

Tarea 8

COVID-19 y teorema de Bayes

Erick Cervantes Mendieta

Matrícula: 2032430

Modelos Probabilistas Aplicados

Octubre 2020

1. Teorema de Bayes

El teorema de Bayes sirve para calcular la probabilidad condicional a *posteriori*[3]. En otras palabras, sirve para determinar la probabilidad de *una* de las causas, puesto que ya se observó el *efecto*, y se calcula de la manera siguiente:

$$P(B_j | A) = \frac{P(B_j)P(A | B_j)}{\sum_{i=1}^n P(B_i)P(A | B_i)},$$

donde B_j es la causa j y A es el efecto, y B_i es la causa i , con $i = 1, 2, \dots, n$. Utilizando la ley de la multiplicación tenemos que:

$$P(B_j | A) = \frac{P(A \cap B_j)}{P(A)}.$$

El denominador del teorema de Bayes se conoce como *probabilidad total*. Este teorema tiene las siguientes características: las causas (B_i) son mutuamente excluyentes y la suma de las probabilidades asociadas a las causas es igual a uno.

Para ejemplificar un ejemplo de este teorema, se consideró el dato del Sistema Especial de Vigilancia Epidemiológica de VIH/SIDA, el cual dice que en México aproximadamente existen 1.1395×10^{-5} jóvenes de entre 15 y 24 años de edad infectados con el virus del VIH en el 2020, existe una prueba llamada Exacto Pro de un solo uso, rápido y de flujo continuo, para la detección de anticuerpos contra el virus de inmunodeficiencia humana tipo 1 y tipo 2 en sangre humana total con EDTA, sangre de punción dactilar, suero o plasma con EDTA. Dicha prueba asegura que se puede determinar si una muestra presenta anticuerpos de VIH en solo un minuto con una sensibilidad igual al 99.52 % y una especificidad del 98.85 % (<https://kabila.mx/pruebasrapidas/prueba-vih-exacto/>). Suponga que un joven en ese rango de edad adquirió dicha prueba y ha dado positivo. ¿Cuál es la probabilidad de que el joven

tenga una infección por VIH?

Sea B_1 = joven con VIH, B_2 = un joven que no tiene VIH y sea A = resultado positivo de la prueba del VIH. Luego entonces, $P(B_1) = 0.00001139$, $P(B_2) = 0.9999886$, $P(A | B_1) = 0.9952$ y $P(A | B_2) = 0.0115$, usando el teorema de Bayes obtenemos:

$$\begin{aligned} P(B_1 | A) &= \frac{P(B_1)P(A | B_1)}{P(B_1)P(A | B_1) + P(B_2)P(A | B_2)} \\ &= \frac{(0.00001139)(0.9952)}{(0.00001139)(0.9952) + (0.9999886)(0.0115)} = 0.00098 \end{aligned}$$

La tasa de falsos positivos es muy baja para este caso, ya que la prueba parece ser muy precisa, independientemente de que el VIH sea considerada todavía como una enfermedad rara.

2. COVID-19

En esta sección se hace una síntesis de diversas fuentes en donde el tema principal es el comportamiento de las pruebas clínicas y/o rápidas para detectar el virus SARS-CoV-2 y algunas fuentes enfatizan la importancia de analizar los falsos negativos, ya que es un factor importante en la propagación del virus. Por ejemplo, Schipper (2020) presenta un análisis mediante dos ejemplos: *pruebas rápidas* con pacientes asintomáticos y con pacientes hospitalizados con síntomas de COVID-19, en su análisis se resalta que muchas pruebas informan un 80 % de sensibilidad y una **especificidad** del 100 %, a pesar de que existen informes de posibles *pruebas positivas falsas*. Los resultados encontrados se muestran enseguida [7]:

Valor predictivo positivo = tener Covid-19 si la prueba es positiva = $8/8 = 100\%$

Valor predictivo negativo = no tener Covid-19 si la prueba es negativa = $990/992 = 99.8\%$

Valor predictivo positivo = tener Covid-19 si la prueba es positiva = $240/240 = 100\%$

Valor predictivo negativo = no tener Covid-19 si la prueba es negativa = $700/760 = 92\%$

Podemos observar que el 99.8 % y 92 % nos indica que se están produciendo *varios* falsos negativos, en la mayoría de personas en donde se supone tienen síntomas pero su prueba salió negativa, lo que nos indicaría que la prueba rápida no es altamente confiable.

