# Probabilidad y Estadística para IA

*"Examen Final"*

Facultad de Ingeniería de la Universidad de Buenos Aires
Laboratorio de Sistemas Embebidos
David Canal
Abril 2024

## 1. Consigna de trabajo

(1) Sonia está trabajando en un proyecto de IA que utiliza redes neuronales para clasificar imágenes. Para entrenar el modelo, utiliza un conjunto de datos de imágenes etiquetadas con 2 clases: "Perro" y "Gato". La distribución de los tamaños de las imágenes en el conjunto de datos sigue una distribución normal con una media de 500 píxeles y una desviación estándar de 50 píxeles para la imagen de "Perro", y una media de 450 píxeles y una desviación estándar de 40 píxeles para las imágenes de "Gato".

Si Sonia toma una nueva imagen y la misma tiene un tamaño de 490 píxeles. ¿Cuál es la probabilidad de que pertenezca a la clase "Perro"?

(2) Un desarrollador web está interesado en determinar si el tiempo promedio en el que los usuarios pasan en un sitio web que él construyó, ha aumentado en los últimos meses. Para ello, recopiló una muestra de 50 usuarios y registró el tiempo que cada uno pasó en el sitio web durante su última visita, obteniendo una media muestral de 6.5 minutos y un desvío estándar muestral de 1.5 minutos.

El desarrollador sabe que la media histórica del tiempo de permanencia en el sitio web ha sido de 6 minutos. Se desea determinar con un nivel de significancia del 5% si el tiempo promedio ha aumentado.

a. Escribir las hipótesis nula y alternativa

b. Calcular el test (definir el estadístico a usar y el valor crítico)

c. ¿Cuál es la conclusión del desarrollador web?

(3) Un mayorista requiere caracterizar el stock diario X de un determinado producto. Se sabe que el stock $X|\Lambda = \lambda$ se distribuye como una Poisson ($\lambda$), para un $\lambda > 0$ determinado. La logística se planifica en base a la distribución a priori $\Lambda \sim Gamma(10,1)$. Se registraron a lo largo de 6 días los siguientes stocks i.i.d x = [20,5,6,30,2,5].

a. Encontrar la distribución a posteriori de $\Lambda$.

b. Dada la muestra, estime la probabilidad de que una nueva observación X sea mayor que 30.

## 2. Respuestas

### 2.1 Ejercicio 1

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

X: “Extraigo una foto de un perro”.

Dado que la distribución de Bernoulli modela experimentos aleatorios con dos resultados posibles: éxito (extraigo una foto de un perro - x = 1) o fracaso (extraigo una foto de un gato - x = 0). Suponemos que ambos eventos son igualmente de probables, es decir, p = 0.5. Luego tenemos:

$$X \sim Bernoulli\ (p),$$

donde:

$$P(X = x) = p^x(1-p)^{1-x}, x = 0 \text{ ó } 1 \ (2.1),$$

por otro lado, tenemos que representar la probabilidad de que dado que extraigo una foto específica (perro o gato), cuál será la probabilidad de que tenga un tamaño en pixeles determinado. Para ellos definimos la siguiente variable:

Px: “Tengo una imagen de px pixeles”,

donde:

$$Px|\ X = 1 \sim N(\mu_{perro}, {\sigma_{perro}}^2),$$

$$P_{Px|\ X\ =\ 1}(px) = \frac{1}{\sqrt{2\pi}\sigma_{perro}} e^{\frac{-(px-\mu_{perro,})^2}{2{\sigma_{perro}}^2}} \ (2.2),$$

$$Px|\ X = 0 \sim N(\mu_{gato}, {\sigma_{gato}}^2),$$

$$P_{Px|\ X\ =\ 0}(px) = \frac{1}{\sqrt{2\pi}\sigma_{gato}} e^{\frac{-(px-\mu_{gato,})^2}{2{\sigma_{gato}}^2}} \ (2.3),$$

siendo: $\sigma_{perro} = 50,\ \sigma_{gato} = 40, \mu_{perro} = 500\ y\ \mu_{gato} = 450.$

