Estadística[1]

Segundo Cuatrimestre 2025

Pau Frangi Mahiques
Ciencias Matemáticas e
Ingenería Informática

[1]basado en los apuntes míos de clase

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:

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

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

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 $(X_1,\dots,X_n)$ 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:*

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

   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.

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

$\implies$ 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 $nS_{n+1}^2 = (n-1)S_n^2 + \left(X_{n+1} - \bar{X}_n\right)^2 \frac{n}{n+1}$
$nS_{n+1}^2 = \sum_{i=1}^{n+1}\left(X_i - \bar{X}_{n+1}^2\right) = \sum_{i=1}^{n}\left(\left(X_i - \bar{X}_n\right) + \left(\bar{X}_n - \bar{X}_{n+1}\right)\right)^2 + \left(X_{n+1} - \bar{X}_{n+1}\right)^2 =$
$(n-1)S_n^2 + n\left(\bar{X}_n - \bar{X}_{n+1}\right)^2 + \left(X_{n+1} - \bar{X}_{n+1}\right)^2 + \sum_{i=0}^{n} 2\left(X_i - \bar{X}_n\right)\left(\bar{X}_n - \bar{X}_{n+1}\right) =$
$(n-1)S_n^2 + n\left(\bar{X}_n - \bar{X}_{n+1}\right)^2 + \left(X_{n+1} - \bar{X}_{n+1}\right)^2 + 2\left(\bar{X}_n - \bar{X}_{n+1}\right)\sum_{i=0}^{n}\left(X_i - \bar{X}_n\right) =$
$(n-1)S_n^2 + n\left(\bar{X}_n - \bar{X}_{n+1}\right)^2 + \left(X_{n+1} - \bar{X}_{n+1}\right)^2 =$
En seste ultimo paso se desarrollan cuadrados, y se aplica la definicion de $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \leftrightarrow n\bar{X}_n = \sum_{i=1}^{n} X_i$
Ahora si se aplica que $\bar{X}_{n+1} = \frac{n\bar{X}_n + X_{n+1}}{n+1} \leftrightarrow \left(X_{n+1} - \bar{X}_{n+1}\right) = \frac{(n+1)X_{n+1} - n\bar{X}_n - X_{n+1}}{n+1}$
$(n-1)S_n^2 + \left(X_{n+1} - \bar{X}_n\right)^2 \frac{n}{(n+1)^2} + \left(X_{n+1} - \bar{X}_n\right)^2 \frac{n^2}{(n+1)^2} =$
$(n-1)S_n^2 + \left(X_{n+1} - \bar{X}_n\right)^2 \frac{n}{n+1}$
Asi ya tenemos el resultado deseado
Ahora veamos: $\left(\frac{X_{n+1} - \bar{X}_n}{\sigma\sqrt{\frac{n+1}{n}}}\right)^2 \sim \chi_1^2$

$$Usemos: X_{n+1} \sim N(\mu, \sigma) \text{ y } \bar{X}_n \sim N(\mu, \frac{\sigma}{\sqrt{n}}) \tag{1}$$

$$X_{n+1} - \bar{X}_n \sim N(0, \sigma\sqrt{\frac{n+1}{n}}) \tag{2}$$

$$\left(\frac{X_{n+1} - \bar{X}_n}{\sigma\sqrt{\frac{n+1}{n}}}\right) \sim N(0,1) \sim \chi_1^2 \tag{3}$$

Finalmente apliquemos induccion: Nuestro caso base es el siguiente: $\frac{1S_2^2}{\sigma^2} = \frac{1}{\sigma^2}\left(\left(X_1 - \bar{X}_2\right)^2 + \left(X_2 - \bar{X}_2\right)^2\right) =$
$\frac{1}{\sigma^2}\left(\left(X_1 - \frac{X_1+X_2}{2}\right)^2 + \left(X_2 - \frac{X_1+X_2}{2}\right)^2\right) =$
$\frac{1}{\sigma^2}\left(\frac{1}{4}\left(X_1 - X_2\right)^2 + \frac{1}{4}\left(X_2 - X_1\right)^2\right) = \frac{1}{\sigma^2}\frac{1}{2}\left(X_2 - \bar{X}_1\right)^2 =$
$\left(\frac{\left(X_2 - \bar{X}_1\right)}{\sigma\sqrt{2}}\right)^2 \sim \chi_1^2$
Por lo visto antes Nuestra hipotesis de induccion es que $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2$
Asi, por inducción (teniendo en cuenta que S_n^2 y $\bar{X}_n$ son independientes por (1))
$\frac{nS_{n+1}^2}{\sigma^2} = \frac{(n-1)S_n^2}{\sigma^2} + \left(\frac{X_{n+1} - \bar{X}_n}{\sigma\sqrt{\frac{n+1}{n}}}\right)^2 \sim \chi_{n-1}^2 + \chi_1^2 \sim \chi_n^2$

3. Se deduce de los anteriores

□

Definición 2.4.1 [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:

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

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

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

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

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

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

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:

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

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.

Ejemplo

Calcular la distribución del rango muestral

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

$$0 < x < y < 1 \Rightarrow \frac{r}{2} < h < 1 - \frac{r}{2}$$

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

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

Ejemplo

Sea $(X_1, \cdots X_n)$ m.a.s. $(n) \sim X$ tal que $F(x) = P(X \leq x)$
Se pide calcular la distribución de $U = F(X)$ y de $U_R = F\left(X_{(r)}\right)$
$G_U(u) = P(U \leq u) = P(F(X) \leq u) = P\left(X \leq F^{-1}(u)\right) = F\left(F^{-1}(u)\right) = u, \quad 0 < u < 1$
$\Rightarrow U = F(X) \sim U(0,1)$
$G_{U_R}(u) = P\left(U_R \leq u\right) = P\left(F\left(X_{(r)}\right) \leq u\right) = P\left(X_{(r)} \leq F^{-1}(u)\right) =$
$F_{X_{(r)}}\left(F^{-1}(u)\right) = \sum_{j=r}^{n} \binom{n}{j} F\left(F^{-1}(u)\right)^j \left(1 - F\left(F^{-1}(u)\right)\right)^{n-j} = \sum_{j=r}^{n} \binom{n}{j} u^j (1-u)^{n-j}$
$\Rightarrow U_R = F\left(X_{(r)}\right)$ es el estadístico ordenado de orden r asociado a una m.a.s. (n) de una población $U = F(X) \sim U(0,1)$

2.5 Estadísticos suficientes

Definición 2.5.1 [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. $(n) \sim X$
$T = T\left(X_1, \cdots X_n\right) : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ es un estadístico suficiente para θ (para la familia de funciones de distribución $\{F_\theta\}_{\theta \in \Theta \subset \mathbb{R}^\ell}$), sí y sólo sí la distribución de probabilidad de la muestra condicionada por T es independiente de θ

