

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 | 7 |
| 2.4 | Resultados de convergencias | 8 |
| 2.5 | Estadísticos ordenados | 16 |
| 2.6 | Ejercicios | 18 |
| 3 | Reducción de Datos | 22 |
| 3.1 | Teorema de Factorización de Fisher | 22 |
| 3.2 | Ejemplos de Factorización de Fisher | 23 |
| 3.3 | Estadístico minimal suficiente | 25 |
| 3.4 | Teorema de caracterización de estadísticos minimales suficientes | 27 |
| 3.5 | Familia exponencial k -paramétrica | 28 |
| 3.6 | Estadísticos Ancilarios y Completos | 31 |
| 3.7 | Principios de reducción de datos | 41 |
| 3.8 | Ejercicios | 42 |
| 4 | Estimación Puntual Paramétrica | 49 |
| 4.1 | Estimadores Bayesianos | 52 |
| 4.2 | Criterios de comparación de estimadores | 56 |
| 4.3 | Cota para la varianza de un estimador | 65 |
| 4.4 | Métodos de construcción de estimadores | 70 |
| 4.4.1 | Método de los momentos | 70 |
| 4.4.2 | Método de máxima verosimilitud | 72 |
| 4.5 | Propiedades asintóticas de los estimadores de máxima verosimilitud | 74 |
| 5 | Estimación por Regiones de Confianza | 77 |
| 5.1 | Región de confianza | 77 |
| 5.2 | Intervalos de confianza | 77 |
| 5.3 | Métodos de obtención de intervalos de confianza | 77 |
| 6 | Contraste de Hipótesis | 83 |
| 6.1 | Principios básicos de un contraste de hipótesis | 83 |
| 6.2 | Errores de tipo I y de tipo II | 84 |
| 6.3 | Test uniformemente más potente de tamaño α | 86 |
| 6.4 | Hipótesis nula simple frente a alternativa simple | 86 |
| 7 | Appendix | 88 |
| 7.1 | Momentos Notables | 88 |
| 7.2 | Función Característica | 91 |

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:

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

Es decir, $\|F_n - F\|_\infty = \sup_x |F_n(x) - F(x)| \xrightarrow[n \rightarrow \infty]{c.s.} 0$

Demostración. Para simplicidad, consideremos el caso de una variable aleatoria continua X . Fijemos $-\infty = x_0 < x_1 < \dots < x_{m-1} < x_m = \infty$ tal que

$$F(x_j) - F(x_{j-1}) = \frac{1}{m} \quad \text{para } j = 1, \dots, m.$$

Ahora, para todo $x \in \mathbb{R}$, existe $j \in \{1, \dots, m\}$ tal que $x \in [x_{j-1}, x_j]$.

$$F_n(x) - F(x) \leq F_n(x_j) - F(x_j) + \frac{1}{m},$$

$$F_n(x) - F(x) \geq F_n(x_{j-1}) - F(x_{j-1}) - \frac{1}{m}.$$

Por lo tanto,

$$\|F_n - F\|_\infty = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \max_{j \in \{1, \dots, m\}} |F_n(x_j) - F(x_j)| + \frac{1}{m}.$$

Dado que

$$\max_{j \in \{1, \dots, m\}} |F_n(x_j) - F(x_j)| \rightarrow 0 \quad \text{c.s.}$$

por la ley fuerte de los grandes números, podemos garantizar que para cualquier $\varepsilon > 0$ y cualquier entero m tal que $1/m < \varepsilon$, podemos encontrar N tal que para todo $n \geq N$,

$$\max_{j \in \{1, \dots, m\}} |F_n(x_j) - F(x_j)| \leq \varepsilon - 1/m \quad \text{c.s.}$$

Combinando con el resultado anterior, esto implica además que

$$\|F_n - F\|_\infty \leq \varepsilon \quad \text{c.s.},$$

lo que es la definición de convergencia casi segura. □

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 $T = 2aY \sim \text{Gamma}\left(\frac{1}{2}, p\right)$.

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 $\left(\frac{1}{2}, p\right)$, lo que demuestra que $T \sim \text{Gamma}\left(\frac{1}{2}, p\right)$. \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

□

Definición 2.4.1 [Distribución t de Student]

Dadas las v.a. independientes X, X_1, \dots, X_n con distribución $N(0, 1)$ la v.a.

$$T_n = X \Bigg/ \sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}$$

tiene una distribución T_n de Student con n grados de libertad. De forma equivalente, una v.a. T_n tiene esta distribución, cuando su función de densidad es

$$f_{T_n}(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n} \Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \quad t \in \mathbb{R}. \quad (1)$$

Es fácil establecer que la media de la distribución T_n es 0 y que la varianza es $V[T_n] = \frac{n}{n-2}$ si $n > 2$.

Corolario 2.4.1

Para m.a.s. de tamaño n de una población $N(\mu, \sigma)$, la distribución de la variable

$$T_{n-1} = \frac{\bar{X} - \mu}{S} \sqrt{n}$$

con S^2 la cuasivarianza muestral, es una distribución de Student con $n - 1$ grados de libertad.

Demostración. La variable $\frac{\bar{X} - \mu}{\sigma} \sqrt{n}$ tiene distribución $N(0, 1)$ y la variable $(n - 1) \frac{S^2}{\sigma^2}$ tiene distribución χ_{n-1}^2 como por el teorema de Fisher, ambas son independientes, se sigue el resultado. \square

Definición 2.4.2 [Distribución F de Snedecor]

Dadas las m.a.s. (X_1, \dots, X_m) de una población $N(0, 1)$ e (Y_1, \dots, Y_n) de una población $N(0, 1)$, si ambas muestras son independientes, la v.a.

$$F_{m,n} = \frac{\frac{1}{m} \sum_{i=1}^m X_i^2}{\frac{1}{n} \sum_{i=1}^n Y_i^2}$$

tiene una distribución F de Snedecor con (m, n) grados de libertad, o de forma equivalente, una v.a. tiene esta distribución cuando su función de densidad es

$$f(x) = \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}} \Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}, \quad \text{para } x > 0.$$

Distribuciones relacionadas con la distribución F de Snedecor:

- Si $X \sim F_{m,n}$ entonces $Y = \lim_{n \rightarrow \infty} mX$ tiene una distribución chi cuadrada χ_m^2 .
- Si $X \sim \chi_m^2$ y $Y \sim \chi_n^2$ son independientes entonces $\frac{X/m}{Y/n} \sim F_{m,n}$.

- Si $X \sim \text{Beta}\left(\frac{\alpha}{2}, \frac{\beta}{2}\right)$ entonces $\frac{\beta X}{\alpha(1-X)} \sim F_{\alpha, \beta}$.
- Si $X \sim F_{m, n}$ entonces $X^{-1} \sim F_{n, m}$.
- Si $X \sim t_{(n)}$ (Distribución t de Student) entonces $X^2 \sim F_{1, n}$.
- Si $X \sim \Gamma(\alpha_1, \beta_1)$ y $Y \sim \Gamma(\alpha_2, \beta_2)$ son independientes entonces

$$\frac{\alpha_2 \beta_1 X}{\alpha_1 \beta_2 Y} \sim F_{2\alpha_1, 2\alpha_2}.$$

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

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

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

$$J = \begin{vmatrix} \frac{\partial X_{(1)}}{\partial H} & \frac{\partial X_{(1)}}{\partial R} \\ \frac{\partial X_{(n)}}{\partial H} & \frac{\partial X_{(n)}}{\partial R} \end{vmatrix} = \begin{vmatrix} 1 & -\frac{1}{2} \\ 1 & \frac{1}{2} \end{vmatrix} = 1.$$

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 *Bernouilli*($p = \frac{1}{2}$) 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 $X \sim \text{Exponencial}(\theta) = \text{Gamma}(\theta, 1) \implies \bar{x} \sim \text{Gamma}(n\theta, n)$.

□

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. $(n) \sim X$

$T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ es un estadístico suficiente para θ cuando $\forall (x_1, \dots, x_n) \in \Omega$ se cumple que:

$$P(X_1 = x_1, \dots, X_n = x_n | T = t) \text{ no depende de } \theta \quad \forall t \in \mathbb{R}^m$$

Ejemplo

Sabiendo que $X \sim \text{Bin}(1, \theta)$, veamos si se cumple que el estadístico $T = \sum_{i=1}^n X_i$ es suficiente para θ . Tenemos que:

$$X_i \sim \text{Bin}(1, \theta) \implies T = \sum_{i=1}^n X_i \sim \text{Bin}(n, \theta)$$

Para saber si un estadístico es suficiente necesitamos además la función de distribución de la muestra, que en este caso:

$$\begin{aligned} X_i &\sim \text{Bin}(1, \theta) \equiv \text{Bernoulli}(\theta) \\ \implies P(X_1 = x_1, \dots, X_n = x_n) &= \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} \prod_{i=1}^n \end{aligned}$$

Para ver si es suficiente, necesitamos calcular la probabilidad condicionada de la muestra dado el estadístico, que en este caso es:

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

3.1 Teorema de Factorización de Fisher

Teorema 3.1.1 [de Factorización de Fisher (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. Esta demostración se desarrolla tomando el caso en el que X sea una v.a. discreta.

- (\Rightarrow) : Supongamos que T es suficiente, entonces como la distribución de la muestra condicionada al estadístico no depende de θ , podemos escribir:

$$f_\theta(x_1, \dots, x_n | t) = \frac{f_\theta(x_1, \dots, x_n)}{f_\theta(t)} \text{ es independiente de } \theta$$

$$\implies f_\theta(x_1, \dots, x_n) = f_\theta(t) \cdot f_\theta(x_1, \dots, x_n | t) \implies$$

Simplemente tomamos $h(x_1, \dots, x_n) = f_\theta(t)$ y $g_\theta(t) = f_\theta(x_1, \dots, x_n | t)$

- (\Leftarrow) : Supongamos que $f_\theta(x_1, \dots, x_n) = h(x_1, \dots, x_n) \cdot g_\theta(T(x_1, \dots, x_n))$, entonces:

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

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

La cual es una expresión no dependiente de θ , por lo que T es suficiente.

□

Proposición 3.1.1

Si T es suficiente para θ y S es una biyección, entonces $S(T)$ es suficiente para θ

Proposición 3.1.2

Si T es suficiente para θ y S es una función medible (en estadística-integrable), entonces $S(T)$ es suficiente para θ

Proposición 3.1.3

Sea X variable aleatoria, con σ y δ parámetros, entonces si T_1 es suficiente para σ y T_2 es suficiente para $\delta \iff T = (T_1, T_2)$ es suficiente para (σ, δ)

3.2 Ejemplos de Factorización de Fisher

Ejemplo

Veamos si θ es un parámetro suficiente para el ejemplo anterior, utilizando la Factorización de Fisher. Como $X \sim \text{Bin}(1, \theta) \equiv \text{Bernoulli}(\theta)$, tenemos que la función de verosimilitud es:

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

donde $I_{\{0,1\}}(x_i)$ es la función indicadora que asegura que $x_i \in \{0, 1\}$. Esto se puede escribir como:

$$f_\theta(x_1, \dots, x_n) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} \prod_{i=1}^n I_{\{0,1\}}(x_i).$$

Por lo que tomando $h(x_1, \dots, x_n) = \prod_{i=1}^n I_{\{0,1\}}(x_i)$ y $g_\theta(T(x_1, \dots, x_n)) = \theta^t (1 - \theta)^{n-t}$, tenemos que $T = \sum_{i=1}^n X_i$ es un estadístico suficiente para θ .

Ejemplo

Sea $X \sim \text{Poisson}(\theta)$ veamos si el estadístico $T = \sum_{i=1}^n X_i$ es suficiente para θ . Calculemos primero la función de densidad asociada a la muestra:

$$X_i \sim \text{Poisson}(\theta) \implies f_{x_i}(x_i) = \frac{e^{-\theta} \theta^{x_i}}{x_i!} \implies f_\theta(x_1, \dots, x_n) = \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!} = e^{-n\theta} \theta^{\sum_{i=1}^n x_i} \prod_{i=1}^n \frac{1}{x_i!}$$

Por lo que tomando $h(x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{x_i!}$ y $g_\theta(T(x_1, \dots, x_n)) = e^{-n\theta} \theta^t$ demostramos que $T(x_1, \dots, x_n) = \sum_{i=1}^n X_i$ es un estadístico suficiente para θ

Ejemplo

Supongamos que tenemos un estadístico T suficiente para el parámetro θ y una biyección S , demostremos que $S(T)$ también es suficiente para θ

Por el Teorema de Caracterización de Fisher tenemos que:

$$f_{\theta}(x_1, \dots, x_n) = h(x_1, \dots, x_n) \cdot g_{\theta}(T(x_1, \dots, x_n))$$

para alguna función h y g_{θ}

$$\iff f_{\theta}(x_1, \dots, x_n) = h(x_1, \dots, x_n) \cdot g_{\theta}(S^{-1}(S(T(x_1, \dots, x_n))))$$

$$\iff f_{\theta}(x_1, \dots, x_n) = h(x_1, \dots, x_n) \cdot g'_{\theta}(S(T(x_1, \dots, x_n))) \implies$$

Entonces por el Teorema de Caracterización de Fisher, $S(T)$ es suficiente para θ

Ejemplo

Veamos si $T = \bar{X} = \sum_{i=1}^n X_i$ es suficiente para μ si $X \sim N(\mu, \sigma_0)$ con σ_0 conocida:

$$X_i \sim N(\mu, \sigma_0) \implies f_{\mu}(x) = \frac{1}{\sigma_0 \sqrt{2\pi}} e^{-\frac{1}{2\sigma_0^2}(x-\mu)^2}$$

$$\implies T = \bar{X} \implies f_{\mu}(x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sigma_0 \sqrt{2\pi}} e^{-\frac{1}{2\sigma_0^2}(x_i-\mu)^2} = \left(\frac{1}{\sigma_0 \sqrt{2\pi}} \right)^n e^{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i-\mu)^2}$$

\implies Tomando $h(x_1, \dots, x_n) = \left(\frac{1}{\sigma_0 \sqrt{2\pi}} \right)^n$ y $g_{\mu}(T(x_1, \dots, x_n)) = e^{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i-\mu)^2}$, tenemos que $T = \bar{X} = \sum_{i=1}^n X_i$ es suficiente para μ

Ejemplo

Sea el estadístico $T = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ suficiente para $\theta = (\mu, \sigma)$ si $X \sim N(\mu, \sigma)$ con μ y σ desconocidas. Veamos si (\bar{X}, S_n^2) también es suficiente para θ :

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

$$\text{Sea la biyección } S: \mathbb{R}^2 \rightarrow \mathbb{R}^2 \text{ dada por } S(x, y) = \left(\frac{1}{n}x, \frac{1}{n-1} \left(y - \frac{1}{n}x^2 \right) \right) \implies$$

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

Entonces por el Teorema de Caracterización de Fisher, (\bar{X}, S_n^2) es suficiente para θ

Ejemplo

Si $X \sim U(0, \theta)$, veamos si el estadístico $T = (X_{(1)}, X_{(n)})$ es suficiente para θ :

$$X \sim U(0, \theta) \implies f_{\theta}(x) = \frac{1}{\theta} \cdot I_{(0, \theta)}(x) \implies f_{\theta}(x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\theta} \cdot I_{(0, \theta)}(x_i) = \frac{1}{\theta^n} \cdot I_{(0, \theta)}(X_{(1)}) \cdot I_{(0, \theta)}(X_{(n)})$$

$$\implies \text{Tomando } h(x_1, \dots, x_n) = 1 \text{ y } g_{\theta}(T(x_1, \dots, X_n)) = \frac{1}{\theta^n} I_{(0, \theta)}(X_{(1)}) \cdot I_{(0, \theta)}(X_{(n)})$$

Ejemplo

Si $X \sim U(0, \theta)$, veamos si el estadístico $T = X_{(n)}$ es suficiente para θ :

$$\begin{aligned} X \sim U(0, \theta) &\implies f_{\theta}(x) = \frac{1}{\theta} \cdot I_{(0, \theta)}(x) \implies f_{\theta}(x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\theta} \cdot I_{(0, \theta)}(x_i) = \frac{1}{\theta^n} \cdot I_{(0, \theta)}(X_{(n)}) \\ &\implies \text{Tomando } h(x_1, \dots, x_n) = 1 \text{ y } g_{\theta}(T(x_1, \dots, X_n)) = \frac{1}{\theta^n} I_{(0, \theta)}(X_{(n)}) \end{aligned}$$

Ejemplo

Si $X \sim U(-\frac{\theta}{2}, \frac{\theta}{2})$, veamos si el estadístico $T = (X_{(1)}, X_{(n)})$ es suficiente para θ :

$$\begin{aligned} X \sim U(-\frac{\theta}{2}, \frac{\theta}{2}) &\implies f_{\theta}(x) = \frac{1}{\theta} \cdot I_{(-\frac{\theta}{2}, \frac{\theta}{2})}(x) \\ \implies f_{\theta}(x_1, \dots, x_n) &= \prod_{i=1}^n \frac{1}{\theta} \cdot I_{(-\frac{\theta}{2}, \frac{\theta}{2})}(x_i) = \frac{1}{\theta^n} \cdot I_{(-\frac{\theta}{2}, \frac{\theta}{2})}(X_{(1)}) \cdot I_{(-\frac{\theta}{2}, \frac{\theta}{2})}(X_{(n)}) \\ \implies \text{Tomando } h(x_1, \dots, x_n) &= 1 \text{ y } g_{\theta}(T(x_1, \dots, X_n)) = \frac{1}{\theta^n} I_{(-\frac{\theta}{2}, \frac{\theta}{2})}(X_{(1)}) \cdot I_{(-\frac{\theta}{2}, \frac{\theta}{2})}(X_{(n)}) \end{aligned}$$

3.3 Estadístico minimal suficiente**Definición 3.3.1** [Órbita]

Dado un estadístico $T = T(X_1, \dots, X_n)$, se define $A_t = \{(x_1, \dots, x_n) \in \chi^n : T(x_1, \dots, x_n) = t\}$ como la órbita de t .

Ejemplo

Dado un m.a.s. de tamaño $n = 3$, definimos el rango muestral como $T(x_1, x_2, x_3) = (x_{\max}, x_{\min})$. Si por ejemplo tomamos el resultado $t = 5$ definimos la órbita de $t = 5$ al conjunto $A_5 = \{(x_1, \dots, x_n) \in \mathbb{R}^3 : \max(x_1, x_2, x_3) - \min(x_1, x_2, x_3) = 5\}$ cuyos elementos podrían ser por ejemplo:

- $(2, 3, 7) \in A_5$ ($7 - 2 = 5$)
- $(10, 12, 15) \in A_5$ ($15 - 10 = 5$)
- $(-1, 4, -5) \notin A_5$ ($4 - (-6) = 10 \neq 5$)

Definición 3.3.2 [Partición Inducida]

Dado un estadístico $T = T(X_1, \dots, X_n)$, se define $K_T = \{A_t : t \in \mathbb{R}^m\}$ como la partición de χ^n inducida por T .

Ejemplo

Sea una muestra aleatoria simple de tamaño $n = 2$ tal que el espacio muestral de cada una de las variables aleatorias es $\chi = \{1, 2, 3\}$ y definimos el estadístico $T(x_1, x_2) = (x_1 + x_2)$. Entonces, podemos definir algunas órbitas de T :

- $A_2 = \{(1, 1), (2, 0), (0, 2)\}$

- $A_3 = \{(1, 2), (2, 1)\}$
- $A_4 = \{(2, 2)\}$
- $A_5 = \{(3, 2), (2, 3)\}$
- ...

Así, una partición inducida por T sería $K_T = \{A_2, A_3, A_4, A_5, \dots\}$

Proposición 3.3.1

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

Proposición 3.3.2

Dado dos estadísticos $T = T(X_1, \dots, X_n)$ y $S = S(X_1, \dots, X_n)$, se dice que K_S es una subpartición de K_T si y solo si:

$$\forall B \in K_S, \exists A \in K_T \text{ tal que } B \subset A.$$

En este caso, se dice que K_T es una partición menos fina que K_S .

Definición 3.3.3 [Estadístico Minimal Suficiente]

Un estadístico T es minimal suficiente si su partición asociada es suficiente y es la menos fina \equiv más gruesa entre todos los estadísticos.

$$\begin{cases} \text{Más fina} \implies \text{Más clases} \implies \text{Más detallada} \\ \text{Más gruesa} \implies \text{Menos clases} \implies \text{Agrupa más elementos} \end{cases}$$

Un estadístico es minimalmente suficiente si es el más "pequeño" o "compacto" posible entre todos los estadísticos suficientes. No hay ningún otro estadístico suficiente que sea más eficiente (en el sentido de que no puede ser simplificado más sin perder información relevante sobre el parámetro). En otras palabras, es el estadístico suficiente más simple y directo que captura toda la información relevante sobre el parámetro.

Proposición 3.3.3

Un estadístico T es minimal suficiente si y sólo si para todo estadístico S suficiente, $\exists \varphi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ medible (integrable) tal que $S = \varphi(T)$.

Demostración.

- (\Rightarrow): Sea S suficiente, si $S(x_1, \dots, x_n) = S(y_1, \dots, y_n) = s \implies (x_1, \dots, x_n), (y_1, \dots, y_n) \in B_s$ (Órbita de S) $\implies \exists A_t \in K_T : B_s \subset A_t \implies T(x_1, \dots, x_n) = T(y_1, \dots, y_n) = t \implies \exists \psi : \psi(S) = T$ y T es suficiente.
- (\Leftarrow): Sea S suficiente y $\psi : \psi(S) = T \implies T$ es suficiente y si $S(x_1, \dots, x_n) = S(y_1, \dots, y_n) = s \implies T(x_1, \dots, x_n) = \psi(S(x_1, \dots, x_n)) = \psi(S) = \psi(S(y_1, \dots, y_n)) = T(y_1, \dots, y_n) \implies B_s \subset A_{\psi(S)} \implies K_S$ es una subpartición de K_T .

3.4 Teorema de caracterización de estadísticos minimales suficientes

Definición 3.4.1 [Relación de equivalencia minimal suficiente]

Sean (x_1, \dots, x_n) y (y_1, \dots, y_n) muestras, definimos la relación de equivalencia minimal suficiente como:

$$(x_1, \dots, x_n) R (y_1, \dots, y_n) \iff \frac{f_\theta(x_1, \dots, x_n)}{f_\theta(y_1, \dots, y_n)} \text{ es independiente de } \theta$$

Esta definición lo que trata de indicar es que ambas muestras del mismo suceso, aportan la misma información sobre el parámetro θ .

Teorema 3.4.1 [Teorema de caracterización de estadísticos minimales (Lehmann-Scheffé)]

A cada clase de equivalencia anterior le asignamos un valor t y definimos un 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 $T(x_1, \dots, x_n) = T(y_1, \dots, y_n) = t$. Entonces, T es minimal suficiente para θ .

Demostración. Supongamos la suficiencia de T para demostrar su minimalidad:

Sea $S = S(X_1, \dots, X_n)$ suficiente y $(x_1, \dots, x_n), (y_1, \dots, y_n) \in B_S \implies$

$$\frac{f_\theta(x_1, \dots, x_n)}{f_\theta(y_1, \dots, y_n)} = \frac{h(x_1, \dots, x_n)g_\theta(s)}{h(y_1, \dots, y_n)g_\theta(s)} = \frac{h(x_1, \dots, x_n)}{h(y_1, \dots, y_n)} \text{ independiente de } \theta$$

$\implies \exists t : (x_1, \dots, x_n), (y_1, \dots, y_n) \in A_t \implies K_S$ es una subpartición de K_T .

Ahora demostremos la suficiencia de T en el caso discreto:

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

$$f_\theta(x_1, \dots, x_n | t) = \frac{f_\theta(x_1, \dots, x_n)}{f_\theta(t)} = \frac{f_\theta(x_1, \dots, x_n)}{\sum_{(y_1, \dots, y_n) : T(y_1, \dots, y_n) = t} f_\theta(y_1, \dots, y_n)} = \frac{1}{\sum_{(y_1, \dots, y_n) \in A_t} \frac{f_\theta(y_1, \dots, y_n)}{f_\theta(x_1, \dots, x_n)}}$$

que es independiente de $\theta \implies T$ es suficiente. □

Ejemplo

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

$$X \sim \text{Bin}(1, \theta) \equiv \text{Bernoulli}(\theta) \implies f_\theta(x) = \theta^x (1 - \theta)^{1-x} \cdot I_{\{0,1\}}(x) \implies$$

$$\implies f_\theta(x_1, \dots, x_n) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} \cdot I_{\{0,1\}}(x_i) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} \cdot \prod_{i=1}^n I_{\{0,1\}}(x_i)$$

\implies Sean dos muestras m.a.s. de tamaño $= n$, (x_1, \dots, x_n) y (y_1, \dots, y_n) , entonces:

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

Observación 3.4.1

Recordemos que $\forall T'$ suficiente $\exists \varphi : \varphi(T) = T'$, por lo que T es minimal suficiente

Ejemplo

Sea $f_\theta(x) = e^{-(x+\theta)} \cdot I_{(\theta,\infty)}(x)$

$$\implies f_\theta(x_1, \dots, x_n) = e^{-\sum_{i=1}^n (x_i + \theta)} \cdot I_{(\theta,\infty)}(x_1) \cdot \dots \cdot I_{(\theta,\infty)}(x_n) = e^{-\sum_{i=1}^n x_i} e^{-n\theta} \cdot I_{(\theta,\infty)}(x_{(1)})$$

Entonces, por el Teorema de Factorización de Fisher, tenemos que:

$$f_\theta(x_1, \dots, x_n) = h(x_1, \dots, x_n) \cdot g_\theta(T(x_1, \dots, x_n)) \implies \begin{cases} h(x_1, \dots, x_n) = e^{-\sum_{i=1}^n x_i} \\ g_\theta(T(x_1, \dots, x_n)) = e^{-n\theta} \cdot I_{(\theta,\infty)}(x_{(1)}) \end{cases}$$

Ahora sean dos muestras m.a.s. de tamaño n , (x_1, \dots, x_n) y (y_1, \dots, y_n) , entonces, tenemos que ver qué expresión hace que el cociente de las funciones de densidad sea independiente de θ :

$$\frac{f_\theta(x_1, \dots, x_n)}{f_\theta(y_1, \dots, y_n)} = \frac{e^{-\sum_{i=1}^n x_i} e^{-n\theta} \cdot I_{(\theta,\infty)}(x_{(1)})}{e^{-\sum_{i=1}^n y_i} e^{-n\theta} \cdot I_{(\theta,\infty)}(y_{(1)})} = \frac{e^{-\sum_{i=1}^n x_i} \cdot I_{(\theta,\infty)}(x_{(1)})}{e^{-\sum_{i=1}^n y_i} \cdot I_{(\theta,\infty)}(y_{(1)})}$$

$$\implies \text{esta expresión es independiente de } \theta \iff X_{(1)} = Y_{(1)} \implies$$

$$\implies \text{el estadístico } T = X_{(1)} \text{ es minimal suficiente para } \theta$$

3.5 Familia exponencial k -paramétrica

Definición 3.5.1 [Familia exponencial k -paramétrica]

Una distribución X pertenece a la familia exponencial k -paramétrica si su función de densidad o de masa, se puede expresar como:

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

donde:

- $c(\theta)$ es una función sólo de θ
- $h(x)$ es una función no negativa sólo de x
- $q_j(\theta)$ son funciones de θ , conocidas como parámetros naturales
- $T_j(x)$ son funciones de x llamadas estadísticos naturales
- k es el número de parámetros de la familia
- θ es el parámetro de la familia

Esto implica a su vez, que si tenemos una m.a.s., la función de densidad conjunta se puede expresar como:

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

Ejemplos de familias exponenciales:

Entre otras, podemos encontrar las siguientes familias exponenciales:

- | | |
|--------------|-----------------|
| 1. Normal | 4. Exponencial |
| 2. Poisson | 5. Geométrica |
| 3. Bernoulli | 6. Chi-cuadrado |

Cabe notar que existen distribuciones que son familias exponenciales, pero solamente cuando ciertos parámetros son fijos y se conocen. Por ejemplo:

1. Binomial (cuando el número de ensayos n es fijo)
2. Multinomial (cuando el número de ensayos n es fijo)
3. Binomial negativa (cuando el número de fallos r es fijo)
4. Laplace (cuando la media es conocida)

Ejemplos de distribuciones comunes que no son familias exponenciales son la t de Student, la mayoría de las distribuciones de mezcla e incluso la familia de distribuciones uniformes cuando los límites no son fijos.

Es importante destacar que el soporte de $f_\theta(x)$ no puede depender de θ en una familia exponencial k -paramétrica. Esto se debe a que los indicadores no se pueden particionar en una parte que dependa de θ y otra que no, como por ejemplo ocurre con la distribución binomial con n desconocido, pues el soporte de la distribución depende de n .

