Clase 9

viernes, 31 de agosto de 2018 11:34

Cuando tenemos la muestra $X = x$ y tal que $X \sim f(x/\theta)$ con $\theta \in \Theta$

Sea a una acción que se toma y sea $\mathcal{A}$ el espacio de las posibles acciones

Podemos definir una función de pérdida $L(\theta, a)$ positiva o nula tal que crece cuando crece la distancia de a a θ

Por ej si $\theta \in \mathbb{R}$ hay 2 funciones de pérdida muy comunes

a) Valor absoluto $L(\theta, a) = |a - \theta|$

b) Pérdida cuadrática $L(\theta, a) = (a - \theta)^2$

Son crecientes en el sentido de cuanto más lejos esta a de θ mayor valor tienen y valen 0 en $\theta = a$

La función de pérdida definida como

$$L(\theta, a) = \begin{cases} (a-\theta)^2 & a < \theta \\ 10(a-\theta)^2 & a \geq \theta \end{cases}$$

esta función penaliza más la sobre estimación

$$L(\theta, a) = \frac{(a-\theta)^2}{|\theta|+1}$$ esta función

penaliza más a los errores cuando θ es cercano a 0 que cuando $|\theta|$ es "grande"

¿Cómo juega esta función para determinar la calidad de un estimador?

Sea $\delta(X)$ un estimador de θ

Def 5.7 Definimos la función de riesgo $R(\theta, \delta)$ como

$$R(\theta, \delta) = E_\theta\left(L(\theta, \delta(x))\right)$$

Puedo a partir de esto comparar estimadores. Sea δ_1 y δ_2 tengo $R(\theta, \delta_1)$ y $R(\theta, \delta_2)$. Si

$$R(\theta, \delta_1) < R(\theta, \delta_2) \text{ si } \forall \theta \in \Theta$$

implica que δ_1 es un estimador más adecuado que δ_2.

si las funciones de riesgo "se cruzan" la elección no es fácil.

¿Se pueden "cruzar"? Sí veamos un ejemplo

sean $X_1, \ldots, X_n \sim Ber(p)$

Sea $\hat{p}$ el estimador máximo verosímil

o sea $\hat{p} = \overline{X}$

Si consideramos como función de pérdida la pérdida cuadrática entonces tenemos que

$$R_1(\theta, \hat{p}) = E_p\left[(\hat{p} - p)^2\right] = ECM =$$

$$= Var_p(\hat{p}) = \frac{p(1-p)}{n}$$

Sabemos que el $ECM = Var + (sesgo)^2$

o sea para un un estimador δ

$$R(\theta, \delta) = E_\theta\left[(\delta(x) - \theta)^2\right] =$$

$$= Var_\theta(\delta(x)) + \left(E_\theta(\delta(x)) - \theta\right)^2 =$$

$$= Var_\theta(\delta(x)) + \left(Sesgo_\theta\, \delta(x)\right)^2$$

Volvamos al ejemplo. En el práctico se verá un estimador del estilo

$$\hat{p}_B = \frac{Y+\alpha}{\alpha+\beta+n} \quad \text{con } Y = \sum x_i$$

α, β ctes.

$$ECM = E_p\left[(\hat{p}_B - p)^2\right] = Var(\hat{p}_B) + (E(\hat{p}_B) - p)^2$$

$$= Var\left(\frac{Y+\alpha}{\alpha+\beta+n}\right) + \left(E\left(\frac{Y+\alpha}{\alpha+\beta+n}\right) - p\right)^2$$

$$= \frac{np(1-p)}{(\alpha+\beta+n)^2} + \left(\frac{np+\alpha}{\alpha+\beta+n} - p\right)^2$$

