

# Composición de evidencia a partir de múltiples pruebas de ADN<sup>1</sup>

Sam C. Saunders, N.

Chris Meyer y Dane W. Wu

Mathematics Magazine; Febrero de 1999; 72, 1; Academic Research Library, pg. 39

## 1. INTRODUCCIÓN

El sensacional juicio del héroe del fútbol americano O. J. Simpson, en 1995, por el brutal asesinato de su ex-esposa Nicole y de Ron Goldman, amigo de Nicole, se ha convertido en uno de los juicios más controversiales en la historia americana, comparable con el juicio de John Scope, en 1925, quien fue llevado a la corte por enseñar la Teoría de la Evolución en su clase de Biología. Ambos juicios son ejemplos notables de decisiones controversiales y divisivas emanadas del sistema jurídico americano.

El sistema jurídico americano requiere que el jurado mantenga al acusado en la condición de inocente hasta que se le compruebe la culpabilidad más allá de una duda razonable. Si los jurados llegan a confundirse acerca del grado de incriminación aportada por ciertas piezas de evidencia, entonces una duda (razonable o no) aparecerá. La decisión del jurado en el juicio de O. J. se basó en la comprensión alcanzada por los miembros del jurado sobre las probabilidades de coincidencia de ADN en muestras de sangre, una técnica para incriminar al acusado que se intentó en este juicio por primera vez. Se trataba de utilizar como argumento de incriminación que si el ADN de muestras de sangre tomadas en la escena del crimen coincidía con el del acusado, esto demostraría que el acusado, por lo menos, estuvo en la escena

del crimen. Algunos expertos afirman que esta prueba de ADN solamente puede usarse para exonerar a personas acusadas de un crimen, pero nunca para establecer su culpabilidad (ver, por ejemplo, [7]). La absolución del jurado fue controversial. Un columnista ([2]) lo atribuye a su analfabetismo numérico. Otros dijeron que los fiscales no estaban en capacidad de explicar lo que ellos mismos no entendían. La justicia requiere que las probabilidades de eventos tales como la coincidencia de huellas dactilares, o de ADN tomado de muestras de sangre, sean comprensibles. En esta nota no nos preocuparemos por los problemas químicos de las coincidencias de ADN. Tampoco consideraremos la posibilidad de contaminación de una muestra de sangre o de la conspiración de los organismos policiales para falsear evidencia. (Estas complicaciones se discuten en [4].) Nos ocuparemos únicamente de algunas de las sutilezas de las probabilidades condicionales en relación con las coincidencias de ADN.

## 2. PROBABILIDADES CONDICIONALES Y CULPABILIDAD

Sea  $M$  el evento de una coincidencia de ADN entre la sangre del acusado y la sangre encontrada en la escena del crimen, sea  $I$  el evento que denota que el acusado es inocente y sea  $I'$  el complemento de  $I$ , es decir el evento en el que el acusado es culpable. Existen medidas confiables para  $P(M|I)$ , la probabilidad condicional de  $M$  dado  $I$ . Si el acusado es inocente la sangre debe ser de alguien más. Por consiguiente  $P(M|I)$  es la probabilidad de una coincidencia de ADN entre muestras de sangre de dos individuos diferentes.

Las firmas humanas de ADN son tan distintivas que la probabilidad de una coincidencia debida al azar es muy baja. Dependiendo en la cantidad de ADN obtenido, comúnmente se encuentra que  $P(M|I)$  está entre  $10^{-8}$  y  $10^{-10}$ . Pero es  $P(I|M)$ , la verosimilitud de inocencia dada la evidencia, la que debería ser considerada por los jurados, no  $P(M|I)$ . Estas dos probabilidades condicionales se relacionan en el *teorema pequeño* de Bayes:

<sup>1</sup> Versión en español de Javier Olaya Ochoa y Daniel Arbeláez Rojas (Universidad del Valle), y Deycy Janeth Sánchez Preciado y Carolina Delgado Hurtado (Universidad del Cauca).

$$P(I|M) = \frac{P(I)}{P(M)} P(M|I) \quad (1)$$

De la ecuación (1) vemos que  $P(I|M) \neq P(M|I)$ , a menos que  $P(I) = P(M)$

La suposición errónea que  $P(I|M)$  es igual a  $P(M|I)$  es llamada algunas veces *la falacia de los fiscales*. (Una discusión sobre la confusión que causa esta falacia puede verse en [3]).

