Probabilidad y estadística

Clase 4

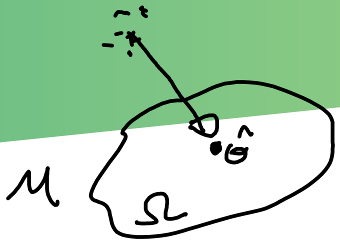
Estimadores puntuales

Estimadores puntuales

$\hat{\Theta}(\underline{X})$ es una v.a.

Def: Dada una muestra aleatoria $\underline{X}_n$, un estimador $\hat{\theta}$ de un parámetro θ es una función de la muestra aleatoria que provee un valor aproximado del parámetro o característica desconocido.

Tipos de estimación



A grandes rasgos, existen dos enfoques para la estimación de parámetros:

θ

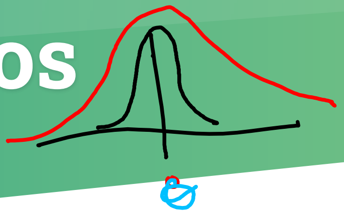
$\hat{\theta} \xrightarrow{n \to \infty}$

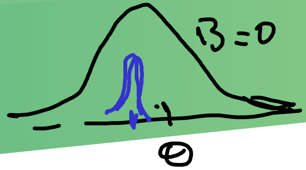
$\hat{\theta}$ es puntual

- **Enfoque frecuentista:** considera que el parámetro a estimar es fijo, es decir que tienen un valor, sólo que yo lo desconozco
- **Enfoque Bayesiano:** considera que el parámetro a estimar tiene una naturaleza aleatoria. Hay una idea de “creencias”. Θ

En este video hay una explicación intuitiva a partir de un ejemplo sencillo de las diferencias entre ambos enfoques.

Bondades de los estimadores





$= \int \hat{\theta}(\underline{x})\, f_{\underline{x}}(\underline{x})\, d\underline{x} - \theta$

Def: $\hat{\theta}$ es un estimador **insesgado** para θ si $B := \mathbb{E}[\hat{\theta} - \theta] = 0 \ \forall\, \theta$. En caso contrario diremos que es **sesgado**. A B se lo conoce como **sesgo**

Def: $\hat{\theta}$ es un estimador **asintóticamente insesgado** para θ si

$$\lim_{n\to\infty} \underbrace{\mathbb{E}[\hat{\theta} - \theta]}_{B(n) \to 0} = 0 \ \forall\, \theta$$

Obs: El ECM se puede descomponer como:

compromiso

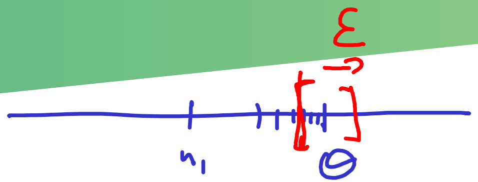
$$E[(\hat{\theta} - \theta)^2] \simeq ECM = \underbrace{var(\hat{\theta})}_{varianza} + \underbrace{B(\hat{\theta})^2}_{sesgo}$$

$(E\hat{\theta} - E\hat{\theta}) = 0$

$$E[(\underbrace{(\hat{\theta} - E\hat{\theta})}_{a} + \underbrace{(E\hat{\theta} - \theta)}_{b})^2] = \underbrace{E[(\hat{\theta} - E\hat{\theta})^2]}_{var(\hat{\theta})} + \underbrace{2\,E[(\hat{\theta} - \underbrace{E\hat{\theta}}_{cte})]\,(E\hat{\theta} - \theta)}_{=0} + \underbrace{(E\hat{\theta} - \theta)^2}_{B(\hat{\theta})^2}$$

$(a + b)^2$

Bondades de los estimadores



$\left[\hat{\Theta} = \hat{\Theta}(\underline{X}_n) = \hat{\Theta}_n\right]$

Def: Dada una sucesión de estimadores $\hat{\theta}_n$ de θ, diremos que $T = \hat{\theta}$ es (débilmente) consistente si

$$\forall\, \varepsilon > 0, \mathbb{P}(|T - \theta| > \varepsilon) \xrightarrow[n\to\infty]{} 0$$

Teorema: Si $var(\hat{\theta}) \to 0$ y $\underbrace{\mathbb{E}[\hat{\theta}] \to \theta}_{\text{asint. insesgado}}$, entonces $\hat{\theta}$ es consistente.

