

## EL PÉNDULO BAYESIANO: CRÓNICA DE UNA POLÉMICA ESTADÍSTICA

LUIS CARLOS SILVA AYÇAGUER  
ALINA BENAVIDES RODRÍGUEZ  
JOSÉ ALMENARA BARRIOS

### RESUMEN

La estadística es la responsable de ayudar a dirimir si un conjunto de datos apoya o no una hipótesis; para ello cuenta con un recurso que se remonta a más de 200 años: el Teorema de Bayes, según el cual se pueden emitir juicios probabilísticos sobre la validez de una hipótesis basándonos en los datos, pero también en la apreciación subjetiva que ella merezca. La preferencia del siglo XIX por los números y por la objetividad incentivó a los investigadores a buscar alternativas sustitutivas de un procedimiento que, al contener un elemento subjetivo, era mirado con suspicacia. Sin embargo, cada día se toma más conciencia de que los esfuerzos por suplantarlos, encarnados en personalidades tan importantes como Fisher, Neyman y Pearson, resultan ineficientes, de modo que en estos momentos asistimos a la irónica situación de que se revitaliza el enfoque bayesiano, el recurso más antiguo de la inferencia estadística.

### ABSTRACT

Statistics are responsible of determining whether a hypothesis could be supported by a set of data; the way to do that, known as Bayes' theorem, is more than 200 years old. According to it, we can give probabilistic judgements about the truth of the hypothesis using not only the data but a subjective idea about its fitness as well. The preference of the 19th century towards numbers and objectivity incentivated scientists to develop alternative approaches which could substitute a procedure that takes into account a subjective element and, therefore, was suspiciously considered. However, it is more and more obvious that the efforts to replace it, made by important personalities such as Fisher, Neyman and Pearson, are unsatisfactory. Thus, we can presently see an ironical situation: the return to the oldest resource to make statistical inferences.

Palabras clave: Estadística bayesiana, Probabilidad, Inferencia inductiva.

*Harán mejores conjeturas los que tengan más experiencia,  
porque tienen la mayoría de las señales con las cuales hacer la conjetura.*

Thomas Hobbes (1650)

## Introducción

La estadística, tal y como la conocemos hoy en día, es el resultado de un devenir evolutivo que, como todo proceso intelectual complejo, ha tenido múltiples hitos, se ha nutrido de diversas fuentes y ha transitado por numerosos conflictos. En este artículo nos detendremos en uno de los más relevantes e interesantes capítulos de ese proceso.

Como es bien sabido, el proceso que vincula el conocimiento subyacente con la realidad observable, conocido como razonamiento inferencial, puede discurrir lógicamente de dos maneras: deductiva e inductivamente. La estadística no escapa a esta dialéctica; por ejemplo, bajo el supuesto de que dos tratamientos son iguales, se puede calcular deductivamente, con relativa facilidad, la frecuencia esperada (probabilidad) de cualquier resultado posible que pudiera arrojar un estudio. Pero, obtenido un resultado particular, procedente o bien de un estudio experimental (por ejemplo, destinado a comparar dos terapéuticas) o bien de un estudio observacional (por ejemplo, que evalúe el efecto del hábito de fumar en adolescentes), no es tan fácil responder a la pregunta inductiva de cuán probable es que los tratamientos sean o no equivalentes [GOODMAN, 1999, pp. 995-1004].

La determinación de cuál de varias hipótesis es más probable sobre la base de los datos constituye un problema de probabilidad inversa, resuelto cuantitativamente hace más de 200 años por el Reverendo Thomas Bayes.

Richard M. Royall, destacado epistemólogo y estadístico de la Universidad Johns Hopkins, escribía recientemente: «la estadística tiene la responsabilidad de responder una pregunta fundamental: ¿cuándo es correcto afirmar que un conjunto de observaciones aporta evidencia a favor de una hipótesis y en contra de otra? Los recursos para dar respuesta a esta pregunta se conocen desde hace más de un siglo;

sin embargo, ni la pregunta ni su respuesta aparecen, salvo aisladas excepciones, en los textos estadísticos modernos» [ROYALL, 1997, p. xi].

Valga esta cita, que para muchos puede resultar en extremo sorprendente por su carácter esencial y a la vez abarcador, como preámbulo de la crónica que nos ocupa; el presente trabajo se propone, precisamente, dar cuenta de los avatares del enfoque bayesiano y sus sucedáneos.

Luego de situar en contexto histórico la propuesta de Bayes y exponer su naturaleza desde el punto de vista epistemológico, nos ocupamos de bosquejar algunos elementos del debate que el siempre inquietante concepto de probabilidad produjo especialmente a lo largo del siglo XIX. La historia bayesiana siempre ha estado asociada a dicho debate, desde la adhesión inicial que consiguió por parte de figuras fundadoras de la teoría probabilística clásica como fue el caso de Laplace (para quien «la probabilidad mide nuestra expectativa subjetiva de la fuerza de una convicción» [NAGEL, 1939, p. 18], aunque su cálculo se asentara en el análisis combinatorio), pasando por las críticas que concitó luego desde la perspectiva frecuentista («la probabilidad es una medida de la frecuencia relativa con la cual una propiedad ocurre en una clase especificada de elementos») que tomara fuerza de la mano de figuras como Ellis y Venn, hasta el enfoque subjetivo del concepto (el que permite hablar, por ejemplo, de que «la probabilidad de que el axioma de Goldbach sea falso es muy alta», o «probablemente el partido X ganará las próximas elecciones») y que fue objeto de formalización matemática en el marco de la estadística por figuras de la talla de Savage (véase, por ejemplo, [SAVAGE, 1954]). En este contexto se da cuenta del advenimiento del largo periodo durante el cual la estadística emplea procedimientos que, a la luz del examen que permiten las varias décadas transcurridas desde su implantación, hoy parecen claramente insuficientes. Y en ese punto se da cuenta de esa especie de resurrección del pensamiento bayesiano que hoy concita creciente atención en el mundo de la estadística aplicada.