Teorema 3.5.1

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. Sean dos muestras $x = (x_1, \dots, x_n)$ e $y = (y_1, \dots, y_n)$, entonces:

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

Lo cual es una expresión independiente de $\theta \iff$

$$\iff \sum_{j=1}^k q_j(\theta) \left(\sum_{i=1}^n [T_j(x_i) - T_j(y_i)] \right) = 0 \iff \sum_{i=1}^n [T_j(x_i) - T_j(y_i)] = 0, \forall j = 1, \dots, k$$

Lo cual implica que el sistema homogéneo $\sum_{i=1}^n [T_j(x_i) - T_j(y_i)] = 0$, $\forall j = 1, \dots, k$ sólo admite la solución $T_j(x_i) - T_j(y_i) = 0$, $\forall j = 1, \dots, k \implies (T_1(x_1), \dots, T_k(x_n))$ es minimal \square

Observación 3.5.1

El estadístico natural de una distribución X perteneciente a una familia exponencial k -paramétrica con $k = 1$ nunca es minimal.

| Distribución | Parámetros | Parámetros naturales | Estadísticos naturales |
|-------------------|-------------------|--|--|
| Normal | (μ, σ^2) | $\begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix}$ | $\begin{bmatrix} \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i^2 \end{bmatrix}$ |
| Normal | μ | $\frac{\mu}{\sigma}$ | $\frac{1}{\sigma} \sum_{i=1}^n X_i$ |
| Poisson | λ | $\log \lambda$ | $\sum_{i=1}^n X_i$ |
| Bernoulli | p | $\log \left(\frac{p}{1-p} \right)$ | $\sum_{i=1}^n X_i$ |
| Exponencial | λ | $-\lambda$ | $\sum_{i=1}^n X_i$ |
| Binomial | p | $\log \left(\frac{p}{1-p} \right)$ | $\sum_{i=1}^n X_i$ |
| Binomial negativa | p | $\log(1-p)$ | $\sum_{i=1}^n X_i$ |
| Chi-cuadrado | k | $\frac{k}{2} - 1$ | $\log \left(\sum_{i=1}^n X_i \right)$ |
| Gamma | (α, β) | $\begin{bmatrix} \alpha - 1 \\ -\beta \end{bmatrix}$ | $\begin{bmatrix} \log \left(\sum_{i=1}^n X_i \right) \\ \sum_{i=1}^n X_i \end{bmatrix}$ |
| Beta | (α, β) | $\begin{bmatrix} \alpha - 1 \\ \beta - 1 \end{bmatrix}$ | $\begin{bmatrix} \log \left(\sum_{i=1}^n X_i \right) \\ \log \left(1 - \sum_{i=1}^n X_i \right) \end{bmatrix}$ |

Ejemplo

Veamos si el estadístico $T = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ es minimal suficiente para $\theta = (\mu, \sigma)$ si $X \sim N(\mu, \sigma)$ con μ y σ desconocidas.

Una distribución perteneciente a la familia 2-paramétrica exponencial es de la forma:

$$\begin{aligned}
f_{\theta}(x) &= c(\theta)h(x)e^{q_1(\theta)T_1(x)+q_2(\theta)T_2(x)} \implies \\
\implies X \sim N(\mu, \sigma) &\implies f_{\theta}(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \implies f_{\theta}(x_1, \dots, x_n) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2} \\
&= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2} e^{\frac{\mu}{\sigma^2}\sum_{i=1}^n x_i} \implies \\
\text{tomando } &\begin{cases} c(\theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \\ h(x) = 1 \\ q_1(\theta) = -\frac{1}{2\sigma^2} \\ q_2(\theta) = \frac{\mu}{\sigma^2} \\ T_1(x) = \sum_{i=1}^n X_i^2 \\ T_2(x) = \sum_{i=1}^n x_i \end{cases}
\end{aligned}$$

La distribución normal pertenece a la familia exponencial 2-paramétrica. Ahora veamos si el estadístico T es minimal suficiente:

Si tomamos como vectores paramétricos $\theta_1 = (0, 1)$ y $\theta_2 = (1, 1)$, entonces 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 $\implies T$ es

minimal suficiente

Ahora veamos si el estadístico $T = (\bar{X}, S_n^2)$ también es minimal suficiente:

Sabemos que:

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

\implies Sea la transformación $S : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dada por $S(x, y) = \left(\frac{1}{n}x, \frac{1}{n-1} \left(y - \frac{1}{n}x^2 \right) \right) \implies$

Analicemos la transformación inversa:

$$W = nF, \quad Z = \frac{n-1}{n}G + nF^2$$

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

Por lo que cómo S es una biyección, y T es minimal suficiente, entonces la imagen de T por S también es minimal suficiente $S(T) = (\bar{X}, S_n^2)$

Proposición 3.5.1

En las familias k -paramétricas exponenciales, es decir, aquellas de la forma:

$$f_\theta(x) = c(\theta)h(x)e^{q(\theta)T(x)}$$

El estadístico $T(x)$ es suficiente.

3.6 Estadísticos Ancilarios y Completos

Definición 3.6.1 [Estadístico Ancilario]

Un estadístico $U(X_1, \dots, X_n)$ es ancilario para θ si su distribución en la muestra es independiente de θ

Ejemplo

Sea una población que sigue una distribución normal:

$$X_i \sim N(\theta, \sigma_0),$$

donde θ es la media desconocida y σ_0 es la desviación estándar conocida. Supongamos que tomamos una muestra aleatoria simple (m.a.s.) de tamaño n :

$$X_1, X_2, \dots, X_n \sim N(\theta, \sigma_0),$$

de manera independiente.

Definimos el siguiente estadístico basado en los primeros dos valores de la muestra:

$$U(X_1, \dots, X_n) = X_1 - X_2$$

Dado que X_1 y X_2 son variables aleatorias normales independientes con la misma distribución, podemos calcular la distribución de U :

$$X_1 - X_2 \sim N(\theta - \theta, \sqrt{\sigma_0^2 + \sigma_0^2}) = N(0, \sqrt{2\sigma_0^2})$$

Por lo tanto, la desviación estándar de U es $\sqrt{2\sigma_0^2} = \sqrt{2}\sigma_0$, lo que implica que:

$$U \sim N(0, \sqrt{2}\sigma_0).$$

Dado que la distribución de U no depende del parámetro θ , concluimos que U es un **estadístico ancilar** para θ .

Este tipo de estadísticos son útiles en inferencia estadística, por ejemplo, en la construcción de estimadores insesgados y en el desarrollo de métodos basados en pivotes.

Definición 3.6.2 [Familia Completa]

La familia $\mathcal{P} = \{f_\theta(x_1, \dots, x_n) : \theta \in \Theta \subset \mathbb{R}^\ell\}$ es completa si para cualquier función real $g(X_1, \dots, X_n)$ tal que $E_\theta[g(X_1, \dots, X_n)] = 0, \forall \theta \in \Theta \implies g(X_1, \dots, X_n) \stackrel{c.s.}{=} 0$

Intuitivamente, una familia completa de estadísticos es un conjunto de funciones de la muestra que, aunque no necesariamente contienen toda la información sobre el parámetro de interés, cumplen una propiedad clave: no permiten la existencia de funciones insesgadas que se anulen para todos los valores del parámetro sin ser trivialmente nulas.

Para visualizarlo, imagina que tienes varias estadísticas distintas calculadas a partir de una muestra. Si dentro de este conjunto hay una combinación de estas estadísticas cuya esperanza matemática es siempre cero para cualquier valor del parámetro, salvo cuando la función es trivialmente cero, entonces la familia es completa. En caso contrario, la familia no sería completa, ya que existiría una función insesgada que no aporta información útil.

Formalmente, una familia de estadísticas es completa si la única función de ellas cuya esperanza es idénticamente cero para todos los valores posibles del parámetro es la función nula. Esto implica que no hay "información redundante" en la familia, en el sentido de que ninguna combinación no trivial de estas estadísticas se anula en promedio para todos los valores del parámetro.

Este concepto es fundamental en inferencia estadística, ya que garantiza que los estimadores insesgados basados en una familia completa son únicos y óptimos, como se establece en el Teorema de Lehmann-Scheffé que veremos más adelante.

Ejemplo

Veamos si la familia de distribuciones $Bin(n, \theta)$ es completa:

Sea $Y \sim Bin(n, \theta)$, entonces:

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

Si definimos la variable $x = \frac{\theta}{1-\theta} \implies$

$$\sum_{i=1}^n h(i) \binom{n}{i} x^i = 0 \text{ es un polinomio de grado } n \text{ en } x \implies h(i) = 0, \forall i = 1, \dots, n \implies h(Y) \stackrel{c.s.}{=} 0$$

Por lo que la familia de distribuciones $Bin(n, \theta)$ es completa

Ejemplo

Veamos si la familia de distribuciones $N(0, \theta)$ es completa:

Sea $Y \sim N(0, \theta)$, entonces: Si tomamos la función real identidad, definida por $g(Y) = Y$, entonces:

$$E_\theta[g(Y)] = E_\theta[Y] = 0, \quad \forall \theta > 0, \quad \text{pero } g(Y) = Y \text{ no es idénticamente nula c.s.}$$

Por lo que la familia de distribuciones $N(0, \theta)$ no es completa

Ejemplo

Sea X una variable aleatoria discreta $D_x = \{1, 2, 3\}$ comprobemos si X es completa: Tenemos que:

$$\begin{cases} P_\theta(X = 1) = \theta - 3 \\ P_\theta(X = 2) = 2\theta - 1 \\ P_\theta(X = 3) = 3 - 3\theta \end{cases} \quad \text{Tomemos el estadístico } T_1(x_1, \dots, x_n) = x_1 \text{ y comprobemos si es completo:}$$

Sea φ función real que actúa sobre el soporte D_x , entonces:

$$E_\theta[\varphi(T_1)] = \varphi(1) \cdot P(X = 1) + \varphi(2) \cdot P(X = 2) + \varphi(3) \cdot P(X = 3)$$

$$= \varphi(1)(\theta - 3) + \varphi(2)(2\theta - 1) + \varphi(3)(3 - 3\theta) = \theta(\varphi(1) + 2\varphi(2) - 3\varphi(3)) - 3\varphi(1) - \varphi(2) + 3\varphi(3) = 0$$

Pero este sistema de ecuaciones es un sistema compatible indeterminado, por lo que no tiene una única solución y por tanto no tiene porqué cumplirse que $\varphi(T_1) = 0$ c.s. $\implies T_1$ no es completo.

Ejemplo

Sea X una variable aleatoria discreta $D_x = \{1, 2, 3\}$ comprobemos si X es completa: Tenemos que:

$$\begin{cases} P_\theta(X = 1) = \theta^2 + \theta \\ P_\theta(X = 2) = 1 - 2\theta^2 \\ P_\theta(X = 3) = \theta^2 - \theta \end{cases} \implies$$

$$E_\theta[\varphi(T)] = \varphi(1) \cdot P(X = 1) + \varphi(2) \cdot P(X = 2) + \varphi(3) \cdot P(X = 3) =$$

$$= \varphi(1)(\theta^2 + \theta) + \varphi(2)(1 - 2\theta^2) + \varphi(3)(\theta^2 - \theta) = \theta^2(\varphi(1) - 2\varphi(2) + \varphi(3)) + \theta(\varphi(1) - \varphi(3)) + \varphi(2) = 0$$

Lo cual se trata de un polinomio de grado 2 en θ por lo que podría tener solución:

$$\begin{cases} \varphi(1) - 2\varphi(2) + \varphi(3) = 0 \\ \varphi(1) - \varphi(3) = 0 \\ \varphi(2) = 0 \end{cases}$$

$$\implies \varphi(1) = \varphi(3) = \varphi(2) = 0 \implies \varphi(T) = 0 \text{ c.s.} \implies T \text{ es completo}$$

Ejemplo

Sea X una variable aleatoria discreta con dominio $D_X = \{-1\} \cup \{0\} \cup \mathbb{N}$. Queremos comprobar si X es completa. Los valores de probabilidad vienen dados por:

$$P(X = -1) = \theta, \quad P(X = x) = \theta^x(1 - \theta)^2, \quad \forall x \in \{0\} \cup \mathbb{N}$$

Por definición, verificamos si $E[\varphi(X)] = 0$ implica $\varphi(X) = 0$ para toda función medible. Calculamos la esperanza:

$$E[\varphi(X)] = \varphi(-1)P(X = -1) + \sum_{k=0}^{+\infty} \varphi(k)P(X = k)$$

$$= \varphi(-1)\theta + \sum_{k=0}^{+\infty} \varphi(k)\theta^k(1-\theta)^2$$

Factorizando:

$$E[\varphi(X)] = \varphi(-1)\theta + (1-\theta)^2 \sum_{k=0}^{+\infty} \varphi(k)\theta^k = 0$$

Para todo θ , lo que nos lleva a estudiar la ecuación funcional:

$$\varphi(-1)\theta + (1-\theta)^2 \sum_{k=0}^{+\infty} \varphi(k)\theta^k = 0, \quad \forall \theta$$

Esta ecuación es un polinomio en θ . Para que se cumpla para todo θ , cada coeficiente debe anularse individualmente. Analizando los coeficientes del desarrollo en serie de potencias:

1. Coeficiente de θ^{-1} : No hay términos de este tipo, por lo que no impone restricciones.
2. Coeficiente de θ^0 :

$$(1-\theta)^2 \varphi(0) = 0 \implies \varphi(0) = 0$$

3. Coeficiente de θ^1 :

$$\varphi(-1) + (1-\theta)^2 \varphi(1) = 0$$

Evalutando en $\theta = 1$:

$$\varphi(-1) + 0 = 0 \implies \varphi(-1) = 0$$

4. Para $k \geq 1$, los coeficientes del polinomio también deben anularse:

$$\varphi(k) = 0, \quad \forall k \in \mathbb{N}$$

Por lo tanto, se concluye que la única solución posible es la trivial:

$$\varphi(k) = 0, \quad \forall k \in D_X$$

Conclusión: La variable aleatoria X es completa.

Definición 3.6.3 [Estadístico Completo]

Un estadístico T (no necesariamente suficiente) se dice completo, cuando:

$$\forall g\text{-medible(integrable) tal que } E_\theta[g(T)] = 0 \implies g(T) \stackrel{c.s.}{=} 0$$

Es decir, se dice que un estadístico es completo si para toda función $g(x_1, \dots, x_n)$ tal que $E_\theta[g(T)] = 0$ se tiene que $g(X) \stackrel{c.s.}{=} 0$

Alternativamente, también se puede decir, que el estadístico es completo si y solo si su distribución en el muestreo es una familia de distribuciones de probabilidad completa

Ejemplo

Dado que cada $X_i \sim \text{Bin}(1, \theta)$, la suma

$$T = \sum_{i=1}^n X_i$$

sigue una distribución binomial con parámetros n y θ :

$$T \sim \text{Bin}(n, \theta).$$

Para demostrar que T es un estadístico completo, consideremos una función $g(T)$ tal que:

$$E_\theta[g(T)] = 0, \quad \forall \theta \in (0, 1).$$

Calculando la esperanza,

$$E_\theta[g(T)] = \sum_{i=0}^n g(i) \binom{n}{i} \theta^i (1-\theta)^{n-i}.$$

Factorizando $(1-\theta)^n$:

$$E_\theta[g(T)] = (1-\theta)^n \sum_{i=0}^n g(i) \binom{n}{i} \left(\frac{\theta}{1-\theta}\right)^i.$$

Definiendo $x = \frac{\theta}{1-\theta}$, obtenemos:

$$\sum_{i=0}^n g(i) \binom{n}{i} x^i = 0, \quad \forall x \in (0, \infty).$$

Dado que el lado izquierdo es un polinomio de grado n en x que se anula para todo x , por el **teorema fundamental del álgebra**, los coeficientes del polinomio deben ser todos cero:

$$g(i) \binom{n}{i} = 0, \quad \forall i = 0, \dots, n.$$

Como los coeficientes binomiales $\binom{n}{i}$ son distintos de cero, se sigue que:

$$g(i) = 0, \quad \forall i.$$

Por lo tanto, $g(T) = 0$ casi seguramente, lo que implica que T es un estadístico **completo**

Proposición 3.6.1

Si T_1 y T_2 son estadísticos completos, entonces (T_1, T_2) también es completo

Ejemplo

Veamos si dado $X \sim N(\theta, \theta)$ el estadístico $T = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ es completo:
Por las propiedades de las esperanzas tenemos que:

$$\sum_{i=1}^n E[X_i] = n\theta, \quad \sum_{i=1}^n E[X_i^2] = n(E[X_i^2]) = n(\theta + \theta^2)$$

Sea $h(t) = (2(\sum_{i=1}^n X_i)^2 - (n+1) \sum_{i=1}^n X_i^2)$, entonces:

$$E_\theta[h(T)] = 2E \left[\left(\sum_{i=1}^n X_i \right)^2 \right] - (n+1)E \left[\sum_{i=1}^n X_i^2 \right] = 2E \left[\left(\sum_{i=1}^n X_i \right)^2 \right] - (n+1) \sum_{i=1}^n E[X_i^2]$$

Recordemos que $\text{Var}(Y) = E[Y^2] - (E[Y])^2$, entonces:

$$E \left[\left(\sum_{i=1}^n X_i \right)^2 \right] = \text{Var} \left(\sum_{i=1}^n X_i \right) + \left(E \left[\sum_{i=1}^n X_i \right] \right)^2 = n\theta^2 + n^2\theta^2 = n\theta^2(n+1)$$

$$E[X_i^2] = \text{Var}(X_i) + (E[X_i])^2 = \theta^2 + \theta^2 = 2\theta^2 \implies E_\theta \left[\sum_{i=1}^n X_i^2 \right] = 2n\theta^2$$

$$\implies E_\theta[h(T)] = 2n\theta^2(n+1) - (n+1)2n\theta^2 \implies E_\theta[h(T)] = 0, \forall \theta \in \Theta$$

Pero $h(T) = 2(\sum_{i=1}^n X_i)^2 - (n+1)\sum_{i=1}^n X_i^2 \neq 0$ por lo que T no es completo

Observación 3.6.1

Recordemos que:

- **Suficiencia:** Tiene toda la información sobre el parámetro θ pero puede tener información de más.
- **Complejitud:** No tiene información de más, pero puede no tener toda la información necesaria.
- **Minimalidad:** Tiene toda la información necesaria y no tiene información de más.

Teorema 3.6.1 [Teorema de Bahadur]

Un estadístico $T = T(X_1, \dots, X_n)$ es suficiente y completo \implies es minimal suficiente.

Demostración. Por hipótesis del ejercicio sabemos que el estadístico T es suficiente y completo. Además, supongamos que tenemos otro estadístico S suficiente y definamos un tercero $H = E[T|S]$.

Observación 3.6.2

Recordemos la propiedad de las esperanzas, propiedad de la torre de la expectativa o propiedad de la iteración de la esperanza que dice que:

$$E[E[T|\mathcal{G}]] = E[T]$$

Donde \mathcal{G} es una σ -álgebra.

No obstante, dado que mayormente se trabaja con estadísticos, podemos denotar:

$$E[E[T|S]] = E[T]$$

Donde S denota en realidad $\sigma(S)$ -álgebra (la menor de todas las generadas por S , en particular).

Definamos ahora, el estadístico L dado por $L = E[H|T] \implies$

$$\begin{cases} H = E[T|S] \implies E[H] = E[E[T|S]] = E[T] \\ L = E[H|T] \implies E[L] = E[E[H|T]] = E[H] \end{cases} \implies E[L] = E[H] = E[T]$$

Sea la función real $g(T) = T - E[H|T]$, entonces, intentemos calcular su esperanza:

$$E[g(T)] = E[T - E[H|T]] = E[T] - E[E[H|T]] = E[T] - E[H] = E[T] - E[T] = 0$$

Entonces, por la completitud de T , ésto implica que:

$$g(T) \stackrel{c.s.}{=} 0 \implies T = E[H|T]$$

Esto nos dice que tras intentar estimar T a través de $H = (T|S)$, volvemos a obtener T , por lo que contiene toda la información necesaria y no tiene información de más, por lo que es minimal suficiente. \square

Observación 3.6.3

El recíproco no se cumple, veamos un contraejemplo:

Sean $X \sim N(\mu, \sigma_X^2)$ e $Y \sim N(\mu, \sigma_Y^2)$ y tomemos el estadístico $T = (\bar{X}, \bar{Y}, S_X^2, S_Y^2)$. Veamos que el estadístico es minimal suficiente para $\theta = (\mu, \sigma_X^2, \sigma_Y^2)$, pero no es completo:

Por el Teorema de Caracterización de los estadísticos minimales, tenemos que:

$$\frac{f_\theta(x_1, \dots, x_n)}{f_\theta(y_1, \dots, y_n)} = \frac{\left(\frac{1}{\sigma_X \sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma_X}\right)^2}}{\left(\frac{1}{\sigma_Y \sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - \mu}{\sigma_Y}\right)^2}} = \left(\frac{\sigma_Y}{\sigma_X}\right)^n \cdot e^{-\frac{1}{2\sigma_X^2} \sum_{i=1}^n (x_i - \mu)^2 + \frac{1}{2\sigma_Y^2} \sum_{i=1}^n (y_i - \mu)^2} =$$

Sabiendo que $(x_i - \mu)^2 = (x_i^2 + \mu^2 - 2x_i\mu) \implies \sum (x_i - \mu)^2 = \sum x_i^2 + \sum \mu^2 - \sum 2x_i\mu = \sum x_i^2 + n\mu^2 - 2n\mu\bar{X}$
Ahora, desarrollemos $\sum x_i^2$:

$$\begin{aligned} \begin{cases} S^2 = \frac{1}{n} \sum (x_i - \bar{X})^2 \\ \bar{X} = \frac{1}{n} \sum x_i \end{cases} &\implies \begin{cases} \sum x_i^2 = n(S_X^2 + \bar{X}^2) \\ \sum y_i^2 = n(S_Y^2 + \bar{Y}^2) \end{cases} \implies \\ \implies \begin{cases} n(S_X^2 + \bar{X}^2) + n\mu^2 - 2n\mu\bar{X} \\ n(S_Y^2 + \bar{Y}^2) + n\mu^2 - 2n\mu\bar{Y} \end{cases} &\implies \begin{cases} n(S_X^2 + \bar{X}^2 + \mu^2 - 2\mu\bar{X}) \\ n(S_Y^2 + \bar{Y}^2 + \mu^2 - 2\mu\bar{Y}) \end{cases} \\ \frac{f_\theta(\vec{x})}{f_\theta(\vec{y})} &= \left(\frac{\sigma_Y}{\sigma_X}\right)^n \cdot e^{-\frac{n}{2\sigma_X^2} (S_X^2 + \bar{X}^2 + \mu^2 - 2\mu\bar{X}) + \frac{n}{2\sigma_Y^2} (S_Y^2 + \bar{Y}^2 + \mu^2 - 2\mu\bar{Y})} \end{aligned}$$

Pero esta expresión no depende de θ si y solo si $\bar{X} = \bar{Y}, S_X^2 = S_Y^2$ por lo que el estadístico T es minimal suficiente.

No obstante, si tomamos la función $g(x, y) = x - y$, entonces:

$$E[g(T)] = E[\bar{X} - \bar{Y}] = E[\bar{X}] - E[\bar{Y}] = \mu - \mu = 0$$

Sin embargo,

$$g(T) = \bar{X} - \bar{Y} \neq 0 \text{ c.s.}$$

Por lo que el estadístico T no es completo

Teorema 3.6.2

El estadístico natural $(\sum_{i=1}^n T_1(X_i), \dots, \sum_{i=1}^n T_k(X_i))$ de la familia exponencial k -paramétrica

$$\left\{ f_\theta(x) = c(\theta)h(x)e^{\sum_{j=1}^k q_j(\theta)T_j(x)} \right\}_{\theta \in \Theta \subset \mathbb{R}^k}$$

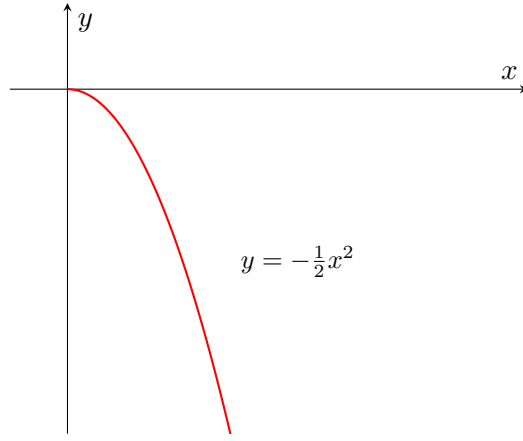
es **completo** si la imagen de la aplicación $q = (q_1(\theta), \dots, q_k(\theta)) : \Theta \rightarrow \mathbb{R}^k$ contiene un rectángulo abierto de \mathbb{R}^k no trivial.

Ejemplo

Recordando el ejemplo anterior en el que probabamos que la familia de distribuciones $N(\theta, \theta)$ no era completa, podemos ver que la imagen de la aplicación $q = \left(\frac{1}{\theta}, -\frac{1}{2\theta^2}\right)$ no contiene ningún rectángulo abierto de \mathbb{R}^2 :

$$q_2(\theta) = -\frac{1}{2}q_1(\theta)^2, \forall \theta > 0$$

Lo cual es una rama de parábola, y por lo tanto no contiene ningún abierto de \mathbb{R}^2



Ejemplo

Sea $X \sim \text{Bernoulli}(\theta) \implies P(X = x) = \theta^x(1 - \theta)^{1-x} \implies$

Sea una m.a.s. de tamaño $n \implies P(\vec{X} = \vec{x}) = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i} = \left(\frac{\theta}{1 - \theta}\right)^{\sum x_i} \cdot (1 - \theta)^n =$

En este caso, la distribución no tiene exponencial, pero se puede refactorizar para obtenerlo: Recordemos que $a^b = e^{b \cdot \ln(a)}$, entonces, aplicando ésto aquí obtenemos que:

$$= e^{\sum x_i \cdot \ln(\frac{\theta}{1-\theta})} \cdot (1 - \theta)^n$$

En donde tenemos por la forma general de las familias exponenciales k-paramétricas, que es de la forma:

- $T(\vec{x}) = \sum x_i$
- $q(\theta) = \ln(\frac{\theta}{1-\theta})$
- $c(\theta) = (1 - \theta)^n$
- $h(\vec{x}) = 1$

Entonces, podemos obtener que como (X_1, \dots, X_n) pertenece a la familia exponencial k-paramétrica, obtenemos que su estadístico natural $(T(\vec{x}) = \sum x_i)$ es suficiente. Y como

$$g(\theta) = \frac{\theta}{1 - \theta} \neq 0 \implies \text{el estadístico natural } \sum x_i \text{ es minimal}$$

Ya que un vector de una dimensión es siempre linealmente independiente si y sólo si no es nulo.

Por último, veamos que el estadístico natural es completo:

Para ello aplicaremos el teorema anterior y para ello analizaremos el conjunto imagen de $q(\theta)$:

$$\text{Im}(q(\theta)) = \{\ln(\frac{\theta}{1-\theta}) : \theta \in \Theta = (0, 1)\} \text{ ya que se trata de una distribución Bernoulli}$$

Por último dado que la función tiene una gráfica así:



La imagen coincide con los números reales \mathbb{R} por lo que contiene un conjunto abierto no trivial y por tanto el estadístico natural $\sum x_i$ es completo.

Ahora veamos otro ejemplo en el que no se puede provocar la pertenencia a la familia exponencial:

Ejemplo

Sea $X \sim U(\theta, 7\theta) \implies f_\theta(x) = \frac{1}{6\theta} \cdot I_{(\theta, 7\theta)}(x) \implies$

$$f_\theta(\vec{x}) = \left(\frac{1}{6\theta}\right)^n \cdot I_{(\theta, 7\theta)}(x_{(1)}) \cdot I_{(\theta, 7\theta)}(x_{(n)})$$

Entonces por el Teorema de Factorización de Fisher, tenemos que el estadístico $T(\vec{x}) = (x_{(1)}, x_{(n)})$ es suficiente.