Como $V(\bar{x}) = \frac{p(1-p)}{n}$

$$V\left(\frac{\sum x_i}{n}\right) = \frac{1}{n^2} n V(x_1) = \frac{pq}{n} = \frac{p(1-p)}{n}$$

$$\Rightarrow V(Y) = np(1-p)$$

Ej Ver que se minimiza el ECM cuando

$$\alpha = \beta = \sqrt{n}/2$$

Se obtiene $\hat{p}_B = \frac{Y + \sqrt{n}/2}{n+\sqrt{n}}$

$$\hat{p}_B = \frac{Y+\alpha}{} = \frac{Y+\sqrt{n}/2}{} = \frac{Y+\sqrt{n}/2}{} = \frac{Y+\sqrt{n}/2}{}$$

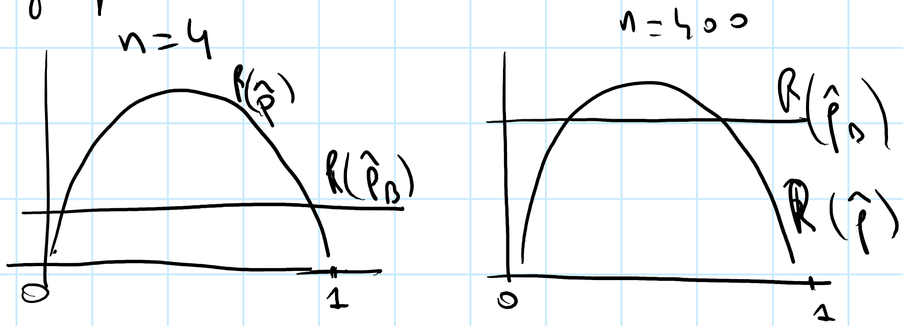
$$\hat p_B = \frac{y+\alpha}{\alpha+\beta+n} = \frac{y+\sqrt{n}/2}{\sqrt{n}/2+\sqrt{n}/2+n} = \frac{y+\sqrt{n}/2}{\frac{2\sqrt{n}}{2}+n} = \frac{y+\sqrt{n}/2}{\sqrt{n}+n}$$

Tenemos $\hat p = \bar X$ y $\hat p_B = \dfrac{y+\sqrt{n}/2}{n+\sqrt{n}}$

Si uno grafica las funciones de Riesgo se ve que cuando tomamos $\hat p_B$ el ECM es cte o sea el riesgo es cte

En tanto el riesgo cuadrático es una parábola

si tomamos por ej $n=4$ y $n=400$ y graficamos el riesgo obtendríamos gráficos del estilo



Se ve que si tengo muestra chica es mejor elegir como estimador $\hat p_B$ y si la muestra es grande elegir $\hat p$

<u>Ejemplo</u> $X_1, \dots, X_n \sim N(\mu, \sigma^2)$

Nuestro problema es estimar σ^2. Me voy a restringir a estimadores del tipo $\delta_b = b S^2$

Sabemos $E(S^2) = \sigma^2$ $\quad Var(S^2) = \frac{2\sigma^4}{n-1}$

$$R(\sigma^2, \delta_b) = ECM(\delta_b) = Var(\delta_b) + (E(\delta_b) - \sigma^2)^2$$

$\downarrow$

uso pérdida cuadrática

$$= Var(b S^2) + (E(b S^2) - \sigma^2)^2 =$$

$$= b^2 \underbrace{Var(S^2)}_{2\sigma^4/n-1} + (b \underbrace{E(S^2)}_{\sigma^2} - \sigma^2)^2 =$$

$$= b^2 \frac{2\sigma^4}{n-1} + ((b-1)\sigma^2)^2 =$$

$$\underbrace{\left[\frac{2b^2}{n-1} + (b-1)^2\right]}_{c_b} \sigma^4$$

El estimador δ_b será mejor que el $\delta_{b'}$

si $c_b < c_{b'}$ o sea voy a minimizar

$\left[\frac{2b^2}{n-1} + (b-1)^2\right]$ para hallar el mejor c_b

$$\left(\frac{2b^2}{n-1} + (b-1)^2\right)' = \frac{4b}{n-1} + 2(b-1) = 0$$

$$4b + 2b(n-1) = 2(n-1)$$

$2b + 2bn = 2(n-1)$

$$b = \frac{n-1}{n+1}$$

Por lo tanto el estimador sería $\delta_b = \frac{n-1}{n+1} S^2$

$$\delta_b = \frac{(\cancel{n-1})}{n+1} \frac{\sum (x_i - \bar{x})^2}{(\cancel{n-1})} = \frac{\sum (x_i - \bar{x})^2}{n+1}$$