Ejemplo

Demostrar que si $X \sim \text{Bin}(1, \theta)$, entonces $T = \sum_{i=1}^{n} X_i \sim \text{Bin}(n, \theta)$ es suficiente para θ
$P_\theta(X = x) = \theta^x (1-\theta)^{1-x}\big|_{\{0,1\}}(x)$
$P_\theta\left(\sum_{i=1}^{n} X_i = t\right) = \binom{n}{t} \theta^t (1-\theta)^{n-t} \left\{_{\{0,1,\ldots,n\}}(t)\right.$

$P_\theta(X_1=x_1,\ldots,X_n=x_n)=\prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i} I_{\{0,1\}}(x_i)=$
$\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}\prod_{i=1}^n l_{\{0,1\}}(x_i)$
Si $\sum_{i=1}^n x_i=t, P_\theta\left(X_1=x_1,\ldots,X_n=x_n \mid \sum_{i=1}^n X_i=t\right)=$
$\frac{P_\theta\left(X_1=x_1,\ldots,x_{n-1}=x_{n-1},X_n=t-\sum_{i=1}^{n-1}x_i\right)}{P\left(\sum_{i=1}^n X_i=t\right)}=\frac{\theta^t(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}}=\frac{1}{\binom{n}{t}}, t=0,1,\ldots,n$

Teorema de Factorización de Fisher

Teorema 2.5.1 [de Factorización de Fisher]

Teorema de Factorización (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. $\Leftrightarrow$) Si $T(x_1,\ldots,x_n)=t$,
$f_\theta(x_1,\ldots,x_n\mid t)=\frac{f_\theta(x_1,\ldots,x_n)}{f_\theta(t)}=\frac{f_\theta(x_1,\ldots,x_n)}{\sum_{\{(y_1,\ldots,y_n):T(y_1,\ldots,y_n)=t\}}f_\theta(y_1,\ldots,y_n)}=$
$\frac{h(x_1,\ldots,x_n)g_\theta(t)}{\sum_{\{(y_1,\ldots,y_n):T(y_1,\ldots,y_n)=t\}}h(y_1,\ldots,y_n)g_\theta(t)}=\frac{h(x_1,\ldots,x_n)}{\sum_{\{y_1,\ldots,y_n):T(y_1,\ldots,y_n)=t\}}(y_1,\ldots,y_n)}$ es
independiente de $\theta\Rightarrow T$ es suficiente para θ
$\Rightarrow$) Sea $T=T(X_1,\cdots X_n)$ un estadístico suficiente para $\theta\Rightarrow f_\theta(x_1,\ldots,x_n\mid t)$ es independiente de θ
Si $T(x_1,\ldots,x_n)=t, f_\theta(x_1,\ldots,x_n)=f_\theta(x_1,\ldots,x_n\mid t)\,f_\theta(t)=h(x_1,\ldots,x_n)\,g_\theta(T(x_1,\ldots,x_n))$ □

Teorema de Factorización de Fisher (continuación)

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

Ejercicios propuestos

1$T=\sum_{i=1}^n X_i$ es suficiente para θ si $X\sim\text{Poisson}(\theta)$
2 La propia muestra $(X_1,\ldots,X_n)$ y el estadístico ordenado $\left(X_{(1)},\ldots,X_{(n)}\right)$ son suficientes para θ
3 Si $T=T(X_1,\cdots X_n)$ es suficiente para θ entonces cualquier biyección $S=S(T)$ también es suficiente para θ
4 Si $T=T(X_1,\cdots X_n)$ y $S=S(X_1,\cdots X_n)$ son dos estadísticos suficientes para θ, entonces también es suficiente para θ el estadístico $(\ T,S\ )$

Teorema de Factorización de Fisher (continuación)

5$T=\bar{X}$ es suficiente para μ si $X\sim N(\mu,\sigma_0),\sigma_0$ conocida
6$T=\sum_{i=1}^n(X_i-\mu_0)^2$ es suficiente para σ si $X\sim N(\mu_0,\sigma),\mu_0$ conocida
7$T=\left(\sum_{i=1}^n X_i,\sum_{i=1}^n X_i^2\right)$ es suficiente para $\theta=(\mu,\sigma)$ si $X\sim N(\mu,\sigma),\mu$ y σ desconocidas $\Rightarrow\left(\bar{X},S_n^2\right)$ también es suficiente

8 Los estadísticos $\left(X_{(1)},X_{(n)}\right)$ y $X_{(n)}$ son suficientes para θ si $X\sim U(0,\theta)$
9 El estadístico $\left(X_{(1)},X_{(n)}\right)$ es suficiente para θ si $X\sim U\left(-\frac{\theta}{2},\frac{\theta}{2}\right)$

Estadístico minimal suficiente

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

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

Definición 2.5.2 [Estadístico Minimal Suficiente]

---

1. *$T = T(X_1, \cdots X_n)$ es minimal suficiente sí y sólo sí K_T es suficiente y $\forall S = S(X_1, \cdots X_n)$ suficiente, K_S es una subpartición de K_T*

2. *$T = T(X_1, \cdots X_n)$ es minimal suficiente sí y sólo sí T es suficiente y $\forall S = S(X_1, \cdots X_n)$ suficiente, $\exists \psi$ tal que $\psi(S) = T$*

Demostración. $\Rightarrow$) Sea S suficiente, si $S(x_1, \ldots, x_n) = S(y_1, \ldots, y_n) = s \Rightarrow (x_1, \ldots, x_n), (y_1, \ldots, y_n) \in B_s \Rightarrow \exists A_t \in K_T$ tal que $B_s \subset A_t \Rightarrow T(x_1, \ldots, x_n) = T(y_1, \ldots, y_n) = t \Rightarrow \exists \psi$ tal que $\psi(S) = T$ y T es suficiente
$\Leftrightarrow$) Sea S suficiente y φ tal que $\psi(S) = T \Rightarrow T$ es suficiente y si $S(x_1, \ldots, x_n) = S(y_1, \ldots, y_n) = s \Rightarrow T(x_1, \ldots, x_n) = \psi(S(x_1, \ldots, x_n)) = \psi(s) = \psi(S(y_1, \ldots, y_n)) = T(y_1, \ldots, y_n) \Rightarrow B_s \subset A_{\psi(s)} \Rightarrow K_S$ es una subpartición de K_T □

Teorema de caracterización de estadísticos

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

Demostración

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

Teorema de caracterización de estadísticos minimales suficientes (continuación)

Demotremos ahora que T es suficiente (caso discreto)
Si $T(x_1, \ldots, x_n) = t, f_\theta(x_1, \ldots, x_n \mid t) = \frac{f_\theta(x_1, \ldots, x_n)}{f_\theta(t)} =$
$\frac{f_\theta(x_1, \ldots, x_n)}{\sum_{\{(y_1, \ldots, y_n): T(y_1, \ldots, y_n) = t\}} f_\theta(y_1, \ldots, y_n)} = \frac{1}{\sum_{(y_1, \ldots, y_n) \in A_t} \frac{f_\theta(y_1, \ldots, y_n)}{f_\theta(x_1, \ldots, x_n)}}$ es independiente de $\theta \Rightarrow T$ es suficiente para θ

