

Estadística Bayesiana

**Fundamentos  
de Estadística Bayesiana**

Lizbeth Naranjo Albarrán

Facultad de Ciencias, UNAM

March 7, 2022

# Índice

<b>1</b>	<b>Fundamentos</b>	<b>1</b>
1.1	Intercambiabilidad . . . . .	1
1.2	Teorema de representación de <i>de Finetti</i> . . . . .	5
1.3	Principio de Verosimilitud . . . . .	7
1.4	Suficiencia y Familias Exponenciales . . . . .	8
1.4.1	Estadística Suficiente . . . . .	8
1.4.2	Familias Exponenciales . . . . .	12

# Capítulo 1

## Fundamentos

En este capítulo estudiamos los conceptos de intercambiabilidad, el teorema de *de Finetti* y las propiedades de las estadísticas suficientes y las familias exponenciales.

### 1.1 Intercambiabilidad

Consideramos a  $X_1, \dots, X_n$  variables aleatorias, cuyo comportamiento se describe a través de la especificación de una distribución conjunta, digamos  $p(x_1, \dots, x_n)$ . Esta distribución define de manera implícita otras especificaciones que pueden ser de gran interés. Por ejemplo, para  $1 \leq m \leq n$ ,

$$p(x_1, \dots, x_m) = \int \cdots \int p(x_1, \dots, x_m, x_{m+1}, \dots, x_n) dx_{m+1} \cdots dx_n,$$

es la distribución marginal de  $(X_1, \dots, X_m)$ , mientras que

$$p(x_{m+1}, \dots, x_n | x_1, \dots, x_m) = \frac{p(x_1, \dots, x_m, x_{m+1}, \dots, x_n)}{p(x_1, \dots, x_m)}$$

es la distribución condicional de las variables,  $X_{m+1}, \dots, X_n$ , dados los datos  $X_1 = x_1, \dots, X_m = x_m$ .

La especificación directa de  $p(x_1, \dots, x_n)$  puede llegar a ser muy difícil en la práctica. Es por eso que conviene examinar con cuidado el proceso de selección de una medida de probabilidad específica que describa adecuadamente nuestro conocimiento.

**Definición 1.1.** Un modelo predictivo para una sucesión de variables aleatorias  $X_1, X_2, \dots$  es una medida de probabilidad  $P$ , la cual especifica la forma de la distribución conjunta que describe nuestro conocimiento acerca de cualquier subconjunto de la sucesión.

Considere una sucesión de variables aleatorias  $X_1, X_2, \dots$ , y supongamos un modelo predictivo que especifica, para todo  $n \in \mathbb{N}$ , una distribución conjunta de la forma

$$p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i),$$

de manera que las variables aleatorias  $X_i$  son independientes. Claramente, para cualquier  $1 \leq m \leq n$  se tiene entonces que

$$p(x_{m+1}, \dots, x_n | x_1, \dots, x_m) = p(x_{m+1}, \dots, x_n),$$

por lo que bajo este modelo las observaciones  $x_1, \dots, x_m$  no proporcionan información alguna sobre las observaciones futuras; no hay aprendizaje a partir de la experiencia dada.

Un modelo predictivo con una estructura de independencia es inapropiado en contextos donde creemos que los datos darán información acerca de eventos futuros. En estos casos, la densidad conjunta debe tener estructura de dependencia entre las variables aleatorias.

Veremos ahora una forma particular de juicio subjetivo acerca de ciertas estructuras simples de dependencia pero que pueden corresponder a juicios reales en situaciones de interés práctico.

**Definición 1.2** (Intercambiabilidad finita). Se dice que las variables aleatorias  $X_1, \dots, X_n$  son (finitamente) intercambiables bajo una medida de probabilidad  $P$  si la distribución inducida por  $P$  satisface

$$p(x_1, \dots, x_n) = p(x_{\pi(1)}, \dots, x_{\pi(n)})$$

para toda permutación  $\pi$  definida sobre el conjunto  $\{1, 2, \dots, n\}$ .

En otras palabras, las “etiquetas” que identifican a cada una de las variables no proporcionan información alguna.

Es claro que si las variables  $X_1, \dots, X_n$  son independientes e idénticamente distribuidas entonces son intercambiables. El inverso no siempre es verdad.

**Ejemplo 1.1** (Intercambiabilidad: Normal). Sea  $\mathbf{X} = (X_1, \dots, X_n)$  un vector aleatorio con distribución Normal multivariada  $N_n(\mathbf{0}, \Sigma)$  y supongamos que los elementos de la diagonal de  $\Sigma$  son todos idénticos. Consideremos los siguientes dos casos:

- i) La matriz  $\Sigma$  es diagonal. Entonces las variables aleatorias  $X_1, \dots, X_n$  son independientes e idénticamente distribuidas, y por lo tanto intercambiables.
- ii) La matriz  $\Sigma$  *no* es diagonal. En este caso las variables aleatorias  $X_1, \dots, X_n$  no son independientes, pero siguen siendo intercambiables.



**Definición 1.3** (Intercambiabilidad infinita). La sucesión infinita de variables aleatorias  $X_1, X_2, \dots$  es (infinitamente) intercambiable si toda subsucesión finita es intercambiable en el sentido de la definición 1.2.

