

Probabilidad y Estadística para Inteligencia Artificial

Resolución de examen PEIA 2024 a15

Nombre: Juan Pablo Schamun

1. Voy a definir a \mathbf{X} como la variable aleatoria del tamaño de una imagen en píxeles. Si bien los píxeles son números enteros, y en rigor \mathbf{X} debería ser una variable discreta, para efectos de este problema \mathbf{X} se puede considerar como una variable continua.

Las imágenes con las que Sonia está trabajando pueden ser del tipo perros o gatos. Voy a definir a \mathbf{Y} como la variable aleatoria del tipo de la imagen. \mathbf{Y} es una variable discreta con dos posibles valores: perro y gato.

Conozco las distribuciones condicionales del tamaño de la imagen dado si la imagen es del tipo perro o del tipo gato:

$$f(X = x|Y = \text{perro}) \sim N(\mu = 500, 50);$$

$$f(X = x|Y = \text{gato}) \sim N(\mu = 450, 40);$$

$$X \in \mathbb{N}(0, \infty);$$

$$Y \in \{\text{perro}, \text{gato}\};$$

La probabilidad de que una imagen sea de un determinado tipo dado que \mathbf{X} toma un valor \mathbf{x} lo voy a determinar con el teorema de Bayes:

$$\frac{P(Y = y|X = x) = f(X = x|Y = y) * P(Y = y)}{f(X = x)}$$

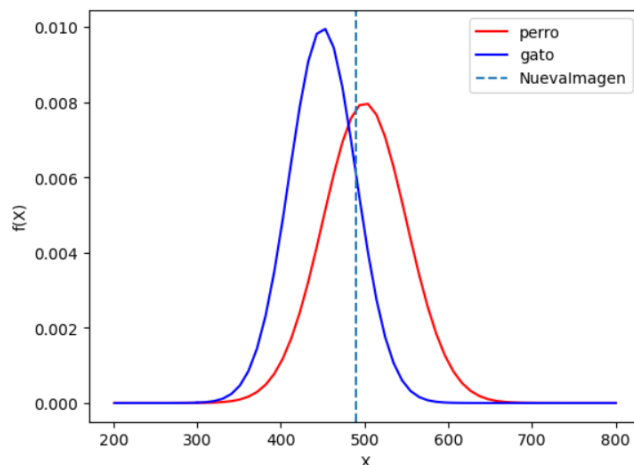
Siendo $P(Y = y)$ la probabilidad marginal de que \mathbf{Y} adopte un valor de \mathbf{y} , es decir la proporción de imágenes que son del tipo \mathbf{y} entre todas las que toma Sonia.

$f(X = x)$ es la probabilidad marginal de que \mathbf{X} adopte un valor de \mathbf{x} , la cual calculo como sigue:

$$f(X = x) = \sum_Y f(X = x|Y = y) * P(Y = y)$$

La nueva imagen que toma Sonia tiene un tamaño de 490 píxeles.

Gráficamente, hasta ahora tenemos algo así:



Entonces la probabilidad de que la nueva imagen sea del tipo perro es:

$$P(Y = \text{perro} | X = 490) = \frac{f(X = x | Y = \text{perro}) * P(Y = \text{perro})}{f(X = 490 | Y = \text{perro}) * P(Y = \text{perro}) + f(X = 490 | Y = \text{gato}) * P(Y = \text{gato})}$$

Voy a suponer que la proporción de imágenes que toma Sonia esta balanceada, es decir $P(Y) = 0.5$. Haciendo los cálculos:

$$P(Y = \text{perro} | X = 490) = 0.563$$

2. Defino la variable aleatoria X como el tiempo que puede pasar un usuario en la página del desarrollador. No se conoce la distribución de X , pero se sabe que la media histórica es 6 minutos y un desvío estándar maestral de 1.5 minutos

$$X \sim \text{desconocida}; X > 0$$

- a. Defino la hipótesis nula como que la media poblacional actual no ha aumentado respecto de la media histórica y la hipótesis alternativa como que la media poblacional si ha aumentado

$$H_0: \mu \leq \mu_0$$

$$H_1: \mu > \mu_0$$

- b. No conozco la distribución de X , pero si se tienen muchas muestras se puede aproximar a una normal.

$$\text{Definiré un estadístico de prueba } U = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

$$\text{Con } \bar{X} = 6.5; \mu_0 = 6; S = 1.5; n = 50$$

Si U es mayor a un valor crítico, entonces rechazare la hipótesis nula. Caso contrario no podré rechazarla.

