MAT^2
MATerials MATemàtics
Volum 2021, treball no. 6, 5 pp. ISSN: 1887-1097
Publicació electrònica de divulgació del Departament de Matemàtiques
de la Universitat Autònoma de Barcelona
www.mat.uab.cat/web/matmat

Las pruebas PCR con elevada sensibilidad y especificidad, en condiciones de alta prevalencia o bajo prescripción médica, son fiables*

J. M. Almira

Un ejemplo bien conocido y citado con frecuencia, que los matemáticos utilizan para motivar la importancia del Teorema de Bayes, está relacionado con la interpretación de la sensibilidad y especificidad de las pruebas médicas [6, 10]. La sensibilidad es la probabilidad de un resultado positivo dado que se está enfermo y la especificidad es la probabilidad de un resultado negativo dado que no se está enfermo. Por lo general, una prueba (o test médico) se considera buena si tiene una alta sensibilidad y una alta especificidad. El argumento sugiere que, a pesar de gozar de tan buenas especificaciones, un buen test médico no garantiza que, cuando el resultado de la prueba es positivo, la probabilidad de estar realmente enfermo sea elevada.

En particular, podemos aplicar el argumento a las pruebas PCR que se utilizan para detectar la Covid-19. Suponga, por ejemplo, que dicha prueba tiene una sensibilidad igual a 97.2 % y una especificidad igual a 98.6 %. Hasta donde sabemos, estos son los valores publicados más optimistas para estos parámetros. Dichos valores significan que la probabilidad de un test positivo cuando se está infectado es de 0.972 y la probabilidad de un test negativo cuando no se está infectado es de 0.986. Por lo tanto, las probabilidades de un verdadero positivo y un verdadero negativo son altas. ¿Significa esto que, si se obtiene un resultado positivo, la probabilidad de estar infectado es alta? ¿Y qué hay de la probabilidad de estar sano cuando el resultado de la prueba es negativo?

Obviamente, el Teorema de Bayes ayuda a responder estas preguntas. Para nuestro argumento, asumimos que tanto la sensibilidad como la espe-

*Versión en castellano del artículo "PCR tests with high sensitivity and specificity are truly trustworthy under high prevalence or medical prescription".

cificidad son menores que 100 %, que es el caso de la mayoría de las pruebas médicas y, en particular, es una hipótesis verdadera para las pruebas PCR. Bajo este supuesto, si la prevalencia de la enfermedad es pequeña, la probabilidad de un verdadero positivo puede ser inesperadamente pequeña. La prevalencia es la probabilidad de estar infectado. Suele ser un valor pequeño. Por ejemplo, podemos suponer que una prevalencia de 2 % es alta para la Covid-19. En tal caso, tendremos que

$$\begin{aligned} P(\text{infec}|\text{pos}) &= \frac{P(\text{pos}|\text{infec})P(\text{infec})}{P(\text{pos}|\text{infec})P(\text{infec}) + P(\text{pos}|\text{no infec})P(\text{no infec})} \\ &= \frac{0.972 \cdot 0.02}{0.972 \cdot 0.02 + (1-0.986) \cdot 0.98} = \frac{0.01944}{0.03316} \\ &= 0.58624849. \end{aligned}$$

Obviamente, una probabilidad del 58.62 % de estar infectado después de conocer que el resultado del test es positivo parece sorprendentemente pequeña. ¿Significa esto, entonces, que las pruebas PCR no son confiables? No. Con los mismos datos, la probabilidad de un verdadero negativo es realmente alta:

$$\begin{aligned} P(\text{no infec}|\text{neg}) &= \frac{P(\text{neg}|\text{no infec})P(\text{no infec})}{P(\text{neg}|\text{no infec})P(\text{no infec}) + P(\text{neg}|\text{infec})P(\text{infec})} \\ &= \frac{0.986 \cdot 0.98}{0.986 \cdot 0.98 + (1-0.972) \cdot 0.02} = \frac{0.96628}{0.96684} \\ &= 0.99942079. \end{aligned}$$

Esto significa que, con estos datos, un test negativo es verdaderamente digno de confianza.

Además, el argumento no tiene en cuenta que la prevalencia podría ser grande. De hecho, la prevalencia de la Covid-19 aún se desconoce y podría ser superior al 2 %. Supongamos, por un momento, una prevalencia del 30 %. Entonces, el mismo cálculo de la probabilidad de un verdadero positivo da un resultado bastante diferente:

$$\begin{aligned} P(\text{infectado}|\text{positivo}) &= \frac{0.972 \cdot 0.3}{0.972 \cdot 0.3 + (1-0.986) \cdot 0.7} = \frac{0.2916}{0.3014} \\ &= 0.96748507. \end{aligned}$$

En tal caso, la probabilidad es casi igual a la sensibilidad de la prueba (sería igual si la especificidad fuera del 100 %) y el resultado de la prueba parece realmente fiable.

Por supuesto, se puede suponer que una prevalencia del 30 % es, para la mayoría de las enfermedades, un valor muy elevado, que no es realista asumir. Pero los cálculos anteriores también ocultan un factor importante:

las personas que se someten a un examen médico no se seleccionan de la población al azar. Suelen tener prescripción médica en función de sus síntomas o —para una enfermedad infecciosa como la Covid-19— teniendo en cuenta el hecho de que son contactos estrechos de personas infectadas. En ambos casos, la prevalencia del sujeto elegido aumenta drásticamente, hasta el punto de que asumir una prevalencia del 20 % o el 30 % para ellos es una suposición muy realista. Así, la conclusión es clara: las pruebas PCR con alta sensibilidad y especificidad son verdaderamente fiables cuando se administran bajo prescripción médica. Además, un test masivo tendrá muchos falsos positivos, mientras que el uso de un test masivo en zonas de alta prevalencia facilitará la detección y el aislamiento de personas infectadas asintomáticas.