No toda colección finita de variables aleatorias intercambiables puede anidarse en una sucesión infinita de variables aleatorias intercambiables definidas de manera similar. Más aún, una colección finita de variables aleatorias intercambiables no necesariamente puede anidarse en una colección finita más grande de variables aleatorias intercambiables.

**Ejemplo 1.2** (Intercambiabilidad: Bernoulli). Sean  $X_1$  y  $X_2$  variables aleatorias que toman valores en el conjunto  $\{0, 1\}$  y con función de probabilidad conjunta dada por

$$\begin{aligned} P(X_1 = 0, X_2 = 1) &= P(X_1 = 1, X_2 = 0) = \frac{1}{2} \\ P(X_1 = 0, X_2 = 0) &= P(X_1 = 1, X_2 = 1) = 0 \end{aligned}$$

de manera que  $X_1$  y  $X_2$ . Se puede mostrar que No existe una variable  $X_3$  que tome valores en  $\{0, 1\}$  y tal que  $X_1, X_2$  y  $X_3$  sean intercambiables.

Para demostrar esto, deduciremos por reducción al absurdo.

Suponga que existe  $X_3$  tal que  $X_1, X_2$  y  $X_3$  sí son intercambiables. En particular, si la propiedad de intercambiabilidad se cumple, entonces se tiene que,

$$P(X_1 = 0, X_2 = 1, X_3 = 0) = P(X_1 = 0, X_2 = 0, X_3 = 1).$$

Sin embargo,

$$\begin{aligned} &P(X_1 = 0, X_2 = 1, X_3 = 0) \\ &= P(X_1 = 0, X_2 = 1) - P(X_1 = 0, X_2 = 1, X_3 = 1) \\ &= \frac{1}{2} - P(X_1 = 0, X_2 = 1, X_3 = 1) \\ &= \frac{1}{2} - P(X_1 = 1, X_2 = 1, X_3 = 0) \end{aligned}$$

Pero,

$$P(X_1 = 1, X_2 = 1, X_3 = 0) \leq P(X_1 = 1, X_2 = 1) = 0$$

Entonces,

$$P(X_1 = 0, X_2 = 1, X_3 = 0) = \frac{1}{2}$$

Por otro lado,

$$P(X_1 = 0, X_2 = 0, X_3 = 1) \leq P(X_1 = 0, X_2 = 0) = 0$$

Por lo tanto,

$$P(X_1 = 0, X_2 = 0, X_3 = 1) \neq P(X_1 = 0, X_2 = 1, X_3 = 0)$$

Lo cual se contradice con el supuesto de intercambiabilidad.

Por lo tanto,  $X_1, X_2, X_3$  y  $X_4$  No son intercambiables. ■

**Ejemplo 1.3** (Intercambiabilidad: Bernoulli). Sean  $X_1, X_2$  y  $X_3$  variables aleatorias que toman valores en el conjunto  $\{0, 1\}$  y con función de probabilidad conjunta dada por

$$\begin{aligned} P(X_1 = 0, X_2 = 1, X_3 = 1) &= P(X_1 = 1, X_2 = 0, X_3 = 1) \\ &= P(X_1 = 1, X_2 = 1, X_3 = 0) \\ &= 1/3, \end{aligned}$$

y donde cualquier otra combinación tiene probabilidad cero, de manera que  $X_1, X_2$  y  $X_3$  son intercambiables. Se puede mostrar que no existe una variable  $X_4$  que tome valores en  $\{0, 1\}$  y tal que  $X_1, X_2, X_3$  y  $X_4$  sean intercambiables.

Para demostrar esto, deduciremos por reducción al absurdo.

Suponga que existe  $X_4$  tal que  $X_1, X_2, X_3$  y  $X_4$  sí son intercambiables. En particular, si la propiedad de intercambiabilidad se cumple, entonces se tiene que,

$$P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0) = P(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1).$$

Sin embargo,

$$\begin{aligned} &P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0) \\ &= P(X_1 = 0, X_2 = 1, X_3 = 1) - P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 1) \\ &= \frac{1}{3} - P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 1) \\ &= \frac{1}{3} - P(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0) \end{aligned}$$

Pero,

$$P(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0) \leq P(X_1 = 1, X_2 = 1, X_3 = 1) = 0$$

Entonces,

$$P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0) = \frac{1}{3}$$

Por otro lado,

$$P(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1) \leq P(X_1 = 0, X_2 = 0, X_3 = 1) = 0$$

Por lo tanto,

$$P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0) \neq P(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1)$$

Lo cual se contradice con el supuesto de intercambiabilidad.

Por lo tanto,  $X_1, X_2, X_3$  y  $X_4$  No son intercambiables. ■

**Ejemplo 1.4** (Urna de Pólya). Se tiene una urna con  $b$  bolas blancas y  $n$  bolas negras. Seleccionar una bola, se observa su color, y se regrese a la urna, además se agrega a la urna  $k$  bolas del mismo color. Sea  $X = 1$  denota a la bola negra, y  $X = 0$  bola blanca.