Ejercicio $\sum_{i=1}^n X_i$ es minimal suficiente para θ si $X \sim \text{Bin}(1, \theta)$
$f_\theta(x_1, \ldots, x_n) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1 - x_i} \big|_{\{0,1\}} (x_i) =$
$\theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^{n=1} x_i} \prod_{i=1}^n I_{\{0,1\}} (x_i)$

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

Familia exponencial k-paramétrica

Definición 2.5.3 [Familia exponencial k-paramétrica]

Sea $(\Omega, \mathcal{A}, \mathcal{P})$ el espacio probabilístico asociado a un experimento aleatorio $\mathcal{E}$, una variable aleatoria observable $X : \Omega \longrightarrow \mathbb{R}$ 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. $(n) \sim X$

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

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

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

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

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

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

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

Familia exponencial k-paramétrica (continuación)

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

Demostración

$\frac{f_f(x_1,\dots,x_n)}{f_\theta(y_1,\dots,y_n)} = \frac{c(\theta)^n \prod_{i=1}^n h(x_i) \sum^{\sum_{j=1}^k q_j(\theta) \sum_{i=1}^n T_j(x_i)}}{c(\theta)^n \prod_{i=1}^n h(y_i) e^{\sum_{j=1}^k q_j(\theta) \sum_{i=1}^n T_j(y_i)}}$
$= \frac{\prod_{i=1}^n h(x_i)}{\prod_{i=1}^i h(y_i)} e^{\sum_{j=1}^k q_j(\theta)\left(\sum_{i=1}^n T_j(x_i) - \sum_{i=1}^n T_j(y_i)\right)}$ es independiente de θ sí y sólo sí $\sum_{j=1}^k q_j(\theta)\left(\sum_{i=1}^n T_j\left(x_i\right) - \sum_{i=1}^n T_j\left(y_i\right)\right)$ 0. En este caso, el sistema homogéneo $\sum_{j=1}^k q_j\left(\theta_r\right)\left(\sum_{i=1}^n T_j\left(x_i\right) - \sum_{i=1}^n T_j\left(y_i\right)\right) = 0$, $r = 1, \dots, k$, sólo admite la solución $\sum_{i=1}^n T_j\left(x_i\right) - \sum_{i=1}^n T_j\left(y_i\right) = 0$, $r = 1, \dots, k$. Entonces $\left(\sum_{i=1}^n T_1\left(X_i\right), \dots, \sum_{i=1}^n T_k\left(X_i\right)\right)$ es minimal (Teorema de caracterización de estadísticos minimales suficientes)

Familia exponencial k-paramétrica (continuación)

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

$N(\mu,\sigma)$ pertenece a la familia exponencial k-paramétrica con $k=2$ $f_\theta(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}=c(\theta)h(x)e^{-\frac{1}{2\sigma^2}x^2+\frac{\mu}{\sigma^2}x}$
$\Rightarrow T=\left(\sum_{i=1}^n X_i,\sum_{i=1}^n X_i^2\right)$ es natural suficiente para θ
Además, $q_1(\theta)=\frac{\mu}{\sigma^2}$ y $q_2(\theta)=-\frac{1}{2\sigma^2}$ y tomando $\theta_1=(0,1)$ y $\theta_2=(1,1)$, los vectores $c_1=(q_1(\theta_1),q_2(\theta_1))=\left(0,-\frac{1}{2}\right)$ y $c_2=(q_1(\theta_2),q_2(\theta_2))=\left(1,-\frac{1}{2}\right)$ son linealmente independientes $\Rightarrow T$ es minimal

Familia exponencial k-paramétrica (continuación)

$S_n^2=\frac{n}{n-1}\left(\sum_{i=1}^n X_i^2-n\bar{X}^2\right)$
Denotemos por $(W,Z)=\left(\sum_{i=1}^n X_i,\sum_{i=1}^n X_i^2\right)$ y
$(F,G)=\left(\bar{X},S_n^2\right)=\left(\frac{w}{n},\frac{n}{n-1}\left(Z-\frac{W^2}{n}\right)\right)$
Transformación inversa ($W=nF, Z=\frac{n-1}{n}G+nF^2$)
$J=\begin{vmatrix} n & 0 \\ 2nF & \frac{n-1}{n}\end{vmatrix}=n-1\neq 0$ si $n\geq 2$
Por lo tanto existe una transformación biyectiva ψ_1 tal que $(F,G)=\psi_1(W,Z)$ y como (W,Z) es minimal suficiente, $\forall S$ suficiente, existe una transformación ψ_2 tal que $\psi_2(S)=(W,Z)$. Por lo tanto, $\psi_1\psi_2(S)=\psi_1(W,Z)=(F,G)$ y (F,G) es minimal suficiente

Estadísticos Ancilarios y Completos

Sea $(\Omega,\mathcal{A},\mathcal{P})$ el espacio probabilístico asociado a un experimento aleatorio $\mathcal{E}$, una variable aleatoria observable $X:\Omega\longrightarrow\mathbb{R}$ 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. $(n)\sim X$

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

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

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

Estadísticos Ancilarios y Completos (continuación)

Ejercicio La familia de distribuciones de probabilidad $\mathrm{Bin}(n,\theta)$ es completa
Sea $Y\sim\mathrm{Bin}(n,\theta)$,
$E_\theta[h(Y)]=\sum_{i=1}^n h(i)\binom{n}{i}\theta^i(1-\theta)^{n-i}=$
$(1-\theta)^n\sum_{i=1}^n h(i)\binom{n}{i}\left(\frac{\theta}{1-\theta}\right)^i=0,\forall\theta\in(0,1)$,
que es un polinomio de grado n en $\frac{\theta}{1-\theta}\in(0,\infty)$, luego para que sea nulo, ha de ser $h(i)=0,\forall i=1,\ldots,n$.
Por lo tanto $h(Y)\overset{cs}{=}0$

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

Estadísticos Ancilarios y Completos (continuación)

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

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

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

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

Estadísticos Ancilarios y Completos (continuación)

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

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

Estadísticos Ancilarios y Completos (continuación)

Teorema de Basu Si $T=T\left(X_1,\ldots,X_n\right)$ es un estadístico suficiente y completo y $U=U\left(X_1,\ldots,X_n\right)$ es un estadístico ancilario, entonces T y U son independientes

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

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

Principios de reducción de datos

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

Principio de verosimilitud

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

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

Principios de reducción de datos (continuación)

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

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

Principios de reducción de datos (continuación)