Observación 3.6.4

Como la función indicadora depende de θ , entonces la distribución no pertenece a la familia exponencial

Recordemos además cuáles son las funciones de densidad de los máximos y lo mínimos (aunque en la práctica real hace falta demostrarlo siempre):

$$\begin{cases} f_\theta(x_{(n)}) = n(F_X(x))^{n-1} \cdot f_X(x) \\ f_\theta(x_{(1)}) = n(1 - F_X(x))^{n-1} \cdot f_X(x) \end{cases} \implies \text{Saquemos primero la esperanza del máximo:}$$

$$E_\theta[x_{(n)}] = \int_\theta^{7\theta} x \cdot n \left(\frac{x - \theta}{6\theta}\right)^{n-1} \cdot \frac{1}{6\theta} dx =$$

Para calcularla introduzcamos el cambio de variable, para usar la regla de la multiplicación:

$$\begin{aligned} \begin{cases} x = u \\ dv = f_\theta(x_{(n)}) \end{cases} &\implies \\ &= \left[x \cdot \left(\frac{x - \theta}{6\theta}\right)^n \right]_{x=\theta}^{7\theta} - \int_\theta^{7\theta} \left(\frac{x - \theta}{6\theta}\right)^n dx = 7\theta - \int_{y=0}^{y=1} y^n \cdot 6\theta dy = 7\theta - 6\theta \cdot \frac{1}{n+1} = \theta \frac{7n+1}{n+1} \theta \end{aligned}$$

Ahora, realicemoslo para el mínimo:

$$E_\theta[x_{(1)}] = \int_\theta^{7\theta} x \cdot n \left(1 - \frac{x - \theta}{6\theta}\right)^{n-1} \cdot \frac{1}{6\theta} dx = \int_\theta^{7\theta} x \cdot n \left(\frac{7\theta - x}{6\theta}\right)^{n-1} \cdot \frac{1}{6\theta} dx =$$

Al igual que antes introduzcamos un cambio de variable para calcularla:

$$= \left[x \cdot \left(1 - \frac{x - \theta}{6\theta} \right)^n \right]_{x=\theta}^{7\theta} - \int_{\theta}^{7\theta} \left(\frac{7\theta - x}{6\theta} \right)^n dx = -\theta - \int_1^0 y^n (-6\theta) dy = -\theta - 6\theta \left[\frac{y^{n+1}}{n+1} \right]_0^1 = -\theta - 6\theta \frac{1}{n+1}$$

HAY QUE TERMINAR ESTE EJEMPLO

Teorema 3.6.3 [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. Sea la muestra (x_1, \dots, x_n) , un parámetro θ y sea un estadístico T suficiente, entonces por definición tenemos que:

$$f(\vec{x}|T=t) \text{ es independiente de } \theta \implies$$

$$f(U=u|T=t) = \sum_{\vec{x}: U(\vec{x})=u} f(\vec{x}|T=t) \text{ es independiente de } \theta$$

Aplicando las propiedades de los condicionales, tenemos que:

$$f_{\theta}(u) = f_{\theta}(t)f_{\theta}(u|t) \text{ pero como } U \text{ es ancilar y } T \text{ es completo, entonces: } \implies f(u) = f_{\theta}(t)f(u|t)$$

Tomemos la función $h(t) = f(u|t) - f(u)$, entonces:

$$E[h(T)] = E[f(u|T) - f(u)] = E[f(u|T)] - E[f(u)] = \int f(u|t)f(t)dt - \int f(u)f(t)dt = f(u) - f(u) = 0$$

Entonces como T es completo, tenemos que $h(T) = 0$ c.s. $\implies f(u|t) = f(u) \implies$

$$f(u|t) = \frac{f(u, t)}{f(t)} \iff f(t, u) = f(u)f(t) \implies U \text{ y } T \text{ son independientes}$$

□

Ejemplo

Sea $X \sim U(0, \theta)$, veamos que $X_{(n)}$ y $\frac{X_{(1)}}{X_{(n)}}$ son independientes:

$$\begin{aligned} X \sim U(0, \theta) &\implies f_{\theta}(x) = \frac{1}{\theta} \cdot I_{(0, \theta)}(x) \implies F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = \\ &= P(X_1 \leq x) \cdot \dots \cdot P(X_n \leq x) = (F(x))^n = \left(\frac{x}{\theta}\right)^n \implies f_{\theta}(x) = n \left(\frac{x}{\theta}\right)^{n-1} \cdot \frac{1}{\theta} = \frac{n}{\theta^n} x^{n-1} \end{aligned}$$

Dado que por el Teorema de Factorización de Fisher, esta distribución es suficiente, consideraremos que $X_{(n)}$ es el estadístico suficiente y completo.

Veamos ahora la función de densidad de $\frac{X_{(1)}}{X_{(n)}}$, para ello calculemos la función de densidad de $X_{(1)}$:

$$\begin{aligned} F_{X_{(1)}}(x) &= P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x) = 1 - P(X_1 > x, \dots, X_n > x) = \\ &= 1 - P(X_1 > x) \cdot \dots \cdot P(X_n > x) = 1 - (1 - F(x))^n = 1 - \left(1 - \frac{x}{\theta}\right)^n \implies f_{\theta}(x) = n \left(1 - \frac{x}{\theta}\right)^{n-1} \cdot \frac{1}{\theta} = \frac{n}{\theta^n} (\theta - x)^{n-1} \end{aligned}$$

Ejemplo

Sea una m.a.s. de tamaño n de una población $X \sim N(\theta, \sigma)$ se tiene que:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{n}\right) \text{ ya que } \begin{cases} E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} n\theta = \theta \\ \text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} n\sigma = \frac{\sigma}{n} \end{cases}$$

Y como se trata de una distribución perteneciente a la familia exponencial 2-paramétrica, entonces \bar{X} es suficiente. Queda demostrar que sea completo, para ello, consideremos la función de densidad de \bar{X} :

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{n}\right) \implies f_{\theta}(x) = \sqrt{\frac{n}{2\pi\sigma}} e^{-\frac{n}{2\sigma}(x-\theta)^2}$$

La completitud se da por definición. No obstante es una integral demasiado difícil de calcular, por lo que se puede deducir que \bar{X} es completo. Por el Teorema de Fisher sabemos que:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \implies S^2 \sim \frac{\sigma^2}{n-1} \chi^2(n-1) \implies S^2 \text{ es un estadístico ancilario} \implies$$

Por el teorema de Basu, \bar{X} y S^2 son independientes

3.7 Principios de reducción de datos

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

Definición 3.7.1 [Función de verosimilitud]

Dada la familia $\mathcal{P} = \{f(\vec{x}|\theta) | \theta \in \Theta\}$ de distribuciones de probabilidad donde $f(\cdot|\theta)$ es una función de densidad o de masa y supuesto que se ha observado el valor muestral $\vec{x} = (x_1, \dots, x_n)$ la función de θ definida mediante

$$L(\theta|\vec{x}) = f(\vec{x}|\theta)$$

se llama función de verosimilitud.

Definición 3.7.2 [Principio de verosimilitud]

Sea $E = (\chi, \theta, \mathcal{P} = \{f(\vec{x}|\theta) : \theta \in \Theta\})$ un experimento aleatorio y $L_1(\theta|x)$ y $L_2(\theta|y)$ las funciones de verosimilitud de dos muestras $X = (x_1, \dots, x_n)$ e $Y = (y_1, \dots, y_m)$ respectivamente.

Si existe una función $c(x, y)$ que no depende de θ tal que:

$$L_1(\theta|x) = c(x, y) L_2(\theta|y) \quad \forall \theta \in \Theta$$

entonces se dice que ambas muestras suministran la misma evidencia estadística sobre el parámetro θ , es decir, $Ev(E_1, \vec{x}) = Ev(E_2, \vec{y})$.

Ejemplo

Sea un modelo $\text{Binomial}(n, \theta)$ la evidencia que se obtiene del mismo, sobre θ , cuando se observan t éxitos en n repeticiones, es la misma que la obtenida por un modelo $\text{Binomialnegativo}(t, \theta)$, siempre que se suponga que se han realizado n repeticiones hasta obtener t éxitos.

$$\text{En el primer caso tenemos que: } f(t|\theta) = \binom{n}{t} \theta^t (1-\theta)^{n-t}$$

y en el segundo caso: $g(n \text{ repeticiones hasta obtener } t \text{ éxitos} | \theta) = \binom{n-1}{t-1} \theta^t (1-\theta)^{n-t}$

Por tanto, se puede aplicar el principio de verosimilitud, tomando la constante $c(t, n) = \frac{n}{t}$. Es decir, si para estimar la probabilidad con la que sale cara una moneda, se tira ésta n veces, y es t el número de caras que se han presentado (modelo binomial), se obtiene la misma evidencia para θ , que si se repite el experimento hasta observar t caras y para ello ha habido que realizar n repeticiones (modelo binomial negativo)

Definición 3.7.3 [Principio de suficiencia]

En un experimento $E = (\chi, \theta, \mathcal{P} = \{f(\vec{x}|\theta) : \theta \in \Theta\})$, si $T = T(\vec{X})$ es un estadístico suficiente para θ y se tiene que $T(\vec{x}) = T(\vec{y})$, entonces las dos muestras \vec{x} e \vec{y} suministran la misma evidencia estadística, es decir, $Ev(E, \vec{x}) = Ev(E, \vec{y})$

Ejemplo

En un experimento consistente en la repetición de n pruebas de Bernoulli, la evidencia estadística que se obtiene de dos puntos muestrales con el mismo número de éxitos es la misma.

Definición 3.7.4 [Principio de condicionalidad]

El principio de condicionalidad dice que si un mecanismo aleatorio no depende del valor a determinar θ , no proporciona evidencia sobre él. Es decir, si $T = T(\vec{X})$ es un estadístico suficiente para θ y $T(\vec{x}) = T(\vec{y})$, entonces las dos muestras \vec{x} e \vec{y} suministran la misma evidencia estadística, es decir, $Ev(E, \vec{x}) = Ev(E, \vec{y})$

Ejemplo

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

Teorema 3.7.1 [Teorema de Birnbaum]

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

3.8 Ejercicios

Ejercicio 2.1. Sea X muestra de una población $N(0, \sigma)$, ¿Es $T(X) = |X|$ un estadístico suficiente?

Solución: Recordemos la definición de estadístico suficiente por definición: Dado un estadístico T en este caso $T(X) = |X|$ y una muestra \vec{x} se dice que T es suficiente si $f_\theta(\vec{X} = \vec{x} | T = t)$

Empecemos definiendo la distribución que sigue el estadístico T :

$$P(|X| \leq a) = P(-a \leq X \leq a) = P(X \leq a) - P(X \leq -a) = F(a) - F(-a) = 2F(a) = 2 \int_0^a f_X(x) dx$$

$$\implies f_{|X|}(x) = 2f_X(x) = 2 \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \implies \text{Si } X \sim N(0, \sigma) \implies f_{|X|}(x) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}}$$

Entonces veamos el cociente de las dos funciones de densidad:

$$P(X = x|T = t) = \frac{f_\sigma(x)}{f_\sigma(t)} = \frac{\frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}}}{\frac{\sqrt{2}}{\sigma\sqrt{\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}}} = \frac{1}{2}$$

Lo cual no depende del parámetro σ y por tanto por definición, el estadístico es suficiente. \square

Ejercicio 2.2. Encuentra un estadístico suficiente para una m.a.s. de tamaño n de cada una de las siguientes poblaciones:

1. Beta(α, β) : $\alpha > 0, \beta > 0$
2. Gamma(a, p) : $a > 0, p > 0$
3. $p_{N_1 N_2}(x) = \frac{1}{N_2 - N_1} \cdot I_{N_1+1, N_1+2, \dots, N_2}(x)$: $N_1 < N_2$
4. $f_\theta(x) = e^{-x+\theta} \cdot I_{\theta, \infty}(x)$
5. $f_{\mu, \sigma}(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln(x-\mu))^2} \cdot I_{0, \infty}$
6. $p_{\theta, p}(x) = (1-p)p^{x-\theta} \cdot I_{\theta, \theta+1, \dots}(x)$

Solución:

1. Sea $X \sim \text{Beta}(\alpha, \beta) \implies$

$$f_{\alpha, \beta}(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \implies f_{\alpha, \beta}(x_1, \dots, x_n)(x) = \prod_{i=1}^n f_{\alpha, \beta}(x_i) = \frac{(\prod_{i=1}^n x_i)^{\alpha-1} (\prod_{i=1}^n (1-x_i))^{\beta-1}}{B(\alpha, \beta)^n}$$

Entonces por el Teorema de Factorización de Fisher, tenemos que

$$f_\theta(x) = h(x) \cdot g_\theta(T(x)) \implies T(\vec{x}) = (\prod_{i=1}^n x_i, \prod_{i=1}^n (1-x_i))$$

2. Sea $X \sim \text{Gamma}(\alpha, \beta) \implies$

$$f_{\alpha, \beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \cdot e^{-x\beta} \implies f_{\alpha, \beta}(x_1, \dots, x_n) = \prod_{i=1}^n f_{\alpha, \beta}(x_i) = \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} (\prod_{i=1}^n x_i)^{\alpha-1} \cdot e^{-\beta \sum_{i=1}^n x_i}$$

Entonces por el Teorema de Factorización de Fisher:

$$f_\theta = h(x) \cdot g_\theta(T(x)) \implies T(\vec{x}) = (T_1(\vec{x}), T_2(\vec{x})) = (\sum_{i=1}^n x_i, \prod_{i=1}^n x_i)$$

3. Sea $X \sim \text{Uniforme Discreta}(N_1, N_2) \implies$

$$\begin{aligned} f_{(N_1, N_2)}(x) &= \frac{1}{N_2 - N_1} \cdot I_{N_1+1, N_1+2, \dots, N_2}(x) \implies f_{(N_1, N_2)}(x_1, \dots, x_n) = \prod_{i=1}^n f_{(N_1, N_2)}(x_i) = \\ &= \frac{1}{(N_2 - N_1)^n} \cdot I_{\{N_1+1, N_1+2, \dots, +\infty\}}(x_{(1)}) \cdot I_{\{-\infty, \dots, N_2-1, N_2\}}(x_{(n)}) \end{aligned}$$

Por el Teorema de Factorización de Fisher, tenemos que:

$$f_\theta(x) = h(x) \cdot g_\theta(T(x)) \implies T(\vec{x}) = (x_{(1)}, x_{(n)})$$

4. Sea una variable aleatoria con función de densidad $f_\theta(x) = e^{-x+\theta} \cdot I_{\theta,\infty}(x) \implies$

$$f_\theta(x_1, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n e^{-x_i+\theta} \cdot I_{\theta,\infty}(x_i) = e^{-\sum_{i=1}^n x_i + n\theta} \cdot I_{\theta,\infty}(x_{(1)})$$

Por el Teorema de Factorización de Fisher, tenemos que:

$$f_\theta(x) = h(x) \cdot g_\theta(T(x)) \implies T(\vec{x}) = (x_{(1)})$$

5. Sea una variable aleatoria X cuya función de densidad es:

$$f_{(\mu,\sigma)}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}(\ln(x)-\mu)^2} \cdot I_{0,\infty}(x) \implies f_{(\mu,\sigma)}(\vec{x}) = \prod_{i=1}^n \frac{1}{x_i} \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \cdot e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (\ln(x_i)-\mu)^2}$$

Desarrollando el cuadrado: $(\ln(x_i) - \mu)^2 = \ln^2(x_i) - 2\mu\ln(x_i) + \mu^2$ y conociendo que la forma general de las funciones de densidad de las familias exponenciales 2-paramétricas es:

$$f_{(\mu,\sigma)}(x) = c(\mu, \sigma) \cdot h(x) \cdot e^{q_1(\mu,\sigma)T_1(x) + q_2(\mu,\sigma)T_2(x)} \implies f_{(\mu,\sigma)}(\vec{x}) = c(\mu, \sigma)^n \cdot h(\vec{x}) \cdot e^{q_1(\mu,\sigma) \sum T_1(x_i) + q_2(\mu,\sigma) \sum T_2(x_i)}$$

Podemos ver que $T_1(\vec{x}) = \sum_{i=1}^n \ln(x_i)$ y $T_2(\vec{x}) = \sum_{i=1}^n \ln^2(x_i)$

Observación 3.8.1

Esta distribución en particular se la conoce como log-normal o distribución de Tinaut y surge de tomar una distribución normal $N(\mu, \sigma)$ y aplicarle la función logarítmica, es decir: $Y = \ln(X) \sim N(\mu, \sigma)$

6. Sea una variable aleatoria X cuya función de densidad es:

$$f_{(\theta,p)}(x) = (1-p)p^{x-\theta} \cdot I_{\theta,\theta+1,\dots}(x) \implies f_{(\theta,p)}(\vec{x}) = (1-p)^n \cdot p^{\sum_{i=1}^n x_i - n\theta} \cdot I_{\theta,\theta+1,\dots}(x_{(1)})$$

Entonces por el Teorema de Factorización de Fisher, tenemos que:

$$f_\theta(x) = h(x) \cdot g_\theta(T(x)) \implies T(\vec{x}) = \left(\sum_{i=1}^n x_i, x_{(1)} \right)$$

□

Ejercicio 2.3. Sea una m.a.s. de tamaño n de una población con función de distribución $X \sim \text{Uniforme}(\theta, 4\theta)$ pruébese que $(X_{(1)}, X_{(n)})$ es un estadístico suficiente para θ pero no es completo.

Solución:

$$X \sim U(\theta, 4\theta) \implies f_\theta(x) = \frac{1}{3\theta} \cdot I_{(\theta,4\theta)}(x) \implies f_\theta(x_1, \dots, x_n) = \frac{1}{(3\theta)^n} \cdot I_{(\theta,4\theta)}(x_{(1)}) \cdot I_{(\theta,4\theta)}(x_{(n)})$$

Entonces, por el Teorema de Factorización de Fisher, tenemos que: $f_\theta(x) = h(x) \cdot g_\theta(T(x)) \implies T(\vec{x}) = (x_{(1)}, x_{(n)})$ Ahora veamos que el estadístico no es completo: Para ello tenemos que sacar cuales son las distribuciones del mínimo y el máximo:

1.

$$\begin{aligned} f_\theta(x_{(n)}) &= n(F(x))^{n-1} \cdot f_\theta(x_{(n)}) = n \left(\frac{x-\theta}{3\theta} \right)^{n-1} \cdot \frac{1}{3\theta} = \frac{n}{3\theta^n} (x-\theta)^{n-1} \implies \\ E[T_1] &= \int_{\theta}^{4\theta} x \cdot n \left(\frac{x-\theta}{3\theta} \right)^{n-1} \cdot \frac{1}{3\theta} dx = \left[x \cdot \left(\frac{x-\theta}{3\theta} \right)^n \right]_{\theta}^{4\theta} - \int_{\theta}^{4\theta} \left(\frac{x-\theta}{3\theta} \right)^n dx = 4\theta - 3\theta \int_0^1 y^n dy = \\ &= 4\theta - 3\theta \frac{1}{n+1} = \theta \frac{4n+1}{n+1} \end{aligned}$$

2.

$$f_{\theta}(x_{(1)}) = n(1 - F(x))^{n-1} \cdot f_{\theta}(x) = n \left(1 - \frac{x - \theta}{3\theta}\right)^{n-1} \cdot \frac{1}{3\theta} = \frac{n}{3\theta^n} (4\theta - x)^{n-1} \implies$$

$$E[T_2] = \int_{\theta}^{4\theta} x \cdot n \left(1 - \frac{x - \theta}{3\theta}\right)^{n-1} \cdot \frac{1}{3\theta} dx = \left[x \cdot - \left(1 - \frac{x - \theta}{3\theta}\right)^n \right]_{\theta}^{4\theta} - \int_{\theta}^{4\theta} \left(\frac{4\theta - x}{3\theta}\right)^n dx =$$

$$= \theta - 3\theta \int_0^1 y^n dy = \theta - 3\theta \frac{1}{n+1} = \theta \frac{4n+1}{n+1}$$

□

Ejercicio 2.4. Sea una m.a.s. de una población que sigue una variable aleatoria X con función de densidad:

$$f_{\theta}(x) = e^{-(x-\theta)} \cdot I_{(\theta, +\infty)}(x)$$

1. Pruébese que $T(\vec{X}) = X_{(1)}$ es un estadístico suficiente y completo para θ
2. Mediante el Teorema de Basu, pruébese que $X_{(1)}$ y S^2 son independientes.

Solución: Antes de comenzar con la resolución del problema, calculemos la función de densidad de la muestra:

$$f_{\theta}(\vec{x}) = e^{-\sum_{i=1}^n x_i + n\theta} \cdot I_{\theta, +\infty}(x_{(1)})$$

1. Para demostrar la suficiencia usaremos el Teorema de Factorización de Fisher con: $f_{\theta}(\vec{x}) = h(\vec{x}) \cdot g_{\theta}(T(\vec{x})) = e^{-\sum_{i=1}^n x_i + n\theta} \cdot I_{\theta, +\infty}(x_{(1)})$ y $T(\vec{x}) = (x_{(1)}, \sum x_i)$
Ahora veamos la completitud: Sea una función real $g(x)$ que cumpla que:

$$E[g(x_{(1)})] = 0 = \int_{\theta}^{+\infty} g(x) f_{\theta}(x) dx = \int_{\theta}^{+\infty} g(x) e^{-x+\theta} dx$$

Si tomamos como función $H(x) = \int_{t>0}^{+\infty} e^{-y}$ sabemos que dicha función es positiva en todo su dominio, por lo que la única manera de que la integral anterior sea nula es que $g(x) = 0$ c.s. \implies el estadístico es completo.

2. El Teorema de Basu, nos dice que si T es un estadístico suficiente y completo y U es un estadístico ancilario, entonces T y U son independientes. En este caso, sabemos que $X_{(1)}$ es completo y suficiente, y por el Teorema de Basu, sabemos que $X_{(1)}$ y S^2 son independientes. Entonces debemos comprobar que S^2 es ancilario:

$$f_{\theta} = e^{-(x-\theta)} \implies X \sim \text{Exponencial}(1) + \theta \implies S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 =$$

$$= \frac{1}{n-1} \sum_{i=1}^n (X_i - \theta + \theta - \bar{X})^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \theta - (\bar{X} - \theta))^2 = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})^2 \implies$$

$$\implies Z \sim \text{Exponencial}(1) \text{ no depende de } \theta \implies S^2 \text{ es ancilario}$$

Por tanto por el Teorema de Basu, ambos estadísticos son independientes.

□

Ejercicio 2.5. Sea una población que sigue una variable aleatoria X con funciones de masa:

$$P_{\theta}(X = -1) = \theta \quad P_{\theta}(X = x) = (1 - \theta)^2 \cdot \theta^x \cdot I_{0 \cup \mathbb{N}}(x)$$

Con $\theta \in (0, 1)$ y el tamaño de la muestra $n = 1$ ¿El estadístico X es completo?

Solución:

$$\begin{aligned}
 E[X] &= \sum_{i=0}^{+\infty} P(X=i) \cdot i = -1 \cdot P(X=-1) + \sum_{i=1}^{+\infty} P_{\theta}(X=i) \cdot i = -1 \cdot \theta + \sum_{i=1}^n (1-\theta)^2 \cdot i\theta^i = \\
 &= -\theta + (1-\theta) \sum_{i=1}^n i\theta^i = -\theta + (1-\theta)t'(x) \text{ donde } t(x) = \sum_{i=1}^{\infty} \theta^i = \frac{\theta}{1-\theta} \implies t'(x) = \frac{1}{(1-\theta)^2} \\
 &\implies E[X] = -\theta + (1-\theta)^2 \theta \left(\frac{1}{(1-\theta)^2} \right) = -\theta + \theta = 0
 \end{aligned}$$

Entonces, como $E[X] = 0$, X no es completo. \square

Ejercicio 2.6. Sea m.a.s. de tamaño n de una población $N(\alpha\sigma, \sigma) : \alpha \in \mathbb{R}$ veamos si el estadístico $T(\vec{X}) = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ es suficiente y completo para σ .

Solución:

$$X \sim N(\alpha\sigma, \sigma) \implies f_{\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\alpha\sigma)^2}{2\sigma^2}} \implies f_{\sigma}(\vec{x}) = \prod_{i=1}^n f_{\sigma}(x_i) = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \cdot e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \alpha\sigma)^2}$$

Entonces descomponiendo: $(x_i - \alpha\sigma)^2 = x_i^2 + \alpha^2\sigma^2 - 2\alpha\sigma x_i \implies$

$$f_{\sigma}(\vec{x}) = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \cdot e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 + \frac{\alpha}{\sigma} \sum_{i=1}^n x_i - \frac{n\alpha^2}{2}}$$

Entonces por el Teorema de Factorización de Fisher, tenemos que el estadístico $T(\vec{x}) = (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ es suficiente.

Ahora veamos si es completo:

$$E \left[\sum_{i=1}^n X_i \right] = n\alpha\sigma \quad E \left[\sum_{i=1}^n X_i^2 \right] = n\sigma^2 + n\alpha^2\sigma^2 = (1 + n\alpha^2)n\sigma^2$$

Pero la función

$$g(\vec{T}) = \frac{1}{1 + n\alpha^2} \left(\sum_{i=1}^n X_i \right)^2 - \frac{1}{1 + \alpha^2} \sum_{i=1}^n X_i^2 \implies E[g(\vec{T})] = 0$$

Pero la función no es nula en casi todo punto, por lo que no es completo. \square

Ejercicio 2.7. Sea una población con variable aleatoria $X \sim Uniforme(\theta, \theta+1)$ y una m.a.s. de tamaño n de dicha población. Entonces,

1. Encuentra un estadístico suficiente para θ
2. Comprueba si el recorrido muestral $R = X_{(n)} - X_{(1)}$ es un estadístico ancilar
3. Si la función de distribución sufriera un desplazamiento de $F_{\theta}(x) = F(x - \theta)$ Comprueba si se conserva la propiedad del apartado anterior

Solución: 1

1. $X \sim Uniforme(\theta, \theta+1) \implies f_{\theta}(x) = \frac{1}{\theta+1-\theta} \cdot I_{(\theta, \theta+1)} = 1 \cdot I_{(\theta, \theta+1)}(x) \implies f_{\theta}(\vec{x}) = I_{(\theta, \theta+1)}(x_{(1)}) \cdot I_{(\theta, \theta+1)}(x_{(n)}) \implies$ por el Teorema de Factorización de Fisher, el estadístico $T(\vec{x}) = (x_{(1)}, x_{(n)})$ es suficiente.
2. Calculemos las funciones de distribución del máximo y el mínimo:

•

$$f_{\theta}(x_{(n)}) = n(F(x))^{n-1} \cdot f_{\theta}(x_{(n)}) = n \left(\frac{x - \theta}{1} \right)^{n-1} \cdot 1 = n(x - \theta)^{n-1}$$

•

$$f_{\theta}(x_{(1)}) = n(1 - F(x))^{n-1} \cdot f_{\theta}(x_{(1)}) = n \left(1 - \frac{x - \theta}{1} \right)^{n-1} \cdot 1 = n(\theta + 1 - x)^{n-1} \implies$$

Cuando intentamos calcular la probabilidad $P(R \leq r)$ estamos calculando la probabilidad de que el máximo y el mínimo estén a menos de una distancia r , entonces:

$$P(R \leq r) = \int_{\theta}^{\theta+1} f_{X_{(n)}}(x) \cdot P(x - X_{(1)} \leq r) dx$$

Intuitivamente, esta expresión nos dice la probabilidad de que un punto x del intervalo sea el máximo por la probabilidad de que la distancia de ese número menos el mínimo esté a una distancia menor que $r \implies$

$$\begin{aligned} P(R \leq r) &= \int_{\theta}^{\theta+1} n(x - \theta)^{n-1} \cdot P(x - X_{(1)} \leq r) dx = \int_{\theta}^{\theta+1} n(x - \theta)^{n-1} \cdot P(X_{(1)} \geq x - r) dx = \\ &= \int_{\theta}^{\theta+1} n(x - \theta)^{n-1} \cdot (1 - P(X_{(1)} \leq x - r)) dx = \int_{\theta}^{\theta+1} n(x - \theta)^{n-1} \cdot (\theta + 1 - x + r)^n dx = \\ &= \int_0^1 ny^{n-1} \cdot (1 + r - y)^n dy \end{aligned}$$

Donde $y = x - \theta$ y $r = r - \theta$ y dado que es una expresión que no depende de θ llegamos a que el estadístico es ancilario.

- Si tenemos en cuenta el caso general del desplazamiento en una distribución uniforme:

$$F_{\theta}(x) = F(x - \theta) \implies P : \theta(r) = P(X_{(n)} - X_{(1)} \leq r) \implies$$

Si hacemos el cambio de varibale $Z_i = X_i - \theta$ obtenemos que $Z_{(n)} - Z_{(1)} = X_{(n)} - X_{(1)}$ y por tanto, el recorrido muestral sigue siendo un estadístico ancilario.

□

Ejercicio 2.8. Dada una m.a.s. de tamaño n de una población donde se sigue la siguiente función de densidad:

$$f_{\theta}(x) = \frac{2x}{\theta^2} \cdot I_{(0,\theta)}(x)$$

Entonces:

1. Comprueba si $T(\vec{X}) = (X_{(n)}, \prod_{i=1}^n X_i)$ es suficiente para θ y si además es minimal.
2. Comprueba también si $T_2(\vec{x}) = X_{(n)}$ es un estadístico completo.

Solución:

- 1.

$$f_{\theta}(x) = \frac{2x}{\theta^2} \cdot I_{(0,\theta)}(x) \implies f_{\theta}(\vec{x}) = \prod_{i=1}^n f_{\theta}(x_i) = \frac{2^n}{\theta^{2n}} \cdot \prod_{i=1}^n x_i \cdot I_{(0,\theta)}(x_{(n)})$$

Entonces por el Teorema de Factorización de Fisher, el estadístico $T(\vec{x}) = (X_{(n)}, \prod_{i=1}^n X_i)$ es suficiente. Para comprobar si es minimal, veamos que:

Dadas dos muestras \vec{x} y $\vec{y} \implies$

$$\frac{f_{\theta}(\vec{x})}{f_{\theta}(\vec{y})} = \frac{\frac{2^n}{\theta^{2n}} \cdot \prod_{i=1}^n x_i \cdot I_{(0,\theta)}(x_{(n)})}{\frac{2^n}{\theta^{2n}} \cdot \prod_{i=1}^n y_i \cdot I_{(0,\theta)}(y_{(n)})} = \frac{\prod_{i=1}^n x_i \cdot I_{(0,\theta)}(x_{(n)})}{\prod_{i=1}^n y_i \cdot I_{(0,\theta)}(y_{(n)})}$$

Entonces, el cociente no depende de $\theta \iff T(\vec{x}) = T(\vec{y})$ por lo que el estadístico es minimal.