El valor crítico para un nivel de significancia del 5% se corresponde con el $t_{0.95}$ de una distribución t de Student con 49 grados de libertad. **Este valor es 1.677**

$$\text{Entonces calculo el } U = \frac{6.5-6}{1.5/\sqrt{50}} = 2.357$$

- c. Se observa que U es mayor al valor crítico, entonces con un nivel de significancia del 5% se puede rechazar la hipótesis nula y afirmar que de acuerdo a los datos que el tiempo de permanencia de los usuarios ha aumentado respecto al histórico.

3. El stock X sigue una distribución de Poisson con parámetro λ .

$$P(X = x) = Po(\lambda);$$

$$X \in \mathbb{N} [0, \infty);$$

$$\lambda > 0$$

El valor de este parámetro es una variable aleatoria θ que sigue una distribución gamma,

La distribución a priori definida del parámetro será del parámetro:

$$\pi(\theta = \theta) \sim \Gamma(\alpha = 10, \beta = 1) * I(\theta > 0)$$

Para la distribución a posteriori, aplicaré el teorema de Bayes para calcular

$$f(\theta = \theta | X = x) \approx L(\theta) * \pi(\theta)$$

Donde $L(\theta)$ es la verosimilitud de los datos dado que θ toma el valor θ

$$L(\theta) \approx \prod_{i=1}^n Po(x_i; \theta) * I(\theta > 0)$$

$$L(\theta) \approx \prod_{i=1}^n \frac{\theta^{x_i} e^{-\theta}}{x_i!} * I(\theta > 0)$$

$$\pi(\theta = \theta) = \frac{1^{10} * \theta^9 * e^{-\theta}}{\Gamma(10)} * I(\theta > 0)$$

Quitando los términos constantes respecto del parámetro queda:

$$f(\theta = \theta | X = x) \approx \prod_{i=1}^n \frac{\theta^{x_i} e^{-\theta}}{x_i!} * \frac{1^{10} * \theta^9 * e^{-\theta}}{\Gamma(10)} * I(\theta > 0)$$

$$f(\theta = \theta | X = x) \approx \theta^{\sum_{i=1}^n x_i} * e^{-n\theta} * \theta^9 * e^{-\theta} * I(\theta > 0)$$

$$f(\theta = \theta | X = x) \approx \theta^{\sum_{i=1}^n x_i + 9} * e^{-\theta(n+1)} * I(\theta > 0)$$

