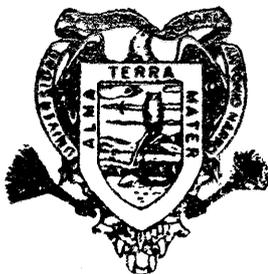


UN ENFOQUE ALGEBRAICO PARA LA PRUEBA DE  
INDEPENDENCIA DE CARACTERISTICAS EN EL  
ANALISIS DE DATOS CATEGORICOS

OLIVIA GARCIA CALVILLO

T E S I S

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

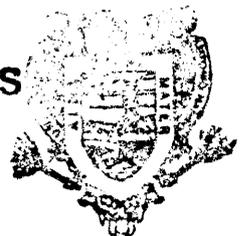


**Universidad Autónoma Agraria  
Antonio Narro**

**PROGRAMA DE GRADUADOS**

**Buenavista. Saltillo. Coah.**

**JUNIO DE 1999**



**BIBLIOTECA  
EGIDIO G. REBONATO  
BANCO DE TESIS  
U.A.A.A.N.**

11027

UNIVERSIDAD AUTÓNOMA AGRARIA

ANTONIO NARRO

SUBDIRECCIÓN DE POSTGRADO

Un Enfoque Algebraico Para la Prueba de Independencia  
de Características en el Análisis de Datos Categóricos

TESIS

por

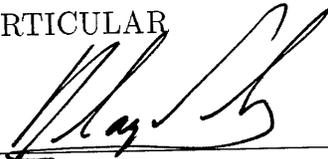
OLIVIA GARCÍA CALVILLO

Elaborada bajo la supervisión del Comité Particular de Asesoría y aprobada  
como requisito parcial para optar al grado de

MAESTRO EN CIENCIAS  
en Estadística Experimental

COMITÉ PARTICULAR

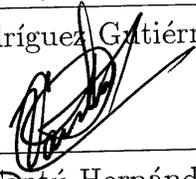
Asesor Principal:

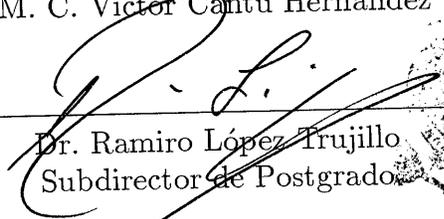
  
Dr. Rolando Cavazos Cadena

Asesor:

  
M. C. Luis Rodríguez Gutiérrez

Asesor:

  
M. C. Víctor Cantú Hernández

  
Dr. Ramiro López Trujillo  
Subdirector de Postgrado

Buenavista, Saltillo, Coahuila. Junio, 1999

## **Agradecimientos**

### **A Dios,**

Quien ha sido mi guía, sustento y fortaleza, en este y todos los momentos de mi vida.

### **A mi hijo Juan Pablo,**

Con todo mi amor. A él le pido disculpas por haberme permitido tomar parte de su tiempo, esperando que este trabajo sea un estímulo en su vida.

### **A mi esposo José Juan,**

Por el amor, comprensión y apoyo incondicional que siempre me ha dado.

### **A mis Padres, Irma y Juvenal,**

Porque me ayudaron a superarme, brindándome siempre un invaluable apoyo, además de que de ellos aprendí que la superación profesional es el mejor camino para lograr el éxito.

### **Al Doctor Rolando Cavazos,**

Gracias por contribuir enormemente en mi formación profesional, pues su experiencia, capacidad y conocimientos, me condujeron al final de esta etapa de mi vida. Su ayuda, consejos y atenciones perdurarán por siempre en mi corazón, como un regalo de altísimo valor.

### **A mis hermanos Dary, Irma, Norma y Juvenal, y a mis suegros, María Luisa y Gerónimo,**

Por el apoyo incondicional que siempre me han brindado.

A todos ellos...

*¡con afecto y agradecimiento!*

COMPENDIO

Un Enfoque Algebraico Para la Prueba de Independencia  
de Características en el Análisis de Datos Categóricos

por

*OLIVIA GARCÍA CALVILLO*

TESIS

**MAESTRÍA EN CIENCIAS**  
**en Estadística Experimental**

UNIVERSIDAD AUTÓNOMA AGRARIA  
ANTONIO NARRO

Buenavista, Saltillo, Coahuila, México. Junio, 1999

Dr. Rolando Cavazos Cadena — Asesor

**Palabras Clave:** Bondad de Ajuste, Distribución Asintótica, Clasificación Cruzada, Grados de Libertad, Teoremas de Invarianza, Matrices Idempotentes.

Este trabajo trata sobre dos problemas que se presentan cuando los resultados de un experimento aleatorio se clasifican en un número finito de categorías,

a saber, juzgar la bondad con que una muestra de la distribución multidimensional de Bernoulli se ajusta a un parámetro específico, y probar la hipótesis de independencia de dos características que originan la clasificación en una tabla de contingencia. El énfasis es sobre el papel que la distribución Ji-cuadrada, como límite distribucional de los correspondientes estadísticos de prueba, desempeña en estos problemas. La principal contribución técnica de esta tesis consiste en deducir, por métodos algebraicos, la expresión para los grados de libertad en la prueba de independencia de los rasgos de interés en una clasificación cruzada.

# ABSTRACT

An Algebraic Approach for the Test of Independence  
of Characteristics in the Analysis of Categorical Data

by

*OLIVIA GARCÍA CALVILLO*

## THESIS

**MASTER of SCIENCE**

**Experimental Statistics**

UNIVERSIDAD AUTÓNOMA AGRARIA

ANTONIO NARRO

Buenavista, Saltillo, Coahuila, México. June, 1999

Dr. Rolando Cavazos Cadena — Advisor

**Key Words:** Goodness of Fit, Asymptotic Distribution, Cross Classification,  
Degrees of Freedom, Invariance Theorems, Idempotent Matrices.

This work is concerned with two problems arising when the outcomes of a random experiment are classified into a finite number of categories, namely, assessing the goodness of fit of a sample of the multidimensional Bernoulli distribution

to a fixed parameter, and testing the independence of two characteristics yielding the categories in a contingency table. The exposition focuses on the role of the Chi-squared distribution as the limit in distribution of the corresponding test statistics for these problems. The main technical contribution of the thesis consists in providing, via an elementary algebraic analysis, a justification for the number of degrees of freedom used to test the independence hypothesis in the analysis of contingency tables.

## CONTENIDO

	Página
Perspectiva Global . . . . .	1
Introducción . . . . .	1
Objetivos . . . . .	2
La Organización . . . . .	4
La Prueba de Bondad de Ajuste . . . . .	6
Introducción . . . . .	6
Momentos de una Distribución de Bernoulli . . . . .	8
Los Teoremas de Invarianza . . . . .	12
Resultados Técnicos Preliminares . . . . .	15
Demostración de Teorema 3.2 . . . . .	19
Bondad de Ajuste . . . . .	22
Intervalos de Confianza . . . . .	26
Observaciones Adicionales . . . . .	33
Una extensión del Método Scheffè . . . . .	37
Conclusión . . . . .	45
Clasificación por Medio de dos Características . . . . .	47
Introducción . . . . .	47
Clasificación Cruzada . . . . .	48
Tablas de Contingencia . . . . .	54
Función de Comparación . . . . .	58
La Derivada de $f$ . . . . .	64
Distribución Límite . . . . .	71
Conclusión . . . . .	72
Prueba de la Hipótesis de Independencia . . . . .	73

Introducción . . . . .	73
El Teorema Principal . . . . .	74
Resultados Técnicos Auxiliares . . . . .	76
Idempotencia de $W$ . . . . .	83
Demostración del Teorema 2.2 . . . . .	87
El Procedimiento de Prueba . . . . .	90
Conclusión . . . . .	95
Epílogo . . . . .	96
Literatura Citada . . . . .	98

# Capítulo 1

## Perspectiva Global

El propósito de este capítulo es proporcionar una visión general del contenido de esta tesis, plantear los objetivos que se pretenden alcanzar, y describir la estructura con que se ha organizado la presentación del trabajo.

### Introducción

Este trabajo trata sobre el análisis de datos catogóricos, los cuales surgen cuando los resultados de un experimento aleatorio se clasifican en grupos, genéricamente referidos como categorías (Agresti, 1984, Bishop *et al.*, 1975). La agrupación de resultados de un experimento aleatorio es frecuente en la aplicaciones, y surge en los contextos más diversos. Por ejemplo, en un proceso de fabricación en serie, los artículos producidos pueden clasificarse como aceptables, medios o defectuosos, dando origen a tres categorías, mientras que en una terminal aérea, los vuelos que están por partir se clasifican en los monitores como ‘a tiempo’ o ‘con retraso’. Frecuentemente, las categorías se forman tomando en cuenta dos características. Un ejemplo de este agrupamiento se encuentra, por ejemplo, en estudios que se realizan en ciencias sociales, donde se trata de investigar si una característica económica—como nivel de ingreso familiar catalogado como bajo, medio y alto—tiene relación con algún rasgo de tipo ‘conductual’—como aprovechamiento académico promedio en los estudios universitarios, clasificado mediante niveles, A, B y C. En este caso, ‘cruzando’ los niveles de las dos características se obtienen diversas categorías, las cuales pueden describirse, por ejemplo, como ‘alto-C’; a esta última clase pertenece un estudiante cuya familia posee un ingreso alto, mientras que su aprovechamiento escolar lo hace acreedor a una ‘C’.

Cuando se realiza un experimento aleatorio y sus resultados se agrupan en un número  $k$  de categorías, el resultado puede condificarse mediante un objeto aleatorio  $\mathbf{X}$  cuyos valores posibles conforman un conjunto de  $k$  miembros, y su distribución se conoce como la *distribución  $k$ -dimensional de Bernoulli*; vea, por ejemplo, Mood *et al.*, (1985), Dudewicz y Mishra, (1988), o Lehman, (1991). Por conveniencia, se supone que  $\mathbf{X}$  asume valores en  $\mathbb{R}^k$ , el espacio Euclideo de dimensión  $k$ , y que los valores posibles del vector conforman la base canónica  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k\}$ , i.e.,  $\mathbf{e}_i$  tiene  $i$ -ésima componente igual a uno, mientras que las restantes son cero. Luego, al realizar  $n$  repeticiones independientes del experimento, el observador termina con una muestra  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  de la distribución de Bernoulli, y la suma  $\mathbf{S} = \sum_{i=1}^k \mathbf{X}_i$  es un estadístico suficiente, y por lo tanto ‘captura’ toda la información que la sucesión contiene sobre el vector  $\mathbf{p} = [p_1, p_2, \dots, p_k]$ , donde  $p_i$  es la probabilidad de que el resultado de un ensayo pertenezca a la  $i$ -ésima clase (Lehman, 1991). La distribución de  $\mathbf{S}$  es multinomial con parámetros  $k$  y  $\mathbf{p}$ , y la componente  $j$  del vector  $\mathbf{S}$  es la frecuencia con que se observa que el resultado pertenece a la  $j$ -ésima categoría en los  $n$  ensayos realizados. El estudio que se realiza en esta tesis tiene, como objeto principal, al vector  $\mathbf{S}$ .

## Objetivos

Dada una muestra de la distribución  $k$ -dimensional de Bernoulli correspondiente a un vector probabilístico  $\mathbf{p}$ —esto es,  $p_i \geq 0$  para todo  $i$  y  $\sum_{i=1}^k p_i = 1$ —se presentan dos problemas fundamentales: El primero de ellos, es utilizar la información de la muestra para decidir si los datos observados pueden considerarse como provenientes de la distribución de Bernoulli asociada a un vector fijo  $\mathbf{p}_0$ , esto es, se pretende probar la hipótesis

$$\mathcal{H}_0: \mathbf{p} = \mathbf{p}_0;$$

en otras palabras, se trata de valorar la *bondad de ajuste* de los datos a una

distribución específica. El otro problema, surge cuando la clasificación categórica se construye ‘cruzando’ dos características, como en la situación descrita en la Introducción. En este caso, se busca averiguar si el nivel de uno de los rasgos de interés influye sobre el que asume la otra característica. Por ejemplo, en el contexto de la sección precedente, se podría tener interés en investigar si el nivel de ingreso familiar de un estudiante, influye sobre su nivel de aprovechamiento escolar o no. De nueva cuenta, se arriba al problema de probar una hipótesis, la cual, de forma verbal, se expresa como

$\mathcal{H}_0$ : Las características que inducen una clasificación cruzada  
son independientes.

Ambos problemas admiten soluciones que son bien conocidas, las cuales se refieren como pruebas Ji-cuadrada (Feinberg, 1980, Mandansky, 1988), pues los procedimientos involucran la construcción de un estadístico cuya distribución límite (conforme el tamaño de la muestra crece) se aproxima a una distribución Ji-cuadrada. Sin embargo, a pesar de que ambos procedimientos son clásicos, hay dos razones fundamentales para analizarlos en esta tesis: (i) Desde un punto de vista formativo, el análisis de los métodos de prueba para las dos hipótesis representa una oportunidad para conjuntar resultados importantes sobre convergencia de variables aleatorias, y (ii) Desde una perspectiva técnica, el estudio de ambos problemas permite obtener, de manera rigurosa y mediante argumentos simples, una expresión para el número de grados de libertad  $\mathcal{L}$  que se utilizan para probar la hipótesis de independencia de características en el análisis de datos categóricos. En este último aspecto se centra *la principal aportación de este trabajo*, pues se proporciona una *deducción, algebraica y elemental, del valor apropiado de  $\mathcal{L}$* , en contraste con el enfoque, analítico y sofisticado, que está implícito en la aproximación al problema mediante el método de verosimilitud máxima (Kendall y Stuart, 1973, Lindgren, 1976). Desde una perspectiva histórica, este desarrollo es interesante en si mismo, pues originalmente se llegó a utilizar un valor incorrecto para el número de grados de libertad; vea, por ejemplo, Agresti (1984).

Después de este preámbulo, los objetivos que se pretenden alcanzar en esta tesis pueden establecerse como sigue:

- *Aplicar* la teoría de normalidad asintótica al problema de bondad de ajuste a una distribución específica de Bernoulli;
- *Formular*, en términos vectoriales, la hipótesis de independencia en el análisis de datos categóricos obtenidos mediante clasificación cruzada de dos rasgos de interés, los cuales se organizan en forma matricial;
- *Determinar*, mediante métodos algebraicos, el número de grados de libertad en la prueba Ji-cuadrada para la independencia de dos rasgos de interés.

Alcanzar estas metas, además de permitir unificar conocimientos adquiridos en los últimos años, representa, particularmente en el caso del tercero de los objetivos planteados, un importante reto. En efecto, como se muestra en el desarrollo posterior, la consecución de estos propósitos requiere un estudio detallado que involucra una gran variedad de ideas probabilísticas y algebraicas.

## La Organización

Para alcanzar los objetivos propuestos, el trabajo que se desarrolla en esta tesis ha sido organizado de la siguiente manera: La exposición técnica inicia en el Capítulo 2, en el cual se presentan resultados fundamentales sobre las nociones estadísticas de *convergencia en distribución y probabilidad*, enfatizando los teoremas de invarianza para estas ideas (Ash, 1975, Dudewicz y Mishra, 1988, Serfling, 1980). Los resultados analizados, los cuales desempeñan un importante papel en todo el trabajo, son abordados de manera rigurosa, y se utilizan para estudiar el problema de bondad de ajuste para un conjunto de datos provenientes de una distribución de Bernoulli, alcanzando, al concluir el capítulo, el primero de los objetivos propuestos.

En el Capítulo 3 se inicia el estudio de datos categóricos obtenidos registrando los niveles de dos características de interés; los datos experimentales que se

obtienen en este contexto se organizan en forma matricial, dando origen a lo que se denomina *tabla de contingencia*. En este capítulo se introduce la idea de *función de comparación*, la cual permite formular, en términos vectoriales, la hipótesis de independencia de los dos rasgos de interés que originan una clasificación cruzada. Además de ejemplificar la construcción de dicha función en forma detallada, se determina su matriz derivada, así como la distribución asintótica del estadístico que se obtiene evaluando la función de comparación en la media muestral de los datos observados. Al concluir la presentación se habrá alcanzado el segundo de los objetivos que se plantearon.

Por otro lado, el Capítulo 4 es la espina dorsal de este trabajo, en el sentido de que contiene la principal aportación técnica de la tesis. La exposición descansa en los resultados establecidos previamente, y se concentra en el análisis de una distribución asintótica, a saber, aquella que corresponde al estadístico de prueba de la hipótesis de independencia de los dos rasgos que inducen una clasificación cruzada. El estudio permite establecer, algebraicamente, la expresión para el número de grados de libertad que utiliza el método clásico, basado en la distribución Ji-cuadrada, alcanzando con ello el tercero de los objetivos de esta tesis. Finalmente, la presentación culmina en el Capítulo 5 con conclusiones y algunos comentarios breves.

## Capítulo 2

### La Prueba de Bondad de Ajuste

En este capítulo se establecen los resultados sobre convergencia en distribución y convergencia en probabilidad que serán utilizados en el desarrollo de este trabajo. Las ideas se ejemplifican para una muestra de la distribución  $k$ -dimensional de Bernoulli, y como producto del análisis se formula la prueba Ji-cuadrada de bondad de ajuste. A través de la dualidad entre pruebas de hipótesis e intervalos de confianza, se obtienen intervalos de confianza para combinaciones lineales de las componentes del parámetro probabiístico de la distribución de Bernoulli.

### Introducción

El ingrediente principal del trabajo que se desarrolla en este capítulo es una muestra  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  de la distribución de Bernoulli  $\mathcal{B}(p_1, p_2, \dots, p_k)$  de dimensión  $k$ , esto es,

(i)  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  son vectores aleatorios independientes con  $k$  componentes, y

(ii) Para  $i = 1, 2, \dots, n$

$$P[\mathbf{X}_i = \mathbf{e}_j] = p_j, \tag{1.1}$$

donde  $\mathbf{e}_j$  es el vector de  $\mathbb{R}^k$  con la todas sus componentes nulas, excepto la  $j$ -ésima, que es uno; note que  $\mathbf{p} = (p_1, p_2, \dots, p_k)'$  es un vector probabilístico, i.e.,  $p_1, p_2, \dots, p_k$  son números positivos cuya suma es la unidad. Intuitivamente, la sucesión  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  se asocia a  $n$  repeticiones independientes de un experimento aleatorio cuyos resultados se clasifican en  $k$  categorías; si el resultado del

ensayo  $i$  arroja un resultado de la categoría  $j$ , éste se codifica a través de  $\mathbf{X}_i$  como

$$X_{i1} = 0, X_{i2} = 0, \dots, X_{i,j-1} = 0, X_{i,j} = 1, X_{i,j+1} = 0, X_{i,k} = 0. \quad (1.2)$$

Suponiendo que el tamaño de la muestra es lo ‘suficientemente grande’, el principal objetivo de este capítulo es presentar una justificación rigurosa de la prueba  $\chi^2$  (asintótica) para distinguir entre las alternativas

$$\mathcal{H}_0: \mathbf{p} = \mathbf{q}, \quad \text{vs} \quad \mathcal{H}_a: \mathbf{p} \neq \mathbf{q}, \quad (1.3)$$

donde  $\mathbf{q} = (q_1, q_2, \dots, q_k)$  es un vector de probabilidad fijo. Este es el problema de determinar si la muestra  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  se ajusta a una distribución de Bernoulli dada. Además, se determinará una región de confianza para el vector de probabilidad  $\mathbf{p}$ , así como intervalos de confianza simultáneos para combinaciones lineales de las componentes de  $\mathbf{p}$ . El material subsecuente se encuentra organizado de la siguiente manera. En la Sección 2 se encuentran los momentos de una distribución de Bernoulli, y se utiliza el Teorema Central de Límite para establecer la normalidad asintótica de la media muestral de los vectores observados. La Sección 3 trata sobre dos resultados fundamentales en la teoría de la convergencia en distribución de vectores aleatorios, a saber, los teoremas de invarianza, los cuales establecen que la propiedad de convergencia en distribución de una sucesión de vectores aleatorios es invariable bajo transformaciones apropiadas de los elementos de la sucesión. El primero de estos resultados es una versión general de una serie de aserciones genéricamente conocidos como teoremas de Slutsky, mientras que el segundo se refiere a la invarianza de la normalidad asintótica de una sucesión de vectores aleatorios cuando ésta se somete a una transformación diferenciable. Este último resultado se demuestra en la Sección 5 después de establecer los preliminares técnicos necesarios en la Sección 4. Posteriormente, en la Sección 6 se utilizan los teoremas de invarianza para obtener la prueba  $\chi^2$  que resuelve el problema (1.3), y en la Sección 7 se combina ese resultado con ideas comunes

en el análisis de varianza para obtener una región de confianza para  $\mathbf{p}$  con nivel asintótico de confianza predeterminado. En la sección 8 se presentan algunas observaciones adicionales, y el capítulo concluye con una extensión del Método de Scheffè en la Sección 9 antes de establecer las conclusiones en la Sección 10.

## Momentos de una Distribución de Bernoulli

El propósito de esta sección es determinar los momentos de orden menor o igual a dos de la distribución  $\mathcal{B}(p_1, p_2, \dots, p_k)$ . El punto de partida es la introducción de una notación adecuada.

**Definición 2.1.** (a) Sea  $\mathbf{r} = (r_1, r_2, \dots, r_k)'$  un vector en  $\mathbb{R}^k$ . En este caso,  $\mathbb{D}(\mathbf{r})$  es la matriz diagonal de orden  $k \times k$  con elementos  $r_1, r_2, \dots, r_k$  a lo largo de la diagonal principal:

$$\mathbb{D}(\mathbf{p}) := \begin{bmatrix} r_1 & 0 & 0 & \cdots & 0 \\ 0 & r_2 & 0 & \cdots & 0 \\ 0 & 0 & r_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & r_k \end{bmatrix}.$$

(b) Si  $\mathbf{p}$  es un vector de probabilidad, para cualquier  $\alpha \in \mathbb{R}$

$$\mathbf{p}^\alpha := (p_1^\alpha, p_2^\alpha, \dots, p_k^\alpha).$$

Note que si  $\mathbf{p}$  es un vector de probabilidad

$$\mathbb{D}(\mathbf{p}^\alpha) = \begin{bmatrix} p_1^\alpha & 0 & 0 & \cdots & 0 \\ 0 & p_2^\alpha & 0 & \cdots & 0 \\ 0 & 0 & p_3^\alpha & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & p_k^\alpha \end{bmatrix}, \quad (2.1)$$

mientras que a un momento de reflexión muestra que

$$\mathbb{D}(\mathbf{p}^\alpha)\mathbb{D}(\mathbf{p}^\beta) = \mathbb{D}(\mathbf{p}^{\alpha+\beta}), \quad \alpha, \beta \in \mathbb{R}. \quad (2.2)$$

Los momentos de orden menor o igual a dos de la distribución de  $\mathcal{B}(p_1, p_2, \dots, p_k)$  se obtienen a continuación.

**Lema 2.1.** Suponga que  $\mathbf{Y}$  es un vector aleatorio con distribución  $k$ -dimensional de Bernoulli con vector de probabilidades  $\mathbf{p}$ , i.e.,

$$P[\mathbf{Y} = \mathbf{e}_j] = p_j, \quad j = 1, 2, \dots, k.$$

En este caso,

$$(a) \quad E[\mathbf{Y}] = \mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{bmatrix};$$

$$(b) \quad E[\mathbf{Y}\mathbf{Y}'] = \mathbb{D}(\mathbf{p}),$$

y entonces

$$(c) \quad \text{Cov}[\mathbf{Y}] = \mathbb{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}'.$$

**Demostración.** (a) Sólo observe que  $E[\mathbf{Y}] = \sum_{i=1}^k p_i \mathbf{e}_i = \mathbf{p}$ .

(b) Debido a que solamente una componente de  $\mathbf{Y}$  es no nula, es claro que  $P[Y_j Y_i = 0] = 1$  siempre que  $i \neq j$ , y entonces

$$E[Y_i Y_j] = 0, \quad i \neq j.$$

Por otro lado, como cada componente de  $\mathbf{Y}$  es cero o uno, se tiene que  $Y_i^2 = Y_i$ , de manera que

$$E[Y_i^2] = E[Y_i] = p_i, \quad i = 1, 2, \dots, k,$$

donde la segunda igualdad se desprende de la parte (a). Observando que la componente  $i, j$  de la matriz  $E[\mathbf{Y}\mathbf{Y}']$  es  $(E[\mathbf{Y}\mathbf{Y}'])_{i,j} = E[(\mathbf{Y}\mathbf{Y}')_{i,j}] = E[Y_i Y_j]$ , al combinar las igualdades desplegadas líneas arriba se obtiene que

$$E[\mathbf{Y}\mathbf{Y}'] = \mathbb{D}(\mathbf{p}).$$

(c) Usando las partes (a) y (b) se obtiene que  $\text{Cov}[\mathbf{Y}] = E[\mathbf{Y}\mathbf{Y}'] - E[\mathbf{Y}](E[\mathbf{Y}])' = \mathbb{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}'$ .  $\square$

El Lema 2.1 permite encontrar la distribución asintótica de la media muestral correspondiente a una sucesión de vectores independientes obtenidos de la distribución de Bernoulli. El siguiente teorema es una consecuencia directa del Teorema Central de Límite (Serfling, 1980).

**Teorema 2.1.** Suponga que  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  es una sucesión de variables aleatorias independientes con distribución común  $\mathcal{B}(p_1, p_2, \dots, p_k)$ , y sea

$$\bar{\mathbf{X}}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$$

la media de los primeros  $n$  vectores. Con esta notación, la siguiente afirmación es válida: Conforme  $n$  tiende a  $\infty$ ,  $\sqrt{n}(\bar{\mathbf{X}}_n - \mathbf{p})$  converge en distribución a un vector  $\mathbf{W}$  con distribución normal  $\mathcal{N}(\mathbf{0}, \mathbb{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}')$ , esto es,

$$\sqrt{n}(\bar{\mathbf{X}}_n - \mathbf{p}) \xrightarrow{\mathcal{D}} \mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathbb{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}').$$

**Observación 2.1.** (a) En este punto es conveniente establecer de manera clara el significado de la noción de convergencia en distribución. Dados el vector aleatorio  $\mathbf{W}$  y la sucesión de vectores  $\{\mathbf{W}_n\}$ —todos de dimensión  $k$ —se tiene que  $\{\mathbf{W}_n\}$  converge en distribución a  $\mathbf{W}$  (i.e.,  $\mathbf{W}_n \xrightarrow{\mathcal{D}} \mathbf{W}$ ) si para cada vector  $\mathbf{w} \in \mathbb{R}^k$  en el cual la función de distribución de  $\mathbf{W}$  es continua, se satisface que

$$P[\mathbf{W}_n \leq \mathbf{w}] \rightarrow P[\mathbf{W} \leq \mathbf{w}] \quad \text{conforme } n \rightarrow \infty;$$

vea, por ejemplo, Ash (1975), Lindgren (1975), Lehmann (1991), o Dudewicz y Mishra (1988). Con frecuencia es conveniente tener caracterizaciones alternativas de un concepto, y la idea de convergencia en distribución no escapa a esta

regla. A continuación se describen tres caracterizaciones alternativas que permiten, tanto entender mejor la noción de convergencia recientemente introducida, como demostrar en casos específicos que una sucesión dada converge efectivamente en distribución; estas caracterizaciones serán utilizadas en las siguientes secciones; detalles sobre los siguientes resultados (i)—(iii), pueden encontrarse en Ash (1975), o en Serfling (1980).

(i)  $\mathbf{W}_n \xrightarrow{\mathcal{D}} \mathbf{W} = (W_1, W_2, \dots, W_k)'$  si y sólo si para cada punto  $\mathbf{w} = (w_1, w_2, \dots, w_k)$  tal que la función de distribución de  $W_i$  es continua en  $w_i$ —i.e.,  $P[W_i = w_i] = 0$ —se tiene que

$$\lim_{n \rightarrow \infty} P[\mathbf{W}_n \leq \mathbf{w}] = P[\mathbf{W} \leq \mathbf{w}].$$

(ii)  $\mathbf{W}_n \xrightarrow{\mathcal{D}} \mathbf{W}$  si y sólo si para todo conjunto  $\mathcal{C} \subset \mathbb{R}^k$  tal que su frontera  $\partial\mathcal{C}$  satisface que  $P[\mathbf{W} \in \partial\mathcal{C}] = 0$ , se tiene que

$$\lim_{n \rightarrow \infty} P[\mathbf{W}_n \in \mathcal{C}] = P[\mathbf{W} \in \mathcal{C}].$$

(iii)  $\mathbf{W}_n \xrightarrow{\mathcal{D}} \mathbf{W}$  si y sólo si la función característica de  $\mathbf{W}_n$  converge puntualmente a la función característica de  $\mathbf{W}$ , i.e, para todo  $\mathbf{t} \in \mathbb{R}^k$ ,

$$\phi_{\mathbf{W}_n}(\mathbf{t}) = E[e^{i\mathbf{t}'\mathbf{W}_n}] \rightarrow \phi_{\mathbf{W}}(\mathbf{t}) = E[e^{i\mathbf{t}'\mathbf{W}}] \quad \text{conforme } n \rightarrow \infty.$$

Por otro lado, en el resto del capítulo se utilizará repetidamente la siguiente propiedad básica que relaciona la ideas convergencia en distribución y convergencia en probabilidad (Dudewicz y Mishra, 1988): Suponga que  $\{\mathbf{W}_n\}$  y  $\{\mathbf{Z}_n\}$  son dos sucesiones de vectores aleatorios de dimensión  $k$ , las cuales satisfacen que

$$\mathbf{W}_n \xrightarrow{\mathcal{D}} \mathbf{W}, \quad \text{y} \quad \mathbf{Z}_n \xrightarrow{P} \mathbf{0},$$

donde  $\mathbf{W}$  es un vector aleatorio fijo. En este caso,

$$\mathbf{W}_n + \mathbf{Z}_n \xrightarrow{\mathcal{D}} \mathbf{W}.$$

En palabras, esta propiedad establece que si una sucesión convergente en distribución se sujeta a una perturbación (aditiva) convergente a cero, entonces la sucesión modificada continúa convergiendo en distribución al mismo límite.

(b) La caracterización (ii) líneas arriba, permite señalar de manera clara la ‘ventaja’ de establecer la convergencia en distribución de una sucesión  $\mathbf{W}_n$  a un vector  $\mathbf{W}$ :  $P[\mathbf{W}_n \in \mathcal{C}]$  puede aproximarse, generalmente, por  $P[\mathbf{W} \in \mathcal{C}]$ , y esta última probabilidad no depende de  $n$ .

## Los Teoremas de Invarianza

En esta sección se presentan dos resultados sobre la invarianza de la convergencia en distribución. En términos generales, éstos permiten concluir que si una sucesión  $\{\mathbf{W}_n\}$  converge en distribución a un vector  $\mathbf{W}$ , y si  $f$  es una función dada que satisface condiciones poco restrictivas, entonces la sucesión transformada  $\{f(\mathbf{W}_n)\}$  también converge en distribución. Más aún, si  $\{\mathbf{W}_n\}$  es asintóticamente normal, lo mismo ocurre con la sucesión  $\{f(\mathbf{W}_n)\}$ . En otras palabras, se establecerá que la propiedad de convergencia en distribución, así como la normalidad asintótica, son generalmente *invariables* bajo transformaciones de la sucesión de vectores aleatorios. Estos resultados se presentan de manera precisa en los siguientes Teoremas 3.1 y 3.2.

**Teorema 3.1.** [Teorema General de Invarianza.] Suponga que  $\{\mathbf{W}_n\}$  es una sucesión de vectores aleatorios de dimensión  $k$  tales que

$$\mathbf{W}_n \xrightarrow{\mathcal{D}} \mathbf{W}.$$

Entonces, para toda función continua  $f: \mathbb{R}^k \rightarrow \mathbb{R}^s$ ,

$$f(\mathbf{W}_n) \xrightarrow{\mathcal{D}} f(\mathbf{W}).$$

Este teorema es una versión general de los resultados conocidos genéricamente como teoremas de Slutsky (Dudewicz y Mishra, 1988), y puede resumirse de la siguiente manera: La propiedad de convergencia en distribución es invariable bajo la aplicación de funciones continuas. Si, por ejemplo,  $f(x) = \sum_{i=1}^k x_i$ , este teorema implica que a partir de la condición  $\mathbf{W}_n \xrightarrow{\mathcal{D}} \mathbf{W}$ , se desprende que

$$\sum_{i=1}^k W_{ni} \xrightarrow{\mathcal{D}} \sum_{i=1}^k W_i,$$

mientras que definiendo  $f$  mediante  $f(\mathbf{x}) = \prod_{i=1}^k x_i$ , se obtiene que

$$\prod_{i=1}^k W_{ni} \xrightarrow{\mathcal{D}} \prod_{i=1}^k W_i.$$

El Teorema 3.1 desempeñará un importante papel en la obtención de la prueba  $\chi^2$  de bondad de ajuste presentada en la Sección 6, en donde se aplicará este resultado tomando a  $f(\mathbf{x})$  como el cuadrado de la norma Euclideana en  $\mathbb{R}^k$ .