2. Comprobemos que el estadístico $T_2(\vec{x}) = X_{(n)}$ es completo:
 Cómo es un máximo la función de densidad sigue una fórmula:

$$f_{X_{(n)}} = n(F(x))^{n-1} \cdot f_X(x) = n \left(\frac{x^2}{\theta^2} \right)^{n-1} \cdot \frac{2x}{\theta^2} = 2n \left(\frac{x^{2n-1}}{\theta^{2n}} \right)$$

Entonces, para una función real $g(x)$ queremos ver si que $E[g(X_{(n)})] = 0$ implica que $g(x) = 0$ c.s.

$$E[g(X_{(n)})] = \int_0^\theta g(x) 2n \left(\frac{x^{2n-1}}{\theta^{2n}} \right) dx = \frac{2n}{\theta^{2n}} \int_0^\theta g(x) \cdot x^{2n-1} dx = 0 \implies$$

Aplicando la Regla de Leibniz para derivadas de integrales:

$$\frac{d}{d\theta} \int_{a(\theta)}^{b(\theta)} f(x, \theta) dx = f(b(\theta), \theta) \cdot \frac{d}{d\theta}(b(\theta)) - f(a(\theta), \theta) \cdot \frac{d}{d\theta}(a(\theta)) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial \theta} f(\theta, x) dx$$

En nuestro caso, podemos sustituir usando:

- $f(x, \theta) = g(x) \cdot x^{2n-1}$
- $b(\theta) = \theta$
- $a(\theta) = 0$

Entonces tenemos que:

$$\begin{aligned} \frac{d}{d\theta} \int_0^\theta g(x) \cdot x^{2n-1} dx &= g(\theta) \cdot \theta^{2n-1} - g(0) \cdot 0^{2n-1} + \int_0^\theta \frac{\partial}{\partial \theta} (g(x) \cdot x^{2n-1}) dx = \\ &= g(\theta) \cdot \theta^{2n-1} = 0 \implies g(\theta) = 0 \implies g(x) = 0 \text{ c.s.} \end{aligned}$$

Entonces el estadístico $T_2(\vec{x}) = X_{(n)}$ es completo.

□

4 Estimación Puntual Paramétrica

Definición 4.0.1 [Estimador]

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 $(\chi, \mathcal{B}, F_\theta)_{\theta \in \Theta \subset \mathbb{R}^l}$, $\mathcal{B} = \mathcal{B}(\mathbb{R})$, y (X_1, \dots, X_n) m.a.s. $(n) \sim X$

Un estimador del parámetro θ es un estadístico $T = T(X_1, \dots, X_n) : \chi^n \rightarrow \Theta$ que se utiliza para determinar el valor desconocido θ

Definición 4.0.2 [Estimador centrado o insesgado]

Dado un estimador $T = T(X_1, \dots, X_n) : \chi^n \rightarrow \Theta$ se dice que es centrado para $h(\theta)$ cuando $E_\theta[T] = h(\theta)$. Se dice asintóticamente insesgado o centrado cuando $\lim_{n \rightarrow \infty} E_\theta[T] = h(\theta)$

Ejemplo

Para cualquier población que tenga media, la media muestral \bar{X} es un estimador centrado de $h(\theta) = E[X]$, puesto que se cumple:

$$E_\theta[\bar{X}] = E[X]$$

Además, se tiene que $E[a_k] = \alpha_k$, lo que implica que el momento muestral de orden k respecto al origen es un estimador centrado del momento poblacional de orden k respecto al origen de la población. En general, no es cierto un resultado análogo para momentos centrales.

Definición 4.0.3 [Sesgo]

Se llama sesgo de un estimador a la diferencia $b(T, \theta) = E_\theta[T] - \theta$

Ejemplo

Veamos que el estadístico $T = (\bar{X}, S^2)$ es un estimador centrado de $\theta = (\mu, \sigma^2)$:

$$\begin{aligned} E[\bar{X}] &= E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} \sum_{i=1}^n \mu = \mu \\ E[S^2] &= E\left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right] = \frac{1}{n-1} \sum_{i=1}^n E[(X_i - \bar{X})^2] = \frac{1}{n-1} \sum_{i=1}^n \sigma^2 = \sigma^2 \end{aligned}$$

Ejemplo

Demuestra que el estadístico $\sigma_n^2 = b_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ es un estimador centrado de $h(\theta) = \frac{n-1}{n} \sigma^2$ y $b(\sigma_n^2, \sigma^2) = -\frac{\sigma^2}{n}$

$$\begin{aligned} E[\sigma_n^2] &= E\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right] = \frac{1}{n} \sum_{i=1}^n E[(X_i - \bar{X})^2] = \\ &= \frac{1}{n} \sum_{i=1}^n E[X_i^2 - 2X_i\bar{X} + \bar{X}^2] = \frac{1}{n} \sum_{i=1}^n E[X_i^2] - 2E[X_i\bar{X}] + E[\bar{X}^2] = \\ &= \frac{1}{n} \sum_{i=1}^n E[X_i^2] - \frac{2}{n} \sum_{i=1}^n E[X_i\bar{X}] + \frac{1}{n} \sum_{i=1}^n E[\bar{X}^2] = \end{aligned}$$

Sabemos que: $\begin{cases} Var(\bar{X}) = E[\bar{X}^2] - E[\bar{X}]^2 \iff \frac{\sigma^2}{n} = E[\bar{X}^2] - \mu^2 \iff E[\bar{X}^2] = \frac{\sigma^2}{n} + \mu^2 \\ Var(X_i) = E[X_i^2] - E[X_i]^2 \iff \sigma^2 = E[X_i^2] - \mu^2 \iff E[X_i^2] = \sigma^2 + \mu^2 \end{cases}$

$$= \frac{1}{n}n(\sigma^2 + \mu^2) + \frac{1}{n}n\left(\frac{\sigma^2}{n} + \mu^2\right) - \frac{2}{n} \sum_{i=1}^n E[X_i \bar{X}] =$$

Ahora desarrollemos el término que falta:

$$\begin{aligned} E[X_i \bar{X}] &= E\left[X_i \frac{1}{n} \sum_{j=1}^n X_j\right] = \frac{1}{n} \sum_{j=1}^n E[X_i X_j] = \frac{1}{n} \sum_{j=1, j \neq i}^n E[X_i X_j] + \frac{1}{n} E[X_i^2] = \\ &= \frac{1}{n} \sum_{j=1, j \neq i}^n E[X_i] E[X_j] + \frac{1}{n} E[X_i^2] = \frac{1}{n}(n-1)\mu^2 + \frac{1}{n}(\sigma^2 + \mu^2) = \mu^2 + \frac{\sigma^2}{n} \\ \implies &= 2\mu^2 + \sigma^2\left(1 + \frac{1}{n}\right) - \frac{2}{n} \left(\sum_{i=1}^n \mu^2 + \frac{\sigma^2}{n}\right) = 2\mu^2 + \sigma^2\left(1 + \frac{1}{n}\right) - 2\left(\mu^2 + \frac{\sigma^2}{n}\right) = \sigma^2\left(1 - \frac{1}{n}\right) \implies \\ &\implies E[\sigma_n^2] = \sigma^2\left(1 - \frac{1}{n}\right) = h(\theta) \text{ y } b(\sigma_n^2, \sigma^2) = E[\sigma_n^2] - \sigma^2 = -\frac{\sigma^2}{n} \\ &\implies E[\sigma_n^2] = \sigma^2 \left(1 - \frac{1}{n}\right) = h(\theta) \text{ y } b(\sigma_n^2, \sigma^2) = E[\sigma_n^2] - \sigma^2 = -\frac{\sigma^2}{n} \end{aligned}$$

Observación 4.0.1

- Puede ocurrir que no exista un estimador centrado de θ
- Si T es un estimador centrado para θ , en general $h(T)$ no tiene por qué ser centrado para $h(\theta)$
- A pesar de que exista un estimador centrado para θ , puede ser que no tenga sentido

Ejemplo

Sea una m.a.s. de tamaño $n = 1$ de una población que sigue una distribución $Bin(1, \theta)$ demostrar que $T(X) = X^2$ no es un estimador centrado de θ^2 :

$$X \sim Bin(1, \theta) \equiv Bernouilli(\theta) \implies X^2 \sim Bernouilli(\theta) \implies E[X^2] = \theta \neq \theta^2$$

Ejemplo

Sea una m.a.s. de tamaño $n = 1$ de una población que sigue una distribución $Bin(1, \theta)$ demostrar que no existe un estimador centrado de θ^2 :

Sea $X \sim Bin(1, \theta)$ y $T(X)$ un posible estimador de θ^2 . Para que $T(X)$ sea centrado (insesgado), debe satisfacer:

$$E[T(X)] = \theta^2 \quad \forall \theta \in (0, 1)$$

Como $n = 1$, el estimador $T(X)$ solo puede tomar dos valores:

- $T(1)$ cuando $X = 1$ (éxito)
- $T(0)$ cuando $X = 0$ (fracaso)

Calculamos la esperanza:

$$E[T(X)] = T(1) \cdot P(X = 1) + T(0) \cdot P(X = 0) = T(1)\theta + T(0)(1 - \theta)$$

Para que sea insesgado:

$$T(1)\theta + T(0)(1 - \theta) = \theta^2 \quad \forall \theta \in (0, 1)$$

Reordenando términos:

$$\theta^2 - (T(1) - T(0))\theta - T(0) = 0 \quad \forall \theta \in (0, 1)$$

Esta es una ecuación cuadrática en θ . Sin embargo, un polinomio de grado 2 no nulo tiene como máximo 2 raíces reales distintas, mientras que la ecuación debe cumplirse para infinitos $\theta \in (0, 1)$. La única posibilidad sería que los coeficientes fueran idénticamente cero:

$$\begin{cases} 1 = 0 \\ T(1) - T(0) = 0 \\ T(0) = 0 \end{cases}$$

Lo cual es un sistema incompatible (la primera ecuación $1 = 0$ es falsa). Por tanto, **no existe** ninguna elección de $T(0)$ y $T(1)$ que satisfaga la condición de insesgaredad para todo $\theta \in (0, 1)$.

Ejemplo

Sea una m.a.s. de tamaño $n = 1$ que sigue una distribución $X \sim \text{Poisson}(\theta)$ demuestramos que $T(X) = (-2)^x$ es un estimador centrado para $h(\theta) = e^{-3\theta}$, pero $\text{Var}_\theta(T) = e^{4\theta} - e^{-6\theta} \rightarrow \infty$ no es un estimador de $h(\theta)$:

$$X \sim \text{Poisson}(\theta) \implies f_\theta(x) = \frac{e^{-\theta}\theta^x}{x!} \implies E[T] = E[(-2)^x] = \sum_{x=0}^{+\infty} (-2)^x \frac{e^{-\theta}\theta^x}{x!} = \sum_{x=0}^{+\infty} \frac{(-2\theta)^x}{x!} e^{-\theta} =$$

$$e^{-\theta} \sum_{x=0}^{+\infty} \frac{(-2\theta)^x}{x!} = e^{-\theta} e^{-2\theta} = e^{-3\theta} = h(\theta)$$

$$\text{Var}_\theta(T) = E[T^2] - E[T]^2 = E[(-2)^{2x}] - e^{-6\theta} = \sum_{x=0}^n (-2)^{2x} \frac{e^{-\theta}\theta^x}{x!} - e^{-6\theta} = \sum_{x=0}^n \frac{(4\theta)^x}{x!} e^{-\theta} - e^{-6\theta} =$$

$$e^{-\theta} \sum_{x=0}^n \frac{(4\theta)^x}{x!} - e^{-6\theta} = e^{-\theta} e^{4\theta} - e^{-6\theta} = e^{3\theta} - e^{-6\theta} \rightarrow \infty$$

Este procedimiento ha demostrado que $T(X)$ es centrado para $h(\theta)$, pero su varianza es demasiado grande (infinita) por lo que no es adecuado para la estimación.

Ejemplo

Sea $(T_j)_{j \in \mathbb{N}}$ sucesión de estimadores centrados para θ , entonces $\bar{T}_k = \frac{1}{k} \sum_{j=1}^k T_j$ es un estimador centrado para θ :

$$E[\bar{T}_k] = E\left[\frac{1}{k} \sum_{j=1}^k T_j\right] = \frac{1}{k} \sum_{j=1}^k E[T_j] = \frac{1}{k} \sum_{j=1}^k \theta = \theta$$

Definición 4.0.4 [Estimadores consistentes]

Una sucesión de estimadores $T_n = T(X_1, \dots, X_n)$ se dice **consistente** para el parámetro θ si, para todo $\theta \in \Theta$, se cumple que

$$E_\theta[T_n] \xrightarrow{n \rightarrow \infty} \theta \quad y \quad V_\theta(T_n) \xrightarrow{n \rightarrow \infty} 0.$$

Es decir, T_n es un estimador insesgado asintóticamente y su varianza tiende a cero conforme aumenta el tamaño de la muestra, lo que garantiza que T_n converge en probabilidad a θ .

Proposición 4.0.1

Si $T_n = T(X_1, \dots, X_n)$ es una sucesión de estimadores tales que $\forall \theta \in \Theta$, $E_\theta[T_n] \xrightarrow{n \rightarrow \infty} \theta$, $V_\theta(T_n) \xrightarrow{n \rightarrow \infty} 0$, entonces T_n es consistente para θ

Demostración. $E_\theta[(T_n - \theta)^2] = V_\theta(T_n) + b(T_n, \theta)^2 \xrightarrow{n \rightarrow \infty} 0, \forall \theta \in \Theta$ entonces $T_n \xrightarrow[n \rightarrow \infty]{m.c.} \theta, \forall \theta \in \Theta$ □

Ejemplo

El estimador $a_k = \frac{1}{n} \sum_{i=1}^n X_i^k$ es un estimador consistente para el momento poblacional con respecto al origen de orden k , es decir, es estimador del parámetro $\theta = \alpha_k$.

Ejemplo

Sea una m.a.s. de tamaño n de una población de *Bernoulli*(θ) comprobemos que el estimador $T_n = \frac{1}{n+2} (\sum_{i=1}^n X_i + 1)$ es un estimador consistente para θ :

$$\lim_{n \rightarrow \infty} E[T_n] = \lim_{n \rightarrow \infty} E \left[\frac{1}{n+2} \sum_{i=1}^n X_i + 1 \right] = \lim_{n \rightarrow \infty} \frac{1}{n+2} \sum_{i=1}^n (E[X_i] + 1) = \lim_{n \rightarrow \infty} \frac{n\theta + 1}{n+2} = \theta$$

$$\lim_{n \rightarrow \infty} V[T_n] = \lim_{n \rightarrow \infty} V \left[\frac{1}{n+2} \sum_{i=1}^n X_i + 1 \right] = \lim_{n \rightarrow \infty} \frac{1}{(n+2)^2} \sum_{i=1}^n V[X_i] = \lim_{n \rightarrow \infty} \frac{n\theta(1-\theta)}{(n+2)^2} = 0$$

Ejemplo

Sea una m.a.s. de tamaño n de una población con distribución $N(\mu; \sigma^2)$ siendo $\theta = (\mu, \sigma)$, el estimador $T_1(\vec{X}) = \bar{X}$ es consistente para estimar $h(\theta) = \mu$, ya que:

$$\lim_{n \rightarrow \infty} E[T_1] = \lim_{n \rightarrow \infty} E[\bar{X}] = \lim_{n \rightarrow \infty} E \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E[X_i] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mu = \mu$$

$$\lim_{n \rightarrow \infty} V[T_1] = \lim_{n \rightarrow \infty} V \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n V[X_i] = \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = 0$$

4.1 Estimadores Bayesianos

El enfoque bayesiano trata a los parámetros de las distribuciones como variables aleatorias con su propia función de distribución, a diferencia de considerar que toma valores fijos desconocidos. Para desarrollar este punto de vista, se asigna una distribución a θ llamada distribución inicial o a priori $\pi(\theta)$ y se actualiza esta distribución con la información de la muestra para obtener la distribución final o a posteriori $\pi(\theta | x_1, \dots, x_n)$

$$\pi(\theta | x_1, \dots, x_n) = \frac{\pi(\theta) f(x_1, \dots, x_n | \theta)}{m(x_1, \dots, x_n)}$$

donde m es la distribución predictiva, dada por

$$m(x_1, \dots, x_n) = \int_{\Theta} \pi(\theta) f(x_1, \dots, x_n | \theta) d\theta$$

A la función $f(x_1, \dots, x_n | \theta)$ se le llama función de verosimilitud y es la función de densidad de la muestra condicionada al parámetro θ .

Observación 4.1.1

Antes de tomar la muestra, la información sobre θ viene dada por $\pi(\theta)$ y tras la experimentación se debe utilizar $\pi(\theta | x_1, \dots, x_n)$. El estimador bayesiano de θ es toda la distribución final y por extensión cualquier medida de centralización correspondiente a esta distribución

Ejemplo

Sea una muestra aleatoria simple (m.a.s.) de tamaño n proveniente de una población que sigue una distribución $Bin(1, \theta)$

$$X_i \sim Bernoulli(\theta), \quad i = 1, \dots, n.$$

Esto significa que cada X_i puede tomar valores 0 o 1 con probabilidades:

$$P(X_i = 1 | \theta) = \theta, \quad P(X_i = 0 | \theta) = 1 - \theta.$$

Buscamos encontrar la distribución a posteriori de θ usando inferencia Bayesiana.

Antes de observar los datos, suponemos que θ sigue una distribución uniforme en el intervalo $(0, 1)$:

$$\theta \sim U(0, 1).$$

Dado que la densidad de la distribución uniforme en $(0, 1)$ es:

$$\pi(\theta) = \frac{1}{1-0} = 1, \quad 0 \leq \theta \leq 1,$$

esto implica que cualquier valor de θ dentro del intervalo $[0, 1]$ es igualmente probable antes de ver los datos.

Dado que la muestra es independiente, la función de verosimilitud se obtiene como el producto de las probabilidades individuales:

$$f(x_1, \dots, x_n | \theta) = \prod_{i=1}^n P(X_i = x_i | \theta).$$

Usando la función de masa de probabilidad de la distribución Bernoulli:

$$P(X_i = x_i | \theta) = \theta^{x_i} (1 - \theta)^{1-x_i},$$

tenemos:

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

Factorizando los exponentes:

$$f(x_1, \dots, x_n | \theta) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}.$$

Definiendo $S = \sum_{i=1}^n x_i$, que es la cantidad de éxitos en la muestra, la función de verosimilitud se expresa como:

$$f(x_1, \dots, x_n | \theta) = \theta^S (1 - \theta)^{n-S}.$$

Aplicando el Teorema de Bayes, la distribución a posteriori es proporcional a:

$$\pi(\theta|x_1, \dots, x_n) \propto \pi(\theta)f(x_1, \dots, x_n|\theta).$$

Sustituyendo $\pi(\theta) = 1$ y la función de verosimilitud:

$$\pi(\theta|x_1, \dots, x_n) \propto \theta^S(1 - \theta)^{n-S}.$$

Reconociendo que esta es la forma de la función de densidad de una distribución Beta:

$$Beta(a, b) \propto \theta^{a-1}(1 - \theta)^{b-1},$$

identificamos que la posteriori es una distribución Beta con parámetros:

$$\pi(\theta|x_1, \dots, x_n) \sim Beta(S + 1, n - S + 1).$$

En general, si en lugar de una uniforme usamos una distribución a priori Beta con parámetros α y β :

$$\theta \sim Beta(\alpha, \beta),$$

cuya función de densidad es:

$$\pi(\theta) = \frac{\theta^{\alpha-1}(1 - \theta)^{\beta-1}}{B(\alpha, \beta)},$$

siguiendo el mismo procedimiento, obtenemos que la distribución a posteriori también sigue una Beta con parámetros actualizados:

$$\pi(\theta|x_1, \dots, x_n) \sim Beta(\alpha + S, \beta + n - S).$$

Esto muestra que **la familia de distribuciones Beta es conjugada para la distribución Bernoulli-Binomial**.

El valor esperado de θ dado los datos es:

$$E[\theta|x_1, \dots, x_n] = \frac{S + \alpha}{n + \alpha + \beta}.$$

Si tomamos la esperanza como una combinación convexa:

$$E[\theta|x_1, \dots, x_n] = \frac{n}{n + \alpha + \beta} \bar{x} + \frac{\alpha + \beta}{n + \alpha + \beta} \frac{\alpha}{\alpha + \beta}.$$

Esto indica que la esperanza de la distribución a posteriori es una mezcla entre la media muestral \bar{x} y la esperanza a priori $\frac{\alpha}{\alpha + \beta}$, ponderada por los tamaños relativos de la muestra y la información previa.

Ejemplo

Sea una m.a.s. de tamaño n de una población $N(\mu, \sigma)$ y con $\mu \sim N(\mu_0, \sigma_0)$ y σ conocida entonces $\pi(\mu|x_1, \dots, x_n) \sim N(\mu_1, \sigma_1)$:

$$\mu \sim N(\mu_0, \sigma_0) \implies \pi(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}}$$

$$X \sim N(\mu, \sigma) \implies f(x|\mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x - \mu)^2}{2\sigma^2}} \implies f(x_1, \dots, x_n|\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} =$$

$$\begin{aligned}
&= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}} \\
\pi(\theta|x_1, \dots, x_n) &= \frac{\frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}} \cdot \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}}}{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}} \cdot \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}} d\mu} = \\
&= \frac{e^{-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}} \cdot e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}}}{\int_{\mathbb{R}} e^{-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}} \cdot e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}} d\mu} = \frac{e^{-\frac{\mu^2 - \mu_0^2 + 2\mu\mu_0}{2\sigma_0^2}} \cdot e^{-\frac{\sum_{i=1}^n x_i^2 - n\mu^2 + 2\mu \sum_{i=1}^n x_i}{2\sigma^2}}}{\int_{\mathbb{R}} e^{-\frac{\mu^2 - \mu_0^2 + 2\mu\mu_0}{2\sigma_0^2}} \cdot e^{-\frac{\sum_{i=1}^n x_i^2 - n\mu^2 + 2\mu \sum_{i=1}^n x_i}{2\sigma^2}} d\mu} = \\
&= \frac{e^{-\frac{\mu^2 + 2\mu\mu_0}{2\sigma_0^2}} \cdot e^{-\frac{-n\mu^2 + 2\mu \bar{X}_n}{2\sigma^2}}}{\int_{\mathbb{R}} e^{-\frac{\mu^2 + 2\mu\mu_0}{2\sigma_0^2}} \cdot e^{-\frac{-n\mu^2 + 2\mu \bar{X}_n}{2\sigma^2}} d\mu} = \frac{e^{-\frac{1}{2}\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right)\mu^2 + \left(\frac{\mu_0}{\sigma_0^2} + \frac{\bar{X}}{\sigma^2/n}\right)\mu}}{\int_{\mathbb{R}} e^{-\frac{1}{2}\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right)\mu^2 + \left(\frac{\mu_0}{\sigma_0^2} + \frac{\bar{X}}{\sigma^2/n}\right)\mu} d\mu} = \frac{e^{-\frac{1}{2V}\mu^2 + \frac{E}{V}\mu}}{\int_{\mathbb{R}} e^{-\frac{1}{2V}\mu^2 + \frac{E}{V}\mu} d\mu}
\end{aligned}$$

donde $V = \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}$ y $E = \frac{\mu_0}{\sigma_0^2} + \frac{\bar{X}}{\sigma^2/n}$.

Considerando la normal $N(E, \sqrt{V})$ y la integral de su función de densidad que es igual a 1:

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi V}} e^{-\frac{1}{2V}(\mu - E)^2} d\mu = 1$$

entonces factorizando el exponente obtenemos:

$$\begin{aligned}
1 &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi V}} e^{-\frac{1}{2V}(\mu^2 - 2E\mu + E^2)} d\mu = \frac{1}{\sqrt{2\pi V}} e^{-\frac{E^2}{2V}} \int_{\mathbb{R}} e^{-\frac{1}{2V}\mu^2 + \frac{E}{V}\mu} d\mu \\
&\Rightarrow \int_{\mathbb{R}} e^{-\frac{1}{2V}\mu^2 + \frac{E}{V}\mu} d\mu = \sqrt{2\pi V} e^{\frac{E^2}{2V}}
\end{aligned}$$

Luego entonces tenemos que

$$\pi(\mu|x_1, \dots, x_n) = \frac{e^{-\frac{1}{2V}\mu^2 + \frac{E}{V}\mu}}{\sqrt{2\pi V} e^{\frac{E^2}{2V}}} = \frac{1}{\sqrt{2\pi V}} e^{-\frac{1}{2V}(\mu - E)^2}$$

Definición 4.1.1 [Estadístico suficiente bayesiano]

$T = T(X_1, \dots, X_n)$ es un estadístico suficiente bayesiano para θ para la familia $\mathcal{P} = \{f(\vec{x}|\theta) : \theta \in \Theta\}$ si cualquiera que sea la distribución inicial $\pi(\theta)$, se tiene que la distribución final dada por la muestra y por el valor del estadístico son la misma. Es decir:

$$\pi(\theta|x_1, \dots, x_n) = \pi(\theta|t) \quad \text{tal que } t = T(x_1, \dots, x_n)$$

Teorema 4.1.1 [Versión bayesiana del Teorema de Factorización de Fisher]

$T = T(X_1, \dots, X_n)$ es un estadístico suficiente para θ si y sólo si T es un estadístico suficiente bayesiano para θ respecto a $\pi(\theta)$, cualquiera que sea la distribución inicial $\pi(\theta)$

Demostración.

$$\begin{aligned}\Rightarrow \pi(\theta|x_1, \dots, x_n) &= \frac{\pi(\theta)f(x_1, \dots, x_n|\theta)}{m(x_1, \dots, x_n)} = \frac{\pi(\theta)g(t|\theta)f(x_1, \dots, x_n|t, \theta)}{m(x_1, \dots, x_n)} = \frac{\pi(\theta)g(t|\theta)}{\int_{\Theta} \pi(\theta)g(t|\theta)d\theta} = \pi(\theta|t) \\ \Rightarrow f(x_1, \dots, x_n|\theta) &= \frac{\pi(\theta|x_1, \dots, x_n)m(x_1, \dots, x_n)}{\pi(\theta)} = \frac{\pi(\theta|t)}{\pi(\theta)}m(x_1, \dots, x_n)\end{aligned}$$

□

4.2 Criterios de comparación de estimadores

Definición 4.2.1 [Error cuadrático medio]

Dado un estimador $T(X_1, \dots, X_n)$ de θ , se denomina error cuadrático medio ECM a la expresión en función de θ :

$$ECM_T(\theta) = E_{\theta}[(T - \theta)^2]$$

Conceptualmente, el error cuadrático medio es una medida que indica qué tan cerca está un estadístico del parámetro verdadero que se intenta estimar.

Proposición 4.2.1

Dado un estimador $T(X_1, \dots, X_n)$ de θ , se tiene que:

$$ECM_T(\theta) = V_{\theta}(T) + b(T, \theta)^2$$

Demostración.

$$ECM_T(\theta) = E_{\theta}[(T - \theta)^2] = E_{\theta}[T^2 - 2T\theta + \theta^2] = E_{\theta}[T^2] - 2\theta E_{\theta}[T] + \theta^2 = V_{\theta}(T) + b(T, \theta)^2$$

□

Observación 4.2.1

El sesgo mide qué tanto se desvía, en promedio, el estimador del valor verdadero del parámetro.

La varianza del estimador mide cómo varían las estimaciones (del estimador) si tomamos diferentes muestras.

Es decir, responden a las preguntas de ¿Apunta al lugar correcto? y ¿Qué tan dispersas están las estimaciones? respectivamente.

Ejemplo

Dada una m.a.s. de tamaño n de una población $X \sim \text{Bernouilli}(\theta)$ el error cuadrático medio del estimador bayesiano \bar{X} :

$$\begin{aligned}E_{\theta}[(T - \theta)^2] &= E_{\theta}[(\bar{X} - \theta)^2] = E_{\theta}[\bar{X}^2 - 2\bar{X}\theta + \theta^2] = E_{\theta}[\bar{X}^2] - 2\theta E_{\theta}[\bar{X}] + \theta^2 = \\ &= \frac{1}{n^2} \sum_{i=1}^n E_{\theta}[X_i^2] - 2\theta \frac{1}{n} \sum_{i=1}^n E_{\theta}[X_i] + \theta^2 = \frac{1}{n^2} n(\theta^2 + \theta(1-\theta)) - 2\theta^2 + \theta^2 = \frac{\theta(1-\theta)}{n} = \frac{\theta}{n} - \frac{\theta^2}{n} = \frac{\theta(1-\theta)}{n}\end{aligned}$$

Ejemplo

Dada una m.a.s. de tamaño n de una población $X \sim N(\mu, \sigma)$, se sabe que los estimadores centrados de ambos parámetros son \bar{X} para μ y S^2 para σ^2 , respectivamente. Y sus errores cuadráticos medios son:

$$ECM_{\bar{X}}(\mu) = Var[\bar{X}] = \frac{\sigma^2}{n} \quad ECM_{S^2}(\sigma^2) = Var[S^2] = \frac{2\sigma^4}{n-1}$$

Sea $b_2 = \sigma_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ otro estimador centrado para σ^2 , calculemos su varianza y sesgo: Gracias al cálculo realizado en un ejemplo anterior tenemos que $b(\sigma_n^2, \sigma) = -\frac{\sigma^2}{n}$, por lo que sólo queda calcular la varianza, la cual es de la forma

$$V_{\theta}(\sigma_n^2) = \frac{2\sigma^4}{n} \implies ECM_{\sigma_n^2}(\sigma^2) = V_{\theta}(\sigma_n^2) + b(\sigma_n^2, \sigma)^2 = \frac{2\sigma^4}{n} - \frac{\sigma^4}{n^2} = \frac{2n-1}{n^2} \sigma^4 < \frac{2\sigma^4}{n-1} = ECM_{S^2}(\sigma^2)$$

Por lo que se puede concluir que S^2 es un estimador más eficiente que σ_n^2 para σ^2 . A pesar de que matemáticamente se parezcan más, la corrección para mejorar la eficiencia se la conoce como corrección de Bessel.

