Capítulo 16 Introducción a los métodos de Bayes para inferencia

16.1 Introducción

En los métodos vistos hasta este capítulo se ha designado una distribución de probabilidad de acuerdo a distintos experimentos que modelan fenómenos naturales. En la estadística bayesiana se supone que el parámetro o parámetros de estas distribuciones no son fijos sino que son variables aleatorias que se pueden modelar con alguna distribución de probabilidad. Entonces se puede designar una función de probabilidad *apriori* o *distribución previa* para los datos dados un parámetro y una distribución *aposteriori* que es la distribución del parámetro en cuestión dado los datos. En pocas palabras la estadística bayesiana propone que es más lógico utilizar la información de alguna muestra para determinar la distribución del parámetro que se desea estimar.

16.2 Bayesianos previos, posteriores y estimadores

Si $Y_1, Y_2, \ldots, Y_n$ denotan las variables aleatorias asociadas con una muestra de tamaño n, ya previamente usamos la notación $L(y_1, y_2, \ldots, y_n|\theta)$ para denotar la verosimilitud de la muestra.

El parámetro θ está incluido entre los argumentos de $L(y_1, y_2, \ldots, y_n|\theta)$ para denotar que esta función depende explícitamente del valor de algún parámetro θ. En el método bayesiano, el parámetro desconocido θ se ve como una variable aleatoria con una distribución de probabilidad, llamada *distribución previa* de θ. Esta distribución previa se especifica antes de recolectar cualquier información y da una buena descripción teórica de la información acerca de θ de la que se disponía antes de obtener cualquier dato. En nuestro análisis inicial supondremos que el parámetro θ tiene una distribución continua con densidad $g(\theta)$ que no tiene parámetros desconocidos.

Usando la probabilidad de los datos y la previa sobre θ, se deduce que la probabilidad conjunta de $Y_1, Y_2, \ldots, Y_n, \theta$ es

$$f(y_1, y_2, \ldots, y_n, \theta) = L(y_1, y_2, \ldots, y_n|\theta) \times g(\theta)$$

y que la densidad marginal o función de masa de $Y_1, Y_2, \ldots, Y_n$ es

$$m(y_1, y_2, \ldots, y_n) = \int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n|\theta) \times g(\theta) d\theta$$

Finalmente, la densidad posterior de $\theta|y_1, y_2, \ldots, y_n$ es

$$g^*(\theta|y_1, y_2, \ldots, y_n) = \frac{L(y_1, y_2, \ldots, y_n|\theta) \times g(\theta)}{\int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n|\theta) \times g(\theta) d\theta}$$

La densidad posterior resume toda la información pertinente acerca del parámetro θ al hacer uso de la información contenida en la densidad previa para θ y la información de los datos.

Ejemplo

Denotemos con $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución de Bernoulli donde $P(Y_i = 1) = p$ y $P(Y_i = 0) = 1 - p$ y supongamos que la distribución previa para p es beta(α, β). Se encontrara la distribución posterior para el parámetro p.

Como la función de probabilidad de Bernoulli se puede escribir como

$$p(y_i|p) = p^{y_i}(1-p)^{1-y_i} \qquad y_i = 0, 1$$

La probabilidad $L(y_1, y_2, \ldots, y_n|p)$ es

$$\begin{aligned} L(y_1, y_2, \ldots, y_n|p) &= p(y_1, y_2, \ldots, y_n|p) \\ &= p^{y_1}(1-p)^{1-y_1} \times p^{y_2}(1-p)^{1-y_2} \times \\ &\cdots \times p^{y_n}(1-p)^{1-y_n} \\ &= p^{\sum y_i}(1-p)^{n-\sum y_i} \end{aligned}$$

donde $y_i = 0, 1$ y $0 < p < 1$. Entonces se tiene,

$$\begin{aligned} f(y_1, \ldots, y_n, p) &= L(y_1, y_2, \ldots, y_n|\theta) \times g(p) \\ &= p^{\sum y_i}(1-p)^{n-\sum y_i} \times \\ &\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\sum y_i+\alpha-1}(1-p)^{n-\sum y_i+\beta-1} \end{aligned}$$

y

$$\begin{aligned} m(y_1, \ldots, y_n) &= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\sum y_i+\alpha-1} \times \\ &(1-p)^{n-\sum y_i+\beta-1} dp \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\sum y_i+\alpha)\Gamma(n-\sum y_i+\beta)}{\Gamma(n+\alpha+\beta)} \end{aligned}$$

