
Matemáticas para la Ciencia de Datos

Profesor: Dra. Briceyda B. Delgado López

Alumno: Miroslava Sandria Yong

Tarea 6. Teorema de Bayes.

> Los siguientes ejercicios se resuelven aplicando el Teorema de Bayes. Sugerencia: Defina los eventos y realice un diagrama que muestre las diferentes casos.
>
> 1. Un análisis de sangre de laboratorio es efectivo en un 99 % en la detección de una determinada enfermedad cuando, de hecho, está presente. Sin embargo, el análisis también da un resultado de falso positivo para el 1 % de las personas sanas analizadas. Si el 0.5 % de la poblaci´on padece realmente la enfermedad, ¿cuál es la probabilidad de que una persona la padezca si el resultado de la prueba es positivo?

El teorema de Bayes establece que:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Donde:

- $P(A|B)$ es la probabilidad de que ocurra A dado que ha ocurrido B.
- $P(B|A)$ es la probabilidad de que ocurra B dado que A ocurre.
- $P(A)$ es la probabilidad de que ocurra A.
- $P(B)$ es la probabilidad de que ocurra B, y se puede descomponer como:

$$P(B) = P(B|A) \cdot P(A) + P(B|\neg A) \cdot P(\neg A)$$

Ejercicio 1

Nos dan los siguientes datos:

- La prueba es efectiva en un 99% si la persona tiene la enfermedad, es decir:

$$P(positivo|enfermo) = 0.99$$

- El 1% de las personas sanas obtienen un falso positivo, es decir:

$$P(positivo|sano) = 0.01$$

- El 0.5% de la población realmente padece la enfermedad, es decir:

$$P(enfermo) = 0.005$$

  Por lo tanto, la probabilidad de que una persona esté sana es:

$$P(sano) = 1 - P(enfermo) = 0.995$$

Queremos encontrar la probabilidad de que una persona esté enferma dado que el resultado de la prueba es positivo, es decir, $P(enfermo|positivo)$.

Usando el teorema de Bayes:

$$P(enfermo|positivo) = \frac{P(positivo|enfermo) \cdot P(enfermo)}{P(positivo)}$$

Primero, necesitamos calcular $P(positivo)$, que se obtiene sumando los dos posibles casos en los que la prueba puede ser positiva (que la persona esté enferma o que no lo esté):

$$P(positivo) = P(positivo|enfermo) \cdot P(enfermo) + P(positivo|sano) \cdot P(sano)$$

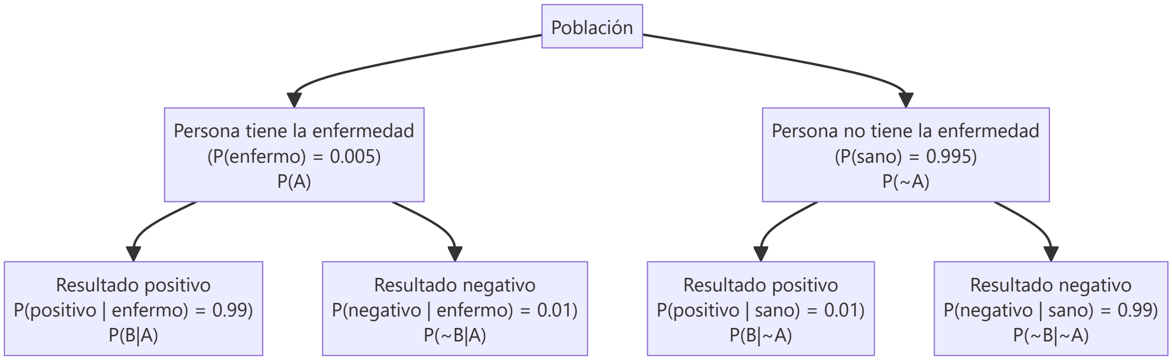
Sustituyendo los valores:

$$P(positivo) = (0.99 \cdot 0.005) + (0.01 \cdot 0.995) = 0.00495 + 0.00995 = 0.0149$$

Ahora calculamos $P(enfermo|positivo)$:

$$P(enfermo|positivo) = \frac{0.99 \cdot 0.005}{0.0149} = \frac{0.00495}{0.0149} \approx 0.332$$

Por lo tanto, la probabilidad de que una persona esté enferma dado que el resultado de la prueba es positivo es aproximadamente **0.332** o **33.2%**.