Por otra parte, Larremore et al. (2020) menciona que debido a que el SARS-CoV-2 se puede propagar desde individuos con infecciones presintomáticas, sintomáticas y asintomáticas, la reapertura de sociedades y el control de la propagación del virus se verán vigilados, para lo cual el realizar pruebas para detectar el virus serán fundamentales. En esta investigación se considera el patrón de cinética de la carga viral, se modela la efectividad de la vigilancia considerando las *sensibilidades* de las pruebas, la *frecuencia* y el *tiempo* de notifi-

cación de la muestra a la respuesta.

Tres características como lo es el aumento, la infectividad y la disminución del virus durante la infección por SARS-CoV-2 los llevó a plantear una hipótesis: puede haber diferencias mínimas en la vigilancia eficaz utilizando pruebas de detección viral de diferentes sensibilidades (RT-qPCR o ensayos más rápidos con límites de detección más altos (como las *pruebas rápidas*)).

Se realizó una simulación con una muestra inicial de 100 individuos [4], los resultados arrojaron que las pruebas de vigilancia para individuos asintomáticos pueden usarse para limitar la propagación del virus, sin embargo, existen muchos factores que aún afectan el comportamiento de su modelo, como lo es la sensibilidad de las pruebas, ya que puede haber diferentes fabricantes o el que haya un muestreo clínico inadecuado, por otra parte, las diferencias exactas de rendimiento entre los esquemas de prueba dependerán de si el modelo realmente captura la cinética viral y los perfiles de infecciosidad durante la fase de aceleración entre la exposición y la carga viral máxima. Se enfatiza que las pruebas de vigilancia son totalmente distintas a las pruebas clínicas, ya que estas últimas se dirigen a individuos sintomáticos, necesitan alta precisión y sensibilidad y tienen un costo alto. Debido a que se enfocan en individuos sintomáticos, esos individuos pueden aislarse de tal manera que un retraso en el diagnóstico no dé lugar a infecciones adicionales. Por el contrario, los resultados de las pruebas de vigilancia de personas asintomáticas deben devolverse rápidamente, ya que incluso un retraso en el diagnóstico de un solo día compromete la eficacia del programa de vigilancia.

Se recomienda que los gobiernos estatales fomenten el desarrollo y uso de pruebas alternativas más rápidas y de menor costo con fines de vigilancia, incluso si tienen límites de detección más bajos. El uso poco frecuente de una prueba sensible no solo identificará a aquellos con una carga viral baja al comienzo de la infección, que deben ser aislados para limitar la propagación viral, sino aquellos en el período de recuperación, que todavía tienen virus o ARN detectables pero están por debajo del umbral infeccioso. Estos resultados demuestran que la vigilancia eficaz depende en gran medida de la *frecuencia de las pruebas* y la *velocidad de la notificación*, y solo mejora marginalmente con una alta *sensibilidad de las pruebas*.

Algunas pruebas tienen una exactitud del 98% para detectar infecciones por COVID. Pero, ¿qué significa realmente esta exactitud?, esta respuesta la trata de explicar Ranjan (2000), para lo cual necesita definir lo siguiente [6]:

Verdadero positivo: una persona con COVID-19 da positivo por COVID-19.

Falso positivo: una persona sin COVID-19 da positivo por COVID-19.

Falso negativo: una persona con COVID-19 da negativo para COVID-19.

Verdadero negativo: una persona sin COVID-19 da negativo en las pruebas de COVID-19.

Exactitud = (verdaderos positivos + verdaderos negativos) / todos los resultados.