Si en la ecuación (1) sustituimos  $P(M)$  utilizando el Teorema de Probabilidades Totales, obtenemos el *teorema grande de Bayes*:

$$P(I|M) = \frac{P(I)P(M|I)}{P(I)P(M|I) + P(I')P(M|I')}$$

Se sigue, dado que  $(1+x^{-1})^{-1} < x$  cuando  $x > 0$ , que

$$P(I|M) < \frac{P(I)P(M|I)}{P(I')P(M|I')} \approx \frac{P(I)P(M|I)}{P(I')} \quad (2)$$

La igualdad aproximada se sigue porque  $P(M|I')$ , la probabilidad de una concordancia dado que el acusado es culpable, es virtualmente igual a uno. Aquí,  $P(I)$  y  $P(I')$  son las probabilidades a priori de inocencia y culpabilidad, antes de que considerar la evidencia. Así, se deduce de (2), que la falacia de los fiscales no es tanto una falacia si no la ignorancia de un factor de proporcionalidad (la disparidad a priori,  $P(I)/P(I')$  en un límite superior de la verdadera probabilidad de inocencia dada una concordancia de ADN.

### 3. LA EVIDENCIA COMPUESTA

Er muchos casos, como en el juicio de O. J., más de un análisis de concordancia de ADN está involucrado. Nuestros cálculos hasta aquí se refieren a una pieza de evidencia,  $M$ . El jurado necesita considerar el efecto combinado de toda la evidencia (ver [6]). Supongamos que los eventos  $M_1, \dots, M_K$  son (todos) piezas de evidencia presentadas contra el acusado.

Deseamos calcular  $P(I|\bigcap_{i=1}^K M_i)$ , la probabili-

dad de inocencia dada toda la evidencia. El Teorema 1 provee un límite superior para esta probabilidad. Introduciremos primero algunas definiciones necesarias.

Para un evento dado  $I$ , denotamos  $P_I(\cdot) = P(\cdot | I)$ . Dos eventos  $M_1$  y  $M_2$  se dice que son condicionalmente independientes con respecto a  $I$  si  $P_I(M_2 | M_1) = P_I(M_2)$ . Note que los eventos  $M_1$  y  $M_2$  pueden ser intercambiados en esta definición y que la independencia condicional se convierte en independencia mutua si  $I$  es un evento seguro.

El cociente  $P(M_2 | M_1) / P(M_2)$  da una medida de qué tan fuertemente asociados están  $M_2$  y  $M_1$ . Si la razón es menor que uno, entonces  $P(M_2 | M_1) < P(M_2)$ , lo cual significa que  $M_2$  es menos probable de ocurrir dado  $M_1$ . De la misma forma,  $M_2$  es más probable de ocurrir dado  $M_1$  si la razón es mayor a 1. La razón es 1 si y solo si  $M_1$  y  $M_2$  son mutuamente independientes.

Con esto en mente, decimos que dos eventos  $M_1$  y  $M_2$  están más fuertemente asociados condicionalmente dado  $I'$  que dado  $I$  si

$$\frac{P_{I'}(M_2 | M_1)}{P_{I'}(M_2)} \geq \frac{P_I(M_2 | M_1)}{P_I(M_2)};$$

lo que es equivalente a

$$\frac{P_{I'}(M_1 M_2)}{P_{I'}(M_1)P_{I'}(M_2)} \geq \frac{P_I(M_1 M_2)}{P_I(M_1)P_I(M_2)} \quad (3)$$

por definición de probabilidad condicional.