Conclusiones Ejercicio 1:

Este ejercicio nos muestra cómo la **sensibilidad** de una prueba (la capacidad de detectar correctamente una enfermedad) y la **tasa de falsos positivos** (cuando la prueba da positivo para personas sanas) pueden influir fuertemente en la interpretación del resultado de una prueba diagnóstica.

- Aunque la prueba tiene una alta precisión para detectar la enfermedad (99% de sensibilidad), la prevalencia de la enfermedad es muy baja (0.5% de la población).
- La probabilidad de que una persona esté realmente enferma dado que su prueba resultó positiva es **33.22%**. Esto significa que, aunque el resultado sea positivo, solo aproximadamente un tercio de las personas con un resultado positivo realmente padecen la enfermedad.
- Esto ocurre porque el pequeño porcentaje de falsos positivos (1%) aplicado a la gran mayoría de personas sanas (99.5%) genera una cantidad considerable de resultados positivos en personas sanas. En consecuencia, la tasa de falsos positivos influye mucho cuando la enfermedad es poco común.

Conclusión: Aun cuando una prueba tiene una alta precisión, si la enfermedad es rara, un resultado positivo no necesariamente implica que la persona esté enferma. Esto subraya la importancia de considerar la prevalencia de la enfermedad al interpretar los resultados de una prueba.

> El cáncer de próstata es el más frecuente entre los varones. Como indicador de si un varón padece cáncer de próstata, los médicos suelen realizar una prueba que mide el nivel de la proteína PSA (antígeno prostático espec´ıfico) que s´olo produce la gl´andula prostática. Aunque unos niveles m´as altos de PSA son indicativos de cáncer, la prueba es muy poco fiable. De hecho, la probabilidad de que un hombre no canceroso tenga un nivel elevado de PSA es de aproximadamente 0,135, probabilidad que aumenta a aproximadamente 0,268 si el hombre tiene cáncer. Si, bas´andose en otros factores, un mpédico está seguro en un 70 % de que un var´on tiene c´ancer de pr´ostata, ¿cuál es la probabilidad condicional de que tenga cáncer si
>
> (a) la prueba indica un nivel elevado de PSA?, (b) la prueba no indica un nivel elevado de PSA?
>
> Repita lo anterior, esta vez suponiendo que el m´edico cree inicialmente que existe hay un 30 % de probabilidades de que el hombre tenga cáncer de próstata

Ejercicio 2

Tenemos los siguientes datos:

- La probabilidad de que un hombre no canceroso tenga un nivel elevado de PSA es $P(elevado|no\ cáncer) = 0.135$.
- La probabilidad de que un hombre con cáncer tenga un nivel elevado de PSA es $P(elevado|cáncer) = 0.268$.
- La probabilidad de que un hombre tenga cáncer (según el médico) es $P(cáncer) = 0.70$ en el caso (a) y $P(cáncer) = 0.30$ en el caso (b).

Parte (a): Probabilidad condicional dado un nivel elevado de PSA

Usamos el teorema de Bayes para calcular $P(cáncer|elevado)$:

$$P(cáncer|elevado) = \frac{P(elevado|cáncer) \cdot P(cáncer)}{P(elevado)}$$

Primero, calculamos $P(elevado)$, que se obtiene sumando los dos posibles casos en los que los niveles de PSA pueden ser elevados:

$$P(elevado) = P(elevado|cáncer) \cdot P(cáncer) + P(elevado|no\ cáncer) \cdot P(no\ cáncer)$$