Ejercicio $\mathrm{Ev}\,(E_1, t) = \mathrm{Ev}\,(E_2, n)$, si
$E_1 = (\{0, 1, \ldots, n\}, \mathrm{Bin}(n, \theta))_{\theta \in (0,1)}, f(t \mid \theta) = \binom{n}{t} \theta^t (1-\theta)^{n-t}$
$E_2 = (\mathbb{N}, BN(t, \theta))_{\theta \in (0,1)}, g(n \mid \theta) = \binom{n-1}{t-1} \theta^t (1-\theta)^{n-t}$
Ejercicio Los procedimientos bayesianos, por estar basados en la distribución de probabilidad final o a posteriori, satisfacen el principio de verosimilitud, ya que si $\mathbf{x} = (x_1, \ldots, x_n)$ e $\mathbf{y} = (y_1, \ldots, y_m)$ satisfacen el principio de verosimilitud, entonces $\pi_1(\theta \mid \mathbf{x}) = \pi_2(\theta \mid \mathbf{y})$

Principio de suficiencia

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

Principios de reducción de datos (continuación)

Principio de condicionalidad

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

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

Teorema de Birnbaum

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

Observación El Teorema de Birnbaum es importante desde el punto de vista de los fundamentos de la Estadística. Muchos de los procedimientos estadísticos usuales violan el principio de verosimilitud, en concreto los procedimientos que se basan en la distribución en el muestreo de un estadístico pueden hacerlo.

Por ejemplo, si se pasa de un modelo binomial a uno binomial negativo, la función de masa cambia y por lo tanto los IC pueden cambiar. Sin embargo, a la luz del teorema, esto significa contradecir el principio de suficiencia, que es compartido por toda aproximación a la inferencia, o el principio de condicionalidad, que parece bastante aséptico. El teorema de Birnbaum constituye uno de los motivos por los que el principio de verosimilitud no es universalmente aceptado, a pesar de que como se verá, la función de verosimilitud posee muchas buenas propiedades estadísticas

Propiedades de los estimadores

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. $(n) \sim X$

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

Estimadores centrados

$T = T\left(X_1, \cdots X_n\right) : \chi^n \longrightarrow \Theta$ es un estimador centrado para θ cuando $E_\theta[T] = \theta$. En general, se llama sesgo de un estimador a la diferencia $b(T, \theta) = E_\theta[T] - \theta$

Propiedades de los estimadores (continuación)

Ejercicio ($\bar{X}, S^2$) es un estimador centrado de $\theta = \left(\mu, \sigma^2\right)$
Ejercicio $\sigma_n^2 = b_2 = \frac{1}{n} \sum_{i=1}^n \left(X_i - \bar{X}\right)^2$ es un estimador centrado de $h(\theta) = \frac{n-1}{n} \sigma^2$ y $b\left(\sigma_n^2, \sigma^2\right) = -\frac{\sigma^2}{n}$

Observaciones

(1) Puede ser que no exista un estimador centrado de θ
(2) Si T es un estimador centrado para θ, en general $h(T)$ no tiene por qué ser centrado para $h(\theta)$
(0) A pesar de que exista un estimador centrado para θ, puede ser que no tenga sentido
(- A pesar de que exista un estimador centrado para θ, puede se que su varianza sea muy grande y no sea adecuado para la estimación

Propiedades de los estimadores (continuación)

Ejercicio Para una m.a.s. (n = 1) de $X \sim \operatorname{Bin}(1, \theta), T(X) = X^2$ no es un estimador centrado de θ^2

Ejercicio Para una m.a.s. (n = 1) de $X \sim \operatorname{Bin}(1, \theta)$, no existe un estimador centrado de θ^2

Ejercicio Para una m.a.s. (n = 1) de $X \sim \operatorname{Poisson}(\theta), T(X) = (-2)^X$ es centrado para $h(\theta) = e^{-3\theta}$ y $V_\theta(T) = e^{4\theta} - e^{-6\theta} \underset{\theta \to \infty}{\longrightarrow} \infty$, pero no es un estimador de $h(\theta)$

Ejercicio Si$T_j, j = 1, 2, \ldots$ es una 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
Si además, los estimadores T_j son independientes y $V_\theta\left(T_j\right) < \sigma^2 < \infty, j = 1, 2, \ldots$, entonces $V_\theta\left(\bar{T}_k\right) \leq \frac{\sigma^2}{k} \underset{k \to \infty}{\longrightarrow} 0 \Rightarrow \bar{T}_k \xrightarrow[k \to \infty]{P} \theta$ (Teorema de Markov)

Propiedades de los estimadores (continuación)

Estimadores consistentes

Una sucesión de estimadores $T_n = T(X_1, \ldots, X_n)$ es consistente para el parámetro θ, si $T_n \xrightarrow[n\to\infty]{P} \theta, \forall \theta \in \Theta$, es decir si $\forall \varepsilon > 0$ se tiene que $\lim_{n\to\infty} P_\theta(|T_n - \theta| \leq \varepsilon) = 1, \forall \theta \in \Theta$

Proposición 1 (condición suficiente) Si $T_n = T(X_1, \ldots, X_n)$ es una sucesión de estimadores tales que $\forall \theta \in \Theta, E_\theta[T_n] \underset{n\to\infty}{\longrightarrow} \theta$, $V_\theta(T_n) \underset{n\to\infty}{\longrightarrow} 0$, entonces T_n es consistente para θ

Demostración $E_\theta\left[(T_n - \theta)^2\right] = V_\theta(T_n) + b(T_n, \theta)^2 \underset{n\to\infty}{\longrightarrow} 0, \forall \theta \in \Theta$ entonces $T_n \xrightarrow[n\to\infty]{\text{m.c.}} \theta, \forall \theta \in \Theta$

Propiedades de los estimadores (continuación)

Estimadores bayesianos

En la aproximación bayesiana θ es una v.a. con distribución inicial o a priori dada por una función de masa o de densidad $\pi(\theta)$. Cuando se observa una muestra se calcula la distribución final o a posteriori,

$$\pi(\theta \mid x_1, \ldots, x_n) = \frac{\pi(\theta) f(x_1, \ldots, x_n \mid \theta)}{m(x_1, \ldots, x_n)}$$
$$m(x_1, \ldots, x_n) = \int_\Theta \pi(\theta) f(x_1, \ldots, x_n \mid \theta)\, d\theta$$

donde m se llama distribución predictiva

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

Propiedades de los estimadores (continuación)

Ejercicio (Familias conjugadas)

Si $(X_1, \ldots, X_n)$ es una m.a.s. (n) de $X \sim \text{Bin}(1, \theta)$ y $\theta \sim U(0,1)$, entonces $\pi(\theta \mid x_1, \ldots, x_n) \sim \text{Beta}\left(\sum_{i=1}^n x_i + 1, n -\right.$

En general, $\pi(\theta \mid x_1, \ldots, x_n) \sim \text{Beta}\left(\sum_{i=1}^n x_i + \alpha, n - \sum_{i=1}^n x_i + \beta\right)$ si $(X_1, \ldots, X_n)$ es una m.a.s. (n) de $X \sim \text{Bin}(1, \theta)$ y $\theta \sim \text{Beta}(\alpha, \beta)$

