf{y}}$ e $\mathbf{y}_i - \bar{\mathbf{y}}$ son no correlacionadas, pero además, por ser normales, son independientes, como se explicó en la sección 5.4.2. Definiendo el vector $\mathbf{Y} = [\mathbf{y}_1 - \bar{\mathbf{y}}, \dots, \mathbf{y}_n - \bar{\mathbf{y}}]$, se cumple entonces que $\bar{\mathbf{y}}$ e \mathbf{Y} son independientes, y también que $\bar{\mathbf{y}}$ y

$$\frac{1}{n-1} \mathbf{Y} \mathbf{Y}^t = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})^2 = \mathbf{s}_Y^2$$

son independientes, como indica la ecuación 5.52. Se concluye que $\bar{\mathbf{y}}$ y \mathbf{s}_Y^2 son variables aleatorias independientes si $\mathbf{y}_i \sim N(0, 1)$. Se generalizará el resultado para el caso de n variables aleatorias \mathbf{x}_i i.i.d. con distribución $N(\eta, \sigma^2)$. Se comienza definiendo las variables aleatorias $\mathbf{y}_i = (\mathbf{x}_i - \eta)/\sigma$. Así, $\mathbf{y}_i \sim N(0, 1)$, y como $\mathbf{x}_i = \sigma \mathbf{y}_i + \eta$, se tiene que

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \frac{1}{n} \sum_{i=1}^n (\sigma \mathbf{y}_i + \eta) = \frac{\sigma}{n} \sum_{i=1}^n \mathbf{y}_i + \eta = \sigma \bar{\mathbf{y}} + \eta,$$

y

$$\begin{aligned} \mathbf{s}_X^2 &= \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n [(\sigma \mathbf{y}_i + \eta) - (\sigma \bar{\mathbf{y}} + \eta)]^2 \\ &= \frac{\sigma^2}{n-1} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})^2 \\ &= \sigma^2 \mathbf{s}_Y^2. \end{aligned}$$

Por lo tanto, como \mathbf{s}_Y^2 y $\bar{\mathbf{y}}$ son independientes por el argumento dado antes, $\bar{\mathbf{x}} = \sigma\bar{\mathbf{y}} + \eta$ y $\mathbf{s}_X^2 = \sigma^2\mathbf{s}_Y^2$ también son independientes.

Se demostrará ahora que la variable aleatoria $(n-1)\mathbf{s}^2/\sigma^2$ es $\chi^2(n-1)$. Para esto, se parte considerando la identidad

$$(\mathbf{x}_i - \eta)^2 = [(\mathbf{x}_i - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \eta)]^2 = (\mathbf{x}_i - \bar{\mathbf{x}})^2 + (\bar{\mathbf{x}} - \eta)^2 + 2(\mathbf{x}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}} - \eta).$$

Dividiendo entre σ^2 y sumando en i desde 1 a n , se obtiene que

$$\frac{1}{\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \eta)^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2 + \frac{1}{\sigma^2} \sum_{i=1}^n (\bar{\mathbf{x}} - \eta)^2 + \frac{2}{\sigma^2} \sum_{i=1}^n (\bar{\mathbf{x}} - \eta)(\mathbf{x}_i - \bar{\mathbf{x}}).$$

Considerando que $(\bar{\mathbf{x}} - \eta)$ no depende de i , esta igualdad queda

$$\begin{aligned} \sum_{i=1}^n \left(\frac{\mathbf{x}_i - \eta}{\sigma} \right)^2 &= \sum_{i=1}^n \left(\frac{\mathbf{x}_i - \bar{\mathbf{x}}}{\sigma} \right)^2 + n \left(\frac{\bar{\mathbf{x}} - \eta}{\sigma} \right)^2 + \frac{2}{\sigma^2} (\bar{\mathbf{x}} - \eta) \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) \\ &\stackrel{(a)}{=} \sum_{i=1}^n \left(\frac{\mathbf{x}_i - \bar{\mathbf{x}}}{\sigma} \right)^2 + n \left(\frac{\bar{\mathbf{x}} - \eta}{\sigma} \right)^2 \end{aligned}$$

donde en (a) se tuvo en cuenta que el tercer sumando es nulo, ya que

$$\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) = \sum_{i=1}^n \mathbf{x}_i - n\bar{\mathbf{x}} = n \left(\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \right) - n\bar{\mathbf{x}} = n\bar{\mathbf{x}} - n\bar{\mathbf{x}} = 0.$$

Se concluye que

$$\sum_{i=1}^n \left(\frac{\mathbf{x}_i - \eta}{\sigma} \right)^2 = \frac{(n-1)\mathbf{s}^2}{\sigma^2} + \left(\frac{\bar{\mathbf{x}} - \eta}{\sigma/\sqrt{n}} \right)^2. \quad (6.46)$$

Como se demostró previamente, los dos términos de la derecha de la igualdad son independientes. Además, respecto al término a la izquierda de la igualdad, se cumple que

$$\frac{\mathbf{x}_i - \eta}{\sigma} \sim N(0, 1) \quad \stackrel{(a)}{\Rightarrow} \quad \left(\frac{\mathbf{x}_i - \eta}{\sigma} \right)^2 \sim \chi^2(1) \quad \stackrel{(b)}{\Rightarrow} \quad \sum_{i=1}^n \left(\frac{\mathbf{x}_i - \eta}{\sigma} \right)^2 \sim \chi^2(n),$$

donde en (a) se usó el resultado de la sección 4.1.2 y en (b) se usó que la suma de variables aleatorias independientes chi-cuadrado también es chi-cuadrado con grado de libertad la suma de los grados de libertad de las variables que se suman, como se mostró antes (ver *Formas cuadráticas normales*). Respecto al segundo sumando de la derecha de la igualdad de la ecuación 6.46, se cumple que

$$\bar{\mathbf{x}} \sim N(\eta, \sigma^2/n) \quad \Rightarrow \quad \frac{\bar{\mathbf{x}} - \eta}{\sigma/\sqrt{n}} \sim N(0, 1) \quad \Rightarrow \quad \left(\frac{\bar{\mathbf{x}} - \eta}{\sigma/\sqrt{n}} \right)^2 \sim \chi^2(1).$$

Como el lado de la izquierda de la igualdad 6.46 es una variable aleatoria con distribución $\chi^2(n)$ y los sumandos de la derecha de la igualdad son independientes y uno de ellos tiene distribución $\chi^2(1)$, el otro tiene distribución $\chi^2(n-1)$, con lo que se completa la prueba.

Debido a que la variable aleatoria $(n-1)\mathbf{s}^2/\sigma^2$ tiene distribución $\chi^2(n-1)$, su media es $n-1$ y su varianza es $2(n-1)$, como se indicó en la sección 4.5.4. Teniendo esto en cuenta, se calculará $E\{\mathbf{s}^2\}$ y $\text{Var}\{\mathbf{s}^2\}$. Para calcular la media, se observa que

$$E \left\{ \frac{(n-1)\mathbf{s}^2}{\sigma^2} \right\} = \frac{n-1}{\sigma^2} E \{ \mathbf{s}^2 \} = n-1,$$

por lo que despejando, se obtiene que

$$E \{ \mathbf{s}^2 \} = \sigma^2.$$

Para calcular la varianza, se ve que

$$\text{Var} \left\{ \frac{(n-1)\mathbf{s}^2}{\sigma^2} \right\} = \frac{(n-1)^2}{\sigma^4} \text{Var} \{ \mathbf{s}^2 \} = 2(n-1),$$

y por lo tanto,

$$\text{Var}\{\mathbf{s}^2\} = \frac{2\sigma^4}{n-1}.$$

Este último resultado puede comprobarse de la siguiente forma. Como indica la ecuación 6.14, si las variables aleatorias \mathbf{x}_i son independientes e idénticamente distribuidas con momento central cuarto $\mu_4 = E\{(\mathbf{x}_i - \eta)^4\}$, la varianza de la varianza muestral es

$$\text{Var}\{\mathbf{s}^2\} = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right).$$

En este caso, $\mathbf{x}_i \sim N(\eta, \sigma^2)$, por lo que el momento cuarto de \mathbf{x}_i es $\mu_4 = 3\sigma^4$, como se demostró en la sección 4.4.3. Sustituyendo este resultado en la ecuación anterior, se ve que

$$\begin{aligned} \text{Var}\{\mathbf{s}^2\} &= \frac{1}{n} \left(3\sigma^4 - \frac{n-3}{n-1} \sigma^4 \right) \\ &= \frac{[3(n-1) - (n-3)]\sigma^4}{n(n-1)} \\ &= \frac{2n\sigma^4}{n(n-1)} \\ &= \frac{2\sigma^4}{n-1}, \end{aligned}$$

coincidiendo con el resultado obtenido antes.

6.3. Estimación cuadrática media

En esta sección, se explican las ideas principales para estimar una variable aleatoria \mathbf{y} en función de otra variable aleatoria \mathbf{x} . El criterio de estimación óptima es que el error cuadrático medio de la estimación sea mínimo.

Para comenzar, supóngase que se desea estimar el valor de una variable aleatoria $\mathbf{y}(\zeta)$ mediante un número c . Naturalmente, el resultado ζ del experimento no se puede predecir. Sin embargo, como la función distribución $F(y)$ de la variable aleatoria \mathbf{y} determina completamente la estadística de \mathbf{y} , su conocimiento sirve como guía en la elección de c .

Al estimar \mathbf{y} mediante una constante c , en un ensayo particular se obtiene un error de estimación $\mathbf{y}(\zeta) - c$. Así, el problema de determinar c consiste en minimizar el error de estimación en algún sentido. Un criterio razonable en la selección de c podría ser que en una secuencia larga de ensayos, el error promedio se acerque a cero, es decir

$$\frac{[\mathbf{y}(\zeta_1) - c] + \cdots + [\mathbf{y}(\zeta_n) - c]}{n} \simeq 0.$$

Este criterio conduce a que c debe ser la *media* de \mathbf{y} , ya que

$$c = \frac{\mathbf{y}(\zeta_1) + \cdots + \mathbf{y}(\zeta_n)}{n} \rightarrow E\{\mathbf{y}\}.$$

Otro criterio para la elección de c podría ser la minimización del promedio de $|\mathbf{y}(\zeta) - c|$. En este caso, el valor de c óptimo es la *mediana* de \mathbf{y} , como se verá mas adelante.

En el presente análisis se considerarán únicamente estimadores que minimizan el error cuadrático medio de la estimación. En el caso de la elección de c , esto implica minimizar el promedio de $|\mathbf{y}(\zeta) - c|^2$. Este criterio es muy empleado en la práctica no solo por su utilidad, sino que además es matemáticamente tratable y conduce a resultados simples. Como se verá, el valor de c óptimo también es $E\{\mathbf{y}\}$.

Supóngase ahora que en cada ensayo se observa el valor $\mathbf{x}(\zeta)$ que toma la variable aleatoria \mathbf{x} . En este caso, sería mejor emplear como estimación de \mathbf{y} un valor que dependa de la observación $\mathbf{x}(\zeta)$ en cada ensayo en lugar de una constante c para todos los ensayos. En otras palabras, se podría estimar \mathbf{y} como una función $c(\mathbf{x})$ de la variable aleatoria \mathbf{x} . El problema a resolver en este caso es la determinación de la función que conduce al resultado óptimo.

Podría pensarse que si en cierto ensayo se observa $\mathbf{x}(\zeta)$, se puede determinar el resultado ζ del ensayo y consecuentemente, el valor correspondiente $\mathbf{y}(\zeta)$ de \mathbf{y} , pero esto no es así. El mismo valor $\mathbf{x}(\zeta) = x$ es observado para cada resultado ζ en el conjunto $\{\mathbf{x} = x\}$. Si este conjunto contiene varios elementos y los valores de \mathbf{y} son distintos para cada elemento del conjunto, el valor observado $\mathbf{x}(\zeta)$ no determina univocamente a $\mathbf{y}(\zeta)$. Sin embargo, al observar $\mathbf{x}(\zeta)$, se tiene conocimiento que ζ es un elemento del subconjunto $\{\mathbf{x} = x\}$, por lo que la incertidumbre respecto al valor de \mathbf{y} se reduce.

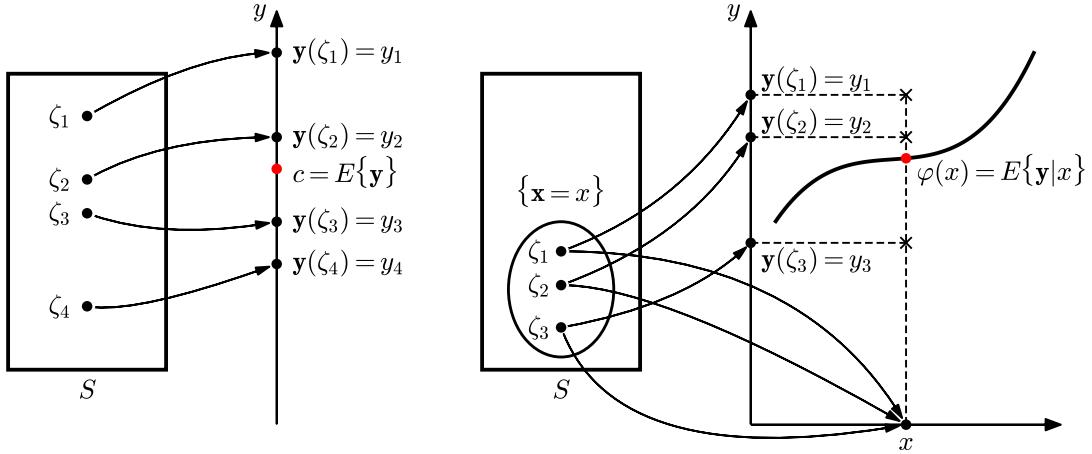


Figura 6.1: Estimación media cuadrática. A la izquierda se esquematiza la estimación media cuadrática de la variable aleatoria \mathbf{y} mediante una constante c . El valor de c que minimiza el error cuadrático medio de estimación es la media de \mathbf{y} . A la derecha se ilustra la estimación de \mathbf{y} a partir de la observación de otra variable aleatoria \mathbf{x} . Si se observa que la variable aleatoria \mathbf{x} toma el valor x , se sabe que el resultado del experimento pertenece al conjunto $\{\mathbf{x} = x\}$, que en el ejemplo es $\{\zeta_1, \zeta_2, \zeta_3\}$. De esta forma, los resultados posibles de la variable aleatoria \mathbf{y} son $\mathbf{y}(\zeta_1), \mathbf{y}(\zeta_2)$ o $\mathbf{y}(\zeta_3)$, y la estimación óptima es $\varphi(x) = E\{\mathbf{y} | x\}$.

Considérese el siguiente ejemplo. Supóngase que el espacio muestral S es el conjunto de n niños en una comunidad y la variable aleatoria \mathbf{y} es la altura de cada niño. Un resultado particular ζ es un niño específico. Supóngase además que cada niño seleccionado es pesado, y la variable aleatoria \mathbf{x} es el peso de cada niño. Al seleccionar un niño ζ_i y observar su peso $\mathbf{x}(\zeta_i) = x$, no queda determinada su altura $\mathbf{y}(\zeta_i) = y_i$ porque puede haber varios niños con igual peso x pero distinta altura. Pero al observar el peso x , se sabe que el niño seleccionado es alguno del subconjunto $\{\mathbf{x} = x\}$ y la incertidumbre se reduce. Por ejemplo, si hubiera solo un niño mas ζ_j de peso x , es decir, $\mathbf{x}(\zeta_j) = x$, al observar el peso x se sabe que el niño puede ser alguno del subconjunto $\{\zeta_i, \zeta_j\}$ del espacio muestral S , es decir, solo pueden ser dos niños de los n niños de la comunidad. De esta forma, el valor de la variable aleatoria \mathbf{y} puede ser solo $\mathbf{y}(\zeta_i) = y_i$ o $\mathbf{y}(\zeta_j) = y_j$. El estimador óptimo de \mathbf{y} es la media condicional $E\{\mathbf{y} | x\}$ de \mathbf{y} dado $\mathbf{x} = x$, donde x es el peso observado. En el caso del ejemplo, el estimador óptimo de la altura es el promedio de y_i e y_j . En el caso de estimar la altura mediante una constante sin observar el peso, el estimador óptimo es el promedio de todas las alturas de los niños, es decir, $E\{\mathbf{y}\}$. Estas consideraciones se ilustran en la figura 6.1.

En el contexto de la teoría de probabilidad, el problema de la estimación cuadrática media (MS, *mean square*) de \mathbf{y} mediante una constante c se plantea como sigue: determinar c de forma tal que el segundo momento e del error de estimación $\mathbf{y} - c$ sea mínimo. El segundo momento del error de estimación es

$$e = E\{(\mathbf{y} - c)^2\} = \int_{-\infty}^{\infty} (y - c)^2 f(y) dy. \quad (6.47)$$

El valor de c que minimiza e se obtiene igualando la derivada de e respecto a c a cero. Esta derivada es

$$\begin{aligned}\frac{de}{dc} &= \frac{d}{dc} \left[\int_{-\infty}^{\infty} (y - c)^2 f(y) dy \right] \\ &= \int_{-\infty}^{\infty} \frac{d}{dc} [(y - c)^2] f(y) dy \\ &= - \int_{-\infty}^{\infty} 2(y - c) f(y) dy,\end{aligned}$$

e igualando a cero, se obtiene que

$$\int_{-\infty}^{\infty} (y - c) f(y) dy = 0 \quad \Leftrightarrow \quad \int_{-\infty}^{\infty} y f(y) dy = c \int_{-\infty}^{\infty} f(y) dy = c.$$

Se concluye que el valor de c que minimiza el error MS de estimación es

$$c = E\{\mathbf{y}\} = \int_{-\infty}^{\infty} y f(y) dy. \quad (6.48)$$

Notar que el mismo razonamiento podría haberse realizado directamente con la esperanza. Efectivamente, la derivada de e respecto a c es

$$\frac{de}{dc} = \frac{d}{dc} E\{(\mathbf{y} - c)^2\} \stackrel{(a)}{=} E \left\{ \frac{d}{dc} [(\mathbf{y} - c)^2] \right\} = E\{-2(\mathbf{y} - c)\} = 2c - 2E\{\mathbf{y}\},$$

donde en (a) se intercambió el orden de la operación derivada y esperanza, que equivale a intercambiar el orden de integración y diferenciación.

Se considera ahora el caso en que se desea estimar \mathbf{y} a partir de la observación de otra variable aleatoria \mathbf{x} . El objetivo es encontrar una función $c(\mathbf{x})$ de forma tal que el error MS sea mínimo. Este problema es referido como *estimación MS no lineal*. El segundo momento e del error de estimación es

$$e = E\{[\mathbf{y} - c(\mathbf{x})]^2\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\mathbf{y} - c(\mathbf{x})]^2 f(x, y) dx dy.$$

Para minimizar e , se considera que $f(x, y) = f(y|x)f(x)$ (ver sección 5.6) y por lo tanto

$$\begin{aligned}e &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\mathbf{y} - c(\mathbf{x})]^2 f(y|x)f(x) dx dy \\ &= \int_{-\infty}^{\infty} f(x) \int_{-\infty}^{\infty} [\mathbf{y} - c(\mathbf{x})]^2 f(y|x) dy dx.\end{aligned}$$

Como los integrandos son positivos, e es mínimo si la integral interna es mínima para todo x . Por lo tanto, para minimizar e alcanza con considerar x fijo y minimizar la integral interna. Como la integral interna es de la forma de la integral 6.47 cambiando c por $c(x)$ y $f(y)$ por $f(y|x)$, el mínimo es de la forma de 6.48 cambiando $f(y)$ por $f(y|x)$, lo que resulta en

$$c(x) = E\{\mathbf{y} | x\} = \int_{-\infty}^{\infty} y f(y|x) dy. \quad (6.49)$$

De esta forma, la función $c(x)$ óptima es la línea de regresión $\varphi(x)$, definida en la ecuación 5.84.

Observar que si $\mathbf{y} = g(\mathbf{x})$, se cumple que $E\{\mathbf{y} | x\} = g(x)$, ya que

$$E\{\mathbf{y} | x\} = \int_{-\infty}^{\infty} y f(y|x) dy = \int_{-\infty}^{\infty} g(x) f(y|x) dy = g(x) \int_{-\infty}^{\infty} f(y|x) dy = g(x).$$

Por lo tanto, $c(x) = g(x) = y$, y el error MS de estimación es nulo. Esto es de esperarse, ya si $\mathbf{y} = g(\mathbf{x})$, \mathbf{y} queda únicamente determinado a partir de \mathbf{x} . Por otro lado, si \mathbf{x} e \mathbf{y} son independientes, $E\{\mathbf{y} | x\} = E\{\mathbf{y}\}$ es constante. En este caso, el conocimiento de \mathbf{x} no aporta nada en la estimación de \mathbf{y} .

6.3.1. Estimación cuadrática media lineal

La solución de la estimación MS no lineal requiere el conocimiento de la función de regresión lineal $\varphi(x)$. Un problema más simple, cuya solución solo requiere el conocimiento de los momentos de segundo orden, es la *estimación MS lineal* de \mathbf{y} en términos de \mathbf{x} , en donde la estimación de \mathbf{y} se realiza a partir de una función lineal $A\mathbf{x} + B$ de \mathbf{x} . En este problema hay que determinar las constantes A y B de forma de minimizar el error MS

$$e = E\{[\mathbf{y} - (A\mathbf{x} + B)]^2\}. \quad (6.50)$$

Se demostrará que $e = e_m$ es mínimo si

$$A = \frac{\mu_{11}}{\mu_{20}} = \frac{r\sigma_y}{\sigma_x}, \quad B = \eta_y - A\eta_x,$$

y el error mínimo es

$$e_m = \mu_{02} - \frac{\mu_{11}^2}{\mu_{20}} = \sigma_y^2(1 - r^2),$$

En estas ecuaciones,

$$\mu_{ij} = E\{(\mathbf{x} - \eta_x)^i(\mathbf{y} - \eta_y)^j\}$$

y r es el coeficiente de correlación, definido en la ecuación 5.50. Además, la segunda igualdad en la ecuación de A se verifica porque

$$A = \frac{\mu_{11}}{\mu_{20}} = \frac{E\{(\mathbf{x} - \eta_x)(\mathbf{y} - \eta_y)\}}{E\{(\mathbf{x} - \eta_x)^2\}} \stackrel{(a)}{=} \frac{C_{xy}}{\sigma_x^2} \stackrel{(b)}{=} \frac{r\sigma_x\sigma_y}{\sigma_x^2} = \frac{r\sigma_y}{\sigma_x},$$

donde en (a) se usó la definición de covarianza, dada en la ecuación 5.49 y en (b) se usó la ecuación 5.50. De forma similar, la segunda igualdad en la ecuación del error MS mínimo e_m se verifica porque

$$e_m = \mu_{02} - \frac{\mu_{11}^2}{\mu_{20}} = \sigma_y^2 - \frac{C_{xy}^2}{\sigma_x^2} = \sigma_y^2 - \frac{r^2\sigma_x^2\sigma_y^2}{\sigma_x^2} = \sigma_y^2 - r^2\sigma_y^2.$$

Para demostrar estos resultados, se parte observando en la ecuación 6.50 del error MS e , que para un valor dado de A , e es el error MS de la estimación de la variable aleatoria $\mathbf{y} - A\mathbf{x}$ mediante de la constante B . Por lo tanto, como indica la ecuación 6.48, e es mínimo si

$$B = E\{\mathbf{y} - A\mathbf{x}\} = E\{\mathbf{y}\} - AE\{\mathbf{x}\} = \eta_y - A\eta_x,$$

con lo que queda determinado B . Para determinar A , se sustituye el valor de B en la ecuación 6.50,

$$\begin{aligned} e &= E\{[\mathbf{y} - (A\mathbf{x} + \eta_y - A\eta_x)]^2\} \\ &= E\{[(\mathbf{y} - \eta_y) - A(\mathbf{x} - \eta_x)]^2\} \\ &= E\{(\mathbf{y} - \eta_y)^2\} + A^2E\{(\mathbf{x} - \eta_x)^2\} - 2AE\{(\mathbf{x} - \eta_x)(\mathbf{y} - \eta_y)\} \\ &= \sigma_y^2 + A^2\sigma_x^2 - 2AC_{xy}, \end{aligned}$$

resultando en

$$e = \sigma_y^2 + A^2\sigma_x^2 - 2Ar\sigma_x\sigma_y. \quad (6.51)$$

Igualando la derivada de e respecto a A a cero,

$$\frac{de}{dA} = 2A\sigma_x^2 - 2r\sigma_x\sigma_y = 0,$$

y despejando A , se obtiene que

$$A = \frac{r\sigma_y}{\sigma_x}.$$

Finalmente, el error mínimo se obtiene sustituyendo el valor de A óptimo en la ecuación 6.51,

$$e_m = \sigma_y^2 + \left(\frac{r\sigma_y}{\sigma_x}\right)^2\sigma_x^2 - 2\left(\frac{r\sigma_y}{\sigma_x}\right)r\sigma_x\sigma_y = \sigma_y^2 + r^2\sigma_y^2 - 2r^2\sigma_y^2 = \sigma_y^2(1 - r^2).$$

Terminología El estimador $A\mathbf{x} + B$ se dice estimador lineal *no homogéneo* de \mathbf{y} en términos de \mathbf{x} . Si \mathbf{y} es estimada por una línea recta $A\mathbf{x}$ que pasa por el origen, el estimador se dice *homogéneo*.

La variable aleatoria \mathbf{x} es el dato de la estimación, la variable aleatoria $\varepsilon = \mathbf{y} - (A\mathbf{x} + B)$ es el *error* de estimación y el valor $e = E\{\varepsilon^2\}$ es el *error cuadrático medio* de estimación.

Variable aleatorias conjuntamente normales En general, el estimador no lineal $\varphi(x) = E\{\mathbf{y} | x\}$ de \mathbf{y} en términos de \mathbf{x} no es una recta, y el error MS de estimación $E\{[\mathbf{y} - \varphi(\mathbf{x})]^2\}$ resultante es menor que el error MS e_m del estimador lineal $A\mathbf{x} + B$. Sin embargo, si las variables aleatorias \mathbf{x} e \mathbf{y} son conjuntamente normales, se cumple que

$$\varphi(x) = E\{\mathbf{y} | x\} = \frac{r\sigma_y x}{\sigma_x} + \eta_y - \frac{r\sigma_y \eta_x}{\sigma_x},$$

como se puede deducir de la ecuación 5.79 de $f(x | y)$ cambiando x por y . Observando de $\varphi(x)$ es una recta, se concluye que si las variables aleatorias son conjuntamente normales, el estimador no lineal y el estimador lineal son idénticos.

6.3.2. Principio de ortogonalidad

El *principio de ortogonalidad* indica que el estimador lineal óptimo $A\mathbf{x} + B$ de la variable \mathbf{y} es tal que el error de estimación $\mathbf{y} - (A\mathbf{x} + B)$ es ortogonal al dato \mathbf{x} . Matemáticamente, esto se expresa como

$$E\{[\mathbf{y} - (A\mathbf{x} + B)]\mathbf{x}\} = 0. \quad (6.52)$$

Para demostrarlo, se observa que los valores de A y B que conducen al error MS mínimo se obtienen resolviendo el sistema de ecuaciones $\partial e / \partial A = 0$ y $\partial e / \partial B = 0$, con e dado por la ecuación 6.50. Derivando el error MS respecto a A , se obtiene que

$$\begin{aligned} \frac{\partial e}{\partial A} &= \frac{\partial}{\partial A} (E\{[\mathbf{y} - (A\mathbf{x} + B)]^2\}) \\ &\stackrel{(a)}{=} E\left\{\frac{\partial}{\partial A} [\mathbf{y} - (A\mathbf{x} + B)]^2\right\} \\ &= E\{2[\mathbf{y} - (A\mathbf{x} + B)](-A\mathbf{x})\} \\ &= -2AE\{[\mathbf{y} - (A\mathbf{x} + B)]\mathbf{x}\}, \end{aligned}$$

donde en (a) se intercambió el orden de las operaciones de diferenciación y esperanza. Igualando el resultado a 0, se obtiene la ecuación 6.52.

Estimación lineal MS homogénea Se pretende encontrar la constante a de forma tal que si \mathbf{y} es estimada a partir de $a\mathbf{x}$, el error MS

$$e = E\{(\mathbf{y} - a\mathbf{x})^2\}. \quad (6.53)$$

sea mínimo. Derivando respecto a a e igualando el resultado a cero, se deduce que a debe cumplir que

$$E\{(\mathbf{y} - a\mathbf{x})\mathbf{x}\} = 0. \quad (6.54)$$

Despejando a , se obtiene que

$$E\{\mathbf{xy}\} = aE\{\mathbf{x}^2\} \quad \Rightarrow \quad a = \frac{E\{\mathbf{xy}\}}{E\{\mathbf{x}^2\}}.$$

El estimador MS lineal de \mathbf{y} en términos de \mathbf{x} se denominará como $\hat{E}\{\mathbf{y} | \mathbf{x}\}$. Por lo tanto, en la estimación lineal homogénea se tiene que

$$\hat{E}\{\mathbf{y} | \mathbf{x}\} = a\mathbf{x}, \quad a = \frac{E\{\mathbf{xy}\}}{E\{\mathbf{x}^2\}}.$$

Se deducirá a continuación la expresión del error MS mínimo. Desarrollando la ecuación 6.53, el error MS puede expresarse por un lado como

$$e = E\{(\mathbf{y} - a\mathbf{x})(\mathbf{y} - a\mathbf{x})\} = E\{(\mathbf{y} - a\mathbf{x})\mathbf{y}\} - aE\{(\mathbf{y} - a\mathbf{x})\mathbf{x}\}.$$

Por otro lado, e también se puede escribir como

$$e = E\{\mathbf{y}^2\} - E\{(a\mathbf{x})^2\} - 2aE\{(\mathbf{y} - a\mathbf{x})\mathbf{x}\},$$

ya que, operando sobre el lado derecho de la igualdad, se observa que

$$\begin{aligned} E\{\mathbf{y}^2\} - E\{(a\mathbf{x})^2\} - 2aE\{(\mathbf{y} - a\mathbf{x})\mathbf{x}\} &= E\{\mathbf{y}^2\} - E\{(a\mathbf{x})^2\} - E\{2\mathbf{y}(a\mathbf{x})\} + 2E\{(a\mathbf{x})^2\} \\ &= E\{\mathbf{y}^2\} + E\{(a\mathbf{x})^2\} - E\{2\mathbf{y}(a\mathbf{x})\} \\ &= E\{\mathbf{y}^2 + (a\mathbf{x})^2 - 2\mathbf{y}(a\mathbf{x})\} \\ &= E\{(\mathbf{y} - a\mathbf{x})^2\} \\ &= e. \end{aligned}$$

Igualando ambas ecuaciones, se tiene que el error MS cumple que

$$e = E\{(\mathbf{y} - a\mathbf{x})\mathbf{y}\} - aE\{(\mathbf{y} - a\mathbf{x})\mathbf{x}\} = E\{\mathbf{y}^2\} - E\{(a\mathbf{x})^2\} - 2aE\{(\mathbf{y} - a\mathbf{x})\mathbf{x}\}.$$

Como el error MS mínimo se obtiene con el valor de a que verifica la ecuación 6.54, se concluye que el error MS mínimo cumple que

$$e = E\{(\mathbf{y} - a\mathbf{x})\mathbf{y}\} = E\{\mathbf{y}^2\} - E\{(a\mathbf{x})^2\}. \quad (6.55)$$

Se realizarán algunas consideraciones sobre el principio de ortogonalidad. La ecuación 6.54 se trata del principio de ortogonalidad, ya que indica que el error de estimación $\mathbf{y} - a\mathbf{x}$ es ortogonal al dato \mathbf{x} . Para una interpretación geométrica del principio de ortogonalidad, observar la figura 6.2 de la representación vectorial de las variables aleatorias involucradas. Por ser a una constante, la estimación $a\mathbf{x}$ de \mathbf{y} es colineal con \mathbf{x} . La diferencia $\mathbf{y} - a\mathbf{x}$ es el vector que va desde $a\mathbf{x}$ hasta \mathbf{y} . El largo de este vector es $\sqrt{E\{(\mathbf{y} - a\mathbf{x})^2\}} = \sqrt{e}$. Claramente, esta distancia es mínima si se elige a de forma tal que $\mathbf{y} - a\mathbf{x}$ es perpendicular con \mathbf{x} , como indica la ecuación 6.54. Observar además que el lado derecho de la ecuación 6.55 es teorema de Pitágoras. Efectivamente, en el triángulo rectángulo en la figura, el cuadrado del largo de la hipotenusa es $E\{\mathbf{y}^2\}$ y el cuadrado del largo de los catetos es $E\{(a\mathbf{x})^2\}$ y e . La primera igualdad de la ecuación 6.55 indica que el cuadrado del largo de $\mathbf{y} - a\mathbf{x}$ es el producto interno de \mathbf{y} con el error $\mathbf{y} - a\mathbf{x}$.

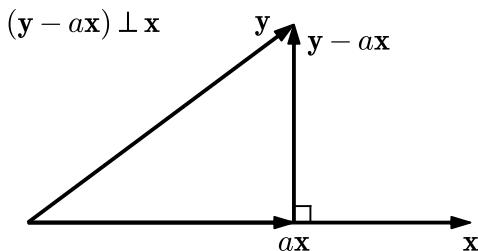


Figura 6.2: Representación vectorial de las variables aleatorias en la estimación MS lineal homogénea de \mathbf{y} en términos de \mathbf{x} . El valor de a que minimiza el largo del vector $\mathbf{y} - a\mathbf{x}$, que representa el error MS de estimación, es el que hace que el vector $\mathbf{y} - a\mathbf{x}$ sea perpendicular con \mathbf{x} .

Funciones de riesgo y de pérdida Se realizarán algunas consideraciones sobre otros criterios de optimización en la estimación de una variable aleatoria \mathbf{y} mediante una constante c . Se elige una función $L(x)$ y se determina c de forma de minimizar la media de la variable aleatoria $L(\mathbf{y} - c)$,

$$R = E\{L(\mathbf{y} - c)\} = \int_{-\infty}^{\infty} L(y - c)f(y) dy.$$

La función $L(x)$ se llama *función de pérdida* y la constante R se llama *riesgo promedio*. La elección de $L(x)$ depende de la aplicación. En el caso en que $L(x) = x^2$, $R = E\{(\mathbf{y} - c)^2\}$ es el error MS, el cual se minimiza si $c = E\{\mathbf{y}\}$.

Si $L(x) = |x|$, $R = E\{|y - c|\}$. Se mostrará que en este caso, el valor de c que minimiza R es la *mediana* $y_{0.5}$ de y . Para ver esto, se comienza planteando el riesgo promedio,

$$R = \int_{-\infty}^{\infty} |y - c| f(y) dy \stackrel{(a)}{=} \int_{-\infty}^c (c - y) f(y) dy + \int_c^{\infty} (y - c) f(y) dy,$$

donde en (a) se tuvo en cuenta que

$$|y - c| = \begin{cases} y - c, & y \geq c \\ c - y, & c > y. \end{cases}$$

Aplicando la regla de integración de Leibnitz (ecuaciones 5.15 y 5.16) para diferenciar R respecto a c , se tiene que

$$\frac{dR}{dc} = \int_{-\infty}^c f(y) dy - \int_c^{\infty} f(y) dy = F(c) - [1 - F(c)] = 2F(c) - 1,$$

y al igualar la derivada a cero, se llega a que R es mínimo si $F(c) = 1/2$, es decir, si $c = y_{0.5}$.

6.3.3. Estimación cuadrática media lineal: caso general

El estimador lineal MS de la variable aleatoria \mathbf{s} en términos de las n variables aleatorias \mathbf{x}_i es

$$\hat{\mathbf{s}} = a_1 \mathbf{x}_1 + \cdots + a_n \mathbf{x}_n,$$

donde a_1, \dots, a_n son constantes tales que el valor MS del error de estimación

$$P = E\{(\mathbf{s} - \hat{\mathbf{s}})^2\} = E\{[s - (a_1 \mathbf{x}_1 + \cdots + a_n \mathbf{x}_n)]^2\} \quad (6.56)$$

sea mínimo.

Se demostrará el *principio de ortogonalidad* para este caso, que indica que P es mínimo si el error $\mathbf{s} - \hat{\mathbf{s}}$ es ortogonal a los datos \mathbf{x}_i , es decir,

$$E\{[s - (a_1 \mathbf{x}_1 + \cdots + a_n \mathbf{x}_n)] \mathbf{x}_i\} = 0, \quad i = 1, \dots, n. \quad (6.57)$$

Para demostrarlo, basta igualar a cero las derivadas de P respecto a las constantes a_i ,

$$\frac{dP}{da_i} = E\{-2[\mathbf{s} - (a_1 \mathbf{x}_1 + \cdots + a_n \mathbf{x}_n)] \mathbf{x}_i\} = 0,$$

resultando en la ecuación 6.57. Este resultado se denomina *teorema de la proyección*. Para obtener las constantes que minimizan el error MS de estimación, se establece $i = 1, \dots, n$ en la ecuación 6.57, resultando en el sistema de ecuaciones

$$\begin{cases} R_{11}a_1 + R_{21}a_2 + \cdots + R_{n1}a_n = R_{01} \\ R_{12}a_1 + R_{22}a_2 + \cdots + R_{n2}a_n = R_{02} \\ \vdots \quad \vdots \quad \vdots \quad \vdots \\ R_{1n}a_1 + R_{2n}a_2 + \cdots + R_{nn}a_n = R_{0n} \end{cases}, \quad (6.58)$$

donde $R_{ij} = E\{\mathbf{x}_i \mathbf{x}_j\}$ y $R_{0j} = E\{\mathbf{s} \mathbf{x}_j\}$. Definiendo los vectores fila

$$\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n], \quad A = [a_1, \dots, a_n], \quad R_0 = [R_{01}, \dots, R_{0n}]$$

y la matriz de correlación de los datos $R = \mathbf{X}^t \mathbf{X}$, el sistema de ecuaciones se puede expresar como

$$AR = R_0,$$

y su solución es

$$A = R_0 R^{-1}. \quad (6.59)$$

Sustituyendo la ecuación 6.59 en la ecuación 6.56 se obtiene el error MS mínimo de estimación (LMS, *least mean square*). Para hacerlo, se parte expresando la ecuación 6.56 de forma vectorial y se opera como sigue,

$$\begin{aligned}
 P &= E\{(s - A\mathbf{X}^t)^2\} \\
 &= E\{s^2\} - 2AE\{s\mathbf{X}^t\} + E\{(A\mathbf{X}^t)^2\} \\
 &\stackrel{(a)}{=} E\{s^2\} - 2AR_0^t + AE\{\mathbf{X}^t\mathbf{X}\}A^t \\
 &= E\{s^2\} - 2AR_0^t + ARA^t,
 \end{aligned}$$

donde en (a) se empleó que $E\{s\mathbf{X}^t\} = R_0^t$ en el segundo sumando, como fue definido previamente, y en el tercer sumando se observó que como $A\mathbf{X}^t$ es un escalar, se cumple que $(A\mathbf{X}^t)^2 = (A\mathbf{X}^t)(A\mathbf{X}^t)^t = A\mathbf{X}^t\mathbf{X}A^t$. Para obtener el error LMS, se sustituye A^t en el tercer sumando de P por los coeficientes óptimos, dados por la ecuación 6.59. De esta forma, ese tercer sumando queda

$$ARA^t = AR(R_0R^{-1})^t = AR(R^{-1})^t R_0^t = AR(R^t)^{-1}R_0^t \stackrel{(a)}{=} ARR^{-1}R_0^t = AR_0^t,$$

donde en (a) se consideró que la matriz R es simétrica y por lo tanto $R^t = R$. Se concluye que el error LMS es

$$P = E\{s^2\} - AR_0^t. \quad (6.60)$$

Otra forma de obtener el error LMS, es observando que como $s - \hat{s} \perp \mathbf{x}_i$ para todo i , se cumple que $s - \hat{s} \perp \hat{s}$. Efectivamente,

$$\begin{aligned}
 E\{(s - \hat{s})\hat{s}\} &= E\{(s - \hat{s})(a_1\mathbf{x}_1 + \dots + a_n\mathbf{x}_n)\} \\
 &= a_1E\{(s - \hat{s})\mathbf{x}_1\} + \dots + a_nE\{(s - \hat{s})\mathbf{x}_n\} \\
 &= 0.
 \end{aligned}$$

Teniendo esto en cuenta, la ecuación 6.56 del error MS queda

$$P = E\{(s - \hat{s})(s - \hat{s})\} = E\{(s - \hat{s})s\} - E\{(s - \hat{s})\hat{s}\} = E\{(s - \hat{s})s\},$$

y considerando que

$$E\{\hat{s}\hat{s}\} = E\{A\mathbf{X}^t s\} = AE\{\mathbf{X}^t s\} = AR_0^t,$$

se obtiene que

$$P = E\{(s - \hat{s})s\} = E\{s^2\} - AR_0^t.$$

Como interpretación geométrica de este resultado, se observa que en la representación de las variables aleatorias en un espacio vectorial, la suma $a_1\mathbf{x}_1 + \dots + a_n\mathbf{x}_n$ es un vector del subespacio S_n de los datos \mathbf{x}_i . El error $\varepsilon = s - \hat{s}$ es el vector que va desde \hat{s} a s , como se muestra en la figura 6.3. El teorema de la proyección indica que el largo de ε es mínimo si ε es ortogonal a los datos \mathbf{x}_i , es decir, si es perpendicular al subespacio S_n . Por lo tanto, el estimador \hat{s} es la proyección de s en el subespacio S_n .

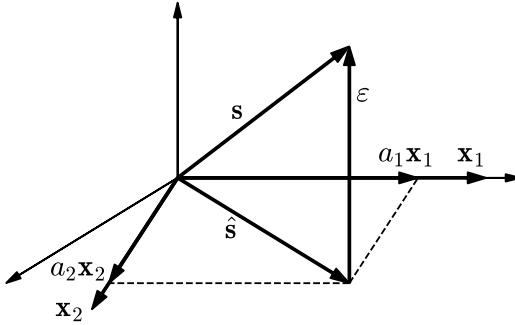


Figura 6.3: Representación vectorial de las variables aleatorias en el caso general de la estimación MS lineal. En el ejemplo de la figura, se estima s en términos de \mathbf{x}_1 y \mathbf{x}_2 . El estimador LMS es la proyección de s en el plano determinado por \mathbf{x}_1 y \mathbf{x}_2 . El error de estimación $\varepsilon = s - \hat{s}$ es perpendicular a dicho plano.

