Un caso de probabilidad engañosa

Hoy vamos a ver otro de esos ejemplos en los que la intuición sobre el valor de determinadas probabilidades nos juega malas pasadas. Y, para ello, vamos a utilizar nada menos que el teorema de Bayes, jugando un poco con las probabilidades condicionadas. Vamos a ver paso a paso cómo funciona.

¿Cuál es la probabilidad de que se produzcan dos sucesos? La probabilidad de que ocurra un suceso A es P(A) y la de que ocurra B, P(B). Pues bien, la probabilidad de que ocurran los dos es P(A∩B) que, si los dos sucesos son independientes, es igual a P(A) x P(B).

Imaginemos que tenemos un dado con seis caras. Si lo lanzamos una vez, la probabilidad de sacar, por ejemplo, un cinco es de 1/6 (un resultado entre los seis posibles). La de sacar un cuatro es, igualmente, 1/6. ¿Cuál será la probabilidad de sacar un cuatro, una vez que en la primera tirada sacamos un cinco?. Como las dos tiradas son independientes, la probabilidad de la combinación cinco seguida de cuatro será de 1/6 x 1/6 = 1/36.

Ahora pensemos otro ejemplo. Supongamos que en un grupo de 10 personas hay cuatro médicos, dos de los cuáles son cirujanos. Si tomamos uno al azar, la probabilidad de que sea médico es de 4/10 = 0,4 y la de que sea cirujano es de 2/10 = 0,2. Pero, si sacamos a uno y sabemos que es médico, la probabilidad de que sea cirujano ya no será de 0,2, porque los dos sucesos, ser médico y cirujano, no son independientes. Si es médico, la probabilidad de que sea cirujano será de 0,5 (la mitad de los médicos de nuestro grupo son cirujanos).

Cuando dos sucesos son dependientes, la probabilidad de que ocurran los dos será la probabilidad de ocurrir el primero, una vez que ocurre el segundo, por la probabilidad de ocurrir el segundo. Así que la P(médico∩cirujano) = P(cirujano|médico) x P(médico). Podemos generalizar la expresión de la siguiente manera:

P(A∩B) = P(A|B) x P(B), y cambiando de orden los componentes de la expresión, obtenemos la llamada regla de Bayes, de la siguiente forma:

P(A|B) = P(A∩B) / P(B).

La P(A∩B) será la probabilidad de B, una vez que se produce A, por la probabilidad de A = P(B|A) x P(A). Por otra parte, la probabilidad de B será igual a la suma de la probabilidad de producirse B una vez que se produzca A más la probabilidad de producirse B sin que ocurra A, lo que puesto de forma matemática queda de la siguiente forma:

P(B|A) x P(A) + P(B|A^c) x P(A^c), siendo P(A^c) la probabilidad de que no ocurra A.

Si sustituimos la regla inicial por sus valores desarrollados, obtendremos la expresión más conocida del teorema de Bayes:

$$P(A|B)=\frac{P(B|A)\times P(A)}{P(B|A)\times P(A)+P(B|A^c)\times P(A^c)}$$

Vamos a ver cómo se aplica el teorema de Bayes con un ejemplo práctico. Pensemos en el caso de la fildulastrosis aguda, una grave enfermedad cuya prevalencia en la población es, afortunadamente, bastante baja, de uno por cada 1000 habitantes. Luego, la P(F) = 0,001.

Por suerte tenemos una buena prueba diagnóstica, con una sensibilidad del 98% y una especificidad del 95%. Supongamos ahora que yo me hago la prueba y me da un resultado positivo. ¿Tengo que asustarme mucho? ¿Cuál es la probabilidad de que realmente tenga la enfermedad? ¿Os parece que será alta o baja? Veámoslo.

Una sensibilidad del 98% quiere decir que la probabilidad de dar positivo cuando se tiene la enfermedad es de 0,98. Matemáticamente, P(POS|F) = 0,98. Por otra parte, una especificidad del 95% quiere decir que la probabilidad de que dé un resultado negativo estando sano es de 0,95. O sea, P(NEG|F^c) = 0,95. Pero nosotros lo que queremos saber no es ninguna de estas dos cosas, sino que realmente buscamos cuál es la probabilidad de estar enfermo una vez que damos positivo en la prueba, o sea, la P(F|POS).

Para calcularla, no tenemos más que aplicar el teorema de Bayes:

$$P(F|POS)=\frac{P(POS|F)\times P(F)}{P(POS|F)\times P(F)+P(POS|F^c)\times P(F^c)}$$

A continuación, sustituimos los símbolos con sus valores y resolvemos la ecuación:

$$P(F|POS)=\frac{0{,}98\times 0{,}001}{0{,}98\times 0{,}001+[(1-0{,}95)\times(1-0{,}001)]}=0{,}02$$

Así que vemos que, en principio, no tengo que asustarme mucho cuando la prueba me da un resultado positivo, ya que la probabilidad de estar enfermo es solo de un 2%. Como veis, mucho más baja de lo que la intuición nos diría con una sensibilidad y una especificidad tan altas. ¿Por qué ocurre esto? Muy sencillo, porque la prevalencia de la enfermedad es muy baja. Vamos a repetir el experimento suponiendo ahora que la prevalencia es del 10%

(0,1):

$$P(F|POS) = \frac{0{,}98 \times 0{,}1}{0{,}98 \times 0{,}1 + [(1-0{,}95) \times (1-0{,}1)]} = 0{,}68$$

Como veis, en este caso la probabilidad de estar enfermo si doy positivo sube hasta el 68%. Esta probabilidad es el conocido valor predictivo positivo que, como podemos comprobar, puede variar enormemente según la frecuencia del efecto que estemos estudiando.

Y aquí lo dejamos por hoy. Antes de terminar, dejadme advertiros que no busquéis qué es la fildulastrosis. Me sorprendería mucho que alguien la encontrase en algún libro de medicina. Además, tened cuidado de no confundir P(POS|F) con P(F|POS), ya que incurriríais en un pecado llamado falacia inversa o falacia de la transposición de los condicionales, que es un error grave.

Hemos visto como el cálculo de probabilidades se complica un poco cuando los sucesos no son independientes. También hemos aprendido lo poco de fiar que son los valores predictivos cuando cambia la prevalencia de la enfermedad. Por eso se inventaron los cocientes de probabilidades, que no dependen tanto de la prevalencia de la enfermedad que se diagnostica y permiten valorar mejor de forma global la potencia de la prueba diagnóstica. Pero esa es otra historia…