### Bayes y la inferencia inductiva

En 1838, en el prefacio de su ensayo sobre probabilidades *The Cabinet Cyclopaedia*, el célebre probabilista Augustus de Morgan escribió:

«Había también otra circunstancia que era válida para los primeros investigadores; esto es, no haber considerado, o al menos no haber descubierto, el método de razonamiento que

va de la ocurrencia de un evento a la probabilidad de una u otra causa [...] Dada una hipótesis que entraña la necesidad de una u otra consecuencia (dentro de cierto conjunto no muy grande de ellas), ellos podían determinar la probabilidad de que ocurriera cualquiera de estas consecuencias. Pero dado que un evento ocurrió, lo cual puede ser explicado por diversas causas o ser explicado por alguna de las diferentes hipótesis, ellos no podían inferir la probabilidad con la cual ese acontecimiento que estaban observando pudiera ser explicado por dichas hipótesis. Pero, tal y como ocurre en la filosofía natural, optar por una hipótesis a partir de hechos observados es siempre anterior a cualquier intento de hacer descubrimientos deductivamente; por lo tanto, en la aplicación de la noción de probabilidad a los hechos reales de la vida, el proceso de razonamiento que va de los hechos observados a sus antecedentes más probables debe ser previo al uso directo de cualquiera de tales antecedentes, causas, hipótesis o como se les quiera llamar. Estos dos obstáculos (la dificultad matemática y el deseo de contar con un método inverso), obstaculizaron que la ciencia extendiera su mirada más allá de los simples problemas encarnados por los juegos de azar [...] Esa forma de razonar fue usada por primera vez por el Reverendo Thomas Bayes, y el autor, aunque ahora está casi olvidado, merece el más honorable recuerdo de todos aquellos que escriban sobre la historia de la ciencia».

Este texto, tomado de una de las obras más notables de Ronald Fisher [FISHER, 1949, pp. 4-5], puede resultar algo enrevesado debido a su estilo ampuloso (nótese que fue escrito hace ya un siglo y medio); pero puede ser más fácilmente comprendido cuando se advierte que el «método inverso» al que alude de Morgan no es otro que la inferencia inductiva. Efectivamente, tal forma de inferencia permite la inducción probabilística; es decir, a partir de los acontecimientos observados se trata de establecer juicios probabilísticos sobre las hipótesis que pudieran explicarlos; esto sólo es posible a través del teorema de Bayes, pues toda la teoría frecuentista es hipotético deductiva (estima la probabilidad de los hechos observados suponiendo para ello la validez de una hipótesis).

Afortunadamente la historia no ha olvidado a Bayes. Se conoce que nació en Londres en 1702 y falleció el 17 de abril de 1761 en Tunbridge Wells, Kent. Su padre fue uno de los primeros seis ministros disidentes ordenados en Inglaterra. Thomas fue educado por un maestro privado y, aunque no se conoce nada de su tutor, se cree que pudo haber sido Abraham De Moivre, el autor del famoso libro *La doctrina de las probabilidades*. Bayes fue ordenado, al igual que su padre, como ministro disidente, y en 1731 se convirtió en ministro de la iglesia presbiteriana en Tunbridge Wells; aparentemente trató de retirarse en 1749, pero continuó ejerciendo hasta 1752, y permaneció viviendo en ese lugar hasta su muerte [O' CONNOR, ROBERTSON, 2000]. Sus primeros estudios son de tipo teológico (en Edimburgo entre 1719 y 1722),

pero no demoró en interesarse en las matemáticas, las ciencias naturales y por último la inferencia estadística [HALD, 1998, p. 112].

Se ha dicho que el célebre ensayo de Bayes era en parte una respuesta a los trabajos del filósofo escocés David Hume, fundamentalmente a su conocido *Tratado de la naturaleza humana* publicado por primera vez en 1739 y también a otros como *Sobre los milagros*. En la obra de Hume se plantea una consistente crítica a la inducción, encarando con escepticismo las *generalizaciones*: ¿Puede algún número de instancias observadas, a falta de un examen completo, hacer que sea razonable creer en una generalización? [HACKING, 1995a, pp. 214-215]. Este pensamiento se resume en lo que se ha venido a llamar el *problema escéptico* acerca del futuro, que se percibe bien en sus propias palabras: « [...] el argumento inductivo no porta una fuerza lógica; es decir, tal argumento no representa más que una *suposición* de que ciertos eventos seguirán en el futuro el mismo patrón que tenían en el pasado» [ROTHMAN, GREENLAND, 1998, pp. 17]. La preocupación por los milagros era en aquella época una cuestión de vida y de ultratumba, y se consideraba que el pueblo podía estar bajo sus efectos, algo que era inaceptable entre las personas ilustradas. Para éstos, por tanto, era de suma importancia dirimir si determinados testimonios, entonces en boga, constituían o no evidencias reales acerca de la existencia de ciertos fenómenos en principio poco creíbles. Se necesitaba establecer procedimientos que permitieran enjuiciar la veracidad implícita en las afirmaciones de una serie de testigos. Ocasionalmente se tenía una información procedente de alguien que era digno de cierta credibilidad, quien informaba sobre el testimonio de otro testigo cuya credibilidad también era imperfecta. La probabilidad que cupiese atribuir a «realidades» así testimoniadas se presentaba entonces como «el escudo con el que un hombre podía guardarse del entusiasmo» [HACKING, 1995b, pp. 134-135]; es decir, como un recurso que cuantificaba el grado de aceptación que pudiera conferirsele a las afirmaciones sobre acontecimientos aparentemente milagrosos.

Se reconoce a Bayes como el primero que captó la importancia de desarrollar una teoría cuantitativa y exacta de razonamiento inductivo. Se dice que desarrolla su teorema para demostrar la existencia de Dios pero, como al final necesita conocer la probabilidad a priori de la existencia de Dios, no pudo demostrar dicha existencia y abandona el trabajo [COBO, 1993, pp. 59-60]. Lo cierto es que, si bien pudo percibir el problema y desarrollar un teorema por cuyo conducto se soluciona el problema de la inferencia inversa en el ámbito de la teoría de probabilidad matemática, fue sufi-

suficientemente cauteloso sobre su validez y aplicabilidad como para retener la publicación de todos sus trabajos mientras sus dudas no fueran disipadas.

El teorema de Bayes produce una probabilidad conformada a partir de dos componentes: una que ocasionalmente (pero no siempre) se delimita subjetivamente, conocida como «probabilidad a priori», y otra objetiva, la llamada verosimilitud, basada exclusivamente en los datos. A través de la combinación de ambas, el analista, por ejemplo, emite un juicio de probabilidad sobre una hipótesis que sintetiza su nuevo grado de convicción al respecto. Esta probabilidad *a priori*, una vez incorporada la evidencia que aportan los datos, se transforma así en una probabilidad *a posteriori*.