- $X_1, X_2, \dots$  Sí es intercambiable.
- $X_1, X_2, \dots$  No es independiente.

$$\begin{aligned} P(1, 1, 0, 1) &= \frac{n}{b+n} \frac{n+k}{b+n+k} \frac{b}{b+n+2k} \frac{n+2k}{b+n+3k} \\ P(1, 0, 1, 0) &= \frac{n}{b+n} \frac{b}{b+n+k} \frac{n+k}{b+n+2k} \frac{n+2k}{b+n+3k} \end{aligned}$$

■

## 1.2 Teorema de representación de *de Finetti*

La importancia del concepto de intercambiabilidad queda expresada en el teorema de Representación de De Finetti, el cual proporciona una representación de la función de probabilidad conjunta de  $n$  variables aleatorias de una sucesión infinita de variables aleatorias intercambiables.

**Teorema 1.1** (Teorema de Representación de De Finetti). Si  $X_1, X_2, \dots$  es una sucesión infinita de variables aleatorias definidas sobre  $\mathfrak{R}$  e intercambiables con respecto a la medida de probabilidad  $P$ , entonces existe una función de distribución  $Q$  definida sobre  $\mathcal{F}$  (el espacio de todas las distribuciones sobre  $\mathfrak{R}$ ) tal que la distribución conjunta de  $X_1, \dots, X_n$  tiene la forma

$$P(x_1, \dots, x_n) = \int_{\mathcal{F}} \left\{ \prod_{i=1}^n F(x_i) \right\} dQ(F),$$

donde  $Q(F) = \lim_{n \rightarrow \infty} P(F_n)$ , y  $F_n$  es la función de distribución empírica de la muestra  $x_1, \dots, x_n$ .

A manera de ilustración, a continuación se presenta el caso más simple del Teorema de Representación.

**Teorema 1.2.** Si  $X_1, X_2, \dots$  es una sucesión infinita de variables aleatorias definidas sobre  $\{0, 1\}$  e intercambiables con respecto a la medida de probabilidad  $P$ , entonces existe una función de distribución  $Q$  tal que la función de probabilidad  $p(x_1, \dots, x_n)$  tiene la forma

$$p(x_1, \dots, x_n) = \int_0^1 \left\{ \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} \right\} dQ(\theta),$$

donde  $Q(\theta) = \lim_{n \rightarrow \infty} P(Y_n/n \leq \theta)$ , con  $Y_n = X_1 + \dots + X_n$ , y  $\theta = \lim_{n \rightarrow \infty} Y_n/n$  (c.s.).

Este teorema tiene un significado muy profundo desde el punto de vista de la modelación subjetiva. El teorema nos dice que el modelo predictivo para una sucesión intercambiable de variables aleatorias binarias puede describirse en términos de una situación en la que, condicional en el valor de una variable aleatoria  $\theta$ , las variables aleatorias  $X_i$  se consideran independientes con distribución Bernoulli y a  $\theta$  se le asigna una distribución de probabilidad  $Q$ .

El teorema muestra que sucesiones binarias intercambiables son mezclas de sucesiones Bernoulli.

Por la ley Fuerte de los Grandes Números,  $\theta = \lim_{n \rightarrow \infty} Y_n/n$  (c.s.), de manera que la distribución de probabilidad  $Q$  puede interpretarse como una descripción de los *juicios acerca del límite de la frecuencia relativa de los "éxitos" en la sucesión de ensayos Bernoulli*.

**Corolario 1.1.** Si  $X_1, X_2, \dots$  es una sucesión infinita de variables aleatorias definidas sobre  $\{0, 1\}$  e intercambiables con respecto a la medida de probabilidad  $P$ , entonces la distribución condicional  $p(x_{m+1}, \dots, x_n | x_1, \dots, x_m)$  es de la forma

$$p(x_{m+1}, \dots, x_n | x_1, \dots, x_m) = \int_0^1 \left\{ \prod_{i=m+1}^n \theta^{x_i} (1 - \theta)^{1-x_i} \right\} dQ(\theta | x_1, \dots, x_m)$$

donde  $1 \leq m < n$ ,

$$dQ(\theta | x_1, \dots, x_m) = \frac{\{\prod_{i=1}^m \theta^{x_i} (1 - \theta)^{1-x_i}\} dQ(\theta)}{\int_0^1 \{\prod_{i=1}^m \theta^{x_i} (1 - \theta)^{1-x_i}\} dQ(\theta)} \quad (1.1)$$

y  $Q(\theta) = \lim_{n \rightarrow \infty} P(Y_n/n \leq \theta)$ .

La expresión (1.1) no es más que una versión del Teorema de Bayes. Notemos que la forma de la representación no cambia. En la terminología usual, la distribución inicial  $Q(\theta)$  ha sido actualizada a través del Teorema de Bayes, obteniéndose la distribución final  $Q(\theta | x_1, \dots, x_m)$ .

La distribución predictiva (final)  $p(x_{m+1}, \dots, x_n | x_1, \dots, x_m)$  nos permite derivar la correspondiente distribución predictiva de cualquier otra variable definida en términos de las observaciones futuras.