**Demostración del Teorema 3.1.** Sea  $\mathbf{x} = (x_1, x_2, \dots, x_k)' \in \mathbb{R}^k$  un punto con la siguiente propiedad: Para cada  $i = 1, 2, \dots, k$  la función de distribución de  $f_i(\mathbf{V})$ —la  $i$ -ésima componente de  $f(\mathbf{V})$ —es continua en  $x_i$ , i.e.,

$$P[f_i(\mathbf{V}) = x_i] = 0. \quad (3.1)$$

Como se señaló en la Observación 2.1, para establecer que  $f(\mathbf{V}_n) \xrightarrow{\mathcal{D}} f(\mathbf{V})$ , es suficiente verificar que

$$\lim_{n \rightarrow \infty} P[f(\mathbf{V}_n) \leq \mathbf{x}] = P[f(\mathbf{V}) \leq \mathbf{x}], \quad (3.2)$$

y con esto en mente observe que (3.1) implica que

$$P[f(\mathbf{V}) \in \partial\mathcal{C}(\mathbf{x})] = 0, \quad (3.3)$$

donde

$$\mathcal{C}(\mathbf{x}) := \prod_{i=1}^k (-\infty, x_i],$$

y para cada conjunto  $A$ , el símbolo ' $\partial A$ ' denota la frontera de  $A$ . Como  $f$  es continua,  $P[f(\mathbf{V}) \in \partial \mathcal{C}(\mathbf{x})] = 0$  implica que

$$P[\mathbf{V} \in \partial f^{-1}\mathcal{C}(\mathbf{x})] = 0,$$

condición que, via la convergencia en distribución de  $\mathbf{V}_n$  hacia  $\mathbf{V}$ , permite concluir, usando la parte (ii) en la Observación 2.1(a), que

$$P[f(\mathbf{V}_n) \in \mathcal{C}(\mathbf{x})] = P[\mathbf{V}_n \in f^{-1}(\mathcal{C}(\mathbf{x}))] \rightarrow P[\mathbf{V} \in f^{-1}(\mathcal{C}(\mathbf{x}))] = P[f(\mathbf{V}) \in \mathcal{C}(\mathbf{x})],$$

convergencia que combinada con (3.3) produce (3.2).  $\square$

**Teorema 3.2.** [Teorema de Invarianza de la Normalidad Asintótica.] Suponga que  $\{\mathbf{W}_n\}$  es una sucesión de vectores aleatorios de dimensión  $k$  tales que

$$\sqrt{n}(\mathbf{W}_n - \mu) \xrightarrow{\mathcal{D}} \mathbf{W} \sim \mathcal{N}(\mathbf{0}, M), \quad (3.4)$$

donde  $\mu \in \mathbb{R}^k$  es un vector fijo. Entonces, si  $f: \mathbb{R}^k \rightarrow \mathbb{R}^s$  es una función continuamente diferenciable,

$$\sqrt{n}(f(\mathbf{W}_n) - f(\mu)) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, Df(\mu)MDf(\mu)'),$$

donde  $Df(\mu)$  es la (matriz) derivada de la función  $f$  evaluada en  $\mu$ .

Este resultado establece que la normalidad asintótica es invariable bajo la aplicación de funciones (continuamente) diferenciables en el vector  $\mu$ , el cual puede denominarse la media asintótica de la sucesión  $\{\mathbf{W}_n\}$ . Como ilustración, suponga que  $X_1, X_2, \dots, X_n$  es una sucesión de variables aleatorias independientes con densidad exponencial común  $\beta^{-1}e^{-x/\beta}I_{(0, \infty)}$ . En este caso, el Teorema Central de Límite implica que

$$\sqrt{n}(\bar{X}_n - \beta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \beta^2).$$

Si  $f(x) = \ln(x)$ , es claro que  $f$  es continuamente diferenciable en  $\mu = \beta$ , y que  $Df(\beta) = 1/\beta$ , de manera que el Teorema 3.2 permite concluir que

$$\sqrt{n}(\ln(\bar{X}_n) - \ln(\beta)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, Df(\beta)\beta^2 Df(\beta)) = \mathcal{N}(0, 1),$$

esto es,  $\{\sqrt{n}(\ln(\bar{X}_n) - \ln(\beta))\}$  converge en distribución a la distribución normal estándar. La demostración del Teorema 3.1 es algo técnica y se presentará en la Sección 5 después de establecer las herramientas auxiliares necesarias.

## Resultados Técnicos Preliminares

Esta sección contiene los preliminares técnicos que se utilizarán para demostrar el Teorema 3.2, los que por claridad han sido divididos en cuatro componentes establecidos a continuación como los Lemas 4.1–4.4. La exposición sigue la presentación clásica sobre este tema (Ash, 1975), y buena parte de la argumentación gira alrededor de la siguiente idea de acotamiento de un vector aleatorio.

**Definición 4.1.** Una sucesión  $\{\mathbf{Z}_n\}$  de vectores aleatorios es acotada en probabilidad si para todo  $\delta > 0$  existe  $K(\delta) > 0$  tal que

$$P[|\mathbf{Z}_n| > K(\delta)] \leq \delta.$$

Intuitivamente, la sucesión  $\{\mathbf{Z}_n\}$  es acotada en probabilidad si la probabilidad de que los vectores  $\mathbf{Z}_n$  tengan una norma que exceda a  $K$  puede hacerse tan pequeña como se quiera con tan solo tomar  $K$  lo suficientemente grande. El siguiente lema relaciona esta noción de acotamiento con la idea de convergencia en distribución.

**Lema 4.1.** Suponga que la sucesión  $\{\mathbf{V}_n\}$  de vectores aleatorios  $k$ -dimensionales es tal que

$$\mathbf{V}_n \xrightarrow{\mathcal{D}} \mathbf{V}.$$

Entonces las siguientes propiedades (a) y (b) son válidas.

(a) La sucesión  $\{\mathbf{V}_n\}$  es acotada en probabilidad, y

(b) Si  $\{c_n\}$  es una sucesión de números reales que convergen a cero, entonces  $c_n \mathbf{V}_n \xrightarrow{P} \mathbf{0}$  conforme  $n \rightarrow \infty$ .

**Demostración.** (a) Sea  $\delta > 0$ , y note que  $P[|\mathbf{V}| > N] \rightarrow 0$  conforme  $N \rightarrow \infty$ .

De esta manera, es posible seleccionar  $K_1 > 0$  tal que

$$P[|\mathbf{V}| > K_1] < \delta; \quad (4.1)$$

además, debido a que  $P[|\mathbf{V}| = N]$  se anula excepto posiblemente para una cantidad numerable de valores de  $N$ , puede suponerse, sin pérdida de generalidad alguna, que  $K_1$  en (4.1) es tal que  $P[|\mathbf{V}| = K_1] = 0$ . En este caso, ya que  $\mathbf{V}_n \xrightarrow{D} \mathbf{V}$ , se tiene que  $\lim_{n \rightarrow \infty} P[|\mathbf{V}_n| > K_1] = P[|\mathbf{V}| > K_1] < \delta$  (vea la Observación 2.1), y entonces existe  $N$  tal que

$$P[|\mathbf{V}_n| > K_1] < \delta, \quad n > N. \quad (4.2)$$

Por otro lado observe que, para cada  $n$  fijo,  $P[|\mathbf{V}_n| > K] \rightarrow 0$  conforme  $K \rightarrow \infty$ , de manera que existe  $K_2 > 0$  tal que

$$P[|\mathbf{V}_n| > K_2] < \delta, \quad n = 1, 2, \dots, N$$

y definiendo  $K(\delta) = \max\{K_1, K_2\}$ , a partir de la última desigualdad y (4.2) se obtiene que

$$P[|\mathbf{V}_n| > K] < \delta \quad \text{para todo } n = 1, 2, \dots,$$

lo cual establece que la sucesión  $\{\mathbf{V}_n\}$  es acotada en probabilidad.

(b) Sea  $\varepsilon > 0$  fijo. A continuación se verificará que  $P[|c_n \mathbf{V}_n| > \varepsilon] \rightarrow 0$  conforme  $n \rightarrow \infty$ , convergencia que equivale al siguiente enunciado: Para todo  $\delta > 0$  existe  $N$  tal que

$$P[|c_n \mathbf{V}_n| > \varepsilon] < \delta, \quad n > N. \quad (4.3)$$

Para establecer esta afirmación seleccione  $K(\delta)$  como en la parte (a), y sea  $N$  un entero tal que  $K(\delta) < \varepsilon/(|c_n| + 1/n)$  para todo  $n > N$ ; tal entero  $N$  existe, pues  $c_n \rightarrow \mathbf{0}$ . En este caso,

$$\begin{aligned} n > N \quad \implies \quad P[|c_n \mathbf{V}_n| > \varepsilon] &\leq P\left[|\mathbf{V}_n| > \frac{\varepsilon}{|c_n| + 1/n}\right] \\ &\leq P[|\mathbf{V}_n| > K(\delta)] \quad (\text{pues } n > N) \\ &\leq \delta \end{aligned}$$

donde la tercera desigualdad se desprende de la parte (a). Este argumento establece (4.3) y concluye la demostración del lema.  $\square$

**Lema 4.2.** Suponga que  $\{\mathbf{Z}_n\}$  es una sucesión de vectores aleatorios de dimensión  $k$ , tales que  $\sqrt{n}(\mathbf{Z}_n - \mu) \xrightarrow{P} \mathbf{Z}$ , donde  $\mathbf{Z}$  es un vector aleatorio dado y  $\mu \in \mathbb{R}^k$  es fijo. Entonces

$$\mathbf{Z}_n \xrightarrow{P} \mu.$$

**Demostración.** Por la parte (a) del Lema 4.1 se tiene que  $\sqrt{n}(\mathbf{Z}_n - \mu)$  es una sucesión acotada en probabilidad, y en este caso el Lema 4.1(b) con  $V_n = \sqrt{n}(\mathbf{Z}_n - \mu)$  y  $c_n = 1/\sqrt{n}$  implica que

$$\mathbf{Z}_n - \mu = \frac{1}{\sqrt{n}} \sqrt{n}(\mathbf{Z}_n - \mu) \xrightarrow{P} \mathbf{0},$$

convergencia que equivale a  $\mathbf{Z}_n \xrightarrow{P} \mu$ .  $\square$

**Lema 4.3.** Sea  $\{\mathbf{W}_n\}$  una sucesión de vectores aleatorios de dimensión  $k$  acotada en probabilidad (vea la Definición 4.1), y suponga que  $\{Z_n\}$  es una sucesión de variables aleatorias tales que  $Z_n \xrightarrow{P} 0$ . Entonces,

$$Z_n \mathbf{W}_n \xrightarrow{P} \mathbf{0}. \tag{4.4}$$

**Demostración.** Sea  $\varepsilon > 0$  fijo. A continuación se mostrará que para todo  $\delta > 0$  existe un entero positivo  $N$  tal que

$$P[|Z_n \mathbf{W}_n| > \varepsilon] < \delta \quad \text{para todo } n > N,$$

lo cual es equivalente a la convergencia (4.4). El acotamiento en probabilidad de la sucesión  $\{\mathbf{W}_n\}$  implica que existe una constante  $K(\delta/2) > 0$  tal que

$$P[|\mathbf{W}_n| > K(\delta/2)] < \frac{\delta}{2}, \quad n = 1, 2, \dots, \quad (4.5)$$

y entonces

$$\begin{aligned} P[|Z_n \mathbf{W}_n| > \varepsilon] &= P[|\mathbf{W}_n| > K(\delta/2), |Z_n \mathbf{W}_n| > \varepsilon] \\ &\quad + P[|\mathbf{W}_n| \leq K(\delta/2), |Z_n \mathbf{W}_n| > \varepsilon] \\ &\leq P[|\mathbf{W}_n| > K(\delta/2)] + P[|\mathbf{W}_n| \leq K(\delta/2), |Z_n \mathbf{W}_n| > \varepsilon] \quad (4.6) \\ &\leq P[|\mathbf{W}_n| > K(\delta/2)] + P[K(\delta/2)|Z_n| > \varepsilon] \\ &\leq \frac{\delta}{2} + P\left[|Z_n| > \frac{\varepsilon}{K(\delta/2)}\right] \end{aligned}$$

donde la última desigualdad se desprende de (4.5). Como  $Z_n \xrightarrow{P} 0$  se sigue que  $\lim_{n \rightarrow \infty} P\left[|Z_n| > \frac{\varepsilon}{K(\delta/2)}\right] \rightarrow 0$  y entonces existe  $N$  tal que

$$P\left[|Z_n| > \frac{\varepsilon}{K(\delta/2)}\right] < \frac{\delta}{2}, \quad n > N$$

y en combinación con (4.6) esta desigualdad implica que  $P[|Z_n \mathbf{W}_n| > \varepsilon] < \delta$  si  $n > N$  y, como se mencionó anteriormente, esto establece el lema.  $\square$

**Lema 4.4.** Suponga que  $\{\mathbf{W}_n\}$  es una sucesión de vectores aleatorios  $k$ -dimensionales tal que

$$\sqrt{n}(\mathbf{W}_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, M).$$

Entonces, si  $A$  es una matriz real de orden  $s \times k$ ,

$$\sqrt{n}(A\mathbf{W}_n - A\mu) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, AMA').$$

**Demostración.** La conclusión que se pretende establecer es equivalente a la siguiente convergencia: Para todo  $\mathbf{r} \in \mathbb{R}^s$ ,

$$\phi_{\sqrt{n}(A\mathbf{W}_n - A\mu)}(\mathbf{r}) = E[e^{i\mathbf{r}'\sqrt{n}(A\mathbf{W}_n - A\mu)}] \rightarrow e^{-\mathbf{r}'AMA'\mathbf{r}/2} \quad \text{conforme } n \rightarrow \infty; \quad (4.7)$$

vea la parte (iii) de la Observación 2.1. Para verificarla, note que  $\sqrt{n}(\mathbf{W}_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, M)$  equivale a

$$\lim_{n \rightarrow \infty} \phi_{\sqrt{n}(\mathbf{W}_n - \mu)}(\mathbf{t}) = E[e^{i\mathbf{t}'\sqrt{n}(\mathbf{W}_n - \mu)}] = e^{-\mathbf{t}'M\mathbf{t}/2}, \quad \mathbf{t} \in \mathbb{R}^k, \quad (4.8)$$

y entonces

$$\begin{aligned} \phi_{\sqrt{n}(A\mathbf{W}_n - A\mu)}(\mathbf{r}) &= E[e^{i\mathbf{r}'\sqrt{n}(A\mathbf{W}_n - A\mu)}] \\ &= E[e^{i\mathbf{r}'A\sqrt{n}(\mathbf{W}_n - \mu)}] \\ &= \phi_{\sqrt{n}(\mathbf{W}_n - \mu)}(A'\mathbf{r}) \end{aligned}$$

de tal modo que usando usando (4.8) con  $\mathbf{t} = A'\mathbf{r}$ , se obtiene que

$$\begin{aligned} \lim_{n \rightarrow \infty} \phi_{\sqrt{n}(A\mathbf{W}_n - A\mu)}(\mathbf{r}) &= \lim_{n \rightarrow \infty} \phi_{\sqrt{n}(\mathbf{W}_n - \mu)}(A'\mathbf{r}) \\ &= e^{-(A'\mathbf{r})'MA'\mathbf{r}/2} = e^{-\mathbf{r}'AMA'\mathbf{r}/2}, \end{aligned}$$

lo cual establece (4.7) y concluye el argumento.  $\square$

## Demostración de Teorema 3.2

En esta sección se presenta una demostración del teorema de invarianza de la normalidad asintótica. Además de los resultados de la Sección 4, el argumento requiere la siguiente herramienta.

**Lema 5.1.** Sea  $\{\mathbf{W}_n\}$  es una sucesión de vectores aleatorios de dimensión  $k$  tales que

$$\sqrt{n}(\mathbf{W}_n - \mu) \xrightarrow{\mathcal{D}} \mathbf{W},$$

donde  $\mu \in \mathbb{R}^k$  es un vector fijo. Suponga que  $f: \mathbb{R}^k \rightarrow \mathbb{R}^s$  es una función continuamente diferenciable, y defina

$$Q_n := \max\{|\partial_j f_i(\mathbf{y}) - \partial_j f_i(\mu)| : |\mathbf{y} - \mu| \leq |\mathbf{W}_n - \mu|, i = 1, 2, \dots, s, j = 1, 2, \dots, k\}. \quad (5.1)$$

Entonces,

(a)  $Q_n \xrightarrow{P} 0$ , y

(b)  $\sqrt{n}Q_n(\mathbf{W}_n - \mu) \xrightarrow{P} 0$ .

**Demostración.** (a) Primero note que

$$\mathbf{W}_m - \mu \xrightarrow{P} \mathbf{0}, \quad (5.2)$$

por el Lema 4.2. Ahora seleccione un número  $\delta > 0$  y observe que, debido a la continuidad de las derivadas parciales de  $f$ , existe  $\rho > 0$  tal que

$$\max\{|\partial_j f_i(\mathbf{y}) - \partial_j f_i(\mu)| : |\mathbf{y} - \mu| \leq \rho, i = 1, 2, \dots, s, j = 1, 2, \dots, k\} \leq \delta,$$

por lo que a partir de la definición de  $Q_n$  se desprende que el evento  $[Q_n > \delta]$  está contenido en  $[|\mathbf{W}_n - \mu| > \rho]$ , de manera que

$$P[Q_n > \delta] \leq P[|\mathbf{W}_n - \mu| > \rho],$$

y entonces (5.2) implica que  $P[Q_n > \delta] \rightarrow 0$  conforme  $n \rightarrow \infty$ .

(b) Por el Lema 4.1,  $\{\sqrt{n}(\mathbf{W}_n - \mu)\}$  es una sucesión acotada en probabilidad, y entonces la parte (a) y el Lema 4.3 implican que  $\sqrt{n}Q_n(\mathbf{W}_n - \mu) \xrightarrow{P} 0$ .  $\square$

En este punto se está en posibilidad de demostrar el Teorema 3.2; antes de seguir adelante, el lector puede encontrar benéfico repasar su enunciado.

**Demostración del Teorema 3.2.** El punto de partida es la siguiente igualdad:

$$\begin{aligned} \sqrt{n}(f(\mathbf{W}_n) - f(\mu)) &= \sqrt{n}(Df(\mu)(\mathbf{W}_n) - f(\mu)) \\ &+ \sqrt{n}(f(\mathbf{W}_n) - f(\mu) - Df(\mu)(\mathbf{W}_n) - f(\mu)). \end{aligned} \quad (5.3)$$

Ahora observe que el Lema 4.4 garantiza que

$$\sqrt{n}(Df(\mu)(\mathbf{W}_n) - f(\mu)) \xrightarrow{D} \mathcal{N}(\mathbf{0}, Df(\mu)MDf(\mu)'). \quad (5.4)$$

A continuación se establecerá la siguiente convergencia:

$$\sqrt{n}(f(\mathbf{W}_n) - f(\mu) - Df(\mu)(\mathbf{W}_n) - \mu) \xrightarrow{P} 0, \quad (5.5)$$

de la cual se desprende el resultado buscado. En efecto, hecho, usando (5.3), las convergencias (5.4) y (5.5) implican, via la parte (iv) en la Observación 2.1, que

$$\sqrt{n}(f(\mathbf{W}_n) - f(\mu)) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, Df(\mu)MDf(\mu)'),$$

la cual es la conclusión deseada. Luego, para concluir es suficiente establecer (5.5). Con esto en mente, considere una componente del vector  $f(\mathbf{V}_n)$ , digamos la  $i$ -ésima, y observe que por el teorema del valor medio se tiene que

$$f_i(\mathbf{W}_n) - f_i(\mu) = \text{grad } f_i(\widehat{\mathbf{W}}_n)(\mathbf{W}_n - \mu) \quad (5.6)$$

donde  $\widehat{\mathbf{W}}_n$  se encuentra en el segmento que une  $\mu$  con  $\mathbf{W}_n$ , y por lo tanto,

$$|\widehat{\mathbf{W}}_n - \mu| \leq |\mathbf{W}_n - \mu|,$$

desigualdad que implica que

$$\begin{aligned} & |\text{grad } f_i(\widehat{\mathbf{W}}_n) - \text{grad } f_i(\mu)| \\ & \leq k \max\{|\partial_j f_i(\widehat{\mathbf{W}}_n) - \partial_j f_i(\mu)| : j = 1, 2, \dots, k\} \\ & \leq k \max\{|\partial_j f_i(\mathbf{y}) - \partial_j f_i(\mu)| : |\mathbf{y} - \mu| \leq |\mathbf{W}_n - \mu|, j = 1, 2, \dots, k\} \\ & \leq kQ_n, \end{aligned} \quad (5.7)$$

donde  $Q_n$  es la variable aleatoria definida en (5.1). Note ahora que (5.6) puede escribirse equivalentemente como

$$\begin{aligned} & f_i(\mathbf{W}_n) - f_i(\mu) - \text{grad } f_i(\mu)(\mathbf{W}_n - \mu) \\ & = (\text{grad } f_i(\widehat{\mathbf{W}}_n) - \text{grad } f_i(\mu))(\mathbf{W}_n - \mu) \end{aligned}$$

y entonces (5.7) implica que

$$\begin{aligned} & |f_i(\mathbf{W}_n) - f_i(\mu) - \text{grad } f_i(\mu)(\mathbf{W}_n - \mu)| \\ & \leq |\text{grad } f_i(\widehat{\mathbf{W}}_n) - \text{grad } f_i(\mu)| \cdot |\mathbf{W}_n - \mu| \\ & \leq kQ_n |\mathbf{W}_n - \mu|; \end{aligned}$$

como esta desigualdad es válida para todo  $i = 1, 2, \dots, s$ , se desprende que

$$\begin{aligned}
 & |f(\mathbf{W}_n) - f(\mu) - Df(\mu)(\mathbf{W}_n - \mu)| \\
 &= \left| \begin{bmatrix} f_1(\mathbf{W}_n) - f_1(\mu) - \text{grad } f_1(\mu)(\mathbf{W}_n - \mu) \\ f_2(\mathbf{W}_n) - f_2(\mu) - \text{grad } f_2(\mu)(\mathbf{W}_n - \mu) \\ \vdots \\ f_s(\mathbf{W}_n) - f_s(\mu) - \text{grad } f_s(\mu)(\mathbf{W}_n - \mu) \end{bmatrix} \right| \\
 &\leq skQ_n|\mathbf{W}_n - \mu|.
 \end{aligned}$$

En resumen, se ha mostrado que

$$|\sqrt{n}(f(\mathbf{W}_n) - f(\mu)) - \sqrt{n}Df(\mu)(\mathbf{W}_n - \mu)| \leq skQ_n|\mathbf{W}_n - \mu|,$$

donde  $Q_n$  es la variable aleatoria definida en (5.1). Aplicando ahora el Lema 5.1, la anterior desigualdad permite concluir que

$$\sqrt{n}f(\mathbf{W}_n) - f(\mu) - \sqrt{n}Df(\mu)(\mathbf{W}_n - \mu) \xrightarrow{P} \mathbf{0},$$

lo cual establece (5.5) y, como se mencionó anteriormente, esto finaliza la demostración del Teorema 5.1.  $\square$

## Bondad de Ajuste

En esta sección se establecerán dos resultados acerca de la distribución asintótica de la norma de la desviación de  $\bar{\mathbf{X}}_n$  respecto a  $\mathbf{p}$ , los cuales se utilizarán para resolver los problemas de prueba de hipótesis y de construcción de regiones de confianza planteados en la introducción. El punto de partida es el siguiente resultado.

**Teorema 6.1.** Sea  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  es una sucesión de variables independientes con distribución común  $\mathcal{B}(p_1, p_2, \dots, p_k)$ , donde  $p_i > 0$  para  $i = 1, 2, \dots, k$ . Entonces,

$$n \sum_{i=1}^k \frac{(\bar{X}_{ni} - p_i)^2}{p_i} \xrightarrow{\mathcal{D}} Y \sim \chi_{k-1}^2 \quad \text{conforme } n \rightarrow \infty.$$

**Demostración.** Por el Teorema 2.1 se tiene que

$$\sqrt{n}(\bar{\mathbf{X}}_n - \mathbf{p}) \xrightarrow{\mathcal{D}} \mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathbb{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}'),$$

y entonces el Lema 4.4 o el Teorema 3.1 implican que

$$\begin{aligned} \sqrt{n}\mathbb{D}(\mathbf{p}^{-1/2})(\bar{\mathbf{X}}_n - \mathbf{p}) &\xrightarrow{\mathcal{D}} \mathbb{D}(\mathbf{p}^{-1/2})\mathbf{W} \\ \text{donde } \mathbb{D}(\mathbf{p}^{-1/2})\mathbf{W} &\sim \mathcal{N}\left(\mathbf{0}, \mathbb{D}(\mathbf{p}^{-1/2})(\mathbb{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}')\mathbb{D}(\mathbf{p}^{-1/2})'\right) \end{aligned} \quad (6.1)$$

vea (2.1) para la definición de  $\mathbb{D}(\mathbf{p}^{-1/2})$ . Por otro lado, un cálculo directo usando (2.2) muestra que

$$\mathbb{D}(\mathbf{p}^{-1/2})(\mathbb{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}')\mathbb{D}(\mathbf{p}^{-1/2})' = I - \mathbf{v}\mathbf{v}', \quad \text{donde } \mathbf{v} = \begin{bmatrix} \sqrt{p_1} \\ \sqrt{p_2} \\ \vdots \\ \sqrt{p_k} \end{bmatrix},$$

por lo que (6.1) es equivalente a

$$\sqrt{n}\mathbb{D}(\mathbf{p}^{-1/2})(\bar{\mathbf{X}}_n - \mathbf{p}) \xrightarrow{\mathcal{D}} \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, I - \mathbf{v}\mathbf{v}'); \quad (6.2)$$

note, además, que  $\mathbf{v}$  es un vector de norma unitaria, de donde se desprende que  $I - \mathbf{v}\mathbf{v}'$  es una matriz *idempotente* de rango  $k - 1$  (Searle, 1982, Graybill, 1985, Harville, 1997). Por otro lado, considere ahora la función  $f: \mathbb{R}^k \rightarrow \mathbb{R}$  dada por  $f(\mathbf{x}) = \mathbf{x}'\mathbf{x}$ . En este caso, la convergencia (6.2) implica, via el Teorema 3.1, que

$$f\left(\sqrt{n}\mathbb{D}(\mathbf{p}^{-1/2})(\bar{\mathbf{X}}_n - \mathbf{p})\right) \xrightarrow{\mathcal{D}} f(\mathbf{Z}) = \mathbf{Z}'\mathbf{Z};$$

como  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, I - \mathbf{v}\mathbf{v}')$  y la matriz  $I - \mathbf{v}\mathbf{v}'$  es idempotente de rango  $k - 1$ , se tiene que  $\mathbf{Z}\mathbf{Z}' \sim \chi_{k-1}^2$  (Graybill, 1985, Christensen, 1996, Fisher y McDonald, 1981), y entonces la convergencia anterior equivale a

$$f\left(\sqrt{n}\mathbb{D}(\mathbf{p}^{-1/2})(\bar{\mathbf{X}}_n - \mathbf{p})\right) \xrightarrow{\mathcal{D}} \chi_{k-1}^2. \quad (6.3)$$

Ahora observe que

$$\begin{aligned} f\left(\sqrt{n}\mathbb{D}(\mathbf{p}^{-1/2})(\bar{\mathbf{X}}_n - \mathbf{p})\right) &= \left(\sqrt{n}\mathbb{D}(\mathbf{p}^{-1/2})(\bar{\mathbf{X}}_n - \mathbf{p})\right)' \left(\sqrt{n}\mathbb{D}(\mathbf{p}^{-1/2})(\bar{\mathbf{X}}_n - \mathbf{p})\right) \\ &= n(\bar{\mathbf{X}}_n - \mathbf{p})'\mathbb{D}(\mathbf{p}^{-1/2})\mathbb{D}(\mathbf{p}^{-1/2})'(\bar{\mathbf{X}}_n - \mathbf{p}) \\ &= n(\bar{\mathbf{X}}_n - \mathbf{p})'\mathbb{D}(\mathbf{p}^{-1})(\bar{\mathbf{X}}_n - \mathbf{p}) \quad (\text{vea (2.1) y (2.2)}) \\ &= n \sum_{i=1}^k \frac{(\bar{\mathbf{X}}_{ni} - p_i)^2}{p_i} \end{aligned}$$

y en combinación con (6.3) se obtiene

$$n \sum_{i=1}^k \frac{(\bar{X}_{ni} - p_i)^2}{p_i} \xrightarrow{\mathcal{D}} \chi_{k-1}^2,$$

lo cual concluye la demostración del teorema.  $\square$

**Observación 6.1.** El vector  $\bar{\mathbf{X}}_n$  es la media de la muestra  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  de la distribución de Bernoulli dada en (1.1), de manera que  $\bar{X}_{ni}$  es la media de las  $i$ -ésimas componentes de los vectores  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ , i.e.,

$$\bar{X}_{ni} = \frac{1}{n} \sum_{j=1}^n X_{ji}.$$

Por otro lado, usando (1.1), no es difícil ver que  $\sum_{i=1}^k \bar{X}_{ni} = 1$ , de manera que

$$\begin{aligned} n \sum_{i=1}^k \frac{(\bar{X}_{ni} - p_i)^2}{p_i} &= n \sum_{i=1}^k \frac{\bar{X}_{ni}^2 - 2\bar{X}_{ni}p_i + p_i^2}{p_i} \\ &= n \left[ \sum_{i=1}^k \frac{\bar{X}_{ni}^2}{p_i} - 2 \sum_{i=1}^k \bar{X}_{ni} + \sum_{i=1}^k p_i \right] \\ &= n \sum_{i=1}^k \frac{\bar{X}_{ni}^2}{p_i} - n, \end{aligned}$$

de donde es claro que la conclusión del Teorema 6.1 equivale a

$$n \sum_{i=1}^k \frac{\bar{X}_{ni}^2}{p_i} - n \xrightarrow{\mathcal{D}} \chi_{k-1}^2,$$

o bien,

$$\sum_{i=1}^k \frac{T_{ni}^2}{E_i} - n \xrightarrow{\mathcal{D}} \chi_{k-1}^2,$$

donde  $T_{ni} = \sum_{j=1}^n X_{ji}$  es la cantidad de veces que se observa un resultado de la clase  $i$  en las primeras  $n$  repeticiones del experimento, mientras que  $E_i = np_i$  es el número de observaciones que *a priori* se espera obtener dentro de la clase  $i$  al realizar  $n$  ensayos.

**El Procedimiento de Prueba.** El Teorema 6.1 permite establecer un procedimiento asintótico para discernir entre las dos hipótesis en (1.1) sobre los parámetros de una distribución de Bernoulli. Sea  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  una muestra de la distribución  $\mathcal{B}(p_1, p_2, \dots, p_k)$  y, como en la introducción, sea  $\mathbf{q}$  un vector de probabilidad en  $\mathbb{R}^k$ . Considere el problema de discernir entre las hipótesis

$$\mathcal{H}_0: \mathbf{p} = \mathbf{q}, \quad \text{vs} \quad \mathcal{H}_a: \mathbf{p} \neq \mathbf{q}. \quad (6.4)$$

Bajo la hipótesis nula, el Teorema 6.1 garantiza que

$$n \sum_{i=1}^k \frac{(\bar{X}_{ni} - q_i)^2}{q_i} \xrightarrow{\mathcal{D}} \chi_{k-1}^2, \quad \text{conforme } n \rightarrow \infty,$$

y entonces, para cada  $\alpha \in (0, 1)$ ,

$$P \left[ n \sum_{i=1}^k \frac{(\bar{X}_{ni} - q_i)^2}{q_i} \leq \chi_{k-1, \alpha}^2 \mid \mathcal{H}_0 \right] \rightarrow 1 - \alpha$$

de manera que una prueba para (6.4) con nivel *aproximado* de significancia  $\alpha$ , está dada como sigue:

**Rechace  $\mathcal{H}_0$**  si y sólo si el estadístico de prueba  $\mathcal{T}$  dado por

$$\mathcal{T} := n \sum_{i=1}^k \frac{(\bar{X}_{ni} - q_i)^2}{q_i}$$

satisface

$$\mathcal{T} > \chi_{k-1, \alpha}^2.$$

El siguiente ejemplo ilustra este procedimiento.

