Estadística

Contents

1 Estimación Puntual Paramétrica **2**
1.1 Estimadores Bayesianos 5
1.2 Criterios de comparación de estimadores 10
1.3 Cota para la varianza de un estimador 17
1.4 Métodos de construcción de estimadores 20
1.5 Propiedades asintóticas de los estimadores de máxima verosimilitud 22

2 Estimación por Regiones de Confianza **25**
2.1 Región de confianza 25
2.2 Intervalos de confianza 25
2.3 Métodos de obtención de intervalos de confianza 25

3 Contraste de Hipótesis **29**
3.1 Principios básicos de un contraste de hipótesis 29
3.2 Errores de tipo I y de tipo II 30

4 Appendix **31**
4.1 Momentos Notables 31
4.2 Función Característica 34

1 Estimación Puntual Paramétrica

Definición 1.0.1 [Estimador]

Sea $(\Omega, \mathcal{A}, \mathcal{P})$ el espacio probabilístico asociado a un experimento aleatorio $\mathcal{E}$, una variable aleatoria observable $X : \Omega \longrightarrow \mathbb{R}$ 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 $T = T(X_1, \cdots X_n) : \chi^n \longrightarrow \Theta$ que se utiliza para determinar el valor desconocido θ

Definición 1.0.2 [Estimador centrado o insesgado]

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

Ejemplo

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

$$E_\theta\left[\overline{X}\right] = E[X]$$

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

Definición 1.0.3 [Sesgo]

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

Ejemplo

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

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

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

Ejemplo

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

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

$$= \frac{1}{n}\sum_{i=1}^{n} E[X_i^2 - 2X_i\bar{X} + \bar{X}^2] = \frac{1}{n}\sum_{i=1}^{n} E[X_i^2] - 2E[X_i\bar{X}] + E[\bar{X}^2] =$$

$$= \frac{1}{n}\sum_{i=1}^{n} E[X_i]^2 - \frac{2}{n}\sum_{i=1}^{n} E[X_i\bar{X}] + \frac{1}{n}\sum_{i=1}^{n} E[\bar{X}^2] =$$

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

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

Ahora desarrollemos el término que falta:

$$E[X_i\bar{X}] = E[X_i\frac{1}{n}\sum_{j=1}^{n} X_j] = \frac{1}{n}\sum_{j=1}^{n} E[X_iX_j] = \frac{1}{n}\sum_{j=1 j\neq i}^{n} E[X_iX_j] + \frac{1}{n}E[X_i^2] =$$

$$= \frac{1}{n}\sum_{j=1 j\neq i}^{n} E[X_i]E[X_j] + \frac{1}{n}E[X_i^2] = \frac{1}{n}(n-1)\mu^2 + \frac{1}{n}(\sigma^2 + \mu^2) = \mu^2 + \frac{\sigma^2}{n}$$

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

$$\implies E[\sigma_n^2] = \sigma^2(1 - \frac{1}{n}) = h(\theta) \text{ y } b(\sigma_n^2, \sigma^2) = E[\sigma_n^2] - \sigma^2 = -\frac{\sigma^2}{n}$$

$$\implies E[\sigma_n^2] = \sigma^2\left(1 - \frac{1}{n}\right) = h(\theta) \text{ y } b(\sigma_n^2, \sigma^2) = E[\sigma_n^2] - \sigma^2 = -\frac{\sigma^2}{n}$$

Observación 1.0.1

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

Ejemplo

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

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

Ejemplo

Sea una m.a.s. de tamaño $n = 1$ de una población que sigue una distribución $Bin(1, \theta)$ demostrar que no existe un estimador centrado de θ^2:
Sea $X \sim Bin(1, \theta)$ y $T(X)$ un posible estimador de θ^2. Para que $T(X)$ sea centrado (insesgado), debe satisfacer:

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

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

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

Calculamos la esperanza:

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

Para que sea insesgado:

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

Reordenando términos:

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

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

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

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

Ejemplo

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

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

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

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

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

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

Ejemplo

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

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

Definición 1.0.4 [Estimadores consistentes]

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

$$E_\theta[T_n] \underset{n\to\infty}{\longrightarrow} \theta \quad y \quad V_\theta(T_n) \underset{n\to\infty}{\longrightarrow} 0.$$

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

Proposición 1.0.1

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 $T_n \xrightarrow[n\to\infty]{\text{m.c.}} \theta, \forall\theta\in\Theta$ □

Ejemplo

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

Ejemplo

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

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

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

1.1 Estimadores Bayesianos

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

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

donde m es la distribución predictiva, dada por

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

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

Observación 1.1.1

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

Ejemplo

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

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

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

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

Buscamos encontrar la distribución a posteriori de θ usando inferencia Bayesiana.
Antes de observar los datos, suponemos que θ sigue una distribución uniforme en el intervalo $(0, 1)$:

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

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

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

esto implica que cualquier valor de θ dentro del intervalo $[0, 1]$ es igualmente probable antes de ver los datos.
Dado que la muestra es independiente, la función de verosimilitud se obtiene como el producto de las probabilidades individuales:

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

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

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

tenemos:

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

Factorizando los exponentes:

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

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

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

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

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

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

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

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

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

identificamos que la posteriori es una distribución Beta con parámetros:

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

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

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

cuya función de densidad es:

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

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

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

Esto muestra que **la familia de distribuciones Beta es conjugada para la distribución Bernoulli-Binomial**.
El valor esperado de θ dado los datos es:

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

Si tomamos la esperanza como una combinación convexa:

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

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

Ejemplo

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

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

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

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

$$\pi(\theta|x_1,\ldots,x_n) = \frac{\frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} \cdot \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}}{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} \cdot \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}} d\mu} =$$

$$= \frac{e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} \cdot e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}}{\int_{\mathbb{R}} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} \cdot e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}} d\mu} = \frac{e^{\frac{-\mu^2-\mu_0^2+2\mu\mu_0}{2\sigma_0^2}} \cdot e^{\frac{-\sum_{i=1}^{n} x_i^2 - n\mu^2 + 2\mu\sum_{i=1}^{n} x_i}{2\sigma^2}}}{\int_{\mathbb{R}} e^{\frac{-\mu^2-\mu_0^2+2\mu\mu_0}{2\sigma_0^2}} \cdot e^{\frac{-\sum_{i=1}^{n} x_i^2 - n\mu^2 + 2\mu\sum_{i=1}^{n} x_i}{2\sigma^2}} d\mu} =$$

$$= \frac{e^{\frac{-\mu^2+2\mu\mu_0}{2\sigma_0^2}} \cdot e^{\frac{-n\mu^2+2\mu\bar{X}n}{2\sigma^2}}}{\int_{\mathbb{R}} e^{\frac{-\mu^2+2\mu\mu_0}{2\sigma_0^2}} \cdot e^{\frac{-n\mu^2+2\mu\bar{X}n}{2\sigma^2}} d\mu} = \frac{e^{-\frac{1}{2}\left(\frac{1}{\sigma_0^2}+\frac{1}{\frac{\sigma^2}{n}}\right)\mu^2+\left(\frac{\mu_0}{\sigma_0^2}+\frac{\bar{X}}{\frac{\sigma^2}{n}}\right)\mu}}{\int_{\mathbb{R}} e^{-\frac{1}{2}\left(\frac{1}{\sigma_0^2}+\frac{1}{\frac{\sigma^2}{n}}\right)\mu^2+\left(\frac{\mu_0}{\sigma_0^2}+\frac{\bar{X}}{\frac{\sigma^2}{n}}\right)\mu} d} = \frac{e^{-\frac{1}{2V}\mu^2+\frac{E}{V}\mu}}{\int_{\mathbb{R}} e^{-\frac{1}{2V}\mu^2+\frac{E}{V}\mu} d\mu}$$

