Tema

1

Introducción a la Inferencia Bayesiana

Índice

1. Introducción

Sea X una variable aleatoria con función de distribución $F(\cdot;\theta)$ donde θ es un parámetro desconocido y fijo que pertenece a un subconjunto de $\mathbb{R}^k$, $(\theta \in \Theta \subset \mathbb{R}^k)$.

Ejemplo 1.1. *Sea X una variable aleatoria que sigue una distribución exponencial negativa de parámetro $\theta > 0, (X \rightsquigarrow Exp(\theta))$. En este caso se tiene que $F(x;\theta) = 1 - e^{-\theta x}$ con $x > 0$ y $\theta > 0$. Por tanto, es obvio que $\Theta = (0, \infty)$. Se sabe que el parámetro θ tiene una interpretación muy sencilla en esta distribución ya que $E(X) = 1/\theta$ y $Var(X) = \mathsf{v} = 1/\theta^2$. En la Figura 1 se puede observar claramente el efecto al variar el parámetro sobre la función de densidad exponencial.*

En la Inferencia Estadística Clásica se ha estudiado como realizar estimaciones puntuales, estimaciones por intervalos de confianza e incluso realizar contrastes de hipótesis

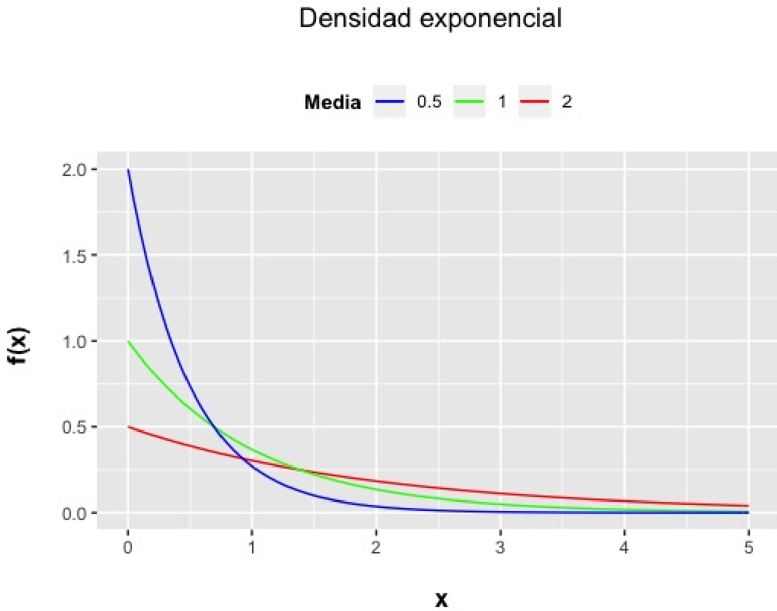

Figura 1: Densidades exponenciales

acerca de los parámetros de una distribución a partir de una muestra aleatoria de tamaño n de la variable aleatoria observada. En concreto, en el caso del Ejemplo 1.1 se vió que el estimador de máxima verosimilitud del parámetro θ de la exponencial era:

$$\hat{\theta}_{MV} = \left(\overline{X}\right)^{-1}.$$

Podemos decir que el punto de partida de la inferencia bayesiana es la relajación de las condiciones impuestas inicialmente sobre el parámetro. En la estadística frecuentista se asume que es desconocido y fijo. La cuestión es que si se realiza la estimación de un parámetro desconocido por qué no cuantificar la incertidumbre acerca de dicho desconocimiento a partir de la información que aporta la muestra. Este planteamiento obviaría la condición de ser un valor fijo impuesto en la inferencia clásica pero a su vez conlleva una complejidad que sería la cuantificación de esa incertidumbre mediante una función de probabilidad. Esto conlleva al tratamiento de los parámetros distribucionales como variables aleatorias.

Este planteamiento no significa que la Estadística Bayesiana mejora los resultados aportados por la Estadística Clásica ya que ambos métodos presentas ventajas e inconvenientes que se irán analizando a lo largo de este curso en la medida de lo posible.

2. Teorema de Bayes

La base del cálculo de la estadística bayesiana radica en el conocido Teorema de Bayes.