Caso 1 (70% de probabilidad de tener cáncer):

$$P(elevado) = (0.268 \cdot 0.70) + (0.135 \cdot 0.30) = 0.1876 + 0.0405 = 0.2281$$

Ahora, calculamos $P(cáncer|elevado)$:

$$P(cáncer|elevado) = \frac{0.268 \cdot 0.70}{0.2281} = \frac{0.1876}{0.2281} \approx 0.822$$

Por lo tanto, la probabilidad de que un hombre tenga cáncer dado que tiene un nivel elevado de PSA es aproximadamente **82.2%** si inicialmente se cree que hay un 70% de probabilidades de que tenga cáncer.

Caso 2 (30% de probabilidad de tener cáncer):

Primero, calculamos $P(elevado)$:

$$P(elevado) = (0.268 \cdot 0.30) + (0.135 \cdot 0.70) = 0.0804 + 0.0945 = 0.1749$$

Ahora, calculamos $P(cáncer|elevado)$:

$$P(cáncer|elevado) = \frac{0.268 \cdot 0.30}{0.1749} = \frac{0.0804}{0.1749} \approx 0.46$$

Por lo tanto, si inicialmente se cree que hay un 30% de probabilidades de que el hombre tenga cáncer, la probabilidad de que lo tenga si tiene un nivel elevado de PSA es aproximadamente **46%**.

Parte (b): Probabilidad condicional dado que no tiene un nivel elevado de PSA

Usamos el teorema de Bayes para calcular $P(cáncer|no\ elevado)$:

$$P(cáncer|no\ elevado) = \frac{P(no\ elevado|cáncer) \cdot P(cáncer)}{P(no\ elevado)}$$

Primero, necesitamos calcular $P(no\ elevado)$, que es la probabilidad de que no tenga niveles elevados de PSA. Esto se puede calcular como:

$$P(no\ elevado) = 1 - P(elevado)$$

Caso 1 (70% de probabilidad de tener cáncer):

$$P(no\ elevado) = 1 - 0.2281 = 0.7719$$

Ahora calculamos $P(no\ elevado|cáncer)$:

$$P(no\ elevado|cáncer) = 1 - P(elevado|cáncer) = 1 - 0.268 = 0.732$$

Finalmente, calculamos $P(cáncer|no\ elevado)$:

$$P(cáncer|no\ elevado) = \frac{0.732 \cdot 0.70}{0.7719} = \frac{0.5124}{0.7719} \approx 0.664$$

Por lo tanto, si inicialmente se cree que hay un 70% de probabilidades de que el hombre tenga cáncer, la probabilidad de que lo tenga si no tiene un nivel elevado de PSA es aproximadamente **66.4%**.

Caso 2 (30% de probabilidad de tener cáncer):

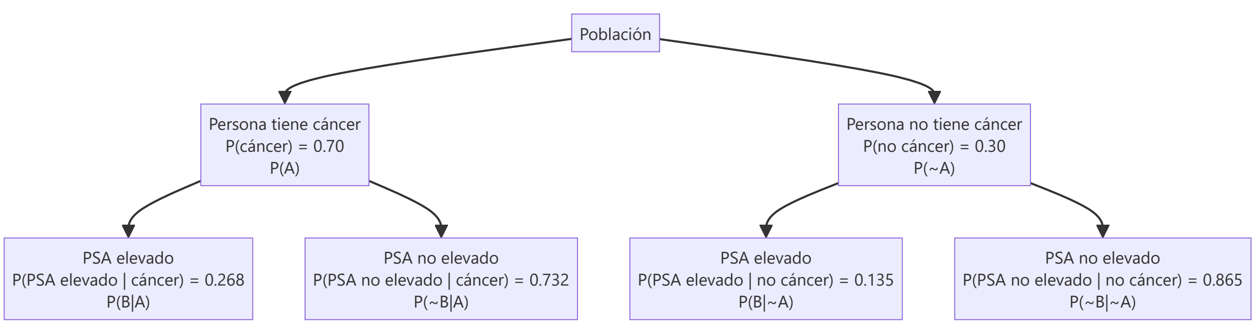
Primero, calculamos $P(no\ elevado)$:

$$P(no\ elevado) = 1 - 0.1749 = 0.8251$$

Ahora calculamos $P(cáncer|no\ elevado)$:

$$P(cáncer|no\ elevado) = \frac{0.732 \cdot 0.30}{0.8251} = \frac{0.2196}{0.8251} \approx 0.266$$

