

Teorema de Bayes

Gabriela Sánchez Y.

5064

1. Teorema de Bayes

Sean H_1, H_2, \dots, H_m un conjunto de eventos disjuntos por pares llamados *hipótesis*, tales que el espacio muestral satisface que $\Omega = H_1 \cup H_2 \cup \dots \cup H_m$. Además sea E un evento llamado *evidencia* que proporciona información sobre cuál hipótesis es correcta.

Antes de recibir la evidencia se tiene el conjunto de probabilidades previas $P(H_1), \dots, P(H_m)$ para las hipótesis. Si se conoce la hipótesis correcta entonces se conoce $P(E | H_i)$ para todo i . Es de interés calcular las probabilidades para las hipótesis dada la evidencia. Es decir, se desea encontrar las probabilidades posteriores. Estas probabilidades se encuentran mediante la fórmula de Bayes [6] expresada en la ecuación (1)

$$P(H_i | E) = \frac{P(H_i)P(E | H_i)}{\sum_{k=1}^m P(H_k)P(E | H_k)}. \quad (1)$$

En este trabajo se desea analizar esta fórmula para el caso específico de pruebas para la detección de alguna enfermedad. En el contexto del problema se tienen dos hipótesis: estar enfermo que se denotará como C^+ y no estar enfermo denotado por C^- . La evidencia, será el resultado de la prueba, *positivo* o *negativo*.

Si se desea saber cuál es la probabilidad de que una persona realmente esté enferma dado que el resultado de la prueba es positivo (también llamado *valor predictivo positivo*), se reescribe la fórmula (1) y se obtiene la ecuación (2)

$$P(C^+ | +) = \frac{P(C^+)P(+ | C^+)}{P(C^+)P(+ | C^+) + P(C^-)P(+ | C^-)}, \quad (2)$$

donde $P(+ | C^+)$ indica la probabilidad de que la prueba dé un resultado positivo cuando la persona tiene la enfermedad, es decir, la probabilidad de obtener un *verdadero positivo* y, $P(+ | C^-)$ indica la probabilidad de que una persona sin la enfermedad obtenga un resultado positivo en la prueba, es decir, la probabilidad de obtener un *falso positivo*.

De manera análoga, la probabilidad de que una persona no esté enferma dado que el resultado de la prueba es negativo, (*valor predictivo negativo*) se puede obtener mediante la ecuación (3)

$$P(C^- | -) = \frac{P(C^-)P(- | C^-)}{P(C^-)P(- | C^-) + P(C^+)P(- | C^+)}, \quad (3)$$

donde $P(- | C^-)$ es la probabilidad de que una persona sin la enfermedad obtenga un resultado negativo en la prueba (*verdadero negativo*) y $P(- | C^+)$ es la probabilidad de que una persona enferma obtenga un resultado negativo en la prueba (*falso negativo*).

Las probabilidades condicionadas $P(+ | C^+)$ y $P(- | C^-)$ expresan la *sensibilidad* (capacidad de la prueba para detectar la enfermedad) y *especificidad* de la prueba (capacidad de la prueba de detectar a los individuos sanos) [7]. De manera que aún se pueden reescribir [4] las ecuaciones (2) y (3) como (4) y (5)

$$P(C^+ | +) = \frac{(\text{prevalencia}) \cdot \text{sensibilidad}}{(\text{prevalencia}) \cdot \text{sensibilidad} + (1 - \text{prevalencia}) \cdot (1 - \text{especificidad})}, \quad (4)$$

$$P(C^- | -) = \frac{(1 - \text{prevalencia}) \cdot \text{especificidad}}{(1 - \text{prevalencia}) \cdot \text{especificidad} + \text{prevalencia} \cdot (1 - \text{sensibilidad})}. \quad (5)$$

Además se puede determinar la *precisión* y *exactitud* [7] de la prueba mediante las ecuaciones (6) y (7)

$$\text{Precisión} = \frac{\text{verdaderos positivos}}{\text{verdaderos positivos} + \text{falsos positivos}}, \quad (6)$$

$$\text{Exactitud} = \frac{\text{verdaderos positivos} + \text{verdaderos negativos}}{\text{total de resultados}}. \quad (7)$$

2. Pruebas Covid-19

Existen dos diferentes tipos de pruebas para la detección del virus SARS-CoV-2, causante de la enfermedad Covid-19: las pruebas de diagnóstico, como la prueba RT-PCR que diagnostica una infección activa de coronavirus y las pruebas de anticuerpos, que muestran si una persona ha sido infectada por el coronavirus en el pasado [3].