Se consideran a continuación dos casos particulares. Si s es un vector del subespacio S_n , $\hat{s} = s$, resultando en $P = 0$. En este caso, las $n+1$ variables aleatorias $s, \mathbf{x}_1, \dots, \mathbf{x}_n$ son linealmente dependientes y el determinante Δ_{n+1} de la matriz de correlación es cero. Por otro lado, si s es perpendicular a S_n , $\hat{s} = 0$, resultando en $P = E\{|s|^2\}$. En esta situación, s es ortogonal a los datos \mathbf{x}_i , y $R_{0j} = 0$ si $j \neq 0$.

6.3.4. Estimación cuadrática media no lineal: caso general

El problema de estimación MS no lineal consiste en determinar la función $g(\mathbf{x}_1, \dots, \mathbf{x}_n) = g(\mathbf{X})$ de los datos \mathbf{x}_i de forma de minimizar el error MS

$$P = E\{[\mathbf{s} - g(\mathbf{X})]^2\}. \quad (6.61)$$

Se demostrará que P es mínimo si

$$g(X) = E\{\mathbf{s} | X\} = \int_{-\infty}^{\infty} s f_s(s | X) ds. \quad (6.62)$$

$E\{\mathbf{s} | X\}$, la media condicional de la variable \mathbf{s} asumiendo que $\mathbf{X} = X$, es una función de X y se denomina superficie de regresión.

Para demostrar que el estimador MS está dado por la ecuación 6.62 se parte de la siguiente identidad

$$P = E\{[\mathbf{s} - g(\mathbf{X})]^2\} = E\{E\{[\mathbf{s} - g(\mathbf{X})]^2 | \mathbf{X}\}\},$$

la cual se justifica con lo indicado en las secciones 5.7.1 y 6.2.1. El último término en esta igualdad expresado explícitamente como una integral es

$$P = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} [s - g(X)]^2 f_s(s | X) ds \right\} f(X) dX,$$

donde la integral interna corresponde a la esperanza condicional interna, y la integral externa es una integral múltiple de multiplicidad n , es decir, la cantidad de datos \mathbf{x}_i . Como todas las funciones del integrando son positivas, P es mínimo si la integral interna es mínima para todo X . Por lo tanto, para minimizar P , se considera X fijo y se minimiza el error MS condicional

$$E\{[\mathbf{s} - g(\mathbf{X})]^2 | X\} = \int_{-\infty}^{\infty} [s - g(X)]^2 f_s(s | X) ds.$$

Como esta integral es análoga a la integral que se minimizó en la deducción de la ecuación 6.49, se concluye que la integral es mínima si $g(X)$ está dada por la ecuación 6.62.

Principio de ortogonalidad general Del teorema de la proyección dado por la ecuación 6.57 se deduce que para el estimador MS lineal $\hat{\mathbf{s}}$ se cumple que

$$E\{[\mathbf{s} - \hat{\mathbf{s}}](c_1 \mathbf{x}_1 + \dots + c_n \mathbf{x}_n)\} = 0$$

para cualquier conjunto de valores c_1, \dots, c_n . Esto indica que si $\hat{\mathbf{s}}$ es el estimador lineal de \mathbf{s} , el error de estimación $\mathbf{s} - \hat{\mathbf{s}}$ es ortogonal a cualquier función lineal $\mathbf{y} = c_1 \mathbf{x}_1 + \dots + c_n \mathbf{x}_n$ de los datos \mathbf{x}_i .

Se demostrará que si $g(\mathbf{X})$ es el estimador MS no lineal de \mathbf{s} , el error de estimación $\mathbf{s} - g(\mathbf{X})$ es ortogonal a cualquier función $w(\mathbf{X})$, lineal o no lineal, de los datos \mathbf{x}_i , es decir,

$$E\{[\mathbf{s} - g(\mathbf{X})]w(\mathbf{X})\} = 0. \quad (6.63)$$

Se demostrará previamente la identidad

$$E\{[\mathbf{s} - g(\mathbf{X})]w(\mathbf{X})\} = E\{w(\mathbf{X})E\{[\mathbf{s} - g(\mathbf{X})] | \mathbf{X}\}\}, \quad (6.64)$$

que es una generalización de las esperanzas condicionales deducidas en la sección 5.7.1. En esta identidad, la esperanza condicional interna del lado derecho es la variable aleatoria

$$E\{[\mathbf{s} - g(\mathbf{X})] | \mathbf{X}\} = \int_{-\infty}^{\infty} [s - g(X)] f_s(s | \mathbf{X}) ds.$$

Al multiplicar por $w(\mathbf{X})$ y tomar esperanza, se tiene que

$$\begin{aligned} E\{w(\mathbf{X})E\{[\mathbf{s} - g(\mathbf{X})] | \mathbf{X}\}\} &= \int_{-\infty}^{\infty} w(X) \left\{ \int_{-\infty}^{\infty} [s - g(X)] f_s(s | X) ds \right\} f(X) dX \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w(X) [s - g(X)] f_s(s | X) f(X) ds dX \end{aligned}$$

y considerando que $f_s(s|X)f(X) = f(s, X)$ (ecuación 6.21), se obtiene que

$$\begin{aligned} E\{w(\mathbf{X})E\{[\mathbf{s} - g(\mathbf{X})]|\mathbf{X}\}\} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w(X)[s - g(X)]f(s, X) ds dX \\ &= E\{[\mathbf{s} - g(\mathbf{X})]w(\mathbf{X})\}, \end{aligned}$$

que es lo que se quería demostrar. Para demostrar la ecuación 6.63 se opera sobre la esperanza interna en el lado derecha de la identidad 6.64,

$$E\{[\mathbf{s} - g(\mathbf{X})]|\mathbf{X}\} \stackrel{(a)}{=} E\{\mathbf{s}|X\} - E\{g(\mathbf{X})|X\} \stackrel{(b)}{=} E\{\mathbf{s}|X\} - g(X) \stackrel{(c)}{=} 0,$$

donde en (a) se aplicó la propiedad de linealidad de la esperanza, en (b) se consideró que $E\{g(\mathbf{X})|X\} = E\{g(X)\} = g(X)$ y en (c) se consideró que $g(X)$ es el estimador MS no lineal y por lo tanto se cumple la ecuación 6.62. Se concluye que el término de la derecha de la identidad 6.64 es nulo, resultando en la ecuación 6.63.

Normalidad Se demostrará que si las variables aleatorias $\mathbf{s}, \mathbf{x}_1, \dots, \mathbf{x}_n$ son conjuntamente normales con media nula, el estimador lineal y el estimador no lineal de \mathbf{s} son iguales,

$$\hat{\mathbf{s}} = a_1 \mathbf{x}_1 + \dots + \mathbf{x}_n = g(\mathbf{X}) = E\{\mathbf{s}|\mathbf{X}\}. \quad (6.65)$$

Para demostrarlo, hay que demostrar que $\hat{\mathbf{s}} = E\{\mathbf{s}|\mathbf{X}\}$. Como $\mathbf{s} - \hat{\mathbf{s}}$ es una combinación lineal de variables aleatorias conjuntamente normales, las variables aleatorias $\mathbf{s} - \hat{\mathbf{s}}$ y \mathbf{x}_i son conjuntamente normales y $\mathbf{s} - \hat{\mathbf{s}}$ tiene media nula. Además, como $\mathbf{s} - \hat{\mathbf{s}}$ es el error de estimación mínimo, es ortogonal a los datos \mathbf{x}_i y se cumple que $E\{(\mathbf{s} - \hat{\mathbf{s}})\mathbf{x}_i\} = 0$. Esto implica que su covarianza es nula y por lo tanto son no correlacionadas. Se concluye que $\mathbf{s} - \hat{\mathbf{s}}$ y \mathbf{x}_i son variables aleatorias no correlacionadas, ya que su covarianza es nula, y conjuntamente normales, y por lo tanto, independientes, como se explicó en la sección 5.4.2. Entonces, por un lado se tiene que

$$E\{\mathbf{s} - \hat{\mathbf{s}}|X\} = E\{\mathbf{s}|X\} - E\{\hat{\mathbf{s}}|X\} = E\{\mathbf{s}|X\} - \hat{\mathbf{s}},$$

donde en la última igualdad se consideró que $\hat{\mathbf{s}}$ es una función de \mathbf{X} , por lo que $E\{\hat{\mathbf{s}}(\mathbf{X})|X\} = E\{\hat{\mathbf{s}}(X)\} = \hat{\mathbf{s}}$. Además, por otro lado se tiene que

$$E\{\mathbf{s} - \hat{\mathbf{s}}\} = E\{\mathbf{s} - \hat{\mathbf{s}}\} = E\{\mathbf{s}\} - E\{\hat{\mathbf{s}}\} = 0,$$

donde en la primera igualdad se consideró que $\mathbf{s} - \hat{\mathbf{s}}$ y \mathbf{X} son variables aleatorias independientes y en la última igualdad, que $E\{\mathbf{s}\} = E\{\hat{\mathbf{s}}\} = 0$. Combinando los dos resultados, se obtiene que $\hat{\mathbf{s}} = E\{\mathbf{s}|X\}$, que es lo que se quería demostrar.

Densidades condicionales de variables aleatorias normales Se emplearán los resultados anteriores para determinar de forma simple las densidades de probabilidad condicional de variables aleatorias normales. La densidad de probabilidad condicional de \mathbf{s} dado \mathbf{X} es

$$f(s|X) = \frac{f(s, X)}{f(X)},$$

y como consiste en el cociente de dos exponentiales de exponentes cuadráticos, también es una densidad normal. Para establecerla, basta con determinar la media $\mu_{s|X}$ y la varianza $\sigma_{s|X}^2$. La media es

$$\mu_{s|X} = E\{\mathbf{s}|X\} = \hat{\mathbf{s}},$$

como indica la ecuación 6.65. La varianza es

$$\sigma_{s|X}^2 = E\{(\mathbf{s} - \hat{\mathbf{s}})^2|X\} \stackrel{(a)}{=} E\{(\mathbf{s} - \hat{\mathbf{s}})^2\} \stackrel{(b)}{=} P,$$

donde en (a) se consideró que $\mathbf{s} - \hat{\mathbf{s}}$ y \mathbf{X} son variables aleatorias independientes y en (b) se empleó la ecuación 6.56, donde P es el error MS mínimo. Por lo tanto, se concluye que

$$f(s|x_1, \dots, x_n) = \frac{1}{\sqrt{2\pi P}} e^{-[s - (a_1 x_1 + \dots + a_n x_n)]^2/2P},$$

donde los coeficientes a_i se obtienen mediante la ecuación 6.59 y el error MS mínimo mediante la ecuación 6.60.

Ejemplo (7-12) Se consideran las variables aleatorias \mathbf{x}_1 , \mathbf{x}_2 y \mathbf{x}_3 conjuntamente normales y de media nula. Se quiere determinar la densidad de probabilidad $f(x_3 | x_1, x_2)$.

Por la argumentación indicada previamente, $f(x_3 | x_1, x_2)$ es una densidad de probabilidad normal, y para determinarla, alcanza con calcular la media $\mu_{x_3|x_1,x_2}$ y la varianza $\sigma_{x_3|x_1,x_2}^2$. La media es el estimador MS lineal,

$$\mu_{x_3|x_1,x_2} = E\{\mathbf{x}_3 | \mathbf{x}_1, \mathbf{x}_2\} = \hat{\mathbf{s}} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2,$$

por lo que para determinar los coeficientes a_1 y a_2 hay que resolver el sistema 6.58, que en el caso de dos coeficientes queda

$$\begin{cases} R_{11}a_1 + R_{12}a_2 = R_{13} \\ R_{12}a_1 + R_{22}a_2 = R_{23} \end{cases},$$

donde $R_{ij} = E\{\mathbf{x}_i \mathbf{x}_j\} = R_{ji}$. Al resolver el sistema, se obtiene que

$$a_1 = \frac{R_{12}R_{23} - R_{22}R_{13}}{R_{12}^2 - R_{11}R_{22}}, \quad a_2 = \frac{R_{12}R_{13} - R_{11}R_{23}}{R_{12}^2 - R_{11}R_{22}}.$$

La varianza se obtiene mediante la ecuación 6.60, y en este caso queda

$$\sigma_{x_3|x_1,x_2}^2 = P = R_{33} - (a_1 R_{13} + a_2 R_{23}).$$

De esta forma la densidad buscada es

$$f(x_3 | x_1, x_2) = \frac{1}{\sqrt{2\pi P}} e^{-[x_3 - (a_1 x_1 + a_2 x_2)]^2 / 2P}.$$

6.4. Convergencia estocástica y teoremas de límite

Un problema importante en la teoría de la probabilidad es el estudio de las propiedades asintóticas de secuencias de variables aleatorias. En esta sección se introducen los conceptos asociados, haciendo énfasis en las definiciones de los distintos modos de convergencia y la relación entre ellos.

Ejemplo Considérese el caso en que se quiere medir el largo a de un objeto. Debido a los errores de medida, la lectura del instrumento es

$$\mathbf{x} = a + \mathbf{v},$$

donde \mathbf{v} es un término de error. Si no hay errores sistemáticos de medición, \mathbf{v} es una variable aleatoria de media nula. Si la desviación estándar σ de \mathbf{v} es pequeña en comparación a a , el valor observado $\mathbf{x}(\zeta)$ de \mathbf{x} en una única medida es un buen estimador del parámetro desconocido a . En el contexto de la teoría de probabilidad, esta conclusión se puede expresar cuantitativamente como sigue: la media de la variable aleatoria \mathbf{x} es a , ya que

$$E\{\mathbf{x}\} = E\{a + \mathbf{v}\} \stackrel{(a)}{=} E\{a\} + E\{\mathbf{v}\} \stackrel{(b)}{=} a,$$

donde en (a) se empleó la propiedad de linealidad de la esperanza y en (b) que como a es una constante, $E\{a\} = a$, y también que el error de medición es de media nula, $E\{\mathbf{v}\} = 0$. Además, la varianza σ_x^2 de \mathbf{x} es σ^2 , ya que

$$\sigma_x^2 = E\{(\mathbf{x} - E\{\mathbf{x}\})^2\} = E\{(a + \mathbf{v} - a)^2\} = E\{\mathbf{v}^2\} = \sigma^2,$$

donde en la última igualdad se empleó que \mathbf{v} es de media nula. Por lo tanto, a partir del complemento de la desigualdad de Chebyshev (ecuación 4.22), se tiene que

$$P\{|\mathbf{x} - a| < \epsilon\} \geq 1 - \frac{\sigma^2}{\epsilon^2}.$$

De esta forma, si $\sigma \ll \epsilon$, la probabilidad de que $|\mathbf{x} - a|$ sea menor que ϵ es cercana a 1. Esto significa que “casi certamente” el valor observado $\mathbf{x}(\zeta)$ estará entre $a - \epsilon$ y $a + \epsilon$ o equivalentemente, el parámetro desconocido a estará entre $\mathbf{x}(\zeta) - \epsilon$ y $\mathbf{x}(\zeta) + \epsilon$. En resumen, una única medida $\mathbf{x}(\zeta)$ es un estimador de a “casi certamente” satisfactorio si $\sigma \ll a$. Si σ no es pequeño en comparación a a , una única medida no es suficiente para obtener una buen estimador de a . En ese caso, para incrementar la exactitud de

la medición, pueden realizarse varias lecturas y promediar los resultados. Supóngase que se realizan n lecturas independientes. La lectura i -ésima es

$$\mathbf{x}_i = a + \mathbf{v}_i,$$

donde los componentes de error \mathbf{v}_i son variables aleatorias independientes de media nula y varianza σ^2 . La media muestral

$$\bar{\mathbf{x}} = \frac{\mathbf{x}_1 + \cdots + \mathbf{x}_n}{n}$$

de las lecturas tiene media a y varianza σ^2/n , como se explicó en la sección 6.1.3. Por lo tanto, si n es suficientemente grande de forma que $\sigma^2 \ll na$, el valor $\bar{\mathbf{x}}(\zeta)$ de la media muestral $\bar{\mathbf{x}}$ en un único experimento de n medidas es un buen estimador de a . Supóngase que la cantidad de lecturas n es tal que $\sigma^2/na^2 = 10^{-4}$ y se quiere calcular la probabilidad de que $\bar{\mathbf{x}}$ esté entre $0.9a$ y $1.1a$. Para hacerlo, se recurre a la desigualdad de Chebyshev (ecuación 4.22) y se observa que

$$|\bar{\mathbf{x}} - a| < \epsilon \Leftrightarrow -\epsilon < \bar{\mathbf{x}} - a < \epsilon \Leftrightarrow a - \epsilon < \bar{\mathbf{x}} < a + \epsilon,$$

y como se busca que $0.9a < \bar{\mathbf{x}} < 1.1a$, $\epsilon = 0.1a$ en la desigualdad. De esta forma,

$$P\{0.9a < \bar{\mathbf{x}} < 1.1a\} \geq 1 - \frac{\sigma^2/n}{(0.1a)^2} = 1 - \frac{100\sigma^2}{na^2} = 1 - 100 \times 10^{-4} = 1 - 0.01 = 0.99.$$

A continuación se definen los distintos modos de convergencia de secuencias de variables aleatorias.

Definición: secuencia aleatoria Una *secuencia aleatoria* o un *proceso aleatorio en tiempo discreto* es una secuencia de variables aleatorias

$$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \dots$$

Para un ζ específico, $\mathbf{x}_n(\zeta)$ es una secuencia de números que puede converger o no. Esto sugiere que la noción de convergencia de secuencias aleatorias puede tener varias interpretaciones, que se explican a continuación.

6.4.1. Modos de convergencia de secuencias aleatorias

- *Convergencia segura (s).* Una secuencia de números x_n converge a cierto valor x si para todo $\epsilon > 0$, existe un número n_0 tal que

$$|x_n - x| < \epsilon, \quad \forall n > n_0. \quad (6.66)$$

Se dice que una secuencia \mathbf{x}_n de variables aleatorias converge *de forma segura, en todos lados o puntualmente* si la secuencia de números $\mathbf{x}_n(\zeta)$ converge como en 6.66 para todo ζ , es decir,

$$\{\zeta \in S : \lim_{n \rightarrow \infty} \mathbf{x}_n(\zeta) = \mathbf{x}(\zeta)\} = S.$$

El límite es un número que en general depende de ζ , es decir, el límite de la secuencia aleatoria \mathbf{x}_n es una variable aleatoria \mathbf{x} ,

$$\mathbf{x}_n \rightarrow \mathbf{x}, \quad \text{con } n \rightarrow \infty.$$

La convergencia segura se denota como

$$\mathbf{x}_n \xrightarrow{s} \mathbf{x}.$$

- *Convergencia casi segura (a.s.).* Si el conjunto de resultados ζ tales que

$$\lim_{n \rightarrow \infty} \mathbf{x}_n(\zeta) = \mathbf{x}(\zeta)$$

existe y tiene probabilidad 1, se dice que la secuencia \mathbf{x}_n converge *casi seguramente* o *con probabilidad 1*. Esto se puede expresar como

$$P\{\zeta \in S : \lim_{n \rightarrow \infty} \mathbf{x}_n(\zeta) = \mathbf{x}(\zeta)\} = 1,$$

donde $\{\zeta \in S : \lim_{n \rightarrow \infty} \mathbf{x}_n(\zeta) = \mathbf{x}(\zeta)\}$ es el evento que consiste en todos los resultados ζ tales que $\mathbf{x}_n(\zeta) \rightarrow \mathbf{x}(\zeta)$. Otra forma de expresar la definición de convergencia casi segura es considerando el evento

$$\liminf_{n \rightarrow \infty} \{\zeta \in S : |\mathbf{x}_n(\zeta) - \mathbf{x}(\zeta)| < \epsilon\},$$

que consiste en el conjunto de resultados ζ tales que $|\mathbf{x}_n(\zeta) - \mathbf{x}(\zeta)| < \epsilon$ se cumple eventualmente siempre en n (ver la definición de \liminf en la sección 1.2.1). Teniendo esto en cuenta, la secuencia converge casi seguramente si se cumple que⁵

$$\Pr \left(\liminf_{n \rightarrow \infty} \{\zeta \in S : |\mathbf{x}_n(\zeta) - \mathbf{x}(\zeta)| < \epsilon\} \right) = 1.$$

Es útil expresar la definición a partir de su complemento [2]: el conjunto de resultados ζ tales que $\mathbf{x}_n(\zeta)$ no converge a $\mathbf{x}(\zeta)$ tiene medida nula. $\mathbf{x}_n(\zeta)$ no converge a $\mathbf{x}(\zeta)$ si existe $\epsilon > 0$ tal que $|\mathbf{x}_n(\zeta) - \mathbf{x}(\zeta)| \geq \epsilon$ para infinitos valores de n . Considérese la familia de eventos

$$B_n^\epsilon = \{\zeta \in S : |\mathbf{x}_n(\zeta) - \mathbf{x}(\zeta)| \geq \epsilon\},$$

que consiste en los resultados ζ tales que $|\mathbf{x}_n(\zeta) - \mathbf{x}(\zeta)| \geq \epsilon$ para valores de n y ϵ específicos. La secuencia no converge casi seguramente si existe algún ϵ tal que el evento $\{B_n^\epsilon\}$ ocurre infinitamente a menudo en n tiene probabilidad no nula,

$$P\{B_n^\epsilon \text{ i.o.}\} = P\{\limsup_{n \rightarrow \infty} B_n^\epsilon\} > 0, \quad (6.67)$$

o equivalentemente, la secuencia converge absolutamente si para todo $\epsilon > 0$

$$P\{B_n^\epsilon \text{ i.o.}\} = P\{\limsup_{n \rightarrow \infty} B_n^\epsilon\} = 0. \quad (6.68)$$

La notación para indicar convergencia casi segura es

$$\mathbf{x}_n \xrightarrow{a.s.} \mathbf{x},$$

donde la sigla *a.s.* proviene de *almost surely* (casi seguramente).

Observar que la diferencia entre convergencia segura y convergencia casi segura es solamente un conjunto de probabilidad nula, y en la teoría de la probabilidad no hay beneficio en emplear convergencia segura frente a convergencia casi segura. Es por esto que el concepto de convergencia segura casi no se usa en la práctica.

- *Convergencia en el sentido cuadrático medio (m.s.).* La secuencia \mathbf{x}_n tiende a la variable aleatoria \mathbf{x} en el sentido cuadrático medio si

$$E\{|\mathbf{x}_n - \mathbf{x}|^2\} \rightarrow 0 \quad \text{con} \quad n \rightarrow \infty.$$

La convergencia en media cuadrática se denota

$$\mathbf{x}_n \xrightarrow{m.s.} \mathbf{x},$$

donde la sigla *m.s.* proviene de *mean square* (media cuadrática).

- *Convergencia en probabilidad (p).* La probabilidad $P\{\zeta \in S : |\mathbf{x}_n(\zeta) - \mathbf{x}(\zeta)| > \epsilon\}$ del evento $\{\zeta \in S : |\mathbf{x}_n(\zeta) - \mathbf{x}(\zeta)| > \epsilon\}$ es una secuencia de números en n que depende de ϵ . Si esta secuencia tiende a 0,

$$P\{\zeta \in S : |\mathbf{x}_n(\zeta) - \mathbf{x}(\zeta)| > \epsilon\} \rightarrow 0 \quad \text{con} \quad n \rightarrow \infty$$

para todo $\epsilon > 0$ se dice que la secuencia \mathbf{x}_n tiende a la variable aleatoria \mathbf{x} *en probabilidad*. La convergencia en probabilidad se denota

$$\mathbf{x}_n \xrightarrow{p} \mathbf{x}.$$

⁵Fuente: https://en.wikipedia.org/wiki/Convergence_of_random_variables#Almost_sure_convergence

- *Convergencia en distribución (d).* Sean $F_n(x)$ y $F(x)$ las funciones de distribución de probabilidad de \mathbf{x}_n y \mathbf{x} respectivamente. Si

$$F_n(x) \rightarrow F(x) \quad \text{con} \quad n \rightarrow \infty$$

para cada punto x en donde $F(x)$ es continua, se dice que la secuencia \mathbf{x}_n converge en distribución a la variable aleatoria \mathbf{x} . Se denota como

$$\mathbf{x}_n \xrightarrow{d} \mathbf{x}.$$

A diferencia de los otros modos de convergencia, para convergencia en distribución no es necesario que las variables aleatorias estén definidas en un único espacio de probabilidad.

En la figura 6.4, se muestra la relación de implicancia de cada modo de convergencia. Las demostraciones de las implicancias que se incluyen a continuación están basadas en los textos [2, 15].

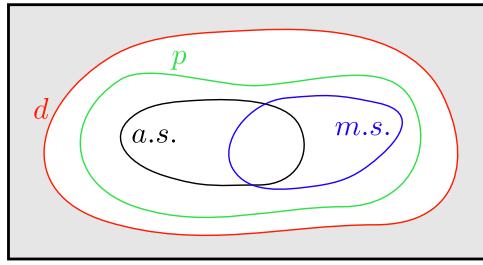


Figura 6.4: Relación de implicancia de los modos de convergencia. Cada punto del rectángulo representa una secuencia aleatoria. La letra de cada curva indica el modo de convergencia de todas las secuencias dentro de la curva. La parte sombreada contiene las secuencias que no convergen en ningún modo. Tanto convergencia casi segura como convergencia cuadrática media implican convergencia en probabilidad, y convergencia en probabilidad implica convergencia en distribución. Además convergencia casi segura no implica convergencia media cuadrática y viceversa.

Previo a la demostración de las implicancias de los modos de convergencia, se dará una explicación intuitiva de la diferencia entre convergencia en probabilidad y convergencia casi segura, que tienen particular importancia porque son los modos de convergencia que diferencian a la *ley débil de los grandes números* con la *ley fuerte de los grandes números*, como se verá más adelante. En el esquema de la figura 6.5 se muestra la diferencia $|\mathbf{x}_n - \mathbf{x}|$ en función de n . Cada curva representa una realización particular $|\mathbf{x}_n(\zeta) - \mathbf{x}(\zeta)|$. Convergencia en probabilidad significa que para cierto $n > n_0$ específico, solo un pequeño porcentaje de las curvas ordenadas que superan ϵ . Sin embargo, es posible que incluso ninguna de las curvas se mantenga bajo ϵ para todo $n > n_0$. Por otro lado, convergencia casi segura requiere que la mayoría de las curvas se mantengan bajo ϵ para todo $n > n_0$. Teniendo en cuenta estas consideraciones, se intuye que convergencia casi segura implica convergencia en probabilidad, y convergencia en probabilidad no implica convergencia casi segura.

Teorema: *Convergencia casi segura implica convergencia en probabilidad,*

$$\mathbf{x}_n \xrightarrow{a.s.} \mathbf{x} \quad \Rightarrow \quad \mathbf{x}_n \xrightarrow{p} \mathbf{x}.$$

Para demostrarlo, se considera la familia de eventos

$$B_n^\epsilon = \{\zeta \in S : |\mathbf{x}_n(\zeta) - \mathbf{x}(\zeta)| > \epsilon\},$$

y por definición, para que $\mathbf{x}_n \xrightarrow{a.s.} \mathbf{x}$, se tiene que cumplir que para todo $\epsilon > 0$,

$$P\{\limsup_{n \rightarrow \infty} B_n^\epsilon\} = 0.$$

El lema de Fatou reverso, dado por la ecuación 1.16, indica que

$$\limsup_{n \rightarrow \infty} P\{B_n^\epsilon\} \leq P\{\limsup_{n \rightarrow \infty} B_n^\epsilon\} = 0,$$

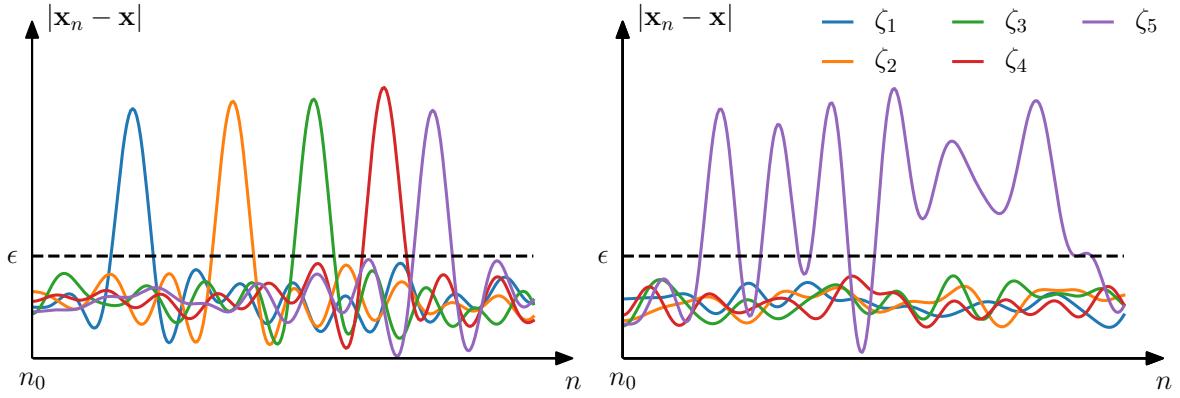


Figura 6.5: Ilustración esquemática de la diferencia de convergencia en probabilidad (izquierda) y convergencia casi segura (derecha). Convergencia en probabilidad implica que para cierto $n > n_0$ específico, solo un pequeño porcentaje de las curvas $|\mathbf{x}_n - \mathbf{x}|$ tenga ordenadas que superen ϵ . En el esquema con cinco realizaciones, solo a lo sumo una curva supera ϵ para cada n . Sin embargo, todas las realizaciones superan ϵ para algún $n > n_0$. Convergencia casi segura implica que la mayoría de las curvas no superan ϵ para todo $n > n_0$. En el esquema, solo una realización supera ϵ , mientras que el resto se mantienen bajo ϵ para todo $n > n_0$.

y como

$$\lim_{n \rightarrow \infty} P\{B_n^\epsilon\} \leq \lim_{n \rightarrow \infty} \left[\sup_{k \geq n} P\{B_k^\epsilon\} \right] = \limsup_{n \rightarrow \infty} P\{B_n^\epsilon\},$$

se concluye que

$$\lim_{n \rightarrow \infty} P\{B_n^\epsilon\} = \lim_{n \rightarrow \infty} P\{\zeta \in S : |\mathbf{x}_n(\zeta) - \mathbf{x}(\zeta)| > \epsilon\} = 0,$$

que es la definición de convergencia en probabilidad, y por lo tanto,

$$\mathbf{x}_n \xrightarrow{p} \mathbf{x}.$$

Teorema: *Convergencia media cuadrática implica convergencia en probabilidad,*

$$\mathbf{x}_n \xrightarrow{m.s.} \mathbf{x} \quad \Rightarrow \quad \mathbf{x}_n \xrightarrow{p} \mathbf{x}.$$

Esto es una consecuencia de la desigualdad de Markov. Planteando la desigualdad de Markov, dada por la ecuación 4.23, para la variable aleatoria $|\mathbf{x}_n - \mathbf{x}|^2$ y parámetro $\alpha = \epsilon^2$, se tiene que

$$P\{|\mathbf{x}_n - \mathbf{x}|^2 > \epsilon^2\} \leq \frac{E\{|\mathbf{x}_n - \mathbf{x}|^2\}}{\epsilon^2},$$

y como

$$P\{|\mathbf{x}_n - \mathbf{x}|^2 > \epsilon^2\} = P\{|\mathbf{x}_n - \mathbf{x}| > \epsilon\},$$

se obtiene que

$$P\{|\mathbf{x}_n - \mathbf{x}| > \epsilon\} \leq \frac{E\{|\mathbf{x}_n - \mathbf{x}|^2\}}{\epsilon^2}.$$

Tomando el límite con $n \rightarrow \infty$, se concluye que

$$\lim_{n \rightarrow \infty} P\{|\mathbf{x}_n - \mathbf{x}| > \epsilon\} \leq \lim_{n \rightarrow \infty} \frac{E\{|\mathbf{x}_n - \mathbf{x}|^2\}}{\epsilon^2} = 0,$$

donde la última igualdad se cumple por hipótesis.

Teorema: *Convergencia media cuadrática no implica convergencia casi segura.*

Para demostrarlo, alcanza con encontrar un ejemplo de secuencia aleatoria que converja en media cuadrática pero no converja casi seguramente. Considérese una familia de variables aleatorias Bernoulli independientes de distribución

$$p_{x_n}(x) = \begin{cases} \frac{1}{n}, & \mathbf{x}_n = 1 \\ 1 - \frac{1}{n}, & \mathbf{x}_n = 0 \end{cases},$$

y sea $\mathbf{x} = 0$. Se observa primero que $\mathbf{x}_n \xrightarrow{m.s.} \mathbf{x}$, ya que (ver sección 4.3.5)

$$E\{|\mathbf{x}_n - \mathbf{x}|^2\} = E\{\mathbf{x}_n^2\} = 1^2 \times \frac{1}{n} + 0^2 \times \left(1 - \frac{1}{n}\right) = \frac{1}{n} \rightarrow 0 \quad \text{con } n \rightarrow \infty.$$

Para demostrar que la secuencia no converge casi seguramente, se recurre al segundo lema de Borel-Cantelli (ver sección 1.3.5), que indica que para eventos mutuamente independientes,

$$\text{si } \sum_n P(A_n) = \infty \quad \Rightarrow \quad P\left(\limsup_n A_n\right) = P(\{A_n \text{ i.o.}\}) = 1.$$

En este caso, tomando $\epsilon = 1/2$, se tiene que

$$\sum_{n=1}^{\infty} P\{|\mathbf{x}_n - \mathbf{x}| \geq \epsilon\} = \sum_{n=1}^{\infty} P\left\{\mathbf{x}_n \geq \frac{1}{2}\right\} = \sum_{n=1}^{\infty} P\{\mathbf{x}_n = 1\} = \sum_{n=1}^{\infty} \frac{1}{n} = \infty,$$

donde en la última igualdad se empleó que la sumatoria es la serie armónica y por lo tanto, diverge⁶. Cumpliendo las hipótesis del segundo lema de Borel-Cantelli, se tiene que

$$\Pr\left(\limsup_{n \rightarrow \infty} \{|\mathbf{x}_n - \mathbf{x}| > \epsilon\}\right) = \Pr(\{|\mathbf{x}_n - \mathbf{x}| > \epsilon\} \text{ i.o.}) = 1$$

para cierto valor de ϵ , concluyendo que \mathbf{x}_n no converge casi seguramente, como indica la ecuación 6.67.

Notar además, que la secuencia converge en probabilidad, ya que

$$P\{|\mathbf{x}_n - 0| > \epsilon\} = P\{\mathbf{x}_n = 1\} = \frac{1}{n} \rightarrow 0 \quad \text{con } n \rightarrow \infty,$$

de lo que se deduce que *convergencia en probabilidad no implica convergencia casi segura*.

Teorema: *Convergencia casi segura no implica convergencia media cuadrática.*

Nuevamente, alcanza con encontrar un ejemplo de secuencia aleatoria que converja casi seguramente pero no converja en media cuadrática. Considérese la familia de variables aleatorias independientes con distribución

$$p_{x_n}(x) = \begin{cases} \frac{1}{n^2}, & \mathbf{x}_n = n^3 \\ 1 - \frac{1}{n^2}, & \mathbf{x}_n = 0 \end{cases},$$

y sea $\mathbf{x} = 0$. De esta forma, cuando $n = 10$ por ejemplo, $\mathbf{x}_n = 1000$ con probabilidad $1/100$, y $\mathbf{x}_n = 0$ con probabilidad $1 - 1/100 = 99/100$.

Se verifica primero que \mathbf{x}_n no converge en media cuadrática, ya que

$$E\{|\mathbf{x}_n - \mathbf{x}|^2\} = E\{\mathbf{x}_n^2\} = n^6 \times \frac{1}{n^2} + 0^2 \times \left(1 - \frac{1}{n^2}\right) = n^4 \rightarrow \infty \quad \text{con } n \rightarrow \infty.$$

Para demostrar que la secuencia converge casi seguramente, se recurre al lema de Borel-Cantelli (ver sección 1.3.3), que indica que

$$\text{si } \sum_n P(A_n) < \infty \quad \Rightarrow \quad P\left(\limsup_n A_n\right) = P(\{A_n \text{ i.o.}\}) = 0.$$

⁶Ver [https://en.wikipedia.org/wiki/Harmonic_series_\(mathematics\)](https://en.wikipedia.org/wiki/Harmonic_series_(mathematics))

En este caso, para cualquier $\epsilon > 0$, se tiene que⁷

$$\sum_{n=1}^{\infty} P\{|\mathbf{x}_n - \mathbf{x}| \geq \epsilon\} = \sum_{n=1}^{\infty} P\{\mathbf{x}_n \geq \epsilon\} = \sum_{n=1}^{\infty} P\{\mathbf{x}_n = n^3\} = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

Estando en las hipótesis del lema de Borel-Cantelli, se cumple que

$$\Pr\left(\limsup_{n \rightarrow \infty} \{|\mathbf{x}_n - \mathbf{x}| > \epsilon\}\right) = \Pr(\{|\mathbf{x}_n - \mathbf{x}| > \epsilon\} \text{ i.o.}) = 0,$$

concluyendo que \mathbf{x}_n converge casi seguramente, como indica la ecuación 6.68

Notar además que la secuencia \mathbf{x}_n converge en probabilidad, ya que

$$P\{|\mathbf{x}_n - \mathbf{x}| > \epsilon\} = P\{|\mathbf{x}_n| > \epsilon\} = P\{\mathbf{x}_n = n^3\} = \frac{1}{n^2} \rightarrow 0 \quad \text{con} \quad n \rightarrow \infty,$$

lo que indica que *convergencia en probabilidad no implica convergencia en media cuadrática*.

Teorema: *Convergencia en probabilidad implica convergencia en distribución,*

$$\mathbf{x}_n \xrightarrow{p} \mathbf{x} \quad \Rightarrow \quad \mathbf{x}_n \xrightarrow{d} \mathbf{x}.$$

Sean $a \in \mathbb{R}$ y $\epsilon > 0$. La distribución de probabilidad $F_{x_n}(a)$ de \mathbf{x}_n en a cumple que

$$\begin{aligned} F_{x_n}(a) &= P\{\mathbf{x}_n \leq a\} \\ &\stackrel{(a)}{=} P\{\mathbf{x}_n \leq a, \mathbf{x} \leq a + \epsilon\} + P\{\mathbf{x}_n \leq a, \mathbf{x} > a + \epsilon\} \\ &\stackrel{(b)}{=} P\{\mathbf{x}_n \leq a \mid \mathbf{x} \leq a + \epsilon\} P\{\mathbf{x} \leq a + \epsilon\} + P\{\mathbf{x}_n \leq a, a < \mathbf{x} - \epsilon\} \\ &\stackrel{(c)}{\leq} P\{\mathbf{x} \leq a + \epsilon\} + P\{\mathbf{x}_n < \mathbf{x} - \epsilon\} \\ &= F_x(a + \epsilon) + P\{\mathbf{x} - \mathbf{x}_n > \epsilon\} \\ &\stackrel{(d)}{\leq} F_x(a + \epsilon) + P\{|\mathbf{x}_n - \mathbf{x}| > \epsilon\}, \end{aligned}$$

donde en (a) se empleó que el evento $\{\mathbf{x}_n \leq a\}$ se puede expresar como

$$\begin{aligned} \{\mathbf{x}_n \leq a\} &= \{\mathbf{x}_n \leq a\} \cap (\{\mathbf{x} \leq a + \epsilon\} \cup \{\mathbf{x} > a + \epsilon\}) \\ &= (\{\mathbf{x}_n \leq a\} \cap \{\mathbf{x} \leq a + \epsilon\}) \cup (\{\mathbf{x}_n \leq a\} \cap \{\mathbf{x} > a + \epsilon\}), \end{aligned}$$

y como se trata de la unión de eventos disjuntos, la probabilidad es la suma de las probabilidades, como indica la ecuación 1.3, en (b) se empleó la definición de probabilidad condicional, dada por la ecuación 1.20, en (c) se empleó que $P\{\mathbf{x}_n \leq a \mid \mathbf{x} \leq a + \epsilon\} \leq 1$ y además que

$$\text{como} \quad \{\mathbf{x}_n \leq a\} \cap \{a < \mathbf{x} - \epsilon\} \subseteq \{\mathbf{x}_n < \mathbf{x} - \epsilon\} \quad \Rightarrow \quad P(\mathbf{x}_n \leq a, a < \mathbf{x} - \epsilon) \leq P(\mathbf{x}_n < \mathbf{x} - \epsilon)$$

por la propiedad de monotonía de la probabilidad (ecuación 1.10), y en (d) se empleó el mismo argumento notando que

$$\{\mathbf{x} - \mathbf{x}_n > \epsilon\} \subseteq \{|\mathbf{x}_n - \mathbf{x}| > \epsilon\} = \{\mathbf{x}_n - \mathbf{x} > \epsilon\} \cup \{\mathbf{x} - \mathbf{x}_n > \epsilon\}.$$

Por otro lado, empleando un razonamiento análogo, se tiene que

$$\begin{aligned} F_x(a - \epsilon) &= P\{\mathbf{x} \leq a - \epsilon\} \\ &= P\{\mathbf{x} \leq a - \epsilon, \mathbf{x}_n \leq a\} + P\{\mathbf{x} \leq a - \epsilon, \mathbf{x}_n > a\} \\ &= P\{\mathbf{x} \leq a - \epsilon \mid \mathbf{x}_n \leq a\} P\{\mathbf{x}_n \leq a\} + P\{\mathbf{x} + \epsilon \leq a, a < \mathbf{x}_n\} \\ &\leq P\{\mathbf{x}_n \leq a\} + P\{\mathbf{x} + \epsilon < \mathbf{x}_n\} \\ &= F_{x_n}(a) + P\{\mathbf{x}_n - \mathbf{x} > \epsilon\} \\ &\leq F_{x_n}(a) + P\{|\mathbf{x}_n - \mathbf{x}| > \epsilon\}. \end{aligned}$$

⁷Ver https://en.wikipedia.org/wiki/Basel_problem

Combinando ambos resultados, se obtiene que

$$F_x(a - \epsilon) - P\{|\mathbf{x}_n - \mathbf{x}| > \epsilon\} \leq F_{x_n}(a) \leq F_x(a + \epsilon) + P\{|\mathbf{x}_n - \mathbf{x}| > \epsilon\}.$$

Tomando el límite cuando $n \rightarrow \infty$, se tiene que

$$F_x(a - \epsilon) \leq \lim_{n \rightarrow \infty} F_{x_n}(a) \leq F_x(a + \epsilon),$$

ya que por hipótesis \mathbf{x}_n converge en probabilidad y por lo tanto $P\{|\mathbf{x}_n - \mathbf{x}| > \epsilon\} \rightarrow 0$ con $n \rightarrow \infty$. También por hipótesis, la distribución $F_x(x)$ de \mathbf{x} es continua en a y por lo tanto $F_x(a - \epsilon)$ y $F_x(a + \epsilon)$ tienden a $F_x(a)$ con $\epsilon \rightarrow 0^+$. Por lo tanto, al tomar este límite se concluye que

$$\lim_{n \rightarrow \infty} F_{x_n}(a) = F_x(a),$$

y por lo tanto \mathbf{x}_n converge en distribución.