**Ejemplo 6.1.** Para verificar si la gente tiene alguna inclinación distinta del azar al elegir un entero entre uno y diez, 1000 personas fueron seleccionadas aleatoriamente y se les pidió que escogieran un número entero entre uno y diez. Las frecuencias observadas se indican en las siguiente tablas.

Número seleccionado	1	2	3	4	5
Frecuencia observada	126	108	60	62	104

Número seleccionado	6	7	8	9	10
Frecuencia observada	98	92	114	116	120

En este caso, se observó una muestra de tamaño  $n = 1000$  de una distribución de Bernoulli de dimensión  $k = 10$ , y  $p_i$  representa la probabilidad de seleccionar el número  $i$  en cada ensayo. Las hipótesis entre las cuales se quiere discernir son

$$\mathcal{H}_0: p_i = .1 \text{ para todo } i, \quad \text{vs} \quad \mathcal{H}_a: p_i \neq .1 \text{ para algún } i.$$

Las medias observadas para cada clase son

$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_3$	$\bar{X}_4$	$\bar{X}_5$
.126	.108	.060	.062	.104

$\bar{X}_6$	$\bar{X}_7$	$\bar{X}_8$	$\bar{X}_9$	$\bar{X}_{10}$
.098	.092	.114	.116	.120

El valor observado del estadístico de prueba es  $T = 47.2$ , mientras que  $P[T > 47.2] \gg .995$ , de modo que la hipótesis nula se rechaza a cualquier nivel de significación mayor o igual a .005, e incluso para niveles menores; de hecho, la evidencia en contra de la hipótesis nula es contundente en este ejemplo.  $\square$

## Intervalos de Confianza

El Teorema 6.1 establecido en la sección precedente permite, con un poco de esfuerzo, construir (i) una región de confianza con nivel asintótico deseado para el vector de probabilidades  $\mathbf{p}$  de una distribución de Bernoulli, así como (ii) intervalos de confianza para combinaciones lineales de las componentes de  $\mathbf{p}$ .

Antes de delinear el procedimiento, se probará una versión ‘ligeramente distinta’ del Teorema 6.1.

**Teorema 7.1.** Si  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  es una sucesión de variables independientes con distribución común  $\mathcal{B}(p_1, p_2, \dots, p_k)$ , donde  $p_i > 0$  para  $i = 1, 2, \dots, k$ . Entonces,

$$n \sum_{i=1}^k \frac{(\bar{X}_{ni} - p_i)^2}{\bar{X}_{ni}} \xrightarrow{\mathcal{D}} \chi_{k-1}^2 \quad \text{conforme } n \rightarrow \infty. \quad (7.1)$$

La diferencia entre la conclusión de este resultado y aquella en el Teorema 6.1, consiste en que los denominadores  $p_i$  han sido sustituidos por sus estimaciones  $\bar{X}_{ni}$ , y es necesario hacer la siguiente convención para cubrir el caso en el que  $\bar{X}_{ni} = 0$ . En el resto del capítulo,  $a/0 \equiv 0$  para todo  $a \in \mathbb{R}$ ; esta convención que será tratada con sumo cuidado, facilita la escritura de diversas expresiones que aparecen a través de esta y la siguiente sección. La demostración del Teorema 7.1 se basa en la siguiente variante del Teorema 3.1.

**Lema 7.1.** Sea  $\{W_n\}$  una sucesión de variables aleatorias tales que  $W_n \xrightarrow{P} \mu \in \mathbb{R}$ , y sea  $g : \mathbb{R} \rightarrow \mathbb{R}$  una función continua en  $\mu$ . Entonces,

$$g(W_n) \xrightarrow{P} g(\mu).$$

**Demostración.** Como  $g$  es continua en  $\mu$ , dado  $\varepsilon > 0$  existe  $\rho > 0$  tal que  $|g(y) - g(\mu)| > \varepsilon \implies |y - \mu| > \rho$ . Luego,  $P[|g(W_n) - g(\mu)| > \varepsilon] \leq P[|W_n - \mu| > \rho] \rightarrow 0$ .  $\square$

**Demostración de Teorema 7.1.** Defina  $g(x) = \frac{1}{x}$ ,  $x \in \mathbb{R}$ . Note que de acuerdo a la convención establecida líneas arriba, esta definición puede expresarse equivalentemente como  $g(x) = \frac{1}{x}$ ,  $x \neq 0$  y  $g(0) = 0$ . Observe ahora que  $g$  es

continua fuera del origen, y que  $\bar{X}_{ni} \xrightarrow{P} p_i$ , por la ley débil de los grandes números. Luego, del Lema 7.1 se desprende que conforme  $n \rightarrow \infty$

$$g(\bar{X}_{ni}) \xrightarrow{P} g(p_i), \quad i = 1, 2, \dots, k. \quad (7.2)$$

Por otro lado,

$$\begin{aligned} n \sum_{i=1}^k \frac{(\bar{X}_{ni} - p_i)^2}{p_i} &= n \sum_{i=1}^k \frac{(\bar{X}_{ni} - p_i)^2}{\bar{X}_{ni}} + n \sum_{i=1}^k (\bar{X}_{ni} - p_i)^2 \left( \frac{1}{\bar{X}_{ni}} - \frac{1}{p_i} \right) \\ &= n \sum_{i=1}^k \frac{(\bar{X}_{ni} - p_i)^2}{\bar{X}_{ni}} + n \sum_{i=1}^k (\bar{X}_{ni} - p_i)^2 (g(\bar{X}_{ni}) - g(p_i)). \end{aligned} \quad (7.3)$$

Adicionalmente, el Teorema 6.1 y el Lema 4.1(b) implican que  $n \sum_{i=1}^k (\bar{X}_{ni} - p_i)^2 / p_i$  es acotada en probabilidad, y por lo tanto lo mismo ocurre con  $n(\bar{X}_{ni} - p_i)^2$  para cada  $i$ . Entonces, combinando (7.2) y el Lema 4.3 se concluye que  $n(\bar{X}_{ni} - p_i)^2 (g(\bar{X}_{ni}) - g(p_i)) \xrightarrow{P} 0$  para cada  $i = 1, 2, \dots, k$ , y por lo tanto

$$n \sum_{i=1}^k (\bar{X}_{ni} - p_i)^2 (g(\bar{X}_{ni}) - g(p_i)) \xrightarrow{P} 0.$$

Combinando esta convergencia con (7.3) y el Teorema 6.1, se concluye que

$$\begin{aligned} n \sum_{i=1}^k \frac{(\bar{X}_{ni} - p_i)^2}{\bar{X}_{ni}} &= n \sum_{i=1}^k \frac{(\bar{X}_{ni} - p_i)^2}{p_i} - n \sum_{i=1}^k (\bar{X}_{ni} - p_i)^2 \left( \frac{1}{\bar{X}_{ni}} - \frac{1}{p_i} \right) \\ &\xrightarrow{\mathcal{D}} Y + 0 = Y \sim \chi_{k-1}^2, \end{aligned}$$

donde se ha utilizado la parte (iv) en la Observación 2.1 para obtener la convergencia. Esto concluye la demostración del teorema.  $\square$

**Regiones e Intervalos de Confianza.** El Teorema 7.1 permite obtener una región de confianza para el vector de parámetros  $\mathbf{p}$  con nivel asintótico deseado. En efecto, si  $\alpha \in (0, 1)$ , entonces (vea (7.1))

$$\lim_{n \rightarrow \infty} P \left[ n \sum_{i=1}^k \frac{(\bar{X}_{ni} - p_i)^2}{\bar{X}_{ni}} \leq \chi_{k-1, \alpha}^2 \right] = 1 - \alpha.$$

Por lo tanto, cuando  $n$  es suficientemente grande, hay una probabilidad aproximadamente igual a  $1 - \alpha$ , de que el siguiente enunciado sea válido:

$$n \sum_{i=1}^k \frac{(\bar{X}_{ni} - p_i)^2}{\bar{X}_{ni}} \leq \chi_{k-1, \alpha}^2,$$

el cual es equivalente a

$$\sum_{i=1}^k \frac{(\bar{X}_{ni} - p_i)^2}{(\chi_{k-1, \alpha}^2 \bar{X}_{ni})/n} \leq 1. \quad (7.4)$$

Una vez que se han observado los valores asumidos por  $\bar{X}_{ni}$  para  $i = 1, 2, \dots, k$ , esta desigualdad prescribe que  $\mathbf{p} = (p_1, p_2, \dots, p_k)'$  pertenece al elipsoide con centro en  $\bar{\mathbf{X}}_n$  y semiejes de longitud  $(\chi_{k-1, \alpha}^2 \bar{X}_{ni})/n$  a lo largo de los ejes coordenados  $p_i$  para  $i = 1, 2, \dots, k$ . Por el Teorema 7.1, este elipsoide es una **región de confianza** para  $\mathbf{p}$  con nivel de confianza asintótico  $1 - \alpha$ . Por otro lado, en las aplicaciones es de interés comparar las diversas probabilidades  $p_i$ , y de manera concreta es deseable obtener intervalos de confianza para funciones lineales  $\mathbf{c}'\mathbf{p}$  donde  $\mathbf{c} = (c_1, c_2, \dots, c_k)$ , los cuales se pueden obtener combinando (7.4) con el Lema 7.2 que se establecerá más adelante. Primero defina el vector  $\psi$  mediante

$$\psi = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_k \end{bmatrix} := \begin{bmatrix} \frac{1}{\chi_{k-1, \alpha}^2 \bar{X}_{n1}/n} \\ 1 \\ \frac{1}{\chi_{k-1, \alpha}^2 \bar{X}_{n2}/n} \\ \vdots \\ 1 \\ \frac{1}{\chi_{k-1, \alpha}^2 \bar{X}_{nk}/n} \end{bmatrix} \quad (7.5)$$

(recuerde que la convención  $a/0 = 0$  aún está en pie), y considere  $\text{ID}(\psi)$ , la cual es una matriz diagonal de orden  $k \times k$  con componentes  $\psi_1, \psi_2, \dots, \psi_k$  a lo largo de la diagonal principal (vea (2.1)). Como  $\psi_i$  es siempre no negativo, es claro que  $\text{ID}(\psi)$  es una matriz no negativa definida. El siguiente resultado es una versión de la desigualdad de Cauchy–Schwartz para matrices no negativas definidas, y se ha utilizado en la construcción de intervalos de confianza simultáneos por medio del método de Scheffè en el análisis de modelos lineales (Graybill, 1985, Christensen, 1996).

**Lema 7.2.** Sea  $M$  una matriz no negativa definida de orden  $k \times k$ . Entonces para todo  $\mathbf{v}$  y  $\mathbf{w}$  en  $\mathbb{R}^k$

$$|\mathbf{w}'M\mathbf{v}| \leq \sqrt{\mathbf{v}'M\mathbf{v}}\sqrt{\mathbf{w}'M\mathbf{w}}.$$

Por lo tanto,

$$|\mathbf{v}'M\mathbf{v}| \leq 1 \iff |\mathbf{w}'M\mathbf{v}| \leq \sqrt{\mathbf{w}'M\mathbf{w}} \text{ para todo } \mathbf{w} \in \mathbb{R}^k.$$

Este resultado puede demostrarse a través de argumentos similares a los utilizados en el caso en que la matriz  $M$  es positiva definida, en cuyo caso  $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}'M\mathbf{y}$  determina un producto interno en  $\mathbb{R}^k$ . Note ahora que (7.4) es equivalente a  $(\bar{\mathbf{X}}_n - \mathbf{p})'\text{ID}(\psi)(\bar{\mathbf{X}}_n - \mu) \leq 1$  y entonces el Lema 7.2 permite establecer que

$$(\bar{\mathbf{X}}_n - \mathbf{p})'\text{ID}(\psi)(\bar{\mathbf{X}}_n - \mu) \leq 1 \iff |\mathbf{w}'\text{ID}(\psi)(\bar{\mathbf{X}}_n - \mu)| \leq \sqrt{\mathbf{w}'\text{ID}(\psi)\mathbf{w}}, \quad \mathbf{w} \in \mathbb{R}^k. \quad (7.6)$$

Sea ahora  $\mathbf{c} = (c_1, c_2, \dots, c_k)$  un vector dado en  $\mathbb{R}^k$  y defina

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix} := \begin{bmatrix} c_1 \chi_{k-1, \alpha}^2 \bar{X}_{n1}/n \\ c_2 \chi_{k-1, \alpha}^2 \bar{X}_{n2}/n \\ \vdots \\ c_k \chi_{k-1, \alpha}^2 \bar{X}_{nk}/n \end{bmatrix} \quad (7.7)$$

Usando la convención  $a/0 = 0$ , conjuntamente con la definición de  $\psi$ , no es difícil ver que

$$w_i \psi_i = \begin{cases} c_i, & \text{si } \bar{X}_{ni} \neq 0 \\ 0, & \text{de otro modo,} \end{cases} \quad (7.8)$$

y entonces (vea (7.7) para la definición de  $w_i$ )  $w_i \psi_i w_i = c_i w_i = c_i^2 \chi_{k-1, \alpha}^2 \bar{X}_{ni}/n$  si  $\bar{X}_{ni} \neq 0$ , mientras que  $w_i \psi_i w_i = 0 = c_i^2 \chi_{k-1, \alpha}^2 \bar{X}_{ni}/n$  si  $\bar{X}_{ni} = 0$ , de manera que se tiene la siguiente igualdad, válida en cualquier circunstancia:

$$w_i \psi_i w_i = \frac{\chi_{k-1, \alpha}^2}{n} c_i^2 \bar{X}_{ni}. \quad (7.9)$$

Además, (7.8) es claramente equivalente a  $w_i \psi_i = c_i I[\bar{X}_{ni} \neq 0]$ , de manera que

$$w_i \psi_i \bar{X}_{ni} = c_i I[\bar{X}_{ni} \neq 0] \bar{X}_{ni} = c_i \bar{X}_{ni}, \quad \text{y} \quad w_i \psi_i p_i = c_i p_i I[\bar{X}_{ni} \neq 0]. \quad (7.10)$$

Por lo tanto,

$$\mathbf{w}'\mathbb{D}(\psi)\mathbf{w} = \sum_{i=1}^k w_i \psi_i w_i = \frac{\chi_{k-1, \alpha}^2}{n} \sum_{i=1}^k \mathbf{c}_i^2 \bar{X}_{ni},$$

y

$$\mathbf{w}'\mathbb{D}(\psi)(\bar{\mathbf{X}}_n - \mathbf{p}) = \sum_{i=1}^k w_i \psi_i (X_{ni} - p_i) = \sum_{i=1}^k c_i X_{ni} - \sum_{i=1}^k c_i p_i I[X_{ni} \neq 0],$$

igualdades que usando (7.6) permiten establecer la siguiente afirmación:

$$(\bar{\mathbf{X}}_n - \mathbf{p})'\mathbb{D}(\psi)(\bar{\mathbf{X}}_n - \mu) \leq 1$$

$$\Updownarrow \tag{7.11}$$

$$\left| \sum_{i=1}^k c_i X_{ni} - \sum_{i=1}^k c_i \mathbf{p}_i I[X_{ni} \neq 0] \right| \leq \sqrt{\frac{\chi_{k-1, \alpha}^2}{n}} \sqrt{\sum_{i=1}^k \mathbf{c}_i^2 \bar{X}_{ni}} \quad \text{para todo } \mathbf{c} \in \mathbb{R}^k.$$

Como ya ha sido observado, (7.4) equivale a  $(\bar{\mathbf{X}}_n - \mathbf{p})'\mathbb{D}(\psi)(\bar{\mathbf{X}}_n - \mathbf{p}) \leq 1$ , y el Teorema 7.1 implica que

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left[ n \sum_{i=1}^k \frac{(\bar{X}_{ni} - p_i)^2}{\bar{X}_{ni}} \leq \chi_{k-1, \alpha}^2 \right] \\ = \lim_{n \rightarrow \infty} P[(\bar{\mathbf{X}}_n - \mathbf{p})'\mathbb{D}(\psi)(\bar{\mathbf{X}}_n - \mathbf{p})] = 1 - \alpha \end{aligned}$$

y a partir de (7.11) se obtiene la siguiente conclusión.

### Intervalos de Confianza Simultáneos para las Componentes de $\mathbf{p}$ .

Con cada  $\mathbf{c} \in \mathbb{R}^k$ , asocie las siguiente desigualdades:

$$\begin{aligned} \sum_{i=1}^k c_i \bar{X}_{ni} - \sqrt{\frac{\chi_{k-1, \alpha}^2}{n}} \sqrt{\sum_{i=1}^k \mathbf{c}_i^2 \bar{X}_{ni}} \\ \leq \sum_{i=1}^k c_i \mathbf{p}_i I[X_{ni} \neq 0] \leq \sum_{i=1}^k c_i \bar{X}_{ni} + \sqrt{\frac{\chi_{k-1, \alpha}^2}{n}} \sqrt{\sum_{i=1}^k \mathbf{c}_i^2 \bar{X}_{ni}}. \end{aligned} \tag{7.12}$$

Entonces, la probabilidad de que estas desigualdades sean *simultáneamente ciertas* para todo  $\mathbf{c} \in \mathbb{R}^k$  converge a  $1 - \alpha$  conforme  $n \rightarrow \infty$ . Esto es, los intervalos

determinados por (7.11) tienen un coeficiente de confianza asintótico igual a  $1 - \alpha$ . En la práctica, (7.12) puede utilizarse, después de observar los datos, para obtener intervalos de confianza para combinaciones lineales de las componentes de  $\mathbf{p}$  que corresponden a medias muestrales no nulas. Con el propósito de ilustrar el procedimiento, se utilizarán los datos del Ejemplo 6.1.

**Continuación del Ejemplo 6.1.** Después de obtener los datos, en el Ejemplo 6.1 se tiene que todas las medias muestrales son no nulas. La mayor media muestral es  $\bar{X}_1 = .126$ , mientras que la menor es  $\bar{X}_3 = .060$ , la cual tiene un valor de menos del 50 por ciento de  $\bar{X}_1$ , y entre otras cuestiones, es interesante preguntarse cual es la diferencia real entre  $p_1$  y  $p_3$ , así que se determinará un intervalo de confianza para  $p_1 - p_3 = \mathbf{c}'\mathbf{p}$  con nivel aproximado de  $95\% = 1 - \alpha$ . Note que en este ejemplo  $k = 10$  y  $n = 1000$ , de manera que

$$\sqrt{\frac{\chi_{k-1, \alpha}^2}{n}} = \sqrt{\frac{16.9190}{1000}} = .130073$$

Además,  $c_i = 0$  para todo  $i \neq 1, 3$ , mientras que  $c_1 = 1$  y  $c_3 = -1$ , de manera que

$$\sum_{i=1}^{10} c_i \bar{X}_{ni} = .1260 - .0600 = .0660$$

y

$$\sqrt{\sum_{i=1}^{10} c_1^2 \bar{X}_i} = \sqrt{\bar{X}_1 + \bar{X}_3} = \sqrt{.126 + .060} = \sqrt{.186} = .4313,$$

entonces

$$\sqrt{\frac{\chi_{k-1, \alpha}^2}{n}} \sqrt{\sum_{i=1}^{10} c_1^2 \bar{X}_i} = .0561.$$

Usando (7.12) el intervalo de confianza que se obtiene para  $p_1 - p_3$  es

$$.0660 - .0561 = .0099 \leq p_1 - p_3 \leq .0660 + .0561 = .1221. \quad (7.13)$$

El aspecto relevante es que es posible continuar estableciendo intervalos de confianza de este tipo para todas las funciones lineales  $\mathbf{c}'\mathbf{p}$  que se consideren interesantes, y se tendrá que la confianza en la veracidad simultánea de todos ellos será, de al menos 95%.

## Observaciones Adicionales

En esta sección se explora un posible refinamiento sobre el tema tratado en la Sección precedente. Se trata de explorar la posibilidad de construir intervalos de confianza con nivel asintótico predeterminado que resultan más cortos que aquellos determinados por (7.12).

**Una Nota Sobre los Intervalos de Confianza Simultáneos.** Los intervalos de confianza simultáneos en (7.12) se constuyeron, via el Lemma 7.2, tomando como punto de partida la convergencia

$$\begin{aligned} \lim_{n \rightarrow \infty} P[(\bar{\mathbf{X}}_n - \mathbf{p})' \mathbb{D}(\psi)(\bar{\mathbf{X}}_n - \mathbf{p}) \leq 1] &= \lim_{n \rightarrow \infty} P \left[ n \sum_{i=1}^k \frac{(\bar{X}_{ni} - p_i)^2}{\bar{X}_{ni}} \leq \chi_{k-1, \alpha}^2 \right] \\ &= 1 - \alpha, \end{aligned} \tag{8.1}$$

la cual implica que después de evaluar  $\bar{\mathbf{X}}_n$ , el elipsoide

$$(\bar{\mathbf{X}}_n - \mathbf{p})' \mathbb{D}(\psi)(\bar{\mathbf{X}}_n - \mathbf{p}) \leq 1 \tag{8.2}$$

es una región de confianza para  $\mathbf{p}$  con nivel asintótico de confianza  $1 - \alpha$ . Sin embargo, se sabe que  $\mathbf{1}'\mathbf{p} = 1$ , así que en realidad, la región de confianza (8.2) puede sustituirse por

$$(\bar{\mathbf{X}}_n - \mathbf{p})' \mathbb{D}(\psi)(\bar{\mathbf{X}}_n - \mathbf{p}) \leq 1, \quad \mathbf{1}'\mathbf{p} = 1. \tag{8.3}$$

La pregunta natural ahora es si al utilizar estas condiciones como punto inicial en lugar de (8.2) es posible obtener intervalos de confianza simultáneos para las combinaciones lineales de  $\mathbf{p}$  que sean más cortos que aquellos determinados en (7.12). Como se muestra en el Teorema 8.1 establecido más adelante, la respuesta es afirmativa, y el argumento que conduce a esta conclusión está basado en el siguiente lema.

**Lema 8.1.** Sea  $M$  una matriz no negativa definida de dimensión  $k \times k$ , y sea  $\mathbf{u} \in \mathbb{R}^k$  un vector que satisface  $\mathbf{u}'M\mathbf{u} > 0$ . Entonces, para cada  $\mathbf{w} \in \mathbb{R}^k$ , las siguientes afirmaciones (i) y (ii) son equivalentes.

(i)  $\mathbf{w}'M\mathbf{w} \leq 1$  y  $\mathbf{u}'M\mathbf{w} = 0$ .

(ii) Para todo  $\mathbf{c} \in \mathbb{R}^k$ ,

$$|\mathbf{c}'M\mathbf{w}| \leq \sqrt{\mathbf{c}'M\mathbf{c} - \frac{(\mathbf{u}'M\mathbf{c})^2}{\mathbf{u}'M\mathbf{u}}}. \quad (8.4)$$

**Demostración.** (ii)  $\implies$  (i). Suponga que (8.4) es válida para todo  $\mathbf{c} \in \mathbb{R}^k$ . En este caso, seleccionando  $\mathbf{c} = \mathbf{w}$ , la desigualdad se transforma en

$$\mathbf{w}'M\mathbf{w} \leq \sqrt{\mathbf{w}'M\mathbf{w} - \frac{(\mathbf{u}'M\mathbf{w})^2}{\mathbf{u}'M\mathbf{u}}}$$

de donde se desprende que  $\mathbf{w}'M\mathbf{w} \leq \sqrt{\mathbf{w}'M\mathbf{w}}$  y entonces  $|\mathbf{w}'M\mathbf{w}| \leq 1$ . Ahora ponga  $\mathbf{c} = \mathbf{u}$  en (8.4), para obtener

$$|\mathbf{u}'M\mathbf{w}| \leq \sqrt{\mathbf{u}'M\mathbf{u} - \frac{(\mathbf{u}'M\mathbf{u})^2}{\mathbf{u}'M\mathbf{u}}} = \sqrt{\mathbf{u}'M\mathbf{u} - (\mathbf{u}'M\mathbf{u})} = 0,$$

completando la demostración.

(i)  $\implies$  (ii). Suponga que  $\mathbf{w}'M\mathbf{w} \leq 1$  y que  $\mathbf{u}'M\mathbf{w} = 0$ . Como  $\mathbf{u}'M\mathbf{w} = 0$ ,

$$\left( \mathbf{c} - \frac{(\mathbf{u}'M\mathbf{c})}{\mathbf{u}'M\mathbf{u}} \mathbf{u} \right)' M \mathbf{w} = \mathbf{c}'M\mathbf{u} - \frac{(\mathbf{u}'M\mathbf{c})}{\mathbf{u}'M\mathbf{u}} \mathbf{u}'M\mathbf{w} = \mathbf{c}'M\mathbf{u},$$

y entonces, combinando la desigualdad de Cauchy-Schwartz con  $\mathbf{w}'M\mathbf{w} \leq 1$ , se obtiene

$$\begin{aligned} |\mathbf{c}'M\mathbf{u}| &= \left| \left( \mathbf{c} - \frac{(\mathbf{u}'M\mathbf{c})}{\mathbf{u}'M\mathbf{u}} \mathbf{u} \right)' M \mathbf{w} \right| \\ &\leq \sqrt{\left( \mathbf{c} - \frac{(\mathbf{u}'M\mathbf{c})}{\mathbf{u}'M\mathbf{u}} \mathbf{u} \right)' M \left( \mathbf{c} - \frac{(\mathbf{u}'M\mathbf{c})}{\mathbf{u}'M\mathbf{u}} \mathbf{u} \right)} \sqrt{\mathbf{w}'M\mathbf{w}} \\ &\leq \sqrt{\left( \mathbf{c} - \frac{(\mathbf{u}'M\mathbf{c})}{\mathbf{u}'M\mathbf{u}} \mathbf{u} \right)' M \left( \mathbf{c} - \frac{(\mathbf{u}'M\mathbf{c})}{\mathbf{u}'M\mathbf{u}} \mathbf{u} \right)}; \end{aligned}$$

observando que

$$\left( \mathbf{c} - \frac{(\mathbf{u}'M\mathbf{c})}{\mathbf{u}'M\mathbf{u}} \mathbf{u} \right)' M \left( \mathbf{c} - \frac{(\mathbf{u}'M\mathbf{c})}{\mathbf{u}'M\mathbf{u}} \mathbf{u} \right) = \mathbf{c}'M\mathbf{c} - \frac{(\mathbf{u}'M\mathbf{c})^2}{\mathbf{u}'M\mathbf{u}}$$

se concluye que (8.4) ocurre.  $\square$

**Teorema 8.1.** Sea  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  una muestra de la distribución  $k$ -dimensional de Bernoulli con parámetro  $\mathbf{p} = (p_1, p_2, \dots, p_k)$ . En este caso, para todo  $\alpha \in (0, 1)$  se tiene que conforme  $n \rightarrow \infty$ , la probabilidad de que las siguientes desigualdades (8.5) sean (simultáneamente) válidas para todo  $\mathbf{c} \in \mathbb{R}^k$  converge a  $1 - \alpha$ .

$$\begin{aligned} \sum_{i=1}^k c_i \bar{X}_{ni} - \sqrt{\frac{\chi_{k-1, \alpha}^2}{n}} \sqrt{\sum_{i=1}^k c_i^2 \bar{X}_{ni} - \left( \sum_{i=1}^k c_i \bar{X}_{ni} \right)^2} \\ \leq \sum_{i=1}^k c_i p_i \leq \sum_{i=1}^k c_i \bar{X}_{ni} + \sqrt{\frac{\chi_{k-1, \alpha}^2}{n}} \sqrt{\sum_{i=1}^k c_i^2 \bar{X}_{ni} - \left( \sum_{i=1}^k c_i \bar{X}_{ni} \right)^2}. \end{aligned} \quad (8.5)$$

**Demostración.** Como  $\{\bar{X}_{ni}\}$  converge a  $p_i > 0$  casi en toda parte (por la ley fuerte de los grandes números), y el interés se centra en la probabilidad de que (8.5) ocurra conforme  $n \rightarrow \infty$ , no se pierde generalidad al suponer a continuación que  $\bar{X}_{ni} > 0$  para todo  $i = 1, 2, \dots, k$ . Debido a que la probabilidad de que (8.3) ocurra converge a  $1 - \alpha$ , para establecer el teorema es suficiente verificar la equivalencia de (8.3) con la validez de (8.5) para todo  $\mathbf{c} \in \mathbb{R}^k$ . Para establecer esta propiedad, defina

$$\mathbf{u} = \mathbb{D}(\psi^{-1})\mathbf{1} \quad (8.6)$$

y note que

$$\mathbf{1}'\mathbf{v} = (\mathbb{D}(\psi^{-1})\mathbf{1})'\mathbb{D}(\psi)\mathbf{v} = \mathbf{u}'\mathbb{D}(\psi)\mathbf{v} \quad (8.7)$$

para todo vector  $\mathbf{v} \in \mathbb{R}^k$ . Como  $\mathbf{1}'\bar{\mathbf{X}}_n = 1$ , se tiene que  $\mathbf{1}'\mathbf{p} = 1$  puede escribirse como  $\mathbf{1}'(\bar{\mathbf{X}}_n - \mathbf{p}) = 0$ , de manera que usando (8.6) y (8.7), la relación (8.3) equivale a

$$(\bar{\mathbf{X}}_n - \mathbf{p})'\mathbb{D}(\psi)(\bar{\mathbf{X}}_n - \mathbf{p}) \leq 1, \quad \mathbf{u}'\mathbb{D}(\psi)(\bar{\mathbf{X}}_n - \mathbf{p}) = 0,$$

y, via el Lema 8.1, este enunciado es equivalente a

$$|\mathbf{v}'\mathbb{D}(\psi)(\bar{\mathbf{X}}_n - \mathbf{p})| \leq \sqrt{\mathbf{v}'\mathbb{D}(\psi)\mathbf{v} - \frac{(\mathbf{u}'\mathbb{D}(\psi)\mathbf{v})^2}{\mathbf{u}'\mathbb{D}(\psi)\mathbf{u}}}, \quad \mathbf{v} \in \mathbb{R}^k,$$

o, sustituyendo  $\mathbb{D}(\psi)\mathbf{v}$  por  $\mathbf{c}$ , a

$$|\mathbf{c}'(\bar{\mathbf{X}}_n) - \mathbf{p}| \leq \sqrt{\mathbf{c}'\mathbb{D}(\psi^{-1})\mathbf{c} - \frac{(\mathbf{u}'\mathbf{c})^2}{\mathbf{u}'\mathbb{D}(\psi)\mathbf{u}}}, \quad \mathbf{c} \in \mathbb{R}^k, \quad (8.8)$$

donde se a utilizado que  $\mathbb{D}(\psi)$  es inversible. Usando (2.2), (7.5) y (8.6), es claro que

$$\begin{aligned} \mathbf{c}'\mathbb{D}(\psi^{-1})\mathbf{c} &= \sum_{i=1}^k \mathbf{c}_i^2 \psi_i^{-1} = \frac{\chi_{k-1, \alpha}^2}{n} \sum_{i=1}^k c_i^2 \bar{X}_{ni}; \\ \mathbf{u}'\mathbf{c} &= \mathbf{1}'\mathbb{D}(\psi^{-1})\mathbf{c} = \frac{\chi_{k-1, \alpha}^2}{n} \sum_{i=1}^k c_i \bar{X}_{ni}, \end{aligned}$$

y

$$\mathbf{u}'\mathbb{D}(\psi)\mathbf{u} = \mathbf{1}'\mathbb{D}(\psi^{-1})\mathbb{D}(\psi)\mathbb{D}(\psi^{-1})\mathbf{1} = \mathbf{1}'\mathbb{D}(\psi^{-1})\mathbf{1} = \sum_{i=1}^k \frac{\chi_{k-1, \alpha}^2}{n} \bar{X}_{ni} = \frac{\chi_{k-1, \alpha}^2}{n},$$

de manera que (8.8) es equivalente a

$$|\mathbf{c}'(\bar{\mathbf{X}}_n) - \mathbf{p}| \leq \sqrt{\frac{\chi_{k-1}^2}{n} \sqrt{\sum_{i=1}^k c_i^2 \bar{X}_{ni} - \left(\sum_{i=1}^k c_i \bar{X}_{ni}\right)^2}},$$

desigualdad que es una forma alternativa de (8.5). En resumen, se ha mostrado que (8.3) y (8.5) son enunciados equivalentes cuando  $\bar{X}_{ni} \neq 0$ , y como se mencionó anteriormente, esto concluye la demostración.  $\square$

Es claro que los intervalos (8.5) tienen un nivel de confianza asintótico igual a  $1 - \alpha$ , y que tienen menor longitud que aquellos en (7.12). El decremento relativo del tamaño de los intervalos es

$$1 - \sqrt{1 - \left(\sum_{i=1}^k c_i \bar{X}_{ni}\right)^2 / \sum_{i=1}^k c_i^2 \bar{X}_{ni}},$$

el cual, a menos que  $\left|\sum_{i=1}^k c_i \bar{X}_{ni}\right|$  sea ‘relativamente grande’ comparado con  $\sum_{i=1}^k c_i^2 \bar{X}_{ni}$ , no significa una disminución espectacular. Por ejemplo, considere el contexto del Ejemplo 6.1. Al construir el intervalo de confianza (8.5) para  $p_1 - p_3$  correspondiente a  $1 - \alpha = .95$ , se obtiene

$$.0106 \leq p_1 - p_3 \leq .1214,$$

cuya longitud es .1108, mientras que usando (7.12), la longitud del correspondiente intervalo es (vea (7.13)), .1122 y el decremento relativo de las longitudes es  $(.1122 - .1108)/.1122 = .0014/.1122 = 1.4/112.2$  apenas superior a uno por ciento. En resumen, los intervalos de confianza simultáneos en (8.5), tienen longitud menor a la de aquellos en 7.12, pero generalmente la reducción no es sustancial.

