

Breve Introducción a la Inferencia Bayesiana

Joshua Kunst

jbkunst@gmail.com

28 de Junio, 2007

Bien ya sabemos, los académicos y estudiantes estadísticos, que desde hace tiempo a tomado fuerza el punto de vista Bayesiano para realizar inferencia. Lo que pretendo con este documento es dar una muy breve introducción a este tema para quienes les motiva el tema y no lo conocen.

El procedimiento a grandes rasgos para realizar inferencia sobre una cantidad desconocida θ (perteneciente a un subconjunto de los números reales) es proponer, antes de cierto experimento, un conocimiento previo (que puede ser: desconocimiento total previo!) a través de una distribución *a priori* en θ , el parámetro, $\pi(\theta)$. Luego del experimento se procede a realizar inferencia a través de la distribución *a posteriori* $\pi(\theta|x)$. Algo más formal se ve en la siguiente definición:

Definición: Un *Modelo Estadístico Bayesiano* está hecho por un modelo paramétrico:

$$x \sim f(x|\theta) ; \theta \sim \pi(\theta)$$

Esto es que se obtienen datos u observaciones según cierta distribución $f(\cdot)$ dependiente del parámetro en cuestión. También, como se dijo antes, debemos proponer una distribución a priori la cual obviamente debe tener un soporte de acuerdo a la cantidad de interés.

Luego, dado el modelo y los datos observados, la inferencia se hace completamente sobre la distribución del parámetro condicionado en el vector de observaciones, la cual se obtiene a partir del teorema de Bayes, llamada distribución a posteriori:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}$$

Cabe notar que:

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta),$$

puesto que la integral del denominador no depende de θ , es solamente una constante de normalización para la distribución a posteriori.

En términos estadísticos más técnicos, el Teorema de Bayes actualiza la información del parámetro a través de las observaciones obtenidas (experimento). Lo que se puede hacer debido a que los parámetros y observaciones tienen, conceptualmente, el mismo nivel; ambos tienen distribuciones.

Aplicación: Daremos como aplicación el ejemplo *Histórico* de Bayes de 1973. El experimento es como sigue: Se lanza una bola de billar W sobre una recta de largo 1, con probabilidad uniforme en parar en cualquier parte. Sea θ el punto de parada de la bola W . Una segunda bola O es tirada n veces bajo las mismas condiciones de la bola W . Sean X el número de veces que la bola O cae a la derecha de W .

Ahora la pregunta en cuestión es: ¿Dado X , que inferencia se puede hacer respecto de θ ? La respuesta es sí, ya que se tiene un modelo como el de la definición. Luego el problema se reduce a calcular la distribución a posteriori de θ . Notemos que particularmente el modelo es:

$$X|\theta \sim \text{Bin}(n, \theta) \ ; \ \theta \sim \mathbb{U}(0, 1) \equiv \pi(\theta)$$

Luego, la distribución posteriori queda:

$$\pi(\theta|x) = \frac{\binom{n}{x} \theta^x (1 - \theta)^{n-x}}{\int_0^1 \binom{n}{x} \lambda^x (1 - \lambda)^{n-x} d\lambda}; \ \theta \in [0, 1]$$

Ahora debemos notar que los coeficientes binomiales se cancelan y que:

$$\int_0^1 \lambda^x (1 - \lambda)^{n-x} d\lambda = \frac{\Gamma(x + 1) \Gamma(n - x + 1)}{\Gamma(x + 1 + (n - x + 1))}$$

Y por tanto:

$$\theta|X = x \sim \text{Beta}(x + 1, n - x + 1)$$

Y dada la distribución a posteriori, como se dijo anteriormente, ahora se puede realizar todo tipo de inferencias tales como :

$$\Pr\{a < \theta < b|x\} \ ; \ \mathbb{E}[g(\theta)]$$

Que se traducen en intervalos de probabilidad y esperanzas de cierta función de interés sobre θ .

Nota: Cabe señalar que este particular se habla de una familia conjugada puesto que la distribución posteriori pertenece a la misma que la priori. Esto es por que la distribución uniforme en el intervalo $[0,1]$ es un caso particular de una distribución Beta (Beta de parámetros 1 y 1). De hecho, el modelo de ejemplo, Binomial-Beta, es varias veces utilizado ya que la distribución a posteriori es conocida. Sin embargo, esto no ocurre la mayoría de las veces puesto que uno puede reflejar el conocimiento previo con otras distribuciones. Póngase en el ejemplo y piense si hubieramos puesto como densidad a priori una distrbución Triangular en el intervalo en cuestión.