Teorema: *Convergencia en distribución a una constante implica convergencia en probabilidad a esa constante,*

$$\mathbf{x}_n \xrightarrow{d} c \quad \Rightarrow \quad \mathbf{x}_n \xrightarrow{p} c.$$

La variable aleatoria constante c tiene distribución de probabilidad degenerada⁸

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

y como por hipótesis $\mathbf{x}_n \xrightarrow{d} c$, se debe cumplir que para todo $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} F_{x_n}(c - \epsilon) = 0, \quad \lim_{n \rightarrow \infty} F_{x_n}(c + \epsilon) = 1.$$

Por otro lado, se observa que

$$\begin{aligned} \lim_{n \rightarrow \infty} P\{|\mathbf{x}_n - c| \geq \epsilon\} &\stackrel{(a)}{=} \lim_{n \rightarrow \infty} [P\{\mathbf{x}_n \geq c + \epsilon\} + P\{\mathbf{x}_n \leq c - \epsilon\}] \\ &= \lim_{n \rightarrow \infty} P\{\mathbf{x}_n \geq c + \epsilon\} + \lim_{n \rightarrow \infty} P\{\mathbf{x}_n \leq c - \epsilon\} \\ &= \lim_{n \rightarrow \infty} [1 - P\{\mathbf{x}_n < c + \epsilon\}] + \lim_{n \rightarrow \infty} P\{\mathbf{x}_n \leq c - \epsilon\} \\ &= 1 - \lim_{n \rightarrow \infty} F_{x_n}(c + \epsilon) + \lim_{n \rightarrow \infty} F_{x_n}(c - \epsilon) \\ &= 0, \end{aligned}$$

donde en (a) se tuvo en cuenta que

$$|\mathbf{x}_n - c| \geq \epsilon \Leftrightarrow \{\mathbf{x}_n - c \geq \epsilon\} \cup \{-\mathbf{x}_n + c \geq \epsilon\} \Leftrightarrow \{\mathbf{x}_n \geq c + \epsilon\} \cup \{\mathbf{x}_n \leq c - \epsilon\}.$$

Se concluye que $\mathbf{x}_n \xrightarrow{p} c$.

Ejemplo Sea $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots$ una secuencia de variables aleatorias independientes e idénticamente distribuidas con distribución uniforme en $(0, 1)$. Se define la secuencia \mathbf{y}_n como

$$\mathbf{y}_n = \min(\mathbf{x}_1, \dots, \mathbf{x}_n),$$

y se quiere ver que⁹

$$\mathbf{y}_n \xrightarrow{d} 0, \quad \mathbf{y}_n \xrightarrow{p} 0, \quad \mathbf{y}_n \xrightarrow{m.s.} 0, \quad \mathbf{y}_n \xrightarrow{a.s.} 0.$$

Se demostrará la convergencia en cada modo de forma independiente sin emplear las relaciones de implicancias.

⁸Ver: https://en.wikipedia.org/wiki/Degenerate_distribution

⁹Fuente: https://www.probabilitycourse.com/chapter7/7_2_8_solved_probs.php.

- *Convergencia en distribución:* para ver que la secuencia converge en distribución, se calculará la distribución $F_{y_n}(y)$ de la variable aleatoria \mathbf{y}_n . Las variables aleatorias \mathbf{x}_n tienen distribución uniforme en $(0, 1)$,

$$F_{x_n}(x) = \begin{cases} 0, & x \leq 0 \\ x, & 0 < x < 1 \\ 1, & x \geq 1 \end{cases}$$

por lo tanto, la distribución de \mathbf{y}_n es

$$\begin{aligned} F_{y_n}(y) &= P\{\mathbf{y}_n \leq y\} \\ &= P\{\min(\mathbf{x}_1, \dots, \mathbf{x}_n) \leq y\} \\ &= 1 - P\{\min(\mathbf{x}_1, \dots, \mathbf{x}_n) > y\} \\ &= 1 - P\{\mathbf{x}_1 > y, \dots, \mathbf{x}_n > y\} \\ &= 1 - P\{\mathbf{x}_1 > y\} \times \dots \times P\{\mathbf{x}_n > y\}, \end{aligned}$$

donde en la última igualdad se empleo que las variables aleatorias \mathbf{x}_n son independientes. Continuando con el razonamiento, teniendo en cuenta que $P\{\mathbf{x}_1 > y\} = 1 - P\{\mathbf{x}_1 \leq y\}$, se tiene que

$$\begin{aligned} F_{y_n}(y) &= 1 - (1 - P\{\mathbf{x}_1 \leq y\}) \times \dots \times (1 - P\{\mathbf{x}_n \leq y\}) \\ &= 1 - [1 - F_{x_1}(y)] \times \dots \times [1 - F_{x_n}(y)] \\ &= 1 - (1 - y) \times \dots \times (1 - y) \\ &= 1 - (1 - y)^n \end{aligned}$$

para $0 < y < 1$. Teniendo en cuenta además que $\mathbf{y}_n \in (0, 1)$, se concluye que

$$F_{y_n}(y) = \begin{cases} 0, & y \leq 0 \\ 1 - (1 - y)^n, & 0 < y < 1 \\ 1, & y \geq 1 \end{cases} \quad (6.69)$$

Notar que la densidad de probabilidad $f_{y_n}(y)$ de \mathbf{y}_n es

$$f_{y_n}(y) = \frac{dF_{y_n}(y)}{dy} = \begin{cases} n(1 - y)^{n-1}, & 0 < y < 1 \\ 0, & \text{en otro caso} \end{cases} \quad (6.70)$$

Este resultado ya se había obtenido en la ecuación 6.8 para la densidad de probabilidad del estadístico de orden 1 de n variables aleatorias independientes de distribución uniforme. Volviendo al estudio de la convergencia en distribución, para $0 < y < 1$ se cumple que $0 < 1 - y < 1$ y por lo tanto

$$\lim_{n \rightarrow \infty} 1 - (1 - y)^n = 1.$$

De esta forma, se tiene que

$$\lim_{n \rightarrow \infty} F_{y_n}(y) = \begin{cases} 0, & y \leq 0 \\ 1, & 0 < y < 1 \\ 1, & y \geq 1 \end{cases} = \begin{cases} 0, & y \leq 0 \\ 1, & y > 0 \end{cases}.$$

Observando además que la variable aleatoria constante 0 tiene distribución de probabilidad degenerada

$$F_0(y) = \begin{cases} 0, & y < 0 \\ 1, & y \geq 0 \end{cases},$$

se concluye que

$$\lim_{n \rightarrow \infty} F_{y_n}(y) = F_0(y), \quad \forall y \in \mathbb{R}, \quad y \neq 0,$$

pero el punto $y = 0$ es el único punto de discontinuidad de $F_0(y)$. En otras palabras, $F_{y_n}(y) \rightarrow F_0(y)$ con $n \rightarrow \infty$ en todo punto donde $F_0(y)$ es continua, y por lo tanto,

$$\mathbf{y}_n \xrightarrow{d} 0.$$

- *Convergencia en probabilidad:* para demostrar que $\mathbf{y}_n \xrightarrow{p} 0$ hay que demostrar que

$$\lim_{n \rightarrow \infty} P\{|\mathbf{y}_n - 0| > \epsilon\} = \lim_{n \rightarrow \infty} P\{|\mathbf{y}_n| > \epsilon\} = 0.$$

Partiendo de que $\mathbf{y}_n \in (0, 1)$, se observa que

$$\begin{aligned} P\{|\mathbf{y}_n| > \epsilon\} &= P\{\mathbf{y}_n > \epsilon\} \\ &= 1 - P\{\mathbf{y}_n \leq \epsilon\} \\ &= 1 - F_{y_n}(\epsilon). \end{aligned}$$

Para $\epsilon \in (0, 1)$, sustituyendo la distribución de probabilidad $F_{y_n}(\epsilon)$, dada por la ecuación 6.69, se obtiene que

$$P\{|\mathbf{y}_n| > \epsilon\} = (1 - \epsilon)^n, \quad (6.71)$$

y por lo tanto,

$$\lim_{n \rightarrow \infty} P\{|\mathbf{y}_n| > \epsilon\} = \lim_{n \rightarrow \infty} (1 - \epsilon)^n = 0, \quad \forall \epsilon \in (0, 1],$$

que es lo que se quería demostrar.

- *Convergencia en media cuadrática:* para demostrar que $\mathbf{y}_n \xrightarrow{m.s.} 0$ hay que demostrar que

$$\lim_{n \rightarrow \infty} E\{|\mathbf{y}_n - 0|^2\} = \lim_{n \rightarrow \infty} E\{\mathbf{y}_n^2\} = 0.$$

Partiendo de la definición de esperanza (ver sección 4.3.1), se tiene que

$$\begin{aligned} E\{\mathbf{y}_n^2\} &= \int_{-\infty}^{\infty} y^2 f_{y_n}(y) dy \\ &\stackrel{(a)}{=} n \int_0^1 y^2 (1-y)^{n-1} dy \\ &\stackrel{(b)}{\leq} n \int_0^1 y (1-y)^{n-1} dy \\ &\stackrel{(c)}{=} -y(1-y)^n \Big|_0^1 + \int_0^1 (1-y)^n dy \\ &= 0 - \frac{(1-y)^{n+1}}{n+1} \Big|_0^1 \\ &= \frac{1}{n+1}, \end{aligned}$$

donde en (a) se sustituyó la densidad de probabilidad de \mathbf{y}_n , dada por la ecuación 6.70, en (b) se consideró que $y^2 < y$ para todo $y \in (0, 1)$ y en (c) se aplicó integración por partes, $\int u dv = uv - \int v du$, con $u = y$ y $dv = (1-y)^{n-1}$, y entonces $du = 1$ y $v = -(1-y)^n/n$. Se concluye que

$$\lim_{n \rightarrow \infty} E\{|\mathbf{y}_n - 0|^2\} \leq \lim_{n \rightarrow \infty} \frac{1}{n+1} = 0.$$

- *Convergencia casi segura:* para demostrar que $\mathbf{y}_n \xrightarrow{a.s.} 0$, hay que demostrar que

$$P(\{|\mathbf{y}_n - 0| \geq \epsilon\} \text{ i.o.}) = 0,$$

y como indica el lema de Borel-Cantelli (ver sección 1.3.3), es equivalente a demostrar que

$$\sum_{n=1}^{\infty} P\{|\mathbf{y}_n - 0| \geq \epsilon\} < \infty.$$

Usando el valor de $P\{|\mathbf{y}_n - 0| \geq \epsilon\}$ que ya fue calculado en la ecuación 6.71, se tiene que

$$\begin{aligned} \sum_{n=1}^{\infty} P\{|\mathbf{y}_n - 0| \geq \epsilon\} &= \sum_{n=1}^{\infty} (1 - \epsilon)^n \\ &= \left[\sum_{n=0}^{\infty} (1 - \epsilon)^n \right] - 1 \\ &\stackrel{(a)}{=} \frac{1}{1 - (1 - \epsilon)} - 1 \\ &= \frac{1 - \epsilon}{\epsilon} < \infty \end{aligned}$$

donde en (a) se empleó el resultado de la convergencia de una serie geométrica, dado en la ecuación 4.54.

6.4.2. La ley de los grandes números

La *ley de los grandes números* (LLN, *law of large numbers*) describe el resultado al realizar el mismo experimento reiteradas veces. De acuerdo a la ley, el promedio de los resultados en una cantidad grande de ensayos se aproxima al valor esperado teórico del experimento, y la aproximación es mejor a medida que se incrementa la cantidad de ensayos. Hay dos versiones de la ley, la *débil* y la *fuerte*. La versión débil indica que la convergencia del promedio al valor esperado es en probabilidad, y la versión fuerte indica que la convergencia es casi segura. La primera demostración de la ley débil de los grandes números fue realizada por Jacob Bernoulli en 1713 para variables aleatorias binarias. En 1907, Émile Borel formuló y demostró la versión fuerte de la ley.

Se comenzará con la ley en el caso particular de ensayos de Bernoulli, que es la que se incluye en [1], y luego se planteará la forma general de la ley.

La ley débil de los grandes números (Bernoulli)

En la sección 2.3 se demostró el teorema de Bernoulli, que indica que si la probabilidad de ocurrencia de cierto evento A en un experimento es p y el número de ocurrencias de A en n ensayos es k , se cumple que (ecuación 2.7),

$$P \left\{ \left| \frac{k}{n} - p \right| < \epsilon \right\} > 1 - \frac{pq}{n\epsilon^2},$$

y por lo tanto,

$$P \left\{ \left| \frac{k}{n} - p \right| < \epsilon \right\} \rightarrow 1, \quad \text{con} \quad n \rightarrow \infty. \quad (6.72)$$

Se demostrará nuevamente este resultado como el límite de una secuencia de variables aleatorias. Para esto, se definen las variables aleatorias

$$\mathbf{x}_i = \begin{cases} 1, & \text{si } A \text{ ocurre en el } i\text{-ésimo ensayo} \\ 0, & \text{en otro caso,} \end{cases}$$

y se demostrará que la media muestral,

$$\bar{\mathbf{x}}_n = \frac{\mathbf{x}_1 + \cdots + \mathbf{x}_n}{n}$$

converge a p en probabilidad cuando $n \rightarrow \infty$. Para hacerlo, se empleará la desigualdad de Chebyshev, dada por la ecuación 4.22, que indica que

$$P\{|\bar{\mathbf{x}}_n - \eta_{x_n}| < \epsilon\} \geq 1 - \frac{\sigma_{x_n}^2}{\epsilon^2}.$$

En este caso, $\eta_{x_i} = p$ y $\sigma_{x_i}^2 = pq$, como se dedujo en la sección 4.3.5, y por lo tanto,

$$\eta_{x_n} \stackrel{(a)}{=} \frac{n\eta_{x_i}}{n} = \eta_{x_i} = p \quad \text{y} \quad \sigma_{x_n}^2 \stackrel{(b)}{=} \frac{n\sigma_{x_i}^2}{n^2} = \frac{\sigma_{x_i}^2}{n} = \frac{pq}{n},$$

donde en (a) se empleó la propiedad de linealidad de la esperanza y en (b) que la varianza de la suma de variables aleatorias independientes es la suma de las varianzas (ver ecuación 5.58) y además que la multiplicación de una variable aleatoria por una constante incrementa a la varianza como el cuadrado de la constante (ecuación 4.11). Sustituyendo estos resultados en la desigualdad de Chebyshev, se obtiene que

$$P\{|\bar{x}_n - p| < \epsilon\} \geq 1 - \frac{pq}{n\epsilon^2}.$$

Observando además que $pq = p(1-p) \leq 1/4$, ya que se trata de un polinomio de segundo grado en p de concavidad negativa cuyo valor máximo se da en $p = 1/2$ y es $1/4$, se cumple que

$$P\{|\bar{x}_n - p| < \epsilon\} \geq 1 - \frac{pq}{n\epsilon^2} \geq 1 - \frac{1}{4n\epsilon^2} \rightarrow 1 \quad \text{con } n \rightarrow \infty,$$

es decir,

$$\bar{x}_n \xrightarrow{p} p.$$

Además, si el evento A ocurre k veces se cumple que $\bar{x}_n = k/n$, obteniendo la ecuación 6.72.

La ley fuerte de los grandes números (Borel)

La versión fuerte de ley indica que \bar{x}_n converge a p no solo en probabilidad, sino también casi seguramente o con probabilidad 1. La demostración de este resultado es difícil y no se incluirá. Se realiza a continuación una interpretación de la diferencia de implicancias entre la ley débil y la ley fuerte en términos de la frecuencia relativa.

Supóngase que cierto evento tiene una probabilidad de ocurrencia p , y se pretende estimar p con un error menor a $\epsilon = 0.1$. Para eso, se repite el experimento $n = 1000$ veces y se usa como estimador la media muestral \bar{x}_n . De esta forma, se cumple que

$$P\{|\bar{x}_n - p| < \epsilon\} \geq 1 - \frac{1}{4n\epsilon^2} = 1 - \frac{1}{4 \times 1000 \times 0.01} = 1 - \frac{1}{40} = \frac{39}{40}.$$

Este resultado indica que en 39 cada 40 realizaciones ζ_i de $\bar{x}_{1000}(\zeta_i)$, el error $|\bar{x}_{1000}(\zeta_i) - p|$ será menor que 0.1.

Supóngase ahora que el experimento se repite $n = 2000$ veces, y se determina la media muestral \bar{x}_n no solo para un valor de n , sino para todo n entre 1000 y 2000. La ley débil de los grandes números permite concluir lo siguiente: si se realiza el experimento ζ_i con $n = 2000$ una cantidad i grande de veces, se cumple que para un valor específico de $n \geq 1000$, el error $|\bar{x}_n(\zeta_i) - p|$ será mayor que 0.1 una de cada 40 realizaciones ζ_i . En otras palabras, el 97.5% de las realizaciones serán “buenas” en el sentido en que el error será menor que ϵ para ese n específico. Lo que la ley débil no permite concluir es que en esas buenas realizaciones el error se mantenga menor que ϵ para todo n entre 1000 y 2000. Esto último sin embargo es cierto, pero solo se puede concluir a partir de la ley fuerte de los grandes números.

La ley débil de los grandes números (Khinchin, caso general)

Sea $\mathbf{x}_1, \mathbf{x}_2, \dots$ una secuencia de variables aleatorias i.i.d. con valor esperado finito η , es decir,

$$E\{\mathbf{x}_1\} = E\{\mathbf{x}_2\} = \dots = \eta < \infty,$$

y sea la media muestral,

$$\bar{x}_n = \frac{\mathbf{x}_1 + \dots + \mathbf{x}_n}{n}.$$

La ley débil de los grandes números indica que

$$\bar{x}_n \xrightarrow{p} \eta.$$

La demostración puede realizarse a partir de la inecuación de Chebyshev de forma idéntica al caso de variables aleatorias binarias visto previamente¹⁰. Se incluye a continuación otra demostración empleando las funciones características, que tiene interés por la similitud a la demostración del teorema central del límite, el que se verá más adelante. La demostración desarrollada a continuación está basada en [16].

Para la prueba, se tendrán en cuenta las siguientes tres consideraciones:

¹⁰ver https://en.wikipedia.org/wiki/Law_of_large_numbers, por ejemplo.

1. *Desarrollo de Taylor de primer orden:* sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función que tiene derivada en cero, y sea la expansión de Taylor de primer orden en torno a cero

$$g(x) = g(0) + g'(0)x + h(x),$$

donde $h(x)$ representa el error en la aproximación de primer orden. Por definición de la derivada, se cumple que

$$g'(0) = \lim_{x \rightarrow 0} \frac{g(x) - g(0)}{x} = \lim_{x \rightarrow 0} \frac{g'(0)x + h(x)}{x} = g'(0) + \lim_{x \rightarrow 0} \frac{h(x)}{x},$$

resultando en que

$$\lim_{x \rightarrow 0} \frac{h(x)}{x} = 0.$$

Una función h con esta propiedad se denota $o(x)$ ¹¹. Estas observaciones también son válidas para funciones complejas considerando separadamente la parte real y la parte imaginaria.

2. *Función exponencial:* recordar el hecho bien conocido de que

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a}{n}\right)^n = e^a, \quad a \in \mathbb{R}.$$

La identidad se cumple además en el caso en que a es un número complejo. Adicionalmente, se puede probar que si $\{a_n\}$ es una secuencia de números complejos que converge a a , se cumple que

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a_n}{n}\right)^n = e^a.$$

3. *Teorema de continuidad de Levy.* El teorema de continuidad de Levy o de convergencia de Levy vincula la convergencia en distribución de una secuencia de variables aleatorias con la convergencia puntual de sus funciones características. A continuación se incluye su formulación, pero no se demostrará.

Sea \mathbf{x}_n una secuencia de variables aleatorias con distribuciones de probabilidad $F_n(x)$ y funciones características $\Phi_n(\omega)$. Sea \mathbf{x} una variable aleatoria con distribución de probabilidad $F(x)$ y función característica $\Phi(\omega)$. Si se cumple que

$$\Phi_n(\omega) \rightarrow \Phi(\omega) \quad \forall \omega \in \mathbb{R},$$

es decir, la secuencia $\Phi_n(\omega)$ de funciones características converge puntualmente a $\Phi(\omega)$, se cumple que

$$F_n(x) \rightarrow F(x)$$

en todos los puntos de continuidad de $F(x)$. En otras palabras, \mathbf{x}_n converge en distribución a \mathbf{x} ,

$$\mathbf{x}_n \xrightarrow{d} \mathbf{x}.$$

Se parte considerando un desarrollo de Taylor de primer orden de la función característica. Teniendo en cuenta que la función característica de una variable aleatoria \mathbf{x} cumple que

$$\begin{aligned} \Phi_x(\omega) &= E(e^{j\omega\mathbf{x}}) & \Rightarrow \quad \Phi_x(0) &= E(1) &= 1 \\ \Phi'_x(\omega) &= jE(\mathbf{x}e^{j\omega\mathbf{x}}) & \Rightarrow \quad \Phi'_x(0) &= jE(\mathbf{x}) &= j\eta, \end{aligned}$$

el desarrollo de Taylor de primer orden en torno a cero es

$$\Phi_x(\omega) = 1 + j\eta\omega + o(\omega), \quad \omega \rightarrow 0.$$

Como las variables aleatorias $\mathbf{x}_1, \mathbf{x}_2, \dots$ son i.i.d., todas tienen la misma función característica, que se denotará $\Phi_x(\omega)$. Considerando las siguientes propiedades de la función característica,

$$\Phi_{\frac{x}{n}}(\omega) = \Phi_x\left(\frac{\omega}{n}\right) \quad \text{y} \quad \Phi_{x+y}(\omega) = \Phi_x(\omega)\Phi_y(\omega)$$

¹¹ver https://en.wikipedia.org/wiki/Big_O_notation#Little-o_notation, por ejemplo

si \mathbf{x} y \mathbf{y} son independientes (ver las ecuaciones 4.27 y 5.63), la función característica de $\bar{\mathbf{x}}_n$ se puede expresar en términos de $\Phi_x(\omega)$ como

$$\Phi_{\bar{x}_n}(\omega) = \left[\Phi_x\left(\frac{\omega}{n}\right) \right]^n = \left[1 + j\frac{\eta\omega}{n} + o\left(\frac{\omega}{n}\right) \right]^n,$$

donde en la segunda igualdad se empleó el desarrollo de Taylor de primer orden considerado previamente, con la función o cumpliendo que $\lim_{x \rightarrow 0} o(x)/x = 0$. Tomando el límite $n \rightarrow \infty$ se obtiene que

$$\lim_{n \rightarrow \infty} \Phi_{\bar{x}_n}(\omega) = e^{j\eta\omega}, \quad \forall \omega.$$

$e^{j\eta\omega}$ se reconoce como la función característica de una variable aleatoria degenerada que toma el valor η con probabilidad uno. Esto es porque la densidad de probabilidad de una variable aleatoria constante que toma el valor η es $\delta(x - \eta)$, por lo que la función característica es

$$\Phi(\omega) = \int_{-\infty}^{\infty} \delta(x - \eta) e^{j\omega x} dx = e^{j\eta\omega}.$$

Aplicando el teorema de continuidad de Levy, se obtiene que

$$\mathbf{x}_n \xrightarrow{d} \eta.$$

Pero como convergencia en distribución a una constante implica convergencia en probabilidad, como se mostró en la sección 6.4.1, se concluye que

$$\mathbf{x}_n \xrightarrow{P} \eta.$$

A continuación se incluyen los llamados *teorema de Markov* y *condición de Chebyshev* en [1]. Sin embargo, no se encontró otra referencia a los teoremas con dichos nombres en la bibliografía.

Teorema de Markov

A partir de una secuencia de \mathbf{x}_i de variables aleatorias, se construye su media muestral

$$\bar{\mathbf{x}}_n = \frac{\mathbf{x}_1 + \cdots + \mathbf{x}_n}{n}.$$

De esta forma, $\bar{\mathbf{x}}_n$ es una variable aleatoria cuyos valores $\bar{\mathbf{x}}_n(\zeta)$ dependen del resultado ζ del experimento. Se demostrará que si las variables aleatorias \mathbf{x}_i son tales que la media $\bar{\eta}_n$ de $\bar{\mathbf{x}}_n$ tiende a un límite η y su varianza tiende a 0 con $n \rightarrow \infty$, es decir,

$$E\{\bar{\mathbf{x}}_n\} = \bar{\eta}_n \xrightarrow{n \rightarrow \infty} \eta \quad \text{y} \quad \bar{\sigma}_n^2 = E\{(\bar{\mathbf{x}}_n - \bar{\eta}_n)^2\} \xrightarrow{n \rightarrow \infty} 0,$$

se cumple que la variable aleatoria $\bar{\mathbf{x}}_n$ tiende a η en el sentido cuadrático medio,

$$E\{(\bar{\mathbf{x}}_n - \eta)^2\} \xrightarrow{n \rightarrow \infty} 0 \quad \text{o} \quad \bar{\mathbf{x}}_n \xrightarrow{m.s.} \eta$$

La demostración se basa en la desigualdad

$$|\bar{\mathbf{x}}_n - \eta|^2 \leq 2|\bar{\mathbf{x}}_n - \bar{\eta}_n|^2 + 2|\bar{\eta}_n - \eta|^2.$$

Esta desigualdad proviene de que por un lado

$$|\bar{\mathbf{x}}_n - \eta| = |\bar{\mathbf{x}}_n - \bar{\eta}_n + \bar{\eta}_n - \eta| \leq |\bar{\mathbf{x}}_n - \bar{\eta}_n| + |\bar{\eta}_n - \eta|,$$

donde la desigualdad proviene de aplicar la desigualdad triangular. Por lo tanto,

$$|\bar{\mathbf{x}}_n - \eta|^2 \leq (|\bar{\mathbf{x}}_n - \bar{\eta}_n| + |\bar{\eta}_n - \eta|)^2 = |\bar{\mathbf{x}}_n - \bar{\eta}_n|^2 + |\bar{\eta}_n - \eta|^2 + 2|\bar{\mathbf{x}}_n - \bar{\eta}_n||\bar{\eta}_n - \eta|.$$

Por otro lado, se cumple que

$$(|\bar{\mathbf{x}}_n - \bar{\eta}_n| - |\bar{\eta}_n - \eta|)^2 = |\bar{\mathbf{x}}_n - \bar{\eta}_n|^2 + |\bar{\eta}_n - \eta|^2 - 2|\bar{\mathbf{x}}_n - \bar{\eta}_n||\bar{\eta}_n - \eta| \geq 0,$$

es decir,

$$2|\bar{\mathbf{x}}_n - \bar{\eta}_n||\bar{\eta}_n - \eta| \leq |\bar{\mathbf{x}}_n - \bar{\eta}_n|^2 + |\bar{\eta}_n - \eta|^2.$$

Combinando ambos resultados, se obtiene que

$$|\bar{\mathbf{x}}_n - \eta|^2 \leq |\bar{\mathbf{x}}_n - \bar{\eta}_n|^2 + |\bar{\eta}_n - \eta|^2 + 2|\bar{\mathbf{x}}_n - \bar{\eta}_n||\bar{\eta}_n - \eta| \leq 2|\bar{\mathbf{x}}_n - \bar{\eta}_n|^2 + 2|\bar{\eta}_n - \eta|^2,$$

que es la desigualdad que se empleará. Tomando los valores esperados en ambos lados resulta en

$$E\{(\bar{\mathbf{x}}_n - \eta)^2\} \leq 2E\{(\bar{\mathbf{x}}_n - \bar{\eta}_n)^2\} + 2(\bar{\eta}_n - \eta)^2 \xrightarrow{n \rightarrow \infty} 0,$$

donde la convergencia a 0 es por las hipótesis.

Corolario: condición de Tchebycheff Si las variables aleatorias \mathbf{x}_i son no correlacionadas y además

$$\frac{\sigma_1^2 + \cdots + \sigma_n^2}{n^2} \xrightarrow{n \rightarrow \infty} 0, \quad (6.73)$$

se cumple que

$$\mathbf{x}_n \xrightarrow{m.s.} \eta = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E\{\mathbf{x}_i\}.$$

Esto es una consecuencia del teorema de Markov, ya que como las variables aleatorias \mathbf{x}_i son no correlacionadas, el lado izquierdo de la ecuación 6.73 es $\bar{\sigma}_n^2$. La ecuación 6.73 es la condición de Tchebycheff, y es satisfecha si $\sigma_i < K < \infty$ para todo i . Ese es el caso si las variables aleatorias \mathbf{x}_i son i.i.d. y de varianza finita.

6.4.3. Teorema central del límite

La suma de n variables aleatorias \mathbf{x}_i independientes,

$$\mathbf{x} = \mathbf{x}_1 + \cdots + \mathbf{x}_n,$$

es una variable aleatoria con media y varianza

$$\eta = \eta_1 + \cdots + \eta_n \quad \text{y} \quad \sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2 \quad (6.74)$$

respectivamente. El teorema central del límite (CLT, *central limit theorem*) establece que en ciertas condiciones generales, la distribución $F(x)$ de \mathbf{x} se aproxima a una distribución normal con la misma media y varianza a medida que n crece, es decir,

$$F(x) \approx G\left(\frac{x - \eta}{\sigma}\right),$$

donde $G(x)$ es la distribución normal de media nula y varianza unidad. Adicionalmente, si las variables aleatorias \mathbf{x}_i son continuas, la densidad $f(x)$ de \mathbf{x} se aproxima a una densidad normal,

$$f(x) \approx \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\eta)^2/2\sigma^2}. \quad (6.75)$$

El teorema puede ser formulado como un límite. Si $\mathbf{z} = (\mathbf{x} - \eta)/\sigma$, se cumple que

$$F_z(z) \xrightarrow{n \rightarrow \infty} G(z) \quad \text{o} \quad f_z(z) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

para el caso general y el caso continuo respectivamente.

El CLT es consecuencia de una propiedad del producto convolución: la convolución de una cantidad grande de funciones positivas (ver la ecuación 6.23) tiende a una función normal. El siguiente ejemplo es una ilustración de esto.

Ejemplo 7-15 Sean las variables \mathbf{x}_i i.i.d. con distribución uniforme en el intervalo $(0, T)$. Se comparará la densidad $f_n(x)$ de su suma con la densidad normal para el caso $n = 2$ y $n = 3$. Es fácil ver que la media es $\eta_i = T/2$ y la varianza es

$$\begin{aligned}\sigma_i^2 &= \int_{-\infty}^{\infty} (x - \eta_i)^2 f(x) dx \\ &\stackrel{(a)}{=} \frac{1}{T} \int_0^T \left(x - \frac{T}{2}\right)^2 dx \\ &\stackrel{(b)}{=} \frac{1}{T} \int_{-T/2}^{T/2} u^2 du \\ &= \frac{1}{T} \frac{u^3}{3} \Big|_{-T/2}^{T/2} \\ &= \frac{1}{T} \left(\frac{T^3}{24} + \frac{T^3}{24}\right) \\ &= \frac{T^2}{12}\end{aligned}$$

donde en (a) se consideró que $f(x) = 1/T$ en $(0, T)$ y 0 en otro caso, y en (b) se realizó el cambio de variable $u = x - T/2$. Por lo tanto, en este problema se tiene que (ver la ecuación 6.74)

$$\eta_i = \frac{T}{2}, \quad \sigma_i^2 = \frac{T^2}{12}, \quad \eta_n = n \frac{T}{2}, \quad \sigma_n^2 = n \frac{T^2}{12}.$$

De la ecuación 5.18, la densidad de la suma de variables aleatorias i.i.d. es la convolución de las densidades.

- Caso $n = 2$: la densidad es

$$f_2(x) = \int_{-\infty}^{\infty} f_1(y) f_1(x - y) dy,$$

y se distinguen los siguientes casos (ver la figura 5.17):

- $0 \leq x \leq T$:

$$f_2(x) = \int_0^x \frac{1}{T^2} dy = \frac{1}{T^2} y \Big|_0^x = \frac{x}{T^2}$$

- $T \leq x \leq 2T$:

$$f_2(x) = \int_{x-T}^T \frac{1}{T^2} dy = \frac{1}{T^2} y \Big|_{x-T}^T = \frac{1}{T^2} [T - (x - T)] = \frac{-x + 2T}{T^2}.$$

Se concluye que

$$f_2(x) = \begin{cases} \frac{x}{T^2}, & 0 \leq x \leq T \\ \frac{-x + 2T}{T^2}, & T \leq x \leq 2T \\ 0 & \text{en otro caso.} \end{cases}$$

- Caso $n = 3$: la densidad es

$$f_3(x) = \int_{-\infty}^{\infty} f_2(y) f_1(x - y) dy,$$

y se distinguen los siguientes casos:

- $0 \leq x \leq T$:

$$f_3(x) = \int_0^x \frac{y}{T^2} \times \frac{1}{T} dy = \frac{1}{T^3} \times \frac{y^2}{2} \Big|_0^x = \frac{x^2}{2T^3}$$

- $T \leq x \leq 2T$:

$$\begin{aligned}
f_3(x) &= \int_{x-T}^T \frac{y}{T^2} \times \frac{1}{T} dy + \int_T^x \frac{-y+2T}{T^2} \times \frac{1}{T} dy \\
&= \frac{y^2}{2T^3} \Big|_{x-T}^T + \frac{1}{T^3} \left(-\frac{y^2}{2} + 2Ty \right) \Big|_T^x \\
&= \frac{1}{2T^3} [T^2 - (x-T)^2] + \frac{1}{T^3} \left[\left(-\frac{x^2}{2} + 2Tx \right) - \left(-\frac{T^2}{2} + 2T^2 \right) \right] \\
&= \frac{1}{2T^3} [(-x^2 + 2Tx) + (-x^2 + 4Tx - 3T^2)] \\
&= \frac{1}{2T^3} (-2x^2 + 6Tx - 3T^2)
\end{aligned}$$

- $2T \leq x \leq 3T$:

$$\begin{aligned}
f_3(x) &= \int_{x-T}^{2T} \frac{-y+2T}{T^2} \times \frac{1}{T} dy \\
&= \frac{1}{T^3} \left(-\frac{y^2}{2} + 2Ty \right) \Big|_{x-T}^{2T} \\
&= \frac{1}{T^3} \left\{ \left[-\frac{4T^2}{2} + 4T^2 \right] - \left[-\frac{(x-T)^2}{2} + 2T(x-T) \right] \right\} \\
&= \frac{1}{2T^3} \{ 4T^2 - [(-x^2 + 2Tx - T^2) + (4Tx - 4T^2)] \} \\
&= \frac{1}{2T^3} (x^2 - 6Tx + 9T^2)
\end{aligned}$$

resultando en

$$f_3(x) = \begin{cases} \frac{x^2}{2T^3}, & 0 \leq x \leq T \\ \frac{1}{2T^3} (-2x^2 + 6Tx - 3T^2), & T \leq x \leq 2T \\ \frac{1}{2T^3} (x^2 - 6Tx + 9T^2), & 2T \leq x \leq 3T \\ 0 & \text{en otro caso.} \end{cases} \quad (6.76)$$

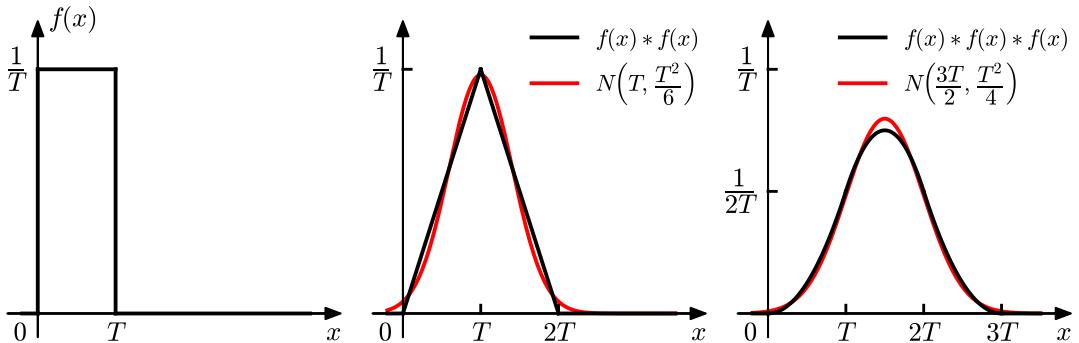


Figura 6.6: Ilustración del teorema central del límite. Se comparan las densidades de $\mathbf{x}_1 + \mathbf{x}_2$ y $\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3$, donde $\mathbf{x}_i \sim U(0, T)$, con las correspondientes aproximaciones gaussianas.

En la figura 6.6 se muestra la comparación de las densidades $f_2(x)$ y $f_3(x)$ de las variables aleatorias $\mathbf{x}_1 + \mathbf{x}_2$ y $\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3$ respectivamente con su correspondiente aproximación gaussiana, $N(T, T^2/6)$ y

$N(3T/2, T^2/4)$. Como se aprecia en la figura, el error es pequeño incluso para los valores chicos de n empleados.