## Una extensión del Método Scheffè

Los resultados sobre intervalos de confianza simultáneos en las secciones precedentes se obtuvieron aplicando las ideas del procedimiento conocido como método de Scheffè en el análisis de modelos lineales. Al llegar a este punto, surge naturalmente una interrogante: Considere los intervalos de confianza en (8.5), y suponga que  $c$  se restringe a un subespacio  $\mathcal{S} \subset \mathbb{R}^k$  de dimensión  $r$ . ¿Es posible en estas circunstancias disminuir el número de grados de libertad utilizado para evaluar el coeficiente  $\sqrt{\chi_{k-1, \alpha}^2/n}$  y de esta forma producir intervalos más cortos? El objetivo de esta sección es establecer la siguiente respuesta afirmativa a esta pregunta.

**Teorema 9.1.** Sea  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n \dots$  una sucesión de vectores aleatorios independientes con distribución  $\beta(p_1, p_2, \dots, p_k)$ . Considere un subespacio  $\mathcal{S} \subset \mathbb{R}^k$  de dimensión  $r$ , tal que  $\mathcal{S}$  es ortogonal al vector  $\mathbf{1} = (1, 1, \dots, 1)' \in \mathbb{R}^k$ . En este caso, para todo  $\alpha \in (0, 1)$  se tiene que, conforme  $\mathbf{n} \rightarrow \infty$ , cualquier punto límite de la probabilidad de que *todas* las siguientes desigualdades sean válidas es mayor o igual a  $1 - \alpha$ :

$$\begin{aligned} \sum_{i=1}^k c_i \bar{X}_{ni} - \sqrt{\frac{\chi_{k-1, \alpha}^2}{n}} \sqrt{\sum_{i=1}^k c_i^2 \bar{X}_{ni} - \left( \sum_{i=1}^k c_i \bar{X}_i \right)^2} \\ \leq \sum_{i=1}^k c_i \mathbf{p}_i \leq \sum_{i=1}^k c_i \bar{X}_{ni} + \sqrt{\frac{\chi_{k-1, \alpha}^2}{n}} \sqrt{\sum_{i=1}^k c_i^2 \bar{X}_{ni} - \left( \sum_{i=1}^k c_i \bar{X}_i \right)^2}. \end{aligned} \quad (9.1)$$

En palabras, este resultado establece que (9.1) determina intervalos de confianza para las combinaciones lineales  $\mathbf{c}'\mathbf{p}$  cuyo coeficiente de confianza simultáneo no crecerá - cuando  $n$  sea 'suficientemente grande'- sustancialmente por debajo de  $1 - \alpha$  siempre que  $\mathbf{c}$  se restrinja al subespacio  $\mathcal{S}$ , el cual tiene dimensión  $r$  y es perpendicular a  $\mathbf{1}$ . Además de las ventajas técnicas de restringirse a subespacios perpendiculares a  $\mathbf{1}$ , hay un sustento práctico en esta elección, el cual puede resumirse diciendo que ya se sabe, de antemano, que  $\mathbf{1}'\mathbf{p} = 1$ , por lo que no es natural plantearse la necesidad de construir un intervalo de confianza para  $\mathbf{c}'\mathbf{p}$  si acaso  $\mathbf{c}$  no es ortogonal a  $\mathbf{1}$ . En efecto, si  $\mathbf{c} = \mathbf{c}_1 + \beta\mathbf{1}$ , donde  $\mathbf{c}_1$  es ortogonal a  $\mathbf{1}$ , entonces  $\mathbf{c}'\mathbf{p} = \mathbf{c}_1'\mathbf{p} + \beta$ , de manera que un intervalo de confianza para  $\mathbf{c}'\mathbf{p}$  se tendrá tan pronto se construya uno para  $\mathbf{c}_1'\mathbf{p}$ . Esto significa que al tratar de estimar la función  $\mathbf{c}'\mathbf{p}$  mediante un intervalo, siempre se puede extraer de  $\mathbf{c}$  la componente a lo largo del vector  $\mathbf{1}$  y restringirse, por lo tanto, a vectores perpendiculares a  $\mathbf{1}$ . Por otro lado, note que la diferencia entre los intervalos (9.1) y (8.5), radica en el número de grados de libertad utilizados para calcular el percentil de la distribución  $\chi^2$ .

**Continuación del Ejemplo 6.1** Es claro que el percentil utilizado en el Teorema 9.1 es menor al empleado en (8.5), de manera que los intervalos (9.1) son, naturalmente, más cortos que aquellos en (8.5). La disminución porcentual en la longitud esta dada por

$$1 - \sqrt{\chi_{r,\alpha}^2 / \chi_{k-1,\alpha}^2}$$

En la situación del Ejemplo 6.1, suponga que se desea construir intervalos de confianza para combinaciones lineales  $\mathbf{c}'\mathbf{p}$  donde  $\mathbf{c} \in \mathcal{S}$  y  $\mathcal{S}$  tiene dimensión 4. Suponga además que  $p_1 - p_3$  es una de esas combinaciones. En este caso, usando (9.1) se obtiene que  $0.486 \leq p_1 - p_3 \leq 0.0836$ , un intervalo cuya longitud es 0.0356, alrededor de 30% de la longitud del correspondiente intervalo calculado en la Sección 8. □

La demostración del Teorema 9.1 se dará a continuación y es bastante

técnica. La razón fundamental es que en la convergencia del Teorema 2.1, la matriz de covarianza de la distribución límite, a saber,  $\mathbb{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}'$  no es homoscedástica, esto es, sus términos diagonales no son iguales. El argumento se ha dividido en cuatro partes que se presentan a continuación como los Lemas 9.1-9.4. Antes de continuar es conveniente puntualizar algunos hechos básicos que serán utilizados en el resto de la sección.

**Observación 9.1** Defina la matriz  $\mathbb{M}(\mathbf{p})$  mediante

$$\mathbb{M}(\mathbf{p}) = \mathbb{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}',$$

de manera que  $\mathbb{M}(\mathbf{p})$  es la matriz de covarianza de la distribución  $\beta(p_1, p_2, \dots, p_k)$ . En este caso, es claro que  $\mathbb{M}(\mathbf{p})$  tiene rango  $k - 1$ , y que su núcleo está generado por el vector  $\mathbf{1}$ ; de hecho como  $\mathbf{p}'\mathbf{1} = 1$ ,

$$\mathbb{M}(\mathbf{p})\mathbf{1} = \mathbb{D}(\mathbf{p})\mathbf{1} - \mathbf{p}\mathbf{p}'\mathbf{1} = \mathbf{p} - \mathbf{p} = 0.$$

Como con toda matriz no negativa definida, las potencias  $\mathbb{M}(\mathbf{p})^\gamma$  están definidas (y son matrices no negativas) para todo  $\gamma \in \mathbb{R}$ , y se tiene que

$$\mathbb{M}(\mathbf{p})^\gamma \mathbb{M}(\mathbf{p})^{-\gamma} = Q.$$

donde  $Q$  es la (matriz de) proyección ortogonal sobre el espacio de columnas de  $\mathbb{M}(\mathbf{p})$ ; éste último, es el complemento ortogonal al núcleo de  $\mathbb{M}(\mathbf{p})$ , el cual, como ya se mencionó, está generado por el vector  $\mathbf{1}$ . Por lo tanto,

$$\mathbb{M}(\mathbf{p})^\gamma \mathbb{M}(\mathbf{p})^{-\gamma} = Q = I - \frac{1}{n} \mathbf{1}\mathbf{1}'.$$

Particularmente, como se supone que  $\mathcal{S}$  es un subespacio ortogonal a  $\mathbf{1}$ , se tiene que

$$\mathbf{c} = Q\mathbf{c} = \mathbb{M}(\mathbf{p})^{-1/2} \mathbb{M}(\mathbf{p})^{1/2} \mathbf{c},$$

y entonces

$$\mathbf{c}'\mathbf{Z} = (\mathbb{M}(\mathbf{p})^{-1/2} \mathbb{M}(\mathbf{p})^{1/2} \mathbf{c})'\mathbf{Z} = (\mathbb{M}(\mathbf{p})^{1/2} \mathbf{c})' \mathbb{M}(\mathbf{p})^{-1/2} \mathbf{Z},$$

para todo  $\mathbf{Z} \in \mathbb{R}^k$ ; esta igualdad desempeña un papel importante, aunque sutil, en la Demostración del Lema 9.3.

El punto de arranque hacia la demostración del Teorema 9.1 es el siguiente.

**Lema 9.1.** Sea  $\mathbf{W}$  un vector aleatorio de dimensión  $k$  tal que  $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, I - \mathbf{v}\mathbf{v}')$  donde  $\mathbf{v}\mathbf{v}' = 1$  y sea  $Y$  una variable aleatoria con distribución  $\chi_r^2$ . Suponga además que  $\nu \subset \mathbb{R}^k$  es un subespacio de dimensión  $r$ . En este caso, si  $P_\nu$  es la proyección ortogonal sobre  $\nu$ , se tiene que

$$|P_\nu \mathbf{W}|^2 \text{ es asintóticamente menor a } Y,$$

esto es, para todo  $z > 0$ ,

$$P[|P_\nu \mathbf{W}|^2 \leq z] \geq P[Y \leq z];$$

en particular

$$P[|P_\nu \mathbf{W}|^2 \leq \chi_{r,\alpha}^2] \geq 1 - \alpha, \quad \alpha \in (0,1).$$

**Demostración.** Primero observe que

$$P_\nu \mathbf{W} \sim \mathcal{N}(\mathbf{0}, P_\nu(I - \mathbf{v}\mathbf{v}')P_\nu) = \mathcal{N}(\mathbf{0}, P_\nu - \mathbf{w}\mathbf{w}'), \quad (9.2)$$

donde  $\mathbf{w} = \mathbb{P}_\nu \mathbf{v} \in \nu$ . A continuación se determinarán los valores propios de  $P_\nu - \mathbf{w}\mathbf{w}'$ . Con este fin considere las siguientes posibilidades (i) y (ii):

(i)  $\mathbf{w} \neq \mathbf{0}$ .

Bajo esta condición, sea  $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{r-1}, \mathbf{w}_r\}$  una base ortonormal del espacio  $\nu$ , donde  $\mathbf{w}_r = \mathbf{w} / |\mathbf{w}|$ . De esta manera,

$$(P_\nu - \mathbf{w}\mathbf{w}')\mathbf{w}_i = P_\nu \mathbf{w}_i - \mathbf{w}\mathbf{w}'\mathbf{w}_i = \mathbf{w}_i, \quad i = 1, 2, \dots, r-1,$$

donde se han utilizado los siguientes hechos: (a)  $P_\nu \mathbf{w}_i = \mathbf{w}_i$ , pues  $\mathbf{w}_i$  pertenece a  $\nu$ , y (b)  $\mathbf{w}'\mathbf{w}_i = 0$  para  $i = 1, 2, \dots, r-1$ , pues  $\mathbf{w}$  es un múltiplo de  $\mathbf{w}_r$ , el cual es ortogonal a  $\mathbf{w}_i$  para  $i < r$ . Por otro lado,

$$(P_\nu - \mathbf{w}\mathbf{w}')\mathbf{w}_r = P_\nu \mathbf{w}_r - \mathbf{w}\mathbf{w}'\mathbf{w}_r = (1 - |\mathbf{w}|^2)\mathbf{w}_r,$$

pues  $\mathbf{w}_r = \mathbf{w} / |\mathbf{w}|$ . Note finalmente que  $P_\nu \mathbf{v} = \mathbf{0}$  para todo vector  $\mathbf{v}$  en el complemento ortogonal a  $\nu$ . Esto permite establecer la siguiente conclusión: Los valores propios de  $P_\nu - \mathbf{w}\mathbf{w}'$  son

$$\lambda_1 = \lambda_2 = \cdots = \lambda_{r-1} = 1, \quad \lambda_r = 1 - |\mathbf{w}|^2, \quad \lambda_t = 0, t = r+1, \dots, k. \quad (9.3)$$

(ii)  $\mathbf{w} = \mathbf{0}$ .

En este caso  $P_\nu - \mathbf{w}\mathbf{w}' = P_\nu$ , de manera que sus valores propios son 1, con multiplicidad  $r$ , y 0 con multiplicidad  $k - r$ , así que (9.3) continua siendo un enunciado válido. En conclusión, los valores propios de  $P_\nu - \mathbf{w}\mathbf{w}'$  están dados en (9.3) bajo cualquier circunstancia. Como  $P_\nu \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, P_\nu - \mathbf{w}\mathbf{w}')$ , existen variables aleatorias independientes  $U_1, U_2, \dots, U_r$ , cada una con distribución  $\chi_i^2$ , tales que

$$|P_\nu \mathbf{Z}|^2 \stackrel{\mathcal{D}}{\rightarrow} \sum_{i=1}^{r-1} U_i + (1 - |\mathbf{w}|^2)U_r \leq \sum_{i=1}^r U_i =: Y,$$

de donde se desprende que para todo  $z \in (0, \infty)$

$$\begin{aligned} P[|P_\nu \mathbf{Z}|^2 \leq z] &= P\left[\sum_{i=1}^r \lambda_i U_i \leq z\right] \\ &= P\left[\sum_{i=1}^{r-1} U_i + (1 - |\mathbf{w}|^2)U_r \leq z\right] \\ &\geq P\left[\sum_{i=1}^r U_i \leq z\right] = P[Y \leq z]. \end{aligned}$$

Finalmente,  $Y = \sum_{i=1}^r U_i$  implica que  $Y \sim \chi_r^2$ , y entonces  $P[|P_\nu \mathbf{Z}|^2 \leq \chi_{r,\alpha}^2] \geq P[Y \leq \chi_{r,\alpha}^2] = 1 - \alpha$ , para todo  $\alpha \in (0, 1)$ .  $\square$

**Lema 9.2.** Sea  $Y$  una variable aleatoria con función de distribución continua y suponga que  $\{V_n\}$  y  $\{Z_n\}$  son dos sucesiones de variables aleatorias no negativas tales que para todo  $z > 0$ ,

$$\lim_{n \rightarrow \infty} P[V_n \leq z] \geq P[Y \leq z],$$

y

$$Z_n \xrightarrow{P} 0.$$

Entonces,

$$\liminf_{n \rightarrow \infty} P[V_n + Z_n \leq z] \geq P[Y \leq z].$$

**Demostración** Dado  $z > 0$  seleccione  $\varepsilon \in (0, z)$  y observe que  $[V_n \leq z - \varepsilon, Z_n \leq \varepsilon] \subset [V_n + Z_n \leq z]$ , de donde se desprende que

$$\begin{aligned} P[V_n + Z_n \leq \varepsilon] &\geq P[V_n \leq z - \varepsilon, Z_n \leq \varepsilon] \\ &= P[V_n \leq z - \varepsilon] - P[V_n \leq z - \varepsilon, Z_n > \varepsilon] \\ &\geq P[V_n \leq z - \varepsilon] - P[Z_n > \varepsilon], \end{aligned}$$

y como  $Z_n \xrightarrow{P} 0$ , la última desigualdad conduce a

$$\begin{aligned} \liminf_{n \rightarrow \infty} P[V_n + Z_n \leq \varepsilon] &\geq \liminf_{n \rightarrow \infty} P[V_n \leq z - \varepsilon] - \liminf_{n \rightarrow \infty} P[Z_n > \varepsilon] \\ &\geq \liminf_{n \rightarrow \infty} P[V_n \leq z - \varepsilon] - 0 \\ &\geq P[Y \leq z - \varepsilon], \end{aligned}$$

y dejando que  $\varepsilon$  decrezca a cero se concluye que

$$\lim_{n \rightarrow \infty} P[V_n + Z_n \leq z] \geq \lim_{\varepsilon \searrow 0} P[Y \leq z - \varepsilon] = P[Y \leq z],$$

donde la igualdad se debe a la continuidad de la distribución de  $Y$ .  $\square$

**Lema 9.3.** Sea  $\mathcal{S} \subset \mathbb{R}^k$  un subespacio de dimensión  $r$  y suponga que  $\mathcal{S}$  es ortogonal a  $\mathbf{1}$ , esto es,  $\mathbf{1}'\mathbf{v} = 0$  para todo  $\mathbf{v} \in \mathcal{S}$ . Entonces, si  $Y$  es una variable aleatoria con distribución  $\chi_r^2$ ,

$$\liminf_{n \rightarrow \infty} P \left[ \max_{\mathbf{c} \in \mathcal{S}} \max_{\mathbf{c}'\mathbf{M}(\mathbf{p})\mathbf{c}=1} n|\mathbf{c}'(\bar{X}_n - \mathbf{p})|^2 \geq z \right] \geq P[Y \leq z].$$

**Demostración** Defina  $f : \mathbb{R}^k \rightarrow \mathbb{R}$  mediante

$$f(\mathbf{x}) = \max_{\mathbf{c} \in \mathcal{S}} \max_{\mathbf{c}'\mathbf{M}(\mathbf{p})\mathbf{c}=1} |\mathbf{c}'\mathbf{x}|^2, \quad \mathbf{x} \in \mathbb{R}^k,$$

y note que  $f$  es una función continua. Combinando los Teoremas 2.1 y 3.1, se obtiene que

$$\begin{aligned} \max_{\mathbf{c} \in \mathcal{S}, \mathbf{c}'\mathbb{M}(\mathbf{p})\mathbf{c}=1} n|\mathbf{c}'(\bar{X}_n - \mathbf{p})|^2 &= f(\sqrt{n}(\bar{X}_n - \mathbf{p})) \xrightarrow{\mathcal{D}} f(\mathbf{Z}) \\ &= \max_{\mathbf{c} \in \mathcal{S}, \mathbf{c}'\mathbb{M}(\mathbf{p})\mathbf{c}=1} |\mathbf{c}'\mathbf{Z}|^2, \end{aligned} \quad (9.4)$$

donde  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbb{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}') = \mathcal{N}(\mathbf{0}, \mathbb{M}(\mathbf{p}))$ . Note además que

$$\mathbf{W} := \mathbb{M}(\mathbf{p})^{-1/2}\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I} - \mathbf{w}\mathbf{w}'), \quad \text{donde } \mathbf{w} := \frac{\mathbf{1}}{\sqrt{n}},$$

lo cual se desprende de los comentarios en la Observación 9.1. Por otro lado, sea  $\nu$  el espacio vectorial que se obtiene al aplicar la matriz  $\mathbb{M}(\mathbf{p})^{1/2}$  a  $\mathcal{S}$ , esto es,

$$\nu = \{\mathbf{v} \mid \mathbf{v} = \mathbb{M}(\mathbf{p})^{1/2}\mathbf{c}, \text{ para algún } \mathbf{c} \in \mathcal{S}\};$$

claramente,  $\nu$  tiene dimensión  $r$ , pues  $\mathcal{S}$  es ortogonal al núcleo de  $\mathbb{M}(\mathbf{p})^{1/2}$ . Usando esta notación se tiene lo siguiente: Si  $\mathbf{c} \in \mathcal{S}$  y  $\mathbf{c}'\mathbb{M}(\mathbf{p})\mathbf{c} = 1$ ,

$$\mathbf{c}'\mathbf{Z} = (\mathbb{M}(\mathbf{p})^{1/2}\mathbf{c})'\mathbb{M}(\mathbf{p})^{-1/2}\mathbf{Z} = (\mathbb{M}(\mathbf{p})^{1/2}\mathbf{c})'\mathbf{W} = (\mathbb{M}(\mathbf{p})^{1/2}\mathbf{c})'P_\nu\mathbf{W},$$

donde la tercera igualdad se debe a que  $\mathbb{M}(\mathbf{p})^{1/2}\mathbf{c}$  es un elemento de  $\nu$ . Aplicando ahora la desigualdad de Cauchy-Schwartz, se obtiene que

$$|\mathbf{c}'\mathbf{Z}| \leq \sqrt{(\mathbb{M}(\mathbf{p})^{1/2}\mathbf{c})'\mathbb{M}(\mathbf{p})^{1/2}\mathbf{c}} \sqrt{(P_\nu\mathbf{W})'P_\nu\mathbf{W}} = \sqrt{\mathbf{c}'\mathbb{M}(\mathbf{p})\mathbf{c}} |P_\nu\mathbf{W}| = |P_\nu\mathbf{W}|,$$

y entonces

$$f(\mathbf{Z}) = \max_{\mathbf{c} \in \mathcal{S}, \mathbf{c}'\mathbb{M}(\mathbf{p})\mathbf{c}=1} |\mathbf{c}'\mathbf{Z}|^2 \leq |P_\nu\mathbf{W}|^2 \quad (9.5)$$

Combinando (9.4) y (9.5) se concluye que

$$\liminf_{n \rightarrow \infty} P \left[ \max_{\mathbf{c} \in \mathcal{S}, \mathbf{c}'\mathbb{M}(\mathbf{p})\mathbf{c}=1} n|\mathbf{c}'(\bar{X}_n - \mathbf{p})|^2 \leq z \right] = P[f(\mathbf{Z}) \leq z] \geq P[|P_\nu\mathbf{W}|^2 \leq z],$$

y entonces el Lema 9.1 implica que

$$\liminf_{n \rightarrow \infty} P \left[ \max_{\mathbf{c} \in \mathcal{S}, \mathbf{c}'\mathbb{M}(\mathbf{p})\mathbf{c}=1} n|\mathbf{c}'(\bar{X}_n - \mathbf{p})|^2 \leq z \right] \geq P[Y \leq z],$$

donde  $Y \sim \chi_r^2$ .

**Lema 9.4.** Sea  $Y \sim \chi_r^2$  y  $\mathcal{S} \subset \mathbb{R}^k$  un subespacio de dimensión  $r$  el cual es ortogonal a  $\mathbf{1}$ . En este caso, para todo  $z > 0$ ,

$$\liminf_{n \rightarrow \infty} P \left[ \max_{\mathbf{c} \in \mathcal{S}, \mathbf{c}'\mathbf{M}(\mathbf{p})\mathbf{c}=1} \{n|\mathbf{c}'(\bar{X}_n - \mathbf{p})|^2 - z\mathbf{c}'\mathbf{M}(\bar{X}_n)\mathbf{c}\} \leq 0 \right] \geq P[Y \leq z].$$

**Demostración** Primero observe que

$$\mathbf{M}(\bar{X}_n) - \mathbf{M}(\mathbf{p}) \xrightarrow{P} 0$$

de donde se desprende que

$$Z_n := \max_{\mathbf{c} \in \mathcal{S}, \mathbf{c}'\mathbf{M}(\mathbf{p})\mathbf{c}=1} |\mathbf{c}'(\mathbf{M}(\bar{X}_n) - \mathbf{M}(\mathbf{p}))| \xrightarrow{P} 0 \quad (9.6)$$

puesto que la ortogonalidad de  $\mathcal{S}$  y el vector  $\mathbf{1}$  implica que el conjunto de todos los vectores  $\mathbf{c} \in \mathcal{S}$  que satisfacen  $\mathbf{c}'\mathbf{M}(\mathbf{p})\mathbf{c} = 1$  es un conjunto acotado. Observe ahora que las siguientes implicaciones son válidas.

$$\begin{aligned} & \max_{\mathbf{c} \in \mathcal{S}, \mathbf{c}'\mathbf{M}(\mathbf{p})\mathbf{c}=1} \{n|\mathbf{c}'(\bar{X}_n - \mathbf{p})|^2\} + \max_{\mathbf{c} \in \mathcal{S}, \mathbf{c}'\mathbf{M}(\mathbf{p})\mathbf{c}=1} z|\mathbf{c}'(\mathbf{M}(\bar{X}_n) - \mathbf{M}(\mathbf{p}))\mathbf{c}| \leq z \\ \implies & \max_{\mathbf{c} \in \mathcal{S}, \mathbf{c}'\mathbf{M}(\mathbf{p})\mathbf{c}=1} [n|\mathbf{c}'(\bar{X}_n - \mathbf{p})|^2 - z\mathbf{c}'(\mathbf{M}(\bar{X}_n) - \mathbf{M}(\mathbf{p}))\mathbf{c}] \leq z \\ \implies & \max_{\mathbf{c} \in \mathcal{S}, \mathbf{c}'\mathbf{M}(\mathbf{p})\mathbf{c}=1} [n|\mathbf{c}'(\bar{X}_n - \mathbf{p})|^2 - z\mathbf{c}'\mathbf{M}(\bar{X}_n)\mathbf{c} + z\mathbf{c}'\mathbf{M}(\mathbf{p})\mathbf{c}] \leq z \\ \implies & \max_{\mathbf{c} \in \mathcal{S}, \mathbf{c}'\mathbf{M}(\mathbf{p})\mathbf{c}=1} [n|\mathbf{c}'(\bar{X}_n - \mathbf{p})|^2 - z\mathbf{c}'\mathbf{M}(\bar{X}_n)\mathbf{c} + z] \leq z \\ \implies & \max_{\mathbf{c} \in \mathcal{S}, \mathbf{c}'\mathbf{M}(\mathbf{p})\mathbf{c}=1} [n|\mathbf{c}'(\bar{X}_n - \mathbf{p})|^2 - z\mathbf{c}'\mathbf{M}(\bar{X}_n)\mathbf{c}] \leq 0 \end{aligned}$$

y entonces

$$P \left[ \max_{\mathbf{c} \in \mathcal{S}, \mathbf{c}'\mathbf{M}(\mathbf{p})\mathbf{c}=1} \{n|\mathbf{c}'(\bar{X}_n - \mathbf{p})|^2 - z\mathbf{c}'\mathbf{M}(\bar{X}_n)\mathbf{c}\} \leq 0 \right] \geq P[V_n + Z_n \leq z] \quad (9.7)$$

donde

$$V_n = \max_{\mathbf{c} \in \mathcal{S}, \mathbf{c}'\mathbf{M}(\mathbf{p})\mathbf{c}=1} n|\mathbf{c}'(\bar{X}_n - \mathbf{p})|^2$$

Por el Lema 9.3,  $\liminf_{n \rightarrow \infty} P[V_n \leq z] \geq P[T \leq z]$ , donde  $Y \sim \chi_r^2$ . Luego, combinando la convergencia (9.5) con (9.6) y el Lema 9.2 se obtiene que

$$\begin{aligned} \liminf_{n \rightarrow \infty} P \left[ \max_{\mathbf{c} \in \mathcal{S}, \mathbf{c}'\mathbf{M}(\mathbf{p})\mathbf{c}=1} \{n|\mathbf{c}'(\bar{\mathbf{X}}_n - \mathbf{p})|^2 - z\mathbf{c}'\mathbf{M}(\bar{\mathbf{X}}_n)\mathbf{c}\} \leq 0 \right] \\ \geq \liminf_{n \rightarrow \infty} P[V_n + Z_n \leq z] \\ \geq P[Y \leq z], \end{aligned}$$

y esto concluye con la demostración del Lema.  $\square$

Finalmente, se está en posibilidad de establecer el resultado principal de esta sección.

**Demostración del Teorema 9.1.** Considere el siguiente enunciado:

$$|\mathbf{c}'(\bar{\mathbf{X}}_n - \mathbf{p})| \leq \sqrt{\frac{\chi_{r,\alpha}^2}{n}} \sqrt{\mathbf{c}'\mathbf{M}(\bar{\mathbf{X}}_n)\mathbf{c}}, \quad \mathbf{c} \in \mathcal{S} \quad (9.8)$$

Ambos lados de la desigualdad son (absolutamente) homogéneos como función de  $\mathbf{c}$ , de manera que (9.7) equivale a

$$|\mathbf{c}'(\bar{\mathbf{X}}_n - \mathbf{p})| \leq \sqrt{\frac{\chi_{r,\alpha}^2}{n}} \sqrt{\mathbf{c}'\mathbf{M}(\bar{\mathbf{X}}_n)\mathbf{c}}, \quad \mathbf{c} \in \mathcal{S} \quad \mathbf{c}'\mathbf{M}(\mathbf{p})\mathbf{c} = 1 \quad (9.9)$$

enunciado que, por el Lema 9.4 aplicado con  $z = \chi_{r,\alpha}^2$ , es tal que todos los puntos límite de la probabilidad de que sea válido son mayores o iguales a  $1 - \alpha$ , y por lo tanto lo mismo ocurre con (9.7). Esto demuestra el teorema, pues observando que

$$\mathbf{c}'\mathbf{M}(\bar{\mathbf{X}}_n)\mathbf{c} = \sum_{i=1}^k c_i^2 \bar{\mathbf{X}}_{ni} - \left( \sum_{i=1}^k c_i \bar{\mathbf{X}}_{ni} \right)^2,$$

se desprende que (9.8) es equivalente a (9.1).  $\square$

## Conclusión

En este capítulo se ha realizado un estudio formal del procedimiento estadístico conocido como la prueba Ji-cuadrada de bondad de ajuste, cuyo contexto

básico supone que se observa una muestra de una distribución de Bernoulli de dimensión  $k$ . El estudio realizado descansa en los teoremas de invarianza presentados en la Sección 3, los cuales establecen que, al transformar una sucesión convergente en distribución se produce una nueva serie que continúa siendo convergente con tal solo que la transformación aplicada satisfaga condiciones poco restrictivas. Después de establecer la prueba Ji-cuadrada en la Sección 6, se utilizó la dualidad entre pruebas de hipótesis y regiones de confianza en la Sección 7 para construir intervalos de confianza para todas las combinaciones lineales de los parámetros  $p_1, p_2, \dots, p_k$  cuyo coeficiente de confianza simultáneo coincide, conforme se incrementa el tamaño de la muestra, con el nivel deseado. Posteriormente, en la Sección 8 se presentó un refinamiento que permite obtener intervalos de confianza más cortos que los originalmente establecidos. En la Sección 9 se presentó una extensión del Método de Scheffè que permite tomar en cuenta la dimensión del espacio de combinaciones lineales de interés, permitiendo acortar la longitud de los intervalos de confianza que se obtienen.

## Capítulo 3

### Clasificación por Medio de dos Características

El propósito de este capítulo es discutir la forma en que una clasificación de los resultados de un experimento puede obtenerse por medio de dos características, las cuales se presentan en varios grados o niveles. Cuando las observaciones aleatorias se ubican en categorías obtenidas combinando los niveles de dos rasgos de interés, las frecuencias con que los diversos tipos de resultados ocurren determinan un arreglo rectangular conocido como tabla de contingencia, para la cual el problema fundamental consiste en probar estadísticamente el supuesto de independencia de las dos características que le dan origen. Con relación a este problema, se introduce una función de comparación  $f$ , se determina su derivada, así como la distribución asintótica del estadístico  $f(\bar{X}_n)$ , donde  $\bar{X}_n$  es la media de una muestra de la distribución multidimensional de Bernoulli. Estos resultados son la base para abordar, en el siguiente capítulo, el problema básico asociado a una tabla de contingencia.

### Introducción

El tema de este capítulo es la construcción de categorías de resultados por medio de dos características de interés. La idea central es que cada uno de esos rasgos se presenta en diversos ‘niveles’ o ‘grados’—de forma similar a lo que ocurre con los niveles de los factores en el diseño de experimentos—y las categorías relevantes para el investigador se obtienen mediante combinaciones de los diversos grados (Fisher y McDonald, 1981, Christensen, 1996). Por ejemplo, los cepillos dentales pueden clasificarse de acuerdo a las características *color* y *dureza*, y las

categorías de cepillos puede describirse, por ejemplo, como ‘blanco-suave’, ‘verde-duro’ o alguna combinación similar de los niveles de color y dureza. De hecho, las categorías en las que los posibles resultados del experimento pueden ubicarse surgen, frecuentemente, de esta forma, es decir, ‘cruzando’ o ‘combinando’ los diversos niveles de dos rasgos de interés. La exposición introduce formalmente la idea de *clasificación cruzada*, así como la noción de tabla de contingencia y el problema fundamental asociado a la misma, a saber, la construcción de una prueba para el supuesto de que las dos características que la originan son independientes (Agresti, 1984, Bishop *et al.*, 1975).