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

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

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

entonce factorizando el exponente obtenemos:

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

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

Luego entonces tenemos que

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

Definición 1.1.1 [Estadístico suficiente bayesiano]

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

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

Teorema 1.1.1 [Versión bayesiana del Teorema de Factorización de Fisher]

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

Demostración.

$$\Rightarrow \pi(\theta|x_1, \ldots, x_n) = \frac{\pi(\theta)f(x_1, \ldots, x_n|\theta)}{m(x_1, \ldots, x_n)} = \frac{\pi(\theta)g(t|\theta)f(x_1, \ldots, x_n|t, \theta)}{m(x_1, \ldots, x_n)} = \frac{\pi(\theta)g(t|\theta)}{\int_{\Theta} \pi(\theta)g(t|\theta)d\theta} = \pi(\theta|t)$$

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

□

Definición 1.1.2 [Error cuadrático medio]

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

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

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

Proposición 1.1.1

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

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

Demostración.

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

□

Observación 1.1.2

El sesgo mide qué tanto se desvía, en promedio, el estimador del valor verdadero del parámetro.
La varianza del estimador mide cómo varían las estimaciones (del estimador) si tomamos diferentes muestras.
Es decir, responden a las preguntas de ¿Apunta al lugar correcto? y ¿Qué tan dispersas están las estimaciones? respectivamente.

Ejemplo

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

$$E_\theta[(T-\theta)^2] = E_\theta[(\bar{X}-\theta)^2] = E_\theta[\bar{X}^2 - 2\bar{X}\theta + \theta^2] = E_\theta[\bar{X}^2] - 2\theta E_\theta[\bar{X}] + \theta^2 =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} E_\theta[X_i^2] - 2\theta\frac{1}{n}\sum_{i=1}^{n} E_\theta[X_i] + \theta^2 = \frac{1}{n^2} n(\theta^2 + \theta(1-\theta)) - 2\theta^2 + \theta^2 = \frac{\theta(1-\theta)}{n} = \frac{\theta}{n} - \frac{\theta^2}{n} = \frac{\theta(1-\theta)}{n}$$

Ejemplo

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

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

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

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

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

Observación 1.1.3

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

1.2 Criterios de comparación de estimadores

Definición 1.2.1 [Pérdida final esperada]

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

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

Proposición 1.2.1

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

1. *Si $\mathcal{L}(\theta, t) = (\theta - t)^2$ entonces $PFE(t) = V(\theta|x_1, \ldots, x_n) + b(T, \theta)^2$ y la pérdida final esperada se minimimia en $t^* = E[\theta|x_1, \ldots, x_n]$*
2. *Si $\mathcal{L}(\theta, t) = |\theta - t|$ entonces $PFE(t) = E[|\theta - t||x_1, \ldots, x_n]$ y la pérdida final esperada se minimiza en la mediana de la distribución final (estimador bayesiano)*

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

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

Si expandimos el cuadrado:

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

Calculemos cada una de las esperanzas por separado:

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

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

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

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

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

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

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

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

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

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

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

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

□

Ejemplo

EJEMPLO DE LA PÉRDIDA FINAL ESPERADA

Definición 1.2.2 [Estimador centrado uniformemente de mínima varianza]

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

Proposición 1.2.2

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

Demostración. Sean T_1 y T_2 dos estimadores centrados uniformemente de mínima varianza para θ, demostremos que entones $T_1 \overset{c.s.}{\equiv} T_2$.
Sea $T = \frac{T_1+T_2}{2} \implies E_\theta[T] = \theta$

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

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

$$\rho_\theta(T_1, T_2) = \frac{Cov(T_1, T_2)}{\sqrt{V_\theta(T_1)V_\theta(T_2)}} \leq 1 \iff Cov(T_1, T_2) \leq V_\theta(T_1) \implies$$

$$\implies V_\theta(T) \leq V_\theta(T_1)$$

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

Ejemplo

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

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

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

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

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

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

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

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

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

Teorema 1.2.1

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

Ejemplo

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

Observación 1.2.1

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

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

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

Teorema 1.2.2 [Teorema de caracterización del ECUMV]

Sea $T_1 = T_1(X_1, \dots, X_n)$ un estimador centrado para θ ($E_\theta[T_1] = \theta$) y $V_\theta(T_1) < \infty$ entonces T_1 es el ECUMV para θ si y sólo si para cualquier otro estimador $T_2 = T_2(X_1, \dots, X_n)$ con $E_\theta[T_2] = 0$ y $V_\theta(T_2) < \infty$ se tiene que $E_\theta[T_1T_2] = 0$

Corolario 1.2.1

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

Teorema 1.2.3 [Teorema de Rao-Blackwell]

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

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

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

Ahora, considerando la varianza de T:

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

Expandiendo el cuadrado y usando la linealidad de la esperanza:

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

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

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

Descomponiendo en términos de la distribución condicional:

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

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

Por lo tanto,

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

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

Teorema 1.2.4 [Teorema de Lehmann-Schefeé]

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

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

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

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

Ejemplo

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

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

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

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

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

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

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

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

ya que los coeficientes binomiales son no nulos.
Por último queda ver que S es un estadístico centrado para θ, i.e. $E_\theta[h(S)] = \theta$ y $V_\theta(h(S)) < \infty$:

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

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

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

Ejemplo

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

Ejemplo

Sea una m.a.s. de tamaño n de una población que sigue una distribución $Poisson(\theta)$, tenemos que encontrar el ECUMV para θ:
Si tomamos el estadístico $S = \sum_{i=1}^n X_i$, para poder aplicar el Teorema de Lehmann-Scheffé necesitamos ver que el estadístico sea completo y suficiente:

Veamos primero que S es suficiente para θ:

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

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

Siguiendo con lo anterior:

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

Entonces, debemos ver que $ln(\theta)$ contiene un abierto $(0, +\infty) \subset \mathbb{R}$, por lo tanto s es completo para θ. Además, tomando el estadístico $T = \frac{1}{n}\sum_{i=1}^n x_i$ tenemos que:

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

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

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

Por el Teorema de Lehmann-Scheffé, $T = h(S)$ es el ECUMV para θ.
Si en lugar de haber tomado $d(\theta) = \theta$ hubieramos querido el estimador centrado uniformemente de mínima varianza para $d(\theta) = e^{-\theta}$ HAY QUE INSERTAR LO DE DIEGO

Ejemplo

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

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

Por el Teorema de Factorización de Fisher, $S = \sum_{i=1}^n X_i$ es suficiente para θ. Veamos ahora su completitud:
Como se sigue una distribución exponencial, podemos ver que pertenece a una familia exponencial uniparamétrica:

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

Por lo que evidentemente en la imagen de la función $q_1(\theta) = -\theta$ tiene un abierto en su imagen por lo que el estadístico S es completo para θ.
Por último, sabemos que $\bar{X} = \frac{S}{n} \implies E[\bar{X}] = \frac{1}{\theta} \implies \bar{X}$ es centrado para $\frac{1}{\theta}$ y por el Teorema de Lehmann-Scheffé, $\bar{X}$ es el ECUMV para $\frac{1}{\theta}$.
Pero nosotros lo que queríamos es un estimador centrado para θ. Por lo que puede parecer intuitivo

pensar que el estadístico que podría estar centrado para θ es $\frac{1}{\bar{X}}$:

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

Por lo que S' es el ECUMV para θ

Ejemplo

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

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

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

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

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

1. Primero veamos el caso para $d_1(\theta) = \theta$: Sabemos que $E[S] = \sum_{i=1}^{n} E[X_i] = n\theta \implies \bar{X} = \frac{S}{n} \implies E[\bar{X}] = \theta \implies \bar{X}$ es centrado para θ. Por el Teorema de Lehmann-Scheffé, $\bar{X}$ es el ECUMV para θ

2. Ahora veamos el caso para $d_2(\theta) = \theta(1-\theta) = \theta - \theta^2$: A pesar de que pueda parecer lógico tomar la varianza muestral S^2 como estimador, no se cumplen las condiciones del Teorema de Lehmann-Scheffé para que sea un ECUMV, ya que tiene que depender únicamente de $S = \sum X_i$ y en S^2 aparecen términos cuadráticos, por lo que no es un ECUMV. Podemos intentar buscar otro:
NO LO ENTIENDO

Ejemplo

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

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

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

Por el Teorema de Factorización de Fisher, $S = \sum_{i=1}^{n}(X_i - \mu)^2$ es suficiente para σ^2. Veamos ahora su completitud:
Sabemos que la distribución normal, en este caso, pertenece a la familia exponencial uniparamétrica. Entonces sabemos que los estadísticos naturales asociados a las funciónes paramétricas $q_i(\theta)$ son completos, si éstas ultimas contienen un abierto en su imagen. En este caso, $q_1(\sigma^2) = \frac{1}{2\sigma^2}$ contiene un abierto en su imagen, por lo que S es completo para σ^2.

Veamos la esperanza de S:

$$E[S]=E[\sum_{i=1}^{n}(X_i-\mu)^2]=\sum_{i=1}^{n}E[(X_i^2+\mu^2-2X_i\mu)]=\sum_{i=1}^{n}E[X_i^2]+\sum_{i=1}^{n}\mu^2-\sum_{i=1}^{n}2\mu E[X_i]=$$

$$=n(\sigma^2+\mu^2)+n\mu^2-2n\mu^2=n\sigma^2 \implies \frac{S}{n} \text{ es el ECUMV para } \sigma^2 \text{ por el T. de Lehmann-Schefeé}$$

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

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

$$=f(x_1,\dots,x_n|\mu,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}e^{-\frac{\sum x_i^2+n\mu^2-\mu\sum x_i}{2\sigma^2}}$$

Por el Teorema de Factorización de Fisher, $T(\vec{x})=(\sum x_i,\sum x_i^2)$ es suficiente para (μ,σ^2).
Veamos ahora su completitud:
Sabemos que la distribución normal, en este caso, pertenece a la familia exponencial uniparamétrica. Entonces sabemos que los estadísticos naturales asociados a las funciónes paramétricas $q_i(\theta)$ son completos, si éstas ultimas contienen un abierto en su imagen. En este caso, $q_1(\mu,\sigma^2)=\mu$ y $q_2(\mu,\sigma^2)=\frac{1}{2\sigma^2}$ contienen un abierto en su imagen, por lo que T es completo para (μ,σ^2).
Veamos la esperanza de T:

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

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

3. Ahora suponamos que μ es desconocido y σ^2 es conocido, por lo que queremos encontrar el ECUMV para μ: Por lo visto en el apartado anterior, podemos que por el Teorema de Fisher, podemos ver que el estadítico $T=\sum_{i=1}^{n}X_i$ es suficiente para μ.
También podríamos preguntarnos porqué no sería el estadístico $\vec{T}=(\sum X_i,\sum X_i^2)$, la respuesta está en que esta distribución normal pertenece a la familia exponencial uniparamétrica y por tanto tomamos la función dependiente del parámetro $q_1(\mu)=\mu$ y su estadístico natural asociado $T_1(\vec{x})=\sum X_i$, el cual por las propiedades que vimos de las familias exponenciales es suficiente, y además completo. Veamos ahora si es insesgado:

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

1.3 Cota para la varianza de un estimador

Hemos visto hasta ahora el estimador centrado uniformemente de mínima varianza, que cómo su nombre indica, es el que tiene varianza más pequeña.
Si definieramos una varianza mínima para los estimadores de un parámetro, es decir, si encontrásemos una cota para los estimadores, podríamos encontrar más fácilmente un estimador que tenga dicha varianza y por tanto seria el centrado uniformemente de mínima varianza.

Definición 1.3.1 [Condiciones de regularidad de Wolfowitz]

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

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

Teorema 1.3.1 [Cota de Fréchet-Cramér-Rao]

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

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

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

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

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

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

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

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

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

Entonces, $\mathrm{Cov}_\theta\left[T, \frac{\partial}{\partial\theta} \log f_\theta\right] = E\left[T \frac{\partial}{\partial\theta} \log f_\theta\right] = d'(\theta)$, 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 $T \overset{\text{c.s.}}{=} a + b\frac{\partial}{\partial\theta} \log f_\theta$, es decir, si y sólo si existe una función $k(\theta)$ tal que $P_\theta\left(T = d(\theta) + k(\theta)\frac{\partial}{\partial\theta} f_\theta\right) = 1$ En efecto, si $T \overset{\text{c.s.}}{=} a + b\frac{\partial}{\partial\theta} \log f_\theta$, entonces $d(\theta) = E_\theta[T] = ay$

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

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

$\square$

Proposición 1.3.1

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

Proposición 1.3.2

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: $f_\theta\left(x_1, \cdots x_n\right) = \prod_{i=1}^n f_\theta\left(x_i\right)$

Proposición 1.3.3

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 $\frac{1}{n}\sum_{i=1}^n T_1\left(X_i\right)$ alcanza la cota de FCR para $d(\theta) = -\frac{c'(\theta)}{c(\theta)q_1'(\theta)}$

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

$\square$

Ejemplo

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

Ejemplo

Si se cumplen las condiciones de regularidad y además
(1) $\forall\left(x_1, \cdots, x_n\right) \in \chi^n y \forall\theta \in \Theta, \exists\frac{\partial^2}{\partial\theta^2} f_\theta\left(x_1, \cdots, x_n\right)$
(2) $\int_{\chi^n}\frac{\partial^2}{\partial\theta^2} f_\theta\left(x_1, \cdots, x_n\right) dx_1 \cdots dx_n = 0$
Entonces, $I_n(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta\left(x_1, \cdots, x_n\right)\right]$
Indicación: $\frac{\partial}{\partial\theta}\log f_\theta\left(x_1, \cdots, x_n\right) = \frac{\partial}{\partial\theta} f_\theta\left(x_1, \cdots, x_n\right)\frac{1}{f_\theta\left(x_1, \cdots, x_n\right)}$

