Método de los momentos

Sea $X \sim f(x, \theta)$ $x \in \mathbb{R}^k$ $\theta \in \Theta \subset \mathbb{R}^m$

$X_1, \ldots, X_n$ i.i.d. X.

(Observemos que $X_i = (X_{i1}, \ldots, X_{ik})$ en el caso general, k puede ser estrictamente mayor que 1)

Sean

$$u_{jp} = \frac{1}{n} \sum_{i=1}^{n} X_{ij}^{p}$$ es decir u_{jp} es el

momento muestral de orden p de la componente j de la variable vectorial X. Si $k=1$, escribiremos simplemente u_p como el momento muestral de orden p de X.

Sean, por otra parte,

$$g_{jp}(\theta) = E(X_j^p)$$ es decir $g_{jp}(\theta)$ es el

momento poblacional de orden p de la variable X_j (la componente j-ésima de X). (Suponemos la existencia de estos momentos)

Se formará, a continuación, un sistema de ecuaciones igualando los momentos muestrales a los poblacionales

$$u_{jp} = g_{jp}(\theta)$$ donde $\theta = (\theta_1, \ldots, \theta_m)$ son las "incógnitas".

escogiendo los valores de j y p de forma conveniente hasta lograr que dicho sistema sea determinado.

La solución de dicho sistema, $\hat{\theta}$, dependerá de las u_{jp} (y por tanto de la muestra) y definirá un estimador de θ.

A dicho método lo llamaremos método de los momentos.

Bajo condiciones muy generales de existencia de momentos poblacionales y continuidad de las $g_{jp}(\theta)$, puede demostrarse que dicho método proporciona estimadores consistentes (debido a las leyes de los grandes números)

NOTA: Los momentos, tanto muestrales como poblacionales, pueden sustituirse por sus correspondientes versiones centradas (varianzas o covarianzas, etc...) obteniendo el mismo resultado.

Veamos algunos ejemplos:

Ejemplo 1

Sea $X \sim \text{Uniforme}(\alpha, \beta)$ $\alpha < \beta$.

Sea $X_1 \dots X_n$ i.i.d X. Hallar un estimador de $\theta = (\alpha, \beta)$ basado en el método de los momentos.

Tengamos en cuenta que $E(X) = \dfrac{\alpha+\beta}{2}$ y $var(X) = \dfrac{(\beta-\alpha)^2}{12}$

El momento de segundo orden centrado (varianza) tiene una expresión más simple que el correspondiente momento de segundo orden sin centrar, por tanto usaremos la versión centrada. Los momentos muestrales correspondientes son

$$u_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n \quad \text{y} \quad u_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = S_n^2$$

(media y varianza muestral), por tanto formaremos el sistema:

$$\left.\begin{aligned} \bar{X}_n &= \frac{\alpha+\beta}{2} \\ S_n^2 &= \frac{(\beta-\alpha)^2}{12} \end{aligned}\right\}$$ donde hemos igualado los momentos muestrales a los poblacionales.

Sistema equivalente a:

$$\left.\begin{aligned} \alpha+\beta &= 2\bar{X}_n \\ \beta-\alpha &= \sqrt{12\, S_n^2} \end{aligned}\right\}$$ (puesto que $\beta > \alpha$)

Sumando y restando ambas ecuaciones, resulta:

$$\begin{aligned} 2\beta &= 2\bar{X}_n + \sqrt{12 S_n^2} \\ 2\alpha &= 2\bar{X}_n - \sqrt{12 S_n^2} \end{aligned} \Rightarrow \begin{aligned} \hat{\beta} &= \bar{X}_n + \sqrt{3}\sqrt{S_n^2} \\ \hat{\alpha} &= \bar{X}_n - \sqrt{3}\sqrt{S_n^2} \end{aligned} \qquad \hat{\theta} = (\hat{\alpha}, \hat{\beta})$$

Ejemplo 2

Sea $X \sim \text{Gamma}(\alpha, p)$ $\alpha>0$, $p>0$

$X_1 \dots X_n$ i.i.d X hallar un estimador de $\theta=(\alpha,p)$ basado en el método de los momentos.

