

¿POR QUÉ EL LR EN ADN?

Coronel D. José Juan Lucena Molina
Doctor en Ciencias Forenses por la Universidad de Lausana (Suiza)
Director de la Escuela de Especialización de la Jefatura de Enseñanza de la Guardia Civil

0. Introducción

El presente trabajo recoge y sintetiza información de reconocidos expertos en la comunidad científica internacional especializada tanto en técnicas de ADN forenses como en valoración estadística de ese tipo de pruebas (consúltase la bibliografía referenciada al final del documento). Puede ser útil a quienes, en el ámbito de su trabajo profesional y con independencia de su actuación en un proceso judicial, necesiten profundizar en aspectos básicos de carácter estadístico que pueden encontrarse en informes periciales de ADN emitidos por laboratorios forenses.

Quien lo escribe no es experto en ADN sino en valoración estadística de cotejos criminalísticos. No obstante, el autor ha querido que su contenido fuera revisado por expertos en ADN como la Dra. Lourdes Prieto Solla, Facultativo responsable de I+D del Laboratorio de ADN de la Comisaría General de Policía Científica, investigadora colaboradora del Catedrático de Medicina Legal de la Universidad de Santiago de Compostela Dr. D. Angel María Carracedo Álvarez, quienes se caracterizan, entre otras muchas cosas, por llevar años realizando un más que notable esfuerzo pedagógico entre juristas españoles para facilitarles una adecuada comprensión de la prueba de ADN, y por el biólogo forense Dr. D. Antonio Alonso Alonso, Facultativo del Instituto Nacional de Toxicología, a los cuales les estoy especialmente agradecido. La Dra. Prieto es autora de numerosos comentarios que han acabado insertados en diversos pasajes del documento durante su proceso de revisión.

1. Una comprensión intuitiva de la relación de verosimilitudes (LR)

Parece un signo distintivo de nuestra época tener que familiarizarse con acrónimos relacionados con palabras escritas en inglés en todas las ciencias. También ocurre eso en el ámbito de un cotejo criminalístico de perfiles genéticos de origen humano: DNA, PCR, STR, LR, etc. De todas esas siglas sólo la última se relaciona directamente con la valoración estadística de las coincidencias encontradas en los marcadores genéticos comparados y obtenidos de una muestra dubitada y otra indubitada. El contexto de este artículo es el de una investigación de un hecho criminal a partir de indicios recogidos en la escena del crimen.

LR se corresponde con las letras iniciales de las palabras inglesas “Likelihood Ratio”, que traducimos por “Relación de verosimilitudes”.

Hay que tener cuidado con el término “verosimilitud” empleado en la expresión del LR porque se trata de un término técnico con significado específico. Se predica de una proposición - forma en que se llama en lógica a los juicios, estructuralmente caracterizada por un sujeto y un predicado nominal unidos por el verbo ser - susceptible de ser verificable, es decir, de la que puede comprobarse su veracidad o falsedad, pero no aisladamente porque se predica de *una proposición condicionada a otra*. Si no fuera así, en nada se distinguiría esencialmente de su significado común.

Pongamos un ejemplo: *el perfil dubitado procede del sospechoso* puede transformarse en *el sospechoso es la fuente del perfil dubitado*. Se trata de una proposición que pudiera sostener la

acusación en un procedimiento penal. Tiene carecer hipotético porque es lo que hay que probar. El sujeto de la frase es “el sospechoso”, el verbo ser tiene función copulativa y, en última instancia, colocamos el predicado “la fuente del perfil dubitado”. Se trata de un juicio formal en toda regla. En este ejemplo no se enuncia la proposición como una verosimilitud porque falta condicionarla a otra.

Siguiendo con el ejemplo, vamos encaminándonos para entender el término “verosimilitud” empleado en las proposiciones utilizadas en los LRs dando el siguiente paso: *el perfil dubitado procede del sospechoso dado que se ha verificado la coincidencia de los perfiles genéticos comparados*. No sólo enunciamos la procedencia del perfil hallado en el vestigio sino que añadimos que se ha verificado previamente que los perfiles comparados son coincidentes. Pero aún alcanzando este nivel de comprensión hay un serio peligro de entender mal el término “verosimilitud” en el contexto considerado porque no es equivalente a probabilidad condicional, sin más. No es lo mismo decir que queremos hallar la *verosimilitud* de que *el perfil dubitado proceda del sospechoso dado que se ha verificado la coincidencia de los perfiles genéticos comparados*, que decir que queremos hallar la *probabilidad* de exactamente lo mismo, porque la verosimilitud enunciada es equivalente a querer hallar una específica probabilidad condicional: la probabilidad de *verificar la coincidencia de los perfiles genéticos comparados dado que el perfil dubitado procede del sospechoso*. Se ha producido una inversión en el orden de los términos de la frase compuesta que nos indica que aunque la verosimilitud tiene estructura lógica de una probabilidad condicional, no es cualquier tipo de probabilidad condicional. Verosimilitud y probabilidad condicional no son términos intercambiables en todo caso.

La estructura lógica especial de una verosimilitud en ciencia forense con respecto a la de una probabilidad condicional se entiende bien si se percibe que lo verosímil se predica de una determinada proposición de carácter hipotético. En este contexto, las proposiciones de las que queremos conocer su verosimilitud son las siguientes: “el perfil dubitado procede del sospechoso” o “el perfil dubitado no procede del sospechoso”, ambas claramente con el carácter mencionado porque enuncian lo que se quiere probar, pero la diferencia primordial entre el término verosimilitud empleado en los LRs respecto al sentido del término en el lenguaje común es que la probabilidad que interesa asignar es la de lo observado, expresado mediante el término “evidence” (evidencia), dadas cada una de las proposiciones hipotéticas consideradas.

El término “evidence” en inglés es polisémico en extremo, mucho más que el término español “evidencia”. En el contexto de un cotejo de perfiles de ADN se expresa con ese término una determinada coincidencia entre los marcadores genéticos de los perfiles que se comparan. Los ingleses lo llaman “match”.

Una coincidencia de marcadores genéticos comparando dos perfiles de ADN entre sí es una proposición fáctica. Si se observa, constatamos un hecho, no una hipótesis. Las proposiciones que son susceptibles de que puedan ser o no verosímiles son las que tienen carácter hipotético, la coincidencia de marcadores observada no. Por eso, lo que condiciona en una verosimilitud son las proposiciones hipotéticas, no lo observado.

Lo explicado hace que enunciar el significado de un LR como relación de verosimilitudes no sea algo intuitivo, ni siquiera sencillo. Si se explica qué es un LR diciendo que es una relación entre probabilidades condicionales no se comete error alguno, porque ciertamente lo es, pero si lo que condiciona en las probabilidades condicionales constitutivas de una relación de

verosimilitudes no son proposiciones hipotéticas, se cometería un grave error porque esas probabilidades condicionales no serían realmente verosimilitudes.

Esclarece el concepto de verosimilitud empleado en ciencia forense pensar que se trata de una probabilidad de lo observado por el experto condicionado a una proposición hipotética planteada por una de las partes del proceso penal. Si se habla de relación de verosimilitudes, con ello se expresa que el LR es esencialmente relativo. Existe una comparación entre verosimilitudes que, matemáticamente expresamos mediante una división.

La relatividad esencial del concepto técnico de verosimilitud empleado en ciencia forense obedece a la idea de que la fuerza probatoria de una proposición fáctica, como es la expresión del resultado de un cotejo criminalístico de cualquier naturaleza, no puede hallarse si no se contemplan, al menos, dos hipótesis explicativas del hallazgo alternativas. Simplificando, podemos ejemplificar esto diciendo que la fuerza probatoria de un “match” entre perfiles de ADN no es posible hallarla si no contemplamos la posibilidad de observar ese “match” considerando que el sospechoso sea la fuente del perfil hallado en la muestra dubitada frente a que otra persona diferente al sospechoso – de la misma población de referencia - lo sea. Dejemos para más adelante qué es eso de la población de referencia.

2. Profundizando en el significado de la probabilidad

El término probabilidad en las ciencias forenses tiene dos significados principales [1]:

- a) Probabilidad aleatoria o deductiva: cuando para medir la incertidumbre de un suceso aplicamos un razonamiento deductivo. Como ejemplo paradigmático podemos citar el lanzamiento de dados y monedas, supuestamente equilibrados. En estos casos cabe deducir resultados partiendo del conocimiento que se tiene del funcionamiento ideal de esos objetos.
- b) Probabilidad epistémica o inductiva: cuando para medir la incertidumbre de un suceso aplicamos un razonamiento inductivo. Como ejemplos paradigmáticos podemos citar un muestreo aleatorio (incautaciones de drogas), el cálculo de probabilidades de sucesos no repetibles (que un determinado sospechoso sea el autor de un determinado crimen) o los estudios probabilísticos de fenómenos a partir de observaciones sistemáticas (con qué probabilidad podría llover un determinado día en una ciudad).

El estudio de las frecuencias alélicas de los marcadores de ADN en una población es netamente inductivo. Sin embargo, la atribución de una frecuencia alélica a una determinada persona perteneciente a una población de referencia exige un razonamiento de naturaleza deductiva.

Resulta muy relevante ser conscientes de que en el cálculo de probabilidades epistémicas o inductivas es preciso asumir el principio de uniformidad que puede enunciarse diciendo que se asume que los procesos que se observan en el presente, sucedieron en el pasado y sucederán en el futuro con similares características. Sin este principio no serían posibles las ciencias basadas en la observación. Sería imposible realizar inferencias o estimaciones a partir de los datos observados.

En el cálculo de la probabilidad de coincidencia aleatoria entre un perfil de ADN hallado en una muestra dubitada y el de una persona distinta al sospechoso y seleccionada al azar en una

población de referencia acorde con el caso, se realiza un razonamiento probabilístico deductivo partiendo de las frecuencias alélicas de cada uno de los marcadores comparados.

Podemos utilizar, para cada frecuencia alélica - igualmente podría aplicarse a un perfil completo -, el denominado "silogismo estadístico" que, siguiendo a los autores de la obra titulada *Bayesian Networks and Probabilistic Inference in Forensic Science* [2] puede enunciarse así:

1. La frecuencia relativa del alelo Q en la población R es γ .
2. a_i es el autor del crimen y un individuo de la población R ($i = 1, 2, \dots, n$).
3. No se conoce de a_i nada más que sea relevante – además de pertenecer a la población R - con respecto a la posibilidad de que tenga el alelo Q.
4. La probabilidad de que el autor del crimen a_i tenga el alelo Q es γ .

El ejemplo sigue el esquema de un silogismo aristotélico simple o categórico a pesar de que su enunciado no facilita esa percepción. Recordemos que en un silogismo de este tipo, el antecedente relaciona dos términos – uno denominado *mayor* y otro *menor* - con un tercero – denominado *medio* -, y se deduce, necesariamente, una conclusión que une o separa los dos primeros términos:

Las proposiciones 1, 2 y 4 enumeradas más arriba tienen el esquema clásico mínimo de un juicio: sujeto, verbo ser y predicado nominal. La proposición 3 no parece tener ese esquema pero puede enunciarse de esta otra forma para que advirtamos más fácilmente que también lo sigue:

3. Cualquier otra razón relevante que pueda explicar que a_i tenga el alelo Q – además del hecho de pertenecer a la población R – es desconocida.

El antecedente del silogismo está conformado por el término menor "*la probabilidad de que a_i , autor del crimen y perteneciente a la población R, tenga el alelo Q*", el término medio "*la frecuencia relativa del alelo Q en la población R*", y el término mayor " γ ".

Recordemos, igualmente, que el término medio es el que permite realizar la inferencia y el término mayor es el que tiene un carácter más universal. La premisa mayor incluye el término mayor y la premisa menor el término menor.

Por tanto, en forma estrictamente clásica se enunciaría la argumentación de esta otra forma (omitiendo, de momento, la proposición 3):

Premisa menor: La probabilidad de que a_i , autor del crimen y perteneciente a la población R, tenga el alelo Q es la frecuencia relativa del alelo Q en la población R.

Premisa mayor: La frecuencia relativa del alelo Q en la población R es γ .

Conclusión: La probabilidad de que a_i , autor del crimen y perteneciente a la población R, tenga el alelo Q es γ .

La referida proposición 3, ahora omitida, está implícita en la premisa menor y hace depender la veracidad de esa premisa de un estado de conocimiento personal. Esta premisa no puede

considerarse cierta por sí misma. Por tanto, el silogismo no concluye: la conclusión no se deriva necesariamente de las premisas.

Lo que acaba de referirse es aplicable a cualquier propiedad observable de la que pueda obtenerse una frecuencia relativa de aparición en la naturaleza o en la realidad fruto de la actividad humana. Por tanto, lo descrito es extensible a cualquier otra clase de cotejos criminalísticos como los de huellas dactilares, cristales, fibras, pinturas, etc.

Durante el siglo XX se ha abordado esta problemática en profundidad y se ha dado una respuesta satisfactoria a este problema dentro del bayesianismo subjetivo. El denominado silogismo estadístico bayesiano solventa la dificultad apuntada consiguiendo que el silogismo sea realmente deductivo.

Esta escuela estadística lo ha resuelto gracias a su concepto de probabilidad subjetiva que puede enunciarse diciendo que se trata del grado de creencia que tiene un sujeto sobre la ocurrencia de un suceso. En el presente caso, la proposición 3 se enunciaría de esta manera:

3. Para mí, la probabilidad de que cualquier individuo de la población R tenga el alelo Q es la misma.

Puede demostrarse que con este concepto de probabilidad (probabilidad subjetiva propia de la escuela estadística bayesiana) se logra que el silogismo estadístico concluya. El razonamiento se fundamenta en el conocido concepto de esperanza matemática (consúltese Anexo I para su desarrollo matemático).

A continuación se inserta la traducción de una parte de la obra titulada *“Introducción a la Estadística para Expertos Forenses”* [1] que esclarece aún más las diferencias entre probabilidades aleatorias o deductivas y las epistémicas o inductivas:

“Hasta qué punto las probabilidades de sistemas deducidos o inducidos son iguales es algo abierto a debate. La probabilidad deducida no puede aplicarse más que a un sistema notional. Un dado puede ser calificado como equilibrado, pero un dado real siempre tendrá pequeñas imperfecciones y defectos que no lo harán perfectamente equilibrado. En alguna medida, la posición aleatoria es artificial y tautológica. Cuando definimos qué es un dado, conocemos las propiedades del dado en sentido absoluto. No es posible tener esa clase de conocimiento en un sistema real observable. Simplemente, utilizamos la noción como un entorno de trabajo adecuado para desarrollar el cálculo de probabilidades, el cual, cada vez que se utiliza se aplica sobre sistemas probabilísticos fundamentalmente epistémicos. Asimismo, debido a que las inferencias hechas sobre la población están basadas en la observación de unos pocos miembros de esas poblaciones, como parte de esa inferencia es inevitable que se deduzca algún grado de incertidumbre aleatoria.

Como todas las probabilidades reales son inducidas por observación y son, esencialmente, frecuencias, ¿eso significa que las probabilidades sólo pueden consistir en afirmaciones sobre las proporciones relativas de las observaciones en la población? Y, si eso es así, ¿tiene sentido hablar de probabilidad respecto a un suceso único del que tengamos especial interés?

La idea de frecuencia en un resultado de un suceso único es ridícula, pues el resultado de interés ocurre o no ocurre. Como resultado de la tirada de un dado de seis lados no podemos

obtener un resultado igual a $1/6$ con la cara del seis boca arriba; o bien está la cara del seis boca arriba o no lo está. No existe un estado físico que se corresponda con la probabilidad de $1/6$ para un suceso único. Si uno lanza un dado de seis lados 12 veces, entonces el estado físico correspondiente a la probabilidad de $1/6$ se correspondería con la observación de dos veces la cara del seis boca arriba. Pero no puede haber ningún suceso físico singular que se corresponda con la probabilidad de $1/6$.

La única forma en la que un suceso único puede cuantificarse mediante una probabilidad es concebirla como un producto de la mente, en resumen, manteniendo una interpretación idealista de la probabilidad. Esto es lo que los estadísticos llaman probabilidad subjetiva y se trata de una interpretación de la probabilidad que establece que la probabilidad es una función de, y sólo existe en, la mente del interesado en un determinado suceso. Ese es el motivo de por qué se le llama subjetiva, no porque sea, de alguna manera, infundada o inventada, sino porque descansa en una interpretación idealista de la probabilidad.

Una interpretación realista de la probabilidad es la que está relacionada con frecuencias y número de resultados en una larga repetición de sucesos, haciendo inferencias sobre las proporciones de los resultados en poblaciones más amplias. Una interpretación realista de la probabilidad no sería capaz de pronunciarse sobre el resultado de un suceso único puesto que lo que puede afirmarse es una creencia, que no puede existir en el mundo real y, por tanto, requiere alguna noción idealista de la probabilidad. Las posiciones realistas implican que hay algo en el mundo observado que causa incertidumbre, considerándose tal incertidumbre como una propiedad externa a la mente del observador. Alguien quizá sostenga que esas probabilidades externas son propensiones del sistema en cuestión a comportarse de una específica forma. Desafortunadamente, la teoría propensiva de la probabilidad genera los mismos problemas que la concepción realista cuando se aplica a un suceso único porque una propensión no puede observarse directamente y sería un producto de la mente. En muchos aspectos, las interpretaciones realistas pueden ser más productivas para el científico como consecuencia de las demandas de hipotetizar o encontrar algún factor explicativo subyacente. Esto está en contraste con las posiciones idealistas donde encontrar las causas que expliquen la incertidumbre es deseable, pero no algo absolutamente necesario, pues la incertidumbre reside en la mente.