Def: Un estimador es consistente en media cuadrática si

$$\lim_{n\to\infty} ECM(\hat{\theta}_n) = 0, \forall\theta$$

$var(\theta) + B(\hat{\theta})^2 \to 0$

Ejercicio 6

Se desea estimar la media de una variable con distribución N(μ,9) a partir del promedio de n realizaciones. Analizar las bondades de las que goza dicho estimador.

$X \sim N(\mu, \sigma)$, Tenemos una m.a. $\underline{X} = (X_1, \ldots, X_n)$

$\Rightarrow X_1, \ldots, X_n$ son iid y $X_i \sim X$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$B = E[\hat{\mu}] - \mu = E\left[\frac{1}{n} \sum X_i\right] - \mu = \frac{1}{n} \sum \underbrace{E X_i}_{\mu} - \mu = \frac{1}{\cancel{n}} \cancel{n} \cdot \mu - \mu = 0$$

$\therefore$ $\hat{\mu}$ es insesgado.

$$\operatorname{var}(\hat{\mu}) = \operatorname{var}\left(\frac{1}{n} \sum x_i\right) = \frac{1}{n^2} \operatorname{var}\left(\sum x_i\right) \overset{\text{iid}}{=} \frac{1}{n^2}\, n \cdot \operatorname{var}(x_1)$$

$$= \frac{\sigma}{n} \xrightarrow[n \to \infty]{} 0$$

y como es insesgado, por el teo, es consistente.

también es consistente en media cuadrática.

Estimación Bayesiana

Enfoque Bayesiano

El enfoque Bayesiano supone que se tiene alguna información previa sobre el parámetro, o una ‘creencia’. Esta información previa se expresa en forma de una distribución sobre θ, llamada distribución a priori.

El objetivo del enfoque Bayesiano es ir actualizando nuestra creencia a partir de las muestras observadas.

Distribución a priori

Según el problema se pueden tener distintas interpretaciones acerca de la distribución a priori $\pi(\theta)$:

- La distribución a priori está basada en experiencias previas similares
- La distribución a priori expresa una creencia subjetiva.

A la variable aleatoria la llamaremos Θ , para distinguirla del valor que toma θ.

También cambia la interpretación de familia de distribución, ya que ahora $\underline{F_\theta(x)}$ es la distribución condicional de X dado que $\underline{\Theta = \theta}$

Distribución a posteriori

Una vez observada la m.a. $\underline{X} = (X_1, \ldots, X_n)$ se puede calcular la distribución de Θ dada $\underline{X} = \underline{x}$. A esta distribución se la conoce como distribución a posteriori y está dada por

$$f_{\Theta|\underline{X}=\underline{x}}(\theta) = \frac{\overbrace{f_{\underline{X}|\Theta=\theta}(\underline{x})}^{\text{verosimilitud}}\overset{\text{a priori}}{\pi(\theta)}}{\int_{-\infty}^{\infty} f_{\underline{X}|\Theta=\theta}(\underline{x})\pi(\theta)d\theta} \quad \begin{array}{l} \pi\{\theta \in A\} \\ \text{cte para Normalizar} \end{array}$$

Obs: Si alguna de las variables ($\underline{X}$ o Θ) son discretas reemplazaremos la función de densidad por una función de probabilidad, y si Θ reemplazaremos la integral por una sumatoria.