Precisión = verdaderos positivos / (verdaderos positivos + falsos positivos).

Sensibilidad = verdaderos positivos / (verdaderos positivos + falsos negativos).

Especificidad = verdaderos negativos / (verdaderos negativos + falsos positivos).

En este caso se puede definir a la *sensibilidad* como la probabilidad de que, dada una persona que tiene covid, las pruebas sean positivas, y está vendrá dada por la probabilidad condicional, $P(+ | C^+)$. De manera similar, la *especificidad* se define como la probabilidad de que un paciente sano dé negativo en la prueba, y está dada por $P(- | C^-)$. Suponga ahora que se elige a una persona al azar de una población y se le aplica una prueba para detectar el virus. Suponga ahora que dio positivo, pero ¿cuál es la probabilidad de que realmente tenga el coronavirus? En otras palabras, aquí se está interesado en encontrar $P(C^+ | +)$. De acuerdo al teorema de Bayes, si conocemos $P(C^+)$ (simplemente dividir el número de casos positivos del país por la población total), entonces se puede calcular fácilmente $P(C^+ | +)$.

Lo interesante del aporte que hace Ranjan es que al calcular la probabilidad condicional de que si una persona da positivo en una prueba dado que en verdad tiene el virus en su país (China) fue muy baja (0.5%). Esto significa que incluso si una persona al azar da positivo en la prueba, hay menos del 0.5% de probabilidades de que *realmente* tenga el virus. Se enfatiza que esto es un *problema con todas las enfermedades raras*.

Un análisis similar a lo que hizo Ranjan lo hace Lewis (2020), pero ahora considerando a Estados Unidos de América, ya que ese país había alcanzado la tasa más grande de casos confirmados de COVID-19 (lunes 6 de abril de 2020 había 336,830 casos confirmados [5]). Debido a la limitada capacidad de prueba de Estados Unidos, el autor sospecha que hay casos que han escapado a la atención de las autoridades. Lewis define a la **sensibilidad** de cualquier prueba biomédica como la probabilidad de que una persona dé positivo dado que tiene la enfermedad y a la **especificidad** de una prueba como la probabilidad de que una persona dé negativo dado que no tiene la enfermedad. Se indica que aquellos que no tienen la enfermedad dan positivo en la prueba; estos constituyen los *falsos positivos*.

Entonces, con base a números puramente hipotéticos, y con una prueba puramente hipotética con una sensibilidad del 99%, se determinó con la ayuda del teorema de Bayes que la probabilidad de que alguien tenga COVID-19 dado que dan positivo en la prueba sería de aproximadamente el 9%. Es decir, solo alrededor de nueve de cada 100 personas que dan positivo en la prueba serían en realidad casos de COVID-19. *Esto implica muchos falsos positivos*.

Finalmente Lewis recomienda algo similar de lo que menciona Larremore, ya que se debe asegurar de que aquellos que creemos que son negativos realmente lo sean, y que puede ser mucho más importante que asegurarse de que aquellos que creemos que son positivos sean realmente positivos. Pero que no se debe olvidar que decirle a la gente que es positivo

para una enfermedad cuando no lo es tiene un costo. No es solo la posibilidad de ser puesto en cuarentena innecesariamente. También es la ansiedad de pensar que uno podría tener una enfermedad que parece matar alrededor del 1–2 % de quienes la contraen .

Chan (2020) comenta que las buenas pruebas de detección con alta **sensibilidad** permiten un alto grado de *verdaderos positivos* mientras permiten un número permisivo de *falsos positivos* para descartar la enfermedad y que las pruebas con alta **especificidad** (baja tasa de falsos positivos) son una simplificación excesiva de una dinámica compleja. Aunque la sensibilidad y la especificidad son atributos de la prueba que son importantes comprender, el teorema de Bayes permite que el proveedor de atención médica convierta los resultados de una prueba en la probabilidad de tener una enfermedad.

