

La falacia del fiscal y el caso Sonia Carabantes

Es actualidad estos días (ver, por ejemplo, El País 02-09-03) el resultado del análisis del ADN de los restos hallados en el cuerpo de Sonia Carabantes o en el lugar donde se encontró su cadáver. Se afirma en las noticias que de la información facilitada por los laboratorios se deduce una altísima probabilidad de que estos restos pertenecieran a la misma persona que había dejado sus huellas genéticas en un cigarrillo encontrado junto al cadáver de Rocío Wanninkhof. En la noticia publicada en El País se especifica el valor de esa probabilidad al decir "...un margen de certeza del 99.999997%, uno entre más de treinta y tres millones". Semejante valor despeja cualquier duda que pudiera tenerse sobre si corresponden, o no, al mismo individuo. Y sin embargo estas cantidades pudieran no ser lo que parecen.

La primera cuestión es ¿cuál de esas cifras proviene directamente de la información proporcionada por la muestra de ADN?. Mucho nos tememos que la probabilidad realmente calculada a partir de los datos del laboratorio sea la que se expresa en letras (uno entre treinta y tres millones, es decir 0.000003%) y no la que se expresa en cifras (99.999997%) que, lógicamente, capta más nuestra atención. El proceso seguido para obtener esa probabilidad debe haber sido el siguiente: los laboratorios han detectado una serie de marcadores genéticos, 17 según la noticia, de los que se conoce su frecuencia (probabilidad) en la población y, suponiendo que los marcadores son independientes, se ha calculado la probabilidad de que todos ellos aparezcan conjuntamente multiplicando las frecuencias (probabilidades) de cada uno de ellos. Este producto, expresado en porcentaje, es el que vale 0.000003%. Aclaremos que la frecuencia en la población de un marcador representa la probabilidad de que en un individuo elegido al azar esté presente el marcador.

La segunda cuestión es que, en realidad, la probabilidad calculada es una probabilidad condicionada. ¿Qué significa esto? Un pequeño ejemplo ayudará a comprender mejor de qué hablamos. Todo el mundo sabe que al lanzar un dado tenemos 1 posibilidad frente a 6 de que salga una cualquiera de sus caras. Con un poco más de rigor diríamos que la probabilidad de "salir" para cualquiera de las caras es $1/6$. Supongamos ahora que, lanzado el dado y sin conocer nosotros el resultado, se nos informa de que la cara que "salió" era par y se nos pregunta por la probabilidad de que haya sido un 2. Responder de nuevo $1/6$ sería poco razonable, porque deberíamos tener en cuenta que las caras pares del dado son 3 y la cara con el 2 es 1 de ellas. La respuesta correcta a la probabilidad de que haya salido un 2 es $1/3$, respuesta que viene condicionada por la información que se nos proporcionó. En Probabilidad se habla de probabilidad condicionada del suceso A (salir un 2) a otro suceso B (salir una cara par) y se la representa por $P(A|B)$. Tenemos pues un instrumento para calcular probabilidades en situaciones en las que disponemos de alguna información adicional. Pero éste es un instrumento que no siempre se utiliza o interpreta bien. Dados dos sucesos cualesquiera, A y B, no suele ser lo mismo $P(A|B)$ que $P(B|A)$, así que hay que ser cuidadoso en la forma en la que un resultado es "leído" o interpretado. Por ejemplo, si A representa el suceso "tener dos brazos y dos piernas" y B "ser un mono", $P(A|B)$ representa la probabilidad de que "si tenemos" un mono, éste tenga dos brazos y dos piernas y, como esto seguro que es cierto, $P(A|B)=1$ (100%). Sin embargo, $P(B|A)$ representaría la probabilidad de que "si tenemos" un individuo con 2 brazos y dos piernas, éste sea un mono y esto, estarán de acuerdo con nosotros, no siempre es cierto.

Volvamos ahora a nuestro caso. Si por A designamos el suceso “poseer la combinación de marcadores encontrada en el análisis del ADN de los restos encontrados en o junto a Sonia” y por B “un individuo elegido al azar”, el valor 0.000003% obtenido es $P(A|B)$, que representa la probabilidad de que un individuo elegido al azar presente esa combinación de marcadores genéticos.

Una vez hemos establecido cuál debe ser, de entre los varios datos manejados y publicados, el resultado obtenido a partir del análisis genético de los restos y que, en realidad, corresponde a una probabilidad condicionada establecida en unos términos precisos, pasemos a discutir dónde puede estar el problema. Al rastrear la base de datos que los laboratorios criminológicos poseen se ha encontrado un individuo C, relacionado con el caso Rocío Wanninkhof, que posee la misma combinación de marcadores (obsérvese que ya no estamos hablando de un individuo elegido al azar). La pregunta inmediata de los investigadores es si este individuo y el que dejó sus restos junto a Sonia son el mismo. Su respuesta, o lo que los periodistas han entendido como tal, es que sí lo son con una probabilidad del 99.999997%. Una cifra apabullante, pero ¿de dónde ha salido? Creemos que el valor 0.000003%, que vimos correspondía a $P(A|B)$, se ha interpretado como $P(B|A)$, es decir, como la probabilidad de que poseyendo la combinación de marcadores encontrada, el individuo que los posee sea un individuo cualquiera (y, por tanto, no el individuo C). Realizada esta errónea interpretación de la probabilidad inicialmente calculada, la complementaria ($100 - 0.000003 = 99.999997$) representará la probabilidad de que, con esa combinación de marcadores genéticos, C y el individuo relacionado con Sonia sean la misma persona. Así pues, se habría cometido el error que señalábamos antes de confundir una probabilidad condicionada con su “opuesta”. Este error de interpretación es más corriente de lo deseable, hasta el punto ser conocido en estadística forense como “la falacia del fiscal”.

Alguien podrá pensar que cuanto precede es un mero ejercicio académico y que, a pesar de todo, con una probabilidad tan baja difícilmente se tratará de individuos distintos. Dos razones para disuadir a quien así piense. La primera es que en asuntos tan delicados todo rigor es poco. La segunda que en la literatura forense pueden encontrarse casos de sentencias anuladas por estar dictadas tomando en consideración “la falacia del fiscal”. Uno de ellos, el caso de un asalto con robo a una mujer ocurrido en 1964 en Los Ángeles (“El Pueblo contra Collins”, 68 Cal. 2d 319, 438 P. 2d 33, 66 Cal. Rptr. 497), es un ejemplo clásico que aparece en algunos textos de introducción a la Probabilidad. En esta ocasión, la Corte Suprema de California anuló la sentencia de culpabilidad dictada por la Corte Superior del Condado de Los Ángeles aun cuando la probabilidad involucrada era de 1 entre 12 millones, del mismo orden de magnitud que la manejada aquí.

Angel Corberán y Francisco Montes
Profesores del Departament d'Estadística i Investigació Operativa
Universitat de València