INFOTEC - MCDI

Materia: Matemáticas para la Ciencia de Datos
Profesor: Dra. Briceyda B. Delgado
Alumno: David Rodriguez Gutierrez

Tarea 6: Introducción a la probabilidad

> 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 población padece realmente la enfermedad, ¿cuál es la probabilidad de que una persona la padezca si el resultado de la prueba es positivo?

Ejercicio 1: Análisis de sangre y detección de la enfermedad

Definimos los siguientes eventos:

- A: La persona padece la enfermedad.
- B: El resultado de la prueba es positivo.

Datos proporcionados:

- $P(A) = 0.005$ (la prevalencia de la enfermedad en la población es del 0.5%).
- $P(B|A) = 0.99$ (la prueba detecta correctamente la enfermedad en un 99% de los casos).
- $P(B|\neg A) = 0.01$ (la prueba da un falso positivo en un 1% de los casos cuando la persona no tiene la enfermedad).
- $P(\neg A) = 1 - P(A) = 0.995$ (la probabilidad de que la persona no tenga la enfermedad).

Aplicamos el **Teorema de Bayes**:

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

Primero, calculamos $P(B)$ usando la ley de probabilidad total:

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

Sustituyendo los valores conocidos:

$$P(B) = (0.99 \cdot 0.005) + (0.01 \cdot 0.995)$$

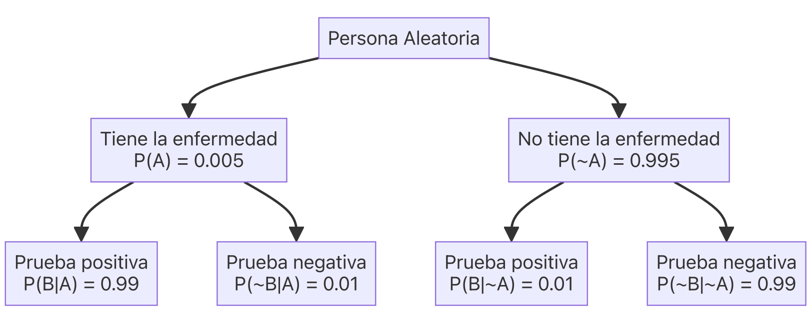
$$P(B) = 0.00495 + 0.00995 = 0.0149$$

Ahora, aplicamos el Teorema de Bayes:

$$P(A|B) = \frac{0.99 \cdot 0.005}{0.0149}$$

$$P(A|B) \approx \frac{0.00495}{0.0149} \approx 0.3322$$

Por lo tanto, la probabilidad de que una persona tenga la enfermedad si el resultado de la prueba es positivo es aproximadamente **33.22%**.



> 2. 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 específico) que sólo produce la glándula prostática. Aunque unos niveles más 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, basándose en otros factores, un médico está seguro en un 70% de que un varón tiene cáncer de próstata, ¿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 médico cree inicialmente que existe hay un 30% de probabilidades de que el hombre tenga cáncer de próstata.

Ejercicio 2: Cáncer de próstata y niveles elevados de PSA

Parte (a): Probabilidad de cáncer si la prueba indica un nivel elevado de PSA

Definimos los siguientes eventos:

- A: El hombre tiene cáncer.
- B: El nivel de PSA es elevado.

Datos proporcionados:

- $P(A) = 0.70$ (el médico cree que hay un 70% de probabilidad de que el hombre tenga cáncer).
- $P(B|A) = 0.268$ (la probabilidad de que el nivel de PSA sea elevado si el hombre tiene cáncer).
- $P(B|\neg A) = 0.135$ (la probabilidad de que el nivel de PSA sea elevado si el hombre no tiene cáncer).
- $P(\neg A) = 0.30$ (la probabilidad de que el hombre no tenga cáncer).

Aplicamos el **Teorema de Bayes**:

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

Primero, calculamos $P(B)$:

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

$$P(B) = (0.268 \cdot 0.70) + (0.135 \cdot 0.30)$$

$$P(B) = 0.1876 + 0.0405 = 0.2281$$

Ahora aplicamos el Teorema de Bayes:

$$P(A|B) = \frac{0.268 \cdot 0.70}{0.2281}$$

$$P(A|B) \approx \frac{0.1876}{0.2281} \approx 0.822$$

Por lo tanto, la probabilidad de que el hombre tenga cáncer si la prueba indica un nivel elevado de PSA es aproximadamente **82.2%**.