Para más detalles de esta sección ver Bernardo y Smith (1994) y Gutiérrez Peña (1998).



## 1.3 Principio de Verosimilitud

El principio de verosimilitud establece que si dos experimentos producen la misma verosimilitud para  $\theta$ , excepto tal vez por una constante de proporcionalidad, entonces las inferencias obtenidas a partir de cada uno de los experimentos deben ser las mismas.

Toda la inferencia se basa en la verosimilitud. Si las verosimilitudes son proporcionales:

$$f_A(\underline{x}|\theta) \propto f_B(\underline{x}|\theta)$$

Entonces, las distribuciones finales, denotadas por:

$$\begin{aligned} f_A(\theta|\underline{x}) &\propto f_A(\underline{x}|\theta)f(\theta) \\ f_B(\theta|\underline{x}) &\propto f_B(\underline{x}|\theta)f(\theta) \end{aligned}$$

también son proporcionales, es decir,

$$f_A(\theta|\underline{x}) \propto f_A(\underline{x}|\theta)f(\theta) \propto f_B(\underline{x}|\theta)f(\theta) \propto f_B(\theta|\underline{x})$$

Lo que implica que  $f_A(\theta|\underline{x}) = f_B(\theta|\underline{x})$ . En los métodos de estadística frecuentista esto No siempre se cumple.

NOTA: El principio de verosimilitud sólo es válido cuando: (1) La inferencia se hace sobre el mismo parámetro; (2)  $\theta$  incluye cada factor desconocido del modelo.

**Ejemplo 1.5** (Principio de Verosimilitud).  $\theta$  prevalencia de una enfermedad. Se encuesta a 12, en los cuales 9 resultan positivos, y 3 negativos. Se tienen dos posibles verosimilitudes:

- $X \sim \text{Binomial}(n, \theta)$ , con  $E[X] = n\theta$ ,  $\text{Var}[X] = n\theta(1 - \theta)$ ,  $f(x|\theta) = \binom{n}{x}\theta^x(1 - \theta)^{n-x}$ . 'Se encuestan  $n$  y se cuenta el número de positivos'.

Bajo las observaciones  $n = 12$ ,  $x = 9$ ,

$$f(x|\theta) = \binom{12}{9}\theta^9(1 - \theta)^3$$

- $X \sim \text{Binomial} - \text{Negativa}(r, 1 - \theta)$ , con  $E[X] = r\frac{\theta}{1-\theta}$ ,  $\text{Var}[X] = r\frac{\theta}{(1-\theta)^2}$ ,  $f(x|\theta) = \binom{x+r-1}{x}(1 - \theta)^r\theta^x$ . 'Se encuestan  $n = x + r$  hasta que resultan  $r$  fallas'

Bajo las observaciones  $r = 3$ ,  $x = 9$ ,

$$f(x|\theta) = \binom{11}{9}(1 - \theta)^3\theta^9$$

En estadística frecuentista (clásica) los intervalos de confianza serán distintos para cada posible modelo. ■

## 1.4 Suficiencia y Familias Exponenciales

En esta sección se presenta una breve introducción al concepto de suficiencia y familias exponenciales, así como sus principales propiedades (ver Bernardo y Smith , 1994). Esto es de gran utilidad para el estudio de datos categóricos debido a que las distribuciones binomial, Poisson y multinomial pertenecen a la familia exponencial.

### 1.4.1 Estadística Suficiente

Una estadística suficiente es una función de los datos que resume toda la información de la muestra disponible referente al parámetro  $\theta$ .

#### Estadística

Comenzaremos con una definición formal, que nos permite discutir el proceso de resumir la información contenida en una sucesión, o *muestra*, de variables aleatorias,  $X_1, \dots, X_m$ .

**Definición 1.4** (Estadística). Dados los valores  $x_1, \dots, x_m$ , de las cantidades aleatorias  $X_1, \dots, X_m$ , un vector aleatorio  $\mathbf{t}_m(x_1, \dots, x_m)$ , con  $\mathbf{t}_m : \mathcal{X}_1 \times \dots \times \mathcal{X}_m \rightarrow \mathbb{R}^{k(m)}$  ( $k(m) \leq m$ ), es una estadística  $k(m)$ -dimensional.

Para reducir la dimensión de los datos claramente necesitamos que  $k(m) < m$ ; como en los casos anteriores la dimensión de la estadística es fija e independiente del tamaño de muestra, es decir,  $k(m) = k$ .

En la siguiente sección estudiaremos estadísticas que nos permiten resumir la mayor información posible.

#### Suficiencia Predictiva y Suficiencia Paramétrica

La densidad de las observaciones futuras,  $x_{m+1}, \dots, x_n$ , condicional en una muestra aleatoria observada,  $x_1, \dots, x_m$ , es  $p(x_{m+1}, \dots, x_n | x_1, \dots, x_m)$ . La siguiente definición describe una forma posible de reducir los datos e incorporarlos en la estructura de esta densidad condicional.