En este caso, se dice que la familia de distribuciones iniciales $\text{Beta}(\alpha, \beta)$ es conjugada de la familia de distribuciones de probabilidad $X \sim \text{Bin}(1, \theta)$

Además, $E[\theta \mid x_1, \ldots, x_n] = \frac{\sum_{i=1}^n x_i + \alpha}{n+\alpha+\beta} = \frac{n}{n+\alpha+\beta}\bar{x} + \frac{\alpha+\beta}{n+\alpha+\beta}\frac{\alpha}{\alpha+\beta}$

Propiedades de los estimadores (continuación)

Ejercicio Si $(X_1, \ldots, X_n)$ es una m.a.s. (n) de $X \sim N(\mu, \sigma)$ con σ conocida y $\mu \sim N(\mu_0, \sigma_0)$, entonces $\pi(\mu \mid x_1, \ldots, x_n) \sim N(\mu_1, \sigma_1)$,

$$\mu_1 = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{\bar{x}}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{1}{\frac{\sigma^2}{\frac{\sigma n}{n}}}} = \frac{\sigma^2}{\sigma^2 + n\sigma_0^2}\mu_0 + \frac{n\sigma_0^2}{\sigma^2 + n\sigma_0^2}\bar{x}$$

$$\sigma_1 = \sqrt{\frac{1}{\frac{1}{\sigma_0^2} + \frac{1}{\frac{\sigma^2}{n}}}}$$

Propiedades de los estimadores (continuación)

Estadístico suficiente bayesiano

$T = T\left(X_1, \cdots X_n\right) : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ es un estadístico suficiente bayesiano para θ (para la familia de funciones de distribución $\left\{F_\theta\right\}_{\theta \in \Theta \subset \mathbb{R}^\ell}$), respecto a la distribución inicial $\pi(\theta)$ sí $\pi\left(\theta \mid x_1, \ldots, x_n\right) = \pi(\theta \mid t)$, con $T\left(x_1, \ldots, x_n\right) = t$Teorema 3$T = T\left(X_1, \cdots X_n\right) : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ es un estadístico suficiente para θ sí y sólo sí T es un estadístico suficiente bayesiano para θ respecto a $\pi(\theta)$, cualquiera que sea la distribución inicial $\pi(\theta)$

Demostración

$$\Rightarrow \pi\left(\theta \mid x_1, \ldots, x_n\right) = \frac{\pi(\theta) f\left(x_1, \ldots, x_n \mid \theta\right)}{\int_\Theta \pi(\theta) f\left(x_1, \ldots, x_n \mid \theta\right) d\theta} = \frac{\pi(\theta) g(t \mid \theta) f\left(x_1, \ldots, x_n \mid t, \theta\right)}{\int_\Theta \pi(\theta) g(t \mid \theta) f\left(x_1, \ldots, x_n tt, \theta\right) d\theta} =$$

$$\frac{\pi(\theta) g(t \mid \theta)}{\int_\Theta \pi(\theta) g(t \mid \theta) d\theta} = \pi(\theta \mid t)$$

$$\Leftrightarrow) f\left(x_1, \ldots, x_n \mid \theta\right) = \frac{\pi\left(\theta \mid x_1, \ldots, x_n\right) m\left(x_1, \ldots, x_n\right)}{\pi(\theta)} = \frac{\pi(\theta \mid t)}{\pi(\theta)} m\left(x_1, \ldots, x_n\right)$$

Criterios de comparación de estimadores

Error cuadrático medio

Dado un estimador $T = T\left(X_1, \ldots, X_n\right)$ de θ, se denomina error cuadrático medio $\mathrm{ECM}(T, \theta) = E_\theta\left[\left(T_n - \theta\right)^2\right]$

Ejercicio Si $\left(X_1, \ldots, X_n\right)$ es una m.a.s.(n) de $X \sim N(\mu, \sigma)$, entonces

$$ECM(\bar{X}, \mu) = \frac{\sigma^2}{n}$$

$$ECM\left(S^2, \sigma^2\right) = V_\theta\left(S^2\right) = \frac{2\sigma^4}{n-1}$$

$$ECM\left(\sigma_n^2, \sigma^2\right) = V_\theta\left(\sigma_n^2\right) + b\left(\sigma_n^2, \sigma\right) = \frac{2n-1}{n^2}\sigma^4 < \frac{2\sigma^4}{n-1}$$

Criterios de comparación de estimadores

Pérdida final esperada

Dado el estimador $T=T\left(X_1, \ldots, X_n\right)$, la distribución inicial $\pi(\theta)$ y la función de pérdida $\mathcal{L}(\theta, t)$ en la que se incurre por estimar θ mediante $T\left(x_1, \ldots, x_n\right)=t$, se define la pérdida final esperada $\circ$ riesgo a posteriori,

$$
\operatorname{PFE}(t)=E\left[\mathcal{L}(t, \theta) \mid x_1, \ldots, x_n\right]=\int_{\Theta} \mathcal{L}(\theta, t) \pi\left(\theta \mid x_1, \ldots, x_n\right) d \theta
$$

Si $\mathcal{L}(\theta, t)=(\theta-t)^2$, entonces la pérdida final esperada se minimiza en $t^*=E\left[\theta \mid x_1, \ldots, x_n\right]$ (estimador bayesiano) y $\operatorname{PFE}(t)=V\left(\theta \mid x_1, \ldots, x_n\right)$

Si $\mathcal{L}(\theta, t)=|\theta-t|$, entonces la pérdida final esperada se minimiza en la mediana de la distribución final (estimador bayesiano)

Estimadores centrados de mínima varianza

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

Proposición 1 Si existe un ECUMV para θ, entonces es único c.s.

Demostración

Sean T_1 y T_2 ECUMV para θ y demostremos que entonces $T_1 \stackrel{\text{c.s.}}{=} T_2$ Sea $T=\frac{T_1+T_2}{2}$. Entonces, $E_\theta[T]=\theta$ y $V_\theta(T) \leq V_\theta\left(T_1\right)$

En efecto, $V_\theta(T)=\frac{1}{4}\left(V_\theta\left(T_1\right)+V_\theta\left(T_2\right)+2 \operatorname{Cov}_\theta\left(T_1, T_2\right)\right)=\frac{V_\theta\left(T_1\right)}{2}+\frac{\operatorname{Cov}_\theta\left(T_1, T_2\right)}{2} \leq V_\theta\left(T_1\right)$

$1 \geq \rho_\theta\left(T_1, T_2\right)=\frac{\operatorname{Cov}\left(T_1, T_2\right)}{\sqrt{V_\theta\left(T_1\right) V_\theta\left(T_2\right)}}=\frac{\operatorname{Cov}\left(T_1, T_2\right)}{V_\theta\left(T_1\right)} \Rightarrow \operatorname{Cov}\left(T_1, T_2\right) \leq V_\theta\left(T_1\right)$

