Ejercicios de Teorema de Bayes

Análisis Estadístico Multivariado

Emmanuel Alcalá

En uno de sus típicos alardes de superioridad intelectual, Elon Musk escribió en un tweet sobre pruebas de antígenos a COVID:

«Technically, I tested positive, then negative twice, then positive again… The “rapid antigen test’ ’ from BD seems to be about as useful as a flipping a coin.»

Básicamente, afirma que si 2/4 pruebas dan positivo, el test que usó es igual de informativo que lanzar una moneda. ¿Es verdad?

Como no somos Musk y no creemos que la verdad brota de nuestras bocotas, veamos qué nos dice la teoría de la probabilidad. Esta es la información con la que disponemos:

1. La especificidad del test BD usado es del 0.995.
2. La sensibilidad del test del 0.84.
3. Al momento de su aseveración, la prevalencia (la proporción de personas infectadas sin importar su resultado) era del 0.0121 (aprox 1.21 %).

Sea X la variable que representa si se tiene COVID, con valores de $\{0, 1\}$ si no tiene y si tiene, respectivamente; y sea T_i el resultado de la $i-$ésima prueba prueba, con $\{-, +\}$ si es negativo y si es positivo, respectivamente.

La prevalencia nos dice la probabilidad *a priori* $p(X=1)=0.0121 $. La especificidad nos dice la probabilidad de que la prueba dé negativa *dado* que la enfermedad esté ausente. Para $i = 1$, $p(T_i = - \mid X = 0) = 0.995$. La sensibilidad nos dice la probabilidad de que la prueba dé positiva dado que la enfermedad esté presente, $p(T_i = + \mid X = 1) = 0.84$.

Ahora bien, ¿cuál es la probabilidad de que Elon tenga COVID *dado* que tuvo dos negativas y dos positivas?

Es decir, necesitamos

1. $p(X=1 \mid T_1=+)$
2. $p(X=1 \mid T_1=+, T_2=-)$
3. $p(X=1 \mid T_1=+, T_2=-, T_3=-)$
4. $p(X=1 \mid T_1=+, T_2=-, T_3=-, T_4=+)$

Para el primer caso, la probabilidad de que tenga COVID dado que la prueba 1 salió positiva, usaremos el teorema de Bayes como sigue:

$$p(X=1 \mid T_1=+) = \frac{p(T_1=+ \mid X=1)p(X=1)}{p(T_1=+)}$$

En $p(T_1=+)$, la probabilidad de tener un resultado positivo sin importar si se tiene COVID o no, usaremos la ley de probabilidad total:

$$p(T_i=+) = \sum_{\mathcal{X}} P(T_i \mid X)p(X) = p(T_1=+ \mid X=1)p(X=1) + p(T_1=+ \mid X=0)p(X=0)$$

$p(T_1=+ \mid X=1)$ es la sensibilidad, y $p(T_1=+ \mid X=0)$ es la probabilidad complementaria de $p(T_1=- \mid X=0)$ (esto es por lo siguiente: *dado* que $X=0$, solo pueden ocurrir dos cosas: $T=+$, ó $T=-$); $p(X=0)$ y $p(X=1)$ son complementario. Recordar que si A^c es el complemento de A, $p(A) = 1 - p(A^c)$.

En palabras

$$p(X=1 \mid T_1=+) = \frac{sensibilidad \times prevalencia}{sensibilidad \times prevalencia + (1 - especificidad) \times (1 - prevalencia)}$$

Por lo tanto, $p(T_i) = 0.015$. La probabilidad de tener COVID dado que la prueba fue positiva, $p(X=1 \mid T=+)$ es de 0.67, aproximadamente.

¿Cómo calculamos $p(X=1 \mid T_1=+, T_2=-)$? Usando la misma ecuación, teniendo en cuenta que la sensibilidad y la especificidad no cambian, pero ahora nuestra probabilidad *a priori* no será $p(X=1)$ sino $p(X=1 \mid T=+)$.

Para simplificar un poco las cosas, supongamos que A es $X=1$, y B es \$T_1=+ T_2=-) \$.El teorema de Bayes nos dice que

$$p(A \mid B) = \frac{p(B)}{}$$

$$p(X=1 \mid T_1=+, T_2=-) = \frac{p(T_2=- \mid T_1=+, X=1)p(X=1 \mid T_1=+)}{p(T_2=- \mid T_1=+)}$$

Sabemos que $p(X=1 \mid T_1=+) = 0.67$, ¿y $p(T_2=- \mid T_1=+, X=1)$? Dado que la prueba T_1 es independiente de T_2, la probabilidad de que salga negativo *dado* que tiene covid es simplemente $p(T_2=- \mid X=1) = 1-p(T_2=+ \mid X=1) = 1-sensibilidad = 1-0.84 = 0.16$. Por lo tanto, en el numerador tenemos $0.67 \times 0.16 = 0.1072$. ¿Cómo calculamos $p(T_2=- \mid T_1=+)$?

$$p(T_2=- \mid T_1=+) = \frac{p(T_1=+, T_2=-)}{p(T_1)}$$

Sabemos que $p(T_1) = 0.015$, $p(T_1=+, T_2=-)$ se puede obtener con la ley de probabilidad total

$$p(T_1=+, T_2=-) = p(T_1=+, T_2=- \mid X=1)p(X=1) + p(T_1=+, T_2=- \mid X=0)p(X=0)$$

Dado que tenemos eventos independientes, $p(T_1=+, T_2=- \mid X=1)p(X=1) = p(T_1=+)p(T_2=-)p(X=1)$; con $p(T_1=+, T_2=- \mid X=0)p(X=0)$ usaremos los eventos complementarios.

$$p(T_1=+, T_2=-) = 0.995 * 0.995 * 0.0121 + (1-0.995) * (1-0.995) * (1-0.0121) = 0.012$$

Por lo tanto

$$p(T_2=- \mid T_1=+) = \frac{p(T_1=+, T_2=-)}{p(T_1)} = \frac{0.012}{0.015} = 0.8$$

y

$$p(X=1 \mid T_1=+, T_2=-) =$$