Estadística

Contents

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

Etimación 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 $X : \omega \to \mathbb{R}$ y su espacio medible asociado $(\chi, \mathcal{B})$.
*Un **modelo estadístico** es una terna $(\chi, \mathcal{B}, F)$, 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 anteior, una terna $(\chi, \mathcal{B}, F)$, 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 $(X_1, \ldots, X_n)$ 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 $(\chi, \mathcal{B}, F)$, 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:*

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

Definición 2.1.4 [Función de Distibución Empírica]

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

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

Observemos que

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

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

Ejemplo

Supongamos que tenemos una muestra aleatoria ordenada de tamaño $n = 5$:
$x_{(1)} = 2, x_{(2)} = 3, x_{(3)} = 5, x_{(4)} = 7x_{(5)} = 9$.
Entones, por como se define la FDE $F_n(x) = \frac{1}{5} \sum_{i=1}^{5} I_{(-\infty, x]}(X_i)$, tenemos que:

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

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 $X_1, X_2, \ldots, X_n$ *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:*

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

Por tanto también podemos afirmar que:

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

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

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

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:

$$V[F_n(x)] = \frac{F(x)(1-F(x))}{n} \underset{n\to\infty}{\longrightarrow} 0$$

3. ***Convergencia:***

   (a) ***Convergencia casi segura:***

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

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

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

4. ***Intervalo de confianza para $F(x)$:*** *Dada una realización particular de la muestra $(x_1, \ldots, x_n)$, 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 $(X_1, \ldots, X_n)$ m.a.s.$(n) \sim X$ con función de distribucióm empírica $F_n(x)$ y sea $F(x)$ la función de distribución de X, es decir, de la población total. Entonces se cumple que:

$$\lim_{n\to\infty} P\left(w : \text{Sup}_x \left|F_n(x) - F(x)\right| < \epsilon\right) = 1, \ \forall \epsilon > 0$$

Corolario 2.1.1

El Teorema de Glivenko-Cantelli permite realizar una técnica estadística denominda ***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 $\bar{X} = \int x \partial F_n(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$.*

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

Ejemplo

Desacamos los siguientes estadísticos muestrales:

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

2.3 Momentos

Sea $(X_1,\cdots X_n)$ 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*

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

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*

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

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

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

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

1. *Momentos muestrales respecto al origen:*

   (a)
   $$a_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow[n\to\infty]{c.s.} \alpha_k = E\left[X^k\right]$$

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

2. *Momentos muestrales respecto a la esperanza/media:*

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

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

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

Propiedades asintóticas de los momentos muestrales

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

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

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

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

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

2.4 Resultados de convergencias

Teorema 2.4.1 [de Slutsky]

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

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

Lema 2.4.1

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

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

entonces para cualquier función $g:\mathbb{R}\to\mathbb{R}$ *con derivada continua y no nula en* a *se tiene que* $a_n\left(g\left(X_n\right)-g(a)\right) \xrightarrow[n\to\infty]{d} g'(a)X$

Ejemplo

Sea $(X_1,\cdots X_n)$ m.a.s. (n) de $X\sim N(\mu,\sigma)$, se pide calcular la distribución de la media muestral:
Tenemos que $\varphi_X(t)=e^{it\mu-\frac{1}{2}t^2\sigma^2} \implies$
$E\left[e^{\bar{X}}\right]=\varphi_{\bar{X}}(t)=E\left[e^{it\frac{1}{n}\sum_{i=1}^n X_i}\right]=\varphi_{\sum_{i=1}^n X_i}\left(\frac{t}{n}\right)=\left(\varphi_X\left(\frac{t}{n}\right)\right)^n=e^{it\mu-\frac{1}{2}\frac{\sigma^2}{n}t^2}$.
Por lo tanto $\bar{X}\sim N\left(\mu,\frac{\sigma}{\sqrt{n}}\right)$

Ejemplo

Sea $X_1,\ldots,X_n$ 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:

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

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 $(X_1, \ldots, X_n)$ 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:

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

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:

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

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

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

Teorema 2.4.2 [de Fisher]

Sea $(X_1,\ldots,X_n)$ 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:*

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

son variables aleatorias independientes.