Corrección del error

En la aproximación de $f(x)$ por una curva normal $N(\eta, \sigma^2)$, el error es

$$\epsilon(x) = f(x) - \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}, \quad (6.77)$$

donde sin perder generalidad se asumió que $\eta = 0$, lo que implica únicamente desplazar el origen de coordenadas. Se verá que este error se puede expresar en función de los momentos

$$m_n = E\{\mathbf{x}^n\}$$

de \mathbf{x} y los polinomios de Hermite [17], que se definen como

$$H_k(x) = (-1)^k e^{x^2/2} \frac{d^k}{dx^k} e^{-x^2/2}. \quad (6.78)$$

Estos polinomios forman una base ortogonal en el eje real respecto a la función de ponderación $e^{-x^2/2}$,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} H_n(x) H_m(x) dx = \begin{cases} n! & n = m \\ 0 & n \neq m, \end{cases}$$

o cambiando la variable x por x/σ ,

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-x^2/2\sigma^2} H_n\left(\frac{x}{\sigma}\right) H_m\left(\frac{x}{\sigma}\right) dx = \begin{cases} n! & n = m \\ 0 & n \neq m. \end{cases} \quad (6.79)$$

Empleando la base de polinomios de Hermite, la densidad de probabilidad $f(x)$ puede escribirse como la serie

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} \sum_{k=0}^{\infty} C_k H_k\left(\frac{x}{\sigma}\right). \quad (6.80)$$

Para encontrar los coeficientes C_n , se multiplica ambos lados de la igualdad por $H_n(x/\sigma)$ y se integra,

$$\int_{-\infty}^{\infty} f(x) H_n\left(\frac{x}{\sigma}\right) dx = \sum_{k=0}^{\infty} C_k \left[\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-x^2/2\sigma^2} H_k\left(\frac{x}{\sigma}\right) H_n\left(\frac{x}{\sigma}\right) dx \right].$$

Observando que el lado izquierdo de la igualdad es la esperanza de $H_n(x/\sigma)$ y empleando la ecuación 6.79 en el lado derecho, se obtiene que

$$E\left\{H_n\left(\frac{x}{\sigma}\right)\right\} = C_n n! \quad \Rightarrow \quad C_n = \frac{1}{n!} E\left\{H_n\left(\frac{x}{\sigma}\right)\right\}. \quad (6.81)$$

Se verá ahora que los coeficientes C_n pueden expresarse en función de los momentos m_n de \mathbf{x} . Para hacerlo, se calculará $H_n(x)$ para $n = 0, \dots, 4$ con la ecuación 6.78. Primero se parte observando que

$$\begin{aligned} \frac{d^0}{dx^0} e^{-x^2/2} &= e^{-x^2/2} \\ \frac{d}{dx} e^{-x^2/2} &= -\frac{2x}{2} e^{-x^2/2} = -x e^{-x^2/2} \\ \frac{d^2}{dx^2} e^{-x^2/2} &= -e^{-x^2/2} - x(-x e^{-x^2/2}) = (x^2 - 1) e^{-x^2/2} \\ \frac{d^3}{dx^3} e^{-x^2/2} &= 2x e^{-x^2/2} + (x^2 - 1)(-x e^{-x^2/2}) = (-x^3 + 3x) e^{-x^2/2} \\ \frac{d^4}{dx^4} e^{-x^2/2} &= (-3x^2 + 3) e^{-x^2/2} + (-x^3 + 3x)(-x e^{-x^2/2}) = (x^4 - 6x^2 + 3) e^{-x^2/2}, \end{aligned}$$

y sustituyendo estos resultados en la ecuación 6.78 se obtiene que

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = x^2 - 1, \quad H_3(x) = x^3 - 3x, \quad H_4(x) = x^4 - 6x^2 + 3.$$

Finalmente, la ecuación 6.81 conduce a que los coeficientes son

$$\begin{aligned} C_0 &= E\left\{H_0\left(\frac{x}{\sigma}\right)\right\} = 1 \\ C_1 &= E\left\{H_1\left(\frac{x}{\sigma}\right)\right\} = E\left\{\frac{x}{\sigma}\right\} = \frac{\eta}{\sigma} = 0 \\ C_2 &= \frac{1}{2!}E\left\{H_2\left(\frac{x}{\sigma}\right)\right\} = \frac{1}{2}E\left\{\frac{x^2}{\sigma^2} - 1\right\} = \frac{1}{2}\left(\frac{\sigma^2}{\sigma^2} - 1\right) = 0 \\ C_3 &= \frac{1}{3!}E\left\{H_3\left(\frac{x}{\sigma}\right)\right\} = \frac{1}{6}E\left\{\frac{x^3}{\sigma^3} - \frac{3x}{\sigma}\right\} = \frac{m_3}{6\sigma^3} \\ C_4 &= \frac{1}{4!}E\left\{H_4\left(\frac{x}{\sigma}\right)\right\} = \frac{1}{24}E\left\{\frac{x^4}{\sigma^4} - \frac{6x^2}{\sigma^2} + 3\right\} = \frac{1}{24}\left(\frac{m_4}{\sigma^4} - \frac{6\sigma^2}{\sigma^2} + 3\right) = \frac{1}{24}\left(\frac{m_4}{\sigma^4} - 3\right), \end{aligned}$$

en donde se tuvo en cuenta que la media $\eta = E\{\mathbf{x}\}$ es nula por hipótesis. Considerando hasta el tercer coeficiente en la ecuación 6.80 de la expansión de Hermite, la densidad puede aproximarse como

$$\begin{aligned} f(x) &\approx \frac{1}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}\left[C_0H_0\left(\frac{x}{\sigma}\right) + C_3H_3\left(\frac{x}{\sigma}\right)\right] \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}\left[1 + \frac{m_3}{6\sigma^3}\left(\frac{x^3}{\sigma^3} - \frac{3x}{\sigma}\right)\right], \end{aligned}$$

y de la ecuación 6.77 del error de aproximación, el error puede aproximarse como

$$\epsilon(x) = f(x) - \frac{1}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2} \approx \frac{1}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}\frac{m_3}{6\sigma^3}\left(\frac{x^3}{\sigma^3} - \frac{3x}{\sigma}\right).$$

Si $f(x)$ es una función par, $m_3 = 0$ y la densidad se puede aproximar como

$$\begin{aligned} f(x) &\approx \frac{1}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}\left[C_0H_0\left(\frac{x}{\sigma}\right) + C_4H_4\left(\frac{x}{\sigma}\right)\right] \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}\left[1 + \frac{1}{24}\left(\frac{m_4}{\sigma^4} - 3\right)\left(\frac{x^4}{\sigma^4} - \frac{6x^2}{\sigma^2} + 3\right)\right]. \end{aligned} \quad (6.82)$$

Ejemplo 7-17 Sean las variables aleatorias \mathbf{x}_i i.i.d con distribución $f_i(x)$ uniforme en el intervalo $(-T/2, T/2)$. La densidad $f(x)$ de $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3$ es de la forma de la obtenida en el ejemplo 7-15, que se muestra en la figura 6.6, pero centrada en cero, es decir, desplazada una cantidad $3T/2$ a la izquierda. Por lo tanto, su ecuación está dada por la ecuación 6.76 evaluda en $x + 3T/2$, y es

$$f(x) = \begin{cases} \frac{1}{8T^3}(4x^2 + 12Tx + 9T^2), & -\frac{3T}{2} \leq x \leq -\frac{T}{2} \\ \frac{1}{4T^3}(-4x^2 + 3T^2), & -\frac{T}{2} \leq x \leq \frac{T}{2} \\ \frac{1}{8T^3}(4x^2 - 12Tx + 9T^2), & \frac{T}{2} \leq x \leq \frac{3T}{2} \\ 0 & \text{en otro caso.} \end{cases}$$

Considerando el caso particular con $T = 1$, es decir, $\mathbf{x}_i \sim U(-1/2, 1/2)$, la densidad queda

$$f(x) = \begin{cases} \frac{1}{8}(4x^2 + 12x + 9), & -\frac{3}{2} \leq x \leq -\frac{1}{2} \\ \frac{1}{4}(-4x^2 + 3), & -\frac{1}{2} \leq x \leq \frac{1}{2} \\ \frac{1}{8}(4x^2 - 12x + 9), & \frac{1}{2} \leq x \leq \frac{3}{2} \\ 0 & \text{en otro caso.} \end{cases}$$

y la aproximación gaussiana es $N(0, 1/4)$, ya que $\sigma^2 = 3\sigma_i^2 = 3/12 = 1/4$. Además, como $f(x)$ es simétrica, $m_3 = 0$ y la corrección de primer orden está dada por la ecuación 6.82. Para calcularla, se observa que el momento cuarto de \mathbf{x} es

$$\begin{aligned} m_4 &= E\{(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3)^4\} \\ &\stackrel{(a)}{=} E\{\mathbf{x}_1^4\} + E\{\mathbf{x}_2^4\} + E\{\mathbf{x}_3^4\} + 6E\{\mathbf{x}_1^2\}E\{\mathbf{x}_2^2\} + 6E\{\mathbf{x}_1^2\}E\{\mathbf{x}_3^2\} + 6E\{\mathbf{x}_2^2\}E\{\mathbf{x}_3^2\} \\ &= 3m_{i4} + 18\sigma_i^4, \end{aligned}$$

donde en (a) se tuvo en cuenta que luego de expandir el trinomio a la cuarta¹², aplicar la propiedad de linealidad de la esperanza y considerar que las variables aleatorias son independientes y por lo tanto $E\{\mathbf{x}_1^i \mathbf{x}_2^j \mathbf{x}_3^k\} = E\{\mathbf{x}_1^i\}E\{\mathbf{x}_2^j\}E\{\mathbf{x}_3^k\}$, todos los sumandos que contienen algún factor $E\{\mathbf{x}_i\}$ o $E\{\mathbf{x}_i^3\}$ se anulan debido a que $f(x)$ es simétrica. Considerando que la varianza de \mathbf{x}_i es $\sigma_i^2 = 1/12$, y el momento cuarto es

$$m_{i4} = \int_{-\infty}^{\infty} x^4 f_i(x) dx = \int_{-1/2}^{1/2} x^4 dx = \frac{x^5}{5} \Big|_{-1/2}^{1/2} = \frac{1}{5} \left(\frac{1}{32} + \frac{1}{32} \right) = \frac{1}{80},$$

el momento cuarto de \mathbf{x} es

$$m_4 = \frac{3}{80} + \frac{18}{12^2} = \frac{13}{80}.$$

De esta forma, con $\sigma^2 = 1/4$ y $m_4 = 13/80$, la corrección de primer orden $\bar{f}(x)$ es

$$\begin{aligned} \bar{f}(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} \left[1 + \frac{1}{24} \left(\frac{m_4}{\sigma^4} - 3 \right) \left(\frac{x^4}{\sigma^4} - \frac{6x^2}{\sigma^2} + 3 \right) \right] \\ &= \sqrt{\frac{2}{\pi}} e^{-2x^2} \left[1 + \frac{1}{24} \left(\frac{\frac{13}{80}}{\frac{1}{16}} - 3 \right) \left(\frac{x^4}{\frac{1}{16}} - \frac{6x^2}{\frac{1}{4}} + 3 \right) \right] \\ &= \sqrt{\frac{2}{\pi}} e^{-2x^2} \left[1 + \frac{1}{24} \left(\frac{13}{5} - 3 \right) (16x^4 - 24x^2 + 3) \right] \\ &= \sqrt{\frac{2}{\pi}} e^{-2x^2} \left[1 - \frac{1}{60} (16x^4 - 24x^2 + 3) \right], \end{aligned}$$

resultando en

$$\bar{f}(x) = \sqrt{\frac{2}{\pi}} e^{-2x^2} \left(1 - \frac{4x^4}{15} + \frac{2x^2}{5} - \frac{1}{20} \right).$$

En la figura 6.7 se compara la densidad verdadera $f(x)$ con la aproximación gaussiana dada por el teorema central del límite, que es $N(0, 1/4)$, y la corrección $\bar{f}(x)$.

Un comentario adicional sobre este ejemplo es que el momento cuarto de \mathbf{x} puede calcularse empleando la función generadora de momentos. Como las variables aleatorias \mathbf{x}_i son i.i.d., del teorema de la convolución (ver la ecuación 5.63), la función generadora de momentos de $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3$ es

$$\Phi(s) = \Phi_i^3(s),$$

donde $\Phi_i(s)$ es la función generadora de momentos de \mathbf{x}_i . La función generadora de momentos de una variable aleatoria uniforme en el intervalo $(-T/2, T/2)$ se calculó en la sección 4.5.5 y su expansión de Taylor está dada por la ecuación 4.43, que con $T = 1$ es

$$\begin{aligned} \Phi_i(s) &= \sum_{k=0}^{\infty} \frac{1}{2^{2k}(2k+1)!} s^{2k} \\ &= 1 + \frac{1}{2^2 \times 3!} s^2 + \frac{1}{2^4 \times 5!} s^4 + \dots \\ &= 1 + \frac{1}{24} s^2 + \frac{1}{1920} s^4 + \dots \end{aligned}$$

¹²Ver https://en.wikipedia.org/wiki/Trinomial_expansion, por ejemplo.

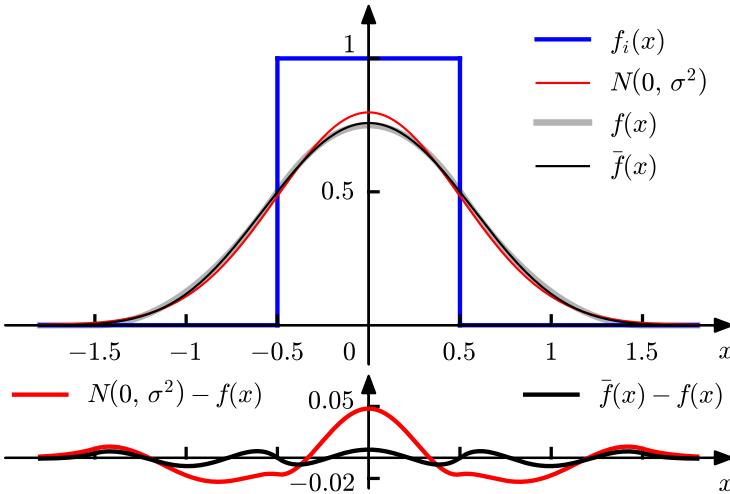


Figura 6.7: Corrección del error mediante la expansión Hermite. Se compara la densidad verdadera $f(x)$ de la variable aleatoria $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3$, con $\mathbf{x}_i \sim f_i(x)$ independientes, con la aproximación gaussiana dada por el teorema central del límite y la corrección $\bar{f}(x)$.

Combinando los resultados, la función generadora de momentos de \mathbf{x} es

$$\Phi(s) = \left(1 + \frac{1}{24}s^2 + \frac{1}{1920}s^4 + \dots\right)^3. \quad (6.83)$$

De la ecuación 4.38, el momento cuarto de \mathbf{x} es la derivada cuarta de la función generadora de momentos evaluada en $s = 0$,

$$m_4 = E\{\mathbf{x}^4\} = \Phi^{(4)}(0).$$

Al derivar n veces un polinomio y evaluar en cero, el único término que no se anula es el que corresponde al grado n . Al desarrollar la potencia al cubo en la ecuación 6.83, el término de grado 4 es

$$\Phi(s) = \dots + \frac{3}{24^2}s^4 + \frac{3}{1920}s^4 + \dots = \dots + \frac{1}{192}s^4 + \frac{3}{1920}s^4 + \dots = \dots + \frac{13}{1920}s^4 + \dots$$

y al derivar cuatro veces y evaluar en cero, se obtiene que

$$m_4 = \Phi^{(4)}(0) = \frac{13 \times 4!}{1920} = \frac{13}{80}.$$

Sobre la prueba del teorema central del límite Se justificará la aproximación de la ecuación 6.75 empleando funciones características. La demostración puede hacerse de forma muy similar a la demostración de la ley débil de los grandes números incluida en la sección 6.4.2¹³, pero se seguirá el razonamiento realizado en [1]. Por simplicidad, se asumirá que $\eta_i = 0$. Denotando como $\Phi_i(\omega)$ y $\Phi(\omega)$ a las funciones características de las variables aleatorias \mathbf{x}_i y $\mathbf{x} = \mathbf{x}_1 + \dots + \mathbf{x}_n$ respectivamente, por la hipótesis de independencia se cumple que

$$\Phi(\omega) = \Phi_1(\omega) \dots \Phi_n(\omega).$$

Las funciones características segundas $\Psi_i(\omega) = \ln \Phi_i(\omega)$ en torno al origen pueden aproximarse por una parábola. Efectivamente, realizando un desarrollo de Taylor en torno a cero, se tiene que

$$\Psi_i(\omega) = \Psi_i(0) + \frac{\Psi'_i(0)}{1!}\omega + \frac{\Psi''_i(0)}{2!}\omega^2 + O(\omega^3).$$

De la ecuación 4.24 de la definición de la función característica se deduce que

$$\Phi_i(0) = 1, \quad \Phi'_i(0) = jE\{\mathbf{x}_i\} = j\eta_i = 0, \quad \Phi''_i(0) = -E\{\mathbf{x}_i^2\} = -(\eta_i^2 + \sigma_i^2) = -\sigma_i^2$$

¹³Ver https://en.wikipedia.org/wiki/Central_limit_theorem#Proof_of_classical_CLT, por ejemplo.

y como

$$\begin{aligned}\Psi_i(\omega) &= \ln \Phi_i'(\omega) & \Psi_i(0) &= 0 \\ \Psi_i'(\omega) &= \frac{\Phi_i'(\omega)}{\Phi_i(\omega)} & \Rightarrow & \Psi_i'(0) = 0 \\ \Psi_i''(\omega) &= \frac{\Phi_i''(\omega)\Phi_i(\omega) - \Phi_i'^2(\omega)}{\Phi_i^2(\omega)} & \Psi_i''(0) &= -\sigma_i^2,\end{aligned}$$

el desarrollo de Taylor queda

$$\Psi_i(\omega) = -\frac{1}{2}\sigma_i^2\omega^2 + O(\omega^3). \quad (6.84)$$

Por lo tanto, las funciones características se pueden aproximar por

$$\Phi_i(\omega) \approx e^{-\sigma_i^2\omega/2}, \quad \text{para } |\omega| < \epsilon,$$

resultando en que

$$\Phi(\omega) \approx e^{-\sigma_1^2\omega/2} \dots e^{-\sigma_n^2\omega/2} = e^{-\sigma^2\omega/2}. \quad (6.85)$$

Observar que si las variables aleatorias son continuas, se cumple que (ver la ecuación 4.25)

$$\Phi_i(0) = 1 \quad \text{y} \quad |\Phi_i(\omega)| < 1, \quad \text{para } |\omega| \neq 0.$$

Esto sugiere que para ϵ pequeño y n grande, la función $\Phi(\omega)$ es despreciable para $|\omega| > \epsilon$. Lo mismo ocurre con la exponencial $e^{-\sigma^2\omega/2}$ si $\sigma^2 \rightarrow \infty$. En ese caso, la aproximación de la ecuación 6.85 es válida para todo ω , y considerando que, como indica la ecuación 4.29, se trata de la función característica de una variable aleatoria con densidad $N(0, \sigma^2)$, el resultado es acorde a la aproximación del teorema central del límite dado por la ecuación 6.75.

La forma exacta del teorema establece que la variable aleatoria normalizada

$$\mathbf{z} = \frac{\mathbf{x}_1 + \dots + \mathbf{x}_n}{\sigma}, \quad \text{donde } \sigma^2 = \sigma_1^2 + \dots + \sigma_n^2,$$

tiende a una variable aleatoria $N(0, 1)$ cuando $n \rightarrow \infty$, es decir,

$$f_z(z) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}. \quad (6.86)$$

A continuación, se esboza una prueba bajo la hipótesis de que las variables aleatorias \mathbf{x}_i son i.i.d. En ese caso,

$$\Phi_1(\omega) = \dots = \Phi_n(\omega) \quad \text{y} \quad \sigma = \sigma_i\sqrt{n},$$

por lo que

$$\Phi_z(\omega) = \Phi_i^n \left(\frac{\omega}{\sigma_i\sqrt{n}} \right),$$

donde se empleó el hecho de que la función característica de la variable aleatoria \mathbf{x}_i/σ es $\Phi_i(\omega/\sigma)$, como indica la ecuación 4.27. Tomando el logaritmo, se obtiene que

$$\Psi_z(\omega) = n\Psi_i \left(\frac{\omega}{\sigma_i\sqrt{n}} \right).$$

Empleando la ecuación 6.84 de la expansión de Taylor de la función característica segunda $\Psi_i(\omega) = \ln \Phi_i(\omega)$ en torno al origen, se tiene que

$$\begin{aligned}\Psi_z(\omega) &= n\Psi_i \left(\frac{\omega}{\sigma_i\sqrt{n}} \right) = n \left\{ -\frac{\sigma_i^2}{2} \left(\frac{\omega}{\sigma_i\sqrt{n}} \right)^2 + O \left[\left(\frac{\omega}{\sigma_i\sqrt{n}} \right)^3 \right] \right\} \\ &\stackrel{(a)}{=} -\frac{\omega^2}{2} + O \left(\frac{\omega^3}{\sqrt{n}} \right),\end{aligned}$$

donde en (a) se consideró que $nO(1/n^{3/2}) = O(n/n^{3/2}) = O(1/\sqrt{n})$. Se concluye que

$$\Psi_z(\omega) = -\frac{\omega^2}{2} + O\left(\frac{\omega^3}{\sqrt{n}}\right) \xrightarrow[n \rightarrow \infty]{} -\frac{\omega^2}{2},$$

lo que muestra que

$$\Phi_z(\omega) \xrightarrow[n \rightarrow \infty]{} e^{-\omega^2/2},$$

obteniendo el resultado de la ecuación 6.86. Observar que lo que permite concluir que se cumple la ecuación 6.86 a partir de la convergencia de la función característica, es el teorema de continuidad de Levy, como en la demostración de la ley débil de los grandes números de la sección 6.4.2.

Capítulo 7

Estadística

7.1. Introducción

Probabilidad es una disciplina matemática desarrollada como un modelo abstracto y sus conclusiones son deducciones basadas en axiomas. La estadística trata con las aplicaciones de la teoría de probabilidad en problemas reales y sus conclusiones son inferencias basadas en observaciones. La conexión entre los conceptos probabilísticos y la realidad reside en la aproximación

$$p \approx \frac{n_A}{n},$$

que relaciona la probabilidad $p = P(A)$ de un evento A con el número n_A de ocurrencias de A en n ensayos del experimento físico subyacente. Esta fórmula empírica se emplea para brindar la interpretación de frecuencia relativa de todos los conceptos probabilísticos.

En una investigación estadística se trata con dos clases generales de problemas. En la primera clase, se asume que el modelo probabilístico es conocido y el objetivo es realizar predicciones vinculadas a observaciones futuras. Por ejemplo, se conoce la distribución de probabilidad $F(x)$ de una variable aleatoria \mathbf{x} y se desea predecir el promedio \bar{x} de n muestras futuras, o se conoce la probabilidad p de un evento A y se quiere predecir la cantidad de ocurrencias n_A de A en n ensayos. En ambos casos, se parte del modelo para inferir sobre las observaciones. En la segunda clase, uno o mas parámetros θ_i del modelo son desconocidos y el objetivo es estimar sus valores (estimación de parámetros) o decidir si θ_i es un conjunto de constantes conocidas θ_{0i} (test de hipótesis). Por ejemplo, se observan los valores x_i de una variable aleatoria \mathbf{x} y se desea estimar su media η , o decidir si aceptar la hipótesis de que $\eta = 5.3$. Una moneda es lanzada al aire 1000 veces, y se obtiene cara 465 veces. Con esta información, se quiere estimar la probabilidad p de la cara, o decidir si la moneda es justa. En ambos casos, se parte de las observaciones para realizar inferencias sobre el modelo. Este capítulo se concentra en los problemas de estimación de parámetros y test de hipótesis. Como preparación, a continuación se comenta brevemente sobre el problema de predicción.

Predicción Sea una variable aleatoria \mathbf{x} con distribución conocida de la cual se quiere predecir su valor en un ensayo futuro. Una *predicción puntual* de \mathbf{x} es la determinación de una constante c elegida de forma de minimizar el error $\mathbf{x} - c$ en algún sentido. En un ensayo específico, \mathbf{x} puede tomar muchos valores, por lo que el valor que tomará no puede predecirse, solo puede estimarse. Si el criterio para la elección de c es la minimización del error $MS E\{(\mathbf{x} - c)^2\}$, se cumple que $c = E\{\mathbf{x}\}$, como se mostró en la sección 6.3.

Un *intervalo de predicción* de \mathbf{x} es la determinación de dos constantes c_1 y c_2 de forma de que

$$P\{c_1 < \mathbf{x} < c_2\} = \gamma = 1 - \delta, \quad (7.1)$$

donde γ es una constante dada llamada *coeficiente de confianza*. Esta ecuación indica que si se predice que el valor x que tomará \mathbf{x} en el próximo ensayo va a estar en el intervalo (c_1, c_2) , la predicción va a ser correcta el $100\gamma\%$ de los casos. El problema en intervalo de predicción es encontrar c_1 y c_2 de forma de minimizar la diferencia $c_2 - c_1$ sujeta a la restricción dada por la ecuación 7.1. La elección de γ se

rigue por dos requerimientos en conflicto. Si γ es cercano a 1, la predicción de que x estará en el intervalo (c_1, c_2) es confiable, pero la diferencia $c_2 - c_1$ es grande y brinda poca información sobre x . Si se reduce γ , se reduce $c_2 - c_1$, pero la estimación es menos confiable. Se mostrará a continuación que si la densidad de probabilidad $f(x)$ de x es unimodal, $c_2 - c_1$ es mínimo para un γ dado si $f(c_1) = f(c_2)$ (teorema 9.3.2 en [18]):

Teorema: Sea $f(x)$ una densidad unimodal. Si el intervalo (c_1, c_2) satisface que

1. $P\{c_1 < x < c_2\} = \int_{c_1}^{c_2} f(x) dx = \gamma$,
2. $f(c_1) = f(c_2) > 0$, y
3. $c_1 \leq x^* \leq c_2$, donde x^* es la moda de $f(x)$ (ubicación del máximo),

el intervalo (c_1, c_2) es el mas corto entre todos los intervalos que cumplen 1.

Demuestra: Sea (c'_1, c'_2) cualquier intervalo tal que $c'_2 - c'_1 < c_2 - c_1$. Se demostrará que esto implica que

$$P\{c'_1 < x < c'_2\} = \int_{c'_1}^{c'_2} f(x) dx < \gamma.$$

El caso en que $(c'_1, c'_2) \subset (c_1, c_2)$ es trivial, ya que inmediatamente se observa que se cumple que

$$\int_{c'_1}^{c'_2} f(x) dx < \int_{c_1}^{c_2} f(x) dx = \gamma.$$

La prueba se realizará únicamente para el caso en que $c'_1 \leq c_1$, dado que la prueba para el caso en que $c'_1 > c_1$ es análoga. Asumiendo entonces que $c'_1 \leq c_1$, hay que distinguir dos casos, $c'_2 \leq c_1$ y $c'_2 > c_1$, que son las situaciones que se muestran en la figura 7.1. En el ejemplo de esta figura, los intervalos (c_1, c_2) y (c'_1, c'_2) tienen la misma probabilidad γ y (c_1, c_2) es el mas corto de todos los posibles, por lo que no se cumple que $c'_2 - c'_1 < c_2 - c_1$, como se asume en el teorema.

- Caso $c'_2 \leq c_1$: se cumple que $c'_1 \leq c'_2 \leq c_1 \leq x^*$. Por lo tanto,

$$\int_{c'_1}^{c'_2} f(x) dx \stackrel{(a)}{\leq} f(c'_2)(c'_2 - c'_1) \stackrel{(b)}{\leq} f(c_1)(c'_2 - c'_1) \stackrel{(c)}{<} f(c_1)(c_2 - c_1) \stackrel{(d)}{\leq} \int_{c_1}^{c_2} f(x) dx = \gamma,$$

donde en (a) se empleó que como $x \leq c'_2 \leq x^*$, $f(x) \leq f(c'_2)$ para todo $x \in (c'_1, c'_2)$, en (b) se empleó que $c'_2 \leq c_1 \leq x^*$ y por lo tanto, $f(c'_2) \leq f(c_1)$, en (c) se tuvo en cuenta que $c'_2 - c'_1 < c_2 - c_1$ y $f(c_1) > 0$ y en (d) se consideró que por 2., 3. y la unimodalidad, $f(x) \geq f(c_1)$ para todo $x \in (c_1, c_2)$, completando la prueba para este caso.

- Caso $c'_2 > c_1$: se cumple que $c'_1 \leq c_1 < c'_2 < c_2$. No puede ocurrir que $c'_2 \geq c_2$ por la condición de que $c'_2 - c'_1 > c_2 - c_1$. En este caso, se tiene que

$$\begin{aligned} \int_{c'_1}^{c'_2} f(x) dx &= \int_{c_1}^{c_2} f(x) dx + \left[\int_{c'_1}^{c_1} f(x) dx - \int_{c'_2}^{c_2} f(x) dx \right] \\ &= \gamma + \left[\int_{c'_1}^{c_1} f(x) dx - \int_{c'_2}^{c_2} f(x) dx \right], \end{aligned}$$

por lo que la demostración del teorema se completa demostrando que el término entre paréntesis rectos es negativo. Considerando 3., la unimodalidad y el orden $c'_1 \leq c_1 < c'_2 < c_2$, se deduce que

$$\int_{c'_1}^{c_1} f(x) dx \leq f(c_1)(c_1 - c'_1) \quad \text{y} \quad \int_{c'_2}^{c_2} f(x) dx \geq f(c_2)(c_2 - c'_2),$$

y por lo tanto,

$$\begin{aligned}
\int_{c'_1}^{c_1} f(x) dx - \int_{c'_2}^{c_2} f(x) dx &\leq f(c_1)(c_1 - c'_1) - f(c_2)(c_2 - c'_2) \\
&\stackrel{(a)}{\leq} f(c_1) [(c_1 - c'_1) - (c_2 - c'_2)] \\
&= f(c_1) [(c'_2 - c'_1) - (c_2 - c_1)] \\
&\stackrel{(b)}{<} 0,
\end{aligned}$$

donde en (a) se tuvo en cuenta 2. y en (b) que $c'_2 - c'_1 < c_2 - c_1$ y $f(c_1)$, concluyendo la prueba.

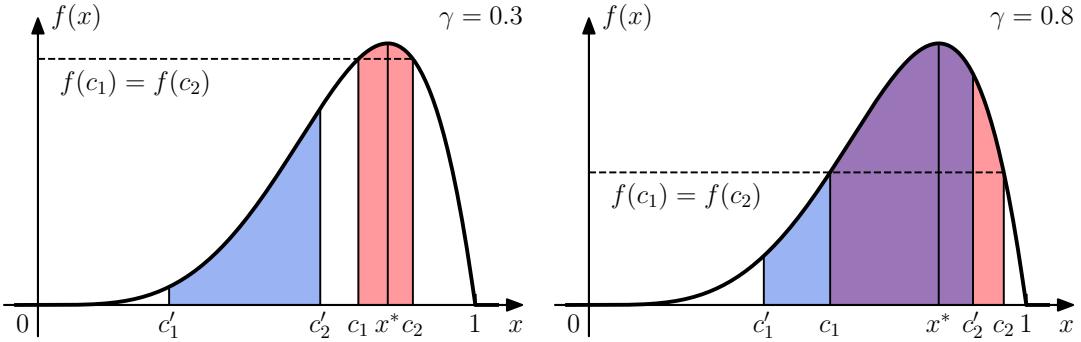


Figura 7.1: Comparación del intervalo de confianza mas corto (c_1, c_2) con otro intervalo (c'_1, c'_2) con la misma probabilidad γ para valores de γ de 0.3 y 0.8 en el caso una variable aleatoria con distribución $\beta(5, 2)$. El intervalo con $c_2 - c_1$ mínimo es el que cumple que $f(c_1) = f(c_2)$.

En la figura 7.1 se muestra el intervalo (c_1, c_2) mas corto para dos valores de γ y se compara con otro intervalo (c'_1, c'_2) con la misma probabilidad en el caso de una variable aleatoria con distribución beta. Se observa que para cierto coeficiente de confianza γ , el intervalo mas corto es el que cumple que $f(c_1) = f(c_2)$. Los valores de c_1 y c_2 pueden obtenerse mediante ensayo y error. En la figura 7.2 se muestra el código en el lenguaje Python para obtener el intervalo de confianza mas corto dado el coeficiente de confianza γ ¹. Una solución subóptima mas simple se obtiene fácilmente si se determina c_1 y c_2 tal que

$$P\{\mathbf{x} < c_1\} = \frac{\delta}{2} \quad \text{y} \quad P\{\mathbf{x} > c_2\} = \frac{\delta}{2}.$$

De esta forma,

$$c_1 = F^{-1}\left(\frac{\delta}{2}\right) = x_{\delta/2} \quad \text{y} \quad c_2 = F^{-1}\left(1 - \frac{\delta}{2}\right) = x_{1-\delta/2},$$

donde x_u es el percentil u y $F(x)$ es la distribución de probabilidad de \mathbf{x} . El intervalo obtenido se denomina *intervalo de confianza de colas iguales* y es óptimo si la densidad $f(x)$ es simétrica respecto a su media η , ya que en ese caso, $f(c_1) = f(c_2)$. Si \mathbf{x} es normal, $x_u = \eta + z_u\sigma$, donde z_u es el percentil normal estándar. Efectivamente, sea $\mathbf{x} \sim N(\eta, \sigma^2)$. De la definición de percentil (ver la sección 3.2.1), el percentil x_u de \mathbf{x} es $u = P\{\mathbf{x} < x_u\}$ y por lo tanto, normalizando, se cumple que

$$P\left\{\frac{\mathbf{x} - \eta}{\sigma} < \frac{x_u - \eta}{\sigma}\right\} = u,$$

o $P\{\mathbf{z} < z_u\} = u$, donde $\mathbf{z} = (\mathbf{x} - \eta)/\sigma \sim N(0, 1)$ y $z_u = (x_u - \eta)/\sigma$.

¹El código fué extraído de <https://stackoverflow.com/a/25777507>

```

from scipy.optimize import fmin

def shortest_confidence_interval(dist_name, confidence_coef=0.95, **args):
    """ Cálculo del intervalo de confianza más corto.

    :param dist_name: nombre scipy de la distribución (str)
    :param confidence_coef: coeficiente de confianza (float entre 0 y 1)
    :param args: argumentos scipy de la distribución
    :return: límite inferior y superior del intervalo (array [low, high])

    Ejemplo de uso:

    c1, c2 = shortest_confidence_interval(beta, confidence_coef=0.9, a=5, b=3)
    """
    # configuración de la distribución con los argumentos dados
    distri = dist_name(**args)
    # estimación inicial de la probabilidad (área) de la cola inferior
    low_tail_pr_ini = 1.0 - confidence_coef

    def interval_width(low_tail_pr):
        """
        Cálculo del intervalo de confianza cuando el área de la cola inferior es
        low_tail_pr, por lo que el límite inferior es  $F^{-1}(low\_tail\_pr)$ .

        :param low_tail_pr: probabilidad de la cola inferior (float entre 0 y 1)
        :return: largo del intervalo (float)
        """
        return distri.ppf(confidence_coef + low_tail_pr) - distri.ppf(low_tail_pr)

    # Busqueda de la probabilidad de la cola inferior (low_tail_pr) que minimiza el
    # intervalo de confianza (función interval_width) fijando el coeficiente de confianza.
    low_tail_pr_short = fmin(interval_width, low_tail_pr_ini, ftol=1e-8, disp=False)[0]
    # Se retorna el intervalo como un array([low, high])
    return distri.ppf([low_tail_pr_short, confidence_coef + low_tail_pr_short])

```

Figura 7.2: Implementación en el lenguaje Python del cálculo del intervalo (c_1, c_2) con $c_2 - c_1$ mínimo para un coeficiente de confianza γ dado. Se emplea la función `fmin` de la biblioteca Scipy para minimizar la función `interval_width`. La función `interval_width` recibe como argumento el área de la cola inferior `low_tail_pr` y retorna el largo $c_2 - c_1$ del intervalo dado el coeficiente de confianza `confidence_coef`. El largo del intervalo se calcula como $c_2 - c_1 = F^{-1}(\gamma + \alpha) - F^{-1}(\alpha)$, donde $F(x)$ es la distribución de probabilidad y α el área de la cola inferior (variable `low_tail_pr`). La condición inicial del algoritmo de optimización es `low_tail_pr_ini = 1 - confidence_coef`, es decir, $\alpha = 1 - \gamma$. De esta forma, al inicio, $c_1 = F^{-1}(1 - \gamma)$, y $c_2 = F^{-1}(1)$ que es la frontera superior de la densidad si es de soporte acotado o infinito si no es de soporte acotado. Observar que el algoritmo no emplea el hecho de que el intervalo de largo mínimo es el que cumple que $f(c_1) = f(c_2)$.

7.2. Estimación

Supóngase que la distribución de una variable aleatoria \mathbf{x} es una función $F(x, \theta)$ de forma conocida que depende de un parámetro θ , escalar o vectorial, y se quiere estimar θ . Para hacerlo, se repite el experimento subyacente n veces, y denotando con x_i a los valores observados de \mathbf{x} , hay que encontrar un estimador puntual y un intervalo de estimación de θ empleando esas observaciones.

Una *estimación puntual* es una función $\hat{\theta} = g(X)$ del vector de observación $X = [x_1, \dots, x_n]$. La variable aleatoria correspondiente $\hat{\theta} = g(\mathbf{X})$ es el *estimador puntual* de θ . Cualquier función del vector de observación $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ se denomina *estadístico*.

Se dice que $\hat{\theta}$ es un estimador *insesgado* del parámetro θ si $E\{\hat{\theta}\} = \theta$. En caso contrario, se dice que el estimador es sesgado con sesgo $b = E\{\hat{\theta}\} - \theta$. Si la función $g(X)$ se elige adecuadamente, el error de estimación $\hat{\theta} - \theta$ decrece cuando n crece. Si tiende a 0 en probabilidad cuando $n \rightarrow \infty$, se dice que $\hat{\theta}$ es un estimador *consistente*. La media muestral $\bar{\mathbf{x}}$ de \mathbf{x} es un estimador insesgado de la media η . Adicionalmente, su varianza σ^2/n tiende a 0 con $n \rightarrow \infty$ (ver el ejemplo de la sección 6.1.3). Esto implica que $\bar{\mathbf{x}}$ tiende a η en media cuadrática, y por lo tanto, también en probabilidad (ver la sección 6.4.1). En otras palabras, $\bar{\mathbf{x}}$ es un estimador consistente de η . El problema de estimación consiste por lo tanto en la elección de una función $g(X)$ que minimice en algún sentido el error de estimación $g(\mathbf{X}) - \theta$. Si $g(X)$

se elige de forma de minimizar el error cuadrático medio

$$e = E \{ [g(\mathbf{X}) - \theta]^2 \} = \int_R [g(X) - \theta]^2 f(X, \theta) dX, \quad (7.2)$$

el estimador $\hat{\theta} = g(\mathbf{X})$ se denomina *mejor estimador*. Es importante notar que $f(X, \theta)$ en la ecuación 7.2 no es la densidad de probabilidad conjunta entre \mathbf{X} y θ , si no que es la densidad de \mathbf{X} paramétrica con θ . De hecho, θ no es una variable aleatoria. Determinar el mejor estimador no es sencillo en general, ya que el integrando de la ecuación 7.2, además de depender de X , depende de θ . El problema correspondiente de predicción involucra la misma integral, pero tiene una solución mas simple, ya que la densidad de probabilidad es conocida (ver las ecuaciones 6.47 y 6.61).

7.2.1. Estimación de intervalos

En el problema de estimación de un parámetro θ , ¿es posible extraer con certeza alguna conclusión sobre el valor verdadero de θ ? No es posible afirmar que θ es idéntico al estimador puntual $\hat{\theta}$ o a alguna otra constante. Sin embargo, se podría concluir casi con certeza que θ iguala a $\hat{\theta}$ con ciertos límites de tolerancia especificados. Esto conduce al concepto de *estimación de intervalos*.

Una *estimación de intervalo* de un parámetro θ es un intervalo (θ_1, θ_2) cuyos extremos son funciones $\theta_1 = g_1(X)$ y $\theta_2 = g_2(X)$ del vector de observación X . El intervalo aleatorio (θ_1, θ_2) correspondiente es el *estimador de intervalo* de θ . Se dice que (θ_1, θ_2) es un intervalo de confianza de γ del parámetro θ si

$$P\{\theta_1 < \theta < \theta_2\} = \gamma. \quad (7.3)$$

La constante γ es el *coeficiente de confianza* del estimador y la diferencia $\delta = 1 - \gamma$ es el *nivel de confianza*. De esta forma, γ es una medida de la confianza de que el parámetro desconocido θ pertenezca al intervalo (θ_1, θ_2) . Si γ es cercano a 1, puede afirmarse casi con certeza que eso es cierto, y el estimador del intervalo es correcto el $100\gamma\%$ de los casos. El objetivo en la estimación de intervalo es determinar las funciones $g_1(X)$ y $g_2(X)$ de forma de minimizar el largo $\theta_2 - \theta_1$ del intervalo (θ_1, θ_2) sujeto a la restricción de la ecuación 7.3. A continuación se desarrollan estimadores de parámetros empleados frecuentemente en la práctica. La elección de $\hat{\theta}$ se basa en que la que media muestral es un estimador insesgado de la media, y en muchos casos es el mejor estimador. Se asume además que n es grande.

Media

Como estimador puntual de la media, se usa la media muestral

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Para encontrar un estimador de intervalo, hay que determinar la distribución de \bar{x} , lo cual en general es un problema difícil. Para simplificarlo, se asumirá que \bar{x} es normal. Esto es cierto si \mathbf{x} es normal y es aproximadamente verdadero para cualquier \mathbf{x} si n es grande por el teorema central del límite (ver la sección 6.4.3).

Varianza conocida Supóngase en primer lugar que la varianza σ^2 de \mathbf{x} es conocida. La hipótesis de normalidad implica que el estimador puntual \bar{x} tiene distribución $N(\eta, \sigma^2/n)$. Dado $\gamma = 1 - \delta$, el estimador de intervalo cumple que

$$P\{x_{\delta/2} < \bar{x} < x_{1-\delta/2}\} = F(1 - \delta/2) - F(\delta/2) = 1 - \delta/2 - \delta/2 = 1 - \delta = \gamma, \quad (7.4)$$

donde $x_{\delta/2}$ y $x_{1-\delta/2}$ son respectivamente los percentiles $\delta/2$ y $1 - \delta/2$ de $\mathbf{x} \sim N(\eta, \sigma^2/n)$. Observar que como la distribución es simétrica, el intervalo es de la forma $\eta \pm a$. Normalizando, se observa que

$$P\left\{ \frac{x_{\delta/2} - \eta}{\sigma/\sqrt{n}} < \frac{\bar{x} - \eta}{\sigma/\sqrt{n}} < \frac{x_{1-\delta/2} - \eta}{\sigma/\sqrt{n}} \right\} = P\{z_{\delta/2} < z < z_{1-\delta/2}\} = 1 - \delta = \gamma,$$

donde z_u es el percentil u de la densidad normal estándar. De esta forma

$$\begin{aligned}\frac{x_{\delta/2} - \eta}{\sigma/\sqrt{n}} &= z_{\delta/2} \stackrel{(a)}{=} -z_{1-\delta/2} & \Rightarrow & & x_{\delta/2} &= \eta - z_{1-\delta/2} \frac{\sigma}{\sqrt{n}} \\ \frac{x_{1-\delta/2} - \eta}{\sigma/\sqrt{n}} &= z_{1-\delta/2} & \Rightarrow & & x_{1-\delta/2} &= \eta + z_{1-\delta/2} \frac{\sigma}{\sqrt{n}}\end{aligned}$$

donde en (a) se consideró que $z_u = -z_{1-u}$ por ser los percentiles de una variable aleatoria de media nula con densidad simétrica. Sustituyendo los resultados en la ecuación 7.4 se obtiene que

$$P\left\{\eta - z_{1-\delta/2} \frac{\sigma}{\sqrt{n}} < \bar{x} < \eta + z_{1-\delta/2} \frac{\sigma}{\sqrt{n}}\right\} = 1 - \delta = \gamma,$$

y por lo tanto,

$$P\left\{\bar{x} - z_{1-\delta/2} \frac{\sigma}{\sqrt{n}} < \eta < \bar{x} + z_{1-\delta/2} \frac{\sigma}{\sqrt{n}}\right\} = 1 - \delta = \gamma. \quad (7.5)$$

Esto indica que se puede afirmar con un coeficiente de confianza γ que la media η está en el intervalo $\bar{x} \pm z_{1-\delta/2} \sigma / \sqrt{n}$. La determinación del intervalo de confianza para la media se realiza de la siguiente forma: se observan las muestras x_i de \mathbf{x} y se calcula el promedio \bar{x} . Se elige el número $\gamma = 1 - \delta$ y se obtiene el percentil normal estándar z_u para $u = 1 - \delta/2$. El intervalo es $\bar{x} \pm z_{1-\delta/2} \sigma / \sqrt{n}$.

Supóngase ahora que no se conoce la distribución de \mathbf{x} . Para encontrar el intervalo de confianza de la media puede emplearse la desigualdad de Chebyshev, dada por la ecuación 4.22, la cual indica que

$$P\{x - \epsilon < \eta < x + \epsilon\} > 1 - \frac{\sigma^2}{\epsilon^2}$$

Reemplazando x por \bar{x} y σ por σ/\sqrt{n} y tomando $\epsilon = \sigma/\sqrt{n\delta}$, se obtiene que

$$P\left\{\bar{x} - \frac{\sigma}{\sqrt{n\delta}} < \eta < \bar{x} + \frac{\sigma}{\sqrt{n\delta}}\right\} > 1 - \delta = \gamma. \quad (7.6)$$

El resultado indica que el intervalo de confianza γ de η está contenido en el intervalo $\bar{x} \pm \sigma/\sqrt{n\delta}$.