Teorema 2.1. *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y $\{A_1, \cdots, A_n\}$ un sistema completo de sucesos con probabilidades no nulas y B otro suceso no nulo se tiene que*

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\displaystyle\sum_{i=1}^{n} P(A_k)P(B|A_k)} \text{ para } k = 1, \cdots, n.$$

Este teorema permite cuantificar la incertidumbre en la relación *causa-efecto*. Es decir, lo común en la observaciones empíricas en Ciencia es obtener los efectos de determinadas causas. Una misma causa puede provocar diferentes efectos, y viceversa, un mismo efecto puede ser provocado por diferentes causas. En definitiva, estaríamos interesados en calcular:

$$P(Causa_i|Efecto_j).$$

El cálculo de la probabilidad anterior no es tan sencillo como parece ya que en principio la información conocida o condicionada ($Efecto_j$) sucede cronológicamente después de la $Causa_i$. El teorema de Bayes permite invertir esa condicionada. Observando el Teorema 2.1 se ve claramente que para poder aplicar dicha fórmula necesitamos el conocimiento de unas probabilidades. En concreto, es necesario saber el valor de $P(A_k)$ para todo valor de k. Es decir, necesitamos conocer las probabilidades de las causas inicialmente. A estas probabilidades se les conoce con el nombre de *Probabilidad a priori*. Y es aquí donde radica la mayor crítica hacia la Estadística Bayesina ya que hasta cierto punto la obtención de estas probabilidades se hacen de manera *subjetiva*. No obstante, existen numerosos métodos en la literatura que permiten la construcción de distribuciones a priori a partir de la información que se disponga de experimentos previos.

En definitiva, la apliación de la fórmula de Bayes nos permite modificar las probabilidades a priori ($P(A_k)$) a partir del conocimiento o información aportada por la experimentación en otras probabilidades ($P(A_k|B)$) que a partir de ahora las denominaremos *probabilidades a posteriori*.

Lo interesante de la Estadística Bayesiana es que el uso del Teorema de Bayes permite ir modificando de manera recursiva las probabilidades a posteriori a medida que vamos obteniendo información del experimento en cuestión.

3. Bayes para variables aleatorias

En esta sección vamos a formular el Teorema 2.1 para variables aleatorias. Aunque se puede plantear de manera general, lo haremos distinguiendo dos casos: v.a. discretas y v.a. continuas.

3.1. Variables Aleatorias Discretas

Sean X e Y dos variables aleatorias discretas finitas con $Img(X) = \{x_1, \cdots, x_m\}$ e $Img(Y) = \{y_1, \cdots, y_n\}$. Supongamos que la v.a. X origina la partición de Ω dada por $A_1, \ldots, A_m$ y la v.a. Y la partición $B_1, \ldots, B_n$. Entonces, aplicando el teorema 2.1 se obtiene que

$$P(Y = y_i|X = x_j) = \frac{P(Y = y_i)P(X = x_j|Y = y_i)}{\sum_{i=1}^{n} P(Y = y_i)P(X = x_j|Y = y_i)},$$

para todo $i = 1, \cdots, n$ y para cada $j = 1, \cdots, m$.

Hay que tener en cuenta que se ha empleado la siguiente notación por simplificar:

$$P(A_i) = P(\omega \in \Omega : X(\omega) = x_i) = P(X = x_i).$$

En este caso la distribución de Y se le conoce como distribución a priori, y la distribución de $Y|X = x_j$ como la distribución a posteriori de Y dado x_j. El numerador, que consiste en el producto de la a priori por la verosimilitud, es proporcional a la distribución a posteriori ya que el denominador sólo juega el papel de constante normalizadora.

3.2. Variables Aleatorias Continuas

Sea (X, Y) una variable aleatoria bidimensional absolutamente continua con función de densidad $f(x, y)$, h la densidad marginal de X y g la densidad marginal de Y. Se tiene que

$$\begin{aligned} f(y|x) &= \frac{f(x, y)}{h(x)} \\ &= \frac{g(y)f(x|y)}{h(x)} \\ &= \frac{g(y)f(x|y)}{\int f(x, y)dy} \\ &= \frac{g(y)f(x|y)}{\int g(y)f(x|y)dy}. \end{aligned}$$

A partir de ahora, y sin pérdida de generalidad, adoptaremos la formulación en el caso absolutamente continuo teniendo siempre en cuenta la versión discreta en el caso de que fuese necesario.

Ver Práctica-R-1

Ejercicios de autoaprendizaje.

1. Sea X una variable aleatoria discrete finita e Y una variable aleatoria discreta numerable infinita. Enunciar el teorema de Bayes para $X|Y = y_j$ y para $Y|X = x_i$.

2. Supongamos que $X|Y = y \sim Poi(y)$, y la distribución a priori de Y es una gamma de parámetros (a, b). Obtener la distribución a posteriori de Y.

3. Supongamos que $X|Y = y \sim Ge(y)$ y que Y tiene una densidad a priori proporcional a $yI_{(0,1)}(y)$. Obtener la distribución a posteriori de Y.