Definición 1.3.2 [Estimador Eficiente]

Diremos que un estimador es eficiente para $d(\theta)$ si es centrado para $d(\theta)$ y su varianza alcanza la cota de FCR
En general, se llama eficiencia de un estimador centrado de $d(\theta)$ a

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

1.4 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 ,

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

Ejemplo

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)

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

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

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

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

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

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

es el estimador de máxima verosimilitud (EMV)

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) $\hat{\theta}\left(x_1, \cdots, x_n\right) \in \Theta, \forall\left(x_1, \cdots, x_n\right) \in \chi^n$
(2) $\log L\left(\hat{\theta} \mid x_1, \cdots, x_n\right)=\sup_{\theta \in \Theta} \log L\left(\theta \mid x_1, \cdots, x_n\right), \forall\left(x_1, \cdots, x_n\right) \in \chi^n$

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

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

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

Observación 1.4.1

1. *El EMV $\hat{\theta}$ no tiene por qué existir*
2. *El EMV $\hat{\theta}$ no tiene por qué ser único*
3. *El EMV $\hat{\theta}$ no tiene por qué ser centrado*
4. *El EMV $\hat{\theta}$ no tiene por qué ser suficiente, pero si $S=S\left(X_1, \cdots, X_n\right)$ es suficiente para θ, entonces $\hat{\theta}=\hat{\theta}(S)$*
5. *Invariancia: Si $\hat{\theta}$ es el EMV de θ, entonces $h(\hat{\theta} s)$ es el EMV de $h(\theta)$*
6. *Bajo ciertas condiciones de regularidad, si $\left(X_1, \cdots X_n\right)$ es m.a.s. (n) y $\theta \in \mathbb{R}$, entonces $\sqrt{n}(\hat{\theta}-\theta) \underset{n \rightarrow \infty}{d} N\left(0, \frac{1}{\sqrt{l_1(\theta)}}\right)$ y por lo tanto, $\hat{\theta}$ es asintóticamente insesgado para θ y asintóticamente eficiente*

Ejemplo

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

Ejemplo

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

Ejemplo

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

Ejemplo

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

Propiedades del EMV

Proposición 1.4.1

Si $S=S\left(X_{1}, \cdots, X_{n}\right)$ es suficiente para θ, entonces el EMV $\hat{\theta}=\hat{\theta}(S)$ es función de S

Demostración. $f_{\theta}\left(x_{1}, \ldots, x_{n}\right)=h\left(x_{1}, \ldots, x_{n}\right) g_{\theta}\left(S\left(x_{1}, \ldots, x_{n}\right)\right)$ (teorema de factorización) □

Proposición 1.4.2

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

$$M\left(\lambda \mid x_{1}, \ldots, x_{n}\right)=\operatorname{Sup}_{\theta \in \Theta_{\lambda}} L\left(\theta \mid x_{1}, \ldots, x_{n}\right)$$
$$\Theta_{\lambda}=\{\theta \in \Theta: h(\theta)=\lambda\}$$

Demostración.

$$\begin{aligned}&\operatorname{Sup}_{\lambda \in \Lambda} M\left(\lambda \mid x_{1}, \ldots, x_{n}\right)=\operatorname{Sup}_{\lambda \in \Lambda} \operatorname{Sup}_{\theta \in \Theta_{\lambda}} L\left(\theta \mid x_{1}, \ldots, x_{n}\right)=\operatorname{Sup}_{\theta \in \Theta} L\left(\theta \mid x_{1}, \ldots, x_{n}\right) \\ &=L\left(\hat{\theta} \mid x_{1}, \ldots, x_{n}\right)\end{aligned}$$

Finalmente, como $\hat{\theta} \in \Theta_{\hat{\lambda}}$, se tiene que
$M\left(\hat{\lambda} \mid x_{1}, \ldots, x_{n}\right)=\operatorname{Sup}_{\theta \in \Theta_{\hat{\lambda}}} L\left(\theta \mid x_{1}, \ldots, x_{n}\right) \geq L\left(\hat{\theta} \mid x_{1}, \ldots, x_{n}\right)$, y por lo tanto $\operatorname{Sup}_{\lambda \in \Lambda} M\left(\lambda \mid x_{1}, \ldots, x_{n}\right)=M\left(\hat{\lambda} \mid x_{1}, \ldots, x_{n}\right)$

□

1.5 Propiedades asintóticas de los estimadores de máxima verosimilitud

Proposición 1.5.1 [Desigualdad de Yelsin]

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

Consideremos $X \approx\left(\chi, \beta_{\chi}, F_{\theta}\right)_{\theta \in \Theta \subset \mathbb{R}}$ modelo estadístico uniparamétrico contínuo (o discreto) y sea $\left(X_{1}, \cdots X_{n}\right)$ muestra de $\left\{F_{\theta}, \theta \in \Theta\right\}$ siendo $f_{\theta}\left(x_{1}, \cdots, x_{n}\right)$ su función de densidad (o de masa). Supongamos que se verifican las condiciones de regularidad

Denotemos por θ_{0} el verdadero valor del parámetro (desconocido).

Lema 1.5.1

(1) $E_{\theta_{0}}\left[\log f_{\theta}\left(x_{1}, \ldots, x_{n}\right)\right]<E_{\theta_{0}}\left[\log f_{\theta_{0}}\left(x_{1}, \ldots, x_{n}\right)\right], \forall \theta \neq \theta_{0}$
(2) Existe un entorno de $\theta_{0}, E\left(\theta_{0}\right)$, donde la función de θ $E_{\theta_{0}}\left[\frac{\partial}{\partial \theta} \log f_{\theta}\left(x_{1}, \ldots, x_{n}\right)\right]=J\left(\theta_{0}, \theta\right)$ es estrictamente decreciente

Demostración. 1. Como el logaritmo es una función estrictamente cóncava y el soporte de la distribución es independiente del parámetro, si $\theta \neq \theta_{0}, E_{\theta_{0}}\left[\log \frac{f_{\theta}\left(x_{1}, \ldots, x_{n}\right)}{f_{\theta_{0}}\left(x_{1}, \ldots, x_{n}\right)}\right]<\log E_{\theta_{0}}\left[\frac{f_{\theta}\left(x_{1}, \ldots, x_{n}\right)}{f_{\theta_{0}}\left(x_{1}, \ldots, x_{n}\right]}\right]=0$. Por lo tanto,
$E_{\theta_{0}}\left[\log f_{\theta}\left(x_{1}, \ldots, x_{n}\right)\right]-E_{\theta_{0}}\left[\log f_{\theta_{0}}\left(x_{1}, \ldots, x_{n}\right)\right]<0, y$
$E_{\theta_{0}}\left[\log f_{\theta}\left(x_{1}, \ldots, x_{n}\right)\right]$ presenta un máximo extricto en θ_{0}