Varianza desconocida Si la varianza es desconocida no puede emplearse la ecuación 7.5. Para estimar el intervalo de confianza de la media η se considera la varianza muestral (ecuación 6.10)

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

la cual es un estimador insesgado de σ^2 , como se mostró en la sección 6.1.3, y tiende a σ^2 con $n \rightarrow \infty$. Por lo tanto, si n es grande, se puede emplear la aproximación $\sigma \approx s$ en la ecuación 7.5, lo que conduce al intervalo de confianza aproximado

$$\bar{x} - z_{1-\delta/2} \frac{s}{\sqrt{n}} < \eta < \bar{x} + z_{1-\delta/2} \frac{s}{\sqrt{n}}.$$

Es posible encontrar un intervalo de confianza exacto para el caso en que \mathbf{x} es normal. Bajo esa hipótesis, el cociente

$$\frac{\bar{x} - \eta}{s/\sqrt{n}}$$

tiene distribución t Student con $n - 1$ grados de libertad. Efectivamente, notando que dicho cociente se puede escribir como

$$\frac{\bar{x} - \eta}{s/\sqrt{n}} = \frac{\frac{\bar{x} - \eta}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s^2/\sigma^2}{n-1}}} = \frac{\mathbf{w}}{\sqrt{\mathbf{y}/(n-1)}},$$

con

$$\mathbf{w} = \frac{\bar{\mathbf{x}} - \eta}{\sigma/\sqrt{n}} \stackrel{(a)}{\sim} N(0, 1) \quad \text{y} \quad \mathbf{y} = \frac{(n-1)\mathbf{s}^2}{\sigma^2} \stackrel{(b)}{\sim} \chi^2(n-1),$$

donde en (a) se tuvo en cuenta que $\bar{\mathbf{x}} \sim N(\eta, \sigma^2/n)$ y en (b) se empleó el resultado de la ecuación 6.45, se concluye por el ejemplo 6-28 de la sección 5.3.1 que el cociente tiene distribución $t(n-1)$. Denotando por t_u su percentil u se concluye que

$$P \left\{ -t_u < \frac{\bar{\mathbf{x}} - \eta}{\mathbf{s}/\sqrt{n}} < t_u \right\} = u - (1 - u) = 2u - 1 = \gamma.$$

Con $u = 1 - \delta/2$, esto conduce al intervalo

$$\bar{x} - t_{1-\delta/2} \frac{s}{\sqrt{n}} < \eta < \bar{x} + t_{1-\delta/2} \frac{s}{\sqrt{n}}.$$

En el siguiente ejemplo, la distribución de \mathbf{x} es especificada en términos de un único parámetro, y por lo tanto, no puede emplearse la ecuación 7.6 para obtener el intervalo de confianza debido a que las constantes η y σ están relacionadas.

Distribución de Poisson Considérese el caso en que la variable aleatoria \mathbf{x} tiene distribución de Poisson con parámetro λ , es decir,

$$P\{\mathbf{x} = k\} = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots.$$

En este caso, como se mostró en la sección 4.3.6, $\eta = \lambda$ y $\sigma^2 = \lambda$, y por lo tanto, si n es grande, la distribución de $\bar{\mathbf{x}}$ es aproximadamente $N(\lambda, \lambda/n)$. De esta forma, de la ecuación 7.5, se obtiene que

$$P \left\{ |\bar{\mathbf{x}} - \lambda| < z_u \sqrt{\frac{\lambda}{n}} \right\} = \gamma.$$

Los puntos del plano $\bar{x}\lambda$ que satisfacen la desigualdad $|\bar{x} - \lambda| < z_u \sqrt{\lambda/n}$ se encuentran en el interior de la parábola

$$(\lambda - \bar{x})^2 = \frac{z_u^2}{n} \lambda, \quad (7.7)$$

como se observa en la figura 7.3. El intervalo de confianza γ de λ es el segmento vertical (λ_1, λ_2) mostrado en la figura, donde λ_1 y λ_2 son las raíces de la forma cuadrática de la ecuación 7.7 dado el valor de \bar{x} .

Varianza

Se quiere estimar la varianza $v = \sigma^2$ de una variable aleatoria normal \mathbf{x} a partir de n muestras x_i de \mathbf{x} .

Media conocida Se asume primero que la media η de \mathbf{x} es conocida, por lo que se emplea como estimador puntual de v el promedio

$$\hat{\mathbf{v}} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \eta)^2.$$

De esta forma, la media es

$$E\{\hat{\mathbf{v}}\} = \frac{1}{n} \sum_{i=1}^n E\{(\mathbf{x}_i - \eta)^2\} = \frac{1}{n} n \sigma^2 = \sigma^2,$$

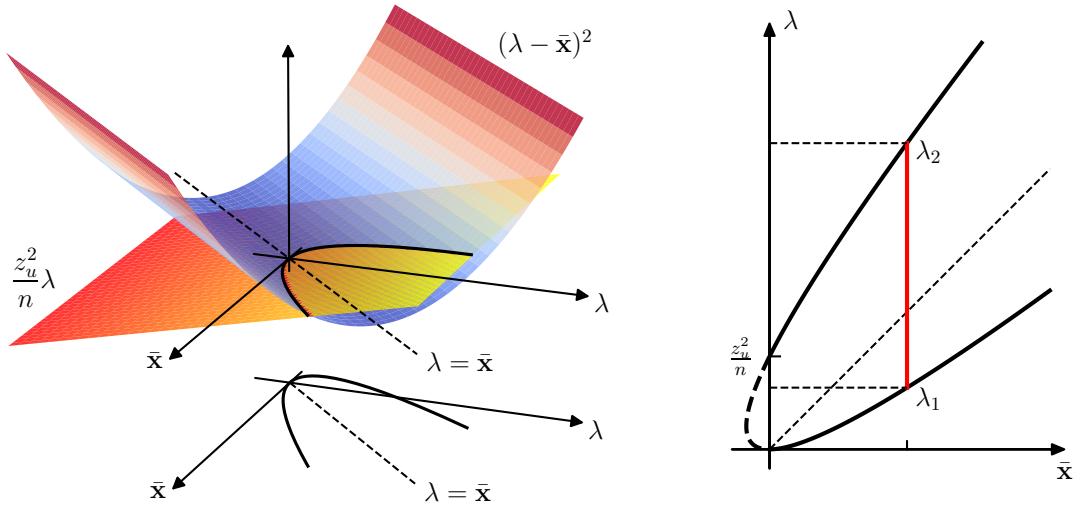


Figura 7.3: Intervalo de confianza del parámetro λ de una variable aleatoria de Poisson. Los puntos del plano $\bar{x}\lambda$ que cumplen $(\lambda - \bar{x})^2 < z_u^2 \lambda/n$ se encuentran en el interior de la parábola determinada por la intersección de la forma cuadrática $(\lambda - \bar{x})^2$ y el plano $z_u^2 \lambda/n$. La forma cuadrática se anula en la recta $\lambda = \bar{x}$ y el plano pasa por el eje \bar{x} . La parábola está dada por la ecuación 7.7.

y la varianza es

$$\begin{aligned}
 \sigma_{\hat{v}}^2 = \text{var}\{\hat{v}\} &\stackrel{(a)}{=} \frac{1}{n^2} \sum_{i=1}^n \text{var}\{(\mathbf{x}_i - \eta)^2\} \\
 &= \frac{1}{n} \text{var}\{(\mathbf{x}_i - \eta)^2\} \\
 &= \frac{1}{n} (E\{(\mathbf{x}_i - \eta)^4\} - E^2\{(\mathbf{x}_i - \eta)^2\}) \\
 &\stackrel{(b)}{=} \frac{1}{n} (3\sigma^4 - \sigma^2) \\
 &= \frac{2\sigma^4}{n},
 \end{aligned}$$

donde en (a) se consideró que las muestras son independientes y en (b) que el momento central cuarto de una variable aleatoria normal es $3\sigma^2$, como indica la ecuación 4.19. Se cumple por lo tanto que

$$E\{\hat{v}\} = v \quad \text{y} \quad \sigma_{\hat{v}}^2 = \frac{2\sigma^4}{n} \xrightarrow{n \rightarrow \infty} 0,$$

lo que implica que \hat{v} es un estimador consistente de σ^2 . Se buscará un estimador de intervalo. Considérese la variable aleatoria

$$\frac{n\hat{v}}{\sigma^2} = \sum_{i=1}^n \frac{(x_i - \eta)^2}{\sigma^2}.$$

Por tratarse de la suma de n variables aleatorias $(x_i - \eta)/\sigma \sim N(0, 1)$ al cuadrado, se cumple que $n\hat{v}/\sigma^2 \sim \chi^2(n)$, como se mostró en la sección 6.2.3. Debido a que la densidad $\chi^2(n)$ no es simétrica, el estimador de intervalo de σ^2 no está centrado en σ^2 (ver la figura 7.4). Para determinarlo, se definen las constantes c_1 y c_2 tal que

$$P \left\{ \frac{n\hat{v}}{\sigma^2} < c_1 \right\} = \frac{\delta}{2} \quad \text{y} \quad P \left\{ \frac{n\hat{v}}{\sigma^2} > c_2 \right\} = \frac{\delta}{2}.$$

De esta forma,

$$c_1 = \chi_{\delta/2}^2(n) \quad \text{y} \quad c_2 = \chi_{1-\delta/2}^2(n)$$

y

$$P \left\{ \chi_{\delta/2}^2(n) < \frac{n\hat{\mathbf{v}}}{\sigma^2} < \chi_{1-\delta/2}^2(n) \right\} = \gamma = 1 - \delta,$$

resultando en

$$P \left\{ \frac{n\hat{\mathbf{v}}}{\chi_{1-\delta/2}^2(n)} < \sigma^2 < \frac{n\hat{\mathbf{v}}}{\chi_{\delta/2}^2(n)} \right\} = \gamma = 1 - \delta.$$

Se concluye que la estimación de intervalo de la varianza es

$$\frac{n\hat{\mathbf{v}}}{\chi_{1-\delta/2}^2(n)} < \sigma^2 < \frac{n\hat{\mathbf{v}}}{\chi_{\delta/2}^2(n)}.$$

Este intervalo no tiene largo mínimo. El intervalo de largo mínimo es aquel que cumple que $f(c_1) = f(c_2)$, donde $f(x)$ es la densidad de probabilidad, como se demostró en la sección 7.1, pero no es fácil de determinar analíticamente. En la figura 7.4 se comparan ambos casos.

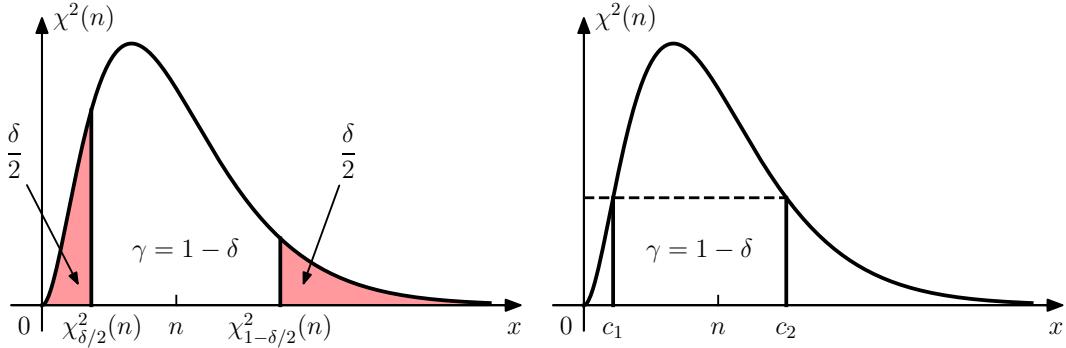


Figura 7.4: Estimación de intervalo de la varianza con la media conocida. Se muestra la estimación de intervalo de la variable aleatoria $n\bar{\mathbf{v}}/\sigma^2 \sim \chi^2(n)$ con un coeficiente de confianza $\gamma = 1 - \delta$. Se compara el intervalo con áreas de cola iguales con el intervalo de largo mínimo. Los parámetros empleados en la figura son $n = 6$ y $\gamma = 0.8$. Con estos valores, el intervalo de áreas de cola iguales es $\chi_{1-\delta/2}^2(n) - \chi_{\delta/2}^2(n) \approx 8.44$ mientras que el intervalo de largo mínimo es $c_2 - c_1 \approx 7.73$. Se observa además que el intervalo no está centrado en la media n (ver la sección 4.5.4).

Media desconocida Si además la media η es desconocida, se emplea como estimador puntual de σ^2 a la varianza muestral \mathbf{s}^2 (ver la ecuación 6.10). Considerando que la variable aleatoria $(n-1)\mathbf{s}^2/\sigma^2 \sim \chi^2(n-1)$, se cumple que

$$P \left\{ \chi_{\delta/2}^2(n-1) < \frac{(n-1)\mathbf{s}^2}{\sigma^2} < \chi_{1-\delta/2}^2(n-1) \right\} = \gamma = 1 - \delta,$$

lo que conduce al intervalo

$$\frac{(n-1)\mathbf{s}^2}{\chi_{1-\delta/2}^2(n-1)} < \sigma^2 < \frac{(n-1)\mathbf{s}^2}{\chi_{\delta/2}^2(n-1)}.$$

Percentiles

El percentil u de una variable aleatoria \mathbf{x} es por definición un número x_u tal que $F(x_u) = u$. Por lo tanto, x_u es la función inversa $F^{-1}(u)$ de la distribución $F(x)$ de x . Se desea estimar x_u a partir de las muestras x_i de \mathbf{x} . Para hacerlo, se ordenan las n observaciones x_i de forma ascendente, y se denota

por y_k la muestra en la posición k -ésima. Las variables aleatorias y_k son los estadísticos de orden de la variable aleatoria \mathbf{x} (ver la ecuación 6.6).

De la definición, $y_k < x_u$ si la menos k muestras x_i son menores que x_u . De forma similar, $y_{k+r} < x_u$ si al menos $k+r$ muestras son menores que x_u , o $y_{k+r} > x_u$ si a lo sumo $k+r-1$ muestras son menores que x_u . Por lo tanto, $y_k < x_u < y_{k+r}$ si al menos k muestras y a lo sumo $k+r-1$ muestras son menores que x_u . Se concluye que ocurre el evento $\{y_k < x_u < y_{k+r}\}$ si la cantidad de ocurrencias del evento $\{\mathbf{x} < x_u\}$ en n repeticiones del experimento S es al menos k y a lo sumo $k+r-1$. Como $P\{\mathbf{x} < x_u\} = u$, de la ecuación 2.5 con $p = u$ se tiene que

$$P\{\mathbf{y}_k < x_u < \mathbf{y}_{k+r}\} = \sum_{m=k}^{k+r-1} \binom{n}{m} u^m (1-u)^{n-m}.$$

Empleando esta relación se buscará el intervalo de confianza γ de x_u para un valor de u específico. Para eso, hay que encontrar un entero k tal que la sumatoria iguale γ para el valor de r mas pequeño posible. Este es un problema difícil que se resuelve mediante ensayo y error. Sin embargo, puede encontrarse una solución simple asumiendo que n es grande empleando la aproximación normal dada por la ecuación 3.38 con $p = u$,

$$P\{\mathbf{y}_k < x_u < \mathbf{y}_{k+r}\} \approx G\left(\frac{k+r+0.5-nu}{\sqrt{nu(1-u)}}\right) - G\left(\frac{k-0.5-nu}{\sqrt{nu(1-u)}}\right) = \gamma.$$

Para un valor dado de γ , r es mínimo si nu está cercano al centro del intervalo $(k, k+r)$. Esto conduce a

$$k \approx nu - z_{1-\delta/2} \sqrt{nu(1-u)} \quad \text{y} \quad k+r \approx nu + z_{1-\delta/2} \sqrt{nu(1-u)}.$$

7.3. Estimación de parámetros

Se considera el caso en que $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ denota n variables aleatorias que representan las observaciones $\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \dots, \mathbf{x}_n = x_n$, y supóngase que se desea estimar el parámetro θ a partir de esos datos. Concretamente, la densidad de probabilidad conjunta de $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ dada por $f_x(x_1, x_2, \dots, x_n; \theta)$ depende del parámetro θ que se quiere estimar, y se construye un estimador como una función de las observaciones. El problema de estimación puntual consiste en elegir un estadístico unidimensional $T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ que mejor estime θ en algún sentido óptimo. El método de máxima verosimilitud es usualmente empleado en este contexto.

7.3.1. Método de máxima verosimilitud

En el método de máxima verosimilitud se elige el valor de θ mas probable de haber ocasionado las observaciones. Una vez que se dispone de las observaciones x_1, x_2, \dots, x_n , $f_x(x_1, x_2, \dots, x_n; \theta)$ es una función únicamente de θ . El valor que maximiza dicha p.d.f. es el valor mas probable de θ , y es el escogido como estimador $\hat{\theta}_{\text{ML}}$ de máxima verosimilitud (ML, *maximum likelihood*).

Dadas las observaciones $\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \dots, \mathbf{x}_n = x_n$, la p.d.f. conjunta $f_x(x_1, x_2, \dots, x_n; \theta)$ se define como la *función de verosimilitud*, y el estimador ML puede determinarse tanto a partir de la función de verosimilitud como

$$\hat{\theta}_{\text{ML}} = \sup_{\theta} f_x(x_1, x_2, \dots, x_n; \theta), \quad (7.8)$$

o a partir de la función de verosimilitud logarítmica

$$L(x_1, x_2, \dots, x_n; \theta) \triangleq \log f_x(x_1, x_2, \dots, x_n; \theta).$$

Si $L(x_1, x_2, \dots, x_n; \theta)$ es diferenciable y existe un supremo $\hat{\theta}_{\text{ML}}$, debe satisfacer la ecuación

$$\frac{\partial \log f_x(x_1, x_2, \dots, x_n; \theta)}{\partial \theta} \bigg|_{\theta=\hat{\theta}_{\text{ML}}} = 0. \quad (7.9)$$

Ejemplo 8-10 Sean $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ variables aleatorias uniformemente distribuidas en el intervalo $(0, \theta)$, es decir,

$$f_{x_i}(x_i, \theta) = \frac{1}{\theta}, \quad 0 < x_i < \theta,$$

donde θ es desconocido. En este caso, la función de verosimilitud está dada por

$$\begin{aligned} f(\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \dots, \mathbf{x}_n = x_n; \theta) &= \frac{1}{\theta^n}, \quad 0 < x_i < \theta, \quad i = 1, \dots, n, \\ &= \frac{1}{\theta^n}, \quad 0 < \min(\mathbf{x}_1, \dots, \mathbf{x}_n) < \max(\mathbf{x}_1, \dots, \mathbf{x}_n) < \theta. \end{aligned}$$

La función de verosimilitud es maximizada por el mínimo valor posible de θ , y como $\theta > \max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, se obtiene que

$$\hat{\theta}_{\text{ML}}(\mathbf{x}) = \max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$$

es el estimador ML para θ .

En general, la función de verosimilitud puede tener mas de una solución o no tener soluciones. Además, puede no ser diferenciable o difícil de resolver analíticamente. Como indica la ecuación 7.8, el método de máxima verosimilitud busca la moda de la p.d.f. $f_x(x_1, x_2, \dots, x_n; \theta)$. Como en general la moda es un peor estimador que la media o la mediana, el estimador ML suele ser pobre cuando se emplean pocas muestras.

7.3.2. Estadísticos suficientes

Se dice que una función $T(\mathbf{x}) = T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ es un *estadístico suficiente* para θ si $T(\mathbf{x})$ brinda la misma información sobre θ que está contenida en el conjunto de datos. De esta forma, dada la densidad de probabilidad

$$P\{\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \dots, \mathbf{x}_n = x_n; \theta\},$$

si la densidad condicional

$$P\{\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \dots, \mathbf{x}_n = x_n; \theta \mid T(\mathbf{x}) = t\},$$

no depende de θ para todos los valores posibles de x_i , $T(\mathbf{x})$ debe contener toda la información sobre θ que está contenida en el conjunto de datos \mathbf{x} . Una vez que se conoce $T(\mathbf{x})$, el conjunto de datos \mathbf{x} no contiene información adicional sobre θ , y por lo tanto, es un estadístico suficiente para θ . En la figura

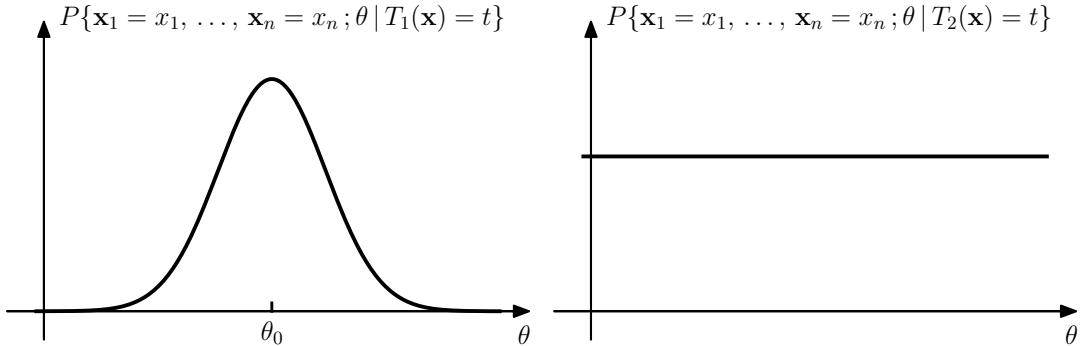


Figura 7.5: Definición de estadístico suficiente. El estadístico $T_1(\mathbf{x})$ no es suficiente, ya que luego de observar $T_1(\mathbf{x})$, si $\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \dots, \mathbf{x}_n = x_n$, los valores de θ cercanos a θ_0 son más probables. En la densidad condicional al estadístico $T_2(\mathbf{x})$, todos los valores de θ son equiprobables, y por lo tanto, luego de observar $T_2(\mathbf{x})$, los datos no brindan información adicional sobre θ .

7.5 se muestra la densidad condicional en el caso de un estadístico que no es suficiente y otro que es suficiente.

Observar que el conjunto de datos $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, es siempre trivialmente suficiente, pero si una función $T(\mathbf{x})$ es suficiente, alcanza con concentrarse solo en ella, ya que agota toda la información sobre θ que contiene el conjunto de datos.

Ejemplo 8-12 Sean las variables aleatorias $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ i.i.d. con distribución de Poisson con parámetro λ , y se considera la función

$$T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \sum_{i=1}^n x_i.$$

De esta forma,

$$\begin{aligned} & P\{\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \dots, \mathbf{x}_n = x_n \mid T(\mathbf{x}) = t\} \\ & \stackrel{(a)}{=} \frac{P\{\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \dots, \mathbf{x}_n = x_n, T(\mathbf{x}) = t\}}{P\{T(\mathbf{x}) = t\}} \\ & = \begin{cases} \frac{P\{\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \dots, \mathbf{x}_n = t - (x_1 + \dots + x_{n-1})\}}{P\left\{T = \sum_{i=1}^n x_i\right\}} & \text{si } t = \sum_{i=1}^n x_i \\ 0 & \text{si } t \neq \sum_{i=1}^n x_i, \end{cases} \end{aligned}$$

donde en (a) se empleó la ecuación 5.76. Como $T(\mathbf{x}) \sim P(n\lambda)$, como indican las ecuaciones 5.30 y 5.64, se tiene que (ver la ecuación 3.23)

$$P\left\{T = \sum_{i=1}^n x_i\right\} = e^{-n\lambda} \frac{(n\lambda)^{\sum_{i=1}^n x_i}}{\left(\sum_{i=1}^n x_i\right)!} = e^{-n\lambda} \frac{(n\lambda)^t}{t!}.$$

Por lo tanto

$$\begin{aligned} & \frac{P\{\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \dots, \mathbf{x}_n = t - (x_1 + \dots + x_{n-1})\}}{P\left\{T = \sum_{i=1}^n x_i\right\}} \\ & = \frac{e^{-\lambda} \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \frac{\lambda^{x_2}}{x_2!} \dots e^{-\lambda} \frac{\lambda^{t-x_1-x_2-\dots-x_{n-1}}}{(t-x_1-x_2-\dots-x_{n-1})!}}{e^{-n\lambda} \frac{(n\lambda)^t}{t!}} \\ & = \frac{e^{-n\lambda} \frac{\lambda^t}{x_1! x_2! \dots x_n!}}{e^{-n\lambda} \frac{(n\lambda)^t}{t!}} \\ & = \frac{\left(\sum_{i=1}^n x_i\right)!}{x_1! x_2! \dots x_n! n^t}. \end{aligned}$$

Se concluye que $P\{\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \dots, \mathbf{x}_n = x_n \mid T(\mathbf{x}) = t\}$ es independiente de $\theta = \lambda$, y por lo tanto, $T(\mathbf{x}) = \sum_{i=1}^n x_i$ es suficiente para λ .

El siguiente teorema permite determinar si un estadístico es suficiente.

El teorema de factorización de Fisher-Neyman El teorema de factorización de Fisher-Neyman provee una caracterización conveniente de un estadístico suficiente. Si la p.d.f. de los datos es $f_x(x_1, \dots, x_n; \theta)$, $T(\mathbf{x})$ es suficiente para θ si y solo si existen las funciones g y h no negativas tal que

$$f_x(x_1, \dots, x_n; \theta) = h(x_1, \dots, x_n) g_\theta(T(\mathbf{x}) = t),$$

es decir, la p.d.f. puede ser factorizada en un producto donde un factor no depende de θ y otro factor depende de θ y de los datos, pero solo a través de $T(\mathbf{x})$. A continuación se incluye una prueba simple pero ilustrativa para el caso de variables aleatorias discretas. Una prueba más general puede encontrarse en [19].

Sean $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ un conjunto de variables aleatorias discretas con p.m.f. $P\{\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \dots, \mathbf{x}_n = x_n; \theta\}$. Entonces, $T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ es suficiente para θ si y solo si

$$P\{\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \dots, \mathbf{x}_n = x_n; \theta\} = h(x_1, x_2, \dots, x_n)g_\theta(T(\mathbf{x}) = t),$$

donde h es una función no negativa de $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ que no depende de θ , y g_θ es una función no negativa de θ y de los datos solo a través de T y no de alguna otra forma.

- *Suficiencia:* supóngase que

$$P\{\mathbf{x}; \theta\} = h(\mathbf{x})g_\theta(T(\mathbf{x})).$$

De esta forma,

$$\begin{aligned} P\{\mathbf{x} = x; \theta \mid T(\mathbf{x}) = t\} &\stackrel{(a)}{=} \begin{cases} \frac{P\{\mathbf{x} = x; \theta, T(\mathbf{x}) = t\}}{P\{T(\mathbf{x}) = t\}} & T(\mathbf{x}) = t \\ 0 & T(\mathbf{x}) \neq t \end{cases} \\ &\stackrel{(b)}{=} \begin{cases} \frac{P\{\mathbf{x} = x; \theta\}}{P\{T(\mathbf{x}) = t\}} & T(\mathbf{x}) = t \\ 0 & T(\mathbf{x}) \neq t, \end{cases} \end{aligned}$$

donde en (a) se empleó la generalización de la ecuación 5.76 y en (b) se consideró que como $T(\mathbf{x})$ es una función de \mathbf{x} , $P\{\mathbf{x} = x; \theta, T(\mathbf{x}) = t\} = P\{\mathbf{x} = x; \theta\}$ siempre y cuando $T(\mathbf{x}) = t$ y cero en otro caso. Si $T(\mathbf{x}) = t$, el denominador es

$$P\{T(\mathbf{x}) = t\} = \sum_{T(\mathbf{x})=t} P\{\mathbf{x} = x; \theta\} = \sum_{T(\mathbf{x})=t} h(\mathbf{x})g_\theta(T(\mathbf{x}) = t) = g_\theta(T(\mathbf{x}) = t) \sum_{T(\mathbf{x})=t} h(\mathbf{x}).$$

Se concluye que

$$P\{\mathbf{x} = x; \theta \mid T(\mathbf{x}) = t\} = \begin{cases} \frac{h(\mathbf{x})}{\sum_{T(\mathbf{x})=t} h(\mathbf{x})} & T(\mathbf{x}) = t \\ 0 & T(\mathbf{x}) \neq t, \end{cases}$$

que no depende de θ , y por lo tanto, T es suficiente.

- *Necesariedad:* supóngase que $T(\mathbf{x}) = t$ es suficiente. Por lo tanto, si $T(\mathbf{x}) = t$, por los mismos argumentos de antes, se cumple que

$$\begin{aligned} P\{\mathbf{x} = x; \theta\} &= P\{\mathbf{x} = x; \theta, T(\mathbf{x}) = t\} \\ &= P\{\mathbf{x} = x; \theta \mid T(\mathbf{x}) = t\}P\{T(\mathbf{x}) = t\}. \end{aligned}$$

Como T es suficiente, $P\{\mathbf{x} = x; \theta \mid T(\mathbf{x}) = t\}$ es independiente de θ , digamos $h(\mathbf{x})$. Así,

$$P\{\mathbf{x} = x; \theta\} = h(\mathbf{x})P\{T(\mathbf{x}) = t\}.$$

Además,

$$P\{T(\mathbf{x}) = t\} = \sum_{T(\mathbf{x})=t} P\{\mathbf{x} = x; \theta\} = g_\theta(T(\mathbf{x}) = t).$$

Se concluye que si T es suficiente, se cumple que

$$P\{T(\mathbf{x}) = t\} = h(\mathbf{x})g_\theta(T(\mathbf{x}) = t).$$

Ejemplo 8-14 Sean $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ variables aleatorias i.i.d. con distribución $N(\mu, \sigma^2)$. La p.d.f. conjunta es

$$f(\mathbf{x}; \mu, \sigma^2) = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} e^{-\sum_{i=1}^n (x_i - \mu)^2 / 2\sigma^2}$$

y se puede expresar como

$$f(\mathbf{x}; \mu, \sigma^2) = \underbrace{1}_{h(\mathbf{x})} \times \underbrace{\left(\frac{1}{2\pi\sigma^2} \right)^{n/2} e^{-\left(\sum_{i=1}^n x_i^2 / 2\sigma^2 - \mu \sum_{i=1}^n x_i / \sigma^2 + n\mu / 2\sigma^2 \right)}}_{g_\theta(T(\mathbf{x}))}.$$

Por lo tanto, el conjunto

$$\left(\sum_{i=1}^n \mathbf{x}_i, \sum_{i=1}^n \mathbf{x}_i^2 \right)$$

es suficiente para el par de parámetros (μ, σ^2) . Observar que si σ^2 es conocido, $\sum_{i=1}^n \mathbf{x}_i$ es suficiente para μ y si μ es conocido, $\sum_{i=1}^n \mathbf{x}_i^2$ es suficiente para σ^2 .

Ejemplo 8-15 Sean $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ variables aleatorias i.i.d. con distribución $U(0, \theta)$. La p.d.f. conjunta es

$$f_x(x_1, x_2, \dots, x_n; \theta) = \begin{cases} \frac{1}{\theta^n} & 0 < x_1, x_2, \dots, x_n < \theta \\ 0 & \text{en otro caso.} \end{cases}$$

Se define la función

$$I_{(a, b)} = \begin{cases} 1 & b \geq a \\ 0 & \text{en otro caso,} \end{cases}$$

y las variables aleatorias

$$\begin{aligned} \max \mathbf{x} &= \max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \\ \min \mathbf{x} &= \min(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n). \end{aligned}$$

De esta forma

$$f_x(x_1, x_2, \dots, x_n; \theta) = \frac{1}{\theta^n} I_{(0, \min \mathbf{x})} I_{(\max \mathbf{x}, \theta)} = h(\mathbf{x}) g_\theta(T),$$

donde

$$h(\mathbf{x}) = I_{(0, \min \mathbf{x})} \quad \text{y} \quad g_\theta(T) = \frac{1}{\theta^n} I_{(\max \mathbf{x}, \theta)}.$$

Por lo tanto,

$$T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$$

es suficiente para θ para variables aleatorias uniformes.

De forma similar, si \mathbf{x}_i representa una variable aleatoria uniforme con

$$P\{\mathbf{x}_i = k\} = \frac{1}{N}, \quad k = 1, 2, \dots, N,$$

y $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ son i.i.d., la p.m.f. de los datos es

$$\begin{aligned} P\{\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \dots, \mathbf{x}_n = x_n; N\} &= \begin{cases} \frac{1}{N^n} & 1 \leq x_1, x_2, \dots, x_n \leq N \\ 0 & \text{en otro caso.} \end{cases} \\ &= \frac{1}{N^n} I_{(1, \min \mathbf{x})} I_{(\max \mathbf{x}, N)} \end{aligned}$$

y por lo tanto,

$$T(\mathbf{x}) = \max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n), \tag{7.10}$$

es suficiente para el parámetro N discreto.

7.3.3. Estimadores insesgados

Un estimador $T(\mathbf{x})$ se dice insesgado para θ si

$$E\{T(\mathbf{x})\} = \theta.$$

Si los estimadores $T_1(\mathbf{x})$ y $T_2(\mathbf{x})$ son ambos insesgados para θ , es preferible el que tiene menor varianza. Por lo tanto, sería útil encontrar el estimador insesgado con la menor varianza posible. La cota inferior de Cramér-Rao provee una expresión de la cota inferior de la varianza de todos los estimadores insesgados.

Cota inferior de Cramér-Rao Sea $T(\mathbf{x})$ cualquier estimador insesgado para el parámetro desconocido θ basado en las observaciones de las variables aleatorias $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ con densidad de probabilidad conjunta $f(\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \dots, \mathbf{x}_n = x_n; \theta)$, denotada como $f(\mathbf{x}; \theta)$. Entonces, se cumple que

$$\text{var}\{T(\mathbf{x})\} \geq \frac{1}{E\left\{\left[\frac{\partial}{\partial\theta} \log f(\mathbf{x}; \theta)\right]^2\right\}} = \frac{1}{E\left\{\frac{\partial^2}{\partial\theta^2} \log f(\mathbf{x}; \theta)\right\}}$$

sujeto a que se satisfagan las siguientes condiciones de regularidad:

$$\frac{\partial}{\partial\theta} \int f(\mathbf{x}; \theta) dx = \int \frac{\partial f(\mathbf{x}; \theta)}{\partial\theta} dx = 0 \quad (7.11)$$

$$\frac{\partial}{\partial\theta} \int T(\mathbf{x})f(\mathbf{x}; \theta) dx = \int T(\mathbf{x}) \frac{\partial f(\mathbf{x}; \theta)}{\partial\theta} dx = 0, \quad (7.12)$$

donde las integrales representan integrales de multiplicidad n .

Para la demostración, se parte de la propiedad de insesgado del estimador $T(\mathbf{x})$,

$$E\{T(\mathbf{x}) - \theta\} = \int_{-\infty}^{\infty} (T(\mathbf{x}) - \theta)f(\mathbf{x}; \theta) dx = 0.$$

Diferenciando respecto a θ se obtiene que

$$\int_{-\infty}^{\infty} (T(\mathbf{x}) - \theta) \frac{\partial f(\mathbf{x}; \theta)}{\partial\theta} dx - \int_{-\infty}^{\infty} f(\mathbf{x}; \theta) dx = 0,$$

resultando en

$$\int_{-\infty}^{\infty} (T(\mathbf{x}) - \theta) \frac{\partial f(\mathbf{x}; \theta)}{\partial\theta} dx = 1, \quad (7.13)$$

en donde se emplearon las condiciones de regularidad dadas por las ecuaciones 7.11 y 7.12 para intercambiar el orden de diferenciancia y integración. Considerando que

$$\frac{\partial \log f(\mathbf{x}; \theta)}{\partial\theta} = \frac{1}{f(\mathbf{x}; \theta)} \frac{\partial f(\mathbf{x}; \theta)}{\partial\theta}$$

y por lo tanto

$$\frac{\partial f(\mathbf{x}; \theta)}{\partial\theta} = f(\mathbf{x}; \theta) \frac{\partial \log f(\mathbf{x}; \theta)}{\partial\theta}, \quad (7.14)$$

se obtiene que

$$\int_{-\infty}^{\infty} (T(\mathbf{x}) - \theta) f(\mathbf{x}; \theta) \frac{\partial \log f(\mathbf{x}; \theta)}{\partial\theta} dx = 1. \quad (7.15)$$

Esta expresión se puede escribir como

$$\int_{-\infty}^{\infty} \left[(T(\mathbf{x}) - \theta) \sqrt{f(\mathbf{x}; \theta)} \right] \left[\sqrt{f(\mathbf{x}; \theta)} \frac{\partial \log f(\mathbf{x}; \theta)}{\partial\theta} \right] dx = 1,$$

Empleando la desigualdad de Cauchy-Schwarz, que indica que²

$$\left| \int_{\mathbb{R}^n} f(x) \overline{g(x)} dx \right|^2 \leq \int_{\mathbb{R}^n} |f(x)|^2 dx \int_{\mathbb{R}^n} |g(x)|^2 dx, \quad (7.16)$$

²Ver https://en.wikipedia.org/wiki/Cauchy-Schwarz_inequality, por ejemplo.

resulta en

$$\int_{-\infty}^{\infty} \left[(T(\mathbf{x}) - \theta) \sqrt{f(\mathbf{x}; \theta)} \right]^2 dx \int_{-\infty}^{\infty} \left[\sqrt{f(\mathbf{x}; \theta)} \frac{\partial \log f(\mathbf{x}; \theta)}{\partial \theta} \right]^2 dx \geq 1, \quad (7.17)$$

o

$$\int_{-\infty}^{\infty} (T(\mathbf{x}) - \theta)^2 f(\mathbf{x}; \theta) dx \int_{-\infty}^{\infty} \left(\frac{\partial \log f(\mathbf{x}; \theta)}{\partial \theta} \right)^2 f(\mathbf{x}; \theta) dx \geq 1.$$

Teniendo en cuenta que

$$\int_{-\infty}^{\infty} (T(\mathbf{x}) - \theta)^2 f(\mathbf{x}; \theta) dx = \text{var}\{T(\mathbf{x})\}$$

y

$$\int_{-\infty}^{\infty} \left(\frac{\partial \log f(\mathbf{x}; \theta)}{\partial \theta} \right)^2 f(\mathbf{x}; \theta) dx = E \left\{ \left(\frac{\partial \log f(\mathbf{x}; \theta)}{\partial \theta} \right)^2 \right\}$$

resulta en

$$\text{var}\{T(\mathbf{x})\} E \left\{ \left(\frac{\partial \log f(\mathbf{x}; \theta)}{\partial \theta} \right)^2 \right\} \geq 1,$$

obteniendo la cota buscada

$$\text{var}\{T(\mathbf{x})\} \geq \frac{1}{E \left\{ \left(\frac{\partial \log f(\mathbf{x}; \theta)}{\partial \theta} \right)^2 \right\}}.$$

Además, integrando la ecuación 7.14, se obtiene que

$$\int_{-\infty}^{\infty} f(\mathbf{x}; \theta) \frac{\partial \log f(\mathbf{x}; \theta)}{\partial \theta} dx = \int_{-\infty}^{\infty} \frac{\partial f(\mathbf{x}; \theta)}{\partial \theta} dx \stackrel{(a)}{=} \frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} f(\mathbf{x}; \theta) dx = \frac{\partial 1}{\partial \theta} = 0, \quad (7.18)$$

donde en (a) se empleó la condición de regularidad dada por la ecuación 7.11. Diferenciando respecto a θ resulta en

$$\int_{-\infty}^{\infty} \frac{\partial f(\mathbf{x}; \theta)}{\partial \theta} \frac{\partial \log f(\mathbf{x}; \theta)}{\partial \theta} dx + \int_{-\infty}^{\infty} f(\mathbf{x}; \theta) \frac{\partial^2 \log f(\mathbf{x}; \theta)}{\partial \theta^2} dx = 0,$$

donde se asumió que $\partial f(\mathbf{x}; \theta)/\partial \theta$ es nuevamente diferenciable. Sustituyendo la ecuación 7.14 en esta expresión se obtiene que

$$\int_{-\infty}^{\infty} f(\mathbf{x}; \theta) \left(\frac{\partial \log f(\mathbf{x}; \theta)}{\partial \theta} \right)^2 dx + \int_{-\infty}^{\infty} f(\mathbf{x}; \theta) \frac{\partial^2 \log f(\mathbf{x}; \theta)}{\partial \theta^2} dx = 0,$$

es decir,

$$E \left\{ \left(\frac{\partial \log f(\mathbf{x}; \theta)}{\partial \theta} \right)^2 \right\} = -E \left\{ \frac{\partial^2 \log f(\mathbf{x}; \theta)}{\partial \theta^2} \right\} \triangleq J_{11}(\theta). \quad (7.19)$$

$J_{11}(\theta)$ es referida como la *información de Fisher* contenida en el conjunto de datos sobre el parámetro θ . Se obtuvo que

$$\text{var}\{T(\mathbf{x})\} \geq \frac{1}{E \left\{ \left(\frac{\partial \log f(\mathbf{x}; \theta)}{\partial \theta} \right)^2 \right\}} = \frac{-1}{E \left\{ \frac{\partial^2 \log f(\mathbf{x}; \theta)}{\partial \theta^2} \right\}} = \frac{1}{J_{11}}. \quad (7.20)$$

Si $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ son variables aleatorias i.i.d., se cumple que

$$\log f(x_1, x_2, \dots, x_n; \theta) = \log f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta) = \sum_{i=1}^n \log f(x_i; \theta),$$

y por lo tanto,

$$\begin{aligned}
E \left\{ \frac{\partial^2 \log f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n; \theta)}{\partial \theta^2} \right\} &= \sum_{i=1}^n E \left\{ \frac{\partial^2 \log f(\mathbf{x}_i; \theta)}{\partial \theta^2} \right\} \\
&= nE \left\{ \frac{\partial^2 \log f(\mathbf{x}_i; \theta)}{\partial \theta^2} \right\} \\
&= -nE \left\{ \left(\frac{\partial \log f(\mathbf{x}_i; \theta)}{\partial \theta} \right)^2 \right\},
\end{aligned} \tag{7.21}$$

por lo que la cota se simplifica a

$$\text{var}\{T(\mathbf{x})\} \geq \frac{1}{nE \left\{ \left(\frac{\partial \log f(\mathbf{x}_i; \theta)}{\partial \theta} \right)^2 \right\}} = \frac{-1}{nE \left\{ \frac{\partial^2 \log f(\mathbf{x}_i; \theta)}{\partial \theta^2} \right\}}. \tag{7.22}$$

Retornando a la situación general de la ecuación 7.20, en la desigualdad de Cauchy-Schwarz de la ecuación 7.16 la igualdad se cumple si y solo si $g(x) = af(x)$, donde a es una constante. Por lo tanto, la condición de igualdad en las ecuaciones 7.17 y 7.20 es

$$\frac{\partial \log f(\mathbf{x}; \theta)}{\partial \theta} = [T(\mathbf{x}) - \theta]a(\theta). \tag{7.23}$$

En ese caso, el estimador $T(\mathbf{x})$ tiene la menor varianza posible y se dice que es un estimador *eficiente* para θ . Claramente, si existe un estimador eficiente, es el mejor estimador respecto al punto de vista de la varianza. En general, un estimador eficiente puede existir o no, según el problema.