2. *Sus distribuciones son:*

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

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

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.

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

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

$$Cov(X_i-\bar{X},\bar{X})=E[(X_i-\bar{X})\bar{X}]-E[X_i-\bar{X}]E[\bar{X}]=E[(X_i-\bar{X}\bar{X})]-(E[X_i]-E[\bar{X}])E[\bar{X}]=$$
$$=E[(X_i-\bar{X})\bar{X}]-(\mu-\mu)\mu=E[X_i\cdot\bar{X}]-E[\bar{X}^2]=E[X_i\cdot\bar{X}]-(V[X]+E[\bar{X}]^2)=E[X_i\cdot\bar{X}]-(\frac{\sigma^2}{n}+\mu^2)=$$
$$=E[\frac{1}{n}X_1\cdot X_i+\ldots+\frac{1}{n}X_i^2+\ldots+\frac{1}{n}X_n\cdot X_i]-(\frac{\sigma^2}{n}-\mu^2)=$$
$$=\frac{1}{n}(E[X_1X_i]+\ldots+E[X_nX_i])+\frac{1}{n}E[X_i^2]-(\frac{\sigma^2}{n}+\mu^2)=$$
$$=\frac{n-1}{n}(\mu^2)+\frac{1}{n}(\sigma^2+\mu^2)-(\frac{\sigma^2}{n}+\mu^2)=$$
$$=\frac{n-1}{n}\mu^2+\frac{\mu^2}{n}-\mu^2+\frac{\sigma^2}{n}-\frac{\sigma^2}{n}=0\implies$$
$$Cov(X_i-\bar{X},\bar{X})=0\implies \text{son independientes}$$

.

2. Denotemos por
$$S_{n+1}^2 = \frac{1}{n}\sum_{i=1}^{n+1}\left(X_i - \bar{X}_{n+1}\right)^2.$$
Demostremos que:
$$nS_{n+1}^2 = (n-1)S_n^2 + \left(X_{n+1} - \bar{X}_n\right)^2 \frac{n}{n+1}.$$
Desarrollando la expresión:
$$\begin{aligned}
nS_{n+1}^2 &= \sum_{i=1}^{n+1}\left(X_i - \bar{X}_{n+1}\right)^2 \\
&= \sum_{i=1}^{n}\left(\left(X_i - \bar{X}_n\right) + \left(\bar{X}_n - \bar{X}_{n+1}\right)\right)^2 + \left(X_{n+1} - \bar{X}_{n+1}\right)^2 \\
&= (n-1)S_n^2 + n\left(\bar{X}_n - \bar{X}_{n+1}\right)^2 + \left(X_{n+1} - \bar{X}_{n+1}\right)^2 \\
&\quad + \sum_{i=1}^{n} 2\left(X_i - \bar{X}_n\right)\left(\bar{X}_n - \bar{X}_{n+1}\right).
\end{aligned}$$
Como $\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right) = 0$, el término cruzado se anula y obtenemos:
$$nS_{n+1}^2 = (n-1)S_n^2 + n\left(\bar{X}_n - \bar{X}_{n+1}\right)^2 + \left(X_{n+1} - \bar{X}_{n+1}\right)^2.$$
En este último paso, se desarrollan los cuadrados y se aplica la definición de la media muestral:
$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \Longrightarrow \quad n\bar{X}_n = \sum_{i=1}^{n} X_i.$$
Ahora, utilizando la relación:
$$\bar{X}_{n+1} = \frac{n\bar{X}_n + X_{n+1}}{n+1} \quad \Longrightarrow \quad X_{n+1} - \bar{X}_{n+1} = \frac{(n+1)X_{n+1} - n\bar{X}_n - X_{n+1}}{n+1},$$
obtenemos:
$$\begin{aligned}
nS_{n+1}^2 &= (n-1)S_n^2 + \left(X_{n+1} - \bar{X}_n\right)^2 \frac{n}{(n+1)^2} + \left(X_{n+1} - \bar{X}_n\right)^2 \frac{n^2}{(n+1)^2} \\
&= (n-1)S_n^2 + \left(X_{n+1} - \bar{X}_n\right)^2 \frac{n}{n+1}.
\end{aligned}$$
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:

$$
\left(\frac{X_{n+1} - \bar{X}_n}{\sigma\sqrt{\frac{n+1}{n}}}\right) \sim N(0,1) \sim \chi^2_1.
$$