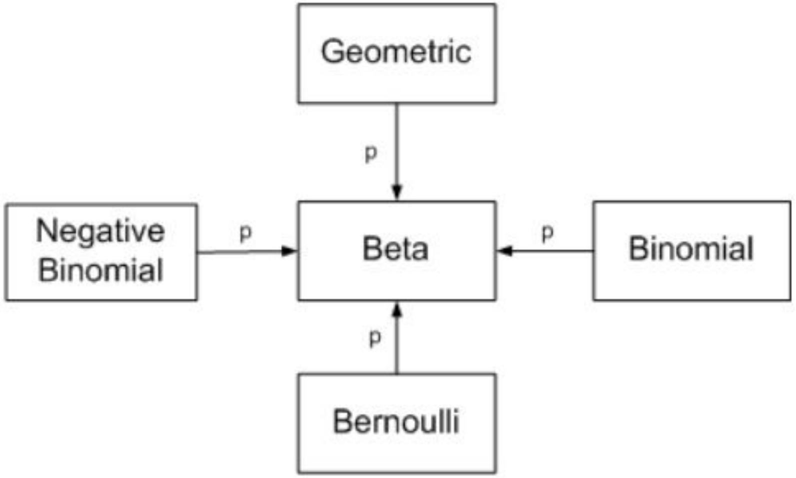
Cómo elegir la distribución a priori

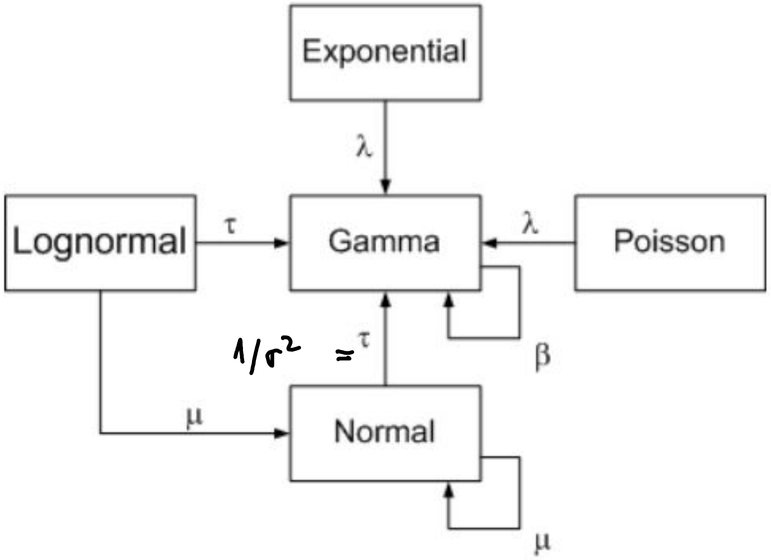
- Cuando no tenemos información acerca del parámetro vamos a usar distribuciones a priori que no favorezcan ningún valor, como por ejemplo una distribución uniforme.

- Hay que prestar atención al soporte que tiene la distribución a priori elegida, ya que los valores de θ que no pertenezcan al soporte original quedan "eliminados".

- Por simplicidad, suelen elegirse **distribuciones conjugadas**. Las familias de distribuciones conjugadas son aquellas para las cuales la función de dist. a posteriori va a pertenecer a la misma familia de distribuciones que la dist. a priori.

Ejemplos de familias de dist. conjugadas





[Fuente]

Ejercicio 1

La posición del impacto en un tiro al blanco (en decímetros) respecto del cero sobre el eje x es una variable aleatoria X con distribución normal de media cero y varianza 1/θ, donde θ representa la precisión del tirador.
A priori, la precisión θ tiene una distribución Chi-cuadrado de 8 grados de libertad. Lucas tiro 10 veces al blanco y observó que $\sum_{i=1}^{10} x_i^2 = 17$. Hallar la distribución a posteriori de θ.

La posición del impacto en un tiro al blanco (en decímetros) respecto del cero sobre el eje x es una variable aleatoria X con distribución normal de media cero y varianza 1/θ, donde θ representa la precisión del tirador.
A priori, la precisión θ tiene una distribución Chi-cuadrado de 8 grados de libertad. Lucas tiró 10 veces al blanco y observó que $\sum_{i=1}^{10} x_i^2 = 17$. Hallar la distribución a posteriori de θ.

$X \sim N(0, \frac{1}{\theta})$ $\quad \theta = \frac{1}{\sigma^2}$

$\pi(\theta) = \chi^2_8 = \Gamma(4, 2)$

Gamma	$\Gamma(\nu, \lambda)$	$\frac{\lambda^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\lambda x}$

$$f_{\Theta|\underline{X}=x}(\theta) \propto f_{\underline{X}|\Theta=\theta}(x) \cdot \pi(\theta)$$

$$= \prod_{i=1}^{n} \frac{\sqrt{\theta}}{\sqrt{2\pi}} e^{-\frac{\theta}{2} x_i^2} \cdot \frac{\theta^3}{2^4 \Gamma(4)} e^{-\frac{\theta}{2}} \quad \mathbb{1}\{\theta > 0\}$$

