Teorema de Bayes

Introducción

La probabilidad condicional surge de problemas del siguiente tipo: se realizan en forma sucesiva dos experimentos aleatorios tales que, una vez hecho el primero, las probabilidades de los posibles resultados del segundo dependen del resultado de aquél; entonces, si A y B son dos eventos relativos al primero y segundo experimento, respectivamente, nos interesa conocer la probabilidad de que ocurra el evento B dado que ocurre el evento A. Sin embargo, la definición de probabilidad condicional es general y, en la misma situación descrita, tiene sentido calcular la probabilidad de ocurrencia del evento A dado que ocurre el evento B. A este tipo de probabilidades se les conoce como probabilidades de causas pues si pensamos en los eventos A y B como relativos a un primero y segundo experimentos, respectivamente, podemos pensar también a los resultados del primer experimento como las "causas" de los resultados del segundo; evidentemente el nombre "causas" expresa únicamente que hay un orden en la realización de los experimentos y no una relación real de causa-efecto.

Al inicio del Cálculo de Probabilidades se calculaban probabilidades condicionales en el caso en que su significado es intuitivo. Por ejemplo, supongamos que se tienen dos urnas, la primera con 5 bolas blancas y 3 negras, la segunda con 4 bolas blancas y 8 negras. Consideremos entonces el experimento aleatorio consistente primero, en elegir al azar una urna y después, en elegir al azar una bola de la urna seleccionada. En este caso, sabiendo que la urna seleccionada es la segunda, la probabilidad de seleccionar una bola blanca, en la segunda parte del experimento, es igual a $\frac{4}{12} = \frac{1}{3}$.

La probabilidad condicional se utilizaba para expresar una de las reglas del Cálculo de Probabilidades, la del producto:

Si A y B son dos eventos relativos a un experimento aleatorio, entonces:

$$P(A \cap B) = P(B \mid A)\, P(A)$$

Esta regla ya se encuentra enunciada en el libro de Abraham de Moivre, *The doctrine of chances*, publicado en el año 1717. Además, cabe mencionar que de Moivre enunció esta regla de manera general; es decir, para cualquier par de eventos, independientemente de que uno de ellos dependiera de la primera parte de un experimento y el otro de una segunda parte. En otros términos, también se puede escribir que:

$$P(A \cap B) = P(A \mid B)\, P(B)$$

De Moivre demostró este resultado para el caso de experimentos aleatorios cuyos posibles resultados son igualmente probables.

Aunque no fue hecho por de Moivre, de las dos relaciones anteriores se sigue la siguiente:

$$P\left(A \mid B\right) P\left(B\right) = P\left(B \mid A\right) P\left(A\right)$$

Así que, asumiendo que $P\left(B\right)$ es positiva, se tiene:

$$P\left(A \mid B\right) = \frac{P(B|A)P(A)}{P(B)}$$

Esta igualdad es conocida como la regla de Bayes por haber sido Thomas Bayes el primero en usarla explícitamente.

Un ejemplo típico de la aplicación de la regla de Bayes es el siguiente:

Supongamos que una prueba de sangre de una cierta enfermedad contagiosa es tal que las probabilidades de un falso negativo (la prueba es negativa para una persona que tiene la enfermedad) y de un falso positivo (la prueba es positiva para una persona que no tiene la enfermedad) son 0,02 y 0,03, respectivamente. Supongamos que cuando la prueba resulta negativa en una persona seleccionada al azar, la probabilidad de que la persona esté contagiada es igual a 0,005. Si la prueba resulta positiva en una persona seleccionada al azar, ¿cuál es la probabilidad de que la persona esté contagiada?

Para resolver este problema, definamos los eventos:

A: La persona está contagiada.

B: La prueba es positiva.

Por los datos que se dan, sabemos que:

$P(B \mid A^c) = 0{,}03$

$P(B^c \mid A) = 0{,}02$

$P(A \mid B^c) = 0{,}005$

De aquí se sigue que:

$P(B^c \mid A^c) = 0{,}97$

$P(B \mid A) = 0{,}98$

$P(A^c \mid B^c) = 0{,}995$

Y la probabilidad que buscamos es $P(A \mid B)$.