**Definición 1.5** (Suficiencia predictiva). Dada una sucesión de cantidades aleatorias  $x_1, x_2, \dots$ , con medida de probabilidad  $P$ , donde  $x_i$  toma valores en  $\mathcal{X}_i$ ,  $i = 1, 2, \dots$ , la sucesión de estadísticas  $\mathbf{t}_1, \mathbf{t}_2, \dots$ , con  $\mathbf{t}_j$  definida sobre  $\mathcal{X}_1 \times \dots \times \mathcal{X}_j$ , se dice que es suficiente predictiva

para la sucesión  $x_1, x_2, \dots$  si, para toda  $m \geq 1$ ,  $r \geq 1$  y  $\{i_1, \dots, i_r\} \cap \{1, \dots, m\} = \emptyset$ , se tiene que

$$p(x_{i_1}, \dots, x_{i_r} | x_1, \dots, x_m) = p(x_{i_1}, \dots, x_{i_r} | \mathbf{t}_m)$$

donde  $p(\cdot | \cdot)$  es la densidad condicional inducida por  $P$ .

La definición captura la idea de que, dada  $\mathbf{t}_m = \mathbf{t}_m(x_1, \dots, x_m)$ , los valores individuales de  $x_1, \dots, x_m$  no contribuyen con más información para las probabilidades de los eventos futuros. Las observaciones futuras  $(x_{i_1}, \dots, x_{i_r})$  y los valores observados  $(x_1, \dots, x_m)$  son condicionalmente independientes dada  $\mathbf{t}_m$ .

Suponiendo que las variables son intercambiables, otra forma de definir la estadística  $\mathbf{t}_m = \mathbf{t}_m(x_1, \dots, x_m)$  como suficiente es la siguiente.

**Definición 1.6** (Suficiencia paramétrica). Si  $x_1, x_2, \dots$  es una sucesión infinita intercambiable de cantidades aleatorias, donde  $x_i$  toma valores en  $\mathcal{X}_i$ ,  $i = 1, 2, \dots$ , la sucesión de estadísticas  $\mathbf{t}_1, \mathbf{t}_2, \dots$ , con  $\mathbf{t}_j$  definida en  $\mathcal{X}_1 \times \dots \times \mathcal{X}_j$ , es suficiente paramétrica para  $x_1, x_2, \dots$  si, para cualquier  $n \geq 1$ ,

$$dQ(\theta | x_1, \dots, x_n) = dQ(\theta | \mathbf{t}_n),$$

para cualquier  $dQ(\theta)$  que defina un modelo de probabilidad predictiva intercambiable vía la representación

$$p(x_1, \dots, x_n) = \int \prod_{i=1}^n p(x_i | \theta) dQ(\theta).$$

A partir de estas definiciones se pueden establecer las siguientes proposiciones.

**Proposición 1.1** (Equivalencia de suficiencias predictiva y paramétrica). Dada una sucesión infinita intercambiable de cantidades aleatorias  $x_1, x_2, \dots$ , donde  $x_i$  toma valores en  $\mathcal{X}_i$ ,  $i = 1, 2, \dots$ , la sucesión de estadísticas  $\mathbf{t}_1, \mathbf{t}_2, \dots$  con  $\mathbf{t}_j$  definida en  $\mathcal{X}_1 \times \dots \times \mathcal{X}_j$  es suficiente predictiva si, y sólo si, es suficiente paramétrica.

**Proposición 1.2** (Criterio de factorización de Neyman-Fisher). La sucesión de estadísticas  $\mathbf{t}_1, \mathbf{t}_2, \dots$  es suficiente paramétrica para  $x_1, x_2, \dots$  (intercambiable infinitamente con representación paramétrica infinita) si y sólo si, para cualquier  $m \geq 1$ , la densidad conjunta para  $x_1, \dots, x_m$  dada  $\theta$  tiene la forma

$$p(x_1, \dots, x_m | \theta) = h_m(\mathbf{t}_m, \theta) g(x_1, \dots, x_m),$$

para algunas funciones  $h_m \geq 0$ ,  $g > 0$ .

**Proposición 1.3** (Suficiencia e independencia condicional). La sucesión  $\mathbf{t}_1, \mathbf{t}_2, \dots$  es suficiente paramétrica para  $x_1, x_2, \dots$  (infinitamente intercambiable) si, y sólo si, para cualquier  $m \geq 1$ , la densidad  $p(x_1, \dots, x_m | \theta, \mathbf{t}_m)$  es independiente de  $\theta$ .

De la definición general de estadística suficiente (paramétrica o predictiva),  $\mathbf{t}_n(x_1, \dots, x_n) = (x_1, \dots, x_n)$  siempre es una estadística suficiente. Dado que podemos encontrar varias estadísticas suficientes para el mismo modelo paramétrico, cabe preguntarse si existe una estadística que sea suficiente y además resuma la información de manera óptima.

**Definición 1.7** (Estadística suficiente minimal). Si  $x_1, x_2, \dots$ , es una sucesión intercambiabile infinita de cantidades aleatorias, donde  $x_i$  toma valores en  $\mathcal{X}_i$ , la sucesión de estadísticas  $\mathbf{t}_1, \mathbf{t}_2, \dots$ , con  $\mathbf{t}_j$  definida en  $\mathcal{X}_1 \times \dots \times \mathcal{X}_j$ , es suficiente minimal para  $x_1, x_2, \dots$  si dada cualquier otra sucesión de estadísticas suficientes,  $\mathbf{s}_1, \mathbf{s}_2, \dots$ , existen funciones  $g_1(\cdot), g_2(\cdot), \dots$  tales que  $\mathbf{t}_i = g_i(\mathbf{s}_i)$ ,  $i = 1, 2, \dots$ .