Parte (b): Probabilidad de cáncer si la prueba **no** indica un nivel elevado de PSA

Definimos los siguientes eventos:

- A: El hombre tiene cáncer.
- $\neg B$: El nivel de PSA no es elevado.

Queremos encontrar $P(A|\neg B)$.

Usamos la probabilidad complementaria:

$$P(\neg B|A) = 1 - P(B|A) = 1 - 0.268 = 0.732$$

$$P(\neg B|\neg A) = 1 - P(B|\neg A) = 1 - 0.135 = 0.865$$

Primero, calculamos $P(\neg B)$:

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

$$P(\neg B) = (0.732 \cdot 0.70) + (0.865 \cdot 0.30)$$

$$P(\neg B) = 0.5124 + 0.2595 = 0.7719$$

Ahora aplicamos el Teorema de Bayes:

$$P(A|\neg B) = \frac{P(\neg B|A) \cdot P(A)}{P(\neg B)} = \frac{0.732 \cdot 0.70}{0.7719}$$

$$P(A|\neg B) \approx \frac{0.5124}{0.7719} \approx 0.664$$

Por lo tanto, la probabilidad de que el hombre tenga cáncer si la prueba no indica un nivel elevado de PSA es aproximadamente **66.4%**.

Caso alternativo: Si el médico cree inicialmente que hay un 30% de probabilidades de que el hombre tenga cáncer

Si $P(A) = 0.30$ y $P(\neg A) = 0.70$, repetimos los cálculos:

Parte (a): Probabilidad de cáncer si el nivel de PSA es elevado

Calculamos $P(B)$:

$$P(B) = (0.268 \cdot 0.30) + (0.135 \cdot 0.70)$$

$$P(B) = 0.0804 + 0.0945 = 0.1749$$

Ahora aplicamos el Teorema de Bayes:

$$P(A|B) = \frac{0.268 \cdot 0.30}{0.1749} = \frac{0.0804}{0.1749} \approx 0.460$$

Por lo tanto, la probabilidad de que el hombre tenga cáncer si la prueba indica un nivel elevado de PSA es aproximadamente **46%**.

Parte (b): Probabilidad de cáncer si el nivel de PSA no es elevado

Calculamos $P(\neg B)$:

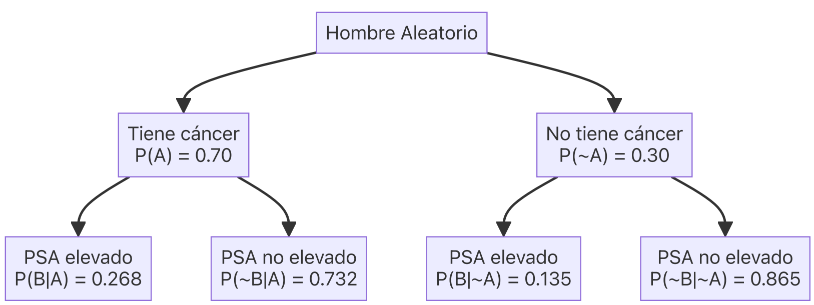
$$P(\neg B) = (0.732 \cdot 0.30) + (0.865 \cdot 0.70)$$

$$P(\neg B) = 0.2196 + 0.6055 = 0.8251$$

Ahora aplicamos el Teorema de Bayes:

$$P(A|\neg B) = \frac{0.732 \cdot 0.30}{0.8251} = \frac{0.2196}{0.8251} \approx 0.266$$

Por lo tanto, la probabilidad de que el hombre tenga cáncer si la prueba no indica un nivel elevado de PSA es aproximadamente **26.6%**.



Conclusiones del Ejercicio 1: Análisis de sangre y detección de la enfermedad

1. **Efectividad del análisis**: Aunque el análisis es muy preciso (99% de efectividad) en detectar la enfermedad cuando está presente ($P(B|A) = 0.99$), la baja prevalencia de la enfermedad en la población ($P(A) = 0.005$) juega un papel importante al calcular la probabilidad condicional de tener la enfermedad si el resultado es positivo.