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

$$
\begin{aligned}
\frac{S_2^2}{\sigma^2} &= \frac{1}{\sigma^2}\left((X_1 - \bar{X}_2)^2 + (X_2 - \bar{X}_2)^2\right) \\
&= \frac{1}{\sigma^2}\left(\left(X_1 - \frac{X_1 + X_2}{2}\right)^2 + \left(X_2 - \frac{X_1 + X_2}{2}\right)^2\right) \\
&= \frac{1}{\sigma^2}\left(\frac{1}{4}(X_1 - X_2)^2 + \frac{1}{4}(X_2 - X_1)^2\right) \\
&= \frac{1}{\sigma^2}\frac{1}{2}(X_2 - \bar{X}_1)^2.
\end{aligned}
$$

Como hemos demostrado antes,

$$
\left(\frac{X_2 - \bar{X}_1}{\sigma\sqrt{2}}\right)^2 \sim \chi^2_1.
$$

Hipótesis de inducción: Para n - 1 supongamos que

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

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

$$
\frac{X_{n+1} - \bar{X}_n}{\sigma\sqrt{\frac{n+1}{n}}} \sim \chi^2_1,
$$

obtenemos la suma de dos variables chi-cuadrado:

$$
\chi^2_{n-1} + \chi^2_1 \sim \chi^2_n.
$$

Conclusión: Por inducción, se ha demostrado que

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

3. Se deduce de los anteriores

$\square$

2.5 Estadísticos ordenados

Definición 2.5.1 [Estadísticos ordenados]

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

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

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

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

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

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

$$f_{\left(X_{(1)},\ldots,X_{(n)}\right)}(y_1,\ldots,y_n) = n!\prod_{j=1}^n f(y_j) \ \textit{si } y_1 < \ldots < y_n$$

Ejemplo

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

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

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 $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$. Se quiere encontrar la función de densidad de estos valores ordenados.
Para el máximo, $X_{(n)}$, se tiene que su función de distribución es:

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

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, $X_{(n)} \sim Beta(n, 1)$.
Para el mínimo, $X_{(1)}$, la función de distribución se obtiene como:

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

Por independencia,

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

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 $X_{(r)} \sim Beta(r, n - r + 1)$.
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 ditribución del rango muestral: $R = X_{(n)} - X_{(1)}$:
Sea el cambio de variable

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

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:

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

Ejemplo

Sea $(X_1, \ldots, X_n)$ 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

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

Para la variable U, tenemos:

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

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

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

Por lo tanto,

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

Ahora, para la distribución de U_R:

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

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

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

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

$$U_R = F\left(X_{(r)}\right),$$

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 $X\sim Gamma(a,b) \implies \varphi_X(t)=(1-\frac{it}{a})^{-b} \implies \varphi_{S_n}(t)=E[e^{i\cdot\frac{t}{n}\sum_{i=1}^{n}X_i}]=E[e^{i\frac{t}{n}X_1}]\dots E[e^{i\frac{t}{n}X_n}]=\left[\varphi(\frac{t}{n})\right]^n=(1-\frac{it}{an})^{-nb} \implies \bar{X}\sim Gamma(na,nb)$

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

$\square$

Ejercicio 2.4. Dada una sucesion $\{X_n\}$ 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

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

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

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

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,

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

$\square$

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. $(X_1, X_2, \ldots, X_n)$.
Sea entonces el conjunto de variables aleatorias

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

con una probabilidad $P(I_j = 1) = P(X_j \leq x) = F_j(x)$
Al ser variables aleatorias independientes con distirbución Bernouilli, se cumple que

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

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:

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

$\square$

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 $X : \Omega \longrightarrow \mathbb{R}$ y su modelo estadístico asociado $(\chi, \mathcal{B}, F_\theta)_{\theta \in \Theta \subset \mathbb{R}^\ell}, \mathcal{B} = \mathcal{B}(\mathbb{R})$, y $(X_1, \cdots X_n)$ m.a.s. $(n) \sim X$
$T : \mathbb{R}^n \to \mathbb{R}^m$ es un estadístico suficiente para θ cuando $\forall (x_1, \ldots, x_n) \in \Omega$ se cumple que:

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

Ejemplo

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

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

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

$$X_i \sim Bin(1, \theta) \equiv Bernoulli(\theta)$$

$$\implies P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n - \sum_{i=1}^n x_i} \prod_{i=1}^n$$

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

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

3.1 Teorema de Factorización de Fisher

Teorema 3.1.1 [de Factorización de Fisher (Caracterización de estadísticos suficientes])

$T = T(X_1, \cdots X_n) : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ es un estadístico suficiente para θ sí y sólo sí existen funciones reales positivas $h : \mathbb{R}^n \longrightarrow \mathbb{R}$ y $g_\theta : \mathbb{R}^m \longrightarrow \mathbb{R}$ tales que $f_\theta(x_1, \ldots, x_n) = h(x_1, \ldots, x_n) g_\theta(T(x_1, \ldots, x_n))$, donde $f_\theta(x_1, \ldots, x_n)$ 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:

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

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

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

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

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

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

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

Ssea 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 $X \sim Poisson(\theta)$ 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:

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

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 $T(x_1,\dots,x_n) = \sum_{i=1}^n$ 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

$$\iff f_\theta(x_1,\dots,x_n)=h(x_1,\dots,x_n)\cdot g_\theta(S^{-1}(S(T(x_1,\dots,x_n))))$$
$$\iff f_\theta(x_1,\dots,x_n)=h(x_1,\dots,x_n)\cdot g'_\theta(S(T(x_1,\dots,x_n)))\implies$$

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

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

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

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

Entonces por el Teorema de Caracterización de Fisher, $\left(\bar{X},S_n^2\right)$ 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:

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

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

Ejemplo

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

$$X \sim U(-\frac{\theta}{2},\frac{\theta}{2}) \implies f_\theta(x) = \frac{1}{\theta} \cdot I_{(-\frac{\theta}{2},\frac{\theta}{2})}(x)$$

$$\implies f_\theta(x_1,\ldots,x_n) = \prod_{i=1}^{n} \frac{1}{\theta} \cdot I_{(-\frac{\theta}{2},\frac{\theta}{2})}(x_i) = \frac{1}{\theta^n} \cdot I_{(-\frac{\theta}{2},\frac{\theta}{2})}(X_{(1)}) \cdot I_{(-\frac{\theta}{2},\frac{\theta}{2})}(X_{(n)})$$

$$\implies \text{Tomando } h(x_1,\ldots,x_n) = 1 \text{ y } g_\theta(T(x_1,\ldots,X_n)) = \frac{1}{\theta^n} I_{(-\frac{\theta}{2},\frac{\theta}{2})}(X_{(1)}) \cdot I_{(-\frac{\theta}{2},\frac{\theta}{2})}(X_{(n)})$$

3.3 Estadístico minimal suficiente

Definición 3.3.1 [Órbita]

Dado un estadístico $T = T(X_1, \cdots X_n)$, se define $A_t = \{(x_1,\ldots,x_n) \in \chi^n : T(x_1,\ldots,x_n) = t\}$ cómo la órbita de $t = 5$

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 $A_5 = \{(x_1,\ldots,x_n) \in \mathbb{R}^3 : max(x_1,x_2,x_3) - min(x_1,x_2,x_3) = 5\}$ cuyos elementos podrían ser por ejemplo:

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

Definición 3.3.2 [Partición Inducida]

Dado un estadístico $T = T(X_1, \cdots X_n)$, se define $K_T = \{A_t : t \in \mathbb{R}^m\}$ cómo la partición 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 deinir 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 $K_T = \{A_2, A_3, A_4, A_5, ...\}$

Proposición 3.3.1

Se dice que $\mathcal{K}_T$ es suficiente sí y sólo si T es suficiente

Proposición 3.3.2

Dado dos estadísticos $T = T(X_1, \ldots, X_n)$ y $S = S(X_1, \ldots, X_n)$, se dice que $\mathcal{K}_S$ es una subpartición de $\mathcal{K}_T$ si y solo si:

$$\forall B \in \mathcal{K}_S,\ \exists A \in \mathcal{K}_T \text{ tal que } B \subset A.$$

En este caso, se dice que $\mathcal{K}_T$ es una partición menos fina que $\mathcal{K}_S$.