Intuitivamente hablando, una vez que se conoce el valor de  $t$ , no hay otras cantidades calculadas que den nueva información acerca de  $\theta$  si el modelo muestral  $p(x) = p(x|\theta)$  es verdadero. El vector completo  $\mathbf{x}$  siempre es suficiente para  $\theta$ . Una estadística suficiente minimal siempre posee la dimensión más pequeña posible, entre las diferentes estadísticas suficientes.

**Ejemplo 1.6** (Suficiencia: Bernoulli). Si  $x_1, x_2, \dots$  es una sucesión infinitamente intercambiabile con valores en 0 y 1, tenemos la representación general

$$\begin{aligned} p(x_1, \dots, x_n) &= \int_0^1 p(x_1, \dots, x_n|\theta) dQ(\theta) \\ &= \int_0^1 \prod_{i=1}^n Br(x_i|\theta) dQ(\theta) \\ &= \int_0^1 \theta^{s_n} (1 - \theta)^{n-s_n} dQ(\theta), \end{aligned}$$

donde  $s_n = x_1 + \dots + x_n$ . Notemos que podemos escribir

$$p(x_1, \dots, x_n|\theta) = h_n(s_n, \theta) g(x_1, \dots, x_n),$$

con

$$h_n(s_n, \theta) = \theta^{s_n} (1 - \theta)^{n-s_n}, \quad g(x_1, \dots, x_n) = 1,$$

de esta manera, por el criterio de factorización de Neyman-Fisher, la sucesión  $s_1, s_2, \dots$  es suficiente predictiva y paramétrica para  $x_1, x_2, \dots$  y además es una estadística suficiente minimal. ■

**Definición 1.8** (Estadística ancilar). Una estadística  $a(\mathbf{x})$  se dice **ancilar** con respecto a  $\theta$  en un modelo paramétrico  $p(\mathbf{x}|\theta)$ , si  $p(a(\mathbf{x})|\theta) = p(a(\mathbf{x}))$  para todos los valores de  $\theta$ .

**Ejemplo 1.7** (Suficiencia y ancilaridad). Sea  $(X_1, X_2, X_3)$  un vector aleatorio con distribución multinomial con tres categorías, Multinomial( $x_1, x_2, x_3|\theta_1, \theta_2, \theta_3, n$ ), con función de masa de probabilidad

$$p(x_1, x_2, x_3|\theta_1, \theta_2, \theta_3, n) = \frac{n!}{x_1! x_2! x_3!} \theta_1^{x_1} \theta_2^{x_2} \theta_3^{x_3}$$

donde  $0 \leq x_i \leq n$ ,  $0 < \theta_i < 1$ , para  $i = 1, 2, 3$ , con  $n = \sum_{i=1}^3 x_i$  y  $1 = \sum_{i=1}^3 \theta_i$ .

Dado que podemos reescribir una variable y un parámetro en términos de los otros por ser linealmente dependientes, sean  $x_3 = n - x_1 - x_2$  y  $\theta_3 = 1 - \theta_1 - \theta_2$ . Entonces la función de masa de probabilidad es

$$p(x_1, x_2 | \theta_1, \theta_2, n) = \frac{n!}{x_1! x_2! (n - x_1 - x_2)!} \theta_1^{x_1} \theta_2^{x_2} (1 - \theta_1 - \theta_2)^{n - x_1 - x_2}.$$

Sea  $\varphi_2 = \frac{\theta_2}{1 - \theta_1}$ . Entonces

$$\begin{aligned} p(x_1, x_2 | \theta_1, \varphi_2, n) &= \binom{n}{x_1} \theta_1^{x_1} (1 - \theta_1)^{n - x_1} \binom{n - x_1}{x_2} \varphi_2^{x_2} (1 - \varphi_2)^{(n - x_1) - x_2} \\ &= \text{Bin}(x_1 | \theta_1, n) \text{Bin}(x_2 | \varphi_2, n - x_1). \end{aligned}$$

Por lo tanto  $x_1$  es una estadística suficiente para  $\theta_1$  y una estadística ancilar para  $\theta_2$ . En otras palabras,  $x_1$  es totalmente informativa acerca de  $\theta_1$ , mientras que  $x_2$  no contiene información sobre  $\theta_1$ .

Por otro lado,  $x_1$  no es directamente informativa sobre  $\varphi_2$ , simplemente afecta la distribución condicional de  $x_2$  dada  $x_1$  (que sólo depende de  $\varphi_2$ ). Por lo tanto, las inferencias sobre  $\varphi_2$  sólo se basan en dicha distribución condicional. Las inferencias hacen uso del valor observado de  $x_1$ , pero ignoran el mecanismo aleatorio que generó ese valor. Se obtendrían las mismas inferencias sobre  $\varphi_2$  si hubiéramos fijado el valor de  $x_1$  desde el principio. ■

**Teorema 1.3** (Equivalencia entre suficiencia clásica y suficiencia Bayesiana). La suficiencia clásica es equivalente a la suficiencia Bayesiana.