Por lo tanto, si inicialmente se cree que hay un 30% de probabilidades de que el hombre tenga cáncer, la probabilidad de que lo tenga si no tiene un nivel elevado de PSA es aproximadamente **26.6%**.



Conclusiones Ejercicio 2:

Este ejercicio analiza cómo las creencias previas del médico sobre la probabilidad de que un paciente tenga cáncer (es decir, la probabilidad inicial) afectan las conclusiones que se sacan después de una prueba de PSA.

Parte (a): Nivel elevado de PSA

- Si el médico cree inicialmente que hay un **70%** de probabilidades de que el paciente tenga cáncer, la probabilidad de que tenga cáncer dado un nivel elevado de PSA es **82.2%**.
- Si el médico cree inicialmente que hay un **30%** de probabilidades de que el paciente tenga cáncer, la probabilidad de que tenga cáncer dado un nivel elevado de PSA es **46%**.

Conclusión: Las creencias iniciales del médico (conocidas como las **probabilidades a priori**) afectan considerablemente la interpretación de un resultado elevado de PSA. Un nivel elevado de PSA aumenta la probabilidad de cáncer, pero cuánto depende de la probabilidad inicial que el médico asigne al cáncer.

Parte (b): Nivel no elevado de PSA

- Si el médico cree inicialmente que hay un **70%** de probabilidades de que el paciente tenga cáncer, la probabilidad de que tenga cáncer dado que no tiene un nivel elevado de PSA es **66.4%**.
- Si el médico cree inicialmente que hay un **30%** de probabilidades de que el paciente tenga cáncer, la probabilidad de que tenga cáncer dado que no tiene un nivel elevado de PSA es **26.6%**.

Conclusión: Un nivel no elevado de PSA reduce la probabilidad de que el paciente tenga cáncer, pero no la elimina por completo, especialmente si el médico tiene una alta probabilidad inicial de que el paciente padezca la enfermedad.

Conclusión general:

En ambos ejercicios, el **Teorema de Bayes** muestra cómo las probabilidades iniciales y los datos de pruebas diagnósticas interactúan para actualizar nuestras creencias sobre un evento (enfermedad o cáncer). Las **probabilidades condicionales** pueden cambiar significativamente dependiendo de la **prevalencia** de la condición y de las creencias previas sobre el paciente. Estos ejemplos destacan la importancia de usar correctamente la probabilidad en la interpretación de pruebas médicas.

Referencias bibliográficas

Dr. Juliho Castillo Colmenares. (2020). Matemáticas para las Ciencias de Datos. Cálculo de probabilidades básicas. https://aulavirtual.infotec.mx/pluginfile.php/105669/mod_bootstrapelements/intro/Calculo%20de%20probabilidades%20basicas.pdf

Rincon Luis, (2013). Introducción a la probabilidad, UNAM https://aulavirtual.infotec.mx/pluginfile.php/105669/mod_bootstrapelements/intro/Rincon%2C%20Luis.pdf

OpenAI. (2024). ChatGPT (GPT-4) LLM. https://chat.openai.com/

3blue1Brown (2019), Bayes theorem, the geometry of changing beliefs. [Youtube] https://www.youtube.com/watch?v=HZGCoVF3YvM&t

---

URL al repositorio de codigo de este documento (Github) : https://github.com/mirossy29/Maestria/blob/main/Matematicas/MiroslavaSandria-TAREA-6-Teorema_Bayes.ipynb