Probabilidad y Estadística para Inteligencia Artificial

Resolución de examen PEIA 2024 a15

Nombre: Juan Pablo Schamun

1. Voy a definir a **X** como la variable aleatoria del tamaño de una imagen en pixeles. Si bien los píxeles son números enteros, y en rigor **X** debería ser una variable discreta, para efectos de este problema **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 **Y** como la variable aleatoria del tipo de la imagen. **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:

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

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

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

$$Y \epsilon \{perro, gato\};$$

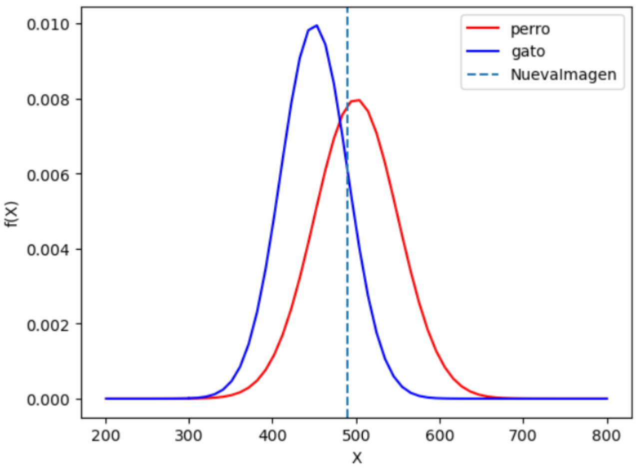
   La probabilidad de que una imagen sea de un determinado tipo dado que **X** toma un valor **x** lo voy a determinar con el teorema de Bayes:

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

   Siendo $P(Y = y)$ la probabilidad marginal de que **Y** adopte un valor de **y**, es decir la proporción de imágenes que son del tipo **y** entre todas las que toma Sonia.
   $f(X = x)$ es la probabilidad marginal de que **X** adopte un valor de **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 pixeles.
   Gráficamente, hasta ahora tenemos algo así:



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

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

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

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

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

$$X \sim desconocida; X > 0$$

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

$$\boldsymbol{H_0: \mu \leq \mu_0}$$
$$\boldsymbol{H_1: \mu > \mu_0}$$

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

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

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

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**

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

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.

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

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

$$\lambda > 0$$

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

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

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:

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

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

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

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

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

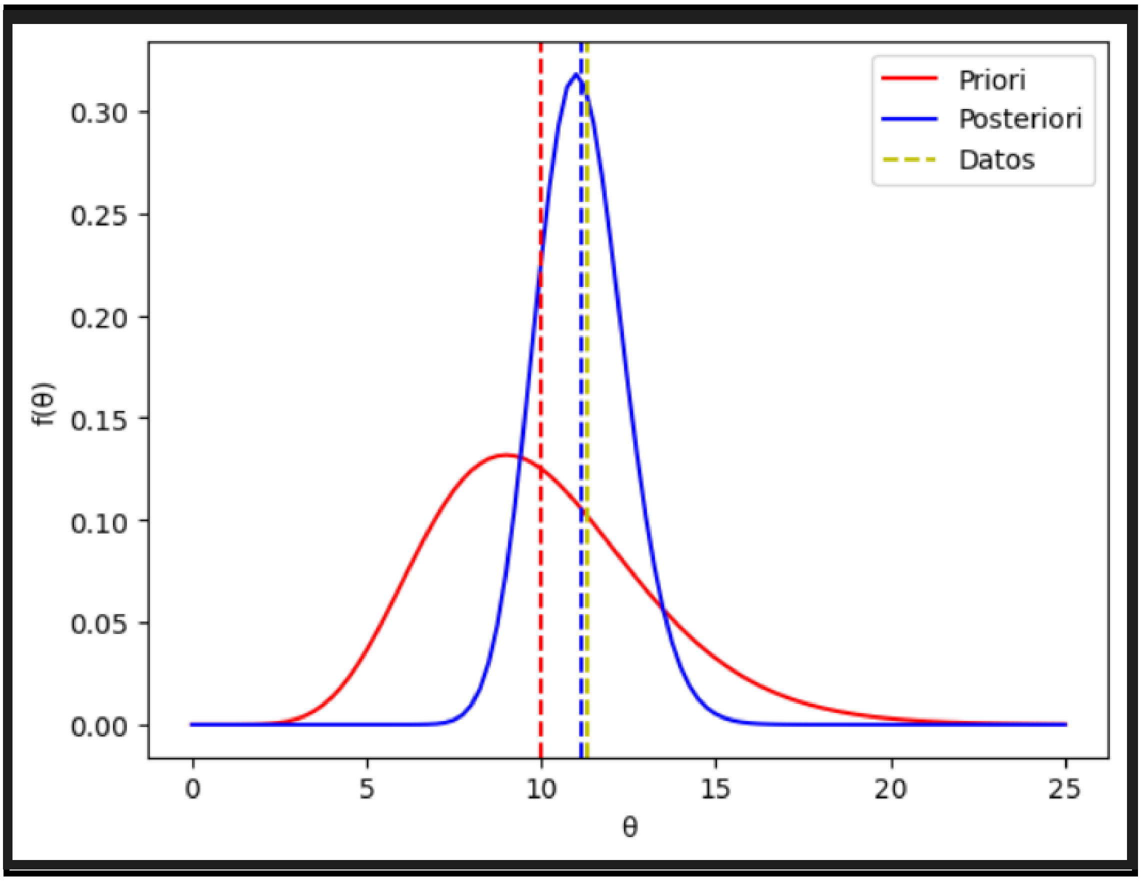
$$f(\Theta = \theta \mid X = x) \approx \theta^{77} * e^{-7\theta} * I(\theta > 0)$$

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:

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



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:

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

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

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

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

No llego con el tiempo a calcular la integral. Dejo planteado.