*La organización del material es la siguiente:* En la Sección 2 se introduce la idea formal de clasificación cruzada de dos rasgos de interés. La noción de tabla de contingencia se presenta en la Sección 3, en donde también se establece la forma que adopta el supuesto de la independencia de las dos características de interés. Posteriormente, en la Sección 4 se introduce la función de comparación, la cual es el medio fundamental para el análisis básico de una tabla de contingencia, dedicando la Sección 5 a calcular su derivada. La Sección 6 trata sobre la distribución asintótica del estadístico que se obtiene componiendo la función de comparación, previamente introducida, con la media muestral de una distribución multidimensional de Bernoulli, y el capítulo concluye en la Sección 7 con algunos comentarios breves.

## **Clasificación Cruzada**

Esta sección persigue el siguiente objetivo: *relacionar la clasificación categórica de las respuestas posibles en un experimento aleatorio, con los niveles asumidos por dos características de interés de los resultados observados.* Hasta ahora se ha supuesto que se dispone de una división categórica dada de antemano, y la discusión que se desarrollará busca entender la forma en que dicha clasificación surge en las aplicaciones.

Como punto de partida, denote mediante  $\mathcal{C} = \{1, 2, \dots, k\}$  a la clase de categorías en que se dividen los posibles resultados del experimento, y suponga que  $F_1, F_2, \dots, F_r$  es una partición  $\mathcal{C}$ , lo cual significa que las siguientes condiciones (i)–(iii) se satisfacen:

- (i) Cada conjunto  $F_i$  es no vacío,  
(ii) La intersección de dos conjuntos  $F_i$  y  $F_j$  es nula cuando  $i \neq j$ , i.e.,

$$F_i \cap F_j = \emptyset \quad \text{si } i \neq j, \quad (2.1)$$

y

- (iii) Los conjuntos  $F_i$  agotan al conjunto de índices, esto es,

$$\cup_{i=1}^r F_i = \{1, 2, \dots, k\}. \quad (2.2)$$

Las categorías que conforman el conjunto  $F_i$  se interpretan como aquellas que corresponden al nivel  $i$  de cierta característica. Por ejemplo, considere el experimento de escoger aleatoriamente un estudiante universitario, y suponga que una característica de interés es el *lugar de origen* del alumno seleccionado, rasgo con el cual se asocian  $r = 4$  ‘niveles’: Norte ( $i = 1$ ), Centro ( $i = 2$ ), Sur ( $i = 3$ ) y Extranjero ( $i = 4$ ). Luego,  $F_1$  consiste de las categorías correspondientes a estudiantes cuyo estado de origen se ubica en el norte del país,  $F_4$  se compone de categorías asociadas a estudiantes extranjeros; por supuesto,  $F_2$  y  $F_3$  consisten de tipos de resultados correspondientes a estudiantes originarios del centro y sur del país, respectivamente.

Considere ahora otra partición  $C_1, C_2, \dots, C_s$  del conjunto  $\{1, 2, \dots, k\} = \mathcal{C}$ , la cual es inducida por otra característica de interés que tiene  $s$  niveles. Por ejemplo, en el caso anterior en que se selecciona un estudiante al azar, dicho rasgo podría ser el *nivel de aprovechamiento*, en el cual se consideran  $s = 3$  niveles: bajo, medio y alto, correspondientes a  $C_1, C_2$  y  $C_3$ , respectivamente.

En el resto de este trabajo se supone que las dos particiones consideradas satisfacen la siguiente condición: Para cada  $i = 1, 2, \dots, r$  y  $j = 1, 2, \dots, s$ ,

$$F_i \cap C_j \text{ consiste de un solo índice.} \quad (2.3)$$

**Proposición 2.1.** Considere dos particiones  $F_1, F_2, \dots, F_r$  y  $C_1, C_2, \dots, C_s$  del conjunto  $\mathcal{C} = \{1, 2, \dots, k\}$ , las cuales satisfacen la condición (2.3). En este caso,

- (i) Cada conjunto  $F_1, F_2, \dots, F_r$  tiene exactamente  $s$  elementos;
- (ii) Cada uno de los conjuntos  $C_1, C_2, \dots, C_s$  consta de  $r$  miembros, y
- (iii) El número total de categorías es  $k = rs$ .

**Demostración.** (i) El argumento depende de las tres propiedades (2.4)–(2.6) establecidas a continuación. Para empezar, observe que para cada  $i = 1, 2, \dots, r$ , la inclusión  $F_i \subset \{1, 2, \dots, k\}$  implica

$$F_i = F_i \cap \{1, 2, \dots, k\}. \quad (2.4)$$

Por otro lado, debido a que  $C_1, C_2, \dots, C_s$  es una partición de  $\{1, 2, \dots, k\}$ , se tiene que

$$\{1, 2, \dots, k\} = C_1 \cup C_2 \cup \dots \cup C_s, \quad (2.5)$$

y

$$C_m \cap C_t = \emptyset, \quad \text{si } m \neq t. \quad (2.6)$$

A continuación, combine (2.4) y (2.5) para obtener

$$\begin{aligned} F_i &= F_i \cap \{1, 2, \dots, k\} \\ &= F_i \cap (C_1 \cup C_2 \cup \dots \cup C_s) \\ &= (F_i \cap C_1) \cup (F_i \cap C_2) \cup \dots \cup (F_i \cap C_s) \end{aligned} \quad (2.7)$$

Por otro lado, de acuerdo a (2.3), cada intersección que aparece entre paréntesis en la última igualdad contiene un sólo elemento, mientras que (2.6) implica que

$$(F_i \cap C_m) \cap (F_i \cap C_t) = F_i \cap (C_m \cap C_t) = \emptyset \quad \text{si } m \neq t. \quad (2.8)$$

En resumen: De acuerdo a (2.8) y (2.9),  $F_i$  es la unión de  $s$  conjuntos disjuntos, cada uno de los cuales contiene un único elemento. Por lo tanto  $F_i$  consta de  $s$  miembros, estableciendo la parte (i), mientras que la parte (ii) se demuestra de manera similar. Finalmente, observe que (2.1) y (2.2) implican que  $\{1, 2, \dots, k\}$  se expresa como la unión de los  $r$  conjuntos disjuntos  $F_1, F_2, \dots, F_r$  y que, por la parte (i), cada uno de ellos tiene  $s$  elementos, de donde se desprende que  $k = rs$ .  $\square$

Cuando las diversas categorías de resultados de un experimento aleatorio admiten dos particiones  $F_1, F_2, \dots, F_r$  y  $C_1, C_2, \dots, C_s$  que satisfacen (2.3), se dice que dicha categorización se obtiene *por clasificación cruzada* de las características que dan origen a las particiones. Es costumbre ubicar las  $r$  categorías que conforman al conjunto  $F_i$  en la  $i$ -ésima fila de una matriz, mientras que los índices que componen a  $C_j$  se ubican en la  $j$ -ésima columna; de esta forma, la categoría que aparece en la posición  $(i, j)$  se obtiene ‘cruzando’ el nivel  $i$  de la característica que induce la partición  $F_1, F_2, \dots, F_r$ , con el nivel  $j$  del rasgo asociado a la partición  $C_1, C_2, \dots, C_s$ . Esta terminología se discute en detalle para la selección aleatoria de un estudiante, experimento que fue considerado al inicio de la sección.

**Discusión.** Considere las dos características *origen* y *aprovechamiento* las cuales, como se describió anteriormente, tienen cuatro y tres niveles, respectivamente. De acuerdo a la Proposición 2.1, si una categorización de los resultados admite las particiones  $F_1, F_2, F_3, F_4$  y  $C_1, C_2, C_3$  entonces el número de categorías es  $k = 4 \times 3 = 12$ . Los índices que conforman a los conjuntos  $F_i$  se colocan en filas, mientras que los que componen a los conjuntos  $C_j$  se ubican en columnas. A continuación se ilustra una forma de etiquetar a las doce categorías acomodadas

de esta forma:

$$\begin{array}{rcccl}
 & C_1(\text{bajo}) & C_2(\text{medio}) & C_3(\text{alto}) & \\
 F_1(\text{Norte}) & 1 & 2 & 3 & \\
 F_2(\text{Centro}) & 4 & 5 & 6 & (2.9) \\
 F_3(\text{Sur}) & 7 & 8 & 9 & \\
 F_4(\text{Extranjero}) & 10 & 11 & 12 & 
 \end{array}$$

En este caso, las doce categorías existentes se particionan en cuatro conjuntos  $F_i$  de acuerdo a la característica *origen*. Por ejemplo,  $F_1$  consta de los índices 1, 2 y 3, mientras que  $F_4$  se compone de las categorías 10, 11 y 12. Similarmente,  $C_2 = \{2, 5, 8, 11\}$ . Los cuatro niveles para el origen del estudiante, y los tres niveles para el grado de aprovechamiento, inducen 12 categorías correspondientes a las entradas de la anterior matriz de orden  $4 \times 3$ . Estas dos características permiten describir las categorías de forma tal que tengan una interpretación clara. Por ejemplo, el hecho de que un estudiante seleccionado al azar pertenezca a la categoría 6, significa que proviene del centro del país y que su aprovechamiento es alto. Note que la categoría 6 se obtiene, efectivamente, ‘cruzando’ el nivel dos de la característica *origen* (centro), con el nivel tres del rasgo *aprovechamiento* (alto). Además, las doce categorías pueden nombrarse, de manera más descriptiva, mediante un par ordenado que haga explícita la posición que ocupan en la matriz. Por ejemplo, la categoría 10 puede representarse mediante  $(4, 1)$ , indicando que corresponde al nivel cuatro del rasgo *origen*, y al nivel uno de la característica *aprovechamiento*. Aunque no se seguirá la idea de denotar a una categoría mediante un par ordenado, si se adoptará la siguiente convención: Cuando un índice represente una de las  $k = rs$  categorías posibles, se utilizará una letra mayúscula para representarlo; la intención es recordar que la categoría  $I$  bajo consideración proviene de un cruzamiento de los niveles de dos características de interés.  $\square$

A continuación, se introducirá una notación que será básica para el análisis que se realizará en el resto de esta tesis.

**Definición 2.1.** Suponga que  $F_1, F_2, \dots, F_r$  y  $C_1, C_2, \dots, C_s$  son dos particiones del conjunto de categorías  $\{1, 2, \dots, k\}$ .

(i) Dado  $I \in \{1, 2, \dots, k\}$ ,  $F(I)$  denota al único conjunto  $F_m$  que contiene a  $I$ , mientras que

(ii)  $C(I)$  representa al único conjunto  $C_t$  que contiene a  $I$ .

Para ilustrar esta idea, considere la situación descrita en la discusión precedente, en la cual los resultados del experimento aleatorio se clasifican en 12 categorías. Considere, por ejemplo, la categoría  $I = 7$ . En este caso, el único conjunto  $F_m$  que contiene a 7 es  $F_3 = \{7, 8, 9\}$ , y por lo tanto,  $F(7) = F_3 = \{7, 8, 9\}$ ; similarmente,  $7 \in C_1 = \{1, 4, 7, 10\}$ , y entonces  $C(7) = C_1 = \{1, 4, 7, 10\}$ . La siguiente propiedad asociada a las particiones de una clasificación cruzada será utilizada en la demostración del Lema 5.1.

**Lema 2.1.** Suponga que  $F_1, F_2, \dots, F_r$  y  $C_1, C_2, \dots, C_s$  dos particiones del conjunto de categorías  $\{1, 2, \dots, k\}$ , las cuales satisfacen (2.3). En este caso, para todo  $J \in \{1, 2, \dots, k\}$ ,

$$\sum_{I \in C(J)} \sum_{i \in F(I)} x_i = \sum_{i=1}^k x_i.$$

**Demostración.** El punto de partida es la relación

$$C(J) = C(J) \cap \{1, 2, \dots, k\} = (C(J) \cap F_1) \cup (C(J) \cap F_2) \cup \dots \cup (C(J) \cap F_r),$$

la cual se desprende combinando la inclusión  $C(J) \subset \{1, 2, \dots, k\}$  con la igualdad (2.2). Por otro lado,  $C(J)$  es uno de los conjuntos  $C_m$ , de manera que (2.3) implica que cada intersección  $C(J) \cap F_t$  consiste de un único elemento, el cual se denotará mediante  $I_t$ , esto es,  $C(J) \cap F_t = \{I_t\}$ ,  $t = 1, 2, \dots, r$ , de manera que

$$C(J) = \{I_1\} \cup \{I_2\} \cup \dots \cup \{I_r\} \quad \text{y} \quad F(I_t) = F_t, \quad t = 1, 2, \dots, r.$$

Por lo tanto,

$$\begin{aligned}
 \sum_{I \in C(J)} \sum_{i \in F(I)} x_i &= \sum_{I \in \{I_1, I_2, \dots, I_r\}} \sum_{i \in F(I)} x_i \\
 &= \sum_{m=1}^r \sum_{i \in F(I_m)} x_i \\
 &= \sum_{m=1}^r \sum_{i \in F_m} x_i \\
 &= \sum_{i \in F_1 \cup F_2 \cup \dots \cup F_r} x_i
 \end{aligned}$$

donde la última igualdad se debe a que los conjuntos  $F_i$  son disjuntos. Ahora combine el argumento anterior con la ecuación (2.2) para obtener

$$\sum_{I \in C(J)} \sum_{i \in F(I)} x_i = \sum_{i \in F_1 \cup F_2 \cup \dots \cup F_r} x_i = \sum_{i \in \{1, 2, \dots, k\}} x_i = \sum_{i=1}^k x_i,$$

concluyendo la demostración del lema.  $\square$

## Tablas de Contingencia

Considere un experimento cuyos resultados se clasifican en  $k = rs$  categorías, las cuales se obtienen cruzando dos características de interés a través de las particiones correspondientes  $F_1, F_2, \dots, F_r$  y  $C_1, C_2, \dots, C_s$ . Como ya se ha mencionado, las categorías se pueden colocar en un arreglo rectangular de  $r$  filas y  $s$  columnas. Sin pérdida de generalidad, se supondrá que los diversos tipos de resultados se etiquetan como se indica a continuación:

	$C_1$	$C_2$	$\dots$	$C_s$	
$F_1$	1	2	$\dots$	$s$	
$F_2$	$s + 1$	$s + 2$	$\dots$	$2s$	
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	
$F_r$	$s(r - 1) + 1$	$s(r - 1) + 2$	$\dots$	$rs$	(3.1)

Cuando el experimento se repite  $n$  veces de forma independiente, cada uno de los resultados observados se ubicará en alguna de las  $k = rs$  categorías, y el observador anota la frecuencia  $n_t$  con que se observa que el resultado pertenece a la categoría

$t$ , acción que conduce a la siguiente tabla:

$$\begin{array}{ccccccc}
 & C_1 & C_2 & \cdots & C_s & & \\
 F_1 & n_1 & n_2 & \cdots & n_s & & \\
 F_2 & n_{s+1} & n_{s+2} & \cdots & n_{2s} & & \\
 \vdots & \vdots & \vdots & \ddots & \vdots & & \\
 F_r & n_{s(r-1)+1} & n_{s(r-1)+2} & \cdots & n_{rs} & & 
 \end{array} \tag{3.2}$$

Este arreglo se denomina la *tabla de contingencia* correspondiente a las  $n$  repeticiones del experimento (Agresti, 1984, Bishop *et al.*, 1975). Puesto que cada resultado se ubica en una sólo de las  $k = rs$  categorías, siempre se tiene que  $n = \sum_{t=1}^k n_t$ . Antes de plantear el problema básico en el análisis del arreglo (3.2), es conveniente relacionar la tabla de contingencia con el otro medio para capturar la información contenida en las repeticiones del experimento que se consideró en el Capítulo 2, a saber, los vectores  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  que constituyen una muestra de la distribución  $k$ -dimensional de Bernoulli. Recordando que  $\mathbf{X}_i = \mathbf{e}_j$  si y sólo si el  $i$ -ésimo ensayo produce un resultado dentro de la categoría  $j$ , se desprende que la componente  $j$  de  $\sum_{i=1}^n \mathbf{X}_i$  es el número de ocasiones en que el resultado observado es del tipo  $j$ , y entonces

$$\sum_{i=1}^n \mathbf{X}_i = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_{rs} \end{bmatrix}. \tag{3.3}$$

El *problema fundamental* en el estudio de una tabla de contingencia puede ahora establecerse de la siguiente manera:

Utilizar la información experimental, representada por las frecuencias  $n_t$ , para decidir si las características de interés asociadas a las particiones  $F_1, F_2, \dots, F_r$  y  $C_1, C_2, \dots, C_s$  son independientes o no.

Antes de plantear formalmente esta cuestión en un contexto general, se discutirá su formulación en un caso especial.

**Ejemplo 3.1.** De nueva cuenta, considere el experimento de seleccionar un estudiante universitario al azar, en el cual las 12 categorías surgen mediante clasi-

ficación cruzada de las características *origen* y *aprovechamiento*, las cuales tienen cuatro y tres niveles, respectivamente; vea la discusión en la Sección 2, particularmente la ecuación (2.9). Las probabilidades de que el resultado pertenezca a cada una de las doce categorías se indican a continuación:

	$C_1(\text{bajo})$	$C_2(\text{medio})$	$C_3(\text{alto})$	
$F_1(\text{Norte})$	$p_1$	$p_2$	$p_3$	(3.4)
$F_2(\text{Centro})$	$p_4$	$p_5$	$p_6$	
$F_3(\text{Sur})$	$p_7$	$p_8$	$p_9$	
$F_4(\text{Extranjero})$	$p_{10}$	$p_{11}$	$p_{12}$	

En el contexto de este experimento aleatorio, una pregunta fundamental es si acaso el lugar de origen del estudiante influye o no sobre el nivel de aprovechamiento que presenta, esto es, si la característica *origen* es independiente del rasgo *aprovechamiento*, y lo que se pretende ahora es encontrar un medio para expresar esta condición. Inicialmente, considere los eventos determinados por las siguientes condiciones:

$A$  : El estudiante seleccionado proviene del norte del país (3.5)

$B$  : El estudiante seleccionado tiene aprovechamiento alto

El primero de estos requerimientos, significa que el ‘resultado del experimento’ pertenece a las categorías que conforman al conjunto de índices  $F_1 = \{1, 2, 3\}$ , lo cual ocurre con probabilidad

$$P[A] = \sum_{i \in F_1} p_i = p_1 + p_2 + p_3, \quad (3.6)$$

mientras que la segunda de las condiciones establece que el estudiante seleccionado se ubica dentro de una categoría perteneciente a  $C_3 = \{3, 6, 9, 12\}$ , lo cual ocurre con probabilidad

$$P[B] = \sum_{i \in C_3} p_i = p_3 + p_6 + p_9 + p_{12}. \quad (3.7)$$

Por otro lado, ambas condiciones en (3.5) ocurren sólo cuando el estudiante seleccionado pertenece a la categoría 3, evento que acontece con probabilidad  $p_3$ , i.e.,  $P[A \cap B] = p_3$ . Cuando la condición de independencia entre la característica

‘origen’ y el rasgo de ‘aprovechamiento’ son independientes, debe tenerse que  $P[A \cap B] = P[A]P[B]$ , igualdad que, vía (3.6) y (3.7), equivale a

$$p_3 = \left( \sum_{i \in F_1} p_i \right) \left( \sum_{i \in C_3} p_i \right) = (p_1 + p_2 + p_3)(p_3 + p_6 + p_9 + p_{12}). \quad (3.8)$$

Hay una manera alternativa de expresar esta ecuación: Observe que la categoría 3 se obtiene cruzando el primer nivel de la característica *origen*, con el tercer nivel del rasgo *aprovechamiento*, esto es,  $3 \in F_1$  y  $3 \in C_3$ , de manera que  $F(3) = F_1$  y  $C(3) = C_3$ , y por lo tanto la ecuación (3.8) puede expresarse como

$$p_3 = \left( \sum_{i \in F(3)} p_i \right) \left( \sum_{i \in C(3)} p_i \right).$$

El argumento precedente, que se aplicó a la categoría 3, se generaliza a cualquier otro valor del índice  $I = 1, 2, \dots, 12$ : Si las características *origen* y *aprovechamiento* son independientes, se tiene que

$$p_I = \left( \sum_{i \in F(I)} p_i \right) \left( \sum_{i \in C(I)} p_i \right), \quad I = 1, 2, \dots, 12.$$

Recíprocamente, cuando estas doce igualdades ocurren simultáneamente, las características de fila y columna (*origen* y *aprovechamiento*) son independientes.  $\square$

La discusión del Ejemplo 3.1 puede extenderse a un contexto general: Cuando los resultados de un experimento aleatorio se clasifican en categorías asociadas a dos particiones  $F_1, F_2, \dots, F_r$  y  $C_1, C_2, \dots, C_s$ , la condición de independencia de las características cuyos niveles dan origen a las particiones, equivale a las igualdades

$$p_I = \left( \sum_{i \in F(I)} p_i \right) \left( \sum_{i \in C(I)} p_i \right), \quad I = 1, 2, \dots, k = rs. \quad (3.9)$$

Luego, el problema fundamental en el análisis de una tabla de contingencia, es utilizar los datos muestrales para decidir si *todas* estas igualdades son válidas.

Como se verá posteriormente, (3.9) puede verificarse utilizando las frecuencias observadas  $\{n_t\}$  para construir un estadístico cuya distribución es, aproximadamente, Ji-cuadrada, y el propósito que se persigue en el resto del trabajo es establecer dicho método de prueba, así como proporcionar un argumento riguroso que valide la expresión para los grados de libertad que se emplean al aplicar el procedimiento. Note que el problema fundamental consiste en *comparar* la probabilidad  $p_I$  de que el resultado de un ensayo pertenezca a la categoría  $I$ , con el producto  $\left(\sum_{i \in F(I)} p_i\right) \left(\sum_{i \in C(I)} p_i\right)$ , tarea que se llevará a cabo mediante una función que se introduce a continuación.

## Función de Comparación

En esta sección se inicia el camino hacia la obtención de un método estadístico para probar la validez de las igualdades en la expresión (3.9). El propósito que se persigue es definir la función que permite realizar la comparación entre la probabilidad de que un resultado se ubique en una categoría dada, y el producto de las probabilidades asociadas a la fila y la columna que contienen a dicha categoría. El punto de partida, es la introducción de una clase especial de funciones lineales.

**Definición 4.1.** Dado un conjunto de índices  $A \subset \{1, 2, 3, \dots, k\}$ , la función  $\mathbb{P}_A: \mathbb{R}^k \rightarrow \mathbb{R}^k$  se define mediante

$$\mathbb{P}_A(\mathbf{x}) = \sum_{i \in A} x_i. \quad (4.1)$$

Esta notación desempeña un papel importante en el análisis del resto de capítulo. Note que  $\mathbb{P}_A(\mathbf{x})$  es una función lineal, y que

$$\frac{\partial \mathbb{P}_A(\mathbf{x})}{\partial x_j} = \begin{cases} 1, & \text{si } j \in A \\ 0, & \text{si } j \notin A. \end{cases} \quad (4.2)$$

Para entender esta fórmula, note que si  $j \notin A$ , entonces la suma que define a  $\mathbb{P}_A(\mathbf{x})$  no incluye a  $x_j$ , de manera que  $\mathbb{P}_A(\mathbf{x})$  no depende de  $x_j$ , y por lo tanto,

$\frac{\partial \mathbb{P}_A(\mathbf{x})}{\partial x_j} = 0$ . Por otro lado, cuando  $j \in A$ , la suma en el lado derecho de (4.1) contiene al término  $x_j$ , de manera que  $\frac{\partial \mathbb{P}_A(\mathbf{x})}{\partial x_j} = 1$ . Recordando que la función indicadora del conjunto  $A$  se define mediante

$$\mathcal{I}_A(j) = \begin{cases} 1, & \text{si } j \in A \\ 0, & \text{si } j \notin A \end{cases} \quad (4.3)$$

se desprende que la expresión (4.2) para las derivadas parciales de (4.1) puede expresarse en forma compacta de la siguiente manera:

$$\frac{\partial \mathbb{P}_A(\mathbf{x})}{\partial x_j} = \mathcal{I}_A(j). \quad (4.4)$$

**Definición 4.2.** Suponga que  $F_1, F_2, \dots, F_r$  y  $C_1, C_2, \dots, C_s$  son dos particiones del conjunto de índices  $\{1, 2, \dots, k\}$ , las cuales están asociadas a una clasificación cruzada como se describió en la Sección 2. La función de comparación correspondiente a dicha clasificación se denota mediante

$$\mathbf{f}: \mathbb{R}^k \rightarrow \mathbb{R}^k,$$

y se define a través de la siguiente expresión para cada una de sus funciones coordenadas  $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_k$ :

$$\mathbf{f}_I(\mathbf{x}) = x_I - P_{F(I)}(\mathbf{x})P_{C(I)}(\mathbf{x}), \quad I = 1, 2, \dots, k; \quad (4.5)$$

recuerde que, para cada índice  $I$ ,  $F(I)$  denota al único conjunto  $F_t$  que contiene a  $I$ , mientras que  $C(I)$  es el miembro de la partición  $C_1, C_2, \dots, C_s$  que contiene a  $I$ .

El análisis subsecuente está principalmente relacionado con la función  $\mathbf{f}$  en esta definición, y para clarificar las ideas detrás de su construcción, se presentará un cálculo detallado de sus componentes en un caso especial.

**Ejemplo 4.1.** Considere el caso en que las observaciones se clasifican en  $k = 6$  tipos de resultados, los cuales se clasifican en forma cruzada utilizando las siguientes particiones:

$$F_1 = \{1, 2, 3\}, \quad y \quad F_2 = \{4, 5, 6\};$$

$$C_1 = \{1, 4\}, \quad C_2 = \{2, 5\}, \quad y \quad C_3 = \{3, 6\}.$$

Equivalentemente, las categorías  $1, 2, \dots, 6$ , se determina cruzando los niveles de las dos características de interés como se indica en la siguiente matriz:

	$C_1$	$C_2$	$C_3$
$F_1$	1	2	3
$F_2$	4	5	6

A continuación se determinan las componentes  $\mathbf{f}_I$  de la función de comparación  $\mathbf{f}$  para cada índice  $I = 1, 2, 3, 4, 5, 6 = k$ . En cada caso, el procedimiento que se sigue involucra los siguientes pasos: Primero, se determinan los conjuntos  $F(I)$  y  $C(I)$  que contienen al índice  $I$ , posteriormente se calculan las funciones  $\mathbb{P}_{F(I)}(\mathbf{x})$  y  $\mathbb{P}_{C(I)}(\mathbf{x})$  sumando las componentes que pertenecen a  $F(I)$  y  $C(I)$ , respectivamente, y entonces  $\mathbf{f}_I(\mathbf{x})$  se encuentra restando a  $x_I$  el producto de las dos funciones anteriores:

- Cálculo de  $\mathbf{f}_1$ . En esta situación se tiene que  $I = 1$ . Para evaluar  $\mathbf{f}_1$ , note que  $1 \in F_1 = \{1, 2, 3\}$ , y  $1 \in C_1 = \{1, 4\}$ . Por lo tanto,

$$F(1) = \{1, 2, 3\}, \quad y \quad C(1) = \{1, 4\}.$$

A continuación se utiliza la Definición 4.1 para determinar

$$\mathbb{P}_{F(1)}(\mathbf{x}) = \sum_{i \in F(1)} x_i = \sum_{i \in \{1, 2, 3\}} x_i = x_1 + x_2 + x_3$$

$$\mathbb{P}_{C(1)}(\mathbf{x}) = \sum_{i \in C(1)} x_i = \sum_{i \in \{1, 4\}} x_i = x_1 + x_4$$

y finalmente  $\mathbf{f}_1(x)$  se obtiene restando de  $x_1$  el producto de las dos funciones anteriores (vea (4.5)):

$$\mathbf{f}_1(\mathbf{x}) = x_1 - \mathbb{P}_{F(1)}(\mathbf{x})\mathbb{P}_{C(1)}(\mathbf{x}) = x_1 - (x_1 + x_2 + x_3)(x_1 + x_4).$$

- Determinación de  $\mathbf{f}_2$ . En este caso  $I = 2$ , y observando que  $2 \in F_1 = \{1, 2, 3\}$ , y  $2 \in C_2 = \{2, 5\}$ , se desprende que

$$F(2) = \{1, 2, 3\}, \quad \text{y} \quad C(2) = \{2, 5\}.$$

Utilizando la Definición 4.1, se obtiene

$$\begin{aligned} \mathbb{P}_{F(2)}(\mathbf{x}) &= \sum_{i \in F(2)} x_i = \sum_{i \in \{1, 2, 3\}} x_i = x_1 + x_2 + x_3 \\ \mathbb{P}_{C(2)}(\mathbf{x}) &= \sum_{i \in C(2)} x_i = \sum_{i \in \{2, 5\}} x_i = x_2 + x_5 \end{aligned}$$

y  $\mathbf{f}_2(x)$  se obtiene tomando la diferencia entre  $x_2$  y el producto de estas dos últimas funciones:

$$\mathbf{f}_2(\mathbf{x}) = x_2 - \mathbb{P}_{F(2)}(\mathbf{x})\mathbb{P}_{C(2)}(\mathbf{x}) = x_2 - (x_1 + x_2 + x_3)(x_2 + x_5).$$

- Cálculo de  $\mathbf{f}_3$ . Ahora se tiene que  $I = 3$ , y observando que  $3 \in F_1 = \{1, 2, 3\}$ , y  $3 \in C_3 = \{3, 6\}$ , se desprenden las igualdades

$$F(3) = \{1, 2, 3\}, \quad \text{y} \quad C(3) = \{3, 6\}.$$

Por lo tanto, la Definición 4.1 implica

$$\begin{aligned} \mathbb{P}_{F(3)}(\mathbf{x}) &= \sum_{i \in F(3)} x_i = \sum_{i \in \{1, 2, 3\}} x_i = x_1 + x_2 + x_3 \\ \mathbb{P}_{C(3)}(\mathbf{x}) &= \sum_{i \in C(3)} x_i = \sum_{i \in \{3, 6\}} x_i = x_3 + x_6 \end{aligned}$$

y a partir de (4.5) se obtiene  $\mathbf{f}_3(x)$  como la diferencia entre  $x_3$  y el producto de las dos funciones anteriores:

$$\mathbf{f}_3(\mathbf{x}) = x_3 - \mathbb{P}_{F(3)}(\mathbf{x})\mathbb{P}_{C(3)}(\mathbf{x}) = x_3 - (x_1 + x_2 + x_3)(x_3 + x_6).$$

- La componente  $\mathbf{f}_4$ . Ahora se trata del caso  $I = 4$ . A partir de la especificación de las particiones, se obtiene que  $4 \in F_2 = \{4, 5, 6\}$ , y  $4 \in C_1 = \{1, 4\}$ , y entonces

$$F(3) = \{4, 5, 6\}, \quad \text{y} \quad C(3) = \{1, 4\}.$$

Luego, a partir de la Definición 4.1 se concluye que

$$\begin{aligned}\mathbb{P}_{F(4)}(\mathbf{x}) &= \sum_{i \in F(4)} x_i = \sum_{i \in \{4,5,6\}} x_i = x_4 + x_5 + x_6 \\ \mathbb{P}_{C(4)}(\mathbf{x}) &= \sum_{i \in C(4)} x_i = \sum_{i \in \{1,4\}} x_i = x_1 + x_4\end{aligned}$$

y utilizando (4.5) se llega a la expresión

$$\mathbf{f}_4(\mathbf{x}) = x_4 - \mathbb{P}_{F(4)}(\mathbf{x})\mathbb{P}_{C(4)}(\mathbf{x}) = x_4 - (x_4 + x_5 + x_6)(x_1 + x_4).$$

• Cálculo de la componente  $\mathbf{f}_5$ . En esta situación se tiene  $I = 5$ . Ahora observe que  $5 \in F_2 = \{4, 5, 6\}$ , y  $5 \in C_2 = \{2, 5\}$ , y entonces

$$F(5) = \{4, 5, 6\}, \quad \text{y} \quad C(5) = \{2, 5\}.$$

Por lo tanto, utilizando la Definición 4.1 se obtiene

$$\begin{aligned}\mathbb{P}_{F(5)}(\mathbf{x}) &= \sum_{i \in F(5)} x_i = \sum_{i \in \{4,5,6\}} x_i = x_4 + x_5 + x_6 \\ \mathbb{P}_{C(5)}(\mathbf{x}) &= \sum_{i \in C(5)} x_i = \sum_{i \in \{2,5\}} x_i = x_2 + x_5\end{aligned}$$

y a través de (4.5) resulta que

$$\mathbf{f}_5(\mathbf{x}) = x_5 - \mathbb{P}_{F(5)}(\mathbf{x})\mathbb{P}_{C(5)}(\mathbf{x}) = x_5 - (x_4 + x_5 + x_6)(x_2 + x_5).$$

• Determinación de  $\mathbf{f}_6$ . En esta situación se tiene  $I = 6$ , y como  $6 \in F_2 = \{4, 5, 6\}$ , y  $6 \in C_3 = \{3, 6\}$ , entonces

$$F(6) = \{4, 5, 6\}, \quad \text{y} \quad C(6) = \{3, 6\}.$$

En consecuencia, las funciones  $\mathbb{P}_{F(6)}(\mathbf{x})$  y  $\mathbb{P}_{C(6)}(\mathbf{x})$  están dadas por

$$\begin{aligned}\mathbb{P}_{F(6)}(\mathbf{x}) &= \sum_{i \in F(6)} x_i = \sum_{i \in \{4,5,6\}} x_i = x_4 + x_5 + x_6 \\ \mathbb{P}_{C(6)}(\mathbf{x}) &= \sum_{i \in C(6)} x_i = \sum_{i \in \{3,6\}} x_i = x_3 + x_6\end{aligned}$$

y por lo tanto (4.5) implica que

$$\mathbf{f}_6(\mathbf{x}) = x_6 - \mathbb{P}_{F(6)}(\mathbf{x})\mathbb{P}_{C(6)}(\mathbf{x}) = x_6 - (x_4 + x_5 + x_6)(x_3 + x_6).$$

Estos cálculos permiten establecer la siguiente expresión para la función de comparación  $\mathbf{f}: \mathbb{R}^6 \rightarrow \mathbb{R}^6$ :

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} \mathbf{f}_1(\mathbf{x}) \\ \mathbf{f}_2(\mathbf{x}) \\ \mathbf{f}_3(\mathbf{x}) \\ \mathbf{f}_4(\mathbf{x}) \\ \mathbf{f}_5(\mathbf{x}) \\ \mathbf{f}_6(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} x_1 - (x_1 + x_2 + x_3)(x_1 + x_4) \\ x_2 - (x_1 + x_2 + x_3)(x_2 + x_5) \\ x_3 - (x_1 + x_2 + x_3)(x_3 + x_6) \\ x_4 - (x_4 + x_5 + x_6)(x_1 + x_4) \\ x_5 - (x_4 + x_5 + x_6)(x_2 + x_5) \\ x_6 - (x_4 + x_5 + x_6)(x_3 + x_6) \end{bmatrix}.$$

□

Después de ilustrar el cálculo de la función de comparación asociada a una clasificación cruzada, es oportuno señalar una propiedad fundamental que se desprende de inmediato de (4.5):

$$\begin{aligned} \mathbf{f}(\mathbf{x}) = \mathbf{0} &\iff \mathbf{f}_I(\mathbf{x}) = 0 \quad \text{para todo } I \\ &\iff x_I = \mathbb{P}_{F(I)}(\mathbf{x})\mathbb{P}_{C(I)}(\mathbf{x}) \quad \text{para todo } I. \end{aligned}$$

Aplicando esta equivalencia con el vector  $\mathbf{p}$  reemplazando a  $\mathbf{x}$ , se desprende que

$$\mathbf{f}(\mathbf{p}) = \mathbf{0} \iff p_I = \mathbb{P}_{F(I)}(\mathbf{p})\mathbb{P}_{C(I)}(\mathbf{p}) \quad \text{para todo } I.$$

Recordando que las igualdades (3.9) son la expresión formal de la condición de independencia de las características de interés, se desprende que

$$p_I = \mathbb{P}_{F(I)}(\mathbf{p})\mathbb{P}_{C(I)}(\mathbf{p}), \quad I = 1, 2, \dots, k,$$

equivale a la independencia de las características de fila y columna que originan la clasificación cruzada. Puesto que la anterior ecuación desplegada equivale a que  $\mathbf{f}(\mathbf{p})$  sea el vector nulo, se desprende que, para decidir si los rasgos de fila y columna son independientes, debe compararse el valor asumido por  $\mathbf{f}(\mathbf{p})$  con el vector  $\mathbf{0}$ ; esta observación justifica el nombre de *función de comparación* para  $\mathbf{f}$ , y

da cuenta de la relevancia que el análisis de sus propiedades tendrá en el desarrollo subsecuente.