Finalmente, la densidad posterior de p es

$$\begin{aligned} g^*(p|y_1, \ldots, y_n) &= \frac{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\sum y_i+\alpha-1}(1-p)^{n-\sum y_i+\beta-1}}{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\sum y_i+\alpha)\Gamma(n-\sum y_i+\beta)}{\Gamma(n+\alpha+\beta)}} \\ &= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\sum y_i+\alpha)\Gamma(n-\sum y_i+\beta)} \times \\ &p^{\sum y_i+\alpha-1}(1-p)^{n-\sum y_i+\beta-1} \end{aligned}$$

donde $0 < p < 1$. Con esto se tiene que la densidad posterior de p es una densidad beta con parámetros $\alpha^* = \sum y_i + \alpha$ y $\beta^* = n - \sum y_i + \beta$

Definición 16.1

Las distribuciones previas que resulten en distribuciones posteriores que sean la misma forma funcional que la previa, pero con valores de parámetros alterados, se denominan *previas conjugadas*.

Cualquier distribución beta es una distribución previa conjugada para una distribución de Bernoulli (o una binomial). Cuando la previa se actualiza (usando los datos), el resultado es una beta posterior con valores de parámetro alterados.

Como la posterior es una función de densidad de probabilidad de buena fe, algún resumen característico de esta densidad da una estimación para θ.

Definición 16.2

Sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria con función de verosimilitud $L(y_1, y_2, \ldots, y_n|\theta)$, y que θ tenga densidad previa $g(\theta)$. El estimador posterior de Bayes para $t(\theta)$ está dado por

$$\widehat{t(\theta)}_B = E(t(\theta)|Y_1, Y_2, \ldots, Y_n).$$

Ejemplo

En el Ejemplo anterior, encontramos que la densidad posterior de p es una densidad bera con parámetros $\alpha^* = \sum y_i + \alpha$ y $\beta^* = n - \sum y_i + \beta$, entonces

$$g^*(p|y_1, y_2, \ldots, y_n) = \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} p^{\alpha^*-1}(1-p)^{\beta^*-1}$$

donde $0 < p < 1$. La estimación de p es la media posterior de p. De nuestro estudio previo de la distribución beta, sabemos que

$$\begin{aligned} \hat{p}_B &= E(p|y_1, y_2, \ldots, y_n) \\ &= \frac{\alpha^*}{\alpha^*+\beta^*} \\ &= \frac{\sum y_i + \alpha}{\sum y_i + \alpha + n - \sum y_i + \beta} = \frac{\sum y_i + \alpha}{n+\alpha+\beta} \end{aligned}$$

Del mismo modo,

$$\begin{aligned} [\widehat{p(1-p)}]_B &= E[p(1-p)|y_1, y_2, \ldots, y_n] \\ &= \int_0^1 p(1-p)\frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} p^{\alpha^*-1}(1-p)^{\beta^*-1} dp \end{aligned}$$

Ejemplo

$$
\begin{aligned}
[\widehat{p(1-p)}]_B &= \int_0^1 \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} p^{\alpha^*}(1-p)^{\beta^*}dp \\
&= \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} \times \frac{\Gamma(\alpha^*+1)\Gamma(\beta^*+1)}{\Gamma(\alpha^*+\beta^*+2)} \\
&= \frac{\Gamma(\alpha^*+\beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} \times \\
&\frac{\alpha^*\Gamma(\alpha^*)\beta^*\Gamma(\beta^*)}{(\alpha^*+\beta^*+1)(\alpha^*+\beta^*)\Gamma(\alpha^*+\beta^*)} \\
&= \frac{\alpha^*\beta^*}{(\alpha^*+\beta^*+1)(\alpha^*+\beta^*)} \\
&= \frac{(\sum y_i+\alpha)(n-\sum y_i+\beta)}{(n+\alpha+\beta+1)(n+\alpha+\beta)}
\end{aligned}
$$

Por lo tanto los estimadores de Bayes para p y $(1-p)$ son

$$
\begin{aligned}
\hat{p}_B &= \frac{\sum Y_i+\alpha}{n+\alpha+\beta} \\
[\widehat{p(1-p)}]_B &= \frac{(\sum Y_i+\alpha)(n-\sum Y_i+\beta)}{(n+\alpha+\beta+1)(n+\alpha+\beta)}
\end{aligned}
$$

Un examen más a fondo del estimador de Bayes para p dado en el ejemplo anterior da como resultado,

$$
\begin{aligned}
\hat{p}_B &= \frac{\sum Y_i+\alpha}{n+\alpha+\beta} \\
&= \left(\frac{n}{n+\alpha+\beta}\right)\left(\frac{\sum y_i}{n}\right) + \left(\frac{\alpha+\beta}{n+\alpha+\beta}\right)\left(\frac{\alpha}{\alpha+\beta}\right) \\
&= \left(\frac{n}{n+\alpha+\beta}\right)\overline{Y} + \left(\frac{\alpha+\beta}{n+\alpha+\beta}\right)\left(\frac{\alpha}{\alpha+\beta}\right)
\end{aligned}
$$

Entonces, vemos que el estimador de Bayes para p es un promedio ponderado de la media muestral, $\overline{Y}$ (la MLE para p) y la media de la previa beta asignada a p. Observe que la media previa de p recibe menos valor para tamaños muestrales más grandes, mientras que el valor dado a la media muestral aumenta para tamaños muestrales más grandes. Además, es claro ver que el estimador de bayes para p no es insesgado y en términos generales los estimadores de Bayes no son insesgados.