Entonces, si logramos calcular $P(A)$, tendríamos:

$P(B) = P(B \cap A) + P(B \cap A^c) = P(B \mid A)P(A) + P(B \mid A^c)P(A^c)$

$= 0{,}98P(A) + 0{,}03P(A^c) = 0{,}98P(A) + 0{,}03\,[1 - P(A)]$

$= 0{,}03 + 0{,}95P(A)$

$P(B^c) = P(B^c \cap A) + P(B^c \cap A^c) = P(B^c \mid A)P(A) + P(B^c \mid A^c)P(A^c)$

$= 0{,}02P(A) + 0{,}97P(A^c) = 0{,}02P(A) + 0{,}97\,[1 - P(A)]$

$= 0{,}97 - 0{,}95P(A)$

Además, tendríamos otro dato que no hemos utilizado, a saber, $P(A \mid B^c)$. Éste lo podemos utilizar en la siguiente relación:

$$P\left(A \mid B^c\right) = \frac{P(B^c|A)P(A)}{P(B^c)}$$

Sustituyendo valores, obtenemos:

$$0{,}005 = \frac{0{,}02P(A)}{0{,}97-0{,}95P(A)}$$

Despejando $P\left(A\right)$ en esta última igualdad, obtenemos:

$$P\left(A\right) = \frac{(0{,}005)(0{,}97)}{0{,}02+(0{,}005)(0{,}95)} = \frac{97}{495}$$

Así que ya podemos obtener la probabilidad que buscamos:

$$P(A \mid B) = \frac{P(B|A)P(A)}{P(B)} = \frac{(0{,}98)\left(\frac{97}{495}\right)}{0{,}03+(0{,}95)\left(\frac{97}{495}\right)} = \frac{\frac{98}{100}\frac{97}{495}}{\frac{95}{100}\frac{97}{495}+\frac{3}{100}} = \frac{4753}{5350} \approx 0{,}888411$$

Por lo tanto, si la prueba resulta positiva en una persona seleccionada al azar, la probabilidad de que la persona esté contagiada es, aproximadamente, igual a 0,888411.

A finales del siglo XIX, esta forma de trabajar con probabilidades condicionales constituía un problema para poder establecer una teoría matemática de la probabilidad ya que para calcular una probabilidad condicional se recurría a su sentido intuitivo; esto y otras consideraciones llevaron a considerar al Cálculo de Probabilidades como una rama de la Física, no de la Matemática.

Cuando la teoría de la probabilidad se formula en forma axiomática, los 3 elementos que conforman la base de la teoría matemática son los mismos con los que se define una medida: 1) un conjunto Ω cualquiera, al cual se le llama espacio muestral; 2) una σ-álgebra $\Im$ de subconjuntos de Ω, cuyos elementos son llamados eventos y 3) una medida P definida sobre $\Im$, la cual tiene como particularidad que la medida de Ω es igual a 1, razón por la cual se llama una medida de probabilidad.

A partir de esta base se desarrolla una teoría puramente matemática sin necesidad de recurrir a las características de los fenómenos que se estudian con ésta.

El tema de la probabilidad condicional se resuelve de una manera muy simple; se da una definición formal de este concepto a partir de lo que se tiene:

Definición 1 (**Probabilidad condicional**). *Sean A y B dos eventos y supongamos $P(A) > 0$, se define la probabilidad condicional de B, dada la ocurrencia de A, $P(B|A)$, mediante la fórmula:*

$$P(B \mid A) = \frac{P(A \cap B)}{P(A}$$

De aquí se sigue lo mismo que teníamos antes; es decir, la regla del producto y la regla de Bayes:

Si A y B son dos eventos cualesquiera y $P(A) > 0$, entonces:

$$P(A \cap B) = P(B \mid A)\, P(A)$$

Si A y B son dos eventos cualesquiera, ambos de probabilidad positiva, entonces:

$$P(A \mid B) = \frac{P(B|A)P(A)}{P(B)}$$

De esta forma, la regla del producto y la regla de Bayes consisten simplemente en diferentes maneras de expresar la definición de probabilidad condional.

Teorema de Bayes

Con lo expuesto anteriormente, la regla de Bayes se obtiene con un manejo algebraico elemental de la definición de probabilidad condicional y puede quedar la idea de que el trabajo de Bayes es una trivialidad.

Sin embargo, la aportación central de Bayes no es esta regla que lleva su nombre. Incluso, como lo mencionamos antes, se deriva fácilmente de lo que ya había sido expuesto por de Moivre.