Por lo tanto, $V_\theta(T)=V_\theta\left(T_1\right)=\operatorname{Cov}_\theta\left(T_1, T_2\right)=$ y $\rho_\theta\left(T_1, T_2\right)=1 \Rightarrow T_1 \stackrel{\text{c.s.}}{=} a+b T_2 \Rightarrow \theta=a+b \theta \Rightarrow a=0$ y $b=1$

Estimadores centrados de mínima varianza (continuación)

Teorema 4 El ECUMV es función simétrica de las observaciones

Ejercicio Para muestras de tamaño $n=2, T_1=\frac{X_1}{X_2}$ no puede puede ser ECUMV para θ

Si lo fuera, llamando $T_2=\frac{X_2}{X_1}$, se tendría $E_\theta\left[T_2\right]=E_\theta\left[T_1\right]$ y

$V_\theta\left(T_2\right)=V_\theta\left(T_1\right)$ y el estimador $T=\frac{T_1+T_2}{2}=\frac{x_1^2+X_2^2}{X_1 X_2}$ sería tal que $V_\theta(T)<V_\theta\left(T_1\right)$ (contradicción)

En efecto, si $V_\theta(T)=V_\theta(T_1)$, entonces $T \stackrel{c.s.}{\equiv} T_1$ **(contradicción)**

En general si T es centrado y consideramos las n ! permutaciones posibles de la muestra, podemos definir los estimadores centrados T_j al evaluar T en la permutación j-ésima. Entonces, el estimador $\bar{T}=\frac{1}{n!} \sum_{j=1}^{n!} T_j$ es tal que $V_\theta(\bar{T}) \leq V_\theta\left(T_j\right), j=1, \ldots, n!$, siendo la desigualdad estricta si T no es simétrico

Estimadores centrados de mínima varianza

(continuación)Teorema (caracterización del ECUMV)
$T_1 = T_1(X_1, \ldots, X_n)$ con $E_\theta[T_1] = \theta$ y $V_\theta(T_1) < \infty$ es el ECUMV para θ si y solo sí, cualquiera que sea $T_2 = T_2(X_1, \ldots, X_n)$ con $E_\theta[T_2] = 0$ y $V_\theta(T_2) < \infty, \forall\theta \in \Theta$, se tiene que $E_\theta[T_1 T_2] = 0$, $\forall\theta \in \Theta$

Corolario 1

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

Estimadores centrados de mínima varianza (continuación)

Teorema de Rao-Blackwell Si$T = T(X_1, \ldots, X_n)$ es un estimador centrado para θ con varianza finita y $S = S(X_1, \ldots, X_n)$ es un estadístico suficiente, entonces $g(S) = E[T \mid S]$ es un estimador centrado para θ con varianza finita tal que $V_\theta(g(S)) \leq V_\theta(T)$

Demostración

Como S es suficiente, $g(S) = E[T \mid S]$ no depende de θ y es un estadístico. Además, $E_\theta[g(S)] = E_\theta[E[T \mid S]] = E_\theta[T] = \theta$ y
$V_\theta(T) = E_\theta\left[(T-\theta)^2\right] = E_\theta\left[(T - g(S) + g(S) - \theta)^2\right] = E_\theta\left[(T - g(S))^2\right] + E_\theta\left[(g(S)-\theta)^2\right] + 2E_\theta[(T - g(S))(g(S) - \theta)] \geq V_\theta(g(S))$
En efecto, $E_\theta[(T - g(S))(g(S) - \theta)] = \iint (t - g(s))(g(s) - \theta) dF_\theta(t, s) = \int (g(s) - \theta) \int (t - g(s)) dF(t \mid s) dF_\theta(s) = 0$

Estimadores centrados de mínima varianza (continuación)

Teorema de Lehmann-Schefeé Si $S = S(X_1, \ldots, X_n)$ es un estadístico suficiente y completo para θ y $T = T(X_1, \ldots, X_n)$ es un estimador centrado para θ tal que $T = h(S)$, entonces T es ECUMV para θ

Demostración

Como S es suficiente y T es centrado para $\theta, g(S) = E[T \mid S]$ es centrado para θ y $V_\theta(g(S)) \leq V_\theta(T)$. Además, para cualquier otro estimador T_1 centrado para $\theta, g_1(S) = E[T_1 \mid S]$ es centrado para θ y $V_\theta(g_1(S)) \leq V_\theta(T_1)$. Por lo tanto, al ser T completo y $E_\theta[g(S) - g_1(S)] = \theta - \theta = 0, \forall\theta \in \Theta$, se tiene que $g(S) \overset{\text{c.s.}}{=} g_1(S)$. En particular, para $T_1 = T = h(S), g_1(S) = E[h(S) \mid S] = h(S) = T$, y $V_\theta(T) \leq V_\theta(T_1)$, cualquiera que sea T_1 centrado para θ

Estimadores centrados de mínima varianza

(continuación)Ejercicio Si $X \sim \operatorname{Bin}(1, \theta), T = \sum_{i=1}^n X_i$ es suficiente y completo, y $\bar{X} = h(T)$ es el ECUMV para θ

Ejercicio Si $X \sim$ Poisson (θ), encontrar el ECUMV para $d(\theta)=e^{-2\theta}$ basado en una m.a.s.(n)
Indicación: $T=T\left(X_1,\ldots,X_n\right)=\left\{\begin{array}{ll}1 & \text{si } X_1+X_2=0 \\ 0 & \text{resto}\end{array}\right.$ es centrado para $d(\theta)=e^{-2\theta}$ y $S=\sum_{i=1}^n X_i$ es suficiente y completo

Cota para la varianza de un estimador

Consideremos $X \approx \left(\chi, \beta_\chi, F_\theta\right)_{\theta\in\Theta\subset\mathbb{R}}$ modelo estadístico uniparamétrico contínuo (o discreto) y sea ($X_1,\cdots X_n$) muestra de $\{F_\theta, \theta\in\Theta\}$ siendo $f_\theta\left(x_1,\cdots,x_n\right)$ 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) $\operatorname{Sop}\left(f_\theta\right)=\left\{\left(x_1,\cdots,x_n\right)\in\chi^n: f_\theta\left(x_1,\cdots,x_n\right)>0\right\}$ no depende de θ

- $\forall\left(x_1,\cdots,x_n\right)\in\chi^n$ y $\forall\theta\in\Theta, \exists\frac{\partial}{\partial\theta} f_\theta\left(x_1,\cdots,x_n\right)$
  (-) $\int_{\chi^n}\frac{\partial}{\partial\theta} f_\theta\left(x_1,\cdots,x_n\right) dx_1\cdots dx_n=0$
  (0) $I_n(\theta)=E\left[\left(\frac{\partial}{\partial\theta}\log f_\theta\left(x_1,\cdots,x_n\right)\right)^2\right]<\infty$(cantidad de información de Fisher)