## La Derivada de $f$

El propósito de esta sección es determinar la derivada de la función de comparación introducida en la Definición 4.2. Como punto de partida para alcanzar este objetivo, es conveniente introducir la siguiente notación.

**Definición 5.1.** Dadas dos particiones  $F_1, F_2, \dots, F_r$  y  $C_1, C_2, \dots, C_s$  del conjunto de índices  $\{1, 2, \dots, k\}$  asociadas a una clasificación cruzada, defina las matrices  $\mathbb{M}^{FC}(\mathbf{x})$  y  $\mathbb{M}^{CF}(\mathbf{x})$  de orden  $k \times k$  mediante las siguientes expresiones: Para todos los índices  $I, J = 1, 2, \dots, k$ ,

$$\begin{aligned}\mathbb{M}_{IJ}^{FC}(\mathbf{x}) &= \mathbb{P}_{F(I)}(\mathbf{x})\mathcal{I}_{C(I)}(J) \\ \mathbb{M}_{IJ}^{CF}(\mathbf{x}) &= \mathbb{P}_{C(I)}(\mathbf{x})\mathcal{I}_{F(I)}(J)\end{aligned}\tag{5.1}$$

Como se verá en el Teorema 5.1 establecido más adelante, estas matrices son dos componentes de la (matriz) derivada de la función de comparación. En seguida se ilustra la construcción de  $\mathbb{M}^{FC}(\mathbf{x})$  y  $\mathbb{M}^{CF}(\mathbf{x})$ .

**Ejemplo 5.1.** Como en la Sección 4, considere el caso en que las observaciones se clasifican en  $k = 6$  tipos de resultados, clasificados en forma cruzada utilizando las siguientes particiones:

$$\begin{aligned}F_1 &= \{1, 2, 3\}, & y & & F_2 &= \{4, 5, 6\}; \\ C_1 &= \{1, 4\}, & C_2 &= \{2, 5\}, & y & & C_3 &= \{3, 6\}.\end{aligned}$$

A continuación se determina la matriz  $\mathbb{M}^{FC}$ . El procedimiento que se sigue determina las componentes de  $\mathbb{M}^{FC}$  fila por fila, e involucra los siguientes pasos: Para determinar la fila  $I$  de  $\mathbb{M}^{FC}$ , primero se determinan los miembros  $F(I)$  y  $C(I)$  de las particiones que contienen al índice  $I$ , posteriormente se calcula la función

$\mathbb{P}_{F(I)}(\mathbf{x})$ , y en la posición  $J$  de la  $I$ -ésima fila se coloca dicho valor si  $J \in C(I)$ , mientras que la componente  $J$  se iguala a cero siempre que  $J \notin C(I)$ . Note que los conjuntos  $F(I)$  y  $C(I)$  fueron determinados en el Ejemplo 4.1.

- Cálculo de la primera fila  $\mathbb{M}^{FC}$ . Para  $I = 1$  se tiene que  $F(1) = \{1, 2, 3\}$  y  $C(1) = \{1, 4\}$ , pues  $1 \in F_1 = \{1, 2, 3\}$  y  $1 \in C_1 = \{1, 4\}$ . Por lo tanto,

$$\mathbb{P}_{F(1)}(\mathbf{x}) = \sum_{i \in F(1)} x_i = x_1 + x_2 + x_3,$$

y entonces

$$M_{1J}^{FC}(\mathbf{x}) = \mathbb{P}_{F(1)}(\mathbf{x})\mathcal{I}_{C(1)}(J) = (x_1 + x_2 + x_3)\mathcal{I}_{\{1,4\}}(J), \quad J = 1, 2, \dots, 6,$$

de tal manera que la primera filas de  $\mathbb{M}^{FC}(\mathbf{x})$  está dada por

$$[x_1 + x_2 + x_3, 0, 0, x_1 + x_2 + x_3, 0, 0].$$

Similarmente, la segunda y tercera fila de  $\mathbb{M}^{FC}(\mathbf{x})$  están dadas por

$$[0, x_1 + x_2 + x_3, 0, 0, x_1 + x_2 + x_3, 0], \quad \text{y} \quad [0, 0, x_1 + x_2 + x_3, 0, 0, x_1 + x_2 + x_3].$$

- Cálculo de la cuarta fila  $\mathbb{M}^{FC}$ . Para  $I = 4$  se tiene que  $F(4) = \{4, 5, 6\}$  y  $C(4) = \{1, 4\}$ , pues  $4 \in F_2 = \{4, 5, 6\}$  y  $4 \in C_1 = \{1, 4\}$ . Por lo tanto,

$$\mathbb{P}_{F(4)}(\mathbf{x}) = \sum_{i \in F(4)} x_i = x_4 + x_5 + x_6.$$

En consecuencia,

$$M_{4J}^{FC}(\mathbf{x}) = \mathbb{P}_{F(4)}(\mathbf{x})\mathcal{I}_{C(4)}(J) = (x_4 + x_5 + x_6)\mathcal{I}_{\{1,4\}}(J), \quad J = 1, 2, \dots, 6,$$

y entonces la cuarta fila de  $\mathbb{M}^{FC}(\mathbf{x})$  está dada por

$$[x_4 + x_5 + x_6, 0, 0, x_4 + x_5 + x_6, 0, 0].$$

De igual manera, la quinta y sexta filas de  $\mathbb{M}^{FC}(\mathbf{x})$  son

$$[0, x_4 + x_5 + x_6, 0, 0, x_4 + x_5 + x_6, 0], \quad \text{y} \quad [0, 0, x_4 + x_5 + x_6, 0, 0, x_4 + x_5 + x_6].$$

A partir de estos cálculos, se obtiene la siguiente expresión:

$$M^{FC}(\mathbf{x}) = \begin{bmatrix} x_1 + x_2 + x_3 & 0 & 0 & x_1 + x_2 + x_3 & 0 & 0 \\ 0 & x_1 + x_2 + x_3 & 0 & 0 & x_1 + x_2 + x_3 & 0 \\ 0 & 0 & x_1 + x_2 + x_3 & 0 & 0 & x_1 + x_2 + x_3 \\ x_4 + x_5 + x_6 & 0 & 0 & x_4 + x_5 + x_6 & 0 & 0 \\ 0 & x_4 + x_5 + x_6 & 0 & 0 & x_4 + x_5 + x_6 & 0 \\ 0 & 0 & x_4 + x_5 + x_6 & 0 & 0 & x_4 + x_5 + x_6 \end{bmatrix} \quad (5.1)$$

Ahora se encontrará la matriz  $\mathbb{M}^{CF}(\mathbf{x})$ :

- Determinación de la primera fila de  $\mathbb{M}^{CF}$ . Para  $I = 1$  se tiene que  $C(1) = \{1, 4\}$  y  $F(1) = \{1, 2, 3\}$  pues  $1 \in F_1 = \{1, 2, 3, \}$  y  $1 \in C_1 = \{1, 4\}$ . Luego,

$$\mathbb{P}_{C(1)}(\mathbf{x}) = \sum_{i \in C(1)} x_i = x_1 + x_4.$$

Por lo tanto,

$$M_{1J}^{CF}(\mathbf{x}) = \mathbb{P}_{C(1)}(\mathbf{x})\mathcal{I}_{F(1)}(J) = (x_1 + x_4)\mathcal{I}_{\{1,2,3\}}(J), \quad J = 1, 2, \dots, 6,$$

de tal manera que la primera fila de  $\mathbb{M}^{CF}(\mathbf{x})$  está dada por

$$[x_1 + x_4, x_1 + x_4, x_1 + x_4, 0, 0, 0].$$

- La segunda fila de  $\mathbb{M}^{CF}$ . Para  $I = 2$  se tiene que  $C(2) = \{2, 5\}$  y  $F(2) = \{1, 2, 3\}$ .

Luego,

$$\mathbb{P}_{C(2)}(\mathbf{x}) = \sum_{i \in C(2)} x_i = x_2 + x_5,$$

y entonces

$$M_{2J}^{CF}(\mathbf{x}) = \mathbb{P}_{C(2)}(\mathbf{x})\mathcal{I}_{F(2)}(J) = (x_2 + x_5)\mathcal{I}_{\{1,2,3\}}(J), \quad J = 1, 2, \dots, 6,$$

y la segunda fila de  $\mathbb{M}^{CF}(\mathbf{x})$  es

$$[x_2 + x_5, x_2 + x_5, x_2 + x_5, 0, 0, 0].$$

Similarmente, la tercera fila de  $\mathbb{M}^{CF}(\mathbf{x})$  está dada por

$$[x_3 + x_6, x_3 + x_6, x_3 + x_6, 0, 0, 0],$$

mientras que la cuarta, quinta y sexta filas son

$$[0, 0, 0, x_1 + x_4, x_1 + x_4, x_1 + x_4]$$

$$[0, 0, 0, x_2 + x_5, x_2 + x_5, x_2 + x_5]$$

$$[0, 0, 0, x_3 + x_6, x_3 + x_6, x_3 + x_6]$$

En consecuencia,

$$\mathbb{M}^{CF}(\mathbf{x}) = \begin{bmatrix} x_1 + x_4 & x_1 + x_4 & x_1 + x_4 & 0 & 0 & 0 \\ x_2 + x_5 & x_2 + x_5 & x_2 + x_5 & 0 & 0 & 0 \\ x_3 + x_6 & x_3 + x_6 & x_3 + x_6 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1 + x_4 & x_1 + x_4 & x_1 + x_4 \\ 0 & 0 & 0 & x_2 + x_5 & x_2 + x_5 & x_2 + x_5 \\ 0 & 0 & 0 & x_3 + x_6 & x_3 + x_6 & x_3 + x_6 \end{bmatrix} \quad (5.2)$$

**Teorema 5.1.** Sea  $\mathbf{f}: \mathbb{R}^k \rightarrow \mathbb{R}^k$  la función de comparación introducida en la Definición 4.2. En este caso, la matriz derivada de  $\mathbf{f}$  es

$$D\mathbf{f}(x) = \mathbf{I} - \mathbb{M}^{FC}(\mathbf{x}) - \mathbb{M}^{CF}(\mathbf{x}). \quad (5.3)$$

**Demostración.** La matriz derivada de  $\mathbf{f}$  es

$$D\mathbf{f}(\mathbf{x}) = \left[ \frac{\partial \mathbf{f}_I(\mathbf{x})}{\partial x_J} \right]_{I, J=1,2,\dots,k}$$

y a continuación se verificará que las matrices en ambos lados de (5.3) tienen las mismas componentes. Utilizando (4.5), se obtiene que el elemento  $(I, J)$  de la matriz  $D\mathbf{f}(\mathbf{x})$  es

$$\begin{aligned} \frac{\partial \mathbf{f}_I(\mathbf{x})}{\partial x_J} &= \frac{\partial [x_I - \mathbb{P}_{F(I)}(\mathbf{x})\mathbb{P}_{C(I)}(\mathbf{x})]}{\partial x_J} \\ &= \frac{\partial x_I}{\partial x_J} - \frac{\partial \mathbb{P}_{F(I)}(\mathbf{x})}{\partial x_J} \mathbb{P}_{C(I)}(\mathbf{x}) - \mathbb{P}_{F(I)}(\mathbf{x}) \frac{\partial \mathbb{P}_{C(I)}(\mathbf{x})}{\partial x_J} \end{aligned}$$

y entonces, a partir de (4.4), se desprende que

$$\begin{aligned}\frac{\partial \mathbf{f}_I(\mathbf{x})}{\partial x_J} &= \frac{\partial x_I}{\partial x_J} - \mathcal{I}_{F(I)}(J)\mathbb{P}_{C(I)}(\mathbf{x}) - \mathbb{P}_{F(I)}(\mathbf{x})\mathcal{I}_{C(I)}(J) \\ &= \delta_{IJ} - \mathcal{I}_{F(I)}(J)\mathbb{P}_{C(I)}(\mathbf{x}) - \mathbb{P}_{F(I)}(\mathbf{x})\mathcal{I}_{C(I)}(J)\end{aligned}\quad (5.4)$$

donde se ha utilizado que  $\partial x_I/\partial x_J = \delta_{IJ}$ . Por otro lado, la componente  $(I, J)$  en el lado derecho de (5.3) es

$$\begin{aligned}[\mathbf{I} - \mathbb{M}^{FC}(\mathbf{x}) - \mathbb{M}^{CF}(\mathbf{x})]_{I, J} &= \mathbf{I}_{I, J} - [\mathbb{M}^{FC}(\mathbf{x})]_{I, J} - [\mathbb{M}^{CF}(\mathbf{x})]_{I, J} \\ &= \delta_{I, J} - \mathbb{M}_{I, J}^{FC}(\mathbf{x}) - \mathbb{M}_{I, J}^{CF}(\mathbf{x}) \\ &= \delta_{I, J} - \mathbb{P}_{F(I)}(\mathbf{x})\mathcal{I}_{C(I)}(J) - \mathbb{P}_{C(I)}(\mathbf{x})\mathcal{I}_{F(I)}(J)\end{aligned}\quad (5.5)$$

y una comparación de las expresiones (5.4) y (5.5) arroja que

$$[D\mathbf{f}(\mathbf{x})]_{I, J} = \frac{\partial \mathbf{f}_I(\mathbf{x})}{\partial x_J} = [\mathbf{I} - \mathbb{M}^{FC}(\mathbf{x}) - \mathbb{M}^{CF}(\mathbf{x})]_{I, J}\quad (5.6)$$

igualdad que es válida para todos los valores posibles de los índices  $I$  y  $J$ . Esto muestra que las componentes correspondientes de las matrices en ambos lados de (5.3) coinciden, y de esta forma concluye la demostración del Teorema 5.1.  $\square$

**Ejemplo 5.2.** Aunque la expresión (5.3) para la derivada de la función de comparación ha sido rigurosamente validada, es instructivo comprobarla directamente en un caso particular. Con este fin, considere el contexto de los Ejemplos 4.1 y 5.1; la correspondiente función de comparación se encontró en el Ejemplo 4.1 y las matrices  $\mathbb{M}^{FC}$  y  $\mathbb{M}^{CF}$  se determinaron en el Ejemplo 5.1. A continuación se verificará directamente la igualdad (5.6) para algunos valores de los subíndices.

• Las componentes  $(2, 3)$ : En esta situación  $I = 2$  y  $J = 3$ , y la componente de la matriz  $D\mathbf{f}(\mathbf{x})$  correspondiente a estos índices es

$$[D\mathbf{f}(\mathbf{x})]_{2, 3} = \frac{\partial \mathbf{f}_2}{\partial x_3} = \frac{\partial [x_2 - (x_1 + x_2 + x_3)(x_2 + x_5)]}{\partial x_3} = -(x_2 + x_5).\quad (5.7)$$

Por otro lado, el elemento  $(2, 3)$  de la matriz identidad es cero, mientras que a partir de (5.1) y (5.2) se obtiene que

$$\mathbb{M}_{2, 3}^{FC}(\mathbf{x}) = 0, \quad \text{y} \quad \mathbb{M}_{2, 3}^{CF}(\mathbf{x}) = x_2 + x_5.$$

por lo tanto,

$$[\mathbf{I} - \mathbb{M}^{FC}(\mathbf{x}) - \mathbb{M}^{CF}(\mathbf{x})]_{2,3} = 0 - 0 - (x_2 + x_5),$$

y una comparación de esta expresión con (5.7) arroja que (5.5) ocurre para  $I = 2$  y  $J = 3$ .

- Las componentes (4, 4): En este caso  $I = 4$  y  $J = 4$ , y es

$$\begin{aligned} [D\mathbf{f}(\mathbf{x})]_{4,3} &= \frac{\partial \mathbf{f}_4}{\partial x_4} \\ &= \frac{\partial [x_4 - (x_4 + x_5 + x_6)(x_1 + x_4)]}{\partial x_4} \\ &= 1 - (x_4 + x_5 + x_6) - (x_1 + x_4). \end{aligned} \quad (5.8)$$

Además, el miembro (4, 4) de la matriz identidad es uno, y a partir de (5.1) y (5.2) se obtiene que

$$\mathbb{M}_{2,3}^{FC}(\mathbf{x}) = x_4 + x_5 + x_6, \quad \text{y} \quad \mathbb{M}_{2,3}^{CF}(\mathbf{x}) = x_1 + x_4.$$

Consecuentemente,

$$[\mathbf{I} - \mathbb{M}^{FC}(\mathbf{x}) - \mathbb{M}^{CF}(\mathbf{x})]_{4,4} = 1 - (x_4 + x_5 + x_6) - (x_1 + x_4),$$

y de nueva cuenta, esta expresión combinada con (5.8) implica que (5.5) ocurre para  $I = 4$  y  $J = 4$ .  $\square$

La fórmula (5.3) para la derivada de la función de comparación desempeña un papel central en el análisis de tablas de contingencia, el cual se desarrollará en el siguiente capítulo. Para concluir esta sección, es oportuno establecer una propiedad de las matrices  $\mathbb{M}^{FC}(\mathbf{x})$  y  $\mathbb{M}^{CF}(\mathbf{x})$  que será de utilidad posteriormente.

**Lema 5.1.** Dadas dos particiones  $F_1, F_2, \dots, F_r$  y  $C_1, C_2, \dots, C_s$  del conjunto de índices  $\{1, 2, \dots, k\}$ , sean  $\mathbb{M}^{FC}(\mathbf{x})$  y  $\mathbb{M}^{CF}(\mathbf{x})$  las matrices introducidas en la

Definición 5.1. En este caso, si  $\mathbf{1} \in \mathbb{R}^k$  es el vector cuyas componentes son todas iguales a uno, se tiene que

$$\mathbf{1}'\mathbf{M}^{FC}(\mathbf{x}) = (\mathbf{1}'\mathbf{x})\mathbf{1}', \quad \text{y} \quad \mathbf{1}'\mathbf{M}^{CF}(\mathbf{x}) = (\mathbf{1}'\mathbf{x})\mathbf{1}'.$$

**Demostración.** Utilizando la igualdad (5.1), se desprende que la componente  $J$  de del vector  $\mathbf{1}'\mathbf{M}^{FC}(\mathbf{x})$  es

$$\begin{aligned} [\mathbf{1}'\mathbf{M}^{FC}(\mathbf{x})]_J &= \sum_{I=1}^k \mathbf{1}_I \mathbf{M}_{I,J}^{FC}(\mathbf{x}) \\ &= \sum_{I=1}^k \mathbf{1}_I \mathbb{P}_{F(I)}(x) \mathcal{I}_{C(I)}(J) \\ &= \sum_{I=1}^k \mathbb{P}_{F(I)}(x) \mathcal{I}_{C(I)}(J). \end{aligned}$$

Ahora observe que  $J \in C(I)$  equivale a  $I \in C(J)$ , de manera que  $\mathcal{I}_{C(I)}(J) = \mathcal{I}_{C(J)}(I)$ , y utilizando las ecuaciones (4.1) y (4.3) se desprende que

$$\begin{aligned} [\mathbf{1}'\mathbf{M}^{FC}(\mathbf{x})]_J &= \sum_{I=1}^k \mathbb{P}_{F(I)}(x) \mathcal{I}_{C(J)}(I) \\ &= \sum_{I \in C(J)} \mathbb{P}_{F(I)}(x) \\ &= \sum_{I \in C(J)} \sum_{i \in F(I)} x_i. \end{aligned}$$

Para concluir, utilice el Lema 2.1 para obtener

$$[\mathbf{1}'\mathbf{M}^{FC}(\mathbf{x})]_J = \sum_{i=1}^k x_i = \mathbf{1}'\mathbf{x}$$

igualdad que muestra que cada componente de  $\mathbf{1}'\mathbf{M}^{FC}(\mathbf{x})$  es igual a  $\mathbf{1}'\mathbf{x}$ , y por lo tanto  $\mathbf{1}'\mathbf{M}^{FC}(\mathbf{x}) = (\mathbf{1}'\mathbf{x})\mathbf{1}'$ . La ecuación  $\mathbf{1}'\mathbf{M}^{CF}(\mathbf{x}) = (\mathbf{1}'\mathbf{x})\mathbf{1}'$  puede establecerse de manera similar.  $\square$

## Distribución Límite

Como se mencionó en la Sección 1, el objetivo del trabajo subsecuente es analizar la independencia de los dos criterios de clasificación cruzada que dan origen a una tabla de contingencia. Por otro lado, se observó al final de la Sección 4 que los criterios de fila y columna son independientes si y sólo si  $\mathbf{f}(\mathbf{p}) = \mathbf{0}$ , donde  $\mathbf{p}$  es el vector cuya componente  $i$  es la probabilidad de que al realizar el experimento el resultado sea del  $i$ -ésimo tipo. Este hecho sugiere el siguiente procedimiento para probar la independencia de las características de fila y columna: Construya un estimador (consistente)  $\hat{\mathbf{p}}$  de  $\mathbf{p}$ , y evalúe  $\mathbf{f}(\hat{\mathbf{p}})$ , y rechace el supuesto de independencia cuando el vector calculado  $\mathbf{f}(\hat{\mathbf{p}})$  es ‘sustancialmente’ diferente del vector cero, mientras que la hipótesis se acepta en otro caso. Para formalizar este procedimiento, es necesario disponer de un estimador del vector  $\mathbf{p}$  y, posteriormente, determinar la distribución de  $\mathbf{f}(\hat{\mathbf{p}})$ . Como se mencionó en el capítulo precedente, al disponer de una muestra de tamaño  $n$  de la distribución  $k$ -dimensional de Bernoulli con parámetro  $\mathbf{p}$ , la media  $\bar{\mathbf{X}}_n$  correspondiente a la muestra es un estimador consistente de  $\mathbf{p}$ , así que se tomará  $\hat{\mathbf{p}} = \bar{\mathbf{X}}_n$ . Por otro lado, debido a la naturaleza no lineal de la función de comparación, no es sencillo determinar la distribución exacta de  $\mathbf{f}(\hat{\mathbf{p}})$ . Sin embargo, su distribución límite puede obtenerse a través de los teoremas de invarianza establecidos anteriormente.

**Teorema 6.1.** Sea  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  una muestra de la distribución  $k$ -dimensional de Bernoulli con parámetro  $\mathbf{p}$ , y denote mediante  $\bar{\mathbf{X}}_n$  a la media muestral correspondiente. En este caso, conforme el tamaño de la muestra crece sin límite,

$$\sqrt{n} [\mathbf{f}(\bar{\mathbf{X}}_n) - \mathbf{f}(\mathbf{p})] \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, D\mathbf{f}(\mathbf{p})[\mathbb{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}']D\mathbf{f}(\mathbf{p})').$$

**Demostración.** A partir del Teorema 2.1 del Capítulo 2 se desprende que

$$\sqrt{n} [\bar{\mathbf{X}}_n - \mathbf{p}] \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbb{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}'),$$

y utilizando la invarianza de la normalidad asintótica establecida en el Teorema 4.2 del Capítulo 2, se obtiene que

$$\sqrt{n} [\mathbf{f}(\bar{\mathbf{X}}_n) - \mathbf{f}(\mathbf{p})] \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, D\mathbf{f}(\mathbf{p})[\mathbb{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}']D\mathbf{f}(\mathbf{p})')$$

completando la demostración del teorema.  $\square$

Este resultado será el punto de partida para analizar una tabla de contingencia en el siguiente capítulo.

## Conclusión

La exposición de este capítulo se concentró en la ideas de clasificación cruzada por medio de dos características de interés, así como en la noción de tabla de contingencia y en la formulación del problema básico asociado a la misma, el cual consiste en probar la hipótesis de que los dos rasgos que le dan origen son independientes. Se mostró que dicho supuesto equivale a la igualdad vectorial  $\mathbf{f}(\mathbf{p}) = \mathbf{0}$ , la cual muestra la importancia de  $\mathbf{f}$  en el estudio del problema fundamental. Dada una muestra de la distribución mutidimensional de Bernoulli, en el Teorema 6.1 se estableció la normalidad asintótica del estadístico  $\mathbf{f}(\bar{\mathbf{X}}_n)$ , resultado que será la plataforma desde la cual se emprenderá el análisis estadístico de una tabla de contingencia en el siguiente capítulo.

## Capítulo 4

### Prueba de la Hipótesis de Independencia

En este capítulo se establece la prueba Ji-cuadrada para el supuesto de independencia de los rasgos que originan una clasificación cruzada. El enfoque que se sigue combina los teoremas de invarianza establecidos en el Capítulo 2, con ideas algebraicas que permiten determinar el número de grados de libertad del estadístico de prueba. Como ya se ha mencionado, en este aspecto se ubica la principal contribución técnica de este trabajo.

### Introducción

En este capítulo se considera que un observador ha colectado datos que son clasificados cruzando dos características de interés, y se analiza el problema de probar la independencia de los dos rasgos que inducen dicha clasificación. El procedimiento clásico para abordar esta cuestión es la denominada prueba Ji-cuadrada, la cual se construye siguiendo las ideas expuestas en el Capítulo 4, esto es, comparando el vector  $\mathbf{f}(\bar{\mathbf{X}}_n)$  con el vector  $\mathbf{0}$ , y rechazando el supuesto de independencia cuando su norma, medida respecto a un producto interno adecuado, es ‘sustancialmente grande’ (Agresti, 1984, Bishop *et al.*, 1975). El principal objetivo de este capítulo es establecer de forma rigurosa el método de prueba, y determinar, mediante métodos algebraicos, el número de grados de libertad que el procedimiento utiliza, cristalizando la principal contribución técnica de este trabajo.

La organización del material es la siguiente: La Sección 2 contiene el resultado técnico básico, el cual se establece en el Teorema 2.1 referente a la distribución asintótica de la norma cuadrática, adecuadamente medida, del vector  $\mathbf{f}(\bar{\mathbf{X}}_n)$ . Dicha

distribución límite es Ji-cuadrada, y la determinación rigurosa de su número de grados de libertad es analizada en el resto del capítulo. El problema de determinar el número de grados de libertad se reduce, en el Teorema 2.2, a mostrar que cierta matriz  $\mathbf{W}$  es idempotente y a determinar su rango (Hoffman y Kunze, 1975, Grossman, 1983). Esta tarea se realiza en varias etapas, la primera de las cuales se efectúa en la Sección 3, en donde se demuestran propiedades básicas de las matrices  $\mathbb{M}^{FC}$  y  $\mathbb{M}^{CF}$  que son parte de la matriz derivada de la función de comparación. En la Sección 4 se demuestra la idempotencia de la matriz  $\mathbf{W}$ , mientras que su rango se encuentra en la Sección 5, en la cual se completa la demostración del Teorema 2.2. El procedimiento para probar el supuesto de independencia se formula en la Sección 6, utilizando tanto la notación vectorial, la cual es adecuada para el análisis teórico, como la notación matricial, la cual es conveniente en el análisis práctico de datos. Finalmente, la presentación culmina en la Sección 7 con algunos comentarios breves.

## El Teorema Principal

El propósito de esta sección es establecer un resultado fundamental sobre ‘el tamaño’ del vector  $\mathbf{f}(\bar{\mathbf{X}}_n)$ , a partir del cual se obtendrá, posteriormente, la prueba Ji-cuadrada para decidir si los datos que conforman una tabla de contingencia concuerdan con la hipótesis de que los rasgos que la originan son independientes. El enunciado formal es el siguiente.

**Teorema 2.1.** Bajo el supuesto

$$\mathcal{H}_0: \mathbf{f}(\mathbf{p}) = \mathbf{0}, \quad (2.1)$$

la norma cuadrada del vector  $\sqrt{n}\mathbb{D}(\mathbf{p})^{-1/2}\mathbf{f}(\bar{\mathbf{X}}_n)$  converge en distribución hacia la distribución Ji-cuadrada con  $(r-1)(s-1)$  grados de libertad, esto es,

$$\left\| \sqrt{n}\mathbb{D}(\mathbf{p})^{-1/2}\mathbf{f}(\bar{\mathbf{X}}_n) \right\|^2 = n\mathbf{f}(\bar{\mathbf{X}}_n)'\mathbb{D}(\mathbf{p})^{-1}\mathbf{f}(\bar{\mathbf{X}}_n) \xrightarrow{\mathcal{D}} \chi_{(r-1)(s-1)}^2. \quad (2.2)$$

La demostración de este resultado consiste, esencialmente, en combinar tres instrumentos: Los primeros dos son herramientas clásicas, a saber, los Teoremas 3.1 y 3.2 del Capítulo 2 referentes a propiedades de invarianza de la idea de convergencia en distribución, mientras que el tercer ingrediente es un resultado algebraico que se enuncia más adelante como el Teorema 2.2. Como ya se ha mencionado, la prueba Ji-cuadrada en el análisis de una tabla de contingencia es un medio bien establecido, y la principal contribución técnica de esta tesis, se ubica en el enfoque empleado para determinar los grados de libertad en la convergencia (2.2).