En realidad, la "regla de Bayes"es únicamente algo accesorio que utilizó Bayes para resolver un problema importante de estimación y, en un sentido, puede decirse que, con la solución que dio a ese problema, se adelantó a su época ya que trabajó con la distribución condicional de una variable aleatoria Y, dado que se conoce el valor de otra variable aleatoria X; una de ellas discreta y la otra absolutamente continua.

El trabajo de Bayes se publicó en el año 1763, dos años después de su muerte, bajo el título *An essay towards solving a problem in the doctrine of chances* (Philos. Trans. Roy. Soc. London Ser. A 53, reproducido en Biometrika 45, 1958).

Comenzó su artículo planteando el siguiente problema:

Dado el número de veces en el cual un evento desconocido ha ocurrido y fallado, encontrar la probabilidad de que su probabilidad de ocurrencia en un ensayo esté comprendida entre dos valores dados.

Para resolver este problema, demostró primero las reglas básicas del Cálculo de Probabilidades.

Proposición 1 (Regla de la suma). *Cuando varios eventos son inconsistentes (mutuamente excluyentes) la probabilidad de que ocurra alguno de ellos es igual a la suma sus probabilidades.*

Proposición 2 (Regla del producto). *La probabilidad de que dos eventos subsecuentes ocurran es igual al producto de la probabilidad del primero por la probabilidad del segundo bajo el supuesto de la ocurrencia del primero.*

Corolario 1 (Probabilidad condicional). *Dados dos eventos subsecuentes, la probabilidad del segundo bajo el supuesto de que el primero ocurre es igual al cociente de la probabilidad de que ambos ocurran entre la probabilidad de que el primero ocurra.*

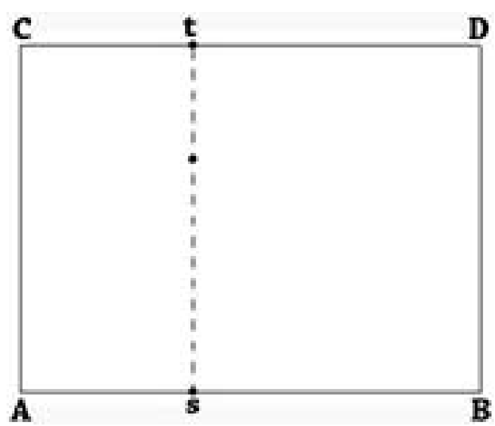
Proposición 3 (**Probabilidad de causas**). *Si dados dos eventos subsecuentes se descubre que el segundo ha ocurrido y de aquí yo conjeturo que el primero ha ocurrido, la probabilidad de estar en lo cierto es igual al cociente de la probabilidad de que ambos ocurran entre la probabilidad de que el segundo ocurra.*

Proposición 4 (**Distribución binomial**). *Si la probabilidad de ocurrencia de un evento es p en un ensayo, la probabilidad de que el evento ocurra k veces y falle $n-k$ veces, en n ensayos, es igual a $\binom{n}{k}p^k\,(1-p)^{n-k}$.*

Inmediatamente después de demostrar estas proposiciones, Bayes formuló el experimento aleatorio que serviría para dar solución al problema original.

Consideró un plano $ABCD$, el cual está hecho de tal manera que si una bola es lanzada sobre él, entonces habrá la misma probabilidad de que permanezca en cualquiera de dos partes iguales del plano y necesariamente permanecerá sobre éste. Formuló entonces el siguiente experimento aleatorio:

Una bola $\mathcal{W}$ es lanzada primero y, a través del punto en donde cae, se traza una recta paralela a AD, la cual corta a los segmentos AB y CD en los puntos s y t, respectivamente.



Después de lanzar la bola $\mathcal{W}$, se lanza una bola $\mathcal{O}$ n veces sobre el plano. En cada lanzamiento se dirá que el evento $\mathcal{M}$ ocurre si la bola $\mathcal{O}$ cae en el rectángulo $AstC$.