$$E(X^k)=\int_0^\infty x^k \frac{\alpha^p}{\Gamma(p)} x^{p-1} e^{-\alpha x}\,dx = \frac{1}{\alpha^k\Gamma(p)}\int_0^\infty t^{p+k-1} e^{-t}\,dt$$

$$\alpha x = t$$
$$\alpha\,dx = dt$$

$$= \frac{1}{\alpha^k}\frac{\Gamma(p+k)}{\Gamma(p)}$$

Así, para $k=1$ resultá:

$$E(X)=\frac{1}{\alpha}\frac{\Gamma(p+1)}{\Gamma(p)} = \frac{1}{\alpha}\frac{p\,\Gamma(p)}{\Gamma(p)} = \frac{p}{\alpha}$$

y para $k=2$ resultará:

$$E(X^2)=\frac{1}{\alpha^2}\frac{\Gamma(p+2)}{\Gamma(p)} = \frac{1}{\alpha^2}\frac{(p+1)p\,\Gamma(p)}{\Gamma(p)} = \frac{(p+1)p}{\alpha^2}$$

y $\operatorname{var}(X)=E(X^2)-E(X)^2=\frac{(p+1)p}{\alpha^2}-\frac{p^2}{\alpha^2}=\frac{p}{\alpha^2}$

Por tanto, escribiremos:

$$\left.\begin{aligned} \frac{p}{\alpha} &= \bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \\ \frac{p}{\alpha^2} &= S_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i-\bar{X}_n)^2 \end{aligned}\right\}$$

dividiendo la primera por la segunda:

$$\frac{\frac{p}{\alpha}}{\frac{p}{\alpha^2}} = \frac{\bar{X}_n}{S_n^2} \Rightarrow \hat{\alpha} = \frac{\bar{X}_n}{S_n^2} \qquad (\text{para } n\geq 2)$$

$$\frac{p}{\alpha^2}=S_n^2 \Rightarrow p=\alpha^2 S_n^2 \Rightarrow \hat{p}=\frac{(\bar{X}_n)^2}{(S_n^2)^2}\cdot S_n^2 = \frac{(\bar{X}_n)^2}{S_n^2}$$

El estimador obtenido por el método de los momentos será:

$$\boxed{\hat{\theta} = (\hat{\alpha}, \hat{p}) = \left(\frac{\bar{X}_n}{S_n^2}, \frac{(\bar{X}_n)^2}{S_n^2} \right)}$$

Ejemplo 3

Si $X \sim \mathcal{U}(-\theta, \theta)$ $\theta \in \mathbb{R}$; hallar el estimador de θ por el método de los momentos, a partir de una muestra aleatoria simple de tamaño n.

Observemos en este caso que $E(X) = \frac{-\theta + \theta}{2} = 0$

por tanto el momento de primer orden (poblacional) no nos sirve para el método de los momentos. Consideraremos el de segundo orden, centrado:

$$\mathrm{var}(X) = \frac{(\theta - (-\theta))^2}{12} = \frac{(2\theta)^2}{12} = \frac{1}{3}\theta^2$$

por tanto:

$$\frac{1}{3}\theta^2 = S_n^2 \implies \boxed{\hat{\theta} = \sqrt{3 S_n^2}}$$

estimador obtenido por el método de los momentos

Estimadores máximo-verosímiles

En el contexto de un modelo estadístico paramétrico:

$$X \sim f(x,\theta) \qquad x \in \mathbb{R}^k \qquad \theta \in \Theta \subset \mathbb{R}^m$$

← densidad de probabilidad

$X_1, \ldots, X_n$ iid X (variables aleatorias muestrales correspondientes a una muestra de tamaño n)

Dados unos resultados muestrales $x_1, \ldots, x_n$, la función de verosimilitud $L_x(\theta)$ se define como:

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

El método de la máxima-verosimilitud consiste en encontrar un estimador de θ, $\overset{*}{\theta}(X_1, \ldots, X_n)$ que verifique:

$$L_x(\overset{*}{\theta}(x_1, \ldots, x_n)) = \sup_{\theta \in \Theta} L_x(\theta) \qquad \text{(MLE)}$$

a dicho estimador le llamaremos el estimador máximo-verosímil de θ.
Si $f(x,\theta)$ es positivo (salvo en un conjunto de probabilidad 0 independiente de θ), y si además L_x es diferenciable con respecto θ, un estimador máximo-verosímil satisface las ecuaciones de verosimilitud

$$\left[\begin{array}{c} \left. \dfrac{\partial \ln L_x}{\partial \theta_j} \right|_{\theta = \theta^*} = 0 \\ j = 1, \ldots m \end{array} \right]$$

Propiedades:

1.- Si $T(X_1, \ldots, X_n)$ es un estadístico suficiente es posible expresar el estimador máximo-verosímil de θ, θ^*, como una función de T
$\theta^*(X_1, \ldots, X_n) = \theta^*(T(X_1, \ldots, X_n))$ como consecuencia del Teorema de factorización de Neyman - Fisher

2.- Invariancia funcional:
Si g es una función (medible) 1-1 de Θ en un abierto $\Lambda \subset \mathbb{R}^m$ y $\theta^*(X_1 \ldots X_n)$ es el estimador máximo-verosímil de θ, entonces $g(\theta^*(X_1, \ldots, X_n))$ es el estimador máximo-verosímil de $g(\theta)$.