2. Como $f_\theta(x_1,\dots,x_n)$ es derivable respecto a θ, $\frac{\partial}{\partial\theta}E_{\theta_0}[\log f_\theta(x_1,\dots,x_n)]=0$ en $\theta=\theta_0 \Rightarrow J(\theta_0,\theta_0)=0$ $\frac{\partial^2}{\partial\theta^2}E_{\theta_0}[\log f_\theta(x_1,\dots,x_n)]<0$ en $\theta=\theta_0 \Rightarrow \frac{\partial}{\partial\theta}J(\theta_0,\theta)<0$ en $\theta=\theta_0$ Por lo tanto, existe un entorno de $\theta_0, E(\theta_0)$, donde la función $J(\theta_0,\theta)$ es estrictamente decreciente

$\square$

Teorema 1.5.1 [Primer teorema de convergencia del EMV (consistencia])

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

Demostración. Para una m.a.s. $(n), f_\theta(x_1,\dots,x_n)=\prod_{i=i}^n f_\theta(x_i)$
Entonces, $\frac{\partial}{\partial\theta}\log L(\theta\mid x_1,\dots,x_n)=\sum_{i=1}^n\frac{\partial}{\partial\theta}\log f_\theta(x_i)=nz_n(\theta)$
Además, $z_n(\theta)\xrightarrow[n\to\infty]{\text{c.s.}}E_{\theta_0}\left[\frac{\partial}{\partial\theta}\log f_\theta(x)\right]=J(\theta_0,\theta)$ que es estrictamente decreciente en el entorno $E(\theta_0)$ y $J(\theta_0,\theta_0)=0$
Por lo tanto, podemos elegir $\varepsilon>0$ tal que $[\theta_0-\varepsilon,\theta_0+\varepsilon]\subset E(\theta_0)$ y $J(\theta_0-\varepsilon,\theta_0)>0, J(\theta_0+\varepsilon,\theta_0)<0$
Entonces,
$z_n(\theta_0-\varepsilon)\xrightarrow[n\to\infty]{\text{c.s.}}J(\theta_0,\theta_0-\varepsilon)>0$ y $z_n(\theta_0+\varepsilon)\xrightarrow[n\to\infty]{\text{c.s.}}J(\theta_0,\theta_0+\varepsilon)<0$
Como $\frac{\partial}{\partial\theta}f_\theta$ es continua, $\exists\hat{\theta}_n\in(\theta_0-\varepsilon,\theta_0+\varepsilon)$ tal que $z_n\left(\hat{\theta}_n\right)=0$. Por lo tanto, $\hat{\theta}_n$ es una solución de la ecuación de verosimilitud tal que $\hat{\theta}_n\xrightarrow[n\to\infty]{\text{c.s.}}\theta_0$ $\square$

Observación 1.5.1

El teorema anterior es útil, sólo cuando existe una única raíz de la ecuación de verosimilitud, ya que entonces esa raíz ha de ser el estimador consistente. En caso contrario, no se sabrá, cual de las raíces al variar n, va a dar lugar a la sucesión consistente
Además, bajo ciertas condiciones de regularidad,
$Y_n(\theta)=\frac{1}{n}\sum_{i=1}^n\frac{\partial^2}{\partial\theta^2}\log f_\theta(x_i)\xrightarrow[n\to\infty]{c.s.}E_{\theta_0}\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(x)\right]=I(\theta_0,\theta)$, *con* $I(\theta_0,\theta_0)=-I_1(\theta_0)<0$
Como $\hat{\theta}_n\xrightarrow[n\to\infty]{c.s.}\theta_0$, *si* $\frac{\partial^2}{\partial\theta^2}f_\theta(x)$ *es continua en θ uniformemente $\forall x$, entonces* $Y_n\left(\hat{\theta}_n\right)-Y_n(\theta_0)\xrightarrow[n\to\infty]{c.s.}0$, *y por lo tanto* $Y_n\left(\hat{\theta}_n\right)\xrightarrow[n\to\infty]{c.s.}-I_1(\theta_0)<0$ *y a partir de un n en adelante la sucesión de v.a. solución de la ecuación de verosimilitud son una sucesión de máximos*

Teorema 1.5.2 [Segundo teorema de convergencia del EMV (convergencia en ley])

Si se cumplen las condiciones de regularidad y $\frac{\partial^2}{\partial\theta^2}f_\theta(x)$ es continua en θ uniformemente $\forall x$, entonces

$$\sqrt{n}\left(\hat{\theta}_n-\theta\right)\xrightarrow[n\to\infty]{d}N\left(0,\frac{1}{\sqrt{I_1(\theta)}}\right)$$

Demostración. Si denotamos por $\varphi(x,\theta)=\frac{\partial}{\partial\theta}\log f_\theta(x)$, entonces la ecuación de verosimilitud
$\frac{\partial}{\partial\theta}\log L(\theta\mid x_1,\dots,x_n)=\sum_{i=1}^n\frac{\partial}{\partial\theta}\log f_\theta(x_i)=\sum_{i=1}^n\varphi(x_i,\theta)=nz_n(\theta)=0$ en $\hat{\theta}_n$, con $\hat{\theta}_n\xrightarrow[n\to\infty]{\text{c.s.}}\theta_0$, y por lo tanto $\sum_{i=1}^n\varphi\left(x_i,\hat{\theta}_n\right)=0$

$\sum_{i=1}^n\varphi(x_i,\theta_0)=\sum_{i=1}^n\varphi\left(x_i,\hat{\theta}_n\right)+\left(\theta_0-\hat{\theta}_n\right)\sum_{i=1}^n\frac{\partial}{\partial\theta}\varphi(x_i,\theta^*)$ con $\theta^*\in\left(\theta_0,\hat{\theta}_n\right)$
Entonces, $\hat{\theta}_n-\theta_0=\frac{\sum_{i=1}^n\varphi(x_i,\theta_0)}{-\sum_{i=1}^n\frac{\partial}{\partial\theta}\varphi(x_i,\theta^*)}$

$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi(x_i,\theta_0)}{-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\varphi(x_i,\theta^*)} = \frac{A_n}{B_n}$
$E_{\theta_0}\left[A_n\right] = \frac{1}{\sqrt{n}}E_{\theta_0}\left[\sum_{i=1}^{n}\varphi\left(x_i,\theta_0\right)\right] = \frac{1}{\sqrt{n}}nE_{\theta_0}\left[\varphi\left(x,\theta_0\right)\right] = 0$
$V_{\theta_0}\left(A_n\right) = \frac{1}{n}V_{\theta_0}\left(\sum_{i=1}^{n}\varphi\left(x_i,\theta_0\right)\right) = \frac{1}{n}nV_{\theta_0}\left(\varphi\left(x,\theta_0\right)\right)$
$= E_{\theta_0}\left[\varphi\left(x,\theta_0\right)^2\right] = I_1\left(\theta_0\right)$. Por lo tanto, $A_n \xrightarrow[n\to\infty]{d} N\left(0,\sqrt{I_1\left(\theta_0\right)}\right)$
Además, $-B_n - E_{\theta_0}\left[\frac{\partial}{\partial\theta}\varphi\left(x,\theta^*\right)\right] \xrightarrow[n\to\infty]{\text{c.s.}} 0$ y como $\hat{\theta}_n \xrightarrow[n\to\infty]{\text{c.s.}} \theta_0$ y $\frac{\partial^2}{\partial\theta^2}f_\theta(x)$ es continua en θ uniformemente $\forall x$, se tiene que $-B_n+I_1\left(\theta_0\right) \xrightarrow[n\to\infty]{\text{c.s.}} 0$ Por lo tanto, $\frac{A_n}{B_n} \xrightarrow[n\to\infty]{d} \frac{1}{I_1(\theta_0)}N\left(0,\sqrt{I_1\left(\theta_0\right)}\right) \sim N\left(0,\sqrt{\frac{1}{I_1(\theta_0)}}\right)$

