TÉCNICAS BAYESIANAS

Juan Carlos Flores López y Marisol Paredes

Elementos de Teoría Bayesiana

Se asume las observaciones $x = (x_1,...,x_n)$ han sido generados de una distribución de probabilidad $f(x/\theta)$, tal que el parámetro θ es desconocido y la función f es conocida. Este modelo es representado por $x\sim f(x/\theta)$> donde x puede ser un vector, lo mismo que θ.

Distribución a priori

Consideremos un problema de inferencia estadística en el que se van a seleccionar observaciones de una distribución cuya función de probabilidad es $f(x/\theta)$, donde θ es un parámetro de valor desconocido. Se supone que el valor desconocido del parámetro θ debe pertenecer a un espacio paramétrico Ω . Lo que se pretende es el de intentar determinar donde es probable que se encuentre el verdadero de θ en el espacio paramétrico Ω , partiendo de las observaciones de la densidad $f(x/\theta)$. E Esta distribución se denomina *distribución a priori de* θ, que se denota por $\pi(\theta)$

Distribución posteriori

Supongamos que $x = (x_1,...,x_n)$ constituyen una muestra, donde los x_i son independientes e idénticamente distribuidas (i.i.d.), de una distribución cuya densidad es $f(x/\theta)$. Luego:

$$f(x/\theta) = \subseteq_{i=1}^{n} f(x_i/\theta) = L(\theta/x)$$

Cuando la función de densidad conjunta $f(x/\theta)$ de las observaciones, donde los x_i son i.i.d, se considera con una función de θ, para valores dados de $x_1,...,x_n$, se llama de verosimilitud y la demostraremos por $L(\theta/x)$. Asumimos, en resumen, para la distribución muestral $f(x/\theta)$, una distribución inicial en θ y que $\pi(\theta)$ esta disponible. Dado lo anterior, podemos construir diferentes distribuciones, llamadas:

(a) la distribución conjunta de (x,θ)

$$f(x/\theta) = f(x/\theta)\pi(\theta)$$

(b) la distribución marginal de x, también llamada la distribución predictiva

$$m(x) = f(x/\theta)\mathrm{d}(\theta)$$

$$= (x/\theta)\pi(\theta)\mathrm{d}(\theta)$$

(c) la distribución posterior de θ, obtenida por la formula de Bayes

$$\pi\left(\frac{\theta}{x}\right) = \frac{f(x/\theta)\pi(\vartheta)}{f(x/\theta)\pi(\theta)d\theta} = \frac{f(x/\theta)\pi(\vartheta)}{m(x)}$$

notemos que $\pi(\theta/x)$ es proporcional a la distribución de x condicional en θ, es decir la función de verosimilitud, multiplicada por la distribución a priori de θ, ya que el denominador de $m(x)$ no depende de θ, puede ser considerado como una constante, así tenemos:

$$\pi(\theta/x)\alpha f(x/\theta)\pi(\theta)$$

esta simple expresión encierra el núcleo de la inferencia Bayesiana.

Principio de in varianza de Jeffreys

Esta introducción introducida por Jeffreys, esta basada en considerar transformaciones uno a uno del parámetro θ, es decir una función del parámetro θ llamada ϕ: $\phi = h(\theta)$. Si realizamos transformaciones de variables, la densidad a priori $\pi(\theta)$ e s equivalente, en términos de expresar la misma confianza, a la siguiente densidad a priori en ϕ:

$$\pi(\phi) = \pi(\theta)\left|\frac{d\theta}{d\phi}\right| = \pi(\theta)|h'(\theta)|^{-1}$$

Definición.- Una familia *F* de distribuciones de probabilidad en Θ se dice que es conjugada (o cerrada bajo muestreo)si, para cada π ∟ *F*, la distribución posteriori $\pi(\theta/x)$ también pertenece a *F* .

Modelo normal con varianza conocida

Dada la importancia de la distribución normal en el desarrollo de la inferencia, es necesario mostrar desde el punto de vista bayesiano. Dado $x = (x_1, \ldots, x_n)$ una muestra, (i.i.d.)de una distribución normal con media θ desconocida, pero con varianza σ^2 conocida, la distribución conjugada de θ es la distribución normal. Luego sea $\pi(\theta) \sim N(\mu_0, \tau_0^2)$ donde μ_0, τ_0^2 están dados. Luego la distribución posterior también es normal, y tiene como media y varianza

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{x}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \quad y \quad \frac{1}{\tau_n^2} = -\frac{1}{\tau_0^2} + \frac{n}{\sigma^2} \; {}^{-1}\surd$$

por otro lado la distribución normal multivariante esta dado por

$$f(x) = (2\pi)^{-2k/2}|\Sigma|^{-\frac{1}{2}}\exp(-\frac{1}{2}(x-\theta)'\Sigma^{-1}(x-\theta))$$

y la función de verosimilitud para una muestra de n observaciones i.i.d. de x_i, $x_{i1}, \ldots, x_{in}$, $i = 1,2,\ldots,k$, es:

$$f(x/\theta, \Sigma) = \underset{r=1}{\overset{n}{\subseteq}} (2\pi)^{-\frac{k}{2}}|\Sigma|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(x-\theta)'\Sigma^{-1}(x-\theta)\right)$$

$$=(2\pi)^{-\frac{nk}{2}}|\Sigma|^{-\frac{n}{2}}\exp\left(-\frac{1}{2}\sum_{r=1}^{n}(x_r-\theta)'\Sigma^{-1}(x_r-\theta)\right)\alpha|\Sigma|^{-\frac{n}{2}}\exp\left(-\frac{1}{2}\sum_{r=1}^{n}(x_r-\theta)'\Sigma^{-1}(x_r-\theta)\right)$$

como $\sum_{r=1}^{n}(x_r-\theta)'\Sigma^{-1}(x_r-\theta)$ es un escalar, tomando la traza y por propiedad de esta ultima, lo anterior se convierte en:

$$= |\Sigma|^{-\frac{n}{2}}\exp(-\frac{1}{2}tr\,(\Sigma^{-1}S))$$

donde S_{kxk} es la matriz de "suma de cuadrados" relativo a θ.

$$S_{kxk} = (x_r - 0)(x_r - 0)'$$

Definición.- un modelo jerárquico de Bayes es un modelo estocástico Bayesiano ($f(x/\theta)$, $\pi(\theta)$), donde la distribución a priori $\pi(\theta)$ esta descompuesta en distribuciones condicionales:

$$\pi_1(\theta/\theta_1), \pi_2(\theta_1/\theta_2), \ldots\ldots\ldots, \pi_n(\theta_{n-1}/\theta_n)$$

y una distribución marginal $\pi_{n+a}(\theta_n)$ tal que:

$$\pi(\theta) = {}_{\Theta 1x\ldots x\Theta n}\, \pi_1(\theta/\theta_1), \pi_2(\theta_1/\theta_2), \ldots\ldots\ldots, \pi_{n'+}(\theta)\, d\theta_1 \ldots\ldots d\theta_{n+1}$$

los parámetros θ_1 son llamados hiperparametros de nivel i.