Tema

3

Inferencia Bayesiana para una Proporción

Índice

1. Introducción

Hay muchas situaciones reales donde la probabilidad o proporción de un suceso respecto a otro es desconocida. Por ejemplo, la proporción de alumnos que responden correctamente a una pregunta determinada, la proporción de votos en una candidatura, la proporción de coincidencia de un texto con otro, la calidad de un tiro de gol en función de determinadas covariables.

En la Inferencia Bayesiana el parámetro a estudiar se considera como una v.a. debido a la incertidumbre que tenemos y se determina la distribución a posteriori a medida que se van obteniendo datos.

La organización de este tema es como sigue. En primer lugar se recordarán algunas distribuciones elementales vistas en la asignatura de Teoría de la Probabilidad I que son de gran utilidad en la Inferencia Bayesiana. A continuación, en la sección 3, calcularemos la función de verosimilitud de la proporción de un ensayo de tipo Bernoulli que sería equivalente a lo realizado en la asignatura de Inferencia Estadística de segundo curso. En la sección cuatro, propondremos diferentes distribuciones a priori para la proporción con sus correspondientes a posteriori según la información que se tenga y utilizando los métodos propuestos en el Tema 2. En la sección 5, propondremos los estimadores de Bayes según las funciones de pérdidas usadas. En la sección 6 calcularemos el factor de Bayes para contrastes unilaterales. Y por último, en la sección 7 estudiaremos la distribución predictiva.

2. La distribución Beta

2.1. La función Gamma

Una de las funciones más utilizadas en Matemáticas es la función gamma. Esta función aparece como parte de muchas distribuciones en Estadística, y aunque no existe una expresión determinada de ella sí se puede calcular su valor por desarrollo en serie.

Definición 2.1. *Para cualquier número positivo p, la función gamma está definida mediante la siguiente integral impropia*

$$\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx.$$

Si $p \in \mathbb{N}$ se obtiene de manera inmediata por partes que

$$\Gamma(p) = (p-1)\Gamma(p-1) = (p-1)!$$

También se puede demostrar que $\Gamma(p+1) = p\Gamma(p)$; y si $a > 0$ entonces

$$\int_0^\infty x^{p-1} e^{-ax} dx = \frac{\Gamma(p)}{a^p}.$$

Mediante un cambio de variable reduciendo a la distribución normal estándard se obtiene que: $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

2.2. La función Beta

Definición 2.2. *Para cualquier par de números positivos a y b, se define la función Beta de parámetros a y b a la función dada por:*

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx.$$

Mediante el análisis matemático se demuestra que:

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b).}$$

2.3. La distribución Beta

Definición 2.3. *Se dice que la variable aleatoria X sigue una distribución beta de parámetros a y b positivos (y se denota por $X \sim Be(a,b)$) si su función de densidad es de la forma:*

$$f(x;a,b) = \frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1} \text{ para } x \in (0,1).$$

Proposición 2.4. *Si $X \sim Be(a,b)$ entonces*

$$E(X) = \frac{a}{a+b}, \quad var(X) = \frac{ab}{(a+b)^2(a+b+1)}, \quad Moda(X) = \frac{a-1}{a+b-2},$$

para la moda tiene que cumplirse que $a > 1$ y $b > 1$.

Nota 1. *Un caso trivial es $Be(1,1) \equiv U(0,1)$.*

3. La función de verosimilitud

Supongamos un suceso aleatorio A de un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ cuya probabilidad de ocurrir es $p = P(A)$, la cual es un valor desconocido y objetivo de la inferencia a realizar (estimaciones puntuales, intervalos de credibilidad y contraste de hipótesis). Para llevar a cabo esta inferencia partiremos de un conjunto de datos provenientes de la realización experimental correspondiente. Es decir, cada dato muestral valdrá 1 o 0 en función de si se verifica o no el suceso A en cada realización muestral.