□

Observación 1.5.2 [Simplificación de las condiciones de regularidad]

Las condiciones de regularidad del teorema anterior pueden simplificarse a las siguientes:

1. $\phi(x,\theta) = \frac{\partial}{\partial\theta}\log f_\theta(x)$ *tiene dos derivadas continuas respecto de* θ.
2. $\hat{\theta}_n \xrightarrow[n\to\infty]{P} \theta_0$
3. $E_{\theta_0}[\varphi^2(x,\theta_0)] = I_1(\theta_0) < \infty$
4. $E_{\theta_0}\left[\frac{\partial}{\partial\theta}\varphi(x,\theta_0)\right] = -I_1(\theta_0) \neq 0$
5. $\exists\,\varepsilon > 0$ *y* $M(x)$ *tales que* $\sup_{|\theta-\theta_0|\leq\varepsilon}\left|\frac{\partial^2}{\partial\theta^2}\varphi(x,\theta)\right| \leq M(x)$ *con* $M(x) < \infty$

2 Estimación por Regiones de Confianza

2.1 Región de confianza

Sea $X \approx (\chi, \beta_\chi, F_\theta)_{\theta \in \Theta \subset \mathbb{R}^k}$ modelo estadístico k-paramétrico y $(X_1, \cdots X_n)$ muestra de $\{F_\theta, \theta \in \Theta\}$. Denotemos por P_θ a la ley de probabilidad de la muestra

Región de confianza
Sea $C(X_1, \cdots, X_n) \subset \Theta$ una región aleatoria del espacio paramétrico tal que

$$P_\theta \{\theta \in C(X_1, \cdots, X_n)\} \geq 1 - \alpha, \forall \theta \in \Theta$$

Entonces, para cada $(x_1, \cdots, x_n) \in \chi^n, C(x_1, \cdots, x_n)$ se denomina región de confianza para θ de nivel $1 - \alpha$

2.2 Intervalos de confianza

Definición 2.2.1

Sea $h : \Theta \to \mathbb{R}, \alpha \in (0, 1)$ y $T_1 = T_1(X_1, \cdots, X_n) : \chi^n \to \mathbb{R}$ y $T_2 = T_2(X_1, \cdots, X_n) : \chi^n \to \mathbb{R}, T_1 \leq T_2$, dos estadísticos unidimensionales tales que

$$P_\theta \{T_1(X_1, \cdots, X_n) \leq h(\theta) \leq T_2(X_1, \cdots, X_n)\} \geq 1 - \alpha, \forall \theta \in \Theta$$

Entonces, para cada $(x_1, \cdots, x_n) \in \chi^n, (T_1(x_1, \cdots, x_n), T_2(x_1, \cdots, x_n))$ se denomina intervalo de confianza para $h(\theta)$ de nivel $1 - \alpha$

Observación 2.2.1

Siempre es deseable hacer que la medida de la región de confianza sea mínima, entre todas las del mismo grado de confianza $1 - \alpha$

2.3 Métodos de obtención de intervalos de confianza

Método de la cantidad pivotal

Una v.a. $T = T(X_1, \cdots, X_n, \theta) : \chi^n \times \Theta \to \mathbb{R}$ es una cantidad pivotal sí y sólo sí su distribución en el muestreo no depende de θ

Si $T = T(X_1, \cdots, X_n, \theta)$ es una cantidad pivotal, fijado cualquier nivel de confianza $1 - \alpha, \alpha \in (0, 1)$, se pueden determinar dos constantes, $c_1(\alpha)$ y $c_2(\alpha) \in \mathbb{R}$ (que no son únicas), tales que

$$P_\theta \{c_1(\alpha) \leq T(X_1, \cdots, X_n, \theta) \leq c_2(\alpha)\} \geq 1 - \alpha, \forall \theta \in \Theta$$

Si para cada $(x_1, \cdots, x_n) \in \chi^n, c_1(\alpha) \leq T(x_1, \cdots, x_n, \theta) \leq c_2(\alpha) \Leftrightarrow T_1(x_1, \cdots, x_n, \alpha) \leq h(\theta) \leq T_2(x_1, \cdots, x_n, \alpha)$, entonces $(T_1(x_1, \cdots, x_n, \alpha), T_2(x_1, \cdots, x_n, \alpha))$ es un intervalo de confianza para $h(\theta)$ de nivel $1 - \alpha$

Intervalos de confianza asociados a la distribución normal

Ejemplo

Para una m.a.s.(n) de $X \sim N(\mu, \sigma), \sigma$ conocida,

$$IC_{1-\alpha}(\mu) = \left(\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

Ejemplo

Para una m.a.s. (n) de de X, si $\theta \in \mathbb{R}$ y la función de distribución de la población $F_\theta(x)$, como función en x es continua y estrictamente monótona $\forall\theta$, y como función de θ es continua y estrictamente monótona $\forall x$, entonces $T = -2\sum_{i=1}^n \ln F_\theta(X_i) \sim \chi^2_{2n}$ constituye una cantidad pivotal y permite obtener un intervalo de confianza para θ

Ejemplo

Construir un intervalo de confianza para θ por el método de la cantidad pivotal basado en una m.a.s. (n) de $f_\theta(x) = \theta x^{\theta-1} I_{(0,1)}(x)$, con $\theta > 0$

Método de Neyman

Sea $T = T(X_1, \cdots X_n) : \chi^n \to \mathbb{R}$ un estimador de $h(\theta)$ y denotemos por $g_\theta(t)$ a su función de densidad (o de masa)

Fijado un nivel de confianza $1-\alpha, \alpha \in (0,1)$ y $\alpha_1, \alpha_2 \in (0,1)$ tales que $\alpha = \alpha_1 + \alpha_2$, para cada $\theta \in \Theta$ se pueden determinar dos valores $c_1(\theta, \alpha_1)$ y $c_2(\theta, \alpha_2) \in \mathbb{R}$, tales que $P_\theta\{T < c_1(\theta, \alpha_1)\} \leq \alpha_1$ y $P_\theta\{T > c_2(\theta, \alpha_2)\} \leq \alpha_2$. Entonces, para cada $\theta \in \Theta$

$$P_\theta\{c_1(\theta, \alpha_1) \leq T \leq c_2(\theta, \alpha)\} \geq 1 - \alpha_2 - \alpha_1 = 1 - \alpha$$