El planteamiento esencial de teorema de Bayes es formalmente el siguiente [MORGAN, 1971, pp. 19-20]: sea un conjunto de  $R$  sucesos mutuamente exclusivos y complementarios  $r_i$ ,  $i=1,2,\dots,R$ , y sea un segundo conjunto de sucesos del mismo tipo  $s_j$ ,  $j=1,2,\dots,S$ . Entonces para  $r_i$  y  $s_j$  particulares cualesquiera,

$$P(r_i | s_j) = \frac{P(s_j | r_i)P(r_i)}{\sum_{h=1}^R [P(s_j | r_h)P(r_h)]} \quad [1]$$

donde  $P(a|b)$  es la «probabilidad de  $a$  dado  $b$ » y es la probabilidad condicional de que se produzca el suceso  $a$  supuesto que rige la condición  $b$ .

A modo de ilustración, supongamos que se tienen dos dados, uno de 4 caras y otro de 20. Llamemos 1, 2, 3, 4 a las caras del primero y 1, 2, ... 20 a las del segundo. Consideremos cada dado por separado y supongamos que, al lanzarlos, sus caras son igualmente probables. Ahora supongamos que se escoge al azar uno de los dados; si es el que tiene 4 caras lo llamamos evento  $F$ , si es el que tiene 20, evento  $\bar{F}$ . Como es obvio:  $P(F) = P(\bar{F}) = 1/2$ . Se lanza el dado escogido; si el resultado fuera 13 no se necesitaría de la regla de Bayes para conocer que el dado escogido es el que tiene 20 caras, pues  $P(F|\text{salió } 13) = 0$ .

Pero si, por ejemplo, el resultado fue 3, ¿cuál de los dos fue el dado escogido? Es imposible saberlo con certeza. La verosimilitud de  $F$  es  $P(3|F) = 1/4$  y la verosimilitud de  $\bar{F}$  es  $P(3|\bar{F}) = 1/20$ . Al sustituir estos valores en la fórmula de Bayes se obtiene:

$$P(F|3) = \frac{P(3|F)P(F)}{[P(3|F)P(F) + P(3|\bar{F})P(\bar{F})]} = \frac{5}{6}$$

Como vemos, una simple operación algebraica nos permite conocer cuál es la probabilidad de cada uno de los dados (hipótesis) [BERRY, 1996, pp. 149-150].

En realidad, no cabe discutir la validez del teorema de Bayes en tanto ecuación matemática; de hecho, se emplea cotidianamente como fundamento para el análisis de diversas situaciones que se presentan en los juegos de azar y en la valoración de las pruebas de pesquijaje médico. Sin embargo, como modelo acerca del modo de operar científicamente, el teorema de Bayes ha sido criticado debido a que requiere asignar una probabilidad *a priori* a la verdad de la afirmación que se valora, un número cuyo significado científico objetivo es a veces borroso, y siempre sujeto a controversia. Se especula que ésta debe ser la razón por la cual el Reverendo Bayes «eligió la más terrible entre las opciones *publicar o perecer*» [GOODMAN, 1999, pp. 995-1004]. Sin embargo, a nuestro juicio, esto no tiene mayor sentido, pues la legitimidad (a nuestro juicio indiscutible) del empleo de probabilidades subjetivamente determinadas es un asunto que merece una discusión por sí misma mientras que, por otra parte, el teorema de Bayes como tal no tiene ningún compromiso especial con el origen que tengan las probabilidades con que opera.

Bayes fue distinguido como *Fellow* de la Royal Society en 1742. Hasta ese momento no había dado publicidad a trabajo alguno bajo su nombre: *Divine Benevolence: Or, an attempt to prove that the principal end of the divine providence and government is the happiness of his creatures*, en 1731 y *An introduction to the Doctrine of Fluxions and a defense of the mathematicians against the objections of the author of the analyst*, de 1736, fueron publicados anónimamente, aunque probablemente era conocida su autoría [HALD, 1998, p. 112].

Bayes lega 100 libras a su amigo Richard Price y la encomienda de examinar sus papeles matemáticos. Price es conocido por los historiadores como el escritor de un folleto que constituía un elogio de la Revolución Francesa, predicador, además, contra la guerra en América y a quien se debe la designación misma de «Estados Unidos de América»; y por los probabilistas, precisamente, como el hombre que presentó al público el famoso ensayo de Thomas Bayes, divulgado en 1763 [HACKING, 1995b, pp. 82-83]. El artículo fue enviado a la *Royal Society* con la siguiente nota:

«Le envío ahora un ensayo que encontré entre los papeles de nuestro difunto amigo Mr. Bayes, el cual, en mi opinión, tiene un gran mérito... En una introducción que bosquejé para este ensayo dice que su diseño, cuando empezó a pensar en el tema, fue hallar un método por conducto del cual pudiéramos juzgar acerca de la probabilidad de que ocurra cierto

evento, dadas ciertas circunstancias, suponiendo que no conocemos nada concerniente a él, salvo que, bajo esas circunstancias, él hubiera ocurrido un cierto número de veces, y hubiese fracasado otro cierto número de veces» [O' CONNOR ROBERTSON, 2000].

El artículo en el cual Bayes expone su teoría sobre probabilidades se titula *Ensayo hacia la solución de un problema en la doctrina del azar (Essay towards solving a problem in the doctrine of chances)* y fue publicado en *Philosophical Transactions of Royal Society of London* en 1763.

### La probabilidad como expresión de lo subjetivo y como síntesis de lo objetivo

Probablemente fue Condorcet, quien interesó a Laplace en las probabilidades como disciplina, lo cual lo convierte en el padrino de la moderna teoría de las probabilidades. Las conclusiones de Bayes fueron aceptadas por Laplace en una memoria de 1781. A partir de ese trabajo de Laplace, Condorcet conoció el modo de inferencia propuesto por Thomas Bayes, y luego lo empleó como el principal instrumento de su análisis sobre los procedimientos de votación.