$M_1$  y  $M_2$  se dice que son igualmente asociados condicionalmente bajo  $I$  e  $I'$  si en (3) se satisface la igualdad (ver [1]). Dejamos al lector la prueba del siguiente hecho: si  $M_1$  y  $M_2$  son condicionalmente independientes con respecto a ambos eventos  $I$  e  $I'$ , entonces ellos son igualmente asociados condicionalmente bajo  $I$  e  $I'$ .

El siguiente ejemplo ilustra la asociación igual. Consideremos lanzar dos dados balanceados, y denotemos  $(X, Y)$  el resultado observado. Sean  $M_1$  y  $M_2$  los eventos:  $M_1 = \{(x, y): y = 3, 4, \text{ ó } 5\}$  y  $M_2 = \{(x, y): x = 1 \text{ ó } 2\}$ . Sea además  $I$  el evento:  $I = \{(x, y): x + y = 7\}$ . En este caso  $M_1$  y  $M_2$  son igualmente asociados condicionalmente bajo  $I'$  e  $I$ , porque:

$$\frac{P_I(M_1 M_2)}{P_I(M_1)P_I(M_2)} = \frac{P_{I'}(M_1 M_2)}{P_{I'}(M_1)P_{I'}(M_2)} = 1$$

Si modificamos el evento  $M_2$ , de tal forma que  $M_2 = \{(x, y): x = 1 \text{ ó } 2\} \cup \{(3, 3)\}$ , entonces  $M_1$  y  $M_2$  están más fuertemente asociados condicionalmente dado  $I'$  que dado  $I$ , pues, en este caso:

$$\frac{P_I(M_1 M_2)}{P_I(M_1)P_I(M_2)} = 1 \quad \text{y} \quad \frac{P_{I'}(M_1 M_2)}{P_{I'}(M_1)P_{I'}(M_2)} = \frac{12}{11}$$

En general, los eventos  $M_1, \dots, M_k$  se dice que son más fuertemente asociados condicionalmente dado  $I'$  que dado  $I$  si:

$$\frac{P_{I'}\left(\bigcap_{i=1}^k M_i\right)}{\prod_{i=1}^k P_{I'}(M_i)} \geq \frac{P_I\left(\bigcap_{i=1}^k M_i\right)}{\prod_{i=1}^k P_I(M_i)} \quad (4)$$

Tenemos una asociación condicional igual cuando la cualidad en (4) se mantiene. (Esta definición está relacionada con el concepto de asociación; para más detalles, ver [1]).

**Teorema 1:** Si los eventos  $M_1, \dots, M_k$  son más fuertemente asociados condicionalmente dado  $I'$  que  $I$ , entonces

$$P\left(I \mid \bigcap_{i=1}^k M_i\right) \leq \frac{P(I)}{P(I')} \prod_{i=1}^k \frac{P(M_i | I)}{P(M_i | I')} \quad (5)$$

**Prueba:** Del teorema grande de Bayes obtenemos:

$$P\left(I \mid \bigcap_{i=1}^k M_i\right) = \frac{P(I) \times P\left(\bigcap_{i=1}^k M_i \mid I\right)}{P(I) \times P\left(\bigcap_{i=1}^k M_i \mid I\right) + P(I') \times P\left(\bigcap_{i=1}^k M_i \mid I'\right)} \quad (6)$$

$$\leq \frac{P(I)P\left(\bigcap_{i=1}^k M_i \mid I\right)}{P(I')P\left(\bigcap_{i=1}^k M_i \mid I'\right)}$$