Esta distinción entre realistas e idealistas no es algo que se trate en las ciencias estadísticas y, ciertamente, esos términos no se utilizan. No hay estadísticos puramente realistas; todos los estadísticos desean realizar afirmaciones probabilísticas sobre sucesos únicos, por tanto, todos los estadísticos son, en algún grado, idealistas en su concepción de la probabilidad. Sin embargo, un debate en la ciencia estadística que refleja las posiciones realista/idealista es el existente entre frequentistas y bayesianos. Hay un teorema matemático sobre probabilidad denominado Teorema de Bayes y una escuela de estadísticos que se denomina bayesiana aparecida mucho después del enunciado del teorema. Las diferencias entre bayesianos y frequentistas no son matemáticas, el Teorema de Bayes es un teorema matemático y, de acuerdo con los principios de la teoría de la probabilidad, el Teorema de Bayes es correcto. Las diferencias entre ellos se encuentran en las distintas interpretaciones que tienen sobre la naturaleza de la probabilidad. Los frequentistas tienden a argumentar en contra de las probabilidades subjetivas, y defienden las interpretaciones de la probabilidad basadas en las frecuencias obtenidas mediante repetición de experimentos. Los bayesianos están a favor de las nociones subjetivas de la probabilidad y piensan que todas las incógnitas que son inciertas pueden expresarse probabilísticamente.

Esto conduce a una bastante interesante posición para los expertos forenses. Por una parte, ellos realizan trabajo experimental en el laboratorio donde es posible obtener resultados repetitivos; por otra parte han de interpretar datos como evidencia relacionados con sucesos únicos. Este último aspecto del trabajo del experto forense es, explícitamente, idealista, porque los sucesos relacionados con hechos criminales ocurren sólo una vez y sólo una, y requieren una interpretación subjetiva de la probabilidad para interpretar las probabilidades como grados de creencia. La faceta experimental de la ciencia forense puede acomodarse fácilmente a un punto de vista más realista de la probabilidad.

El punto de vista subjetivo de la probabilidad es uno de los que más fácilmente encaja entre las nociones de sentido común sobre la probabilidad, y el único que puede ser utilizado para cuantificar la incertidumbre sobre sucesos singulares. Existen algunos temores, entre los científicos, de que una probabilidad subjetiva es una probabilidad indemostrable, sin fundamento y sin apoyo empírico, y ciertamente, una probabilidad subjetiva podría ser todo eso. Pero la mayoría de las probabilidades subjetivas están basadas en frecuencias observadas empíricamente, y no son, como el término subjetivo quizá implique, algo sacado de la manga o inventado.

Hay un punto de vista sobre la naturaleza de la probabilidad que puede dejar a un lado muchos de los problemas y debates sobre el significado del término probabilidad. Se trata de la posición instrumentalista (Hacking, 1965) en la que simplemente no nos importa la interpretación exacta que demos a la probabilidad, sino que simplemente la vemos como un instrumento intelectual conveniente para elaborar cálculos sobre incertidumbre. La posición instrumentalista implica una cierta concepción idealista de fondo, en la que la probabilidad es un producto de la mente y no un componente fundamental del mundo material”.

La probabilidad de la ocurrencia de un suceso condicionado a la previa y efectiva ocurrencia de otro es algo que nos planteamos los seres humanos de manera bastante común. Por ejemplo cuando nos preguntamos si subirá hoy la tasa de polen si observamos un día soleado o, en un caso análogo, si subirá esa tasa si observamos un día lluvioso. Lo que observamos es el estado de la climatología y consideramos como proposiciones de las que nos interesa hallar la probabilidad la subida o no de la tasa de polen.

Las verosimilitudes de los LR utilizados en la ciencias experimentales consideran a las proposiciones de las que se predicen como algo ya sucedido o asumido como sucedido, algo de lo que se conoce su ocurrencia con certeza o se asume como tal, y a lo observado – expresable también como una proposición - como aquello de lo que quiere conocerse la probabilidad de su ocurrencia. Se predicen de proposiciones pero al añadirles, por ejemplo, “dado lo observado” o “condicionado a lo observado”, se está haciendo referencia a que se trata de probabilidades condicionales. Se desea conocer la probabilidad de lo observado condicionado a la ocurrencia de cada una de las proposiciones.

Siguiendo el ejemplo atmosférico, si empleamos el término verosimilitud de forma análoga a como se emplea con los LR de ADN puede afirmarse lo siguiente: preguntarse por la verosimilitud de que suba la tasa de polen si observamos un día soleado es decir exactamente lo mismo que preguntarse por la probabilidad de observar un día soleado si sube la tasa de polen.

Centrándonos de nuevo en el ejemplo del cotejo de perfiles de ADN, la verosimilitud de que *el perfil dubitado proceda del sospechoso dado que se ha verificado la coincidencia de los perfiles*

genéticos comparados, es lo mismo que decir, traducido en términos de probabilidad condicional, lo siguiente: la probabilidad *de que se verifique la coincidencia de los perfiles genéticos comparados dado que el perfil dubitado proceda del sospechoso*.

La probabilidad de que *el perfil dubitado proceda del sospechoso dado que se ha verificado la coincidencia de los perfiles genéticos comparados* no es la misma probabilidad condicional que la probabilidad *de que se verifique la coincidencia de los perfiles genéticos comparados dado que el perfil dubitado proceda del sospechoso*. Hay que fijarse, detenidamente, en el orden en que están escritas las proposiciones. Se puede apreciar, sin especial dificultad, que están invertidas de orden en las frases que comparamos.

No es difícil demostrar, matemáticamente, que esas probabilidades, con el orden de las proposiciones o, si se prefiere, de los sucesos cambiados entre sí, no son iguales, ordinariamente. Utilizando terminología matemática: $P(A | B) \neq P(B | A)$, generalmente, siendo A y B dos sucesos cualesquiera, P significando probabilidad y la barra vertical significando que un suceso está condicionado por el otro.

¿Qué credibilidad merece la proposición *“el perfil dubitado procede del sospechoso”* dadas las coincidencias encontradas en los marcadores genéticos comparados? Realmente eso es justamente lo que al Tribunal o Jurado les interesa como resultado de la práctica de la pericial de ADN: ¿qué debe creer el Tribunal o Jurado sobre la veracidad o falsedad de la proposición formulada – en este caso por la acusación - una vez proporcionado por el perito el resultado de su informe pericial?

La frase mencionada en el anterior párrafo no tiene la misma estructura lógica que las empleadas en una relación de verosimilitudes. Aunque se parezca mucho y tenga una estructura de probabilidad condicional – en eso coincide con las verosimilitudes matemáticamente formuladas -, los sucesos de la probabilidad condicional que en este ejemplo se contemplan están cambiados de orden: se pretende calcular la probabilidad de cada proposición condicionada a lo observado, no la probabilidad de lo observado condicionado a cada proposición. Lo que se considera cierto es lo observado, no las proposiciones.

Cuando se confunde la verosimilitud tal y como se emplea en los LR con la probabilidad condicional de una proposición condicionada a lo observado se comete la denominada *falacia de transposición del condicional*.

El procedimiento de la prueba de los hechos consiste en *verificar enunciados sobre hechos*. Esos enunciados coinciden con lo que hemos denominado proposiciones y la verificación consiste en comprobar su veracidad o falsedad. Sin embargo, la verificación de una proposición depende del tipo de razonamiento subyacente necesario: deductivo o inductivo. En el primer caso podremos llegar a la certeza a través de las reglas de la lógica ya enunciadas por Aristóteles en el siglo IV a.C., en el segundo caso será inevitable contar siempre con un cierto grado de incertidumbre.

La comprensión de la naturaleza epistemológica de los razonamientos deductivo e inductivo resulta clave para entender a qué tipo de certezas pueden llegar los científicos cuando practican las diversas ciencias. Con las denominadas ciencias exactas se logran certezas que se denominan metafísicas. Con las experimentales, tan sólo es posible lograr certezas físicas. Las primeras permiten formular conclusiones categóricas, las segundas sólo conclusiones

probabilísticas. Este es el marco lógico de referencia básico en el que las conclusiones de los informes periciales han de insertarse: salirse de él es abandonar el marco de la racionalidad y sumergirse en el de una intuición que no suele ser buena consejera para calcular probabilidades correctamente.

3. Origen histórico del LR [3]

Durante el siglo XX los epistemólogos lograron discernir y formular axiomáticamente en qué consiste valorar unos datos como prueba. Las distintas escuelas epistemológicas predominantes sostuvieron un enriquecedor debate que acabó sacando a la luz la denominada *ley de verosimilitud* [4] en 1965.

Las escuelas estadísticas vinculadas a la teoría de Neyman-Pearson, formulada en 1933, desarrollaron un concepto de valoración de los datos como prueba vinculado a una sola proposición. Trataban de ver hasta qué punto los resultados obtenidos eran compatibles con ella. Se propusieron principalmente dos leyes: la del cambio de probabilidad y la de la improbabilidad (consúltese el Anexo II si se quieren conocer más en profundidad). La segunda, ampliamente utilizada en la ciencia experimental por las propiedades de los *p-valores* y tests de rechazo de la teoría de Fisher, fue predominante durante el siglo XX.

Ian Hacking formuló su *ley de verosimilitud* vinculando la valoración de los datos como prueba a un par de hipótesis mutuamente excluyentes.

La diferencia descrita - valorar los datos como prueba respecto a una sola hipótesis o respecto a dos hipótesis mutuamente excluyentes - es la esencia del cambio de paradigma que la ciencia estadística experimentó a lo largo del siglo XX en esta materia y que se ha consolidado en el siglo actual.

De una manera informal, la ley de cambio de probabilidad puede resumirse diciendo que la forma de valorar los datos como prueba consiste en comparar la probabilidad de la veracidad de una proposición - una vez obtenidos los datos - con la probabilidad de la veracidad de esa misma proposición antes de obtenerlos. A medida que la primera probabilidad crece, mayor valor de prueba tienen los datos.

Igualmente, la ley de improbabilidad puede resumirse diciendo que la forma de valorar los datos como prueba consiste en averiguar en qué medida es posible obtener esos datos e, incluso, más extremos, si la proposición es cierta. A medida que sean menos probables, mayor valor de prueba tienen esos datos para rechazar la proposición.

A pesar de la aparente razonabilidad de las dos leyes expuestas, puede demostrarse matemáticamente que la ley de cambio de probabilidad es subjetiva y que la ley de improbabilidad es errónea. La subjetividad de la primera se refiere a la necesidad de contar con otras proposiciones y con las denominadas probabilidades a priori (antes de obtener los datos) de esas proposiciones para poder valorar los datos como prueba. La falsedad de la segunda estriba en que es dependiente del espacio muestral, es decir, de los datos observados disponibles, puesto que el valor de tales datos como prueba cambia en función de los datos de los que se disponga en cada caso siguiendo esta ley.

La *ley de verosimilitud* supera las dificultades apuntadas en las leyes anteriores: no depende de probabilidades a priori de proposiciones ni del espacio muestral disponible, pero necesita dos

proposiciones mutuamente excluyentes para que los datos puedan ser evaluados como prueba. La valoración de los datos como prueba es, por tanto, esencialmente relativa conforme a esta ley.

La *ley de verosimilitud* permite formular el *principio de verosimilitud* que consiste en afirmar que una serie de datos tienen el mismo valor de prueba si sus relaciones de verosimilitudes (LRs) se igualan, es decir, aunque los datos observados sean distintos entre sí y las parejas de proposiciones estén relacionadas con realidades totalmente diversas, si sus LRs son numéricamente iguales tienen el mismo valor de prueba.

No es infrecuente que en la traducción de textos de literatura anglosajona especializada se emplee la expresión “fuerza de la evidencia” en lugar de “valor de los datos como prueba” o “interpretar los datos como evidencia” en lugar de “interpretar los datos como prueba”. El término *evidence* en inglés recoge los significados de los términos prueba y evidencia en español.

La *ley de verosimilitud* la formuló Hacking utilizando categorías matemáticas. Informalmente puede enunciarse diciendo que consiste en sostener que los datos se interpretan como prueba correctamente desde la ciencia de la lógica si obtenemos la relación (transformada matemáticamente como una división) entre la probabilidad de observar esos datos si fuera cierta una de las proposiciones con la probabilidad de observar esos mismos datos si fuera cierta la proposición alternativa. La valoración de los datos como prueba – como ya se ha subrayado – es así esencialmente relativa. No cabe valorarlos bajo una sola proposición, como lo hacían las otras leyes.

En situaciones determinísticas se cumple que si la proposición A implica que, bajo ciertas condiciones, se observará x; y la proposición B implica que, bajo idénticas condiciones, se observará algo distinto a x, si esas condiciones se dan de hecho y se observa x, esa observación es evidencia a favor de A sobre B. Veamos el siguiente ejemplo:

Proposición A: el sospechoso es la fuente del perfil de ADN hallado en la muestra dubitada recogida en la escena del crimen.

Proposición B: otra persona distinta al sospechoso, procedente de una población de referencia acorde con el caso y no relacionado biológicamente con él, es la fuente del perfil de ADN hallado en la muestra dubitada recogida en la escena del crimen.

Observación x: coincidencia plena entre los marcadores comparados de los perfiles de ADN procedentes de una muestra dubitada, recogida en la escena del crimen, y una muestra indubitada obtenida del sospechoso u otra persona distinta al sospechoso y no relacionado biológicamente con él, procedente de una población de referencia acorde con el caso. Esa coincidencia plena suele expresarse en la literatura científica especializada mediante la palabra inglesa *match*.

Los expertos en ADN consideran que la probabilidad de observar una coincidencia plena entre los marcadores comparados de los perfiles de ADN procedentes de una muestra dubitada recogida en la escena del crimen y otra del sospechoso, si éste hubiera sido el autor del crimen, es igual a uno (generalmente, porque puede haber excepciones como la existencia de una mutación que implicase alguna discrepancia o una sospecha de posible pérdida de alelos). Consecuentemente, la probabilidad de observar una coincidencia no plena bajo esas mismas

condiciones de generalidad es igual a cero. Esta argumentación es la que se presenta en una situación determinística en la que los valores probabilísticos extremos se justifican por la fiabilidad de la tecnología disponible si se superan ciertos requisitos mínimos perfectamente definidos y por la capacidad de los expertos para describir acertadamente los perfiles en esas condiciones.

Si al aplicar la *ley de verosimilitud*, el numerador de la relación de verosimilitudes fuera igual a uno y el denominador igual a cero, tendríamos un ejemplo de situación determinística en el marco de esa ley. Cuando se realiza un análisis comparativo con un patrón perfectamente definido, por ejemplo, si se quiere determinar si una sustancia recogida en una inspección ocular de un posible delito de incendio es una gasolina, cuando se ha verificado un *match* con el patrón, la verosimilitud de que no sea gasolina lo observado mediante las técnicas cromatográficas puede considerarse nula siguiendo los protocolos estandarizados. La observación del *match* es evidencia a favor de la proposición de que se trata de gasolina frente a que no lo sea.

En el ejemplo de la prueba de ADN es fácil percibir que la relación de verosimilitudes de la ley enunciada proporciona una extensión de la forma de razonar determinística. La extensión consiste en decir que si la coincidencia observada entre los marcadores (*match*) es más probable si la proposición A de que el perfil de la muestra dubitada procede del sospechoso se considera cierta frente a que lo sea la proposición alternativa B, la ocurrencia de esa coincidencia constituye evidencia a favor de la proposición A sobre B. La fuerza de la evidencia está determinada por la medida de observar con mayor probabilidad la coincidencia bajo la proposición A que bajo la proposición B.

La proposición que asigna mayor probabilidad a la observación (coincidencia) es la que realiza mejor el trabajo predictivo, por tanto, resulta más apoyada por la observación. Si, por ejemplo, la relación de verosimilitudes es muy grande, la proposición A realiza un trabajo predictivo mucho mejor que B en cuanto a la ocurrencia de la observación x, por lo que la ocurrencia de esa observación constituye una muy fuerte evidencia a favor de A frente a B.

A pesar de lo descrito sobre la búsqueda histórica llevada a cabo por la humanidad de una ley que permita valorar los datos como prueba, la dificultad de su aplicación en las ciencias forenses ajenas al ADN - verificable aún en nuestros días en todo el mundo - se debe a factores claramente extracientíficos.

La dactiloscopia tiene mucho que ver con esas dificultades prácticas debido a que durante el siglo XX fue la prueba científica reina en un procedimiento penal en cualquier parte del mundo. A pesar de los avances experimentados en el desarrollo y extensión de tecnologías biométricas relacionadas con huellas dactilares y palmares en el siglo XX, no sucedió lo mismo en el aspecto interpretativo, en la valoración de los cotejos verificados como prueba.

Bajo la creencia generalizada de la capacidad de los expertos en dactiloscopia de identificar a una persona mediante el uso de cotejos dactiloscópicos y ante la abrumadora experiencia de la capacidad identificativa de las huellas dactilares y palmares entre quienes han practicado esa disciplina forense a lo largo del tiempo en todas partes, el papel de un experto en criminalística identificativa en un juicio ha sido entendido como quien presta auxilio a un Tribunal o Jurado proporcionando conclusiones sobre lo que compara en forma categórica (sí / no) o en forma inconclusa ante dificultades insoslayables a juicio del perito.

Decíamos que los factores aducidos son extracientíficos porque lo descrito son intuiciones o creencias generalizadas no fundamentadas en la metodología científica.

Los perfiles de ADN o los patrones de huellas dactilares tienen un evidente poder discriminador cuando se utilizan para distinguir individuos. Cada uno de los seres humanos es un individuo singular y posee sus propiedades de esa naturaleza específicas, pero la capacidad de distinguirlos mediante procedimientos experimentales tiene el límite que marca la metodología científica. Si no se respeta la metodología científica se pasa al dominio de la fe humana. Por eso algunos científicos han puesto de relieve la tendencia del ser humano a considerar como conocimiento cierto de alguna parte de la realidad estudiada mediante las ciencias experimentales lo que sólo puede llegar a ser conocimiento probable, por la propia limitación del método de conocimiento empleado, y han calificando esa tendencia como *un salto de fe*.

