

Estadística

Contents

1	Estadística Descriptiva	2
1.1	Introducción	2
2	Muestreo	3
2.1	Modelos estadísticos	3
2.2	Estadísticos muestrales	6
2.3	Momentos	6
2.4	Resultados de convergencias	8
2.5	Estadísticos ordenados	14
2.6	Ejercicios	16
3	Reducción de Datos	20

1 Estadística Descriptiva

1.1 Introducción

El Cálculo de Probabilidades proporciona una Teoría Matemática que permite analizar las propiedades de los Experimentos Aleatorios

”La velocidad del movimiento caótico de las moléculas de un gas sigue una distribución normal de parámetros...”

”La vida de un determinado tipo de componente eléctrica tiene distribución exponencial de media...”

Construir un Espacio Probabilístico que sirva de Modelo Estadístico asociado a una determinada Variable Aleatoria real para la Deducción de Consecuencias

Para tratar de averiguar si una moneda está trucada no hay mejor procedimiento que lanzarla un buen número de veces y verificar si estadísticamente los resultados obtenidos confirman o invalidan la hipótesis $p = 0.5$, siendo p la probabilidad de cara. Desde el Cálculo de Probabilidades sólo se podrá actuar en función del parámetro p sin alcanzar soluciones numéricas

Disponer de un Conjunto de Observaciones del fenómeno considerado (en lugar de un espacio probabilístico totalmente especificado) hace abandonar los dominios del Cálculo de Probabilidades para introducirse en el terreno de la Estadística Matemática o Inferencia Estadística, cuya finalidad es obtener información sobre la Ley de Probabilidad de dicho fenómeno a partir del Análisis e Interpretación de las observaciones recolectadas

Estadística Descriptiva o Análisis de Datos: Recolección de la Información y su Tratamiento Numérico

Métodos Estadísticos e Inferencia Estadística: Conjunto de Técnicas que utilizan la Información para construir Modelos Matemáticos en situaciones prácticas de incertidumbre y Análisis e Interpretación de las observaciones como método para obtener conclusiones sobre la Ley de Probabilidad del fenómeno en estudio

Inferencia Frecuentista e Inferencia Bayesiana

Modelos Estadísticos Paramétricos y No Paramétricos

Estimación Puntual: Pronóstico de un determinado parámetro de la distribución mediante un único valor numérico

Estimación por Intervalo: Intervalo numérico de valores en el que se pueda afirmar razonablemente que varía el parámetro en cuestión

Contraste de Hipótesis: Corroborar o Invalidar una determinada afirmación acerca de la distribución del fenómeno estudiado.

Concepto de Población y Muestra Aleatoria

2 Muestreo

2.1 Modelos estadísticos

Definición 2.1.1 [Modelo Estadístico]

Sea $(\omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad asociado a un experimento aleatorio \mathcal{E} , una variable aleatoria observable $X : \omega \rightarrow \mathbb{R}$ y su espacio medible asociado (χ, \mathcal{B}) .

Un **modelo estadístico** es una terna (χ, \mathcal{B}, F) , donde:

- χ es el espacio donde la variable aleatoria X toma valores.
- \mathcal{B} es la σ -álgebra asociada a χ . Generalmente se toma que $\mathcal{B} = \mathcal{B}(\mathbb{R})$.
- F es el conjunto de todas las posibles funciones de distribución que podemos considerar sobre \mathcal{B} .

Definición 2.1.2 [Modelo Estadístico Paramétrico]

Un **modelo estadístico paramétrico** es, al igual que el anterior, una terna (χ, \mathcal{B}, F) , pero en este caso F depende de un parámetro θ desconocido.

Se define $F = \{F_\theta : \theta \in \Theta\}$, donde Θ es el espacio paramétrico, es decir, el conjunto de todos los posibles valores que puede tomar θ para que F_θ sea una función de distribución. De forma general se tiene que $\theta \in \Theta \subseteq \mathbb{R}^k$.

Además se tienen dos enfoques según cómo se tome el comportamiento de θ :

- **Enfoque frecuentista:** θ es un valor fijo pero desconocido.
- **Enfoque bayesiano:** θ es una variable aleatoria.

Ejemplo

1. Sea X una variable aleatoria con distribución $N(\theta, 1)$, donde θ es un parámetro desconocido. Entonces, el modelo estadístico asociado a este experimento es $(\mathbb{R}, \mathcal{B}, F_\theta, \theta)$, donde $F_\theta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}$. En este caso $\Theta = \mathbb{R}$.
2. Sea X -variable aleatoria con distribución *Bernouilli*, de parámetro θ . Entonces en este caso, el modelo estadístico asociado es $(\{0, 1\}, \mathcal{B}, F_\theta, \theta)$, donde $F_\theta(x) = \theta^x(1 - \theta)^{1-x}$. En este caso $\Theta = [0, 1]$.

Definición 2.1.3 [Muestra aleatoria simple]

Sea (X_1, \dots, X_n) un conjunto de variables aleatorias independientes e idénticamente distribuidas (i.i.d.) entonces a dicho conjunto se le conoce como **muestra aleatoria simple** de tamaño n .

Por tanto tenemos que el modelo estadístico asociado es una terna (χ, \mathcal{B}, F) , donde:

- $\chi = \mathbb{R}^n$.
- $\mathcal{B} = \mathcal{B}(\mathbb{R}^n)$.
- $F = \{F_\theta^n : \theta \in \Theta\}$, donde Θ es el espacio paramétrico, donde:

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

Definición 2.1.4 [Función de Distribución Empírica]

Sea (X_1, \dots, X_n) m.a.s. $(n) \sim X$ y denotemos por $(X_{(1)}, \dots, X_{(n)})$ a la muestra ordenada de menor a mayor. $\forall x \in \mathbb{R}$ fijo definimos la función distribución empírica como la variable aleatoria

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i)$$

Observemos que

$$F_n(x) = \begin{cases} 0 & \text{si } x < X_{(1)} \\ \frac{k}{n} & \text{si } X_{(k)} < x < X_{(k+1)}, k = 1, \dots, n-1 \\ 1 & \text{si } x \geq X_{(n)} \end{cases}$$

Dada una realización particular de la muestra (x_1, \dots, x_n) , $F_n(x)$ es una función de distribución asociada a una variable aleatoria discreta que toma valores $(x_{(1)}, \dots, x_{(n)})$ con función de masa $(\frac{1}{n}, \dots, \frac{1}{n})$

Ejemplo

Supongamos que tenemos una muestra aleatoria ordenada de tamaño $n = 5$:

$$x_{(1)} = 2, x_{(2)} = 3, x_{(3)} = 5, x_{(4)} = 7, x_{(5)} = 9.$$

Entonces, por como se define la FDE $F_n(x) = \frac{1}{5} \sum_{i=1}^5 I_{(-\infty, x]}(X_i)$, tenemos que:

$$F_n(x) = \begin{cases} 0 & \text{si } x < 2 \\ \frac{1}{5} & \text{si } 2 \leq x < 3 \\ \frac{2}{5} & \text{si } 3 \leq x < 5 \\ \frac{3}{5} & \text{si } 5 \leq x < 7 \\ \frac{4}{5} & \text{si } 7 \leq x < 9 \\ 1 & \text{si } x \geq 9 \end{cases}$$

De manera que cada vez que x alcanza un valor de la muestra, la función de distribución empírica aumenta en $\frac{1}{n} = 0,2$.

Proposición 2.1.1 [Propiedades de la Función de Distribución Empírica]

Sea una muestra aleatoria X_1, X_2, \dots, X_n de una variable aleatoria X con función de distribución $F(x)$. Definimos la función de distribución empírica como:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \chi_{(-\infty, x]}(X_i)$$

donde $\chi_{(-\infty, x]}(X_i)$ es la función indicadora que toma el valor 1 si $X_i \leq x$ y 0 en caso contrario.

1. **Interpretación probabilística:** La función indicadora $\chi_{(-\infty, x]}(X_i)$ sigue una distribución Bernoulli con parámetro $F(x)$, es decir:

$$\chi_{(-\infty, x]}(X_i) \sim \text{Bernoulli}(F(x))$$

Por tanto también podemos afirmar que:

$$\chi_{(-\infty, x]}(X_i) \sim \text{Bin}(1, F(x))$$

Además, la suma de estas variables sigue una distribución binomial:

$$\sum_{i=1}^n \chi_{(-\infty, x]}(X_i) \sim \text{Bin}(n, F(x))$$

2. **Esperanza y varianza:** Para un valor fijo de x , se cumple que:

$$E[F_n(x)] = F(x)$$

lo que indica que $F_n(x)$ es un estimador insesgado de $F(x)$. La varianza está dada por:

$$V[F_n(x)] = \frac{F(x)(1 - F(x))}{n} \xrightarrow{n \rightarrow \infty} 0$$

3. **Convergencia:**

(a) **Convergencia casi segura:**

$$F_n(x) \xrightarrow[n \rightarrow \infty]{c.s.} F(x)$$

(b) **Convergencia en distribución:** Se cumple la normalidad asintótica:

$$\frac{F_n(x) - F(x)}{\sqrt{\frac{F(x)(1-F(x))}{n}}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1)$$

4. **Intervalo de confianza para $F(x)$:** Dada una realización particular de la muestra (x_1, \dots, x_n) , se puede construir un intervalo de confianza asintótico para $F(x)$ de nivel $1 - \alpha$:

$$IC_{1-\alpha}(F(x)) = \left(F_n(x) - \frac{z_{\alpha/2}}{2\sqrt{n}}, F_n(x) + \frac{z_{\alpha/2}}{2\sqrt{n}} \right)$$

donde $z_{\alpha/2}$ es el cuantil de la distribución normal estándar.