Del hecho que  $M_1, \dots, M_k$  son más fuertemente asociados condicionalmente dado  $I'$  que dado  $I$  tenemos

$$\frac{P(I)P\left(\bigcap_{i=1}^k M_i \mid I\right)}{P(I')P\left(\bigcap_{i=1}^k M_i \mid I'\right)} \leq \frac{\prod_{i=1}^k P(M_i | I)}{\prod_{i=1}^k P(M_i | I')} \quad (7)$$

Entonces (5) resulta de sustituir (7) en (6)

#### 4. UNA APLICACIÓN DEL TEOREMA

En relación con el juicio de O. J. Simpson, estamos interesados en evaluar la probabilidad de culpabilidad dada la totalidad de la evidencia de ADN. Sea  $M_1$  el evento que una gota de sangre encontrada cerca de los cuerpos de las víctimas es concordante con la sangre del acusado. Diagnósticos Cellmark, el laboratorio de ADN empleado en el juicio, dijo que solamente una persona en 170 millones podría esperarse que coincidiera con las marcas genéticas identificadas en la gota de sangre. Por lo tanto  $P(M_1 | I)$  es igual a  $(1.7 \times 10^8)^{-1}$  lo cual es igual a  $5.88 \times 10^{-9}$ . Sea  $M_2$  el evento que una muestra de sangre encontrada en una media del acusado, y encontrada en su cuarto, es consistente con la sangre de la víctima. Cellmark dijo que la probabilidad era de uno en 6.8 millones que otra persona coincidiera con las marcas genéticas que ellos encontraron en la sangre de la víctima (ver [3]). Por lo tanto  $P(M_2 | I)$  es igual a  $(6.8 \times 10^6)^{-1}$  lo cual es igual a  $1.47 \times 10^{-7}$ . Más aún, como ya se mencionó,  $P(M_1 | I') \approx 1$  y  $P(M_2 | I') \approx 1$ .

Para poder aplicar el teorema, uno debe argumentar que  $M_1$  y  $M_2$  están más fuertemente asociados condicionalmente dada la culpabilidad que dada la inocencia. Si  $I$  es verdad, entonces  $M_1$  y  $M_2$  son independientes. Pero asumiendo la culpabilidad, uno puede imaginarse muchos escenarios en los cuales la ocurrencia de uno incrementaría la probabilidad del otro. Los dos eventos de concordancia de la sangre serían razonablemente juzgados como más fuertemente asociados condicionalmente dado  $I'$  que dado  $I$ . Sustituyendo estos números en (5) da

$$P(I | M_1 M_2) \leq \frac{P(I)}{P(I')} \times P(M_1 | I) \times P(M_2 | I) \approx \frac{P(I)}{P(I')} \times 8.65 \times 10^{-19}$$

Busquemos ahora un límite superior para la razón  $P(I)/P(I')$ . Si nos vamos al extremo de decir que O. J. Simpson no era más probable de ser el asesino que cualquier otra persona en el mundo, entonces  $P(I')$  equivale a  $10^{-10}$  (realmente,  $P(I')$  es mucho más grande que esto), y por supuesto  $P(I) \approx 1$ . Entonces,  $P(I)/P(I')$  equivale a  $10^{10}$  y  $P(I | M_1 M_2)$  es menor o igual a  $8.65 \times 10^{-9}$ . La probabilidad condicional de inocencia del acusado dadas ambas coincidencias de ADN es lo suficientemente pequeña como para declarar culpable al acusado más allá de cualquier duda razonable.