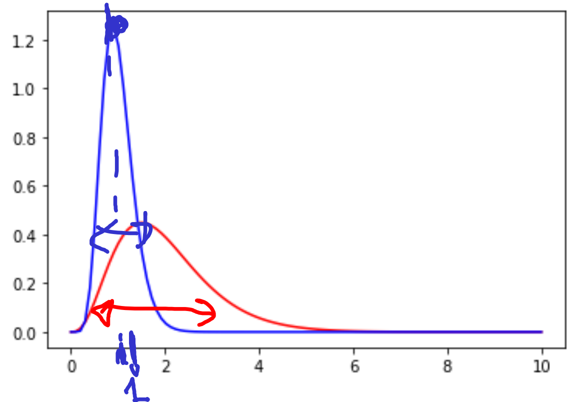
$$\propto \theta^{\frac{n}{2}+3} e^{-\frac{\theta}{2}\sum x_i^2} \cdot e^{-\frac{\theta}{2}}$$

$\nu - 1 = \frac{n}{2} + 3 = 8$

$\nu = 9$

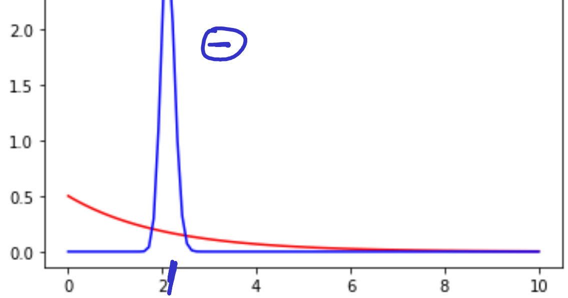
$$= \theta^{\frac{n}{2}+3} e^{-\frac{\theta}{2}\left(\sum x_i^2 + 1\right)} \quad \mathbb{1}$$

$$\Theta \sim \Gamma\left(9, \frac{\sum x_i^2 + 1}{2}\right) \quad 9 = \frac{\sum x_i^2+1}{2}$$



Ejercicio 2

$X \sim \text{Poisson}(\mu)$



La cantidad de accidentes semanales en una planta industrial tiene una distribución de Poisson de media μ.

$n=100$

En una muestra de 100 semanas se observaron las frecuencias:

Cantidad de accidentes	0	1	2	3	4	5
Frecuencia	10	29	25	17	13	6

(hubo 10 semanas sin accidentes, hubo 29 semanas con 1 accidente...)

$$\sum_{i=1}^{100} x_i = 0 \cdot 10 + 1 \cdot 29 + \cdots + 5 \cdot 6 = 212$$

A priori, μ tiene una distribución exponencial de media 2. Hallar la distribución a posteriori de μ.

$$\mu \sim \mathcal{E}\left(\tfrac{1}{2}\right)$$

Gamma	$\Gamma(\nu,\lambda)$	$\frac{\lambda^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\lambda x}$

$$f_{M \mid \underline{X}=\underline{x}}(\mu) \propto f_{\underline{X} \mid M=\mu}(\underline{x}) \; \pi(\mu)$$

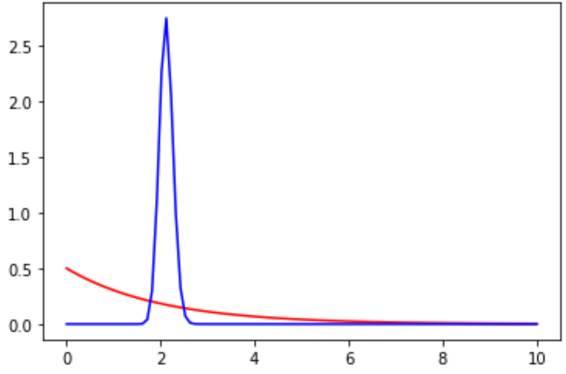
$$\overset{\text{iid}}{=} \prod_{i=1}^{n} \frac{\mu^{x_i} e^{-\mu}}{x_i!} \; \frac{1}{2} e^{-\frac{1}{2}\mu} \; \mathbb{1}\{\mu \geq 0\}$$

$$= \mu^{\sum x_i} e^{-n\mu} e^{-\frac{1}{2}\mu} \; \mathbb{1}\{\mu \geq 0\}$$

$\nu - 1 = 212$

$$= \mu^{\overset{212}{\sum x_i}} e^{-(n+\frac{1}{2})\mu} \; \mathbb{1}\{\mu \geq 0\}$$

$$M \mid \underline{X}=\underline{x} \sim \Gamma(213 \, , \, 100{,}5)$$



Estimadores puntuales Bayesianos

Una ventaja de este método es que se pueden definir de manera natural estimadores óptimos. Según que se considere como función de pérdida se obtendrán distintos estimadores puntuales.