Teorema 2.1.1 [de Glivenko-Cantelli]

Sea (X_1, \dots, X_n) m.a.s.(n) $\sim X$ con función de distribución empírica $F_n(x)$ y sea $F(x)$ la función de distribución de X , es decir, de la población total. Entonces se cumple que:

$$\lim_{n \rightarrow \infty} P(w : \sup_x |F_n(x) - F(x)| < \epsilon) = 1, \forall \epsilon > 0$$

Corolario 2.1.1

El Teorema de Glivenko-Cantelli permite realizar una técnica estadística denominada **método de sustitución (Plug-In)** la cual se basa en la sustitución de parámetros desconocidos por sus estimaciones sobre una muestra. Por ejemplo:

1. Se puede estimar la media poblacional μ por la media muestral $\bar{X} = \int x dF_n(x) = \frac{1}{n} \sum_{i=1}^n x_i$.

2. Se puede estimar la varianza poblacional σ^2 por la varianza muestral $S^2 = \int (x - \bar{x}) dF_n(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \sigma_n^2$.

2.2 Estadísticos muestrales

Definición 2.2.1 [Estadístico muestral]

Sea (X_1, \dots, X_n) m.a.s.(n) $\sim X$ y sea $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ medible (integrable), bien definida y no dependiente de parámetros desconocidos, se le llama **estadístico muestral**

Ejemplo

Desacamos los siguientes estadísticos muestrales:

1. Media muestral $T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$
2. Cuasivarianza muestral $T(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = S_n^2$
3. $T(X_1, \dots, X_n) = (\bar{X}, S_n^2)$

2.3 Momentos

Sea (X_1, \dots, X_n) m.a.s.(n) de X , $\mu = E[X]$ y $\sigma = \sqrt{V(X)}$.

Definición 2.3.1 [Momento Muestral]

Se define el momento muestral de orden k respecto al origen como

$$a_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

y el momento muestral de orden k respecto a la media como

$$b_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$$

Observación 2.3.1

- El momento muestral de orden 1 respecto al origen es la media muestral ($a_1 = \bar{X}$).
- El momento muestral de orden 2 respecto a la media es la varianza muestral ($b_2 = \sigma_n^2$).

Definición 2.3.2 [Momento Poblacional]

Se define el momento poblacional de orden k respecto al origen como

$$\alpha_k = E[X^k]$$

y el momento poblacional de orden k respecto a la media como

$$\beta_k = E[(X - \mu)^k]$$

Observación 2.3.2

- El momento poblacional de orden 1 respecto al origen es la media poblacional ($\alpha_1 = \mu$).
- El momento poblacional de orden 2 respecto a la media es la varianza poblacional ($\beta_2 = \sigma^2$).

Proposición 2.3.1 [Propiedades asintóticas de los momentos muestrales]

Sea (X_1, \dots, X_n) m.a.s.(n) de $X \sim N(\mu, \sigma)$, entonces se cumple que:

1. Momentos muestrales respecto al origen:

(a)

$$a_k = \frac{1}{n} \sum_{i=1}^n X_i^k \xrightarrow[n \rightarrow \infty]{c.s.} \alpha_k = E[X^k]$$

(b)

$$\bar{X} \xrightarrow[n \rightarrow \infty]{c.s.} \mu \quad (\text{Ley Fuerte de Kintchine, } \mu < \infty)$$

2. Momentos muestrales respecto a la esperanza/media:

(a)

$$b_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k \xrightarrow[n \rightarrow \infty]{c.s.} \beta_k = E[(X - \mu)^k]$$

(b)

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \xrightarrow[n \rightarrow \infty]{c.s.} \sigma^2$$

(c)

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \xrightarrow[n \rightarrow \infty]{c.s.} \sigma^2$$

Propiedades asintóticas de los momentos muestrales

$$\sqrt{n} \frac{a_k - \alpha_k}{\sqrt{a_{2k} - a_k^2}} \xrightarrow[n \rightarrow \infty]{d} Z \sim N(0, 1)$$

$$\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \xrightarrow[n \rightarrow \infty]{d} Z \sim N(0, 1), \quad \sigma = \sqrt{E[X^2] - \mu^2}$$

(Teorema Central del Límite de Levy-Lindeberg, $\mu < \infty, \sigma < \infty$)

$$\sqrt{n} \frac{b_k - \beta_k}{\sqrt{\beta_{2k} - \beta_k^2 - 2k\beta_{k-1}\beta_{k+1} + k^2\beta_{k-1}^2\beta_2}} \xrightarrow[n \rightarrow \infty]{d} Z \sim N(0, 1)$$

$$\sqrt{n} \frac{\sigma_n^2 - \sigma^2}{\sqrt{\beta^4 - \sigma^4}} \xrightarrow[n \rightarrow \infty]{d} Z \sim N(0, 1)$$

2.4 Resultados de convergencias

Teorema 2.4.1 [de Slutsky]

Si $X_n \xrightarrow[n \rightarrow \infty]{d} X$ y $Y_n \xrightarrow[n \rightarrow \infty]{P} a$ entonces

1. $Y_n X_n \xrightarrow[n \rightarrow \infty]{d} aX$
2. $X_n + Y_n \xrightarrow[n \rightarrow \infty]{d} X + a$
3. $\frac{X_n}{Y_n} \xrightarrow[n \rightarrow \infty]{d} \frac{X}{a}$ siempre que $a \neq 0$

Lema 2.4.1

Si $\{a_n\}$ es una sucesión de constantes con $\lim_{n \rightarrow \infty} a_n = +\infty$ y a es un número fijo tal que

$$a_n (X_n - a) \xrightarrow[n \rightarrow \infty]{d} X$$

entonces para cualquier función $g : \mathbb{R} \rightarrow \mathbb{R}$ con derivada continua y no nula en a se tiene que

$$a_n (g(X_n) - g(a)) \xrightarrow[n \rightarrow \infty]{d} g'(a)X$$

Ejemplo

Sea (X_1, \dots, X_n) m.a.s. (n) de $X \sim N(\mu, \sigma)$, se pide calcular la distribución de la media muestral:

Tenemos que $\varphi_X(t) = e^{it\mu - \frac{1}{2}t^2\sigma^2} \implies$

$$E[e^{\bar{X}}] = \varphi_{\bar{X}}(t) = E\left[e^{it\frac{1}{n}\sum_{i=1}^n X_i}\right] = \varphi_{\sum_{i=1}^n X_i}\left(\frac{t}{n}\right) = \left(\varphi_X\left(\frac{t}{n}\right)\right)^n = e^{it\mu - \frac{1}{2}\frac{\sigma^2}{n}t^2}.$$

Por lo tanto $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Ejemplo

Sea X_1, \dots, X_n una muestra aleatoria simple de una variable aleatoria $X \sim N(0, \sigma)$. La función de densidad de X es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}x^2}$$

Calculemos la distribución de $a_2 = \frac{1}{n} \sum_{i=1}^n X_i^2$

Definimos la variable estandarizada:

$$Z_i = \frac{X_i}{\sigma} \sim N(0, 1).$$

Entonces, la suma de los cuadrados sigue una distribución chi-cuadrado:

$$\sum_{i=1}^n Z_i^2 \sim \chi_n^2.$$

La distribución chi-cuadrado con n grados de libertad es un caso particular de la distribución Gamma:

$$\chi_n^2 \sim \text{Gamma} \left(a = \frac{1}{2}, p = \frac{n}{2} \right).$$

Donde: - $a = \frac{1}{2}$ es el **parámetro de forma**. - $p = \frac{n}{2}$ es el **parámetro de escala**.

La función de densidad de la suma de cuadrados es:

$$f_{\sum_{i=1}^n Z_i^2}(y) = \frac{1}{2^{n/2} \Gamma(n/2)} y^{n/2-1} e^{-y/2}.$$

Como $X_i^2 = \sigma^2 Z_i^2$, al tomar la media muestral obtenemos:

$$a_2 = \frac{1}{n} \sum_{i=1}^n X_i^2 = \sigma^2 \frac{1}{n} \sum_{i=1}^n Z_i^2.$$

Sustituyendo la distribución Gamma de la suma de Z_i^2 , se tiene:

$$a_2 \sim \text{Gamma} \left(a = \frac{n}{2\sigma^2}, p = \frac{n}{2} \right).$$

Para muestras grandes, usando el **Teorema Central del Límite**, la variable estandarizada:

$$\sigma^2 \sqrt{\frac{n}{2\sigma^2}} (a_2 - \sigma^2)$$

converge en distribución a una normal estándar:

$$N(0, 1).$$

Este resultado es fundamental en inferencia estadística, ya que muestra que la varianza muestral puede aproximarse por una normal para muestras grandes.