Utilizando sólo la información de la prueba RT-PCR, se desea analizar las implicaciones del teorema de Bayes para el caso específico de la detección de Covid-19. Usando la notación descrita previamente, C^+ representará la población enferma y C^- la población sana. De esta manera, conociendo la especificidad y sensibilidad de esta prueba, se pueden determinar el valor predictivo positivo y negativo.

No se encontró un valor específico para la especificidad de la prueba sin embargo se estima que es muy cercana al 100 %. Para fines prácticos se fijará en un 99 %. Tampoco se encontró un valor para la sensibilidad y el rango de valores es variado además de que depende del lugar de la muestra y de la carga viral. De acuerdo a la información de diversas fuentes [1, 5], este valor oscila entre 31 % – 70 %.

Con esta información y los casos de Covid-19 en el estado de Michoacán [2] se determinarán los valores predictivos positivo y negativo.

La información proporcionada al 26 de octubre indica que de un total de 57050 pruebas realizadas, 24,499 han resultado positivas y 32,551 han resultado negativas. Si la población del estado es alrededor de 5 millones, la prevalencia será entonces de un 0.049. Usando estos datos y considerando una especificidad del 70 %, se calcula el valor predictivo positivo y negativo a partir de las

ecuaciones (4) y (5):

$$P(C^+ | +) = \frac{(0.049) \cdot (0.7)}{(0.049) \cdot (0.7) + (0.951) \cdot (0.01)} = 0.78,$$

$$P(C^- | -) = \frac{(0.951) \cdot (0.99)}{(0.951) \cdot (0.99) + (0.049) \cdot (0.3)} = 0.98.$$

Lo que está indicando el primer resultado es la probabilidad de que una persona que haya dado positivo en la prueba realmente esté enferma, mientras que el segundo resultado indica la probabilidad de que una persona que haya dado negativo en la prueba realmente esté sana.

Note que según los resultados obtenidos, la probabilidad de obtener un falso negativo $[1 - P(C^- | -)] = 0.02$ es baja, mientras que la probabilidad de obtener un falso positivo es más alta $[1 - P(C^+ | +)] = 0.22$.

Referencias

- [1] Gar Ming Chan. Bayes' theorem, COVID19, and screening tests. *The American Journal of Emergency Medicine*, 2020.
- [2] Gobierno del estado de Michoacán. Michoacán coronavirus Covid-19. <https://michoacancoronavirus.com/>.
- [3] FDA. Conceptos básicos de las pruebas para el coronavirus. <https://www.fda.gov/consumers/articulos-en-espanol/conceptos-basicos-de-las-pruebas-para-el-coronavirus>, 2020.
- [4] Raúl Fernández Regalado. El teorema de Bayes y su utilización en la interpretación de las pruebas diagnósticas en el laboratorio clínico. *Revista Cubana de Investigaciones Biomédicas*, 28, 2009.
- [5] Chester B Good, Inmaculada Hernandez, and Kenneth Smith. Interpreting COVID-19 Test Results: a Bayesian Approach. *Journal of General Internal Medicine*, 2020.
- [6] Charles M. Grinstead and J. Laurie Snell. *Introduction to Probability*. AMS, 2003.
- [7] Towards Data Science. COVID-19, Bayes' theorem and taking probabilistic decisions. <https://towardsdatascience.com/covid-19-bayes-theorem-and-taking-data-driven-decisions-part-1-b61e2c2b3bea>, 2020.