Ejemplo: si $X \sim N(0,\sigma^2)$ es bien sabido que el estimador máximo verosímil de σ^2 es: $(\sigma^2)^* = \frac{1}{m}\sum_{i=1}^{m} x_i^2$

Entonces escogiendo $g(u)=\sqrt{u}$, resulta que podemos afirmar que $g\left(\frac{1}{m}\sum_{i=1}^{m} x_i^2\right)$ es el estimador máximo-verosímil de $\sigma=\sqrt{\sigma^2}$

por tanto:

$$\sigma^* = \sqrt{\frac{1}{m}\sum_{i=1}^{m} x_i^2}$$

3.- Si se dan ciertas condiciones de regularidad:

(a)- Θ abierto de $\mathbb{R}^m$

(b)- Modelo identificable (es decir si $\theta_1 \neq \theta_2$ entonces $f(x,\theta_1) \neq f(x,\theta_2)$)

(c) - $f(x,\theta)>0$ excepto en un conjunto de probabilidad 0 independiente de θ.

(d) - $\frac{\partial f}{\partial \theta}$ existe para todo θ (excepto conjunto de probabilidad cero independiente de θ)

entonces si θ_0 es el verdadero valor del parámetro, existe una sucesión de raíces de las ecuaciones de verosimilitud θ_m^* que convergen casi seguramente a θ_0 (por tanto definen un estimador consistente en el sentido de la convergencia casi-segura, y por tanto también consistente en sentido ordinario).

4.- Si junto con las condiciones (a), (b) añadimos

(e) - f dos veces derivable respecto θ y además las $\frac{\partial^2 \log L_x(\theta)}{\partial\theta_i\,\partial\theta_j}$ son continuas en θ uniformemente en x.

(f) $I(\theta)$ la matriz información de Fisher está definida y $I(\theta)>0$ estrictamente definida positiva

(g) Se puede derivar respecto θ bajo el signo de integral respecto x al menos dos veces, a la función de densidad $f(x,\theta)$

Entonces, si θ_m^* es el estimador-máximo verosímil de θ, resulta:

$$\boxed{\sqrt{m}\,(\theta_m^* - \theta) \xrightarrow{\mathcal{L}} Y \sim N_m\left(0,\ I^{-1}(\theta)\right)}$$

es decir, la sucesión $\sqrt{m}\,(\theta_m^* - \theta)$ converge en ley (o en distribución) a una normal multivariante de esperanza 0 y covarianza $I^{-1}(\theta)$

Por tanto, la distribución asintótica de θ_n^* está centrada en θ (insesgado) y su matriz de covarianzas coincide con la cota de Cramer-Rao.

Por ello suele decirse, de forma un tanto imprecisa, que θ_n^* es asintoticamente insesgado y eficiente.

Ejemplo 1

Hallar el estimador máximo-verosímil de una distribución normal $N(\mu, \sigma^2)$, a partir de una muestra aleatoria simple de tamaño n.

La función de verosimilitud será:

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

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

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{n}{2\sigma^2}\left(s_n^2 + (\bar{x}_n-\mu)^2\right)} \qquad \text{con } \bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

ya que

$$\sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}(x_i - \bar{x}_n + \bar{x}_n - \mu)^2 =$$

$$= \sum_{i=1}^{n}(x_i-\bar{x}_n)^2 + n(\bar{x}_n-\mu)^2 + (\bar{x}_n-\mu)\underbrace{\sum_{i=1}^{n}(x_i-\bar{x}_n)}_{=0}$$

$$= n\,s_n^2 + n(\bar{x}_n-\mu)^2$$

Por tanto:

$$L_x(\mu,\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{n}{2\sigma^2}s_n^2} \underbrace{e^{-\frac{n}{2\sigma^2}(\bar{x}_n-\mu)^2}}_{\leq 1} \leq \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{n}{2\sigma^2}s_n^2}$$

(igual a 1 cuando $\mu = \bar{x}_n$)

Por otra parte observar que si introducimos la función:

$$g_a(y) = \frac{1}{y^{m/2}} e^{-\frac{m}{2y}a} \qquad \text{con } a>0, y>0$$

resulta que si tomamos logaritmos:

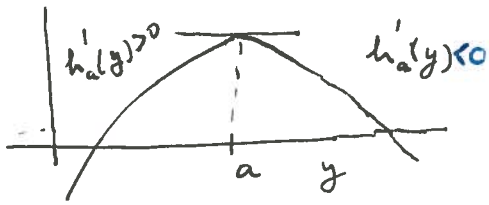
$$h_a(y) = \ln g_a(y) = -\frac{m}{2}\ln y - \frac{m}{2y}a$$

$$h'_a(y) = -\frac{m}{2}\frac{1}{y} + \frac{m}{2}\frac{a}{y^2} = \frac{m}{2y}\left(-1+\frac{a}{y}\right)$$

$h'_a(y) = 0 \iff y = a$

$h'_a(y) > 0 \iff 0 < y < a$

$h'_a(y) < 0 \iff a < y$



Por tanto $h_a(y)$ es creciente antes de $y=a$ y decreciente después de $y=a$. Por tanto h_a presenta un máximo absoluto en a.

Lo mismo ocurre con g_a, ya que $g_a(y) = e^{h_a(y)}$ tiene un máximo en $y=a$

Observar que

$$L_x(\mu,\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{m}{2}} e^{-\frac{m}{2\sigma^2}s_m^2} = \left(\frac{1}{2\pi}\right)^{\frac{m}{2}} g_{s_m^2}(\sigma^2) \leq$$

$$= \left(\frac{1}{2\pi}\right)^{\frac{m}{2}} g_{s_m^2}(s_m^2) =$$

$$= \left(\frac{1}{2\pi}\right)^{\frac{m}{2}} \frac{1}{(s_m^2)^{\frac{m}{2}}} e^{-\frac{m}{2}}$$

La verosimilitud esta acotada superiormente por

$$\left(\frac{1}{2\pi}\right)^{\frac{m}{2}} \frac{1}{(s_m^2)^{\frac{m}{2}}} e^{-\frac{m}{2}}$$

pero esta cota superior es accesible: corresponde al valor de la función de verosimilitud cuando $\mu = \bar{x}_m$ y $\sigma^2 = s_m^2$, por tanto,

$$\boxed{L_x(\mu,\sigma^2) \leq L_x(\bar{x}_n, s_n^2)}$$

y el estimador máximo-verosímil de $\theta = (\mu, \sigma^2)$ es:

$$\boxed{\theta^* = (\bar{x}_n, s_n^2)}$$

Ejemplo 2

Hallar el estimador máximo-verosímil de los parámetros de una uniforme $\mathcal{U}(\alpha,\beta)$, $\alpha < \beta$, basados en una muestra de tamaño n.

La función de densidad de una uniforme en $[\alpha,\beta]$ es:

$$f(x,\alpha,\beta) = \frac{1}{\beta-\alpha} \mathbb{1}_{[\alpha,\beta]}(x)$$

La función de verosimilitud es:

$$L_x(\alpha,\beta) = \prod_{i=1}^{n} \left\{ \frac{1}{\beta-\alpha} \mathbb{1}_{[\alpha,\beta]}(x_i) \right\} = \frac{1}{(\beta-\alpha)^n} \mathbb{1}_{[\alpha,\infty)}(x_{(1)}) \mathbb{1}_{(-\infty,\beta]}(x_{(n)})$$

es decir:

$$L_x(\alpha,\beta) = \begin{cases} \dfrac{1}{(\beta-\alpha)^n} & \text{si} \quad \alpha \leq x_{(1)} \text{ y } x_{(n)} \leq \beta \\ 0 & \text{en caso contrario.} \end{cases}$$

por tanto:

$$\boxed{L_x(\alpha,\beta) \leq \underbrace{\frac{1}{(x_{(n)}-x_{(1)})^n}}_{= L_x(x_{(1)},x_{(n)})}} \qquad (\text{ya que } \beta-\alpha \geq x_{(n)}-x_{(1)})$$

por tanto el estimador máximo-verosímil de $\theta = (\alpha,\beta)$, es:

$$\boxed{\theta^* = (x_{(1)}, x_{(n)})}$$

Ejemplo 3

Sea $\mathbf{X} = (X_1, \dots X_k)$ una variable aleatoria k-dimensional, $k > 1$, con distribución de Bernouilli multivariante de parámetros $p_1, \dots, p_k > 0$, tales que $\sum_{j=1}^{k} p_j = 1$.

Disponemos de una muestra aleatoria simple de tamaño n $\mathbf{X}_1, \dots, \mathbf{X}_n$ (como notación, $\mathbf{X}_i = (X_{i1}, \dots, X_{ik})$).