La creencia de que el perito, con su técnica y metodología, puede llegar a identificar a una persona u objeto a partir de vestigios es una creencia generalizada, incluso en ámbitos profesionales. Por eso el LR supone un cambio de mentalidad que produce intuitivamente rechazo, con un nivel de oposición proporcional al grado de convicción sobre la creencia apuntada. Sin embargo, en el campo científico del ADN, la medida del valor probatorio de los cotejos mediante el LR fue defendida por los investigadores sin especiales problemas desde la aparición de la técnica, si bien la variedad de situaciones que se presentan en la realidad ha conllevado fuertes discusiones científicas sobre la forma correcta de calcularlos en cada caso.

4. Las preguntas correctas [3, 5]

Una forma pedagógicamente brillante de entender mejor el LR es situarlo en el contexto de la valoración de la prueba de los hechos que realiza un Tribunal antes de dictar sentencia. Hay autores que han propuesto las preguntas que formulamos en este apartado como inspiradoras de lo que deben hacer los peritos, por una parte, y los miembros de un Tribunal, por otra, en el proceso mental que conduce a la declaración de hechos probados en una sentencia.

Partimos de la base de que ha habido una investigación judicial sobre la supuesta comisión de un crimen y se dispone de vestigios que pueden ser valorados como prueba. Un perito ha realizado un informe pericial sobre uno de los vestigios y como conclusión aporta un resultado de forma concluyente.

Supongamos que el informe versa sobre un cotejo entre una huella dactilar transplantada desde la escena del crimen a un soporte que permite su examen (dubitada) y una huella dactilar rodada obtenida por la policía tras la detención del sospechoso (indubitada). El resultado de la comparación realizada por el perito se traduce en la coincidencia de doce puntos característicos del mismo tipo y localizados en los mismos lugares en ambas huellas. La conclusión del informe pericial es que la huella dubitada pertenece al sospechoso.

¿Qué debe creer el Tribunal sobre la autoría de la huella dubitada una vez que el perito refleja en su informe que la ha identificado plenamente? La mayoría de los preguntados dirían, con absoluta convicción, de que la cuestión es una obviedad, de que el sospechoso es el autor de la huella encontrada en la escena del crimen tras el resultado alcanzado en el informe.

Lo que acaba de ejemplificarse es un ejemplo arquetípico de falacia de transposición del condicional en el ámbito de la valoración de una prueba científica por parte de un Tribunal. La

pregunta está bien formulada desde el punto de vista de la lógica que el Tribunal tiene que aplicar para valorar una prueba pericial. Tiene que valorar la prueba después de conocer el resultado del informe pericial. Sin embargo, el perito ha proporcionado una información que no se corresponde con su rol. La autoría de la huella es una decisión que corresponde al Tribunal a la luz de la información técnica que el perito aporta y cualquier otra información, incluyendo otras pruebas científicas, que proporcionan al Tribunal elementos de juicio a priori sobre la relación de la huella dubitada con el sospechoso. Esta argumentación podría calificarse de contraintuitiva pero está fundamentada en el Teorema de Bayes, uno de los más importantes de la teoría de la probabilidad.

Veamos cómo puede entenderse lo que acaba de decirse, para tantos quizá sorprendente. Si el sospechoso (por ejemplo, un varón) tuvo relación frecuente con la víctima (por ejemplo, una mujer) en el pasado no sería extraño que en el domicilio de la víctima pudieran encontrarse sus huellas. Sin embargo, imaginemos que la huella dubitada se recoge del cuchillo clavado en la espalda de la víctima – que yace en el dormitorio de su vivienda - y que se trata de una huella parcial de la que sólo es posible observar claramente siete puntos característicos.

Mientras que un perito de ADN tiene en cuenta la probabilidad de que otra persona, distinta al sospechoso, pudiera tener el mismo perfil genético que el hallado en una muestra dubitada y coincidente con el del sospechoso al valorar científicamente su cotejo, un experto en huellas dactilares no tiene en cuenta esta posibilidad porque la considera nula. El primero valora la prueba mediante un LR, que permite una valoración equilibrada entre las proposiciones que plantean las partes en el proceso, el segundo la valora bajo el exclusivo punto de vista de la acusación. El primero valora la probabilidad del cotejo observado condicionado a las proposiciones alternativas existentes, el segundo la probabilidad de la proposición de la acusación condicionada al resultado del cotejo.

La coincidencia entre las características comparadas entre las huellas dubitada e indubitada puede encontrarse bajo dos hipótesis alternativas: que las huellas procedan del sospechoso o que procedan de personas distintas produciendo esa misma coincidencia. Esta proposición alternativa sí se considera en el cotejo de ADN mediante el LR.

¿Por qué en el cotejo de ADN se tienen en cuenta dos proposiciones alternativas entre sí y en los cotejos de huellas dactilares no?

Propiedades semejantes entre las técnicas de cotejo comparadas:

- Los perfiles genéticos y las huellas dactilares presentan características individuales normalmente invariables con el tiempo.
- Las técnicas que permiten examinarlas son robustas y pueden aplicarse normalmente más de una vez por otros expertos sobre las mismas muestras.
- Los expertos tienen una preparación específica y los métodos de comparación están bien descritos.
- En ambas técnicas es posible realizar comparaciones con mayor o menor información identificativa. En el ámbito forense no es infrecuente que la información disponible sea parcial como consecuencia del estado del vestigio examinado.

Diferencias entre los cotejos descritos:

- Valoración estadística de la coincidencia observada: en ADN sí, en cotejos dactiloscópicos no.

Mientras que en el ámbito de las propiedades individualizadoras de los perfiles de ADN se han realizado, durante décadas, profundos estudios de naturaleza estadística para hallar las frecuencias relativas de aparición de los alelos de cada uno de los marcadores de un perfil, así como para la comprobación de la independencia probabilística entre los alelos, no ha sido así en el ámbito de la dactiloscopia (existen algunos estudios de indudable mérito pero el estado de la investigación en esta materia sigue muy lejos de lo que podría conseguirse con la tecnología y bases de datos disponibles en la actualidad).

Decimos que no es meramente una cuestión de disponer de tecnología y grandes bases de datos para la efectiva realización de los necesarios estudios estadísticos. Desde la aparición hace dos décadas de los sistemas AFIS, en los países que cuentan con esos sistemas existe tanta posibilidad de realizar estudios de frecuencias de puntos característicos y sus propiedades probabilísticas como los realizados en el ámbito del ADN en su momento.

La comunidad científica internacional ha subrayado en las dos últimas décadas una falta clara de rigor científico de los cotejos dactiloscópicos por esta razón. Sólo en los últimos diez años están publicándose estudios de naturaleza estadística en el ámbito de la dactiloscopia que permitirán evaluar las pruebas dactiloscópicas con el mismo rigor científico que con los cotejos de perfiles de ADN desde hace dos décadas.

En definitiva, la respuesta a la pregunta formulada es que mientras en el ámbito científico del ADN sus primeros investigadores vieron la importancia de realizar una valoración del resultado de un cotejo de perfiles de ADN bajo una estructura de razonamiento lógico robusta, y por eso acabaron eligiendo el LR, los dactiloscopistas pensaron siempre que ya disponían de ella. Sólo a raíz de la progresiva extensión de esa nueva estructura de razonamiento lógico en otros ámbitos de la ciencia forense, se suscitó en la comunidad científica relacionada con la dactiloscopia la preocupación por preguntarse si estaban valorando acertadamente sus cotejos.

Hoy día no existe foro científico forense serio en el mundo que no sea consciente de que está cuestionada la forma en que clásicamente se ha valorado el cotejo dactiloscópico y que se están realizando esfuerzos muy notables para asemejar la valoración de esa prueba a como se realiza en el ámbito del ADN, es decir, mediante LR.

Por tanto, ¿qué debe creer el Tribunal sobre la autoría de la huella dubitada una vez que el perito refleja en su informe el valor de la prueba mediante un LR? El Teorema de Bayes ayuda a combinar la información disponible al Tribunal sobre la cuestión antes de conocer el resultado del informe con la valoración en forma de LR aportada por el perito, así como a calcular las probabilidades de cada una de las proposiciones a la luz del resultado del informe pericial. Son estas últimas probabilidades condicionales las que tienen directamente que ver con la pregunta formulada.

El perito tiene, pues, que plantearse la siguiente pregunta: ¿qué nos dice lo observado como prueba a favor de una proposición frente a la otra? La clave está en hallar la probabilidad de lo observado condicionada a cada una de las proposiciones y relacionar las probabilidades condicionales entre sí mediante su división: eso es el LR. Al perito le está vetado proporcionar

conclusiones fuera de ese contexto porque puede demostrarse que atentaría contra las reglas de la lógica.

Y al Tribunal también le competen preguntas como la siguiente: ¿qué debe hacerse una vez que se dispone de toda la información necesaria para conocer en qué medida una huella dubitada procede de un sospechoso? La teoría de la decisión aplicable por un Tribunal en una declaración de hechos probados en una sentencia tiene suficiente margen de maniobra como para adaptarse a cada caso según los criterios de justicia aplicables. Estamos en el ámbito de los denominados *estándares de prueba*.

En resumen, podemos enunciar tres preguntas clave:

- 1.- ¿Qué nos dice lo observado como prueba a favor de una proposición frente a la otra? (eso es el LR): la debe responder un perito en las conclusiones de su informe.
- 2.- ¿Qué debe creer el Tribunal sobre la autoría de un vestigio una vez que el perito refleja en su informe el valor de la prueba mediante un LR?
- 3.- ¿Qué debe hacer un Tribunal una vez que dispone de toda la información necesaria para conocer en qué medida un vestigio procede de un sospechoso?

Las tres preguntas son acordes con el Teorema de Bayes en forma de apuestas. Eso quiere decir que existe un marco de referencia lógico robusto – se trata de un razonamiento deductivo que permite llegar a conclusiones ciertas - que sostiene la racionalidad de la propuesta.

Volvamos ahora al ejemplo del cotejo de la huella dactilar encontrada en la hoja del cuchillo clavado en la espalda de la víctima. Si cotejada la parte de la huella transplantada con la huella correspondiente del sospechoso se encontrasen siete puntos característicos coincidentes - siguiendo la metodología de cotejos dactilares vigente en la actualidad en laboratorios oficiales -, el perito muy probablemente se decantaría a favor de que el informe se emitiera con el resultado de “inconcluso”.

Normalmente utilizaría otras huellas halladas en el lugar de los hechos para realizar un informe donde pudiera afirmar que se ha conseguido una identificación plena e, incluso, no sería extraño que la huella parcial no fuera expresamente mencionada en el informe como objeto de pericia al considerarla “no apta para cotejo”. Cualquier huella que pueda claramente relacionar al imputado con la acción criminal – algo que el Tribunal tendrá que verificar en todo caso – puede llegar a constituir prueba de cargo contra él, pero el descarte de huellas relevantes no parece justificado que quede a merced, exclusivamente, del criterio del perito.

El cotejo de la huella parcial con la del sospechoso es técnicamente viable y su valoración estadística podría realizarse, al menos, de forma subjetiva. La vinculación de la huella parcial con el sospechoso y su relación con el crimen podría ser evaluada por el Tribunal si el perito informara de ambos extremos. Si al Tribunal se le proporciona un informe concluyente, desde el punto de vista identificativo, donde la huella transplantada pudiera tener una explicación inocente, podría no sólo ser inútil para resolver la causa sino perjudicial por el sesgo que esa información pudiera producir en quien tiene la misión de valorarla y tomar una decisión sobre la autoría de la huella en el contexto del caso.

5. Rendimiento de los sistemas de reconocimiento y valoración estadística de una prueba científica

La evaluación del rendimiento de un sistema biométrico necesita bases de datos perfectamente etiquetadas, es decir, cada dato utilizado en la prueba ha de estar plenamente identificado en cuanto a su origen. Si no fuera así, no podrían calcularse las tasas de error que clasificamos en dos grupos: falsos positivos y falsos negativos.

Las tasas de acierto y error calculables con bases de datos son datos que se consiguen mediante experimentación, es decir, son datos conseguidos con un control de la parte de la realidad estudiada que permita al investigador evaluar cómo de bueno es su sistema de reconocimiento con carácter general.

Sin embargo, cuando se realiza un informe pericial, el cotejo criminalístico que ha de llevarse a cabo es único y la evaluación de su resultado como prueba no es calculable experimentalmente como se ha mencionado en el párrafo anterior. En los anteriores apartados se ha dado ya suficiente información sobre cómo resolver este problema de acuerdo con la lógica aplicable.

A pesar de todo, no es infrecuente que se confunda la tasa de falsos positivos experimentalmente calculada para un sistema de reconocimiento con la probabilidad de que la muestras comparadas procedan de fuentes diferentes si el cotejo ha sido positivo. O, si se prefiere, que se confunda la tasa de falsos negativos experimentalmente calculada para un sistema de reconocimiento con la probabilidad de que la muestras comparadas procedan de la misma fuente si el cotejo ha sido negativo. Análogos razonamientos podrían realizarse respecto a las tasas de verdaderos positivos y negativos.

Las confusiones mencionadas son ejemplos clásicos de la falacia de transposición del condicional.

6. El Teorema de Bayes en forma de apuestas

En la cultura anglosajona está arraigada la costumbre de pensar haciendo apuestas. Por ejemplo, imaginemos un combate de boxeo donde hay dos púgiles, A y B, sobre el que se organizan apuestas. Podríamos apostar 2 a 1 a favor de A. Esto puede decirse, también, que *a priori* pensamos que el púgil A tiene el doble de posibilidades de ser el ganador que B. El apriorismo es algo que formulamos *antes* de que el combate empiece.

Imaginemos que no podemos asistir al combate y que nos tendremos que conformar con la información que algún testigo presencial nos comuniqué por el móvil.

Una vez comenzado el combate recibimos un primer mensaje diciéndonos que el púgil B ha caído a la lona pero que se ha logrado recuperar tras unos segundos. Pasan pocos minutos más y recibimos un segundo mensaje en el mismo sentido, si bien la recuperación del púgil B ha sido algo más lenta que en la primera caída.

Está claro que la información recibida no sólo confirma nuestra apuesta *a priori* sino que nos inclina a reforzarla. Si cayera el púgil una tercera vez, podríamos realizar una apuesta a favor de que el púgil A será el ganador mucho más ambiciosa que la inicialmente realizada. Por ejemplo, podríamos apostar 5 a 1, si bien esta apuesta sería ya una apuesta *a posteriori*

porque se nos ha transmitido información valiosa sobre la evolución del combate antes de que concluya.

Nuestro entorno cultural está más acostumbrado a pensar en términos de probabilidades que en términos de apuestas. Una apuesta puede transformarse en probabilidades y al revés. En nuestro ejemplo, apostar sobre dos opciones, A y B, es equivalente a relacionar dos probabilidades condicionales: que gane el púgil A teniendo en cuenta los méritos precedentes de carácter deportivo de los boxeadores frente a que gane el púgil B dada la misma información previa.

Si decimos que *a priori* hacemos una apuesta de 2 a 1 a favor del púgil A, decimos lo mismo que si afirmamos que el púgil A tiene una probabilidad del 66,66% de ser el ganador frente al 33,33% del púgil B (dejamos las cifras con dos decimales para simplificar el ejemplo porque, en realidad, los números tendrían infinitos decimales). La clave está en observar que la probabilidad de que gane A ha de ser doble de que gane B cuando apostamos 2 a 1 a favor de A. Con esos valores probabilísticos lo aseguramos: 66,66% es el doble que el 33,33%. Además, si sólo hay dos jugadores, A y B, no es posible ofrecer otras probabilidades que cumplan esa propiedad y que sumen entre sí el 100%.

Lo que acaba de ejemplificarse respeta el concepto de apuesta que se utiliza en el Teorema de Bayes en forma de apuestas.

Si actualizamos la apuesta después de recibir la información de los mensajes mencionados, decíamos que podíamos apostar 5 a 1 a favor de A. En ese caso, las probabilidades que podríamos emplear cumpliendo todos los requisitos mencionados serían el 83,33% y el 16,66%, que sumados con infinitos decimales llegarían al 100%. Esta apuesta sería una apuesta *a posteriori* respecto a la primera porque se ha actualizado con información recibida mediante mensajes sobre la evolución del combate. Obsérvese que si multiplicamos 16,66% con infinitos decimales por 5 obtenemos la cifra 83,33% con infinitos decimales. No existe otra combinación de números enteros decimales que teniendo entre ellos una proporción 5 a 1 en magnitud sumen 100.

Por otro lado, viene bien advertir aquí que si se compara la apuesta *a posteriori* con la *a priori* mediante la división entre ambas, obtenemos $5/2 = 2.5$, cifra interesante porque nos informa sobre la influencia que las tres caídas a la lona por parte del púgil B ha tenido sobre la persona que realiza la apuesta. Así actúa el LR y de hecho, esa magnitud es un LR.

La relación de verosimilitudes que hemos cifrado en 2.5 puede enunciarse, verbalmente, del siguiente modo: observar tres caídas consecutivas del púgil B respalda 2.5 veces más que sea cierta la proposición de que el púgil A gana el combate que lo sea la alternativa, es decir, que B sea el ganador. O también, que es 2.5 veces más probable observar que el púgil B caiga en la lona tres veces consecutivas si la proposición de que el púgil A es el ganador fuera cierta que si lo fuera la proposición alternativa.

Existe una sencilla regla nemotécnica que permite transformar apuestas en probabilidades. Lógicamente la transformación está fundamentada en la aplicación de las leyes de la probabilidad y de la aritmética, pero evitamos al lector esas demostraciones que son, por otra parte, bastante simples.

La regla nemotécnica es la siguiente: una apuesta entre dos opciones, por ejemplo, de 5 a 1, entendida como una relación entre esos números en forma de una división, puede transformarse en dos probabilidades condicionales:

- a) la del numerador de la apuesta dividiendo el primer término de la apuesta por la suma de los dos, es decir, 5 entre 6 ó, si se prefiere, $5/6$, lo que proporciona una cifra del 83,33%;
- b) la del denominador de la apuesta restando la cifra anterior del 100%, lo que se traduce en restar de $6/6$, $5/6$, que da, como resultado, $1/6$, o sea, el 16,66%.