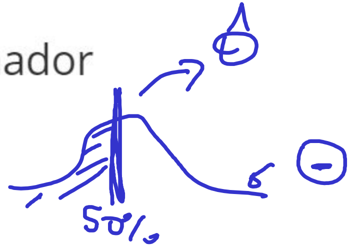
Si consideramos una función de pérdidas $l(\theta, d)$, que es el costo de estimar al parámetro θ por el valor d, si $\hat{\theta} = \varphi(\underline{X})$, entonces la pérdida será una v.a. $l(\Theta, \varphi(\underline{X}))$.

La pérdida esperada es lo que se conoce como riesgo de Bayes. Esto significa que dada una distribución a priori $\pi(\theta)$, un estimador de Bayes será el que minimice $r(\varphi, \pi) = \mathbb{E}[l(\Theta, \varphi(\underline{X}))]$

Distintos estimadores según la función de riesgo

- Si consideramos una función de pérdida cuadrática: $\ell(\theta, d) = (\theta - d)^2$, el estimador de Bayes será es que minimice el ECM. ¿Quién era este estimador? $\varphi(\underline{X}) = \mathbb{E}[\Theta|\underline{X}]$ (con la esperanza tomada respecto de la distribución a posteriori.)

- Si consideramos la pérdida ℓ_1 dada por $\ell(\theta, d) = |\theta - d|$, el estimador de Bayes será la $\boxed{\text{mediana}}$ de la distribución a posteriori de Θ condicionada a $\underline{X} = \underline{x}$.



- Muchas veces se utiliza lo que se conoce como máximo a posteriori (MAP) que se corresponde con la $\boxed{\text{moda}}$ de la distribución a posteriori de $\Theta|\underline{X} = \underline{x}$, es decir el de el valor de θ que maximiza la dist. a posteriori. En general no es un estimador Bayesiano.

Ejercicio 3

Para el ejercicio 1, hallar la estimación de Bayes de θ para el riesgo cuadrático.

$$\Theta \mid \underline{X} = \underline{x} \sim \Gamma(9, 9)$$

$$\hat{\Theta} = E[\Theta \mid \underline{X} = \underline{x}] = \frac{9}{9} = 1$$

para una $\Gamma(\nu, \lambda)$, la esperanza es $\frac{\nu}{\lambda}$

Cálculo de probabilidades

Para poder estimar probabilidades a partir de la distribución a posteriori de los parámetros aleatorios, usaremos la fórmula de probabilidad total. Por ejemplo, si Θ y X son v.a. continuas

$$\mathbb{P}(X \in A) = \int_A \int_{-\infty}^{\infty} f_{X|\Theta=\theta}(x) f_{\Theta|\underline{X}=\underline{x}}(\theta) d\theta dx$$

$f_{\Theta|\underline{X}=\underline{x}}(\theta)$

<u>Prob. total</u>

$$P(X \in A) = \int_A f_X(x) dx = \int_A f_{X|\Theta=\theta}(x) \, \boxed{f_\Theta(\theta)} \, d\theta dx$$

Ejercicio 4

$\Theta | \underline{X} = x \sim \Gamma(213;\ 100{,}5)$

$\qquad\qquad\qquad\qquad\quad \nu \qquad\quad \lambda$

$X \sim \text{Poisson}(\Theta)$

X: # accidentes en 1 mes

Gamma	$\Gamma(\nu, \lambda)$	$\frac{\lambda^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\lambda x}$

Para el ejercicio 2, estimar la probabilidad de que en la semana del 18 de diciembre de 2021 no ocurra ningún accidente en la mencionada planta.