Lema 2.4.2

Si $Y \sim \text{Gamma}(a, p)$, entonces $T = 2aY \sim \text{Gamma}(\frac{1}{2}, p)$.

Demostración. Sabemos que la función de densidad de probabilidad de una variable aleatoria Y que sigue una distribución Gamma con parámetros a y p es:

$$f_Y(y) = \frac{a^p}{\Gamma(p)} e^{-ay} y^{p-1}, \quad y \geq 0.$$

Ahora, definimos la variable T como $T = 2aY$, lo que implica que $Y = \frac{1}{2a}T$. Usamos el cambio de variable para encontrar la función de densidad de probabilidad de T . El jacobiano de este cambio es:

$$J = \left| \frac{dY}{dT} \right| = \frac{1}{2a}.$$

Por lo tanto, la función de densidad de probabilidad de T se obtiene sustituyendo en la fórmula general para el cambio de variable:

$$f_T(t) = f_Y\left(\frac{1}{2a}t\right) \cdot \frac{1}{2a}.$$

Sustituyendo la expresión de $f_Y(y)$, obtenemos:

$$f_T(t) = \frac{a^p}{\Gamma(p)} e^{-a(\frac{1}{2a}t)} \left(\frac{1}{2a}t\right)^{p-1} \cdot \frac{1}{2a}.$$

Simplificando, tenemos:

$$f_T(t) = \frac{\left(\frac{1}{2}\right)^p}{\Gamma(p)} e^{-\frac{1}{2}t} t^{p-1}, \quad t \geq 0.$$

Esta es precisamente la función de densidad de una distribución Gamma con parámetros $(\frac{1}{2}, p)$, lo que demuestra que $T \sim \text{Gamma}(\frac{1}{2}, p)$. \square

Ejemplo

Sea (X_1, \dots, X_n) una m.a.s.(n) de $X \sim N(0, \sigma)$. Entonces, la función de densidad es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

Se pide calcular la distribución de

$$a_2 = \frac{1}{n} \sum_{i=1}^n X_i^2$$

Tipificando las variables aleatorias X_i como:

$$Z_i = \frac{X_i}{\sigma} \sim N(0, 1),$$

se tiene que

$$\sum_{i=1}^n Z_i^2 \sim \chi_n^2 \equiv \text{Gamma}\left(a = \frac{1}{2}, p = \frac{n}{2}\right)$$

La función de densidad de esta suma es:

$$f_{\sum_{i=1}^n Z_i^2}(y) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} e^{-\frac{y}{2}} y^{\frac{n}{2}-1}$$

Dado que

$$a_2 = \frac{1}{n} \sum_{i=1}^n X_i^2 = \frac{\sigma^2}{n} \sum_{i=1}^n Z_i^2,$$

definimos el cambio de variable

$$J = \frac{n}{\sigma^2}.$$

Por lo tanto, la función de densidad de a_2 es:

$$f_{a_2}(t) = f_{\sum_{i=1}^n Z_i^2}\left(\frac{n}{\sigma^2}t\right) \frac{n}{\sigma^2} = \frac{\left(\frac{n}{2\sigma^2}\right)^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} e^{-\frac{n}{2\sigma^2}t} t^{\frac{n}{2}-1}.$$

Por lo tanto,

$$a_2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \sim \text{Gamma}\left(a = \frac{2\sigma^2}{n}, p = \frac{n}{2}\right).$$

Finalmente, bajo el límite

$$\sigma^2 \sqrt{\frac{n}{2}} (a_2 - \sigma^2) \xrightarrow{d} N(0, 1) \quad \text{cuando } n \rightarrow \infty.$$

Teorema 2.4.2 [de Fisher]

Sea (X_1, \dots, X_n) una m.a.s. (n) de $X \sim N(\mu, \sigma)$, con μ y σ desconocidos. Se cumple que:

1. La media muestral y la varianza muestral:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

son variables aleatorias independientes.

2. Sus distribuciones son:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right),$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.$$

3. La siguiente variable aleatoria sigue una distribución t de Student:

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}.$$

Demostración.

1.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \text{ y } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ son independientes } \iff X_i - \bar{X} \text{ y } \bar{X} \text{ son independientes}$$

\implies Veamos la independencia a través de la covarianza: $Cov(X_i - \bar{X}, \bar{X}) = 0$?

$$\begin{aligned} Cov(X_i - \bar{X}, \bar{X}) &= E[(X_i - \bar{X})\bar{X}] - E[X_i - \bar{X}]E[\bar{X}] = E[(X_i - \bar{X})\bar{X}] - (E[X_i] - E[\bar{X}])E[\bar{X}] = \\ &= E[(X_i - \bar{X})\bar{X}] - (\mu - \mu)\mu = E[X_i \cdot \bar{X}] - E[\bar{X}^2] = E[X_i \cdot \bar{X}] - (V[\bar{X}] + E[\bar{X}]^2) = E[X_i \cdot \bar{X}] - \left(\frac{\sigma^2}{n} + \mu^2\right) = \\ &= E\left[\frac{1}{n}X_1 \cdot X_i + \dots + \frac{1}{n}X_i^2 + \dots + \frac{1}{n}X_n \cdot X_i\right] - \left(\frac{\sigma^2}{n} + \mu^2\right) = \\ &= \frac{1}{n}(E[X_1 X_i] + \dots + E[X_n X_i]) + \frac{1}{n}E[X_i^2] - \left(\frac{\sigma^2}{n} + \mu^2\right) = \\ &= \frac{n-1}{n}(\mu^2) + \frac{1}{n}(\sigma^2 + \mu^2) - \left(\frac{\sigma^2}{n} + \mu^2\right) = \\ &= \frac{n-1}{n}\mu^2 + \frac{\mu^2}{n} - \mu^2 + \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0 \implies \\ Cov(X_i - \bar{X}, \bar{X}) &= 0 \implies \text{son independientes} \end{aligned}$$

2. Denotemos por

$$S_{n+1}^2 = \frac{1}{n} \sum_{i=1}^{n+1} (X_i - \bar{X}_{n+1})^2.$$

Demostremos que:

$$nS_{n+1}^2 = (n-1)S_n^2 + (X_{n+1} - \bar{X}_n)^2 \frac{n}{n+1}.$$

Desarrollando la expresión:

$$\begin{aligned} nS_{n+1}^2 &= \sum_{i=1}^{n+1} (X_i - \bar{X}_{n+1})^2 \\ &= \sum_{i=1}^n ((X_i - \bar{X}_n) + (\bar{X}_n - \bar{X}_{n+1}))^2 + (X_{n+1} - \bar{X}_{n+1})^2 \\ &= (n-1)S_n^2 + n(\bar{X}_n - \bar{X}_{n+1})^2 + (X_{n+1} - \bar{X}_{n+1})^2 \\ &\quad + \sum_{i=1}^n 2(X_i - \bar{X}_n)(\bar{X}_n - \bar{X}_{n+1}). \end{aligned}$$

Como $\sum_{i=1}^n (X_i - \bar{X}_n) = 0$, el término cruzado se anula y obtenemos:

$$nS_{n+1}^2 = (n-1)S_n^2 + n(\bar{X}_n - \bar{X}_{n+1})^2 + (X_{n+1} - \bar{X}_{n+1})^2.$$

En este último paso, se desarrollan los cuadrados y se aplica la definición de la media muestral:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \implies n\bar{X}_n = \sum_{i=1}^n X_i.$$

Ahora, utilizando la relación:

$$\bar{X}_{n+1} = \frac{n\bar{X}_n + X_{n+1}}{n+1} \implies X_{n+1} - \bar{X}_{n+1} = \frac{(n+1)X_{n+1} - n\bar{X}_n - X_{n+1}}{n+1},$$

obtenemos:

$$\begin{aligned} nS_{n+1}^2 &= (n-1)S_n^2 + (X_{n+1} - \bar{X}_n)^2 \frac{n}{(n+1)^2} + (X_{n+1} - \bar{X}_n)^2 \frac{n^2}{(n+1)^2} \\ &= (n-1)S_n^2 + (X_{n+1} - \bar{X}_n)^2 \frac{n}{n+1}. \end{aligned}$$

Así, hemos obtenido el resultado deseado.