Ej Función de pérdida de Stein

Si $X_1, \dots, X_n$ v.a con varianza finita

se define

$$L(\sigma^2, a) = \frac{a}{\sigma^2} - 1 - \log \frac{a}{\sigma^2}$$

Si $a = \sigma^2 \Rightarrow L(\sigma^2, a) = 0$

Si σ^2 es fijo si $a \to 0 \Rightarrow L(\sigma^2, a) = +\infty$

$a \to +\infty \Rightarrow L(\sigma^2, a) = +\infty$

Si considero estimadores $\delta_b = b S^2$ ¿cuál es el estimador óptimo?

$$R(\sigma^2, \delta_b) = E\left(\frac{bS^2}{\sigma^2} - 1 - \log \frac{bS^2}{\sigma^2} \right) =$$

$$= b \underbrace{E(S^2/\sigma^2)}_{1 \text{ pues } S^2 \text{ es insesgado}} - 1 - \log b - E(\log S^2/\sigma^2) =$$

$$= b - 1 - \log b - \underbrace{E(\log S^2/\sigma^2)}_{\text{No depende}}$$

$b - 1 - \lg b \quad E(\delta/\sigma^2) \to$ No depende de b

Minimizo $b-1-\lg b$

$(b-1-\lg b)' = 1 - 1/b = 0 \Rightarrow b = 1$

luego $\boxed{\delta_b = S^2}$

Enfoque Bayesiano

Def 58: Riesgo Bayesiano

Definimos el riesgo bayesiano como

$$\int_{\Theta} R(\theta,\delta)\,\pi(\theta)\,d\theta$$

$\pi(\theta)$ distribución a-priori de θ

Regla de Bayes con una a priori π

$X \sim f(x/\theta) \quad \theta \sim \pi(\theta)$

$$\int_{\Theta} R(\theta,\delta)\,\pi(\theta)\,d\theta = \int_{\Theta}\left[\int_{X} L(\theta,\delta(x))\,f(x/\theta)\,dx\right]\pi(\theta)\,d\theta$$

$$\int_{\Theta} R(\theta,\delta)\,\pi(\theta)\,d\theta = \int_{\Theta}\underbrace{\int_{X} L(\theta,\delta(x))\,f(x/\theta)\,dx}_{\text{la esperanza de la función de pérdida}}\,d\theta$$

a-posteriori

$$f(x/\theta)\,\pi(\theta) = \pi(\theta/x)\,m(x) \rightarrow \text{marginal de } X$$

$$= \int_{\Theta}\left[\int_{X} L(\theta,\delta(x))\,\pi(\theta/x)\,m(x)\,dx\right] d\theta = r_\delta$$

$$m(x) = \int_{\Theta} f(x/\theta)\,\pi(\theta)\,d\theta$$

$$r_\delta = \int_{X}\underbrace{\left[\int_{\Theta} L(\theta,\delta(x))\,\pi(\theta/x)\,d\theta\right]}_{\text{depende solo de } X} m(x)\,dx$$

Valor esperado de la pérdida a-posteriori

luego minimizar la pérdida esperada a-posteriori minimiza el riesgo de Bayes

Ejemplos: pérdida cuadrática $\Rightarrow$ el valor esperado de la pérdida a posteriori es

$$\int_{\Theta} (\theta-a)^2\,\pi(\theta/x)\,d\theta = E\left((\theta-a)^2/X=x\right)$$

Θ

se minimiza cuando $a = E(\theta/x)$

Si uso la pérdida con el valor absoluto

$$\Rightarrow \int_{\Theta} |\theta - a| \, \pi(\theta/x) \, dx = E\left(|\theta - a| / X = x\right)$$

y esto se minimiza en la $Med(\theta/x)$