Ian Hacking, en un libro ciertamente excepcional [HACKING, 1995b, pp. 144-145], se refiere ampliamente a los dos aspectos que caracterizan el concepto de probabilidad: *frecuencia y grado de credibilidad*. Según allí se expone, Poisson, en su obra sobre la ley de los grandes números de 1837, y Cournot, en el último libro de ciencia moral de la ilustración publicado seis años más tarde, aclaran que posibilidad y probabilidad debían ser apreciados como dos nociones diferentes; por ello utilizaron las palabras francesas *chances* y *probabilités* para designarlas. La probabilidad aludía al grado de razonable credibilidad: «La probabilidad de un suceso es la razón que tenemos para pensar que ese suceso ocurrió u ocurrirá»; era, por tanto, reflejo de una impresión subjetiva. Pero *chance* aludía a una propiedad objetiva del suceso, la «facilidad» con que pueda ocurrir: «De suerte que un suceso tendrá, por su naturaleza misma, una *chance* mayor o menor, conocida o desconocida». *Probabilité* era una noción relativa a nuestro conocimiento y a nuestra ignorancia, subjetivamente determinada al margen de formales constataciones empíricas, mientras que la frecuencia objetiva de varios desenlaces posibles determinaban los *chances*. Poisson fue cuidadoso al distinguir entre lo «objetivo» y «subjetivo» en materia de probabilidad (o, como más se conoce: frecuencia frente a creencia); en su opinión se trataba de distinguir entre *modelar* e *inferir*. Cuando modelamos procesos desde el punto de vista de la probabilidad, suponemos que existe alguna característica objetiva de las cosas que

las hace comportarse como, digamos, una urna de la cual se extraen bolillas de color con una frecuencia relativamente estable. Cuando inferimos utilizando probabilidades estamos llegando a conclusiones de cuya verdad no estamos completamente seguros, debido a la inexorable participación de un elemento subjetivo. Curiosamente pasó inadvertido tanto el comentario de la carta de Price a John Canton (10 de noviembre de 1763) donde en el último párrafo apunta las ideas de Bayes sobre las palabras *chance* y *probability*, como lo más importante para nosotros y que denota lo que verdaderamente pensaba Bayes sobre estas disputas a la hora de utilizar un término u otro, que también ocurrían en su época. En su trabajo deja claro a nuestro entender una idea de síntesis de lo objetivo y de lo subjetivo, cuando en la Sección I, en la definición 1 en el punto 6 dice: «By *chance* I mean the same as probability» (pongo usar *chance* con el mismo sentido que probabilidad). De hecho en el resto del trabajo de Bayes no aparece más la palabra *chance* [BAYES, 1763, p. 299].

Para Augustus de Morgan «La probabilidad es una sensación del espíritu, no la propiedad inherente a una serie de circunstancias». De Morgan afirmaba que la «probabilidad investiga las reglas de conformidad con las cuales el valor de nuestra creencia en una proposición varía con el valor de nuestra creencia en otras proposiciones con las cuales aquélla tiene relación» [HACKING, 1995b, pp. 186].

A la altura de la década de 1830 ya el mundo no era como en la época de Laplace, el aspecto «objetivo» de la probabilidad pasó a ser mucho más importante que el «subjetivo», sencillamente porque se tenía acceso a muchas más frecuencias. Sin lugar a dudas, el «alud de números impresos», que había comenzado en la década anterior con las *Recherches statistiques* de París y del departamento del Sena, dirigidas por Joseph Fourier, provocaron el cambio a favor de la idea objetiva [HACKING, 1995b, pp. 145-146].

Leslie Ellis, un connotado matemático de Cambridge, afirmó en 1842: «En cuanto a mí, después de haber prestado un penoso grado de atención a este punto, comprobé que era incapaz de separar el juicio de que un suceso tiene más probabilidad que otro de ocurrir (o que se espera con preferencia a otro) de la creencia de que a la larga tal suceso ocurrirá con más frecuencia» [HACKING, 1995b, pp. 186-187].

John Venn, considerado como el creador del enfoque frecuentista de la probabilidad (cuya obra *La lógica del azar* fue publicada 24 años después de la afirmación de Ellis), opinaba que «La concepción fundamental es la de una serie que combina la

regularidad individual con la regularidad colectiva». Para él, la probabilidad de cierto acontecimiento no tiene significado alguno sino en relación con una serie de resultados que lo contengan como posible desenlace; toda probabilidad ha de referirse a una serie y la probabilidad de un suceso es su frecuencia relativa dentro de la serie [HACKING, 1995b, pp. 185-186].

Las técnicas de Bayes se mantuvieron sin que su utilidad fuera puesta mayormente en duda hasta que George Boole, creador del álgebra binaria, las cuestionó en su célebre tratado *Leyes del pensamiento (Law of Thought)* publicado en 1854. Esta obra fue un golpe de gracia, pues vino a consolidar una tendencia frecuentista, de manera que desde entonces, los seguidores del enfoque bayesiano se comportaron cada vez con más frecuencia como una secta, a veces «ofendidos» por la incomprensión de quienes no la integraban, enclaustrados en cierta endogamia teoricista y sobreviviendo en un coto exclusivista, alejado del mundo de las aplicaciones.

### **Crítica de Fisher al bayesianismo y aportes alternativos para suplirlo**

El ilustre estadístico Ronald Fisher no creía que el uso de la fórmula de Bayes estuviera justificado en las investigaciones científicas, en las que generalmente es incierta la probabilidad a priori [GOODMAN, 1995, p. 141-151]. Enumeraba tres consideraciones que, según él, fundamentan por qué no se debe asumir para la aplicación práctica la verdad del «axioma de Bayes», como él le llamaba [FISHER, 1949, pp. 6-7]. En sus propias palabras, tales endebles son:

- a) El axioma conduce a contradicciones matemáticas evidentes. Al tratar de superar esas contradicciones, los defensores de la probabilidad inversa parecen forzados a considerar la probabilidad matemática, no como una cantidad objetiva medida por las frecuencias observadas, sino como una medición meramente determinada por tendencias psicológicas, lo cual es inútil para propósitos científicos.
- b) Es inherente a la naturaleza de un axioma que su verdad debe ser evidente para cualquier mente racional que haya captado su significado. El axioma de Bayes ha sido sin dudas completamente captado por muchas mentes racionales, incluyendo la de su autor, pero sin que ello haya supuesto la convicción de que sea necesariamente cierto.
- c) La probabilidad inversa ha sido escasamente empleada para justificar las conclusiones a partir de un experimento, aunque la teoría ha sido ampliamente divulgada y está extensamente tratada en la literatura sobre probabilidades. Cualesquiera sean las razones que le dan confianza a los experimentadores en el sentido de que puedan

obtener conclusiones válidas a partir de sus resultados, ellos parecen actuar con igual convicción, hayan o no oído hablar de la teoría de probabilidad inversa».