De esta forma, el conjunto de datos lo podemos denotar como $\mathbf{x} = (x_1, \ldots, x_n)$ donde $x_i = 1$ o 0 para cada valor de i.

Como se sabe de Inferencia Estadística, la función de verosimilitud de un parámetro se define como la función de densidad muestral en términos de dicho parámetro. En general, una función de verosimilitud no tiene por qué ser una función de densidad. También, y aunque no se ha especificado, vamos a suponer que la muestra obtenida es aleatoria simple. Entonces se tiene que

$$\begin{aligned} L(p|\mathbf{x}) &= f(\mathbf{x}|p) \\ &= \prod_{i=1}^{n} f(x_i|p) \text{ por ser m.a.s.} \\ &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\ &= p^{\sum x_i}(1-p)^{n-\sum x_i}. \end{aligned}$$

Por tanto, expresándolo correctamente nos queda que

$$L(p|\mathbf{x}) = p^{\sum x_i}(1-p)^{n-\sum x_i} \text{ para } p \in [0,1] \text{ y } 0 \leq \sum x_i \leq n.$$

Hay que aclarar algunos aspectos. El primero es que en el sumatorio de la fórmula anterior no se ha expresado su recorrido desde $i = 1$ hasta $i = n$ para no sobrecargarla pero cuando no se especifique dicho recorrido es porque se supone que es este. Otro aspecto a aclarar es que es posible, en función de los valores del número de éxitos (veces que ocurre el suceso A) y del número de repeticiones de dicho experimento n, la representación gráfica en **R** de la función de verosimilitud no sea muy *visual* por razones de escalas. En este caso se permite una rescalización de los valores de la verosimilitud para que la visualización de la misma sea más efectiva.

4. Distribución a priori y a posteriori

El paso inicial en todo proceso de Inferencia Bayesiana Paramétrica es el cálculo de las distribuciones a priori y a posteriori . Como se vio en el tema 2, la distribución a posteriori recoge la actualización de la información contenida en la distribución a priori debida al nuevo conjunto de datos procedentes de la experimentación. Por tanto es fundamental saber construir una distribución a priori a partir de la información que se disponga (pudiéndose dar el caso de no tener información alguna). Para ello, en esta sección propondremos tres métodos de construcción de la distribución a priori según la información que se tenga y a continuación a partir de un conjunto de datos obtener la correspondiente distribución a posteriori.

4.1. A priori no informativa completa

Este caso es cuando antes de realizar el experimento no se tienen información alguna acerca de los posibles valores del parámetro p salvo su rango que en este caso sí sabemos que es el intervalo unidad $[0,1]$. El hecho de no tener información de sus posibles valores puede interpretarse que a priori, y nunca mejor dicho, cualquier valor tiene las mismas posibilidades de salir. Obviamente los posibles valores pertenecen a un intervalo y por tanto la densidad adecuada para modelar esta información es la uniforme sobre el intervalo unidad. Así tendríamos que la distribución a priori sería:

$$\pi(p) = I_{[0,1]}(p).$$

Nótese que en este caso la distribución a priori es una distribución beta de parámetros $(1,1)$. Por tanto, la distribución a posteriori sería:

$$\pi(p|\mathbf{x}) \propto L(p|\mathbf{x})\pi(p).$$

En este caso se ha puesto el símbolo $\propto$ que significa que la densidad a posteriori es proporcional a dicho valor ya que aplicando el Teorema de Bayes faltaría calcular el denominador correspondiente que en realidad es una constante normalizadora (es una función de los datos muestrales únicamente). En muchos casos podemos proceder así porque reconociendo la estructura del numerador podemos obtener de manera inmediata la constante normalizadora. En este caso se tendría

$$\pi(p|\mathbf{x}) \propto p^{\sum x_i}(1-p)^{n-\sum x_i},$$

que se corresponde con la estructura de una densidad Beta de parámetros $(\sum x_i + 1, n - \sum x_i + 1)$. Luego, en realidad no hay necesidad de calcular esa constante normalizadora puesto que se ha reconocido completamente la distribución a posteriori. Es decir,