Hallar el estimador máximo-verosímil de $p_1, \dots, p_k$.

La densidad de $\mathbf{X}$ (discreta) es:

$$f(x_1, \dots, x_k, p_1, \dots, p_k) = p_1^{x_1} \cdot \ldots \cdot p_k^{x_k} \qquad \text{si } x_j \in \{0,1\} \text{ y } \sum_{j=1}^{k} x_j = 1$$

$$= 0 \qquad \text{en caso contrario.}$$

Para una muestra de tamaño n, la función de verosimilitud será (con probabilidad 1) igual a:

$$L_x(p_1, \dots, p_k) = \prod_{i=1}^{n} \left\{ p_1^{x_{i1}} \cdot \ldots \cdot p_k^{x_{ik}} \right\} =$$

$$= p_1^{\sum_{i=1}^{n} x_{i1}} \cdot \ldots \cdot p_k^{\sum_{i=1}^{n} x_{ik}}$$

Si llamamos $\nu_j \equiv \sum_{i=1}^{n} x_{ij}$ (frecuencia absoluta del número de veces en que $X_j = 1$ en la muestra)

podremos expresar la verosimilitud como:

$$L_x(p_1, \dots, p_k) = p_1^{\nu_1} \cdot \ldots \cdot p_k^{\nu_k} \qquad \left(\text{observar que } \nu_j \in \mathbb{N} \quad j = 1 \dots k \text{ y } \sum_{j=1}^{k} \nu_j = n\right)$$

Habrá que maximizar esta función o, más fácil, su logaritmo, sujeta a la ligadura: $\sum_{j=1}^{k} p_j = 1$ $\quad (p_j > 0 \quad j = 1 - k)$

Aplicaremos el método de los multiplicadores de Lagrange a $\ln L_x$:

$$\mathcal{L}_\lambda(p_1, \dots p_k) = \sum_{j=1}^{k} \nu_j \ln p_j - \lambda \left(\sum_{j=1}^{k} p_j - 1 \right)$$

$$\frac{\partial \mathcal{L}_\lambda}{\partial p_\alpha} = \frac{\nu_\alpha}{p_\alpha} - \lambda$$

$\alpha = 1 \dots k$

Formaremos el sistema:

$$\left.\begin{array}{l} \dfrac{\nu_\alpha}{p_\alpha} - \lambda = 0 \\ \alpha = 1 - k \\ p_1 + \cdots + p_k = 1 \end{array}\right\} \quad \text{equivalente a} \quad \left.\begin{array}{l} \nu_\alpha = \lambda p_\alpha \\ \alpha = 1 - k \\ p_1 + \cdots + p_k = 1 \end{array}\right\}$$

sumando las k primeras ecuaciones resulta:

$$\sum_{\alpha=1}^{k} \nu_\alpha = n = \sum_{\alpha=1}^{k} \lambda p_\alpha = \lambda \sum_{\alpha=1}^{k} p_\alpha = \lambda$$

por tanto $\lambda = n$. Sustituyendo λ por n en las k-primeras ecuaciones resulta:

$$\nu_\alpha = n\, p_\alpha \implies p_\alpha^* = \frac{\nu_\alpha}{n}$$

$$\alpha = 1, \ldots, k$$

Observemos que

$$\ln L_x(p_1^*, \ldots, p_k^*) - \ln L_x(p_1, \ldots, p_k) = \sum_{j=1}^{k} \nu_j \ln \frac{\nu_j}{n} - \sum_{j=1}^{k} \nu_j \ln p_j =$$

$$= \sum_{j=1}^{k} \nu_j \ln \frac{\frac{\nu_j}{n}}{p_j} = n \sum_{j=1}^{k} \frac{\nu_j}{n}\left(-\ln\left(\frac{p_j}{\nu_j/n}\right)\right) \geqslant$$

$$\geqslant n\left(-\ln\left(\sum_{j=1}^{k} \frac{\nu_j}{n} \frac{p_j}{\nu_j/n}\right)\right) = n\left(-\ln\left(\sum_{j=1}^{k} p_j\right)\right)$$

↑ por la desigualdad de Jensen: $E(f(X)) \geqslant f(E(X))$ para f convexa

$$= -n \ln 1 = 0$$

y $-\ln$ es CONVEXA

por tanto:

$$\boxed{\ln L_x(p_1, \ldots, p_k) \leqslant \ln_x L_x(p_1^*, \ldots, p_k^*)}$$

Por tanto $\boxed{p_1^* = \frac{\nu_1}{n}, \ldots, p_k^* = \frac{\nu_k}{n}}$

constituyen la <u>estimación máximo-verosímil de</u> <u>$p_1, \ldots, p_k$</u>.

