Trabajo Práctico N.º 2
"Inferencia Estadística"

Laboratorio de Sistemas Embebidos
Probabilidad y Estadística para IA
David Canal
Abril 2024

1. Consigna de trabajo

Supongamos que estás trabajando en una empresa de tecnología que ha lanzado una nueva aplicación de recomendación de contenido. La aplicación utiliza un algoritmo automático para recomendar posibles series y películas a los usuarios. La empresa cree que el algoritmo es muy efectivo y afirma que la probabilidad de que sus sugerencias sean correctas es del 70%.

Juan ha estado usando la aplicación durante los últimos meses y de las 7 series recomendadas ha visto 5. Él está interesado en saber qué tan efectiva es realmente la aplicación.

Utilizando el método de estimación Bayesiana, estimar la probabilidad de que de las siguientes 10 recomendaciones Juan acepte exactamente 5.

(a) Definir una distribución adecuada para la probabilidad de que una recomendación sea aceptada.
(b) Calcular la distribución a posteriori para la probabilidad de que una recomendación sea aceptada.
(c) A partir de esta distribución a posteriori halla la probabilidad deseada.

2. Resolución

Antes de abordar la resolución del problema planteado, se procede a definir las variables aleatorias que serán objeto de estudio en este análisis.

2.1 Definición de variables aleatorias

$X|_{\Theta=\theta}$:" El número de recomendaciones aceptadas".

Dado que la distribución binomial modela el número de éxitos (recomendaciones aceptadas) en un número fijo de ***n*** intento, donde cada intento tiene una probabilidad ***θ*** de éxito constante e independientes entre intentos, tenemos entonces:

$$X|_{\Theta=\theta} \sim Binomial(n, p),$$

donde:

$$P_{\mathrm{X}|\Theta=\theta}(x) = \binom{n}{x}\theta^{x}(1-\theta)^{n-x} \qquad (2.1).$$

Θ: "Probabilidad de que una recomendación de la aplicación sea aceptada",

Se sabe que $\boldsymbol{\Theta}$ representa una probabilidad, y por ende, debe tomar valores en el intervalo [0,1], asignamos una distribución de densidad de probabilidad Beta a $\boldsymbol{\Theta}$. La distribución Beta es una elección frecuente para modelar variables aleatorias que están limitadas al intervalo **[**0,1]. Esto la hace particularmente adecuada para representar probabilidades. Tenemos entonces:

$$\Theta \sim \mathrm{Beta}(\alpha, \beta),$$

Basándonos en la afirmación de que la empresa tiene una probabilidad de éxito del 70%, se sugiere que la probabilidad de una recomendación exitosa es alta. Sin consideramos 10 eventos posibles, esto significa que se espera que 7 sean exitosos y 3 sean negativos. Por lo tanto, procedemos con la siguiente selección:

$$\alpha = 7\ (Exitos - Eventos\ positivos),$$

$$\beta = 3\ (Fracasos - eventos\ negativos).$$

La distribución a priori es entonces:

$$\pi(\theta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \quad (2.2)$$

2.2 Determinación de la distribución a posteriori

Mediante la inferencia bayesiana podemos plantear los siguiente:

$$P_{\Theta|\mathrm{X=x}}(\theta) \propto P_{\mathrm{X}|\Theta=\theta}(x)\pi(\theta) \quad (2.3),$$

Siendo:

$P_{\mathrm{X}|\Theta=\theta}$: Distribución de probabilidad likelihood,

π: Distribución de probabilidades a priori,

$P_{\Theta|\mathrm{X=x}}$: Distribución de probabilidades a posteriori.

Reemplazando las ecuaciones (2.1), (2.2) en (2.4), se tiene:

$$P_{\Theta|\mathrm{X=x}}(\theta) \propto \binom{n}{x}\theta^{x}(1-\theta)^{n-x}\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\mathbb{I}(0<\theta<1)$$

$$P_{\Theta|\mathrm{X=x}}(\theta) \propto \binom{n}{x}\theta^{x+\alpha-1}(1-\theta)^{n+\beta-x-1}\frac{1}{B(\alpha,\beta)}\mathbb{I}(0<\theta<1)$$

$$P_{\Theta|\mathrm{X=x}}(\theta) \propto \theta^{x+\alpha-1}(1-\theta)^{\beta+n-x-1}\mathbb{I}(0<\theta<1)$$

$$\Theta|X = x \sim Beta(\alpha', \beta')$$

donde:

$$\alpha' = \alpha + x \,; \quad \beta' = \beta + n - x$$

El enunciado del problema nos indica que Juan vio 5 series de las 7 recomendaciones hechas por el programa, lo cual implica:

$$\alpha' = 7 + 5 = 12; \quad \beta' = 3 + 2 = 5$$

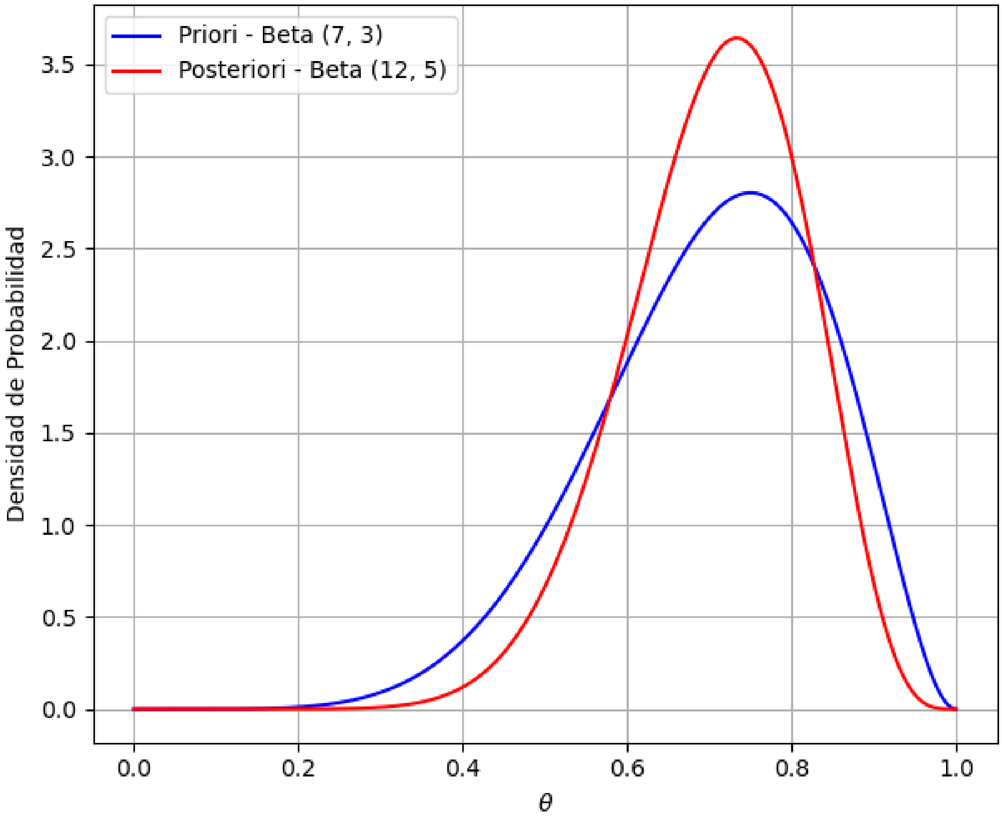

Figura 2.1: "Distribuciones de probabilidades a priori y a posteriori".

2.3 Determinación de la probabilidad deseada

$$P(X = x_1) = \int_{\mathbf{0}}^{\mathbf{1}} P_{X|\Theta=\theta}(x_1)\, P_{\Theta|X=x}(\theta) d\theta$$

$$= \int_{\mathbf{0}}^{\mathbf{1}} \binom{n}{x_1} \theta^{x_1}(1-\theta)^{n-x_1} \frac{1}{B(\alpha', \beta')} \theta^{\alpha' - 1}(1-\theta)^{\beta' - 1}\, \mathbb{I}(0 < \theta < 1) d\theta$$

$$= \binom{n}{x_1} \frac{1}{B(\alpha', \beta')} \int_{\mathbf{0}}^{\mathbf{1}} \theta^{\alpha' + x_1 - 1}(1-\theta)^{\beta' + n - x_1 - 1}\, \mathbb{I}(0 < \theta < 1) d\theta,$$

Por definición de la función beta, sabemos que:

$$B(x, y) = \int_{\mathbf{0}}^{\mathbf{1}} t^{x-1}(1-t)^{y-1} \quad (2.4),$$

Luego por (2.4), tenemos finalmente que:

$$P(X = x_1) = \binom{n}{x_1} \frac{B(\alpha' + x_1, \beta' + n - x_1)}{B(\alpha', \beta')} \quad (2.5)$$

Ahorta tenemos $x_1 = 5\ y\ \ n = 10, entonces$:

$$P(X = 5) = \binom{10}{5} \frac{B(12 + 5, 5 + 10 - 5)}{B(12,5)}$$

$$= \binom{10}{5} \frac{B(17, 10)}{B(12,5)},$$

Por lo tanto:

$$P(X = 5) = 0.1036.$$

3. Conclusiones

A partir del análisis expuesto en secciones anteriores, se deduce que la probabilidad de que Juan acepte 5 de las 10 recomendaciones siguientes es aproximadamente del 10.4%