$$p|\mathbf{x} \sim Be(\sum x_i + 1, n - \sum x_i + 1).$$

4.2. A priori de Jeffrey

En este caso, tal como se vio en tema 2, hay que calcular la Información de Fisher del parámetro p. Para ello tenemos que

$$f(x|p) = p^x(1-p)^{1-x} \text{ con } x \in \{0,1\} \text{ y } p \in [0,1].$$

Por tanto, $\ln f(x|p) = x\ln(p) + (1-x)\ln(1-p)$. Mediante un cálculo sencillo se obtiene que la cantidad de información de Fisher para p vale

$$I(p) = \frac{1}{p(1-p)}.$$

Luego, la distribución a priori de Jeffrey sería

$$\pi(p) = \frac{1}{\sqrt{p(1-p)}} \text{ para } p \in (0,1).$$

Esta distribución a priori se podría interpretar como que modela la información cuando son más probables los valores extremos de p que los centrales. En este caso en particular tendríamos la distribución beta de parámetros $(1/2, 1/2)$.

Luego,

$$\pi(p|\mathbf{x}) \propto p^{\sum x_i}(1-p)^{n-\sum x_i} p^{-1/2}(1-p)^{-1/2}.$$

Es decir, $p|\mathbf{x} \sim Be(\sum x_i + 1/2, n - \sum x_i + 1/2)$.

4.3. A priori por el método de las conjugadas

Posiblemente este sea el método más criticable de construcción de la distribución a priori ya que supone un modelo completamente determinado. La forma de calcularlo depende de la estructura de la función de verosimilitud. Hemos visto en el apartado 3 que la función de verosimilitud tiene estructura de una distribución beta por tanto, la distribución a priori propuesta por el método de las conjugadas debe ser una distribución beta. La razón es muy simple. Cuando se calcule la distribución a posteriori mediante el paso a proporcionalidad, tendríamos el producto de dos funciones con la misma estructura con lo cual nos conllevaría a la distribución conjugada.

Esta selección se debe a una distribución *informativa.* Supongamos que la distribución a priori es una beta de parámetros a y b. Entonces,

$$\pi(p|\mathbf{x}) \propto p^{\sum x_i}(1-p)^{n-\sum x_i} p^{a-1}(1-p)^{b-1}.$$

Es decir, $p|\mathbf{x} \sim Be(\sum x_i + a, n - \sum x_i + b)$. Por tanto, la distribución beta es conjugada respecto a la distribución Bernoulli. Todo lo desarrollado hasta aquí se puede generalizar de manera sencilla al caso de que la distribución muestral sea una binomial, es decir, $X|p \sim Bi(n_0, p)$ donde el valor de n_0 es conocido. Es por esta razón que este modelo Bayesiano se le suele conocer con el nombre de modelo Beta-Binomial.

Para seleccionar la distribución beta a priori debemos especificar los valores de los dos parámetros a y b. Una forma sencilla, no exenta de crítica, es estableciendo a priori dos valores fijos para ciertas medidas asociadas a la distribución beta como pueden ser la media y la moda o mediante dos cuantiles. Para su solución requiere del uso de la programación en **R** para resolver sistemas de ecuaciones o bien usando el paquete `LearnBayes` en el caso de que dispongamos de dos cuantiles.

5. Estimadores Bayes e Intervalos de Credibilidad

5.1. Estimadores de Bayes

Se ha visto en el tema anterior que el estimador de Bayes usando función de pérdida cuadrática es la media de la distribución a posteriori. Por tanto, supongamos que la distribución a posteriori de p es una beta de parámetros $(a(\mathbf{x}), b(\mathbf{x}))$ entonces se tiene que

$$\hat{p}_B = \frac{a(\mathbf{x})}{a(\mathbf{x}) + b(\mathbf{x})}.$$

Sin embargo, si usamos la función de pérdida valor absoluto sabemos que el estimador de Bayes es la mediana. En el caso de la distribución beta no existe una fórmula explícita para la mediana con lo cual para su cálculo tendremos que recurrir al **R**.