Como resultado del criterio de factorización (Teorema 9.4) un estimador de Bayes es siempre una función de un estadístico suficiente.

Demostración

Si U es un estadístico suficiente para el parámetro θ basado en una muestra aleatorias $Y_1, Y_2, \ldots, Y_n$, entonces

$$L(y_1, y_2, \ldots, y_n|\theta) = k(u,\theta) \times h(y_1, y_2, \ldots, y_n)$$

Demostración

donde $k(u,\theta)$ es una función sólo de u y θ y $h(y_1, y_2, \ldots, y_n)$ no es una función de θ. Además, la función $k(u,\theta)$ puede (pero no necesita) ser escogida para ser la función de probabilidad de masa o densidad del estadístico U. De acuerdo con la notación de este capítulo, escribimos la densidad condicional de $U|\theta$ como $k(u|\theta)$. Entonces, como $h(y_1, y_2, \ldots, y_n)$ no es una función de θ,

$$
\begin{aligned}
g^\star(\theta|y_1, y_2, \ldots, y_n) &= \frac{L(y_1, y_2, \ldots, y_n|\theta) \times g(\theta)}{\int_{-\infty}^{\infty} L(y_1, y_2, \ldots, y_n|\theta) \times g(\theta)d\theta} \\
&= \frac{k(u|\theta) \times h(y_1, y_2, \ldots, y_n) \times g(\theta)}{\int_{-\infty}^{\infty} k(u|\theta) \times h(y_1, y_2, \ldots, y_n) \times g(\theta)d\theta} \\
&= \frac{k(u|\theta) \times g(\theta)}{\int_{-\infty}^{\infty} k(u|\theta) \times g(\theta)d\theta}
\end{aligned}
$$

Por lo tanto, en casos donde se conozca la distribución de un estadístico U suficiente, la posterior se puede determinar usando la densidad condicional de $U|\theta$.

16.3 Intervalos creíbles de Bayes

De acuerdo a capítulo anteriores se puede decir que un intervalo de confianza es aleatorio y el parámetro es fijo. Decimos que es un intervalo de confianza de 95% porque *si se tomaran muestras independientes y separadas, cada una de tamaño n* y se calculan los intervalos resultantes, a la larga, 95% de *los intervalos contendrían la verdadera media*. El parámetro es fijo, los puntos del intervalo son aleatorios y diferentes muestras darán intervalos realizados diferentes.

En el contexto de Bayes, el parámetro θ es una *variable aleatoria* con función de densidad posterior $g^\star(\theta)$. Si consideramos el intervalo (a,b), la probabilidad posterior de que la variable aleatoria θ se encuentre en este intervalo es

$$P^\star(a \le \theta \le b) = \int_a^b g^\star(\theta)d\theta$$

Si la probabilidad posterior $P^\star(a \le \theta \le b) = 0.90$, decimos que (a,b) es un *intervalo creíble* de 90% para θ.

16.4 Pruebas de hipótesis de Bayes

Si, como en la Sección 10.11 en la que consideramos pruebas de razón de verosimilitud, estamos interesados en probar que el parámetro θ se encuentra en uno de dos conjuntos de valores, Ω_0 y Ω_a, podemos usar la distribución posterior de θ para calcular la probabilidad posterior de que θ esté en cada uno de estos conjuntos de valores. Cuando se pruebe $H_0 : \theta \in \Omega_0$ contra $H_a : \theta \in \Omega_a$, un método que se usa con frecuencia es calcular las probabilidades posteriores $P^*(\theta \in \Omega_0)$ y $P^*(\theta \in \Omega_a)$ y aceptar la hipótesis con la más alta probabilidad posterior. Esto es, para probar $H_0 : \theta \in \Omega_0$ contra $H_a : \theta \in \Omega_a$,

aceptar H_0 si $P^\star(\theta \in \Omega_0) > P^\star(\theta \in \Omega_a)$

aceptar H_a si $P^\star(\theta \in \Omega_a) > P^\star(\theta \in \Omega_0)$

Diferentes analistas con distintas opciones de valores para los parámetros de la distribución posterior pueden llegar a conclusiones diferentes.