En realidad, los argumentos de Fisher en contra del teorema de Bayes son, en nuestra opinión, poco convincentes. La probabilidad *a priori* que se utiliza en el teorema, suele estar, en efecto, determinada de forma «subjetiva». Pero esta subjetividad no obedece solamente a «tendencias psicológicas», sino que, en buena ley, ha de derivarse también de la experiencia y de los conocimientos previos de los investigadores. Por otra parte, resulta inadecuado descalificar el carácter científico de un cuerpo teórico por la única razón de que en su conformación intervenga un elemento subjetivo, pues la ciencia es permeable y, con mucha frecuencia, de manera inevitable, a la impronta personal de sus creadores y sus intérpretes. En otro sitio [SILVA, 1997, pp. 263-264] hemos desarrollado y fundamentado detalladamente no sólo esta realidad sino también la idea de que tal componente subjetivo, si se asume con responsabilidad y rigor, puede ser más productivo que muchas conductas vertebradas en torno a una objetividad a ultranza que es, en el fondo, quimérica.

Por otra parte, no queda del todo claro por qué Fisher clasifica al teorema de Bayes de «axioma», pues estos términos no son equivalentes. Quizás ello se deba a que la versión más simple del teorema se deriva casi directamente de la definición de probabilidad condicional. Su segundo argumento contradice la afirmación de que el teorema de Bayes es indiscutible desde el punto de vista matemático.

Finalmente, cabe apuntar que el hecho de que la probabilidad inversa no sea ampliamente usada no puede esgrimirse como una endebles intrínseca del procedimiento; en última instancia, son los investigadores los que podrían estar desdeñando esta herramienta, en muchos casos, quizás, por mera inercia o porque desconocen sus potencialidades.

Ronald Fisher es considerado uno de los padres de la estadística moderna. Nació en Londres en 1890 y murió en Australia en 1962; desde niño mostró un talento singular para las matemáticas, ganando varios premios. Se mostró muy interesado por la teoría de la evolución y la genética, y combinó estos intereses con el que le despertaba el movimiento eugenésico, hoy muy desacreditado por su carácter reaccionario, pero acreedor entonces de notables influencia e implantación. La corriente eugenésica alentaba a la procreación por parte de aquellos que poseyeran «calidades excelentes». Según esta teoría, las personas más inteligentes saben que para mejorar sus oportunidades en la vida y sus condiciones económicas deben tener pocos hijos, lo cual

conducía a un «desperdicio genético». Fisher proponía que el estado subsidiara a las personas «adecuadas» para así alentarlos a procrear, y a la postre, supuestamente, mejorar la especie. Se mantuvo como miembro activo de la Sociedad Eugenésica Londinense a lo largo de 20 años y tuvo él mismo muchos hijos [STOLLEY, 1991, pp. 416-425].

Su influencia en la estadística y el diseño experimental fue crucial; entre sus aportes procede destacar especialmente los siguientes tres: el desarrollo del análisis de la varianza, el principio de la aleatorización y la idea de la replicación. Además de estas tres contribuciones medulares, hizo un notable servicio a la comprensión de los factores de confusión. Más puntualmente, resolvió el problema de establecer la distribución muestral del coeficiente de correlación descubierto por su contemporáneo (y enemigo) Karl Pearson, así como de otros estadígrafos de uso común, tales como el coeficiente de correlación múltiple. En 1925 publicó *Statistical Methods for Research Workers*, y 10 años más tarde *The Design of Experiments*. En 1938, junto a Frank Yates, publicó *Statistical Tables for Biological, Agricultural and Medical Research*, usadas en la práctica hasta hace muy pocos años [STOLLEY, 1991, pp. 416-425].

A pesar de las numerosas e importantes contribuciones de Fisher a la estadística, nunca tuvo un nombramiento académico como estadístico, pues siempre rechazó trabajar en lo que sería su lugar natural, el departamento que dirigía Karl Pearson, debido a las desavenencias surgidas entre ambos a raíz de una crítica que hiciera Pearson a uno de sus trabajos, antipatía que más tarde también profesó a Egon, el hijo de Karl [STOLLEY, 1991, pp. 416-425].

Fisher mostró un constante interés por crear un método cuantitativo de naturaleza objetiva para facilitar el proceso de inferencia inductiva por conducto del cual se pudieran derivar conclusiones a partir de las observaciones [GOODMAN, 1995, pp. 141-151]. Concebido como alternativa al planteamiento bayesiano, su enfoque comienza por establecer una hipótesis estadística nula según la cual los grupos comparados son esencialmente equivalentes. Una vez que se obtiene la diferencia observada entre los grupos, se calcula el valor  $p$  [FEINSTEIN, 1998, pp. 355-360]. Si bien es cierto que Fisher no fue el primero que lo usó, cabe destacar que fue él quien elaboró un esquema formal de la lógica en que se basa su empleo y quien explicó los métodos para calcularlo en muy diversas situaciones. Su definición del valor  $p$  o «probabilidad de significación» es esencialmente la misma que se emplea en la actualidad. Se definió como la probabilidad, bajo el supuesto de no efecto o no diferencia (hipótesis nula), de obtener un resultado igual a, o más extremo que, lo efectivamente observado. Si

este valor era pequeño, se podía «rechazar» la posible validez de la hipótesis nula. Al uso de un valor  $p$ , en comparación con una cota por debajo de la cual se establezca el rechazo, se le llamó *prueba de significación* [GOODMAN, 1995, pp. 141-151].

Fisher plantea que la hipótesis nula nunca puede ser demostrada, pero sí refutada en el curso de un experimento; por tanto, un experimento nos puede conducir a desaprobársela, pero es inútil para valorar la hipótesis opuesta, que aunque pueda ser razonable, no es elegible como hipótesis nula debido a que es inexacta [FISHER, 1949, p. 16].

