Estadística Bayesiana

Stefano Molina Martínez

Septiembre 2016

Introducción

Una intuición inicial

La estadística bayesiana hereda su nombre del teorema más famoso del estadístico Thomas Bayes, el cual fue planteado para describir la probabilidad de ocurrencia de un evento sujeta a condiciones relacionadas al evento. El paradigma Bayesiano se basa en el Teorema de Bayes para llevar a cabo un aprendizaje adicional sobre los conocimientos que se tienen sobre un evento, la intuición es la que sigue.

Definición. Dados A y B dos eventos tales que $P(B) > 0$, la probabilidad condicional A dado B se define como

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Suponer que se tiene **H**, una hipótesis que quiere ser probada y se posee evidencia para poner a prueba la veracidad de ella, la cual se denota como **E**, por la definición de probabilidad condicional, se tiene que

$$P(E \cap H) = P(E|H)P(H) = P(H|E)P(E)$$

Si la evidencia es verídica, es decir $P(E) > 0$, se obtiene la fórmula del Teorema de Bayes:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

Inferencia Bayesiana

En la práctica se utilizan las variables θ, una característica de la población en cuestión que sustituye a la hipótesis **H**, la cual es considerada variable aleatoria por ser desconocida y $\bar{x} = \{x_1, x_2, ..., x_n\}$, un conjunto de observaciones de la población, que sustituyen a la evidencia **E**, sobre la cual se busca hacer la inferencia.

En el sentido Bayesiano, los términos de esta expresión son conocidos como sigue:

$$f(\theta|\bar{x}) = \frac{f(\bar{x}|\theta)f(\theta)}{f(\bar{x})}$$

- $f(\bar{x})$ es llamada distribución *a priori*, el conocimiento muchas veces empírico que se tiene del evento antes de conocer la validez de la hipótesis.
- $f(\bar{x}|\theta)$ es la verosimilitud $f(\bar{x}|\theta) = L(\theta;\bar{x})$, asumiendo que $\bar{x}$ sea fijo dado θ.
- $f(\theta)$ es la constante de propocionalidad.
- $f(\theta|\bar{x})$ se llama distribución posterior o *a posteriori*, es la probabilidad de la hipótesis dada la evidencia $\bar{x}$. También es conocido como *kernel* de la distribución, el cual se dice que es proporcional a la verosimilitud por la *a priori*:

$$f(\theta|\bar{x}) \propto L(\theta;\bar{x})f(\theta)$$

Distribuciones *a priori*

Una distribución *a priori* para una variable θ posee todas las características de la distribución de una variable aleatoria con la excepción de que θ no es observable. Debido a lo anterior, la distribución de la variable θ está construida con los conocimientos o suposiciones que se tienen acerca de su verdadero valor. Las distribuciones *a priori* de una misma variable pueden variar según la cantidad de información que se posea sobre de ella o las fuentes que se consulten para obtenerla, por lo que cada una de éstas influirá en la forma y valores de la distribución posterior obtenida después del proceso de inferencia. Existen dos métodos populares para determinar las distribuciones *a priori*, uno es el de distribuciones conjugadas y el otro es el de distribuciones iniciales no informativas.

Usando distribuciones conjugadas se puede añadir información adicional a un modelo, el método consiste en basar la distribución *a priori* en la forma de la verosimilitud. Se dice que se tiene un par conjugado cuando la distribución *a priori* y la posterior pertecener a la misma familia.

Existen ocasiones en las que es poco conveniente usar las creencias o información *a priori* debido a un posible sesgo en éstas que podrían aportar a la distribución final o la forma distribucional no es conjugable, por lo que en situaciones como éstas se adoptan distribuciones *a priori* que aportan información mínima o nula y reducen el sesgo subjetivo, se conocen como distribuciones no informativas.