El argumento para demostrar el Teorema 2.1, inicia con la convergencia establecida en el Teorema 6.1 del Capítulo 3:

$$\sqrt{n}[\mathbf{f}(\bar{\mathbf{X}}_n) - \mathbf{f}(\mathbf{p})] \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, D\mathbf{f}(\mathbf{p})[\mathbb{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}']D\mathbf{f}(\mathbf{p})').$$

Luego, definiendo la transformación  $\mathbf{T}: \mathbb{R}^k \rightarrow \mathbb{R}^k$  mediante

$$\mathbf{T}\mathbf{y} = \mathbb{D}(\mathbf{p})^{-1/2}\mathbf{y}, \quad \mathbf{y} \in \mathbb{R}^k,$$

a través del Teorema 3.2 del Capítulo 2 se obtiene que

$$\sqrt{n}[\mathbb{D}(\mathbf{p})^{-1/2}\mathbf{f}(\bar{\mathbf{X}}_n) - \mathbb{D}(\mathbf{p})^{-1/2}\mathbf{f}(\mathbf{p})] = \sqrt{n}[\mathbf{T}(\mathbf{f}(\bar{\mathbf{X}}_n)) - \mathbf{T}(\mathbf{f}(\mathbf{p}))] \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{W}), \quad (2.3)$$

donde

$$\mathbf{W} = \mathbb{D}(\mathbf{p})^{-1/2}D\mathbf{f}(\mathbf{p})[\mathbb{D}(\mathbf{p}) - \mathbf{p}\mathbf{p}']D\mathbf{f}(\mathbf{p})'\mathbb{D}(\mathbf{p})^{-1/2}; \quad (2.4)$$

note que la matriz derivada de  $\mathbf{T}$  es  $\mathbb{D}(\mathbf{p})^{-1/2}$ , la cual es simétrica. El siguiente resultado algebraico es el ingrediente final que permite demostrar el Teorema 2.1.

**Teorema 2.2.** Bajo la hipótesis (2.1), la matriz  $\mathbf{W}$  es idempotente, esto es,  $\mathbf{W}^2 = \mathbf{W}$ . Más aún, el rango de  $\mathbf{W}$  es  $(r-1)(s-1)$ .

La demostración de este teorema será proporcionada en la Sección 5, después de establecer los preliminares técnicos necesarios. A continuación se mostrará

el papel relevante que el Teorema 2.2 desempeña en el argumento que conduce a establecer la convergencia (2.2).

**Demostración del Teorema 2.1.** Cuando  $\mathbf{f}(\mathbf{p}) = \mathbf{0}$ , la convergencia (2.3) es equivalente a

$$\sqrt{n}\mathbf{D}(\mathbf{p})^{-1/2}\mathbf{f}(\bar{\mathbf{X}}_n) \xrightarrow{\mathcal{D}} \mathbf{Z},$$

donde

$$\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{W}). \quad (2.4)$$

Por lo tanto, el teorema general de invarianza establecido en el Teorema 3.1 del Capítulo 2, implica que

$$\left\| \sqrt{n}\mathbf{D}(\mathbf{p})^{-1/2}\mathbf{f}(\bar{\mathbf{X}}_n) \right\|^2 = n\mathbf{f}(\bar{\mathbf{X}}_n)'\mathbf{D}(\mathbf{p})^{-1}\mathbf{f}(\bar{\mathbf{X}}_n) \xrightarrow{\mathcal{D}} \|\mathbf{Z}\|^2. \quad (2.5)$$

Por otro lado, debido a la idempotencia de  $\mathbf{W}$ , (2.4) implica que  $\|\mathbf{Z}\|^2$  tiene una distribución Ji-cuadrada; este resultado es fundamental en la teoría de los modelos lineales, y puede encontrarse, por ejemplo, en Christensen (1996), Box y Draper (1987), o Graybill (1985). Más aún, el número de grados de libertad de la distribución de  $\|\mathbf{Z}\|^2$  es igual al rango de  $\mathbf{W}$ , el cual, de acuerdo al Teorema 2.2, es igual a  $(r-1)(s-1)$ . Por lo tanto,

$$\|\mathbf{Z}\|^2 \sim \chi_{(r-1)(s-1)}^2,$$

y entonces (2.5) equivale a  $n\mathbf{f}(\bar{\mathbf{X}}_n)'\mathbf{D}(\mathbf{p})^{-1}\mathbf{f}(\bar{\mathbf{X}}_n) \xrightarrow{\mathcal{D}} \chi_{(r-1)(s-1)}^2$ , estableciendo la conclusión deseada.  $\square$

El Teorema 2.1 será utilizado en la Sección 6 para formular la prueba Ji-cuadrada en el análisis de una tabla de contingencia.

## Resultados Técnicos Auxiliares

Como se mostró en el argumento que condujo a la demostración del Teorema 2.1, la idempotencia de la matriz  $\mathbf{W}$  en (2.4), y el hecho de que su rango

sea igual a  $(r - 1)(s - 1)$ , son dos puntos importantes que permiten establecer la convergencia asintótica en (2.1). El objetivo de ésta y las siguientes dos secciones, es introducir los resultados algebraicos preliminares que se requieren para demostrar el Teorema 2.1 en la Sección 5. Los resultados que se presentarán involucran un análisis detallado de las propiedades de las matrices  $M^{FC}$  y  $M^{CF}$  correspondientes al vector probabilístico  $\mathbf{p}$ , de la matriz derivada  $D\mathbf{f}(\mathbf{p})$  de la función de comparación evaluada en  $\mathbf{p}$ , así como de sus productos con la matriz diagonal  $\mathbb{D}(\mathbf{p})$ .

**Lema 3.1.** Sean  $\mathbb{M}^{FC}(\mathbf{p})$  y  $\mathbb{M}^{CF}(\mathbf{p})$  las dos matrices definidas en la relación (5.1) del Capítulo 3. En este caso,

(i)  $\mathbb{M}^{FC}(\mathbf{p})$  y  $\mathbb{M}^{CF}(\mathbf{p})$  son matrices idempotentes (Hoffman y Kunze, 1975, Grossman 1983, Searle 1982);

(ii)  $\mathbb{M}^{FC}(\mathbf{p})\mathbf{p} = \mathbf{p} - \mathbf{f}(\mathbf{p}) = \mathbb{M}^{CF}(\mathbf{p})\mathbf{p}$ , y

(iii)  $D\mathbf{f}(\mathbf{p})\mathbf{p} = 2\mathbf{f}(\mathbf{p}) - \mathbf{p}$ .

Consecuentemente,

(iv) Bajo el supuesto de que los rasgos asociados a las particiones  $F_1, F_2, \dots, F_r$  y  $C_1, C_2, \dots, C_s$  son independientes,

$$\mathbb{M}^{FC}(\mathbf{p})\mathbf{p} = \mathbf{p} = \mathbb{M}^{CF}(\mathbf{p})\mathbf{p}, \quad \text{y} \quad D\mathbf{f}(\mathbf{p})\mathbf{p} = -\mathbf{p}.$$

**Demostración.** (i) Considere una pareja de índices  $I, J$  entre 1 y  $k$ . Utilizando

la primera igualdad en la expresión (5.1) del capítulo precedente, se desprende que

$$\begin{aligned}
[\mathbb{M}^{FC}(\mathbf{p})\mathbb{M}^{FC}(\mathbf{p})]_{I,J} &= \sum_{T=1}^k \mathbb{M}_{I,T}^{FC}(\mathbf{p})\mathbb{M}_{T,J}^{FC}(\mathbf{p}) \\
&= \sum_{T=1}^k \mathbb{P}_{F(I)}(\mathbf{p})\mathcal{I}_{C(I)}(T)\mathbb{P}_{F(T)}(\mathbf{p})\mathcal{I}_{C(T)}(J) \\
&= \mathbb{P}_{F(I)}(\mathbf{p}) \sum_{T=1}^k \mathcal{I}_{C(I)}(T)\mathbb{P}_{F(T)}(\mathbf{p})\mathcal{I}_{C(T)}(J) \\
&= \mathbb{P}_{F(I)}(\mathbf{p}) \sum_{T \in C(I)} \mathbb{P}_{F(T)}(\mathbf{p})\mathcal{I}_{C(T)}(J)
\end{aligned} \tag{3.1}$$

donde en la última igualdad se utilizó el hecho de que  $\mathcal{I}_{C(I)}(T) = 1$  cuando  $T \in C(I)$ , mientras que  $\mathcal{I}_{C(I)}(T) = 0$  en otro caso. Para continuar, recuerde que  $C(T)$  es el miembro de la partición  $C_1, C_2, \dots, C_s$  que contiene a  $T$ , de manera que si  $T \in C(I)$ , entonces se tiene que  $C(T) = C(I)$ , y por lo tanto la última igualdad en (3.1) implica que

$$\begin{aligned}
[\mathbb{M}^{FC}(\mathbf{p})\mathbb{M}^{FC}(\mathbf{p})]_{I,J} &= \mathbb{P}_{F(I)}(\mathbf{p}) \sum_{T \in C(I)} \mathbb{P}_{F(T)}(\mathbf{p})\mathcal{I}_{C(I)}(J) \\
&= \mathbb{P}_{F(I)}(\mathbf{p})\mathcal{I}_{C(I)}(J) \sum_{T \in C(I)} \mathbb{P}_{F(T)}(\mathbf{p});
\end{aligned} \tag{3.2}$$

ahora aplique el Lema 2.1 del Capítulo 3 para concluir que

$$\sum_{T \in C(I)} \mathbb{P}_{F(T)}(\mathbf{p}) = \mathbf{1}'\mathbf{p} = 1,$$

donde la segunda igualdad se debe a que  $\mathbf{p}$  es un vector probabilístico. Por lo tanto, (3.2) equivale a

$$[\mathbb{M}^{FC}(\mathbf{p})\mathbb{M}^{FC}(\mathbf{p})]_{I,J} = \mathbb{P}_{F(I)}(\mathbf{p})\mathcal{I}_{C(I)}(J) = [\mathbb{M}^{FC}(\mathbf{p})]_{I,J};$$

note que la segunda igualdad se desprende de la Definición de la matriz  $\mathbb{M}^{FC}$  (vea la expresión (5.1) en el Capítulo 5). Como los índices  $I$  y  $J$  son miembros arbitrarios de  $\{1, 2, \dots, k\}$ , se concluye que  $(\mathbb{M}^{FC})^2 = \mathbb{M}^{FC}$ , estableciendo la

idempotencia de la matriz  $\mathbb{M}^{FC}$ , mientras que el resultado correspondiente a  $\mathbb{M}^{CF}$  se demuestra de forma similar.

(ii) Para cada  $I \in \{1, 2, \dots, k\}$ , la  $I$ -ésima componente de  $\mathbb{M}^{FC}(\mathbf{p})\mathbf{p}$  es

$$\begin{aligned} [\mathbb{M}^{FC}(\mathbf{p})\mathbf{p}]_I &= \sum_{J=1}^k \mathbb{M}_{I,J}^{FC}(\mathbf{p})p_J \\ &= \sum_{J=1}^k \mathbb{P}_{F(I)}(\mathbf{p})\mathcal{I}_{C(I)}(J)p_J \\ &= \mathbb{P}_{F(I)}(\mathbf{p}) \sum_{J=1}^k \mathcal{I}_{C(I)}(J)p_J \\ &= \mathbb{P}_{F(I)}(\mathbf{p}) \sum_{J \in C(I)} p_J \\ &= \mathbb{P}_{F(I)}(\mathbf{p})\mathbb{P}_{C(I)}(\mathbf{p}) \end{aligned}$$

donde se utilizó la ecuación (4.1) establecida en el Capítulo 3 para obtener la última igualdad. Por otro lado, de acuerdo a la definición de la función de comparación, se tiene que  $\mathbf{f}_I(\mathbf{p}) = p_I - \mathbb{P}_{F(I)}(\mathbf{p})\mathbb{P}_{C(I)}(\mathbf{p})$ , de manera que la anterior expresión desplegada es equivalente a  $[\mathbb{M}^{FC}(\mathbf{p})\mathbf{p}]_I = p_I - \mathbf{f}_I(\mathbf{p}) = [\mathbf{p} - \mathbf{f}(\mathbf{p})]_I$ , y entonces  $\mathbb{M}^{FC}(\mathbf{p})\mathbf{p} = \mathbf{p} - \mathbf{f}(\mathbf{p})$ , pues el índice  $I$  en este argumento es un miembro arbitrario del conjunto de categorías; la ecuación  $\mathbb{M}^{CF}(\mathbf{p})\mathbf{p} = \mathbf{p} - \mathbf{f}(\mathbf{p})$  se demuestra siguiendo las mismas ideas.

(iii) Combinando el Teorema 5.3 del capítulo precedente con la parte (ii), se obtiene

$$\begin{aligned} D\mathbf{f}(\mathbf{p})\mathbf{p} &= [\mathbf{I} - \mathbb{M}^{FC}(\mathbf{p}) - \mathbb{M}^{CF}(\mathbf{p})]\mathbf{p} \\ &= \mathbf{p} - \mathbb{M}^{FC}(\mathbf{p})\mathbf{p} - \mathbb{M}^{CF}(\mathbf{p})\mathbf{p} \\ &= \mathbf{p} - [\mathbf{p} - \mathbf{f}(\mathbf{p})] - [\mathbf{p} - \mathbf{f}(\mathbf{p})] \\ &= 2\mathbf{f}(\mathbf{p}) - \mathbf{p}. \end{aligned}$$

(iv) Las ecuaciones establecidas en este apartado se obtienen combinando las partes (ii) y (iii), con el hecho de que la independencia de los rasgos asociadas a las particiones  $F_1, F_2, \dots, F_r$  y  $C_1, C_2, \dots, C_s$ , equivale a la igualdad  $\mathbf{f}(\mathbf{p}) = \mathbf{0}$ .  $\square$

En los siguientes dos lemas se establecen propiedades de productos que involucran a las cuatro matrices  $\mathbf{M}^{FC}(\mathbf{p})$ ,  $\mathbf{M}^{CF}(\mathbf{p})$ ,  $\mathbf{ID}(\mathbf{p})$  y  $D\mathbf{f}(\mathbf{p})$ .

**Lema 3.2.** Bajo la condición  $\mathbf{f}(\mathbf{p}) = \mathbf{0}$ , las siguientes afirmaciones (i) y (ii) son válidas:

(i) Sea  $\mathbf{1}$  el vector  $k$ -dimensional cuyas componentes son todas iguales a uno. En este caso,

$$\mathbf{M}^{FC}(\mathbf{p})\mathbf{M}^{CF}(\mathbf{p}) = \mathbf{p}\mathbf{1}' \quad \text{y} \quad \mathbf{M}^{CF}(\mathbf{p})\mathbf{M}^{FC}(\mathbf{p}) = \mathbf{p}\mathbf{1}'. \quad (3.3)$$

(ii) Las matrices  $\mathbf{M}^{FC}(\mathbf{p})\mathbf{ID}(\mathbf{p})$  y  $\mathbf{M}^{CF}(\mathbf{p})\mathbf{ID}(\mathbf{p})$  son simétricas, esto es,

$$\begin{aligned} \mathbf{M}^{FC}(\mathbf{p})\mathbf{ID}(\mathbf{p}) &= \mathbf{ID}(\mathbf{p})\mathbf{M}^{FC}(\mathbf{p})' \\ \mathbf{M}^{CF}(\mathbf{p})\mathbf{ID}(\mathbf{p}) &= \mathbf{ID}(\mathbf{p})\mathbf{M}^{CF}(\mathbf{p})' \end{aligned} \quad (3.4)$$

**Demostración.** (i) De acuerdo a la definición de las matrices  $\mathbf{M}^{FC}(\mathbf{p})$  y  $\mathbf{M}^{CF}(\mathbf{p})$  estipulada en la ecuación (5.1) del Capítulo 3, se tiene que la componente  $(I, J)$  del producto  $\mathbf{M}^{FC}(\mathbf{p})\mathbf{M}^{CF}(\mathbf{p})$  está dada por

$$\begin{aligned} [\mathbf{M}^{FC}(\mathbf{p})\mathbf{M}^{CF}(\mathbf{p})]_{I, J} &= \sum_{Q=1}^k \mathbf{M}_{I, Q}^{FC}(\mathbf{p})\mathbf{M}_{Q, J}^{CF}(\mathbf{p}) \\ &= \sum_{Q=1}^k \mathbb{P}_{F(I)}(\mathbf{p})\mathcal{I}_{C(I)}(Q)\mathbb{P}_{C(Q)}(\mathbf{p})\mathcal{I}_{F(Q)}(J) \\ &= \sum_{Q \in C(I)} \mathbb{P}_{F(I)}(\mathbf{p})\mathbb{P}_{C(Q)}(\mathbf{p})\mathcal{I}_{F(Q)}(J) \end{aligned}$$

donde se utilizó que  $\mathcal{I}_{C(I)}(Q) = 1$  si  $Q \in C(I)$ , mientras que  $\mathcal{I}_{C(I)}(Q) = 0$  de otro modo. Note ahora que  $C(Q)$  es el único conjunto en la partición  $C_1, C_2, \dots, C_s$  que contiene a  $Q$ , de tal forma que  $C(Q) = C(I)$  cuando  $Q \in C(I)$ . Luego,

$$\begin{aligned} [\mathbf{M}^{FC}(\mathbf{p})\mathbf{M}^{CF}(\mathbf{p})]_{I, J} &= \sum_{Q \in C(I)} \mathbb{P}_{F(I)}(\mathbf{p})\mathbb{P}_{C(Q)}(\mathbf{p})\mathcal{I}_{F(Q)}(J) \\ &= \sum_{Q \in C(I)} \mathbb{P}_{F(I)}(\mathbf{p})\mathbb{P}_{C(I)}(\mathbf{p})\mathcal{I}_{F(Q)}(J) \\ &= \mathbb{P}_{F(I)}(\mathbf{p})\mathbb{P}_{C(I)}(\mathbf{p}) \sum_{Q \in C(I)} \mathcal{I}_{F(Q)}(J) \end{aligned}$$

y debido a la igualdad  $\sum_{Q \in C(I)} \mathcal{I}_{F(Q)}(J) = \mathcal{I}_{\{1,2,\dots,k\}}(J) = 1$ , la cual fue establecida en el Lema 2.1 del Capítulo 3, se obtiene

$$\begin{aligned} [\mathbf{M}^{FC}(\mathbf{p})\mathbf{M}^{CF}(\mathbf{p})]_{I,J} &= \mathbb{P}_{F(I)}(\mathbf{p})\mathbb{P}_{C(I)}(\mathbf{p}) \sum_{Q \in C(I)} \mathcal{I}_{F(Q)}(J) \\ &= \mathbb{P}_{F(I)}(\mathbf{p})\mathbb{P}_{C(I)}(\mathbf{p}); \end{aligned}$$

sin embargo, la condición  $\mathbf{f}(\mathbf{p}) = \mathbf{0}$  equivale a  $p_I = \mathbb{P}_{F(I)}(\mathbf{p})\mathbb{P}_{C(I)}(\mathbf{p})$  para todo  $I$ , de manera que, bajo el supuesto de independencia de los rasgos de interés, se tiene

$$[\mathbf{M}^{FC}(\mathbf{p})\mathbf{M}^{CF}(\mathbf{p})]_{I,J} = p_I = [\mathbf{p}\mathbf{1}']_{I,J} \quad \text{para todo } I, J = 1, 2, \dots, k.$$

Luego,  $[\mathbf{M}^{FC}(\mathbf{p})\mathbf{M}^{CF}(\mathbf{p})] = \mathbf{p}\mathbf{1}'$ , mientras que la ecuación  $[\mathbf{M}^{CF}(\mathbf{p})\mathbf{M}^{FC}(\mathbf{p})] = \mathbf{p}\mathbf{1}'$  se establece a través de argumentos similares.

(ii) Note que bajo el supuesto  $p_I = \mathbb{P}_{F(I)}(\mathbf{p})\mathbb{P}_{C(I)}(\mathbf{p})$ , se tiene, para todos los índices  $I, J \in \{1, 2, \dots, k\}$ , que

$$\begin{aligned} [\mathbf{D}(\mathbf{p})\mathbf{M}^{FC}(\mathbf{p})']_{IJ} &= \sum_{Q=1}^k \mathbf{D}(\mathbf{p})_{I,Q} M_{Q,J}^{FC}(\mathbf{p})' \\ &= \mathbf{D}(\mathbf{p})_{I,I} M_{I,J}^{FC}(\mathbf{p})' \\ &= p_I M_{J,I}^{FC}(\mathbf{p}) \\ &= p_I \mathbb{P}_{F(J)}(\mathbf{p}) \mathcal{I}_{C(J)}(I) \\ &= \mathbb{P}_{F(I)}(\mathbf{p}) \mathbb{P}_{C(I)}(\mathbf{p}) \mathbb{P}_{F(J)}(\mathbf{p}) \mathcal{I}_{C(J)}(I) \end{aligned} \tag{3.5}$$

Para continuar observe que

$$\mathbb{P}_{C(I)}(\mathbf{p})(\mathbf{p}) \mathcal{I}_{C(J)}(I) = \mathbb{P}_{C(J)}(\mathbf{p})(\mathbf{p}) \mathcal{I}_{C(I)}(J). \tag{3.6}$$

En efecto, cuando  $\mathcal{I}_{C(J)}(I) = 1$ , se tiene que  $I \in C(J)$ , y entonces  $C(I) = C(J)$  y  $\mathcal{I}_{C(I)}(J) = 1$ , de manera que (3.6) se reduce a  $\mathbb{P}_{C(I)}(\mathbf{p})(\mathbf{p}) = \mathbb{P}_{C(J)}(\mathbf{p})(\mathbf{p})$ , igualdad que es cierta. Por otro lado, cuando  $\mathcal{I}_{C(J)}(I) = 0$  el índice  $I$  no pertenece a  $C(J)$ , y entonces  $J \notin C(I)$ , de manera que  $\mathcal{I}_{C(I)}(J) = 0$ , por lo que en esta

situación (3.6) equivale a  $0 = 0$ . Combinando ahora (3.5) con (3.6) se desprende que

$$\begin{aligned}
[\mathbb{D}(\mathbf{p})M^{FC}(\mathbf{p})']_{IJ} &= \mathbb{P}_{F(I)}(\mathbf{p})\mathbb{P}_{C(I)}(\mathbf{p})\mathbb{P}_{F(J)}(\mathbf{p})\mathcal{I}_{C(J)}(I) \\
&= \mathbb{P}_{F(I)}(\mathbf{p})\mathbb{P}_{C(J)}(\mathbf{p})\mathbb{P}_{F(J)}(\mathbf{p})\mathcal{I}_{C(I)}(J) \\
&= (\mathbb{P}_{F(I)}(\mathbf{p})\mathcal{I}_{C(I)}(J)) (\mathbb{P}_{C(J)}(\mathbf{p})\mathbb{P}_{F(J)}(\mathbf{p})) \\
&= \mathbb{M}_{I,J}^{FC}(\mathbf{p})p_J \quad (\text{pues } p_J = \mathbb{P}_{F(J)}(\mathbf{p})\mathbb{P}_{C(J)}(\mathbf{p})) \\
&= \mathbb{M}_{I,J}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})_{J,J} \\
&= \sum_{Q=1}^k \mathbb{M}_{I,Q}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})_{Q,J} \\
&= [\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})]_{I,J}
\end{aligned}$$

de donde se desprende que  $[\mathbb{D}(\mathbf{p})M^{FC}(\mathbf{p})'] = [\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})]$ , mientras que la otra igualdad en (3.4) se demuestra de manera similar.  $\square$

**Lema 3.3.** La siguiente expresión es válida:

$$Df(\mathbf{p})\mathbb{D}(\mathbf{p})Df(\mathbf{p})' = [\mathbf{I} - M^{FC}(\mathbf{p}) - M^{CF}(\mathbf{p})]\mathbb{D}(\mathbf{p}) + 2\mathbf{p}\mathbf{p}'.$$

**Demostración.** Utilizando el Teorema 5.1 del Capítulo 3, se obtienen las igualdades

$$\begin{aligned}
&Df(\mathbf{p})\mathbb{D}(\mathbf{p})Df(\mathbf{p})' \\
&= [\mathbf{I} - \mathbb{M}^{FC}(\mathbf{p}) - \mathbb{M}^{CF}(\mathbf{p})]\mathbb{D}(\mathbf{p})[\mathbf{I} - \mathbb{M}^{FC}(\mathbf{p}) - \mathbb{M}^{CF}(\mathbf{p})]' \\
&= [\mathbf{I} - \mathbb{M}^{FC}(\mathbf{p}) - \mathbb{M}^{CF}(\mathbf{p})]\mathbb{D}(\mathbf{p})[\mathbf{I} - \mathbb{M}^{FC}(\mathbf{p})' - \mathbb{M}^{CF}(\mathbf{p})'] \\
&= [\mathbf{I} - \mathbb{M}^{FC}(\mathbf{p}) - \mathbb{M}^{CF}(\mathbf{p})] \times [\mathbb{D}(\mathbf{p}) - \mathbb{D}(\mathbf{p})\mathbb{M}^{FC}(\mathbf{p})' - \mathbb{D}(\mathbf{p})\mathbb{M}^{CF}(\mathbf{p})']
\end{aligned}$$

y una aplicación del Lema 3.2 (ii) permite escribir

$$\begin{aligned}
&Df(\mathbf{p})\mathbb{D}(\mathbf{p})Df(\mathbf{p})' \\
&= [\mathbf{I} - \mathbb{M}^{FC}(\mathbf{p}) - \mathbb{M}^{CF}(\mathbf{p})] \times [\mathbb{D}(\mathbf{p}) - \mathbb{D}(\mathbf{p})\mathbb{M}^{FC}(\mathbf{p})' - \mathbb{D}(\mathbf{p})\mathbb{M}^{CF}(\mathbf{p})'] \\
&= [\mathbf{I} - \mathbb{M}^{FC}(\mathbf{p}) - \mathbb{M}^{CF}(\mathbf{p})] \times [\mathbb{D}(\mathbf{p}) - \mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p}) - \mathbb{M}^{CF}(\mathbf{p})\mathbb{D}(\mathbf{p})] \\
&= [\mathbf{I} - \mathbb{M}^{FC}(\mathbf{p}) - \mathbb{M}^{CF}(\mathbf{p})] \times [\mathbf{I} - \mathbb{M}^{FC}(\mathbf{p}) - \mathbb{M}^{CF}(\mathbf{p})]\mathbb{D}(\mathbf{p}) \\
&= [\mathbf{I} - \mathbb{M}^{FC}(\mathbf{p}) - \mathbb{M}^{CF}(\mathbf{p})]^2\mathbb{D}(\mathbf{p}).
\end{aligned}$$

(3.7)

A continuación observe que

$$\begin{aligned}
& [\mathbf{I} - \mathbf{M}^{FC}(\mathbf{p}) - \mathbf{M}^{CF}(\mathbf{p})]^2 \\
&= [\mathbf{I} - \mathbf{M}^{FC}(\mathbf{p}) - \mathbf{M}^{CF}(\mathbf{p})] - [\mathbf{I} - \mathbf{M}^{FC}(\mathbf{p}) - \mathbf{M}^{CF}(\mathbf{p})]\mathbf{M}^{FC}(\mathbf{p}) \\
&\quad - [\mathbf{I} - \mathbf{M}^{FC}(\mathbf{p}) - \mathbf{M}^{CF}(\mathbf{p})]\mathbf{M}^{CF}(\mathbf{p}) \\
&= [\mathbf{I} - \mathbf{M}^{FC}(\mathbf{p}) - \mathbf{M}^{CF}(\mathbf{p})] - [\mathbf{M}^{FC}(\mathbf{p}) - \mathbf{M}^{FC}(\mathbf{p})^2 - \mathbf{M}^{CF}(\mathbf{p})\mathbf{M}^{FC}(\mathbf{p})] \\
&\quad - [\mathbf{M}^{CF}(\mathbf{p}) - \mathbf{M}^{FC}(\mathbf{p})\mathbf{M}^{CF}(\mathbf{p}) - \mathbf{M}^{CF}(\mathbf{p})^2]
\end{aligned}$$

y utilizando las propiedades de idempotencia establecidas en el Lema 3.1(i), se obtiene

$$\begin{aligned}
& [\mathbf{I} - \mathbf{M}^{FC}(\mathbf{p}) - \mathbf{M}^{CF}(\mathbf{p})]^2 \\
&= [\mathbf{I} - \mathbf{M}^{FC}(\mathbf{p}) - \mathbf{M}^{CF}(\mathbf{p})] - [\mathbf{M}^{FC}(\mathbf{p}) - \mathbf{M}^{FC}(\mathbf{p}) - \mathbf{M}^{CF}(\mathbf{p})\mathbf{M}^{FC}(\mathbf{p})] \\
&\quad - [\mathbf{M}^{CF}(\mathbf{p}) - \mathbf{M}^{FC}(\mathbf{p})\mathbf{M}^{CF}(\mathbf{p}) - \mathbf{M}^{CF}(\mathbf{p})] \\
&= [\mathbf{I} - \mathbf{M}^{FC}(\mathbf{p}) - \mathbf{M}^{CF}(\mathbf{p})] + \mathbf{M}^{CF}(\mathbf{p})\mathbf{M}^{FC}(\mathbf{p}) + \mathbf{M}^{FC}(\mathbf{p})\mathbf{M}^{CF}(\mathbf{p}),
\end{aligned}$$

de tal forma que las igualdades en (3.3) implican

$$[\mathbf{I} - \mathbf{M}^{FC}(\mathbf{p}) - \mathbf{M}^{CF}(\mathbf{p})]^2 = [\mathbf{I} - \mathbf{M}^{FC}(\mathbf{p}) - \mathbf{M}^{CF}(\mathbf{p})] + 2\mathbf{p}\mathbf{1}',$$

y combinando esta igualdad con (3.7) se obtiene

$$\begin{aligned}
D\mathbf{f}(\mathbf{p})\mathbf{ID}(\mathbf{p})D\mathbf{f}(\mathbf{p})' &= ([\mathbf{I} - \mathbf{M}^{FC}(\mathbf{p}) - \mathbf{M}^{CF}(\mathbf{p})] + 2\mathbf{p}\mathbf{1}')\mathbf{ID}(\mathbf{p}). \\
&= [\mathbf{I} - \mathbf{M}^{FC}(\mathbf{p}) - \mathbf{M}^{CF}(\mathbf{p})]\mathbf{ID}(\mathbf{p}) + 2\mathbf{p}\mathbf{1}'\mathbf{ID}(\mathbf{p})
\end{aligned}$$

y observando que  $\mathbf{1}'\mathbf{ID}(\mathbf{p}) = \mathbf{p}'$  se arriba a la conclusión deseada.  $\square$

## Idempotencia de W

En esta sección se analiza la matriz de covarianza asintótica del vector aleatorio  $[\sqrt{n}\mathbf{ID}(\mathbf{p})]^{-1/2}\mathbf{f}(\bar{\mathbf{X}}_n)$  considerado en la Sección 2. Por conveniencia, la expresión para dicha matriz se reproduce a continuación:

$$\mathbf{W} = [\mathbf{ID}(\mathbf{p})]^{-1/2}D\mathbf{f}(\mathbf{p})[\mathbf{ID}(\mathbf{p}) - \mathbf{p}\mathbf{p}']D\mathbf{f}(\mathbf{p})'[\mathbf{ID}(\mathbf{p})]^{-1/2}. \quad (4.1)$$

**Teorema 4.1.** Bajo el supuesto de que los dos rasgos que dan origen a la clasificación cruzada son independientes, se tiene que la matriz  $\mathbf{W}$  en (4.1) es idempotente.

El argumento que conduce a establecer este resultado utiliza una expresión para la matriz  $\mathbf{W}$  que es más simple de analizar que la igualdad (4.1). La demostración del teorema precedente utiliza los siguientes dos lemas.