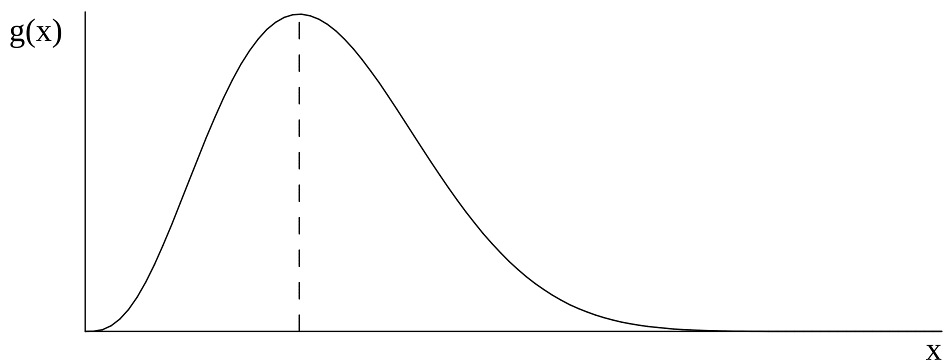
En seguida mostró que, como la bola $\mathcal{W}$ se lanza al azar sobre el plano $ABCD$, la probabilidad de que el punto s caerá dentro de un intervalo I de extremos a y b, pertenecientes al segmento AB, es igual a la razón de la longitud del intervalo I a la longitud de AB, la cual, sin pérdida de generalidad, se puede suponer que es igual a 1.

Después probó uno de los resultados centrales de su artículo, el cual demostraremos siguiendo su razonamiento, aunque utilizando terminología moderna para mayor claridad:

Proposición 5. *Antes de lanzar la bola $\mathcal{W}$, si a y b son dos puntos sobre el segmento AB tales que $a < b$, la probabilidad de que el punto s caiga dentro del intervalo $[a,b]$ y que el evento $\mathcal{M}$ ocurra k veces y falle $n-k$ veces es igual al área bajo la gráfica de la función $g(x) = \binom{n}{k} x^k (1-x)^{n-k}$ entre a y b.*

Demostración

Observemos que función g es cero en $x = 0$, después es creciente, hasta alcanzar un máximo en $x_{máx} = \frac{k}{n}$, después de lo cual decrece, hasta llegar a cero en $x = 1$.



Definamos el evento:

B_k : $\mathcal{M}$ ocurre k veces y falla $n-k$ veces.

Definamos también la variable aleatoria X como el número real $x \in [0,1]$ (el punto s en el enunciado), determinado por el punto donde cae la bola $\mathcal{W}$.

Entonces, la probabilidad que se quiere calcular es:

$$P\left([a \leq X \leq b] \cap B_k\right)$$

Denotemos por γ al área bajo la gráfica de la función g entre a y b.

Supongamos que la probabilidad en consideración es mayor que γ.

γ se puede aproximar mediante sumas de la forma:

$$\sum_{j=1}^{m} g\left(x_{máx}^{(j)}\right)\left(x_j - x_{j-1}\right)$$

donde $\{a = x_0, x_1, \ldots, x_m = b\}$ es una partición del intervalo $[a, b]$ de tal forma que los subintervalos que se forman tengan la misma longitud; y, para cada $j \in \{1, 2, \ldots, m\}$, $x_{máx}^{(j)}$ es el punto del intervalo $[x_{j-1}, x_j]$ en el cual la función g alcanza su valor máximo en ese intervalo.

Consideremos entonces una partición $\{a = x_0, x_1, \ldots, x_m = b\}$ del intervalo $[a, b]$ de la forma mencionada antes, de tal forma que:

$$\sum_{j=1}^{m} g\left(x_{máx}^{(j)}\right)(x_j - x_{j-1}) < P\left([a \leq X \leq b] \cap B_k\right)$$

Consideremos dos puntos consecutivos de la partición, x_{j-1} y x_j.

Si el punto s cayera en algún punto $x \in [x_{j-1}, x_j)$, la probabilidad de B_k sería igual a $g(x)$; pero $g(x) \leq g(x_{máx}^{(j)})$.

Entonces, la probabilidad de que el punto s caiga dentro del intervalo $[x_{j-1}, x_j)$ y que el evento B_k ocurra es menor o igual que el producto $g(x_{máx}^{(j)})\,(x_j - x_{j-1})$.