****Distribución de la razón estandarizada:**** Ahora veamos que:

$$\left(\frac{X_{n+1} - \bar{X}_n}{\sigma \sqrt{\frac{n+1}{n}}} \right)^2 \sim \chi_1^2.$$

Usamos las siguientes propiedades:

$$\begin{aligned}
X_{n+1} &\sim N(\mu, \sigma), \\
\bar{X}_n &\sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right), \\
X_{n+1} - \bar{X}_n &\sim N\left(0, \sigma\sqrt{\frac{n+1}{n}}\right).
\end{aligned}$$

En este desarrollo se ha usado que en las distribuciones normales, se restan las medias y se suman las varianzas. Por lo tanto, la variable estandarizada es:

$$\left(\frac{X_{n+1} - \bar{X}_n}{\sigma\sqrt{\frac{n+1}{n}}}\right) \sim N(0, 1) \sim \chi_1^2.$$

****Prueba por inducción:**** Consideremos el caso base, con $n = 2$:

$$\begin{aligned}
\frac{S_2^2}{\sigma^2} &= \frac{1}{\sigma^2} ((X_1 - \bar{X}_2)^2 + (X_2 - \bar{X}_2)^2) \\
&= \frac{1}{\sigma^2} \left(\left(X_1 - \frac{X_1 + X_2}{2} \right)^2 + \left(X_2 - \frac{X_1 + X_2}{2} \right)^2 \right) \\
&= \frac{1}{\sigma^2} \left(\frac{1}{4}(X_1 - X_2)^2 + \frac{1}{4}(X_2 - X_1)^2 \right) \\
&= \frac{1}{\sigma^2} \frac{1}{2} (X_2 - X_1)^2.
\end{aligned}$$

Como hemos demostrado antes,

$$\left(\frac{X_2 - X_1}{\sigma\sqrt{2}}\right)^2 \sim \chi_1^2.$$

****Hipótesis de inducción:**** Para $n - 1$ supongamos que

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Aplicando la fórmula demostrada:

$$\frac{nS_{n+1}^2}{\sigma^2} = \frac{(n-1)S_n^2}{\sigma^2} + \left(\frac{X_{n+1} - \bar{X}_n}{\sigma\sqrt{\frac{n+1}{n}}}\right)^2.$$

Como los dos términos a la derecha son independientes, y

$$\frac{X_{n+1} - \bar{X}_n}{\sigma\sqrt{\frac{n+1}{n}}} \sim \chi_1^2,$$

obtenemos la suma de dos variables chi-cuadrado:

$$\chi_{n-1}^2 + \chi_1^2 \sim \chi_n^2.$$

****Conclusión:**** Por inducción, se ha demostrado que

$$\frac{nS_{n+1}^2}{\sigma^2} \sim \chi_n^2.$$

3. Se deduce de los anteriores

□

2.5 Estadísticos ordenados

Definición 2.5.1 [Estadísticos ordenados]

Sea (X_1, \dots, X_n) una m.a.s. (n) de X . Podemos ordenar los valores de menor a mayor. A éstos se les llama estadísticos ordenados y se denotan por $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. Sus funciones de distribución y densidad son:

$$F_{X_{(n)}}(x) = F(x)^n \implies f_{X_{(n)}}(x) = \frac{\partial}{\partial x}(F(x)^n) = nF(x)^{n-1}f(x)$$

$$F_{X_{(1)}}(x) = 1 - (1 - F(x))^n \implies f_{X_{(1)}}(x) = \frac{\partial}{\partial x}(1 - (1 - F(x))^n) = n(1 - F(x))^{n-1}f(x)$$

$$F_{X_{(r)}}(x) = P\left(\sum_{i=1}^n I_{(-\infty, x]}(X_i) \geq r\right) = P(\text{Bin}(n, F(x)) \geq r) = \sum_{j=r}^n \binom{n}{j} F(x)^j (1 - F(x))^{n-j} \implies$$

$$f_{X_{(r)}}(x) = \binom{n}{r} r F(x)^{r-1} (1 - F(x))^{n-r} f(x)$$

$$f_{(X_{(r)}, X_{(s)})}(x, y) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} F(x)^{r-1} (F(y) - F(x))^{s-r-1} (1 - F(y))^{n-s} f(x) f(y)$$

$$f_{(X_{(1)}, \dots, X_{(n)})}(y_1, \dots, y_n) = n! \prod_{j=1}^n f(y_j) \text{ si } y_1 < \dots < y_n$$

Ejemplo

Sea X_1, X_2, \dots, X_n una muestra aleatoria simple de una distribución uniforme en $(0, 1)$, es decir:

$$X_1, X_2, \dots, X_n \stackrel{i.i.d.}{\sim} U(0, 1)$$

La función de distribución acumulada de una variable uniforme en $(0, 1)$ es:

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

Ordenando la muestra de menor a mayor, los estadísticos ordenados se denotan como $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. Se quiere encontrar la función de densidad de estos valores ordenados.

Para el máximo, $X_{(n)}$, se tiene que su función de distribución es:

$$P(X_{(n)} \leq x) = P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x)$$

Dado que los datos son independientes, esto se factoriza como:

$$P(X_{(n)} \leq x) = F(x)^n = x^n, \quad 0 < x < 1$$

Derivando se obtiene la función de densidad:

$$f_{X_{(n)}}(x) = \frac{d}{dx}x^n = nx^{n-1}, \quad 0 < x < 1$$

Es decir, $X_{(n)} \sim \text{Beta}(n, 1)$.

Para el mínimo, $X_{(1)}$, la función de distribución se obtiene como:

$$P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x) = 1 - P(X_1 > x, X_2 > x, \dots, X_n > x)$$

Por independencia,

$$P(X_1 > x, X_2 > x, \dots, X_n > x) = (1 - F(x))^n = (1 - x)^n$$

Por lo que,

$$P(X_{(1)} \leq x) = 1 - (1 - x)^n, \quad 0 < x < 1$$

Derivando se obtiene la densidad:

$$f_{X_{(1)}}(x) = n(1 - x)^{n-1}, \quad 0 < x < 1$$

De manera general, para el r -ésimo estadístico ordenado, su densidad es:

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!} x^{r-1} (1-x)^{n-r}, \quad 0 < x < 1$$

lo que implica que $X_{(r)} \sim \text{Beta}(r, n - r + 1)$.

La densidad conjunta del mínimo y el máximo de la muestra es:

$$f_{X_{(1)}, X_{(n)}}(x, y) = n(n-1)(y-x)^{n-2}, \quad 0 < x < y < 1$$

Esto muestra cómo la distribución de los estadísticos ordenados sigue distribuciones beta en función de la posición del orden estadístico dentro de la muestra.

Con lo obtenido, calculemos la distribución del rango muestral: $R = X_{(n)} - X_{(1)}$:

Sea el cambio de variable

$$\left. \begin{array}{l} R = X_{(n)} - X_{(1)} \\ H = \frac{X_{(n)} + X_{(1)}}{2} \end{array} \right\} \begin{array}{l} X_{(n)} = H + \frac{R}{2} \\ X_{(1)} = H - \frac{R}{2} \end{array}$$

Dado que estamos intentando calcular un cambio de variable aleatoria, necesitamos calcular el jacobiano de la transformación. En este caso, el jacobiano es:

$$J = \left| \begin{array}{cc} \frac{\partial X_{(1)}}{\partial H} & \frac{\partial X_{(1)}}{\partial R} \\ \frac{\partial X_{(n)}}{\partial H} & \frac{\partial X_{(n)}}{\partial R} \end{array} \right| = \left| \begin{array}{cc} 1 & -\frac{1}{2} \\ 1 & \frac{1}{2} \end{array} \right| = 1.$$

Además, dado que $X_{(i)}$ son variables aleatorias uniformes en $(0, 1)$, tenemos que:

$$0 < X_{(1)} = H - \frac{R}{2} < 1 \quad 0 < X_{(n)} = H + \frac{R}{2} < 1 \implies \frac{R}{2} < H < 1 - \frac{R}{2}$$

Entonces por el Teorema de la Transformación de Variables, la densidad conjunta de R y H es:

$$g_{(R,H)}(r, h) = f_{(X_{(1)}, X_{(n)})} \left(h - \frac{r}{2}, h + \frac{r}{2} \right) = 1 - n(n-1)r^{n-2}$$

Para obtener la densidad de R , se integra la densidad conjunta respecto a H :

$$f_R(r) = \int_0^1 g_{(R,H)}(r, h) dh = \int_0^1 1 - n(n-1)r^{n-2} dh = h - n(n-1)r^{n-2}h \Big|_0^1 = 1 - n(n-1)r^{n-2}$$

Entonces, la densidad de R es:

$$f_R(r) = 1 - n(n-1)r^{n-2} \sim \text{Beta}(n-1, 2)$$

Ejemplo

Sea (X_1, \dots, X_n) una muestra aleatoria simple (m.a.s.) de tamaño n de una variable aleatoria X , cuya función de distribución es

$$F(x) = P(X \leq x).$$

Se desea calcular la distribución de las variables

$$U = F(X) \quad \text{y} \quad U_R = F(X_{(r)}).$$

Para la variable U , tenemos:

$$G_U(u) = P(U \leq u) = P(F(X) \leq u) = P(X \leq F^{-1}(u)).$$

Utilizando la propiedad de la función de distribución inversa, se obtiene:

$$F(F^{-1}(u)) = u, \quad 0 < u < 1.$$

Por lo tanto,

$$U = F(X) \sim U(0, 1).$$

Ahora, para la distribución de U_R :

$$G_{U_R}(u) = P(U_R \leq u) = P(F(X_{(r)}) \leq u) = P(X_{(r)} \leq F^{-1}(u)).$$

Esto es equivalente a la función de distribución del estadístico ordenado:

$$F_{X_{(r)}}(F^{-1}(u)) = \sum_{j=r}^n \binom{n}{j} F(F^{-1}(u))^j (1 - F(F^{-1}(u)))^{n-j}.$$

Reescribiendo en términos de u :

$$G_{U_R}(u) = \sum_{j=r}^n \binom{n}{j} u^j (1 - u)^{n-j}.$$

Por lo tanto, el estadístico ordenado de orden r asociado a la muestra aleatoria sigue la distribución

$$U_R = F(X_{(r)}),$$

donde la población original $U = F(X)$ sigue una distribución uniforme en el intervalo $(0, 1)$.