$$P(X=0) = \sum_{x=0} \int_{-\infty}^{\infty} p_{X|\Theta=\theta}(x) \cdot f_{\Theta|\underline{X}=x}(\theta)\, d\theta$$

$$= \int_{-\infty}^{\infty} \sum_{x=0} \frac{\theta^{x} e^{-\theta}}{x!} \cdot \frac{100{,}5^{213}}{\Gamma(213)} \theta^{212} e^{-100{,}5\theta}\, \mathbb{1}_{\{\theta>0\}}\, d\theta$$

$$\hat{P}(x=0) = \int_{-\infty}^{\infty} \underbrace{\sum_{x=0} \left(\frac{\theta^{x} e^{-\theta}}{x!}\right)}_{=e^{-\theta}} \cdot \frac{100,5^{213}}{\Gamma(213)} \theta^{212} e^{-100,5\,\theta} \mathbb{1}_{\{\theta>0\}} \, d\theta$$

$$= \int_{0}^{\infty} \frac{100,5^{213}}{\Gamma(213)} \underbrace{\theta^{212} e^{-101,5\,\theta}}_{\longrightarrow \Gamma(213,\, 101,5)} \, d\theta$$

Gamma	$\Gamma(\nu, \lambda)$	$\frac{\lambda^{\nu}}{\Gamma(\nu)} x^{\nu-1} e^{-\lambda x}$

$$= \frac{100,5^{213}}{\cancel{\Gamma(213)}} \cdot \frac{\cancel{\Gamma(213)}}{101,5^{213}} \underbrace{\int_{0}^{\infty} \frac{101,5^{213}}{\Gamma(213)} \theta^{212} e^{-101,5\,\theta} \, d\theta}_{=1}$$

$$= \left(\frac{100,5}{101,5}\right)^{213} \simeq 0,12$$

Enfoque frecuentista

Estadístico

Def: Dada una muestra aleatoria $\underline{X}_n$, un estadístico es cualquier función $T_n = T(\underline{X}_n)$

Def: Sea una muestra aleatoria $\underline{X}_n$, cuya distribución es $F_\theta(\underline{x}), \theta \in \Theta$, se dice que $T = r(\underline{X}_n)$ es un estadístico suficiente para $\boxed{\theta}$ si $\underline{F_{\underline{X}|T=t}(\underline{x})}$ no depende de θ.

Teorema de factorización: Diremos que $T = r(\underline{X}_n)$ es un est. suficiente para θ sii existen funciones h y g tales que:

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} \quad f_\theta(\underline{x}) = \underbrace{g(r(\underline{x}), \theta)}_{\text{ver de } r(\underline{x})}\underbrace{h(\underline{x})}_{\text{cte cuando fijo } \theta}$$

$\Longrightarrow$ la verosimilitud de $\underline{x}$
es $\propto$ " de $r(\underline{x})$

Ejercicio 5

1. Hallar un estadístico suficiente para el parámetro de una v.a. con distribución exponencial
2. Hallar un estadístico suficiente para el parámetro de una v.a. con distribución de Bernoulli
3. Hallar un estadístico suficiente para el parámetro de una v.a. con distribución Uniforme en el intervalo $(0,\theta)$

$X \sim \mathcal{E}(\lambda)$ y $\underline{X}$ es una m.a de tamaño n.

- $f_{\underline{X}}(\underline{x}) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \underbrace{\lambda e^{-\lambda \sum_{i=1}^{n} x_i}}_{g(\underline{x})} \rightarrow r(x)$

- $X \sim Ber(p)$

$f_{\underline{X}}(x) = \prod_{i=1}^{n} p^{x}(1-p)^{1-x} = p^{\sum x}(1-p)^{n-\sum x}$

- $X \sim U(0,\theta)$

$f_{\underline{X}}(x) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}\{0 \leq x_i \leq \theta\}$

$= \frac{1}{\theta^n} \mathbb{1}\{0 \leq \min(\underline{x})\} \; \mathbb{1}\{\underbrace{\max(\underline{x})}_{} \leq \theta\}$ g $r(\underline{x})$

Método de máxima verosimilitud

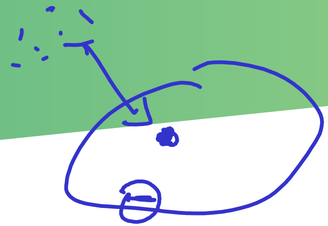
Método de Máxima Verosimilitud

Def: Diremos que $\hat{\theta}(\underline{X})$ es un Estimador de Máxima Verosimilitud (MLE) si se cumple que:

$$f(\underline{X}, \hat{\theta}) = \max_{\theta} f_{\theta}(\underline{X})$$

La idea es que si observé una determinada muestra, entonces esta debería tener alta probabilidad de ocurrir, por lo tanto busco el θ que maximiza esa probabilidad de ocurrencia

