Caso Práctico III

Modelización y tratamiento de la incertidumbre

Inferencia

José Ignacio Escribano

Móstoles, 16 de octubre de 2015

Índice de figuras

Índice

1. Introducción

Este caso práctico consta de dos partes relativas a dos modelos distintos: el primero es acerca del modelo beta-binomial, y el segundo, sobre el modelo normal-normal. En ambas partes se piden calcular intervalos de probabilidad y contrastes de hipótesis, entre otras cosas.

2. Estimando proporciones y predicción de futuras muestras

En esta primera parte, consideramos una población de 29 niños que tenían un contenido en plomo en los dientes de leche superior a 22.22 partes por millón, de los cuales, 22 terminaron la Educación Secundaria, y 7 que no lo hicieron. Consideraremos que la distribución a priori de la proporción p de niños que terminaron la Educación Secundaria sigue una distribución $p \sim \mathcal{B}e(1,1)$.

Lo primero que tenemos que calcular es la función de verosimilitud $\mathcal{L}(p) = P(X = x|p)$. Sabemos que sólo hay dos resultados posibles: terminar o no terminar la Educación Secundaria, por lo que tenemos una distribución binomial con $n = 29$ y $x = 22$.

Por tanto, la función de verosimilitud es:

$$\begin{aligned}
\mathcal{L}(p) &= P(X = 22|p) \\
&= \binom{29}{22} p^{22}(1-p)^{29-22} \\
&\propto p^{22}(1-p)^{29-22} \\
&= p^{22}(1-p)^{7}
\end{aligned}$$

Por otro lado, tenemos como distribución a priori una $\mathcal{B}e(1,1)$. La función de densidad de la distribución $\mathcal{B}e(\alpha,\beta)$ viene dada por

$$f(p|\alpha,\beta) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}, & \text{si } 0 \le p \le 1 \\ 0, & \text{en otro caso} \end{cases}$$

donde $\Gamma(\cdot)$ denota función gamma definida como

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$$

Por tanto,

$$
\begin{aligned}
f(p|\alpha=1,\beta=1) &= \frac{\Gamma(1+1)}{\Gamma(1)+\Gamma(1)} p^0 (1-p)^0 \\
&\propto 1
\end{aligned}
$$

Tenemos una distribución a priori que no nos da ninguna información sobre p.

La distribución a posteriori $f(p|x)$ viene dada por

$$
\begin{aligned}
f(p|x) &\propto f(p) \cdot \mathcal{L}(p) \\
&\propto 1 \cdot p^{22}(1-p)^7 \\
&= p^{22}(1-p)^7
\end{aligned}
$$

Comparando con la definición de función de densidad de una distribución beta, tenemos que la distribución a posteriori $p|x \sim \mathcal{B}e(23, 8)$. Tanto la distribución a posteriori como la distribución empírica son bastante parecidas, por lo que el comando `rbeta` de R aproxima bien la función aún cunado el número de muestras es relativamente pequeño.

La Figura 1 muestra las distribuciones a priori, a posteriori y la distribución empírica. Ésta última se ha generado generando con 1000 muestras de una distribución $\mathcal{B}e(23, 8)$ (nuestra distribución a posteriori) con el comando `rbeta` de R. Tanto la distribución beta

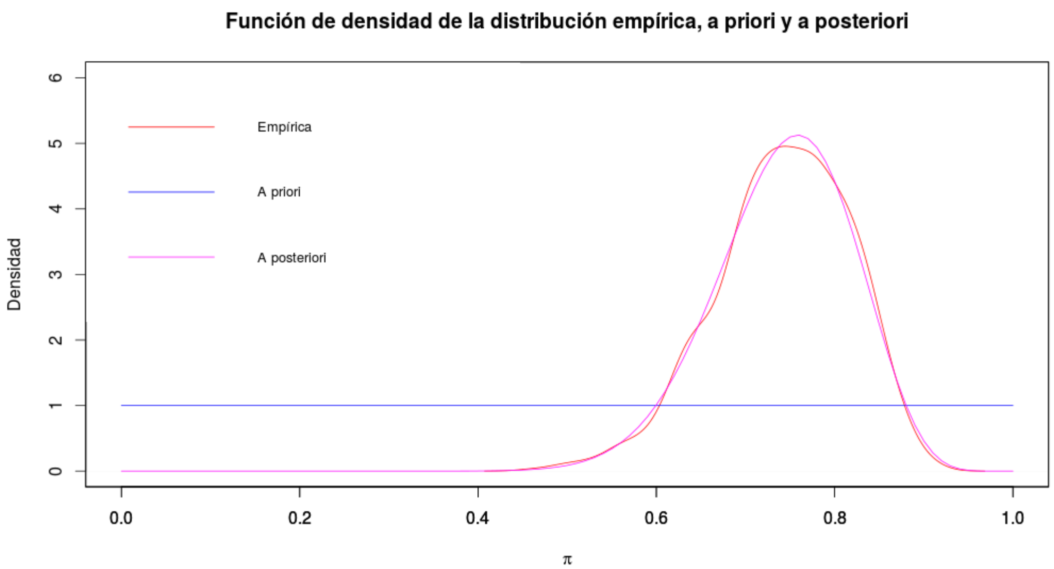

Figura 1: Función de densidad de la distribución empírica, a priori y a posteriori

Como tenemos nuestra distribución a posteriori, podemos calcular la media y la varianza a posteriori de esta distribución, ya que las fórmulas para la media y la varianza son conocidas:

$$E(x|p) = \frac{\alpha}{\alpha + \beta}$$
$$Var(x|p) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Sustituyendo en las fórmulas anteriores α por 23 y β por 8, se tiene que

$$E(x|p) = 0.7419$$
$$Var(x|p) = 0.0059$$
$$\sigma(x|p) = \sqrt{Var(x|p)} = 0.0773$$

3. Estimando una media normal con una a priori discreta

4. Conclusiones

5. Código R