Se mostrará que si existe un estimador eficiente, es el estimador ML. Para hacerlo, se observa que el estimador ML cumple la ecuación 7.9, que indica que

$$\left. \frac{\partial \log f(\mathbf{x}; \theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_{\text{ML}}} = 0.$$

Sustituyendo este resultado en la condición de eficiencia, dada por la ecuación 7.23, se obtiene que

$$\left. \frac{\partial \log f(\mathbf{x}; \theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_{\text{ML}}} = [T(\mathbf{x}) - \theta]a(\theta) \Big|_{\theta=\hat{\theta}_{\text{ML}}} = 0.$$

Considerando que $a(\theta) > 0$, como se mostrará, se concluye que

$$\hat{\theta}_{\text{ML}} = T(\mathbf{x}),$$

lo que muestra que si existe un estimador eficiente, necesariamente es el estimador ML. Para completar la prueba, hay que mostrar que $a(\theta) > 0$. La diferenciación de la ecuación 7.23 resulta en

$$\frac{\partial^2 \log f(\mathbf{x}; \theta)}{\partial \theta^2} = -a(\theta) + [T(\mathbf{x}) - \theta]a'(\theta),$$

y tomando esperanza se obtiene que

$$a(\theta) = -E \left\{ \frac{\partial^2 \log f(\mathbf{x}; \theta)}{\partial \theta^2} \right\} = E \left\{ \left(\frac{\partial \log f(\mathbf{x}; \theta)}{\partial \theta} \right)^2 \right\} = J_{11} > 0,$$

donde se empleó que $E\{T(\mathbf{x}) - \theta\} = 0$ por la hipótesis de no sesgo de $T(\mathbf{x})$, además de la ecuación 7.19.

Sustituyendo $a(\theta) = J_{11}$ en la ecuación 7.23, se obtiene que la p.d.f. de un estimador eficiente tiene la forma

$$f(\mathbf{x}; \theta) = h(\mathbf{x})e^{\int J_{11}[T(\mathbf{x}) - \theta] d\theta}. \tag{7.24}$$

Efectivamente, integrando la ecuación 7.23 respecto a θ se obtiene que

$$\log f(\mathbf{x}; \theta) = \int J_{11}[T(\mathbf{x}) - \theta] d\theta + g(\mathbf{x}),$$

y al tomar exponencial se obtiene la ecuación 7.24. La ecuación 7.24 representa la familia de p.d.f. exponencial³. En particular, si J_{11} es independiente de θ , se reduce a una p.d.f. gaussiana. Para verlo, se parte observando que en ese caso, el exponente es

$$J_{11} \int [T(\mathbf{x}) - \theta] d\theta = J_{11} \left[T(x)\theta - \frac{\theta^2}{2} \right],$$

y se puede escribir como

$$J_{11} \left[T(x)\theta - \frac{\theta^2}{2} \right] = -\frac{J_{11}}{2}[T(\mathbf{x}) - \theta]^2 + \frac{J_{11}}{2}T^2(\mathbf{x}),$$

por lo que la ecuación 7.24 queda

$$f(\mathbf{x}; \theta) = h(\mathbf{x})e^{J_{11}T^2(\mathbf{x})/2}e^{-J_{11}[T(\mathbf{x})-\theta]^2/2}$$

La función $h(\mathbf{x})$ debe ser tal que la p.d.f. integre a 1, y por lo tanto,

$$h(\mathbf{x})e^{J_{11}T^2(\mathbf{x})/2} = \frac{J_{11}}{\sqrt{2\pi}},$$

resultando en que

$$f(\mathbf{x}; \theta) = \frac{J_{11}}{\sqrt{2\pi}}e^{-J_{11}[T(\mathbf{x})-\theta]^2/2},$$

concluyendo que $T(\mathbf{x}) \sim N(\theta, 1/J_{11})$.

No necesariamente existe un estimador eficiente en todos los casos. Sin embargo, el parámetro puede tener diversos estimadores insesgados, y entre ellos, alguno tiene menor varianza que el resto. Surge la pregunta de como proseguir para obtener dicho estimador.

Ejemplo 8-17 Sea $\mathbf{x}_i \sim N(\mu, \sigma^2)$, donde μ es desconocido y σ^2 es conocido. Además, las muestras \mathbf{x}_i son i.i.d. para $i = 1, \dots, n$. Es fácil verificar que

$$T(\mathbf{x}) = \frac{\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_n}{n} = \bar{\mathbf{x}}$$

es un estimador insesgado para μ , y también eficiente. De hecho, $\bar{\mathbf{x}} \sim N(\mu, \sigma^2/n)$ (ver la ecuación 6.11). Pero supóngase que el parámetro de interés no es μ sino μ^2 . De esta forma, $\theta = \mu^2$ es desconocido y se necesita un estimador para θ . Se comenzará calculando la cota de Cramér-Rao. La p.d.f. de los datos es

$$f(x_1, x_2, \dots, x_n; \theta) = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} e^{-\sum_{i=1}^n (x_i - \sqrt{\theta})^2 / 2\sigma^2} \quad (7.25)$$

y tomando logaritmo y diferenciando respecto a θ se obtiene que

$$\frac{\partial \log f(x_1, x_2, \dots, x_n; \theta)}{\partial \theta} = \frac{\sum_{i=1}^n (x_i - \sqrt{\theta})^2}{2\sigma^2\sqrt{\theta}},$$

y elevar al cuadrado y tomar esperanza resulta en

$$E \left\{ \left(\frac{\partial \log f(x_1, x_2, \dots, x_n; \theta)}{\partial \theta} \right)^2 \right\} \stackrel{(a)}{=} \frac{\sum_{i=1}^n E \{(x_i - \mu)^2\}}{4\sigma^4\mu^2} = \frac{n\sigma^2}{4\sigma^4\mu^2} = \frac{n}{4\sigma^2\mu^2},$$

donde (a) se tuvo en cuenta que como las variables aleatorias \mathbf{x}_i son independientes, los términos de la forma $E\{(x_i - \mu)(x_j - \mu)\}$ con $i \neq j$ que surgen al elevar al cuadrado se anulan. Se concluye que la cota de CR para $\theta = \mu^2$ es

$$\sigma_{\text{CR}}^2 \triangleq \frac{4\sigma^2\mu^2}{n} \leq \text{var}\{\hat{\theta}\},$$

³Ver https://en.wikipedia.org/wiki/Exponential_family, por ejemplo.

donde $\hat{\theta}$ representa un estimador insesgado para $\theta = \mu^2$. Para obtener un estimador insesgado, se examinará el estimador \bar{x}^2 . Como

$$E\{\bar{x}^2\} = \text{var}\{\bar{x}\} + E^2\{\bar{x}\} = \frac{\sigma^2}{n} + \mu^2, \quad (7.26)$$

no es un estimador insesgado de $\theta = \mu^2$, pero

$$\hat{\theta}(\mathbf{x}) = \bar{x}^2 - \frac{\sigma^2}{n}$$

es insesgado. Su varianza es

$$\begin{aligned} \text{var}\{\hat{\theta}(\mathbf{x})\} &= E\left\{[\hat{\theta} - \mu^2]^2\right\} \\ &= E\left\{\left[\bar{x}^2 - \left(\frac{\sigma^2}{n} + \mu^2\right)\right]^2\right\} \\ &= E\left\{\left[\bar{x}^2 - E\{\bar{x}^2\}\right]^2\right\} \\ &= E\{\bar{x}^4\} - E^2\{\bar{x}^2\}. \end{aligned}$$

Para continuar, se considera que el momento cuarto de $\bar{x} \sim N(\mu, \sigma^2/n)$ está dado por la ecuación 4.20 reemplazando σ^2 por σ^2/n , y el momento segundo por la ecuación 7.26. Por lo tanto,

$$\begin{aligned} \text{var}\{\hat{\theta}(\mathbf{x})\} &= \left(\mu^4 + \frac{6\sigma^2\mu^2}{n} + \frac{3\sigma^4}{n^2}\right) - \left(\mu^4 + \frac{2\sigma^2\mu^2}{n} + \frac{\sigma^4}{n^2}\right) \\ &= \frac{4\sigma^2\mu^2}{n} + \frac{2\sigma^4}{n^2}. \end{aligned}$$

Se concluye que

$$\text{var}\{\hat{\theta}(\mathbf{x})\} = \frac{4\sigma^2\mu^2}{n} + \frac{2\sigma^4}{n^2} > \frac{4\sigma^2\mu^2}{n} = \sigma_{\text{CR}}^2.$$

Se obtuvo que la varianza del estimador insesgado para $\theta = \mu$ supera la cota de CR, y por lo tanto, no es un estimador eficiente. ¿Será posible obtener un mejor estimador para $\theta = \mu^2$? Para responder esta pregunta, se introduce el concepto de estimador insesgado de varianza mínima.

7.3.4. Estimador insesgado de varianza mínima

Llámese $\{T(\mathbf{x})\}$ al conjunto de todos los estimadores insesgados de θ y supóngase que $T_0(\mathbf{x}) \in \{T(\mathbf{x})\}$. Si

$$E\{(T_0(\mathbf{x}) - \theta)^2\} \leq E\{(T(\mathbf{x}) - \theta)^2\}$$

para todo $T(\mathbf{x}) \in \{T(\mathbf{x})\}$, $T_0(\mathbf{x})$ se denomina el estimador insesgado de varianza mínima (UMVUE, *uniformly minimum variance unbiased estimator*) para θ . El UMVUE para θ es el estimador insesgado con la menor varianza posible. Es fácil mostrar que el UMVUE es único. El teorema de Rao-Blackwell indica la forma de encontrar el UMVUE para un parámetro θ .

Teorema de Rao-Blackwell Sea $h(\mathbf{x})$ un estimador insesgado para θ y $T(\mathbf{x})$ un estadístico suficiente para θ . Bajo estas hipótesis, la esperanza condicional $E[h(\mathbf{x}) | T(\mathbf{x})]$ es independiente de θ y es el UMVUE para θ .

Se realiza a continuación la demostración basada en [1], pero se incluye la expresión explícita de las esperanzas como integrales para mayor claridad. La demostración también puede encontrarse en [19].

Sea el estimador

$$T_0(\mathbf{x}) = E[h(\mathbf{x}) | T(\mathbf{x})]. \quad (7.27)$$

Observar que de esta forma, $T_0(\mathbf{x})$ solo depende de $T(\mathbf{x}) = t$ y no del parámetro θ ni de otra función de los datos, ya que

$$\begin{aligned} T_0(\mathbf{x}) &= E[h(\mathbf{x}) | T(\mathbf{x})] \\ &= \int h(\mathbf{x}) f(\mathbf{x}; \theta | T(\mathbf{x})) d\mathbf{x} \\ &\stackrel{(a)}{=} \int h(\mathbf{x}) f(\mathbf{x} | T(\mathbf{x})) d\mathbf{x}, \end{aligned}$$

y al integrar en \mathbf{x} , el resultado solo depende de los datos a través de $T(\mathbf{x}) = t$. Notar que en (a) se empleó la hipótesis de que $T(\mathbf{x})$ es suficiente, por lo que la p.d.f. condicional no depende del parámetro θ . Además, se cumple que

$$E\{T_0(\mathbf{x})\} = E\{E[h(\mathbf{x}) | T(\mathbf{x})]\} \stackrel{(a)}{=} E\{h(\mathbf{x})\} \stackrel{(b)}{=} \theta.$$

donde en (a) se aplicó la ley de la esperanza total, dada por la ecuación 5.86, y en (b) se empleó la hipótesis de que el estimador $h(\mathbf{x})$ es insesgado. Esto indica que $T_0(\mathbf{x})$ también es un estimador insesgado para θ . Para hacer el razonamiento mas explícito se incluye la misma expresión en términos de integrales:

$$\begin{aligned} E\{T_0(\mathbf{x})\} &\stackrel{(a)}{=} \int T_0(\mathbf{x}) f(T(\mathbf{x}); \theta) dt \\ &= \int E[h(\mathbf{x}) | T(\mathbf{x})] f(T(\mathbf{x}); \theta) dt \\ &= \int \left(\int h(\mathbf{x}) f(\mathbf{x} | T(\mathbf{x})) d\mathbf{x} \right) f(T(\mathbf{x}); \theta) dt \\ &= \int h(\mathbf{x}) \int f(\mathbf{x} | T(\mathbf{x})) f(T(\mathbf{x}); \theta) dt d\mathbf{x} \\ &= \int h(\mathbf{x}) \int f(\mathbf{x}; \theta, T(\mathbf{x})) dt d\mathbf{x} \\ &= \int h(\mathbf{x}) f(\mathbf{x}; \theta) d\mathbf{x} \\ &= E\{h(\mathbf{x})\}, \end{aligned}$$

donde en (a) se tuvo en cuenta que $T_0(\mathbf{x})$ solo depende de los datos a través de $T(\mathbf{x}) = t$ y no de θ o de otra función de los datos, como se comentó previamente, y en el resto del razonamiento se emplearon definiciones y propiedades básicas.

Continuando con la demostración, al momento se tiene que $T_0(\mathbf{x})$ es un estimador válido ya que no depende del parámetro θ y además es insesgado. Para completar la demostración hay que probar que la varianza de $T_0(\mathbf{x})$ es menor a la varianza de cualquier otro estimador insesgado, es decir,

$$\text{var}\{T_0(\mathbf{x})\} \leq \text{var}\{h(\mathbf{x})\}. \quad (7.28)$$

Como

$$\text{var}\{T_0(\mathbf{x})\} = E\{T_0^2(\mathbf{x})\} - \theta^2 \quad \text{y} \quad \text{var}\{h(\mathbf{x})\} = E\{h^2(\mathbf{x})\} - \theta^2,$$

alcanza con probar que

$$E\{T_0^2(\mathbf{x})\} \leq E\{h^2(\mathbf{x})\}. \quad (7.29)$$

Además

$$E\{h^2(\mathbf{x})\} = E\{E[h^2(\mathbf{x}) | T(\mathbf{x})]\}, \quad (7.30)$$

donde nuevamente se empleó la ley de esperanza total. Al igual que antes, se incluye el razonamiento

expresando las esperanzas como integrales:

$$\begin{aligned}
E\{h^2(\mathbf{x})\} &= \int h^2(\mathbf{x})f(\mathbf{x}; \theta) d\mathbf{x} \\
&= \int h^2(\mathbf{x}) \left(\int f(\mathbf{x}; \theta | T(\mathbf{x}))f(T(\mathbf{x}); \theta) dt \right) d\mathbf{x} \\
&\stackrel{(a)}{=} \int \left(\int h^2(\mathbf{x})f(\mathbf{x} | T(\mathbf{x})) d\mathbf{x} \right) f(T(\mathbf{x}); \theta) dt \\
&= \int E[h^2(\mathbf{x}) | T(\mathbf{x})]f(T(\mathbf{x}); \theta) dt \\
&\stackrel{(b)}{=} E\{E[h^2(\mathbf{x}) | T(\mathbf{x})]\},
\end{aligned}$$

donde en (a) se tuvo en cuenta que $T(\mathbf{x})$ es suficiente y por lo tanto la densidad condicional no depende del parámetro y en (b) se consideró que $E[h^2(\mathbf{x}) | T(\mathbf{x})]$ solo depende de los datos a través de $T(\mathbf{x})$ por el mismo argumento dado antes para $T_0(\mathbf{x}) = E[h(\mathbf{x}) | T(\mathbf{x})]$. A partir de este resultado y la definición de $T_0(\mathbf{x})$, la desigualdad de la ecuación 7.29 que se quiere probar es equivalente a

$$E\{E^2[h(\mathbf{x}) | T(\mathbf{x})]\} \leq E\{E[h^2(\mathbf{x}) | T(\mathbf{x})]\}.$$

Como esta expresión en términos de integrales es

$$\begin{aligned}
\int \left(\int h(\mathbf{x})f(\mathbf{x} | T(\mathbf{x})) d\mathbf{x} \right)^2 f(T(\mathbf{x}); \theta) dt &\leq \int \left(\int h^2(\mathbf{x})f(\mathbf{x} | T(\mathbf{x})) d\mathbf{x} \right) f(T(\mathbf{x}); \theta) dt, \\
\text{o} \quad \int E^2[h(\mathbf{x}) | T(\mathbf{x})]f(T(\mathbf{x}); \theta) dt &\leq \int E[h^2(\mathbf{x}) | T(\mathbf{x})]f(T(\mathbf{x}); \theta) dt
\end{aligned}$$

la desigualdad se cumple si se cumple que

$$E^2[h(\mathbf{x}) | T(\mathbf{x})] \leq E[h^2(\mathbf{x}) | T(\mathbf{x})]$$

o

$$\left(\int h(\mathbf{x})f(\mathbf{x} | T(\mathbf{x})) d\mathbf{x} \right)^2 \leq \int h^2(\mathbf{x})f(\mathbf{x} | T(\mathbf{x})) d\mathbf{x}.$$

Pero esta desigualdad efectivamente se cumple, ya que

$$\begin{aligned}
\left(\int h(\mathbf{x})f(\mathbf{x} | T(\mathbf{x})) d\mathbf{x} \right)^2 &= \left(\int h(\mathbf{x})\sqrt{f(\mathbf{x} | T(\mathbf{x}))}\sqrt{f(\mathbf{x} | T(\mathbf{x}))} d\mathbf{x} \right)^2 \\
&\stackrel{(a)}{\leq} \left(\int h^2(\mathbf{x})f(\mathbf{x} | T(\mathbf{x})) d\mathbf{x} \right) \left(\int f(\mathbf{x} | T(\mathbf{x})) d\mathbf{x} \right) \quad (7.31) \\
&= \int h^2(\mathbf{x})f(\mathbf{x} | T(\mathbf{x})) d\mathbf{x},
\end{aligned}$$

donde en (a) se aplicó la desigualdad de Cauchy-Schwarz, dada por la ecuación 7.16. Esto concluye la prueba. El mismo razonamiento en términos de esperanzas es

$$E^2[h(\mathbf{x}) | T(\mathbf{x})] \leq E[h^2(\mathbf{x}) | T(\mathbf{x})]E[1 | T(\mathbf{x})] = E[h^2(\mathbf{x}) | T(\mathbf{x})].$$

La igualdad en las ecuaciones 7.28 y 7.29 es

$$E\{T_0^2(\mathbf{x})\} = E\{h^2(\mathbf{x})\},$$

y a partir de la definición de $T_0(\mathbf{x})$ de la ecuación 7.27 y de la ecuación 7.30 se puede expresar como

$$E\{E^2[h(\mathbf{x}) | T(\mathbf{x})]\} = E\{E[h^2(\mathbf{x}) | T(\mathbf{x})]\} \quad \Leftrightarrow \quad E\{E[h^2(\mathbf{x}) | T(\mathbf{x})]\} - E\{E^2[h(\mathbf{x}) | T(\mathbf{x})]\} = 0.$$

Además, como

$$\begin{aligned} E\{E[h^2(\mathbf{x}) | T(\mathbf{x})]\} - E\{E^2[h(\mathbf{x}) | T(\mathbf{x})]\} &= E\{E[h^2(\mathbf{x}) | T(\mathbf{x})] - E^2[h(\mathbf{x}) | T(\mathbf{x})]\} \\ &= E\{\text{var}[h(\mathbf{x}) | T(\mathbf{x})]\}, \end{aligned}$$

la condición de igualdad es

$$E\{\text{var}[h(\mathbf{x}) | T(\mathbf{x})]\} = 0 \Leftrightarrow \text{var}[h(\mathbf{x}) | T(\mathbf{x})] = 0 \Leftrightarrow E[h^2(\mathbf{x}) | T(\mathbf{x})] = E^2[h(\mathbf{x}) | T(\mathbf{x})] = h^2,$$

lo cual se cumple si y solo si

$$h = E[h(\mathbf{x}) | T(\mathbf{x})].$$

El resultado indica que si h es una función únicamente del estadístico suficiente $T(\mathbf{x}) = t$, es el UMVUE.

La condición expresando las esperanzas explícitamente como integrales es la igualdad en la desigualdad de Cauchy-Schwarz,

$$\left(\int h(\mathbf{x}) f(\mathbf{x} | T(\mathbf{x})) d\mathbf{x} \right)^2 = \int h^2(\mathbf{x}) f(\mathbf{x} | T(\mathbf{x})) d\mathbf{x},$$

que se cumple si $h(\mathbf{x}) = h$ solo depende de los datos a través de $T(\mathbf{x}) = t$, y en ese caso,

$$h^2 \left(\int f(\mathbf{x} | T(\mathbf{x})) d\mathbf{x} \right)^2 = h^2 \int f(\mathbf{x} | T(\mathbf{x})) d\mathbf{x} = h^2.$$

De forma equivalente, la condición de igualdad en la desigualdad de Cauchy-Schwarz en la ecuación 7.16 es $f(x) = hg(x)$, por lo que la igualdad en 7.31 se da si

$$h(\mathbf{x}) \sqrt{f(\mathbf{x} | T(\mathbf{x}))} = h \sqrt{f(\mathbf{x} | T(\mathbf{x}))},$$

donde h depende solo de los datos \mathbf{x} a través de $T(\mathbf{x}) = t$.

De acuerdo al teorema de Rao-Blackwell, un estimador insesgado para θ que es función de su estadístico suficiente, es el UMVUE para θ . Además, la esperanza condicional de cualquier estimador insesgado dado el estadístico suficiente provee el UMVUE.

Ejemplo 8-17 (continuación) En el ejemplo 8-17 se obtuvo que un estimador insesgado de $\theta = \mu^2$ es

$$\hat{\theta}(\bar{\mathbf{x}}) = \bar{\mathbf{x}}^2 - \frac{\sigma^2}{n}.$$

La p.d.f de los datos está dada por la ecuación 7.25, y observando que el exponente se puede escribir como

$$-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \sqrt{\theta})^2 = -\frac{1}{2\sigma^2} \left(\sum_{i=1}^n x_i^2 - 2\sqrt{\theta} \sum_{i=1}^n x_i + n\theta \right) = -\frac{1}{2\sigma^2} \left(\sum_{i=1}^n x_i^2 - 2\sqrt{\theta}n\bar{\mathbf{x}} + n\theta \right),$$

la p.d.f. queda

$$f(x_1, x_2, \dots, x_n; \theta) = \underbrace{\left(\frac{1}{2\pi\sigma^2} \right)^{n/2} e^{-\sum_{i=1}^n x_i^2 / 2\sigma^2}}_{h(\mathbf{x})} \underbrace{e^{-n(\theta - 2\sqrt{\theta}\bar{\mathbf{x}}) / 2\sigma^2}}_{g_\theta(T(\mathbf{x}))} = h(\mathbf{x})g_\theta(T(\mathbf{x})),$$

y por el teorema de factorización de Fisher-Neyman (ver la sección 7.3.2), $T(\mathbf{x}) = \bar{\mathbf{x}}$ es un estadístico suficiente para $\theta = \mu^2$. Como el estimador es una función del estadístico suficiente, ya que

$$\hat{\theta}(\bar{\mathbf{x}}) = T^2(\mathbf{x}) - \frac{\sigma^2}{n},$$

por el teorema de Rao-Blackwell, es el UMVUE para $\theta = \mu^2$.

El teorema de Rao-Blackwell establece que todos los UMVUEs dependen únicamente del estadístico suficiente. Esto es consistente con la noción de suficiencia, ya que en ese caso, $T(\mathbf{x})$ contiene la misma información sobre θ que el conjunto de datos $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, y por lo tanto, el mejor estimador posible debe depender únicamente del estadístico suficiente. De esta forma, si $h(\mathbf{x})$ es un estimador insesgado para θ , por el teorema de Rao-Blackwell es posible reducir su varianza condicionándolo en $T(\mathbf{x})$. Comparado con $h(\mathbf{x})$,

$$\begin{aligned} T_0 &= E\{h(\mathbf{x}) \mid T(\mathbf{x}) = t\} \\ &= \int h(\mathbf{x})f(\mathbf{x}; \theta \mid T(\mathbf{x}) = t) d\mathbf{x} \end{aligned}$$

es un mejor estimador para θ . Como se mencionó previamente, $f(\mathbf{x}; \theta \mid T(\mathbf{x}) = t)$ no depende de θ debido a que $T(\mathbf{x})$ es suficiente. Por lo tanto, $T_0 = E\{h(\mathbf{x}) \mid T(\mathbf{x}) = t\}$ es independiente de θ y depende de los datos solo a través de $T(\mathbf{x}) = t$, y como se demostró en el teorema, es insesgado. Otra forma de ver que reduce la varianza es aplicando la identidad

$$\text{var}\{\mathbf{x}\} = E\{\text{var}[\mathbf{x} \mid \mathbf{y}]\} + \text{var}\{E[\mathbf{x} \mid \mathbf{y}]\} \quad (7.32)$$

a $h(\mathbf{x})$ condicionado con $T(\mathbf{x})$. De esta forma,

$$\text{var}\{h(\mathbf{x})\} = E\{\text{var}[h(\mathbf{x}) \mid T(\mathbf{x})]\} + \text{var}\{E[h(\mathbf{x}) \mid T(\mathbf{x})]\} = E\{\text{var}[h(\mathbf{x}) \mid T(\mathbf{x})]\} + \text{var}\{T_1\},$$

y como $\text{var}[h(\mathbf{x}) \mid T(\mathbf{x})] \geq 0$, se concluye que

$$\text{var}\{T_1\} \leq \text{var}\{h(\mathbf{x})\}.$$

La ecuación 7.32 se denomina *ley de la varianza total*⁴. Para demostrarla, se parte de que de la definición de varianza, se cumple que

$$\text{var}\{\mathbf{x} \mid \mathbf{y}\} = E[\mathbf{x}^2 \mid \mathbf{y}] + E^2[\mathbf{x} \mid \mathbf{y}],$$

y tomando esperanza, se tiene que

$$E\{\text{var}[\mathbf{x} \mid \mathbf{y}]\} = E\{E[\mathbf{x}^2 \mid \mathbf{y}]\} - E\{E^2[\mathbf{x} \mid \mathbf{y}]\}.$$

Por otro lado, también de la definición de varianza, se cumple que

$$\text{var}\{E[\mathbf{x} \mid \mathbf{y}]\} = E\{E^2[\mathbf{x} \mid \mathbf{y}]\} - E^2\{E[\mathbf{x} \mid \mathbf{y}]\}.$$

Sumando estas dos ecuaciones, se obtiene que

$$\begin{aligned} E\{\text{var}[\mathbf{x} \mid \mathbf{y}]\} + \text{var}\{E[\mathbf{x} \mid \mathbf{y}]\} &= E\{E[\mathbf{x}^2 \mid \mathbf{y}]\} - E^2\{E[\mathbf{x} \mid \mathbf{y}]\} \\ &\stackrel{(a)}{=} E(\mathbf{x}^2) - E^2(\mathbf{x}) \\ &= \text{var}\{\mathbf{x}\}, \end{aligned}$$

donde en (a) se empleó la ley de la esperanza total, dada por la ecuación 5.86.

7.3.5. El problema de los taxis

Se considera el problema de estimar la cantidad total de taxis en una ciudad. Para hacerlo, un observador estacionario parado en una esquina, observa los números de serie de algunos de ellos. Si x_1, x_2, \dots, x_n son los números de serie de n taxis observados, el objetivo es estimar la cantidad total N de taxis en base a las observaciones. Este problema también es referido en la literatura como *el problema de los tanques alemanes*⁵.

Es razonable asumir que la probabilidad de observar un número de serie particular es $1/N$, por lo que las variables aleatorias $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ observadas pueden asumirse discretas con distribución uniforme en el intervalo $[1, N]$, donde N es la constante desconocida que se quiere determinar. De esta forma,

$$P\{\mathbf{x}_i = k\} = \frac{1}{N}, \quad \text{con} \quad k = 1, 2, \dots, N \quad \text{y} \quad i = 1, 2, \dots, n.$$

⁴Ver https://en.wikipedia.org/wiki/Law_of_total_variance, por ejemplo.

⁵Ver https://en.wikipedia.org/wiki/German_tank_problem, por ejemplo.

En este caso, como se vio en el ejemplo 8-15 (ecuación 7.10), el estadístico

$$T(\mathbf{x}) = \max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n),$$

es suficiente para N . En la búsqueda del UMVUE, se comienza calculando la media del estadístico suficiente con el fin de estudiar si existe una transformación que lo haga insesgado. La p.m.f. del estadístico suficiente es

$$\begin{aligned} P\{T = k\} &= P\{\max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = k\} \\ &= P\{\max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \leq k\} - P\{\max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \leq k-1\} \\ &= P\{\mathbf{x}_1 \leq k, \mathbf{x}_2 \leq k, \dots, \mathbf{x}_n \leq k\} - P\{\mathbf{x}_1 \leq k-1, \mathbf{x}_2 \leq k-1, \dots, \mathbf{x}_n \leq k-1\} \\ &\stackrel{(a)}{=} (P\{\mathbf{x}_i \leq k\})^n - (P\{\mathbf{x}_i \leq k-1\})^n \\ &\stackrel{(b)}{=} \left(\frac{k}{N}\right)^n - \left(\frac{k-1}{N}\right)^n, \quad k = 1, 2, \dots, N, \end{aligned}$$

donde en (a) se tuvo en cuenta que las variables aleatorias \mathbf{x}_i son independientes y en (b) que

$$P\{\mathbf{x}_i \leq k\} = \sum_{j=1}^k P\{\mathbf{x}_i = j\} = \sum_{j=1}^k \frac{1}{N} = \frac{k}{N}.$$

La media es

$$\begin{aligned} E\{T(\mathbf{x})\} &= \sum_{k=1}^N k P\{T = k\} \\ &= \frac{1}{N^n} \sum_{k=1}^N k [k^n - (k-1)^n] \\ &= \frac{1}{N^n} \sum_{k=1}^N [k^{n+1} - (k-1+1)(k-1)^n] \\ &= \frac{1}{N^n} \sum_{k=1}^N [k^{n+1} - (k-1)^{n+1} - (k-1)^n] \\ &\stackrel{(a)}{=} \frac{1}{N^n} \left[N^{n+1} - \sum_{k=1}^N (k-1)^n \right], \end{aligned} \tag{7.33}$$

donde en (a) se tuvo en cuenta que

$$\sum_{k=1}^N k^{n+1} - \sum_{k=1}^N (k-1)^{n+1} = \sum_{k=1}^N k^{n+1} - \sum_{j=0}^{N-1} j^{n+1} = \sum_{k=1}^N k^{n+1} - \sum_{k=1}^{N-1} k^{n+1} = N^{n+1}.$$

Continuando con el cálculo de la media, para N grande, la sumatoria se puede aproximar como

$$\sum_{k=1}^N (k-1)^n \approx \int_0^N y^n dy = \frac{y^{n+1}}{n+1} \Big|_0^N = \frac{N^{n+1}}{n+1},$$

resultando en

$$E\{\max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)\} \approx \frac{1}{N^n} \left(N^{n+1} - \frac{N^{n+1}}{n+1} \right) = \frac{nN^n}{n+1}.$$

Se concluye que

$$T_1(\mathbf{x}) = \left(\frac{n+1}{n}\right) T(\mathbf{x}) = \left(\frac{n+1}{n}\right) \max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$$

es un estimador aproximadamente insesgado para N . Además, como depende únicamente del estadístico suficiente, por el teorema de Rao-Blackwell es también aproximadamente el UMVUE para N . En realidad,

al no ser un estimador exactamente insesgado, no puede aplicarse el teorema de Rao-Blackwell, por lo que se continuará en la búsqueda de un estimador insesgado. Para eso, considérese el estadístico

$$N(\mathbf{x}) = \min(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n),$$

que representa el menor número de serie observado entre las n observaciones. Este estadístico contiene información relevante sobre el número total de taxis, ya que indica que hay $N(\mathbf{x}) - 1$ taxis de número menor al menor número $N(\mathbf{x})$ observado. Es razonable asumir que hay la misma cantidad de taxis con número mayor al mayor observado, por lo que se elige como estimador para N a

$$T_0(\mathbf{x}) = \max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) + \min(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) - 1.$$

Se estudiará el sesgo de este estimador, para lo cual hay que calcular su media. La media exacta de $\max(\mathbf{x})$ está dada por la ecuación 7.33 y es

$$E\{\max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)\} = N - \frac{1}{N^n} \sum_{k=1}^{N-1} k^n. \quad (7.34)$$

Para calcular la media de $\min(\mathbf{x})$ hay que calcular su p.m.f.,

$$P\{\min(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = j\} = P\{\min(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \leq j\} - P\{\min(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \leq j-1\}.$$

Pero

$$\begin{aligned} P\{\min(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \leq j\} &= 1 - P\{\min(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) > j\} \\ &= 1 - P\{\mathbf{x}_1 > j, \mathbf{x}_2 > j, \dots, \mathbf{x}_n > j\} \\ &= 1 - \prod_{i=1}^n P\{\mathbf{x}_i > j\} \\ &= 1 - \prod_{i=1}^n (1 - P\{\mathbf{x}_i \leq j\}) \\ &= 1 - \prod_{i=1}^n \left(1 - \frac{j}{N}\right) \\ &= 1 - \frac{(N-j)^n}{N^n}, \end{aligned}$$

por lo que

$$\begin{aligned} P\{\min(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = j\} &= 1 - \frac{(N-j)^n}{N^n} - \left(1 - \frac{(N-j+1)^n}{N^n}\right) \\ &= \frac{(N-j+1)^n - (N-j)^n}{N^n}, \end{aligned}$$

y la esperanza es

$$\begin{aligned} E\{\min(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)\} &= \sum_{j=1}^N j P\{\min(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = j\} \\ &= \sum_{j=1}^N j \frac{(N-j+1)^n - (N-j)^n}{N^n} \end{aligned}$$

Realizando el cambio de variable $k = N - j + 1$ se tiene que

$$\begin{aligned}
E\{\min(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)\} &= \frac{1}{N^n} \sum_{k=1}^N (N - k + 1)[k^n - (k - 1)^n] \\
&= \frac{1}{N^n} \left\{ N \sum_{k=1}^N [k^n - (k - 1)^n] - \sum_{k=1}^N (k - 1)[k^n - (k - 1)^n] \right\} \\
&= \frac{1}{N^n} \left\{ NN^n - \sum_{k=1}^N [(k - 1)k^n - (k - 1)^{n+1}] \right\} \\
&= \frac{1}{N^n} \left\{ N^{n+1} - \sum_{k=1}^N [k^{n+1} - (k - 1)^{n+1}] + \sum_{k=1}^N k^n \right\} \\
&= \frac{1}{N^n} \left(N^{n+1} - N^{n+1} + \sum_{k=1}^N k^n \right) \\
&= \frac{1}{N^n} \left(N^n + \sum_{k=1}^{N-1} k^n \right),
\end{aligned}$$

resultando en

$$E\{\min(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)\} = 1 + \frac{1}{N^n} \sum_{k=1}^{N-1} k^n.$$

Finalmente, combinando este resultado con el de la ecuación 7.34 se obtiene que

$$\begin{aligned}
E\{T_0(\mathbf{x})\} &= E\{\max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)\} + E\{\min(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)\} - 1 \\
&= \left(N - \frac{1}{N^n} \sum_{k=1}^{N-1} k^n \right) + \left(1 + \frac{1}{N^n} \sum_{k=1}^{N-1} k^n \right) - 1 \\
&= N.
\end{aligned}$$

Se concluye que $T_0(\mathbf{x})$ es un estimador insesgado para N . Sin embargo, no es una función únicamente del estadístico suficiente $\max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, y por lo tanto, no es el estimador UMVUE para N . Teniendo en cuenta que $T_0(\mathbf{x})$ es insesgado, por el teorema de Rao-Blackwell, el UMVUE es

$$E\{T_0 | T(\mathbf{x}) = \max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)\},$$

pero debido a la presencia de la función $\min(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ en $T_0(\mathbf{x})$, el cálculo de esta esperanza es difícil. Para simplificar el cálculo de la esperanza condicional involucrada en el teorema de Rao-Blackwell, conviene emplear un estimador insesgado que consista en una función más simple de los datos. Considérese por ejemplo el estimador más simple

$$T_2(\mathbf{x}) = \mathbf{x}_1.$$

La media es

$$E\{T_2(\mathbf{x})\} = E\{\mathbf{x}_1\} = \sum_{k=1}^N k P\{\mathbf{x}_1 = k\} = \sum_{k=1}^N k \frac{1}{N} \stackrel{(a)}{=} \frac{1}{N} \times \frac{N(N+1)}{2} = \frac{N+1}{2},$$

donde en (a) se empleó que⁶

$$\sum_{k=1}^m k = \frac{m(m+1)}{2}. \quad (7.35)$$

Por lo tanto,

$$h(\mathbf{x}) = 2\mathbf{x}_1 - 1$$

⁶Ver https://en.wikipedia.org/wiki/List_of_mathematical_series, por ejemplo.

es un estimador insesgado para N , al cual se le puede aplicar el teorema de Rao-Blackwell para obtener el UMVUE,

$$\begin{aligned} E\{h(\mathbf{x}) \mid T(\mathbf{x})\} &= E\{2\mathbf{x}_1 - 1 \mid \max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = t\} \\ &= \sum_{k=1}^N (2k-1)P\{\mathbf{x}_1 = k \mid \max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = t\}. \end{aligned}$$

Pero

$$\begin{aligned} P\{\mathbf{x}_1 = k \mid T(\mathbf{x})\} &= \frac{P\{\mathbf{x}_1 = k, T(\mathbf{x}) = t\}}{P\{T(\mathbf{x}) = t\}} \\ &= \frac{P\{\mathbf{x}_1 = k, T(\mathbf{x}) \leq t\} - P\{\mathbf{x}_1 = k, T(\mathbf{x}) \leq t-1\}}{P\{T(\mathbf{x}) \leq t\} - P\{T(\mathbf{x}) \leq t-1\}} \\ &= \frac{P\{\mathbf{x}_1 = k, \mathbf{x}_2 \leq t, \dots, \mathbf{x}_n \leq t\} - P\{\mathbf{x}_1 = k, \mathbf{x}_2 \leq t-1, \dots, \mathbf{x}_n \leq t-1\}}{P\{\mathbf{x}_1 \leq t, \mathbf{x}_2 \leq t, \dots, \mathbf{x}_n \leq t\} - P\{\mathbf{x}_1 \leq t-1, \mathbf{x}_2 \leq t-1, \dots, \mathbf{x}_n \leq t-1\}} \\ &\stackrel{(a)}{=} \begin{cases} \frac{1}{N} \left(\frac{t}{N}\right)^{n-1} - \frac{1}{N} \left(\frac{t-1}{N}\right)^{n-1} & \text{si } k = 1, 2, \dots, t-1 \\ \frac{1}{N} \left(\frac{t}{N}\right)^{n-1} & \text{si } k = t \\ 0 & \text{si } k > t \end{cases} \end{aligned}$$

donde en (a) se empleó que las variables aleatorias \mathbf{x}_i son independientes con $P\{\mathbf{x}_i = k\} = 1/N$ y $P\{\mathbf{x}_i \leq k\} = k/N$ y además que $P\{\mathbf{x}_1 = t, T(\mathbf{x}) \leq t-1\} = 0$, ya que no puede ocurrir que $\mathbf{x}_1 = t$ y que $\max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \leq t-1$. Finalmente,

$$P\{\mathbf{x}_1 = k \mid \max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = t\} = \begin{cases} \frac{t^{n-1} - (t-1)^{n-1}}{t^n - (t-1)^n} & \text{si } k = 1, 2, \dots, t-1 \\ \frac{t^{n-1}}{t^n - (t-1)^n} & \text{si } k = t \\ 0 & \text{si } k > t. \end{cases}$$

Continuando,

$$\begin{aligned} E\{h(\mathbf{x}) \mid T(\mathbf{x})\} &= \sum_{k=1}^N (2k-1)P\{\mathbf{x}_1 = k \mid \max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = t\} \\ &= \frac{t^{n-1} - (t-1)^{n-1}}{t^n - (t-1)^n} \sum_{k=1}^{t-1} (2k-1) + \frac{t^{n-1}}{t^n - (t-1)^n} (2t-1) \\ &\stackrel{(a)}{=} \frac{t^{n-1} - (t-1)^{n-1}}{t^n - (t-1)^n} (t-1)^2 + \frac{t^{n-1}}{t^n - (t-1)^n} (2t-1) \\ &= \frac{t^{n-1}(t^2 - 2t + 1) - (t-1)^{n+1} + t^{n-1}(2t-1)}{t^n - (t-1)^n} \\ &= \frac{t^{n+1} - (t-1)^{n+1}}{t^n - (t-1)^n}, \end{aligned}$$

es el UMVUE para N , donde $t = \max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$. La igualdad en (a) proviene de que

$$\sum_{k=1}^{t-1} (2k-1) = 2 \sum_{k=1}^{t-1} k - \sum_{k=1}^{t-1} 1 = 2 \frac{(t-1)t}{2} - (t-1) = (t-1)(t-1) = (t-1)^2,$$

donde nuevamente se empleó la identidad de la ecuación 7.35. Se concluye que si $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ representan n observaciones de una población numerada,

$$\hat{N}_n = \frac{[\max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)]^{n+1} - [\max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) - 1]^{n+1}}{[\max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)]^n - [\max(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) - 1]^n}$$

es el estimador insesgado de varianza mínima para el total de la población. Si bien la estimación de la varianza de este UMVUE es difícil, se tiene la certeza de que es el que tiene la menor varianza entre todos los estimadores insesgados.