Si la apuesta fuera justamente la inversa, es decir, de 1 a 5, se intercambiarían de posición las reglas enunciadas siendo el numerador un 16,66% y el denominador el 83,33%.

¿Puede funcionar el Teorema de Bayes con apuestas cuyos valores no estén referidos al 1, como en los ejemplos anteriores, es decir, del tipo 3 a 2 o cualquiera que fuera el par numérico? Se interpretaría como que apostamos 3 a 2 a favor de que el púgil A gane al B. La respuesta es afirmativa. La relación de 3 a 2 entre dos probabilidades que sumadas entre sí han de sumar el 100% se transformaría en el 60% y el 40%, respectivamente. La regla nemotécnica también funciona en este contexto: $3/5$ y $2/5$, respectivamente.

La situación cambiaría si en lugar de dos opciones posibles (en el ejemplo hay dos púgiles en combate) hubiera una multiplicidad de opciones que se pudieran elegir (por ejemplo, si tuviéramos que apostar en una carrera de caballos de hasta 10 concursantes). En situaciones como esa resultaría en nuestro entorno cultural mucho más complejo el empleo numérico de apuestas que de probabilidades. Tendríamos muchas posibilidades de realizar apuestas no respetuosas con la teoría de la probabilidad y, por consiguiente, con riesgo evidente de jugar a favor de un caballo perdedor.

Afortunadamente, en un entorno forense suelen existir dos proposiciones (opciones) mutuamente excluyentes (no pueden ser ciertas las dos a la vez) y normalmente exhaustivas (cubren todas las posibilidades de explicación sobre lo que estamos interesados), la de la acusación y la de la defensa, por lo que el empleo de las apuestas en relación con la unidad facilita enormemente la transformación en probabilidades, como se ha podido ver en los ejemplos.

La transformación de probabilidades en apuestas resulta más sencilla. En un contexto de dos opciones mutuamente excluyentes y exhaustivas si conocemos una de las probabilidades condicionales, la segunda se calcula restándola del 100%. La apuesta se construye dividiendo una por otra. Por ejemplo, si la probabilidad a priori de que gane el combate el púgil A teniendo en cuenta los méritos precedentes de carácter deportivo de ambos boxeadores es del 60%, la de B será el 40%. La apuesta es de 6 a 4 a favor de A, o reduciéndola numéricamente hasta encontrar los números enteros más pequeños posibles, de 3 a 2 a favor de A.

Esta explicación carecería de interés especial si no fuera porque el LR figura insertado en la expresión matemática del Teorema de Bayes en forma de apuestas.

Las preguntas que interesan al Tribunal referidas en el anterior apartado se responden con la ayuda del Teorema de Bayes, por lo que los conceptos de apuesta a priori y apuesta a

posteriori se necesitan comprender y se han de utilizar. El Teorema se enuncia diciendo que *la apuesta a posteriori es igual al LR multiplicado por la apuesta a priori*.

En términos que faciliten su comprensión, puede enunciarse diciendo que la apuesta a priori que el Tribunal pueda hacer antes de que el perito aporte su información con un LR sobre en qué medida, dada la información disponible del caso, es más probable la proposición que defiende la acusación sobre la que sostiene la defensa, se multiplica por el LR – forma en que el informe pericial valora la prueba científica realizada – para obtener la apuesta a posteriori. Esta última apuesta responde a la pregunta:

¿Qué debe creer el Tribunal sobre la autoría de un vestigio una vez que el perito refleja en su informe el valor de la prueba mediante un LR?

La creencia de la que se habla se refiere a la probabilidad de cada proposición condicionada al resultado de la pericia.

Concluimos este apartado resaltando que el concepto de LR como interpretación de lo observado como prueba es coherente con la teoría de la probabilidad, como ha quedado de manifiesto al explicar el Teorema de Bayes en forma de apuestas, lo que le dota de especial valor científico como solución a ese problema. Puede afirmarse, sin temor alguno al error, que si una observación constituye una prueba que favorece a una proposición A frente a otra B utilizando el LR, si las proposiciones tuvieran probabilidades previas de ocurrencia, el efecto de esa observación nunca disminuirá la probabilidad de A respecto a B.

7. La ley de verosimilitud conduce hacia la verdad [3]

Nos preguntamos en este apartado si es posible que los LR puedan favorecer la proposición equivocada. La respuesta es afirmativa, sin que esto signifique que sea contradictoria con lo descrito al término del apartado anterior sobre la coherencia del LR con la teoría de la probabilidad.

Una observación correctamente interpretada conforme a la ley de verosimilitud puede favorecer la proposición falsa frente a la verdadera. La razón de que eso pueda ocurrir está relacionada con la limitación que la metodología científica aplicable al problema real al que se enfrenta tiene por sí misma. Todo sistema real de reconocimiento de patrones tiene tasas de error de falsos positivos y falsos negativos.

Si el patrón con el que comparamos lo observado está perfectamente definido en una biblioteca de patrones como sucede, por ejemplo, con muchos de los análisis que se realizan en la química forense (detección de acelerantes de la combustión, determinación de la naturaleza de sustancias estupefacientes, etc.), el interés del análisis no se centra en la determinación de la fuente original de donde procede la sustancia sino en su naturaleza específica como tal sustancia, es decir, en la clase o tipo de sustancia en la que se clasifica lo observado. En estos casos las tasas de error están principalmente relacionadas con la sensibilidad de la técnica para detectarlas correctamente.

Sin embargo, si el patrón observado procedente de una fuente desconocida tiene una variabilidad intrínseca por su propia naturaleza (por ejemplo, la existente en la voz), a la hora de discriminar entre fuentes por cualquier sistema de reconocimiento nos encontraremos con tasas de error de falsos positivos y negativos donde influye no sólo el problema de la

sensibilidad referido en el párrafo precedente sino el problema de la siempre limitada capacidad discriminativa entre patrones del sistema utilizado.

Los LR que favorezcan la proposición equivocada se denominan “engañosos”. Si se modelan matemáticamente las distribuciones de los datos generados por el sistema de reconocimiento cuando se comparan patrones procedentes de la misma y de distintas fuentes en todas sus formas, es posible calcular probabilidades de obtener LR “engañosos”.

Por tanto, un único LR, por sí mismo, y aunque esté perfectamente calculado, no garantiza que conduzca a la verdad. Puede demostrarse matemáticamente que existe un límite universal de probabilidad máxima de LR “engañoso” por cada valor de LR calculado por un sistema de reconocimiento que lo haga correctamente. Coincide con el inverso del valor de LR (por ejemplo, un LR igual a 10 tiene una probabilidad máxima de ser “engañoso” del 10%, y un LR = 100, del 1%). A mayor valor numérico de LR, menor probabilidad máxima de LR “engañoso”. Igualmente, puede demostrarse que a medida que se realizan más observaciones, independientes entre sí, del mismo fenómeno y en idénticas circunstancias, lo que implica multiplicar los LR de cada una de ellas y obtener LR globales cada vez más altos numéricamente, se consigue estar cada vez más cerca de la verdad, es decir, que el LR global favorezca la proposición verdadera.

Esta última propiedad del LR es extremadamente interesante porque garantiza la utilidad práctica de los LR para resolver los problemas en los que se aplica.

8. Hacia la comprensión correcta de los LR de ADN

8.1 Distinguiendo probabilidades de interés [6, 7]

La capacidad discriminativa de un perfil de ADN (utilizamos la notación Υ para denominarlo) con los kits comerciales disponibles es enorme. La probabilidad de que hallemos un perfil genético determinado con 24 marcadores del kit *GlobalFiler*¹ puede llevarnos a cifras, teóricas, de 10^{-30} y aún menores. Son cifras teóricas porque no son verificables. Están fundamentadas en modelos estadísticos robustos pero también en frecuencias alélicas experimentales obtenidas en las tres las últimas décadas y en la asunción de uniformidad que siempre se necesita para realizar inferencias, es decir, en la asunción de que esas frecuencias se hubieran obtenido en el pasado y seguirán obteniéndose en el futuro sin especiales variaciones.

Un primer concepto básico e importante que se necesita distinguir es la frecuencia de un determinado perfil genético en una población. Si se considera la población como un conjunto de individuos genéticamente independientes, la frecuencia del perfil se calcula multiplicando las frecuencias de cada uno de los alelos. Si la población está determinada por un cierto grado de parentesco, las frecuencias alélicas serán mayores que las anteriores, más altas a medida que sea más próximo el grado de parentesco.

Analicemos el caso más simple posible en el que disponemos de un perfil genético obtenido de una muestra recogida en una inspección ocular - asumimos que se ha producido un crimen y que la muestra procede del criminal - y observamos que coincide con el perfil genético de un sospechoso. En este caso decimos que se ha producido un *match*. Llamamos probabilidad de

¹ Aunque el kit GlobalFiler tiene 24 marcadores, sólo se usan 21 en la valoración estadística. No se tienen en cuenta los siguientes: amelogenina (indicador de género), YInDel y DYS391 (ambos localizados en el cromosoma Y, sólo presente en varones).

coincidencia aleatoria (*random match probability*) a la probabilidad de observar el mismo perfil genético del caso entre los miembros de la población de la que procede el criminal. Si asumimos que el sospechoso procede de esa población, la probabilidad de observar que su perfil coincida con el del caso suponiendo que no fuera el criminal es la misma que la probabilidad de que cualquier individuo de la población de la que procede el criminal, escogido al azar, pudiera tener el mismo perfil que el del caso, es decir, la probabilidad de una coincidencia aleatoria.

Pero si queremos valorar la coincidencia entre perfiles (el *match*) como prueba, nos conviene conocer en qué medida, una persona distinta al sospechoso, perteneciente a la población de referencia de la que suponemos que procede el criminal y biológicamente independiente del sospechoso, pudiera tener el mismo perfil que el hallado en la muestra cuestionada. Esta probabilidad – netamente diferente a las anteriormente referidas y no fácilmente distinguible en el lenguaje ordinario - se calcula de la siguiente forma:

Llamemos γ a la probabilidad del perfil Υ hallado en la muestra dubitada en la población de referencia acorde con el caso. La probabilidad de que un individuo de la referida población, escogido al azar, tenga el perfil Υ buscado es también γ . La probabilidad de que un individuo de esa población no tenga el perfil Υ es $1 - \gamma$, la probabilidad complementaria. Si la población de donde puede proceder el criminal tiene N habitantes genéticamente independientes, la probabilidad de que ningún individuo de esa población tenga el perfil Υ buscado es $(1 - \gamma)^N$. Se ha utilizado una conocida propiedad de la teoría de la probabilidad que afirma que la probabilidad de una intersección de sucesos probabilísticamente independientes se calcula multiplicando cada una de las probabilidades de esos sucesos entre sí. Cada habitante de la población de donde puede provenir el criminal es un suceso y la probabilidad de que no tenga el perfil Υ es $(1 - \gamma)$. Como hay N habitantes, se llega finalmente a esta expresión: $(1 - \gamma)^N$.

En el ejemplo se ha hablado de dos poblaciones: la población de referencia y la relacionada con el lugar de origen del criminal. Son netamente distintas, la primera se denomina “de referencia” porque hace alusión al cálculo estadístico de las frecuencias alélicas acorde al caso; la segunda hace alusión al lugar habitual de residencia del criminal.

Pues bien, la probabilidad de que, al menos, haya un habitante de la población de donde puede proceder el criminal que tenga el mismo perfil genético que el hallado en la muestra dubitada es $1 - (1 - \gamma)^N$.

Fijémonos en el siguiente ejemplo:

Se ha producido un crimen en Coín (Málaga) en enero de 2003. Por las circunstancias del caso se conoce que el criminal reside en España, es varón y más bien joven. Se consulta la base de datos del Instituto Nacional de Estadística que proporciona la siguiente información sobre la población de varones residentes en Coín (Málaga) en esas fechas: 3.370 varones entre 15 y 39 años. Se ha extraído un perfil genético de una muestra recogida en el lugar de los hechos que se considera que se relaciona con el crimen.

La frecuencia del perfil hallado, utilizando un kit de marcadores SGM+, es 10^{-9} . Para hallar la frecuencia del perfil se han utilizado frecuencias alélicas de la población de referencia publicadas en revistas científicas y, por tanto, sometidas a revisión entre iguales.

Se quiere hallar la probabilidad de que, al menos, haya un habitante de la población de donde puede proceder el criminal – en un primer momento la centramos en Coín - que tenga el mismo perfil genético que el hallado en la muestra dubitada: $1 - (1 - \gamma)^N = 1 - (1 - 10^{-9})^{3.370} = 3,37 \times 10^{-6}$. Esta cifra considera a todos los habitantes varones de Coín de esa fecha independientes genéticamente, lo cual es evidente que no es cierto. Con la aparición del kit SGM+, Foreman y Evett² propusieron limitar las probabilidades de coincidencia aleatoria aplicables a un caso concreto. Aplicando proporcionalmente su argumentación a la población de Coín (reducción de 6 a 1 en el ámbito de la población mundial que hacemos extrapolable a la Coín), la cifra de habitantes varones en la mencionada franja de edad genéticamente independientes se estima que es de 561. Utilizando de nuevo la fórmula: $1 - (1 - \gamma)^N = 1 - (1 - 10^{-9})^{561} = 5,6 \times 10^{-7}$. Esta cifra es seis veces menor, aproximadamente, que la anterior. No es nada intuitivo descubrirlo cuando se manejan cifras de este tenor y se comparan entre sí. Sin embargo, si la comparativa entre habitantes se realiza directamente, sí se percibe fácilmente la comparación de magnitudes: de 561 a 3.370. Como puede verse, la relación de proporcionalidad se guarda en ambos casos, lo que simplifica la interpretación de los datos.

Si alguien objetara que reducir la procedencia del criminal a la ciudad de Coín sería una decisión más que discutible puesto que resulta probable que el criminal proceda de otras localidades de la provincia de Málaga, quizá fuera más sensato tomar como población de procedencia del criminal la totalidad de la provincia de Málaga con las mismas características. En esas fechas asciende a 241.134 habitantes varones entre 15 y 39 años.

Aplicando de nuevo los mismos razonamientos, incluido el de la proporcionalidad de consanguinidad, y cálculos pertinentes, podemos argumentar lo siguiente:

La probabilidad de que, al menos, haya un habitante de la población de donde puede proceder el criminal – nos referimos ahora a la provincia de Málaga - que tenga el mismo perfil genético que el hallado en la muestra dubitada: $1 - (1 - \gamma)^N = 1 - (1 - 10^{-9})^{40.189} = 4,01 \times 10^{-5}$.

² Foreman y Evett, en “*Statistical analysis to support forensic interpretation of a new ten-locus STR profiling system*”, publicado en la revista *International Journal of Legal Medicine*, en el año 2001, publicado tras la aparición del kit SGM+ con 10 marcadores, propusieron que las probabilidades de coincidencia aleatoria entre perfiles de ADN en casos concretos no debieran específicamente calcularse por principio. En su lugar, propusieron cifras límite de acuerdo con las diferentes casuísticas: diferentes grados de parentescos, pertenencia a la misma subpoblación o en el supuesto más genérico. Cuando se llegaron a esas cifras de probabilidades de coincidencia aleatoria tan bajas (por ejemplo, 10^{-10} o inferiores) se plantearon si era posible verificar experimentalmente la corrección de las mismas de acuerdo con las teorías de inferencia estadística utilizadas. El problema que advirtieron es que no era posible realizar esas pruebas adicionales de verificación si se tiene en cuenta que la población real de la Tierra era – entonces - de 6.000.000.000 de habitantes y la existencia de parentesco – en sus diversos grados en la población de la Tierra - era más que evidente. Garantizar que hubiera 1.000.000.000 de pobladores genéticamente independientes era un límite que pensaban que no se podía traspasar para preservar lo que la teoría estadística utilizada asumía como cierto en los cálculos obtenidos: que es posible encontrar ese número de personas en la Tierra genéticamente independientes – lógicamente se tiene en cuenta la posibilidad de seleccionarlas aleatoriamente entre todos los que la habitan garantizando que se cumpla siempre esa propiedad -. Por eso establecieron como límite máximo de probabilidad de coincidencia aleatoria la expresión “one in a billion”, es decir, 10^{-9} . Cuando apareció el kit de 15 marcadores no cambiaron su parecer. Los cálculos estadísticos no los podían en duda si lo que se asumía en ellos se consideraba cierto, pero no pensaban que estuviera justificado aplicarlos a la población real de la Tierra. Lo que esos cálculos estadísticos asumían no era posible verificarlo con experimentos reales a posteriori. Aunque la argumentación de ambos científicos sigue siendo sustancialmente válida, otros científicos - reconocidos expertos en ADN - han recomendado que se proporcionen las cifras reales calculadas advirtiendo explícitamente que no pueden ser verificadas experimentalmente (por ejemplo, en la obra *Forensic DNA Evidence Interpretation*, capítulo 3, cuyo autor es John Buckleton, editado en el año 2005 por CRC Press).

E incluso, si alguno dijera que la reducción a la provincia de Málaga pudiera también ser sesgada y que sería más razonable extender la procedencia a toda Andalucía, en esas fechas la población de la comunidad andaluza era de 1.353.114 varones entre 15 y 39 años.

Aplicando de nuevo los mismos razonamientos, incluido el de la proporcionalidad de consanguinidad, y cálculos pertinentes:

La probabilidad de que, al menos, haya un habitante de la población de donde puede proceder el criminal – nos referimos ahora a toda Andalucía - que tenga el mismo perfil genético que el hallado en la muestra dubitada: $1 - (1 - \gamma)^N = 1 - (1 - 10^{-9})^{225.519} = 2,25 \times 10^{-4}$.

Y si extrapolamos la población a toda España en las mismas circunstancias: $1 - (1 - \gamma)^N = 1 - (1 - 10^{-9})^{1.261.762} = 0.00126$, es decir, aproximadamente una persona por cada mil habitantes no relacionados entre sí biológicamente.

