Estadística Bayesiana: Clase 8

Juan Carlos Correa

23 de marzo de 2021

Inferencia Bayesiana

Estimación Puntual

- Dada una distribución sobre un parámetro particular, digamos θ, requerimos seleccionar un mecanismo para escoger un "buen" un estimador $\hat{\theta}$.
- Supongamos que θ_0 es el verdadero parámetro, desconocido. Sea d nuestra adivinanza de este valor.
- Debemos de alguna forma medir el error que cometemos (digamos que esto puede ser una multa o un pago) al adivinar a θ_0 mediante d.
- Esto puede ser medido por $(d-\theta_0)^2$ o por $|d-\theta_0|$ o mediante alguna otra función.

Un problema estadístico puede resumirse como (S, Ω, D, L), donde

S: Es el espacio muestral de un experimento relevante que tiene asociada una variable aleatoria X cuya distribución de probabilidad está parametrizada por un elemento de Ω.

Ω: Espacio parametral (en un sentido amplio)

D: Un espacio de decisiones

L: Una función de pérdida.

Una vez un problema estadístico ha sido especificado, el problema de inferencia estadística es seleccionar un procedimiento (estadístico), a veces llamado una función de decisión, que nos describe la forma de tomar una decisión una vez un resultado muestral ha sido obtenido.

- Una función de decisión o procedimiento estadístico es una función o estadístico d que mapea de S a D.
- Sea D un espacio arbitrario de decisiones. Una función no negativa L que mapea de $\Omega \times D$ a $\mathcal{R}$ es llamada una función de pérdida.
- El valor esperado de $L(\theta, d(X))$ cuando θ es el verdadero valor es llamada la *función de riesgo*

$$R(\theta, d) = E_\theta \left[L\left(\theta, d(X)\right)\right] = \int L\left(\theta, d(x)\right)\, dP_\theta(x)$$

Función de Pérdida Cuadrática:

$$L(d,\theta) = (d-\theta)^2$$

Miremos el riesgo para esta función de pérdida. Sea

$$b = E_{\xi(\theta|x)}(\theta) = \int \theta\, \xi\left(\theta|x\right)\, d\theta$$

el promedio de la distribución aposteriori. Entonces

$$
\begin{aligned}
E\left[L(d,\theta)\right] &= \int L(a,\theta)\,\xi\left(\theta|x\right)\,d\theta \\
&= \int (a-b+b-\theta)^2\,\xi\left(\theta|x\right)\,d\theta \\
&= (a-b)^2 + \int (b-\theta)^2\,\xi\left(\theta|x\right)\,d\theta \\
&\geq \int (b-\theta)^2\,\xi\left(\theta|x\right)\,d\theta
\end{aligned}
$$
,

para cualquier valor de d. La desigualdad anterior se convierte en igualdad cuando $d = b$.
El estimador bayesiano bajo una función de pérdida cuadrática es la media de la distribución posterior.

Función de Pérdida Error Absoluto:

$$L(d, \theta) = |d - \theta|$$

El riesgo es minimizado tomando d como la mediana de la distribución posterior, digamos d^*.
O sea, *la mediana es el estimador bayesiano cuando la función de pérdida es el valor absoluto.*

$$|\theta - d| - |\theta - d^*| = \begin{cases} d^* - d & \text{si } \theta \geq d, \\ d + d^* - 2\theta & \text{si } d^* < \theta < d, \\ d - d^* & \text{si } \theta \leq d^*. \end{cases}$$

Ya que $(d + d^* - 2\theta) > (d^* - d)$ cuando $d^* < \theta < d$, entonces el siguiente resultado se consigue

$$E(|\theta - d| - |\theta - d^*|) \geq (d^* - d)P(\theta \geq d) + (d^* - d)P(d^* < \theta < d)$$
$$+(d - d^*)P(\theta \leq d^*)$$
$$= (d - d^*)\left[P(\theta \leq d^*) - P(\theta > d^*)\right] \geq 0$$

Esta última desigualdad sigue del hecho que d^* es la mediana de la distribución de θ. La primera desigualdad en este conjunto de ecuaciones será una igualdad si, y solo si, $P(d^* < \theta < d) = 0$. La desigualdad final será una igualdad si, y solo si,

$$P(\theta \leq d^*) = P(\theta > d^*) = \frac{1}{2}.$$

Estas condiciones implican que d es también una mediana. Por lo tanto, $E(|\theta - d|) \geq E(|\theta - d^*|)$, y la igualdad se cumple si, y solo si, d es también mediana.
Una prueba similar puede hacerse si $d < d^*$.

Función de Pérdida Escalonada:

$$\begin{aligned} L(d,\theta) &= 0 \text{ si } |d-\theta| \leq \delta \\ &= 1 \text{ si } |d-\theta| > \delta \end{aligned}$$

donde δ es un número predeterminado, usualmente pequeño.

$$
\begin{aligned}
E\left[L(d,\theta)\right] &= \int_{\Theta} I\left(|d-\theta|>\delta\right)\xi\left(\theta|x\right)\,d\theta \\
&= \int_{\Theta} I\left(1-\left(|d-\theta|\leq\delta\right)\right)\xi\left(\theta|x\right)\,d\theta \\
&= 1-\int_{d-\delta}^{d+\delta}\xi\left(\theta|x\right)\,d\theta \\
&\approx 1-2\delta\xi\left(d|x\right)
\end{aligned}
$$

Para minimizar el riesgo es necesario maximizar $\xi\left(d|x\right)$ con respecto a d y el estimador bayesiano es el maximizador. Por lo tanto, el estimador bayesiano será el que maximiza la posterior, esto es, el valor modal. Este estimador es llamado el estimador máximo-aposteriori (MAP).

Una estimación que puede ser utilizada en una o más dimensiones, especialmente cuando la función de pérdida no ha sido definida explícitamente, es el valor del parámetro en el cual se maximiza la distribución posterior. Para cualquier observación de x, sea $\psi(\cdot|x)$ que denota la distribución posterior de W en el espacio parametral Ω. Sea $\hat{w}(x)$ el valor de w que satisface la relación

Ejemplo: Poisson

Sea $y_1, \cdots y_n$ una muestra aleatoria de una $Poisson(\lambda)$. Supongamos también que la apriori es una $Gamma(1,1)$. Por lo tanto la aposterior será $Gamma(1+\sum_{i=1}^n y_i, n+1)$.
El estimador bayesiano para λ

- bajo la función de pérdida cuadrática es

$$\hat{\lambda} = \frac{1+\sum_{i=1}^n y_i}{n+1}$$

- bajo la función de pérdida escalonada

$$\hat{\lambda} = \frac{\alpha^* - 1}{\beta^*} = \frac{\sum_{i=1}^n y_i}{n+1} \text{ si } \alpha^* \geq 1$$

La siguiente función en R calcula los tres estimadores, bajo el supuesto de una aprori $Gamma(\alpha_0, \beta_0)$:

```
calcula.estimadores.poisson<-function(alfa0,beta0,x,n=lenght(x))
{
alfa1<-alfa0+sum(x)
beta1<-beta0+n
estimador.fpc<-alfa1/beta1
estimador.fpa<-qgamma(0.5,alfa1,beta1)
estimador.fpe<-(alfa1-1)/beta1
list(estimador.fpc=estimador.fpc,
      estimador.fpa=estimador.fpa,
      estimador.fpe=estimador.fpe)
}
```

La utilización será

```
> calcula.estimadores.poisson(1,1,16,n=4)

$estimador.fpc
[1] 3.4

$estimador.fpa
[1] 3.333571

$estimador.fpe
[1] 3.2
```