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.

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

Alternativamente, lo podemos definir como que dado un estadístico suficiente T, se dice minimal suficiente cuando $\forall T'$ suficiente, $\exists \varphi : \mathbb{R}^n \to \mathbb{R}^n$ medible (integrable) tal que $T' = \varphi(T)$

Demostración. Demostración de la equivalencia de las definiciones:

- ($\Rightarrow$): Sea S suficiente, si $S(x_1, \ldots, x_n) = S(y_1, \ldots, y_n) = s \implies (x_1, \ldots, x_n), (y_1, \ldots, y_n) \in B_s$(Órbita de S) $\implies \exists A_t \in K_T : B_s \subset A_t \implies T(x_1, \ldots, x_n) = T(y_1, \ldots, y_n) = t \implies \exists \psi : \psi(S) = T$ y T es suficiente.
- ($\Leftarrow$): Sea S suficiente y $\psi : \psi(S) = T \implies T$ es suficiente y si $S(x_1, \ldots, x_n) = S(y_1, \ldots, y_n) = s \implies T(x_1, \ldots, x_n) = \psi(S(x_1, \ldots, x_n)) = \psi(S) = \psi(S(y_1, \ldots, y_n)) = T(y_1, \ldots, y_n) \implies B_s \subset A_{\psi(S)} \implies K_S$ 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 $(x_1, \ldots, x_)$ y $(y_1, \ldots, y_n)$ *muestras, definimos la relación de equivalencia minimal suficiente como:*

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

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 $T = T(X_1, \ldots, X_n)$ tal que $\frac{f_\theta(x_1,\ldots,x_n)}{f_\theta(y_1,\ldots,y_n)}$ es independiete de θ cuando $T(x_1, \ldots, x_n) = T(y_1, \ldots, y_n) = t$. Entonces, T es minimal suficiente para θ.

Demostración. Supongamos la suficiencia de T para demostrar su minimalidad:
Sea $S = S(X_1, \ldots, X_n)$ suficiente y $(x_1, \ldots, x_n), (y_1, \ldots, y_n) \in B_S \implies$

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

$\implies \exists t : (x_1, \ldots, x_n), (y_1, \ldots, y_n) \in A_t \implies K_S$ es una subpartición de K_T.
Ahora demostremos la suficiencia de T en el caso discreto:
Si $T(x_1, \ldots, x_n) = t$,

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

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 $X \sim Bin(1, \theta)$

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

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

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

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

Observación 3.4.1

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

Ejemplo

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

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

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

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

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

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

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

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

3.5 Familia exponencial k-paramétrica

Definición 3.5.1 [Familia exponencial k-paramétrica]

Una distribución X pertence 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 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:

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

Teorema 3.5.1

Sean $\theta_1,\ldots,\theta_k \in \Theta \subset \mathbb{R}^\ell$ tales que los vectores $c_r = (q_1(\theta_r),\ldots,q_k(\theta_r))$, $r = 1,\ldots,k$ 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:

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

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 □

Ejemplo

Veamos si el estadístico $T = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$ 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:

$$f_\theta(x) = c(\theta)h(x)e^{q_1(\theta)T_1(x) + q_2(\theta)T_2(x)} \implies$$

$$\implies X \sim N(\mu, \sigma) \implies f_\theta(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \implies f_\theta(x_1, \dots, x_n) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2} e^{\frac{\mu}{\sigma^2}\sum_{i=1}^n x_i} \implies$$

$$\text{tomando} \begin{cases} c(\theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \\ h(x) = 1 \\ q_1(\theta) = -\frac{1}{2\sigma^2} \\ q_2(\theta) = \frac{\mu}{\sigma^2} \\ T_1(x) = \sum_{i=1}^n X_i^2 \\ T_2(x) = \sum_{i=1}^n x_i \end{cases}$$

La distribución normal pertence 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 $S : \mathbb{R}^2 \to \mathbb{R}^2$ 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

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

Ejemplo

Sea una población que sigue una distribucion normal $N(\theta.\sigma_0)$ con σ_0 conocida, tomemos una m.a.s. de tamaño n tal que $U(X_1, \ldots, X_n) = X_1 - X - 2 \sim N(0, \sqrt{2}\sigma_0)$ que no depende de θ por lo que U es un estadístico ancilario para θ.