Dependiendo de qué población estemos considerando de la que proceda el criminal, es decir, de cuál es su posible lugar de residencia habitual, habiéndose obtenido un *match* entre el perfil hallado en la muestra cuestionada y el perfil indubitado de un sospechoso procedente de la misma población, la probabilidad de que otra persona ajena a él, de la misma población de referencia e independiente genéticamente de él, pueda producir otro *match* varía, en el ejemplo, de $5,6 \times 10^{-7}$ a 0.00126. La relación entre estas últimas cifras es de 2.250, es decir, la variedad de hipótesis de procedencia poblacional del criminal puede hacer que la probabilidad buscada varíe hasta alcanzar un número 2.250 veces mayor que el más bajo posible.

¿Qué relación tienen estas cifras con la probabilidad del perfil de 10^{-9} ? Cuando se confunde la probabilidad del perfil con la probabilidad buscada en este ejemplo, el error numérico que se comete puede llegar a ser de enorme magnitud. Ahora quizá quede más claro que si nos proporcionan la frecuencia del perfil y pensamos que nos informan sobre la probabilidad de que otra persona distinta al sospechoso pueda producir un *match* con el de la muestra dubitada cometemos un error lamentable. En ese modo de pensar discurre el repetido razonamiento intuitivo y falaz, cuando los peritos hablan de las probabilidades de coincidencia aleatoria de perfiles genéticos en informes de ADN, de que se necesitan varios mundos como el nuestro con la misma población que tenemos actualmente para encontrar aleatoriamente a otra persona que produzca un *match* con el perfil buscado y coincidente con el del sospechoso.

Conviene subrayar que la mayoría de los laboratorios forenses optan por utilizar probabilidades de coincidencia aleatoria sin limitación³ y sin advertir lo que, por ejemplo, aconseja Buckleton (véase pie de página número 2).

8.2 Interpretación de LR con perfiles de ADN nuclear

Nos centramos en este apartado en el supuesto más sencillo posible porque existen múltiples soluciones de cálculo de LR en función de los distintos grados de parentesco. Por tanto, la

³ Incluso muchos laboratorios emiten conclusiones con valores de probabilidad de coincidencia aleatoria muy pequeñas con cifras decimales. El tribunal puede interpretar entonces que la medida es muy exacta, lo cual no se corresponde con la realidad. Para estimar la frecuencia de un alelo en una población se selecciona a un número muy limitado de individuos y se suponen una serie de condiciones como ciertas (por ejemplo, que no existe migración, que los emparejamientos se producen al azar, etc.). Por tanto, puede afirmarse que la estimación que se realiza no es la frecuencia real del alelo en cuestión sino una aproximación.

fórmula en la que vamos a fijar nuestra mirada es una sencilla división aritmética entre la unidad y la probabilidad de coincidencia aleatoria. La unidad del numerador se justificó en el apartado 3 y la probabilidad mencionada es la de observar un *match* bajo la tesis que suelen defender las defensas por defecto.

Al ser la probabilidad de coincidencia aleatoria un número muy pequeño y estar en el denominador de la fórmula matemática del LR, pasa a ser numéricamente el valor del LR.

En ADN es muy frecuente encontrarse con valores de LR del orden de 10^{20} , y con los últimos kits, aún mayores. Son números realmente enormes. Deben interpretarse de la siguiente forma:

El *match* es 10^{20} veces más probable observarlo si la proposición de que los perfiles comparados y coincidentes proceden del sospechoso fuera la proposición correcta que si lo fuera la proposición alternativa.

Lo que es ese número tan elevado de veces más probable es el *match*, pero valorándolo de forma relativa, es decir, frente a dos proposiciones distintas entre sí, mutuamente excluyentes.

Al ser un LR una relación entre verosimilitudes que, en algunos casos, pueden ser probabilidades condicionales discretas, es decir, con valores posibles entre 0 y 1, el rango de valores de un LR oscila entre 0 e infinito. Cuando las verosimilitudes sean densidades (con variables continuas), éstas pueden tener valores numéricos entre 0 e infinito y pudieran producirse indeterminaciones en los resultados como ocurre cuando se produjera una división de cero entre cero, o infinito entre infinito. En los demás casos la división también se circunscribe al rango entre 0 e infinito.

El LR puede ser visto también, dentro del Teorema de Bayes en forma de apuestas descrito en el apartado 6, como el resultado de la división entre la apuesta a posteriori y la apuesta a priori. El valor numérico que representa las veces en que se incrementa o decrementa la apuesta a priori para dar lugar a la apuesta a posteriori es precisamente el LR.

Las apuestas se relacionan con probabilidades condicionales de proposiciones condicionadas a lo observado (por ejemplo, un *match* entre perfiles de ADN). Ya se ha indicado en el apartado 6 cómo deben interpretarse las apuestas, ilustrándolo con ejemplos sencillos, pero vale la pena recordar aquí que se relacionan directamente con los verdaderos intereses de un Tribunal o Jurado respecto a la valoración de la prueba científica dentro del procedimiento penal en curso.

Por consiguiente, para un Tribunal o Jurado el valor de LR es un medio que le permite centrar su atención sobre las probabilidades condicionales de las proposiciones que le interesan dentro de la causa después de obtenerse el valor que cuantifica la fuerza de la prueba científica: la magnitud del LR. Son las denominadas probabilidades a posteriori sobre las proposiciones. Probabilidades que relativizándolas entre sí, dividiéndolas, configuran la apuesta a posteriori.

Existen formas distintas de expresar un valor numérico de LR, además de la anteriormente expresada. Por ejemplo, podría decirse lo siguiente: "Lo observado apoya o respalda 10^{20} veces más la proposición A que la proposición B". No resulta superfluo enfatizar que el respaldo o apoyo siempre se realiza en un entorno relativo, es decir, de una proposición frente a otra,

nunca respecto a una única proposición. Por tanto no es lo mismo decir “Lo observado apoya o respalda 10^{20} veces más la proposición A que la proposición B” que decir “Lo observado apoya o respalda 10^{20} veces más la proposición A”. Como quiera que expresemos, verbalmente, un LR, se precisa valorar la prueba siempre de forma relativa. La interpretación de los datos como prueba es esencialmente relativa según la ley de verosimilitud.

8.3 ¿Cómo interpretar el número del LR como fuerza de la prueba? [3]

Hemos señalado que la magnitud del LR expresa la fuerza probatoria de lo observado frente a dos proposiciones mutuamente excluyentes. Hay algunos números de LR que nos permiten realizar una interpretación directa de su magnitud:

- a) Un LR igual a 1 significa que lo observado tiene un valor de prueba que deja a quién recibe esa información en el estado mental de duda escéptica respecto a esa específica información. La prueba científica no apoya más una proposición que la otra dejando a quien recibe su resultado igual que estaba antes de que se realizara.
- b) Un LR igual a infinito significa que lo observado apoya de tal manera una proposición respecto a su alternativa que nos encontramos ante una situación determinística. Como ejemplo podemos citar la clasificación de una sustancia como gasolina.
- c) Un LR igual a cero significa que lo observado apoya la proposición alternativa frente a la primera de tal forma que nos encontramos, de nuevo, ante una situación determinística, en sentido contrario a la expuesta en b). Como ejemplos podrían citarse los cotejos negativos en huellas dactilares o con perfiles de ADN.
- d) Un LR con un valor diferente a los anteriores apoya una proposición frente a la alternativa de forma gradual. Algunos investigadores han propuesto utilizar una escala verbal que exprese el grado de fuerza de la prueba acorde con la magnitud numérica del LR. Existen diversas propuestas, cada una de ellas con ciertas propiedades que pudieran hacerlas más ventajosas o atractivas.
- e) A mayor magnitud del LR, tanto hacia cero como hacia infinito, la fuerza de la prueba a favor de la proposición correspondiente frente a su alternativa es mayor. Y viceversa.

Los apartados anteriores recogen, esencialmente, todo lo que se puede decir sobre el significado de la magnitud del LR. Sin embargo, ¿existe alguna forma de ilustrar la fuerza de la prueba mediante algún ejemplo que nos permitiera valorarla de forma intuitiva? Todos podemos tener una idea intuitiva de lo que significa que la probabilidad de observar algo sea doble si la proposición A es cierta frente a que lo sea una proposición B. Por ejemplo, que la probabilidad de observar un autobús en nuestra parada es doble si es de la línea A que si lo es de la B. Igualmente podríamos captar el significado de la fuerza de la prueba con valores comparativos mayores, pero ¿hasta cuánto? Y no sólo eso, ¿cómo diferenciamos un 1002 de un 1033, por poner un ejemplo? A mayor magnitud, la cercanía de los números entre sí nos impide percibir sus diferencias comparativamente.

Y si los números tienen valores como 10^{20} , es que no podemos ni percibir intuitivamente la entidad de su magnitud salvo el mero hecho de saber que son números enormes. Por ejemplo, Plutón está a más de 7.500.000.000 de km de la Tierra, algo similar, en número, a la población actual de nuestro planeta. Si expresamos la cifra en milésimas de milímetro, tendremos 7.500.000.000.000.000.000, o sea, 7.5×10^{18} . Con ello podemos percibir que con los actuales sistemas de análisis de perfiles de ADN somos teóricamente capaces de distinguir ese número de perfiles diferentes o incluso mayores, como ya se ha indicado. El LR en ADN se calcula

dividiendo la unidad⁴ por la probabilidad de un perfil en una población, por tanto, si fuera esta última de esa magnitud: $1 / 7.5 \times 10^{18}$, el LR sería igual a 7.5×10^{18} .

Por todo ello conviene disponer de ejemplos que nos puedan servir para hacernos una idea del significado de las magnitudes. Uno de los más felices es el relacionado con un experimento sencillo consistente en sacar a ciegas una bola de dos posibles urnas, una llena de bolas blancas (urna B de sólo bolas blancas) y otra, de igual tamaño y forma, con la mitad de sus bolas blancas y la otra mitad negras (urna N que indica que en ella hay bolas negras, además de blancas). Se introduce la mano en la urna que el organizador del juego nos ponga al alcance y sólo podemos conocer el resultado, es decir, si la bola es blanca o negra. Acto seguido, se introduce la bola en la urna y vuelta a empezar.

En el contexto del juego referido, es fácil ver que la probabilidad de sacar una bola blanca de la urna B es 1, y que la probabilidad de sacar una bola blanca de la urna N es 0.5.

Imaginemos que sacamos la primera bola de la urna – recordemos que no sabemos qué urna es porque lo hacemos a ciegas - y resulta ser blanca. Queremos conocer qué valor de LR tiene el hecho de observar una bola blanca. La probabilidad de observarla si fuera cierta la proposición “se ha sacado la bola de la urna B” frente a la proposición “se ha sacado la bola de la urna N”, es $1 / 0.5 = 2$. No es difícil entenderlo. Si la bola se hubiera extraído de la urna en que todas las bolas son blancas, no podía ser de otro color, por lo que la probabilidad es igual a 1 (suceso seguro). Si la bola se hubiera extraído de la otra urna, donde la mitad son blancas y la otra mitad negras, la probabilidad de observar una bola blanca es claramente 0.5. El LR divide entre sí esas probabilidades, obteniéndose el resultado indicado: $LR = 2$. Su significado es el siguiente: observar una bola blanca es dos veces más probable si es extraída de la urna B que si lo es de la urna N.

Imaginemos ahora que sacamos una segunda bola y el resultado es que es nuevamente blanca. Como suponemos que las bolas estaban aleatoriamente ubicadas en las urnas y ambas tenían el mismo número de bolas que al principio, pues tras extraer la primera bola la hemos vuelto a introducir en la urna de donde la sacamos, la extracción de la segunda bola podemos considerarla probabilísticamente independiente de la primera extracción. Aplicando las leyes de la teoría de la probabilidad, la probabilidad de sacar dos bolas blancas consecutivamente es igual a multiplicar la probabilidad de sacar cada una de ellas por separado. Por tanto, si esta segunda bola hubiera sido extraída de la urna B, y la primera también, la probabilidad de sacar las dos bolas blancas de la urna B es $1 \times 1 = 1$. Pero si se hubieran sacado las dos de la urna N, es igual a $0.5 \times 0.5 = 0.25$. Y si se hubieran sacado cruzadas, es decir, una de cada urna, es igual a 1×0.5 ó 0.5×1 , o sea, 0.5.

El LR asociado a sacar dos bolas blancas sería igual a multiplicar los LR asociados a cada extracción si se considerasen las extracciones probabilísticamente independientes, por lo que tendríamos los siguientes valores: la probabilidad de observar dos bolas extraídas consecutivamente si “se han sacado de la urna B” frente a “se han sacado de la urna N” es igual a $1 / 0.25 = 4$. Como vemos, el LR ahora es el doble que el calculado sacando una bola blanca bajo las mismas proposiciones.

⁴ El numerador del LR es igual a 1 sólo cuando se cumple: (i) que el caso en cuestión sea un caso de coincidencia de perfiles (lo que llamamos “match” directo) y no un caso de parentesco o una mezcla; (ii) cuando en la muestra dubitada no haya pérdida de alelos (en inglés se utiliza la expresión “drop-out”), ni alelos extra (en inglés, “drop-in”) y (iii) cuando suponemos que la tasa de error en el análisis sea igual a cero.

Por analogía, podemos intuir que el LR es igual a 2^E (utilizamos la notación E de extracción), donde E expresa el número de bolas blancas consecutivamente extraídas de la urna. En el supuesto de que hubiéramos hecho tres extracciones en las mismas condiciones que la referida, obtendríamos un LR = 8.

Luego, el valor del LR del ejemplo va creciendo según la ley 2^E a medida que observemos E bolas blancas consecutivas. Es decir, observar tres bolas blancas consecutivas en tres extracciones independientes de una misma urna, tiene un valor de prueba que puede expresarse diciendo que es ocho veces más probable observarlo si la urna B es de donde se están extrayendo las bolas que si se están extrayendo de la urna N.

Pensemos en el supuesto de que hubiéramos extraído 6 bolas blancas consecutivas en las mismas condiciones que en el ejemplo. Si nos preguntasen qué significado tiene para nosotros esa observación con respecto a las dos posibilidades conocidas: que las bolas procedan de la urna B o de la urna N, diríamos que esa información inclina significativamente la balanza de nuestra respuesta a favor de la proposición de que “se han extraído de la urna B” frente a que “se han extraído de la urna N”. El LR asociado a las 6 bolas blancas consecutivamente extraídas sería de $2^E = 2^6 = 64$.

Si quisiéramos utilizar este ejemplo como referencia intuitiva, tendríamos que tener un modo de transformar el valor de LR en número de bolas consecutivamente extraídas de la urna B. Puede hacerse aplicando logaritmos en base 2 (porque hay dos urnas, es decir, dos proposiciones) a los dos términos de la igualdad siguiente: $2^E = LR$. Consideremos que el LR = 64.

Primer término de la igualdad: $\log_2(2^E) = E \log_2(2) = E$;

Segundo término de la igualdad: $\log_2(LR = 64) = 6$ (por definición, el logaritmo de un número en una base es el número al que hay que elevar la base utilizada en el logaritmo – en este caso 2 – para obtener el valor del número (en este caso, un LR));

Resultado E = 6 (seis bolas blancas consecutivamente extraídas de la urna B).

Si el LR fuera, por ejemplo, igual a 1500, el número de bolas blancas consecutivas de la urna B que se correspondería con ese valor probatorio expresado mediante el LR sería algo mayor que 10.5. Basta calcular el $\log_2(1500)$ con una calculadora. Esa es la regla de transformación para un caso de dos urnas como el del ejemplo.

Cualquier valor de LR que obtuviéramos en un cotejo criminalístico podría compararse con el mismo número en el contexto del ejemplo de las urnas para hacernos una idea intuitivamente fácil de captar su fuerza como prueba.

9. Hacia una mejor comprensión de la valoración estadística de un cotejo de perfiles de ADN en entorno forense

9.1 El problema de la isla [8]

Expertos en valoración estadística de la prueba de cotejo de perfiles de ADN han hecho célebre un ejemplo que en la literatura especializada se conoce como “el problema de la isla”,

en el que se intenta encontrar una solución matemática rigurosa a un problema que se plantea en un contexto idealizado pero que arroja luz sobre un contexto análogo real.

Eggleston en 1983 planteó, por vez primera, el problema de un crimen imaginario en una isla imaginaria donde la vida y el crimen son mucho más simples que en el mundo real. En lo que sigue nos apoyamos en algunas de las denominadas *lecciones* que sobre el ejemplo imparte Balding en su obra titulada “*Weight-of-Evidence for Forensic DNA Profiles*”:

Consideremos una cualidad muy singular llamada Y . No es visible a simple vista y necesitamos realizar un test para hallarla. Asumimos, por el momento, que el test no falla. Se comete un crimen en una isla con una población de 101 habitantes. Al principio no hay pistas y todos los habitantes de la isla son considerados por la policía equiprobablemente sospechosos. Se asume que un habitante de la isla es el criminal. En un momento dado se averigua que el criminal ha de tener la cualidad Y , y se halla a un sospechoso – habitante de la isla - que la tiene. ¿De qué modo podemos estar convencidos de que hemos hallado al culpable?

La respuesta a esta pregunta depende, entre otros factores, de la rareza de la cualidad Y . Supongamos que el sospechoso y el criminal (que pueden ser o no la misma persona) son las únicas personas en la isla de las que se conoce que tengan o no la cualidad Y . Supongamos, también, que en un reciente estudio llevado a cabo en el continente más cercano a la isla, se indica que el 1% de la población tiene la cualidad Y y asumimos que esa tasa se mantiene, en media, en las islas más cercanas. También asumimos que la probabilidad de que cualquier habitante de la isla tenga la cualidad Y no está afectada por el hecho de que el sospechoso la tenga. Resumiendo:

- Los 101 habitantes de la isla están bajo sospechosa inicialmente y en la misma proporción.
- Uno de los habitantes de la isla es el criminal.
- El criminal y el sospechoso tienen la cualidad Y .
- Se desconoce si el resto de los isleños tienen la cualidad Y o no.
- Se espera que una persona de cada 100 tengan la cualidad Y .