Con terminología moderna, tendríamos, para $x \in [x_{j-1}, x_j)$:

$$P(B_k \mid X = x) = \binom{n}{k} x^k (1-x)^{n-k} = g(x) \leq g(x_{máx}^{(j)})$$

Por lo tanto:

$$P(B_k \mid x_{j-1} \leq X < x_j) \leq g(x_{máx}^{(j)})$$

Así que,

$P\left([x_{j-1} \leq X < x_j] \cap B_k\right)$

$= P(B_k \mid x_{j-1} \leq X < x_j) P(x_{j-1} \leq X < x_j) \leq g(x_{máx}^{(j)})\,(x_j - x_{j-1})$

Aplicando la regla de la suma y tomando en cuenta que $P\,(X = b) = 0$, se obtendría:

$$P(a \leq X \leq b, B_k) \leq \sum_{j=1}^{m} g\left(x_{máx}^{(j)}\right)(x_j - x_{j-1})$$

Pero esto es una contradicción pues se tomó la partición de tal manera que:

$$\sum_{j=1}^{m} g\left(x_{máx}^{(j)}\right)(x_j - x_{j-1}) < P\left([a \leq X \leq b] \cap B_k\right)$$

Así que, la probabilidad de que el punto s caiga entre a y b y que el evento $\mathcal{M}$ ocurra k veces y falle $n-k$ veces no puede ser mayor que el área bajo la la gráfica de la función g entre a y b.

Supongamos ahora que la probabilidad en consideración es menor que γ.

γ se puede aproximar mediante sumas de la forma:

$$\sum_{j=1}^{m} g\left(x_{mín}^{(j)}\right)(x_j - x_{j-1})$$

donde $\{a = x_0, x_1, \ldots, x_m = b\}$ es una partición del intervalo $[a, b]$ de tal forma que los subintervalos que se forman tengan la misma longitud; y, para cada $j \in \{1, 2, \ldots, m\}$, $x_{mín}^{(j)}$ es el punto del intervalo $[x_{j-1}, x_j]$ en el cual la función g alcanza su valor mínimo en ese intervalo.

Consideremos entonces una partición $\{a = x_0, x_1, \ldots, x_m = b\}$ del intervalo $[a, b]$ de la forma mencionada antes, de tal forma que:

$$\sum_{j=1}^{m} g\left(x_{mín}^{(j)}\right)(x_j - x_{j-1}) > P\left([a \leq X \leq b] \cap B_k\right)$$

Consideremos dos puntos consecutivos de la partición, x_{j-1} y x_j.

Si el punto s cayera en algún punto $x \in [x_{j-1}, x_j)$, la probabilidad de B_k sería igual a $g(x)$; pero $g(x) \geq g(x_{mín}^{(j)})$.

Entonces, la probabilidad de que el punto s caiga dentro del intervalo $[x_{j-1}, x_j)$ y que el evento B_k ocurra es mayor o igual que el producto $g(x_{mín}^{(j)})\,(x_j - x_{j-1})$.

Aplicando la regla de la suma y tomando en cuenta que $P\left(X = b\right) = 0$, se obtendría:

$$P(a \leq X \leq b, B_k) \geq \sum_{j=1}^{m} g\left(x_{mín}^{(j)}\right)(x_j - x_{j-1})$$

Pero esto es una contradicción pues se tomó la partición de tal manera que:

$$\sum_{j=1}^{m} g\left(x_{mín}^{(j)}\right)(x_j - x_{j-1}) < P\left([a \leq X \leq b] \cap B_k\right)$$

Así que, la probabilidad de que el punto s caiga entre a y b y que el evento $\mathcal{M}$ ocurra k veces y falle $n-k$ veces no puede ser menor que el área bajo la la gráfica de la función g entre a y b.

Por lo tanto:

La probabilidad de que el punto s caiga entre a y b y que el evento $\mathcal{M}$ ocurra k veces y falle $n-k$ veces es igual al área bajo la gráfica de la función g entre a y b.

■

Puede observarse que, en lenguaje moderno, lo que demostró Bayes es una caso particular de la regla de la probabilidad total en su versión continua:

$$P\left([a \leq X \leq b] \cap B_k\right) = \int_a^b P(B_k \mid X = x) f_X(x) dx$$

En particular, continúa argumentando Bayes, se tiene:

$$P(B_k) = P\left([0 \leq X \leq 1] \cap B_k\right) = \int_0^1 \binom{n}{k} x^k (1-x)^{n-k} dx$$

Esta última integral la evaluó Bayes, obteniendo:

$$\int_0^1 \binom{n}{k} x^k (1-x)^{n-k} dx = \frac{1}{n+1}$$

En efecto, aunque Bayes siguió otro método, el valor de la integral se puede obtener de la siguiente forma:

Para cada $k \in \{0, 1, 2, \ldots, n\}$, definamos:

$$b_k = \int_0^1 x^k (1-x)^{n-k} dx$$

Se tiene:

$$b_0 = \int_0^1 (1-x)^n dx = \frac{1}{n+1}$$

Y, para $k \in \{1, 2, \ldots, n\}$, integrando por partes, con $u = x^k$ y $dv = (1-x)^{n-k} dx$, se obtiene:

$b_k = \int_0^1 \frac{k}{n-k+1} x^{k-1} (1-x)^{n-k+1} dx = \frac{k}{n-k+1} b_{k-1}$

$= \frac{k}{n-k+1} \frac{k-1}{n-k+2} \frac{k-2}{n-k+3} \cdots \frac{2}{n-2+1} \frac{1}{n-1+1} b_0$

$= \frac{1}{n} \frac{2}{n-1} \cdots \frac{k-2}{n-(k-3)} \frac{k-1}{n-(k-2)} \frac{k}{n-(k-1)} b_0$

$= \frac{k!(n-k)!}{n!} \frac{1}{n+1} = \frac{k!(n-k)!}{(n+1)!}$

Definiendo $0! = 1$, la fórmula también es válida para $k = 0$.

Por lo tanto, para cualquier $k \in \{0, 1, 2, \ldots, n\}$, se tiene:

$$\int_0^1 \binom{n}{k} x^k (1-x)^{n-k} dx = \frac{k!(n-k)!}{(n+1)!} \binom{n}{k} = \frac{1}{n+1}$$

Finalmente, Bayes demostró el resultado principal de su artículo:

Teorema 1 (Teorema de Bayes)**.** *Si antes de que se descubra la posición del punto s resulta que el evento $\mathcal{M}$ ha ocurrido k veces y fallado $n-k$ veces, en n ensayos, y de aquí yo conjeturo que el punto s cae entre los puntos a y b dentro del segmento AB, la probabilidad de estar en lo cierto es igual a la razón del área bajo la gráfica de la función $g(x) = \binom{n}{k} x^k (1-x)^{n-k}$ entre a y b y el área bajo la gráfica de la función g entre A y B.*

Demostración

Por la proposición anterior y utilizando las definiciones que se dieron ahí, se tiene:

$$P\left([a \leq X \leq b] \cap B_k\right) = \int_a^b \binom{n}{k} x^k (1-x)^{n-k} dx$$

Así que, para $k \in \{0, 1, 2, \ldots, n\}$, se obtiene:

$$P(a \leq X \leq b \mid B_k) = \frac{P([a \leq X \leq b] \cap B_k)}{P(B_k)} = \frac{\int_a^b \binom{n}{k} x^k (1-x)^{n-k} dx}{\int_0^1 \binom{n}{k} x^k (1-x)^{n-k} dx}$$

$$= (n+1) \binom{n}{k} \int_a^b x^k (1-x)^{n-k} dx$$

■

Obsérvese que en el razonamiento hecho está implícita la que ahora se conoce como regla de Bayes:

$$P(a \leq X \leq b \mid B_k) = \frac{P(B_k | a \leq X \leq b) P(a \leq X \leq b)}{P(B_k)}$$

Actualmente a la función $\beta : (0, \infty) \times (0, \infty) \to \mathbb{R}$ definida por mediante la relación:

$$\beta(\alpha_1, \alpha_2) = \int_0^1 x^{\alpha_1 - 1} (1-x)^{\alpha_2 - 1} dx$$

se le conoce como función beta.

Se tiene la fórmula para su valor cuando $\alpha_1, \alpha_2 \in \mathbb{N}$:

$$\beta(n, m) = \frac{(n-1)!(m-1)!}{(n+m-1)!} \text{ para cualquier pareja } n, m \in \mathbb{N}$$

También se tiene definida la distribución beta:

Definición 2 (Distribución beta). *Se dice que la variable aleatoria X tiene distribución beta con parámetros $\alpha_1 \in (0, \infty)$ y $\alpha_2 \in (0, \infty)$ si la función:*

$$f_X(x) = \begin{cases} \frac{1}{\beta(\alpha_1,\alpha_2)} x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1} & \text{si } x \in (0,1) \\ 0 & \text{en otro caso} \end{cases}$$

es una función de densidad de X.