Observación 4.2.2

En general, si T_1 y T_2 son dos estimadores de θ y $ECM_{T_1}(\theta) < ECM_{T_2}(\theta)$, entonces T_1 es un estimador más eficiente que T_2 para θ

Definición 4.2.2 [Pérdida final esperada]

Dado un estimador $T(X_1, \dots, X_n)$, la distribución inicial $\pi(\theta)$ y la función de pérdida $\mathcal{L}(\theta, t)$ donde t son los valores que toma el estimador, se define la Pérdida Final Esperada o PFE o el riesgo a posteriori como:

$$PFE_T = E[\mathcal{L}(t, \theta) | X_1 = x_1, \dots, X_n = x_n] = \int_{\Theta} \mathcal{L}(\theta, t) \pi(\theta | x_1, \dots, x_n) d\theta$$

Proposición 4.2.2

Puede darse que existan varias funciones de pérdida, veamos las más comunes:

1. **Pérdida cuadrática:** Si $\mathcal{L}(\theta, t) = (\theta - t)^2$ entonces $PFE(t) = V(\theta | x_1, \dots, x_n) + b(T, \theta)^2$ y la pérdida final esperada se minimiza en $t^* = E[\theta | x_1, \dots, x_n]$
2. **Pérdida absoluta:** Si $\mathcal{L}(\theta, t) = |\theta - t|$ entonces $PFE(t) = E[|\theta - t| | x_1, \dots, x_n]$ y la pérdida final esperada se minimiza en la mediana de la distribución final (estimador bayesiano)
3. **Pérdida 0-1:** Si $\mathcal{L}(\theta, t) = \begin{cases} 0 & \text{si } t = \theta \\ 1 & \text{si } t \neq \theta \end{cases}$ entonces $PFE_T = P(\theta \neq t | x_1, \dots, x_n) = 1 - P(\theta = t | x_1, \dots, x_n)$. Se usa en problemas de decisión o clasificación donde sólo importa acertar o errar, no cuán lejos se está del valor correcto, es por ello que es más difícil saber dónde se encuentra el valor que la minimiza.

Demostración. 1. Si $\mathcal{L}(\theta, t) = (\theta - t)^2 \implies$

$$PFE_T = E[(\theta - t)^2 | x_1, \dots, x_n] = E[((\theta - E[\theta | x_1, \dots, x_n]) + (E[\theta | x_1, \dots, x_n] - t))^2 | x_1, \dots, x_n]$$

Si expandimos el cuadrado:

$$(\theta - E[\theta|x_1, \dots, x_n])^2 + 2(\theta - E[\theta|x_1, \dots, x_n])(E[\theta|x_1, \dots, x_n] - t) + (E[\theta|x_1, \dots, x_n] - t)^2 \implies$$

Calculemos cada una de las esperanzas por separado:

$$E[(\theta - E[\theta|x_1, \dots, x_n])^2|x_1, \dots, x_n] = V(\theta|x_1, \dots, x_n)$$

$$E[(\theta - E[\theta|x_1, \dots, x_n])|x_1, \dots, x_n] = 0 \text{ por propiedades de la esperanza condicional}$$

$$E[(E[\theta|x_1, \dots, x_n] - t)^2|x_1, \dots, x_n] = (E[\theta|x_1, \dots, x_n] - t)^2 \text{ dada una muestra, se vuelve una constante}$$

Por lo que se puede concluir que $PFE_T = V(\theta|x_1, \dots, x_n) + (E[\theta|x_1, \dots, x_n] - t)^2$ y se minimiza en $t^* = E[\theta|x_1, \dots, x_n]$

2. Si $\mathcal{L}(\theta, t) = |\theta - t| \implies$

$$PFE_T = E[|\theta - t|x_1, \dots, x_n] = \int_{\Theta} |\theta - t| F(\theta|x) d\theta \implies$$

Dividiendo la integral entre los valores positivos y los negativos de θ , nos queda que la integral es la suma de:

$$\int_{-\infty}^t (t - \theta) f(\theta|x) d\theta + \int_t^{+\infty} (\theta - t) f(\theta|x) d\theta$$

Si derivamos con respecto a t y obtenemos 0 podemos ver un posible punto máximo o mínimo de la función:

$$\frac{d}{dt} PFE_T = \frac{d}{dt} \int_{-\infty}^t (t - \theta) f(\theta|x) d\theta + \frac{d}{dt} \int_t^{+\infty} (\theta - t) f(\theta|x) d\theta =$$

$$= \int_{-\infty}^t f(\theta|x) d\theta - \int_t^{+\infty} f(\theta|x) d\theta = 0 \iff$$

$$F_X(t) = 1 - F_X(t) \iff F_X(t) = \frac{1}{2} \implies t^* = \text{mediana de la distribución final}$$

□

Observación 4.2.3

$$PFE(t) = \int_{\Theta} \underbrace{\mathcal{L}(\theta, t)}_{\text{pérdida si fuera } \theta} \cdot \underbrace{\pi(\theta | x)}_{\text{qué tan probable es } \theta} d\theta$$

Ejemplo

Sea una m.a.s. de tamaño n de una población $X \sim \text{Bernoulli}(\theta)$ y pérdida cuadrática e información inicial dadas por la distribución $\text{Beta}(\alpha_0, \beta_0)$. Por un ejemplo anterior en el apartado de **Estimadores bayesianos**, que la distribución final o a posteriori es $\text{Beta}(\alpha_0 + \sum_{i=1}^n X_i, \beta_0 + n - \sum_{i=1}^n X_i)$

$$\begin{aligned} \implies PFE_T &= E[(t - \theta)^2|\vec{x}] = \int_0^1 (t - \theta)^2 \cdot \frac{\theta^{\alpha_0 + \sum_{i=1}^n X_i - 1} (1 - \theta)^{\beta_0 + n - \sum_{i=1}^n X_i - 1}}{\beta(\alpha_0 + \sum_{i=1}^n X_i, \beta_0 + n - \sum_{i=1}^n X_i)} d\theta \\ &= \int_0^1 (t^2 - 2t\theta + \theta^2) \cdot \text{Beta}(\theta; \alpha_1, \beta_1) d\theta = t^2 \int_0^1 \text{Beta}(\theta) d\theta - 2t \int_0^1 \theta \cdot \text{Beta}(\theta) d\theta + \int_0^1 \theta^2 \cdot \text{Beta}(\theta) \\ &= t^2 - 2t \frac{\alpha_1}{\alpha_1 + \beta_1} + \frac{\alpha_1(\alpha_1 + 1)}{(\alpha_1 + \beta_1)(\alpha_1 + \beta_1 + 1)} = t^2 - 2t \cdot \mu + E[\theta^2] \end{aligned}$$

Expresión que se minimiza cuando $t = E[\theta|\vec{x}] = \frac{\alpha_0 + \sum X_i}{\alpha_0 + \beta_0 + n} = \frac{\alpha_1}{\alpha_1 + \beta_1}$

Definición 4.2.3 [Estimador centrado uniformemente de mínima varianza]

$T^* = T^*(X_1, \dots, X_n)$ es un estimador centrado uniformemente de mínima varianza para θ si y sólo si $E_\theta[T^*] = \theta$ y para cualquier otro estimador $T = T(X_1, \dots, X_n)$ con $E_\theta[T] = \theta$, se tiene que $V_\theta(T^*) \leq V_\theta(T)$, $\forall \theta \in \Theta$

Proposición 4.2.3 [Unicidad de los estimadores centrados uniformemente de mínima varianza]

Si existe un estimador centrado uniformemente de mínima varianza para θ , entonces es único c.s.

Demostración. Sean T_1 y T_2 dos estimadores centrados uniformemente de mínima varianza para θ , demostremos que entonces $T_1 \stackrel{c.s.}{=} T_2$.

Sea $T = \frac{T_1 + T_2}{2} \implies E_\theta[T] = \theta$

$$V_\theta(T) = V_\theta\left(\frac{T_1 + T_2}{2}\right) = \frac{1}{4}(V_\theta(T_1) + V_\theta(T_2) + 2Cov_\theta(T_1, T_2)) = \frac{V_\theta(T_1)}{2} + \frac{Cov_\theta(T_1, T_2)}{2}$$

Sabemos que la correlación de dos variables aleatorias está acotada por 1, entonces:

$$\begin{aligned} \rho_\theta(T_1, T_2) &= \frac{Cov_\theta(T_1, T_2)}{\sqrt{V_\theta(T_1)V_\theta(T_2)}} \leq 1 \iff Cov_\theta(T_1, T_2) \leq V_\theta(T_1) \implies \\ &\implies V_\theta(T) \leq V_\theta(T_1) \end{aligned}$$

Pero además como T_1 es un estimador centrado uniformemente de mínima varianza, ningún otro estimador centrado puede tener un avarianza más pequeña: $V_\theta(T) \leq V_\theta(T_1)$ Por lo tanto $V_\theta(T) = V_\theta(T_1) = Cov_\theta(T_1, T_2) = V_\theta(T_2) \implies \rho_\theta(T_1, T_2) = 1 \implies \exists a, b : T_1 = aT_2 + b \iff E[T_1] = aE[T_2] + b \iff \theta = a\theta + b \iff a = 1, b = 0 \implies T_1 \stackrel{c.s.}{=} T_2$ \square

Ejemplo

Sea una m.a.s. de tamaño n de una población $X \sim N(\mu, \sigma^2)$ vamos a trabajar con la familia de estimadores $T_k = \{k \cdot S_n^2\}$. Calculemos cuál es el menor error cuadrático medio. Y tomemos como función del estimador $d(\theta) = \theta^2$

$$ECM_{T_k}(\theta) = V_\theta(T_k) + b(T_k, \theta)^2 \implies$$

$$\begin{cases} b(T_k, \theta) = E_\theta[T_k] - d(\theta) = \frac{k\theta^2}{n} \cdot E_\theta\left[\frac{n}{\theta^2} S_n^2\right] = \frac{k\theta^2}{n} \cdot E_\theta[\chi_n^2] - \theta^2 = \frac{k\theta^2}{n} \cdot n - \theta^2 = k\theta^2 - \theta^2 = (k-1)\theta^2 \\ V_\theta(T_k) = V_\theta[k \cdot S_n^2] = \frac{k^2\theta^4}{n^2} V_\theta\left[\frac{n}{\theta^2} S_n^2\right] = \frac{k^2\theta^4}{n^2} V_\theta[\chi_n^2] = \theta^4 \cdot \frac{k^2}{n^2} \cdot 2n = \frac{2\theta^4 k^2}{n} \end{cases} \implies$$

$$\implies ECM_{T_k}(\theta) = \frac{2\theta^4 k^2}{n} + (k-1)^2 \theta^4 = \theta^4 \left(\frac{2k^2}{n} + (k-1)^2 \right)$$

Para encontrar el valor de k que minimiza el error cuadrático medio, derivamos con respecto a k e igualamos a 0:

$$\frac{d}{dk} ECM_{T_k}(\theta) = 0 \iff \frac{d}{dk} \left(\theta^4 \left(\frac{2k^2}{n} + (k-1)^2 \right) \right) = 0 \iff \theta^4 \left(\frac{4k}{n} + 2(k-1) \right) = 0 \iff$$

$$\iff \frac{4k}{n} + 2(k-1) = 0 \iff 4k + 2n(k-1) = 0 \iff 4k + 2nk - 2n = 0 \iff 4k + 2nk = 2n \iff$$

$$\iff k(4+2n) = 2n \iff k = \frac{2n}{4+2n} = \frac{n}{2+n}$$

Por lo que el estimador que minimiza el error cuadrático medio es $T_{\frac{n}{2+n}} = \frac{n}{2+n} S_n^2$

Teorema 4.2.1

El estimador centrado uniformemente de mínima varianza es función simétrica de las observaciones

Ejemplo

Sea una m.a.s. de tamaño $n = 2$, entonces el estimador $T_1 = \frac{X_1}{X_2}$ no puede ser un estimador centrado uniformemente de mínima varianza, ya que si lo fuera, entonces para el nuevo estimador $T_2 = \frac{X_2}{X_1}$ se tendría que $E_\theta[T_2] = E_\theta[T_1]$ y $V_\theta(T_2) = V_\theta(T_1)$ con lo que el estimador $T = \frac{T_1+T_2}{2} = \frac{X_1^2+X_2^2}{X_1X_2}$ sería tal que $V_\theta < V_\theta(T_1)$, lo cual es una contradicción

Observación 4.2.4

En general si tienes un estimador T que no es simétrico, puedes promediar sobre todas las permutaciones posibles para crear un nuevo estimador \bar{T} :

$$\bar{T} = \frac{1}{n!} \sum_{i=1}^{n!} T_j$$

Este nuevo estimador es simétrico respecto a las observaciones y $V_\theta(\bar{T}) \leq V_\theta(T_j) \forall j$ donde se cumple que $V_\theta(\bar{T}) < V_\theta(T_j)$ si T_j no es un estimador simétrico. Además, se cumple que $E_\theta[\bar{T}] = E_\theta[T]$

Teorema 4.2.2 [Teorema de caracterización del ECUMV]

Sea $T_1 = T_1(X_1, \dots, X_n)$ un estimador centrado para θ ($E_\theta[T_1] = \theta$) y $V_\theta(T_1) < \infty$ entonces T_1 es el ECUMV para θ si y sólo si para cualquier otro estimador $T_2 = T_2(X_1, \dots, X_n)$ con $E_\theta[T_2] = 0$ y $V_\theta(T_2) < \infty$ se tiene que $E_\theta[T_1 T_2] = 0$

Corolario 4.2.1

Si $T_1 = T_1(X_1, \dots, X_n)$ y $T_2 = T_2(X_1, \dots, X_n)$ son ECUMV para $h_1(\theta)$ y $h_2(\theta)$ respectivamente, entonces $b_1 T_1 + b_2 T_2$ es el ECUMV para $b_1 h_1(\theta) + b_2 h_2(\theta)$

Teorema 4.2.3 [Teorema de Rao-Blackwell]

Si $T = T(X_1, \dots, X_n)$ es un estimador centrado para θ con $V_\theta(T) < \infty$ y $S(X_1, \dots, X_n)$ es un estadístico suficiente, entonces $g(S) = E[T|S]$ es un estimador centrado para θ con $V_\theta(g(S)) \leq V_\theta(T)$

Demostración. Si S es una estadística suficiente para el parámetro θ , entonces $E[T | S]$ no depende de θ . Por las propiedades de la esperanza condicionada, se tiene que:

$$E_\theta[g(S)] = E_\theta[E[T | S]] = E_\theta[T] = \theta$$

Ahora, considerando la varianza de T :

$$V_\theta(T) = E_\theta[(T - \theta)^2] = E_\theta[(T - g(S) + g(S) - \theta)^2]$$

Expandiendo el cuadrado y usando la linealidad de la esperanza:

$$V_\theta(T) = E_\theta[(T - g(S))^2] + E_\theta[(g(S) - \theta)^2] + 2E_\theta[(T - g(S))(g(S) - \theta)]$$

El último término se anula debido a la siguiente propiedad de la esperanza condicionada:

$$E_\theta[(T - g(S))(g(S) - \theta)] = \iint (t - g(s))(g(s) - \theta) dF_\theta(t, s)$$

Descomponiendo en términos de la distribución condicional:

$$= \int (g(s) - \theta) \left(\int (t - g(s)) dF(t | s) \right) dF_\theta(s) = 0$$

Ya que $E[T | S] = g(S)$, la esperanza condicional centrada es cero.

Por lo tanto,

$$V_\theta(T) = V_\theta(g(S)) + E_\theta[(T - g(S))^2] \geq V_\theta(g(S))$$

donde se alcanza la igualdad si y solo si $T = g(S)$ casi seguramente. \square

Teorema 4.2.4 [Teorema de Lehmann-Scheffé]

Si $S(X_1, \dots, X_n)$ es un estadístico suficiente y completo para θ y $T = T(X_1, \dots, X_n)$ es un estimador centrado para θ tal que $T = h(S)$, entonces T es ECUMV para θ

Demostración. $\begin{cases} S \text{ suficiente} \\ T \text{ centrado} \end{cases} \implies g(S) = E[T|S] \text{ es centrado para } \theta \text{ y } V_\theta(g(S)) \leq V_\theta(T)$

Además, se tiene que para cualquier otro estimador T_1 centrado para θ , $g_1(S) = E[T_1|S]$ es centrado para θ y $V_\theta(g_1(S)) \leq V_\theta(T_1)$

Por lo tanto al ser S completo y $E_\theta[g(S) - g_1(S)] = \theta - \theta = 0$ se tiene que $g(S) \stackrel{c.s.}{=} g_1(S)$. En particular, para $T = h(S)$, $g_1(S) = E[h(S)|S] = h(S) = T$ y $V_\theta(T) \leq V_\theta(T_1)$, cualquiera que sea T_1 centrado para θ \square

Ejemplo

Sean una m.a.s. de tamaño n de una población con $X \sim \text{Bin}(1, \theta)$ y un estadístico $T = \sum_{i=1}^n X_i$, comprueba que es suficiente y completo y además, si $h(T) = \bar{X}$, comprueba entonces que $h(T)$ es el ECUMV para θ :

Veamos primero que $T = \sum_{i=1}^n X_i$ es suficiente y completo para θ :

$$X \sim \text{Bin}(1, \theta) \equiv \text{Bernoulli}(\theta) \implies f_\theta(x) = \theta^x(1-\theta)^{1-x} \implies f_\theta(x_1, \dots, x_n) = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n - \sum_{i=1}^n x_i}$$

Entonces por el Teorema de Factorización de Fisher, tenemos que T es suficiente para θ . Veamos ahora su completitud:

$$X \sim \text{Bin}(1, \theta) \implies T \sim \text{Bin}(n, \theta) \implies$$

Sea g función real tal que: $E_\theta[g(T)] = 0, \forall \theta \in [0, 1]$, entonces:

$$E_\theta[g(T)] = \sum_{t=0}^n g(t) \binom{n}{t} \theta^t (1-\theta)^{n-t} = 0, \forall \theta \in [0, 1] \iff$$

$$\iff (1-\theta)^n \cdot \sum_{i=1}^n g(t) \binom{n}{t} \left(\frac{\theta}{1-\theta}\right)^t = 0 \iff$$

$$\iff \sum_{t=0}^n g(t) \binom{n}{t} x^t = 0, \forall x \in \mathbb{R} \iff g(t) = 0, \forall t \in \{0, \dots, n\}$$

ya que los coeficientes binomiales son no nulos.

Por último queda ver que S es un estadístico centrado para θ , i.e. $E_\theta[h(S)] = \theta$ y $V_\theta(h(S)) < \infty$:

$$E_\theta[h(T)] = E_\theta[\bar{X}] = \frac{1}{n} \cdot \sum_{i=1}^n E_\theta[X_i] = \frac{1}{n} \cdot n\theta = \theta$$

$$V_\theta(h(T)) = V_\theta(\bar{X}) = \frac{1}{n^2} \cdot \sum_{i=1}^n V_\theta(X_i) = \frac{1}{n^2} \cdot n\theta(1-\theta) = \frac{\theta(1-\theta)}{n} < \infty$$

Por lo tanto, $T = n\bar{X}$ es el ECUMV para θ

Ejemplo

Sea una m.a.s. de tamaño n de una población con distribución $Poisson(\theta)$ y dado un estadístico $d(\theta) = e^{-2\theta}$. Encuentra el ECUMV para $d(\theta)$:

Ejemplo

Sea una m.a.s. de tamaño n de una población que sigue una distribución $Poisson(\theta)$, tenemos que encontrar el ECUMV para θ :

Si tomamos el estadístico $S = \sum_{i=1}^n X_i$, para poder aplicar el Teorema de Lehmann-Scheffé necesitamos ver que el estadístico sea completo y suficiente:

Veamos primero que S es suficiente para θ :

$$X \sim \text{Poisson}(\theta) \implies f_\theta(x) = \frac{e^{-\theta} \theta^x}{x!} \implies f_\theta(x_1, \dots, x_n) = \frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$$

Entonces, por el Teorema de Factorización de Fisher, tenemos que S es suficiente para θ . Veamos ahora su completitud:

Siguiendo con lo anterior:

$$f_\theta(x_1, \dots, x_n) = \frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} = \frac{e^{-n\theta} e^{\ln(\theta) \sum x_i}}{\prod_{i=1}^n x_i!} \implies \begin{cases} c(\theta)^n = e^{-n\theta} \\ \prod_{i=1}^n h(x_i) = \prod_{i=1}^n x_i! \\ q_1(\theta) = \ln(\theta) \\ T_1(\vec{x}) = \sum_{i=1}^n x_i \end{cases}$$

Entonces, debemos ver que $\ln(\theta)$ contiene un abierto $(0, +\infty) \subset \mathbb{R}$, por lo tanto s es completo para

θ . Además, tomando el estadístico $T = \frac{1}{n} \sum_{i=1}^n x_i$ tenemos que:

$$E[T] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} \cdot n\theta = \theta \implies T \text{ es centrado para } \theta \implies$$

Tomando la función $h(x) = \frac{1}{n} \cdot x$ tenemos que:

$$h(S) = \frac{1}{n} \sum_{i=1}^n X_i = \frac{S}{n} \implies$$

Por el Teorema de Lehmann-Scheffé, $T = h(S)$ es el ECUMV para θ .

Si en lugar de haber tomado $d(\theta) = \theta$ hubieramos querido el estimador centrado uniformemente de mínima varianza para $d(\theta) = e^{-\theta}$ HAY QUE INSERTAR LO DE DIEGO

Ejemplo

Sea una m.a.s. de tamaño n de una población que sigue una distribución exponencial con parámetro θ , queremos encontrar el estimador centrado uniformemente de mínima varianza para $d(\theta) = \theta$.

$$X \sim \text{Exponencial}(\theta) \implies f(x|\theta) = \theta e^{-\theta x} \implies f(x_1, \dots, x_n|\theta) = \theta^n e^{-\theta \sum_{i=1}^n x_i}$$

Por el Teorema de Factorización de Fisher, $S = \sum_{i=1}^n X_i$ es suficiente para θ . Veamos ahora su completitud:

Como se sigue una distribución exponencial, podemos ver que pertenece a una familia exponencial uniparamétrica:

$$f(x_1, \dots, x_n|\theta) = \theta^n e^{-\theta \sum_{i=1}^n x_i} \implies \begin{cases} c(\theta)^n = \theta^n \\ h(\vec{x}) = 1 \\ q_1(\theta) = -\theta \\ T_1(\vec{x}) = \sum_{i=1}^n x_i \end{cases}$$

Por lo que evidentemente en la imagen de la función $q_1(\theta) = -\theta$ tiene un abierto en su imagen por lo que el estadístico S es completo para θ .

Por último, sabemos que $\bar{X} = \frac{S}{n} \implies E[\bar{X}] = \frac{1}{\theta} \implies \bar{X}$ es centrado para $\frac{1}{\theta}$ y por el Teorema de Lehmann-Scheffé, \bar{X} es el ECUMV para $\frac{1}{\theta}$.

Pero nosotros lo que queríamos es un estimador centrado para θ . Por lo que puede parecer intuitivo pensar que el estadístico que podría estar centrado para θ es $\frac{1}{\bar{X}}$:

$$E\left[\frac{1}{\bar{X}}\right] = n \cdot E\left[\frac{1}{S}\right] = n \cdot E\left[\frac{1}{\sum_{i=1}^n X_i}\right] = n \cdot \frac{n-1}{\theta} \implies S' = \frac{n(n-1)}{\sum_{i=1}^n X_i}$$

Por lo que S' es el ECUMV para θ

Ejemplo

Sea una m.a.s. de tamaño n de una población que sigue una distribución $Bernouilli(\theta)$, busquemos cual es el estimador centrado uniformemente de mínima varianza para $d_1(\theta) = \theta$ y para $d_2(\theta) = \theta(1-\theta)$:

$$X \sim \text{Bernouilli}(\theta) \implies f(x|\theta) = \theta^x (1-\theta)^{1-x} \implies f(x_1, \dots, x_n|\theta) = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i}$$

Por el Teorema de Factorización de Fisher, $S = \sum_{i=1}^n X_i$ es suficiente para θ . Veamos ahora su completitud:

$$f(x_1, \dots, x_n | \theta) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} = (1 - \theta)^n \cdot e^{\sum_{i=1}^n x_i \ln(\frac{\theta}{1-\theta})} \implies \begin{cases} c(\theta)^n = (1 - \theta)^n \\ h(\vec{x}) = 1 \\ q_1(\theta) = \ln(\frac{\theta}{1-\theta}) \\ T_1(\vec{x}) = \sum_{i=1}^n x_i \end{cases}$$

Por lo que evidentemente en la imagen de la función $q_1(\theta) = \ln(\frac{\theta}{1-\theta})$ tiene un abierto en su imagen por lo que el estadístico $S = \sum_{i=1}^n X_i$ es completo para θ .

1. Primero veamos el caso para $d_1(\theta) = \theta$: Sabemos que $E[S] = \sum_{i=1}^n E[X_i] = n\theta \implies \bar{X} = \frac{S}{n} \implies E[\bar{X}] = \theta \implies \bar{X}$ es centrado para θ . Por el Teorema de Lehmann-Scheffé, \bar{X} es el ECUMV para θ
2. Ahora veamos el caso para $d_2(\theta) = \theta(1 - \theta) = \theta - \theta^2$: A pesar de que pueda parecer lógico tomar la varianza muestral S^2 como estimador, no se cumplen las condiciones del Teorema de Lehmann-Scheffé para que sea un ECUMV, ya que tiene que depender únicamente de $S = \sum X_i$ y en S^2 aparecen términos cuadráticos, por lo que no es un ECUMV. Podemos intentar buscar otro:
NO LO ENTIENDO

Ejemplo

Sea una m.a.s. de tamaño n de una población que sigue una distribución $Normal(\mu, \sigma^2)$, veamos distintas casuísticas:

1. Tomemos que μ es conocida y σ^2 no, por lo que queremos buscar el ECUMV para σ^2 :

$$X \sim Normal(\mu, \sigma^2) \implies f(x | \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \implies f(x_1, \dots, x_n | \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}}$$

Por el Teorema de Factorización de Fisher, $S = \sum_{i=1}^n (X_i - \mu)^2$ es suficiente para σ^2 . Veamos ahora su completitud:

Sabemos que la distribución normal, en este caso, pertenece a la familia exponencial uni-paramétrica. Entonces sabemos que los estadísticos naturales asociados a las funciones paramétricas $q_i(\theta)$ son completos, si éstas últimas contienen un abierto en su imagen. En este caso, $q_1(\sigma^2) = \frac{1}{2\sigma^2}$ contiene un abierto en su imagen, por lo que S es completo para σ^2 . Veamos la esperanza de S :

$$\begin{aligned} E[S] &= E\left[\sum_{i=1}^n (X_i - \mu)^2\right] = \sum_{i=1}^n E[(X_i^2 + \mu^2 - 2X_i\mu)] = \sum_{i=1}^n E[X_i^2] + \sum_{i=1}^n \mu^2 - \sum_{i=1}^n 2\mu E[X_i] = \\ &= n(\sigma^2 + \mu^2) + n\mu^2 - 2n\mu^2 = n\sigma^2 \implies \frac{S}{n} \text{ es el ECUMV para } \sigma^2 \text{ por el T. de Lehmann-Scheffé} \end{aligned}$$

2. Ahora supongamos que μ y σ^2 son desconocidos, por lo que queremos encontrar el ECUMV para μ y σ^2 :

$$\begin{aligned} f(x | \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \implies f(x_1, \dots, x_n | \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}} = \\ &= f(x_1, \dots, x_n | \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\sum x_i^2 + n\mu^2 - 2\mu \sum x_i}{2\sigma^2}} \end{aligned}$$

Por el Teorema de Factorización de Fisher, $T(\vec{x}) = (\sum x_i, \sum x_i^2)$ es suficiente para (μ, σ^2) . Veamos ahora su completitud:

Sabemos que la distribución normal, en este caso, pertenece a la familia exponencial uniparamétrica. Entonces sabemos que los estadísticos naturales asociados a las funciones paramétricas $q_i(\theta)$ son completos, si éstas últimas contienen un abierto en su imagen. En este caso, $q_1(\mu, \sigma^2) = \mu$ y $q_2(\mu, \sigma^2) = \frac{1}{2\sigma^2}$ contienen un abierto en su imagen, por lo que T es completo para (μ, σ^2) .

Veamos la esperanza de T :

$$E[T] = E\left[\left(\sum X_i, \sum X_i^2\right)\right] = \left(\sum E[X_i], \sum E[X_i^2]\right) = (n\mu, n\sigma^2 + n\mu^2)$$

$$\Rightarrow \frac{\sum X_i}{n} = \bar{X} \text{ es el ECUMV para } \mu \text{ y } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ es el ECUMV para } \sigma^2$$

3. Ahora suponamos que μ es desconocido y σ^2 es conocido, por lo que queremos encontrar el ECUMV para μ : Por lo visto en el apartado anterior, podemos que por el Teorema de Fisher, podemos ver que el estadístico $T = \sum_{i=1}^n X_i$ es suficiente para μ .

También podríamos preguntarnos porqué no sería el estadístico $\vec{T} = (\sum X_i, \sum X_i^2)$, la respuesta está en que esta distribución normal pertenece a la familia exponencial uniparamétrica y por tanto tomamos la función dependiente del parámetro $q_1(\mu) = \mu$ y su estadístico natural asociado $T_1(\vec{x}) = \sum X_i$, el cual por las propiedades que vimos de las familias exponenciales es suficiente, y además completo. Veamos ahora si es insesgado:

$$E[T] = E\left[\sum X_i\right] = \sum E[X_i] = n\mu \Rightarrow \bar{X} = \frac{T}{n} \text{ es el ECUMV para } \mu$$

4.3 Cota para la varianza de un estimador

Definición 4.3.1 [Score]

Sea X variable aleatoria que sigue una distribución uniparamétrica continua o discreta y sea una m.a.s. de tamaño n tales que $f_\theta(x_1, \dots, x_n)$ es su función de densidad o masa. Entonces se define el score como:

$$S_\theta(x_1, \dots, x_n) = \frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) = \frac{1}{f_\theta(x_1, \dots, x_n)} \frac{\partial}{\partial \theta} f_\theta(x_1, \dots, x_n)$$

Observación 4.3.1

El score mide cuánta información hay en una observación sobre θ . Si el score es muy sensible a cambios sobre θ entonces hay una poca variación en θ hace que la verosimilitud cambie mucho y por tanto hay mucha información para estimar θ .