Hay 101 isleños: 100 de ellos son inocentes y 1 culpable. Tenemos la certeza de que el criminal tiene la cualidad Y . Encontramos a un sospechoso con la cualidad Y . Debido a que se espera que una persona de cada 100 tenga la cualidad Y , es posible que uno de los habitantes de la isla pueda tenerla sin ser el criminal. Por tanto, esperamos que haya dos personas con la cualidad Y : el criminal y el sospechoso, cuando éste no sea el criminal, o el sospechoso y uno de los habitantes de la isla (en este caso el sospechoso es el criminal). Diríamos que hay un 1% de tasa de falso positivo, por lo que de 100 personas inocentes en la isla esperamos que una de ellas tenga la cualidad Y . Por tanto, aunque Y sea una cualidad rara y haya habido coincidencia en ella entre la muestra dubitada procedente del criminal y la indubitada procedente del sospechoso, sólo hay un 50% de probabilidad de que el sospechoso sea el criminal.

Aunque parezca poco probable que un individuo tenga una cualidad determinada (1% en nuestro ejemplo), debido a la coincidencia producida con la muestra dubitada se alcanza un 50% de probabilidad de que el sospechoso pueda ser el culpable.

9.1.1 La primera lección del problema de la isla

Lo que resulta relevante para el Tribunal o el Jurado es la probabilidad de que el sospechoso sea culpable dada una coincidencia en la cualidad Y . Esta probabilidad es, en un caso real, bastante difícil de calcular. Sin embargo, trabajamos en un contexto de isla simplificada en el que vamos a poder calcularla. De momento, consideramos que no existen pruebas adicionales y que todos los habitantes de la isla siguen siendo equiprobablemente sospechosos.

En general, si hay N habitantes en la isla, además del sospechoso, y la probabilidad de que tengan la cualidad Y es p , llegamos a la siguiente fórmula de la probabilidad condicional de culpabilidad del sospechoso dada la observación de la cualidad Y : $P(H_f | Y) = 1 / (1 + N p)$, donde H_f significa “hipótesis del Fiscal” e Y se ha de entender como la coincidencia en la cualidad Y entre la muestra dubitada y la indubitada procedente del sospechoso.

Sin que todavía ofrezcamos explicaciones sobre el porqué de esta fórmula, si la aplicamos al caso del problema de la isla enunciado llegamos al resultado del apartado 9.1: $P(H_f | Y) = 1 / (1 + 100 \times (1/100)) = 1 / 2 = 0.5$, o sea, el 50%. Un 50% en la probabilidad condicional mencionada significa quedarse en la duda. No existen razones para decantarse ni a favor de que el sospechoso sea la fuente de la cualidad Y encontrada en la muestra recogida en la escena del crimen ni a favor de la proposición contraria.

Cuanto más rara sea la cualidad Y , mayor será la probabilidad de que el sospechoso sea el criminal. Sin embargo, la fuerza de la prueba contra el sospechoso depende tanto de la rareza de la cualidad Y como del número de posibles fuentes de la muestra dubitada, o sea, N .

Primera lección: *el hecho de que la cualidad Y sea rara (que es lo mismo que decir que p sea muy pequeño) no conlleva, aisladamente considerado, que el sospechoso sea probablemente culpable.* Consúltese el Anexo III para encontrar una explicación pormenorizada de la fórmula matemática empleada para resolver el problema.

9.1.2 Acercamiento a un contexto real

9.1.2.1 Incertidumbre sobre el valor de p

En la práctica, el valor de p – la probabilidad de la cualidad Y – no se conoce con exactitud. Lo que podemos hacer es estimarlo mediante técnicas de muestreo. Por tanto, siempre habrá una incertidumbre que ciframos igual a la varianza σ^2 . En la fórmula de la probabilidad condicional de la culpabilidad del sospechoso dado que tiene la cualidad Y hasta ahora considerada, la varianza fue igual a cero. Existe una fórmula de esa probabilidad condicional, más general, que la contempla:

$$P(H_f | Y) = 1 / [1 + N (p + (\sigma^2 / p))] < 1 / (1 + N p),$$

que, como se señala en la ecuación, producirá un valor de $P(H_f | Y)$ menor que el calculado con la fórmula primigenia con varianza igual a cero.

Segunda lección: *la incertidumbre sobre el valor de p no es despreciable. Si se ignora su efecto se perjudica al sospechoso.* Consúltese el Anexo III para encontrar una explicación pormenorizada de la fórmula matemática empleada para resolver el problema.

Un ejemplo sirve para cuantificar en qué medida puede afectar al sospechoso la ignorancia de la incertidumbre sobre p .

En el original problema de la isla, con $N = 100$ y $p = 0.01$ (1%), si aplicamos la primera fórmula tenemos que $P(H_f | Y) = 1 / (1 + 100 \times (1/100)) = 1 / 2 = 0.5$.

Si suponemos ahora que hay incertidumbre sobre p , digamos que $p = 0.01 \pm 0.005$ (es decir, $\sigma = 0.005$, siendo la desviación típica o estándar), entonces tenemos que:

$P(H_f | Y) = 1 / [1 + N (p + (\sigma^2 / p))] = 1 / [1 + 100 \times (0.01 + ((0.005)^2 / 0.01))] = 1 / (1 + 100 \times (0.01 + 0.0025)) = 1 / (1 + 100 \times (0.0125)) = 1 / (1 + 1.25) = 1 / 2.25 \cong 44\%$. Por tanto, no tener en cuenta la incertidumbre sobre el valor de p sobreestima la probabilidad de la hipótesis del Fiscal (culpabilidad del sospechoso) dada la observación de la cualidad Y . Pasar del 44% (valor más exacto) al 50% no puede considerarse irrelevante.

La incertidumbre sobre p , como ya hemos dicho, sobreviene por el muestreo – cada muestra tendrá un valor diferente estimando p – y siempre tendremos la posibilidad de que las muestras no sean suficientemente representativas (por ejemplo, los isleños pudieran no ser, en términos generales, tan similares a los habitantes del continente más cercano). Respecto al ADN, la incertidumbre sobre las frecuencias alélicas (p) es importante no sólo por el error de muestreo sino porque las proporciones de los perfiles de ADN varían según los grupos étnicos, religiosos y sociales, y nunca se conoce exactamente cuál es la población de referencia en un caso particular o cuáles son las proporciones de los perfiles en los grupos relevantes. Aunque reconocemos grupos humanos como los “Árabes”, “Hispanos”, “Escandinavos”, etc., como consecuencia de las migraciones, casamientos interraciales, y límites imprecisos, no podemos definirlos de forma exacta. La rareza de los perfiles de ADN se apoya, por una parte, en datos aceptados como *muestras de conveniencia* más que en *muestras aleatorias científicamente consideradas*, y por otra parte, los fundamentos de la teoría de genética poblacional descansan, en el mejor de los casos, en las poblaciones humanas actuales. En la práctica, σ^2 / p de nuestra última fórmula será mucho mayor que p , así pues, si se ignora, su influencia será mayor que en el ejemplo que ha servido para ilustrar el presente tema.

9.1.2.2 Incertidumbre sobre N

El tamaño N de la población de la isla, excluyendo al sospechoso, puede también desconocerse y tener una cierta influencia en el cálculo de $P(H_f | Y)$. En este caso, el efecto de la incertidumbre sobre N es, normalmente, pequeño y a favor del sospechoso. Si el tamaño real fuera N y considerásemos $N - 1$ o $N + 1$, cada uno de ellos con probabilidad ϵ , se llega a la fórmula siguiente:

$$P(H_f | Y) \cong [1 / (1 + N p (1 - 4\epsilon / N^3))] > 1 / (1 + N p).$$

Tercera lección: *La incertidumbre sobre N, el número de posibles sospechosos alternativos, tampoco puede despreciarse, pero el efecto de ignorar esta incertidumbre es generalmente pequeño y tiende a favorecer a los sospechosos.* Consúltese el Anexo III para encontrar una explicación pormenorizada de la fórmula matemática empleada para resolver el problema.

9.1.2.3 Posibles errores de catalogación (tipado)

Dejemos a un lado la incertidumbre y asumamos de nuevo que p y N son exactos. Supongamos, sin embargo, que haya una probabilidad ε_1 de que un individuo que no tenga la cualidad Y sea erróneamente catalogado como que la tiene (es decir, se trata de un falso positivo), y una probabilidad ε_2 para el error de falso negativo. Asumimos que ambas probabilidades de error se refieren tanto al sospechoso como a la muestra dubitada y que los errores ocurren independientemente.

Lo que hemos asumido sigue siendo una simplificación de la realidad, pero nos permite averiguar en qué medida la posibilidad de error afecta a la probabilidad de culpabilidad del sospechoso dada la observación de la cualidad Y . La fórmula propuesta es la siguiente:

$$P(H_f | Y) \cong [1 / 1 + N ((p + \varepsilon_1)^2 / p)]$$

Nótese que, como una primera aproximación, ε_2 es irrelevante. Una vez observada la coincidencia en la cualidad Y , la probabilidad de error de falso negativo carece de importancia.

Lección cuarta: *El valor de la cualidad como prueba incluye tanto la probabilidad de que se produzca una coincidencia por azar como la probabilidad de esa coincidencia debido a un error de catalogación.* Consúltese el Anexo III para encontrar una explicación pormenorizada de la fórmula matemática empleada para resolver el problema.

Un ejemplo numérico que ilustra lo que se acaba de mencionar es el siguiente: supongamos que $\varepsilon_1 = 0.005$, es decir, la probabilidad de que un isleño sin la cualidad Y sea catalogado erróneamente como que la tiene. La probabilidad de la hipótesis del Fiscal dada la cualidad Y es la siguiente:

$$P(H_f | Y) \cong [1 / 1 + N ((p + \varepsilon_1)^2 / p)] = [1 / 1 + 100 ((0.01 + 0.005)^2 / 0.01)] \cong 31\%.$$

Por tanto, ignorar la posibilidad de cometer un error en la catalogación puede conducir a sobreestimar la probabilidad de que la hipótesis del Fiscal sea cierta dada la observación de la cualidad Y . En este caso, ¡en lugar del 31% se ha defendido que fuera un 50%!

Los errores que pueden cometerse en un laboratorio que realiza cotejo de perfiles de ADN no son despreciables. Existen, al menos, dos formas de observar una coincidencia de perfiles erróneamente (en el sentido de que conlleve una falsa implicación del sospechoso con el crimen):

- Que se trate de una coincidencia aleatoria.
- Que los perfiles sean distintos pero se ha producido un error de catalogación en uno o en los dos perfiles produciéndose la falsa coincidencia.

Las dos posibilidades descritas son altamente improbables. Sin embargo, como en la relación de verosimilitudes de la fórmula se incluye tanto la probabilidad de coincidencia aleatoria como la de falsas coincidencias por errores de catalogación por cualquier causa, ignorar la posibilidad de una coincidencia falsa va en contra del sospechoso, y en algunas ocasiones de forma importante. Si la probabilidad de una coincidencia falsa fuera mucho mayor que la probabilidad de una coincidencia aleatoria, esta última sería incluso irrelevante en el cálculo de la fuerza de la prueba. La cuestión clave no es si la probabilidad de una falsa coincidencia es

pequeña sino en qué medida se diferencia de la probabilidad de una coincidencia aleatoria. Y, por otra parte, lo que verdaderamente importa no es la probabilidad de error de falsa coincidencia o de catalogación con carácter general sino la probabilidad de un error que pudiera haber conducido a una observación de una coincidencia Y . En la prueba de ADN la probabilidad de una coincidencia por casualidad es tan sumamente improbable que la probabilidad alternativa de error involuntario o fraudulento en el perfil del sospechoso es mucho más probable en comparación con ella, aunque en términos absolutos sea también muy improbable en sí misma⁵.

El valor de ε_1 es difícil de estimar, generalmente suele ser más difícil que estimar el valor de p . Esto es así como consecuencia de que las tasas de error son bajas y se necesitarían muchas pruebas para valorarlas debidamente. Y lo que es peor, ninguna serie de pruebas sería capaz de imitar las especiales circunstancias que concurren en cada caso, incluyendo la cantidad de ADN disponible, así como su nivel de contaminación o degradación. En última instancia corresponde al Tribunal o Jurado valorar en qué medida se ha podido producir un error sobre la base de la prueba presentada en el juicio. Resulta de gran importancia que el Tribunal o Jurado conozca qué errores se pueden cometer, en qué medida pueden presentarse en un caso concreto y qué efectos pueden tener en la fuerza de la prueba.

9.2 Búsquedas en base de datos

Dejemos a un lado, por el momento, las incertidumbres y los errores, y centremos la atención en la cuestión sobre la causa de que el sospechoso haya llamado la atención de los investigadores del crimen en la isla. En la práctica, los sospechosos suelen identificarse por una combinación de factores como antecedentes penales, comportamientos inusuales, relaciones con otros delitos o delincuentes, etc.

Supongamos ahora que el sospechoso fue identificado tras una búsqueda de la cualidad Y en una base de datos⁶: los isleños fueron examinados en orden secuencial aleatorio hasta que apareció el primero con esa cualidad. A ese individuo se le acusó de ser el posible autor del crimen. Además de todo lo que ya sabemos que forma parte del contexto del problema de la isla, añadimos la información de que k isleños han sido investigados y ninguno de ellos tuvo la cualidad Y . Hasta ahora, en el problema que contemplamos no se habían tenido en cuenta las razones por las cuales el sospechoso fue considerado como tal. ¿Cómo afecta a la probabilidad de que el sospechoso sea culpable dada la observación de la cualidad Y (es decir, a $P(H_f | Y)$),

⁵ En un ejercicio de transparencia, el *Netherlands Forensic Institute* (conocido por sus siglas NFI) catalogó y publicó hace unos años sus tasas de error (consúltese la siguiente referencia bibliográfica: Kloosterman A., Sjerps M. and Quak A. "Error rates in forensic DNA analysis: definition, numbers, impact and Communications". *For. Sci. Int. Genet.* 12 (2014): 77-85). Los kits comerciales que se usan en la actualidad para el análisis de muestras biológicas son mucho más sensibles que en el pasado y, por tanto, han aumentado las probabilidades de que se produzcan y detecten contaminaciones entre perfiles genéticos. De hecho, las contaminaciones se consideran hoy en día prácticamente inevitables, si bien los laboratorios acreditados mediante la aplicación de normas de gestión de la calidad tienen cierta facilidad para cuantificarlas, ya que quedan patentes en sus registros de análisis y en las acciones correctivas que deben aplicarse. En otros campos de la genética es habitual que los laboratorios tengan en cuenta sus tasas de error (ej., en laboratorios clínicos), pero aún no ocurre esto en el campo forense. Sería deseable, por tanto, que los laboratorios forenses no sólo tuvieran la capacidad de detectar sus errores, sino también que los cuantificaran, que los pusieran a disposición del tribunal, e incluso que los tuvieran en cuenta en la valoración de sus resultados.

⁶ Para una revisión de la controversia surgida sobre cómo debería evaluarse la evidencia cuando se encuentre al sospechoso a través de una búsqueda en una base de datos, recomendamos la siguiente lectura: <http://folk.uio.no/geirs/publ/Storvik-Egeland-2007.pdf>

el hecho de que se haya identificado tras una búsqueda en una base de datos? La respuesta es que la probabilidad de que un sospechoso sea culpable dada la observación de la cualidad Υ es mayor si se ha llegado hasta él a través de una búsqueda en una base de datos que si se ha realizado sin ella. La siguiente fórmula lo explica, donde k es el número de personas de la base exploradas sin la cualidad Υ hasta llegar al primero que la tiene:

$$P(H_f | \Upsilon) = 1 / (1 + (N - k) p) > 1 / (1 + N p).$$

Para entender intuitivamente la fórmula propuesta pensemos que el sospechoso ha aparecido justamente cuando era la última persona que quedaba en la isla por ser examinada de la cualidad Υ . Ninguno de los isleños examinados anteriormente poseían la cualidad Υ . El hecho de que ese sospechoso tenga la cualidad Υ es una prueba irrefutable contra él, como se refleja en el resultado de $P(H_f | \Upsilon)$ si $N = k$, que conlleva que $P(H_f | \Upsilon) = 1$. La razón que explica que la fórmula aportada es la correcta es que cada individuo examinado que no tiene la cualidad Υ es excluido de estar bajo sospecha (deja de ser equiprobablemente sospechoso con respecto a los isleños antes de que se examinaran por la cualidad Υ) – recuérdese que hemos supuesto que no hay errores en la realización del test -. Cada individuo que deja de ser sospechoso deja un conjunto de posibles sospechosos más reducido y, por consiguiente, cada uno de los que quedan por examinar va incrementando paulatinamente la probabilidad de culpabilidad. Obsérvese que si la primera persona que se examinara tuviera la cualidad Υ , $k = 0$, la fórmula de $P(H_f | \Upsilon)$ volvería a ser la del apartado 9.1.1.

Lección quinta: *En el caso de una búsqueda de posibles culpables en una base de datos que conlleve una coincidencia en una cualidad con la muestra de la escena del crimen, cuanto más larga sea la búsqueda (es decir, cuantos más individuos se hayan examinado sin que hubiera existido una coincidencia) mayor será la fuerza de la prueba contra aquél en el que se haya producido la coincidencia.*

ANEXO I

Silogismo estadístico bayesiano [2]⁷

Sea X una *variable aleatoria discreta* que puede tomar valores x con probabilidades $\Pr(X = x)$. Incluso cuando desconozcamos el valor de X , podemos calcular un valor representativo que se conoce como el *valor esperado o esperanza* de X , que se denota mediante $E(X)$, y que se calcula así:

$$E(X) = \sum_x x \Pr(X = x). \quad (1)$$

Supongamos que X es una variable aleatoria que toma los valores $\{1, 0\}$ (respectivamente, 'verdadero' y 'falso'), asociadas a la proposición A_i : "El individuo a_i tiene la propiedad Q ", donde i pertenece a la población R de n individuos. De (1) se sigue, inmediatamente, que *la esperanza de esta variable aleatoria es la probabilidad de que la proposición sea verdadera*:

$$E(X_i) = \{\Pr(X_i = 1) \times 1\} + \{\Pr(X_i = 0) \times 0\} = \Pr(X_i = 1).$$

Como la esperanza es aditiva (la esperanza de una suma es la suma de las esperanzas), la suma de las probabilidades de las proposiciones A_i es igual al valor esperado de las variables aleatorias A_i :

$$\Pr(X_1 = 1) + \dots + \Pr(X_n = 1) = E(X_1) + \dots + E(X_n) = E(X_1 + \dots + X_n). \quad (2)$$

Pero la suma de las variables es el número de individuos que tienen la propiedad Q entre los n individuos de la población R , es decir, la *frecuencia* de Q en R .