2.6 Ejercicios

Ejercicio 2.1. Sea X una población de *Bernoulli*($p = \frac{1}{2}$) y se consideran todas las m.a.s. posibles de tamaño 3. Para cada muestra calcúlese \bar{X} , s^2 , $\mu y S^2$ y determínese sus distribuciones en el muestreo.

Solución:

Muestras	\bar{X}	s^2	P
(0,0,0)	0	0	1/8
(0,0,1)	1/3	1/3	1/8
(0,1,0)	1/3	1/3	1/8
(0,1,1)	2/3	1/3	1/8
(1,0,0)	1/3	1/3	1/8
(1,0,1)	2/3	1/3	1/8
(1,1,0)	2/3	1/3	1/8
(1,1,1)	1	0	1/8

Distribución de \bar{X} y s^2 :

\bar{x}	0	1/3	2/3	1
P	1/8	3/8	3/8	1/8

s^2	0	1/3
P	1/4	3/4

□

Ejercicio 2.2. De una población con media μ desconocida y varianza 1, se toma una m.a.s. de tamaño n . ¿Cuál debe ser éste para que la media muestral diste en valor absoluto de la media de la población menos que 0,5, con una probabilidad mayor o igual que 0,95?

Solución:

El enunciado nos pide es averiguar la n suficiente para que se cumpla que $P(|\bar{X} - \mu| < 0.5) \geq 0.95 \iff P(-0.5 < \bar{X} - \mu < 0.5) \geq 0.95$.

Sabemos que $\bar{X} \sim N(\mu, \frac{1}{n})$. Por lo que para solucionarlo, haremos uso de la Desigualdad de Chebushev, la cual afirma que:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad \forall k > 0$$

Aplicando la desigualdad a nuestro caso, obtenemos que $k = 0.5$ y $\sigma = \frac{1}{\sqrt{n}}$. Por lo que:

$$P(|\bar{X} - \mu| \leq 0'5) \geq 0'95 \iff P(|\bar{X} - \mu| > 0'5) \leq 0'05 \implies \begin{cases} \frac{1}{k^2} = 0'05 \\ \frac{k}{\sqrt{n}} = 0,5 \end{cases} \implies n \geq 80$$

□

Ejercicio 2.3. Dada una m.a.s. de tamaño n , calcúlese la distribución de la media muestral \bar{X} cuando la población es:

1. Bernouilli
2. Gamma
3. Exponencial

Solución:

1. **Bernouilli:** Si $X \sim Bernouilli(p) \implies \sum_{i=1}^n X_i = S_n \sim Bin(n, p) \implies \bar{X} = \frac{S_n}{n} \implies P(\bar{X} = w) = P(S_n = nw) = \binom{n}{nw} p^{nw} (1-p)^{n(1-w)}$
2. **Gamma:** Si $X \sim Gamma(a, b) \implies \varphi_X(t) = (1 - \frac{it}{a})^{-b} \implies \varphi_{S_n}(t) = E[e^{i \frac{t}{n} \sum_{i=1}^n X_i}] = E[e^{i \frac{t}{n} X_1}] \dots E[e^{i \frac{t}{n} X_n}] = [\varphi(\frac{t}{n})]^n = (1 - \frac{it}{an})^{-nb} \implies \bar{X} \sim Gamma(na, nb)$

3. **Exponencial:** Es un caso particular del apartado anterior dado que $X \sim Exponencial(\theta) = Gamma(\theta, 1) \implies \bar{x} \sim Gamma(n\theta, n)$.

□

Ejercicio 2.4. Dada una sucesión $\{X_n\}$ de variables aleatorias independientes con distribución $N(0, 1)$ y un entero positivo k se define:

$$F_{k,m} = \left(\frac{1}{k} \sum_{i=1}^k X_i^2 \right) / \left(\frac{1}{m} \sum_{i=k+1}^{k+m} X_i^2 \right)$$

Pruébese que

$$F_{k,m} \xrightarrow[m \rightarrow \infty]{\mathcal{L}} \frac{1}{k} X$$

donde $X \sim \chi_k^2$.

Solución: Por las propiedades de las distribuciones normales, tenemos que:

$$X_i \sim N(0, 1) \implies X_i^2 \sim \chi_1^2 \implies \sum_{i=1}^n X_i^2 \sim \chi_n^2$$

De esta manera podemos ver los sumandos como el calculo de la media muestral:

$$\bar{X} = \frac{1}{m} \sum_{i=k+1}^{k+m} \chi_1^2 \implies$$

Por la Ley Fuerte de los Grandes Números tenemos que

$$\bar{X} \xrightarrow[m \rightarrow \infty]{\mathcal{L}} E[\chi_1^2] = 1$$

Finalmente como el segundo sumatorio $\sum_{i=1}^k X_i^2$ se mantiene constante (pues sólo se está modificando k) tenemos que, por el Teorema de Slutsky,

$$F_{k,m} \xrightarrow[m \rightarrow \infty]{\mathcal{L}} \frac{1}{k} \chi_k^2$$

□

Ejercicio 2.5. Demuéstrese que para una m.a.s. de tamaño n tales que

$$P(X = x_i) = p_i \forall i \in \mathbb{N}$$

se cumple que la distribución del estadístico ordenado $X_{(k)}$ es discreta y viene dada por:

$$P(X_{(k)} \leq x_i) = \sum_{j=k}^n \binom{n}{j} (F(x_j))^j (1 - F(x_j))^{n-j}$$

$$P(X_{(k)} = x_i) = \sum_{j=k}^n \binom{n}{j} [(F(x_j))^j (1 - F(x_i))^{n-j} - F(x_{i-1})^j (1 - F(x_{i-1}))^{n-j}]$$

Solución: Sea una m.a.s. (X_1, X_2, \dots, X_n) .

Sea entonces el conjunto de variables aleatorias

$$I_j = \begin{cases} 1, & \text{si } X_j \leq x \\ 0, & \text{si } X_j > x \end{cases} \implies I_j \sim \text{Bernouilli}$$

con una probabilidad $P(I_j = 1) = P(X_j \leq x) = F_j(x)$

Al ser variables aleatorias independientes con distribución Bernouilli, se cumple que

$$\sum_{j=1}^n I_j = S \sim \text{Bin}(n, F(x))$$

Nos piden calcular que $P(X_{(k)} \leq x_i)$, lo cual es equivalente a que nos pregunten la probabilidad de que haya al menos k valores menores o iguales a x_i : $P(S \geq k)$. Por lo que podemos escribir:

$$P(S \geq k) = P(X_{(k)} \leq x_i) = \sum_{j=k}^n \binom{n}{j} (F(x_i))^j (1 - F(x_i))^{n-j}$$

Finalmente, nos queda sólo calcular la probabilidad de que $P(X_{(k)} = x_i)$ pero esto es igual a hacer:

$$P(X_{(k)} = x_i) = P(X_{(k)} \leq x_i) - P(X_{(k)} \leq x_{i-1})$$

□

3 Reducción de Datos

Definición 3.0.1 [Estadístico suficiente]

Sea $(\Omega, \mathcal{A}, \mathcal{P})$ el espacio probabilístico asociado a un experimento aleatorio \mathcal{E} , una variable aleatoria observable $X : \Omega \rightarrow \mathbb{R}$ y su modelo estadístico asociado $(\chi, \mathcal{B}, F_\theta)_{\theta \in \Theta \subset \mathbb{R}^\ell}$, $\mathcal{B} = \mathcal{B}(\mathbb{R})$, y (X_1, \dots, X_n) m.a.s. $(n) \sim X$

$T = T(X_1, \dots, X_n) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ es un estadístico suficiente para θ (para la familia de funciones de distribución $\{F_\theta\}_{\theta \in \Theta \subset \mathbb{R}^\ell}$), si y sólo si la distribución de probabilidad de la muestra condicionada por T es independiente de θ

Ejemplo

Demostrar que si $X \sim \text{Bin}(1, \theta)$, entonces $T = \sum_{i=1}^n X_i \sim \text{Bin}(n, \theta)$ es suficiente para θ

$$P_\theta(X = x) = \theta^x (1 - \theta)^{1-x} I_{\{0,1\}}(x)$$

$$P_\theta(\sum_{i=1}^n X_i = t) = \binom{n}{t} \theta^t (1 - \theta)^{n-t} \{ \{0,1,\dots,n\} \} (t)$$

$$P_\theta(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} I_{\{0,1\}}(x_i) =$$

$$\theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} \prod_{i=1}^n I_{\{0,1\}}(x_i)$$

$$\text{Si } \sum_{i=1}^n x_i = t, P_\theta(X_1 = x_1, \dots, X_n = x_n \mid \sum_{i=1}^n X_i = t) =$$

$$\frac{P_\theta(X_1 = x_1, \dots, x_{n-1} = x_{n-1}, X_n = t - \sum_{i=1}^{n-1} x_i)}{P(\sum_{i=1}^n X_i = t)} = \frac{\theta^t (1 - \theta)^{n-t}}{\binom{n}{t} \theta^t (1 - \theta)^{n-t}} = \frac{1}{\binom{n}{t}}, t = 0, 1, \dots, n$$