Utilizando esta definición, el resultado de Bayes se puede enunciar de la siguiente manera:

Cuando lo único que se conoce de un evento A es que ha ocurrido k veces y fallado $\mathbf{n - k}$ veces, en n ensayos, entonces su probabilidad de ocurrencia puede considerarse como seleccionada de una población con distribución beta de parámetros $\mathbf{k + 1}$ y $\mathbf{n - k + 1}$.

En su artículo, para hacer ver que con lo que demostró se resuelve el problema planteado inicialmente, Bayes argumentó lo siguiente:

En el caso de un evento como el que llamo $\mathcal{M}$, a partir del número de veces en que ocurre y falla en un cierto número de ensayos, sin conocer nada más sobre él, uno puede hacer una suposición acerca de su probabilidad y, por los métodos usuales para calcular las magnitudes de las áreas mencionadas, encontrar la probabilidad de que la suposición sea correcta.

Y que la misma regla es la que debe de ser usada en el caso de un evento para el cual nada es sabido acerca de su probabilidad, antes de cualquier ensayo que se realice, parece provenir de la siguiente consideración:

En relación a tal evento, no tengo razón de pensar que, en un cierto número de ensayos, debería ocurrir algún número de veces en lugar de otro. En efecto, sólo puedo razonar, en lo que a ese evento concierne, como si su probabilidad primero estuviera no fija y después determinada de tal manera que no me da razón para pensar que, en un cierto número de ensayos, debería ocurrir algún número de veces en lugar de otro.

Pero éste es exactamente el caso del evento $\mathcal{M}$. En efecto, antes de que la bola $\mathcal{W}$ sea lanzada, lo cual determina su probabilidad en un ensayo, la probabilidad que tiene de ocurrir k veces y fallar $n - k$ en n ensayos es igual al área bajo la curva entre 0 y 1, la cual es la misma cuando n está dado, cualquiera que sea el número k. Y como consecuencia, antes de que el lugar del punto s sea descubierto o se conozca el número de veces que el evento $\mathcal{M}$ ha ocurrido en n ensayos, no tengo razón para pensar que debería ocurrir algún número de veces en lugar de otro.

Entonces, en lo que sigue tomaré por garantizado que la regla dada en relación al evento $\mathcal{M}$, es también la regla que debe utilizarse en relación a cualquier evento para el cual nada es sabido acerca de su probabilidad antes de realizar u observar cualquier ensayo concerniente a él.

Puede observarse que la base de la argumentación de Bayes radica en el hecho de que la probabilidad del evento B_k ($\mathcal{M}$ ocurre k veces y falla $n-k$ veces) depende exclusivamente de n, en otras palabras, el área bajo la gráfica de g es la misma para cualquier $k \in \{0, 1, 2, \ldots, n\}$.

Este hecho es lo que fundamenta el que el experimento aleatorio que considera Bayes sirva de modelo para el problema planteado pues, antes de realizar cualquier observación, no hay razón para esperar que un evento totalmente desconocido pueda ocurrir cierto número de veces en lugar de otro número distinto al realizar n observaciones.

Esta misma argumentación justifica el que el punto s lo considere como elegido al azar en el segmento AB. Ese punto determina la probabilidad de que el evento $\mathcal{M}$ ocurra en cada ensayo de la segunda parte del experimento. De aquí que el resultado de Bayes también se pueda plantear de la siguiente manera:

Sabiendo que se ha observado que un cierto evento A ha ocurrido k veces y fallado $\mathbf{n-k}$ veces, en n ensayos, se quiere estimar su probabilidad de ocurrencia en ensayos posteriores que se realicen. Antes de las n observaciones no se sabe nada acerca de la probabilidad de ocurrencia de A, de manera que ésta puede ser cualquier número real en el intervalo $[\mathbf{0}, \mathbf{1}]$ y no hay razón para esperar que sea uno u otro número. Es decir, antes de las n observaciones, se puede asumir que la probabilidad de ocurrencia de A está dada por el valor de una variable aleatoria X con distribución uniforme en el intervalo $[\mathbf{0}, \mathbf{1}]$. Una vez que se sabe que A ha ocurrido k veces y fallado $\mathbf{n-k}$ veces, en n ensayos, la distribución de X cambia a una distribución beta de parámetros $\mathbf{k+1}$ y $\mathbf{n-k+1}$.