Dividiendo cada lado de (2) por n , y aplicando la linealidad de E , es decir, $E(X_i) / n = E(X_i / n)$, obtenemos que *la probabilidad promedio es igual a la frecuencia relativa esperada*:

$$\frac{\Pr(X_1 = 1) + \dots + \Pr(X_n = 1)}{n} = E\left(\frac{X_1 + \dots + X_n}{n}\right). \quad (3)$$

Supongamos que la distribución $\Pr(X_i = x_i)$ es nuestra probabilidad subjetiva. Supongamos, también, que conocemos la frecuencia relativa γ : entonces γ puede sustituir a la parte derecha de la ecuación (3), y, si desconocemos qué $n\gamma$ de R tiene la propiedad Q y creemos que todos tienen la misma probabilidad de tener la propiedad Q , entonces se sigue, inmediatamente, que:

$$\Pr(X_i = 1) = \gamma, (i = 1, \dots, n).$$

⁷ Traducción de la página 22 con algunas mínimas aclaraciones adicionales.

Por consiguiente, si a las probabilidades se les da una interpretación subjetiva, se conocen las frecuencias relativas y las probabilidades de pertenecer a Q de todos los miembros de R se consideran iguales, las frecuencias relativas determinan las probabilidades individuales.

ANEXO II

Leyes de cambio de probabilidad y de improbabilidad [3]

A. Ley de cambio de probabilidad

Hemos afirmado que esta ley es subjetiva, es decir, aunque aparenta ser la mejor formulación matemática para interpretar un dato como prueba – sobre todo por su simplicidad al requerir aparentemente una sola proposición -, sin embargo depende de otras proposiciones alternativas y sus probabilidades a priori. Veamos un ejemplo.

Recordemos que esta ley compara la probabilidad de la veracidad de una proposición - una vez obtenidos los datos - con la probabilidad de la veracidad de esa misma proposición antes de obtenerlos. A medida que la primera probabilidad crece, mayor valor de prueba tienen los datos.

Queremos evaluar una observación como prueba a favor de una proposición. Conocemos la probabilidad a priori de la proposición y la verosimilitud de la proposición conocido el valor de la observación, o lo que es lo mismo, la probabilidad del valor de la observación conocida la veracidad de la proposición.

Ocurre un crimen y disponemos de una muestra tomada de un vestigio recogido en la inspección ocular. De la muestra extraemos un perfil de ADN. Hay un sospechoso y obtenemos su perfil de ADN. Los perfiles dubitado e indubitado coinciden. Tenemos un *match*.

Queremos valorar el *match* como prueba y nos acogemos a la ley del cambio de probabilidad. Por tanto, necesitamos la probabilidad a priori de que el sospechoso pudiera ser la fuente del perfil de ADN hallado en la muestra y la probabilidad a posteriori de que una vez verificada la coincidencia de perfiles sea el sospechoso la fuente del perfil dubitado.

Tenemos que utilizar notación matemática para percibir si la ley del cambio de probabilidad es válida para interpretar el *match* como prueba.

Llamamos A a la proposición de que el sospechoso es la fuente del perfil de ADN hallado en la muestra dubitada. Llamamos M al *match*. Esta ley nos dice que la relación de probabilidades $P(A|M) / P(A)$ representa la forma de que el *match* sea valorado como prueba. En las probabilidades condicionales, como la de la izquierda de la igualdad, se coloca una barra vertical entre las proposiciones consideradas: probabilidad de que la proposición A sea cierta condicionada a que se haya verificado un *match* (coincidencia entre los perfiles comparados) por el perito.

Utilizando propiedades de la teoría de la probabilidad, podemos desarrollar la expresión matemática anterior de la siguiente forma:

$$P(A|M) / P(A) = P(M|A) P(A) / P(A) P(M) = P(M|A) / P(M)$$

Y $P(M)$ puede desplegarse del siguiente modo: $P(M) = P(M|A) P(A) + P(M|\bar{A}) P(\bar{A})$, donde \bar{A} es la proposición complementaria a A, es decir, que la fuente del perfil no es el sospechoso.

Por tanto, la expresión de la ley del cambio de probabilidad queda, finalmente, así:

$$P(A|M) / P(A) = P(M|A) / P(M) = P(M|A) / (P(M|A)P(A) + P(M|\bar{A})P(\bar{A}))$$

Esta expresión indica que para interpretar el *match* como prueba no es posible hacerlo desde la perspectiva de una única proposición, es decir, de A. Hace falta conocer, además de $P(M|A)$ y $P(A)$, $P(M|\bar{A})$. $P(\bar{A})$ es calculable a partir de $P(A)$ por complementariedad pero $P(M|\bar{A})$ no es complementaria de $P(M|A)$.

Hace falta, al menos, tener en cuenta la proposición \bar{A} para interpretar el *match* como prueba.

Si en lugar de dos proposiciones hubiera tres, podría fácilmente demostrarse que para interpretar el *match* siguiendo la ley del cambio de probabilidad como prueba dependeríamos de las probabilidades a priori de las tres y de sus respectivas verosimilitudes.

La dependencia de la ley analizada de probabilidades a priori para interpretar lo observado como prueba la convierte en subjetiva per se. Esas probabilidades las asigna cada persona según lo que conoce y asume como cierto.

La dependencia de las verosimilitudes, también hace ver que la interpretación de lo observado como prueba no depende de una proposición sino de todas las posibles aplicables al caso. La simplicidad aparente de la expresión matemática de esta ley, como dependiente de una única proposición, se queda en eso, en apariencia.

La ley de verosimilitud no depende de las probabilidades a priori de las proposiciones que entran en juego. En este sentido es mucho más objetiva, entendiendo por objetividad la propiedad de que pueda ser más fácilmente aceptada por todos.

B. Ley de improbabilidad

Aunque la teoría de Neyman-Pearson se orientó por sus creadores hacia la toma de una decisión, con el tiempo fue derivando hacia la interpretación de lo observado como prueba. El proceso de deriva fue gradual y fue imponiéndose de forma generalizada en la ciencia experimental con las aportaciones de investigadores de la talla de Wald, Chernoff, Moses, Barnard, Fisher, Cox, Cornfield, Hinkley, Lehmann y muchos otros destacados estadísticos del siglo XX.

Los denominados *tests de significancia* de Fisher parecían reunir las condiciones ideales para la interpretación de los datos como prueba. Se conocen por el resultado del procedimiento que permite realizar la interpretación, el *p-valor*. Apoyándose en las conocidas propiedades de los tests de hipótesis de la teoría de Neyman-Pearson, es decir, en la posibilidad de disponer de tests de los que podemos conocer sus tasas de error (falsos positivos y negativos), y en los que podemos fijar el máximo error que toleramos cometer, los tests de significancia se han propuesto mayoritariamente en la ciencia experimental desarrollada en el siglo XX como el procedimiento más apto disponible para interpretar lo observado como prueba. La fuerza de la prueba está relacionada con la rareza de obtener valores iguales o superiores al *p-valor* si la hipótesis que se contrasta fuera cierta. Cuanto mayor sea, más fuerza tendrá como prueba lo observado por su mayor rareza.

Puede demostrarse que los test de significancia están afectados por el espacio muestral disponible, no así cuando se aplica la ley de verosimilitud.

¿Qué significa que una ley que se proponga como modelo interpretativo de lo observado como prueba sea independiente del espacio muestral?

Básicamente significa que la valoración de lo observado como prueba no puede depender de la información muestral que tengamos disponible, que puede ser mayor de la necesaria. En otras palabras, sea cual fuere la información muestral que tengamos para realizar la interpretación, el resultado de tal interpretación ha de ser siempre el mismo.

Veamos con un ejemplo en qué medida esta propiedad invalida a los tests de significancia como procedimientos correctos para interpretar lo observado como prueba.

Necesitamos recordar qué es un experimento de Bernoulli, uno de los más simples pero, al mismo tiempo, más frecuentes que sirven para resolver problemas reales. Estas son sus características:

- Las observaciones se clasifican en DOS categorías.
- La proporción de elementos de ambas categorías en la población es CONSTANTE, cualquiera que sea la cantidad observada.
- Por tanto, si la población es FINITA, los elementos se reemplazan una vez observados.
- Se llama p a la probabilidad de observar una de las categorías y q a la complementaria, o sea, $1 - p$.
- Las observaciones son INDEPENDIENTES. Por tanto, las probabilidades p y q son siempre las mismas, siendo irrelevantes los resultados que se puedan ir obteniendo.
- El proceso de Bernoulli puede aplicarse a poblaciones finitas (tomamos elementos al azar con reemplazamiento) o a poblaciones conceptualmente infinitas (procesos estables en el tiempo y sin memoria). Ejemplos: observar si un recién nacido es niño o niña, si un cliente está o no satisfecho de un servicio, aparición del número 3 en tiradas sucesivas de un dado, aparición de una determinada pieza defectuosa en un proceso de fabricación, etc.
- En el proceso de Bernoulli podemos definir distintas variables aleatorias que darán lugar a distintas distribuciones de probabilidad.

Podemos afirmar, razonablemente, que las monedas que salen de las fábricas oficiales de tales utensilios están bien hechas, es decir, podemos esperar que si las lanzamos al aire repetidas veces y de un modo equilibrado tenderemos a ver tantas caras como cruces. Imaginemos que lanzamos la moneda 20 veces para generar un espacio muestral suficiente para realizar inferencias y queremos saber la probabilidad de que salgan 6 caras en los 20 lanzamientos.

Lo primero que examinamos es si el experimento del ejemplo es realmente un experimento de Bernoulli:

- Las observaciones se clasifican en DOS categorías: cara y cruz.
- La proporción de elementos de ambas categorías en la población es CONSTANTE, cualquiera que sea la cantidad observada. Asumimos como algo cierto que todas las monedas que se fabrican tienen cara y cruz.
- Se llama p a la probabilidad de observar cara en un lanzamiento y q a la complementaria, o sea, $1 - p$.

- Las observaciones son INDEPENDIENTES, es decir, el resultado de un lanzamiento no influye en el resultado del siguiente. Por tanto, las probabilidades p y q son siempre las mismas, siendo irrelevantes los resultados que se puedan ir obteniendo.

Una vez que hemos comprobado que el experimento del ejemplo cumple los requisitos para catalogarse como experimento de Bernoulli, nos interesa formalizar el modelo de probabilidad de sus resultados.

Para ello utilizamos una variable aleatoria X que represente el número de caras que queremos observar en 20 lanzamientos. Si queremos observar, por ejemplo, seis caras, $X = 6$. El espacio de valores posibles de la variable X tiene 21 elementos: $\{0, 1, 2, \dots, 20\}$. $X = 0$ significa que en los 20 lanzamientos se ha observado cruz.

Nuestro espacio muestral consta de información hasta 20 lanzamientos. Imaginemos que alguien distinto a nosotros tiene que calcular la misma probabilidad de que salgan 6 caras y su variable aleatoria Y tuviera sólo estos dos elementos posibles: $Y = 6$, e $Y = \text{no-6}$. La clasificación de los resultados posibles en distintas categorías se suele denominar *codificación*. En la de 21 elementos, cada resultado es una categoría distinta. En la de 2 elementos, clasificamos los resultados en dos categorías.

Lo que queremos decir cuando exigimos que la interpretación de un resultado como prueba sea independiente del espacio muestral es que podamos interpretar el resultado siempre del mismo modo sea cual fuere el número de elementos del espacio muestral disponible, es decir, del número de categorías en que dividamos el espacio muestral o, si se prefiere, de la codificación de los resultados utilizada.

En otras palabras, queremos que la probabilidad de lo observado, es decir, que salgan 6 caras en 20 lanzamientos, sea exactamente el mismo número con independencia de que tengamos información de 20 lanzamientos con la variable X o sólo la información resumida al máximo con la variable Y . Veámoslo:

A la variable X podemos aplicarle un modelo de probabilidad de que salgan 6 caras en 20 lanzamientos como sigue:

$$\binom{20}{n} p^n q^{20-n}$$

El número combinatorio entre corchetes significa combinaciones de 20 elementos tomados de n en n . Nos interesa en el ejemplo que $n = 6$. Por tanto donde se escribe la variable n pondremos el valor numérico por el que estamos interesados.

La variable p representa la probabilidad de que en un lanzamiento cualquiera salga cara y la q de que salga cruz, como anteriormente indicamos.

$$\binom{20}{6} p^6 q^{14} = \binom{20}{6} (0.5)^6 (0.5)^{14} = 0.0369644165 \approx 0.04$$

Si la moneda es equilibrada, la probabilidad de que salgan 6 caras en 20 lanzamientos es del 4%, aproximadamente.

A la variable Y podemos asignarle el mismo modelo de probabilidad que a X pero restringido a la condición de que salgan 6 caras en 20 lanzamientos (la misma fórmula que para $n = 6$):

$$\binom{20}{6} p^6 q^{14} \approx 0.04$$

La probabilidad de que no se cumpla la condición anterior para la variable Y es, sencillamente, la probabilidad complementaria:

$$1 - \binom{20}{6} p^6 q^{14} = 1 - 0.0369644165 \approx 0.96$$

La sencillez de la información disponible del espacio muestral relacionado con la variable Y nos constriñe a dos posibles cálculos probabilísticos: que salgan 6 caras en los 20 lanzamientos frente a la proposición complementaria de que no salgan en esas condiciones.

Sin embargo, la información que nos proporciona la variable X es mucho más amplia pues podemos calcular probabilidades de que salga cualquier número de caras, entre 0 y 20, ambos inclusive, en 20 lanzamientos.

En cualquier caso, si nos interesa lo mismo, es decir, que salgan 6 caras en 20 lanzamientos, el resultado es el mismo tanto si utilizo la variable X como la variable Y. Esto es lo realmente importante para que la interpretación de lo observado como prueba sea válida.

No obstante, la ley de verosimilitud establece que lo observado constituye prueba a favor de una proposición frente a otra. Hasta ahora sólo hemos considerado una proposición porque hemos considerado la moneda ideal. Sin embargo, las monedas reales no están perfectamente equilibradas. Esto produce que tengan cierta propensión por uno de los dos resultados: hacia cara o hacia cruz. Además, el lanzador también influye en este sentido. Por tanto, pudieran existir dos proposiciones donde las probabilidades de cara y cruz fueran ligeramente diferentes a 0.5.

La relación de verosimilitudes sería, entonces, la siguiente:

$$\frac{\binom{20}{n} p_2^n q_2^{20-n}}{\binom{20}{n} p_1^n q_1^{20-n}} = \frac{p_2^n q_2^{20-n}}{p_1^n q_1^{20-n}}$$

Bajo la ley de verosimilitud no habría diferencia alguna en cuanto a la expresión del LR entre las dos variables X e Y:

$$LR = \frac{p_2^6 q_2^{14}}{p_1^6 q_1^{14}}$$

Para interpretar lo observado como prueba, la ley de verosimilitud sólo precisa el LR. El que dispongamos de mayor o menor información del espacio muestral es irrelevante. También es irrelevante cómo cada una de las distribuciones de probabilidad consideradas asignan probabilidades sobre los valores no observados. Lo único que cuenta es la relación entre las verosimilitudes, es decir, entre las probabilidades de lo observado condicionadas a cada una de las proposiciones.

El *p-valor* no es solamente la probabilidad de lo observado sino también de todas las probabilidades de todos los valores más extremos que no han sido observados. Una medida de la fuerza de la prueba no puede depender de probabilidades de valores no observados.

Apliquemos la teoría de Fisher al problema de los 20 lanzamientos de una moneda donde desconocemos la probabilidad de que salgan caras. Recordemos que disponemos de dos formas de codificar el espacio muestral: con 21 elementos o con 2. Si el número de caras observadas en los 20 lanzamientos es 6, con las dos formas de codificar el resultado de la prueba se obtiene el mismo número: 6. Supongamos que se defienden dos proposiciones: H_1 que dice que la probabilidad de que salga cara al lanzar la moneda es 0.5; y H_2 que dice que la probabilidad de que salga cara al lanzarla es 0.45.

El LR, intentando encontrar evidencia a favor de H_2 frente a H_1 , es el siguiente: $(0.45)^6 \times (0.55)^{14} / (0.5)^6 \times (0.5)^{14} = (0.45)^6 \times (0.55)^{14} / (0.5)^{20} \approx 2.08146$. Este valor de LR nos dice que lo observado apoya aproximadamente el doble a la proposición H_2 que a la H_1 .

Vamos ahora a comprobar que los *p-valores* que se obtienen cuando se enfrenta H_1 con H_2 no están de acuerdo con la fuerza de la prueba calculada mediante el LR.

Utilizando el *p-valor* con un espacio muestral de 21 elementos $\{0, 1, 2, \dots, 20\}$, podemos obtener esta probabilidad: $p(X = 6) + p(X = 5) + p(X = 4) + p(X = 3) + p(X = 2) + p(X = 1) + p(X = 0) \approx 0.06$. Hemos tenido en cuenta aquellos resultados (número de caras) que producen un LR que apoya H_2 frente a H_1 con un valor numérico mayor que el que se obtiene con $X = 6$.

Utilizando el *p-valor* con un espacio muestral de sólo dos elementos, $\{6, \text{no-6}\}$, podemos obtener esta otra probabilidad: $p(X = 6) \approx 0.04$.

Los resultados anteriores evidencian que el *p-valor* aplicado para medir la fuerza como prueba de lo observado conduce a resultados distintos en función del espacio muestral disponible.