Cota para la varianza de un estimador

(continuación)Teorema (Cota de Fréchet-Cramér-Rao) Si $T=T\left(X_1,\cdots,X_n\right)$ es un estadístico unidimensional tal que $E_\theta\left[T^2\right]<\infty, E_\theta[T]=d(\theta)$ y $d'(\theta)=\int_{\chi^n} T\left(x_1,\cdots,x_n\right)\frac{\partial}{\partial\theta} f_\theta\left(x_1,\cdots,x_n\right) dx_1\cdots dx_n$, 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

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

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

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

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

Cota para la varianza de un estimador (continuación)

Además, $E_\theta\left[\frac{\partial}{\partial\theta}\log f_\theta\left(x_1,\cdots,x_n\right)\right]=0$ y por lo tanto,
$V_\theta\left[\frac{\partial}{\partial\theta}\log f_\theta\left(x_1,\cdots,x_n\right)\right]=E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_\theta\left(x_1,\cdots,x_n\right)\right)^2\right]=I_n(\theta)$
En efecto, $0=\int_{x^n}\frac{\partial}{\partial\theta} f_\theta\left(x_1,\cdots,x_n\right) dx_1\cdots dx_n=\int_{x^n}\frac{\partial}{\partial\theta}\log\left(f_\theta\left(x_1,\cdots,x_n\right)\right) f_\theta\left(x_1,\cdots,x_n\right) dx_1\cdots dx_n$

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

Entonces, $\operatorname{Cov}_\theta\left[T, \frac{\partial}{\partial\theta}\log f_\theta\right]=E\left[T\frac{\partial}{\partial\theta}\log f_\theta\right]=d'(\theta)$, 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 $P_\theta\left(T=d(\theta)+k(\theta)\frac{\partial}{\partial\theta} f_\theta\right)=1$

Cota para la varianza de un estimador (continuación)

En efecto, si $T \stackrel{\text{c.s.}}{=} a + b\frac{\partial}{\partial\theta}\log f_\theta$, entonces $d(\theta) = E_\theta[T] = ay$

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

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

Proposición 3 Bajo las suposiciones anteriores, si ($X_1, \cdots X_n$) es m.a.s. (n) de $\{F_\theta, \theta \in \Theta\}$, entonces $I_n(\theta) = nI_1(\theta)$ Indicación: $f_\theta(x_1, \cdots x_n) = \prod_{i=1}^n f_\theta(x_i)$

Cota para la varianza de un estimador (continuación)

Proposición 4 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

$f_\theta(x_1, \ldots, x_n) = c(\theta)^n \prod_{i=1}^n h(x_i) e^{q_1(\theta)\sum_{i=1}^n T_1(x_i)}$
$\frac{\partial}{\partial\theta}\log f_\theta = n\frac{c'(\theta)}{c(\theta)} + q_1'(\theta)\sum_{i=1}^n T_1(x_i)$
$\frac{1}{n}\sum_{i=1}^n T_1(x_i) = a(\theta) + b(\theta)\frac{\partial}{\partial\theta}\log f_\theta, a(\theta) = -\frac{c'(\theta)}{c(\theta)q_1'(\theta)}, b(\theta) = \frac{1}{nq_1'(\theta)}$
$\frac{1}{n}\sum_{i=1}^n T_1(x_i)$ es centrado para $d(\theta) = -\frac{c'(\theta)}{c(\theta)q_1'(\theta)}$ y alcanza la cota

Cota para la varianza de un estimador (continuación)

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

Ejercicio Si se cumplen las condiciones de regularidad y además
(1) $\forall(x_1, \cdots, x_n) \in \chi^n y \forall\theta \in \Theta, \exists\frac{\partial^2}{\partial\theta^2}f_\theta(x_1, \cdots, x_n)$
(2) $\int_{\chi^n}\frac{\partial^2}{\partial\theta^2}f_\theta(x_1, \cdots, x_n)\,dx_1\cdots dx_n = 0$

Entonces, $I_n(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(x_1, \cdots, x_n)\right]$
Indicación: $\frac{\partial}{\partial\theta}\log f_\theta(x_1, \cdots, x_n) = \frac{\partial}{\partial\theta}f_\theta(x_1, \cdots, x_n)\frac{1}{f_\theta(x_1, \cdots, x_n)}$

Cota para la varianza de un estimador

(continuación)

Estimadores eficientes

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

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

Métodos de construcción de estimadores

Método de los momentos
Este método consiste en elegir como estimador de un momento poblacional su momento muestral asociado, es decir
(1) El estimador por el método de los momentos del momento poblacional respecto al origen de orden k, $\alpha_k = E\left[X^k\right]$, es $a_k = \frac{1}{n}\sum_{i=1}^n X_i^k$
(2) El estimador por el método de los momentos del momento poblacional respecto a la media de orden k ,

$$\beta_k = E\left[(X-\alpha_1)^k\right], \text{ es } b_k = \frac{1}{n}\sum_{i=1}^n \left(X_i - \bar{X}\right)^k$$

Ejercicio Si $X \sim \text{Gamma}(a, p)$, calcular un estimador por el método de los momentos de $\theta = (a, p)$ basado en una m.a.s. (n)

Métodos de construcción de estimadores

(continuación)

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 $f_\theta(t) = \binom{2}{t}\theta^t(1-\theta)^{2-t}, t = 0, 1, 2$

Métodos de construcción de estimadores

(continuación)

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

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

es el estimador de máxima verosimilitud (EMV)

Métodos de construcción de estimadores

(continuación)Sea $(X_1, \cdots X_n)$ una muestra con $f_\theta(x_1, \cdots, x_n) = f(x_1, \cdots, x_n \mid \theta)$ función de densidad (o de masa), $\theta \in \Theta \subset \mathbb{R}^\ell$ Denotemos por $L(\theta \mid x_1, \cdots, x_n) = f(x_1, \cdots, x_n \mid \theta)$ a la función de verosimilitud de la muestra Un estimador $\hat{\theta} = \hat{\theta}(X_1, \cdots, X_n)$ se denomina estimador de máxima verosimilitud (EMV) de θ, sí y sólo sí
(1) $\hat{\theta}(x_1, \cdots, x_n) \in \Theta, \forall (x_1, \cdots, x_n) \in \chi^n$
(2) $L\left(\hat{\theta} \mid x_1, \cdots, x_n\right) = \sup_{\theta \in \Theta} L(\theta \mid x_1, \cdots, x_n), \forall (x_1, \cdots, x_n) \in \chi^n$
o equivalentemente, sí y sólo sí
(1) $\hat{\theta}(x_1, \cdots, x_n) \in \Theta, \forall (x_1, \cdots, x_n) \in \chi^n$
(2) $\log L\left(\hat{\theta} \mid x_1, \cdots, x_n\right) = \sup_{\theta \in \Theta} \log L(\theta \mid x_1, \cdots, x_n), \forall (x_1, \cdots, x_n) \in \chi^n$

Métodos de construcción de estimadores (continuación)

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:

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

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

Métodos de construcción de estimadores

(continuación)

Observaciones

(1) EI 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 $S = S(X_1, \cdots, X_n)$ 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 $(X_1, \cdots X_n)$ es m.a.s. (n) y $\theta \in \mathbb{R}$, entonces $\sqrt{n}(\hat{\theta} - \theta)\underset{n\to\infty}{d} N\left(0, \frac{1}{\sqrt{l_1(\theta)}}\right)$ y por lo tanto, $\hat{\theta}$ es asintóticamente insesgado para θ y asintóticamente eficiente

3 Integrales de línea: campos escalares y vectoriales

Ejercicio 1 [Ejercicio 2.7.1]

Sea X una observación de una población $N(0, \sigma)$. ¿Es $|X|$ un estadístico suficiente?