Este planteamiento era coherente con la corriente filosófica hipotético-deductiva encabezada en esa época por Popper. La idea de la «falsabilidad» presentada por este filósofo, se vertebra sobre el hecho de que siempre es posible refutar una afirmación que sea falsa (basta un contraejemplo), pero no es posible demostrar su veracidad sobre la base de corroborar con casos particulares que tal afirmación se cumple. La sugerencia de Popper en lo metodológico consiste en formular una hipótesis dada (una afirmación acerca de cómo funciona la naturaleza), predecir lo que ha de observarse si tal hipótesis fuera verdadera y procurar su refutación mediante constataciones incompatibles con dicha predicción. Al establecer la hipótesis, esta corriente de pensamiento concede, incluso, más interés a la que más se aleje de lo que el conocimiento vigente permite esperar, ya que, si no pudiera refutarse, tal resultado sería mucho más informativo. Como es lógico, la negación de una hipótesis explicativa bien definida entraña varias explicaciones alternativas, pero no identifica a ninguna de las muchas otras explicaciones que pueda tener el hecho de haberla observado [SILVA, 1997, pp. 163-165] y ahí radica la limitación fundamental del pensamiento popperiano, pero lo cierto es que éste, al igual que la prueba de hipótesis de Fisher, pone el énfasis en la fuerza que tiene la *refutación* en tanto prueba de algo y como contrapartida de la *corroboración*, mera constatación que no pasa de dar un aval débil a nuestra convicción de que ese algo es cierto.

Según Fisher, si se empleaba un nivel de «significación», éste no debería ser un umbral rígido sino que debería depender del conocimiento previo respecto del fenómeno analizado [FISHER, 1949, p. 12]. Pero lo cierto es que tal advertencia fue rápidamente olvidada, y lo que se generalizó fue el uso de 0.05, al punto de considerar como no significativo a un valor  $p$  igual a 0.051 mientras que, si fuera igual a 0.049, ya se considera significativo [CHIA, 1997, pp. 152-154].

La historia, sin embargo, atribuye al propio Fisher el establecimiento del 0.05, y para ello recurre a una anécdota divertida. En palabras de Chia [CHIA, 1997, pp. 152-154]: «Se especula que Fisher estaba en su bañera meditando muy perplejo sobre este problema. Mientras estaba absorto en sus pensamientos tiró del tapón accidentalmente y el agua comenzó a disminuir en la bañera. Cuando el nivel del agua descendió, él vio que sobresalían los cinco dedos de su pie. Esta observación decidió el asunto: ¡la significación estadística se fija al nivel 5%! Se podría pensar que si Fisher fuera más observador lo habría fijado al 10%, a menos que tuviera una sola pierna o alguna deformidad en el otro pie».

### Procedimiento deductivo de Neyman y Pearson

Los famosos valores  $p$  de Fisher fueron cuestionados, sin que mediara mucho tiempo, por algunos investigadores y por muchos estadísticos debido a las endeblesces que afectan a sus bases lógicas y, sobre todo, a la difusa utilidad que tenían. Su papel era congénitamente orientativo, sin que se estableciera algorítmicamente qué hacer con esos números. La crítica más reiterada ha sido que constituye una medida que depende vitalmente de un elemento exógeno a los datos: el tamaño de muestra. A un pequeño efecto observado en un estudio con un tamaño de muestra grande puede corresponder el mismo valor  $p$  que a un gran efecto registrado a través de una muestra pequeña [SILVA, 1999, p. 64].

Los matemáticos Jerzy Neyman y Egon Pearson consideraron que los valores  $p$  tal y como los concebía Fisher constituían una respuesta incompleta al problema de desarrollar, sin usar el teorema de Bayes, un método inferencial. Empeñados igualmente en prescindir del recurso bayesiano, en 1928 publicaron un documento clave referente a las bases teóricas de un procedimiento al que llamaron *prueba de hipótesis*. Según este enfoque, deben plantearse dos hipótesis para decidirse por una de ellas: una nula (usualmente una afirmación según la cual no hay efecto debido a cierto tratamiento o condición), y una hipótesis alternativa, la cual usualmente se opone a la nula (por ejemplo, que el efecto no es cero). A dichas hipótesis se les denota por lo general  $H_0$  y  $H_1$  respectivamente. La idea de la *hipótesis alternativa* aparece así por primera vez.

El resultado de este proceso fue un procedimiento que, en lugar de una forma de hacer inferencia, pasó a ser un *modo de tomar decisiones*: rechazar una hipótesis y aceptar la otra, solamente sobre la base de los datos. Esto coloca a los investigadores frente a la posibilidad de cometer dos tipos de error: elegir  $H_0$  siendo cierta  $H_1$ , o seleccionar  $H_1$  cuando es falsa. En el marco clínico, por ejemplo, podría tratarse de

afirmar que dos terapias difieren cuando realmente son equivalentes (conocido también como resultado falso positivo o error tipo I), o concluir que son iguales cuando en realidad difieren (también conocido como resultado falso negativo o error tipo II). Este enfoque tiene el atractivo de que, si asumimos una verdad subyacente, la probabilidad de cometer cada uno de estos errores puede ser calculada con fórmulas matemáticas, deductivamente, y por lo tanto «objetivamente» [SILVA, 1999, p. 65].

Estas tasas de error definen una llamada «región crítica». Si el valor del estadístico de la prueba se sitúa en dicha región, se acepta la hipótesis alternativa y se rechaza la hipótesis nula; en su defecto, se decide lo contrario. Como vemos, la afirmación «la hipótesis nula nunca se acepta, sólo es posible no rechazarla» es típica de la prueba de significación de Fisher y no de la prueba de hipótesis tal y como fue formulada por Neyman y Pearson.