El problema nos pide calcular la probabilidad de que una imagen sea de un perro, dado que si tamaño en píxeles es de px = 490. Aplicando el teorema de Bayes tenemos:

$$P(X|Px) = \frac{P(Px|X)P(X)}{P(Px)} \ (2.4),$$

Tenemos luego:

$$P_{Px|\,X=1}(490) = \frac{1}{\sqrt{2\pi} * 50} e^{\frac{-(490-500)^2}{2*50^2}} = 7.821 * 10^{-3},$$

$$P_{Px|\,X=0}(490) = \frac{1}{\sqrt{2\pi} * 40} e^{\frac{-(490-450)^2}{2*40^2}} = 6.049 * 10^{-3}$$

Por otro lado, considerando la probabilidad total, tenemos:

$$\begin{aligned} P(Px = 490) &= P_{Px|\,X=1}(490)P(X = 1) + P_{Px|\,X=0}(490)P(X = 0) \\ &= 7.821 * 10^{-3} * 0.5 + 6.049 * 10^{-3} * 0.5 \\ &= 6.935 * 10^{-3} \end{aligned}$$

Finalmente reemplazando en (2.4), tenemos:

$$P(X = 1\,|Px = 490) = \frac{P_{Px|\,X=1}(490)P(X=1)}{P(Px=490)}$$

$$P(X = 1\,|Px = 490) = \frac{7.821 * 10^{-3} * 0.5}{6.935 * 10^{-3}} = 0.564$$

Por último, llegamos a la conclusión de que la probabilidad de que una imagen de 490 píxeles sea de un perro es de 56.4%.

## 2.2 Ejercicio 2

### 2.2.1 Punto a

Dado que estamos interesados en determinar si el tiempo promedio que los usuarios pasan en el sitio web ha aumentado. Por lo tanto, nuestra hipótesis nula alternativa será las siguientes:

- **Hipótesis Nula ($H_0$):** El tiempo promedio que los usuarios pasan en el sitio web no ha aumentado ($\mu_0 = 6\,minutos$).

$$H_0: \mu = \mu_0$$

- **Hipótesis alternativa ($H_1$):** El tiempo promedio que los usuarios pasan en el sitio web ha aumentado, es decir, la media poblacional **(**$\mu$) es mayor que la media histórica ($\mu > 6\,minutos$).

$$H_1: \mu > \mu_0$$

donde:

$\underline{X} = 6.5$ minutos es la media muestral.

$\mu_0 = 6$ minutos es la media histórica.

s = 1.5 minutos es el desvío estándar muestral.

n = 50 es el tamaño de la muestra.

### 2.2.2 Punto b

Para un nivel de significancia del 5% y con una cola derecha (ya que estamos buscando si la media ha aumentado), el valor crítico de t en una distribución t de Student con n - 1 grados de libertad es:

$$t_{crítico} = t_{\alpha,n-1}\ (2.5),$$

siendo $\alpha = 0.05\ y\ n - 1 = 49$. Lo que sigue es, calcular el valor crítico de t y el estadístico t para toma una decisión:

- **Valor crítico de t**

  Para $\alpha = 0.05\ y\ n - 1 = 49$, el valor crítico de t (calculado en python con la punción t.ppf haciendo "from scipy.stats import t) es:

$$t_{crítico} = 1.676$$

- **Estadístico t**

$$t = \frac{6.5-6}{\frac{1.5}{\sqrt{50}}} = 2.358$$

- **Decisión**

  Como $t = 2.358 > t_{crítico} = 1.676$, rechazamos la hipótesis nula $H_0$.

### 2.2.3 Punto c

Hay evidencia suficiente para concluir que el tiempo promedio que los usuarios pasan en el sitio web ha aumentado.

### 2.3 Ejercicio 3

Se tiene que la variable aleatoria X: "Stock diario de un producto dado", sigue una distribución Poisson ($X \sim Poisson(\lambda)$), y la logística $\lambda$ sigue una distribución gamma ($\Lambda \sim Gamma(10,1)$). Tenemos entonces:

$$P_{X|\Lambda=\lambda}(x) = \frac{\lambda^x e^{-\lambda}}{x!}\ (2.6)$$

$$\pi(\lambda) = \Gamma(\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\lambda\beta}\ (2.7)$$