7.3.6. Cota inferior de Crámer-Rao para el caso de múltiples parámetros

Es posible extender la cota de Crámer-Rao a la situación de múltiples parámetros. Sea $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_m]^t$ el conjunto de parámetros desconocidos bajo la p.d.f. $f(\mathbf{x}; \boldsymbol{\theta})$ y

$$\mathbf{T}(\mathbf{x}) = [T_1(\mathbf{x}), T_2(\mathbf{x}), \dots, T_m(\mathbf{x})]^t$$

un estimador insesgado de $\boldsymbol{\theta}$. La matriz de covarianza de $T(\mathbf{x})$ es

$$\text{cov}\{\mathbf{T}(\mathbf{x})\} = E\{[\mathbf{T}(\mathbf{x}) - \boldsymbol{\theta}][\mathbf{T}(\mathbf{x}) - \boldsymbol{\theta}]^t\},$$

de tamaño $m \times m$ y definida positiva, como toda matriz de covarianza (ver la ecuación 6.18). La cota de Crámer-Rao en este caso es

$$\text{cov}\{\mathbf{T}(\mathbf{x})\} \geq \mathbf{J}^{-1}(\boldsymbol{\theta}), \quad (7.36)$$

donde $\mathbf{J}(\boldsymbol{\theta})$ es la matriz $m \times m$ de información de Fisher asociada al parámetro vectorial $\boldsymbol{\theta}$ bajo la p.d.f. $f(\mathbf{x}; \boldsymbol{\theta})$. La notación $\mathbf{A} \geq \mathbf{B}$ en la ecuación 7.36 indica que la matriz $\mathbf{A} - \mathbf{B}$ es semidefinida positiva, y de forma acorde, la desigualdad estricta significa que la matriz es definida positiva. Los elementos de la matriz de información de Fisher son

$$J_{ij} = E\left(\frac{\partial \log f(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \log f(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_j}\right), \quad \begin{array}{l} i = 1, 2, \dots, m \\ j = 1, 2, \dots, m. \end{array} \quad (7.37)$$

Para probar la ecuación 7.36 se puede emplear la ecuación 7.15,

$$E\left[(T_k(\mathbf{x}) - \theta_k) \frac{\partial \log f(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_k}\right] = 1, \quad (7.38)$$

que es válida para cada parámetro θ_k individual con $k = 1, 2, \dots, m$. Además, como

$$E[T_k(\mathbf{x}) - \theta_k] = \int_{-\infty}^{\infty} [T_k(\mathbf{x}) - \theta_k] f(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x} = 0$$

por la condición de no sesgo, se cumple que

$$\begin{aligned} \frac{\partial E[T_k(\mathbf{x}) - \theta_k]}{\partial \theta_j} &= \int_{-\infty}^{\infty} [T_k(\mathbf{x}) - \theta_k] \frac{\partial f(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_j} d\mathbf{x} \\ &\stackrel{(a)}{=} \int_{-\infty}^{\infty} [T_k(\mathbf{x}) - \theta_k] \frac{\partial \log f(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_j} f(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x} \\ &= 0, \end{aligned}$$

donde en (a) se empleó la ecuación 7.14. Esto es

$$E\left[(T_k(\mathbf{x}) - \theta_k) \frac{\partial \log f(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_j}\right] = 0, \quad \begin{array}{l} j = 1, 2, \dots, m \\ k = 1, 2, \dots, m. \end{array}, \quad k \neq j. \quad (7.39)$$

Considérese el vector de tamaño $2m \times 1$

$$\mathbf{Z} = \begin{pmatrix} T_1(\mathbf{x}) - \theta_1 \\ T_2(\mathbf{x}) - \theta_2 \\ \vdots \\ \frac{T_m(\mathbf{x}) - \theta_m}{\frac{\partial \log f(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_1}} \\ \frac{\partial \log f(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_2} \\ \vdots \\ \frac{\partial \log f(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_m} \end{pmatrix} \triangleq \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}.$$

Debido a la condición de no sesgo del estimador y por la ecuación 7.18, se cumple que

$$E\{\mathbf{Z}\} = \mathbf{0},$$

y por lo tanto

$$\text{cov}\{\mathbf{Z}\} = E\{\mathbf{Z}\mathbf{Z}^t\} = E\left\{\begin{pmatrix} \mathbf{Y}_1 \mathbf{Y}_1^t & \mathbf{Y}_1 \mathbf{Y}_2^t \\ \mathbf{Y}_2 \mathbf{Y}_1^t & \mathbf{Y}_2 \mathbf{Y}_2^t \end{pmatrix}\right\}.$$

Considerando que

$$E\{\mathbf{Y}_1 \mathbf{Y}_1^t\} = \text{cov}\{\mathbf{T}(\mathbf{x})\},$$

y de las ecuaciones 7.38 y 7.39 que

$$E\{\mathbf{Y}_1 \mathbf{Y}_2^t\} = \mathbf{I},$$

es la matriz identidad $m \times m$ y de la ecuación 7.37 que

$$E\{\mathbf{Y}_2 \mathbf{Y}_2^t\} = \mathbf{J},$$

es la matriz de información de Fisher, se obtiene que

$$\text{cov}\{\mathbf{Z}\} = \begin{pmatrix} \text{cov}\{\mathbf{T}(\mathbf{x})\} & \mathbf{I} \\ \mathbf{I} & \mathbf{J} \end{pmatrix} \geq \mathbf{0}.$$

La propiedad de ser una matriz semidefinida positiva se debe a que se trata de una matriz de covarianza. Para continuar, se considera la identidad

$$\underbrace{\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{D}^{-1} & \mathbf{I} \end{pmatrix}}_{\mathbf{S}^T}^T \underbrace{\begin{pmatrix} \mathbf{A} & \mathbf{I} \\ \mathbf{I} & \mathbf{D} \end{pmatrix}}_{\mathbf{R}} \underbrace{\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{D}^{-1} & \mathbf{I} \end{pmatrix}}_{\mathbf{S}} = \begin{pmatrix} \mathbf{A} - \mathbf{D}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix},$$

y se cumple que si $\mathbf{R} \geq \mathbf{0}$, entonces $\mathbf{S}^T \mathbf{R} \mathbf{S} \geq \mathbf{0}$. Esto es porque $\mathbf{x}^T \mathbf{S}^T \mathbf{R} \mathbf{S} \mathbf{x} = \mathbf{y}^T \mathbf{R} \mathbf{y} \geq \mathbf{0}$. Con $\mathbf{R} = \text{cov}\{\mathbf{Z}\} \geq \mathbf{0}$ y por consiguiente, $\mathbf{A} = \text{cov}\{\mathbf{T}(\mathbf{x})\}$ y $\mathbf{D} = \mathbf{J}$, de la identidad se tiene que

$$\begin{pmatrix} \text{cov}\{\mathbf{T}(\mathbf{x})\} - \mathbf{J}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{J} \end{pmatrix} \geq \mathbf{0},$$

y como $\mathbf{J} \geq \mathbf{0}$ por ser una matriz de covarianza, se obtiene que

$$\text{cov}\{\mathbf{T}(\mathbf{x})\} - \mathbf{J}^{-1} \geq \mathbf{0}.$$

Esto se debe a que

$$(\mathbf{u}_1^t \quad \mathbf{u}_2^t) \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} = \mathbf{u}_1^t \mathbf{A} \mathbf{u}_1 + \mathbf{u}_2^t \mathbf{B} \mathbf{u}_2 \geq 0,$$

y si $\mathbf{u}_2^t \mathbf{B} \mathbf{u}_2 \geq 0$ debe ser $\mathbf{u}_1^t \mathbf{A} \mathbf{u}_1 \geq 0$. Se concluye que

$$\text{cov}\{\mathbf{T}(\mathbf{x})\} \geq \mathbf{J}^{-1}(\boldsymbol{\theta}), \quad (7.40)$$

que es lo que se quería demostrar. Como los elementos de la diagonal de una matriz semidefinida positiva son no negativos, se cumple que

$$\text{var}\{T_k(\mathbf{x})\} \geq J^{kk} \triangleq (\mathbf{J}^{-1})_{kk}, \quad k = 1, 2, \dots, m.$$

Las ecuaciones 7.36 y 7.40 pueden emplearse para evaluar el factor de degradación de la cota de Crámer-Rao de un parámetro ocasionada por la presencia de mas parámetros desconocidos. Por ejemplo, cuando θ_1 es el único parámetro desconocido, la cota de Crámer-Rao está dada por $1/J_{11}$, como indican las ecuaciones 7.19 y 7.20, mientras que está dada por J^{11} en presencia de otros parámetros desconocidos $\theta_1, \theta_2, \dots, \theta_m$. En el caso en que $m = 2$ se tiene que

$$\mathbf{J}^{-1} = \begin{pmatrix} J^{11} & J^{12} \\ J^{21} & J^{22} \end{pmatrix} = \frac{1}{J_{11}J_{22} - J_{12}^2} \begin{pmatrix} J_{22} & -J_{12} \\ -J_{12} & J_{11} \end{pmatrix},$$

y por lo tanto,

$$J^{11} = \frac{J_{22}}{J_{11}J_{22} - J_{12}^2} = \frac{1}{J_{11}} \times \frac{1}{1 - \frac{J_{12}^2}{J_{11}J_{22}}}. \quad (7.41)$$

Como \mathbf{J} es semidefinida positiva, se cumple que

$$\det(\mathbf{J}) = J_{11}J_{22} - J_{12}^2 \geq 0 \quad \Rightarrow \quad J_{11}J_{22} \geq J_{12}^2 \quad \Rightarrow \quad 0 \leq \frac{J_{12}^2}{J_{11}J_{22}} \leq 1,$$

por lo que el factor que multiplica a $1/J_{11}$ en la ecuación 7.41, que es el factor de degradación por la presencia de un parámetro desconocido adicional, es mayor que 1. En el caso general, se tiene que

$$\mathbf{J} = \begin{pmatrix} J_{11} & \mathbf{b}^T \\ \mathbf{b} & \mathbf{G} \end{pmatrix}.$$

Considerando la identidad⁷

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}, \quad (7.42)$$

se obtiene que

$$(\mathbf{J}^{-1})_{11} = J^{11} = \frac{1}{J_{11} - \mathbf{b}^T \mathbf{G}^{-1} \mathbf{b}} = \frac{1}{J_{11}} \times \frac{1}{1 - \frac{\mathbf{b}^T \mathbf{G}^{-1} \mathbf{b}}{J_{11}}}. \quad (7.43)$$

Como⁸

$$\det \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \det(\mathbf{A}) \times \det(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}) = \det(\mathbf{D}) \times \det(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}), \quad (7.44)$$

se cumple que

$$\det(\mathbf{J}) = \det(\mathbf{G})(J_{11} - \mathbf{b}^T \mathbf{G}^{-1} \mathbf{b}) \geq 0$$

por ser \mathbf{J} semidefida positiva y además $\det(\mathbf{G}) \geq 0$ por la misma razón, se cumple que

$$J_{11} - \mathbf{b}^T \mathbf{G}^{-1} \mathbf{b} \geq 0 \quad \Rightarrow \quad 0 \leq \frac{\mathbf{b}^T \mathbf{G}^{-1} \mathbf{b}}{J_{11}} \leq 1.$$

Por lo tanto, el factor que multiplica a $1/J_{11}$ en la ecuación 7.43 representa el efecto del resto de los parámetros desconocidos en la cota de θ_1 .

⁷Ver https://en.wikipedia.org/wiki/Invertible_matrix#Blockwise_inversion

⁸Ver https://en.wikipedia.org/wiki/Block_matrix#Block_matrix_determinant,

7.3.7. Mejoras a la cota de Crámer-Rao: la cota de Bhattacharya

El contenido de esta sección en [1] es una reimpresión de [20].

En el caso en que no existe un estimador eficiente, la cota de Crámer-Rao no será ajustada, es decir, la varianza del estimador UMVUE será mayor que la cota de Crámer-Rao, y por lo tanto, es interesante considerar técnicas para mejorarla. Teniendo en cuenta que la cota de Crámer-Rao solo emplea la derivada primera de la p.d.f. conjunta, es natural analizar el uso de derivadas de mayor orden, asumiendo que existen. Para comenzar, se considera la ecuación 7.13

$$\int_{-\infty}^{\infty} (T(\mathbf{x}) - \theta) \frac{\partial f(\mathbf{x}; \theta)}{\partial \theta} dx = 1,$$

que se puede expresar como

$$\int_{-\infty}^{\infty} \left[(T(\mathbf{x}) - \theta) \frac{1}{f(\mathbf{x}; \theta)} \frac{\partial f(\mathbf{x}; \theta)}{\partial \theta} \right] f(\mathbf{x}; \theta) dx = E \left[(T(\mathbf{x}) - \theta) \frac{1}{f(\mathbf{x}; \theta)} \frac{\partial f(\mathbf{x}; \theta)}{\partial \theta} \right] = 1. \quad (7.45)$$

Diferenciando 7.13, se tiene que

$$\int_{-\infty}^{\infty} (T(\mathbf{x}) - \theta) \frac{\partial^2 f(\mathbf{x}; \theta)}{\partial \theta^2} dx - \int_{-\infty}^{\infty} \frac{\partial f(\mathbf{x}; \theta)}{\partial \theta} dx = 0,$$

y como

$$\int_{-\infty}^{\infty} \frac{\partial f(\mathbf{x}; \theta)}{\partial \theta} dx = \frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} f(\mathbf{x}; \theta) dx = \frac{\partial 1}{\partial \theta} = 0, \quad (7.46)$$

resulta en

$$\int_{-\infty}^{\infty} (T(\mathbf{x}) - \theta) \frac{\partial^2 f(\mathbf{x}; \theta)}{\partial \theta^2} dx = 0.$$

En general, al diferenciar k veces la ecuación 7.13 resulta en

$$\int_{-\infty}^{\infty} (T(\mathbf{x}) - \theta) \frac{\partial^k f(\mathbf{x}; \theta)}{\partial \theta^k} dx = 0,$$

que al igual que antes, se puede expresar como

$$\int_{-\infty}^{\infty} \left[(T(\mathbf{x}) - \theta) \frac{1}{f(\mathbf{x}; \theta)} \frac{\partial^k f(\mathbf{x}; \theta)}{\partial \theta^k} \right] f(\mathbf{x}; \theta) dx = E \left[(T(\mathbf{x}) - \theta) \frac{1}{f(\mathbf{x}; \theta)} \frac{\partial^k f(\mathbf{x}; \theta)}{\partial \theta^k} \right] = 0, \quad k \geq 2. \quad (7.47)$$

En el razonamiento se aplicaron las condiciones de regularidad y se asumió que $f(\mathbf{x}; \theta)$ es diferenciable de alto orden. Motivado por las ecuaciones 7.45 y 7.47, se define el vector aleatorio de Bhattacharya como

$$\mathbf{Y}_m = \left(T(\mathbf{x}) - \theta, \frac{1}{f(\mathbf{x}; \theta)} \frac{\partial f(\mathbf{x}; \theta)}{\partial \theta}, \frac{1}{f(\mathbf{x}; \theta)} \frac{\partial^2 f(\mathbf{x}; \theta)}{\partial \theta^2}, \dots, \frac{1}{f(\mathbf{x}; \theta)} \frac{\partial^m f(\mathbf{x}; \theta)}{\partial \theta^m} \right)^t.$$

Considerando que $E\{\mathbf{Y}_m\} = 0$ (ver la ecuación 7.46) y empleando las ecuaciones 7.45 y 7.47, se obtiene que la matriz de covarianza de \mathbf{Y}_m es

$$E\{\mathbf{Y}_m \mathbf{Y}_m^t\} = \begin{pmatrix} \text{var}\{T\} & 1 & 0 & 0 & \dots & 0 \\ 1 & J_{11} & B_{12} & B_{13} & \dots & B_{1m} \\ 0 & B_{12} & B_{22} & B_{23} & \dots & B_{2m} \\ 0 & B_{13} & B_{23} & B_{33} & \dots & B_{3m} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & B_{1m} & B_{2m} & B_{3m} & \dots & B_{mm} \end{pmatrix} \quad (7.48)$$

donde

$$B_{ij} = E \left\{ \left(\frac{1}{f(\mathbf{x}; \theta)} \frac{\partial^i f(\mathbf{x}; \theta)}{\partial \theta^i} \right) \left(\frac{1}{f(\mathbf{x}; \theta)} \frac{\partial^j f(\mathbf{x}; \theta)}{\partial \theta^j} \right) \right\}.$$

En particular,

$$B_{11} = J_{11} = E \left\{ \left(\frac{1}{f(\mathbf{x}; \theta)} \frac{\partial f(\mathbf{x}; \theta)}{\partial \theta} \right)^2 \right\}$$

es la información de Fisher sobre θ contenida en la p.d.f. $f(\mathbf{x}; \theta)$. En lo que sigue, se asume que $(1/f)(\partial^k f / \partial \theta^k)$, con $k = 1, \dots, m$, son linealmente independientes, de forma que la matriz

$$\mathbf{B}(m) = \begin{pmatrix} J_{11} & B_{12} & B_{13} & \dots & B_{1m} \\ B_{12} & B_{22} & B_{23} & \dots & B_{2m} \\ B_{13} & B_{23} & B_{33} & \dots & B_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ B_{1m} & B_{2m} & B_{3m} & \dots & B_{mm} \end{pmatrix} \quad (7.49)$$

es simétrica y de rango completo, y por lo tanto, definida positiva. Empleando la identidad del determinante de una matriz en bloques dada por la ecuación 7.44 a la ecuación 7.48 con

$$\mathbf{A} = \text{var}\{T\}, \quad \mathbf{B} = [1, 0, \dots, 0] = \mathbf{C}^t \quad \text{y} \quad \mathbf{D} = \mathbf{B}(m)$$

se obtiene que

$$\begin{aligned} \det(E\{\mathbf{Y}_m \mathbf{Y}_m^t\}) &= \det[\mathbf{B}(m)][\text{var}\{T\} - \mathbf{B}\mathbf{B}^{-1}(m)\mathbf{C}] \\ &= \det[\mathbf{B}(m)][\text{var}\{T\} - B^{11}(m)] \\ &\stackrel{(a)}{\geq} 0, \end{aligned}$$

donde en (a) se consideró que $\det(E\{\mathbf{Y}_m \mathbf{Y}_m^t\}) \geq 0$ por ser $E\{\mathbf{Y}_m \mathbf{Y}_m^t\}$ semidefinida positiva por tratarse de una matriz de covarianza. Se concluye que

$$\text{var}\{T(\mathbf{x})\} \geq B^{11}(m) \triangleq [\mathbf{B}^{-1}(m)]_{11}. \quad (7.50)$$

Como se cumple que (ver la ecuación 7.43)

$$B^{11}(m) \geq \frac{1}{B_{11}(m)} = \frac{1}{J_{11}},$$

la ecuación 7.50 representa una mejora a la cota de Crámer-Rao.

Se mostrará a continuación que al incrementar m la cota de Bhattacharya no decrece. De la definición de $\mathbf{B}(m)$ de la ecuación 7.49,

$$\mathbf{B}(m+1) = \begin{pmatrix} \mathbf{B}(m) & \mathbf{b} \\ \mathbf{b}^t & B_{m+1, m+1} \end{pmatrix}$$

con

$$\mathbf{b} = (B_{1, m+1}, B_{2, m+1}, \dots, B_{m, m+1})^t.$$

Considerando la identidad matricial⁹

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix},$$

en este caso se tiene que

$$(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} = \frac{1}{B_{m+1, m+1} - \mathbf{b}^t \mathbf{B}^{-1}(m) \mathbf{b}} \triangleq q_m$$

y

$$\mathbf{A}^{-1}\mathbf{B} = (\mathbf{C}\mathbf{A}^{-1})^t = \mathbf{B}^{-1}(m)\mathbf{b} \triangleq \mathbf{c} = (c_1, c_2, \dots, c_m)^t,$$

resultando en que

$$\mathbf{B}^{-1}(m+1) = \begin{pmatrix} \mathbf{B}^{-1}(m) + q_m \mathbf{c} \mathbf{c}^t & -q_m \mathbf{c} \\ -q_m \mathbf{c}^t & q_m \end{pmatrix}.$$

⁹Ver https://en.wikipedia.org/wiki/Invertible_matrix#Blockwise_inversion

Por lo tanto,

$$B^{11}(m+1) = B^{11}(m) + c_1^2 q_m.$$

Como $q_m \geq 0$ por ser un elemento de la diagonal de una matriz semidefinida positiva, se concluye que

$$B^{11}(m+1) \geq B^{11}(m), \quad m = 1, 2, \dots,$$

lo que indica que la cota dada por la ecuación 7.50 es una secuencia no decreciente con m .

Factores de mayor orden de la mejora Para obtener $B^{11}(m)$ explícitamente en la ecuación 7.50, se recurre a la definición de $\mathbf{B}(m)$ de la ecuación 7.49. Si M_{ij} es el menor complementario (i, j) de $\mathbf{B}(m)$, se cumple que

$$B^{11}(m) \stackrel{(a)}{=} \frac{M_{11}}{\det[\mathbf{B}(m)]} \stackrel{(b)}{=} \frac{M_{11}}{J_{11}M_{11} - \sum_{k=2}^m (-1)^k B_{1k}M_{1k}} = \frac{1}{J_{11}} \left(\frac{1}{1-\epsilon} \right),$$

con

$$\epsilon = \sum_{k=2}^m \frac{(-1)^k B_{1k} M_{1k}}{J_{11} M_{11}}.$$

En (a) se empleó la regla de Crámer para calcular la inversa de una matriz¹⁰ y en (b) la expansión de Laplace¹¹ para expresar el determinante. Como $B^{11} \geq 1/J_{11}$, se cumple que $0 \leq \epsilon < 1$. Además, ϵ es una función de m y la cantidad de observaciones n . Realizando un expansión de Taylor en torno a $\epsilon = 0$, se tiene que

$$\frac{1}{1-\epsilon} = \sum_{k=0}^{\infty} \epsilon^k.$$

ya que

$$\text{si } f(\epsilon) = \frac{1}{1-\epsilon} \quad \Rightarrow \quad f^{(n)}(\epsilon) = \frac{n!}{(1-\epsilon)^{n+1}} \quad \text{y} \quad f^{(n)}(0) = n!.$$

De esta forma,

$$B^{11}(m) = \frac{1}{J_{11}} + \frac{\epsilon(n, m)}{J_{11}} + \frac{\epsilon^2(n, m)}{J_{11}} + \dots$$

Se observa que el primer término es independiente de m , y el resto de los términos consisten en la mejora en la cota de Crámer-Rao. Si $T(\mathbf{x})$ es eficiente, $\epsilon = 0$, y los términos de mayor orden están presentes solo si no existe un estimador eficiente. Es posible encontrar una expresión explícita para los términos de mayor orden para valores de m pequeños. Por ejemplo, si $m = 2$,

$$\mathbf{B}(2) = \begin{pmatrix} J_{11} & B_{12} \\ B_{12} & B_{22} \end{pmatrix} \quad \Rightarrow \quad \mathbf{B}^{-1}(2) = \frac{1}{J_{11}B_{22} - B_{12}^2} \begin{pmatrix} B_{22} & -B_{12} \\ -B_{12} & J_{11} \end{pmatrix},$$

y por lo tanto,

$$B^{11}(2) = \frac{B_{22}}{J_{11}B_{22} - B_{12}^2} = \frac{1}{J_{11}} + \frac{B_{12}^2}{J_{11}(J_{11}B_{22} - B_{12}^2)},$$

por lo que la ecuación 7.50 indica que

$$\text{var}\{T(\mathbf{x})\} \geq \frac{1}{J_{11}} + \frac{B_{12}^2}{J_{11}(J_{11}B_{22} - B_{12}^2)}.$$

Si las observaciones son independientes e idénticamente distribuidas, se cumple que

$$\begin{aligned} J_{11} &= nj_{11} \\ B_{12} &= nb_{12} \\ B_{22} &= nb_{22} + 2n(n-1)j_{11}^2, \end{aligned} \tag{7.51}$$

¹⁰Ver https://en.wikipedia.org/wiki/Cramer's_rule#Finding_inverse_matrix.

¹¹Ver https://en.wikipedia.org/wiki/Laplace_expansion.

donde $j_{11}, b_{12}, b_{22}, b_{13}, \dots$ corresponden a los valores $J_{11}, B_{12}, B_{22}, B_{13}, \dots$ en el caso de una única observación, es decir, cuando $n = 1$. Este resultado se demostrará al final del razonamiento. Por lo tanto,

$$\begin{aligned}\text{var}\{T(\mathbf{x})\} &\geq \frac{1}{nj_{11}} + \frac{nb_{12}^2}{nj_{11}[nj_{11}(nb_{22} + 2n(n-1)j_{11}^2) - nb_{12}^2]} \\ &= \frac{1}{nj_{11}} + \frac{b_{12}^2}{2n^2j_{11}^4 + nj_{11}^2(nb_{22} - 2j_{11}^2) - j_{11}b_{12}^2} \\ &= \frac{1}{nj_{11}} + \frac{b_{12}^2}{2n^2j_{11}^4} + o\left(\frac{1}{n^2}\right).\end{aligned}$$

Se obtuvo que si n es grande y las observaciones son i.i.d.,

$$\text{var}\{T(\mathbf{x})\} \geq \frac{1}{nj_{11}} + \frac{b_{12}^2}{2n^2j_{11}^4}. \quad (7.52)$$

Si un estimador $T(\mathbf{x})$ no es eficiente pero su varianza es tal que la ecuación 7.52 se cumple con igualdad, se dice que es *eficiente de segundo orden*.

Falta demostrar la propiedad dada por la ecuación 7.51. Se mostrará que si x y y son variables aleatorias independientes, se cumple que [21]

$$\begin{aligned}J_{11}(x, y) &= J_{11}(x) + J_{11}(y) \\ B_{12}(x, y) &= B_{12}(x) + B_{12}(y) \\ B_{22}(x, y) - 2J_{11}^2(x, y) &= B_{22}(x) - 2J_{11}^2(x) + B_{22}(y) - 2J_{11}^2(y).\end{aligned} \quad (7.53)$$

En la ecuación 7.21 ya se demostró que se cumple para la información de Fisher J_{11} . Se mostrará la propiedad para B_{22} , ya que la demostración para B_{12} es análoga. Por definición,

$$B_{22}(x, y) = E\left\{\left(\frac{1}{f(x, y; \theta)} \frac{\partial^2 f(x, y; \theta)}{\partial \theta^2}\right)^2\right\}.$$

Como las variables aleatorias x y y son independientes, la p.d.f conjunta cumple que $f(x, y; \theta) = f(x; \theta)f(y; \theta)$ y sus derivadas primera y segunda son

$$\frac{\partial f(x, y)}{\partial \theta} = \frac{\partial f(x)f(y)}{\partial \theta} = \frac{\partial f(x)}{\partial \theta}f(y) + f(x)\frac{\partial f(y)}{\partial \theta}$$

y

$$\frac{\partial^2 f(x, y)}{\partial \theta^2} = \frac{\partial^2 f(x)f(y)}{\partial \theta^2} = \frac{\partial^2 f(x)}{\partial \theta^2}f(y) + 2\frac{\partial f(x)}{\partial \theta}\frac{\partial f(y)}{\partial \theta} + f(x)\frac{\partial^2 f(y)}{\partial \theta^2},$$

donde se omitió la parametrización con θ por claridad. Por lo tanto,

$$\begin{aligned}B_{22}(x, y) &= E\left\{\left[\frac{1}{f(x)f(y)} \left(\frac{\partial^2 f(x)}{\partial \theta^2}f(y) + 2\frac{\partial f(x)}{\partial \theta}\frac{\partial f(y)}{\partial \theta} + f(x)\frac{\partial^2 f(y)}{\partial \theta^2}\right)\right]^2\right\} \\ &= E\left\{\frac{1}{f^2(x)f^2(y)} \left[\left(\frac{\partial^2 f(x)}{\partial \theta^2}\right)^2 f^2(y) + 4\left(\frac{\partial f(x)}{\partial \theta}\right)^2 \left(\frac{\partial f(y)}{\partial \theta}\right)^2 + f^2(x) \left(\frac{\partial^2 f(y)}{\partial \theta^2}\right)^2\right.\right. \\ &\quad \left.\left. + 4\frac{\partial f(x)}{\partial \theta}\frac{\partial^2 f(x)}{\partial \theta^2}\frac{\partial f(y)}{\partial \theta}f(y) + 4f(x)\frac{\partial f(x)}{\partial \theta}\frac{\partial f(y)}{\partial \theta}\frac{\partial^2 f(y)}{\partial \theta^2} + 2f(x)\frac{\partial^2 f(x)}{\partial \theta^2}\frac{\partial^2 f(y)}{\partial \theta^2}f(y)\right]\right\} \\ &= E\left\{\left(\frac{1}{f(x)}\frac{\partial^2 f(x)}{\partial \theta^2}\right)^2\right\} + 4E\left\{\left(\frac{1}{f(x)}\frac{\partial f(x)}{\partial \theta}\right)^2\right\}E\left\{\left(\frac{1}{f(y)}\frac{\partial f(y)}{\partial \theta}\right)^2\right\} \\ &\quad + E\left\{\left(\frac{1}{f(y)}\frac{\partial^2 f(y)}{\partial \theta^2}\right)^2\right\} + 4E\left\{\frac{1}{f^2(x)}\frac{\partial f(x)}{\partial \theta}\frac{\partial^2 f(x)}{\partial \theta^2}\right\}E\left\{\frac{1}{f(y)}\frac{\partial f(y)}{\partial \theta}\right\} \\ &\quad + 4E\left\{\frac{1}{f(x)}\frac{\partial f(x)}{\partial \theta}\right\}E\left\{\frac{1}{f^2(y)}\frac{\partial f(y)}{\partial \theta}\frac{\partial^2 f(y)}{\partial \theta^2}\right\} + 2E\left\{\frac{1}{f(x)}\frac{\partial^2 f(x)}{\partial \theta^2}\right\}\left\{\frac{1}{f(y)}\frac{\partial^2 f(y)}{\partial \theta^2}\right\},\end{aligned}$$

donde se empleó el hecho de que x y y son variables aleatorias independientes. Considerando que (ver la ecuación 7.46)

$$E \left\{ \frac{1}{f(x)} \frac{\partial^k f(x)}{\partial \theta^k} \right\} = 0,$$

los tres últimos sumandos son nulos, resultando en que

$$B_{22}(x, y) = B_{22}(x) + 4J_{11}(x)J_{11}(y) + B_{22}(y),$$

y como $J_{11}(x, y) = J_{11}(x) + J_{11}(y)$, resulta en que

$$B_{22}(x, y) - 2J_{11}^2(x, y) = B_{22}(x) - 2J_{11}^2(x) + B_{22}(y) - 2J_{11}^2(y),$$

que es lo que se quería mostrar. Aplicando la ecuación 7.53 repetidamente, se obtiene que

$$\begin{aligned} J_{11}(x_1, x_2, \dots, x_n) &= \sum_{i=1}^n J_{11}(x_i) \\ B_{12}(x_1, x_2, \dots, x_n) &= \sum_{i=1}^n B_{12}(x_i) \\ B_{22}(x_1, x_2, \dots, x_n) &= \sum_{i=1}^n B_{22}(x_i) - 2 \sum_{i=1}^n J_{11}^2(x_i) + 2 \left[\sum_{i=1}^n J_{11}(x_i) \right]^2. \end{aligned}$$

Finalmente, si además de independientes las variables aleatorias son idénticamente distribuidas, se obtiene que

$$J_{11}(x_1, x_2, \dots, x_n) = nj_{11}$$

$$B_{12}(x_1, x_2, \dots, x_n) = nb_{12}$$

$$B_{22}(x_1, x_2, \dots, x_n) = nb_{22} + 2n(n-1)j_{11}^2,$$

donde la ecuación de B_{22} proviene de que

$$B_{22} = nb_{22} - 2nj_{11}^2 + 2(nj_{11})^2 = nb_{22} + 2n(n-1)j_{11}^2.$$

Ejemplo 8-20 Sean $\mathbf{x}_i \sim N(\mu, \sigma^2)$, con $i = 1, 2, \dots, n$, variables aleatorias i.i.d., con μ y σ^2 desconocidos. Del ejemplo de la sección 7.3.2

$$\mathbf{z}_1 = \bar{\mathbf{x}}, \quad \text{y} \quad \mathbf{z}_2 = \sum_{i=1}^n \mathbf{x}_i^2$$

forman un \mathbf{z}_1 estadístico suficiente en este problema, y como los estimadores

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \hat{\mathbf{x}} = \mathbf{z}_1$$

y

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \hat{\mathbf{x}})^2 = \frac{\mathbf{z}_2 - n\mathbf{z}_1^2}{n-1}$$

son insesgados e independientes y son funciones únicamente de \mathbf{z}_1 y \mathbf{z}_2 , son los estimadores UMVUE para μ y σ^2 respectivamente. En la sección 6.1.3 se mostró que los estimadores $\hat{\mu}$ y $\hat{\sigma}^2$ son independientes (ver la ecuación 6.10), pero igualmente se mostrará de forma alternativa. Sea la matriz

$$A = \begin{bmatrix} 1 & -1 & & & & & & \\ 1 & 1 & -2 & & & & & 0 \\ 1 & 1 & 1 & -3 & & & & \\ \vdots & \vdots & \vdots & \ddots & \ddots & & & \\ 1 & 1 & 1 & \cdots & 1 & -(n-1) & & \\ 1 & 1 & 1 & \cdots & 1 & & 1 & \end{bmatrix}$$

de forma que

$$\begin{aligned}
 AA^t &= \begin{bmatrix} 1 & -1 & & & & 0 \\ 1 & 1 & -2 & & & \\ 1 & 1 & 1 & -3 & & \\ \vdots & \vdots & \vdots & \ddots & \ddots & \\ 1 & 1 & 1 & \cdots & 1 & -(n-1) \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ -1 & 1 & 1 & \cdots & 1 & 1 \\ -2 & 1 & 1 & \cdots & 1 & 1 \\ \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & & & -(n-2) & 1 & 1 \\ & & & & -(n-1) & 1 \end{bmatrix} \\
 &= \begin{bmatrix} 2 & & & & 0 & \\ 6 & & & & & \\ \ddots & & & & & \\ & i(i+1) & & & & \\ 0 & & & & n(n-1) & \\ & & & & & n \end{bmatrix}
 \end{aligned}$$

Por lo tanto,

$$\begin{aligned}
 U &= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & & 0 & & \\ & \ddots & & & & \\ & & \frac{1}{\sqrt{i(i+1)}} & & & \\ & 0 & & \frac{1}{\sqrt{n(n-1)}} & & \\ & & & & \frac{1}{\sqrt{n}} & \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & & 0 & & 0 \\ & \ddots & & & & \\ & & \frac{1}{\sqrt{i(i+1)}} & & & \\ & 0 & & \frac{1}{\sqrt{n(n-1)}} & & \\ & & & & \frac{1}{\sqrt{n}} & \end{bmatrix} A \\
 &= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & & 0 & & \\ & \ddots & & & & \\ & & \frac{1}{\sqrt{i(i+1)}} & & & \\ & 0 & & \frac{1}{\sqrt{n(n-1)}} & & \\ & & & & \frac{1}{\sqrt{n}} & \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & & 0 & & 0 \\ & \ddots & & & & \\ & & \frac{1}{\sqrt{i(i+1)}} & & & \\ & 0 & & \frac{1}{\sqrt{n(n-1)}} & & \\ & & & & \frac{1}{\sqrt{n}} & \end{bmatrix} \begin{bmatrix} 1 & -1 & & & & 0 \\ 1 & 1 & -2 & & & \\ 1 & 1 & 1 & -3 & & \\ \vdots & \vdots & \vdots & \ddots & \ddots & \\ 1 & 1 & 1 & \cdots & 1 & -i \\ \vdots & \vdots & \vdots & & \ddots & \ddots \\ 1 & 1 & 1 & \cdots & \cdots & 1 & -(n-1) \\ 1 & 1 & 1 & \cdots & \cdots & 1 & 1 \end{bmatrix} \\
 &= \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & & 0 & & \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & & & \\ \vdots & \vdots & \ddots & & & \\ \frac{1}{\sqrt{i(i+1)}} & \frac{1}{\sqrt{i(i+1)}} & \cdots & \frac{1}{\sqrt{i(i+1)}} & -\frac{i}{\sqrt{i(i+1)}} & \\ \vdots & \vdots & & & \ddots & \ddots \\ \frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & \cdots & \cdots & \cdots & \frac{1}{\sqrt{(n-1)n}} -\frac{n-1}{\sqrt{(n-1)n}} \end{bmatrix}
 \end{aligned}$$

es una matriz ortonormal. Efectivamente,

$$\begin{aligned}
U_{ij} &= \left[\frac{1}{\sqrt{i(i+1)}} \cdots \frac{1}{\sqrt{i(i+1)}} - \frac{i}{\sqrt{i(i+1)}} 0 \cdots 0 \right] \begin{bmatrix} \frac{1}{\sqrt{j(j+1)}} \\ \vdots \\ -\frac{1}{\sqrt{j(j+1)}} \\ -\frac{j}{\sqrt{j(j+1)}} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{array}{l} \leftarrow j \\ \leftarrow j+1 \end{array} \\
&= \begin{cases} j \frac{1}{\sqrt{i(i+1)}} \frac{1}{\sqrt{j(j+1)}} - \frac{1}{\sqrt{i(i+1)}} \frac{j}{\sqrt{j(j+1)}} = 0, & i \neq j \quad (i > j \text{ en este caso.}) \\ j \frac{1}{j(j+1)} + \frac{j^2}{j(j+1)} = \frac{j(j+1)}{j(j+1)} = 1, & i = j. \end{cases}
\end{aligned}$$

Es fácil verificar que el resultado se mantiene en el caso especial en que i o j valen n . Se concluye que $UU^t = I$. Sea $\mathbf{x} = [\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_n]^t$, y se define

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = U\mathbf{x} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & & & & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & & & \\ \vdots & \vdots & \ddots & \ddots & & \\ \frac{1}{\sqrt{i(i+1)}} & \frac{1}{\sqrt{i(i+1)}} & \cdots & \frac{1}{\sqrt{i(i+1)}} & -\frac{i}{\sqrt{i(i+1)}} & \\ \vdots & \vdots & & & \ddots & \\ \frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & \cdots & \cdots & \cdots & \frac{1}{\sqrt{(n-1)n}} -\frac{n-1}{\sqrt{(n-1)n}} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \cdots & \cdots & \frac{1}{\sqrt{n}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_i \\ \vdots \\ \mathbf{x}_{n-1} \\ \mathbf{x}_n \end{bmatrix},$$

es decir,

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = U\mathbf{x} = \begin{bmatrix} \frac{\mathbf{x}_1 - \mathbf{x}_2}{\sqrt{2}} \\ \frac{\mathbf{x}_1 + \mathbf{x}_2 - 2\mathbf{x}_3}{\sqrt{6}} \\ \vdots \\ \frac{\mathbf{x}_1 + \dots + \mathbf{x}_i - i\mathbf{x}_{i+1}}{\sqrt{i(i+1)}} \\ \vdots \\ \frac{\mathbf{x}_1 + \dots + \mathbf{x}_{n-1} - (n-1)\mathbf{x}_n}{\sqrt{(n-1)n}} \\ \frac{\mathbf{x}_1 + \dots + \mathbf{x}_n}{\sqrt{n}} \end{bmatrix}.$$

Observar que

$$\mathbf{y}_i = \frac{1}{\sqrt{i(i+1)}} \left(\sum_{j=1}^i \mathbf{x}_j - i\mathbf{x}_{i+1} \right), \quad i = 1, 2, \dots, n-1,$$

y

$$\mathbf{y}_n = \sqrt{n}\bar{\mathbf{x}} \sim N(\sqrt{n}\mu, \sigma^2). \quad (7.54)$$

Además, para $i = 1, 2, \dots, n-1$,

$$E\{\mathbf{y}_i\} = \frac{1}{\sqrt{i(i+1)}} \left(\sum_{j=1}^i E\{\mathbf{x}_j\} - iE\{\mathbf{x}_{i+1}\} \right) = \frac{iE\{\mathbf{x}_j\} - iE\{\mathbf{x}_i\}}{\sqrt{i(i+1)}} = \frac{i\mu - i\mu}{\sqrt{i(i+1)}} = 0,$$

y

$$\text{var}\{\mathbf{y}_i\} = \frac{1}{i(i+1)} \left(\sum_{j=1}^i \text{var}\{\mathbf{x}_j\} + i^2 \text{var}\{\mathbf{x}_{i+1}\} \right) = \frac{i \text{var}\{\mathbf{x}_j\} + i^2 \text{var}\{\mathbf{x}_{i+1}\}}{i(i+1)} = \frac{i(i+1)}{i(i+1)} \sigma^2 = \sigma^2,$$

resultando en que $\mathbf{y}_i \sim N(0, \sigma^2)$. Teniendo esto en cuenta,

$$E\{\mathbf{y}\mathbf{y}^t\} = UE\{\mathbf{x}\mathbf{x}^t\}U^t = U(\sigma^2 I)U^t = \sigma^2 I,$$

lo que indica que $(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)$ son variables aleatorias no correlacionadas, y como son gaussianas por ser una combinación lineal de variables aleatorias gaussianas, se concluye que son variables aleatorias independientes. Por otro lado,

$$\sum_{i=1}^n \mathbf{y}_i^2 = \mathbf{y}^t \mathbf{y} = \mathbf{x}^t U^t U \mathbf{x} = \mathbf{x}^t \mathbf{x} = \sum_{i=1}^n \mathbf{x}_i^2 \quad (7.55)$$

y por lo tanto,

$$(n-1)\hat{\sigma}^2 = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2 = \sum_{i=1}^n \mathbf{x}_i^2 - n\bar{\mathbf{x}}^2 \stackrel{(a)}{=} \sum_{i=1}^n \mathbf{y}_i^2 - \mathbf{y}_n^2 = \sum_{i=1}^{n-1} \mathbf{y}_i^2,$$

donde en (a) se emplearon los resultados de las ecuaciones 7.54 y 7.55. Se concluye que

$$\hat{\mu} = \bar{\mathbf{x}} = \frac{1}{\sqrt{n}} \mathbf{y}_n \quad \text{y} \quad \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} \mathbf{y}_i^2 \quad (7.56)$$

son variables aleatorias independientes debido a que las variables aleatorias \mathbf{y}_i son independientes.