Si para cada $t \in \mathbb{R}, c_1(\theta, \alpha_1) \leq T(t) \leq c_2(\theta, \alpha) \Leftrightarrow$
$T_1(t, \alpha_1, \alpha_2) \leq h(\theta) \leq T_2(t, \alpha_1, \alpha_2)$, entonces
$(T_1(t, \alpha_1, \alpha_2), T_2(t, \alpha_1, \alpha_2))$ es un intervalo de confianza para $h(\theta)$ de nivel $1-\alpha, \alpha \in (0,1)$

Ejemplo

Construir por el método de Neyman un intervalo de confianza de longitud esperada mínima para θ basado en una m.a.s. (n) de $X \sim U(0, \theta)$, con $\theta > 0$. Indicación: utilizar $T = T(X_1, \cdots, X_n) = X_{(n)}$

Intervalos de confianza para muestras grandes
Si $T_n = T(X_1, \cdots, X_n)$ es un estimador de $h(\theta)$ tal que

$$\frac{T_n - h(\theta)}{\sigma_n(\theta)} \underset{n\to\infty}{d} N(0,1)$$

$$P_\theta\left(-z_{\alpha/2} \leq \frac{T_n - h(\theta)}{\sigma_n(\theta)} \leq z_{\alpha/2}\right) \underset{n\to\infty}{\longrightarrow} 1 - \alpha$$

Por lo tanto, si puede invertirse la desigualdad anterior, despejando $h(\theta)$, se puede obtener un intervalo de confianza para $h(\theta)$, de nivel aproximado $1-\alpha$, cuando el tamaño muestral es suficientemente grande

Observación 2.3.1

Si se cumplen todas las condiciones de regularidad y la ecuación de verosimilitud tiene una única raíz, $\hat{\theta}_n \xrightarrow[n\to\infty]{c.s.} \theta$, puede tomarse $T_n = \hat{\theta}_n$ y $h(\theta) = \theta$, y como

$$\frac{\hat{\theta}_n - \theta}{\sqrt{\frac{1}{nI_1(\theta)}}} \xrightarrow[n\to\infty]{d} N(0,1)$$

entonces $\sigma_n(\theta) = \sqrt{\frac{1}{nI_1(\theta)}}$, *que si es una función continua puede ser aproximada por* $\sigma_n\left(\hat{\theta}_n\right)$, *lo que facilita la inversión*

$$IC_{1-\alpha}(\theta) = \hat{\theta}_n \mp z_{\alpha/2}\sqrt{\frac{1}{nI_1\left(\hat{\theta}_n\right)}}$$

Ejemplo

Comprobar que si $X \sim \text{Bin}(1,\theta)$, entonces

$$IC_{1-\alpha}(\theta) = \bar{x} \mp z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}$$
$$IC_{1-\alpha}(\theta) = \bar{x} \mp z_{\alpha/2}\frac{1}{2\sqrt{n}}$$

son intervalos de confianza para θ basados en el EMV, para muestras grandes
Desigualdad de Tchebychev

$$P(|Y - E[Y]| > k\sqrt{V(Y)}) \leq \frac{1}{k^2}$$

Ejemplo

Comprobar que si $X \sim \text{Bin}(1,\theta)$, entonces

$$IC_{1-\alpha}(\theta) = \bar{x} \mp \frac{1}{\sqrt{\alpha}}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}$$
$$IC_{1-\alpha}(\theta) = \bar{x} \mp \frac{1}{\sqrt{\alpha}}\frac{1}{2\sqrt{n}}$$

son intervalos de confianza para θ basados en la desigualdad de Tchebychev

Observación 2.3.2

Observación Los intervalos que se obtienen mediante el método de la desigualdad de Tchebychev son más amplios que los construídos mediante procedimientos específicos a cada modelo de probabilidad

Región creíble

Dada una familia de distribuciones de probabilidad $\{f(x_1,\ldots,x_n \mid \theta), \theta \in \Theta\}$, si la información inicial sobre θ viene dada por la función de densidad o de masa $\pi(\theta)$, la región $C(x_1,\ldots,x_n) \subset \Theta$ es una región creíble de probabilidad $1-\alpha$ si

$$P(\theta \in C(x_1,\ldots,x_n) \mid x_1,\ldots,x_n) \geq 1-\alpha$$

donde esta probabilidad se calcula mediante la distribución final, es decir

$$\int_{C(x_1,\ldots,x_n)} \pi(\theta \mid x_1,\ldots,x_n)\, d\theta \geq 1-\alpha$$

Ejemplo

Para muestras de tamaño $n=1$ de $X\sim \text{Bin}(1,\theta)$, si $\theta\sim U(0,1)$ y se observa $x=1$, entonces $C_{1-\alpha}(\theta)=\left(\sqrt{\frac{\alpha}{2}},\sqrt{1-\frac{\alpha}{2}}\right)$
En efecto, 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 $\pi\left(\theta\mid x_1,\ldots,x_n\right)\sim \text{Beta}\left(\sum_{i=1}^n x_i+1, n-\sum_{i=1}^n x_i+1\right)$
Por lo tanto, para muestras de tamaño $n=1$
$\pi(\theta\mid x=0)\sim\text{Beta}(1,2), \int_{\ell_1}^{\ell_2}2(1-\theta)d\theta=1-\alpha$
$\pi(\theta\mid x=1)\sim\text{Beta}(2,1), \int_{\ell_1}^{\ell_2}2\theta d\theta=1-\alpha$
Por ejemplo, si $x=1, \ell_1^2=\frac{\alpha}{2}, \ell_2^2=1-\frac{\alpha}{2}$
$C_{1-\alpha}(\theta)=(0.223,0.974)$, para $\alpha=0.1$
Otra posibilidad es la región creíble de más alta probabilidad final o amplitud mínima, $C_{1-\alpha}(\theta)=(\sqrt{\alpha},1)=(0.316,1)$

Ejemplo

Para muestras de tamaño $n=10$ de $X\sim\text{Bin}(1,\theta)$, si $\theta\sim U(0,1)$ y se observa $\sum_{i=1}^n x_i=3$, entonces la distribución final es $\text{Beta}(4,8)$ y $C_{1-\alpha}(\theta)=(0.135,0.564)$, para $\alpha=0.1$
En este caso la región creíble de más alta probabilidad final o amplitud mínima es $C_{1-\alpha}(\theta)=(0.117,0.542)$, para $\alpha=0.1$

Ejemplo

Para una m.a.s. (n) de $X\sim N(\theta,1)$, si $\theta\sim N(0,1)$ y se observa $\bar{x}$, entonces $C_{1-\alpha}(\theta)=\frac{n}{n+1}\bar{x}\mp z_{\alpha/2}\frac{1}{\sqrt{n+1}}$
En efecto, Si $(X_1,\ldots,X_n)$ es una m.a.s. (n) de $X\sim N(\mu,\sigma)$ con σ conocida y $\mu\sim N\left(\mu_0,\sigma_0\right)$, entonces $\pi\left(\mu\mid x_1,\ldots,x_n\right)\sim N\left(\mu_1,\sigma_1\right)$,

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

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

Observación 2.3.3

Recordemos que el intervalo de confianza obtenido desde el punto de vista frecuentista es $/C_{1-\alpha}(\theta)=\bar{x}\mp z_{\alpha/2}\frac{1}{\sqrt{n}}$, *que tiene mayor amplitud, aparte de su diferente interpretación*