Teorema de Factorización de Fisher

Teorema 3.0.1 [de Factorización de Fisher]

Teorema de Factorización (caracterización de estadísticos suficientes) $T = T(X_1, \dots, X_n) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ es un estadístico suficiente para θ si y sólo si existen funciones reales positivas $h : \mathbb{R}^n \rightarrow \mathbb{R}$ y $g_\theta : \mathbb{R}^m \rightarrow \mathbb{R}$ tales que $f_\theta(x_1, \dots, x_n) = h(x_1, \dots, x_n) g_\theta(T(x_1, \dots, x_n))$, donde $f_\theta(x_1, \dots, x_n)$ es la función de densidad o de masa de la muestra

Demostración. \Leftrightarrow Si $T(x_1, \dots, x_n) = t$,

$$f_\theta(x_1, \dots, x_n \mid t) = \frac{f_\theta(x_1, \dots, x_n)}{f_\theta(t)} = \frac{f_\theta(x_1, \dots, x_n)}{\sum_{\{(y_1, \dots, y_n) : T(y_1, \dots, y_n) = t\}} f_\theta(y_1, \dots, y_n)} =$$

$$\frac{h(x_1, \dots, x_n) g_\theta(t)}{\sum_{\{(y_1, \dots, y_n) : T(y_1, \dots, y_n) = t\}} h(y_1, \dots, y_n) g_\theta(t)} = \frac{h(x_1, \dots, x_n)}{\sum_{\{(y_1, \dots, y_n) : T(y_1, \dots, y_n) = t\}} h(y_1, \dots, y_n)} \text{ es}$$

independiente de $\theta \Rightarrow T$ es suficiente para θ

\Rightarrow Sea $T = T(X_1, \dots, X_n)$ un estadístico suficiente para $\theta \Rightarrow f_\theta(x_1, \dots, x_n \mid t)$ es independiente de θ

Si $T(x_1, \dots, x_n) = t$, $f_\theta(x_1, \dots, x_n) = f_\theta(x_1, \dots, x_n \mid t) f_\theta(t) = h(x_1, \dots, x_n) g_\theta(T(x_1, \dots, x_n))$ \square

Teorema de Factorización de Fisher (continuación)

Ejercicio Encontrar un estadístico suficiente para θ si $X \sim \text{Bin}(1, \theta)$ $f_\theta(x_1, \dots, x_n) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} I_{\{0,1\}}(x_i) =$
 $\theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} \prod_{i=1}^n I_{\{0,1\}}(x_i) = g_\theta(\sum_{i=1}^n x_i) h(x_1, \dots, x_n) \Rightarrow T = \sum_{i=1}^n X_i$ es suficiente para θ

Ejercicios propuestos

1 $T = \sum_{i=1}^n X_i$ es suficiente para θ si $X \sim \text{Poisson}(\theta)$

2 La propia muestra (X_1, \dots, X_n) y el estadístico ordenado $(X_{(1)}, \dots, X_{(n)})$ son suficientes para θ

- 3 Si $T = T(X_1, \dots, X_n)$ es suficiente para θ entonces cualquier biyección $S = S(T)$ también es suficiente para θ
- 4 Si $T = T(X_1, \dots, X_n)$ y $S = S(X_1, \dots, X_n)$ son dos estadísticos suficientes para θ , entonces también es suficiente para θ el estadístico (T, S)

Teorema de Factorización de Fisher (continuación)

- 5 $T = \bar{X}$ es suficiente para μ si $X \sim N(\mu, \sigma_0)$, σ_0 conocida
- 6 $T = \sum_{i=1}^n (X_i - \mu_0)^2$ es suficiente para σ si $X \sim N(\mu_0, \sigma)$, μ_0 conocida
- 7 $T = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ es suficiente para $\theta = (\mu, \sigma)$ si $X \sim N(\mu, \sigma)$, μ y σ desconocidas $\Rightarrow (\bar{X}, S_n^2)$ también es suficiente
- 8 Los estadísticos $(X_{(1)}, X_{(n)})$ y $X_{(n)}$ son suficientes para θ si $X \sim U(0, \theta)$
- 9 El estadístico $(X_{(1)}, X_{(n)})$ es suficiente para θ si $X \sim U(-\frac{\theta}{2}, \frac{\theta}{2})$

Estadístico minimal suficiente

Dado un estadístico $T = T(X_1, \dots, X_n)$, se define $A_t = \{(x_1, \dots, x_n) \in \chi^n : T(x_1, \dots, x_n) = t\}$ (órbita) $K_T = \{A_t, t \in \mathbb{R}^m\}$ es una partición disjunta de χ^n y se denomina partición inducida por el estadístico T

Se dice que K_T es suficiente sí y sólo si T es suficiente

Dados dos estadísticos $T = T(X_1, \dots, X_n)$ y $S = S(X_1, \dots, X_n)$, K_S es una subpartición de K_T , sí y sólo si $\forall B \in K_S, \exists A \in K_T$ tal que $B \subset A$. En este caso, se dice que K_T es una partición menos fina que K_S

Definición 3.0.2 [Estadístico Minimal Suficiente]

1. $T = T(X_1, \dots, X_n)$ es minimal suficiente sí y sólo si K_T es suficiente y $\forall S = S(X_1, \dots, X_n)$ suficiente, K_S es una subpartición de K_T
2. $T = T(X_1, \dots, X_n)$ es minimal suficiente sí y sólo si T es suficiente y $\forall S = S(X_1, \dots, X_n)$ suficiente, $\exists \psi$ tal que $\psi(S) = T$

Demostración. \Rightarrow) Sea S suficiente, si $S(x_1, \dots, x_n) = S(y_1, \dots, y_n) = s \Rightarrow (x_1, \dots, x_n), (y_1, \dots, y_n) \in B_s \Rightarrow \exists A_t \in K_T$ tal que $B_s \subset A_t \Rightarrow T(x_1, \dots, x_n) = T(y_1, \dots, y_n) = t \Rightarrow \exists \psi$ tal que $\psi(S) = T$ y T es suficiente

\Leftarrow) Sea S suficiente y φ tal que $\psi(S) = T \Rightarrow T$ es suficiente y si $S(x_1, \dots, x_n) = S(y_1, \dots, y_n) = s \Rightarrow T(x_1, \dots, x_n) = \psi(S(x_1, \dots, x_n)) = \psi(s) = \psi(S(y_1, \dots, y_n)) = T(y_1, \dots, y_n) \Rightarrow B_s \subset A_{\psi(s)} \Rightarrow K_S$ es una subpartición de K_T \square

Teorema de caracterización de estadísticos

minimales suficientes Definamos la siguiente relación de equivalencia $(x_1, \dots, x_n) R(y_1, \dots, y_n) \Leftrightarrow \frac{f_\theta(x_1, \dots, x_n)}{f_\theta(y_1, \dots, y_n)}$ es independiente de θ Asignemos a cada clase del conjunto cociente un valor t y definamos el estadístico $T = T(X_1, \dots, X_n)$ tal que $\frac{f_\theta(x_1, \dots, x_n)}{f_\theta(y_1, \dots, y_n)}$ es independiente de θ cuando $T(x_1, \dots, x_n) = t = T(y_1, \dots, y_n)$. Entonces T es minimal suficiente

Demostración

Supongamos que T es suficiente y demostremos que es minimal Sea $S = S(X_1, \dots, X_n)$ suficiente y $(x_1, \dots, x_n), (y_1, \dots, y_n) \in B_s \Rightarrow \frac{f_\theta(x_1, \dots, x_n)}{f_\theta(y_1, \dots, y_n)} = \frac{h(x_1, \dots, x_n)g_\theta(s)}{h(y_1, \dots, y_n)g_\theta(s)} = \frac{h(x_1, \dots, x_n)}{h(y_1, \dots, y_n)}$ es independiente de $\theta \Rightarrow \exists t$ tal que $(x_1, \dots, x_n), (y_1, \dots, y_n) \in A_t \Rightarrow K_S$ es una subpartición de K_T

Teorema de caracterización de estadísticos minimales suficientes (continuación)

Demostremos ahora que T es suficiente (caso discreto)

Si $T(x_1, \dots, x_n) = t, f_\theta(x_1, \dots, x_n | t) = \frac{f_\theta(x_1, \dots, x_n)}{f_\theta(t)} =$