**2.3.1 Punto a:** "Distribución a posteriori"

$$P_{\Lambda|X=x}(\lambda) \propto \prod_{i=1}^{n} P_{X|\Lambda=\lambda}(x)\,\pi(\lambda)$$

$$P_{\Lambda|X=x}(\lambda) \propto \prod_{i=1}^{n} \left(\frac{\lambda^{x_i} e^{-\lambda}}{x_i!}\right) \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda\beta}$$

$$P_{\Lambda|X=x}(\lambda) \propto \lambda^{\alpha+\sum_{i=1}^{n} x_i - 1} e^{-\lambda(\beta+n)},$$

siendo $\alpha = 10,\ \beta = 1$ y $\sum_{i=1}^{n} x_i = 20 + 5 + 6 + 30 + 2 + 5 = 68$. Tenemos entonces:

$$P_{\Lambda|X=x}(\lambda) \propto \lambda^{10+68-1} e^{-\lambda(1+n)}$$

$$P_{\Lambda|X=x}(\lambda) \propto \lambda^{78-1} e^{-7\lambda} \propto \Gamma(78,7)$$

Se concluye que $\Lambda|X = x \sim \Gamma(78,7)$.

**2.3.2 Punto b**

Para estimar la probabilidad de que una nueva observación X sea mayor a 30, procedemos de la siguiente manera:

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

$$= 1 - \int_0^{\infty} \left(\sum_{i=0}^{30} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}\right) \frac{7^{78}}{\Gamma(78)} \lambda^{77} e^{-7\lambda}\, d\lambda$$

$$= 1 - \sum_{i=0}^{30} \int_0^{\infty} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \frac{7^{78}}{\Gamma(78)} \lambda^{77} e^{-7\lambda}\, d\lambda$$

$$= 1 - \sum_{i=0}^{30} \int_0^{\infty} \frac{7^{78}}{x_i!\,\Gamma(78)} \lambda^{x_i+77} e^{-8\lambda}\, d\lambda$$

$$= 1 - \sum_{i=0}^{30} \frac{7^{78}}{x_i!\,\Gamma(78)} \int_0^{\infty} \lambda^{x_i+77} e^{-8\lambda}\, d\lambda \quad (2.8)$$

Construyendo la distribución $\Gamma(78 + x_i, 8)$:

$$\Gamma(78 + x_i, 8) = \frac{8^{78+x_i}}{\Gamma(78+x_i)} \lambda^{78+x_i-1} e^{-8\lambda},$$

$$\int_0^{\infty} \Gamma(78 + x_i, 8)\, d\lambda = 1 \Rightarrow 1 = \frac{8^{78+x_i}}{\Gamma(78 + x_i)} \int_0^{\infty} \lambda^{78+x_i-1} e^{-8\lambda} d\lambda\,,$$

$$\Rightarrow \frac{\Gamma(78 + x_i)}{8^{78+x_i}} = \int_0^{\infty} \lambda^{78+x_i-1} e^{-8\lambda} d\lambda \quad (2.9),$$

Reemplazando (2.9) y (2.10) en (2.9), obtenemos:

$$P(X > 30) = 1 - \sum_{i=0}^{30} \frac{7^{78}}{x_i!\,\Gamma(78)} \frac{\Gamma(78 + x_i)}{8^{78 + x_i}}$$

Por propiedades de la función gamma, tenemos que:

$$\Gamma(78 + xi) = \prod_{i=0}^{x_i - 1} (78 + i)\,\Gamma(78)$$

Teneos entonces:

$$\begin{aligned} P(X > 30) &= 1 - \sum_{i=0}^{30} \frac{7^{78}}{x_i!\,\Gamma(78)} \frac{\prod_{i=0}^{x_i-1}(78 + i)\,\Gamma(78)}{8^{78 + x_i}} \\ &= 1 - \sum_{i=0}^{30} \frac{7^{78} \prod_{i=0}^{x_i-1}(78 + i)}{x_i!\,8^{78 + x_i}} \\ &= 1 - \frac{7^{78}}{8^{78}} \sum_{i=0}^{30} \frac{\prod_{i=0}^{x_i-1}(78 + i)}{x_i!\,8^{x_i}} \\ &= 1 - 0.999 = 0.00100 \end{aligned}$$

Por lo tanto, se concluye que la probabilidad de que una nueva observación X sea mayor a 30 es del 0.1%.