Método de Máxima Verosimilitud



Def: Definimos la función de verosimilitud como

$L(\theta) = f(\underline{x}, \theta)$ (vista como función de θ) luego,

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\max}\, L(\theta).$$

el dominio de Θ

Si el soporte de X no depende θ, Θ es un conjunto abierto y $f_\theta(x)$ es derivable respecto de θ, entonces para hallar el EMV puedo hallar θ tal que

$$\frac{\partial\, ln(L(\theta))}{\partial\, \theta} = 0$$

Ejercicio 7

La probabilidad de acertar a un blanco es p. Se realizan 10 tiros independientes, en los cuales se observaron 4 aciertos. A partir este valor observado estimar el valor de p por MV.

$X \sim Ber(p)$

$$\mathcal{L}(p) = p^{\sum x_i}(1-p)^{10-\sum x_i} = p^4 \cdot (1-p)^6$$

$$\ell(p) = \log \mathcal{L}(p) = 4 \log p + 6 \log(1-p)$$

$$\frac{d}{dp}\ell = \frac{4}{p} - \frac{6}{1-p} = 0 \Rightarrow 4(1-p) = 6p$$

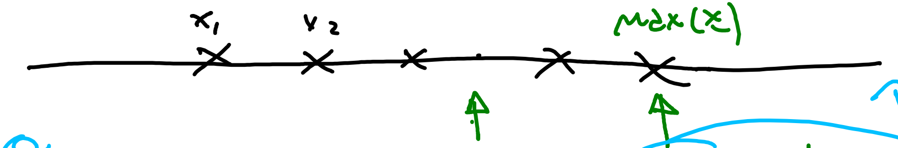
$$4 = 10p \Rightarrow \boxed{\hat{p} = \frac{4}{10}}$$

Ejercicio 8

1. Sea X~U(0, θ). Hallar el EMV para θ basado en una muestra de tamaño n.
2. Si en una muestra de tamaño 3 se observaron los valores (0.756, 1.1, 3,11), hallar el valor estimado de θ.

$\underline{X}$ una m.a.

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} \frac{1}{\theta}\, \mathbb{1}\{0 < x_i < \theta\} = \frac{1}{\theta^n} \cdot \mathbb{1}\{0 \le \min(\underline{x})\}\, \mathbb{1}\underbrace{\{\max(\underline{x}) \le \theta\}}_{=0}$$



θ_1 < $\mathcal{L}(\theta) = \frac{1}{\max(\underline{X})}$ > $\mathcal{L}(\theta_2) = \frac{1}{\max(\underline{X})+\varepsilon}$

∴ el emv de θ es $\max(\underline{X})$.

Principio de invariancia

Supongamos que ahora queremos estimar por máxima verosimilitud a $\boxed{\lambda = q(\theta).}$

MV

Teorema: Si $\hat{\theta}$ es MLE de θ, entonces $\hat{\lambda} = q(\hat{\theta})$.

¿Por qué es útil? Por ejemplo, podría querer estimar una probabilidad de la v.a. X, que en general no puedo porque desconozco el parámetro de la distribución. $X \sim f_\theta$

$\hat{P}(X \in A)$ .. si $\hat{P}(X \in A) = \int_A f_{\hat{\theta}}(x)\,dx = g(\hat{\theta})$ $\hat{\theta}$ era el MV de θ

Ejercicio 9

Siguiendo el ejercicio 7, estimar la probabilidad de que se necesiten al menos 2 tiros para observar el primer acierto.

$Y \sim Geom(p)$

$\hat{P}(Y \geq 2) = ?$

$P(Y \geq 2) = 1 - P(Y < 2) = 1 - P(Y = 1) = 1 - p$

Por ppio de inv. de MV, $\hat{P}(Y \geq 2) = 1 - \hat{p} = 1 - 0,4 = \boxed{0,6}$

Bibliografía

Bibliografía

- Notas de estadística. Graciela Boente y Víctor Yohai
- “All of Statistics : A Concise Course in Statistical Inference”,   Larry Wasserman, Springer-Verlag New York Inc.
- “Mathematical statistics with applications”, Mendenhall, W., Scheaffer, R. L., & Wackerly, D. D. Boston: Duxbury Press.