$\frac{f_\theta(x_1, \dots, x_n)}{\sum_{\{(y_1, \dots, y_n): T(y_1, \dots, y_n)=t\}} f_\theta(y_1, \dots, y_n)} = \frac{1}{\sum_{(y_1, \dots, y_n) \in A_t} \frac{f_\theta(y_1, \dots, y_n)}{f_\theta(x_1, \dots, x_n)}}$ es independiente de $\theta \Rightarrow T$ es suficiente para θ

Ejercicio $\sum_{i=1}^n X_i$ es minimal suficiente para θ si $X \sim \text{Bin}(1, \theta)$

$f_\theta(x_1, \dots, x_n) = \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i} \Big|_{\{0,1\}}(x_i) =$

$\theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i} \prod_{i=1}^n I_{\{0,1\}}(x_i)$

$\frac{f_\theta(x_1, \dots, x_n)}{f_\theta(y_1, \dots, y_n)} = \left(\frac{\theta}{1-\theta}\right)^{\sum_{i=1}^n x_i - \sum_{i=1}^n y_i}$ es independiente de θ cuando

$\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$

Familia exponencial k -paramétrica

Definición 3.0.3 [Familia exponencial k -paramétrica]

Sea $(\Omega, \mathcal{A}, \mathcal{P})$ el espacio probabilístico asociado a un experimento aleatorio \mathcal{E} , una variable aleatoria observable $X: \Omega \rightarrow \mathbb{R}$ y su modelo estadístico asociado $(\chi, \mathcal{B}, F_\theta)_{\theta \in \Theta \subset \mathbb{R}^k}, \mathcal{B} = \mathcal{B}(\mathbb{R})$, y (X_1, \dots, X_n) m.a.s. $(n) \sim X$

La distribución de X pertenece a la familia exponencial k -paramétrica si y sólo si

$$f_\theta(x) = c(\theta)h(x)e^{\sum_{j=1}^k q_j(\theta)T_j(x)}$$

$$f_\theta(x_1, \dots, x_n) = c(\theta)^n \prod_{i=1}^n h(x_i) e^{\sum_{j=1}^k q_j(\theta) \sum_{i=1}^n T_j(x_i)}$$

Entonces, $(\sum_{i=1}^n T_1(X_i), \dots, \sum_{i=1}^n T_k(X_i))$ es suficiente para θ (Teorema de factorización) y se le denomina estadístico natural

$$X \sim N(\sigma, \mu) \quad f_\theta = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

Como $(x-\mu)^2 = x^2 + \mu^2 - 2x\mu$

$$f_\theta = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}x^2} e^{\frac{\mu}{\sigma^2}x} e^{-\frac{\mu^2}{2\sigma^2}}$$

Familia exponencial k -paramétrica (continuación)

Teorema 1 Sean $\theta_1, \dots, \theta_k \in \Theta \subset \mathbb{R}^\ell$ tales que los vectores $c_r = (q_1(\theta_r), \dots, q_k(\theta_r)), r = 1, \dots, k$ son linealmente independientes, entonces el estadístico natural suficiente de la familia exponencial k -paramétrica

es minimal

Demostración

$$\frac{f_{\theta}(x_1, \dots, x_n)}{f_{\theta}(y_1, \dots, y_n)} = \frac{c(\theta)^n \prod_{i=1}^n h(x_i) \sum_{j=1}^k q_j(\theta) \sum_{i=1}^n T_j(x_i)}{c(\theta)^n \prod_{i=1}^n h(y_i) \sum_{j=1}^k q_j(\theta) \sum_{i=1}^n T_j(y_i)}$$

$$= \frac{\prod_{i=1}^n h(x_i)}{\prod_{i=1}^n h(y_i)} e^{\sum_{j=1}^k q_j(\theta) (\sum_{i=1}^n T_j(x_i) - \sum_{i=1}^n T_j(y_i))}$$

es independiente de θ si y sólo si $\sum_{j=1}^k q_j(\theta) (\sum_{i=1}^n T_j(x_i) - \sum_{i=1}^n T_j(y_i)) = 0$, $r = 1, \dots, k$, sólo admite la solución $\sum_{i=1}^n T_j(x_i) - \sum_{i=1}^n T_j(y_i) = 0$, $r = 1, \dots, k$. Entonces $(\sum_{i=1}^n T_1(X_i), \dots, \sum_{i=1}^n T_k(X_i))$ es minimal (Teorema de caracterización de estadísticos minimales suficientes)

Familia exponencial k -paramétrica (continuación)

Ejercicio $T = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ es minimal suficiente para $\theta = (\mu, \sigma)$ si $X \sim N(\mu, \sigma)$, μ y σ desconocidas $\Rightarrow (\bar{X}, S_n^2)$ también es minimal suficiente

$N(\mu, \sigma)$ pertenece a la familia exponencial k -paramétrica con $k = 2$ $f_{\theta}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} = c(\theta)h(x)e^{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x}$

$\Rightarrow T = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ es natural suficiente para θ

Además, $q_1(\theta) = \frac{\mu}{\sigma^2}$ y $q_2(\theta) = -\frac{1}{2\sigma^2}$ y tomando $\theta_1 = (0, 1)$ y $\theta_2 = (1, 1)$, los vectores $c_1 = (q_1(\theta_1), q_2(\theta_1)) = (0, -\frac{1}{2})$ y $c_2 = (q_1(\theta_2), q_2(\theta_2)) = (1, -\frac{1}{2})$ son linealmente independientes $\Rightarrow T$ es minimal

Familia exponencial k -paramétrica (continuación)

$$S_n^2 = \frac{n}{n-1} (\sum_{i=1}^n X_i^2 - n\bar{X}^2)$$

Denotemos por $(W, Z) = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ y

$$(F, G) = (\bar{X}, S_n^2) = \left(\frac{w}{n}, \frac{n}{n-1} \left(Z - \frac{w^2}{n} \right) \right)$$

Transformación inversa ($W = nF, Z = \frac{n-1}{n}G + nF^2$)

$$J = \begin{vmatrix} n & 0 \\ 2nF & \frac{n-1}{n} \end{vmatrix} = n-1 \neq 0 \text{ si } n \geq 2$$

Por lo tanto existe una transformación biyectiva ψ_1 tal que $(F, G) = \psi_1(W, Z)$ y como (W, Z) es minimal suficiente, $\forall S$ suficiente, existe una transformación ψ_2 tal que $\psi_2(S) = (W, Z)$. Por lo tanto, $\psi_1\psi_2(S) = \psi_1(W, Z) = (F, G)$ y (F, G) es minimal suficiente

Estadísticos Ancilarios y Completos

Sea $(\Omega, \mathcal{A}, \mathcal{P})$ el espacio probabilístico asociado a un experimento aleatorio \mathcal{E} , una variable aleatoria observable $X : \Omega \rightarrow \mathbb{R}$ y su modelo estadístico asociado $(\chi, \mathcal{B}, F_{\theta})_{\theta \in \Theta \subset \mathbb{R}^{\ell}}$, $\mathcal{B} = \mathcal{B}(\mathbb{R})$, y (X_1, \dots, X_n) m.a.s. $(n) \sim X$

El estadístico $U = U(X_1, \dots, X_n)$ es ancilario para θ si su distribución en el muestreo es independiente de θ

Ejercicio Si (X_1, \dots, X_n) es una m.a.s. (n) de $X \sim N(\mu, \sigma_0)$, σ_0 conocida, entonces $U(X_1, \dots, X_n) = X_1 - X_2 \sim N(0, \sigma_0\sqrt{2})$ es un estadístico ancilario para μ

La familia de distribuciones de probabilidad $\{G_{\theta}(y_1, \dots, y_m)\}_{\theta \in \Theta \subset \mathbb{R}^e}$ es completa sí y sólo sí para cualquier función real $h(y_1, \dots, y_m)$ con $h(Y_1, \dots, Y_m)$ medible y tal que $E_{\theta}[h(Y_1, \dots, Y_m)] = 0, \forall \theta \in \Theta$, se sigue que $h(Y_1, \dots, Y_m) \stackrel{CS}{=} 0$

Estadísticos Ancilarios y Completos (continuación)

Ejercicio La familia de distribuciones de probabilidad $\text{Bin}(n, \theta)$ es completa

Sea $Y \sim \text{Bin}(n, \theta)$,

$$E_\theta[h(Y)] = \sum_{i=1}^n h(i) \binom{n}{i} \theta^i (1-\theta)^{n-i} =$$

$$(1-\theta)^n \sum_{i=1}^n h(i) \binom{n}{i} \left(\frac{\theta}{1-\theta}\right)^i = 0, \forall \theta \in (0, 1),$$

que es un polinomio de grado n en $\frac{\theta}{1-\theta} \in (0, \infty)$, luego para que sea nulo, ha de ser $h(i) = 0, \forall i = 1, \dots, n$.

Por lo tanto $h(Y) \stackrel{\text{cs}}{=} 0$

Ejercicio La familia de distribuciones de probabilidad $N(0, \theta)$ no es completa

Sea $Y \sim N(0, \theta)$ y $h(Y) = Y$, entonces $E_\theta[h(Y)] = E_\theta[Y] = 0 \forall \theta > 0$, sin embargo $h(Y) = Y$ no es idénticamente nula c.s.