Definición 4.3.2 [Información de Fisher]

Sea X variable aleatoria que sigue una distribución uniparamétrica continua o discreta y sea una m.a.s. de tamaño n tales que $f_\theta(x_1, \dots, x_n)$ es su función de densidad o masa. Entonces se define la información de Fisher como:

$$I_n(\theta) = E_\theta [S_\theta(X_1, \dots, X_n)^2] = E_\theta \left[\left(\frac{\partial}{\partial \theta} \log f_\theta(X_1, \dots, X_n) \right)^2 \right]$$

Hemos visto hasta ahora el estimador centrado uniformemente de mínima varianza, que cómo su nombre indica, es el que tiene varianza más pequeña.

Si definiéramos una varianza mínima para los estimadores de un parámetro, es decir, si encontrásemos una cota para los estimadores, podríamos encontrar más fácilmente un estimador que tenga dicha varianza y por tanto sería el centrado uniformemente de mínima varianza.

Definición 4.3.3 [Condiciones de regularidad de Wolfowitz]

Sea $X \approx (\chi, \beta_\chi, F_\theta)_{\theta \in \Theta \subset \mathbb{R}}$ modelo estadístico uniparamétrico continuo (o discreto) y sea (X_1, \dots, X_n) muestra de $\{F_\theta, \theta \in \Theta\}$ siendo $f_\theta(x_1, \dots, x_n)$ su función de densidad (o de masa). Supongamos que se verifican las siguientes condiciones de regularidad:

1. Θ es un intervalo abierto de \mathbb{R}
2. $\text{Sop}(f_\theta) = \{(x_1, \dots, x_n) \in \chi^n : f_\theta(x_1, \dots, x_n) > 0\}$ no depende de θ
3. $\forall (x_1, \dots, x_n) \in \chi^n$ y $\forall \theta \in \Theta, \exists \frac{\partial}{\partial \theta} f_\theta(x_1, \dots, x_n)$
4. $\int_{\chi^n} \frac{\partial}{\partial \theta} f_\theta(x_1, \dots, x_n) dx_1 \dots dx_n = 0$
5. $I_n(\theta) = E \left[\left(\frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) \right)^2 \right] < \infty$ (cantidad de información de Fisher)

Teorema 4.3.1 [Cota de Fréchet-Cramér-Rao]

Si $T = T(X_1, \dots, X_n)$ es un estadístico unidimensional tal que $E_\theta[T^2] < \infty$, $E_\theta[T] = d(\theta)$ y

$$d'(\theta) = \int_{\chi^n} T(x_1, \dots, x_n) \frac{\partial}{\partial \theta} f_\theta(x_1, \dots, x_n) dx_1 \dots dx_n$$

entonces $d'(\theta)^2 \leq V_\theta(T)I_n(\theta)$, $\forall \theta \in \Theta$, con igualdad si y solo si existe una función $k(\theta)$ tal que

$$P_\theta \left((x_1, \dots, x_n) \in \chi^n : T(x_1, \dots, x_n) = d(\theta) + k(\theta) \frac{\partial}{\partial \theta} f_\theta(x_1, \dots, x_n) \right) = 1, \forall \theta \in \Theta$$

Demostración. $\exists d'(\theta)$ puesto que

$$d'(\theta) = \int_{\chi^n} T(x_1, \dots, x_n) \frac{\partial}{\partial \theta} \log(f_\theta(x_1, \dots, x_n)) f_\theta(x_1, \dots, x_n) dx_1 \dots dx_n = E_\theta \left[T \frac{\partial}{\partial \theta} \log f_\theta \right]$$

$$|d'(\theta)| \leq E_\theta \left[\left| T \frac{\partial}{\partial \theta} \log f_\theta \right| \right] \leq \sqrt{E_\theta[T^2] E_\theta \left[\left(\frac{\partial}{\partial \theta} \log f_\theta \right)^2 \right]} < \infty \text{ (desigualdad de Cauchy-Swartz)}$$

Además, $E_\theta \left[\frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) \right] = 0$ y por lo tanto,

$$V_\theta \left[\frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) \right] = E_\theta \left[\left(\frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) \right)^2 \right] = I_n(\theta)$$

$$\text{En efecto, } 0 = \int_{\chi^n} \frac{\partial}{\partial \theta} f_\theta(x_1, \dots, x_n) dx_1 \dots dx_n = \int_{\chi^n} \frac{\partial}{\partial \theta} \log(f_\theta(x_1, \dots, x_n)) f_\theta(x_1, \dots, x_n) dx_1 \dots dx_n$$

$$= E_\theta \left[\frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) \right]$$

Entonces, $\text{Cov}_\theta \left[T, \frac{\partial}{\partial \theta} \log f_\theta \right] = E \left[T \frac{\partial}{\partial \theta} \log f_\theta \right] = d'(\theta)$, y como $\rho_\theta^2 \left(T, \frac{\partial}{\partial \theta} \log f_\theta \right) = \frac{d'(\theta)^2}{V_\theta(T)V_\theta \left(\frac{\partial}{\partial \theta} \log f_\theta \right)} \leq 1$, se tiene que

$d'(\theta)^2 \leq V_\theta(T)I_n(\theta)$, con igualdad si y sólo si $\rho_\theta^2\left(T, \frac{\partial}{\partial\theta} \log f_\theta\right) = 1$, es decir, si y sólo si $T \stackrel{\text{c.s.}}{=} a + b \frac{\partial}{\partial\theta} \log f_\theta$, es decir, si y sólo si existe una función $k(\theta)$ tal que $P_\theta\left(T = d(\theta) + k(\theta) \frac{\partial}{\partial\theta} \log f_\theta\right) = 1$. En efecto, si $T \stackrel{\text{c.s.}}{=} a + b \frac{\partial}{\partial\theta} \log f_\theta$, entonces $d(\theta) = E_\theta[T] = ay$

$$d'(\theta) = E_\theta \left[T \frac{\partial}{\partial\theta} \log f_\theta \right] = E_\theta \left[a \frac{\partial}{\partial\theta} \log f_\theta + b \left(\frac{\partial}{\partial\theta} \log f_\theta \right)^2 \right] = b I_n(\theta),$$

y $b = \frac{d'(\theta)}{I_n(\theta)} = k(\theta)$

□

Ejemplo

Supongamos que tenemos un termómetro que mide la temperatura siguiendo una distribución normal con media μ y varianza σ^2 . Tomemos que $\sigma^2 = 0.01$, de manera que tenemos que:

| Medición | Valor |
|----------|-------|
| 1 | 21.0 |
| 2 | 21.1 |
| 3 | 20.9 |

Se puede ver que las mediciones están mu cerca entre ellas, por lo que podríamos confiar en la media de las observaciones.

Veamos la información de Fisher:

$$S_\mu(x_1, \dots, x_n) = \frac{\partial}{\partial\mu} \log f_\mu(x_1, \dots, x_n) = \frac{\partial}{\partial\mu} \left(-\frac{1}{2} \log(2\pi\sigma^2) - \frac{(x - \mu)^2}{2\sigma^2} \right) = \frac{x - \mu}{\sigma^2} \Rightarrow$$

Observación 4.3.2

Dado que se da que:

$$E_\mu \left[\frac{x - \mu}{\sigma^2} \right] = \frac{1}{\sigma^2} E_\mu [x - \mu] = \frac{1}{\sigma^2} (\mu - \mu) = 0 \Rightarrow I_n(\mu) = \text{Var}(S_\theta(\vec{x})) = E_\mu [S_\mu(\vec{x})^2]$$

$$\begin{aligned} I_n(\mu) &= E_\mu [S_\mu(X_1, \dots, X_n)^2] = E_\mu \left[\left(\frac{X - \mu}{\sigma^2} \right)^2 \right] = \text{Var}\left(\frac{X - \mu}{\sigma^2}\right) = \\ &= \frac{1}{\sigma^2} \cdot \text{Var}(X - \mu) = \frac{1}{\sigma^4} \cdot \text{Var}(X) = \frac{1}{\sigma^4} \cdot \sigma^2 = \frac{1}{\sigma^2} \end{aligned}$$

Entonces, la información de Fisher es:

$$I_n(\mu) = \frac{1}{\sigma^2} = \frac{1}{0.01} = 100$$

La cual es bastante alta lo que conlleva que podremos realiar una estimación mucho más precisa de μ . Asimismo, también podemos deducir por la cota de Fréchet-Cramér-Rao que ningun estimador centrado o insesgado puede tener varianza menor que:

$$\text{Var}(T) \geq \frac{1}{I_n(\mu)} = \frac{1}{100} = 0.01$$

Para poder estimar $\theta = \mu$ con precisión.

Corolario 4.3.1

Bajo las suposiciones anteriores, si T es un estadístico tal que $E_\theta[T] = d(\theta)$ y $V_\theta(T) = \frac{d'(\theta)^2}{I_n(\theta)}$, entonces T es ECUMV para $d(\theta)$

Corolario 4.3.2

Bajo las suposiciones anteriores, si (X_1, \dots, X_n) es m.a.s. (n) de $\{F_\theta, \theta \in \Theta\}$, entonces $I_n(\theta) = nI_1(\theta)$
Indicación: $f_\theta(x_1, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i)$

Proposición 4.3.1

Bajo las suposiciones anteriores, si la distribución de X pertenece a la familia exponencial uniparamétrica, es decir, $f_\theta(x) = c(\theta)h(x)e^{q_1(\theta)T_1(x)}$, con $q'_1(\theta)$ no nula, entonces el estadístico $\frac{1}{n} \sum_{i=1}^n T_1(X_i)$ alcanza la cota de FCR para $d(\theta) = -\frac{c'(\theta)}{c(\theta)q'_1(\theta)}$

Demostración. Sea una función de verosimilitud de la forma:

$$f_\theta(x_1, \dots, x_n) = c(\theta)^n \prod_{i=1}^n h(x_i) \cdot \exp\left(q_1(\theta) \sum_{i=1}^n T_1(x_i)\right)$$

Entonces, la derivada del logaritmo de la verosimilitud respecto de θ (el **score**) es:

$$\frac{\partial}{\partial \theta} \log f_\theta = n \cdot \frac{c'(\theta)}{c(\theta)} + q'_1(\theta) \sum_{i=1}^n T_1(x_i)$$

Dividiendo ambos lados entre n , se obtiene:

$$\frac{1}{n} \sum_{i=1}^n T_1(x_i) = -\frac{c'(\theta)}{c(\theta)q'_1(\theta)} + \frac{1}{nq'_1(\theta)} \cdot \frac{\partial}{\partial \theta} \log f_\theta$$

Es decir:

$$\frac{1}{n} \sum_{i=1}^n T_1(x_i) = a(\theta) + b(\theta) \cdot \frac{\partial}{\partial \theta} \log f_\theta$$

donde:

$$a(\theta) = -\frac{c'(\theta)}{c(\theta)q'_1(\theta)}, \quad b(\theta) = \frac{1}{nq'_1(\theta)}$$

Como la esperanza del score es cero bajo condiciones regulares, se tiene:

$$\mathbb{E}_\theta \left[\frac{1}{n} \sum_{i=1}^n T_1(x_i) \right] = a(\theta)$$

por lo que el estadístico $\frac{1}{n} \sum_{i=1}^n T_1(x_i)$ es **centrado** para el parámetro:

$$d(\theta) = -\frac{c'(\theta)}{c(\theta)q'_1(\theta)}$$

Además, como está escrito como combinación lineal del score, alcanza la **cota de Cramér-Rao** bajo condiciones regulares. \square

Ejemplo

Si $X \sim \text{Bin}(1, \theta)$, $T = \bar{X}$ alcanza la cota de FCR para $d(\theta) = \theta$ Primero calculemos el score:

$$\begin{aligned} S_\theta(x_1, \dots, x_n) &= \frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) = \frac{\partial}{\partial \theta} \left(\ln \left(\theta^{\sum x_i} \cdot (1 - \theta)^{n - \sum x_i} \right) \right) = \\ &= \frac{\partial}{\partial \theta} \left(\left(\sum x_i \right) \cdot \ln(\theta) + (n - \sum x_i) \cdot \ln(1 - \theta) \right) = \frac{\sum x_i}{\theta} + \frac{\sum x_i - n}{1 - \theta} \\ &= \frac{\sum x_i - n\theta}{\theta(1 - \theta)} = \frac{n\bar{X} - n\theta}{\theta(1 - \theta)} \end{aligned}$$

Ahora calculemos la información de Fisher:

$$\begin{aligned} I_n(\theta) &= E_\theta [S_\theta(X_1, \dots, X_n)^2] = E_\theta \left[\left(\frac{n\bar{X} - n\theta}{\theta(1 - \theta)} \right)^2 \right] = \frac{n^2}{\theta^2(1 - \theta)^2} E_\theta [(\bar{X} - \theta)^2] = \\ &= \frac{n^2}{\theta^2(1 - \theta)^2} \cdot E_\theta [\bar{X}^2 + \theta^2 - 2\theta\bar{X}] = \frac{n^2}{\theta^2(1 - \theta)^2} \cdot (E_\theta[\bar{X}^2] + \theta^2 - 2\theta E_\theta[\bar{X}]) = \\ &= \frac{n^2}{\theta^2(1 - \theta)^2} \cdot \left(\frac{\sigma^2}{n} + \mu^2 + \theta^2 - 2\theta\mu \right) \end{aligned}$$

Ahora sustituimos $\mu = \theta$ y $\sigma^2 = \theta(1 - \theta)$:

$$= \frac{n^2}{\theta^2(1 - \theta)^2} \cdot \left(\frac{\theta(1 - \theta)}{n} + \theta^2 + \theta^2 - 2\theta^2 \right) = \frac{n^2}{\theta^2(1 - \theta)^2} \cdot \frac{\theta(1 - \theta)}{n} = \frac{n}{\theta(1 - \theta)}$$

Entonces tenemos que la cota de FCR es:

$$\text{Var}(T) \geq \frac{1}{I_n(\theta)} = \frac{\theta(1 - \theta)}{n}$$

Por lo que el estimador \bar{X} es el ECUMV para $d(\theta) = \theta$.

Ejemplo

Sea una población que sigue una distribución $\text{Poisson}(\theta)$ y una m.a.s. de tamaño n de esa población. Se quiere estimar θ veamos cual debería ser la varianza del supuesto estimador: Primero calculemos el score:

$$\begin{aligned} S_\theta(x_1, \dots, x_n) &= \frac{\partial}{\partial \theta} \ln(f_\theta(x_1, \dots, x_n)) = \frac{\partial}{\partial \theta} \left(\ln \left(\frac{e^{-n\theta} \cdot \theta^{\sum x_i}}{\prod x_i!} \right) \right) = \\ &= \frac{\partial}{\partial \theta} \left(-n\theta + \left(\sum x_i \right) \ln(\theta) - \sum \ln(x_i!) \right) = \frac{\sum x_i}{\theta} - n = \frac{n\bar{X}}{\theta} - n \end{aligned}$$

Ahora veamos la información de Fisher:

$$\begin{aligned} I_n(\theta) &= E_\theta [S_\theta(X_1, \dots, X_n)^2] = E_\theta \left[\left(\frac{n\bar{X}}{\theta} - n \right)^2 \right] = \\ &= E_\theta \left[n^2 \left(\frac{\bar{X}^2}{\theta^2} + 1 - \frac{2\bar{X}}{\theta} \right) \right] = n^2 \left(\frac{E_\theta[\bar{X}^2]}{\theta^2} + 1 - \frac{2E_\theta[\bar{X}]}{\theta} \right) = n^2 \left(\frac{\mu^2 + \frac{\sigma^2}{n}}{\theta^2} + 1 - \frac{2\mu}{\theta} \right) = \end{aligned}$$

Sustituyendo $\mu = \theta$ y $\sigma^2 = \theta$:

$$= n^2 \left(\frac{\theta^2}{\theta^2} + \frac{1}{n\theta} + 1 - \frac{2\theta}{\theta} \right) = n^2 \left(\frac{1}{n\theta} \right) = \frac{n}{\theta}$$

Por tanto la cota de FCR es:

$$\text{Var}(T) \geq \frac{\theta}{n}$$

Luego el estimador \bar{X} coincide con la varianza y por tanto es el ECUMV para $d(\theta) = \theta$.

Asimismo, existe una forma alternativa de calcular la información de Fisher, y es a través de la fórmula:

$$I_n(\theta) = -E_\theta \left[\frac{\partial^2}{\partial \theta^2} \log f_\theta(X_1, \dots, X_n) \right]$$

En nuestro caso sería:

$$\begin{aligned} I_n(\theta) &= -E_\theta \left[\frac{\partial^2}{\partial \theta^2} \log f_\theta(X_1, \dots, X_n) \right] = -E_\theta \left[\frac{\partial}{\partial \theta} \left(\frac{n\bar{X}}{\theta} - n \right) \right] = -E_\theta \left[\frac{-n\bar{X}}{\theta^2} \right] = \\ &= \frac{n}{\theta^2} E_\theta [\bar{X}] = \frac{n}{\theta^2} \cdot \theta = \frac{n}{\theta} \end{aligned}$$

Por lo que la información de Fisher coincide con la que hemos calculado anteriormente.

Definición 4.3.4 [Estimador Eficiente]

Diremos que un estimador es eficiente para $d(\theta)$ si es centrado para $d(\theta)$ y su varianza alcanza la cota de FCR

En general, se llama eficiencia de un estimador centrado de $d(\theta)$ a

$$ef(T, d(\theta)) = \frac{d'(\theta)^2}{I_n(\theta)V_\theta(T)} \leq 1$$

Ejemplo

Sea $X \sim \text{Poisson}(\theta)$ y $T = \bar{X}$, entonces T es un estimador eficiente para θ

Primero calculemos la varianza de T :

$$V_\theta(T) = E_\theta [\bar{X}^2] - E_\theta [\bar{X}]^2 = \frac{\sigma^2}{n} + \mu^2 - \mu^2 = \frac{\sigma^2}{n} = \frac{\theta}{n}$$

Por lo que la eficiencia de T es:

$$ef(T, d(\theta)) = \frac{d'(\theta)^2}{I_n(\theta)V_\theta(T)} = \frac{\theta^2}{\frac{n}{\theta} \cdot \frac{\theta}{n}} = 1$$

Por lo que el estimador \bar{X} es eficiente para θ .

4.4 Métodos de construcción de estimadores

4.4.1 Método de los momentos

Definición 4.4.1 [Método de los momentos]

Sea X una variable aleatoria que sigue una distribución uniparamétrica continua o discreta y sea una m.a.s. de tamaño n tales que $f_\theta(x_1, \dots, x_n)$ es su función de densidad o masa. Entonces se define el método de los momentos como el método que consiste en igualar los momentos muestrales a los

momentos poblacionales.

1. Si $E[X^k]$ es el momento poblacional respecto al origen de orden k , entonces el estimador por el método de los momentos del momento poblacional respecto al origen de orden k es:

$$\hat{\theta}_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

2. Si $E[(X - \alpha_1)^k]$ es el momento poblacional respecto a la media de orden k , entonces el estimador por el método de los momentos del momento poblacional respecto a la media de orden k es:

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

Ejemplo

Supongamos que $X \sim \text{Gamma}(a, p)$, donde $a > 0$ es el parámetro de forma y $p > 0$ es el parámetro de escala. La función de densidad es:

$$f(x) = \frac{1}{\Gamma(a)p^a} x^{a-1} e^{-x/p}, \quad x > 0$$

Sean X_1, X_2, \dots, X_n una muestra aleatoria simple de esta distribución. Aplicamos el método de los momentos, que consiste en igualar los momentos teóricos con los empíricos.

Momentos teóricos

Sabemos que:

$$\mathbb{E}[X] = ap, \quad \text{Var}(X) = ap^2$$

Momentos muestrales

Calculamos los momentos muestrales:

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

Igualación de momentos

Igualamos los momentos teóricos y muestrales:

$$\begin{cases} \bar{X} = ap \\ S^2 = ap^2 \end{cases}$$

Despejamos a y p :

De la primera ecuación:

$$a = \frac{\bar{X}}{p}$$

Sustituimos en la segunda:

$$S^2 = \left(\frac{\bar{X}}{p}\right) p^2 = \bar{X}p \Rightarrow p = \frac{S^2}{\bar{X}}$$

Luego, sustituimos en la expresión de a :

$$a = \frac{\bar{X}}{p} = \frac{\bar{X}^2}{S^2}$$

Conclusión

El estimador de $\theta = (a, p)$ por el método de los momentos es:

$$\hat{a} = \frac{\bar{X}^2}{S^2}, \quad \hat{p} = \frac{S^2}{\bar{X}}$$

Ejemplo

Para una m.a.s. de tamaño n de una población de $Bernoulli(\theta)$ en este caso $r = 1$, $\alpha_1(\theta) = \theta$ y $a_1 = \bar{X}$ por lo tanto el estimador \bar{X} es el estimador de los momentos para θ .

Ejemplo

Para una población con distribución de $Poisson(\theta)$ se necesita una sola ecuación, $\alpha_1(\theta) = \theta$ y $a_1 = \bar{X}$ por lo tanto el estimador \bar{X} es el estimador de los momentos para θ .

Ejemplo

Para una población X con distribución $N(\mu, \sigma)$ con ambos parámetros desconocidos $\vec{\theta} = (\mu, \sigma)$ es $r = 2$ y se debe resolver el sistema:

$$\begin{cases} \mu = \bar{X} \\ \sigma^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \end{cases}$$

Lo que lleva a que los estimadores por el método de los momentos son: \bar{X} y $\sqrt{b_2}$

4.4.2 Método de máxima verosimilitud

Definición 4.4.2 [Estimador de máxima verosimilitud]

Dada $C(X_1, \dots, X_n) \subset \Theta$ región aleatoria del espacio paramétrico tal que $P_\theta\{\theta \in C(X_1, \dots, X_n)\} \geq 1 - \alpha \forall \theta \in \Theta$, entonces para cada posible valor $(x_1, \dots, x_n) \in \chi^n$, $C(x_1, \dots, x_n)$ se denomina región de confianza para θ de nivel $1 - \alpha$

Método de máxima verosimilitud

Supongamos que una urna contiene 6 bolas entre blancas y negras, no todas del mismo color, pero se ignora cuantas hay de cada uno. Para tratar de adivinar la composición de la urna se permiten dos extracciones con reemplazamiento de la misma y resultó que ninguna de ellas fue blanca. Dar una estimación de la probabilidad θ de que una bola extraída aleatoriamente de dicha urna sea blanca

$$\theta \in \Theta = \left\{ \frac{1}{6}, \frac{2}{6}, \frac{3}{6}, \frac{4}{6}, \frac{5}{6} \right\}$$

$T = X_1 + X_2 \equiv$ n° de blancas en las dos extracciones C.R. de la urna $\sim \text{Bin}(2, \theta)$ y $f_\theta(t) = \binom{2}{t} \theta^t (1-\theta)^{2-t}$, $t = 0, 1, 2$

| | | | | | |
|--------------------------------|-------|-------|------|-------|-------|
| θ | 1/6 | 2/6 | 3/6 | 4/6 | 5/6 |
| $f_\theta(0) = (1 - \theta)^2$ | 0.694 | 0.444 | 0.25 | 0.111 | 0.027 |

Por lo tanto, la estimación $\hat{\theta}(0) = \frac{1}{6}$ y el estimador

$$\hat{\theta} = \hat{\theta}(T) = \begin{cases} 1/6 & \text{si } T = 0 \\ 1/2 & \text{si } T = 1 \\ 5/6 & \text{si } T = 2 \end{cases}$$

es el estimador de máxima verosimilitud (EMV)

Sea (X_1, \dots, X_n) una muestra con $f_\theta(x_1, \dots, x_n) = f(x_1, \dots, x_n | \theta)$ función de densidad (o de masa), $\theta \in \Theta \subset \mathbb{R}^\ell$. Denotemos por $L(\theta | x_1, \dots, x_n) = f(x_1, \dots, x_n | \theta)$ a la función de verosimilitud de la muestra. Un estimador $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ se denomina estimador de máxima verosimilitud (EMV) de θ , sí y sólo sí

- (1) $\hat{\theta}(x_1, \dots, x_n) \in \Theta, \forall (x_1, \dots, x_n) \in \chi^n$
 - (2) $L(\hat{\theta} | x_1, \dots, x_n) = \sup_{\theta \in \Theta} L(\theta | x_1, \dots, x_n), \forall (x_1, \dots, x_n) \in \chi^n$
- o equivalentemente, sí y sólo sí
- (1) $\hat{\theta}(x_1, \dots, x_n) \in \Theta, \forall (x_1, \dots, x_n) \in \chi^n$
 - (2) $\log L(\hat{\theta} | x_1, \dots, x_n) = \sup_{\theta \in \Theta} \log L(\theta | x_1, \dots, x_n), \forall (x_1, \dots, x_n) \in \chi^n$

Si f_θ es una función derivable respecto a θ en el interior del espacio paramétrico Θ , la forma usual de determinar el estimador de máxima verosimilitud es examinar primero los máximos relativos de f_θ , para compararlos después, con los valores sobre la frontera de Θ . Ello conduce a resolver las ecuaciones de verosimilitud:

$$\frac{\partial}{\partial \theta_j} \log L(\theta | x_1, \dots, x_n) = 0, j = 1, \dots, \ell$$

(en el supuesto de que $\theta = (\theta_1, \dots, \theta_\ell)$ sea un parámetro ℓ -dimensional), seleccionando las soluciones correspondientes a un máximo de f_θ , es decir aquellas en las que la matriz hessiana $H = \left(\frac{\partial}{\partial \theta_i} \frac{\partial}{\partial \theta_j} \log L(\theta | x_1, \dots, x_n) \right)_{i,j=1, \dots, \ell}$ sea definida negativa.

Observación 4.4.1

1. El EMV $\hat{\theta}$ no tiene por qué existir
2. El EMV $\hat{\theta}$ no tiene por qué ser único
3. El EMV $\hat{\theta}$ no tiene por qué ser centrado
4. El EMV $\hat{\theta}$ no tiene por qué ser suficiente, pero si $S = S(X_1, \dots, X_n)$ es suficiente para θ , entonces $\hat{\theta} = \hat{\theta}(S)$
5. Invariancia: Si $\hat{\theta}$ es el EMV de θ , entonces $h(\hat{\theta})$ es el EMV de $h(\theta)$
6. Bajo ciertas condiciones de regularidad, si (X_1, \dots, X_n) es m.a.s. (n) y $\theta \in \mathbb{R}$, entonces $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow[n \rightarrow \infty]{d} N\left(0, \frac{1}{\sqrt{I_1(\theta)}}\right)$ y por lo tanto, $\hat{\theta}$ es asintóticamente insesgado para θ y asintóticamente eficiente

Ejemplo

Si $X \sim \text{Bin}(1, \theta)$, $\hat{\theta} = \bar{X}$ es el EMV para $\theta \in [0, 1]$ basado en una m.a.s.(n)

Ejemplo

Si $X \sim N(\mu, \sigma)$, $\hat{\theta} = (\bar{X}, \sigma_n) = \left(\frac{1}{n} \sum_{i=1}^n X_i, \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} \right)$ es el EMV para $\theta = (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+$ basado en una m.a.s. (n). Además, si $n = 1$, no existe el EMV para σ

Ejemplo

Si $X \sim U(0, \theta)$, $\hat{\theta} = X_{(n)}$ es el EMV para $\theta > 0$ basado en una m.a.s. (n) y $E_\theta [X_{(n)}] = \frac{n}{n+1} \theta$

Ejemplo

Si $X \sim U\left(\theta - \frac{1}{2}, \theta + \frac{1}{2}\right)$, entonces cualquier valor en el intervalo $(X_{(n)} - \frac{1}{2}, X_{(1)} + \frac{1}{2})$ es un EMV para θ . En particular, $\frac{X_{(1)} + X_{(n)}}{2}$ no es suficiente y es EMV para θ

Propiedades del EMV**Proposición 4.4.1**

Si $S = S(X_1, \dots, X_n)$ es suficiente para θ , entonces el EMV $\hat{\theta} = \hat{\theta}(S)$ es función de S

Demostración. $f_\theta(x_1, \dots, x_n) = h(x_1, \dots, x_n) g_\theta(S(x_1, \dots, x_n))$ (teorema de factorización) □

Proposición 4.4.2

Si $\hat{\theta}$ es el EMV de θ y $h: \Theta \rightarrow \Lambda$, entonces $\hat{\lambda} = h(\hat{\theta})$ es el EMV de $\lambda = h(\theta)$ respecto de la función de verosimilitud inducida,

$$M(\lambda | x_1, \dots, x_n) = \sup_{\theta \in \Theta_\lambda} L(\theta | x_1, \dots, x_n) \\ \Theta_\lambda = \{\theta \in \Theta : h(\theta) = \lambda\}$$

Demostración.

$$\sup_{\lambda \in \Lambda} M(\lambda | x_1, \dots, x_n) = \sup_{\lambda \in \Lambda} \sup_{\theta \in \Theta_\lambda} L(\theta | x_1, \dots, x_n) = \sup_{\theta \in \Theta} L(\theta | x_1, \dots, x_n) \\ = L(\hat{\theta} | x_1, \dots, x_n)$$

Finalmente, como $\hat{\theta} \in \Theta_{\hat{\lambda}}$, se tiene que

$$M(\hat{\lambda} | x_1, \dots, x_n) = \sup_{\theta \in \Theta_{\hat{\lambda}}} L(\theta | x_1, \dots, x_n) \geq L(\hat{\theta} | x_1, \dots, x_n), \text{ y por lo tanto } \sup_{\lambda \in \Lambda} M(\lambda | x_1, \dots, x_n) = M(\hat{\lambda} | x_1, \dots, x_n)$$

□

4.5 Propiedades asintóticas de los estimadores de máxima verosimilitud

Proposición 4.5.1 [Desigualdad de Yelsin]

Sea X una v.a. integrable y g una función convexa tal que $g(X)$ es integrable. Entonces $g(E[X]) \leq E[g(X)]$. Además, si g es estrictamente convexa entonces $g(E[X]) < E[g(X)]$

Consideremos $X \approx (\chi, \beta_\chi, F_\theta)_{\theta \in \Theta \subset \mathbb{R}}$ modelo estadístico uniparamétrico continuo (o discreto) y sea (X_1, \dots, X_n) muestra de $\{F_\theta, \theta \in \Theta\}$ siendo $f_\theta(x_1, \dots, x_n)$ su función de densidad (o de masa). Supongamos que se verifican las condiciones de regularidad

Denotemos por θ_0 el verdadero valor del parámetro (desconocido).