*Demostración.* ( $\Leftarrow$ )

$$\begin{aligned} f(\underline{x} | \theta, t(\underline{x})) &= \frac{f(\underline{x}, \theta, t(\underline{x}))}{f(\theta, t(\underline{x}))} = \frac{f(\underline{x}, \theta)}{f(\theta, t(\underline{x}))} = \frac{f(\theta | \underline{x}) f(\underline{x})}{f(\theta | t(\underline{x})) f(t(\underline{x}))} \\ &= \frac{f(\underline{x})}{f(t(\underline{x}))} = \frac{f(\underline{x}, t(\underline{x}))}{f(t(\underline{x}))} = f(\underline{x} | t(\underline{x})) \end{aligned}$$

porque por definición de estadística suficiente Bayesiana se tiene que  $f(\theta | \underline{x}) = f(\theta | t(\underline{x}))$ .

( $\Rightarrow$ )

$$\begin{aligned} f(\theta | \underline{x}) &= \frac{f(\underline{x} | \theta) f(\theta)}{f(\underline{x})} = \frac{f(\underline{x}, t(\underline{x}) | \theta) f(\theta)}{f(\underline{x}, t(\underline{x}))} = \frac{f(\underline{x} | t(\underline{x}), \theta) f(t(\underline{x}) | \theta) f(\theta)}{f(\underline{x} | t(\underline{x})) f(t(\underline{x}))} \\ &= \frac{f(\underline{x} | t(\underline{x})) f(t(\underline{x}) | \theta) f(\theta)}{f(\underline{x} | t(\underline{x})) f(t(\underline{x}))} = \frac{f(t(\underline{x}) | \theta) f(\theta)}{f(t(\underline{x}))} = f(\theta | t(\underline{x})) \end{aligned}$$

porque por definición de estadística suficiente clásica se tiene que  $f(\underline{x} | t(\underline{x}), \theta) = f(\underline{x} | t(\underline{x}))$  □

**Definición 1.9** (Suficiencia clásica). Sea  $X_1, \dots, X_n$  m.m con fdp  $f(x|\theta)$ . Una estadística  $t(\underline{X})$  es suficiente si la distribución de  $\underline{X}$  condicional a  $T(\underline{X})$  No depende de  $\theta$ , es decir,

$$f(\underline{X}|\theta, T(\underline{X})) = f(\underline{X}|T(\underline{X}))$$

## 1.4.2 Familias Exponenciales

En la sección anterior se estudiaron las definiciones de estadística predictiva y paramétrica en forma general; en esta sección estudiaremos estadísticas suficientes de dimensiones fijas. Para más detalles ver Bernardo y Smith (1994).

Consideremos la propiedad de intercambiabilidad y la representación con respecto a  $dQ(\theta)$  en una forma paramétrica específica

$$p(x_1, \dots, x_n|\theta) = \prod_{i=1}^n p(x_i|\theta), \quad (x_1, \dots, x_n) \in \mathcal{X}^n \subseteq \mathbb{R}^n$$

donde  $\theta$  es un parámetro de dimensión  $k$ . Si  $p(x|\theta)$  es tal que  $p(x_1, \dots, x_n|\theta)$  se factoriza en  $h_n(\mathbf{t}_n, \theta)g(x_1, \dots, x_n)$ , para algunas funciones  $h_n$  y  $g$ , la estadística  $\mathbf{t}_n = \mathbf{t}_n(x_1, \dots, x_n)$  sería suficiente. Una clase importante de funciones  $p(x|\theta)$  está identificada por la siguiente definición.

**Definición 1.10** (Familia exponencial con  $k$  parámetros). Una densidad de probabilidad (o función de masa)  $p(x|\theta)$ ,  $x \in \mathcal{X}$ , con parámetro  $\theta \in \Theta \subseteq \mathbb{R}^k$ , pertenece a la familia exponencial con  $k$  parámetros si es de la forma

$$p(x|\theta) = f(x)g(\theta) \exp \left\{ \sum_{i=1}^k c_i \phi_i(\theta) h_i(x) \right\},$$

donde  $h = (h_1, \dots, h_k)$ ,  $\phi(\theta) = (\phi_1, \dots, \phi_k)$  y, dadas las funciones  $f$ ,  $h$ ,  $\phi$ , y las constantes  $c_i$ ,

$$\frac{1}{g(\theta)} = \int_{\mathcal{X}} f(x) \exp \left\{ \sum_{i=1}^k c_i \phi_i(\theta) h_i(x) \right\} dx < \infty.$$

**Proposición 1.4** (Estadísticas suficientes para familias exponenciales con  $k$  parámetros). Si  $x_1, \dots, x_i \in \mathcal{X}$ , es una sucesión intercambiable tal que, dada una familia exponencial regular con  $k$  parámetros,

$$p(x_1, \dots, x_n) = \int_{\Theta} \prod_{i=1}^n f(x)g(\theta) \exp \left\{ \sum_{i=1}^k c_i \phi_i(\theta) h_i(x) \right\} dQ(\theta),$$

para alguna  $dQ(\theta)$ , entonces

$$\mathbf{t}_n = \mathbf{t}_n(x_1, \dots, x_n) = \left[ \sum_{i=1}^n h_1(x_i), \dots, \sum_{i=1}^n h_k(x_i) \right] \quad n = 1, 2, \dots,$$

es una sucesión de estadísticas suficientes.