Estadísticos Ancilarios y Completos (continuación)

El estadístico $T = T(X_1, \dots, X_n)$ es completo sí y sólo sí su distribución en el muestreo es una familia de distribuciones de probabilidad completa

Ejercicio Si $X \sim \text{Bin}(1, \theta), T = \sum_{i=1}^n X_i \sim \text{Bin}(n, \theta)$ es completo

Ejercicio Si $X \sim N(\theta, \theta), T = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ no es completo Indicación: $h(T) = \left(2(\sum_{i=1}^n X_i)^2 - (n+1)\sum_{i=1}^n X_i\right)$

Ejercicio Si $S = S(X_1, \dots, X_n)$ es suficiente y completo, entonces es minimal suficiente Indicación: Demostrar que si $T = T(X_1, \dots, X_n)$ es minimal suficiente, entonces $S \stackrel{\text{cs}}{=} E[S | T]$ y por lo tanto S es función de T

Estadísticos Ancilarios y Completos (continuación)

Teorema 2 El estadístico natural $(\sum_{i=1}^n T_1(X_i), \dots, \sum_{i=1}^n T_k(X_i))$ de la familia de distribuciones exponencial k -paramétrica, $\left\{f_\theta(x) = c(\theta)h(x)e^{\sum_{j=1}^k q_j(\theta)T_j(x)}\right\}_{\theta \in \Theta \subset \mathbb{R}^e}$, es completo si la imagen de la aplicación $q = (q_1(\theta), \dots, q_k(\theta)) : \Theta \longrightarrow \mathbb{R}^k$ contiene un rectángulo abierto de \mathbb{R}^k

Observación Si $X \sim N(\theta, \theta), q = (q_1(\theta), q_2(\theta)) = \left(\frac{1}{\theta}, -\frac{1}{2\theta^2}\right) \Rightarrow q_2(\theta) = -\frac{1}{2}q_1(\theta)^2, \forall \theta > 0$, que es una rama de parábola, y que por lo tanto no contiene ningún abierto de \mathbb{R}^2

Estadísticos Ancilarios y Completos (continuación)

Teorema de Basu Si $T = T(X_1, \dots, X_n)$ es un estadístico suficiente y completo y $U = U(X_1, \dots, X_n)$ es un estadístico ancilar, entonces T y U son independientes

Demostración $T = T(X_1, \dots, X_n)$ es un estadístico suficiente $\Rightarrow f(x_1, \dots, x_n | t)$ es independiente de $\theta \Rightarrow f(u | t) = \sum_{(x_1, \dots, x_n) : U(x_1, \dots, x_n) = u} f(x_1, \dots, x_n | t)$ es independiente de θ Además, como $T = T(X_1, \dots, X_n)$ es un estadístico completo y la función $h(t) = f(u | t) - f(u)$ tiene media 0, $\forall \theta$, respecto a la distribución de $T \sim f_\theta(t) \Rightarrow f(u | t) \stackrel{\text{cs}}{=} f(u)$

Ejercicio Si $X \sim U(0, \theta)$, entonces $X_{(n)}$ y $\frac{X_{(1)}}{X_{(n)}}$ son independientes

Principios de reducción de datos

Sea $(\Omega, \mathcal{A}, \mathcal{P})$ el espacio probabilístico asociado a un experimento aleatorio \mathcal{E} , una variable aleatoria observable $X : \Omega \longrightarrow \mathbb{R}$ y su modelo estadístico asociado $(\chi, \mathcal{B}, F_\theta)_{\theta \in \Theta \subset \mathbb{R}^\ell}$, $\mathcal{B} = \mathcal{B}(\mathbb{R})$, y (X_1, \dots, X_n) m.a.s.

$$(n) \sim X$$

Principio de verosimilitud

La idea es considerar la distribución de probabilidad de la muestra, no como función de (x_1, \dots, x_n) sino como función del parámetro θ desconocido

Supuesto que se ha observado un valor muestral (x_1, \dots, x_n) , la función de θ definida mediante $L(\theta | x_1, \dots, x_n) = f(x_1, \dots, x_n | \theta)$, se llama función de verosimilitud, siendo $f(x_1, \dots, x_n | \theta)$ la función de densidad o de masa de la muestra

Principios de reducción de datos (continuación)

Si $\mathbf{x} = (x_1, \dots, x_n)$ e $\mathbf{y} = (y_1, \dots, y_m)$ son dos puntos muestrales, tales que existe una constante $c(\mathbf{x}, \mathbf{y})$ verificando que $L_1(\theta | \mathbf{x}) = c(\mathbf{x}, \mathbf{y})L_2(\theta | \mathbf{y})$, entonces la evidencia estadística que suministran ambos puntos debe ser idéntica

Dos aspectos son importantes en esta definición. El primero es que la evidencia estadística se toma en un sentido amplio y no se define, así puede ser ésta, un estadístico muestral, un estadístico suficiente, un intervalo de confianza, etc. El segundo es que las dos funciones de verosimilitud no tienen por qué estar obligatoriamente definidas en el mismo espacio muestral. Realmente la evidencia estadística depende del experimento bajo estudio E y del punto observado y debe expresarse como $Ev(E, \mathbf{x})$, $E = (\chi^n, f(\mathbf{x} | \theta))_{\theta \in \Theta \subset \mathbb{R}^e}$

Principios de reducción de datos (continuación)

Ejercicio $Ev(E_1, t) = Ev(E_2, n)$, si

$$E_1 = (\{0, 1, \dots, n\}, \text{Bin}(n, \theta))_{\theta \in (0,1)}, f(t | \theta) = \binom{n}{t} \theta^t (1 - \theta)^{n-t}$$

$$E_2 = (\mathbb{N}, \text{BN}(t, \theta))_{\theta \in (0,1)}, g(n | \theta) = \binom{n-1}{t-1} \theta^t (1 - \theta)^{n-t}$$

Ejercicio Los procedimientos bayesianos, por estar basados en la distribución de probabilidad final o a posteriori, satisfacen el principio de verosimilitud, ya que si $\mathbf{x} = (x_1, \dots, x_n)$ e $\mathbf{y} = (y_1, \dots, y_m)$ satisfacen el principio de verosimilitud, entonces $\pi_1(\theta | \mathbf{x}) = \pi_2(\theta | \mathbf{y})$

Principio de suficiencia

En un experimento $E = (\chi^n, f(\mathbf{x} | \theta))_{\theta \in \Theta \subset \mathbb{R}^e}$, si $T = T(\mathbf{X})$ es un estadístico suficiente para θ y se tiene que $T(\mathbf{x}) = T(\mathbf{y})$, entonces $Ev(E, \mathbf{x}) = Ev(E, \mathbf{y})$

Principios de reducción de datos (continuación)

Principio de condicionalidad

Dados dos experimentos $E_1 = (\chi^n, f_1(\mathbf{x} | \theta))_{\theta \in \Theta \subset \mathbb{R}^e}$ y $E_2 = (\chi^m, f_2(\mathbf{y} | \theta))_{\theta \in \Theta \subset \mathbb{R}^e}$, y el lanzamiento de una moneda al aire representado por la v.a. J tal que $P(J = 1) = P(J = 2) = \frac{1}{2}$, si $E = (\chi^n \cup \chi^m \times \{1, 2\}, f(\mathbf{x}, j | \theta))_{\theta \in \Theta \subset \mathbb{R}^e}$ es el experimento mixto representado por la v.a. (Z, J) tal que $Z = \begin{cases} X & \text{si } J = 1 \\ Y & \text{si } J = 2 \end{cases}$, $f(\mathbf{x}, 1 | \theta) = \frac{1}{2}f_1(\mathbf{x} | \theta)$, $f(\mathbf{y}, 2 | \theta) = \frac{1}{2}f_2(\mathbf{y} | \theta)$, entonces $Ev(E, (\mathbf{x}, 1)) = Ev(E_1, \mathbf{x})$ y $Ev(E, (\mathbf{y}, 2)) = Ev(E_2, \mathbf{y})$

El principio de condicionalidad dice algo bastante intuitivo: mecanismos aleatorios que no dependan del valor a determinar θ , no proporcionan evidencia sobre él (aleatorización en los contrastes de hipótesis para conseguir un test de tamaño determinado)

Teorema de Birnbaum

El principio de verosimilitud es equivalente a los principios de suficiencia y condicionalidad

Observación El Teorema de Birnbaum es importante desde el punto de vista de los fundamentos de la Estadística. Muchos de los procedimientos estadísticos usuales violan el principio de verosimilitud, en concreto los procedimientos que se basan en la distribución en el muestreo de un estadístico pueden hacerlo. Por ejemplo, si se pasa de un modelo binomial a uno binomial negativo, la función de masa cambia y por lo tanto los IC pueden cambiar. Sin embargo, a la luz del teorema, esto significa contradecir el principio de suficiencia, que es compartido por toda aproximación a la inferencia, o el principio de condicionalidad, que parece bastante aséptico. El teorema de Birnbaum constituye uno de los motivos por los que el principio de verosimilitud no es universalmente aceptado, a pesar de que como se verá, la función de verosimilitud posee muchas buenas propiedades estadísticas