Muchos otros eventos se podrían utilizar para condicionar los eventos en el juicio. Por ejemplo, el guante encontrado en la casa del acusado coincidía con uno encontrado en la escena del crimen y reveló marcas genéticas no sólo concordantes con las del ADN de las víctimas sino también con las del acusado; designemos a esta coincidencia  $M_3$ . Sea  $M_4$  la concordancia de la sangre del acusado con la de gotas que formaban una huella desde la escena del crimen hasta la puerta. De acuerdo al analista de ADN del Estado de California, esta huella hubiera podido ser dejada por una persona en 240.000, incluyendo a O. J. Sea  $M_5$  la concordancia del ADN extraído de una gota de sangre salpicada en el Bronco blanco, la cual es consistente con la de Ron Goldman. Antes afirmamos que  $M_1$  y  $M_2$  eran indepen-

dientes dado  $I$ . Ahora  $M_1, \dots, M_5$  no son independientes dado  $I$  porque algunos involucran a las mismas personas. Como sea, no necesitamos corroborar dicha independencia para poder aplicar ese teorema; solamente necesitamos aseverar que  $M_1, \dots, M_5$  están más fuertemente asociados condicionalmente dado  $I'$  que dado  $I$ .

## 5. UNA ANALOGÍA INSTRUCTIVA

Según comentarios de algunos de los jurados posteriores al juicio, su opinión es que la prueba de coincidencia de ADN no es más contundente que la coincidencia de las huellas dactilares. Una simple analogía puede ilustrar la confiabilidad de la evidencia provista por la prueba de ADN. Cada persona carga en cada célula de su sangre ADN que contiene información análoga a la información de una permutación de una baraja de cartas. Con una baraja de 52 cartas es posible formar aproximadamente  $52! \approx 8.0658 \times 10^{67}$  permutaciones. Supongamos que, de una gota de sangre en la escena del crimen, hay sólo unos pocos puntos de ADN (correspondientes a las posiciones de una carta en la baraja) de la cual se puede extraer información.

Algunas veces, sin embargo, dependiendo de la cantidad de ADN o del grado de contaminación, no puede determinarse toda la información de cada punto. La situación análoga de cartas puede ser que únicamente el número, la figura, o el color (digamos N (Negro) o R (Rojo)) de la carta puede ser determinada. ¿Qué tanto puede esto incriminar al acusado? Asuma que se dispone de información parcial en sólo 7 posiciones de cartas de la siguiente manera (estas serían las marcas del ADN de la sangre encontrada en la escena del crimen):

(♠, ♦, 2-♥, 3-♣, 10-♥, 4-♦, N)

Correspondientemente, suponga que en las mismas posiciones de ADN en la sangre del acusado encontramos:

(2-♠, 10-♦, 2-♥, 3-♣, 10-♥, 4-♦, K-♠)

Este patrón provee una serie de marcas de ADN comparable con las de la sangre del sospechoso (recogida en la escena del crimen). Asumiendo que todas las permutaciones son igualmente probables, la probabilidad de obtener esta concordancia es:

$$\frac{13}{52} \times \frac{13}{51} \times \frac{1}{50} \times \frac{1}{49} \times \frac{1}{48} \times \frac{1}{47} \times \frac{24}{46} \approx 6.0154 \times 10^{-9}$$

Solamente en unas 6 personas entre mil millones se produciría esta coincidencia.

**Agradecimientos:** Los autores desean agradecer a los evaluadores por sus constructivas sugerencias.

## Referencias

- R. E. Barlow y F. Proschan, *Statistical Theory of Reliability and Life Testing, To Begin With*, Silver Spring, MD, 1081, 31.
- DAVE Barry, Happy New Year, column in the December 31, 1995, Knight-Ridder newspapers.
- ROBIN Clark, Will DNA turn the tide?, *Philadelphia Inquirer*, May 21, 1995.
- R. Gomulkiewicz y N. A. Slade, Legal Standards and the significance of DNA evidence, *Human Biology*, 69, 1997.
- EDWARD J. Larsen, *Summer for the Gods: The Scopes Trial and America's Continuing Debate over Science and Religion*, Basic Books, New York, NY, 1997.
- S. C. Saunders, An application of conditional probability, or, O. J. + DNA = QED, *Mathematics notes from Washington State University*, Vol. 38, No. 2, 1995.
- DAVID Pringle, Who's the DNA fingerprinting pointing to? *New Scientist*, January 29, 1994.