Según este enfoque no se opera con ninguna medida de la evidencia: ningún número sale de los datos con vistas a respaldar o no las hipótesis. Después del experimento, lo único que hay que indicar es si el resultado se sitúa o no en la región crítica [GOODMAN, 1995, pp. 141-151]. La razón para esta ausencia en el enfoque de Neyman y Pearson era que cualquier elemento inductivo conduciría inevitablemente al controvertido teorema de Bayes. Por lo tanto, ellos proponen otro objetivo para la ciencia: no razonar inductivamente a partir de un único experimento sino usar métodos deductivos para limitar el número de errores que se cometen cuando se realizan muchos experimentos diferentes. En sus palabras:

«[...] Ninguna prueba basada en la teoría de las probabilidades puede por sí misma proporcionar alguna evidencia valiosa sobre la veracidad o falsedad de una hipótesis. Pero podemos ver el propósito de estas pruebas desde otro punto de vista. Sin esperar conocer si cada hipótesis por separado es verdadera o falsa, podemos buscar las reglas que gobiernan nuestro comportamiento con respecto a ellas, siguiendo las cuales podemos asegurar que, en una frecuencia grande de repeticiones experimentales, no estaremos frecuentemente equivocados» [NEYMAN PEARSON, 1933, pp. 289-337].

Esta afirmación es notable. Según los propios Neyman y Pearson, el precio que tenemos que pagar para disfrutar los beneficios de la objetividad es abandonar nuestra capacidad para medir evidencias, o juzgar verdades, en un experimento individual. En la práctica esto significa que lo que se puede comunicar si los resultados son estadísticamente significativos o no, y actuar de acuerdo con ese veredicto [GOODMAN, 1999, pp. 995-1004].

Por supuesto, esta propuesta alternativa, que empleaba pero a la vez desvirtuaba el papel de su propia invención, fue muy irritante para Fisher, quien por otra parte había desarrollado una profunda antipatía por Neyman y Pearson desde hacía mucho tiempo [STOLLEY, 1991, pp. 416-425]. Pero lo esencial es que la diferencia entre su valor  $p$  y las pruebas de hipótesis, a las que denominó burlescamente «procedimientos de aceptación» y «funciones de decisión», no sólo era matemática, sino que entrañaba diferentes conceptos acerca del papel de la inferencia estadística en la ciencia, a pesar de que ambos enfoques se inscribían en el mismo propósito: sacar conclusiones de manera «exclusivamente» objetiva.

La prueba de hipótesis representó un cambio dramático con respecto a los métodos anteriores, ya que se trata de un proceder que esencialmente dictaba las acciones del investigador. Matemática y conceptualmente fue un enorme paso de avance, pero como modelo de la práctica científica era problemático. Sin lugar a dudas, no hubiera logrado su actual grado de aceptación si no se hubieran amalgamado con los valores  $p$  de Fisher [GOODMAN, 1999, pp. 995-1004].

En efecto, a pesar de que el valor  $p$  propuesto por Fisher y el método de prueba de hipótesis de Neyman y Pearson eran incompatibles, en el altar del pragmatismo se sacrificó la individualidad de cada enfoque y éstos se unieron de manera anónima en la década del 40 para dar lugar a lo que se conoce hoy como *prueba de significación* (o también, *prueba de hipótesis*, pues ambos términos se usan indistintamente), en lo que constituye un entramado que erróneamente se considera como un único y coherente enfoque de inferencia estadística. Tal síntesis consiste esencialmente en calcular  $p$  y declarar que «la diferencia es significativa al nivel  $p$ » sin hacer declaración explícita de ninguna hipótesis alternativa. A esta simbiosis, y para distinguirla de sus dos fuentes nutricias, Royall [ROYALL, 1997, p. 73] la denomina «pruebas de rechazo» (*rejection trials*).

### Reaparece Bayes

La metodología de las pruebas de hipótesis vive actualmente una crisis notable. Muchos investigadores han alertado sobre sus limitaciones conceptuales y prácticas, que son básicamente las siguientes: a) que el rechazo de la hipótesis nula queda asegurado con un tamaño de muestra suficientemente grande, b) que no toma en cuenta la información proveniente de estudios previos o de la experiencia empírica informalmente acumulada y c) que maneja el resultado como un modo de tomar

decisiones dicotómicas sobre las hipótesis, en vez de ayudarnos a valorar la credibilidad de ellas [GOODMAN, 1999, pp. 995-1004]. Va resultando crecientemente obvia la necesidad de un «nuevo» paradigma: irónicamente, éste no es otro que el viejo enfoque bayesiano, arduamente combatido. Afortunadamente, hay un movimiento creciente en la comunidad estadística orientado a recuperar los métodos bayesianos dentro de la educación estadística básica; junto con la introducción de la lógica probabilística en la enseñanza epidemiológica y los avances en el campo computacional, tal movimiento apresuraría el día en que la escasa familiaridad con este proceder deje de suponer un obstáculo importante [GREENLAND, 1998, pp. 322-332].

El enfoque bayesiano tiene varios atractivos. El más persuasivo es que invita a incorporar formalmente evidencias de experiencias y experimentos previos en el proceso que conduce a las conclusiones. De manera que, a la hora de obtenerlas, el investigador no se circunscribe a la experiencia u observación puntual que ha realizado sino que incorpora la información precedente de manera explícita en el análisis [BERRY, 1996, p. iii].

Permite, además, conformar afirmaciones inferenciales intuitivas que no se pueden hacer con la estadística convencional. Por ejemplo, un intervalo al 95% de confianza se entiende comúnmente como un recorrido de valores dentro de los cuales se encuentra el valor verdadero con una probabilidad del 95%. Esta interpretación, incorrecta en la estadística frecuentista, es válida, sin embargo, bajo el enfoque bayesiano [BURTON, GURRIN, CAMPBELL, 1998, pp. 1318-1323].

Finalmente, es más específico en el sentido de que si la magnitud de un efecto estimado está constreñida por una distribución a priori, será menos probable que la estimación aparezca como sustantiva o estadísticamente significativa que con un análisis convencional. En este sentido, el análisis bayesiano proporciona un recurso más exigente que la prueba de hipótesis [GREENLAND, 1998, pp. 322-332]. También, es en nuestra opinión, más honesto, puesto que la aparición del elemento subjetivo es explícita y transparente, y no oculta y subrepticia como en el enfoque frecuentista, donde se actúa como si se pudiera prescindir de lo subjetivo, cuando en realidad este componente aparece en uno u otro punto, aunque sea de contrabando.