Lema 4.5.1

- (1) $E_{\theta_0} [\log f_\theta(x_1, \dots, x_n)] < E_{\theta_0} [\log f_{\theta_0}(x_1, \dots, x_n)], \forall \theta \neq \theta_0$
- (2) Existe un entorno de $\theta_0, E(\theta_0)$, donde la función de θ $E_{\theta_0} \left[\frac{\partial}{\partial \theta} \log f_\theta(x_1, \dots, x_n) \right] = J(\theta_0, \theta)$ es estrictamente decreciente

Demostración. 1. Como el logaritmo es una función estrictamente cóncava y el soporte de la distribución es independiente del parámetro, si $\theta \neq \theta_0, E_{\theta_0} \left[\log \frac{f_\theta(x_1, \dots, x_n)}{f_{\theta_0}(x_1, \dots, x_n)} \right] < \log E_{\theta_0} \left[\frac{f_\theta(x_1, \dots, x_n)}{f_{\theta_0}(x_1, \dots, x_n)} \right] = 0$. Por lo tanto,

$$E_{\theta_0} [\log f_\theta(x_1, \dots, x_n)] - E_{\theta_0} [\log f_{\theta_0}(x_1, \dots, x_n)] < 0, y$$

$$E_{\theta_0} [\log f_\theta(x_1, \dots, x_n)] \text{ presenta un máximo estricto en } \theta_0$$

2. Como $f_\theta(x_1, \dots, x_n)$ es derivable respecto a θ , $\frac{\partial}{\partial \theta} E_{\theta_0} [\log f_\theta(x_1, \dots, x_n)] = 0$ en $\theta = \theta_0 \Rightarrow J(\theta_0, \theta_0) = 0$

$$\frac{\partial^2}{\partial \theta^2} E_{\theta_0} [\log f_\theta(x_1, \dots, x_n)] < 0 \text{ en } \theta = \theta_0 \Rightarrow \frac{\partial}{\partial \theta} J(\theta_0, \theta) < 0 \text{ en } \theta = \theta_0$$

Por lo tanto, existe un entorno de $\theta_0, E(\theta_0)$, donde la función $J(\theta_0, \theta)$ es estrictamente decreciente

□

Teorema 4.5.1 [Primer teorema de convergencia del EMV (consistencia)]

Para una m.a.s. (n) , existe una sucesión de v.a. $\hat{\theta}_n$, para cada n solución de la ecuación de verosimilitud, $\frac{\partial}{\partial \theta} \log L(\theta | x_1, \dots, x_n) = 0$, que converge c.s. al verdadero valor del parámetro θ_0 .

Demostración. Para una m.a.s. (n) , $f_\theta(x_1, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i)$

Entonces, $\frac{\partial}{\partial \theta} \log L(\theta | x_1, \dots, x_n) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f_\theta(x_i) = n z_n(\theta)$

Además, $z_n(\theta) \xrightarrow[n \rightarrow \infty]{c.s.} E_{\theta_0} \left[\frac{\partial}{\partial \theta} \log f_\theta(x) \right] = J(\theta_0, \theta)$ que es estrictamente decreciente en el entorno $E(\theta_0)$ y $J(\theta_0, \theta_0) = 0$

Por lo tanto, podemos elegir $\varepsilon > 0$ tal que $[\theta_0 - \varepsilon, \theta_0 + \varepsilon] \subset E(\theta_0)$ y $J(\theta_0 - \varepsilon, \theta_0) > 0, J(\theta_0 + \varepsilon, \theta_0) < 0$

Entonces,

$$z_n(\theta_0 - \varepsilon) \xrightarrow[n \rightarrow \infty]{c.s.} J(\theta_0, \theta_0 - \varepsilon) > 0 \text{ y } z_n(\theta_0 + \varepsilon) \xrightarrow[n \rightarrow \infty]{c.s.} J(\theta_0, \theta_0 + \varepsilon) < 0$$

Como $\frac{\partial}{\partial \theta} f_\theta$ es continua, $\exists \hat{\theta}_n \in (\theta_0 - \varepsilon, \theta_0 + \varepsilon)$ tal que $z_n(\hat{\theta}_n) = 0$. Por lo tanto, $\hat{\theta}_n$ es una solución de la ecuación de verosimilitud tal que $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{c.s.} \theta_0$

□

Observación 4.5.1

El teorema anterior es útil, sólo cuando existe una única raíz de la ecuación de verosimilitud, ya que entonces esa raíz ha de ser el estimador consistente. En caso contrario, no se sabrá, cual de las raíces al variar n , va a dar lugar a la sucesión consistente

Además, bajo ciertas condiciones de regularidad,

$$Y_n(\theta) = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \log f_\theta(x_i) \xrightarrow[n \rightarrow \infty]{c.s.} E_{\theta_0} \left[\frac{\partial^2}{\partial \theta^2} \log f_\theta(x) \right] = I(\theta_0, \theta), \text{ con } I(\theta_0, \theta_0) = -I_1(\theta_0) < 0$$

Como $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{c.s.} \theta_0$, si $\frac{\partial^2}{\partial \theta^2} f_\theta(x)$ es continua en θ uniformemente $\forall x$, entonces $Y_n(\hat{\theta}_n) - Y_n(\theta_0) \xrightarrow[n \rightarrow \infty]{c.s.} 0$, y por lo tanto $Y_n(\hat{\theta}_n) \xrightarrow[n \rightarrow \infty]{c.s.} -I_1(\theta_0) < 0$ y a partir de un n en adelante la sucesión de v.a. solución de la ecuación de verosimilitud son una sucesión de máximos

Teorema 4.5.2 [Segundo teorema de convergencia del EMV (convergencia en ley)]

Si se cumplen las condiciones de regularidad y $\frac{\partial^2}{\partial \theta^2} f_\theta(x)$ es continua en θ uniformemente $\forall x$, entonces

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{d} N\left(0, \frac{1}{\sqrt{I_1(\theta)}}\right)$$

Demostración. Si denotamos por $\varphi(x, \theta) = \frac{\partial}{\partial \theta} \log f_\theta(x)$, entonces la ecuación de verosimilitud $\frac{\partial}{\partial \theta} \log L(\theta | x_1, \dots, x_n) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f_\theta(x_i) = \sum_{i=1}^n \varphi(x_i, \theta) = n z_n(\theta) = 0$ en $\hat{\theta}_n$, con $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{c.s.} \theta_0$, y por lo tanto $\sum_{i=1}^n \varphi(x_i, \hat{\theta}_n) = 0$

$$\sum_{i=1}^n \varphi(x_i, \theta_0) = \sum_{i=1}^n \varphi(x_i, \hat{\theta}_n) + (\theta_0 - \hat{\theta}_n) \sum_{i=1}^n \frac{\partial}{\partial \theta} \varphi(x_i, \theta^*) \text{ con } \theta^* \in (\theta_0, \hat{\theta}_n)$$

$$\text{Entonces, } \hat{\theta}_n - \theta_0 = \frac{\sum_{i=1}^n \varphi(x_i, \theta_0)}{-\sum_{i=1}^n \frac{\partial}{\partial \theta} \varphi(x_i, \theta^*)}$$

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi(x_i, \theta_0)}{-\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial \theta} \varphi(x_i, \theta^*)} = \frac{A_n}{B_n}$$

$$E_{\theta_0}[A_n] = \frac{1}{\sqrt{n}} E_{\theta_0} \left[\sum_{i=1}^n \varphi(x_i, \theta_0) \right] = \frac{1}{\sqrt{n}} n E_{\theta_0} [\varphi(x, \theta_0)] = 0$$

$$V_{\theta_0}(A_n) = \frac{1}{n} V_{\theta_0} \left(\sum_{i=1}^n \varphi(x_i, \theta_0) \right) = \frac{1}{n} n V_{\theta_0}(\varphi(x, \theta_0))$$

$$= E_{\theta_0} [\varphi(x, \theta_0)^2] = I_1(\theta_0). \text{ Por lo tanto, } A_n \xrightarrow[n \rightarrow \infty]{d} N\left(0, \sqrt{I_1(\theta_0)}\right)$$

Además, $-B_n - E_{\theta_0} \left[\frac{\partial}{\partial \theta} \varphi(x, \theta^*) \right] \xrightarrow[n \rightarrow \infty]{c.s.} 0$ y como $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{c.s.} \theta_0$ y $\frac{\partial^2}{\partial \theta^2} f_\theta(x)$ es

continua en θ uniformemente $\forall x$, se tiene que $-B_n + I_1(\theta_0) \xrightarrow[n \rightarrow \infty]{c.s.} 0$ Por lo tanto, $\frac{A_n}{B_n} \xrightarrow[n \rightarrow \infty]{d} \frac{1}{I_1(\theta_0)} N\left(0, \sqrt{I_1(\theta_0)}\right) \sim N\left(0, \sqrt{\frac{1}{I_1(\theta_0)}}\right)$

□

Observación 4.5.2 [Simplificación de las condiciones de regularidad]

Las condiciones de regularidad del teorema anterior pueden simplificarse a las siguientes:

1. $\phi(x, \theta) = \frac{\partial}{\partial \theta} \log f_\theta(x)$ tiene dos derivadas continuas respecto de θ .

2. $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{P} \theta_0$

3. $E_{\theta_0}[\varphi^2(x, \theta_0)] = I_1(\theta_0) < \infty$

4. $E_{\theta_0} \left[\frac{\partial}{\partial \theta} \varphi(x, \theta_0) \right] = -I_1(\theta_0) \neq 0$

5. $\exists \varepsilon > 0$ y $M(x)$ tales que $\sup_{|\theta - \theta_0| \leq \varepsilon} \left| \frac{\partial^2}{\partial \theta^2} \varphi(x, \theta) \right| \leq M(x)$ con $M(x) < \infty$

5 Estimación por Regiones de Confianza

5.1 Región de confianza

Sea $X \approx (\chi, \beta_\chi, F_\theta)_{\theta \in \Theta \subset \mathbb{R}^k}$ modelo estadístico k-paramétrico y (X_1, \dots, X_n) muestra de $\{F_\theta, \theta \in \Theta\}$. Denotemos por P_θ a la ley de probabilidad de la muestra

Región de confianza

Sea $C(X_1, \dots, X_n) \subset \Theta$ una región aleatoria del espacio paramétrico tal que

$$P_\theta \{ \theta \in C(X_1, \dots, X_n) \} \geq 1 - \alpha, \forall \theta \in \Theta$$

Entonces, para cada $(x_1, \dots, x_n) \in \chi^n$, $C(x_1, \dots, x_n)$ se denomina región de confianza para θ de nivel $1 - \alpha$

5.2 Intervalos de confianza

Definición 5.2.1

Sea $h : \Theta \rightarrow \mathbb{R}$, $\alpha \in (0, 1)$ y $T_1 = T_1(X_1, \dots, X_n) : \chi^n \rightarrow \mathbb{R}$ y $T_2 = T_2(X_1, \dots, X_n) : \chi^n \rightarrow \mathbb{R}$, $T_1 \leq T_2$, dos estadísticos unidimensionales tales que

$$P_\theta \{ T_1(X_1, \dots, X_n) \leq h(\theta) \leq T_2(X_1, \dots, X_n) \} \geq 1 - \alpha, \forall \theta \in \Theta$$

Entonces, para cada $(x_1, \dots, x_n) \in \chi^n$, $(T_1(x_1, \dots, x_n), T_2(x_1, \dots, x_n))$ se denomina intervalo de confianza para $h(\theta)$ de nivel $1 - \alpha$

Observación 5.2.1

Siempre es deseable hacer que la medida de la región de confianza sea mínima, entre todas las del mismo grado de confianza $1 - \alpha$

5.3 Métodos de obtención de intervalos de confianza

Método de la cantidad pivotal

Definición 5.3.1 [Cantidad pivotal]

Una v.a. $T = T(X_1, \dots, X_n, \theta) : \chi^n \times \Theta \rightarrow \mathbb{R}$ es una cantidad pivotal sí y sólo sí su distribución en el muestreo no depende de θ

Si $T = T(X_1, \dots, X_n, \theta)$ es una cantidad pivotal, fijado cualquier nivel de confianza $1 - \alpha$, $\alpha \in (0, 1)$, se pueden determinar dos constantes, $c_1(\alpha)$ y $c_2(\alpha) \in \mathbb{R}$ (que no son únicas), tales que

$$P_\theta \{ c_1(\alpha) \leq T(X_1, \dots, X_n, \theta) \leq c_2(\alpha) \} \geq 1 - \alpha, \forall \theta \in \Theta$$

Si para cada $(x_1, \dots, x_n) \in \chi^n$, $c_1(\alpha) \leq T(x_1, \dots, x_n, \theta) \leq c_2(\alpha) \Leftrightarrow T_1(x_1, \dots, x_n, \alpha) \leq h(\theta) \leq T_2(x_1, \dots, x_n, \alpha)$, entonces $(T_1(x_1, \dots, x_n, \alpha), T_2(x_1, \dots, x_n, \alpha))$ es un intervalo de confianza para $h(\theta)$ de nivel $1 - \alpha$

Intervalos de confianza asociados a la distribución normal

Ejemplo

Para una m.a.s. (n) de $X \sim N(\mu, \sigma)$, σ conocida,

$$IC_{1-\alpha}(\mu) = \left(\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

es un intervalo de confianza para μ de nivel $1 - \alpha$, $\alpha \in (0, 1)$

Para demostrar esto, utilizaremos el teorema de fisher, segun el cual $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$, entonces:

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

$$P(z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}}) \geq 1 - \alpha \implies P(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) \geq 1 - \alpha$$

Y asi finalmente se deduce el intervalo dado.

Ejemplo

Para una m.a.s. (n) de X , si $\theta \in \mathbb{R}$ y la función de distribución de la población $F_{\theta}(x)$, como función en x es continua y estrictamente monótona $\forall \theta$, y como función de θ es continua y estrictamente monótona $\forall x$, entonces $T = -2 \sum_{i=1}^n \ln F_{\theta}(X_i) \sim \chi_{2n}^2$ constituye una cantidad pivotal y permite obtener un intervalo de confianza para θ

Para ver este primero fijemonos en que como $F_{\theta}(x)$ es continua y estrictamente monótona, entonces: $Y = F(X|\theta) \sim U(0, 1)$. Ahora hagamos el cambio:

$$z = -2 \ln(F(X|\theta)) \implies y = e^{-z/2}, \quad \left| \frac{dy}{dz} \right| = \frac{1}{2} e^{-z/2} \implies g(z) = \frac{1}{2} e^{-z/2}$$

Que coincide con la densidad de una variable aleatoria χ^2 con 2 grados de libertad.

Asi, sumandolas:

$$T = -2 \sum_{i=1}^n \ln F(X|\theta) \sim \chi_{2n}^2$$

Ahora podemos utilizar los cuantiles para obtener el intervalo de confianza:

$$P(\chi_{2n; 1-\frac{\alpha}{2}} \leq T \leq \chi_{2n; \frac{\alpha}{2}}) \geq 1 - \alpha$$

Por tanto utilizando la monotonía estricta en θ de $F_{\theta}(x)$, y la invertibilidad de las ecuaciones, podemos obtener:

$$\chi_{2n; 1-\frac{\alpha}{2}} = T(x_1, \dots, x_n; \underline{\theta}(x_1, \dots, x_n))$$

$$\chi_{2n; \frac{\alpha}{2}} = T(x_1, \dots, x_n; \bar{\theta}(x_1, \dots, x_n))$$

Finalmente obtenemos el intervalo de confianza:

$$(\underline{\theta}(x_1, \dots, x_n), \bar{\theta}(x_1, \dots, x_n))$$

Ejemplo

Construir un intervalo de confianza para θ por el método de la cantidad pivotal basado en una m.a.s. (n) de $f_{\theta}(x) = \theta x^{\theta-1} I_{(0,1)}(x)$, con $\theta > 0$

Utilizando el ejemplo anterior, y como:

$$F_{\theta}(x) = \begin{cases} 0 & x \leq 0 \\ x^{\theta} & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$$

Es una función continua y estrictamente monótona, entonces:

$$P\left(\chi_{2n;1-\frac{\alpha}{2}}^2 < -2\theta \sum_{i=1}^n \ln X_i < \chi_{2n;\frac{\alpha}{2}}^2 \middle| \theta\right) = 1 - \alpha \quad \forall \theta \in \Theta$$

Y por tanto:

$$P\left\{\frac{\chi_{2n;1-\frac{\alpha}{2}}^2}{-2 \sum_{i=1}^n \ln X_i} < \theta < \frac{\chi_{2n;\frac{\alpha}{2}}^2}{-2 \sum_{i=1}^n \ln X_i} \middle| \theta\right\} = 1 - \alpha \quad \forall \theta \in \Theta$$

Método de Neyman

Sea $T = T(X_1, \dots, X_n) : \chi^n \rightarrow \mathbb{R}$ un estimador de $h(\theta)$ y denotemos por $g_{\theta}(t)$ a su función de densidad (o de masa)

Fijado un nivel de confianza $1 - \alpha, \alpha \in (0, 1)$ y $\alpha_1, \alpha_2 \in (0, 1)$ tales que $\alpha = \alpha_1 + \alpha_2$, para cada $\theta \in \Theta$ se pueden determinar dos valores $c_1(\theta, \alpha_1)$ y $c_2(\theta, \alpha_2) \in \mathbb{R}$, tales que $P_{\theta}\{T < c_1(\theta, \alpha_1)\} \leq \alpha_1$ y $P_{\theta}\{T > c_2(\theta, \alpha_2)\} \leq \alpha_2$. Entonces, para cada $\theta \in \Theta$

$$P_{\theta}\{c_1(\theta, \alpha_1) \leq T \leq c_2(\theta, \alpha_2)\} \geq 1 - \alpha_2 - \alpha_1 = 1 - \alpha$$

Si para cada $t \in \mathbb{R}, c_1(\theta, \alpha_1) \leq T(t) \leq c_2(\theta, \alpha_2) \Leftrightarrow$

$T_1(t, \alpha_1, \alpha_2) \leq h(\theta) \leq T_2(t, \alpha_1, \alpha_2)$, entonces

$(T_1(t, \alpha_1, \alpha_2), T_2(t, \alpha_1, \alpha_2))$ es un intervalo de confianza para $h(\theta)$ de nivel $1 - \alpha, \alpha \in (0, 1)$

Ejemplo

Construir por el método de Neyman un intervalo de confianza de longitud esperada mínima para θ basado en una m.a.s. (n) de $X \sim U(0, \theta)$, con $\theta > 0$. Indicación: utilizar $T = T(X_1, \dots, X_n) = X_{(n)}$. Para ello utilizaremos la función de distribución del estadístico suficiente $T = X_{(n)}$:

$$F(t|\theta) = \begin{cases} 0 & t \leq 0 \\ \left(\frac{t}{\theta}\right)^n & 0 < t < \theta \\ 1 & t \geq \theta \end{cases}$$

El método de Neyman nos dice que tenemos que dos funciones $\gamma_1(\theta; \alpha)$ y $\gamma_2(\theta; \alpha)$ tales que:

$$P(\gamma_1(\theta; \alpha) < T < \gamma_2(\theta; \alpha) | \theta) \geq 1 - \alpha, \quad \alpha \in [0, 1]$$

Así deducimos que:

$$F(\gamma_1(\theta; \alpha) | \theta) = \alpha_1$$

$$F(\gamma_2(\theta; \alpha) | \theta) = \alpha - \alpha_1$$

Donde $\alpha_1 \in [0, \alpha]$ Despejando obtenemos que ambas funciones deben ser:

$$\gamma_1(\theta; \alpha) = \theta (\alpha_1)^{\frac{1}{n}}$$

$$\gamma_2(\theta; \alpha) = \theta (1 - (\alpha - \alpha_1))^{\frac{1}{n}}$$

Finalmente, volviendo a la probabilidad y teniendo ya nuestras funciones y nuestro estadístico:

$$P\left(\theta(\alpha_1)^{\frac{1}{n}} < X_{(n)} < \theta(1 - (\alpha - \alpha_1))^{\frac{1}{n}} \mid \theta\right) \geq 1 - \alpha \implies$$

$$P\left(\frac{X_{(n)}}{(1 - (\alpha - \alpha_1))^{\frac{1}{n}}} < \theta < \frac{X_{(n)}}{(\alpha_1)^{\frac{1}{n}}} \mid \theta\right) \geq 1 - \alpha$$

Obteniendo así el intervalo de confianza:

$$\left(\frac{X_{(n)}}{(1 - (\alpha - \alpha_1))^{\frac{1}{n}}}, \frac{X_{(n)}}{(\alpha_1)^{\frac{1}{n}}}\right)$$

El cual claramente se minimiza en $\alpha_1 = \alpha$, obteniendo finalmente el intervalo de confianza:

$$IC_{1-\alpha}(\theta) = \left(X_{(n)}, X_{(n)}\alpha^{-\frac{1}{n}}\right)$$

Intervalos de confianza para muestras grandes

Si $T_n = T(X_1, \dots, X_n)$ es un estimador de $h(\theta)$ tal que

$$\frac{T_n - h(\theta)}{\sigma_n(\theta)} \xrightarrow[n \rightarrow \infty]{d} N(0, 1)$$

$$P_\theta\left(-z_{\alpha/2} \leq \frac{T_n - h(\theta)}{\sigma_n(\theta)} \leq z_{\alpha/2}\right) \xrightarrow[n \rightarrow \infty]{} 1 - \alpha$$

Por lo tanto, si puede invertirse la desigualdad anterior, despejando $h(\theta)$, se puede obtener un intervalo de confianza para $h(\theta)$, de nivel aproximado $1 - \alpha$, cuando el tamaño muestral es suficientemente grande

Observación 5.3.1

Si se cumplen todas las condiciones de regularidad y la ecuación de verosimilitud tiene una única raíz, $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{c.s.} \theta$, puede tomarse $T_n = \hat{\theta}_n$ y $h(\theta) = \theta$, y como

$$\frac{\hat{\theta}_n - \theta}{\sqrt{\frac{1}{nI_1(\theta)}}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1)$$

entonces $\sigma_n(\theta) = \sqrt{\frac{1}{nI_1(\theta)}}$, que si es una función continua puede ser aproximada por $\sigma_n(\hat{\theta}_n)$, lo que facilita la inversión

$$IC_{1-\alpha}(\theta) = \hat{\theta}_n \mp z_{\alpha/2} \sqrt{\frac{1}{nI_1(\hat{\theta}_n)}}$$

Ejemplo

Comprobar que si $X \sim \text{Bin}(1, \theta)$, entonces

$$IC_{1-\alpha}(\theta) = \bar{x} \mp z_{\alpha/2} \sqrt{\frac{\bar{x}(1 - \bar{x})}{n}}$$

$$IC_{1-\alpha}(\theta) = \bar{x} \mp z_{\alpha/2} \frac{1}{2\sqrt{n}}$$

son intervalos de confianza para θ basados en el EMV, para muestras grandes
Desigualdad de Tchebychev

$$P(|Y - E[Y]| > k\sqrt{V(Y)}) \leq \frac{1}{k^2}$$

Teniendo en cuenta que $X \sim \text{Bin}(1, \theta)$, entonces $E[X] = \theta$ y $V(X) = \theta(1 - \theta)$, y por tanto:

$$P(|\bar{X} - \theta| > k\sqrt{\frac{\bar{x}(1 - \bar{x})}{n}}) \leq \frac{1}{k^2} \implies$$

$$P\left(\bar{X} - k\sqrt{\frac{\bar{x}(1 - \bar{x})}{n}} < \theta < \bar{X} + k\sqrt{\frac{\bar{x}(1 - \bar{x})}{n}}\right) \geq 1 - \frac{1}{k^2} \quad \forall k > 0$$

Ahora fijamos $k = \sqrt{\frac{1}{\alpha}}$, y obtenemos el intervalo de confianza:

$$IC_{1-\alpha}(\theta) = \bar{x} \mp \frac{1}{\sqrt{\alpha}} \sqrt{\frac{\bar{x}(1 - \bar{x})}{n}} \quad \forall k > 0$$

Si $\bar{x} = \frac{1}{2}$, lo cual ocurre cuando n es muy grande, entonces:

$$IC_{1-\alpha}(\theta) = \bar{x} \mp \frac{1}{\sqrt{\alpha}} \frac{1}{2\sqrt{n}}$$

Ejemplo

Comprobar que si $X \sim \text{Bin}(1, \theta)$, entonces

$$IC_{1-\alpha}(\theta) = \bar{x} \mp \frac{1}{\sqrt{\alpha}} \sqrt{\frac{\bar{x}(1 - \bar{x})}{n}}$$

$$IC_{1-\alpha}(\theta) = \bar{x} \mp \frac{1}{\sqrt{\alpha}} \frac{1}{2\sqrt{n}}$$

son intervalos de confianza para θ basados en la desigualdad de Tchebychev

Observación 5.3.2

Observación Los intervalos que se obtienen mediante el método de la desigualdad de Tchebychev son más amplios que los construídos mediante procedimientos específicos a cada modelo de probabilidad

Región creíble

Dada una familia de distribuciones de probabilidad $\{f(x_1, \dots, x_n | \theta), \theta \in \Theta\}$, si la información inicial sobre θ viene dada por la función de densidad o de masa $\pi(\theta)$, la región $C(x_1, \dots, x_n) \subset \Theta$ es una región creíble de probabilidad $1 - \alpha$ si

$$P(\theta \in C(x_1, \dots, x_n) | x_1, \dots, x_n) \geq 1 - \alpha$$

donde esta probabilidad se calcula mediante la distribución final, es decir

$$\int_{C(x_1, \dots, x_n)} \pi(\theta | x_1, \dots, x_n) d\theta \geq 1 - \alpha$$

Ejemplo

Para muestras de tamaño $n = 1$ de $X \sim \text{Bin}(1, \theta)$, si $\theta \sim U(0, 1)$ y se observa $x = 1$, entonces $C_{1-\alpha}(\theta) = (\sqrt{\frac{\alpha}{2}}, \sqrt{1 - \frac{\alpha}{2}})$

En efecto, si (X_1, \dots, X_n) es una m.a.s. (n) de $X \sim \text{Bin}(1, \theta)$ y $\theta \sim U(0, 1)$, entonces

$$\pi(\theta \mid x_1, \dots, x_n) \sim \text{Beta}(\sum_{i=1}^n x_i + 1, n - \sum_{i=1}^n x_i + 1)$$

Por lo tanto, para muestras de tamaño $n = 1$

$$\pi(\theta \mid x = 0) \sim \text{Beta}(1, 2), \int_{\ell_1}^{\ell_2} 2(1 - \theta)d\theta = 1 - \alpha$$

$$\pi(\theta \mid x = 1) \sim \text{Beta}(2, 1), \int_{\ell_1}^{\ell_2} 2\theta d\theta = 1 - \alpha$$

Por ejemplo, si $x = 1$, $\ell_1^2 = \frac{\alpha}{2}$, $\ell_2^2 = 1 - \frac{\alpha}{2}$

$$C_{1-\alpha}(\theta) = (0.223, 0.974), \text{ para } \alpha = 0.1$$

Otra posibilidad es la región creíble de más alta probabilidad final o amplitud mínima,

$$C_{1-\alpha}(\theta) = (\sqrt{\alpha}, 1) = (0.316, 1)$$

Ejemplo

Para muestras de tamaño $n = 10$ de $X \sim \text{Bin}(1, \theta)$, si $\theta \sim U(0, 1)$ y se observa $\sum_{i=1}^n x_i = 3$, entonces la distribución final es $\text{Beta}(4, 8)$ y $C_{1-\alpha}(\theta) = (0.135, 0.564)$, para $\alpha = 0.1$

En este caso la región creíble de más alta probabilidad final o amplitud mínima es $C_{1-\alpha}(\theta) = (0.117, 0.542)$, para $\alpha = 0.1$

Ejemplo

Para una m.a.s. (n) de $X \sim N(\theta, 1)$, si $\theta \sim N(0, 1)$ y se observa \bar{x} , entonces $C_{1-\alpha}(\theta) = \frac{n}{n+1}\bar{x} \mp z_{\alpha/2} \frac{1}{\sqrt{n+1}}$

En efecto, Si (X_1, \dots, X_n) es una m.a.s. (n) de $X \sim N(\mu, \sigma)$ con σ conocida y $\mu \sim N(\mu_0, \sigma_0)$, entonces $\pi(\mu \mid x_1, \dots, x_n) \sim N(\mu_1, \sigma_1)$,

$$\mu_1 = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{\bar{x}}{\frac{\sigma^2}{n}}}{\frac{1}{\sigma_0^2} + \frac{\frac{\sigma^2}{n}}{n}} = \frac{\sigma^2}{\sigma^2 + n\sigma_0^2}\mu_0 + \frac{n\sigma_0^2}{\sigma^2 + n\sigma_0^2}\bar{x} = \frac{n}{n+1}\bar{x}$$

$$\sigma_1 = \sqrt{\frac{1}{\frac{1}{\sigma_0^2} + \frac{1}{\frac{\sigma^2}{n}}}} = \frac{1}{\sqrt{n+1}}$$

Observación 5.3.3

Recordemos que el intervalo de confianza obtenido desde el punto de vista frecuentista es $IC_{1-\alpha}(\theta) = \bar{x} \mp z_{\alpha/2} \frac{1}{\sqrt{n}}$, que tiene mayor amplitud, aparte de su diferente interpretación

6 Contraste de Hipótesis

6.1 Principios básicos de un contraste de hipótesis

Sea $X \approx (\chi, \beta_\chi, F_\theta)_{\theta \in \Theta \subset \mathbb{R}^\ell}$ modelo estadístico ℓ -paramétrico y (X_1, \dots, X_n) muestra de $\{F_\theta, \theta \in \Theta\}$

Idea: estudiar si una determinada afirmación sobre $\{F_\theta, \theta \in \Theta\}$ es confirmada o invalidada a partir de los datos muestrales

Ejemplo

Supongamos que en un laboratorio se está estudiando cierta reacción química sobre una determinada sustancia y que el resultado de dicha reacción es una variable observable que se puede modelizar mediante una v.a. X con distribución normal. Por experiencias anteriores se sabe que, si en la sustancia está presente cierto mineral, $X \sim N(\mu = 10, \sigma = 4)$ y si no lo está $X \sim N(\mu = 11, \sigma = 4)$. Se puede comprobar por medio de unos análisis si el mineral está o no presente en la sustancia en estudio, pero dichos análisis son muy costosos, por lo que se procede a realizar la reacción química $n = 25$ veces para decidir, a la luz de los resultados, si $\mu = 10$ o $\mu = 11$

Definición 6.1.1 [Hipótesis Estadística]

Una hipótesis estadística es cualquier afirmación acerca de un modelo estadístico.