El argumento, sin comentarios sobre la prevalencia, aparece con frecuencia en muchos libros de introducción a la teoría de la probabilidad [2, 3] y, con mayor frecuencia, en los libros [5, 6, 10] y artículos [4, 7, 9] dedicados a la divulgación de las matemáticas, pero también en algunos trabajos de investigación [8]. Hice el experimento de escribir “Usos del teorema de Bayes” en el motor de búsqueda de Google y descubrí que este ejemplo se discute, nuevamente sin especial atención al papel de la prevalencia en los cálculos, en varias de las primeras diez páginas web que el sistema muestra por orden de relevancia. Uno de ellos es una carta publicada en Scientific American en 2016 [4], y el mismo argumento también aparece como parte de otros dos artículos en la misma revista en 2012 [7] y 2006 [9].

Como director científico de una colección de libros [11] dedicada a explicar la importancia de las matemáticas en la nueva era del big data y la inteligencia artificial, lo que implica una explicación de su papel en el desarrollo de muchas aplicaciones en ciencia y tecnología —incluyendo temas tan diversos como redes sociales, filtros colaborativos, decisiones políticas, ciencia de redes, big data, sistemas electorales, chatbots, comunicaciones digitales, robótica, modelos matemáticos de epidemias, medicina personalizada, privacidad digital, etc.— he tenido que revisar muchos libros donde se usa el Teorema de Bayes y, en varios de ellos, encontré que los autores presentaron el argumento anterior sin referencias a la prevalencia. En todos los casos sugerí a los autores que era necesaria una mención a la prevalencia, porque de lo contrario el argumento podría resultar confuso. Cuando hice esto por n-ésima vez, le mencioné el caso a un amigo, que es médico en activo, en busca de su empatía. Para mi sorpresa, mi amigo, que es una persona inteligente y un médico experimentado que trabaja en un hospital de referencia, necesitó revisar todos los cálculos y, después de eso, todavía dijo que faltaba algo, ya que su experiencia le decía que un resultado positivo en una prueba médica (y, en particular, esto había sido así con las pruebas PCR) generalmente significa que la persona ya está infectada, y también creía que las prevalencias nunca son tan elevadas [1]. Fue solo después de esta conversación que, con la ayuda de un buen café compartido, me di cuenta de que un factor oculto en la discusión, y que nunca había leído en ninguna parte, es el hecho de que, cuando te sometes a una prueba PCR es porque tienes algunos síntomas o eres un contacto estrecho de alguna persona infectada. Esta información, por supuesto, tiene el efecto de incrementar la prevalencia. De nuevo, este es el resultado de aplicar una concepción Bayesiana de la probabilidad: la información que tenemos sobre un experimento suele ser útil para cambiar nuestros antecedentes. Considero que no incluir el papel de la prevalencia en la argumentación, y no señalar el hecho de que las pruebas médicas no se suelen realizar a personas elegidas al azar sino a personas que presentan algunos síntomas, puede generar confusión. Una persona que haya leído el artículo de Paulos [7] bien podría decidir no hacerse una prueba PCR prescrita por su médico ¡argumentando que el resultado de dicha prueba es como lanzar una moneda al aire!

Referencias

[1] Joseph Lawrence, Theresa W. Gyorkos, and Louis Coupal, Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard, American Journal of Epidemiology , 141 (3) 263-272, 1995.

[2] Paolo L. Gatti, *Probability Theory and Mathematical Statistics for Engineers,* CRC Press, 2004.

[3] Andrew Gelman, Deborah Nolan, *Teaching Statistics. A Bag of Tricks,* Oxford University Press, 2002.

[4] John Horgan, Bayes's Theorem: What's the Big Deal?, Scientific American Blog Network, Jan, 4, 2016.
https://blogs.scientificamerican.com/cross-check/bayes-s-theorem-what-s-the-big-deal/

[5] Bartolo Luque, Juan Mr Parrondo, *Las leyes del azar,* Shackleton books, 2021

[6] John Allen Paulos, *El hombre anumérico El analfabetismo matemático y sus consecuencias,* Metatemas, Tusquets, 2016.

[7] John Allen Paulos, The Math behind Screening Tests. What a positive result really means, Scientific American, Jan.,1, 2012.
https://www.scientificamerican.com/article/weighing-the-positives/

[8] Marco Tommasi, Grazia Ferrara and Aristide Saggino, Application of Bayes' Theorem in Valuating Depression Tests Performance. Front. Psychol. 9:1240, 2018. doi:10.3389/fpsyg.2018.01240

[9] Chris Wiggins, What is Bayes's theorem, and how can it be used to assign probabilities to questions such as the existence of God? What scientific value does it have?, Scientific American, Dec., 4, 2006. https://www.scientificamerican.com/article/what-is-bayess-theorem-an/

[10] Kit Yates, *Los números de la vida,* Blackie Books, 2020.

[11] La matematica che trasforma il mondo, 2020. http://www.matematicarba.it/?utm_source=rba&utm_medium=post2020

Dpt. de Ingeniería y Tecnología de Computadores
Área de Matemática Aplicada
Facultad de Informática
Universidad de Murcia
jmalmira@um.es

Publicat el 2 de desembre de 2021