Estadística Bayesiana

Introducción

Ms Carlos López de Castilla Vásquez

Universidad Nacional Agraria La Molina

2017-1

Introducción

- La estadística Bayesiana le debe su nombre al trabajo pionero del reverendo Thomas Bayes titulado: *An Essay towards solving a Problem in the Doctrine of Chances* publicado en 1764.
- El artículo fue enviado a la Real Sociedad de Londres por Richard Price en 1763 quién escribió: *Yo ahora le mando un ensayo que he encontrado entre los papeles de nuestro fallecido amigo Thomas Bayes y el cual, en mi opinión, tiene un gran mérito y bien merece ser preservado* ...
- Aunque la obra de Thomas Bayes data ya de más de dos siglos la estadística Bayesiana es relativamente nueva y actualmente ostenta un gran desarrollo aunque no ajena también a grandes controversias.

Introducción

- El marco teórico sobre el cual se desarrolla la inferencia Bayesiana es idéntico al de la teoría clásica.
- Se tiene un parámetro θ sobre el cual se desea realizar el proceso de inferencia y un modelo de probabilidad $f(y|\theta)$.
- Un punto importante en la definición clásica de inferencia es que el parámetro θ es tratado como *constante* lo cual conduce a ciertos problemas de interpretación.
- Sin embargo la diferencia fundamental entre la teoría clásica y la Bayesiana está en que θ es considerado una *variable aleatoria*.

Introducción

- El proceso de inferencia Bayesiana se basa en $f(\theta|y)$ conocida como la *distribución posterior* para el parámetro luego de observados los datos.
- Se requiere especificar una *distribución a priori* $f(\theta)$ que representa el conocimiento que se tiene sobre θ antes de observar los datos.
- La distribución a priori constituye el centro del pensamiento Bayesiano y, dependiendo de si se es un defensor o un opositor a esta metodología, su principal ventaja sobre la teoría clásica o su mayor vulnerabilidad.

Teorema de Bayes

- Sean A y B dos eventos con $\Pr(A) > 0$, entonces:

$$\Pr(B|A) = \frac{\Pr(A|B)\Pr(B)}{\Pr(A)}$$

- Si se tiene k eventos $B_1, \cdots, B_k$, los cuales constituyen una partición del espacio muestral Ω:

$$\Pr(B_i|A) = \frac{\Pr(A|B_i)\Pr(B_i)}{\Pr(A)} = \frac{\Pr(A|B_i)\Pr(B_i)}{\sum_{j=1}^{k}\Pr(A|B_j)\Pr(B_j)}$$

donde $\Pr(A)$ se le conoce como *probabilidad total*.

Ejemplo: Transmición de señales

- Un transmisor está enviando un mensaje mediante un código binario con señales de 0 y 1.
- Cada señal transmitida debe pasar por dos relevadores antes de llegar finalmente al receptor.
- En cada relevador, la probabilidad de que la señal enviada sea diferente de la señal recibida (inversión de la señal) es de 0,10. Los relevadores funcionan independientemente uno de otro.
- El 60 % de todas las señales enviadas desde el emisor son 1.
- Si un 1 es recibido por el receptor ¿cuál es la probabilidad de que se haya sido enviado realmente un 1?

Estadística Bayesiana

- Desde el punto de vista Bayesiano, el objetivo es relacionar probabilísticamente al parámetro θ con los datos y.
- El teorema de Bayes, definido anteriormente usando eventos, puede presentarse en términos de funciones de probabilidad:

$$f\left(\theta|y\right) = \frac{f\left(y|\theta\right) f\left(\theta\right)}{f\left(y\right)}$$

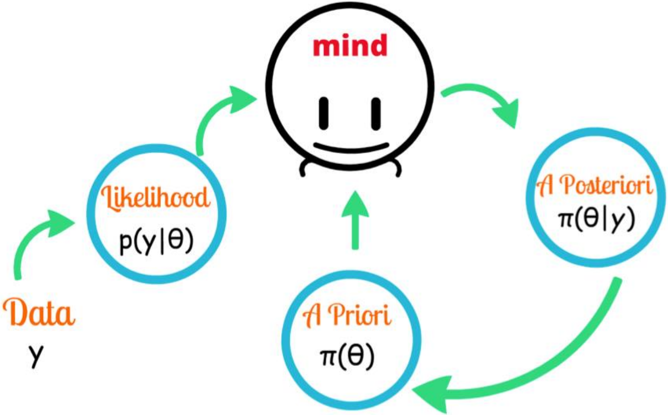
donde $f\left(y\right)$ se puede obtener de la siguiente manera:

$$f\left(y\right) = \begin{cases} \sum_{\theta} f\left(y|\theta\right) f\left(\theta\right) & \text{si } \theta \text{ es discreta} \\ \int_{\theta} f\left(y|\theta\right) f\left(\theta\right) d\theta & \text{si } \theta \text{ es continua} \end{cases}$$