En este caso, para todas las **pruebas de detección**, ya sea para COVID–19 u otros diagnósticos, la comprensión de los valores predictivos y las razones de probabilidad con la ayuda del teorema de Bayes garantizará una interpretación sólida y las recomendaciones y acciones resultantes por parte de los médicos y las partes interesadas. Un resultado de prueba negativo, en este paradigma, nunca es absolutamente negativo. Más bien, ajusta la probabilidad previa a la prueba de tener una enfermedad más baja [1].

A medida que los médicos atienden a pacientes con antecedentes de contacto y síntomas que podrían representar la enfermedad por COVID–19, es crucial interpretar los resultados de los ensayos de reacción en cadena de la polimerasa a partir de hisopos nasales y faríngeos. Si bien un resultado positivo en un paciente gravemente enfermo es sencillo, ¿cómo deberían los médicos interpretar las pruebas negativas en pacientes con sospecha de infección por COVID–19?, esta cuestión la analiza Good (2020), para esto el autor señala que para interpretar con precisión los resultados de la prueba, es necesario conocer los valores **predictivos positivos y negativos** de una prueba en el entorno aplicado, que dependen de su sensibilidad y especificidad, junto con la prevalencia o la probabilidad previa a la prueba. Aunque la especificidad de los ensayos de PCR para COVID-19 parece estar cerca del 100 %, documentar su sensibilidad es sorprendentemente difícil de alcanzar [2].

Good aplicó un análisis bayesiano para interpretar los resultados negativos y positivos del ensayo de PCR COVID–19 para dos escenarios clínicos. Para ambos escenarios, se asumió una *especificidad* del ensayo de PCR del 99.9 % y se consideró valores del 70 al 90 % para la *sensibilidad*. El primer escenario considera una **alta** probabilidad previa a la prueba de infección por COVID–19, el escenario 2 considera una probabilidad de infección por COVID–19 **baja** antes de la prueba.

Dicho análisis permitió comprender que una prueba positiva en escenarios tanto de preprueba alta como de preprueba baja probablemente represente una *infección aguda*. Asimismo, una prueba negativa en un caso de baja probabilidad previa a la prueba indica una *baja* probabilidad de *infección aguda*. Sin embargo, cuando se tiene una probable infección por el virus, una prueba negativa no debería descartar una infección aguda, por lo que se recomienda la posibilidad de repetir otra prueba o realizar una evaluación adicional

3. Teorema de Bayes y COVID-19 en México

En México no se han visto de buena forma las pruebas rápidas para diagnóstico del COVID-19, ya que algunas han mostrado inexactitud en sus resultados, motivo por el cual algunos mexicanos han optado por no realizarse la prueba a menos que su cuadro de salud empeore demasiado. Cofepris aprobó 19 pruebas rápidas, cuyo nombre correcto es *serológicas*, asimismo, a través de un comunicado indicó cuáles son los posibles resultados de dichas pruebas:

- **IgM-/IgG-** No hay evidencia de infección por SARS-Cov 2.
- **IgM+/IgG-** Probable infección reciente sin anticuerpos protectores.
- **IgM+/IgG+** Probable Infección reciente con anticuerpos protectores en desarrollo.
- **IgM-/IgG+** Probable infección pasada con anticuerpos protectores.

Cofepris afirma también que la presencia de anticuerpos tipo IgG sugiere que el sujeto ha sido expuesto a el virus y ha desarrollado una respuesta inmune, típicamente esto ocurre al menos dos semanas después de la exposición y expresión clínica de la enfermedad. No determina en forma categórica que ya no se tiene riesgo de contraer la enfermedad, pero sugiere que es de menor riesgo que quien no tiene anticuerpos. La presencia de anticuerpos IgM indica que el sujeto ha sido expuesto al virus y sugiere que el contacto ha ocurrido en las dos semanas anteriores a la muestra. La presencia de anticuerpos IgG e IgM en forma simultánea, indica que la enfermedad está pasando su forma aguda.