Conclusiones

El trabajo de T. Bayes se ubica en la época en que se estaban sistematizando los resultados relativos al cálculo de probabilidades, todos basados en la definición clásica de probabilidad, la cual se refiere a distribuciones discretas y, si bien Bayes vio la necesidad de demostrar los resultados básicos de la naciente teoría, en realidad ya se contaba con la versión bastante completa que se daba en el libro de de Moivre, el cual parece haber sido del conocimiento de Bayes.

Sin embargo, los resultados de Bayes se refieren también a propiedades de las distribuciones continuas y ésta es una de las aportaciones importantes de su trabajo. El principal resultado en este sentido es la regla de la probabilidad total en su versión continua. La llamada "regla de Bayes" en realidad no se debe a Bayes, se puede obtener inmediatamente de las reglas que da de Moivre pues, en la regla del producto que él demuestra, el orden en que ocurren los eventos es irrelevante, pero su formulación explícita es posterior a de Moivre. Por otra parte, lo que realmente interesa no es la regla misma sino su aplicación para calcular probabilidades de eventos que dependen de la primera parte de un experimento sabiendo lo que ha ocurrido en una segunda parte y esta última es precisamente la aplicación que hace Bayes, pero no en el caso discreto sino en el continuo, de manera que su aportación es aún más significativa.

La segunda aportación importante y, por el desarrollo actual de la teoría, la más significativa, se refiere a la introducción de una distribución a priori para resolver un problema de estimación. Con su método, sorprendentemente original, Bayes parece adelantarse a su tiempo. Es digno de observarse que la introducción de una distribución a priori no aparece como arbitraria en su trabajo, ésta encaja perfectamente en el problema que se plantea pues, mediante ella, todas los posibles resultados de las n observaciones que se han hecho del evento cuya probabilidad se quiere estimar tienen, a priori, la misma probabilidad de ocurrrencia, cosa que - de acuerdo a lo que parece ser la concepción de Bayes de la probabilidad, como algo ligado al conocimiento o ignorancia del fenómeno en consideración - debe cumplirse ya que, a priori, no se tiene información alguna sobre que posible resultado puede esperarse con mayor facilidad que otro.

Finalmente, en la época de Bayes también estaba desarrollándose el Cálculo Diferencial e Integral y aún no se contaba con la definición analítica de la integral, dada por Cauchy en el año 1823 y reformulada por Riemann en el año 1854; de manera que la aproximación de una integral mediante sumas de Riemann no era conocida.

La demostración de la proposición 5 la hicimos utilizando sumas de Riemann. Bayes, en cambio se basó en la interpretación de la integral como el área bajo la gráfica de la función a integrar y lo que hizo fue aproximar esa área acercándose por fuera y por dentro de la región acotada por la gráfica de la función y el eje horizontal, en el intervalo de integración.

En lenguaje moderno, lo que hizo Bayes fue resolver un problema de distribuciones condicionales. Si p es la probabilidad de ocurrencia de un evento al realizar un cierto experimento aleatorio y Y es el número de veces en los cuales el evento ocurre al repetir n veces el experimento, Bayes se planteó el problema de encontrar la distribución condicional de p dado que $Y = k$, asumiendo que p es una cantidad que originalmente se selecciona al azar en el intervalo $[0, 1]$.

Conociendo el valor de p, Y tiene distribución binomial con parámetros n y p. De manera que el problema de Bayes equivale al siguiente:

Si X es una variable aleatoria con distribución uniforme en el intervalo $[\mathbf{0}, \mathbf{1}]$ y suponiendo que, para cada valor x de X, Y es una variable aleatoria con distribución binomial de parámetros n y $\mathbf{p} = \mathbf{x}$, se trata de encontrar la distribución condicional de X dado que $\mathbf{Y} = \mathbf{k}$, para $\mathbf{k} \in \{\mathbf{0}, \mathbf{1}, \mathbf{2}, \ldots, \mathbf{n}\}$. El resultado es que la distribución condicional de X es beta con parámetros $\mathbf{k} + \mathbf{1}$ y $\mathbf{n} - \mathbf{k} + \mathbf{1}$.

Finalmente, obsérvese que la relación $P(B_k) = \frac{1}{n+1}$, para cualquier $k \in \{0, 1, 2, \ldots, n\}$, se puede expresar diciendo que la distribución de Y es uniforme en el conjunto $\{0, 1, 2, \ldots, n\}$.