Solucion

Para una variable aleatoria X con distribución normal $N(0, \sigma)$, la función de densidad de probabilidad es:

$$f(x|\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{|x|^2}{2\sigma^2}} \tag{4}$$

Podemos entonces aplicar el teorema de factorización (2.2.1) para obtener una factorización de la siguiente forma:

$$f(x|\sigma) = h(x)g(|x|, \sigma) \tag{5}$$

Donde nuestro estadístico es $T(X) = |X|$. Podemos tomar:

$$h(x) = \frac{1}{\sqrt{2\pi}} \tag{6}$$

$$g(|x|, \sigma) = \frac{1}{\sigma} e^{-\frac{|x|^2}{2\sigma^2}} \tag{7}$$

Finalmente, podemos ver que:

$$f(x|\sigma) = h(x)g(|x|, \sigma) \tag{8}$$

Por lo que $|X|$ es un estadístico suficiente para σ.

4 Appendix

4.1 Momentos Notables

Media

Definición 4.1.1 [Media]

Distinguimos entre casos discretos y continuos:

- ***Caso discreto:*** *Se define la media de una variable aleatoria discreta X como:*

$$\mu = E(X) = \sum_{i=1}^{n} x_i \cdot p_i \tag{9}$$

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:*

$$\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx \tag{10}$$

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 4.1.1 [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

$$E[X+Y] = \sum_{x,y}(x+y)P(X=x, Y=y)$$

$$= \sum_{x,y} xP(X=x, Y=y) + \sum_{x,y} yP(X=x, Y=y)$$

$$= \sum_{x} xP(X=x, Y=y) + \sum_{y} y \sum_{x} P(X=x, Y=y)$$

$$= \sum_{x} xP(X=x) + \sum_{y} yP(Y=y)$$

$$= E[X] + E[Y]$$

□

Teorema 4.1.2 [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 $\exists E[X_1 \cdots X_n]$, y se verifica:*

$$E[X_1 \cdots X_n] = E[X_1] \cdots E[X_n] = \prod_{i=1}^{n} E[X_i]$$

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).

$$E[XY] = \sum_{x}\sum_{y} xy\, P(X=x, Y=y)$$

$$= \sum_{x}\sum_{y} xy\, P(X=x)P(Y=y)$$

$$= \sum_{x} x\, P(X=x) \sum_{y} y\, P(Y=y)$$

$$= E[X]E[Y]$$

□

Varianza

Definición 4.1.2 [Varianza]

Distinguimos entre casos discretos y continuos:

- ***Caso discreto:*** *Se define la varianza de una variable aleatoria discreta X como:*

$$\sigma^2 = Var(X) = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot p_i \tag{11}$$

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:*

$$\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)\, dx \tag{12}$$

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:

$$\sigma^2 = \mathrm{Var}(X) = E[(X - E[X])^2] = E[X^2 - 2XE[X] + E[X]^2] \tag{13}$$

$$= E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2 \tag{14}$$

La varianza también puede ser vista como covarianza de una variable consigo misma:

$$\sigma^2 = \mathrm{Var}(X) = \mathrm{Cov}(X, X) \tag{15}$$

Propiedades

Si X y Y son **variables aleatorias** con varianza finita entonces

1. $\mathrm{Var}(X) \geq 0$.
2. $\mathrm{Var}(X) = 0$ si y sólo si X es constante.
3. $\mathrm{Var}(X) = 0 \iff \exists c \in \mathbb{R}$ tal que $P(X = c) = 1$.
4. $\mathrm{Var}(X + a) = \mathrm{Var}(X)$.
5. $\mathrm{Var}(aX) = a^2 \,\mathrm{Var}(X)$.
6. $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$.
7. $\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\,\mathrm{Cov}(X, Y)$.

Teorema 4.1.3 [Identidad de Bienaymé]

En general, para la suma de n variables aleatorias $X_1, X_2, \ldots, X_n$ se cumple que

$$\operatorname{Var}(X_1 + X_2 + \cdots + X_n) = \sum_{i,j=1}^{n} \operatorname{Cov}(X_i, X_j) = \sum_{i=1}^{n} \operatorname{Var}(X_i) + \sum_{i \neq j} \operatorname{Cov}(X_i, X_j) \tag{16}$$

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

$$\operatorname{Var}(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} \operatorname{Var}(X_i) \tag{17}$$

Producto de Variables Aleatorias

- **Variables Aleatorias Independientes** Si dos variables X e Y son **independientes**, la varianza de su producto está dada por:

$$\operatorname{Var}(XY) = [E(X)]^2 \operatorname{Var}(Y) + [E(Y)]^2 \operatorname{Var}(X) + \operatorname{Var}(X) \operatorname{Var}(Y).$$

  De manera equivalente, utilizando las propiedades básicas de la esperanza, se expresa como:

$$\operatorname{Var}(XY) = E(X^2)E(Y^2) - [E(X)]^2[E(Y)]^2.$$

- **Variables Aleatorias Correlacionadas** En general, si dos variables son **estadísticamente dependientes**, entonces la varianza de su producto está dada por:

$$\operatorname{Var}(XY) = E[X^2Y^2] - [E(XY)]^2$$

$$= \operatorname{Cov}(X^2, Y^2) + E(X^2)E(Y^2) - [E(XY)]^2$$

$$= \operatorname{Cov}(X^2, Y^2) + (\operatorname{Var}(X) + [E(X)]^2)(\operatorname{Var}(Y) + [E(Y)]^2)$$

$$-[\operatorname{Cov}(X, Y) + E(X)E(Y)]^2$$

4.2 Función Característica

Definición 4.2.1 [Función Característica]

*La **función característica** de una variable aleatoria X es una función $\varphi_X(t) : \mathbb{R} \to \mathbb{C}$ 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 $X_1, ..., X_n$ son variables aleatorias independientes y $a_1, ..., a_n$ son constantes, entonces la función característica de la combinación lineal de las variables X_i es

  $$\varphi_{a_1X_1+\cdots+a_nX_n}(t) = \varphi_{X_1}(a_1t)\cdots\varphi_{X_n}(a_nt).$$

  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

  $$\varphi_{\mathbf{Y}}(t) = e^{it^T\mathbf{B}}\varphi_{\mathbf{X}}(A^Tt).$$