En general, todo el proceso intelectual asociado a la inferencia bayesiana es mucho más coherente con el verdadero modo usual que emplea el científico que los recursos clásicos [ROYALL, 1997, p. 167; GUTIÉRREZ, 1994, p. 217]. Al igual que

hace un médico al admitir con naturalidad que tiene un criterio a priori sobre un paciente, realizar exámenes complementarios y actualizar su visión sobre éste al con-  
jugar las dos cosas, los investigadores pueden encarar sus estudios haciendo uso del  
enfoque bayesiano. Ello explica que para la última década del siglo XX la base de  
datos del *Institute for Scientific Information* registre varios miles de artículos sobre esta  
temática [SPIEGELHALTHER, MYLES, JONES, AABRAMS, 1999, pp. 508-512] y  
no cientos como en el decenio precedente. Muy importantes ya las aportaciones en  
los ensayos clínicos y en el estudio de «áreas pequeñas».

Pongamos por ejemplo el estudio de áreas pequeñas. Debido a que el número  
de casos (enfermos de ciertas dolencias, suicidios, etcétera) en espacios geográficos  
reducidos es bajo, las tasas brutas de mortalidad o morbilidad son muy lábiles y la  
representación de ellas a través de un mapa exhibe una variación amplia y poco infor-  
mativa. Los métodos bayesianos sugieren formas de corregir estos mapas, para que  
emerjan patrones explicativos inalcanzables con los procedimientos estadísticos  
clásicos [ELLIOT, CUZICK, ENGLISH, STERN, 1992, p. 14].

Sin lugar a dudas, los procedimientos bayesianos constituyen una tecnología  
de procesamiento y análisis emergente. El péndulo retorna a la zona original y reco-  
mienda su camino, impulsado ahora por recursos computacionales inexistentes en su  
primera oportunidad. Esta, su segunda aparición, aún está en fase embrionaria, pero  
cabe vaticinarle una presencia creciente y relevante en materia de aplicación de la  
estadística con fines inferenciales en la investigación contemporánea.

## BIBLIOGRAFÍA

- BAYES, T. (1763) «Essay towards solving a problem in the doctrine of chances». [Reproduced  
from *Phil. Trans. Roy. Soc.*, 53(1763), 370-418]. Studies in the history of probability and  
statistics. IX. With a bibliographical note by G.A. Barnard. *Biometrika*. Vol. 45, parts 3  
and 4. 1958, 299.
- BERRY, D.A. (1996) *Statistics: A bayesian perspective*. Belmont, California, Duxbury Press.
- BURTON, P.R.; GURRIN, L.C.; CAMPBELL, M.J. (1998) «Clinical significance not statistical  
significance: a simple bayesian alternative to p values». *Journal of Epidemiology Community  
Health*, 52, 1318-1323.
- COBO, E. (1993) *Estadística para no estadísticos*. Eada Gestion, Barcelona.
- CHIA, K-S. (1997) «Significant-itis -an obsession with the p-value». *Scandinavian Journal of Work  
and Environmental Health*, 23, 152-154.

- ELLIOT, P.; CUZICK, J.; ENGLISH, D., STERN, R. (1992) *Geographical and Environmental Epidemiology. Methods for Small-Area Studies*. Published on Behalf of the World Health Organization Regional Office for Europe by Oxford University Press.
- FEINSTEIN, A.R. (1998) «P-values and confidence intervals: Two sides of the same unsatisfactory coin». *Journal Clinical of Epidemiology*, 51, 355-360.
- FISHER, R.A. (1949) *The design of experiments*. Edinburgh, Oliver and Boyd.
- GOODMAN, S.N. (1999) «Toward evidence-based medical statistics (I): The p value fallacy». *Annals of Internal Medicine*, 130, 995-1004.
- GOODMAN, S.N. (1995) «Valores p, pruebas de hipótesis y verosimilitud: las consecuencias para la epidemiología de un debate histórico ignorado». *Boletín Oficina Sanitaria Panamericana*, 118, 141-155.
- GREENLAND, S. (1998) «Probability logic and probabilistic induction». *Epidemiology*, 9, 322-332.
- GUTIÉRREZ, S. (1994) *Filosofía de la estadística*. Universitat de València, Servei de Publicacions.
- HACKING, I. (1995a) *El surgimiento de la probabilidad*. 1ª edición, Barcelona, Gedisa editorial. Traducción del original en inglés *The emergence of probability*, 1975.
- HACKING, I. (1995b) *La domesticación del azar. La erosión del determinismo y el nacimiento de las ciencias del caos*. Barcelona, Gedisa editorial. Traducción del original en inglés *The Taming of Chance*, 1990.
- HALD, A. (1998) *A History of Mathematical Statistics. From 1750 to 1930*. New York, John Wiley & Sons, INC. Wiley Series in Probability and Statistics.
- MORGAN, B.W. (1971) *Introducción a los procesos bayesianos de decisión estadística*. Madrid, PARANINFO.
- NAGEL, E. (1939) *Principles of the theory of probability*. Foundations of the Unity of Sciences Vol. 1, Number 6. Chicago, The University of Chicago Press.
- NEYMAN, J., PEARSON, E. (1933) «On the problem of the most efficient tests of statistical hypotheses». *Philosophical Transactions of the Royal Society, Series A*, 231, 289-337.
- O'CONNOR, J.J.; ROBERTSON, E.F. (2000)  
<http://WWW-history.mcs.st-andrews.ac.uk/history/Mathematicians/Bayes.html>.
- ROTHMAN, J.K.; GREENLAND, S. (1998) *Modern epidemiology*. 2ª edición, Philadelphia, Lippincott-Raven.
- ROYALL, R.M. (1997) *Statistical evidence: a likelihood paradigm*. Boca Raton, Chapman & Hall/CRC.
- SAVAGE, L.J. (1954) *The foundations of statistics*. New York, Wiley.
- SILVA, L.C. (1997) *Cultura estadística e investigación en el campo de la salud: Una mirada crítica*. Madrid, Díaz de Santos.
- SILVA, L.C. (1999) «La crisis de las pruebas de significación y la alternativa bayesiana». *Memorias del XI Congreso de la Sociedad Gallega de Estadística e Investigación Operativa, Santiago de Compostela*.

- SPIEGELHALTHER, D.J.; MYLES, J.P., JONES, D.R.: AABRAMS, K.R. (1999) «An introduction to bayesian methods in healthy technology assesment». *British Medical Journal*, 319, 508-512.
- STOLLEY, P.D. (1991) «When a genius errs: R.A Fisher and the lung cancer controversy». *American Journal of Epidemiology*, 133, 416-425.