ANEXO III

Explicaciones de las fórmulas matemáticas empleadas en “el problema de la isla” [8]

Apartado 9.1.1

Para que la fórmula $P(H_f | Y) = 1 / (1 + 100 \times (1/100))$ pueda ser deducida con facilidad es necesario recordar algunos conceptos y puntualizar algunos detalles prácticos:

1. La probabilidad de un suceso – por ejemplo que un sospechoso sea culpable de un crimen – puede hallarse a través de la definición de probabilidad de Laplace: casos favorables dividido entre casos posibles. Si se espera que de cada 100 personas que viven en el continente más cercano a la isla, una de ellas tenga la cualidad Y , la probabilidad de que una persona escogida al azar en el continente tenga la cualidad Y será $1 / 100$, es decir, un caso favorable por cada 100 posibles. Se asume esa probabilidad en la isla por su cercanía al continente.
2. La probabilidad de un suceso condicionado a otro es igual a la probabilidad de la intersección de los sucesos dividida por la probabilidad del suceso que condiciona. Al interseccionar dos sucesos puede ocurrir que tengan elementos comunes en sus espacios muestrales, por lo que si esos elementos se observan, harán que los dos sucesos ocurran a la vez. Si no tienen elementos comunes y se observa alguno de ellos, sólo ocurrirá uno u otro suceso. Para esos elementos la intersección de los sucesos es nula.

Centrándonos en el camino hacia la fórmula anterior, queremos deducir la probabilidad de la culpabilidad dada la observación de la cualidad Y , o sea, $P(H_f | Y)$:

$P(H_f | Y) = P(H_f \cap Y) / P(Y)$, donde \cap significa intersección. Aplicando la tercera ley de la teoría de la probabilidad: $P(H_f \cap Y) = P(Y | H_f) P(H_f)$, lo que nos permite llegar hasta esta fórmula:

$P(H_f | Y) = [P(Y | H_f) P(H_f)] / P(Y)$, y aún podemos seguir avanzando en su desarrollo. Para ello se ha de recordar la ley de la probabilidad total de la teoría de la probabilidad.

La ley de la probabilidad total nos dice que la probabilidad de un suceso, por ejemplo Y , puede hallarse de esta forma: $P(Y) = \sum P(Y | H_i) P(H_i)$, donde el sumatorio expresado con la letra griega sigma mayúscula se realiza entre $i = 1$ hasta S (siendo S el número de sucesos que condicionan y que tienen la propiedad de ser disjuntos dos a dos – disjuntos significa que su intersección ha de ser el conjunto vacío - y cuya unión contiene todas las observaciones que configuran el espacio muestral de Y ; además, la intersección de Y con cualquiera de ellos no es el conjunto vacío). A los sucesos condicionados los hemos llamado H para significar que son hipótesis, y en nuestro contexto contemplamos la del Fiscal y la de la Defensa. Así, resulta muy común la expresión siguiente en criminalística:

$P(Y) = \sum P(Y | H_i) P(H_i) = P(Y | H_f) P(H_f) + P(Y | H_d) P(H_d)$, donde el subíndice f indica “Fiscal” y el subíndice d indica “Defensor”. No resulta difícil entender que la probabilidad de una coincidencia en la cualidad Y entre la muestra dubitada e indubitada, o sea, $P(Y)$, puede igualarse a la probabilidad de la observación de la cualidad Y condicionada a la hipótesis del Fiscal, multiplicada por la probabilidad de la hipótesis del Fiscal antes de que la cualidad Y sea conocida, sumada a la probabilidad de la cualidad Y condicionada a la hipótesis de la Defensa, multiplicada por la probabilidad de la hipótesis de la Defensa antes de que la cualidad Y sea

conocida. Las hipótesis son disjuntas claramente: no pueden ser ciertas a la vez – suponemos que el Fiscal sostiene que la muestra dubitada procede del sospechoso y que el Defensor sostiene lo contrario –, ambas unidas conforman el espacio muestral de la cualidad Y y la intersección de la cualidad Y con cualquiera de las hipótesis no es el conjunto vacío.

Lo anterior nos lleva a la siguiente fórmula:

$$P(H_f | Y) = [P(Y | H_f) P(H_f)] / P(Y) = [P(Y | H_f) P(H_f)] / [P(Y | H_f) P(H_f) + P(Y | H_d) P(H_d)].$$

Al igual que hemos calculado $P(H_f | Y)$, podemos calcular $P(H_d | Y)$:

$$P(H_d | Y) = [P(Y | H_d) P(H_d)] / P(Y) = [P(Y | H_d) P(H_d)] / [P(Y | H_d) P(H_d) + P(Y | H_f) P(H_f)].$$

El siguiente paso consiste en hallar las apuestas entre las hipótesis dada la observación de la cualidad Y , que no es otra cosa que dividir entre sí las ecuaciones anteriores:

$$\begin{aligned} [P(H_f | Y) / P(H_d | Y)] &= [P(Y | H_f) / P(Y | H_d)] * [P(H_f) / P(H_d)] \\ P(H_f | Y) &= [P(Y | H_f) / P(Y | H_d)] * [P(H_f) / P(H_d)] * P(H_d | Y) \\ [1 / P(H_f | Y)] &= [P(Y | H_d) / P(Y | H_f)] * [P(H_d) / P(H_f)] * [1 / P(H_d | Y)] \end{aligned}$$

Si llamamos $R = [P(Y | H_d) / P(Y | H_f)]$ y $w = [P(H_d) / P(H_f)]$. Nótese que estas variables se han definido en contra de lo habitual – las hipótesis del Fiscal y de la Defensa están en el lugar contrario al normalmente utilizado por muchos autores – porque para nuestro contexto es conveniente hacerlo así:

$$\begin{aligned} [1 / P(H_f | Y)] &= R * w * [1 / P(H_d | Y)] \\ P(H_f | Y) &= (1 - P(H_f | Y)) / R * w \\ P(H_f | Y) R * w &= 1 - P(H_f | Y) \\ P(H_f | Y) + P(H_f | Y) R * w &= 1 \\ P(H_f | Y) (1 + R * w) &= 1 \\ P(H_f | Y) &= 1 / (1 + R * w) \end{aligned}$$

Si en lugar de que consideremos H_d como un suceso relacionado con una sola persona distinta al sospechoso, lo tratamos como un suceso compuesto de tantos simples como candidatos a ser fuentes de la muestra dubitada, siendo todos ellos distintos al sospechoso – que es contemplado en el suceso H_f –, el desarrollo realizado más arriba se complicaría un poco más por la existencia de sumatorios relacionados con H_d , pero resulta sencillo comprenderlo si consideramos que:

$H_d = \sum H_k$, donde k va de 1 a N , siendo N el número de candidatos a ser fuentes de la muestra dubitada distintos al sospechoso.

$$R_{k,s} = [P(Y | H_k) / P(Y | H_f)] \text{ y } w_{k,s} = [P(H_k) / P(H_f)].$$

De esta forma, podemos llegar a la siguiente expresión de la probabilidad condicional de la culpabilidad del sospechoso dada la observación de la cualidad Y :

$$P(H_f | Y) = 1 / (1 + \sum (R_{k,s} * w_{k,s})), \text{ donde } k \text{ va de } 1 \text{ a } N, \text{ siendo } N \text{ el número de personas que pudieran competir con el sospechoso como fuente de la muestra dubitada.}$$

En el ejemplo del problema de la isla el sumatorio $\sum(R_{k,s} * w_{k,s})$ se realiza desde $k = 1$ hasta $k = 100$, la probabilidad $p = 1/100$ y $N = 100$.

$P(Y | H_k) = p^2$ porque la probabilidad de que dos individuos cualesquiera de la isla (el culpable y un inocente) tengan la cualidad Y se halla multiplicando las probabilidades de que cada uno de ellos pueda tener esa cualidad, que para cada uno es p .

$P(Y | H_f) = p$ que es la probabilidad del perfil entre los que pueblan la isla.

$P(H_k) = P(H_f) = 1/101$, porque hay 101 equiprobales sospechosos en la isla.

$$\sum R_{k,s} = \sum [P(Y | H_k) / P(Y | H_f)] = [(1/100)^2 / (1/100)] * 100 = 1/100 = p.$$

$$\sum w_{k,s} = \sum [P(H_k) / P(H_f)] = [(1/101) / (1/101)] * 100 = 100 = N.$$

$$\sum (R_{k,s} * w_{k,s}) = \sum R_{k,s} * \sum w_{k,s} = (1/100) * 100 = p * N = 1.$$

$$P(H_f | Y) = 1 / (1 + \sum (R_{k,s} * w_{k,s})) = 1 / (1 + p * N) = 0.5.$$

Apartado 9.1.2.1

Conforme a lo desarrollado en el apartado 9.1.1, la relación de verosimilitudes para cualquier posible culpable es la siguiente: $R_{k,s} = [P(Y | H_k) / P(Y | H_f)] = p^2 / p = 1 / p$.

Si tenemos en cuenta la incertidumbre sobre p , ahora \tilde{p} es un estimador del verdadero valor del parámetro p en la población. Se trata de una variable aleatoria con media p y varianza σ^2 .

Calculamos las esperanzas del numerador y denominador de $R_{k,s}$. La esperanza del numerador $E[\tilde{p}^2]$ la deducimos de la definición de la varianza de una variable aleatoria: $\text{Var}[\tilde{p}] = E[\tilde{p}^2] - E[\tilde{p}]^2$, de donde se deduce que $E[\tilde{p}^2]$ es igual a $p^2 + \sigma^2$. La esperanza del denominador $E[\tilde{p}]$ es igual a p porque la probabilidad de la cualidad Y en la población es desconocida y en ese caso utilizamos su valor medio.

Así pues, $R_{k,s} = [p^2 + \sigma^2] / p = p + (\sigma^2 / p)$. Como es frecuente que (σ^2 / p) sea mucho mayor que p por los conocimientos disponibles sobre genética poblacional, el efecto de la incertidumbre sobre p puede tener efectos importantes sobre el valor de $P(H_f | Y)$. Intuitivamente considerado, la observación de un perfil de ADN en particular lo hace mucho más probable debido a la existencia de parentesco en la población, por lo que es probable que exista otra copia del perfil.

Apartado 9.1.2.2

La incertidumbre sobre N afecta a la probabilidad a priori, $P(H_f)$. Sea \tilde{N} el número de isleños inocentes en la isla, una variable aleatoria de media N . $P(H_f | \tilde{N}) = 1 / (1 + \tilde{N})$. Como \tilde{N} es desconocido necesitamos usar la esperanza:

$$P(H_f) = E[H_f | \tilde{N}] = E[1 / (1 + \tilde{N})].$$

Como $1 / (1 + \tilde{N})$ no es una función simétrica de N , aunque es convexa (la función curva cae por debajo de una línea recta entre dos puntos cualesquiera de la curva), se sigue de un resultado fundamental de la teoría de la probabilidad (la desigualdad de Jensen) que

cualquiera que sea la distribución de probabilidad de \tilde{N} , siempre que $E[\tilde{N}] = N$, la probabilidad a priori a favor del Fiscal nunca será menor que en el caso en que \tilde{N} fuera conocido. Es decir:

$$P(H_f) = E[1 / (1 + \tilde{N})] \geq 1 / (1 + N).$$

Con respecto al ejemplo de este apartado, tenemos que:

$\tilde{N} = N - 1$ con probabilidad ε ;

$\tilde{N} = N$ con probabilidad $1 - 2\varepsilon$;

$\tilde{N} = N + 1$ con probabilidad ε ; entonces:

$$E[1 / (1 + \tilde{N})] = (\varepsilon / N) + [(1 - 2\varepsilon) / (1 + N)] + [\varepsilon / (2 + N)] = [1 / (1 + N)] + [2\varepsilon / (N(1 + N)(2 + N))] \geq [1 / (1 + N)].$$

A la fórmula $P(H_f | \Upsilon) \cong [1 / (1 + Np(1 - 4\varepsilon / N^3))]$ se llega utilizando:

$$P(H_f) = [1 / (1 + N)] + [2\varepsilon / (N(1 + N)(2 + N))].$$

Como la incertidumbre sobre \tilde{N} no afecta a la relación de verosimilitudes, si se tiene en cuenta a mayor probabilidad a priori a favor de la hipótesis del Fiscal habrá mayor probabilidad a posteriori a favor de esa hipótesis que si no se tiene en cuenta cuando N sea conocido. Si se ignora la incertidumbre se tiende a favorecer al sospechoso, aunque el efecto sea generalmente pequeño. En la práctica, la incertidumbre sobre \tilde{N} es sustituida por un valor límite superior que también tiende a favorecer al sospechoso.

Apartado 9.1.2.3

Consideremos el problema de la isla en el que suceden errores de catalogación de forma independiente, con probabilidades ε_1 y ε_2 .

Si el sospechoso y el criminal no fueran la misma persona, la evidencia se encuentra en una de estas tres situaciones posibles:

- Tanto el sospechoso como el criminal tienen la cualidad Υ , y no ha ocurrido error alguno de catalogación:
 - Probabilidad de la cualidad Υ en el sospechoso: p
 - Probabilidad de acierto de catalogación: $(1 - \varepsilon_2)$
 - Probabilidad conjunta parcial 1 (sucesos independientes): $p(1 - \varepsilon_2)$
 - Probabilidad de la cualidad Υ en el criminal: p
 - Probabilidad de acierto de catalogación: $(1 - \varepsilon_2)$
 - Probabilidad conjunta parcial 2 (sucesos independientes): $p(1 - \varepsilon_2)$
 - Probabilidad conjunta total (sucesos independientes): $p^2(1 - \varepsilon_2)^2$
- El sospechoso o el criminal tiene la cualidad Υ y el otro no, y ha ocurrido un error de catalogación (falso positivo):
 - Probabilidad de la cualidad Υ en uno de los dos: p
 - Probabilidad de no tener la cualidad Υ en uno de los dos: $(1 - p)$

- Probabilidad conjunta parcial 1 (unión de sucesos tras aplicarse la propiedad de los sucesos independientes): $2 p (1 - p)$
 - Probabilidad de error de catalogación (falso positivo): ε_1
 - Probabilidad de acierto de catalogación: $(1 - \varepsilon_2)$
 - Probabilidad conjunta parcial 2 (sucesos independientes): $\varepsilon_1 (1 - \varepsilon_2)$
 - Probabilidad conjunta total (sucesos independientes): $2 p (1 - p) \varepsilon_1 (1 - \varepsilon_2)$
- Ni el sospechoso ni el criminal tienen la cualidad Υ , y ambos han sido erróneamente catalogados:
 - Probabilidad de no tener la cualidad Υ : $(1 - p)$
 - Probabilidad conjunta parcial 1 (sucesos independientes): $(1 - p)^2$
 - Probabilidad de error de catalogación (falso positivo): ε_1
 - Probabilidad conjunta parcial 2 (falsos positivos) – sucesos independientes -: ε_1^2
 - Probabilidad conjunta total (sucesos independientes): $(1 - p)^2 \varepsilon_1^2$

Si el sospechoso y el criminal fueran la misma persona, la evidencia se encuentra en una de estas dos situaciones posibles:

- El sospechoso tiene la cualidad Υ , y no ha ocurrido error alguno de catalogación:
 - Probabilidad de la cualidad Υ en el sospechoso: p
 - Probabilidad conjunta de acierto de catalogación (en el sospechoso y en la muestra dubitada) – sucesos independientes -: $(1 - \varepsilon_2)^2$
 - Probabilidad conjunta total (sucesos independientes): $p (1 - \varepsilon_2)^2$
- El sospechoso no tiene la cualidad Υ , y ha ocurrido un error de catalogación (falso positivo):
 - Probabilidad de no tener la cualidad Υ : $(1 - p)$
 - Probabilidad conjunta de error de catalogación (falso positivo en el sospechoso y en la muestra dubitada): ε_1
 - Probabilidad conjunta total (sucesos independientes): $(1 - p) \varepsilon_1^2$

Combinando todas las anteriores probabilidades podemos hallar $R_{k,s}$:

$$R_{k,s} = \frac{p^2(1-\varepsilon_2)^2 + 2p\varepsilon_1(1-p)(1-\varepsilon_2) + (1-p)^2\varepsilon_1^2}{p(1-\varepsilon_2)^2 + (1-p)\varepsilon_1^2} = \frac{(p + \varepsilon_1 - p(\varepsilon_1 + \varepsilon_2))^2}{p(1-\varepsilon_2)^2 + (1-p)\varepsilon_1^2} \approx (p + \varepsilon_1)^2 / p$$

La aproximación final se cumple si $p, \varepsilon_1, \varepsilon_2$ son pequeños.

Bibliografía referenciada

- [1] Lucy D, Introduction to Statistics for Forensic Scientists, Wiley, Chichester (UK), 2005.
- [2] Taroni F, Aitken CGG, Garbolino P, Biedermann A, Bayesian Networks and Probabilistic Inference in Forensic Science, Wiley, Chichester (UK), 2006.
- [3] Royall R, Statistical Evidence: A Likelihood Paradigm, Chapman&Hall, London (UK), 1997.
- [4] Hacking I, The Logic of Statistical Inference, Cambridge University Press, Cambridge (UK), 1965.
- [5] Gascón A, Lucena-Molina JJ, González-Rodríguez J, Razones científico-jurídicas para valorar la prueba científica: una argumentación multidisciplinar, Diario La Ley, Nº 7481, Sección Doctrina, 4 Oct. 2010, Año XXXI.
- [6] Aitken CGG, Taroni F, Statistics and the Evaluation of Evidence for Forensic Scientists, 2nd ed., Wiley, Chichester (UK), 2004.
- [7] Taroni F, Aitken CGG, Garbolino P, Biedermann A, Bayesian Networks and Probabilistic Inference in Forensic Science, 2^a ed., Wiley, Chichester (UK), 2014.
- [8] Balding DJ, Weight-of-evidence for Forensic DNA Profiles, Wiley, Chichester (UK), 2005.