Para determinar si los estimadores son eficientes, se calculará su varianza para compararlos con la cota de Crámer-Rao. De las ecuaciones 7.54 y 7.56, la varianza de $\hat{\mu}$ es

$$\text{var}\{\hat{\mu}\} = \frac{1}{n} \text{var}\{\mathbf{y}_n\} = \frac{\sigma^2}{n}. \quad (7.57)$$

Partiendo de la ecuación 7.56 la varianza de $\hat{\sigma}^2$ es

$$\begin{aligned} \text{var}\{\hat{\sigma}^2\} &\stackrel{(a)}{=} \frac{1}{(n-1)^2} \sum_{i=1}^{n-1} \text{var}\{\mathbf{y}_i^2\} \\ &= \frac{1}{n-1} \text{var}\{\mathbf{y}_i^2\} \\ &= \frac{1}{n-1} (E\{\mathbf{y}_i^4\} - E^2\{\mathbf{y}_i^2\}) \\ &\stackrel{(b)}{=} \frac{1}{n-1} (3\sigma^4 - \sigma^4), \end{aligned}$$

donde en (a) se tuvo en cuenta que las variables aleatorias \mathbf{y}_i son independientes y en (b) que $\mathbf{y}_i \sim N(0, \sigma^2)$ para $i = 1, \dots, n-1$. Esto resulta en

$$\text{var}\{\hat{\sigma}^2\} = \frac{2\sigma^4}{n-1}. \quad (7.58)$$

Observar que el resultado coincide con el de la ecuación 6.12 teniendo en cuenta que en este caso $\mu_4 = \sigma^4$ por tratarse de variables aleatorias normales. Para determinar la cota de Crámer-Rao, sea el parámetro vectorial $\boldsymbol{\theta} = [\mu \ \sigma^2]^t$ y defínase

$$\mathbf{Z}_k \triangleq \begin{bmatrix} \frac{1}{f} \frac{\partial^k f}{\partial \mu^k} \\ \frac{1}{f} \frac{\partial^k f}{\partial (\sigma^2)^k} \end{bmatrix}.$$

De esta forma, la matriz de información de Fisher está dada por

$$J = E\{\mathbf{Z}_1 \mathbf{Z}_1^t\}.$$

Para calcular \mathbf{Z}_1 se considera que la p.d.f. conjunta es

$$f(\mathbf{x}; \theta) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right].$$

Diferenciando respecto a μ se tiene que

$$\frac{\partial f}{\partial \mu} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right] \left[\frac{1}{\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \mu) \right] = f(\mathbf{x}; \theta) \left[\frac{1}{\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \mu) \right]. \quad (7.59)$$

resultando en que

$$\frac{1}{f} \frac{\partial f}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \mu). \quad (7.60)$$

Diferenciando respecto a σ^2 se tiene que

$$\begin{aligned} \frac{\partial f}{\partial \sigma^2} &= \frac{1}{(2\pi)^{\frac{n}{2}}} \left\{ -\frac{n}{2(\sigma^2)^{\frac{n}{2}} \sigma^2} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right] \right. \\ &\quad \left. + \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right] \frac{1}{2\sigma^4} \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right\} \\ &= f(\mathbf{x}; \theta) \left[-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]. \end{aligned} \quad (7.61)$$

resultando en que

$$\frac{1}{f} \frac{\partial f}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (\mathbf{x}_i - \mu_i)^2. \quad (7.62)$$

En estos cálculos no se empleó la propiedad de la ecuación 7.14 porque mas adelante se calcularán las derivadas segundas, por lo que se necesita la expresión analítica explícita de las derivadas primeras. Por lo tanto,

$$\mathbf{Z}_1 = \begin{bmatrix} \frac{1}{f} \frac{\partial f}{\partial \mu} \\ \frac{1}{f} \frac{\partial f}{\partial (\sigma^2)} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \mu) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \end{bmatrix}.$$

De esta forma,

$$\begin{aligned} J_{11} &= E \left\{ \left[\frac{1}{f} \frac{\partial f}{\partial \mu} \right]^2 \right\} \\ &= \frac{1}{\sigma^4} E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \right\} \\ &\stackrel{(a)}{=} \frac{1}{\sigma^4} [nE\{(\mathbf{x}_i - \mu)^2\} + n(n-1)E^2\{\mathbf{x}_i - \mu\}] \\ &\stackrel{(b)}{=} \frac{n}{\sigma^2}, \end{aligned}$$

donde en (a) se consideró que

$$E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \right\} = E \left\{ \sum_{i=1}^n (\mathbf{x}_i - \mu) \sum_{j=1}^n (\mathbf{x}_j - \mu) \right\} = nE\{(\mathbf{x}_i - \mu)^2\} + n(n-1)E\{(\mathbf{x}_i - \mu)(\mathbf{x}_j - \mu)\},$$

ya que al multiplicar las sumatorias, hay n sumandos donde el producto cumple que $i = j$ y $n(n - 1)$ sumandos donde el producto cumple que $i \neq j$, y en (b) se tuvo en cuenta que $E\{\mathbf{x}_i - \mu\} = 0$. Además,

$$\begin{aligned}
J_{12} = J_{21} &= E \left\{ \left[\frac{1}{f} \frac{\partial f}{\partial \mu} \right] \left[\frac{1}{f} \frac{\partial f}{\partial \sigma^2} \right] \right\} \\
&= E \left\{ \left[\frac{1}{\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \mu) \right] \left[-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right] \right\} \\
&= -\frac{n}{2\sigma^4} E \left\{ \sum_{i=1}^n (\mathbf{x}_i - \mu) \right\} + \frac{1}{2\sigma^4} E \left\{ \sum_{i=1}^n (\mathbf{x}_i - \mu) \sum_{j=1}^n (\mathbf{x}_j - \mu)^2 \right\} \\
&\stackrel{(a)}{=} -\frac{n^2}{2\sigma^4} E\{\mathbf{x}_i - \mu\} + \frac{1}{2\sigma^4} [nE\{(\mathbf{x}_i - \mu)^3\} + n(n-1)E\{\mathbf{x}_i - \mu\}E\{(\mathbf{x}_j - \mu)^2\}] \\
&\stackrel{(b)}{=} -\frac{n^2}{2\sigma^4} \mu_1 + \frac{1}{2\sigma^4} [n\mu_3 + n(n-1)\mu_1\mu_2] \\
&\stackrel{(c)}{=} 0,
\end{aligned}$$

donde en (a) se consideró nuevamente que al multiplicar las sumatorias hay n sumandos donde el producto cumple que $i = j$ y $n(n - 1)$ sumandos donde el producto cumple que $i \neq j$, en (b) $\mu_k = E\{(\mathbf{x}_i - \mu)^k\}$ denota el momento central k -ésimo de la variable aleatoria \mathbf{x}_i (ver la sección 4.4.1) y en (c) que los momentos centrales impares de una variable aleatoria gaussiana son nulos, como indica la ecuación 4.16. Finalmente,

$$\begin{aligned}
J_{22} &= E \left\{ \left[\frac{1}{f} \frac{\partial f}{\partial \sigma^2} \right]^2 \right\} \\
&= E \left\{ \left[-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 \right\} \\
&= E \left\{ \frac{n^2}{4\sigma^4} - \frac{n}{2\sigma^6} \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 + \frac{1}{4\sigma^8} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 \right\} \\
&= \frac{1}{2\sigma^4} \left[\frac{n^2}{2} - \frac{n}{\sigma^2} \sum_{i=1}^n E\{(\mathbf{x}_i - \mu)^2\} + \frac{1}{2\sigma^4} E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 \right\} \right] \\
&\stackrel{(a)}{=} \frac{1}{2\sigma^4} \left\{ \frac{n^2}{2} - \frac{n}{\sigma^2} nE\{(\mathbf{x}_i - \mu)^2\} + \frac{1}{2\sigma^4} [nE\{(\mathbf{x}_i - \mu)^4\} + n(n-1)E\{(\mathbf{x}_i - \mu)^2\}E\{(\mathbf{x}_j - \mu)^2\}] \right\} \\
&= \frac{1}{2\sigma^4} \left\{ \frac{n^2}{2} - \frac{n^2}{\sigma^2} \mu_2 + \frac{1}{2\sigma^4} [n\mu_4 + n(n-1)\mu_2^2] \right\} \\
&\stackrel{(b)}{=} \frac{1}{2\sigma^4} \left\{ \frac{n^2}{2} - \frac{n^2}{\sigma^2} \sigma^2 + \frac{1}{2\sigma^4} [3n\sigma^4 + n(n-1)\sigma^4] \right\} \\
&= \frac{1}{2\sigma^4} \left(\frac{n^2}{2} - n^2 + \frac{n^2 + 2n}{2} \right) \\
&= \frac{n}{2\sigma^4},
\end{aligned}$$

donde en (a) se emplearon la mismas consideraciones que antes para calcular la esperanza del cuadrado de la sumatoria y en (b) se tuvo en cuenta que $\mu_2 = \sigma^2$ y $\mu_4 = 3\sigma^4$, como indican las ecuaciones 4.16 y 4.19. Por lo tanto, la matriz de información de Fisher es

$$J = E\{\mathbf{Z}_1 \mathbf{Z}_1^t\} = \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & n/2\sigma^4 \end{bmatrix} \quad \Rightarrow \quad J^{-1} = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{bmatrix}.$$

De la ecuación 7.57, se observa que $\text{var}\{\hat{\mu}\} = \sigma^2/n = J^{11}$, mientras que de la ecuación 7.58 se ve que $\text{var}\{\hat{\sigma}^2\} = 2\sigma^4/(n-1) > 2\sigma^2/n = J^{22}$. Se concluye que el estimador $\hat{\mu}$ es eficiente mientras que

el estimador $\hat{\sigma}^2$ no es eficiente. Sin embargo, como $\hat{\sigma}^2$ es una función únicamente de los estadísticos suficientes \mathbf{z}_1 y \mathbf{z}_2 , se estudiará si es eficiente de segundo orden. Para hacerlo, hay que calcular la matriz de Bhattacharya, dada por la ecuación 7.49, para $m = 2$,

$$\mathbf{B}(2) = E \left\{ \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix} \begin{bmatrix} \mathbf{Z}_1^t & \mathbf{Z}_2^t \end{bmatrix} \right\}, \quad \text{donde} \quad \mathbf{Z}_2 = \begin{bmatrix} \frac{1}{f} \frac{\partial^2 f}{\partial \mu^2} \\ \frac{1}{f} \frac{\partial^2 f}{\partial (\sigma^2)^2} \end{bmatrix}.$$

La derivada segunda respecto a μ se obtiene diferenciando la ecuación 7.59 respecto a μ ,

$$\begin{aligned} \frac{\partial^2 f}{\partial \mu^2} &= \frac{\partial f(\mathbf{x}; \theta)}{\partial \mu} \left[\frac{1}{\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \mu) \right] - f(\mathbf{x}; \theta) \frac{1}{\sigma^2} \sum_{i=1}^n 1 \\ &= f(\mathbf{x}; \theta) \left[\frac{1}{\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 - f(\mathbf{x}; \theta) \frac{n}{\sigma^2}, \end{aligned}$$

y por lo tanto,

$$\frac{1}{f} \frac{\partial^2 f}{\partial \mu^2} = -\frac{n}{\sigma^2} + \frac{1}{\sigma^4} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2. \quad (7.63)$$

Además, diferenciando la ecuación 7.61, se obtiene que la derivada segunda respecto a σ^2 es

$$\begin{aligned} \frac{\partial^2 f}{\partial \sigma^2} &= \frac{\partial f(\mathbf{x}; \theta)}{\partial \sigma^2} \left[-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right] + f(\mathbf{x}; \theta) \left[\frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right] \\ &= f(\mathbf{x}; \theta) \left[-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 + f(\mathbf{x}; \theta) \left[\frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]. \end{aligned}$$

Por lo tanto,

$$\begin{aligned} \frac{1}{f} \frac{\partial^2 f}{\partial \sigma^2} &= \left\{ \frac{n^2}{4\sigma^4} - \frac{n}{2\sigma^6} \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 + \frac{1}{4\sigma^8} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \right\} + \left\{ \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right\} \\ &= \frac{1}{\sigma^4} \left\{ \left(\frac{n^2}{4} + \frac{n}{2} \right) - \frac{1}{\sigma^2} \left(\frac{n}{2} + 1 \right) \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 + \frac{1}{4\sigma^4} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \right\} \\ &= \frac{1}{\sigma^4} \left\{ \frac{1}{4} n(n+2) - \frac{1}{2\sigma^2} (n+2) \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 + \frac{1}{4\sigma^4} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \right\}, \end{aligned}$$

resultando en que

$$\frac{1}{f} \frac{\partial^2 f}{\partial \sigma^2} = \frac{1}{2\sigma^4} \left\{ \frac{1}{2} n(n+2) - \frac{1}{\sigma^2} (n+2) \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 + \frac{1}{2\sigma^4} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \right\}. \quad (7.64)$$

Previo al cálculo de los elementos de la matriz $\mathbf{B}(2)$ se consideran los siguientes resultados, que se

necesitarán:

$$E \left\{ \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right\} = n\sigma^2 \quad (7.65)$$

$$E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 \right\} = n(n+2)\sigma^4 \quad (7.66)$$

$$E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^3 \right\} = n(n+2)(n+4)\sigma^6 \quad (7.67)$$

$$E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^4 \right\} = n(n+2)(n+4)(n+6)\sigma^8. \quad (7.68)$$

Además

$$E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \right\} = n\sigma^2 \quad (7.69)$$

$$E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^3 \right\} = 0 \quad (7.70)$$

$$E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^4 \right\} = 3n^2\sigma^4 \quad (7.71)$$

$$E \left\{ \sum_{i=1}^n (\mathbf{x}_i - \mu) \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right\} = 0 \quad (7.72)$$

$$E \left\{ \sum_{i=1}^n (\mathbf{x}_i - \mu) \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 \right\} = 0 \quad (7.73)$$

$$E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right\} = n(n+2)\sigma^4 \quad (7.74)$$

$$E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 \right\} = n(n+2)(n+4)\sigma^6. \quad (7.75)$$

Es fácil verificar que se cumple la ecuación 7.65 y el resultado ya se empleó previamente en el cálculo de la matriz de información de Fisher. El resultado de la ecuación 7.66 también se empleó antes, y proviene de que

$$\begin{aligned} E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 \right\} &\stackrel{(a)}{=} nE\{(\mathbf{x}_i - \mu)^4\} + n(n-1)E^2\{(\mathbf{x}_i - \mu)^2\} \\ &= n\mu_4 + n(n-1)\mu_2^2 \\ &\stackrel{(b)}{=} 3n\sigma^4 + n(n-1)\sigma^4 \\ &= n(n+2)\sigma^4, \end{aligned}$$

donde en (a) se observó que al elevar al cuadrado la sumatoria hay n sumandos de la forma $(\mathbf{x}_i - \mu)^4$ y $n(n-1)$ sumandos de la forma $(\mathbf{x}_i - \mu)^2(\mathbf{x}_j - \mu)^2$ con $i \neq j$ y en (b) se consideró que $\mu_4 = 3\sigma^4$ y

$\mu_2 = \sigma^2$, como indica la ecuación 4.16. Para mostrar la ecuación 7.67, se observa que

$$\begin{aligned}
E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^3 \right\} &\stackrel{(a)}{=} n\mu_6 + 3n(n-1)\mu_4\mu_2 + n(n-1)(n-2)\mu_2^3 \\
&= [15n + 9n(n-1) + n(n-1)(n-2)]\sigma^6 \\
&= n [15 + (9n-9) + (n^2-3n+2)]\sigma^6 \\
&= n (n^2 + 6n + 8)\sigma^6 \\
&= n(n+2)(n+4),
\end{aligned}$$

donde en (a) se consideró que al elevar la sumatoria al cubo hay n sumandos de la forma $(\mathbf{x}_i - \mu)^6$, $3n(n-1)$ sumandos de la forma $(\mathbf{x}_i - \mu)^2(\mathbf{x}_j - \mu)$ con $i \neq j$, es decir, con dos índices iguales y uno distinto, y $n(n-1)(n-2)$ con los tres índices distintos de la forma $(\mathbf{x}_i - \mu)(\mathbf{x}_j - \mu)(\mathbf{x}_k - \mu)$. Para deducir la ecuación 7.68, se emplea el resultado

$$\begin{aligned}
E \left\{ \left(\sum_{i=1}^n \mathbf{x}_i \right)^4 \right\} &= nm_4 + 4n(n-1)m_3m_1 + 3n(n-1)m_2^2 \\
&\quad + 6n(n-1)(n-2)m_2m_1^2 + n(n-1)(n-2)(n-3)m_1^4,
\end{aligned} \tag{7.76}$$

que es válido para variables aleatorias i.i.d., donde m_k es el momento k -ésimo de \mathbf{x}_i . Este resultado se demuestra en el apéndice C de [22]. Aplicando el resultado de la ecuación 7.76 a la ecuación 7.68 cambiando \mathbf{x}_i por $(\mathbf{x}_i - \mu)^2$ y por lo tanto m_k por μ_{2k} , donde μ_k es el momento central k -ésimo de las variables aleatorias gaussianas \mathbf{x}_i , se tiene que

$$\begin{aligned}
E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^4 \right\} &= n\mu_8 + 4n(n-1)\mu_6\mu_2 + 3n(n-1)\mu_4^2 \\
&\quad + 6n(n-1)(n-2)\mu_4\mu_2^2 + n(n-1)(n-2)(n-3)\mu_2^4 \\
&\stackrel{(a)}{=} [105n + 60n(n-1) + 27n(n-1) \\
&\quad + 18n(n-1)(n-2) + n(n-1)(n-2)(n-3)]\sigma^8 \\
&= n \{ 105 + (n-1) [60 + 27 + 18(n-2) + (n-2)(n-3)] \} \\
&= n \{ 105 + (n-1) [87 + (18n-36) + (n^2-5n+6)] \} \\
&= n [105 + (n-1)(n^2+13n+57)] \\
&= n (n^3 + 12n^2 + 44n + 48) \\
&= n(n+2)(n+4)(n+6),
\end{aligned}$$

donde en (a) se tuvo en cuenta que $\mu_6 = 3 \times 5\sigma^6 = 15\sigma^6$ y $\mu_8 = 3 \times 5 \times 7\sigma^8 = 105\sigma^8$ (ver la ecuación 4.16). Para mostrar la ecuación 7.70, se observa que

$$\begin{aligned}
E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^3 \right\} &\stackrel{(a)}{=} n\mu_3 + 3n(n-1)\mu_1\mu_2 + n(n-1)(n-2)\mu_1^3 \\
&\stackrel{(b)}{=} 0,
\end{aligned}$$

donde en (a) se consideró que al elevar la sumatoria al cubo hay n sumandos de la forma $(\mathbf{x}_i - \mu)^3$ con los tres índices iguales, $3n(n-1)$ sumandos de la forma $(\mathbf{x}_i - \mu)^2(\mathbf{x}_j - \mu)$ con dos índices iguales y uno distinto y $n(n-1)(n-2)$ con los tres índices distintos, y en (b) se tuvo en cuenta que los momentos centrales impares son nulos, es decir, $\mu_1 = \mu_3 = 0$, como indica la ecuación 4.16. Las ecuaciones 7.69, 7.72 y 7.73 se obtienen realizando un razonamiento análogo. El resultado de la ecuación 7.71 se obtiene directamente de la ecuación 7.76 reemplazando los momentos m_k por los momentos centrales μ_k y teniendo en cuenta

que los momentos centrales son nulos. La ecuación 7.74 proviene de que

$$\begin{aligned}
 E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right\} &= n\mu_4 + n(n-1)\mu_2^2 + 2n(n-1)\mu_1\mu_3 + n(n-1)(n-2)\mu_1^2\mu_2 \\
 &\stackrel{(a)}{=} [3n + n(n-1)] \sigma^4 \\
 &= n(n+2)\sigma^4,
 \end{aligned}$$

donde en (a) nuevamente se empleó el resultado de los momentos centrales de una variable aleatoria gaussiana de la ecuación 4.16. Finalmente, para mostrar 7.75, se observa que al desarrollar las sumatorias se obtienen los sumandos con las características mostradas en la siguiente tabla:

Sumando	Esperanza	Número de sumandos
$(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^2(\mathbf{x}_i - \mu)^2$	μ_6	n
$(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)(\mathbf{x}_j - \mu)^2(\mathbf{x}_j - \mu)^2$	$\mu_2\mu_4$	$n(n-1)$
$(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^2(\mathbf{x}_j - \mu)^2$	$\mu_4\mu_2$	$2n(n-1)$
$(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)(\mathbf{x}_j - \mu)^2(\mathbf{x}_k - \mu)^2$	μ_2^3	$n(n-1)(n-2)$
$(\mathbf{x}_i - \mu)(\mathbf{x}_j - \mu)(\mathbf{x}_k - \mu)^2(\mathbf{x}_l - \mu)^2$	$\mu_1^2\mu_2^2$	$n(n-1)(n-2)(n-3)$
$(\mathbf{x}_i - \mu)(\mathbf{x}_j - \mu)(\mathbf{x}_i - \mu)^2(\mathbf{x}_i - \mu)^2$	$\mu_5\mu_1$	$2n(n-1)$
$(\mathbf{x}_i - \mu)(\mathbf{x}_j - \mu)(\mathbf{x}_i - \mu)^2(\mathbf{x}_j - \mu)^2$	$\mu_3\mu_3$	$2n(n-1)$
$(\mathbf{x}_i - \mu)(\mathbf{x}_j - \mu)(\mathbf{x}_i - \mu)^2(\mathbf{x}_k - \mu)^2$	$\mu_3\mu_1\mu_2$	$4n(n-1)(n-2)$
$(\mathbf{x}_i - \mu)(\mathbf{x}_j - \mu)(\mathbf{x}_k - \mu)^2(\mathbf{x}_k - \mu)^2$	$\mu_1^2\mu_4$	$n(n-1)(n-2)$

Teniendo en cuenta que los momentos centrales impares son nulos, la esperanza queda,

$$\begin{aligned}
 E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 \right\} &\stackrel{(a)}{=} n\mu_6 + n(n-1)\mu_2\mu_4 + 2n(n+1)\mu_4\mu_2 + n(n-1)(n-2)\mu_2^3 \\
 &= n[15 + 9(n-1) + (n-1)(n-2)]\sigma^6 \\
 &= n(15 + 9n - 9 + n^2 - 3n + 2)\sigma^6 \\
 &= n(n^2 + 6n + 8)\sigma^6 \\
 &= n(n+2)(n+4)\sigma^6.
 \end{aligned}$$

Retornando al cálculo de la matriz $\mathbf{B}(2)$, se observa que el bloque 2×2 en la posición (1, 1) es la matriz de información de Fisher y ya fue calculada. Se calcularán los valores del resto de los elementos. Comenzando,

$$\begin{aligned}
 [\mathbf{B}(2)]_{13} = [\mathbf{B}(2)]_{31} &= E \left\{ \left(\frac{1}{f} \frac{\partial f}{\partial \mu} \right) \left(\frac{1}{f} \frac{\partial^2 f}{\partial \mu^2} \right) \right\} \\
 &\stackrel{(a)}{=} E \left\{ \left(\frac{1}{\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \mu) \right) \left(-\frac{n}{\sigma^2} + \frac{1}{\sigma^4} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \right) \right\} \\
 &= -\frac{n}{\sigma^4} \sum_{i=1}^n E \{(\mathbf{x}_i - \mu)\} + \frac{1}{\sigma^6} E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^3 \right\} \\
 &\stackrel{(b)}{=} 0
 \end{aligned}$$

donde las derivadas en (a) están dadas por las ecuaciones 7.60 y 7.63 y en (b) se empleó el resultado de la ecuación 7.70. Luego,

$$[\mathbf{B}(2)]_{14} = [\mathbf{B}(2)]_{41} = E \left\{ \left(\frac{1}{f} \frac{\partial f}{\partial \mu} \right) \left(\frac{1}{f} \frac{\partial^2 f}{\partial \sigma^2} \right) \right\}$$

De las ecuaciones 7.60 y 7.64 se tiene que

$$\begin{aligned} \left(\frac{1}{f} \frac{\partial f}{\partial \mu} \right) \left(\frac{1}{f} \frac{\partial^2 f}{\partial \sigma^2} \right) &= \frac{1}{2\sigma^6} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right] \left\{ \frac{1}{2} n(n+2) - \frac{1}{\sigma^2} (n+2) \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 + \frac{1}{2\sigma^4} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \right\} \\ &= \frac{1}{2\sigma^6} \left\{ \frac{1}{2} n(n+2) \sum_{i=1}^n (\mathbf{x}_i - \mu) - \frac{1}{\sigma^2} (n+2) \sum_{i=1}^n (\mathbf{x}_i - \mu) \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right. \\ &\quad \left. + \frac{1}{2\sigma^4} \sum_{i=1}^n (\mathbf{x}_i - \mu) \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 \right\}, \end{aligned}$$

y al tomar esperanza, de las ecuaciones 7.72 y 7.73, se obtiene que

$$[\mathbf{B}(2)]_{14} = [\mathbf{B}(2)]_{41} = E \left\{ \left(\frac{1}{f} \frac{\partial f}{\partial \mu} \right) \left(\frac{1}{f} \frac{\partial^2 f}{\partial \sigma^2} \right) \right\} = 0.$$

Continuando,

$$[\mathbf{B}(2)]_{23} = [\mathbf{B}(2)]_{32} = E \left\{ \left(\frac{1}{f} \frac{\partial f}{\partial \sigma^2} \right) \left(\frac{1}{f} \frac{\partial^2 f}{\partial \mu^2} \right) \right\}$$

y de las ecuaciones 7.62 y 7.63,

$$\begin{aligned} \left(\frac{1}{f} \frac{\partial f}{\partial \sigma^2} \right) \left(\frac{1}{f} \frac{\partial^2 f}{\partial \mu^2} \right) &= \left(-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (\mathbf{x}_i - \mu_i)^2 \right) \left(-\frac{n}{\sigma^2} + \frac{1}{\sigma^4} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \right) \\ &= \frac{n^2}{2\sigma^4} - \frac{n}{2\sigma^6} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 - \frac{n}{2\sigma^6} \sum_{i=1}^n (\mathbf{x}_i - \mu_i)^2 + \frac{1}{2\sigma^8} \sum_{i=1}^n (\mathbf{x}_i - \mu_i)^2 \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \end{aligned}$$

y al tomar esperanza, se tiene que

$$\begin{aligned} E \left\{ \left(\frac{1}{f} \frac{\partial f}{\partial \sigma^2} \right) \left(\frac{1}{f} \frac{\partial^2 f}{\partial \mu^2} \right) \right\} &= \frac{1}{2\sigma^4} \left[n^2 - \frac{n}{\sigma^2} E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \right\} - \frac{n}{\sigma^2} \sum_{i=1}^n E \{ (\mathbf{x}_i - \mu_i)^2 \} \right. \\ &\quad \left. + \frac{1}{\sigma^4} E \left\{ \sum_{i=1}^n (\mathbf{x}_i - \mu_i)^2 \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \right\} \right] \\ &\stackrel{(a)}{=} \frac{1}{2\sigma^4} [n^2 - n^2 - n^2 + n(n+2)] \\ &= \frac{n}{\sigma^4}, \end{aligned}$$

donde en (a) se empleó el resultado de la ecuación 7.74. Se obtuvo que

$$[\mathbf{B}(2)]_{23} = [\mathbf{B}(2)]_{32} = E \left\{ \left(\frac{1}{f} \frac{\partial f}{\partial \sigma^2} \right) \left(\frac{1}{f} \frac{\partial^2 f}{\partial \mu^2} \right) \right\} = \frac{n}{\sigma^4}.$$

Además,

$$[\mathbf{B}(2)]_{24} = [\mathbf{B}(2)]_{42} = E \left\{ \left(\frac{1}{f} \frac{\partial f}{\partial \sigma^2} \right) \left(\frac{1}{f} \frac{\partial^2 f}{\partial \sigma^2} \right) \right\},$$

y de las ecuaciones 7.62 y 7.64,

$$\begin{aligned}
\left(\frac{1}{f} \frac{\partial f}{\partial \sigma^2}\right) \left(\frac{1}{f} \frac{\partial^2 f}{\partial \sigma^2}\right) &= \frac{1}{4\sigma^6} \left(-n + \frac{1}{\sigma^2} \sum_{i=1}^n (\mathbf{x}_i - \mu_i)^2 \right) \\
&\times \left(\frac{1}{2}n(n+2) - \frac{1}{\sigma^2}(n+2) \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 + \frac{1}{2\sigma^4} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 \right) \\
&= \frac{1}{4\sigma^6} \left\{ -\frac{1}{2}n^2(n+2) + \frac{1}{\sigma^2}n(n+2) \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 - \frac{1}{2\sigma^4}n \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 \right. \\
&\quad \left. + \frac{1}{2\sigma^2}n(n+2) \sum_{i=1}^n (\mathbf{x}_i - \mu_i)^2 - \frac{1}{\sigma^4}(n+2) \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 + \frac{1}{2\sigma^6} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^3 \right\} \\
&= \frac{1}{4\sigma^6} \left\{ -\frac{1}{2}n^2(n+2) + \frac{3}{2\sigma^2}n(n+2) \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 - \frac{1}{2\sigma^4}(3n+4) \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 \right. \\
&\quad \left. + \frac{1}{2\sigma^6} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^3 \right\}
\end{aligned}$$

y al tomar esperanza, se obtiene que

$$\begin{aligned}
E \left\{ \left(\frac{1}{f} \frac{\partial f}{\partial \sigma^2} \right) \left(\frac{1}{f} \frac{\partial^2 f}{\partial \sigma^2} \right) \right\} &= \frac{1}{8\sigma^6} [-n^2(n+2) + 3n^2(n+2) - (3n+4)n(n+2) + n(n+2)(n+4)] \\
&= \frac{1}{8\sigma^6} n(n+2) [-n + 3n - (3n+4) + (n+4)] \\
&= 0,
\end{aligned}$$

donde se emplearon los resultados de las ecuaciones 7.66 y 7.67. Por lo tanto,

$$[\mathbf{B}(2)]_{24} = [\mathbf{B}(2)]_{42} = E \left\{ \left(\frac{1}{f} \frac{\partial f}{\partial \sigma^2} \right) \left(\frac{1}{f} \frac{\partial^2 f}{\partial \sigma^2} \right) \right\} = 0.$$

Luego,

$$\begin{aligned}
[\mathbf{B}(2)]_{33} &= E \left\{ \left(\frac{1}{f} \frac{\partial^2 f}{\partial \mu^2} \right)^2 \right\} \\
&= E \left\{ \left(-\frac{n}{\sigma^2} + \frac{1}{\sigma^4} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \right)^2 \right\} \\
&= E \left\{ \frac{n^2}{\sigma^4} - \frac{2}{\sigma^6} n \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 + \frac{1}{\sigma^8} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^4 \right\} \\
&= \frac{1}{\sigma^4} \left[n^2 - \frac{2}{\sigma^2} n E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \right\} + \frac{1}{\sigma^4} E \left\{ \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^4 \right\} \right] \\
&\stackrel{(a)}{=} \frac{1}{\sigma^4} [n^2 - 2n^2 + 3n^2] \\
&= \frac{2n^2}{\sigma^4},
\end{aligned}$$

donde en (a) se emplearon las ecuaciones 7.69 y 7.71. Se obtuvo que

$$[\mathbf{B}(2)]_{33} = \frac{2n^2}{\sigma^4}.$$

Continuando,

$$[\mathbf{B}(2)]_{34} = [\mathbf{B}(2)]_{43} = E \left\{ \left(\frac{1}{f} \frac{\partial^2 f}{\partial \mu^2} \right) \left(\frac{1}{f} \frac{\partial^2 f}{\partial \sigma^2} \right) \right\},$$

y de las ecuaciones 7.63 y 7.64,

$$\begin{aligned}
\left(\frac{1}{f} \frac{\partial^2 f}{\partial \mu^2} \right) \left(\frac{1}{f} \frac{\partial^2 f}{\partial \sigma^2} \right) &= \frac{1}{2\sigma^6} \left(-n + \frac{1}{\sigma^2} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \right) \\
&\times \left(\frac{1}{2} n(n+2) - \frac{1}{\sigma^2} (n+2) \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 + \frac{1}{2\sigma^4} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 \right) \\
&= \frac{1}{2\sigma^6} \left\{ -\frac{1}{2} n^2(n+2) + \frac{1}{\sigma^2} n(n+2) \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 - \frac{1}{2\sigma^4} n \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 \right. \\
&\quad + \frac{1}{2\sigma^2} n(n+2) \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 - \frac{1}{\sigma^4} (n+2) \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \\
&\quad \left. + \frac{1}{2\sigma^6} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu) \right]^2 \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 \right\}.
\end{aligned}$$

Al tomar esperanza, empleando las ecuaciones 7.66, 7.69, 7.74 y 7.75, se tiene que

$$\begin{aligned}
E \left\{ \left(\frac{1}{f} \frac{\partial^2 f}{\partial \mu^2} \right) \left(\frac{1}{f} \frac{\partial^2 f}{\partial \sigma^2} \right) \right\} &= \frac{1}{2\sigma^6} \left[-\frac{1}{2} n^2(n+2) + n^2(n+2) - \frac{1}{2} n^2(n+2) + \frac{1}{2} n^2(n+2) \right. \\
&\quad \left. - n(n+2)^2 + \frac{1}{2} n(n+2)(n+4) \right] \\
&= \frac{1}{4\sigma^6} n(n+2) [-n + 2n - 2(n+2) + (n+4)] \\
&= 0,
\end{aligned}$$

es decir,

$$[\mathbf{B}(2)]_{34} = [\mathbf{B}(2)]_{43} = E \left\{ \left(\frac{1}{f} \frac{\partial^2 f}{\partial \mu^2} \right) \left(\frac{1}{f} \frac{\partial^2 f}{\partial \sigma^2} \right) \right\} = 0.$$

Finalmente,

$$[\mathbf{B}(2)]_{44} = E \left\{ \left(\frac{1}{f} \frac{\partial^2 f}{\partial \sigma^2} \right)^2 \right\},$$

y de la ecuación 7.64,

$$\begin{aligned}
\left(\frac{1}{f} \frac{\partial^2 f}{\partial \sigma^2} \right)^2 &= \frac{1}{4\sigma^8} \left(\frac{1}{2} n(n+2) - \frac{1}{\sigma^2} (n+2) \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 + \frac{1}{2\sigma^4} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 \right)^2 \\
&= \frac{1}{4\sigma^8} \left\{ \frac{1}{4} n^2(n+2)^2 + \frac{1}{\sigma^4} (n+2)^2 \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 + \frac{1}{4\sigma^8} \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^4 \right. \\
&\quad \left. - \frac{1}{\sigma^2} n(n+2)^2 \sum_{i=1}^n (\mathbf{x}_i - \mu)^2 + \frac{1}{2\sigma^4} n(n+2) \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^2 \right. \\
&\quad \left. - \frac{1}{\sigma^6} (n+2) \left[\sum_{i=1}^n (\mathbf{x}_i - \mu)^2 \right]^3 \right\}.
\end{aligned}$$

De las ecuaciones 7.66, 7.67 y 7.68, al tomar esperanza se obtiene que

$$\begin{aligned}
E \left\{ \left(\frac{1}{f} \frac{\partial^2 f}{\partial \sigma^2} \right)^2 \right\} &= \frac{1}{4\sigma^8} \left[\frac{1}{4} n^2(n+2)^2 + n(n+2)^3 + \frac{1}{4} n(n+2)(n+4)(n+6) \right. \\
&\quad \left. - n^2(n+2)^2 + \frac{1}{2} n^2(n+2)^2 - n(n+2)^2(n+4) \right] \\
&= \frac{1}{4\sigma^8} \left[-\frac{1}{4} n^2(n+2)^2 + n(n+2)^3 + \frac{1}{4} n(n+2)(n+4)(n+6) - n(n+2)^2(n+4) \right] \\
&= \frac{1}{16\sigma^8} n(n+2) \left[-n(n+2) + 4(n+2)^2 + (n+4)(n+6) - 4(n+2)(n+4) \right] \\
&= \frac{1}{16\sigma^8} n(n+2) \left[-(n^2 + 2n) + 4(n^2 + 4n + 4) + (n^2 + 10n + 24) - 4(n^2 + 6n + 8) \right] \\
&= \frac{1}{16\sigma^8} n(n+2) \left[n^2(-1 + 4 + 1 - 4) + n(-2 + 16 + 10 - 24) + (16 + 24 - 32) \right] \\
&= \frac{1}{2\sigma^8} n(n+2),
\end{aligned}$$

es decir,

$$[\mathbf{B}(2)]_{44} = E \left\{ \left(\frac{1}{f} \frac{\partial^2 f}{\partial \sigma^2} \right)^2 \right\} = \frac{1}{2\sigma^8} n(n+2).$$

Combinando los resultados, se concluye que

$$\mathbf{B}(2) = E \left\{ \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix} \begin{bmatrix} \mathbf{Z}_1^t & \mathbf{Z}_2^t \end{bmatrix} \right\} = \begin{bmatrix} \frac{n}{\sigma^2} & 0 & 0 & 0 \\ 0 & \frac{n}{2\sigma^4} & \frac{n}{\sigma^4} & 0 \\ 0 & \frac{n}{\sigma^4} & \frac{2n^2}{\sigma^4} & 0 \\ 0 & 0 & 0 & \frac{n(n+2)}{2\sigma^8} \end{bmatrix}.$$

La “matriz de información de Fisher extendida” es el bloque 2×2 en la posición $(1, 1)$ de $\mathbf{B}^{-1}(2)$. Para calcularla, puede emplearse la identidad de la ecuación 7.42, siendo

$$\mathbf{A} = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 0 \\ \frac{n}{\sigma^4} & 0 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 0 & \frac{n}{\sigma^4} \\ 0 & 0 \end{bmatrix} \quad \text{y} \quad \mathbf{D} = \begin{bmatrix} \frac{2n^2}{\sigma^4} & 0 \\ 0 & \frac{n(n+2)}{2\sigma^8} \end{bmatrix}$$

los bloques de $\mathbf{B}(2)$. De esta forma,

$$\mathbf{BD}^{-1}\mathbf{C} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix} \quad \text{y} \quad \mathbf{A} - \mathbf{BD}^{-1}\mathbf{C} = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix} = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n-1}{2\sigma^4} \end{bmatrix}$$

y como $[\mathbf{B}^{-1}(2)]_{11} = (\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1}$, se obtiene que

$$[\mathbf{B}^{-1}(2)]_{11} = \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n-1} \end{bmatrix}.$$

Comparando este resultado con las ecuaciones 7.57 y 7.58, se tiene que

$$E \left\{ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^t \right\} = \begin{bmatrix} \text{var}\{\hat{\mu}\} & 0 \\ 0 & \text{var}\{\hat{\sigma}^2\} \end{bmatrix} = [\mathbf{B}^{-1}(2)]_{11}.$$

Por lo tanto, la ecuación 7.50 se cumple con igualdad para $m = 2$, concluyendo que $\hat{\boldsymbol{\theta}}$ es un estimador eficiente de segundo orden.

Bibliografía

- [1] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 4th ed., 2002.
- [2] R. Kupferman, “Lecture notes in probability,” 2009. The Hebrew University of Jerusalem. Institute of Mathematics.
- [3] S. I. Resnick, *A probability path*. Birkhäuser Boston, 1999.
- [4] S. Ross, *A First Course in Probability*, ch. Axioms of Probability. Pearson, 9th ed., 2014.
- [5] F. Mosteller, “Optimal length of play for a binomial game,” *Selected Papers of Frederick Mosteller*, pp. 333–335, 2006.
- [6] D. Panchenko, “Lecture 8: Gamma Distribution, Beta Distribution,” in *18.650 Statistics for Applications*, Massachusetts Institute of Technology, 2003. MIT OpenCourseWare. License: Creative Commons BY-NC-SA. URL: <https://ocw.mit.edu/courses/mathematics/18-443-statistics-for-applications-fall-2003/index.htm> (last visited on 2021-11-16).
- [7] J. O. Smith III, *Spectral audio signal processing*, ch. Gaussian Function Properties. W3K publishing, 2011.
- [8] L. Takács, “The problem of coincidences,” *Archive for History of Exact Sciences*, vol. 21, no. 3, pp. 229–244, 1980.
- [9] H. Hathout, “The old hats problem,” *Rose-Hulman Undergraduate Mathematics Journal*, vol. 4, no. 1, p. 2, 2003.
- [10] S. J. Miller, *The Probability Lifesaver: All the Tools You Need to Understand Chance*, ch. Generating Functions and Convolutions. Princeton University Press, 2017.
- [11] E. W. Weisstein, “Bivariate normal distribution.” From MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/BivariateNormalDistribution.html>. Last visited on 2018-05-16.
- [12] M. D. Landau and W. R. Jones, “A hardy old problem,” *Mathematics Magazine*, vol. 56, no. 4, pp. 230–232, 1983.
- [13] user940, “Variance of sample variance?.” Mathematics Stack Exchange. URL: <https://math.stackexchange.com/q/73080> (version: 2011-10-23).
- [14] C. B. Do, “More on multivariate gaussians,” 2008. URL: http://vision.csee.wvu.edu/~doretto/courses/2016-fall-ml/reading/more_on_gaussians.pdf (last checked: 2018-06-01).
- [15] G. Grimmett and D. Stirzaker, *Probability and random processes*. Oxford University Press, 3rd ed., 2001.
- [16] Y. Polyanskiy, “Lecture 17: Laws of Large Numbers I,” in *6.436J Fundamentals of Probability*, Massachusetts Institute of Technology, 2018. MIT OpenCourseWare. License: Creative Commons BY-NC-SA. URL: <https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-436j-fundamentals-of-probability-fall-2018/index.htm> (last visited on 2021-09-28).

- [17] H. Singer, “Moment equations and Hermite expansion for nonlinear stochastic differential equations with application to stock price models,” *Computational Statistics*, vol. 21, no. 3, pp. 385–397, 2006.
- [18] G. Casella and R. L. Berger, *Statistical inference*. Cengage Learning, 2nd ed., 2001.
- [19] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1st ed., 1993.
- [20] S. U. Pillai and N. D. Sinha, “Improvements on the Cramer-Rao bound,” in *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, pp. 3565–3566, IEEE Computer Society, 1991.
- [21] A. Bhattacharyya, “On some analogues of the amount of information and their use in statistical estimation,” *Sankhyā: The Indian Journal of Statistics*, pp. 1–14, 1946.
- [22] E. López, “Estudiando el libro «Fundamentals of Statistical Signal Processing: Estimation Theory» de Steven M. Kay,” 2021. URL: https://github.com/bor9/estudiando_el_kay/blob/main/kay_statistical_signal_processing.pdf (last visited on 2022-01-15).