Definición 3.6.2

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

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 $x = \frac{\theta}{1-\theta} \implies$

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

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:

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

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:

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

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

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

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

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:

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

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

Ejemplo

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

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

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**:

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

Evaluando 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

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

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

Es decir, Se dice que un estadístico es completo si no existe función $g(x_1, \ldots, x_n)$ tal que $E_\theta[g(T)] = 0$ a menos rque $g(X) \overset{c.s.}{=} 0$
Alternativamente, también se puede decir, que el estadístico es completo si y solo sí 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:

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

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 $T = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$ es completo:
Por las propiedades de las esperanzas tenemos que:

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

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

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

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

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

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

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

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.*
- ***Completitud:*** *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 estadístio 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:

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

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. □

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:

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

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 $\sum x_i^2$*:*

$$\begin{cases} S^2 = \frac{1}{n}\sum(x_i-\bar{X})^2 \\ \bar{X} = \frac{1}{n}\sum x_i \end{cases} \implies \begin{cases} \sum x_i^2 = n(S_X^2+\bar{X}^2) \\ \sum y_i^2 = n(S_Y^2+\bar{Y}^2) \end{cases} \implies$$

$$\implies \begin{cases} n(S_X^2+\bar{X}^2)+n\mu^2-2n\mu\bar{X} \\ n(S_Y^2+\bar{Y}^2)+n\mu^2-2n\mu\bar{Y} \end{cases} \implies \begin{cases} n(S_X^2+\bar{X}^2+\mu^2-2\mu\bar{X}) \\ n(S_Y^2+\bar{Y}^2+\mu^2-2\mu\bar{Y}) \end{cases}$$

$$\frac{f_\theta(\vec{x})}{f_\theta(\vec{y})} = \left(\frac{\sigma_Y}{\sigma_X}\right)^n \cdot e^{-\frac{n}{2\sigma_X^2}(S_X^2+\bar{X}^2+\mu^2-2\mu\bar{X})+\frac{n}{2\sigma_Y^2}(S_Y^2+\bar{Y}^2+\mu^2-2\mu\bar{Y})}$$

Pero esta expresión no depnde 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,

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

Por lo que el estadístico T *no es completo*

Teorema 3.6.2

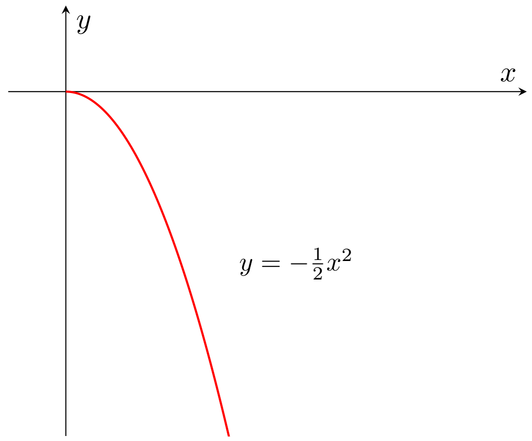
El estadístico natural $(\sum_{i=1}^n T_1(X_i),\dots\sum_{i=1}^n T_k(X_i))$ *de la familia exponencial k-paramétrica,* $\{f_\theta(x) = c(\theta)h(x)e^{\sum_{j=1}^k q_j(\theta)T_j(x)}\}_{\theta\in\Theta\subset\mathbb{R}^1}$*, es completo si la imagen de la aplicación* $q = (q_1(\theta),\dots,q_k(\theta)) : \Theta \to \mathbb{R}^k$ *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 $X \sim Bernoulli(\theta) \implies P(X = x) = \theta^x(1-\theta)^{1-x} \implies$