2. **Probabilidad posterior**: La probabilidad de que una persona padezca la enfermedad dado un resultado positivo, calculada usando el Teorema de Bayes, es de **33.22%**:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} = \frac{0.99 \cdot 0.005}{0.0149} \approx 0.3322$$

   Esto significa que, aunque el resultado de la prueba es positivo, **solo un tercio de las personas con un resultado positivo realmente padecen la enfermedad**, lo que se debe principalmente a la baja prevalencia de la enfermedad en la población.

3. **Conclusión**: Este ejercicio subraya cómo una prueba muy precisa puede, no obstante, arrojar resultados donde la probabilidad de padecer la enfermedad sigue siendo relativamente baja en poblaciones donde la enfermedad es poco frecuente. Esto demuestra la importancia de considerar tanto la efectividad de la prueba como la prevalencia de la enfermedad en la interpretación de los resultados.

Conclusiones del Ejercicio 2: Cáncer de próstata y niveles elevados de PSA

Caso 1: El médico está 70% seguro de que el hombre tiene cáncer

1. **Probabilidad posterior con PSA elevado**: Si la prueba indica un nivel elevado de PSA, la probabilidad de que el hombre tenga cáncer aumenta a **82.2%**:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} = \frac{0.268 \cdot 0.70}{0.2281} \approx 0.822$$

2. **Probabilidad posterior sin PSA elevado**: Si la prueba **no** indica un nivel elevado de PSA, la probabilidad de que el hombre tenga cáncer disminuye a **66.4%**:

$$P(A|\neg B) = \frac{P(\neg B|A) \cdot P(A)}{P(\neg B)} = \frac{0.732 \cdot 0.70}{0.7719} \approx 0.664$$

3. **Conclusión**: La prueba de PSA tiene un impacto en las probabilidades, pero no es concluyente. Incluso sin un nivel elevado de PSA, la probabilidad de tener cáncer sigue siendo relativamente alta (66.4%), lo que indica que la prueba de PSA por sí sola no es suficiente para tomar decisiones definitivas.

Caso 2: El médico cree que hay 30% de probabilidades de que el hombre tenga cáncer

1. **Probabilidad posterior con PSA elevado**: Si la prueba de PSA indica niveles elevados, la probabilidad de que el hombre tenga cáncer es del **46%**:

$$P(A|B) = \frac{0.268 \cdot 0.30}{0.1749} \approx 0.460$$

2. **Probabilidad posterior sin PSA elevado**: Si la prueba **no** indica un nivel elevado de PSA, la probabilidad de cáncer disminuye a **26.6%**:

$$P(A|\neg B) = \frac{0.732 \cdot 0.30}{0.8251} \approx 0.266$$

3. **Conclusión**: Cuando la probabilidad inicial de cáncer es menor (30%), la prueba de PSA tiene un menor impacto en las probabilidades finales. Un resultado positivo aumenta la probabilidad de cáncer a **46%**, mientras que un resultado negativo la reduce a **26.6%**. Esto demuestra que la prueba es más útil en escenarios donde la probabilidad inicial de cáncer es más alta.

Referencias

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

Eliezer S. Yudkowsky (2003), An Intuitive Explanation of Bayes' Theorem. https://www.yudkowsky.net/rational/bayes

Numberphile (2015), How random is a coin toss?. [Youtube] https://www.youtube.com/watch?v=AYnJv68T3MM&

Wireless Philosophy (2016), CRITICAL THINKING - Fundamentals: Bayes' Theorem. [Youtube] https://www.youtube.com/watch?v=OqmJhPQYRc8&t

Numberphile (2016), Are you REALLY sick? (false positives) - Numberphile. [Youtube] https://www.youtube.com/watch?v=M8xlOm2wPAA&t

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

3blue1Brown (2019), The quick proof of Bayes' theorem. [Youtube] https://www.youtube.com/watch?v=U_85TaXbeIo&t

3blue1Brown (2020), Binomial distributions | Probabilities of probabilities, part 1. [Youtube] https://www.youtube.com/watch?v=8idr1WZ1A7Q&t

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

URL al repositorio de codigo de este documento (Github):

https://github.com/davidlobolobo/data_science/blob/main/Math/MCDI_MAT_David_Rodriguez_Tarea_6.ipynb

---