Método de estimación de Bayes

Sea $X \sim f(x,\theta)$ $x \in \mathbb{R}^k$ $\theta \in \Theta \subset \mathbb{R}^m$

y $X_1, \dots, X_n$ i.i.d X

Los métodos Bayesianos tratan de cuantificar el conocimiento previo a la muestra que disponemos referente al parámetro. Esto se efectua introduciendo lo que llamaremos una distribución "a priori" en el espacio de parámetros Θ, corrientemente en términos de una función de densidad de probabilidad en Θ, $h(\theta)$.

A partir de aquí, se construye una distribución de probabilidad en el producto cartesiano $\Omega \times \Theta$ donde Ω es el espacio muestral correspondiente a la muestra aleatoria simple de tamaño n que disponemos $(\Omega = \mathbb{R}^k \times \overset{n}{\cdots} \times \mathbb{R}^k)$

La función de densidad que construimos en $\Omega \times \Theta$ es:

densidad conjunta en $\Omega \times \Theta$

$$h(\theta) \cdot f(x_1,\theta) f(x_2,\theta) \dots f(x_n, \theta)$$

es decir multiplicaremos la densidad "a priori" por la densidad conjunta de la muestra condicionada a un valor determinado del parámetro.

Posteriormente calcularemos la función de densidad del parámetro dada la muestra:

$$g(\theta | x_1 - x_n) = \frac{h(\theta) \prod_{i=1}^{n} f(x_i, \theta)}{\int_\Theta h(\gamma) \left(\prod_{i=1}^{n} f(x_i, \gamma)\right) d\gamma}$$

$(d\gamma \equiv d\gamma_1 d\gamma_2 \cdots d\gamma_m$ si el parámetro es m-dimensional$)$

Esta densidad condicionada a la muestra obtenida es conocida como distribución "a posteriori".

Paralelamente, dada una función de pérdida, se define el Riesgo de Bayes correspondiente a un estimador u como:

$$\boxed{RB(u) = \int_\Theta h(\theta)\, R_\theta(u)\, d\theta}$$

donde $\boxed{R_\theta(u) = E_\theta(\ell(u,\theta))}$ siendo $\ell(u,\theta)$ la función de pérdida correspondiente.

Es decir el Riesgo de Bayes es el promedio del riesgo ordinario correspondiente a una función de pérdida, calculado en términos de la distribución "a priori".

Se define el estimador de Bayes como aquél que minimiza el riesgo de Bayes.

Si usamos la pérdida cuadrática puede demostrarse que el estimador de Bayes es igual al valor medio de la distribución a posteriori.

* Ejemplo 1

Sea $X \sim$ Exponencial(α). Sea $X_1, \dots, X_n$ i.i.d X.

Consideremos la distribución "a priori" dada por $h(\alpha) = e^{-\alpha} \quad \alpha > 0 \quad (\Theta = \mathbb{R}^+)$. Hallar el estimador de Bayes correspondiente a la pérdida cuadrática.

La función de densidad conjunta en el espacio muestral producto cartesiano espacio de parámetros es:

$$e^{-\alpha} \prod_{i=1}^{n} \left\{ \alpha e^{-\alpha x_i} \right\}$$

(suponemos $x_i > 0$, $i = 1 - n$, lo que ocurrirá con probabilidad 1)

por tanto la distribución "a posteriori", teniendo en cuenta que la densidad conjunta es igual a $\alpha^n e^{-\alpha\left(\left(\sum_{i=1}^{n} x_i\right)+1\right)}$, será:

$$g(\alpha | x_1 - x_n) = \frac{\alpha^n e^{-\alpha\left(\left(\sum_{i=1}^{n} x_i\right)+1\right)}}{\int_0^\infty \gamma^n e^{-\gamma\left(\left(\sum_{i=1}^{n} x_i\right)+1\right)} d\gamma}$$

$$= \frac{\left(\left(\sum_{i=1}^{n} x_i\right)+1\right)^{n+1}}{\Gamma(n+1)} \alpha^n e^{-\left(\left(\sum_{i=1}^{n} x_i\right)+1\right)\alpha}$$

es decir, la distribución de α condicionada a $x_1 - x_n$ es una Gamma$\left(\left(\sum_{i=1}^{n} x_i\right)+1, \ n+1\right)$

por tanto el estimador de Bayes será:

$$u(x_1 - x_m) = \int_0^\infty \alpha \, g(\alpha | x_1 - x_m) \, d\alpha =$$

$$= \int_0^\infty \alpha \frac{\left(\left(\sum_{i=1}^m x_i\right)+1\right)^{m+1}}{\Gamma(m+1)} \alpha^m e^{-\left(\left(\sum_{i=1}^m x_i\right)+1\right)\alpha} d\alpha$$

$$= \frac{1}{\left(\left(\sum_{i=1}^m x_i\right)+1\right)\Gamma(m+1)} \int_0^\infty \left(\left(\sum_{i=1}^m x_i\right)+1\right)^{m+2} \alpha^{m+1} e^{-\left(\left(\sum_{i=1}^m x_i\right)+1\right)\alpha} d\alpha$$

con el cambio $\left(\left(\sum_{i=1}^m x_i\right)+1\right)\alpha = t$ resulta

$$= \frac{\Gamma(m+2)}{\left(\left(\sum_{i=1}^m x_i\right)+1\right)\Gamma(m+1)} =$$

$$= \frac{m+1}{\left(\sum_{i=1}^m x_i\right)+1} = \frac{m+1}{m\,\bar{x}_m + 1} = \frac{1}{\frac{m}{m+1}\bar{x}_m + \frac{1}{m+1}}$$

$$\boxed{u(X_1 - X_m) = \frac{1}{\frac{m}{m+1}\bar{X}_m + \frac{1}{m+1}}}$$

que asintóticamente es equivalente al MLE: $\alpha^* = \frac{1}{\bar{X}_m}$

↑

Estimador de Bayes correspondiente a la pérdida cuadrática $\ell(u,\alpha) = (u-\alpha)^2$

Nota: los cálculos se acortan si nos damos cuenta que

$$g(\alpha | x_1 - x_m) \propto \alpha^m e^{-\left(\left(\sum_{i=1}^m x_i\right)+1\right)\alpha}$$

implica que $\alpha | X_1 - X_m \sim \text{Gamma}\left(\left(\sum_{i=1}^m x_i\right)+1, \, m+1\right)$

y que la esperanza de una gamma de parámetros β, k es igual a $\frac{k}{\beta}$.

---

* Ejemplo 2

Sea $X \sim$ Bernouilli (p) ; $X_1 \ldots X_n$ i.i.d X ; $\Theta = [0,1]$.

Consideremos ahora la distribución "a priori"

$h(p) = p^a (1-p)^b$ con $a, b > 0$ y conocidos.

Hallar el estimador de Bayes correspondiente a la pérdida cuadrática.

$$g(p|x_1 \ldots x_n) \propto p^a (1-p)^b \prod_{i=1}^{n} \left\{ p^{x_i} (1-p)^{1-x_i} \right\}$$

$x_i \in \{0,1\}$
$i = 1 \ldots n$
con prob. igual a 1

$$\propto p^{a+\sum_{i=1}^n x_i} (1-p)^{n+b-\sum_{i=1}^n x_i}$$

por tanto

$$g(p|x_1 \ldots x_n) = c\, p^{a+\sum_{i=1}^n x_i} (1-p)^{n+b-\sum_{i=1}^n x_i}$$

para una cierta constante de normalización c. Para hallarla hay que recordar que la función Beta de Euler satisface:

$$\left[B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx \quad \text{para } \alpha,\beta > 0 \right]$$

por tanto, como:

$$1 = \int_0^1 c\, p^{a+\sum_{i=1}^n x_i} (1-p)^{n+b-\sum_{i=1}^n x_i} dp = c\, B\left(a+1+\sum_{i=1}^n x_i,\ n+b+1-\sum_{i=1}^n x_i\right)$$

y

$$c = \frac{1}{B\left(a+1+\sum_{i=1}^n x_i,\ n+b+1-\sum_{i=1}^n x_i\right)} = \frac{\Gamma(n+a+b+2)}{\Gamma\left(a+1+\sum_{i=1}^n x_i\right)\Gamma\left(n+b+1-\sum_{i=1}^n x_i\right)}$$