Estadística Bayesiana

- La función $f(\theta)$ representa lo que es conocido de θ antes de observar los datos y es llamada la *distribución a priori de* θ.
- La función $f(\theta|y)$ representa lo que se conoce de θ después de observar los datos y es llamada la *distribución posterior de* θ *dado* y.
- La función $f(y|\theta)$ es la *distribución fundamental* e incorpora al modelo la información muestral proporcionada por los datos.
- Luego de observar los datos a $f(y|\theta)$ también se le denomina *función de verosimilitud* y se denota por $l(\theta|y_1, \cdots, y_n)$.

Estadística Bayesiana



Estadística Bayesiana

- Se puede presentar $f(\theta|y)$ de forma equivalente omitiendo la distribución marginal $f(y)$ ya que no depende de θ y puede ser considerada como una constante:

$$f(\theta|y) \propto f(y|\theta) f(\theta)$$

- Lo que se obtiene al lado derecho se conoce como la *distribución posterior no normalizada* y constituye el *núcleo* de la distribución posterior.
- El núcleo permite muchas veces determinar la distribución posterior sin necesidad de hallar previamente $f(y)$.

Ejemplos

- Suponga que la distribución fundamental es:

$$y|\theta \sim \mathcal{E}(\theta)$$

  y que $\theta \sim \mathcal{G}(\alpha, \beta)$. Hallar la distribución marginal $f(y)$ y la distribución posterior para θ.

- Suponga que se tiene una muestra:

$$y_1, \cdots, y_n \sim \mathcal{P}(\theta)$$

  y que la distribución a priori para $\theta \sim \mathcal{E}(\lambda)$. Hallar la distribución posterior para θ.

Distribución predictiva

- La distribución marginal de y es llamada la *distribución predictiva a priori*:

$$f(y) = \int_{\theta} f(y|\theta)\, f(\theta)\, d\theta$$

- La distribución de los $\tilde{y}$ luego de observar los datos y es llamada la *distribución predictiva posterior*:

$$\begin{aligned} f(\tilde{y}|y) &= \int_{\theta} f(\tilde{y}, \theta|y)\, d\theta \\ &= \int_{\theta} f(\tilde{y}|\theta)\, f(\theta|y)\, d\theta \end{aligned}$$

Ejemplos

- Suponga que se tiene una muestra:

$$y_1, \cdots, y_n \sim \mathcal{B}(\theta)$$

y que la distribución a priori es $\theta \sim \mathcal{BE}(\alpha, \beta)$. Hallar la distribución predictiva posterior para una nueva observacion $\tilde{y}$.

- Suponga que se tiene una muestra:

$$y_1, \cdots, y_n \sim \mathcal{B}(\theta)$$

y que la distribución a priori es uniforme. Hallar la distribución predictiva posterior para una nueva observacion $\tilde{y}$.

La naturaleza secuencial

- Suponga que una mujer cree que se encuentra embarazada después de un encuentro sexual pero no esta del todo segura.
- Ella decide tomar una prueba de embarazo que sabe es 90 % segura y la prueba resulta ser positiva.
- También se sabe que la prueba da falsos positivos el 50 % de las veces.
- Se podría considerar que sin ninguna información adicional la probabilidad de concepción es aproximadamente 15 % para un encuentro sexual.

La naturaleza secuencial

- Usando el teorema de Bayes se obtiene que la probabilidad que la mujer se encuentre embarazada dado que la prueba resultó positiva es 0.241.
- Usando la terminología Bayesiana la probabilidad anterior es llamada *probabilidad posterior* ya que es obtenida luego de observar el resultado de la prueba.
- Suponga que la mujer decide volverse a hacer la prueba de embarazo. Se puede utilizar la probabilidad calculada en la primera prueba como probabilidad a priori para calcular la probabilidad posterior *actualizada*.

La naturaleza secuencial

- Si la segunda prueba resultó positiva entonces la nueva probabilidad posterior es 0.364.
- Si el resultado anterior no es convincente para ella, puede volver a repetir la prueba y de obtener nuevamente un resultado positivo su probabilidad se incrementa a 0,507.
- Si el proceso nuevamente se repite con el mismo resultado se tendria: prueba 4 $= 0.649$, prueba 5 $= 0.769$, prueba 6 $= 0.857$, prueba 7 $= 0.915$, prueba 8 $= 0.951$, prueba 9 $= 0.972$ y prueba 10 $= 0.984$.
- Este proceso de repetir la prueba y recalcular la probabilidad de interés es el proceso básico en la estadística Bayesiana.