Los datos son $x = [20, 5, 6, 30, 2, 5]$, con lo cual

$$n = 6; \bar{X} = 11.33; \sum_{i=1}^n x_i = 68$$

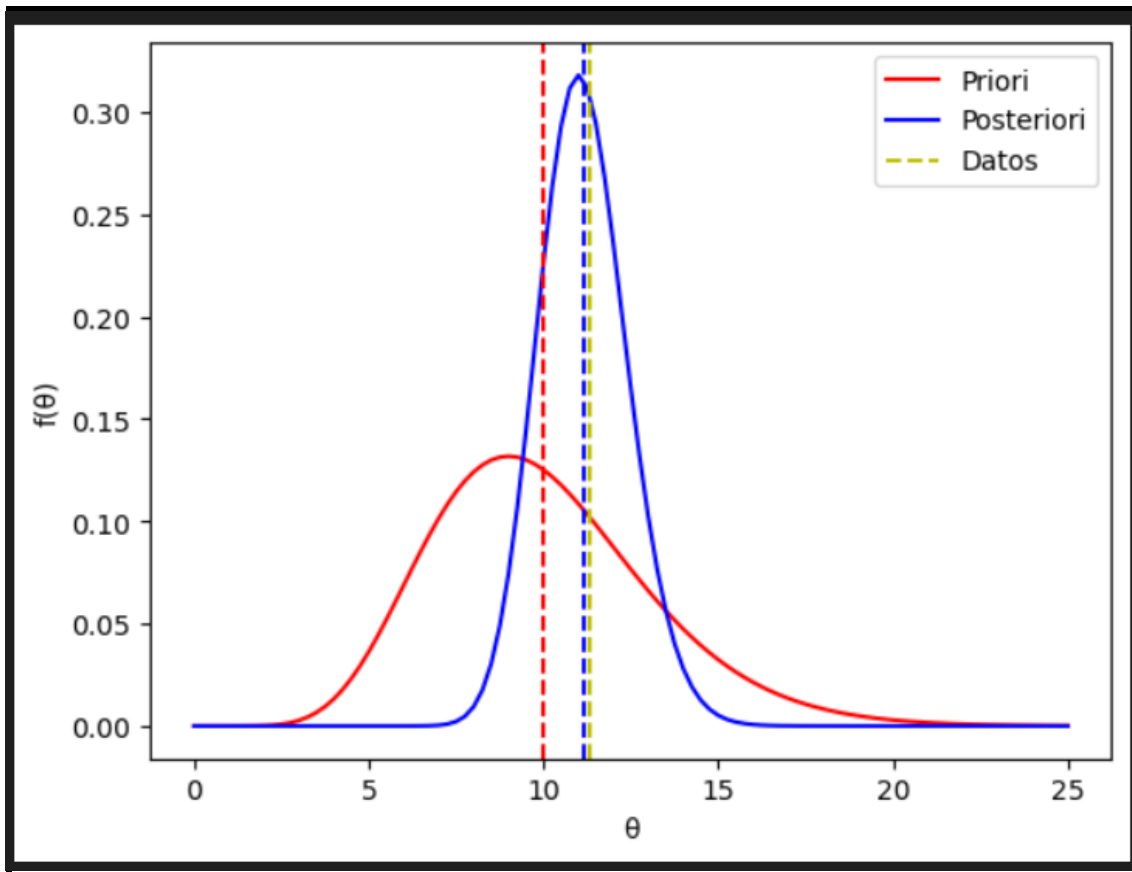
$$f(\theta = \theta | X = x) \approx \theta^{77} * e^{-7\theta} * I(\theta > 0)$$

La idea es llegar a una gamma (familia conjugada de la poisson). La distribución gamma es la siguiente:

$$\Gamma(y; \alpha, \beta) = \frac{\alpha^\beta}{\Gamma(\alpha)} * y^{\alpha-1} * e^{-\beta y} * I(y > 0)$$

- a. Así, la distribución a posteriori queda:

$$f(\theta = \theta | X = x) \sim \Gamma(\alpha = 78, \beta = 7) * I(\theta > 0)$$



- b. La probabilidad marginal de que X tome un mayor a 30 la obtendré integrando la condicional dado un θ en todo el espectro de θ y luego sumando para los valores de X que apliquen:

$$P(X > 30) = 1 - P(X \leq 30)$$

$$P(X > 30) = 1 - \sum_{x=0}^{30} \int_{\theta=0}^{\infty} L(\theta) * f(\theta = \theta | X = x) d\theta$$

$$P(X > 30) = 1 - \sum_{x=0}^{30} \int_{\theta=0}^{\infty} \frac{\theta^x * e^{-\theta}}{x!} \frac{7^{78}}{\Gamma(78)} * \theta^{77} * e^{-7\theta} d\theta$$

$$P(X > 30) = 1 - \sum_{x=0}^{30} \frac{7^{78}}{77! * x!} * \int_{\theta=0}^{\infty} \theta^{77+x} * e^{-8\theta} d\theta$$

Calculando numéricamente da $P(X > 30) = 1 - 0,9999942 = 5.8 * 10^{-6}$.