En cuanto al estimador de Bayes correspondiente a la pérdida cuadrática será:

$$\Bigg[u(x_1 \ldots x_n) = \int_0^1 p\, g(p|x_1 \ldots x_n)\, dp = \int_0^1 \frac{\Gamma(n+a+b+2)\, p^{a+1+\sum_{i=1}^n x_i} (1-p)^{n+b-\sum_{i=1}^n x_i}}{\Gamma\left(a+1+\sum_{i=1}^n x_i\right)\Gamma\left(n+b+1-\sum_{i=1}^n x_i\right)} dp$$

$$= \frac{\Gamma(n+a+b+2)}{\Gamma\left(a+1+\sum_{i=1}^n x_i\right)\Gamma\left(n+b+1-\sum_{i=1}^n x_i\right)} \cdot \frac{\Gamma\left(a+2+\sum_{i=1}^n x_i\right)\Gamma\left(n+b+1-\sum_{i=1}^n x_i\right)}{\Gamma(n+a+b+3)} \Bigg]$$

y, por las propiedades de la función gamma,

$$\mathcal{U}(x_1 - x_n) = \frac{a+1+\sum_{i=1}^{n} x_i}{n+a+b+2} = \frac{\frac{a+1}{n} + \bar{x}_n}{1 + \frac{a+b+2}{n}}$$

y, por tanto, el estimador de Bayes correspondiente es:

$$\boxed{\mathcal{U}(X_1 - X_n) = \frac{\frac{a+1}{n} + \bar{X}_n}{1 + \frac{a+b+2}{n}}}$$

que asintóticamente es equivalente al MLE, $p^* = \bar{X}_n$

* Ejemplo 3

Sea $X \sim N(\mu, \sigma)$ con σ conocida. Sea $X_1, \ldots, X_n$ i.i.d X. Consideremos la apriori: $\mu \sim N(\mu_0, \sigma_0)$ con μ_0 y σ_0 conocidos. Determinar el estimador de Bayes correspondiente a la pérdida cuadrática.

$$g(\mu \mid x_1 - x_n) \propto e^{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2} \prod_{i=1}^{n} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$$

$$\propto e^{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$\propto e^{-\frac{1}{2}\left(\frac{\mu^2}{\sigma_0^2} - \frac{2\mu_0}{\sigma_0^2}\mu + \frac{n\mu^2}{\sigma^2} - \frac{2n\bar{x}_n}{\sigma^2}\mu\right)}$$

(donde $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$

$$\propto e^{-\frac{1}{2}\left(\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)\mu^2 - 2\left(\frac{\mu_0}{\sigma_0^2} + \frac{n\bar{x}_n}{\sigma^2}\right)\mu\right)}$$

$$\propto e^{-\frac{1}{2\frac{1}{\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)}}\left(\mu^2 - 2\frac{\left(\frac{\mu_0}{\sigma_0^2}+n\frac{\bar{x}_n}{\sigma^2}\right)}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}}\mu\right)}$$

$$\propto e^{-\frac{1}{2\frac{\sigma_0^2\sigma^2}{\sigma^2+n\sigma_0^2}}\left(\mu^2 - 2\frac{\mu_0\sigma^2+n\sigma_0^2\bar{x}_n}{\sigma^2+n\sigma_0^2}\mu\right)}$$

observar que :

$$g(\mu \mid x_1 - x_n) = C\, e^{-\frac{1}{2\left(\frac{\sigma_0^2\sigma^2}{\sigma^2+n\sigma_0^2}\right)}\left(\mu - \frac{\mu_0\sigma^2 + n\sigma_0^2\,\bar{x}_n}{\sigma^2+n\sigma_0^2}\right)^2}$$

y por tanto la distribución de μ condicionada a $x_1, \dots, x_n$ es una normal de esperanza $\dfrac{\mu_0\sigma^2 + n\sigma_0^2\,\bar{x}_n}{\sigma^2+n\sigma_0^2}$

y varianza $\dfrac{\sigma_0^2\sigma^2}{\sigma^2+n\sigma_0^2}$. La constante (de normalización)

C será:

$$C = \frac{1}{\sqrt{2\pi}\,\dfrac{\sigma_0\sigma}{\sqrt{\sigma^2+n\sigma_0^2}}} = \frac{\sqrt{\sigma^2+n\sigma_0^2}}{\sqrt{2\pi}\,\sigma_0\sigma}$$

El estimador de Bayes correspondiente a la pérdida cuadrática es:

$$\boxed{u(X_1,\dots,X_n) = \frac{\mu_0\sigma^2 + n\sigma_0^2\,\bar{X}_n}{\sigma^2+n\sigma_0^2} = \frac{\mu_0\sigma^2}{\sigma^2+n\sigma_0^2} + \frac{n\sigma_0^2}{\sigma+n\sigma_0^2}\,\bar{X}_n}$$

observemos que cuando $n \to \infty$ $u \to \bar{X}_n$,
es decir, asintóticamente volvemos a tender al estimador máximo-verosímil.

Observemos, además, que si $\sigma_0 \to 0$ entonces $u \to \mu_0$
es decir: la información "a priori" se vuelve cada vez más importante, y en el límite hace que la muestra sea "irrelevante".