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

En este caso, la distribución no tiene exponencial, pero se puede refactrizar 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:

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

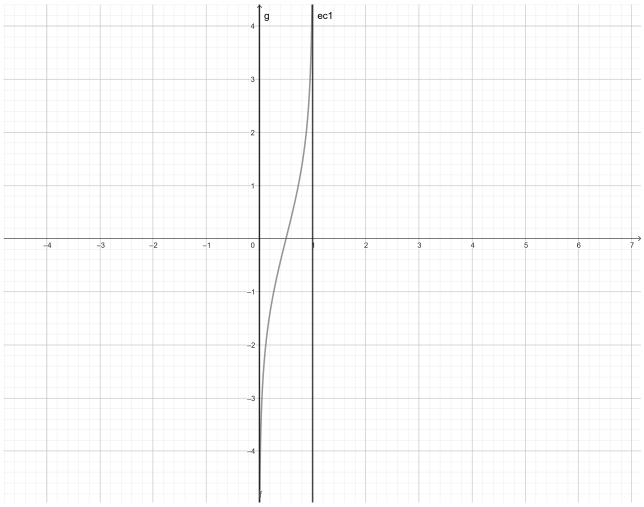
Entonces, podemos obtener que como $(X_1 \ldots, X_n)$ pertenece a la famiia exponencial k-paramétrica, obtenemos que su estadístico natural $(T(\vec{x}) = \sum x_i)$ es suficiente. Y como

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

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 imaagen de $q(\theta)$:

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

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 perteniencia 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 multipicación:

$$\begin{cases} x = u \\ dv = f_\theta(x_{(n)}) \end{cases} \implies$$

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

Ahora, realicemoslo para el mínimo:

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

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 ancilario, 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:

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

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

Aplicando las propiedades de los condicionales, tenemos que:

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

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

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

Entones 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:

$$X \sim U(0,\theta) \implies f_\theta(x) = \frac{1}{\theta} \cdot I_{(0,\theta)}(x) \implies F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = P(X_1 \leq x, \dots, X_n \leq x) =$$

$$= P(X_1 \leq x) \cdot \ldots \cdot P(X_n \leq x) = (F(x))^n = \left(\frac{x}{\theta}\right)^n \implies f_\theta(x) = n\left(\frac{x}{\theta}\right)^{n-1} \cdot \frac{1}{\theta} = \frac{n}{\theta^n}x^{n-1}$$

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 $X_{(1)}$:

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

$$= 1 - P(X_1 > x) \cdot \ldots \cdot P(X_n > x) = 1 - (1 - F(x))^n = 1 - \left(1 - \frac{x}{\theta}\right)^n \implies f_\theta(x) = n\left(1 - \frac{x}{\theta}\right)^{n-1} \cdot \frac{1}{\theta} = \frac{n}{\theta^n}(\theta - x)^{n-1}$$

Ejemplo

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

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

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 $\bar{X}$:

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

La completitud se da por definición. No obstante es una integral demasiado dificil 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

Teorema 3.7.1 [Principio de verosimilitud]

Dada una m.a.s. consideramos que la distribución de la muestra es una función de θ y no de la muestra en sí, es decir, $L(\theta|x_1, \ldots, x_n) = f(x_1, \ldots, x_n|\theta)$

Proposición 3.7.1

Sean dos muestras $X = (x_1, \ldots, x_n)$ e $Y = (y_1, \ldots, y_m)$, si $\exists c(x, y)$ tal que $L_1(\theta|x) = c(x, y)L_2(\theta|y)$, entonces la evidencia estadística que suministran ambas muestras es idéntica. Es decir, $Ev(E_1, \vec{x}) = Ev(E_2, \vec{y})$

Ejemplo

Sea un modelo $Binomial(n, \theta)$ 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 $Binomialnegativo(t, \theta)$, siempre que se suponga que se han realizado n repeticiones hasta obtener t éxitos.

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

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

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 numero de cara s 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)

Proposición 3.7.2 [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 muetras $\vec{x}$ 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.

Proposición 3.7.3 [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 3.7.2 [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 $|X|$ un estadístico suficiente?

Solución:

□