5.2. Intervalos de Credibilidad

Debido a que la distribución a posteriori, independientemente de la a priori seleccionada, es una distribución beta eligiremos el método de las colas igualmente ponderadas para proponer un intervalo de credibilidad asegurándonos que la moda de la distribución a posteriori pertenezca a dicho intervalo. En caso de que no pertenezca propondremos un intervalo que la contenga independientemente del reparto de las probabilidades en la colas pudiendo ser nula una de ellas. Tambén podemos usar el paquete `HDInterval` si queremos usar el método de intervalo de probabilidad mayor. En cualquier caso, también tendremos que recurrir al **R** para su cálculo.

6. Factor de Bayes

Supongamos que desamos realizar el siguiente contraste de hipótesis:

$$H_0 : p \leq p_0 \text{ frente a } H_1 : p > p_0.$$

Partimos de un conjunto de datos $\mathbf{x}$ procedente de la realización n pruebas independientes de tipo Bernoulli. Luego en este caso se tendría que

$$FB_{01} = \frac{P(p \leq p_0|\mathbf{x})}{P(p > p_0|\mathbf{x})} \times \frac{P(p > p_0)}{P(p \leq p_0)}.$$

El primer factor corresponde a las distribuciones a posteriori mientras que el segundo a las distribuciones a priori. También es evidente que necesitaremos el **R** para llevar a cabo este cálculo. Posteriormente decidiremos la evidencia de las hipótesis en función del valor del FB_{01} según la tabla de Jeffrey.

7. Distribución Predictiva

Es evidente que si repetimos de manera independiente e idénticamente distribuidos n experimentos de tipo Bernoulli para intentar predecir la prueba $(n+1)$-ésima, ésta última será también un valor de una Bernoulli. Por tanto estamos interesados en calcular

$$p^* = P(Y = 1|\mathbf{x}).$$

Usando la fórmula que hemos visto en el tema 2, se tiene que

$$\begin{aligned} p^* &= \int_0^1 P(Y = 1|\mathbf{x}, p)\pi(p|\mathbf{x})dp \\ &\text{(supongamos el método conjugado)} \\ &= \int_0^1 p\, \frac{p^{\sum x_i + a_0 - 1}(1-p)^{n-\sum x_i + b_0 - 1}}{B(\sum x_i + a_0, n - \sum x_i + b_0)} dp \\ &= \frac{B(\sum x_i + a_0 + 1, n - \sum x_i + b_0)}{B(\sum x_i + a_0, n - \sum x_i + b_0)} \end{aligned}$$

En definitiva, usando la función beta y simplificando adecuadamente en la función gamma quedaría

$$P(Y = 1|\mathbf{x}) = \frac{\sum x_i + a_0}{n + a_0 + b_0}.$$

De forma similar se puede llegar a la función de probabilidad predictiva para una binomial. Es decir, supondiendo que se repite el experimento de Bernoulli un número conocido de n_0 veces y se pretende calcular

$$P(Y = k|\mathbf{x}) \text{ para } k = 0, 1, \ldots, n_0.$$

Afortunadamente hay un comando del paquete `Learnbayes` que nos permite realizar este cálculo.

Ejercicios de autoaprendizaje.

1. Supongamos que tenemos la distribución muestral binomial de parámetros m y p, con m conocido. Calcular la función de verosimilitd para p a partir de una muestra aleatoria de tamaño n.

2. Suponiendo una a priori conjugada completamente determinada en el apartado anterior, calcular la distribución a posteriori para p.

3. Calcular la distribución predictiva bajo los supuestos anteriores.