**Lema 4.1.** Si las características que determinan la clasificación cruzada son independientes, entonces

$$Df(\mathbf{p})[\mathbf{ID}(\mathbf{p}) - \mathbf{p}\mathbf{p}']Df(\mathbf{p})' = [\mathbf{I} - \mathbf{M}^{FC}(\mathbf{p}) - \mathbf{M}^{CF}(\mathbf{p})]\mathbf{ID}(\mathbf{p}) + \mathbf{p}\mathbf{p}' \quad (4.2)$$

**Demostración.** De acuerdo al Lema 3.1(*iv*) se tiene que  $Df(\mathbf{p})\mathbf{p} = -\mathbf{p}$ , y entonces

$$Df(\mathbf{p})\mathbf{p}\mathbf{p}'Df(\mathbf{p})' = [Df(\mathbf{p})\mathbf{p}][\mathbf{p}'Df(\mathbf{p})'] = [Df(\mathbf{p})\mathbf{p}][Df(\mathbf{p})\mathbf{p}]' = [-\mathbf{p}][-\mathbf{p}]' = \mathbf{p}\mathbf{p}',$$

y la igualdad (4.2) se obtiene combinando esta relación con la expresión obtenida en el Lema 3.3.  $\square$

Utilizando (4.2), es posible obtener una fórmula alternativa para la matriz  $\mathbf{W}$ . Primero, es conveniente introducir la siguiente notación.

**Definición 4.1.** El vector  $\tilde{\mathbf{p}} \in \mathbb{R}^k$  está definido mediante

$$\tilde{\mathbf{p}} = \begin{bmatrix} \tilde{p}_1 \\ \tilde{p}_2 \\ \vdots \\ \tilde{p}_k \end{bmatrix} = \begin{bmatrix} \sqrt{p_1} \\ \sqrt{p_2} \\ \vdots \\ \sqrt{p_k} \end{bmatrix}.$$

Observe que  $[\mathbf{ID}(\mathbf{p})]^{1/2} = \mathbf{ID}(\tilde{\mathbf{p}})$  (vea la Definición 2.2 del Capítulo 2), y entonces no es difícil verificar que

$$\mathbf{1}'\mathbf{ID}(\mathbf{p})^{1/2} = \tilde{\mathbf{p}}', \quad \mathbf{ID}(\mathbf{p})^{-1/2}\mathbf{p} = \tilde{\mathbf{p}}, \quad (4.3)$$

mientras que

$$\mathbb{D}(\mathbf{p})^{1/2}\tilde{\mathbf{p}} = \mathbf{p}. \quad (4.4)$$

Combinando las ecuaciones (4.1) y (4.2) se obtiene que, cuando los rasgos que determinan la partición cruzada son independientes, entonces

$$\begin{aligned} \mathbf{W} &= \mathbb{D}(\mathbf{p})^{-1/2} \times ([\mathbf{I} - \mathbb{M}^{FC}(\mathbf{p}) - \mathbb{M}^{CF}(\mathbf{p})]\mathbb{D}(\mathbf{p}) + \mathbf{p}\mathbf{p}') \times \mathbb{D}(\mathbf{p})^{-1/2} \\ &= \mathbf{I} - \mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2} - \mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{CF}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2} \\ &\quad + \mathbb{D}(\mathbf{p})^{-1/2}\mathbf{p}\mathbf{p}'\mathbb{D}(\mathbf{p})^{-1/2} \\ &= \mathbf{I} - \mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2} - \mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{CF}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2} \\ &\quad + \mathbb{D}(\mathbf{p})^{-1/2}\mathbf{p}[\mathbb{D}(\mathbf{p})^{-1/2}\mathbf{p}]' \end{aligned}$$

y (4.3) implica que

$$\mathbf{W} = \mathbf{I} - \mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2} - \mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{CF}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2} + \tilde{\mathbf{p}}\tilde{\mathbf{p}}' \quad (4.5)$$

**Lema 4.2.** Si las características que inducen la partición cruzada son independientes, entonces las siguientes afirmaciones (i)–(iii) son válidas:

- (i)  $\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2}\mathbf{W} = \mathbf{0}$ ;
- (ii)  $\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{CF}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2}\mathbf{W} = \mathbf{0}$ , y
- (iii)  $\tilde{\mathbf{p}}\tilde{\mathbf{p}}'\mathbf{W} = \mathbf{0}$ .

**Demostración.** (i) Utilizando la expresión (4.5) se desprende que

$$\begin{aligned} &\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2}\mathbf{W} \\ &= [\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2}] \\ &\quad - [\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2}]\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2} \\ &\quad - [\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2}]\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{CF}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2} \\ &\quad + [\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2}]\tilde{\mathbf{p}}\tilde{\mathbf{p}}' \end{aligned} \quad (4.6)$$

A continuación se analizarán los tres últimos términos en el lado derecho de esta ecuación. Observe que  $\mathbb{D}(\mathbf{p})^{1/2}\mathbb{D}(\mathbf{p})^{-1/2} = \mathbf{I}$ , de manera que

$$\begin{aligned} & [\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2}]\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2} \\ &= \mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2} \\ &= \mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2} \end{aligned} \quad (4.7)$$

donde la segunda igualdad se desprende del Lema 3.1(i). Por otro lado, utilizando (3.3) se desprende que

$$\begin{aligned} & [\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2}]\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{CF}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2} \\ &= \mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{M}^{CF}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2} \\ &= \mathbb{D}(\mathbf{p})^{-1/2}\mathbf{p}\mathbf{1}'\mathbb{D}(\mathbf{p})^{1/2} \end{aligned}$$

y entonces (4.3) implica que

$$\begin{aligned} & [\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2}]\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{CF}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2} \\ &= [\mathbb{D}(\mathbf{p})^{-1/2}\mathbf{p}] \times [\mathbf{1}'\mathbb{D}(\mathbf{p})^{1/2}] \\ &= \widetilde{\mathbf{p}}\widetilde{\mathbf{p}}' \end{aligned} \quad (4.8)$$

Considere ahora el último término en el lado derecho de (4.6). Por medio de (4.4) se obtiene

$$\begin{aligned} & \mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2}\widetilde{\mathbf{p}}\widetilde{\mathbf{p}}' \\ &= \mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p}) \times [\mathbb{D}(\mathbf{p})^{1/2}\widetilde{\mathbf{p}}] \times \widetilde{\mathbf{p}}' \\ &= \mathbb{D}(\mathbf{p})^{-1/2} \underbrace{\mathbb{M}^{FC}(\mathbf{p}) \times \mathbf{p}}_{\mathbf{p}} \times \widetilde{\mathbf{p}}' \\ &= \mathbb{D}(\mathbf{p})^{-1/2}\mathbf{p} \times \widetilde{\mathbf{p}}' \quad (\text{vea el Lema 3.1}(iv)) \\ &= \widetilde{\mathbf{p}}\widetilde{\mathbf{p}}' \end{aligned} \quad (4.9)$$

donde se utilizó (4.3) para obtener la última igualdad. Combinando (4.6)–(4.9) se desprende que  $\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2}\mathbf{W} = \mathbf{0}$ , lo cual establece la parte (i), mientras que el apartado (ii) se demuestra de forma similar. En este momento es oportuno observar que la siguiente ecuación, análoga a (4.9), también es válida:

$$\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{CF}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2}\widetilde{\mathbf{p}}\widetilde{\mathbf{p}}' = \widetilde{\mathbf{p}}\widetilde{\mathbf{p}}' \quad (4.10)$$

(ii) La traza de  $\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{CF}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2}$  es  $r$ .

**Demostración.** (i) El argumento utiliza la propiedad básica de la transformación lineal ‘traza’, a saber, si  $\mathbf{A}$  y  $\mathbf{B}$  son dos matrices cuyo producto está bien definido y genera una matriz cuadrada, entonces,

$$\text{Tr}[\mathbf{AB}] = \text{Tr}[\mathbf{BA}];$$

de hecho esta propiedad caracteriza, módulo una constante multiplicativa, a la función traza (Hoffman y Kunze, 1975, Searle, 1982, Grossman, 1983). Utilizando esta relación se tiene que

$$\begin{aligned} \text{Tr}[\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2}] &= \text{Tr}[\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2}\mathbb{D}(\mathbf{p})^{-1/2}] \\ &= \text{Tr}[\mathbb{M}^{FC}(\mathbf{p})] \\ &= \sum_{I=1}^k \mathbb{M}_{I,I}^{FC}(\mathbf{p}) \\ &= \sum_{I=1}^k \mathbb{P}_{F(I)}(\mathbf{p})\mathcal{I}_{C(I)}(I) \end{aligned}$$

donde la última igualdad se debe a la definición de  $\mathbb{M}^{FC}(\mathbf{p})$ . Recordando que  $C(I)$  es el miembro de la partición  $C_1, C_2, \dots, C_s$  que contiene a  $I$ , es claro que  $\mathcal{I}_{C(I)}(I) = 1$ , de tal forma que

$$\text{Tr}[\mathbb{D}(\mathbf{p})^{-1/2}\mathbb{M}^{FC}(\mathbf{p})\mathbb{D}(\mathbf{p})^{1/2}] = \sum_{I=1}^k \mathbb{P}_{F(I)}(\mathbf{p}). \quad (5.1)$$

Para calcular el lado derecho de esta igualdad, recuerde que  $C_1, C_2, \dots, C_s$  es una partición del conjunto de índices  $\{1, 2, \dots, k\}$ , de manera que

$$\begin{aligned} \sum_{I=1}^k \mathbb{P}_{F(I)}(\mathbf{p}) &= \sum_{I \in C_1 \cup C_2 \cup \dots \cup C_s} \mathbb{P}_{F(I)}(\mathbf{p}) \\ &= \sum_{m=1}^s \sum_{I \in C_m} \mathbb{P}_{F(I)}(\mathbf{p}) \end{aligned}$$

A continuación, utilice el Lema 2.1 del Capítulo 3 para obtener que, para todo  $m = 1, 2, \dots, s$ ,

$$\sum_{I \in C_m} \mathbb{P}_{F(I)}(\mathbf{p}) = \mathbf{1}'\mathbf{p} = 1,$$

y entonces la anterior igualdad desplegada permite concluir que  $\sum_{I=1}^k \mathbb{P}_{F(I)}(\mathbf{p}) = \sum_{m=1}^s 1 = s$ ; combinando esta igualdad con (5.1) se obtiene que

$$\text{Tr}[\mathbb{D}(\mathbf{p})^{-1/2} \mathbb{M}^{FC}(\mathbf{p}) \mathbb{D}(\mathbf{p})^{1/2}] = s.$$

Este argumento demuestra la parte (i), mientras que (ii) se establece a través de lineamientos semejantes.  $\square$

**Demostración del Teorema 2.2.** La idempotencia de la matriz  $\mathbf{W}$  fue establecida en el Teorema 4.1, así que para concluir la demostración es suficiente verificar que

$$\text{Rango}[W] = (r - 1)(s - 1). \quad (5.2)$$

Para comprobar que esta igualdad es válida, recuerde que la idempotencia de  $\mathbf{W}$  implica que  $\text{Rango}[\mathbf{W}] = \text{Tr}[\mathbf{W}]$  (Hoffman y Kunze, 1975, Grossman, 1983, Searle, 1982). Teniendo en mente esta propiedad, y combinando (5.2) con (4.5) se obtiene, a través de la linealidad de la función traza, que

$$\begin{aligned} \text{Rango}[\mathbf{W}] &= \text{Tr}[\mathbf{W}] \\ &= \text{Tr}[\mathbf{I}] - \text{Tr}[\mathbb{D}(\mathbf{p})^{-1/2} \mathbb{M}^{FC}(\mathbf{p}) \mathbb{D}(\mathbf{p})^{1/2}] \\ &\quad - \text{Tr}[\mathbb{D}(\mathbf{p})^{-1/2} \mathbb{M}^{CF}(\mathbf{p}) \mathbb{D}(\mathbf{p})^{1/2}] + \text{Tr}[\tilde{\mathbf{p}}\tilde{\mathbf{p}}'] \\ &= rs - s - r + \text{Tr}[\tilde{\mathbf{p}}\tilde{\mathbf{p}}']; \end{aligned} \quad (5.3)$$

recuerde que el número de categorías de resultados es  $k = rs$ , y que la matriz identidad tiene orden  $k \times k$ , de manera que  $\text{Tr}[\mathbf{I}] = k = rs$ . Por otro lado, de acuerdo a (4.10), la matriz  $\mathbf{p}\mathbf{p}'$  es idempotente, de manera que su traza coincide con su rango; puesto que este último es claramente uno, se tiene que  $\text{Tr}[\mathbf{p}\mathbf{p}'] = 1$ , y entonces (5.3) implica

$$\text{Rango}[\mathbf{W}] = rs - s - r + 1 = (r - 1)(s - 1),$$

concluyendo la demostración del Teorema. □

## El Procedimiento de Prueba

En esta sección se utiliza el Teorema 2.1 para establecer la prueba Ji-cuadrada en el análisis de una tabla de contingencia, la cual permite probar la hipótesis de independencia de los rasgos que originan la clasificación cruzada. Como ya se ha mencionado, el supuesto de independencia se rechaza cuando el vector  $\mathbf{f}(\bar{\mathbf{X}}_n)$  asume un valor que es ‘sustancialmente’ distinto de  $\mathbf{0}$ . El resultado en el Teorema 2.1, sugiere una forma de formalizar esta idea: Como medida de la ‘longitud’ de  $\mathbf{f}(\bar{\mathbf{X}}_n)$ , seleccione a  $\|\sqrt{n}\mathbb{D}(\mathbf{p})^{-1/2}\mathbf{f}(\bar{\mathbf{X}}_n)\|$ , y rechace el supuesto de independencia de los rasgos de interés cuando  $\|\sqrt{n}\mathbb{D}(\mathbf{p})^{-1/2}\mathbf{f}(\bar{\mathbf{X}}_n)\|^2 > G$ , donde la constante  $G$  es tal que, si los rasgos de interés son independientes, entonces

$$P \left[ \|\sqrt{n}\mathbb{D}(\mathbf{p})^{-1/2}\mathbf{f}(\bar{\mathbf{X}}_n)\|^2 > G \right] = \alpha,$$

donde  $\alpha$  es el nivel de significancia deseado; por el Teorema 2.1, seleccionando  $G = \chi_{\alpha:(r-1)(s-1)}^2$ , la probabilidad en la anterior igualdad será aproximadamente igual a  $\alpha$ , al menos cuando el tamaño de la muestra  $n$  sea grande. Así, en primera instancia, un procedimiento para probar el supuesto de independencia de los rasgos de interés, podría formularse como sigue:

$$\text{Rechace } \mathcal{H}_0: \mathbf{f}(\mathbf{p}) = \mathbf{0} \text{ cuando } \|\sqrt{n}\mathbb{D}(\mathbf{p})^{-1/2}\mathbf{f}(\bar{\mathbf{X}}_n)\|^2 > \chi_{\alpha:(r-1)(s-1)}^2.$$

Sin embargo, hay una dificultad que emerge de inmediato al reflexionar sobre esta propuesta, a saber, el término a la izquierda de la desigualdad no es un estadístico, pues involucra al vector  $\mathbf{p}$ , el cual se desconoce. Por lo tanto, después de realizar el experimento, el observador no podrá evaluar la cantidad  $\|\sqrt{n}\mathbb{D}(\mathbf{p})^{-1/2}\mathbf{f}(\bar{\mathbf{X}}_n)\|^2$ , y tampoco podrá decidir si excede al percentil  $\chi_{\alpha:(r-1)(s-1)}^2$ . Sin embargo, ya se ha demostrado en el Capítulo 2 que el vector  $\bar{\mathbf{X}}_n$  es un estimador consistente de  $\mathbf{p}$ , de manera que el observador puede, razonablemente, sustituir  $\mathbf{p}$  por  $\bar{\mathbf{X}}_n$ , y adoptar

la siguiente regla de decisión:

$$\text{Rechace } \mathcal{H}_0: \mathbf{f}(\mathbf{p}) = \mathbf{0} \text{ cuando } \left\| \sqrt{n} \mathbb{D}(\bar{\mathbf{X}}_n)^{-1/2} \mathbf{f}(\bar{\mathbf{X}}_n) \right\|^2 > \chi_{\alpha: (r-1)(s-1)}^2. \quad (6.1)$$

Esta es la prueba Ji-cuadrada para probar la independencia de las características de interés que originan la clasificación cruzada. Una pregunta importante que surge de inmediato, es si acaso la probabilidad de rechazar  $\mathcal{H}_0$  con el procedimiento (6.1) es, aproximadamente, igual a  $\alpha$ , de manera que método determine una prueba cuyo tamaño asintótico sea el estipulado. El siguiente teorema muestra que, efectivamente, este es el caso.

**Teorema 6.1.** Bajo el supuesto

$$\mathcal{H}_0: \mathbf{f}(\mathbf{p}) = \mathbf{0},$$

$$\left\| \sqrt{n} \mathbb{D}(\bar{\mathbf{X}}_n)^{-1/2} \mathbf{f}(\bar{\mathbf{X}}_n) \right\|^2 = n \mathbf{f}(\bar{\mathbf{X}}_n)' \mathbb{D}(\bar{\mathbf{X}}_n)^{-1} \mathbf{f}(\bar{\mathbf{X}}_n) \xrightarrow{\mathcal{D}} \chi_{(r-1)(s-1)}^2. \quad (6.2)$$

Consecuentemente, para cada  $\alpha \in (0, 1)$ ,

$$\lim_{n \rightarrow \infty} P \left[ n \mathbf{f}(\bar{\mathbf{X}}_n)' \mathbb{D}(\bar{\mathbf{X}}_n)^{-1} \mathbf{f}(\bar{\mathbf{X}}_n) > \chi_{\alpha: (r-1)(s-1)}^2 \right] = \alpha. \quad (6.3)$$

La demostración de este resultado involucra la noción de acotamiento en probabilidad enunciada en el Capítulo 2, y por claridad, primero se establecen los prerequisites necesarios en el siguiente Lema.

**Lema 6.1.** Para  $I = 1, 2, \dots, k$ , sea  $\mathbf{f}_I(\bar{\mathbf{X}}_n)$  la componente  $I$  de la función de comparación  $\mathbf{f}$  evaluada en  $\bar{\mathbf{X}}_n$ , y sea  $\bar{X}_{n,I}$  el  $I$ -ésimo elemento de la media muestral  $\bar{\mathbf{X}}_n$ .

(i) Para cada índice  $I$ , la sucesión de variables aleatoria  $\{n \mathbf{f}(\bar{\mathbf{X}}_n)^2\}$  es acotada en probabilidad; vea la Definición 4.1 en el Capítulo 2.

(ii) Conforme  $n$  tiende a  $\infty$ ,

$$\frac{1}{\bar{X}_{n,I}} - \frac{1}{p_I} \xrightarrow{P} 0,$$

donde ‘ $\xrightarrow{P}$ ’ representa ‘convergencia en probabilidad’.

Consecuentemente,

$$(iii) \quad \frac{n\mathbf{f}(\bar{\mathbf{X}}_n)^2}{\bar{X}_{nI}} - \frac{n\mathbf{f}(\bar{\mathbf{X}}_n)^2}{p_I} \xrightarrow{P} 0.$$

y

$$(iv). \quad n \sum_{I=1}^k \frac{\mathbf{f}(\bar{\mathbf{X}}_n)^2}{\bar{X}_{nI}} - n \sum_{I=1}^k \frac{\mathbf{f}(\bar{\mathbf{X}}_n)^2}{p_I} \xrightarrow{P} 0$$

**Demostración.** Primero observe que

$$\begin{aligned} \left\| \sqrt{n}\mathbb{D}(\bar{\mathbf{X}}_n)^{-1/2}\mathbf{f}(\bar{\mathbf{X}}_n) \right\|^2 &= n\mathbf{f}(\bar{\mathbf{X}}_n)'\mathbb{D}(\bar{\mathbf{X}}_n)^{-1}\mathbf{f}(\bar{\mathbf{X}}_n) \\ &= n \sum_{I=1}^k \frac{\mathbf{f}_I(\bar{\mathbf{X}}_n)^2}{\bar{X}_{n,I}}, \end{aligned} \quad (6.4)$$

y similarmente,

$$\left\| \sqrt{n}\mathbb{D}(\bar{\mathbf{p}})^{-1/2}\mathbf{f}(\bar{\mathbf{X}}_n) \right\|^2 = n \sum_{I=1}^k \frac{\mathbf{f}_I(\bar{\mathbf{X}}_n)^2}{p_I}. \quad (6.5)$$

(i) Observe que para cada  $I \in \{1, 2, \dots, k\}$  y  $K > 0$ , la igualdad (6.5) implica que

$$\begin{aligned} P [n\mathbf{f}_I(\bar{\mathbf{X}}_n)^2 > K] &= P \left[ n \frac{\mathbf{f}_I(\bar{\mathbf{X}}_n)^2}{p_I} > K \right] \\ &\leq P \left[ n \sum_{I=1}^k \frac{\mathbf{f}_I(\bar{\mathbf{X}}_n)^2}{p_I} > K \right] \\ &= P \left[ \left\| \sqrt{n}\mathbb{D}(\bar{\mathbf{p}})^{-1/2}\mathbf{f}(\bar{\mathbf{X}}_n) \right\|^2 > K \right] \end{aligned} \quad (6.6)$$

A continuación, sea  $Y$  una variable aleatoria con distribución  $\chi_{(r-1)(s-1)}^2$ , y dada  $\varepsilon > 0$ , seleccione un número  $K(\varepsilon)$  tal que  $P[Y > K(\varepsilon)] < \varepsilon/2$ . Debido a que  $\left\| \sqrt{n}\mathbb{D}(\bar{\mathbf{p}})^{-1/2}\mathbf{f}(\bar{\mathbf{X}}_n) \right\|^2 \xrightarrow{D} Y$ , por el Teorema 2.1, se tiene que

$$\lim_{n \rightarrow \infty} P \left[ \left\| \sqrt{n}\mathbb{D}(\bar{\mathbf{p}})^{-1/2}\mathbf{f}(\bar{\mathbf{X}}_n) \right\|^2 > K(\varepsilon) \right] = P[Y > K(\varepsilon)] = \frac{\varepsilon}{2}$$

de tal manera que, para algún entero positivo  $N_0$ ,

$$P \left[ \left\| \sqrt{n}\mathbb{D}(\bar{\mathbf{p}})^{-1/2}\mathbf{f}(\bar{\mathbf{X}}_n) \right\|^2 > K(\varepsilon) \right] < \varepsilon, \quad \text{si } n > N_0;$$

en este caso, (6.6) implica que

$$P [nf_I(\bar{\mathbf{X}}_n)^2 > K(\varepsilon)] < \varepsilon, \quad \text{cuando } n > N_0,$$

de donde se desprende que  $\{nf_I(\bar{\mathbf{X}}_n)^2\}$  es una sucesión acotada en probabilidad.

(ii) Utilizando el Teorema 2.1 y el Lema 4.2 del Capítulo 2, se desprende que

$$\bar{X}_{n,I} \xrightarrow{P} p_I,$$

y entonces el resultado general de invarianza presentado en el Teorema 3.1 del mismo capítulo, implica que  $\frac{1}{\bar{X}_{n,I}} \xrightarrow{P} \frac{1}{p_I}$ , lo cual equivale a

$$\frac{1}{\bar{X}_{n,I}} - \frac{1}{p_I} \xrightarrow{P} 0.$$

(iii) Combinando las partes (i) y (ii) con el Lema 4.3 del Capítulo 2, se obtiene

$$nf(\bar{\mathbf{X}}_n)^2 \left[ \frac{1}{\bar{X}_{nI}} - \frac{1}{p_I} \right] = \frac{nf(\bar{\mathbf{X}}_n)^2}{\bar{X}_{nI}} - \frac{nf(\bar{\mathbf{X}}_n)^2}{p_I} \xrightarrow{P} 0,$$

estableciendo la parte (iii), y entonces la parte (iv) se obtiene utilizando el hecho de que la suma de las variables aleatorias que convergen a cero en probabilidad, también tiene esta propiedad.  $\square$

**Demostración del Teorema 6.1.** Note que

$$\left\| \sqrt{n} \mathbb{D}(\bar{\mathbf{X}}_n)^{-1/2} \mathbf{f}(\bar{\mathbf{X}}_n) \right\|^2 = Z_n + \left\| \sqrt{n} \mathbb{D}(\mathbf{p})^{-1/2} \mathbf{f}(\bar{\mathbf{X}}_n) \right\|^2, \quad (6.7)$$

donde

$$Z_n = \left\| \sqrt{n} \mathbb{D}(\bar{\mathbf{X}}_n)^{-1/2} \mathbf{f}(\bar{\mathbf{X}}_n) \right\|^2 - \left\| \sqrt{n} \mathbb{D}(\mathbf{p})^{-1/2} \mathbf{f}(\bar{\mathbf{X}}_n) \right\|^2.$$

Por el Lema anterior,  $Z_n \xrightarrow{P} 0$ , mientras que

$$\left\| \sqrt{n} \mathbb{D}(\mathbf{p})^{-1/2} \mathbf{f}(\bar{\mathbf{X}}_n) \right\|^2 \xrightarrow{\mathcal{D}} \chi_{(r-1)(s-1)}^2,$$

por el Teorema 2.1. Luego, (6.7) implica que  $\left\| \sqrt{n} \mathbb{D}(\bar{\mathbf{X}}_n)^{-1/2} \mathbf{f}(\bar{\mathbf{X}}_n) \right\|^2 \xrightarrow{\mathcal{D}} 0$ ; vea, por ejemplo, Ash(1975).  $\square$

**Discusión.** En esta sección se ha formulado la prueba Ji-cuadrada para probar el supuesto de independencia de las características de interés que inducen una clasificación cruzada. El análisis realizado utilizó la notación vectorial en vez de la matricial que involucra dobles subíndices; vea el Capítulo 5. Sin embargo, el registro de las frecuencias en forma de una matriz ‘de contingencia’ es práctica común y conveniente, por lo cual es oportuno enunciar la prueba (6.1) en términos de la notación matricial. Si  $N_{i,j}$  es la frecuencia con que los datos observados se ubican en la celda  $(i,j)$  de la matriz de contingencia, y denote mediante  $N_{i,+}$  y  $N_{+,j}$  a la suma de las frecuencias observadas en la  $i$ -ésima fila y en la  $j$ -ésima columna, respectivamente. Si el índice  $I$  corresponde a la pareja  $(i,j)$ , entonces no es difícil ver que

$$P_{F(I)}(\bar{X}_{nI}) = \frac{N_{i,+}}{n}, \quad P_{C(I)}(\bar{X}_{nI}) = \frac{N_{+,j}}{n}.$$

Debido a que  $\bar{X}_{n,I} = \frac{N_{i,j}}{n}$  cuando el índice  $I$  está asociado a la pareja  $(i,j)$ , se desprende que

$$\mathbf{f}_I(\bar{X}_{n,I}) = \frac{N_{i,j}}{n} - \left( \frac{N_{i,+}}{n} \right) \left( \frac{N_{+,j}}{n} \right).$$

Luego,

$$\begin{aligned} \left\| \sqrt{n} \mathbf{D}(\bar{\mathbf{X}}_n)^{-1/2} \mathbf{f}(\bar{\mathbf{X}}_n) \right\|^2 &= n \mathbf{f}(\bar{\mathbf{X}}_n)' \mathbf{D}(\bar{\mathbf{X}}_n)^{-1} \mathbf{f}(\bar{\mathbf{X}}_n) \\ &= n \sum_{I=1}^k \frac{\mathbf{f}_I(\bar{\mathbf{X}}_n)^2}{\bar{X}_{n,I}} \\ &= \sum_{i=1}^r \sum_{j=1}^s \frac{1}{N_{i,j}} \left( N_{i,j} - \frac{N_{i,+} N_{+,j}}{n} \right)^2 \end{aligned}$$

y (6.1) es equivalente al siguiente procedimiento de prueba:

Rechace el supuesto de que los rasgos de interés que inducen la clasificación cruzada son independientes, si y sólo si

$$\sum_{i=1}^r \sum_{j=1}^s \frac{1}{N_{i,j}} \left( N_{i,j} - \frac{N_{i,+} N_{+,j}}{n} \right)^2 > \chi_{\alpha; (r-1)(s-1)}^2. \quad (6.8)$$

## Conclusión

En este capítulo se estudió la prueba Ji-cuadrada para probar la independencia de las características que dan origen a una clasificación cruzada de datos. El análisis realizado contiene la principal contribución técnica de esta tesis, a saber, la determinación, por métodos algebraicos, del número adecuado de grados de libertad que se utilizan en la prueba, la cual fue enunciada en (6.1) utilizando notación vectorial, y en (6.8) empleando la notación matricial que, por su conveniencia, es ampliamente utilizada en las aplicaciones.

## Capítulo 5

### Epílogo

En esta trabajo se han estudiado dos problemas fundamentales en el análisis de datos categóricos, a saber, la valoración del ajuste de una muestra de una distribución multidimensional de Bernoulli a un parámetro específico, y la construcción de una prueba para la hipótesis de independencia de las características que inducen una clasificación cruzada.

La exposición formal inició con una presentación rigurosa de diversos resultados básicos sobre las ideas de convergencia en distribución y probabilidad para una sucesión de vectores aleatorios, en la cual se enfatizaron las propiedades de invarianza; el análisis detallado que se desarrolló en el Capítulo 2 mostró la importancia que, desde una perspectiva formal, tiene la noción de acotamiento en probabilidad. Los resultados obtenidos permitieron abordar el problema de bondad de ajuste, y establecer la correspondiente prueba Ji-cuadrada. Además, a través de la dualidad entre pruebas de hipótesis y regiones de confianza, se construyeron intervalos de confianza simultáneos para funciones lineales del parámetro de una distribución multidimensional de Bernoulli; el enfoque seguido en esta parte del trabajo siguió las ideas que se emplean en el desarrollo del método de Scheffè, el cual es una técnica bien establecida en el estudio de modelos lineales.

En el Capítulo 3 se introdujo la noción de clasificación cruzada de datos por medio de dos características de interés, y se planteó la hipótesis de independencia de los rasgos que originan dicha clasificación por medio de una *función de comparación*, la cual, de manera natural, se convirtió en el objeto central de estudio de esta trabajo; se ejemplificó de manera detallada su construcción, así como el cálculo de la correspondiente matriz derivada.

Posteriormente, en el Capítulo 4 se presentó la principal contribución técnica de esta tesis, a saber, la determinación algebraica del número de grados de libertad en la prueba Ji-cuadrada para el supuesto de independencia de los rasgos de interés de una tabla de contingencia. El estudio que se realizó sobre este aspecto, implicó un análisis matricial detallado.

## Literatura Citada

- Agresti, A. (1984), *Analysis of Ordinal Categorical Data*, Wiley, New York.
- Ash, R. B. (1975), *Real Analysis and Probability*, Academic Press, New York.
- Bishop, Y. M. M., S. E. Feinberg, y P. W. Holland, (1975), *Discrete Multivariate Analysis*, MIT Press, Cambridge, MA.
- Box, G. E. P., y N. R. Draper (1987), *Empirical Model-Building and Response Surfaces*, Wiley, New York.
- Christensen, R. (1996), *Plane Answers to Complex Questions: The Theory of Linear Models*, Springer-Verlag, New York.
- Dudewicz, E., y N. Mishra, (1988), *Modern Mathematical Statistics*, Wiley, New York.
- Feinberg, S. E. (1980), *The Analysis of Cross Classified Statistical Data*, Segunda Edición, MIT Press, Cambridge, MA.
- Fisher, L., y J. McDonald (1981), *Fixed Effects Analysis of Variance*, Academic Press, New York.
- Graybill, F. A. (1985), *Theory and Application of the Linear Model*, 1st. Edition, Wadsworth, Belmont, CA
- Grossman, S. I. (1983), *Algebra Lineal*, Grupo Editorial Iberoamérica, Tercera Edición, México D. F.
- Harville, D. A. (1997), *Matrix Algebra from a Statistician Perspective*, Springer-Verlag, New York
- Hoffman, K., y R. Kunze. (1973) *Algebra Lineal*, Prentice Hall Hispanoamericana, México D. F.
- Kendall, M., y A. Stuart, (1973), *The Advanced Theory of Statistics*, Vol. 2, Segunda Edición, Haffner, New, York.
- Lehman, E. L. (1991), *Testing Statistical Hypothesis*, Wiley, New York.
- Lindgren, B. W. (1976), *Statistical Theory*, Tercera Edición, Macmillan, New York.
- Mandansky, A. (1988), *Prescriptions for Working Statisticians*, Springer-Verlag, New York.
- Mood, A. M., F. A. Graybill y D. C. Boes (1985), *Introduction to the Theory of Statistics*, McGraw-Hill, New York.
- Searle, S. R. (1982), *Matrix Algebra Useful for Statistics*, McGraw-Hill, New York.