La Asociación Nacional de Farmacias de México tiene a la venta un dispositivo de prueba para COVID-19 altamente portátil y estable a temperatura ambiente, ideal para pequeños laboratorios, empresas, clínicas y pacientes posiblemente contagiados, dicha prueba ofrece resultados de diagnóstico positivo/negativo en 15 minutos, mediante un método de plasma y suero. La **sensibilidad** de esta prueba es del 93.5 % y su **especificidad** es del 100 % (<https://www.anafarmex.com.mx/dispositivo-de-prueba-para-covid-19/>).

Con base en esta información y las lecturas realizadas en la sección anterior, no me fue posible establecer una probabilidad que se pudiera obtener con el teorema de Bayes de forma directa, ya que los **resultados** que arrojan las pruebas no parecen ser del todo excluyentes (característica primordial en el teorema), por lo que supuse, en el peor de los casos, que los resultados de la prueba se pueden manipular de tal forma que es posible considerar si se está contagiado (3 casos) o no (un caso).

Es por esto que se consideró la base de datos del número de Casos COVID-19 en Tlaxcala por Municipio, el cual indica que aproximadamente el 5.97×10^{-6} de las personas que viven en el estado fueron contagiados con el virus (http://evaluacion.septlaxcala.gob.mx/maps_esc/). Con la información de la prueba rápida descrita anteriormente, y suponiendo que algún conocido se realizó dicha prueba y salió positiva, me gustaría determinar la probabilidad de que realmente esa persona está contagiada.

Sea B_1 = persona con COVID-19, B_2 = persona que no tiene COVID-19 y sea A = resultado positivo de la prueba del COVID-19. Luego entonces, $P(B_1) = 0.00000597$, $P(B_2) = 0.99999403$, $P(A | B_1) = 0.935$ y $P(A | B_2) = 0$, usando el teorema de Bayes obtenemos:

$$\begin{aligned} P(B_1 | A) &= \frac{P(B_1)P(A | B_1)}{P(B_1)P(A | B_1) + P(B_2)P(A | B_2)} \\ &= \frac{(0.00000597)(0.935)}{(0.00000597)(0.935) + (0.99999403)(0)} = 1 \end{aligned}$$

El resultado encontrado tal vez sea una probabilidad errónea mal calculada, esto al no saber interpretar el manejo de los resultados de las pruebas con los índices de sensibilidad y especificidad que indica la prueba rápida presentada en este análisis, de igual forma me llama mucho la atención el porcentaje de personas que han confirmado positivos de COVID-19 en el estado de Tlaxcala, ya que al comparar con los análisis realizados en las lecturas, el porcentaje presentado es exageradamente pequeño.

Referencias

- [1] G. M. Chan. Bayes' Theorem, COVID19, and screening tests. *The American Journal of Emergency Medicine*, 2020.
- [2] H. I. . S. K. Good, C. B. Interpreting COVID-19 Test Results: a Bayesian Approach. *Journal of general internal medicine*, 2020.
- [3] A. Gutiérrez. *Probabilidad y Estadística, Enfoque por competencias*. McGraw Hill, México, 2012.
- [4] D. B. Larremore, B. Wilder, E. Lester, S. Shehata, J. M. Burke, J. A. Hay, M. Tambe, M. J. Mina, and R. Parker. Test sensitivity is secondary to frequency and turnaround time for covid-19 surveillance. *medRxiv*, 2020.
- [5] M. A. Lewis. Covid-19, Bayes' Theorem and covid-19 testing. <https://www.significancemagazine.com/science/660-bayes-theorem-and-covid-19-testing>, 2020.
- [6] A. Ranjan. Covid-19, Bayes' Theorem and taking probabilistic decisions. <https://towardsdatascience.com/covid-19-bayes-theorem-and-taking-data-driven-decisions-part-1-b61e2c2b3bea>, 2020.
- [7] J. L. Schnipper and P. E. Sax. Covid-19 test accuracy supplement: The math of Bayes' Theorem. <https://www.statnews.com/2020/08/20/covid-19-test-accuracy-supplement-the-math-of-bayes-theorem/>, 2020.