Definición 6.1.2 [Hipótesis Estadística Simple y Compuesta]

Una hipótesis estadística es simple si especifica totalmente el modelo estadístico, en otro caso, se dice que es compuesta

Definición 6.1.3 [Hipótesis Estadística Nula y Alternativa]

Se llama hipótesis nula H_0 a la hipótesis de trabajo y es la hipótesis estadística que vamos a aceptar si no hay suficiente evidencia a partir de los datos para rechazarla, consecuentemente se llama hipótesis alternativa H_1 a la hipótesis estadística que se acepta si hay suficiente evidencia a partir de los datos para rechazar H_0

Definición 6.1.4 [Contraste de Hipótesis Paramétrico]

Un contraste de hipótesis paramétrico es una partición del espacio paramétrico Θ en dos subconjuntos Θ_0 y Θ_1 tales que $\Theta = \Theta_0 \cup \Theta_1$ y $\Theta_0 \cap \Theta_1 = \emptyset$

En el ejemplo anterior $\Theta = \{\mu_0, \mu_1\}$, $\Theta_0 = \{\mu_0\}$, $\Theta_1 = \{\mu_1\}$

En un problema de contraste de hipótesis paramétrico se pretende contrastar $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$. Por supuesto, la decisión debe basarse en la evidencia aportada por la observación de una muestra (X_1, \dots, X_n) , o equivalentemente por la observación de un cierto estadístico $T = T(X_1, \dots, X_n)$, que se denomina estadístico del contraste, y que será usualmente un estimador suficiente del parámetro θ

El contraste entre dos hipótesis basado en un estadístico, exige conocer la distribución en el muestreo de dicho estadístico, para los diversos valores del parámetro. De hecho, la idea del contraste consiste en localizar un suceso que sea muy improbable cuando la hipótesis nula es cierta. Si, una vez observada la muestra, acontece dicho suceso, o bien es que el azar ha jugado la mala pasada de elegir una muestra "muy rara" o, como parece más razonable, la hipótesis nula era falsa

Definición 6.1.5 [Región Crítica]

Sea una partición del espacio muestral χ^n en dos subconjuntos C y C^* tales que $\chi^n = C \cup C^*$ y $C \cap C^* = \emptyset$. C es una región crítica para el contraste $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$ si y sólo si, se rechaza H_0 cuando se observa un valor muestral $(x_1, \dots, x_n) \in C$, en cuyo caso se acepta H_1 . Consecuentemente, C^* se denomina región de aceptación y si $(x_1, \dots, x_n) \in C^*$, se dice que no hay suficiente evidencia estadística para rechazar H_0 , en este sentido se acepta H_0 .

Si $T = T(X_1, \dots, X_n) : \chi^n \rightarrow \tau$ es el estadístico del contraste, sea una partición de τ en dos subconjuntos C_τ y C_τ^* tales que $\tau = C_\tau \cup C_\tau^*$ y $C_\tau \cap C_\tau^* = \emptyset$. C_τ es una región crítica para el contraste $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$ si y sólo si, se rechaza H_0 cuando se observa un valor muestral (x_1, \dots, x_n) tal que $T(x_1, \dots, x_n) \in C_\tau$, en cuyo caso se acepta H_1 . Consecuentemente, C_τ^* se denomina región de aceptación. Así, $C = \{(x_1, \dots, x_n) \in \chi^n : T(x_1, \dots, x_n) \in C_\tau\}$.

En el ejemplo anterior, podemos considerar como región crítica $C = \{(x_1, \dots, x_n) : \bar{x} \geq k\}$.

En un problema de contraste no sólo es importante conocer las probabilidades de cada resultado posible, sino también valorar el riesgo que estamos dispuestos a correr al tomar una decisión equivocada. En el ejemplo anterior, al rechazar que en la sustancia está presente el mineral cuando en realidad si lo está, o bien al aceptar que en la sustancia está presente el mineral cuando en realidad no lo está. Ambos errores tienen consecuencias prácticas distintas.

6.2 Errores de tipo I y de tipo II

Error de tipo I es el error que se comete cuando se rechaza H_0 siendo cierta. Error de tipo II es el error que se comete cuando se acepta H_1 siendo falsa.

En el ejemplo anterior, las probabilidades de cometer error de tipo I y error de tipo II son $P(I) = P(\bar{x} \geq k \mid \mu = 10)$ y $P(II) = P(\bar{x} < k \mid \mu = 11)$.

Lo idóneo sería conseguir un test para el que ambas probabilidades de error fuesen pequeñas. Sin embargo (salvo en casos excepcionales y triviales), la reducción de la probabilidad de error de tipo II se hace a costa del aumento de la probabilidad de error de tipo I y viceversa, con lo que el único procedimiento para disminuir ambas probabilidades de error simultáneamente es aumentar el tamaño muestral, lo que conlleva un incremento del coste del procedimiento que en la práctica puede ser prohibitivo.

Definición 6.2.1 [Función de Potencia]

Si C es una región crítica para el contraste $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$, se define la función de potencia del test como la función $\beta_C : \Theta \rightarrow [0, 1]$ que a cada valor θ del parámetro le asigna el valor $\beta_C(\theta) = P_\theta(C)$, es decir, la probabilidad de rechazar H_0 cuando el valor del parámetro es θ .

Definición 6.2.2 [Nivel de significación y tamaño del test]

Un test C tiene nivel de significación $\alpha \in [0, 1]$ si y sólo si $\sup \beta_C(\theta) \leq \alpha$ y se denomina tamaño del test al valor $\sup \beta_C(\theta) \mid \theta \in \Theta_0$.

Ejemplo

En el ejemplo anterior, con $n = 25$, $C = \{10 < \bar{x} < 10,006\}$, $P(I) = \beta_C(\mu = 10) = P(10 < \bar{x} < 10,006 \mid \mu = 10) = 0.05$ $P(II) = 1 - \beta_C(\mu = 11) = P(10 < \bar{x} < 10,006 \mid \mu = 11) = 0.976$.

En el ejemplo anterior, con $n = 25$, $C = \{\bar{x} \geq k\}$ y $\alpha = 0.05$, $P(I) = \beta_C(\mu = 10) = P(\bar{x} \geq 11.316 \mid \mu = 10) = 0.05$ $P(II) = 1 - \beta_C(\mu = 11) = P(\bar{x} < 11.316 \mid \mu = 11) = 0.6554$.

En el ejemplo anterior, con $n = 100$, $C = \{\bar{x} \geq k\}$ y $\alpha = 0.05$,
 $P(I) = \beta_C(\mu = 10) = P(\bar{x} \geq 10.658 \mid \mu = 10) = 0.05$
 $P(II) = 1 - \beta_C(\mu = 11) = P(\bar{x} < 10.658 \mid \mu = 11) = 0.196$

Definición 6.2.3 [p-valor]

Si para contrastar $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$, el test tiene región crítica

$$C = \{(x_1, \dots, x_n) : T(x_1, \dots, x_n) \geq k\}$$

para T un estadístico conveniente, y se observa la muestra (x_1, \dots, x_n) , se denomina p -valor correspondiente a (x_1, \dots, x_n) al valor

$$p(x_1, \dots, x_n) = \sup_{\theta \in \Theta_0} P\{T(X_1, \dots, X_n) \geq T(x_1, \dots, x_n) \mid \theta\}$$

Si el tamaño del test es α y observada la muestra (x_1, \dots, x_n) el p -valor correspondiente $p(x_1, \dots, x_n) \leq \alpha$, entonces (x_1, \dots, x_n) pertenece a la región crítica y por lo tanto se rechaza H_0 . Si el p -valor $p(x_1, \dots, x_n) > \alpha$, entonces (x_1, \dots, x_n) pertenece a la región de aceptación y por lo tanto no hay suficiente evidencia estadística para rechazar H_0

Ejemplo

En el ejemplo anterior, con $n = 100$, $C = \{\bar{x} \geq k\}$ y $\alpha = 0.05$,
 $P(I) = \beta_c(\mu = 10) = P(\bar{x} \geq 10.658 \mid \mu = 10) = 0.05$
 $P(II) = 1 - \beta_C(\mu = 11) = P(\bar{x} < 10.658 \mid \mu = 11) = 0.196$
 Entonces, observada $\bar{x} = 11$

$$p(11) = P(\bar{X} \geq 11 \mid \mu = 10) = P(Z \geq 2.5) = 0.00621$$

Por lo tanto, se rechaza $H_0 : \mu = 10$ a favor de $H_1 : \mu = 11$

Observación 6.2.1

1. n y α son valores fijados de antemano
2. Las hipótesis nula y alternativa no son intercambiables puesto que el tratamiento que reciben es asimétrico, la asimetría queda matizada por el valor α que se elija como nivel de significación y por la probabilidad de error de tipo II que resulte una vez diseñado el test, pues podría ocurrir que para $\theta \in \Theta_1$, $P_\theta(C^*) = 1 - \beta_C(\theta) \leq \alpha$
3. En el contraste de hipótesis planteado se considera H_0 como la hipótesis de interés, en el sentido que para poder invalidarla es necesario esgrimir una gran evidencia. Por consiguiente, los test de hipótesis se emplean con un carácter conservador, a favor de la hipótesis nula, ya que el nivel de significación que se fija, intenta garantizar que sea muy infrecuente rechazar una hipótesis nula correcta, y la preocupación por dejar vigente una hipótesis nula falsa es menor, pudiéndose aceptar en este último caso riesgos más altos. En este sentido, si el resultado de un contraste de hipótesis es aceptar H_0 , debe interpretarse que las observaciones no han aportado suficiente evidencia para descartarla; mientras que, si se rechaza, es porque se está razonablemente seguro de que H_0 es falsa y, por consiguiente, aceptamos H_1

Proposición 6.2.1 [Criterio de comparación de contrastes]

Si C y C' son dos test con nivel de significación α basados en una muestra (X_1, \dots, X_n) de $\{F_\theta, \theta \in \Theta\}$, para contrastar $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$, tales que $\beta_C(\theta) \geq \beta_{C'}(\theta), \forall \theta \in \Theta_1$, entonces C es uniformemente más potente que C'

6.3 Test uniformemente más potente de tamaño α **Proposición 6.3.1**

Sea C una región crítica para el contraste $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$, basada en una muestra (X_1, \dots, X_n) de $\{F_\theta, \theta \in \Theta\}$. C es un test uniformemente de máxima potencia de tamaño α (TUMP) si y sólo si

1. $\sup_{\theta \in \Theta_0} \beta_C(\theta) = \alpha$
2. Para cualquier otro test basado en (X_1, \dots, X_n) con región crítica C' tal que $\sup_{\theta \in \Theta_0} \beta_{C'}(\theta) \leq \alpha$, es $\beta_C(\theta) \geq \beta_{C'}(\theta), \forall \theta \in \Theta_1$

6.4 Hipótesis nula simple frente a alternativa simple**Teorema 6.4.1** [Teorema de Neyman-Pearson (Parte I)]

Para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta = \theta_1$, si para algún $k \geq 0$ existe un test con región crítica $c = \left\{ (x_1, \dots, x_n) \in \chi^n : \frac{f_{\theta_1}(x_1, \dots, x_n)}{f_{\theta_0}(x_1, \dots, x_n)} \geq k \right\}$ y región de aceptación $c^* = \left\{ (x_1, \dots, x_n) \in \chi^n : \frac{f_{\theta_1}(x_1, \dots, x_n)}{f_{\theta_0}(x_1, \dots, x_n)} < k \right\}$ tal que $\alpha = P_{\theta_0}(C)$, entonces C es uniformemente de máxima potencia de tamaño α

Demostración. Observemos que C es un test de tamaño α ya que $\Theta_0 = \{\theta_0\}$ y por lo tanto $\sup_{\theta \in \Theta_0} \beta_C(\theta) = P_{\theta_0}(C) = \alpha$

Sea C' otro test de nivel α , es decir tal que $\alpha \geq \sup_{\theta \in \Theta_0} \beta_{C'}(\theta) = P_{\theta_0}(C')$ y consideremos la siguiente partición del espacio muestral.

$$\begin{aligned}
S^+ &= \{(x_1, \dots, x_n) \in \chi^n : I_C(x_1, \dots, x_n) > I_{C'}(x_1, \dots, x_n)\}, \\
S^- &= \{(x_1, \dots, x_n) \in \chi^n : I_C(x_1, \dots, x_n) < I_{C'}(x_1, \dots, x_n)\}, \\
\chi^n - S^+ \cup S^- &= \{(x_1, \dots, x_n) \in \chi^n : I_C(x_1, \dots, x_n) = I_{C'}(x_1, \dots, x_n)\} \\
\int_{\chi^n} (I_C(x_1, \dots, x_n) - I_{C'}(x_1, \dots, x_n)) (f_{\theta_1}(x_1, \dots, x_n) - k f_{\theta_0}(x_1, \dots, x_n)) dx_1 \cdots dx_n &= \\
\int_{S^+} (I_C(x_1, \dots, x_n) - I_{C'}(x_1, \dots, x_n)) (f_{\theta_1}(x_1, \dots, x_n) - k f_{\theta_0}(x_1, \dots, x_n)) dx_1 \cdots dx_n &+ \\
\int_{S^-} (I_C(x_1, \dots, x_n) - I_{C'}(x_1, \dots, x_n)) (f_{\theta_1}(x_1, \dots, x_n) - k f_{\theta_0}(x_1, \dots, x_n)) dx_1 \cdots dx_n &+ \\
\int_{\chi^n - S^+ \cup S^-} (I_C(x_1, \dots, x_n) - I_{C'}(x_1, \dots, x_n)) (f_{\theta_1}(x_1, \dots, x_n) - k f_{\theta_0}(x_1, \dots, x_n)) dx_1 \cdots dx_n &\geq 0 \\
\int_{\chi^n} I_C(x_1, \dots, x_n) f_{\theta_1}(x_1, \dots, x_n) - \int_{\chi^n} I_{C'}(x_1, \dots, x_n) f_{\theta_1}(x_1, \dots, x_n) dx_1 \cdots dx_n &\geq \\
k \left(\int_{\chi^n} I_C(x_1, \dots, x_n) f_{\theta_0}(x_1, \dots, x_n) - \int_{\chi^n} I_{C'}(x_1, \dots, x_n) f_{\theta_0}(x_1, \dots, x_n) dx_1 \cdots dx_n \right) & \\
P_{\theta_1}(C) - P_{\theta_1}(C') \geq k (P_{\theta_0}(C) - P_{\theta_0}(C')) \geq k(\alpha - \alpha) = 0 \Rightarrow \beta_C(\theta) \geq \beta_{C'}(\theta), \forall \theta \in \Theta_1 = \{\theta_1\} &
\end{aligned}$$

Observación 6.4.1

De la demostración del teorema se deduce que los puntos para los que $f(x_1, \dots, x_n | \theta_1) = k f(x_1, \dots, x_n | \theta_0)$ pueden ser colocados tanto en la región crítica como en la región de aceptación.

Es importante señalar que el teorema de Neyman-Pearson no dice que el test de la forma dada en su enunciado deba existir cualquiera que sea $\alpha \in [0, 1]$

Ejemplo

Para una muestra de tamaño $n = 12$, extraída de una distribución de Poisson con parámetro θ , donde $\theta \in [0, 0.5]$, se plantea el siguiente contraste de hipótesis:

$$\begin{cases} H_0 : \theta = 0 \\ H_1 : \theta = 0.5 \end{cases}$$

La región crítica para este contraste viene dada por:

$$C = \{(x_1, \dots, x_{12}) : \sum_{i=1}^{12} x_i \geq 2\}$$

En este caso particular, como $\sum_{i=1}^{12} x_i < 2$, se tiene que $\bar{X} < \frac{1}{6}$.

La probabilidad de error de tipo I (α) y la función de potencia ($\beta(\theta)$) se calculan como sigue:

$$\alpha = \beta(0) = P(C | \theta = 0) = P\left(\sum_{i=1}^{12} x_i \geq 2 | \theta = 0\right) = 0$$

$$\begin{aligned} \beta(0.5) &= P(C | \theta = 0.5) = P\left(\sum_{i=1}^{12} x_i \geq 2 | \theta = 0.5\right) \\ &= P(\text{Poisson}(6) \geq 2) \\ &= 1 - P(\text{Poisson}(6) = 0) - P(\text{Poisson}(6) = 1) \\ &= 1 - e^{-6} \left(\frac{6^0}{0!} + \frac{6^1}{1!} \right) \\ &= 1 - e^{-6}(1 + 6) \approx 0.9826 \end{aligned}$$

$$\begin{aligned} \beta(0.25) &= P(C | \theta = 0.25) = P\left(\sum_{i=1}^{12} x_i \geq 2 | \theta = 0.25\right) \\ &= P(\text{Poisson}(3) \geq 2) \\ &= 1 - P(\text{Poisson}(3) = 0) - P(\text{Poisson}(3) = 1) \\ &= 1 - e^{-3} \left(\frac{3^0}{0!} + \frac{3^1}{1!} \right) \\ &= 1 - e^{-3}(1 + 3) \approx 0.8009 \end{aligned}$$

7 Appendix

7.1 Momentos Notables

Media

Definición 7.1.1 [Media]

Distinguimos entre casos discretos y continuos:

- **Caso discreto:** Se define la media de una variable aleatoria discreta X como:

$$\mu = E(X) = \sum_{i=1}^n x_i \cdot p_i \quad (2)$$

donde x_i son los valores que puede tomar la variable aleatoria y p_i son las probabilidades asociadas a cada valor.

- **Caso continuo:** Se define la media de una variable aleatoria continua X como:

$$\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx \quad (3)$$

donde $f(x)$ es la función de densidad de probabilidad de la variable aleatoria.

Propiedades

Si X y Y son **variables aleatorias** con esperanza finita y $a, b, c \in \mathbb{R}$ son constantes entonces

1. $E[c] = c$.
2. $E[cX] = cE[X]$.
3. Si $X \geq 0$ entonces $E[X] \geq 0$.
4. Si $X \leq Y$ entonces $E[X] \leq E[Y]$.
5. Si X está delimitada por dos números reales, a y b , esto es $a < X < b$ entonces también lo está su media, es decir,

$$a < E[X] < b.$$

6. Si $Y = a + bX$, entonces

$$E[Y] = E[a + bX] = a + bE[X].$$

7. En general, $E[XY] \neq E[X]E[Y]$, la igualdad sólo se cumple cuando las variables aleatorias son independientes.

Teorema 7.1.1 [Linealidad de la Esperanza]

El operador esperanza $E[\cdot]$ es una **aplicación lineal**, pues para cualesquiera **variables aleatorias** X y Y y cualquier constante c tal que $c \in \mathbb{R}$, se cumple lo siguiente:

$$E[X + Y] = E[X] + E[Y]$$

$$E[cX] = cE[X]$$

Demostración. Demostrar este resultado es sencillo. Si consideramos que X y Y son **variables aleatorias** discretas, entonces

$$\begin{aligned} E[X + Y] &= \sum_{x,y} (x + y) P(X = x, Y = y) \\ &= \sum_{x,y} x P(X = x, Y = y) + \sum_{x,y} y P(X = x, Y = y) \\ &= \sum_x x P(X = x, Y = y) + \sum_y y \sum_x P(X = x, Y = y) \\ &= \sum_x x P(X = x) + \sum_y y P(Y = y) \\ &= E[X] + E[Y] \end{aligned}$$

□

Teorema 7.1.2 [Multiplicación de las Esperanzas]

Sean X_1, \dots, X_n **variables aleatorias independientes**, tales que $\exists E[X_i] \forall i = 1 \dots n$. Entonces $\exists E[X_1 \dots X_n]$, y se verifica:

$$E[X_1 \dots X_n] = E[X_1] \dots E[X_n] = \prod_{i=1}^n E[X_i]$$

Demostración. La demostración de este resultado es muy sencilla, sólo hay que considerar el concepto de independencia. El resultado se demuestra sólo para el caso discreto bidimensional (la demostración del caso continuo es análoga).

$$\begin{aligned} E[XY] &= \sum_x \sum_y xy P(X = x, Y = y) \\ &= \sum_x \sum_y xy P(X = x) P(Y = y) \\ &= \sum_x x P(X = x) \sum_y y P(Y = y) \\ &= E[X] E[Y] \end{aligned}$$

□

Varianza

Definición 7.1.2 [Varianza]

Distinguimos entre casos discretos y continuos:

- **Caso discreto:** Se define la varianza de una variable aleatoria discreta X como:

$$\sigma^2 = \text{Var}(X) = \sum_{i=1}^n (x_i - \mu)^2 \cdot p_i \quad (4)$$

donde x_i son los valores que puede tomar la variable aleatoria, p_i son las probabilidades asociadas a cada valor y μ es la media de la variable aleatoria.

- **Caso continuo:** Se define la varianza de una variable aleatoria continua X como:

$$\sigma^2 = \text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx \quad (5)$$

donde $f(x)$ es la función de densidad de probabilidad de la variable aleatoria y μ es la media de la variable aleatoria.

La varianza también puede ser expresada como:

$$\sigma^2 = \text{Var}(X) = E[(X - E[X])^2] = E[X^2 - 2XE[X] + E[X]^2] \quad (6)$$

$$= E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2 \quad (7)$$

La varianza también puede ser vista como covarianza de una variable consigo misma:

$$\sigma^2 = \text{Var}(X) = \text{Cov}(X, X) \quad (8)$$

Propiedades

Si X y Y son **variables aleatorias** con varianza finita entonces

1. $\text{Var}(X) \geq 0$.
2. $\text{Var}(X) = 0$ si y sólo si X es constante.
3. $\text{Var}(X) = 0 \iff \exists c \in \mathbb{R}$ tal que $P(X = c) = 1$.
4. $\text{Var}(X + a) = \text{Var}(X)$.
5. $\text{Var}(aX) = a^2 \text{Var}(X)$.
6. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$.
7. $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2 \text{Cov}(X, Y)$.

Teorema 7.1.3 [Identidad de Bienaymé]

En general, para la suma de n variables aleatorias X_1, X_2, \dots, X_n se cumple que

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \sum_{i,j=1}^n \text{Cov}(X_i, X_j) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j) \quad (9)$$

Si las variables aleatorias son independientes, entonces la covarianza entre ellas es nula, y la varianza de la suma de las variables aleatorias es la suma de las varianzas de las variables aleatorias, es decir, se cumple que

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) \quad (10)$$

Producto de Variables Aleatorias

- **Variables Aleatorias Independientes** Si dos variables X e Y son **independientes**, la varianza de su producto está dada por:

$$\text{Var}(XY) = [E(X)]^2 \text{Var}(Y) + [E(Y)]^2 \text{Var}(X) + \text{Var}(X) \text{Var}(Y).$$

De manera equivalente, utilizando las propiedades básicas de la esperanza, se expresa como:

$$\text{Var}(XY) = E(X^2)E(Y^2) - [E(X)]^2[E(Y)]^2.$$

- **Variables Aleatorias Correlacionadas** En general, si dos variables son **estadísticamente dependientes**, entonces la varianza de su producto está dada por:

$$\begin{aligned} \text{Var}(XY) &= E[X^2Y^2] - [E(XY)]^2 \\ &= \text{Cov}(X^2, Y^2) + E(X^2)E(Y^2) - [E(XY)]^2 \\ &= \text{Cov}(X^2, Y^2) + (\text{Var}(X) + [E(X)]^2)(\text{Var}(Y) + [E(Y)]^2) \\ &\quad - [\text{Cov}(X, Y) + E(X)E(Y)]^2 \end{aligned}$$

7.2 Función Característica

Definición 7.2.1 [Función Característica]

La **función característica** de una variable aleatoria X es una función $\varphi_X(t) : \mathbb{R} \rightarrow \mathbb{C}$ definida como:

$$\varphi_X(t) = E[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} f(x) dx$$

donde i es la unidad imaginaria y t es un número real.

Cuando los momentos de una variable aleatoria existen, se pueden calcular mediante las derivadas de la función característica. De modo que se puede obtener derivando formalmente a ambos lados de la definición y tomando $t = 0$:

$$\varphi_X^{(n)}(0) = i^n E[X^n]$$

Propiedades

- La función característica de una variable aleatoria real siempre existe, ya que es una integral de una función continua acotada sobre un espacio cuya medida es finita.
- Una función característica es **uniformemente continua** en todo el espacio.
- No se anula en una región alrededor de cero: $\varphi(0) = 1$.
- Es acotada: $|\varphi(t)| \leq 1$.
- Es **hermítica**: $\varphi(-t) = \overline{\varphi(t)}$. En particular, la función característica de una variable aleatoria simétrica (alrededor del origen) es de valores reales y **par**.
- Existe una **biyección** entre **distribuciones de probabilidad** y funciones características. Es decir, dos variables aleatorias X_1 y X_2 tienen la misma distribución de probabilidad si y solo si $\varphi_{X_1} = \varphi_{X_2}$.
- Si una variable aleatoria X tiene **momentos** hasta orden k , entonces su función característica φ_X es k veces continuamente diferenciable en toda la recta real. En este caso:

$$E[X^k] = i^{-k} \varphi_X^{(k)}(0).$$

- Si una función característica φ_X tiene derivada de orden k en cero, entonces la variable aleatoria X tiene todos los momentos hasta k si k es par, pero solo hasta $k - 1$ si k es impar.

$$\varphi_X^{(k)}(0) = i^k E[X^k].$$

- Si X_1, \dots, X_n son variables aleatorias independientes y a_1, \dots, a_n son constantes, entonces la función característica de la combinación lineal de las variables X_i es

$$\varphi_{a_1 X_1 + \dots + a_n X_n}(t) = \varphi_{X_1}(a_1 t) \cdots \varphi_{X_n}(a_n t).$$

Un caso particular es la suma de dos variables aleatorias independientes X_1 y X_2 , en cuyo caso se cumple

$$\varphi_{X_1 + X_2}(t) = \varphi_{X_1}(t) \cdot \varphi_{X_2}(t).$$

- Sean X y Y dos variables aleatorias con funciones características φ_X y φ_Y . X y Y son independientes si y solo si

$$\varphi_{X,Y}(s,t) = \varphi_X(s) \varphi_Y(t) \quad \text{para todo } (s,t) \in \mathbb{R}^2.$$

- El comportamiento en la cola de la función característica determina la **suavidad** de la función de densidad correspondiente.
- Sea la variable aleatoria $Y = aX + b$, una transformación lineal de la variable aleatoria X . La función característica de Y es

$$\varphi_Y(t) = e^{itb} \varphi_X(at).$$

Para vectores aleatorios \mathbf{X} y $\mathbf{Y} = A\mathbf{X} + \mathbf{B}$ (donde A es una matriz constante y \mathbf{B} un vector constante), se tiene

$$\varphi_{\mathbf{Y}}(t) = e^{it^T \mathbf{B}} \varphi_{\mathbf{X}}(A^T t).$$