3 Contraste de Hipótesis

3.1 Principios básicos de un contraste de hipótesis

Sea $X \approx (\chi, \beta_\chi, F_\theta)_{\theta \in \Theta \subset \mathbb{R}^\ell}$ modelo estadístico ℓ-paramétrico y $(X_1, \cdots X_n)$ muestra de $\{F_\theta, \theta \in \Theta\}$

Idea: estudiar si una determinada afirmación sobre $\{F_\theta, \theta \in \Theta\}$ es confirmada o invalidada a partir de los datos muestrales

Ejemplo

Supongamos que en un laboratorio se está estudiando cierta reacción química sobre una determinada sustancia y que el resultado de dicha reacción es una variable observable que se puede modelizar mediante una v.a. X con distribución normal. Por experiencias anteriores se sabe que, si en la sustancia está presente cierto mineral, $X \sim N(\mu = 10, \sigma = 4)$ y si no lo está $X \sim N(\mu = 11, \sigma = 4)$. Se puede comprobar por medio de unos análisis si el mineral está o no presente en la sustancia en estudio, pero dichos análisis son muy costosos, por lo que se procede a realizar la reacción química $n = 25$ veces para decidir, a la luz de los resultados, si $\mu = 10$ o $\mu = 11$

Definición 3.1.1 [Hiptótesis Estadística]

Una hipótesis estadística es cualquier afirmación acerca de un modelo estadístico.

Definición 3.1.2 [Hipótesis Estadística Simple y Compuesta]

Una hipótesis estadística es simple si especifica totalmente el modelo estadístico, en otro caso, se dice que es compuesta

Definición 3.1.3 [Hipótesis Estadística Nula y Alternativa]

Se llama hipótesis nula H_0 a la hipótesis de trabajo y es la hipótesis estadística que vamos a aceptar si no hay suficiente evidencia a partir de los datos para rechazarla, consecuentemente se llama hipótesis alternativa H_1 a la hipótesis estadística que se acepta si hay suficiente evidencia a partir de los datos para rechazar H_0

Definición 3.1.4 [Contraste de Hipótesis Paramétrico]

Un contraste de hipótesis paramétrico es una partición del espacio paramétrico Θ en dos subconjuntos Θ_0 y Θ_1 tales que $\Theta = \Theta_0 \bigcup \Theta_1$ y $\Theta_0 \bigcap \Theta_1 = \phi$

En el ejemplo anterior $\Theta = \{\mu_0, \mu_1\}, \Theta_0 = \{\mu_0\}, \Theta_1 = \{\mu_1\}$

En un problema de contraste de hipótesis paramétrico se pretende contrastar $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$. Por supuesto, la decisión debe basarse en la evidencia aportada por la observación de una muestra $(X_1, \cdots X_n)$, o equivalentemente por la observación de un cierto estadístico $T = T(X_1, \cdots, X_n)$, que se denomina estadístico del contraste, y que será usualmente un estimador suficiente del parámetro θ

El contraste entre dos hipótesis basado en un estadístico, exige conocer la distribución en el muestreo de dicho estadístico, para los diversos valores del parámetro. De hecho, la idea del contraste consiste en localizar un suceso que sea muy improbable cuando la hipótesis nula es cierta. Si, una vez observada la muestra, acontece dicho suceso, o bien es que el azar ha jugado la mala pasada de elegir una muestra "muy rara" o, como parece más razonable, la hipótesis nula era falsa

Definición 3.1.5 [Región Crítica]

Sea una partición del espacio muestral χ^n en dos subconjuntos C y C^* tales que $\chi^n = C\bigcup C^*$ y $C\bigcap C^* = \phi.C$ es una región crítica para el contraste $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$ sí y sólo sí, se rechaza H_0 cuando se observa un valor muestral $(x_1, \cdots, x_n) \in C$, en cuyo caso se acepta H_1. Consecuentemente, C^* se denomina región de aceptación y si $(x_1, \cdots, x_n) \in C^*$, se dice que no hay suficiente evidencia estadística para rechazar H_0, en este sentido se acepta H_0

Si $T = T(X_1, \cdots, X_n) : \chi^n \to \tau$ es el estadístico del contraste, sea una partición de τ en dos subconjuntos C_τ y C_τ^* tales que $\tau = C_\tau \bigcup C_\tau^*$ y $C_\tau \bigcap C_\tau^* = \phi.C_\tau$ es una región crítica para el contraste $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$ sí y sólo sí, se rechaza H_0 cuando se observa un valor muestral $(x_1, \cdots, x_n)$ tal que $T(x_1, \cdots, x_n) \in C_\tau$, en cuyo caso acepto H_1. Consecuentemente, C_τ^* se denomina región de aceptación
Así, $C = \{(x_1, \cdots, x_n) \in \chi^n : T(x_1, \cdots, x_n) \in C_\tau\}$
En el ejemplo anterior, podemos considerar como región crítica $C = \{(x_1, \cdots, x_n) : \bar{x} \geq k\}$

En un problema de contraste no sólo es importante conocer las probabilidades de cada resultado posible, sino también valorar el riesgo que estamos dispuestos a correr al tomar una decisión equivocada. En el ejemplo anterior, al rechazar que en la sustancia está presente el mineral cuando en realidad si lo está, o bien al aceptar que en la sustancia está presente el mineral cuando en realidad no lo está. Ambos errores tienen consecuencias prácticas distintas

3.2 Errores de tipo I y de tipo II

Error de tipo I es el error que se comete cuando se rechaza H_0 siendo cierta. Error de tipo II es el error que se comete cuando se acepta H_1 siendo falsa

En el ejemplo anterior, las probabilidades de cometer error de tipo I y error de tipo II son $P(\mathrm{I}) = P(\bar{x} \geq k \mid \mu = 10)$ y $P(\mathrm{II}) = P(\bar{x} < k \mid \mu = 11)$

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:*

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

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:*

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

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 $X_1,\ldots,X_n$ **variables aleatorias independientes**, tales que $\exists E[X_i]$ $\forall i = 1\ldots n$. 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:*

$$\sigma^2 = Var(X) = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot p_i \tag{3}$$

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:*

$$\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)\, dx \tag{4}$$

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:

$$\sigma^2 = \mathrm{Var}(X) = E[(X - E[X])^2] = E[X^2 - 2XE[X] + E[X]^2] \tag{5}$$

$$= E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2 \tag{6}$$

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

$$\sigma^2 = \mathrm{Var}(X) = \mathrm{Cov}(X, X) \tag{7}$$

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

$$\text{Var}(X_1 + X_2 + \cdots + X_n) = \sum_{i,j=1}^{n} \text{Cov}(X_i, X_j) = \sum_{i=1}^{n} \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j) \tag{8}$$

Si las variables aleatorias son independientes, entonces la covarianza entre ellas es nula, y la varianza de la suma de las variables aleatorias es la suma de las varianzas de las variables aleatorias, es decir, se cumple que

$$\text{Var}(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} \text{Var}(X_i) \tag{9}$$

Producto de Variables Aleatorias

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

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

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

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

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

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

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

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

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

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

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