*Demostración.* Note que

$$\begin{aligned} \prod_{i=1}^n f(x_i) g(\theta) \exp \left\{ \sum_{j=1}^k c_j \phi_j(\theta) h_j(x_i) \right\} \\ = \prod_{i=1}^n f(x_i) \cdot [g(\theta)]^n \exp \left\{ \sum_{j=1}^k c_j \phi_j(\theta) \sum_{i=1}^n h_j(x_i) \right\}. \end{aligned}$$

El resultado se obtiene aplicando el criterio de factorización de Neyman-Fisher.  $\square$

En ocasiones es más conveniente expresar a la familia exponencial en forma canónica, dada por la siguiente definición.

**Definición 1.11** (Forma Canónica). La densidad de probabilidad (o función de masa)

$$p(\mathbf{y}|\psi) = a(\mathbf{y}) \exp \{ \mathbf{y}'\psi - b(\psi) \}, \quad \mathbf{y} \in \mathcal{Y},$$

obtenida de la familia exponencial con  $k$  parámetros, vía las transformaciones

$$\begin{aligned} \mathbf{y} &= (y_1, \dots, y_k), & \psi &= (\psi_1, \dots, \psi_k), \\ y_i &= h_i(x), & \psi_i &= c_i \phi_i(\theta), \quad i = 1, \dots, k, \end{aligned}$$

se llama la **forma canónica** de la familia exponencial.

**Proposición 1.5** (Primeros dos momentos de la familia exponencial). Para  $\mathbf{y}$  con densidad de probabilidad (o función de masa) perteneciente a la familia exponencial canónica, se tiene que

$$E(\mathbf{y}|\psi) = \nabla b(\psi), \quad V(\mathbf{y}|\psi) = \nabla^2 b(\psi).$$

*Demostración.* Es fácil verificar que la función característica de  $\mathbf{y}$  condicional a  $\psi$  está dada por

$$E(\exp\{i\mathbf{u}'\mathbf{y}\}|\psi) = \exp\{b(i\mathbf{u} + \psi) - b(\psi)\},$$

de la cual se obtienen los momentos directamente.  $\square$

**Proposición 1.6** (Suficiencia en la familia exponencial). Si las cantidades aleatorias  $\mathbf{y}_1, \dots, \mathbf{y}_n$  son independientes con densidad de probabilidad (o función de masa) perteneciente a la familia exponencial, entonces

$$\mathbf{s} = \sum_{i=1}^n \mathbf{y}_i$$

es una estadística suficiente y tiene una distribución

$$p(\mathbf{s}|\psi) = a^{(n)}(\mathbf{s}) \exp\{\mathbf{s}'\psi - nb(\psi)\}$$

donde  $a^{(n)}$  es una convolución de orden  $n$  de  $a$ .

*Demostración.* La suficiencia es inmediata ya que las variables pertenecen a la familia exponencial. La función característica de  $\mathbf{s}$  es

$$\exp\{nb(i\mathbf{u} + \psi) - nb(\psi)\}$$

de tal manera que la distribución de  $\mathbf{s}$  es como se pretendía, donde  $a^{(n)}$  satisface

$$nb(\psi) = \log \int a^{(n)}(\mathbf{s}) \exp\{\psi' \mathbf{s}\} d\mathbf{s}.$$

Examinando la convolución para  $n = 1$ , y por inducción, se establece la forma de  $a^{(n)}$ .  $\square$

**Ejemplo 1.8** (Familia exponencial: Uniforme). Sea  $x$  variable aleatoria con distribución  $U(x|0, \theta)$ , con densidad de probabilidad

$$p(x|\theta) = U(x|0, \theta) = \theta^{-1}, \quad x \in (0, \theta), \quad \theta \in \mathfrak{R}^+.$$

Esta distribución pertenece a la familia exponencial no regular en donde

$$f(x) = 1, \quad g(\theta) = \theta^{-1}, \quad h(x) = 0, \quad \phi(\theta) = \theta, \quad c = 1.$$

Sea  $x_1, \dots, x_n$  una muestra de variables aleatorias independientes con distribución  $U(x|0, \theta)$ ,

$$p(x_1, \dots, x_n|\theta) = \prod_{i=1}^n p(x_i|\theta) = \theta^{-n} I_{(0, \theta)}(\max_{i=1, \dots, n} \{x_i\}), \quad (x_1, \dots, x_n) \in \mathfrak{R}^n.$$

Por el criterio de factorización de Neyman-Fisher

$$t_n = t_n(x_1, \dots, x_n) = \max_{i=1, \dots, n} \{x_i\}, \quad n = 1, 2, \dots,$$

es una sucesión de estadísticas suficientes en este caso.  $\blacksquare$